

S
E
M
I
N
A
L
O
G
A
R
I
O
D

Proceedings IST, III

Dezembro 2006

Editores:

Ana Cannas da Silva

Luís Cruz-Filipe

Jorge Drumond Silva

Nuno Freitas

André Vasconcelos

Título ◇ Seminário Diagonal – Proceedings IST, III ◇ Dezembro 2006

Editores ◇ Ana Cannas da Silva¹ ◇ Luís Cruz-Filipe¹
◇ Jorge Drumond Silva¹ ◇ Nuno Freitas² ◇ André Vasconcelos²

⁽¹⁾ Departamento de Matemática do Instituto Superior Técnico

⁽²⁾ alunos da Licenciatura em Matemática Aplicada e Computação do IST até 2006

www ◇ <http://www.math.ist.utl.pt/diagonal/>

Artigos © dos respectivos autores. Colectânea © 2006 dos editores.

A cópia privada é permitida.

Documento $\text{\LaTeX}$ executado em 9 de Dezembro de 2006, no Departamento de Matemática do Instituto Superior Técnico.

Prefácio

Já com seis anos de actividade regular, o Seminário Diagonal¹ mantém-se fiel ao anúncio que o lançou no Instituto Superior Técnico (IST) em Outubro de 2000:

S
E
M
I
N
Á
R
I
O
D
I
A
G
O
N
A
L

O que é?
Novo seminário de estudantes.

Para quem?
Todos os interessados em Matemática.

Sobre quê?
Matemática, no sentido lato.

Estreia brevemente em todo o país.

O Seminário conserva-se um espaço onde os alunos da Licenciatura em Matemática Aplicada e Computação do IST e muitos outros interessados em Matemática podem trocar ideias e experiências matemáticas. Aqui se mostram novos assuntos – sobretudo alguns estranhos aos currículos das licenciaturas – de modo acessível aos alunos dos primeiros anos, ou seja, *diago-*

¹ O Seminário Diagonal é uma iniciativa de âmbito nacional, e actualmente realiza-se também na Faculdade de Ciências da Universidade do Porto, na Faculdade de Ciências e Tecnologia da Universidade de Coimbra e na Faculdade de Ciências da Universidade de Lisboa.

nalizados. Como dizia João Boavida, tendo em conta que “semi-simples” é sinónimo de “diagonalizável”, isto significa que os assuntos devem ser apresentados como “soma directa” de “partes simples”.

Esta colectânea reúne seis *artigos diagonais* correspondentes a seminários realizados entre Outubro de 2004 e Fevereiro de 2006. Nem todos os oradores nesse período puderam contribuir com artigos. Juntamos uma lista dos resumos de todos os seminários realizados desde Março de 2003 – quando terminou a colectânea anterior – até Fevereiro de 2006. A indicação do ano curricular refere-se ao ano lectivo em que o respectivo seminário foi apresentado.

Agradecimentos

Muitas pessoas têm construído o sucesso do Seminário Diagonal. O primeiro agradecimento vai para os oradores e seus colaboradores. Realçamos os que contribuíram artigos para este volume e seus revisores ou orientadores: Paulo Abrantes, Miguel Abreu, Adérito Araújo, José Félix Costa, José Luís Fachada, Luís Gil, Diogo Gomes, Fernando Machado, João Nuno Mestre, Bruno Montalto, Diogo Oliveira e Silva, Hélio Pais, Roger Picken, Daniela Pontes e Carlos Tamulonis. Um agradecimento vai também para toda a audiência – abrangendo alunos da LMAC, da LEFT, da LCI, da LEC, da LEIC, da LEEC e de outros cursos e professores do Departamento de Matemática do IST – e em especial para os organizadores do Seminário, cuja dedicação o sustenta.

Duas instituições possibilitaram materialmente este projecto. A Fundação Calouste Gulbenkian, através do Programa Gulbenkian Novos Talentos em Matemática, patrocinou o trabalho de vinte e um oradores entre Março de 2003 e Fevereiro de 2006, quatro dos quais são autores de artigos aqui incluídos. O Banco BPI, através do Concurso de Apoio a Actividades Extra-curriculares do IST, financiou ou co-financiou a publicação das duas primeiras colectâneas *Seminário Diagonal – Proceedings IST*, financiou a aquisição de livros para a *Colecção Diagonal* – colecção de livros e revistas de divulgação da Matemática – e financiou a publicação desta colectânea.

O suporte gráfico em $\text{\LaTeX}$ que tornou tão fácil a produção deste documento em PDF e a criação de <http://www.math.ist.utl.pt/diagonal/> devem-se ao generoso empenho do João Boavida e do João Palhoto Matos no lançamento das iniciativas *diagonais* entre 2000 e 2002.

Os Editores,
Lisboa, 3 de Novembro de 2006

Conteúdo

Prefácio	1
Resumos dos Seminários	5
<i>Bruno Montalto — Cadeias de Markov</i>	14
<i>Luís Gil — Modelação de Plantas e Fractais a Partir de Gramáticas</i>	29
<i>Fernando Machado — Geometria em Superfícies</i>	49
<i>Daniela Pontes — Será Possível Encontrar Topologia numa Fábrica?</i>	77
<i>Diogo Oliveira e Silva — Volta ao Mundo em 45 Minutos</i>	99
<i>João Nuno Mestre — Co-Existência de Ciclos</i>	115

Resumos dos Seminários

A UNICIDADE DE SOLUÇÃO DE UMA EQUAÇÃO NÃO LINEAR

Hugo Tavares (3º ano de Matemática, Faculdade de Ciências da Universidade de Lisboa)

24 de Março de 2003

No estudo das equações diferenciais não lineares com condições de fronteira, em geral é difícil deduzir a unicidade da solução. Este seminário é o resultado do estudo de um célebre artigo de M.K. Kwong (1989), onde se estabelece um tal teorema de unicidade.

Vamos dar uma ideia de como se resolve o problema e mostrar que, talvez surpreendentemente, os requisitos para efectuar essa prova são muito poucos, apesar da complexidade de algumas definições.

O GRUPO FUNDAMENTAL

Ricardo Joel (3º ano da LMAC — Análise, Geometria e Álgebra)

12 de Maio de 2003

À primeira vista pode parecer uma observação trivial que uma superfície esférica é essencialmente diferente de um toro (a superfície de um donut). No entanto, como é que podemos distinguir de forma concreta aqueles dois objectos? E como encontrar ideias matemáticas que façam essa distinção?

A topologia algébrica vem em nosso auxílio para responder a questões desta natureza e neste seminário vamos descobrir algumas potencialidades de uma das suas ferramentas mais simples e importantes: o grupo fundamental.

WAVELETS

João Pina (5º ano da Licenciatura em Engenharia Civil)

6 de Junho de 2003

As Wavelets são funções complexas com origem numa simples e elegante construção matemática. Translações e dilatações de uma função mãe permitem representar qualquer tipo de função. Assim, como que um “upgrade” da análise harmónica, retêm informação tanto no domínio do tempo como da frequência. Ao serviço do FBI reduziram 20 vezes o arquivo de impressões digitais e têm sido utilizadas em áreas tão distintas como o processamento de imagens, “noise cleaning”, equações diferenciais ou análise de estruturas.

SERÁ POSSÍVEL OUVIR A FORMA DE UM TAMBOR?

Andreia Hortence Gomes (2º ano da LMAC)

14 de Outubro de 2003

O som de um tambor caracteriza-se por um conjunto de frequências particulares de vibração que dependem da sua geometria. Mas poder-se-á responder à questão inversa? Isto é, será que se conhecermos as frequências de vibração de um tambor, poderemos determinar a sua forma? Ou existirão dois tambores de formas diferentes com o mesmo som? Neste seminário vamos ver como podemos responder a estas perguntas, que já se fazem desde há cerca de cem anos.

COMO SOMAR PONTOS EM CURVAS ELÍPTICAS?

Sérgio Marcelino (3º ano da LMAC — Ciência da Computação)

28 de Outubro de 2003

O estudo de curvas elípticas é fundamental na matemática moderna: na teoria de números, na famosa demonstração do último teorema de Fermat por Andrew Wiles e até na criptografia. Estamos habituados a desenhar curvas no plano do tipo $y^2 = x^3 + ax + b$. Mas será que conseguimos achar inteiros ou racionais x e y tais que (x, y) está sobre a curva? O que podemos saber sobre estes pontos especiais? Serão finitos? Estudaremos uma “soma natural” definida nestes pontos e exploraremos algumas das suas consequências.

DECISÃO DE PRIMALIDADE: A INOVAÇÃO POLINOMIAL

Jorge Vitória (2º ano de Matemática, Faculdade de Ciências da Universidade do Porto)

8 de Novembro de 2003

Há séculos que os números primos são alvo de um enorme fascínio e uma fonte inesgotável de resultados. No entanto, uma questão sempre se colocou: como distingui-los dos números compostos de forma eficiente? Haverá um algoritmo capaz de o fazer em tempo polinomial? A resposta (afirmativa) viria a ser dada por três matemáticos indianos no ano de 2002... Apresentar a solução algorítmica deste problema e clarificar a inovação trazida por este resultado é o objectivo desta apresentação.

MUITO COMPLEXA, MUITO LINEAR

Joana Santos (2º ano da LMAC)

2 de Dezembro de 2003

A transformação de Möbius, $M(z) = (az + b)/(cz + d)$, é uma aplicação entre números complexos com propriedades muito interessantes e bastantes aplicações, não só na Análise Complexa, mas também em áreas tão diversas como a Geometria não Euclidiana e a Teoria da Relatividade de Einstein.

Nesta apresentação o objectivo principal será perceber como esta aplicação, também chamada transformação homográfica, bilinear ou linear fraccional, actua sobre os pontos do plano e ver de que forma uma representação matricial traz para o mundo complexo as mais valias da Álgebra Linear.

HISTÓRIAS DE NÓS

Cláudia Pascoal (3º ano da LMAC — Probabilidades e Estatística)

16 de Março de 2004

Conheces o nó do teu sapato, o nó da tua gravata, ou o nó da trança do teu cabelo? Achas que todos os nós são iguais? A Teoria dos nós ajuda-nos a responder a estas questões que aparentemente são simples. No entanto, a resposta nem sempre é trivial. Neste seminário iremos ver alguns aspectos importantes desta Teoria: como surgiu o seu estudo, como podemos representar os nós, distingui-los e classificá-los.

DO VAZIO AOS INFINITESIMAIS – COMO SE CONSTROEM OS NÚMEROS?

Luís Diogo (3º ano da LMAC — Análise, Geometria e Álgebra)

30 de Março de 2004

Num curso introdutório de Análise Matemática, os números reais são muitas vezes apresentados de forma axiomática, como elementos de um conjunto satisfazendo certas regras. Será que estes axiomas definem os reais de forma única? Será mesmo razoável supor a sua existência? Ideias da teoria de conjuntos permitem-nos estudar estas questões. Veremos até como construir um corpo ordenado que além dos reais contém números infinitesimais e infinitamente grandes. Estas novas entidades permitem-nos codificar de forma natural os comportamentos assintóticos de sucessões.

DISTRIBUIÇÃO QUÂNTICA DE CHAVES

Ana Sofia Graça (3º ano da LMAC — Ciência da Computação)

4 de Maio de 2004

Existe algum método seguro de duas pessoas comunicarem secretamente, usando um computador? Usam-se vários métodos de codificação na internet ou multi-banco, que supomos não interceptáveis. Mas serão eles perfeitamente seguros? É de esperar que a melhoria do suporte tecnológico permita, do ponto de vista quer teórico quer prático, implementar algoritmos de codificação mais sofisticados. Em particular, com o conceito fascinante do computador quântico, põe-se a questão de quão longe se poderá ir? Neste seminário veremos o que são os sistemas criptográficos simétricos e assimétricos usados actualmente; e de que forma os postulados da mecânica quântica, e suas impressionantes consequências, permitirão arranjar uma forma perfeitamente segura de codificar mensagens.

PROBLEMA RESTRITO DOS 3 CORPOS

Ana Knopfli (2º ano da LMAC)

18 de Maio de 2004

Quando estudamos o movimento de 2 corpos, conseguimos obter a solução exacta das suas órbitas a partir das equações do movimento de cada um deles. Porém, basta considerarmos um sistema com apenas mais um corpo para que a situação se complique, e não haja possibilidade de obter soluções exactas, podendo até surgir movimentos caóticos... Neste seminário vamos partir do caso geral de um sistema de n corpos, para os casos em que $n = 2$ e $n = 3$, e analisar em particular o problema restrito dos 3 corpos.

ANÁLISE INFINITESIMAL COM INFINITESIMAIS

Luís Diogo (3º ano da LMAC — Análise, Geometria e Álgebra)

1 de Junho de 2004

Neste seminário vamos apresentar uma extensão do conjunto dos números reais: o conjunto dos números hiperreais. Com ele, podemos obter demonstrações alternativas para os teoremas da Análise Real, em certos casos mais directas que as usuais e que incorporam directamente argumentos que estiveram na génese do Cálculo Infinitesimal. Estruturas deste tipo ampliam a capacidade expressiva da Matemática, podendo ser usadas, por um lado, para caracterizar de forma mais directa conceitos como o de distribuição, e por outro, para introduzir objectos sem paralelo em Análise Standard.

POLÍTOPOS SIMPLES, SOLUÇÕES SIMPLES

Joana Santos (3º ano da LMAC — Análise, Geometria e Álgebra)

12 de Outubro de 2004

Em 1988 Gil Kalai demonstrou de uma maneira “simples”, fazendo apenas uso de geometria elementar e argumentos combinatórios, que se pode reconstruir um polítopo simples a partir do seu grafo, isto é, a partir dos seus vértices e arestas. Este problema, de solução realmente simples em dimensões menores ou iguais a três, mas não tão simples em dimensões superiores, é, tal como outros problemas de combinatória e teoria de grafos, muito importante em programação linear e em optimização de algoritmos computacionais. Nesta apresentação tentaremos perceber o que são polítopos e grafos de polítopos, porque razão o seu estudo é importante e dar algumas ideias da demonstração de Kalai.

PROCESSOS DE MARKOV

Bruno Montalto (2º ano da LMAC)

26 de Outubro de 2004

Os processos (ou cadeias) de Markov modelam evoluções aleatórias – chamam-se rigorosamente processos estocásticos – de sistemas “sem memória”, como por exemplo o dinheiro de que dispomos ao longo de um jogo de roleta, a evolução no jogo da glória ou o número de exemplares de uma espécie numa dada população em estudo. Sendo relativamente simples e bastante comuns, é possível, sem ferramentas matemáticas demasiado sofisticadas, estudar muitas das suas propriedades fundamentais, com várias aplicações de interesse. Neste seminário vamos introduzir estes processos de forma rigorosa, bem como expor algumas das características que os tornam fáceis de estudar. Apresentaremos formas explícitas de “calcular” o seu comportamento, e daremos uma ideia das suas aplicações a vários aspectos da nossa vida.

ÚLTIMO TEOREMA DE FERMAT

Andreia Gomes (3º ano da LMAC — Análise, Geometria e Álgebra)

9 de Novembro de 2004

A equação $x^n + y^n = z^n$ não tem soluções inteiras não nulas para $n > 2$. Um problema de tão simples enunciado resistiu mais de 350 anos aos esforços dos melhores matemáticos. Neste seminário será apresentado o Teorema de Kummer, que prova o Teorema de Fermat para o caso particular do expoente ser um primo regular (conceito a explicar no seminário). Este teorema foi provado em 1850 e ocupa uma posição importante no desenvolvimento da demonstração do UTF. Para acompanhar este seminário bastam conceitos muito elementares de álgebra.

A FUNÇÃO $\pi(x)$ E A HIPÓTESE DE RIEMANN

Nuno Freitas (3º ano da LMAC — Análise, Geometria e Álgebra)

23 de Novembro de 2004

Função Zeta, Teorema dos Números Primos, $\pi(x)$ e Hipótese de Riemann são termos que facilmente se encontram num livro de divulgação, mas o que querem dizer? Neste seminário vamos responder à pergunta anterior através de uma breve visita ao artigo de Bernhard Riemann *On the Number of Prime Numbers Less Than a Given Quantity*. Neste artigo Riemann obtém uma fórmula para a função $\pi(x)$ e levanta uma conjectura que se tornou num dos mais importantes problemas em aberto da Matemática actual: a Hipótese de Riemann.

EQUAÇÕES DIOFANTINAS, TERNOS PITAGÓRICOS E PONTOS RACIONAIS EM CURVAS ELÍPTICAS

Luís Alexandre Pereira (2º ano da LMAC)

7 de Dezembro de 2004

Uma questão importante em Teoria dos Números é encontrar soluções racionais de um polinómio de coeficientes racionais. Os casos linear e de equações a uma variável são de fácil resolução. A duas incógnitas, o caso quadrático (não sendo trivial) está completamente compreendido e sabe-se que equações não-singulares de grau superior ao terceiro só podem ter um número finito de soluções. Neste seminário vamos estudar o caso das cúbicas a duas incógnitas, onde há ainda muitas questões em aberto; por exemplo, não é sequer conhecido um algoritmo para determinar se uma equação tem ou não soluções. No entanto, conhecendo uma delas é possível introduzir no conjunto de pontos solução uma estrutura de grupo abeliano. É essa estrutura de grupo que vamos caracterizar, apresentando alguns resultados relevantes sobre ela conhecidos.

MODELAÇÃO DE PLANTAS E FRACTAIS ATRAVÉS DE GRAMÁTICAS

Luís Gil (4º ano da Licenciatura em Ciências Informáticas)

2 de Março de 2005

Em 1968 o biólogo Aristid Lindenmayer apresentou um modelo matemático para descrever o crescimento e desenvolvimento de plantas. Este modelo utiliza o conceito de gramática, isto é, um conjunto de regras que permite gerar algoritmicamente sequências de símbolos duma dada linguagem. Após a definição de um sistema de interpretação geométrica para estes símbolos torna-se possível gerar imagens de plantas de uma maneira muito realista. Neste seminário vamos descrever este método e utilizá-lo para gerar interactivamente figuras de plantas e fractais.

CONJUNTOS QUE NÃO SE DECIDEM

Tiago Reis (5º ano da LMAC — Ciência da Computação)

16 de Março de 2005

É muito fácil definir conjuntos, por exemplo “o conjunto A dos números primos gémeos”. Mas será assim tão fácil provar propriedades de A? - Será A não vazio? Sim, 3 e 5 são primos gémeos. Conseguirei encontrar o próximo par de elementos de A sem dificuldades? Será que A tem infinitos elementos?...

Neste seminário vamos olhar para conjuntos de números naturais e perceber de onde provêm as dificuldades. Com alguma criatividade poderemos construir conjuntos para os quais não conseguimos responder à simples questão: “Será que este elemento pertence a este conjunto?”.

GEOMETRIA EM SUPERFÍCIES

Jorge Machado (4º ano da Licenciatura em Engenharia Electrotécnica e de Computadores)

30 de Março de 2005

Vista de perto, uma superfície é semelhante a um plano. Generalizando a noção de distância num plano, podemos definir métricas numa superfície. A uma superfície dotada de uma métrica com boas propriedades chamamos uma geometria. Exemplos de geometrias são o espaço euclidiano, a esfera e o plano hiperbólico. Será que existem outros? Será que qualquer superfície admite uma geometria? Será que essa geometria é única? Estas e outras questões serão abordadas neste seminário.

SERÁ POSSÍVEL ENCONTRAR TOPOLOGIA NUMA FÁBRICA?

Daniela Pontes (2º ano da LMAC)

13 de Abril de 2005

Imaginemos uma fábrica e alguns robots que se deslocam por ela para realizarem determinada tarefa, contornando alguns obstáculos. Será que, sabendo o número de robots e qual o seu percurso, poderemos visualizar o seu espaço de configurações? Se sim, qual o aspecto desses espaços? Serão fáceis ou difíceis de visualizar? Haverá maneira de os simplificar? Até que ponto essa simplificação será válida? É no âmbito da topologia que, neste seminário, serão dadas as respostas a estas e outras questões. Para concluir faremos uma introdução à teoria das tranças e veremos como se ajusta a este problema.

LUZ E CÁUSTICAS

Ângela Cardoso (3º ano de Matemática, Faculdade de Ciências da Universidade do Porto)

27 de Abril de 2005

O estudo das propriedades de reflexão e refração da luz remonta à Grécia antiga e tem aplicações que vão da destruição de armadas ao desenho dos faróis dos automóveis. Estas propriedades permitem definir e estudar a formação de cáusticas, curvas que é possível observar usando apenas um copo de vinho. Este seminário será uma oportunidade de redescobrir este tema.

ESTE NÃO É O TÍTULO DESTE SEMINÁRIO!

António Maltsev Santos (1º ano da LMAC)

11 de Maio de 2005

Desde a antiguidade e até aos dias de hoje, têm surgido na matemática, e em particular na lógica, resultados que desafiam a intuição. Alguns, como o título deste seminário, consistem em contradições aparentemente impossíveis, enquanto outros permitem deduzir factos inesperados.

Neste seminário, serão apresentados alguns exemplos de tais resultados e procuraremos compreender quais são realmente paradoxais. A axiomatização de Zermelo-Fraenkel para a teoria dos conjuntos em lógica de primeira ordem será introduzida como ferramenta para desvendar o paradoxo de Russell.

VOLTA AO MUNDO EM 45 MINUTOS

Diogo Oliveira e Silva (3º ano de Matemática, Faculdade de Ciências da Universidade do Porto)

8 de Junho de 2005

Que tal juntarmo-nos a Bernoulli, Catalan, Euler e outros numa intrépida aventura que nos levará a montanhas habitadas por serpentes indomáveis e de onde teremos uma vista privilegiada sobre o mundo da combinatória? Material necessário: alguma vodka, energia q.b. e muita curiosidade!

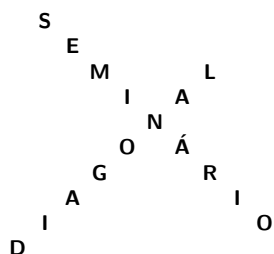
NÃO HÁ 3 SEM 2: O TEOREMA DE SHARKOVSKII

Nuno Mestre (2º ano de Matemática, Faculdade de Ciências e Tecnologia da Universidade de Coimbra)

22 de Fevereiro de 2006

Quando na década de 70 se provou que um sistema dinâmico discreto unidimensional com pontos 3-periódicos tem também pontos de qualquer outro período, “reencontrou-se” o Teorema de Sharkovskii. Este não só contém este facto como o generaliza, criando uma hierarquia completa para a existência de períodos dos pontos de um tal sistema.

Neste seminário vão ser apresentados o ordenamento de Sharkovskii, um esboço da prova do teorema de Sharkovskii e algumas ligações deste ordenamento a outros fenómenos na área dos sistemas dinâmicos.



Cadeias de Markov

Bruno Montalto

4º ano da LMAC — Ciência da Computação

Instituto Superior Técnico

bmontal@math.ist.utl.pt

Palavras Chave

Variável aleatória, estado, classe, tempo de absorção, probabilidade de absorção, recorrência, transitoriedade

Resumo

As cadeias de Markov constituem um dos mais simples e importantes tipos de processos estocásticos, e um importante objecto de estudo nas áreas de Probabilidade e Estatística. A sua simplicidade permite calcular explicitamente vários valores de interesse, o que as torna um objecto matemático prático e útil, capaz de modelar adequadamente as mais diversas situações do dia-a-dia. Neste artigo vamos definir e estudar estas cadeias. Mostraremos como podem ser calculadas explicitamente várias probabilidades a elas associadas, apresentando exemplo simples (teóricos e práticos) de aplicações.

1 Introdução

Numerosos processos aleatórios podem ser descritos como *sem memória*. Isto significa que o que acontece no futuro não depende do que aconteceu no passado, mas apenas do presente. O vulgar jogo da glória, em que os jogadores lançam dados tentando chegar ao fim do tabuleiro antes do adversário, é um exemplo. Outro exemplo é o *Euromilhões*. Um vulgar cidadão, ao deslocar-se aleatoriamente pela cidade, sem querer ir para lado nenhum, pode também ser protagonista de um destes processos.

Tais processos constituem um tipo especial e simples de processos estocásticos de grande interesse: as cadeias de Markov.

É interessante do ponto de vista matemático questionar qual o comportamento destes processos “passado muito tempo”; ou qual a probabilidade de se atingir um determinado estado; ou ainda quando um outro estado será atingido. Quem não se perguntou já quando vai ganhar o *Euromilhões*?

A condição de *amnésia* impõe fortes limitações ao comportamento destes processos, permitindo calcular a resposta a estas e muitas outras perguntas.

Por outro lado, abundam exemplos de situações do mundo real que podem ser assim modeladas, e apresentaremos mesmo algumas neste artigo (será rentável jogar no casino?).

As cadeias de Markov aqui definidas e estudadas são todas de tipo discreto. A bibliografia encontrada sobre este tipo de processo é, contudo, abundante, e os exemplos de aplicações tornam-se ainda mais vastos à medida que alargamos as possibilidades. O estudo da evolução da população de uma espécie e da vida/substituição das lâmpadas de uma loja são apenas alguns exemplos de aplicações de cadeias de Markov em tempo contínuo (algumas das quais podem no entanto ser bem simuladas por cadeias em tempo discreto).

2 Cadeias de Markov

Começamos por apresentar uma definição rigorosa de cadeia de Markov (neste artigo estudaremos apenas o caso *discreto*) e algumas das suas propriedades elementares. A definição a seguir corresponde à ideia intuitiva já exposta, e não deve por isso ser surpreendente.

DEFINIÇÃO 1.

Seja $(X_n)_{n \geq 0}$ uma sucessão de variáveis aleatórias com valores num conjunto contável I (o *espaço de estados*). Diz-se que $(X_n)_{n \geq 0}$ é uma cadeia de Markov com distribuição inicial $\lambda = (\lambda_i, i \in I)$ e matriz de transição $P = (p_{ij}, i, j \in I)^1$ se:

$$(i) \quad P(X_0 = i) = \lambda_i;$$

$$(ii) \quad P(X_{n+1} = i_{n+1} | X_0 = i_0, \dots, X_n = i_n) = p_{i_{n+1}i_n}^2.$$

Nestas condições diz-se também que $(X_n)_{n \geq 0}$ é uma cadeia Markov(λ, P).

Assim, por exemplo, lançar um dado ou jogar no *Euromilhões* são exemplos de cadeias de Markov: em particular, a evolução destes processos nem sequer depende do presente, e a matriz de transição é a identidade. O jogo da Glória também verifica estas condições: ao fim de cada ronda de jogadas, a probabilidade de cada jogador ganhar só depende da posição em que se

1 Esta matriz P pode naturalmente ser infinita, se o conjunto I for numerável; neste caso todas as definições continuam a aplicar-se, substituindo apenas as somas por séries. Da definição é claro que se deverá ter, para todo o $j \in I$, $\sum_{i \in I} p_{ij} = 1$; as matrizes de entradas não-negativas que verificam esta condição dizem-se matrizes *estocásticas*.

2 Ao longo deste artigo designamos por p_{ji} a entrada (ji) da matriz P .

encontra, e não de todos os lançamentos efectuados até ali. Um exemplo de um processo que não é uma cadeia de Markov é a extracção de cartas de um baralho (sem reposição): assim, as cartas que já foram retiradas não podem ser retiradas a seguir, e portanto o próximo "estado", ou a próxima carta retirada, não depende apenas da última carta que se tirou.

TEOREMA 2. *Seja $(X_n)_{n \geq 0}$ uma cadeia Markov (λ, P) . Então, para todo o $n \geq 0$ e todo o $i \in I$, tem-se $P(X_n = i) = (P^n \lambda)_i$.*

Demonstração. Apresentamos uma prova por indução. Por definição de cadeia Markov (λ, P) , temos

$$P(X_0 = i) = \lambda_i = (P^0 \lambda)_i.$$

Suponhamos agora que a asserção é válida para X_n . Então, para cada $i \in I$, tem-se (atendendo ao facto de que "o futuro é independente do passado"),

$$\begin{aligned} P(X_{n+1} = i) &= \sum_{j \in I} (P(X_{n+1} = i | X_n = j) P(X_n = j)) = \\ &= \sum_{j \in I} (p_{ij} (P^n \lambda)_j) = (P^{n+1} \lambda)_i. \end{aligned}$$

□

Este teorema permite-nos desde já resolver um problema interessante, que foi proposto na edição de 2003/2004 das Olimpíadas Ibero-Americanas Universitárias da Matemática. Apresentamos a seguir o enunciado exacto, e uma resolução simples (mas que exige muitos cálculos) proporcionada por este teorema.

Exemplo 1 (O telefone estragado). Algumas crianças estão a brincar ao "telefone sem fios". A criança C_0 sussurra três palavras à criança C_1 , que sussurra o que ouviu à criança C_2 e assim por diante, até uma mensagem chegar à criança C_n . Cada uma das três palavras tem exactamente uma "gémea" errada (por exemplo, as palavras *ração* e *razão* são "gémeas" pois é muito fácil confundi-las). Cada criança C_{i+1} tem 1/2 de probabilidade de ouvir correctamente o que a criança C_i disse, 1/6 de probabilidade de trocar a primeira palavra dita pela criança C_i pela sua "gémea", 1/6 de probabilidade de trocar a segunda palavra e 1/6 de probabilidade de trocar a terceira palavra (e portanto nunca troca mais de uma palavra). Note que numa troca a mensagem pode ser acidentalmente corrigida. Calcule a probabilidade de a criança C_n ouvir exactamente a mensagem original.

Resolução. Para resolver este problema vamos primeiro adaptar a sua linguagem à linguagem mais conveniente das cadeias de Markov. Para $n \geq 0$, seja X_n a variável aleatória que indica o número de palavras erradas que a criança C_n ouviu. Assim, $(X_n)_{n \geq 0}$ é uma cadeia Markov $(\delta_0, P)^3$, onde

$$P = \begin{pmatrix} 1/2 & 1/6 & 0 & 0 \\ 1/2 & 1/2 & 1/3 & 0 \\ 0 & 1/3 & 1/2 & 1/2 \\ 0 & 0 & 1/6 & 1/2 \end{pmatrix}$$

é a matriz de transição.

Pelo teorema anterior, a probabilidade de a criança C_n ouvir a mensagem correcta é então dada por $P(X_n = 0) = (P^n \lambda)_0$. Basta-nos portanto calcular P^n . Para tal usamos o método *standard* de escrever P na forma canónica de Jordan.

Usando técnicas de álgebra linear, concluímos que $P =$

$$\begin{pmatrix} 1 & -1 & 1 & -1 \\ 3 & -1 & -1 & 3 \\ 3 & 1 & -1 & -3 \\ 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 2/3 & 0 & 0 \\ 0 & 0 & 1/3 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1/8 & 1/8 & 1/8 & 1/8 \\ -3/8 & -1/8 & 1/8 & 3/8 \\ 3/8 & -1/8 & -1/8 & 3/8 \\ -1/8 & 1/8 & -1/8 & 1/8 \end{pmatrix},$$

onde a matriz à direita é a inversa da matriz à esquerda. A n -ésima potência de P é então obtida elevando a matriz do meio à n -ésima potência e multiplicando à esquerda e à direita pelas mesmas matrizes. Assim, a probabilidade pedida é

3 Por δ_i designamos o vector que tem 1 na i -ésima entrada e 0 em todas as restantes.

$$\begin{aligned}
P(X_n = 0) &= (P^n \lambda)_0 = \\
&= \left(\begin{pmatrix} 1 & -1 & 1 & -1 \\ 3 & -1 & -1 & 3 \\ 3 & 1 & -1 & -3 \\ 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 2/3 & 0 & 0 \\ 0 & 0 & 1/3 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}^n \right. \\
&\quad \left. \begin{pmatrix} 1/8 & 1/8 & 1/8 & 1/8 \\ -3/8 & -1/8 & 1/8 & 3/8 \\ 3/8 & -1/8 & -1/8 & 3/8 \\ -1/8 & 1/8 & -1/8 & 1/8 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \right)_0 = \\
&= \left(\begin{pmatrix} 1 & -(2/3)^n & (1/3)^n & 0 \\ 3 & -(2/3)^n & -(1/3)^n & 0 \\ 3 & (2/3)^n & -(1/3)^n & 0 \\ 1 & (2/3)^n & (1/3)^n & 0 \end{pmatrix} \begin{pmatrix} 1/8 \\ -3/8 \\ 3/8 \\ -1/8 \end{pmatrix} \right)_0 = \frac{1}{8} \left(1 + \frac{2^n + 1}{3^{n-1}} \right).
\end{aligned}$$

□

3 Probabilidade e Tempo de Absorção

Em certas situações podemos estar interessados em estudar o tempo que uma cadeia de Markov demora a atingir determinado estado, ou a probabilidade de este alguma vez ser atingido. O exemplo seguinte, apresentado de forma informal, é interessante e instrutivo.

Exemplo (A ruína do jogador). Um jogador entra num casino com $i\text{€}$. Em cada jogada aposta 1€ ; se ganhar, o que acontece com probabilidade q , recebe 2€ ; se perder, o que acontece com probabilidade $p = 1 - q$, não recebe nada. Assumindo que os recursos do casino são infinitos⁴ (não havendo assim limite para a fortuna do jogador), qual é a probabilidade de este acabar falido?

Ao longo deste capítulo vamos apresentar resultados e ferramentas que nos permitem responder de forma satisfatória à questão colocada neste exemplo. Começamos por explicitar o que se entende por *classe* de uma cadeia de Markov.

⁴ Este exemplo apresenta uma cadeia de Markov com espaço de estados infinito.

Seja então I o espaço de estados, $(X_n)_{n \geq 0}$ uma cadeia Markov(λ, P), e $i, j \in I$.

DEFINIÇÃO 3. Dizemos que i “leva a” j , e escrevemos $i \rightarrow j$, se existe n tal que $p_{ji}^n = P(X_{m+n} = j | X_m = i) > 0$; dizemos que i “comunica com” j , e escrevemos $i \leftrightarrow j$, se $i \rightarrow j$ e $j \rightarrow i$.

PROPOSIÇÃO 4. $\leftrightarrow$ é uma relação de equivalência.

Demonstração. Há que verificar que $\leftrightarrow$ é reflexiva, simétrica e transitiva.

(i) $\leftrightarrow$ é reflexiva, pois, para $n = 0$ e $i \in I$, $p_{ii}^n = 1 > 0$.

(ii) $\leftrightarrow$ é obviamente simétrica.

(iii) $\leftrightarrow$ é transitiva, pois, se $i \rightarrow j$ e $j \rightarrow k$, existem n_1, n_2 tais que $P(X_{m+n_1} = j | X_m = i) > 0$ e $P(X_{m+n_2} = k | X_m = j) > 0$, e portanto

$$\begin{aligned} P(X_{m+n_1+n_2} = k | X_m = i) &\geq \\ &\geq P(X_{m+n_1+n_2} = k | X_{m+n_1} = j) P(X_{m+n_1} = j | X_m = i) > 0. \end{aligned}$$

□

Esta proposição permite-nos concluir que a relação $\leftrightarrow$ determina uma partição em I . Podemos agora garantir que o conceito de classe de uma cadeia de Markov está bem definido.

DEFINIÇÃO 5. Uma classe de uma cadeia de Markov é qualquer das classes de equivalência em que a relação $\leftrightarrow$ particiona I .

Uma classe C diz-se fechada se $i \in C, i \rightarrow j \Rightarrow j \in C$.

Uma cadeia de Markov diz-se *irredutível* se é constituída por uma única classe.

Introduzimos agora alguma notação que nos auxiliará ao longo do texto. Seja $(X_n)_{n \geq 0}$ uma cadeia de Markov com valores num conjunto I , e $A \subseteq I$. Designamos por H^A a variável aleatória que representa o tempo até a cadeia atingir um elemento de A ; assim,

$$H^A = \inf\{n \geq 0 : X_n \in A\}.$$

Definimos ainda:

- $h_i^A = P_i(H^A < \infty) = P(H^A < \infty | X_0 = i)$. h_i^A é assim a probabilidade de, começando no estado i , a cadeia atingir um elemento de A . Se A é uma classe fechada, h_i^A diz-se a “*probabilidade de absorção*” pela classe A .
- $k_i^A = E_i(H^A) = E(H^A | X_0 = i)$ ⁵. k_i^A é o tempo médio que a cadeia, começando no estado i , demora a atingir um elemento de A . Se A é uma classe fechada, k_i^A diz-se o “*tempo médio de absorção*”.

Antes de enunciarmos o resultado geral que permite calcular explicitamente, mediante a resolução de um sistema de equações lineares, os valores de h_i^A e k_i^A , apresentamos aqui um exemplo simples, que ajuda a compreender a ideia subjacente a este resultado.

Exemplo 2. Considere-se a cadeia de Markov cujo espaço de estados é $I = \{1, 2, 3, 4\}$, que começa no estado 2 e cuja matriz de transição é

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Queremos agora calcular a probabilidade de a cadeia ser “absorvida” pelo estado 4, $h_2^{\{4\}}$. Abreviaremos aqui $h_i^{\{4\}}$ por h_i , para cada $i \in \{1, 2, 3, 4\}$.

Tem-se $h_1 = 0$, $h_4 = 1$; começando em 2 saltamos para 1 ou para 3, com probabilidade $1/2$ em ambos os casos. Uma vez em 3, saltamos para 2 ou para 4 com probabilidade $1/2$ em ambos os casos. Portanto, temos

$$\begin{cases} h_2 = \frac{1}{2}h_1 + \frac{1}{2}h_3 \\ h_3 = \frac{1}{2}h_2 + \frac{1}{2}h_4 \end{cases}$$

Resolvendo este sistema, obtemos então

$$h_2 = \frac{1}{2}h_3 = \frac{1}{2} \left(\frac{1}{2}h_2 + \frac{1}{2} \right) \Rightarrow h_2 = \frac{1}{3}.$$

É fácil intuir que o processo vai eventualmente ser absorvido por $\{1\}$ ou por $\{4\}$. Podemos, de forma igualmente simples, calcular o tempo médio de absorção (pelo conjunto $\{1, 4\}$), notando que

$$\begin{cases} k_1 = k_4 = 0 \\ k_2 = 1 + \frac{1}{2}k_1 + \frac{1}{2}k_3 \\ k_3 = 1 + \frac{1}{2}k_2 + \frac{1}{2}k_4 \end{cases}$$

⁵ Utilizamos aqui a convenção usual de que $E(X)$ representa o valor esperado da variável aleatória X .

Apresentamos agora (sem demonstração) o resultado que nos permite calcular explicitamente estes valores, e vamos ver como este teorema nos permite resolver o problema que colocámos no início desta secção.

TEOREMA 6. *O vector $h^A = (h_i^A : i \in I)$ é a menor⁶ solução não-negativa do sistema linear*

$$\begin{cases} h_i^A = 1 & \text{se } i \in A \\ h_i^A = \sum_{j \in I} p_{ji} h_j^A & \text{se } i \notin A \end{cases}$$

O vector $k^A = (k_i^A : i \in I)$ é a menor solução não-negativa do sistema linear

$$\begin{cases} k_i^A = 0 & \text{se } i \in A \\ k_i^A = 1 + \sum_{j \in I} p_{ji} k_j^A & \text{se } i \notin A \end{cases}$$

Com as ferramentas que desenvolvemos agora podemos então resolver o problema colocado no início desta secção. Recordamos primeiro o seu enunciado.

Exemplo (A ruína do jogador). Um jogador entra num casino com $i\text{€}$. Em cada jogada aposta 1€ ; se ganhar, o que acontece com probabilidade q , recebe 2€ ; se perder, o que acontece com probabilidade $p = 1 - q$, não recebe nada. Assumindo que os recursos do casino são infinitos⁷ (não havendo assim limite para a fortuna do jogador), qual é a probabilidade de este acabar falido?

Resolução. Na linguagem mais precisa das cadeias de Markov, podemos agora reduzir o problema ao de calcular $h_i^{\{0\}}$. Note-se de passagem que $\{0\}$ é uma classe fechada desta cadeia – aliás a única. Escrevendo h_i em vez de $h_i^{\{0\}}$ para simplificar a notação, o sistema que temos de resolver, de acordo com o teorema anterior, é

⁶ Por “menor” entendemos aqui que, se $x = (x_i : i \in I)$ é outra solução do sistema, então $h_i \leq x_i$, para todo o $i \in I$.

⁷ Este exemplo apresenta uma cadeia de Markov com espaço de estados infinito.

$$\begin{cases} h_i^{\{0\}} = 1 & \text{para } i = 0 \\ h_i^{\{0\}} = ph_{i+1} + qh_{i-1} & \text{para } i \neq 0 \end{cases}$$

A solução geral deste sistema é da forma

$$h_i = A + B \left(\frac{q}{p} \right)^i.$$

Se supusermos que o casino é justo, *i.e.*, que $p = q = \frac{1}{2}$, então $\left(\frac{q}{p} \right)^i$ é constante igual a 1 para todo o i ; portanto, $h_i = A + B = h_0 = 1$ para todo o i , e o jogador atinge o estado 0 com probabilidade 1.

□

Este resultado mostra-nos que, mesmo num casino justo, qualquer jogador (e independentemente da fortuna com que comece a jogar) sairá do casino falido, desde que jogue tempo suficiente. Refira-se, a título de curiosidade, que se este jogador dispuser à partida de dinheiro infinito existe uma estratégia muito simples para “assegurar” (ou quase...) que sai sempre a ganhar: basta, em cada jogada, apostar o dobro do que apostou na anterior. Assim, mesmo que tenha perdido todas as jogadas, basta ganhar uma para ganhar mais dinheiro do que todo o que perdeu até então. No entanto, dispondo à partida de dinheiro infinito, é natural que o interesse em ganhar seja bastante reduzido...

Deve ser então intuitivamente claro que, se $q > p$, a conclusão será semelhante. Deixamos ao leitor o exercício de decidir o que fazer, na eventualidade improvável de encontrar um casino para o qual $p > q$...

4 Recorrência e Transitoriedade

Um outro aspecto interessante a estudar numa cadeia de Markov é se um dado estado é *recorrente* ou *transitório*, ou seja, se, começando nesse estado, a cadeia volta a passar por lá infinitas vezes ou não. O exemplo que apresentamos como motivação a seguir é um clássico, e um dos casos mais simples de *Random Walks*.

Exemplo 3 (O Passeio do Bêbado).

Um bêbado incorrigível desloca-se aleatoriamente sobre $\mathbb{Z}$. Se estiver na posição $i \in \mathbb{Z}$, o seu próximo passo leva-o, com probabilidade p , para a posição $i + 1$, e, com probabilidade $q = 1 - p$, para a posição $i - 1$. No início desta história o bêbado encontra-se em casa, na posição 0, pronto para iniciar o seu passeio. Sabendo que o seu passeio continua indefinidamente

independentemente das vezes em que ele regresse a casa, qual a probabilidade de este bêbado se perder definitivamente (ou seja, existir um momento no tempo em que ele sai de casa e nunca mais regressa)?

Nesta secção vamos apresentar resultados que permitem analisar facilmente este problema e decidir a resposta à questão colocada. Alguns dos argumentos aqui utilizados serão de carácter informal; esta opção permite simplificar a exposição sem prejudicar a compreensão intuitiva do que é exposto.

É importante introduzir alguma notação útil.

Denotamos por 1_A a variável aleatória que toma o valor 1 se o acontecimento A se verifica, e 0 caso contrário.

A variável aleatória T_i representa o tempo médio que a cadeia demora, começando no estado i , a voltar a ele.

Definimos ainda a variável aleatória V_i , que representa o número de “visitas” de uma cadeia de Markov ao estado i , *i.e.*,

$$V_i = \#\{n : X_n = i\} = \sum_{n=0}^{\infty} 1_{\{X_n=i\}}.$$

Vamos agora definir estados recorrentes e transitórios. É claro desta definição que para resolver o problema do “*passeio do bêbado*” apenas temos de decidir se o estado 0 é recorrente ou transitório.

DEFINIÇÃO 7. Um estado $i \in I$ diz-se *recorrente* se

$$P_i(X_n = i \text{ para infinitos valores de } n) = P_i(V_i = \infty) = 1$$

. Um estado $i \in I$ diz-se *transitório* se $P_i(V_i = \infty) = 0$.

LEMA. Para $r \in \mathbb{N}$, $P(V_i > r) = (P(T_i < \infty))^r$.

Demonstração. Aplicamos um raciocínio indutivo.⁸ Para $r = 0$ a igualdade é trivial. Assumindo que o resultado é válido para r , temos então

$$\begin{aligned} P_i(V_i > r + 1) &= P_i(V_i > r)P(T_i < \infty) = \\ &= (P_i(T_i < \infty))^r P_i(T_i < \infty) = (P_i(T_i < \infty))^{r+1}. \end{aligned}$$

□

⁸ Para construir um argumento completamente rigoroso seria necessário invocar a chamada “*propriedade forte de Markov*”, que garante que se T é uma variável aleatória tal que o evento $\{T = n\}$ depende apenas de $X_0, \dots, X_n$, e se tem $T < \infty$ e $X_T = i$, então $(X_{T+n})_{n \geq 0}$ é uma cadeia Markov δ_i, P , se $(X_n)_{n \geq 0}$ é uma cadeia Markov (λ, P) .

TEOREMA. Se $P(T_i < \infty) = 1$, i é recorrente. Se $P(T_i < \infty) < 1$, i é transitório.

Demonstração.

$$P(V_i = \infty) = \lim_{r \rightarrow \infty} P(V_i > r) = \lim_{r \rightarrow \infty} (P(T_i < \infty))^n,$$

onde se aplica o lema demonstrado atrás. É fácil ver que o lado direito da igualdade é 1 se $P(T_i < \infty) = 1$ e 0 se $P(T_i < \infty) < 1$, o que implica o enunciado do teorema, por definição de estado recorrente e transitório.

□

É relativamente simples verificar que se C é uma classe então ou todos os elementos de C são recorrentes ou todos os elementos de C são transitórios, o que justifica as expressões “*classe recorrente*” e “*classe transitória*”. Também é simples demonstrar que toda a classe recorrente é fechada e que toda a classe fechada finita é recorrente.

Resta apenas estudar o caso das classes fechadas infinitas, como é o caso da classe (única) $\mathbb{Z}$ no exemplo apresentado no início da secção. Para estudar este exemplo demonstraremos ainda mais um teorema.

TEOREMA. Se V é uma variável aleatória com valores em $\mathbb{N}$, então $E(V) = \sum_{r=0}^{\infty} P(V > r)$.

Demonstração.

$$\begin{aligned} \sum_{r=0}^{\infty} P(V > r) &= \sum_{r=0}^{\infty} \sum_{v=r+1}^{\infty} P(V = v) = \\ &= \sum_{v=1}^{\infty} \sum_{r=0}^{v-1} P(V = v) = \sum_{v=1}^{\infty} v \cdot P(V = v) = E(V). \end{aligned}$$

□

Se V_i é a variável aleatória que conta o número de visitas de uma cadeia de Markov ao estado i , então

$$E_i(V_i) = E_i \left(\sum_{n=0}^{\infty} 1_{\{X_n=i\}} \right) = \sum_{n=0}^{\infty} E_i(1_{\{X_n=i\}}) = \sum_{n=0}^{\infty} P_i(X_n = i).$$

Usando este facto e o lema já provado, obtemos o seguinte

TEOREMA 8. Um estado i é recorrente se e só se $\sum_{n=0}^{\infty} P_i(X_n = i) = \infty$.

Demonstração.

($\Rightarrow$) Se i é recorrente, temos $P_i(V_i = \infty) = 1$. Assim,

$$\infty = E(V_i) = E\left(\sum_{n=0}^{\infty} 1_{\{X_n=i\}}\right) = \sum_{n=0}^{\infty} E(1_{\{X_n=i\}}) = \sum_{n=0}^{\infty} P(X_n = i),$$

como queríamos demonstrar.

($\Leftarrow$) Se i é transitório, então, de acordo com Teorema 3.4, $P(T_i < \infty) < 1$. Combinando os vários resultados desta secção, obtemos

$$\begin{aligned} \sum_{n=0}^{\infty} P_i(X_n = i) &= E_i(V_i) = \sum_{r=0}^{\infty} P_i(V_i > r) = \\ &= \sum_{r=0}^{\infty} (P(T_i < \infty))^r = \frac{1}{1 - P_i(T_i < \infty)} < \infty, \end{aligned}$$

o que conclui a demonstração. $\square$

Concluimos esta exposição com a resposta à questão colocada no início desta secção. Mais uma vez começamos por recordar o enunciado do problema.

Exemplo (O Passeio do Bêbado).

Um bêbado incorrigível desloca-se aleatoriamente sobre $\mathbb{Z}$. Se estiver na posição $i \in \mathbb{Z}$, o seu próximo passo leva-o, com probabilidade p , para a posição $i + 1$, e, com probabilidade $q = 1 - p$, para a posição $i - 1$. No início desta história o bêbado encontra-se em casa, na posição 0, pronto para iniciar o seu passeio. Sabendo que o seu passeio continua indefinidamente independentemente das vezes em que ele regresse a casa, qual a probabilidade de este bêbado se perder definitivamente (ou seja, existir um momento no tempo em que ele sai de casa e nunca mais regressa)?

Resolução. Seja $(X_n)_{n \geq 0}$ a cadeia de Markov em que X_n representa a posição do bêbado após n passos. Temos $X_0 = 0$. Vamos estudar a convergência da série $\sum_{n=0}^{\infty} P(X_n = 0)$ e usar o último teorema para tirar as nossas conclusões.

É claro que, para estar na posição 0, o bêbado teve de dar r passos para a direita e r passos para a esquerda. Portanto, para m ímpar, temos $P(X_m = 0) = 0$.

Para $m = 2n$ par, existem $\binom{2n}{n}$ caminhos que o bêbado pode tomar que consistem em n passos para a direita e n passos para a esquerda e que, por isso, o levam de volta a casa. A probabilidade de cada um destes caminhos é $p^n q^n = (pq)^n$. A série a estudar é, portanto,

$$\sum_{n=0}^{\infty} \left(\binom{2n}{n} (pq)^n \right).$$

Se $p \neq q$, a desigualdade aritmética/geométrica permite-nos concluir que $pq < \frac{1}{4}$. Neste caso, como

$$\lim_{n \rightarrow \infty} \frac{\binom{2(n+1)}{n+1} (pq)^{n+1}}{\binom{2n}{n} (pq)^n} = \frac{(2n+1)(2n+2)}{(n+1)^2} pq < 1,$$

concluimos que a série é convergente e, portanto, o bêbado vai, um dia, perder-se definitivamente.

Suponhamos agora que $p = q = \frac{1}{2}$. A série transforma-se então em $\sum_{n=0}^{\infty} \frac{1}{4^n} \binom{2n}{n}$. Note-se que $\binom{2n}{n}$ é o maior dos $2n+1$ números da forma $\binom{2n}{p}$, e que

$$\sum_{p=0}^{2n} \binom{2n}{p} = 2^{2n} = 4^n.$$

Daqui concluimos que $\binom{2n}{n} > \frac{4^n}{2n+1}$. Portanto, no caso em que $p = q = \frac{1}{2}$, a série

$$\sum_{n=0}^{\infty} \binom{2n}{n} (pq)^n = \sum_{n=0}^{\infty} \frac{1}{4^n} \binom{2n}{n} > \sum_{n=0}^{\infty} \frac{1}{2n+1} > \infty$$

diverge pelo critério de comparação.

□

Concluimos portanto que o bêbado, mesmo deambulando eternamente, acaba sempre por voltar a casa, desde que consiga manter o equilíbrio!⁹

Agradecimentos

Gostaria de agradecer ao professor Diogo Gomes, pela revisão deste artigo e pelo seu excelente trabalho como tutor durante a realização deste projecto, e à Fundação Calouste Gulbenkian, que possibilitou a sua existência através do programa “*Novos Talentos em Matemática*”.

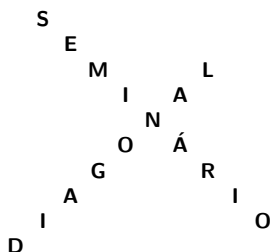
⁹ Refira-se ainda, para concluir, que este mesmo resultado é válido se o bêbado se deslocar em $\mathbb{Z}^2$. No entanto, se o bêbado voasse, deslocando-se assim em $\mathbb{Z}^3$, já não teria a mesma sorte...

Agradeço também à professora Ana Cannas da Silva, por toda a disponibilidade e dinamismo sempre demonstrados, desde o momento de procurar tutor até à apresentação do trabalho realizado, no Encontro Nacional dos “*Novos Talentos em Matemática*”.

Uma palavra de agradecimento ainda para a organização do Seminário Diagonal do IST, por me ter proporcionado a oportunidade de publicar este artigo.

Referências

- [1] J. R. Norris, *Markov Chains*, Cambridge University Press, 1997.
- [2] P. Brémaud, *Markov Chains : Gibbs Fields, Monte Carlo Simulation, and Queues*, Springer, 1999.
- [3] S. M. Ross, *Stochastic Processes*, John Wiley & Sons, 1996.



Modelação de Plantas e Fractais a Partir de Gramáticas

Luís Gil

4º ano da LCI

Instituto Superior Técnico

lcmg@mega.ist.utl.pt

Palavras Chave

L-System, fractal, modelação de plantas, gramática, computação gráfica

Resumo

A natureza sempre foi um fascínio para os poetas e pintores que desde sempre a representaram em versos ou pinturas. Mais recentemente, os matemáticos juntaram-se a este grupo através da representação de árvores e plantas por modelos denominados L-Systems, que permitem descrever uma planta em qualquer fase do seu desenvolvimento e posteriormente obter uma imagem desta.

Neste artigo iremos definir gramáticas e de seguida apresentar uma descrição dos L-Systems e de produção de imagens, bem como mostrar um conjunto de exemplos que demonstram o nível de realismo e a qualidade que podem ser obtidos.

1 Introdução

Aprender é como remar contra a maré: é só parar e anda-se para trás.

Descrever a natureza a partir da Matemática dá a impressão de ser algo extremamente complicado ou despropositado. Os mais desatentos podem perguntar: mas o que uma coisa tem a ver com a outra? Poderemos descrever o crescimento de uma árvore ou de uma flor matematicamente?

A resposta é afirmativa e a investigação começou em 1968 pela mão do biólogo Aristid Lindenmayer, com a utilização de gramáticas – designadas por L-Systems – para descrever o crescimento de plantas, árvores e mais tarde de certos tipos de fractais.

De seguida foi inventado um método para produzir imagens a partir das representações textuais e, com o aumento das capacidades gráficas dos computadores a partir dos anos 80, a qualidade e complexidade das representações aumentou substancialmente, permitindo obter imagens com grande realismo.

No presente os L-Systems são utilizados na produção de ambientes virtuais e em jogos de computador, enquanto a investigação continua e se procuram outras aplicações para estes métodos.

2 Gramáticas

*Para bom mestre não
há má ferramenta.*

2.1 Definição

As gramáticas são utilizadas para descrever a estrutura de uma linguagem, isto é, estabelecem as regras de formação das palavras.

DEFINIÇÃO 1 (PALAVRA). Uma palavra é uma sequência (que pode ser vazia) de símbolos de um dado alfabeto.

DEFINIÇÃO 2 (LINGUAGEM). Uma linguagem é um conjunto finito ou infinito de palavras escritas com símbolos de um determinado alfabeto.

DEFINIÇÃO 3 (GRAMÁTICA). Uma gramática é um quádruplo $\langle \Sigma, V, R, S \rangle$ onde:

- Σ é um conjunto (não vazio e finito) de símbolos terminais;
- V é um conjunto (não vazio) de símbolos variáveis;
- R é um conjunto de regras ou produções;
- $S \in V$ é a variável inicial.

Apresenta-se com detalhe a análise dos elementos da gramática.

- Σ é o alfabeto da linguagem, o conjunto de símbolos que compõem as palavras. Os elementos do alfabeto são normalmente representados por caracteres minúsculos do alfabeto latino.
- V é um conjunto de símbolos que representam as partes de uma palavra que ainda não foram geradas. Os símbolos variáveis, ou simplesmente variáveis, denotam-se por letras maiúsculas do alfabeto latino.

- R é um conjunto de regras que definem o modo de formação das palavras e relacionam os símbolos variáveis com sequências de terminais e variáveis.

Uma produção é constituída por três elementos:

1. uma sequência de símbolos terminais e variáveis, com comprimento maior que zero também conhecida por cabeça da produção;
2. o símbolo de produção $\rightarrow$;
3. uma palavra com a mesma estrutura da cabeça e com comprimento maior ou igual a zero denominada por corpo da produção.

Podem existir várias regras com a mesma cabeça e corpos diferentes.

- S é a variável inicial a partir do qual se geram todas as palavras da linguagem.

NOTAÇÃO 4 (SEQUÊNCIA VAZIA). *A palavra de comprimento zero é normalmente representada pelas letras ϵ ou λ .*

NOTAÇÃO 5 (ABREVIATURA NA REPRESENTAÇÃO DAS PRODUÇÕES). *Quando várias produções possuem a mesma cabeça, podemos representá-las abreviadamente na forma cabeça $\rightarrow$ corpo 1 | ... | corpo N .*

Vejamos dois exemplos de gramáticas.

Exemplo 1 (Gramática dos parêntesis bem encadeados).

$$\Sigma = \{ (,) \}$$

$$V = \{ P \}$$

Regras:

$$P \rightarrow (P)$$

$$P \rightarrow PP$$

$$P \rightarrow \epsilon$$

ou na forma abreviada: $P \rightarrow (P) \mid PP \mid \epsilon$

$$S = P$$

Exemplo 2 (Palavras binárias que começam em 1 e acabam em 0).

$$\Sigma = \{ 0 , 1 \}$$

$$V = \{ A, B, C \}$$

Regras:

$$A \rightarrow 1B$$

$$B \rightarrow C0$$

$$C \rightarrow C0 \mid C1 \mid \epsilon$$

$$S = A$$

2.2 Derivação de uma Palavra

Após ser definida a gramática para uma linguagem, podemos obter todas as palavras desta através do seguinte algoritmo de derivação:

1. escrever a variável inicial;
2. substituir uma cabeça de produção presente na palavra pelo respectivo corpo;
3. se ainda existirem cabeças de produções presentes na palavra, voltar a 2.

Terminado o algoritmo obtemos a palavra, composta apenas por símbolos terminais.

NOTAÇÃO 6. O símbolo utilizado para indicar um passo da derivação é $\Rightarrow$.

Exemplo 3.

A partir da gramática dos parêntesis vamos derivar a palavra $((()))$:

$$P \Rightarrow (P) \Rightarrow (PP) \Rightarrow ((P)P) \Rightarrow ((P)(P)) \Rightarrow ((()))$$

Exemplo 4.

Vamos derivar a sequência 100 a partir da gramática binária:

$$A \Rightarrow 1B \Rightarrow 1C0 \Rightarrow 1C00 \Rightarrow 100$$

No último passo de cada exemplo as variáveis foram substituídas por ϵ .

2.3 Tipos de Gramáticas

As gramáticas podem ser divididas em várias categorias de acordo com a estrutura da produção, que é responsável pela quantidade e tipo de linguagens que podem ser descritas. Apresentam-se os tipos de gramáticas por ordem crescente no poder expressivo.

- **Gramáticas Regulares:**

$$A \rightarrow c|\epsilon$$

$$A \rightarrow Bc$$

$$A \rightarrow cB$$

As duas últimas produções são classificadas de recursivas à esquerda e à direita respectivamente.

- **Gramáticas Livres de Contexto:**

$$A \rightarrow (V \cup \Sigma)^*$$

As produções podem ser qualquer sequência de símbolos variáveis ou terminais.

- **Gramáticas Sensíveis ao Contexto:**

$$aAC \rightarrow aBC$$

É possível indicar que uma variável só pode ser substituída se for precedida ou sucedida por outra(s) variáveis ou terminais. Neste exemplo, A só se pode transformar em aBC se estiver entre um a e um C .

- **Gramáticas Recursivamente Enumeráveis:**

$$ABC \rightarrow aD$$

Nestas gramáticas não existem restrições quanto ao formato da cabeça da produção, havendo a possibilidade de substituir vários símbolos por outros.

Apresentam-se as relações de inclusão entre as linguagens geradas pelas várias gramáticas.

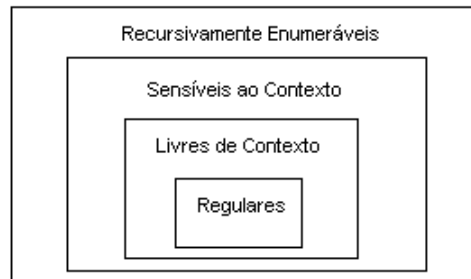


Figura 1: Relação entre as linguagens

Por fim vamos listar aplicações para as diferentes gramáticas.

- Gramáticas Regulares: utilizadas essencialmente em analisadores léxicos de compiladores ou em linhas de comandos dos sistemas operativos para identificar palavras com certos formatos.
- Gramáticas Livres de Contexto: utilizadas para descrever as linguagens de programação e gerar os analisadores sintáticos (*parsers*) que verificam os erros nos programas.
- Gramáticas Sensíveis ao Contexto: descrevem as linguagens humanas.
- Gramáticas Recursivamente Enumeráveis: descrevem todas as linguagens que podem ser reconhecidas por um computador.

3 L-Systems

Da discussão nasce a luz.

Os L-Systems são um modelo, um formalismo matemático baseado em gramáticas, que o biólogo Aristid Lindenmayer apresentou em 1968 para descrever o modo de crescimento das plantas.

A partir dos anos 80 os L-Systems foram aplicados, através da computação gráfica, na modelação realística de plantas e geração de fractais, obtendo-se imagens destes em qualquer fase do crescimento.

A produção de uma imagem passa pelas seguintes fases:

1. criação do L-System com a descrição do desenvolvimento da planta/fractal;
2. obtenção da representação textual da planta/fractal numa certa fase de crescimento;
3. interpretação geométrica da representação para gerar a imagem.

3.1 Definição

DEFINIÇÃO 7 (L-SYSTEM). Um L-System é um quintuplo $\langle N, \Sigma, V, R, A \rangle$ onde:

- N é o número de etapas ou passos de crescimento;
- Σ é um conjunto de símbolos terminais;
- V um conjunto de variáveis;
- R é um conjunto de regras de crescimento;
- A é uma sequência inicial normalmente designado por axioma.

Pela sua estrutura constata-se a semelhança do L-System com a gramática, com o acréscimo das etapas. Mas existem outras diferenças que passamos a apresentar e que serão analisadas em detalhe nas secções seguintes:

- o axioma passa a poder ser uma sequência de símbolos terminais e variáveis, ao contrário de um único símbolo variável permitido pelas gramáticas;
- a palavra final é composta por símbolos terminais e variáveis;

- na derivação as substituições das variáveis pelos corpos das produções são feitas simultaneamente.

Note-se que vamos assumir que o corpo das produções é composto apenas por uma sequência de símbolos. A existência de várias sequências implica a utilização de um critério de escolha apenas existente em L-Systems mais complexos que serão apresentados numa secção ulterior.

Exemplo 5 (Planta simples).

Apresenta-se um exemplos introdutório

$$N = 2$$

$$\Sigma = \{ +, -, [,] \}$$

$$V = \{ F, G \}$$

Regras:

$$G \rightarrow F[+G][-G]FG$$

$$F \rightarrow FF$$

$$\text{Axioma} = G$$

3.2 Derivação de uma Palavra

Apresenta-se o algoritmo de derivação de uma palavra. Difere no das gramáticas em dois aspectos: a condição de terminação e a aplicação das substituições.

1. escrever o axioma;
2. substituir simultaneamente as variáveis pela respectiva sequência associada;
3. se o número de passos ainda não foi atingido, voltar a 2.

Quando o algoritmo termina temos uma palavra composta por símbolos terminais e variáveis.

A substituição das variáveis é simultânea por analogia às plantas cujo crescimento ocorre em todo o lado ao mesmo tempo.

Exemplo 6 (Planta simples – continuação).

Vejamos agora as derivações para os dois primeiros passos do exemplo anterior (para tornar mais clara a derivação temos a *itálico* o corpo da regra de G e a **negrito** o corpo da regra de F).

$$n=0: G$$

$$n=1: F[+G][-G]FG$$

$$n=2: \mathbf{FF} [+ F[+G][-G]FG][- F[+G][-G]FG] \mathbf{FF} F[+G][-G]FG$$

Estas sequências definem a representação textual da planta nos referidos estágios de desenvolvimento.

3.3 Obtenção da Imagem

Para gerar a imagem é necessário interpretar geometricamente a referida representação textual. O modo escolhido foi o gráfico de tartaruga, utilizado pela primeira vez na antiga linguagem de programação LOGO. Baseia-se em dar ordens (codificadas nas letras da palavra produzida) de movimento, rotação e desenho a uma estrutura (a tartaruga) que se desloca ao longo de um plano cartesiano tridimensional.

As ordens de desenho estão expressas nos símbolos da palavra que é derivada e dividem-se em várias categorias. Antes de as apresentarmos exaustivamente, vamos ver as operações principais e um pequeno exemplo.

F	move em frente e desenha uma linha
+	roda para a esquerda em torno do eixo do Z.
−	roda para a direita em torno do eixo do Z.
[	guarda a posição actual numa pilha
]	altera a posição actual para a que está no topo da pilha

Tabela 1: Comandos básicos de desenho a duas dimensões

- Os símbolos variáveis indicam o desenho de uma linha, referindo-se aos troncos e ramos das plantas.
- Os símbolos de rotação $+$ e $-$ regulam a orientação das ramificações.
- $[$ e $]$ são símbolos de controlo que manipulam posição da estrutura de desenho para produzir os ramos das plantas. Servem respectivamente para memorizar a posição actual ou voltar ao último local guardado, sendo a informação registada numa pilha. Correspondem às funções de manipulação de pilhas *push* (inserção no topo da pilha) e *pop* (obtenção do elemento no topo e posterior remoção).

Exemplo 7 (Imagens da Planta Simples).

Vejamos as imagens produzidas a partir das palavras geradas no exemplo da secção anterior. Vamos assumir que a estrutura de desenho está inicialmente orientada segundo o vector $(x,y,z) = (0,1,0)$. Vamos descrever a formação da figura 3, analisando a interpretação geométrica de cada símbolo da palavra.

- F : indica o deslocamento em frente e o desenho do caule.
- [: uma cópia da posição actual é guardada na pilha.
- + : a estrutura roda para a esquerda.



Figura 2: G

Figura 3: $F[+G][-G]FG$ Figura 4: $FF [+ F[+G][-G]FG][- F[+G][-G]FG] FF F[+G][-G]FG$

- G : é desenhado o ramo esquerdo.
-] : a estrutura torna-se a lida da pilha e regressa ao local anterior, readquirindo a orientação vertical.
- [: a cópia da posição actual é novamente guardada na pilha.
- - : a estrutura roda para a direita.
- G : o ramo direito é desenhado.
-] : a estrutura volta a ser a lida da pilha.
- F e G : são desenhadas mais duas linhas verticais

Vejamos tabelas com os símbolos mais comuns e o respectivo significado geométrico.

Esta listagem não é exaustiva nem convencional, no sentido que diferentes autores podem representar uma operação de através de símbolos distintos. Certos autores introduzem outros tipos de interpretações, como outras orientações (voltar 180° ou rodar até à horizontal) ou até o efeito da gravidade nos ramos.

Os parâmetros que sucedem aos símbolos serão explicados na secção seguinte.

$f(s)$	move em frente sem desenhar
--------	-----------------------------

Tabela 2: Símbolo de movimento sem desenho

$\&(a)$	roda para a direita à volta do eixo do Y.
$\wedge(a)$	roda para a esquerda à volta do eixo do Y.
$/ (a)$	roda para a esquerda em torno do eixo do X.
$\backslash (a)$	roda para a direita em torno do eixo do X.

Tabela 3: Símbolos de orientação em 3D

$\{$	guarda as posições seguintes como vértices de um polígono a preencher
$\}$	preenche o polígono
$@O(r)$	desenha uma esfera de raio r

Tabela 4: Superfícies

$\#(w)$	coloca a largura da linha a w, ou incrementa-a (sem parâmetro)
$!(w)$	idem, ou decrementa-a quando invocado sem parâmetro
$;(n)$	incrementa o mapa de cor
$,(n)$	decrementa o mapa de cor

Tabela 5: Atributos de desenho

3.4 Tipos de L-Systems

Tal como as gramáticas, existem várias categorias de L-Systems, também de acordo com a estrutura das produções.

- **Sistemas livres de contexto:** são os mais simples, que apresentam apenas as operações elementares de rotação e desenho em que o corpo de cada regra é composto apenas por uma sequência de símbolos.
- **Sistemas paramétricos:** permitem acoplar às variáveis e aos símbolos terminais parâmetros que indicam o modo como a interpretação destes é feita (desenha uma linha de comprimento x, roda y graus). O corpo das regras já pode conter várias palavras, sendo a escolha feita em função do valor do parâmetro do símbolo variável.

As produções são da forma $V(x) : \text{condição} \rightarrow \text{sequência}$

Por exemplo:

$F(x) : x > 1 \rightarrow F(x-1)G(x-1)$

$F(x) : x = 1 \rightarrow G(5)[+(45)F(5)][-(45)F(5)]F(5)$

- **Sistemas probabilísticos:** neste tipo de sistemas, as sequências dos corpos das produções estão associadas a uma probabilidade de ocorrência. Deste modo, a aplicação de uma regra é determinada durante a derivação da representação textual do sistema por um processo aleatório. O benefício introduzido por este tipo de aplicação das regras é a possibilidade de obter sempre imagens diferentes de um só sistema.

As produções são da forma:

$V : \text{probabilidade } 1 \rightarrow \text{sequência } 1$

$V : \text{probabilidade } N \rightarrow \text{sequência } N$

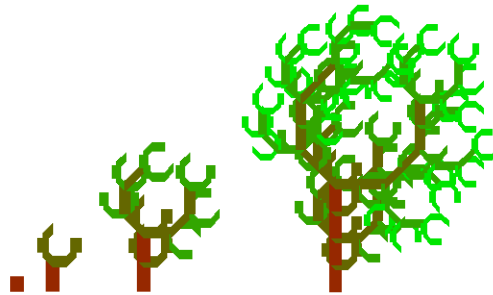
com probabilidade $1 + \dots + \text{probabilidade } N = 1$.

- **Sistemas sensíveis ao contexto:** as produções assemelham-se às das gramáticas da mesma categoria.

4 Exemplos

*Cada árvore que se planta é
um bom legado que se deixa.*

Nesta secção vamos apresentar um conjunto de imagens e alguns dos L-Systems correspondentes (passos, regras e axioma). Todas as imagens foram geradas a partir de exemplos fornecidos com os programas listados nas referências.

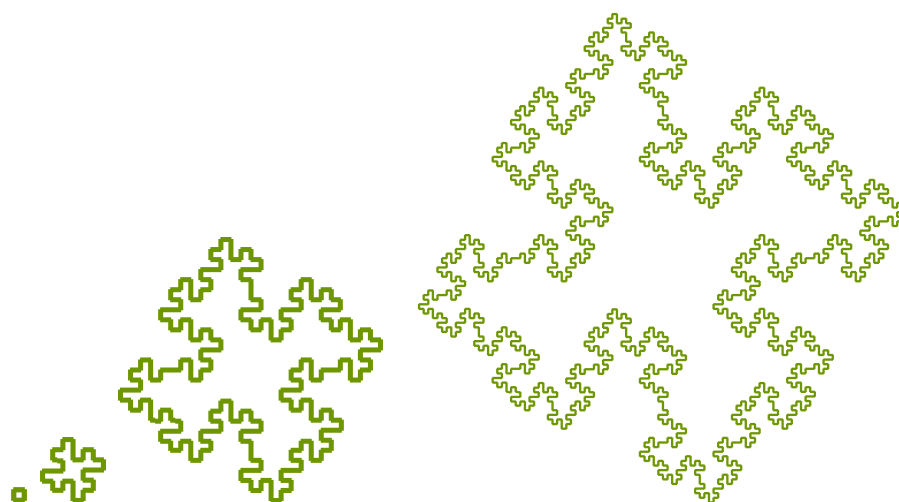


```
passos 3
axioma F
F -> FF-[-F+F+F]+[+F-F-F]
```

Figura 5: Arbusto



Figura 6: Arbusto a três dimensões



passos 3
 axioma $F+F+F+F$
 $F \rightarrow F+F-F-FF+F+F-F$

Figura 7: Ilha de Koch

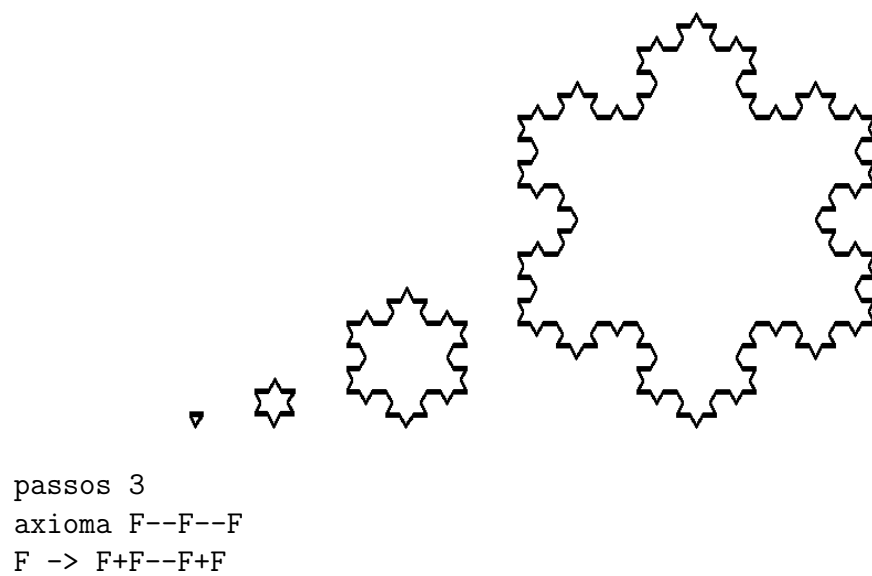


Figura 8: Floco de neve

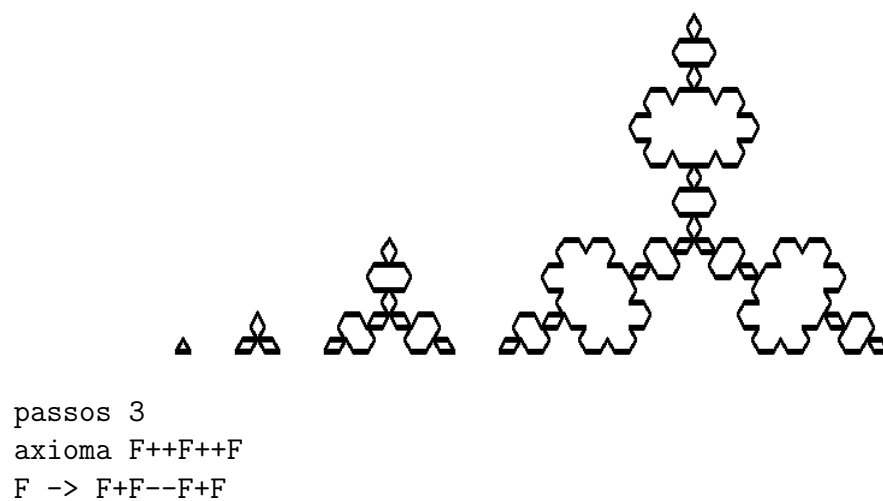
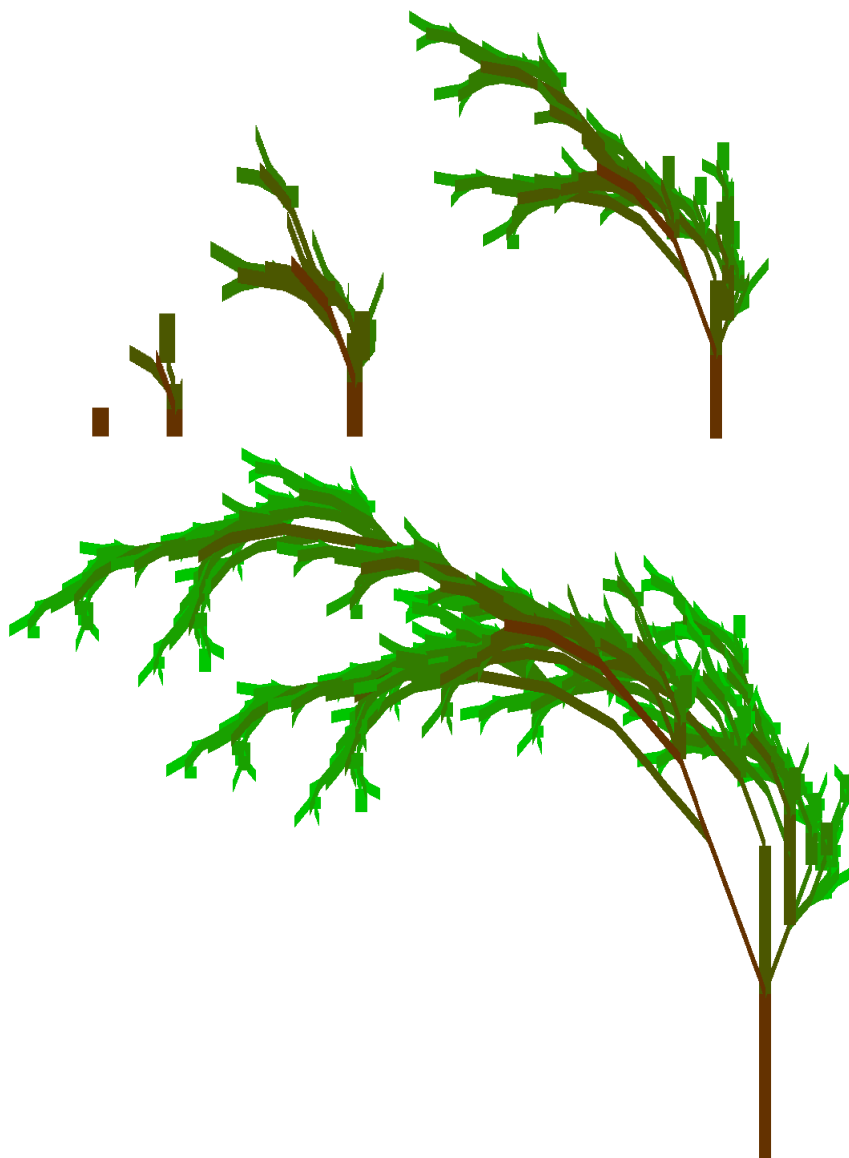
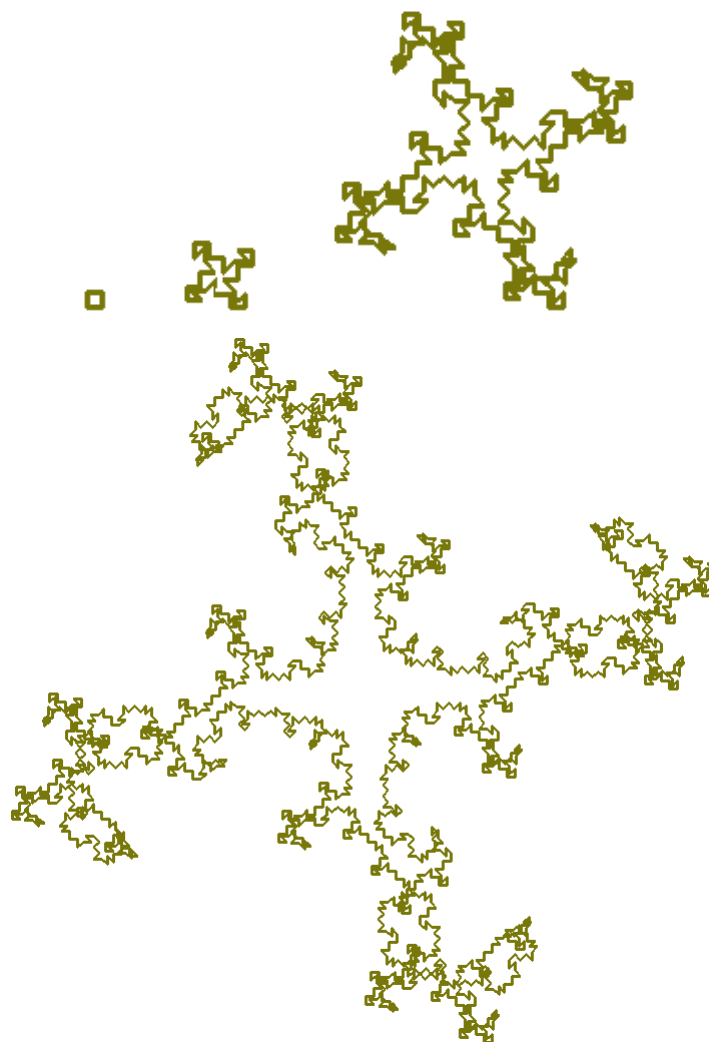


Figura 9: Floco de neve com as rotações do axioma invertidas



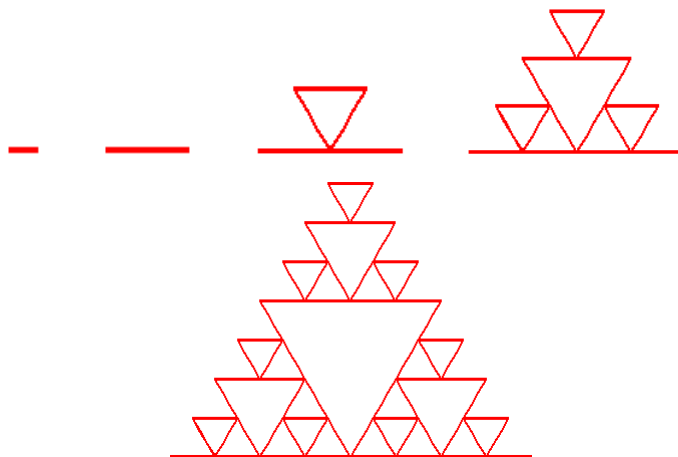
passos 4
 axioma B
 $A \rightarrow AA$
 $B \rightarrow A[-B][A+B-AB]+A[+A+B]B$

Figura 10: Planta curvada



passos 3
axioma F++F++F++F
F -> F+++F---F+F---F++F--F++F

Figura 11: Lace



passos 4
 axioma F
 $F \rightarrow FXF$
 $X \rightarrow +FXF-FXF-FXF+$

Figura 12: Triângulo de Sierpinski



Figura 13: Planta violoncelo



Figura 14: Sistema probabilístico



Figura 15: Árvore nas etapas 2, 4, 6 e 8



Figura 16: Planta em espiral



Figura 17: Planta com flores

5 Conclusões

Resumindo e concluindo.

As gramáticas já não são estudadas com o mesmo fervor dos anos 70 e 80, quando a investigação para produzir compiladores atingiu o seu auge. Pelo contrário, agora os compiladores e analisadores léxicos são utilizados maciçamente numa grande variedade de aplicações. O leitor interessado em Teoria das Linguagens pode consultar [1, 2, 5].

A modelação de plantas através de gramáticas impulsiona a investigação de novos modos de descrição de plantas e de programas informáticos para produzir imagens mais realísticas ou complexas; ou é aplicada em jogos de computador, em simuladores de paisagens ou outros ambientes virtuais e mais recentemente na geração de música [7], onde os símbolos da palavra são interpretados como notas musicais, alteração de ritmos, escalas, instrumentos, etc.

Agradecimentos

Metade do pagar é agradecer.

Quero agradecer à organização do Seminário Diagonal: Ana Rita Pires, Jorge Silva, Luís Cruz-Filipe e Nuno Freitas pela oportunidade para apresentar este trabalho e pelo apoio dado e sugestões feitas durante a preparação do mesmo. Agradeço ao professor José Félix Costa por me ter emprestado o livro *The Algorithmic Beauty of Plants* e ter sugerido fazer um seminário sobre o assunto.

Aos colegas Carlos Tamulonis, Hélio Pais e Paulo Abrantes agradeço a revisão deste artigo.

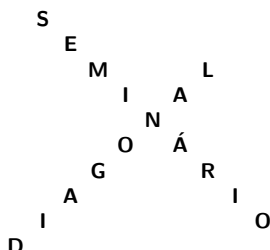
Muito obrigado a todos.

Referências

- [1] R. G. Crespo, *Processadores de Linguagens*, 2ª edição, IST Press, 2001.
- [2] J. Hopcroft, R. Motwani e J. Ullman, *Introduction to Automata Theory, Languages and Computation*, 2nd edition, Addison Wesley, 2001.
- [3] A. Lindenmayer *et al*, *The Algorithmic Beauty of Plants*, Springer, 1990 (disponível em <http://www.algorithmicbotany.org>).

- [4] P. Prusinkiewicz, M. Hammel, J. Hanan e R. Mech, L-Systems: from the theory to visual models of plants, *Proceedings of the 2nd CSIRO Symposium on Computational Challenges in Life Sciences*, 1996.
- [5] M. Sipser, *Introduction to Theory of Computation*, PWS Publishing Company, 1997.
- [6] LinSys3D - <http://www.esuli.it/Software/LinSys3d/>
- [7] LMUSE -
<http://www.geocities.com/Athens/Academy/8764/lmuse/lmuse.html>
- [8] LParser - <http://home.wanadoo.nl/laurens.lapre/>
- [9] SistemaL - <http://web.ist.utl.pt/ist151334/>

*Procura tempo para ler.
É a fonte da inteligência.*



Geometria em Superfícies

Fernando Machado

5º ano da LEEC

Instituto Superior Técnico

jorgemac@gmail.com

Palavras Chave

Geometria, superfície, métrica, isometria, plano euclidiano, superfície esférica, plano hiperbólico.

Resumo

É sabido desde o século XIX que em espaços bidimensionais existe uma relação íntima entre a geometria e a topologia, sendo possível classificar topologicamente os diversos tipos de superfícies com recurso a métricas.

Nas últimas décadas, e ao contrário do que se julgava não ser possível, o trabalho de Thurston permitiu estender esta classificação às variedades-3, e possivelmente solucionando um dos enigmas da Matemática: a conjectura de Poincaré.

Neste artigo analisa-se a noção de geometria e algumas aplicações aos espaços bidimensionais, as superfícies. No final procura-se descrever a generalização dos resultados ao espaço tridimensional.

1 Introdução

1.1 Alguns Conceitos de Geometria

Uma formiga, de uma galáxia distante, habita num dos três planetas do sistema solar representado na figura seguinte. Ao observar em seu redor, a formiga apenas avista alguns quilómetros, e como tal não consegue compreender que o seu planeta é curvo, julgando que se encontra sobre um plano.

Em Matemática, conjuntos com a propriedade de que a vizinhança de um ponto é semelhante (*homeomorfa*) a um plano são denominados de superfícies. Os conjuntos a (esfera), b (toro) e c da Fig. 1 são alguns exemplos

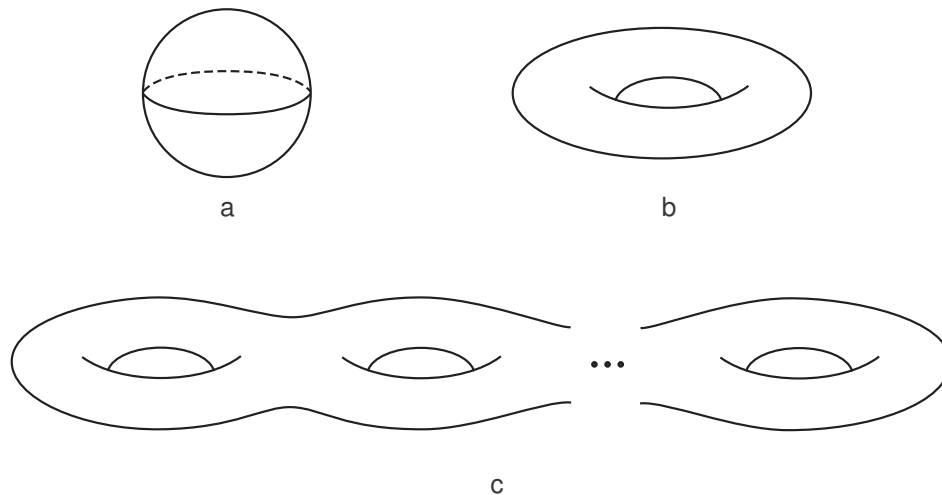


Figura 1: Alguns exemplos de superfícies

de superfícies¹.

Quando a formiga caminha sobre a superfície M do seu planeta, define uma curva $c(t) \in M$. A derivada desta curva no instante $t = 0$ é um vector velocidade $\dot{c}(0)$, e trata-se de um vector tangente no ponto $c(0)$, como se representa na Fig. 2. De facto, os vectores tangentes são formas de indicar uma direcção sobre a superfície.

O conjunto dos vectores tangentes a uma superfície M num dado ponto x forma o espaço tangente a esse ponto, $T_x M$, e a colecção de todos os espaços tangentes constitui o fibrado tangente da superfície, representado por TM .

A geometria não sobrevive apenas com a noção de conjunto, também se tem de apoiar sobre a noção de distância. Por exemplo, se a formiga tivesse acesso a um Mapa-Múndi, nunca conseguiria orientar-se no seu planeta sem ter conhecimento da escala. A noção de distância é dada, em geometria, pela métrica. É sabido de álgebra linear que um produto interno permite quantificar ângulos e normas de vectores. Ao atribuir um produto interno a cada ponto x de uma superfície M , operando no espaço tangente, $\langle \cdot, \cdot \rangle_x : T_x M \times T_x M \mapsto \mathbb{R}$, determina-se a sua métrica.

Consideremos, por exemplo, o Mapa-Múndi do planeta Terra. Tendo sido a superfície terrestre reduzida e deformada para caber num pequeno rectângulo, o mapa vai apresentar uma métrica diferente da que consideramos

¹ Uma generalização do conceito de superfície é o de uma variedade, sendo uma variedade M de dimensão n um conjunto em que a vizinhança de cada ponto é homeomorfa a um aberto de $\mathbb{R}^n$.

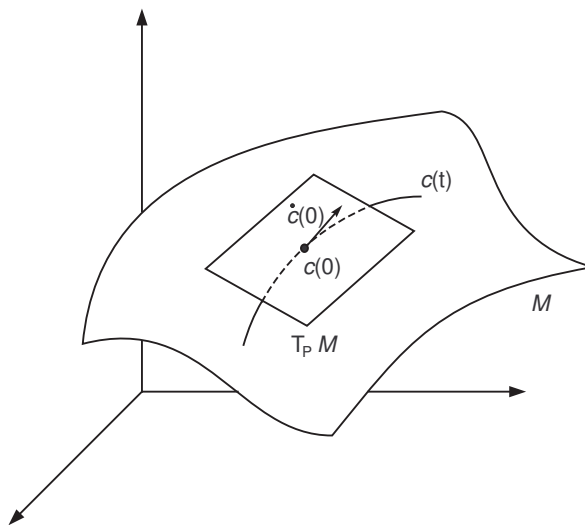


Figura 2: Vector tangente e espaço tangente

como sendo usual. No entanto, do ponto de vista geométrico tratam-se do mesmo objecto.

Uma métrica transmite a noção de normas de vectores tangentes a uma superfície e ângulos formados entre estes, sendo dados respectivamente por:

$$(1) \quad \|v\| = \sqrt{\langle v, v \rangle}$$

$$(2) \quad \alpha = \arccos \left(\frac{\langle v, w \rangle}{\|v\| \cdot \|w\|} \right)$$

O comprimento de uma curva γ sobre a superfície M , $\gamma : [0, 1] \rightarrow M$, é então obtido através de integração da norma do vector velocidade desta:

$$(3) \quad l = \int_0^1 \|\dot{\gamma}(t)\| dt$$

Dados dois pontos numa superfície, x e y , pode-se definir a distância entre ambos como o ínfimo do conjunto dos comprimentos das curvas que unem ambos os pontos. Associado a este conceito vem a ideia de geodésica, que corresponde (localmente) ao caminho mais curto entre pontos no espaço. Trata-se da generalização da linha recta para outras superfícies.

1.2 Estruturas Geométricas

Uma isometria entre dois espaços métricos M e N é uma transformação $f : M \rightarrow N$ que preserva distâncias, isto é:

$$(4) \quad d_M(x, y) = d_N(f(x), f(y))$$

Este conceito torna-se útil pois permite comparar duas superfícies. Quando tal isometria existe, as superfícies têm a mesma forma (são isométricas). O conjunto das isometrias de um espaço métrico nele mesmo designa-se por grupo de isometrias.

Alguns exemplos simples de grupos de isometrias aplicados ao mundo real (espaço euclídeo tridimensional munido da métrica usual) são a translação e a rotação, as quais obviamente não alteram distâncias.

O conceito de grupo de isometria é extremamente importante em geometria, sendo explorado ao longo deste texto. Na generalidade, uma métrica arbitrária numa superfície não apresenta isometrias para além da identidade. Interessa estudar quais as superfícies e as métricas que resultam num grupo de isometrias suficientemente rico.

DEFINIÇÃO. Diz-se que uma métrica numa superfície M é localmente homogénea se dados x e y em M , existem vizinhanças U e V de x e y e uma isometria que transforma U em V e x em y .

Uma métrica localmente homogénea adapta-se à superfície por forma a que esta tenha o mesmo aspecto vista de qualquer ponto. Qualquer raciocínio geométrico apoiado sobre a métrica localmente homogénea aplica-se igualmente a todos os pontos ou vizinhanças.

Diz-se que uma superfície M admite uma estrutura geométrica se M admitir uma métrica localmente homogénea e completa².

Pretende-se neste texto responder à questão de quais superfícies compactas e orientáveis³ admitem uma estrutura geométrica. Para tal recorre-se a noções de acções de grupos sobre superfícies.

1.3 Álgebra Geométrica

O conjunto das isometrias de uma superfície munida de uma métrica é um grupo algébrico, uma vez que a inversa de uma isometria também é uma

² Diz-se que uma métrica é completa se as suas geodésicas não forem limitadas.

³ Uma superfície diz-se orientável se é possível distinguir dois lados nela, ou equivalentemente, se o vector normal quando trasladado ao longo da superfície nunca troca o sentido. Ilustram-se exemplos de superfícies não orientáveis nas figuras 7 e 8.

isometria, assim como a composição de isometrias. É costume representar o grupo de isometrias da superfície M por $\text{Iso}(M)$.

Por forma a simplificar a notação, identificam-se subgrupos de $\text{Iso}(M)$ com grupos algébricos bem conhecidos, dizendo-se que um grupo G actua numa superfície se existir um homomorfismo $\varphi : G \longrightarrow \text{Iso}(M)$. Considere-se como exemplo a esfera munida da métrica usual e as aplicações identidade $x \mapsto x$ e antípoda $x \mapsto -x$, ambas isometrias. Neste caso diz-se que o grupo $\mathbb{Z}_2$, identificando 0 como a aplicação identidade, 1 como a aplicação antípoda e a soma (módulo 2) como a composição de aplicações, actua sobre a esfera.

Dados uma superfície M e um grupo G que nela actua, define-se o estabilizador de um ponto x da superfície como sendo o conjunto de elementos do grupo que preservam o ponto, isto é, tais que $g(x) = x$. Nota-se que o estabilizador é um subgrupo de G .

Um subgrupo de isometrias G permite identificar pares de pontos de uma superfície, quando existe uma isometria em G que transforma um no outro. Diz-se neste caso que os pontos são equivalentes, sendo:

$$(5) \quad x \sim y \Leftrightarrow \exists g \in G : y = gx$$

O conjunto dos pontos de equivalência de um dado ponto x designa-se por classe de equivalência (ou órbita) de x , representando-se por $[x]$. Define-se o quociente de uma superfície M por um grupo G como o conjunto de todas as classes de equivalência, representando-se por M/G .

Para o exemplo anterior, da superfície esférica actuada pelo grupo $\mathbb{Z}_2$, dois pontos extremos são equivalentes pois a aplicação antípoda transforma um no outro. O espaço quociente não é mais que o conjunto das classes de equivalência obtidas, correspondendo a “metade” da superfície esférica.

Nem sempre o quociente de superfícies por um grupo de isometrias produz um conjunto com boas propriedades, podendo existir casos patológicos. Para assegurar que tal não ocorre, vamos concentrar-nos apenas em grupos de isometrias propriamente descontínuos. Diz-se que uma acção de um grupo G numa superfície M é propriamente descontínua se para qualquer subconjunto compacto C de M se verifica que:

$$(6) \quad \{g \in G : gC \cap C \neq \emptyset\}$$

é um conjunto finito. Em particular, se G actua propriamente descontínuamente então os estabilizadores de qualquer ponto de M são finitos. Se todos estes estabilizadores são triviais, diz-se que G actua livremente.

Para qualquer grupo G actuando num espaço M , define-se a *região fundamental para G* como sendo um conjunto fechado P de M tal que:

$$(7) \quad \bigcup_{g \in G} gP = M$$

isto é, na região fundamental está presente pelo menos um ponto de cada classe de equivalência de M , e:

$$(8) \quad P \cap gP = \emptyset$$

para todos os elementos não triviais g em G . Por outras palavras, a região fundamental não apresenta mais do que um elemento da mesma classe de equivalência. É de notar que não existe uma escolha única para a região fundamental.

De seguida vão ser estudados cuidadosamente os subgrupos de isometrias de três superfícies: o plano Euclideano, a superfície esférica e o plano hiperbólico. Como se irá verificar posteriormente, esta análise permite tirar conclusões sobre as restantes métricas em outras superfícies, recorrendo aos conceitos descritos neste capítulo.

2 O Plano Euclideano E^2

O plano Euclideano E^2 consiste no espaço bidimensional $\mathbb{R}^2$ munido da métrica usual:

$$(9) \quad ds^2 = dx^2 + dy^2$$

2.1 Grupo das Isometrias

Vamos começar por analisar o grupo das isometrias de E^2 , também designado por grupo das transformações de Galileu, representado por $\text{Iso}(E^2)$. Uma isometria α de E^2 pode ser escrita na forma:

$$(10) \quad \alpha(x) = Ax + b, A \in O(2), b \in E^2$$

em que $O(2)$ é o grupo das matrizes ortogonais 2×2 . As isometrias do plano Euclideano correspondem a uma translação, rotação, reflexão ou

reflexão com deslizamento segundo uma recta l (que consiste na composição de uma reflexão sobre uma recta l com uma translação na direcção de l). De seguida ilustram-se cada uma destas acções.

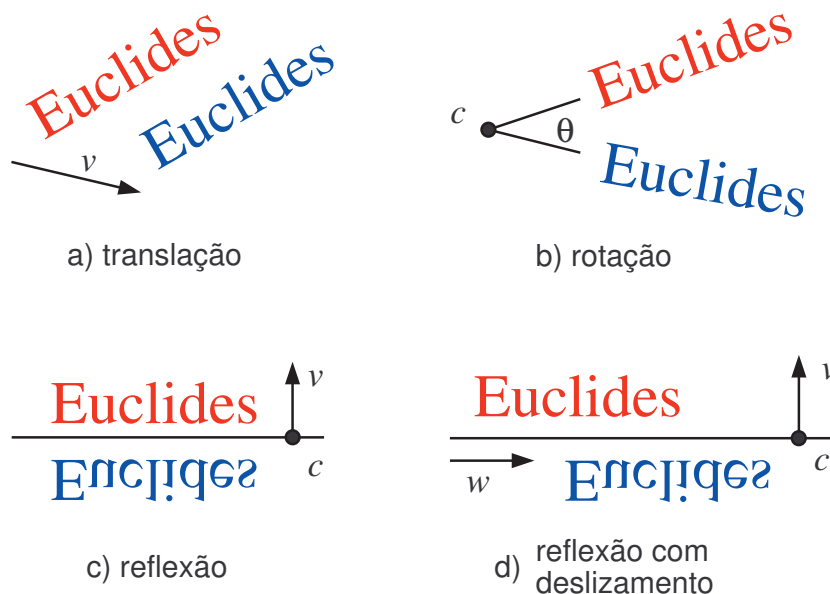


Figura 3: Os quatro tipos de isometrias de E^2

Verifica-se que a aplicação natural $\text{Iso}(E^2) \rightarrow O(2)$ que leva uma isometria α para a correspondente matriz A que a representa é um homomorfismo sobrejectivo, cujo núcleo consiste no subgrupo formado pelas translações. Esta informação pode ser compactada numa sequência exacta. Uma sequência exacta é tal que, para um dado elemento, a imagem do seu antecessor pertence--lhe ao núcleo.

$$(11) \quad 0 \longrightarrow \mathbb{R}^2 \longrightarrow \text{Iso}(E^2) \longrightarrow O(2) \longrightarrow 1$$

Uma propriedade do grupo das isometrias de E^2 é que é gerado por reflexões. Considere-se o produto de duas reflexões α e β segundo rectas l e m respectivamente. Se l e m forem paralelas e distanciarem-se de d , então $\alpha\beta$ é uma translação numa direcção perpendicular às rectas anteriores de distância $2d$. Caso contrário, se l e m se intersectarem num ponto x segundo um ângulo θ então $\alpha\beta$ é uma rotação segundo um ângulo 2θ em torno do ponto x . Segue-se assim que qualquer rotação ou translação é um produto de duas reflexões. As reflexões com deslizamento podem ser obtidas através de composição das isometrias anteriores, ficando provado que qualquer isometria de E^2 pode ser gerada com composições de reflexões.

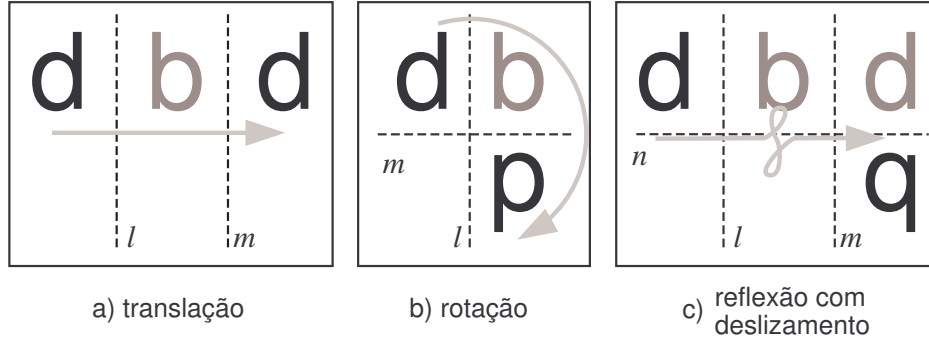


Figura 4: As isometrias de E^2 podem obter-se através de produtos de reflexões

2.2 Grupos Discretos Que Actuam Livremente

Os únicos exemplos de subgrupos G de isometrias de E^2 cuja acção é livre e propriamente descontínua são os gerados por uma translação, por uma reflexão com deslizamento, por duas translações ou por duas reflexões com deslizamento em diferentes direcções. Qualquer rotação ou reflexão do plano Euclideano tem um ponto fixo, pelo que se G for um grupo que actua livre e discretamente em E^2 só poderá conter translações ou reflexões com deslizamento. Assim, o subgrupo G_1 de G que preserva orientações⁴ consiste apenas em translações, pelo que é um subgrupo discreto de $\mathbb{R}^2$, o grupo de todas as translações de $\mathbb{R}^2$. Este subgrupo só pode então ser isomorfo a $\mathbb{Z}$ ou $\mathbb{Z} \oplus \mathbb{Z}$, e G é gerado no máximo por duas translações ou reflexões com deslizamento. Nestas situações, o quociente de E^2 por G é um cilindro aberto, uma banda de Moebius aberta, um torus ou uma garrafa de Klein, como se ilustra nas figuras 5, 6, 7 e 8. Note-se que a banda de Moebius e a garrafa de Klein não são superfícies orientáveis.

O quociente de E^2 por G herda uma métrica natural da métrica de E^2 , tal que a projecção $E^2 \rightarrow E^2/G$ é localmente uma isometria. Isto ocorre pois todos os pontos dentro da mesma classe de equivalência apresentam a mesma métrica (existem isometrias que transformam um ponto noutro dentro da mesma classe de equivalência).

⁴ Diz-se que um grupo de isometrias preserva a orientação quando não contém nenhum elemento que inverta o sinal do ângulo entre dois vectores. No caso do plano Euclideano, as isometrias que preservam a orientação são apenas as translações e as rotações. O produto de quaisquer duas isometrias que não preservam a orientação é uma isometria que preserva a orientação.

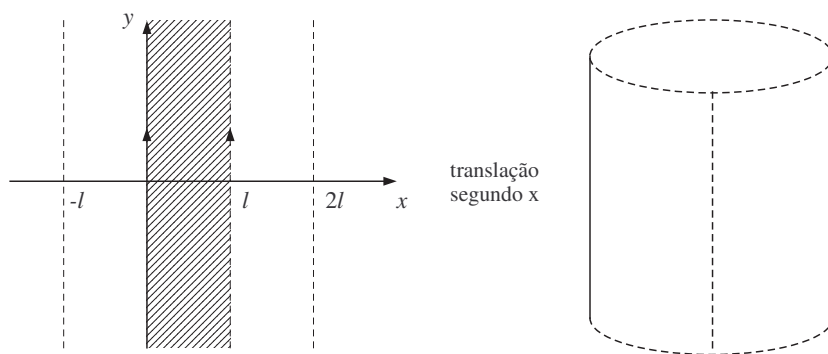


Figura 5: Exemplo do quociente de E^2 pelo grupo gerado por uma translação horizontal

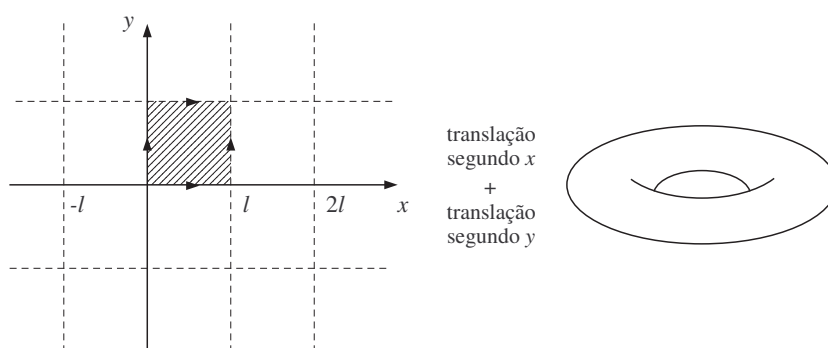


Figura 6: Exemplo do quociente de E^2 pelo grupo gerado por duas translações em diferentes direcções

2.3 Grupos Discretos Que Não Actuam Livrementemente

Se G é um grupo discreto de isometrias de E^2 que não actua livremente, então G tem de conter um elemento de ordem finita, que é uma reflexão ou uma rotação. Antes de considerar a situação geral, vamos analisar os casos em que G é um grupo cíclico⁵ ou um grupo diédrico⁶.

Seja G o grupo cíclico gerado por uma isometria α que é uma rotação de um ângulo $2\pi/n$ em torno de um ponto x em E^2 (ver figura 9). O quociente E^2/G representa-se de seguida, encontrando-se cada órbita representada por um ponto da região W , excepto os pontos de l_1 que têm um ponto da mesma classe de equivalência em l_2 .

O espaço quociente, E^2/G , obtém-se colando l_1 a l_2 , resultando um cone

⁵ Um grupo cíclico é um grupo que pode ser gerado por apenas um elemento.

⁶ O grupo diédrico de ordem $2n$, representado por D_n , é o conjunto das isometrias que preservam um polígono regular de n lados, e é gerado por uma rotação e uma reflexão.

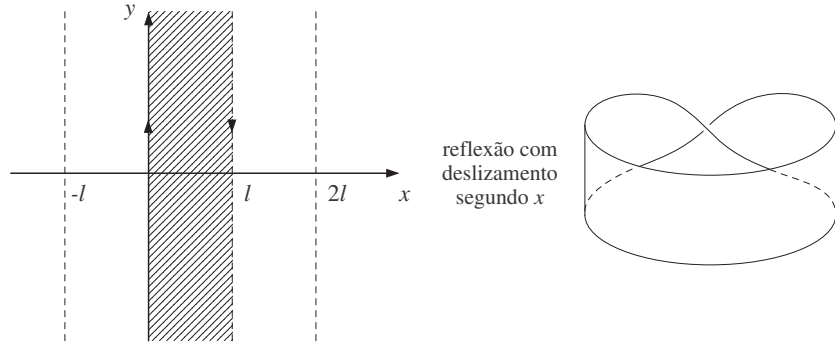


Figura 7: Exemplo do quociente de E^2 pelo grupo gerado por uma reflexão com deslizamento em torno do eixo vertical

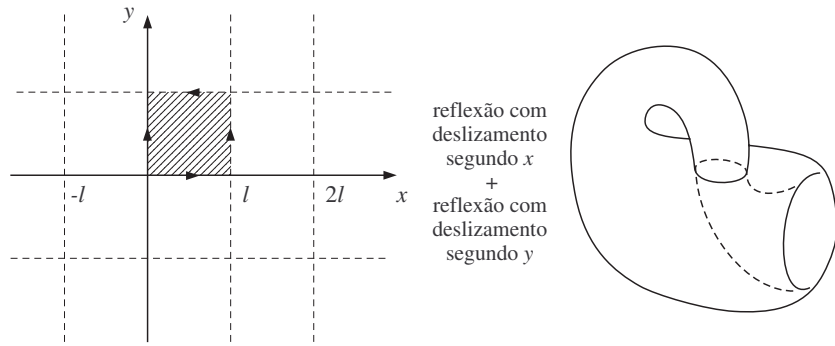
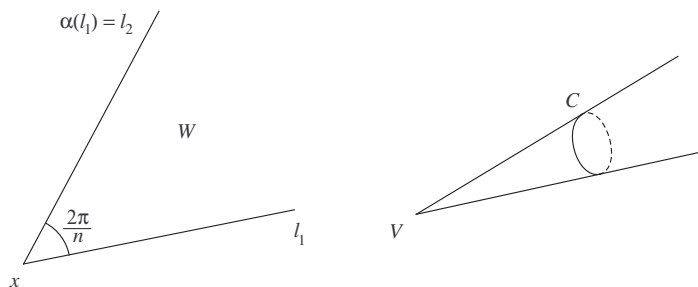


Figura 8: Exemplo do quociente de E^2 pelo grupo gerado por duas reflexões com deslizamento em diferentes direcções

C com ângulo $2\pi/n$ num vértice representado por V . Em termos topológicos, o cone C é isomorfo a $\mathbb{R}^2$, mas a métrica de C não é a métrica usual de $\mathbb{R}^2$. Qualquer ponto x em C tem pelo menos um representante em E^2 (bastando para isso escolher um ponto da classe de equivalência respectiva). Chama-se a este procedimento um levantamento de x . Um caminho l em C poderá ser também levantado para um caminho em E^2 , apesar deste procedimento não ser unicamente determinado pela escolha do levantamento de um ponto extremo. Este fenómeno ocorre sempre que l atravessa o vértice V . No entanto, todas as pré-imagens de l têm o mesmo comprimento, podendo-se assim definir uma métrica em C .

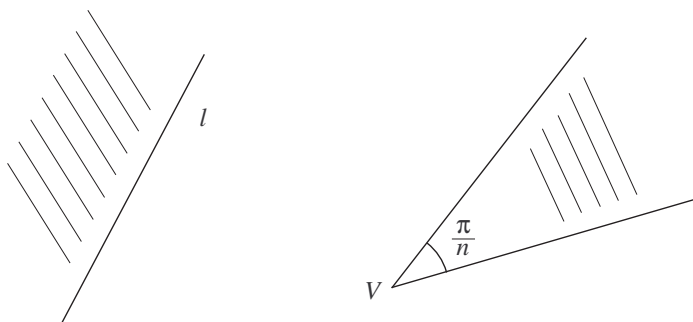
Claramente a projecção $E^2 - \{0\} \rightarrow C - V$ é localmente uma isometria. É também claro que C é isométrico a um cone circular embebido em E^3 . Nota-se que a afirmação de que o cone C tem um ângulo de $2\pi/n$ é equivalente a afirmar que uma circunferência de raio r no cone C com centro V tem


 Figura 9: Quociente de E^2 por um grupo cíclico

comprimento $2\pi/n$. É óbvio que este círculo é a projecção de um círculo de raio r e centro em x (ver figura 9).

Se G é um grupo cíclico de ordem dois gerado por uma reflexão sobre uma recta l , então E^2/G herda uma métrica, sendo isométrico a um semi-plano cuja recta fronteira é a imagem de l (ver figura 10 à esquerda).

Se G é um grupo diédrico de ordem $2n$ gerado por uma rotação de ordem n em torno de x e uma reflexão em torno de uma recta que atravessa esse ponto, então E^2/G volta a herdar uma métrica e é isométrico a um canto infinito com ângulo π/n (ver figura 10 à direita).


 Figura 10: Quociente de E^2 por um grupo cíclico de ordem dois e por um grupo diédrico de ordem $2n$

Em cada um dos casos anteriores, descreveu-se uma métrica natural na superfície quociente E^2/G . No caso em que G é cíclico e gerado por uma rotação, esta métrica tem uma singularidade no vértice V do cone C . Por outras palavras, a métrica restrita a $C - \{V\}$ é Riemanniana, mas a métrica

7 Uma circunferência de raio r e centro num ponto x de uma superfície M é o conjunto de pontos em M a uma distância r de x .

em C não o é. No caso em que G é gerado por uma reflexão, o quociente E^2/G apresenta uma recta fronteira formada por pontos singulares. Na situação em que G é diédrico, existem duas semi-rectas fronteiras formadas por pontos singulares que se encontram no ponto singular V . Estes três tipos de singularidades são designados por cantos, linhas reflectores e cantos reflectores.

Considerando um grupo discreto de isometrias de E^2 , a situação é semelhante à anterior. O quociente E^2/G é uma superfície que herda uma métrica natural da métrica de E^2 . Esta métrica pode apresentar pontos singulares que podem ser de um dos três tipos atrás mencionados. Um ponto de canto em E^2/G é um ponto com uma vizinhança isométrica a uma vizinhança do cone C , para algum ângulo $2\pi/n$. Se E^2/G tiver fronteira não vazia, essa fronteira consiste de linhas reflectoras com cantos reflectores isolados.

Como exemplos vamos considerar os grupos triangulares. Sejam p, q e r três inteiros tais que existe um triângulo Δ em E^2 com ângulos $\pi/p, \pi/q$ e π/r . Isto implica que se verifique a igualdade

$$(12) \quad \frac{1}{p} + \frac{1}{q} + \frac{1}{r} = 1$$

Facilmente se verifica que as únicas soluções para p, q e r são os tripletos $(3, 3, 3)$, $(2, 3, 6)$ e $(2, 4, 4)$. O grupo triangular $\Delta^*(p, q, r)$ é definido como o grupo de isometrias de E^2 gerado pelas reflexões L, M e N segundo os lados YZ, ZX e XY de Δ . Em cada caso é fácil verificar que as sucessivas imagens de Δ por acções de isometrias de $\Delta^*(p, q, r)$ preenchem E^2 . Esta situação é apresentada na figura seguinte para o caso $p = q = r = 3$.

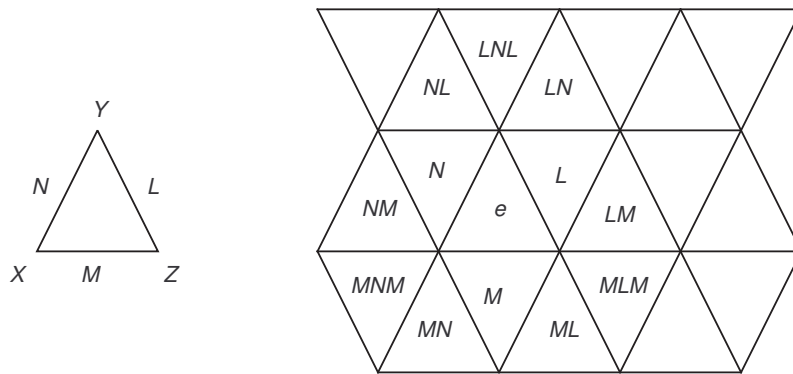


Figura 11: Pavimentação triangular de E^2

É fácil verificar que o estabilizador de Δ é trivial. Existe uma correspondência bijectiva entre os triângulos que pavimentam E^2 e os elementos

de $\Delta^*(p, q, r)$, definido mapeando $g \in \Delta^*(p, q, r)$ no triângulo $g\Delta$. Torna-se claro que $\Delta^*(p, q, r)$ é discreto e que $E^2/\Delta^*(p, q, r)$ é isométrico ao triângulo original Δ .

Podem-se efectuar as mesmas definições começando com um triângulo qualquer em E^2 , no entanto no geral não se obtém um grupo discreto. Se o triângulo tem um ângulo α num vértice X , então o grupo Δ^* gerado por reflexões segundo os lados do triângulo contém uma rotação segundo X de um ângulo 2α . Se Δ^* é discreto, esta rotação terá de ser de ordem finita, sendo assim α um múltiplo racional de π .

Seja $\Delta(p, q, r)$ o subgrupo que preserva orientações de $\Delta^*(p, q, r)$. O produto LM de duas reflexões de $\Delta^*(p, q, r)$ é uma rotação em torno de Z de um ângulo $2\pi/r$, assim como MN e NL são também rotações. Para simplificar a notação escreve-se $LM = z$, $MN = x$ e $NL = y$, de tal forma que x , y e z representam rotações no sentido dos ponteiros do relógio em torno de X , Y e Z respectivamente. É óbvio que estas rotações pertencem a $\Delta(p, q, r)$, e pode-se mostrar que o geram. Verifica-se a relação $xyz = 1$, pelo que $\Delta(p, q, r)$ é gerado por quaisquer duas das três rotações anteriores. Uma vez que $\Delta(p, q, r)$ contém rotações, será natural esperar que $E^2/\Delta(p, q, r)$ tenha pontos de canto. De facto, $E^2/\Delta(p, q, r)$ é topologicamente uma superfície esférica de dimensão 2 e tem três pontos de canto com ângulos $2\pi/p$, $2\pi/q$ e $2\pi/r$. A forma mais fácil de o verificar é observando que dois triângulos adjacentes na pavimentação da figura anterior formam uma região fundamental para $\Delta(p, q, r)$. Por exemplo, tomando $P = \Delta \cup L\Delta$, então as imagens de P por $\Delta(p, q, r)$ cobrem E^2 . Para cada g em $\Delta^*(p, q, r)$, ou g se encontra em $\Delta(p, q, r)$ ou no conjunto $\Delta(p, q, r) \cdot L$, pelo que qualquer triângulo $g\Delta$ está sobre uma imagem de P por um elemento de $\Delta(p, q, r)$. Segue que $E^2/\Delta(p, q, r)$ pode ser obtido através de P identificando as arestas apropriadamente.

Este tipo de construção de uma região fundamental de um subgrupo de índice finito de um dado grupo é muito útil. Se G é um grupo de isometrias de E^2 com região fundamental X e se G_1 é um subgrupo de G de índice n , então uma região fundamental X_1 para G_1 pode ser formada a partir de n imagens de X .

Existem poucos grupos de isometrias de E^2 e estes são conhecidos há vários anos. Tem-se a sequência exacta, já referida:

$$(13) \quad 0 \longrightarrow \mathbb{R}^2 \longrightarrow \text{Iso}(E^2) \longrightarrow O(2) \longrightarrow 1$$

Se G é um subgrupo de $\text{Iso}(E^2)$, seja T o subgrupo de todas as translações em G e H um subgrupo de $O(2)$. Então tem-se a sequência exacta:

$$(14) \quad 0 \longrightarrow T \longrightarrow G \longrightarrow H \longrightarrow 1$$

Se G é discreto, então T é trivial, $\mathbb{Z}$ ou $\mathbb{Z} \oplus \mathbb{Z}$. Se T é trivial, então G fixa um ponto de E^2 e é um grupo cíclico finito ou diédrico. Se T é cíclico, obtém-se os sete padrões de tira e se T for $\mathbb{Z} \oplus \mathbb{Z}$ obtém-se os dezassete grupos cristalográficos. Para uma descrição mais profunda sobre este assunto recomenda-se a consulta de [8].

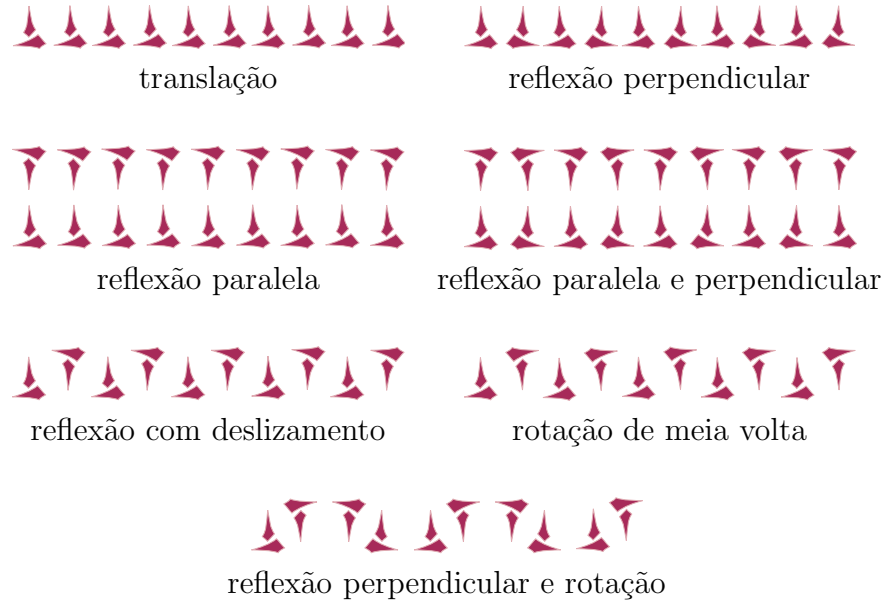


Figura 12: Os sete padrões em tira do plano

3 A Superfície Esférica S^2

Considera-se a superfície esférica S^2 munida com a métrica induzida pela sua inclusão em $\mathbb{R}^3$.

Os círculos máximos de S^2 são as geodésicas de S^2 , desempenhando o mesmo papel para a geometria de S^2 que as rectas para a geometria de E^2 . No entanto existem várias diferenças importantes que devem ser consideradas antes de discutir as isometrias de S^2 . As duas principais diferenças são que quaisquer duas geodésicas em S^2 cruzam-se, não existindo a noção de geodésicas paralelas, e que, enquanto que em E^2 por cada dois pontos distintos atravessa uma única geodésica, em S^2 existem infinitas geodésicas que atravessam pontos diametralmente opostos.

Talvez a diferença mais crucial é o efeito que a curvatura de S^2 tem nas propriedades das áreas de polígonos. Considerando o exemplo de triângulos em S^2 (cujas arestas coincidem com geodésicas), os seus ângulos determinam a área completamente (o que não ocorria em E^2 , onde era possível ter uma área arbitrária). Se um triângulo em S^2 tiver ângulos α , β e γ , então a sua área é $\alpha + \beta + \gamma - \pi$. Conclui-se daqui que a soma dos ângulos internos de um triângulo em S^2 encontra-se entre π e 5π , situações em que a sua área é, respectivamente, nula e máxima (toda a superfície esférica). Note-se que numa superfície esférica de raio R , esta área seria $R^2(\alpha + \beta + \gamma - \pi)$. A prova deste resultado é simples, bastando notar que a área entre duas geodésicas numa superfície esférica depende unicamente da área da superfície e do ângulo que ambas as geodésicas formam.

3.1 Grupo das Isometrias

O grupo das isometrias de S^2 é $O(3)$, o grupo das matrizes ortogonais 3×3 , que é o grupo de isometrias de E^3 que fixa a origem. Um elemento de $SO(3)$, o subgrupo de $O(3)$ que preserva orientações, é uma rotação de E^3 em torno de uma recta que atravessa a origem. A sua restrição a S^2 é também designada por rotação e fixa exactamente dois dos seus pontos, pelo que cada isometria de S^2 que preserva orientações é uma rotação. Quanto aos elementos α de $O(3)$ que não preservam orientação, estes são de dois tipos. Estes elementos terão de apresentar um valor próprio igual a -1 , pelo que associada ao correspondente vector próprio, existe uma recta l em E^3 que atravessa a origem que é invariante pela aplicação de α , mas que inverte o seu sentido. O plano ortogonal a esta recta e que cruza a origem, Π , é também invariante pela aplicação de α , pelo que a restrição de α a Π é uma rotação. Se esta rotação for trivial, então α é uma reflexão sobre o plano Π , caso contrário α fixa apenas a origem. Existem assim dois tipos de isometrias de S^2 que invertem a orientação - reflexões, que fixam um círculo máximo em S^2 , e isometrias sem pontos fixos. Uma vez que os círculos máximos de S^2 são as suas geodésicas, então uma reflexão em S^2 fixa uma única geodésica, da mesma forma que uma reflexão de E^2 fixa uma única recta. Tal como E^2 , o grupo de isometrias de S^2 é gerado por reflexões. Para provar este resultado, sejam α e β reflexões de S^2 em torno das geodésicas l e m respectivamente. Ao contrário da situação de E^2 , duas geodésicas nunca podem ser disjuntas, ou são a mesma ou intersectam-se em dois pontos x e y segundo um ângulo θ . Assim, se α e β são distintas, o produto $\alpha\beta$ é uma rotação de S^2 segundo um ângulo 2θ que fixa x e y . Segue-se que qualquer rotação de S^2 é um produto de duas reflexões. Todas as isometrias sem pontos fixos podem ser obtidas compondo rotações com reflexões apropriadamente.

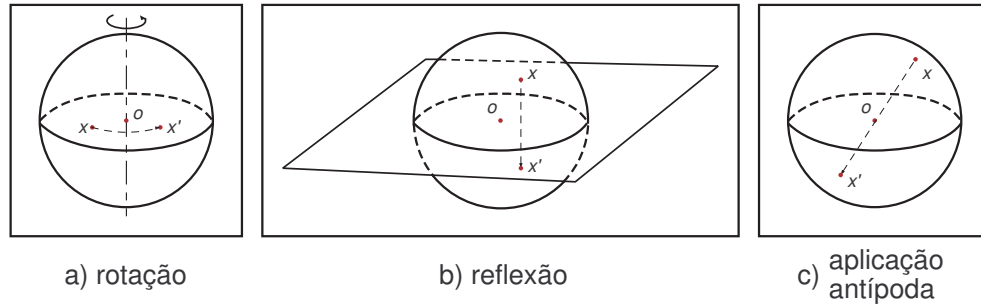


Figura 13: Três exemplos das isometrias da esfera

Vamos agora considerar grupos discretos de isometrias de S^2 .

3.2 Grupos Discretos Que Actuam Livremente

Existe um único subgrupo não trivial de $O(3)$ que actua livremente em S^2 , nomeadamente o grupo de ordem dois gerado pela transformação antipodal de S^2 . Qualquer isometria que preserva a orientação é uma rotação, pelo que apresenta pontos fixos. Desta forma, se G actua livremente em S^2 só poderá conter elementos não triviais que invertem a orientação, pelo que terá de ser um grupo de ordem dois, visto que a composição de duas reflexões origina um elemento de $SO(3)$. Verifica-se facilmente que a única isometria que actua em S^2 sem pontos fixos e com ordem dois é a transformação antipodal. O quociente de S^2 por este grupo é o plano projectivo real P^2 , e, como usual, P^2 herda uma métrica de S^2 , tal que a projecção $S^2 \rightarrow P^2$ é localmente uma isometria.

3.3 Grupos Discretos Que Não Actuam Livremente

Os grupos G discretos que não actuam livremente em S^2 comportam-se de forma análoga aos de E^2 . O quociente S^2/G herda uma métrica que apresenta singularidades de três tipos - cantos, linhas reflectores e cantos reflectores.

Os subgrupos finitos de $SO(3)$ encontram-se completamente classificados, sendo cíclicos, diédricos ou o grupo de simetria de um dos cinco sólidos regulares. O quociente de S^2 por um grupo cíclico é uma superfície isomorfa a S^2 apresentando dois pontos cónicos. O quociente de S^2 por um grupo diédrico ou por qualquer dos grupos de simetria dos sólidos regulares resulta numa superfície isomorfa à superfície esférica mas com três cantos. Estes dois últimos tipos de subgrupos são grupos triangulares e podem ser descritos de forma análoga aos grupos triangulares Euclidianos.

Se p , q e r são inteiros tais que:

$$(15) \quad \frac{1}{p} + \frac{1}{q} + \frac{1}{r} > 1$$

então existe um triângulo em S^2 com ângulos π/p , π/q e π/r . Pode mostrar-se facilmente que o tripleto (p, q, r) tem de ser da forma $(2, 2, n)$, $(2, 3, 3)$, $(2, 3, 4)$ ou $(2, 3, 5)$. Definindo novamente $\Delta^*(p, q, r)$ como sendo o grupo de isometrias de S^2 gerado por reflexões nos lados do triângulo e $\Delta(p, q, r)$ o correspondente subgrupo que preserva orientações, então $\Delta(2, 2, n)$ é o grupo diédrico de ordem $2n$ e $\Delta^*(2, 3, 3)$, $\Delta^*(2, 3, 4)$ e $\Delta^*(2, 3, 5)$ são respectivamente os grupos de simetria do tetraedro, cubo (ou octaedro) e dodecaedro (ou icosaedro).

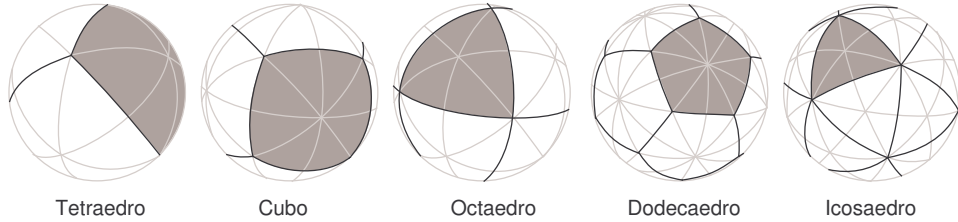


Figura 14: Regiões fundamentais dos vários grupos triangulares da esfera

Vamos analisar o isomorfismo entre $\Delta^*(2, 3, 3)$ e o grupo de simetria do tetraedro. Começando com um tetraedro regular com vértices em S^2 e unido estes pontos com geodésicas de S^2 (círculos máximos), obtém-se uma pavimentação da superfície esférica com quatro triângulos equiláteros. Cada um dos seus ângulos vale $2\pi/3$, uma vez que existem três em torno de cada vértice. Subdividindo cada um destes triângulos através da adição de vértices no centro do triângulo e nos pontos médios das arestas, obtém-se uma pavimentação da superfície esférica com 24 triângulos. Cada um apresenta ângulos de valor $\pi/2$, $\pi/2$ e $\pi/3$, e identifica-se com a região fundamental de $\Delta^*(2, 3, 3)$, compreendendo-se a relação com o grupo de simetria do tetraedro.

4 O Plano Hiperbólico H^2

O plano hiperbólico H^2 consiste no semiplano superior $\mathbb{R}_+^2 = \{(x, y) \in \mathbb{R}^2 : y > 0\}$, munido da métrica seguinte:

$$(16) \quad ds^2 = \frac{1}{y^2} (dx^2 + dy^2)$$

Apesar do semiplano superior ter uma fronteira, quando munido da métrica anterior o espaço é completo. Considere-se por exemplo uma formiga que inicia o seu movimento no ponto $(0, 1)$ e que caminha para Sul. Então a sua posição é parametrizada pela curva $\gamma(t) = (0, y(t))$ com $y(0) = 1$, e sendo a velocidade unitária vem que:

$$(17) \quad \|\dot{\gamma}(t)\| = 1 \Leftrightarrow \sqrt{\frac{\dot{y}^2(t)}{y^2(t)}} = 1 \Leftrightarrow \dot{y}(t) = -y(t) \Leftrightarrow y(t) = e^{-t}$$

pelo que a formiga nunca atinge a ordenada nula em tempo finito.

Ao contrário dos exemplos anteriores, o plano hiperbólico não pode ser visualizado no espaço tridimensional. No entanto é isométrico localmente a uma superfície com um ponto de sela em cada ponto. Um bom exemplo disso é a pseudo-esfera que se trata de uma superfície de revolução.

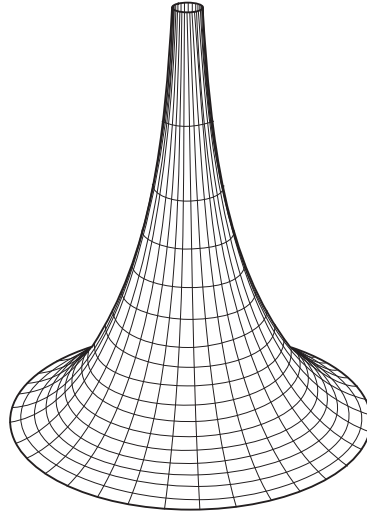


Figura 15: A pseudo-esfera

Esta superfície não é nem simplesmente conexa nem completa, ao contrário do plano hiperbólico, mas é localmente idêntica a este.

4.1 Grupo das Isometrias

Para analisar a geometria de H^2 convém conhecerem-se as suas geodésicas. Começa-se assim por provar que semi-rectas verticais, são geodésicas de H^2 . Este resultado parece ser intuitivo, uma vez que a métrica definida apenas

introduz uma deformação vertical em relação à norma usual (é independente da coordenada horizontal do ponto considerado).

Sejam P_0 e P_1 dois pontos de $\mathbb{R}_+^2$ com a mesma coordenada horizontal, e com coordenadas verticais y_0 e y_1 respectivamente. O comprimento do segmento de recta que une os dois pontos, γ , é então:

$$(18) \quad \int_{\gamma} ds = \int_{y_0}^{y_1} \frac{1}{y} dy = \log \left| \frac{y_1}{y_0} \right|$$

pois sobre γ tem-se $ds = \frac{1}{y} dy$. Se l é um outro caminho qualquer que une P_0 a P_1 , então o comprimento de l será maior que este. Parametrizando l com $t \in [t_0, t_1]$, então este caminho tem comprimento igual a:

$$(19) \quad \int_l ds = \int_{t_0}^{t_1} \frac{1}{y} \left(\left(\frac{dx}{dt} \right)^2 + \left(\frac{dy}{dt} \right)^2 \right)^{1/2} dt$$

Uma vez que a componente da velocidade segundo x é não-nula em determinado ponto, tem-se então:

$$(20) \quad \int_l ds = \int_{t_0}^{t_1} \frac{1}{y} \left(\left(\frac{dx}{dt} \right)^2 + \left(\frac{dy}{dt} \right)^2 \right)^{1/2} dt > \int_{t_0}^{t_1} \frac{1}{y} \left| \frac{dy}{dt} \right| dt \geq \int_{\gamma} ds$$

provando-se desta forma que as semirectas verticais de $\mathbb{R}_+^2$ são geodésicas de $\mathbb{H}^2$. Conclui-se também que se os pontos P_0 e P_1 estão posicionados na mesma semi-recta vertical, então essa é a única geodésica que os atravessa. Nota-se novamente que, apesar de estas curvas parecerem limitadas inferiormente (em termos Euclidianos) pelo eixo dos xx , não o são em termos da métrica do plano hiperbólico, o que se verifica facilmente da equação 18 ao fazer y_0 arbitrariamente pequeno e ao obter comprimentos arbitrariamente elevados.

Por forma a determinar as restantes geodésicas do plano hiperbólico, analisam-se primeiro as suas isometrias, pois a translação de uma geodésica por uma isometria origina uma nova geodésica. Uma isometria óbvia é a reflexão de $\mathbb{R}_+^2$ segundo uma recta vertical. O produto de duas destas corresponde a uma translação horizontal e é também uma isometria. Uma isometria menos evidente é a inversão em torno de uma circunferência de raio r centrada num ponto a do eixo dos xx ⁸. A inversão de uma semirecta é uma semicircunferência com centro no eixo dos xx , tal como se representa na figura seguinte.

8 Estas transformações costumam designar-se por Transformações de Möbius

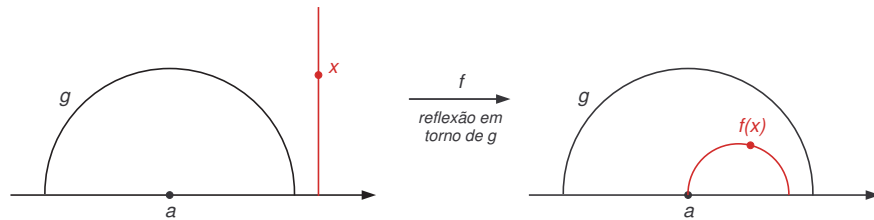


Figura 16: Exemplo de uma inversão em torno de uma circunferência

Utilizando números complexos para representar um elemento de $\mathbb{R}_+^2$, então a transformação anterior é definida por $w = f(z)$, onde:

$$(21) \quad w - a = \frac{r^2}{\bar{z} - a}$$

Deduz-se daqui que todas as semicircunferências com centro no eixo dos xx são geodésicas de H^2 . Vamos provar que estas e as semirectas verticais são as únicas. Considere-se um par de pontos em H^2 . Certamente que por este par de pontos atravessa uma única geodésica do tipo anteriormente descrito, m . Suponha-se que existia também uma outra geodésica a atravessar estes pontos, l . Por aplicação de uma isometria apropriada, pode-se considerar que estes pontos se encontram alinhados verticalmente, sendo m uma semi-recta vertical. Como já se provou anteriormente, esta geodésica é única, pelo que l nunca poderia sê-lo também.

Tal como E^2 , em H^2 existe uma única geodésica a unir dois pontos, e duas geodésicas distintas cruzam-se no máximo num ponto. Uma grande diferença é que dada uma geodésica l em H^2 e um ponto P fora dessa curva, existe uma infinidade de geodésicas que atravessam P e que não cruzam l ⁹.

Quanto aos triângulos em H^2 , cujas arestas são formadas por geodésicas, a situação é semelhante à ocorrida para S^2 , sendo agora a área de um triângulo igual a $\pi - \alpha - \beta - \gamma$. Conclui-se daqui que a soma dos ângulos internos de um triângulo em H^2 é inferior a π . Um triângulo diz-se ideal se todos os seus ângulos internos forem nulos, posicionando-se os seus vértices no infinito. Assim um triângulo ideal tem área π , enquanto que os restantes têm área ligeiramente inferior a esse valor.

As isometrias anteriormente descritas preservam geodésicas, e trocam as regiões que estas separam, designando-se por isso de reflexões. Prova-se que

⁹ Euclides derivou a maior parte da geometria planar a partir de 5 postulados. O 5º postulado indica que através de um ponto exterior a uma recta atravessa uma única recta. A geometria Hiperbólica coincide nos restantes quatro postulados com a Euclideana, diferindo ambas apenas neste último.

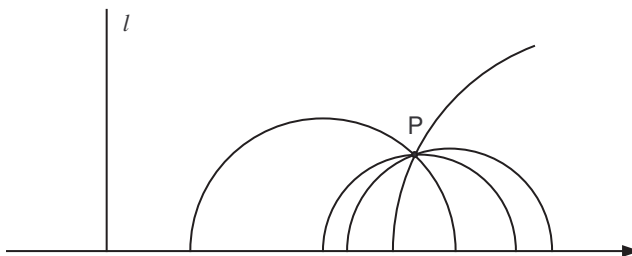


Figura 17: A noção de rectas paralelas não se aplica ao plano hiperbólico.

o grupo de isometrias de H^2 é gerado apenas por reflexões, tal como se viu para E^2 e S^2 .

4.2 Disco de Poincaré

Uma representação equivalente para o plano hiperbólico pode ser obtida mapeando o semiplano superior no interior do círculo de raio unitário através da transformação

$$(22) \quad z \rightarrow \frac{z - i}{z + i}$$

A métrica neste disco induzida pela métrica de H^2 é

$$(23) \quad ds^2 = \frac{4}{1 - (x^2 + y^2)^2} (dx^2 + dy^2)$$

É costume designar esta representação por *disco de Poincaré*¹⁰. As geodésicas nesta representação são arcos de circunferência que encontram a fronteira do círculo ortogonalmente ou os seus diâmetros. Nota-se que as geodésicas continuam a ser infinitamente compridas, apesar de serem limitadas em termos Euclidianos, interpretando-se a fronteira do círculo como o infinito. Na figura 18 apresentam-se alguns exemplos de geodésicas no disco de Poincaré.

Duas geodésicas que se encontram no infinito são chamadas de paralelas, por analogia com a situação Euclideana. No entanto isto é apenas uma

¹⁰ Para além das duas aqui analisadas, existem ainda as representações de Klein-Beltrami e de Minkowski.

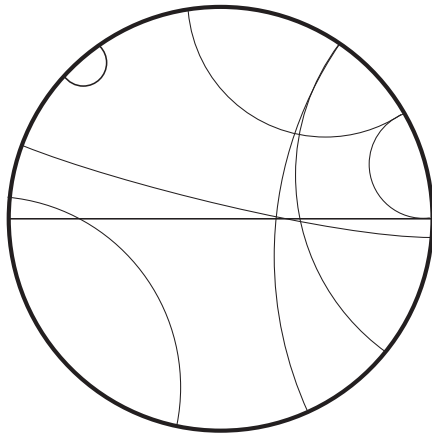


Figura 18: Representação equivalente de H^2 e algumas das suas geodésicas

analogia, pois ambas divergem exponencialmente à medida que nos afastamos do ponto comum no infinito. Duas geodésicas que nunca se cruzam, nem mesmo no infinito, são chamadas de ultra-paralelas, representando-se na figura 19. Nota-se que por duas geodésicas ultra-paralelas atravessa uma única geodésica perpendicular a ambas.

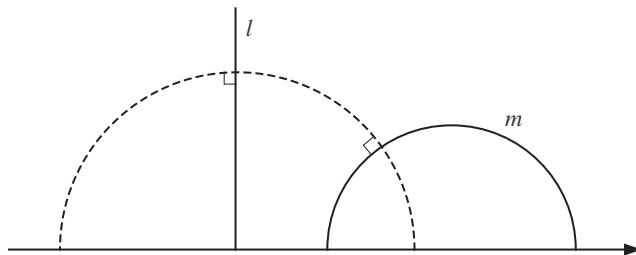


Figura 19: Geodésicas ultra-paralelas no semi-plano superior

No modelo do disco de Poincaré facilmente se compreendem as isometrias de H^2 . Uma isometria α que preserva orientações é obtida através do produto de duas reflexões em torno das geodésicas g_1 e g_2 , existindo assim três situações dependendo das suas posições relativas, como se ilustra na figura 20. A tracejado representam-se as trajectórias dos pontos quando aplicada a isometria respectiva. É de notar que a direcção da trajectória depende da ordem da composição, tendo-se neste caso considerado $g_2 \cdot g_1$. Para além disso apresentam-se três pontos tais que $x_1 = g_1(x)$ e $x_2 = g_2(x_1)$.

As isometrias elípticas surgem quando g_1 e g_2 se intersectam num ponto x de H^2 . A isometria α preserva esse ponto e chama-se uma rotação em

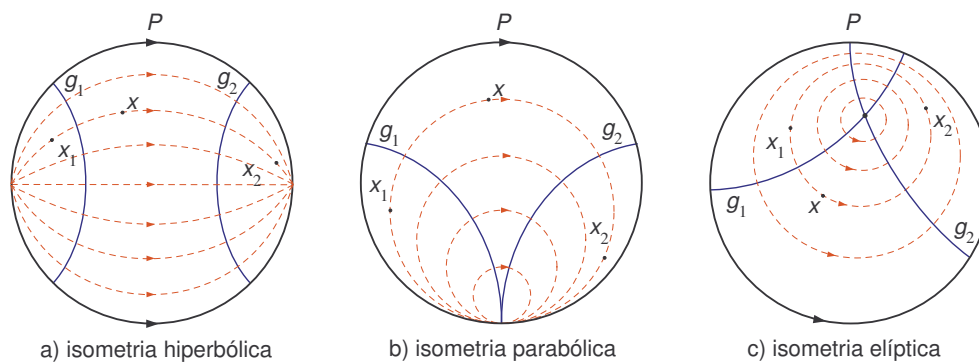


Figura 20: Exemplos de isometrias no disco de Poincaré

torno de x . O ângulo de rotação é o dobro do ângulo formado entre ambas as geodésicas.

Quando g_1 e g_2 se encontram no infinito (são paralelas), a isometria diz-se parabólica, e corresponde no semi-plano superior à composição de duas reflexões em torno de semi-rectas verticais, ou por outras palavras a uma translação horizontal.

Se ambas as geodésicas nunca se cruzam, sendo ultra-paralelas, α deixa invariante a sua perpendicular comum, transladando-a de uma distância $2d$, em que d é a distância da perpendicular que une g_1 a g_2 . Esta isometria fixa exactamente dois pontos no infinito, sendo então a única geodésica preservada a perpendicular.

As isometrias que não preservam orientações induzem um homeomorfismo em S^1 , a fronteira do círculo, o qual inverte orientações. Desta forma a isometria fixa dois pontos no infinito, e logo preserva uma geodésica l . Compondo α com uma reflexão em torno de l , obtém-se uma isometria que preserva a orientação e, da descrição anterior, só poderá ser ou a isometria trivial ou uma isometria hiperbólica. No primeiro caso α consiste numa reflexão e no segundo α é uma reflexão com deslizamento. Para uma análise mais aprofundada sobre as representações equivalentes do plano hiperbólico e as suas isometrias aconselha-se a consulta de [7].

4.3 Grupos Discretos

Para terminar a análise do plano hiperbólico consideremos os grupos discretos de isometrias. Se a acção de G sobre H^2 é livre, então o quociente H^2/G herda a métrica usual do plano hiperbólico. Se G não actuar livremente, então H^2/G herda também uma métrica, mas existindo pontos singulares de três tipos.

Tomemos como exemplo os grupos triangulares. Sejam p , q e r inteiros tais que

$$(24) \quad \frac{1}{p} + \frac{1}{q} + \frac{1}{r} < 1$$

Então, existe um triângulo Δ com ângulos π/p , π/q e π/r em H^2 . Como anteriormente, define-se $\Delta^*(p, q, r)$ como o grupo de isometrias de H^2 gerado por reflexões em torno dos lados de Δ . Prova-se que translações de Δ por acção de elementos de $\Delta^*(p, q, r)$ cobrem todo o H^2 , e que o estabilizador de Δ é o trivial.

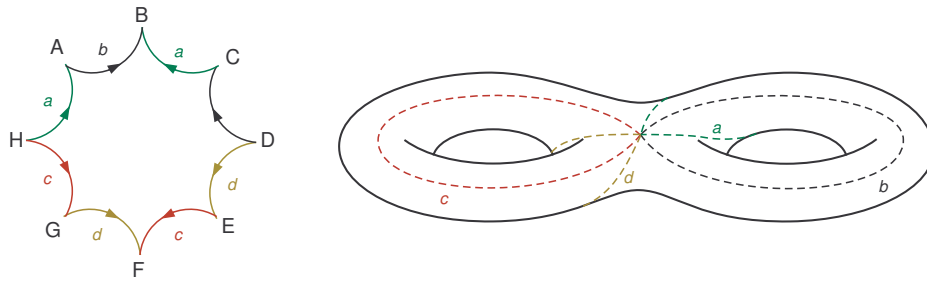


Figura 21: Obtenção da superfície de género 2

Podem-se definir grupos que actuam livremente em H^2 através de um método semelhante, começando com uma região fundamental. Por exemplo, seja X um octógono regular em H^2 com ângulos idênticos a $\pi/4$, como se ilustra na figura 21. Sejam α , β , γ e δ as isometrias de H^2 tais que $\alpha(AB) = DC$, $\beta(BC) = ED$, $\gamma(EF) = HG$ e $\delta(FG) = AH$, e que em todos os casos o interior de X é disjunto das suas translações. O grupo de isometrias Γ gerado por α , β , γ e δ é discreto e a sua região fundamental é X , sendo o quociente H^2/Γ obtido identificando pares de arestas de X . Nota-se que os oito vértices de X são identificados num único ponto, não se tratando de um ponto cónico uma vez que o ângulo total é 2π . O quociente H^2/Γ é isomorfo a uma superfície de género 2, T_2 , tal como se representa à direita.

Prova-se que existem também grupos de isometrias que originam quocientes isomorfos as superfícies com mais de 2 “buracos”. Desta forma qualquer superfície do tipo das listadas na figura 1 apresenta uma estrutura geométrica¹¹.

11 Relembra-se que uma superfície admite uma estrutura geométrica se existe uma métrica que é localmente homogênea, i.e., que apresenta um grupo de isometrias transitivo.

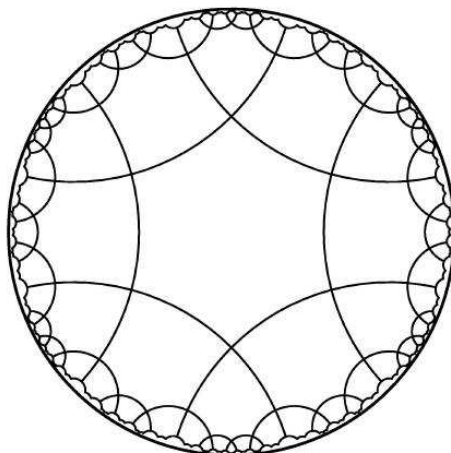


Figura 22: Exemplo de uma pavimentação de H^2 por hexágonos

Considerando regiões fundamentais formadas por polígonos regulares, obtém-se pavimentações do plano hiperbólico, tal como se representa como exemplo na figura 22.

É de salientar que algumas das pinturas de M. C. Escher exploram estas pavimentações do plano hiperbólico; nas figuras seguintes algumas destas estão representadas, nomeadamente o *Circle Limit III* e *Circle Limit IV*.



Figura 23: *Circle Limit III* e *Circle Limit IV* de M. C. Escher

5 Conclusões

Neste artigo foram descritas as estruturas geométricas para superfícies compactas (variedades de dimensão 2), resultados já conhecidos há mais de um século. No final da década de 70, William Thurston elaborou a conjectura de geometrização, propondo uma caracterização completa das estruturas geométricas para variedades de dimensão 3.

As variedades tridimensionais possuem uma propriedade de decomposição em dois níveis:

1. A decomposição prima, onde qualquer variedade-3 compacta é a soma conexa¹² de uma única colecção de variedades primas, e
2. A decomposição de Jaco-Shalen-Johannson, JSJ, para as variedades primas.

A conjectura de Thurston indica que após a desagregação de uma variedade-3 em somas conexas de variedades primas e a desaglutinação destas de acordo com a decomposição de JSJ, as componentes restantes admitem exactamente uma geometria modelada exactamente num dos seguintes espaços:

- Espaço Euclideano E^3
- Espaço hiperbólico H^3
- Esfera S^3
- $S^2 \times \mathbb{R}$
- $H^2 \times \mathbb{R}$
- O grupo de Lie $SL_2\mathbb{R}$
- O espaço Nil
- O espaço Sol

Seis destas geometrias são agora bem conhecidas (excepto a hiperbólica e a esférica) e correspondem aos chamados espaços fibrados de Seifert.

Se a conjectura de Thurston se encontrar correcta, o mesmo ocorrerá com a conjectura de Poincaré, a qual afirma que qualquer variedade tridimensional fechada e com grupo fundamental trivial (i.e. tal que qualquer caminho

12 A soma conexa é uma operação que consiste na aglutinação de duas variedades por remoção de uma bola em cada uma, e respectiva colagem da superfície fronteira.

fechado e simples na variedade pode ser deformado num ponto, continuamente) é homeomorfa a uma esfera tridimensional.

A medalha Fields foi atribuída a Thurston em 1982 parcialmente pelo sua prova da conjectura para variedades de Haken.

Muito progresso foi feito no estudo do espaço hiperbólico H^3 , verificando-se que algumas variedades-3 são modeladas neste espaço.

O estudo da da esfera S^3 foi mais lento, mas o suficiente para Richard Hamilton desenvolver o conceito de fluxo de Ricci. Em 1982, Hamilton provou que dada uma variedade-3 fechada com uma métrica de curvatura de Ricci positiva, o fluxo de Ricci iria suavizá-la, obtendo-se curvatura constante positiva, i.e. métrica esférica. Mais tarde desenvolveu uma forma de provar a conjectura da geometrização através do fluxo de Ricci.

Resultados de Grigori Perelman (em 2002 e 2003) parecem ter demonstrado a conjectura da geometrização, existindo consenso entre profissionais de que a prova está correcta. De acordo com o *Clay Mathematics Institute*, ele poderá ser eleito para o prémio de um milhão de dólares pela resolução da conjectura de Poincaré.

Em 2006, Zhu Xiping e Cao Huaidong, dois matemáticos chineses, publicaram os detalhes finais da prova da Conjectura de Poincaré. O trabalho foi publicado na edição de Junho do “Asian Journal of Mathematics”.

Agradecimentos

Gostaria de deixar os meus agradecimentos ao Professor Miguel Abreu por todo o tempo que me dispensou. Sem ele nunca seria possível a elaboração deste artigo nem a realização do seminário que lhe deu origem.

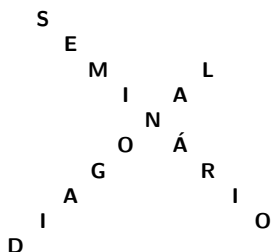
Deixo também uma palavra de apreço a todos os envolvidos na revisão deste documento, em especial ao Nuno Freitas e ao André Vasconcelos pelo apoio cedido durante a revisão do texto.

Referências

- [1] P. Scott, The geometries of 3-manifolds, *Bull. London Math. Soc.* **15** (1983), 401-487.
- [2] M. P. do Carmo, *Geometria Riemanniana*, IMPA, 1979.
- [3] M. P. do Carmo, *Differential Geometry of Curves and Surfaces*, Prentice-

Hall, 1976.

- [4] O. Forster, *Lectures on Riemann Surfaces*, Springer, 1981.
- [5] W. S. Massey, *Algebraic Topology: An Introduction*, Springer-Verlag (1967).
- [6] <http://members.tripod.com/vismath/dunham/index.html>
- [7] <http://www.geom.uiuc.edu/~crobles/hyperbolic/>
- [8] [http://en.wikipedia.org/wiki/Wallpaper\relax\\$\\@@underline{\hbox{}}\mathsurround](http://en.wikipedia.org/wiki/Wallpaper\relax$\\@@underline{\hbox{}}\mathsurround)



Será Possível Encontrar Topologia numa Fábrica?

Daniela Pontes
2º ano da LMAC
Instituto Superior Técnico
danielapontes@gmail.com

Palavras Chave

espaço de configurações, grafos, espaços topológicos, discretizações e tranças.

Resumo

Imaginemos uma fábrica e alguns robots que se deslocam por ela para realizarem uma determinada tarefa, contornando alguns obstáculos. Será que, sabendo o número de robots e qual o seu percurso, poderemos visualizar o seu espaço de configurações? Se sim, qual o aspecto desses espaços? Serão fáceis ou difíceis de visualizar? Haverá maneira de os simplificar? Até que ponto essa simplificação será válida? É no âmbito da topologia que serão dadas as respostas a estas e outras questões. Para concluir, far-se-á uma introdução à teoria das tranças e veremos como se ajusta a este problema.

1 Introdução

Sabe-se hoje que se podem encontrar objectos topológicos bastante interessantes num armazém ou numa fábrica automatizada.

Os exemplos de espaços topológicos construídos neste artigo surgiram simultaneamente de dois campos completamente distintos.

A. Abrams descobriu estes espaços depois de ter trabalhado com H. Landau, Z. Landau, J. Pommersheim e E. Zaslow sobre problemas relacionados com múltiplos percursos aleatórios em grafos.

R. Ghrist descobriu estes mesmos espaços enquanto trabalhava com D. Koditschek no Laboratório de Inteligência Artificial na Universidade de Michigan, o que mostra que também existem objectos topológicos no mundo físico.

Desde 1960 que se desenvolvem ideias sobre espaços de configurações para robots, tanto por matemáticos, como na área da computação e engenharia. Vários tipos de espaços de configurações surgem na topologia e na física tal como no estudo das tranças.

2 Motivação

Imagine-se uma fábrica ou um armazém onde existem robots, cuja função seja transportar objectos de um local para outro.

Obviamente, pretende-se encontrar um algoritmo que permita que estes robots se possam deslocar, enquanto desempenham a sua função, sem chocarem uns com os outros. Caso contrário, esses choques iriam dar prejuízo à fábrica, tanto na manutenção dos robots como na sua própria produção.

Esta ideia não está apenas presente nas fábricas. Pode-se também pensar num hipermercado e nos seus carrinhos de compras: “Por onde se pode deslocar sem se correr o risco de chocar com as outras pessoas?”.

Estes são apenas dois exemplos de locais onde se pode encontrar topologia, mas existem outros. No entanto, para a compreensão desta exposição, bastam estas duas situações.

Para responder à questão anteriormente colocada é necessário introduzir-se o conceito de espaço de configurações.

Ao longo de todo o artigo o leitor estará em contacto simultâneo com os conceitos teóricos necessários e com o exemplo concreto dos robots numa fábrica.

3 Espaço de Configurações

Numa linguagem corrente e mais intuitiva pode-se dizer que o espaço de configurações é composto pelas zonas onde certas partículas se podem encontrar. Neste contexto, são os locais por onde os robots se podem deslocar sem chocarem.

Para representar estes espaços usa-se a notação $C^n(\mathbb{R}^m)$, que define o espaço de configurações de n partículas que se deslocam em $\mathbb{R}^m$.

Para uma melhor compreensão do conceito seguem-se alguns exemplos.

3.1 $C^2(\mathbb{R})$, Espaço de Configurações de Duas Partículas que se Deslocam na Recta Real

Suponha-se que se tem duas partículas (x_1 e x_2) que se deslocam na recta real. Estas podem tomar qualquer posição dessa recta. No entanto não podem estar ao mesmo tempo na mesma posição. No contexto presente nesta exposição, suponha-se que dois robots se deslocam num mesmo corredor infinito da fábrica. Assim, as suas posições vão “desenhar” um plano em $\mathbb{R}^2$ (ver exemplos na Figura 1), excepto a recta $x_1 = x_2$, como se pode observar na Figura 2.

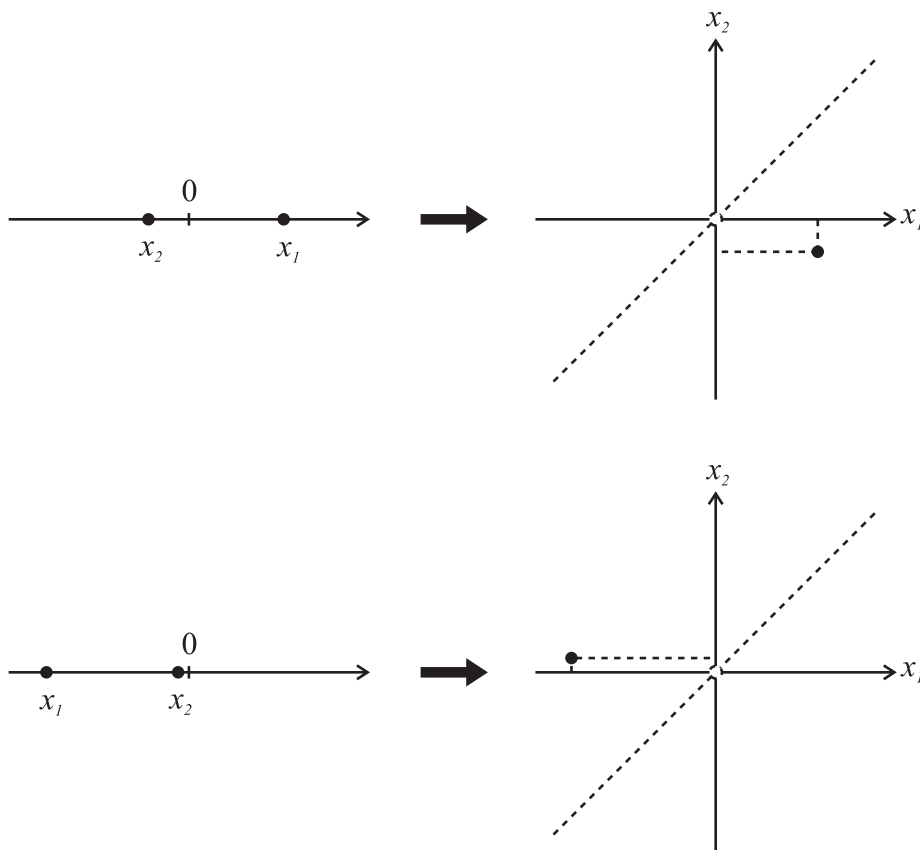


Figura 1: Exemplo de dois pontos do espaço de configurações (à direita) de duas partículas que se deslocam na recta real (à esquerda).

Pode-se então dizer que o espaço de configurações de duas partículas que se deslocam na recta real é da forma

$$C^2(\mathbb{R}) = \mathbb{R} \times \mathbb{R} \setminus \{(x_1, x_2) : x_1 = x_2\}$$

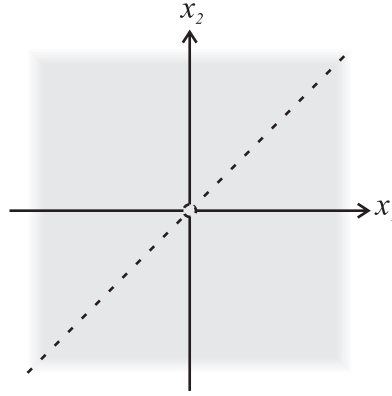


Figura 2: Espaço de configurações gerado pelo movimento de duas partículas na recta real.

3.2 $C^3(\mathbb{R})$, Espaço de Configurações de Três Partículas que se Deslocam na Recta Real

Esta situação é muito semelhante à anterior: três robots $(x_1, x_2$ e $x_3)$ que se deslocam no mesmo corredor infinito da fábrica, mas que não podem chocar entre si. Assim, as suas posições vão desenhar todo o espaço em $\mathbb{R}^3$, excepto os planos $x_1 = x_2$, $x_2 = x_3$ e $x_1 = x_3$, que se encontram na Figura 3.

Pode-se então dizer que o espaço de configurações de três partículas que se deslocam na recta real é da forma

$$C^3(\mathbb{R}) = \mathbb{R} \times \mathbb{R} \times \mathbb{R} \setminus \{(x_1, x_2, x_3) : x_1 = x_2 \vee x_2 = x_3 \vee x_1 = x_3\}$$

3.3 $C^2([a, b])$, Espaço de Configurações de Duas Partículas que se Deslocam no Intervalo $[a, b]$

Este caso é em tudo semelhante ao primeiro. A única diferença reside no facto de em vez de se ter a recta real, tem-se apenas um intervalo desta. Esta situação traduz-se na deslocação de dois robots num corredor da fábrica. O espaço obtido pelas suas posições é o que se encontra na Figura 4.

Pode-se então dizer que o espaço de configurações de duas partículas que se deslocam num intervalo da recta real é da forma

$$C^2([a, b]) = [a, b] \times [a, b] \setminus \{(x_1, x_2) : x_1 = x_2\}$$

Seguindo a mesma linha de raciocínio tem-se o espaço de configurações de duas partículas no plano $\mathbb{R}^2$, $C^2(\mathbb{R}^2)$.

Este é da forma

$$C^2(\mathbb{R}^2) = \mathbb{R}^2 \times \mathbb{R}^2 \setminus \{(x_1, y_1), (x_2, y_2) : x_1 = x_2 \wedge y_1 = y_2\}$$

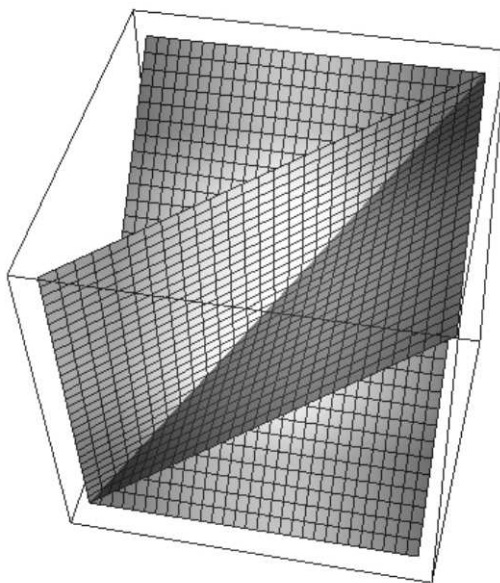


Figura 3: Visualização dos planos em $\mathbb{R}^3$ que não fazem parte do espaço de configurações de três partículas que se deslocam na recta real.

O espaço de configurações de três partículas no plano $\mathbb{R}^2$, $C^3(\mathbb{R}^2)$, é da forma

$$C^3(\mathbb{R}^2) = \mathbb{R}^2 \times \mathbb{R}^2 \times \mathbb{R}^2 \setminus \left\{ \begin{array}{l} (x_1, y_1), (x_2, y_2), (x_3, y_3) : \\ (x_1 = x_2 \wedge y_1 = y_2) \vee \\ (x_2 = x_3 \wedge y_2 = y_3) \vee \\ (x_1 = x_3 \wedge y_1 = y_3) \end{array} \right\}$$

No entanto, uma fábrica não é apenas constituída por um corredor como se viu nestes exemplos e muitas vezes possui obstáculos que impedem a livre circulação dos robots. Desta forma, deve-se fazer outro tipo de análise: parece mais correcto considerar a fábrica como um grafo.

4 Grafos

A teoria de grafos é bastante mais complexa do que os aspectos que vão ser considerados aqui. Não se pretende dar a conhecer em detalhe toda essa teoria, mas sim explicar e exemplificar alguns conceitos que serão úteis no decorrer desta exposição.

Um grafo pode ser considerado como uma rede de arestas e vértices (ver exemplo na Figura 5).

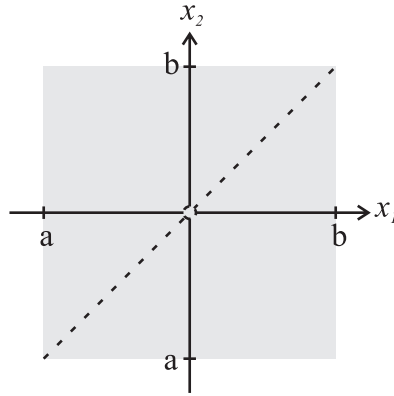
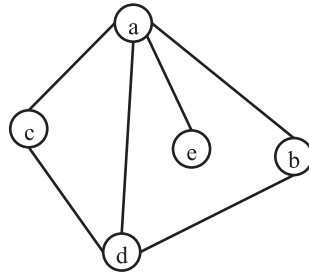


Figura 4: Espaço de configurações gerado pelo movimento de duas partículas num intervalo $[a, b]$ de $\mathbb{R}$.

Cada vértice tem uma determinada valência, ou seja, número de arestas que terminam nesse vértice.



Vértice	Valência
a	4
b	2
c	2
d	3
e	1

Figura 5: Exemplo de grafo com cinco vértices e seis arestas e respectiva valência de cada vértice.

Neste caso, pode-se considerar cada aresta do grafo como um corredor da fábrica e os vértices o encontro de corredores.

5 Espaços de Configurações de Partículas que se Movimentam em Grafos

A notação utilizada para indicar os espaços de configurações de grafos é semelhante à utilizada nos exemplos anteriores. Desta forma, $C^n(\Gamma)$ indica o espaço de configurações de n partículas, neste caso em particular robots, que se deslocam no grafo Γ .

A forma formal deste espaço segue a mesma estrutura dos indicados anteriormente, ou seja, este espaço poderá ser indicado da seguinte forma

$$C^n(\Gamma) = \underbrace{\Gamma \times \Gamma \times \dots \times \Gamma}_{n \text{ vezes}} \setminus \Delta$$

Assim, surgem algumas questões:

1. Qual o aspecto destes espaços?
2. O que é ou qual a forma do conjunto Δ ?
3. Como impedir que os robots colidam entre si?

De seguida estas questões serão respondidas com base em alguns exemplos.

Considere-se, em primeiro lugar, um grafo com três arestas e quatro vértices, G , no qual se deslocam duas partículas e o respectivo espaço de configurações. As partículas podem deslocar-se livremente sem que choquem, assim, quando estas se encontram em arestas distintas o espaço de configurações resultante terá o aspecto de um quadrilátero. Na Figura 6 encontra-se uma possível posição das partículas e o respectivo ponto no espaço de configurações. À medida que elas se movimentam o plano apresentado começará a ser preenchido, terminando como um quadrilátero.

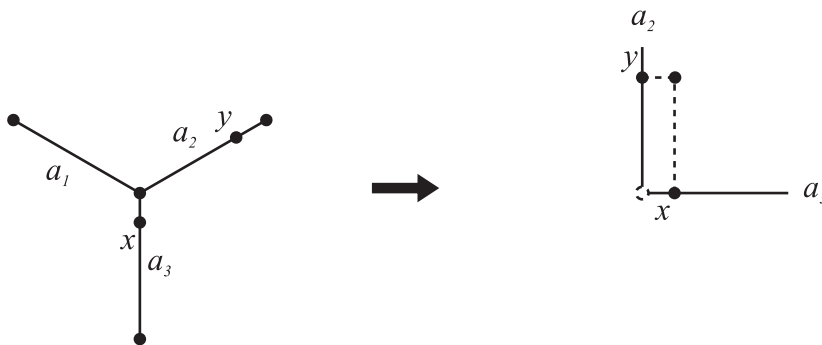


Figura 6: Exemplo de uma configuração de partículas no grafo G , e do respectivo ponto no espaço de configurações.

A Figura 7 representa outra possível posição das partículas e como explicado anteriormente, enquanto as partículas se movimentam no grafo (nas arestas representadas), surge um quadrilátero no espaço de configurações.

Note-se que cada combinação diferente de partículas e arestas irá originar uma parte diferente do espaço. Ou seja, verifica-se que existem seis maneiras

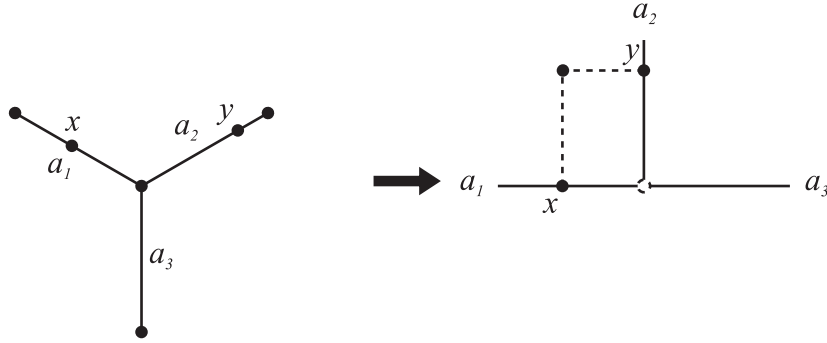


Figura 7: Mais uma configuração de partículas no grafo.

diferentes das partículas estarem em arestas distintas, pelo que o espaço de configurações possuirá seis quadriláteros. Assim, é necessário comprimir os eixos, como se pode observar na Figura 8.

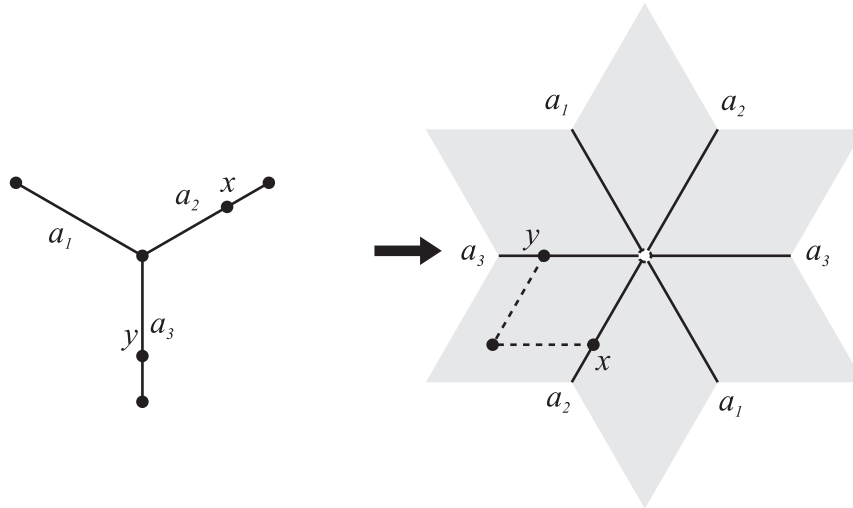


Figura 8: Espaço resultante de todas as combinações entre partículas e arestas.

Falta ainda analisar o caso em que as partículas se encontram na mesma aresta ao mesmo tempo. Como elas não podem estar no mesmo ponto ao mesmo tempo a “recta” $a_3 = a_3$ não poderá fazer parte do espaço de configurações. Esta ideia poderá ser melhor compreendida através da análise da Figura 9.

Para se poder sobrepor esta configuração na obtida na Figura 8, divide-se o espaço em metade (ver Figura 10).

Através da união de dois vértices do grafo G consegue-se construir um

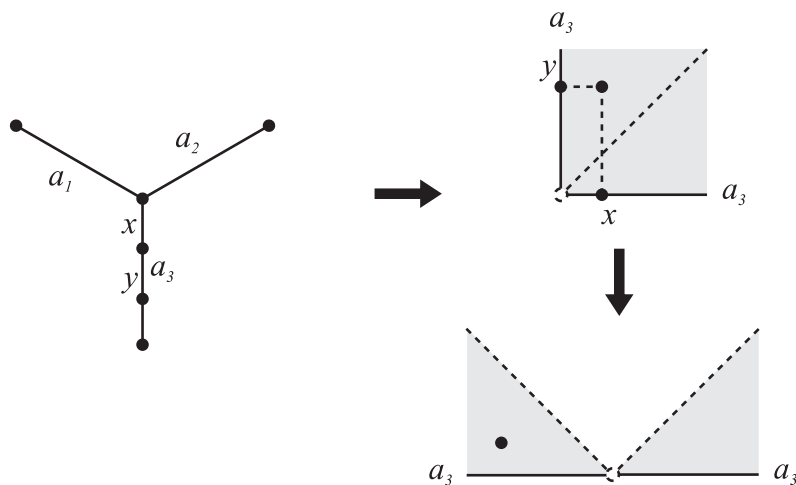


Figura 9: Caso particular de duas partículas na mesma aresta. A linha a tracejado representa um conjunto de pontos que não se pode obter: as duas partículas nunca se podem encontrar no mesmo ponto.

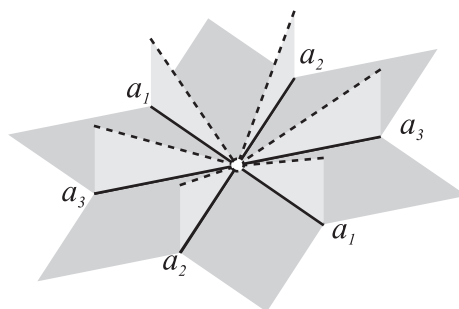


Figura 10: Espaço de configurações obtido para duas partículas que se deslocam num grafo com três arestas e quatro vértices.

novos grafo O , que se encontra na Figura 11. O espaço de configurações de um grafo desta forma, no qual se deslocam duas partículas é mais complicado de se explicar. Este só foi exposto a título de exemplo, uma vez que será referido posteriormente.

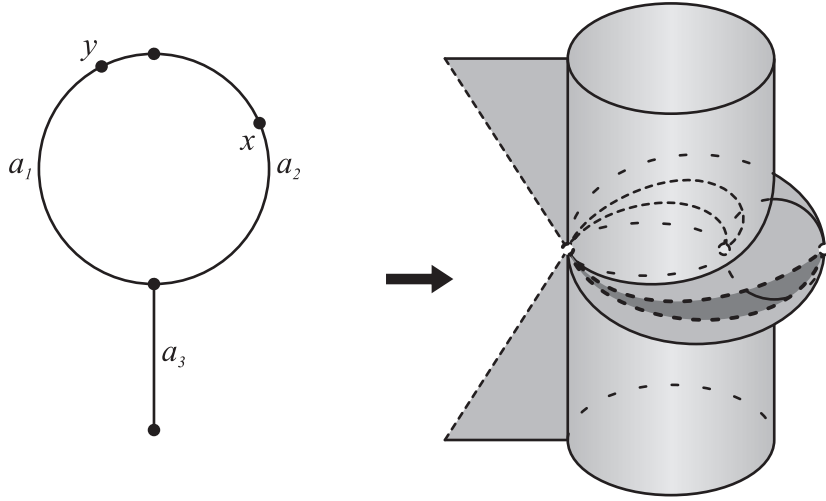


Figura 11: Espaço de configurações de duas partículas que se deslocam num grafo com três arestas e três vértices.

O espaço de configurações de duas partículas que se movimentam num grafo como o exemplificado na Figura 12, constrói-se de forma semelhante ao da Figura 6. No entanto, este é um espaço com maior complexidade, pois o vértice central do grafo possui maior valência.

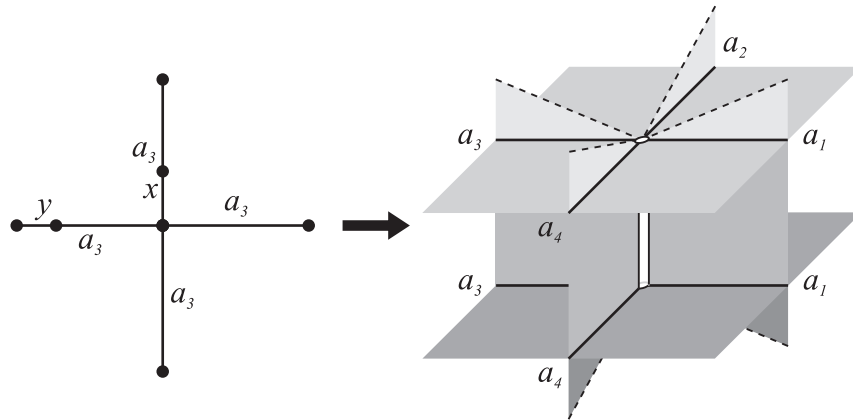


Figura 12: Espaço de configurações para duas partículas que se deslocam num grafo com quatro arestas e cinco vértices.

Como se pode verificar, aumentando a valência do vértice central do grafo aumenta-se a complexidade do espaço de configurações. Surge então a necessidade de tentar simplificar estes espaços. Uma maneira de os simplificar é impondo certas condições. Neste caso, uma forma de evitar que os robots colidam é impedir que eles se desloquem no mesmo corredor ao mesmo tempo, ou seja, em termos de grafos, terá de existir, no mínimo, uma aresta entre cada partícula. Desta forma, eliminam-se os triângulos do espaço de configurações, o que lhe confere um aspecto mais simples. Esta simplificação recebe o nome de discretização.

6 Discretização

Para se referir o espaço de configurações simplificado de partículas que se deslocam no grafo Γ usa-se a notação $D^n(\Gamma)$, ou seja, o espaço que surge impondo a condição de que terá de existir pelo menos uma aresta entre cada dois robots.

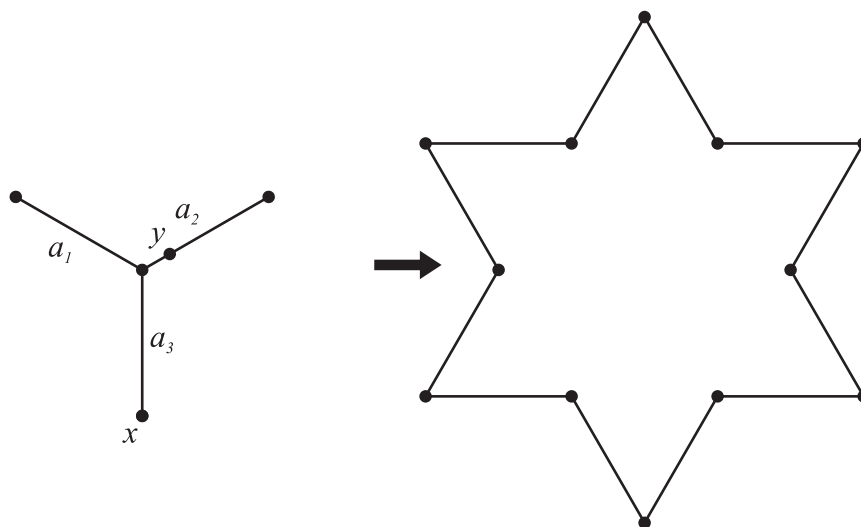


Figura 13: Espaço de configurações simplificado para duas partículas que se deslocam no grafo da Figura 10.

Tendo em conta o grafo da Figura 13 e a condição de que tem de existir pelo menos uma aresta entre as partículas, conclui-se que é necessário que uma delas esteja parada num vértice (excepto o central) para que a outra se possa movimentar nas arestas. Assim, quando se está nessa situação, no espaço de configurações simplificado, surge uma aresta. Como para cada vértice (excepto o central) a partícula que se movimenta pode estar em duas

arestas do grafo e como elas podem trocar de funções, ou seja, a partícula que está em movimento ficar fixa no vértice e a que estava fixa entrar em movimento tem-se que, no espaço de configurações, se irá ter doze arestas. Quando as duas partículas estão em vértices distintos, nesse espaço, aparecerá um ponto. Quando a partícula x está parada num vértice, a y poderá estar nos outros três vértices e fazendo as contas conclui-se que terão de existir doze pontos no espaço de configurações, como se pode ver na Figura 13.

Este é um espaço conexo e orientável.

Para o caso de duas partículas que se deslocam no grafo da Figura 14, quando uma partícula de encontra em movimento na aresta 3, a outra só poderá estar parada no vértice superior e desta situação surge uma aresta no espaço de configurações. Para que a partícula que se encontrava no vértice superior se possa movimentar, a outra terá de estar fixa no vértice inferior. Ora, como as partículas não conseguem trocar entre si sem quebrarem a condição imposta o espaço terá de ser desconexo, como se pode observar na Figura 14.

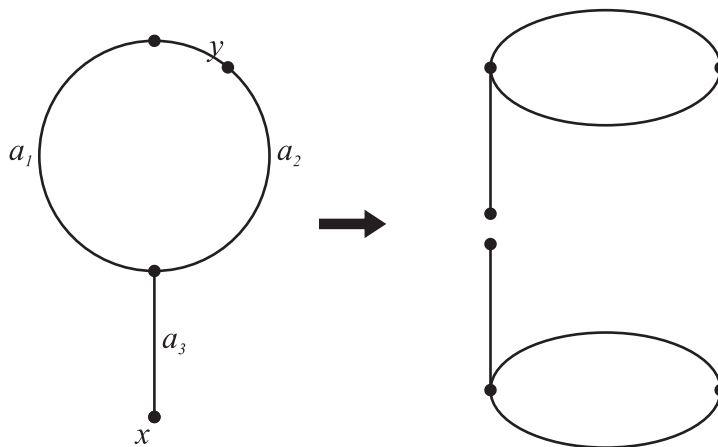


Figura 14: Espaço de configurações simplificado para duas partículas que se deslocam no grafo da Figura 11.

Seguindo a mesma linha de raciocínio que se usou para a construção do espaço de configurações simplificado da Figura 13, de cada vez que as duas partículas se encontram paradas em vértices distintos do grafo surge um ponto no espaço de configurações. Assim, quando a partícula x está parada num vértice, a y poderá estar em qualquer um dos outros quatro, mas a partícula x poderá (quando a y assim o permitir) estar num total de cinco vértices, donde surgem 20 pontos no espaço de configurações simplificado. As arestas deste espaço formam-se quando se tem uma partícula num vértice e outra em movimento nas arestas. Ora para um dado vértice a partícula

em movimento pode estar em três arestas distintas. Como temos quatro vértices e duas partículas que permitem esta situação, segue-se que se tem $4 \times 3 \times 2 = 24$ arestas, como se pode ver na Figura 15 (entre cada dois vértices existe uma aresta). Este espaço não é orientável.

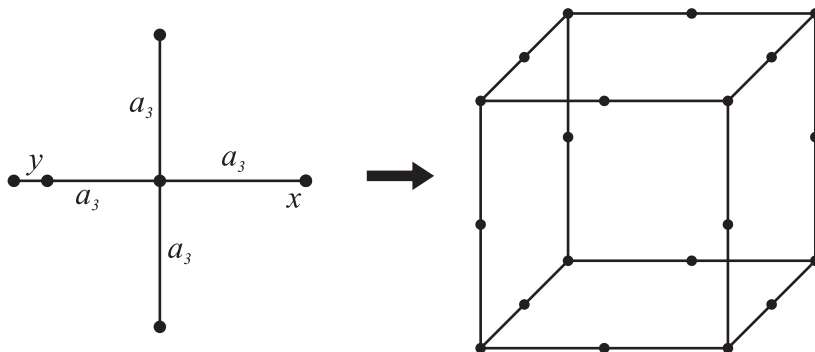


Figura 15: Espaço de configurações simplificado para duas partículas que se deslocam no grafo da Figura 12.

Verifica-se então que através do conhecimento do aspecto do grafo se consegue inferir sobre o número de vértices, arestas e faces do espaço de configurações simplificado.

Pretende-se agora saber quais os constituintes deste espaço para grafos não-planares. Estes são uma rede de arestas e vértices, onde os vértices surgem da intersecção de arestas, mas que ao serem representados no plano aparecem outras intersecções que não são vértices. São, no fundo, grafos que não podem existir no plano.

Através da forma do grafo pretende-se agora contabilizar o número de vértices, arestas e faces do espaço de configurações simplificado.

Análise-se primeiro o grafo K_5 ¹:

Para a partícula x fixa num vértice, a y poderá estar em qualquer um dos outros quatro vértices. No entanto, a partícula x poderá estar num total de cinco vértices, donde se conclui que o espaço de configurações simplificado contém $5 \times 4 = 20$ pontos.

Quando a partícula x está parada num vértice, a y poderá deslocar-se em seis arestas distintas sem no entanto quebrar a condição imposta. Mas a partícula x poderá estar parada num total de cinco vértices e como se tem duas partículas segue-se que o espaço é constituído por $6 \times 5 \times 2 = 60$ arestas.

Neste caso, as partículas podem movimentar-se ao mesmo tempo, em arestas diferentes, mantendo sempre, no mínimo, uma aresta entre elas.

¹ A notação K_n denota um grafo com n vértices, em que cada um está ligado a todos os outros, ou seja, entre qualquer par de vértices distintos existe uma aresta a ligá-los.

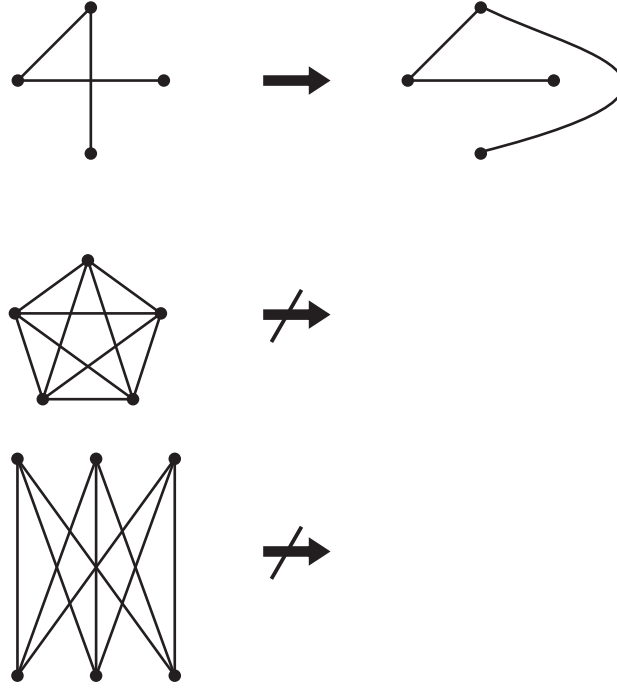


Figura 16: Exemplos de grafos planar (primeiro) e não planares (segundo, denominado por K_5 , e último, denominado por $K_{3,3}$).

Daqui surgirão faces no espaço de configurações. Desta forma, quando a partícula x se movimenta numa aresta, a y pode movimentar-se em três arestas. Determinam-se um total de $3 \times 10 = 30$ faces.

Fazendo o mesmo estudo para o grafo $K_{3,3}$ ², conclui-se que o seu espaço de configurações simplificado é constituído por: 30 pontos, 72 arestas e 36 faces.

Quando o espaço é conexo e orientável, este é determinado pela característica de Euler:

$$\chi(D^n(\Gamma)) = \# \text{faces} - \# \text{arestas} + \# \text{vértices}$$

$$\text{Género} = 1 - \frac{1}{2}\chi$$

Concretizando para o grafo não planar K_5 :

$$\chi(D^2(K_5)) = 30 - 60 + 20 = -10$$

² A notação $K_{m,n}$ denota um grafo com dois conjuntos de vértices, em que m e n são as dimensões respectivas, e entre cada vértice do primeiro se liga a cada um do segundo.

$$\text{Género} = 1 + \frac{10}{2} = 6$$

A partir daqui pode-se concluir que o espaço de configurações simplificado $D^2(K_5)$ pode ser transformado topologicamente na Figura 17.



Figura 17: Variedade de gênero 6.

Concretizando para o grafo $K_{3,3}$

$$\chi(D^2(K_{3,3})) = 36 - 72 + 30 = -6$$

$$\text{Género} = 1 + \frac{6}{2} = 4$$

Donde se conclui que o espaço de configurações simplificado $D^2(K_{3,3})$ pode ser transformado topologicamente na Figura 18.



Figura 18: Variedade de gênero 4.

No entanto, estas simplificações $D^n(\Gamma)$ nem sempre são válidas. Já foi dado um exemplo em que a simplificação de $C^n(\Gamma)$ em $D^n(\Gamma)$ não é credível.

Repare-se que o espaço da Figura 11 é conexo e o da Figura 14 não o é, pelo que não se está perante uma boa simplificação, pois os espaços deveriam ser semelhantes.

Para que $D^n(\Gamma)$ seja uma boa simplificação de $C^n(\Gamma)$ tem-se de conseguir transformar $C^n(\Gamma)$ em $D^n(\Gamma)$ de forma contínua. Neste sentido surge o seguinte teorema:

Teorema 1. *Para $n > 1$ e um grafo Γ com pelo menos n vértices, $D^n(\Gamma)$ é uma boa simplificação de $C^n(\Gamma)$ se e só se:*

1. cada caminho entre dois vértices de valência não igual a 2 passa pelo menos por $n - 1$ arestas;
2. quando se parte de um vértice e se volta a ele passa-se pelo menos por $n + 1$ arestas³.

³ O grafo da Figura 11 falha esta condição. De acordo com o Teorema 1, inserindo um vértice 2-valente à parte circular do grafo garante que a simplificação do respectivo espaço de configurações é válida.

A dimensão do espaço $D^n(\Gamma)$ tem uma grande importância na prática, pois uma grande dimensão faz crescer muito rapidamente a complexidade destes espaços, criando grandes dificuldades de visualização.

Neste sentido surge outro teorema:

Teorema 2. *Dado um grafo Γ com V vértices de valência maior que 2, o espaço $D^n(\Gamma)$ possui no máximo dimensão igual a V .*

Donde se conclui que a dimensão do espaço de configurações simplificado não depende do número de partículas, neste caso robots, que se deslocam pelo grafo.

Seja Υ_k um grafo com $k > 2$ arestas e $k + 1$ vértices, em que todas as arestas estão unidas no mesmo vértice central (relembrando a Figura 6, temos um Υ_3 , e a Figura 12, que é um Υ_4).

Pelo teorema verifica-se que Υ_k , com $k = 3, 4$ possui dimensão igual a 1.

Assim, desde que as propriedades topológicas sejam determinadas pela característica de Euler, pode-se apenas determinar o número de vértices, arestas e faces para se conseguir determinar a dimensão desses espaços.

A partir daqui pode-se provar que $D^n(\Upsilon_k)$ pode deformar-se num grafo tipo *bouquet* com P curvas fechadas que se encontram num ponto, tal como as pétalas numa flor. O número de curvas fechadas P é dado por

$$P = 1 + (nk - 2n - k + 1) \frac{(n + k - 2)!}{(k - 1)!}$$

Por exemplo, $D^2(\Upsilon_3)$ pode ser deformado/simplificado numa curva apenas como se pode ver na Figura 19.

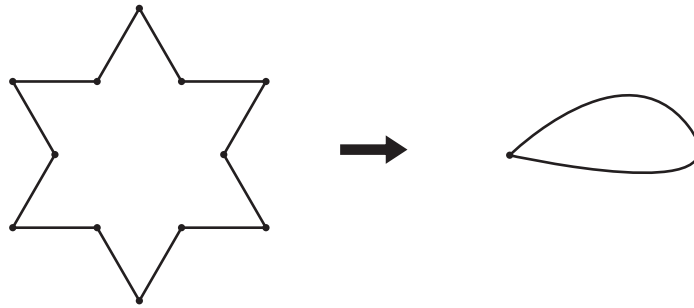


Figura 19: Deformação de $D^2(\Upsilon_3)$.

Por exemplo, $D^2(\Upsilon_4)$ pode ser deformado/simplificado na curva que se pode ver na Figura 20.

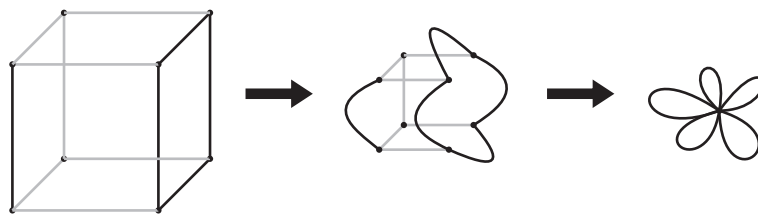


Figura 20: Deformação de $D^2(\Upsilon_4)$.

7 Tranças

Do estudo de tranças surgem vários tipos de espaços de configurações, daí que este tema seja abordado nesta exposição.

Desta forma, convém que se tenha pelo menos uma ideia sobre o que é uma trança e conhecer algumas das suas propriedades.

O grupo das tranças com n fios é gerado pelos elementos da Figura 21 com relações:

- a. $\sigma_i \sigma_i^{-1} = e = \sigma_i^{-1} \sigma_i$
- b. $\sigma_i \sigma_{i+1} \sigma_i = \sigma_{i+1} \sigma_i \sigma_{i+1}$
- c. $\sigma_i \sigma_j = \sigma_j \sigma_i, |i - j| > 1$

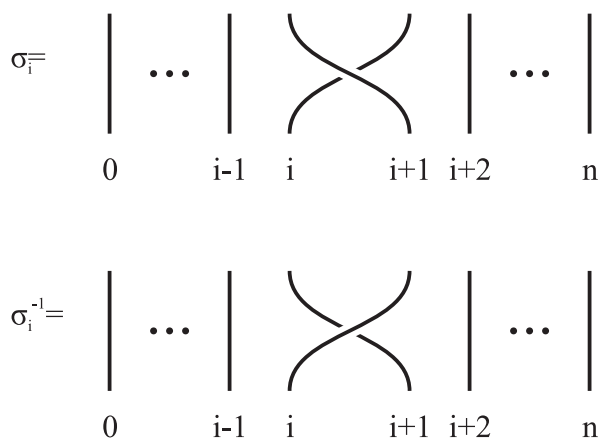


Figura 21: Elementos dos grupos das tranças com n fios.

Nas páginas seguintes ilustram-se as três relações.

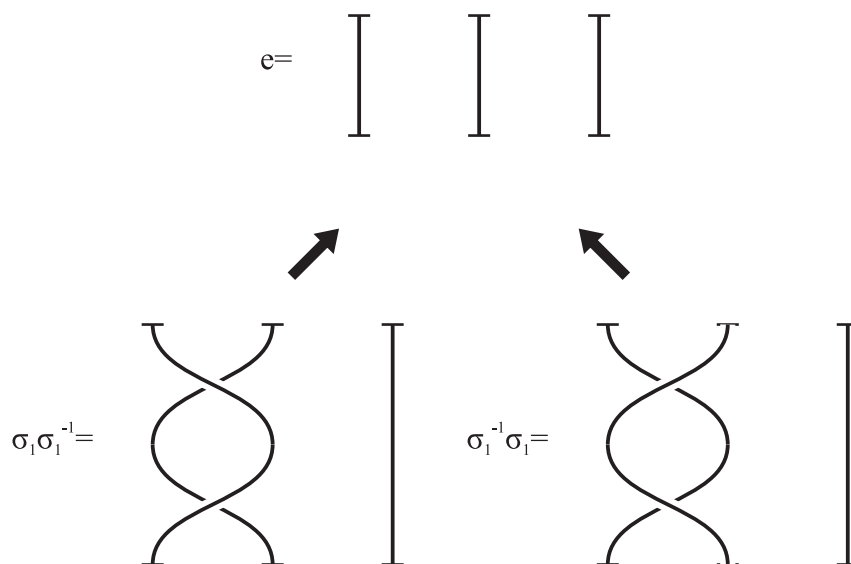


Figura 22: Exemplo da relação a.

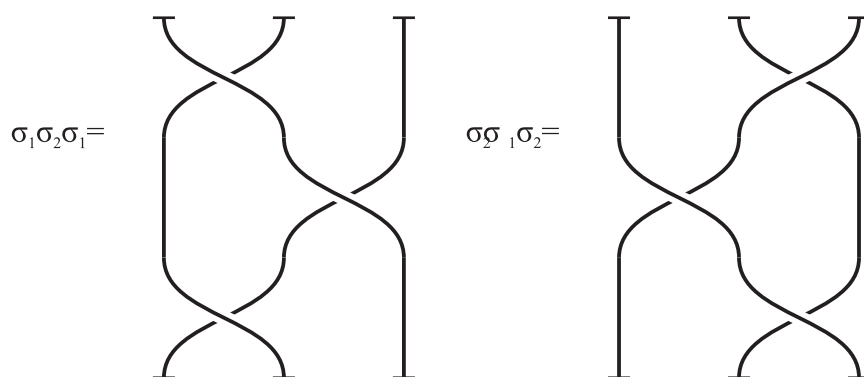


Figura 23: Exemplo da relação b.

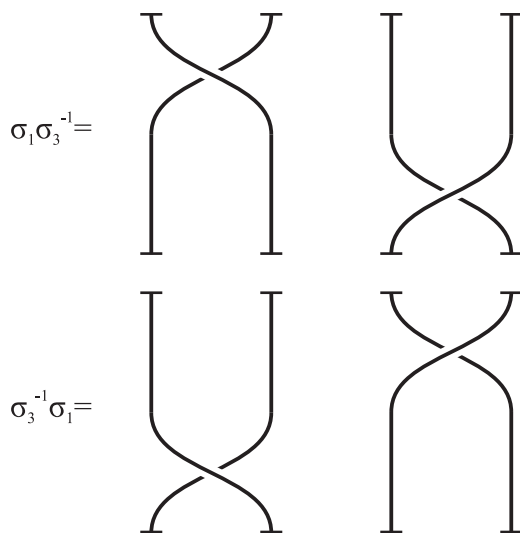


Figura 24: Exemplo da relação c.

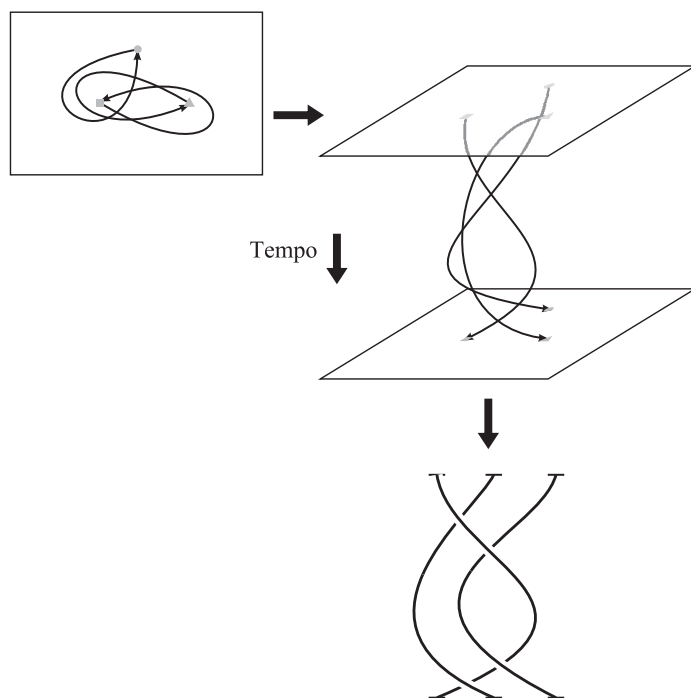


Figura 25: O gráfico de espaço-tempo para uma curva fechada em $C^3(\mathbb{R}^2)$ é uma trança pura de três fios.

Depois desta pequena introdução, pretende-se relacionar as tranças com os espaços de configurações.

Uma trança pura é uma trança onde cada fio volta à sua posição de origem. Na Figura 25 encontra-se um exemplo de uma trança deste tipo. Qualquer curva fechada no espaço $C^m(\mathbb{R}^2)$ é uma trança pura de n fios, pois nas curvas fechadas os pontos inicial e final coincidem, tal como nas tranças puras. Imagine-se três robots distintos que pretendem voltar ao seu ponto inicial de pois de efectuarem a sua tarefa. Ao longo do tempo, estes vão percorrendo o seu caminho. Assim, pode-se representar os seus gráficos espaço-tempo. Uma vez que cada robot volta ao seu ponto inicial, o gráfico obtido poderá ser visto como uma trança pura de três fios. Este processo encontra-se esquematicamente representado na Figura 25.

7.1 Curiosidade

Tem-se que um nó poderá ser obtido se “fecharmos” uma trança, tal como se pode ver na Figura 26.

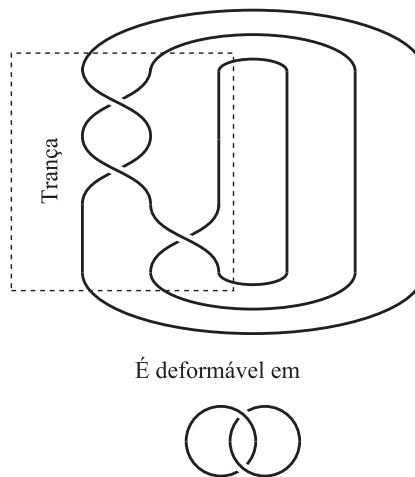


Figura 26: Resultado da deformação de uma trança previamente fechada, o chamado Enlace de Hopf.

Agradecimentos

Em primeiro lugar gostaria de agradecer ao professor Roger Picken do Departamento de Matemática do Instituto Superior Técnico por toda a ajuda e tempo dispensado, tanto no processo de aprendizagem do tema como na sua posterior apresentação.

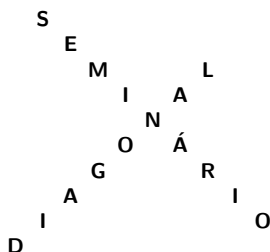
Aproveito para agradecer à Fundação Calouste Gulbenkian pela oportunidade que me foi dada para aprofundar os meus conhecimentos.

Não posso deixar de agradecer à Organização do Seminário Diagonal do Instituto Superior Técnico o convite para a apresentação do seminário e para a elaboração deste artigo, tal como à Organização do Seminário Diagonal da Universidade do Porto pelo convite para a apresentação do mesmo seminário.

Finalmente, um agradecimento muito especial para o João Pedro Maurício e Silva, que me apoiou desde o início, me ajudou bastante na preparação da apresentação do seminário e deste artigo.

Referências

- [1] A. Abrams e R. Ghrist, Finding topology in a factory: configuration spaces, *The American Mathematical Monthly* **109** (2002), 140-150.
- [2] J. S. Birman, *Braids, Links, and Mapping Class Groups*, Princeton University Press, 1975.
- [3] J. S. Carter, *How Surfaces Intersect in Space: an Introduction to Topology*, Singapore World Scientific, 1995.
- [4] R. L. Fernandes e M. Ricou, *Introdução à Álgebra Geral*, IST Press, 2004.
- [5] L. H. Kauffman, *Knots and Physics*, Singapore World Scientific, 1994.



Volta ao Mundo em 45 Minutos

Diogo Oliveira e Silva

3º ano de Matemática

Universidade do Porto

doos85@gmail.com

Palavras Chave

combinatória, função geradora, passeio aleatório, números de Catalan, permutação, problema de André, números de Bernoulli, números de Euler, função ζ de Euler.

Resumo

Que tal juntarmo-nos a Bernoulli, Catalan, Euler e outros numa intrépida aventura que nos levará a montanhas habitadas por serpentes indomáveis e de onde teremos uma vista privilegiada sobre o mundo da combinatória?

Material necessário: alguma vodka, energia q.b. e muita curiosidade! A mistura não pode deixar de ser explosiva...

1 Introdução

A viagem que motiva este artigo durou um pouco mais de 45 minutos. Começou no princípio de um dia quente de Verão, corria o ano de 2004. Estava prestes a embarcar numa aventura. Nela viajaria até à Rússia, país de lendárias dimensões e dono de uma não menos mítica escola de matemáticos. Tinha como objectivo participar no programa “Math in Moscow”¹, oferecido pela Universidade Independente de Moscovo. Aí, e durante quatro meses, vi-me “bombardeado” por quantidades alucinantes de matemática, sob as mais variadas formas e feitios. Um dos cursos que mais me motivou foi entusiasticamente leccionado pelo Professor Yuri Burman e versava a Combinatória. Nele, foram-me propostos vários problemas que, acompanhados pelos típicos biscoitos moscovitas para chá, me proporcionaram agradáveis tardes de domingo.

Debruço-me, de seguida, sobre alguns destes problemas. Falo, na primeira secção, do problema do bêbado unidimensional e resolvo-o completamente.

¹ Informações em <http://www.mccme.ru/mathinmoscow/>.

A segunda secção é dedicada a vários problemas que, aparentemente irrelevantes, têm a mesma solução. Esta é deduzida de forma rigorosa e dela se retiram algumas consequências não triviais. A terceira secção tem um carácter manifestamente diferente das outras duas: aí, aborda-se o problema de André, a título informativo e omitindo quase todas as provas.

2 O Problema do Bêbado

É das ruas mais inóspitas (e mais compridas, também!) da capital russa que vem o primeiro problema:

Exemplo 1 (Problema). Seja n um inteiro. Suponhamos que temos uma partícula material colocada inicialmente na posição n da recta real onde se marcaram as posições correspondentes aos números inteiros, e só a esses. Em cada unidade de tempo a partícula desloca-se uma unidade para a esquerda ou uma unidade para a direita, e fá-lo com probabilidade $1/2$. Qual é a probabilidade (dependente da posição inicial n) de a partícula vir a passar pela origem (isto é, pela posição 0) da recta?

Podemos dar uma côr adicional ao problema se pensarmos num bêbado bem recheado de vodka² como sendo a partícula material, e num poço situado na posição 0. Se em cada minuto o bêbado dá um passo para a esquerda ou para a direita, e se conserva as suas faculdades mentais por forma a fazê-lo com probabilidade $1/2$, coloca-se a questão quanto à probabilidade de sobrevivência do pobre bêbado, que se cair no poço dificilmente se recomporá do choque...

Consideremos então um bêbado B , e seja p_n a probabilidade de B cair no poço, dado que começou o seu passeio aleatório na posição n .

As regras do jogo e o teorema da probabilidade total³ garantem-nos que a seguinte equação se verifica, para todo o inteiro n :

$$(1) \quad p_n = \frac{1}{2}p_{n-1} + \frac{1}{2}p_{n+1}.$$

Vamos calcular os termos da sucessão $(p_n)_{n \in \mathbb{Z}}$ recorrendo a um dos mais proveitosos conceitos na combinatória enumerativa: o de *função geradora*.

DEFINIÇÃO 1. Dada uma sucessão complexa $(z_n)_{n \in \mathbb{N}_0}$, a função geradora associada é

$$Z(t) := \sum_{n=0}^{\infty} z_n x^n.$$

² Recomenda-se a “Moskovskaia” para resultados optimais...

³ Ver, por exemplo, [4].

No fundo, dada uma sucessão $(z_n)_{n \in \mathbb{N}_0}$, construímos uma série (formal) de potências cujos coeficientes são os termos da sucessão de partida.

Uma classe de sucessões para a qual é fácil determinar as correspondentes funções geradoras é a classe das sucessões definidas por recorrências lineares de coeficientes constantes de 2.^a ordem⁴. Acontece que a equação (1) é equivalente à equação

$$p_{n+1} = 2p_n - p_{n-1}.$$

Ilustremos então, com o caso que temos em mãos, o método de determinação da função geradora associada a uma sucessão definida por um tal recorrência. Sem nos preocuparmos com questões de convergência temos, por simples manipulações algébricas:

$$\begin{aligned} p_{n+1} &= 2p_n - p_{n-1} \\ \Leftrightarrow \sum_{n \geq 1} p_{n+1} t^{n+1} &= \sum_{n \geq 1} (2p_n) t^{n+1} + \sum_{n \geq 1} (-p_{n-1}) t^{n+1} \\ \Leftrightarrow P(t) - (p_0 + p_1 t) &= 2t(P(t) - p_0) - t^2 P(t) \\ \Leftrightarrow P(t) &= \frac{p_0 + (p_1 - 2p_0)t}{(1-t)^2} = \frac{1}{1-t} \end{aligned}$$

Na última igualdade foram usadas as seguintes condições iniciais:

- i) $p_0 = 1$ (óbvia: p_0 representa a probabilidade de o bêbado vir a cair no poço dado que...já lá está!);
- ii) $p_1 = 1$ (resulta de $p_1 = p_{-1}$, por simetria do problema, e de $p_1 = 2p_0 - p_{-1}$).

E com isto acabámos de determinar a função geradora $P(t)$ e temos o nosso problema resolvido! De facto,

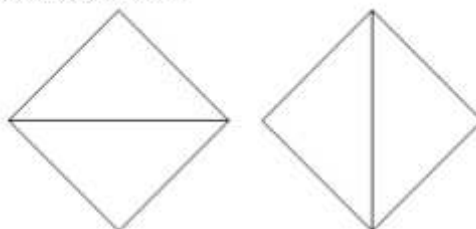
$$P(t) = \frac{1}{1-t} = \sum_{n \geq 0} t^n = \sum_{n \geq 0} 1 \cdot t^n,$$

donde $p_n = 1$, para todo $n \in \mathbb{N}_0$. Pela simetria do problema, $p_n = 1$ para todo $n \in \mathbb{Z}$.

O resultado pode parecer um pouco surpreendente: o que estamos a afirmar é que, *independentemente* da posição onde o bêbado começa o seu passeio aleatório, ele irá cair no poço com probabilidade 1. Coitado do bêbado!

⁴ Isto é, por uma condição do género de $a_n = \alpha a_{n-1} + \beta a_{n-2}$, onde α e β são complexos arbitrários (e β é não nulo).

4 lados, 2 maneiras:



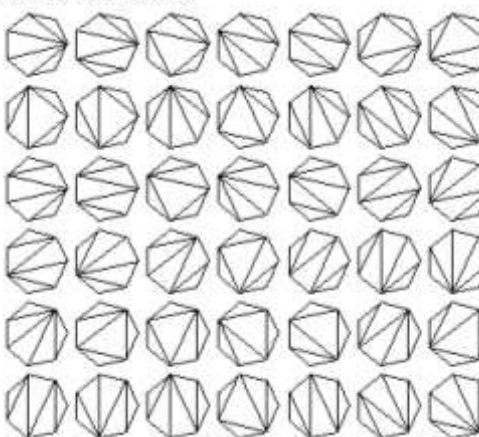
5 lados, 5 maneiras:

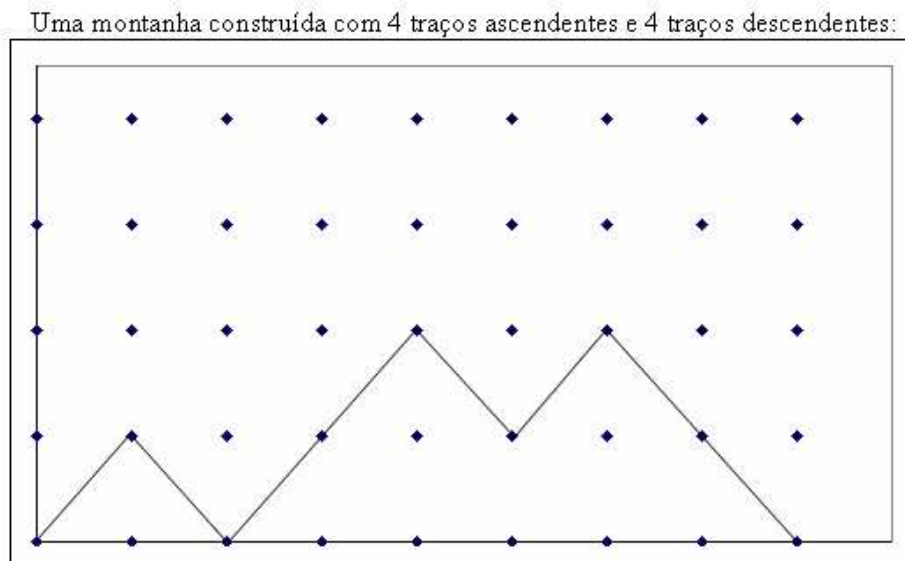


6 lados, 14 maneiras:



7 lados, 42 maneiras:



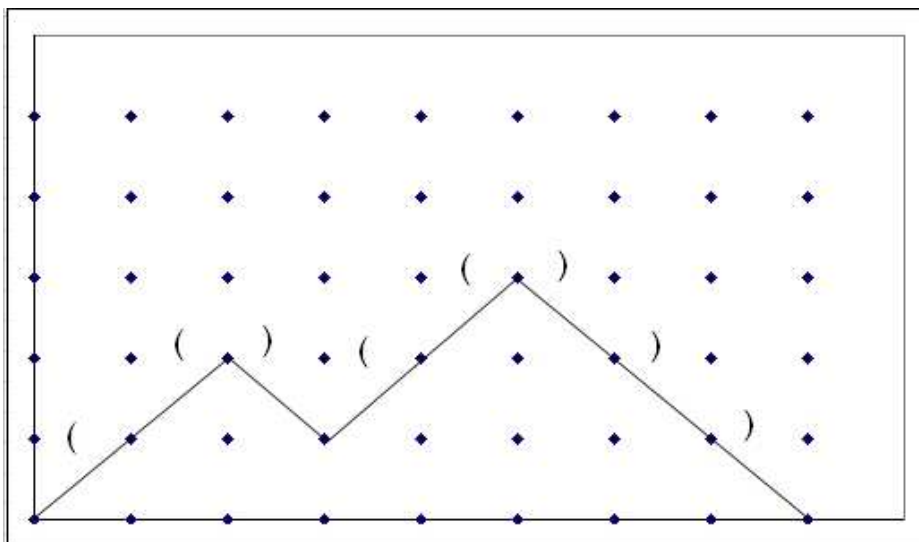


Exemplo 6 (Problema). Quantos apertos de mão sem cruzamento são possíveis com n pares de pessoas?



3 pares de pessoas podem cumprimentar-se de 5 modos distintos.

Mostra-se em [1] que todos estes problemas têm a mesma solução. A título exemplificativo, aqui está uma figura que elucida como se pode construir uma bijecção entre o conjunto das soluções do problema 3 (estruturas regulares de parêntesis) e o conjunto das soluções do problema 5 (montanhas):



A solução para qualquer um dos problemas reside na famosa sequência de Catalan $(c_n)_{n \in \mathbb{N}_0}$, cujos primeiros termos são

$$1, 1, 2, 5, 14, 42, 132, 429, 1430, 4862, \dots$$

Procuramos uma fórmula explícita para o n -ésimo número de Catalan c_n .

Como todos os problemas têm a mesma solução, podemos concentrar-nos num deles: escolhemos o problema 4.

Em geral, c_n vai ser obtido por recorrência. Para isso, a observação crucial é a de que, para $n \neq 0$, existe exactamente uma operação “ $\cdot$ ” fora de todos os parêntesis: é a multiplicação que se faz em último lugar. Suponhamos então que este “ $\cdot$ ” está entre x_k e x_{k+1} , para algum $0 \leq k \leq n-1$. Existem então c_k maneiras de inserir parêntesis no produto $x_0 \cdot x_1 \cdot \dots \cdot x_k$ e existem c_{n-k-1} maneiras de inserir parêntesis no produto $x_{k+1} \cdot x_{k+2} \cdot \dots \cdot x_n$. Ao todo, para este k fixo, obtemos $c_k c_{n-k-1}$ maneiras de colocar parêntesis em $x_0 \cdot x_1 \cdot \dots \cdot x_n$. Como o k pode ser qualquer entre 0 e $n-1$, concluímos que

$$(2) \quad c_n = c_0 c_{n-1} + c_1 c_{n-2} + \dots + c_{n-1} c_0 = \sum_{k=0}^{n-1} c_k c_{n-k-1}.$$

Esta fórmula é válida para todo o número natural $n \geq 1$. Por exemplo, $c_4 = c_0 c_3 + c_1 c_2 + c_2 c_1 + c_3 c_0 = 5 + 2 + 2 + 5 = 14$. Ainda não obtivemos, contudo, a ambicionada fórmula explícita para o n -ésimo número de Catalan. Para tal, recorreremos novamente ao poderoso conceito de função geradora.

Considere-se então a série formal

$$C(t) = \sum_{n=0}^{\infty} c_n t^n.$$

Podemos usar a equação (2) e concluir que

$$C(t)^2 = C(t) \cdot C(t) = \sum_{n=0}^{\infty} \left(\sum_{k=0}^n c_k c_{n-k} \right) t^n = \sum_{n=0}^{\infty} c_{n+1} t^n.$$

Daqui retira-se sem dificuldade que

$$1 + t \cdot C(t)^2 = 1 + \sum_{n=0}^{\infty} c_{n+1} t^{n+1} = 1 + \sum_{n=1}^{\infty} c_n t^n = C(t).$$

Em conclusão, é válida a equação

$$t \cdot C(t)^2 - C(t) + 1 = 0.$$

Esta é uma equação quadrática em $C(t)$. Resolvendo-a (e escolhendo o sinal negativo antes da raiz para assegurar que nos mantemos no “mundo” das séries de potências) obtemos

$$C(t) = \frac{1 - \sqrt{1 - 4t}}{2t}.$$

Expandido a função obtida em série de Taylor e procedendo a algumas simplificações algébricas, temos finalmente que⁷

$$\begin{aligned} C(t) &= \frac{1 - \sqrt{1 - 4t}}{2t} \\ &= -\frac{1}{2} \sum_{n=1}^{\infty} \binom{1/2}{n} (-1)^n 4^n t^{n-1} \\ &= \sum_{n=0}^{\infty} \frac{1}{n+1} \binom{2n}{n} t^n \end{aligned}$$

Exemplo 7 (Conclusão). $c_n = \frac{1}{n+1} \binom{2n}{n}$.

TEOREMA 2 (COROLÁRIO). $n+1 \mid \binom{2n}{n}$, para cada número natural n .

⁷ Se a é um número racional, $\binom{a}{n}$ define-se pela igualdade $(1+t)^a = \sum_{n=0}^{\infty} \binom{a}{n} t^n$. Temos em particular que $\binom{1/2}{0} = 1$ e que $\binom{1/2}{n} = \frac{1}{n!} [\frac{1}{2}(\frac{1}{2}-1) \dots (\frac{1}{2}-n+1)]$ para $n \geq 1$.

Demonstração. A sequência $(c_n)_{n \in \mathbb{N}_0}$ é solução para um problema que nos pergunta “De quantas maneiras é possível...?”. Ora isto implica que todos os termos da sucessão são números inteiros, de onde segue o pretendido. $\square$

O interesse do corolário prende-se mais com o método, muito usado em combinatória, do que com o resultado em si. O método consiste basicamente em visualizar uma determinada entidade numérica como solução de um problema que, por construção, só admite soluções de um determinado tipo (neste caso, do tipo inteiro).

O facto de os números de Catalan serem solução de problemas aparentemente irrelacionados⁸ permite-nos olhá-los de diferentes perspectivas e daí retirar consequências não triviais.

Os números de Catalan vistos como triangulações de $(n+2)$ -ágonos regulares (cf. problema 2) conduzem ao seguinte resultado:

PROPOSIÇÃO 3. *Se $n+2$ é potência de um primo p , digamos, $n+2 = p^k$, e $n > 1$, então o número de Catalan c_n é divisível por p .*

Exemplo 8. $c_2 = 2 \equiv 0(\text{mod}2)$; $c_5 = 42 \equiv 0(\text{mod}7)$; $c_7 = 429 \equiv 0(\text{mod}3)$.

Demonstração. O grupo $\mathbb{Z}_{n+2}$ de resíduos módulo $n+2$ actua por rotações no conjunto das triangulações de um $(n+2)$ -ágono regular. Se $n > 1$, a acção não tem pontos fixos, e o comprimento de cada órbita é divisível por p , uma vez que é um divisor de $n+2$ (pelo teorema de Lagrange). Assim, o número total de triangulações é divisível por p . $\square$

Por outro lado, interpretando os números de Catalan em termos de estruturas regulares de parêntesis (como no problema 3), temos a seguinte

PROPOSIÇÃO 4. *Se n é potência de um primo, $n = p^k$, então $c_n \equiv 2(\text{mod}p)$.*

Exemplo 9. $c_2 = 2 \equiv 2(\text{mod}2)$; $c_3 = 5 \equiv 2(\text{mod}3)$; $c_5 = 42 \equiv 2(\text{mod}5)$.

Demonstração. O grupo $\mathbb{Z}_{2n}$ de resíduos módulo $2n$ actua no conjunto das estruturas regulares de parêntesis com n pares de parêntesis do seguinte modo: o gerador do grupo é representado pelo *shift* cíclico minimal da estrutura. Sob a acção de um tal *shift*:

- 1) o parêntesis situado no extremo esquerdo da estrutura é apagado;
- 2) em vez disso, adiciona-se um parêntesis no extremo direito da estrutura;

⁸ Veja-se o exercício 6.19 de [7] para uma lista de mais de 50 contextos diferentes onde podemos encontrar os números de Catalan.

- 3) finalmente, o parêntesis direito correspondente ao parêntesis esquerdo apagado em 1) é substituído por um parêntesis esquerdo (que agora corresponde ao parêntesis direito adicionado em 2)). Todos os outros parêntesis permanecem inalterados.

Esta transformação não tem pontos fixos se $n > 1$. Exactamente uma das órbitas desta acção tem comprimento 2. Consiste nas estruturas

$$\underbrace{()() \dots ()}_{n \text{ pares}} \text{ e } \underbrace{(() \dots ())}_{n-1 \text{ pares}}.$$

O comprimento de todas as outras órbitas é divisível por p , e o resultado segue. $\square$

4 Cobras que Apreciam a Vida...

O último problema que trataremos neste artigo vem da selva francesa. Isto porque foi originalmente proposto e resolvido em finais do século XIX pelo matemático francês Désiré André, e conta com a participação de alguns dos mais ferozes habitantes de qualquer selva que se preze. Para esclarecer esta última afirmação, introduzimos a seguinte

DEFINIÇÃO 5. Uma *n-cobra* é uma permutação $\pi \in S_n$, na qual a diferença entre cada par de elementos consecutivos muda de sinal.

Exemplo 10. Em S_5 , as permutações $(1)(23)(45)$ e $(12)(34)(5)$ são 5-cobras.

A população de cobras tem dois tipos de espécimens: cobras optimistas e pessimistas. Temos uma óbvia preferência pelo primeiro tipo de cobras, pelo que introduzimos a

DEFINIÇÃO 6. Uma *n-cobra optimista* é uma *n-cobra* π que verifica $\pi(1) < \pi(2)$.

Exemplo 11. Mais uma vez em S_5 , a 5-cobra $(1)(23)(45)$ é optimista enquanto que a 5-cobra $(12)(34)(5)$ não o é.

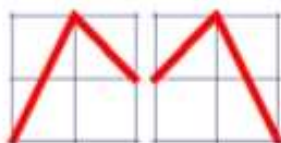
Estamos finalmente em condições de formular o

Exemplo 12 (Problema). Quantas *n-cobras optimistas* existem?

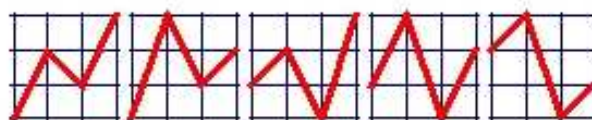
Apresentamos de seguida todas as *n-cobras optimistas*⁹ para alguns valores pequenos de n :

⁹ Por vezes, a uma cobra optimista chama-se uma *permutação zigue-zague*. Depois de ver as figuras, não será difícil imaginar porquê...

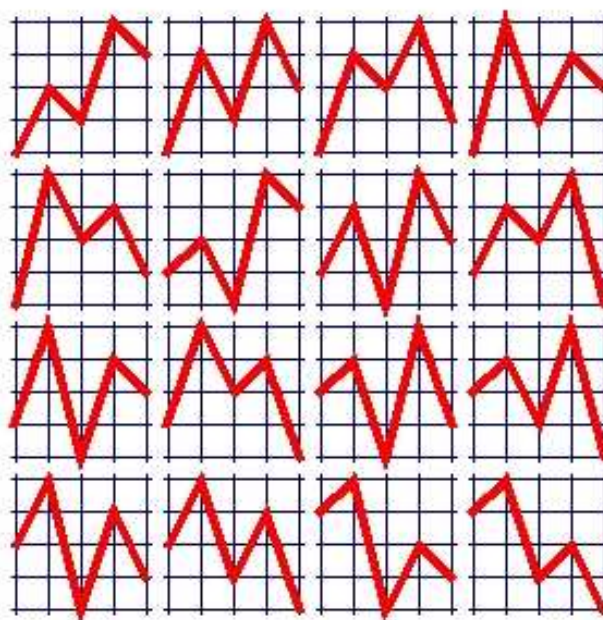
Existem duas 3-cobras optimistas,



cinco 4-cobras optimistas



e dezasseis 5-cobras optimistas.



A solução para o problema de André não é fácil, e encontra-se totalmente explicada em [6]. A estratégia consiste em proceder a uma distinção de casos, consoante a paridade de n . Temos, então:

- 1) (*n ímpar*) Seja b_n o número de n -cobras optimistas. Prova-se que a seguinte relação de recorrência é satisfeita:

$$b_{n+1} = \sum_{k \in 2\mathbb{N}+1} \binom{n}{k} b_k b_{n-k}.$$

Isto corresponde à seguinte equação satisfeita pela função geradora exponencial¹⁰ associada:

$$B'(x) = B(x)^2 + 1.$$

Podemos resolver a equação diferencial anterior por separação de variáveis, obtendo

$$\begin{aligned} dB &= (B^2 + 1)dx, \\ \int \frac{dB}{B^2 + 1} &= \int dx, \\ \arctg(B) &= x, \\ B(x) &= tg(x). \end{aligned}$$

A expansão em série de Taylor da função tangente é bem conhecida e faz uso dos chamados números de Bernoulli:

$$(3) \quad B(x) = tg(x) = \sum_{n=1}^{\infty} \frac{B_n}{(2n-1)!} x^{2n-1} = x + \frac{2}{3!}x^3 + \frac{16}{5!}x^5 + \frac{272}{7!}x^7 + \dots$$

A solução para os valores ímpares de n é por conseguinte dada pela sucessão dos números de Bernoulli (também conhecidos por *números zigue*):

$$1, 2, 16, 272, 7936, \dots$$

- 2) (*n par*) Seja agora e_n o número de n -cobras optimistas. Desta vez, a relação de recorrência tem o seguinte aspecto:

$$e_{n+1} = \sum_{k \in 2\mathbb{N}+1} \binom{n}{k} e_k b_{n-k}.$$

¹⁰ Dada uma sucessão $(z_n)_{n \in \mathbb{N}}$, a função geradora exponencial associada é dada por $Z(t) := \sum_{n=0}^{\infty} \frac{z_n}{n!} t^n$.

A equação satisfeita por $E(y)$ é

$$E'(y) = E(y)B(y)$$

e a resolução da mesma conduz ao resultado

$$E(y) = \frac{1}{\cos(y)} (= \sec(y)).$$

A expansão em série de potências da função secante é feita nos cursos de análise, e recorre aos números de Euler (ou *números zague*):

$$E(y) = \sec(y) = 1 + \sum_{n=1}^{\infty} \frac{E_n}{(2n)!} y^{2n} = 1 + \frac{5}{2!} y^2 + \frac{61}{4!} y^4 + \frac{1385}{6!} y^6 + \dots$$

A solução para os valores pares de n é portanto dada pela sucessão dos números de Euler:

$$1, 5, 61, 1385, 50521 \dots$$

A solução geral do problema de André é pois uma sucessão que consiste num “entrelaçado” de duas sequências famosas: a dos números de Bernoulli B_n e a dos números de Euler E_n .

Para terminar, vou tornar (ainda mais!) produtivo o conhecimento que temos da expansão em série de Taylor da função tangente.

A função complexa $z \mapsto tg(z)$ é meromorfa em $\mathbb{C}$, com pólos simples nos pontos da forma $\pm \frac{\pi}{2} + 2k\pi, k \in \mathbb{Z}$. Não é, portanto, de estranhar que em todo o domínio de definição da função tangente, e em particular nos pontos $x \in \mathbb{R}$ onde faz sentido falar de $tg(x)$, se tenha que

$$(4) \quad tg(x) = \sum_{n \geq 0} -\frac{1}{x - (\frac{\pi}{2} + n\pi)} - \frac{1}{x + (\frac{\pi}{2} + n\pi)}.$$

Podemos desenvolver o segundo membro de (4) em série de Taylor. Já tínhamos visto, em (3), uma expansão da tangente. Pela unicidade da série de Taylor, temos a fórmula seguinte, que se obtém igualando os coeficientes correspondentes aos termos do mesmo grau nas duas expansões:

$$(5) \quad \sum_{k \geq 0} \frac{1}{(2k+1)^{2n+2}} = \frac{\pi^{2n+2}}{2^{2n+3}(2n+1)!} B_n, n \geq 0.$$

Podemos retirar várias consequências interessantes da fórmula (5). Dá-nos, em primeiro lugar, boas aproximações numéricas do número π . Por

outro lado, este foi o método usado por Euler para calcular os valores que a função zeta¹¹ toma nos números pares positivos. O raciocínio é o seguinte:

$$\begin{aligned}
 \zeta(2n+2) &:= \sum_{k \geq 1} \frac{1}{k^{2n+2}} \\
 &= 1 + \frac{1}{2^{2n+2}} + \frac{1}{3^{2n+2}} + \frac{1}{4^{2n+2}} \dots \\
 &= (1 + \frac{1}{3^{2n+2}} + \dots) + (\frac{1}{2^{2n+2}} + \frac{1}{4^{2n+2}} + \dots) \\
 &= \sum_{k \geq 0} \frac{1}{(2k+1)^{2n+2}} + \frac{1}{2^{2n+2}} \zeta(2n+2) \\
 &= \frac{\pi^{2n+2}}{2^{2n+3}(2n+1)!} B_n + \frac{1}{2^{2n+2}} \zeta(2n+2).
 \end{aligned}$$

Daqui retiramos finalmente que

$$(6) \quad \zeta(2n+2) = \frac{\pi^{2n+2}}{2(2^{2n+2}-1)(2n+1)!} B_n.$$

Gostava de notar que os valores que a função ζ toma nos números pares positivos são os únicos casos não triviais para os quais é conhecida uma fórmula explícita. O facto de só em 1994 se ter provado que $\zeta(3)$ é um número transcendente, por um lado, e o facto de a localização dos zeros não triviais da função ζ ser objecto de pesquisa actual e constituir um dos mais famosos problemas em aberto de toda a matemática¹², por outro lado, faz-nos imaginar qual a magnitude das dificuldades de que se reveste uma compreensão satisfatória da função zeta de Riemann...

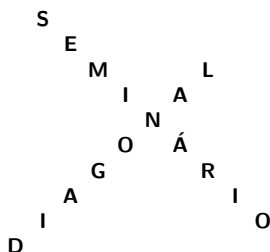
Referências

- [1] J. Conway e R. Guy, *The Book of Numbers*, Springer-Verlag, Nova Iorque, 1996.
- [2] R. Durrett, *Probability: Theory and Examples*, Duxbury, 2004.
- [3] P. Flajolet e R. Sedgewick, *Analytic Combinatorics*, 2ª Impressão, Zeroth Edition, 2004.
- [4] E. Gonçalves e N. M. Lopes, *Probabilidades*, Escolar Editora, Lisboa, 2000.

11 É a função real de variável real $s \mapsto \zeta(s) := \sum_{n \geq 1} \frac{1}{n^s}$, definida para $s > 1$. A função complexa correspondente admite um prolongamento analítico em $\mathbb{C} \setminus \{1\}$ e foi estudada um século mais tarde por Riemann.

12 Refiro-me, claro está, à hipótese de Riemann. Mais informações em http://www.claymath.org/millennium/Riemann_Hypothesis/.

- [5] T. Guttman, *Random and Self-Avoiding Walks*, University of Melbourne, 2003.
- [6] S. Lando, *Lectures on Generating Functions*, American Mathematical Society, Providence, 2003.
- [7] R. P. Stanley, *Enumerative Combinatorics*, 2^o vol., Cambridge University Press, Cambridge, 1999.
- [8] H. Wilf, *generatingfunctionology*, Academic Press, 1994.
- [9] www.math.harvard.edu/random/walks.html
- [10] www.sosmath.com/tables/bernoulli/bernoulli.html



Co-Existência de Ciclos num Sistema Dinâmico

João Nuno Mestre
2º ano de Matemática
Universidade de Coimbra
nunomestre@msn.com

Palavras Chave

sistema dinâmico, órbita periódica, ordenação, ponto fixo, grafo de transição, ciclo.

Resumo

Os sistemas dinâmicos são modelos matemáticos que traduzem uma evolução ao longo do tempo e representam inúmeros fenómenos do nosso mundo (físicos, biológicos, económicos ou meramente abstractos). É frequente que se pretenda estudar aspectos que sejam comuns a uma grande classe de sistemas.

Aqui surge o Teorema de Sharkovskii, sobre a co-existência de pontos periódicos num dado sistema. É um resultado espantoso, quer por nos dar tanta informação em troca de tão poucas hipóteses, quer por poder ser provado apenas com argumentos de matemática elementar, brincando com pequenas noções topológicas e teoria de grafos.

1 Introdução

A teoria dos sistemas dinâmicos surge já em traços gerais com Poincaré, o qual desenvolveu uma abordagem qualitativa a complicados modelos da mecânica celeste (como o famoso problema dos três corpos) que dependem duma lei determinística e evoluem em função do tempo. Desde então desenvolveu-se imenso dando origem a um largo leque de temas, abordados das mais diversas formas, num cruzamento de áreas como análise, geometria, topologia e probabilidades. A abordagem feita neste artigo parte de um ponto de vista determinístico, ou seja, não se dá lugar a aleatoriedade no comportamento dos sistemas.

Um sistema dinâmico discreto é uma colecção de estados avaliados em tempos discretos, em que um estado é dado em função de um ou mais dos

anteriores. O caso aqui estudado será aquele em que o sistema é do tipo

$$x_n = f(x_{n-1}),$$

onde o estado x_n é obtido a partir de x_{n-1} aplicando uma função f , de variável real, e n varia discretamente.

Mas como se comporta o sistema à medida que n cresce? Para o descobrir podemos, em vez de o estudar com um todo, tomar um valor no seu domínio como condição inicial e estudar a evolução desse valor sob acção de f . E de que forma depende essa evolução da escolha do valor? O que acontece ao alterar ligeiramente a função do sistema, f ? É com questões como esta que se preocupa a teoria dos sistemas dinâmicos, a qual pode ser abordada usando métodos muito diferentes. No caso unidimensional real, que iremos tratar, não são precisas ferramentas complicadas, mas isso não nos impede de encontrar fenómenos já bastante sofisticados.

Um exemplo particularmente interessante é o sistema logístico, que obtemos tomando a função

$$f(x) = rx(1 - x)$$

onde r é um parâmetro fixo. Este sistema surge muito ligado à dinâmica populacional, com o parâmetro r a representar uma taxa de crescimento que varia entre 0 e 4 para que $f(x)$ esteja em $[0, 1]$. Pensando em termos de modelo biológico, quanto maior o valor de r mais acentuada a oscilação do valor da população.

Precisamos agora de algumas noções para estudar o comportamento do sistema a partir de um valor fixado.

Definição 3. *Se x for um valor no domínio do sistema f , chamamos:*

- *trajectória de x segundo f à sucessão $x, f(x), f^2(x), \dots$*
- *órbita de x segundo f ao conjunto $\{x, f(x), f^2(x), \dots\}$.*

Vemos na Figura 1 a trajectória e a órbita de $x = 0.6$, no sistema logístico, para diferentes valores de r .

Pode acontecer que uma órbita seja finita, com se vê na Figura 2. Isto implica que ao fim de algumas iterações de f sobre x se repitam os pontos da órbita na sua trajectória. Este tipo de órbitas (também chamadas de ciclos) assume especial importância no que iremos tratar.

Definição 4. *Um ponto x diz-se um ponto periódico de período n ou n -periódico se $f^n(x) = x$ e $f^j(x) \neq x$, para $1 \leq j \leq n - 1$. A sua órbita diz-se n -periódica.*

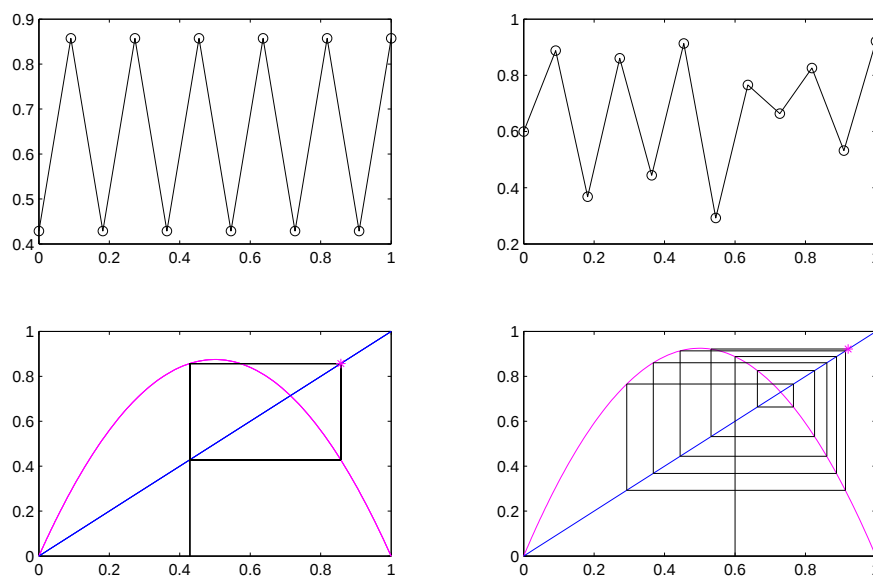


Figura 1: Sistema logístico, com $r = 3.5$ e $r = 3.7$

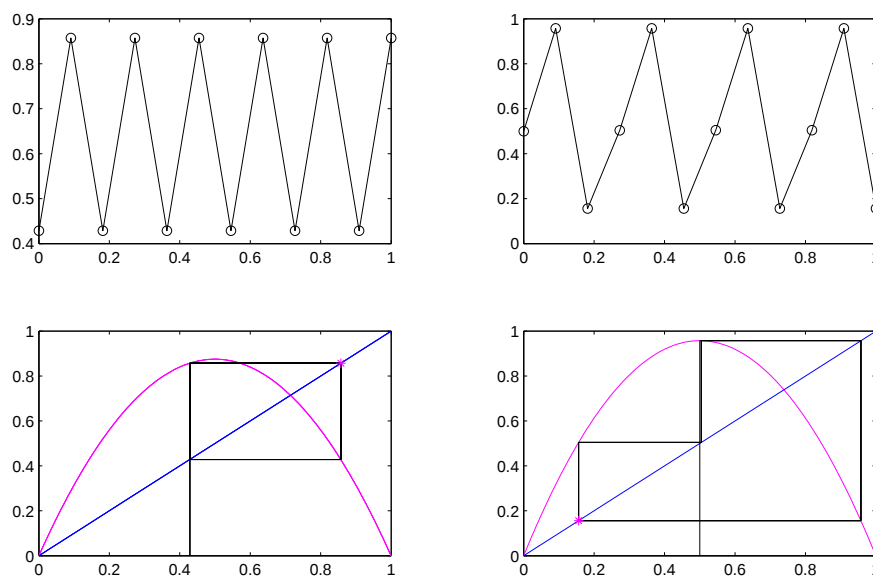


Figura 2: Sistema logístico, com $r = 3.5$, $x = 3/7$ e $r = 3.83$, $x = 0.5$

Vem então da definição que uma órbita n -periódica é um conjunto de n pontos fixos de f^n que formam, sob acção de f , um ciclo: $(x, f(x), \dots, f^{n-1}(x))$.

O Teorema de Sharkovskii¹ é precisamente relativo a estas órbitas, estabelecendo relações de “hierarquia” entre os diferentes períodos possíveis de um sistema dinâmico.

2 O Teorema de Sharkovskii

Em 1975, Yorke e Li publicaram o artigo *Period three implies chaos* [5], o qual se tornou famoso por ter utilizado pela primeira vez o termo *caos* no sentido matemático, dando assim nome à Teoria do Caos. Nesse artigo, era afirmado um caso particular do resultado por eles obtido, particularmente relevante: que a existência de um ponto 3-periódico num sistema definido num intervalo real implicava a existência de pontos k -periódicos para todo o inteiro positivo k .

Descobriu-se então que esse caso particular já tinha sido provado, no âmbito de um teorema muito mais geral (teorema da co-existência de ciclos, ou teorema de Sharkovskii, a quem é devido), em 1964, mas o artigo em que tinha sido publicado, na Ucrânia, passou despercebido no Ocidente. Assim reencontrado, tornou-se desde então um dos mais famosos teoremas da área dos sistemas dinâmicos.

A seguinte ordenação completa² dos números inteiros positivos é a ordenação de Sharkovskii, no qual o teorema se baseia, mas que também aparece noutros contextos.

$$3 \prec 5 \prec 7 \prec 9 \prec \dots \prec 2 \cdot 3 \prec 2 \cdot 5 \prec \dots \prec 2^2 \cdot 3 \prec 2^2 \cdot 5 \prec \dots \prec 2^3 \prec 2^2 \prec 2 \prec 1$$

Teorema 5 (da co-existência de ciclos). *Seja f uma função contínua num intervalo, com uma órbita p -periódica. Se $p \prec q$ então f tem uma órbita q -periódica.*

Este teorema diz que se x_0 é um ponto p -periódico então existem condições iniciais no domínio de f que levam a órbitas q -periódicas, para todo o q tal que $p \prec q$. Começaremos por provar dois casos particulares, os casos $p = 2$ e $p = 3$, pois a demonstração desses casos ilustra as técnicas fundamentais utilizadas para os restantes. Alguns dos resultados auxiliares são acompanhados apenas de um esboço de demonstração. Para as ver com todos os detalhes podem ser consultados [3] ou [6].

-
- 1 Aleksandr Nikolaevich Sharkovskii, matemático ucraniano, nascido a 1936 em Kiev. As suas principais áreas de trabalho são a teoria dos sistemas dinâmicos, equações funcionais e equações às diferenças.
 - 2 Completa pois para todos os inteiros positivos x, y temos $x \prec y$ ou $y \prec x$.

Pretendemos então provar que:

- se f tem órbita 2-periódica, tem um ponto fixo;
- se f tem órbita 3-periódica, tem órbita de qualquer período.

2.1 O Caso $p = 2$

Na demonstração do caso $p = 2$ utilizamos o seguinte resultado, que nos será muito útil para provar o teorema.

Proposição 6 (Ponto fixo). *Seja $f : J \rightarrow J$ contínua e $I = [a, b] \subseteq J$ um intervalo tal que $I \subseteq f(I)$. Então f tem um ponto fixo em I .*

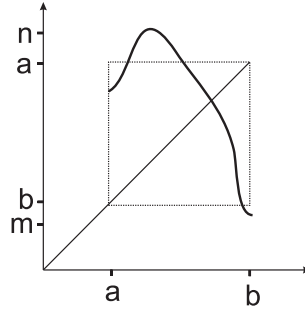


Figura 3: ponto fixo em $I = [a, b]$ ($f(I) = [m, n]$).

Suponhamos então que f tem órbita 2-periódica $\{x_1, x_2\}$, ou seja, $f(x_1) = x_2$ e $f(x_2) = x_1$.

Então $\{x_1, x_2\} \subseteq f([x_1, x_2])$, e como a imagem por uma função contínua de um intervalo é ainda um intervalo temos $[x_1, x_2] \subseteq f([x_1, x_2])$.

A existência de um ponto fixo é então imediata, pelo resultado anterior.

2.2 O Caso $p = 3$

Antes de provar o caso $p = 3$ vamos introduzir a notação e as ferramentas necessárias. Suponhamos que f tem uma órbita p -periódica, com $p \geq 3$.

Sejam $x_1 < \dots < x_p$ os pontos da órbita. Estes pontos dividem o intervalo $[x_1, x_p]$ em $p - 1$ subintervalos. A cada subintervalo $[x_i, x_{i+1}]$, com $i = 1, 2, \dots, p - 1$ chamamos S_i .

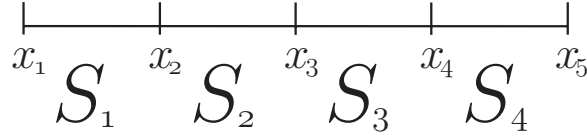


Figura 4: órbita 5-periódica

À órbita p -periódica de f vamos associar um grafo dirigido ao qual chamaremos *grafo de transição*. Este grafo tem $p - 1$ vértices, que são os subintervalos S_i , e um arco $S_i \rightarrow S_j$ se e só se $S_j \subseteq f(S_i)$. Dizemos então que S_i cobre (f -cobre) S_j .

De agora em diante usaremos a notação $A \rightarrow B$ para substituir $B \subseteq f(A)$.

Por exemplo, suponhamos que f tem órbita 3-periódica, $x_1 < x_2 < x_3$, com $f(x_2) = x_1$, $f(x_1) = x_3$ (o caso em que $f(x_2) = x_3$ é semelhante). Então temos como consequência do comportamento da órbita e da continuidade de f o grafo:

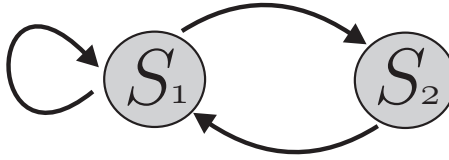


Figura 5: grafo de órbita 3-periódica

Conseguimos então traduzir o comportamento da órbita através de um grafo, o que embora não reduza a dificuldade do problema, torna-o mais visual. Conjugando esta ferramenta com o seguinte resultado auxiliar, obtemos a ideia essencial da demonstração do teorema:

Proposição 7. *Se $I_1, I_2, \dots, I_n$ são intervalos fechados e tivermos um ciclo (caminho fechado) $I_1 \rightarrow I_2 \rightarrow \dots \rightarrow I_n \rightarrow I_1$ então f^n tem um ponto fixo e f uma órbita periódica, com período m , divisor de n .*

Mais, se o ciclo não puder ser decomposto em vários ciclos iguais (designamo-lo de ciclo primitivo), então a órbita periódica tem efectivamente período n .

A demonstração deste facto, embora elementar, é longa e pode ser consultada em [3]. Mas estamos agora em posse de uma ajuda preciosa... dada uma órbita periódica num sistema, basta-nos construir o grafo de transição associado a essa órbita e procurar os seus ciclos primitivos para poder garantir a existência de outras órbitas periódicas. Vamos já aplicar esta tática para provar o caso $p = 3$ do teorema de Sharkovskii.

No grafo de transição da órbita 3-periódica encontramos os ciclos primitivos:

- $S_1 \rightarrow S_1$
- $S_1 \rightarrow S_2 \rightarrow S_1$
- $S_1 \rightarrow S_2 \rightarrow S_1 \rightarrow S_1 \rightarrow \dots \rightarrow S_1$,

correspondentes a um ponto fixo de f , a uma órbita 2-periódica e a órbitas de período m , para todo o m maior do que 3.

Temos portanto órbitas de período n , para todo o n natural.

2.3 Como Atacar o Caso Geral?

O segredo, tal como no caso anterior, está em conseguir analisar o grafo de transição da órbita que estamos a tratar. Interessa-nos portanto descobrir algumas propriedades que sejam comuns ao grafo de qualquer órbita, e é isso que faremos de seguida.

Vamos assumir que a função f tem uma órbita p -periódica, com $p > 3$. Sejam $x_1 < \dots < x_p$ os pontos da órbita. Na partição resultante, designamos por A_1 o intervalo mais à direita cujo extremo inferior é transformado num ponto superior a si próprio.

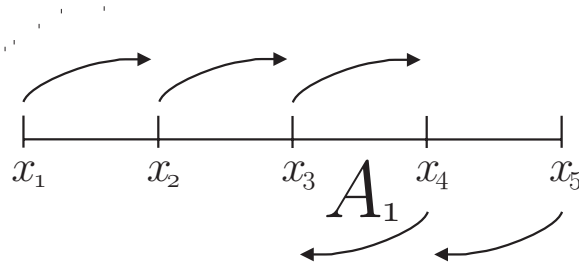


Figura 6: escolha de A_1

Pela escolha de A_1 , temos $A_1 \subseteq f(A_1)$ e aplicando f aos dois termos sucessivamente, obtemos:

$$A_1 \subseteq f(A_1) \subseteq f^2(A_1) \subseteq \dots$$

Lema 8. *O número de pontos x_i da órbita que estão contidos em $f^j(A_1)$ cresce estritamente com j , até que todos os p pontos da órbita estejam contidos em $f^K(A_1)$, para um certo K .*

Esta propriedade decorre da nossa escolha de A_1 e de que, embora os pontos da órbita permutem entre si por acção de f , não pode haver um subconjunto próprio dos pontos da órbita que permute entre si por acção de f , pela definição de órbita periódica.

Definimos então K como sendo o menor inteiro positivo para o qual $f^K(A_1)$ contém $[x_1, x_p]$. $K \leq p - 2$, pois A_1 contém 2 pontos da órbita e, com cada incremento de j , $f^j(A_1)$ contém pelo menos mais um dos p pontos.

Já sabemos então que a imagem de A_1 contém não só A_1 mas pelo menos mais um subintervalo S_i , que chamaremos de A_2 . Ou seja, $A_1 \rightarrow A_1$ e $A_1 \rightarrow A_2$.

Lema 9. *Dado um S_n fixo temos uma de duas hipóteses:*

1. *existe um subintervalo S_m distinto de S_n tal que $S_m \rightarrow S_n$;*
2. *p é par e f tem uma órbita 2-periódica.*

Caso p seja ímpar, verifica-se que, por causa de os x_i terem de percorrer toda a órbita sob acção de f e por a paridade do número de x_i ser diferente em cada lado de S_n , terá de existir pelo menos um subintervalo $S_m = [x_m, x_{m+1}]$ distinto de S_n tal que $f(x_m)$ e $f(x_{m+1})$ estão de lados opostos de S_n . Tem-se que $S_m \rightarrow S_n$.

No caso de p ser par acontece o mesmo, excepto quando todos os x_i têm a sua imagem do lado oposto de S_n . Caso isto aconteça, o intervalo contendo todos os pontos x_i à esquerda de A_1 está contido na sua imagem por f^2 e portanto f tem uma órbita 2-periódica.

Recapitulando:

- pela nossa escolha de A_1 , temos $A_1 \rightarrow A_1$;
- pelo lema 8 existe um S_n , distinto de A_1 tal que $A_1 \rightarrow S_n$;
- pelo lema 9 existe um S_m , distinto de A_1 tal que $S_m \rightarrow A_1$;
- ou então p é par e f tem uma órbita 2-periódica.

Lema 10. *Temos uma das seguintes alternativas:*

1. *f tem uma órbita $(p - 2)$ -periódica;*
2. *p é par e f tem uma órbita 2-periódica;*
3. *$K = p - 2$.*

Como se tem sempre $K \leq p - 2$, basta provar que se $K < p - 2$ e não estivermos na segunda opção, então f tem uma órbita $(p - 2)$ -periódica. Para o fazer, começa-se por concluir dos lemas 8 e 9 que existe pelo menos um ciclo ao qual A_1 pertence. Tomando de seguida o menor destes ciclos (exceptuando o ciclo $A_1 \rightarrow A_1$) não é difícil ver que o seu comprimento terá de ser menor ou igual a $p - 2$ (analisando o comportamento da órbita vê-se que se o comprimento do menor ciclo ao qual A_1 pertence fosse $p - 1$ entrariamos em contradição com a hipótese $K < p - 2$). Como este ciclo $A_1 \rightarrow A_2 \rightarrow \dots \rightarrow A_n \rightarrow A_1$ é primitivo, se tem comprimento $p - 2$ corresponde a uma órbita $(p - 2)$ -periódica; se tem comprimento menor do que $p - 2$, o ciclo $A_1 \rightarrow A_2 \rightarrow \dots \rightarrow A_n \rightarrow A_1 \rightarrow A_1 \rightarrow \dots \rightarrow A_1$, também primitivo, de comprimento $p - 2$ dá-nos a órbita pretendida.

2.4 Atacando o Caso Geral: p é Ímpar

Vamos tratar um caso particular, para ser mais fácil visualizar o que se passa, mas o caso geral é já completamente análogo. Consideremos então por exemplo que o maior número ímpar para o qual f tem uma órbita periódica é 7. Pelo lema 10 concluímos que $K = p - 2 = 5$. Como K é o maior possível, em cada passo de iteração de f sobre o subintervalo A_1 apenas um ponto da órbita é acrescentado ao intervalo anterior.

Seja $A_1 = [x_m, x_{m+1}]$; então:

1. ou $f(x_m) = x_{m+1}$ e $f(x_{m+1}) = x_{m-1}$,
2. ou $f(x_m) = x_{m+2}$ e $f(x_{m+1}) = x_m$.

Sem perda de generalidade consideremos o caso 1. Como $f(x_m) = x_{m+1}$, podemos por uma questão de simplicidade denotar $y = x_m$ e determinar a imagem de cada um dos pontos da órbita acompanhando as sucessivas iterações de y .

O comportamento da órbita sob a acção de f é então o visível na Figura 7.

Vemos então que $A_1 \rightarrow A_1$, $A_1 \rightarrow A_2 \rightarrow \dots \rightarrow A_6 \rightarrow A_1$ e $A_6 \rightarrow A_j$, qualquer que seja j ímpar, $j < 7$.

Podemos agora construir o grafo de transição da órbita:

Temos então órbitas periódicas de período:

- 1: $A_1 \rightarrow A_1$
- todos os períodos pares menores do que 7:

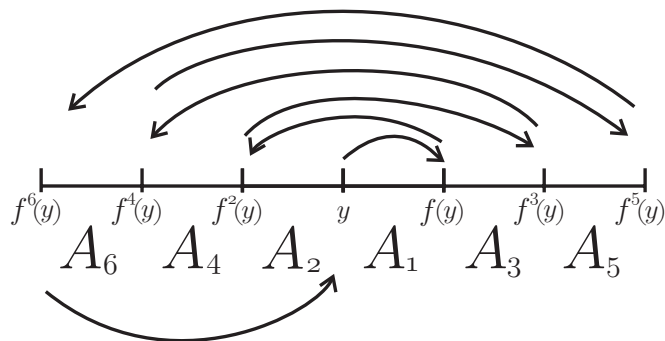


Figura 7: comportamento de uma órbita 7-periódica

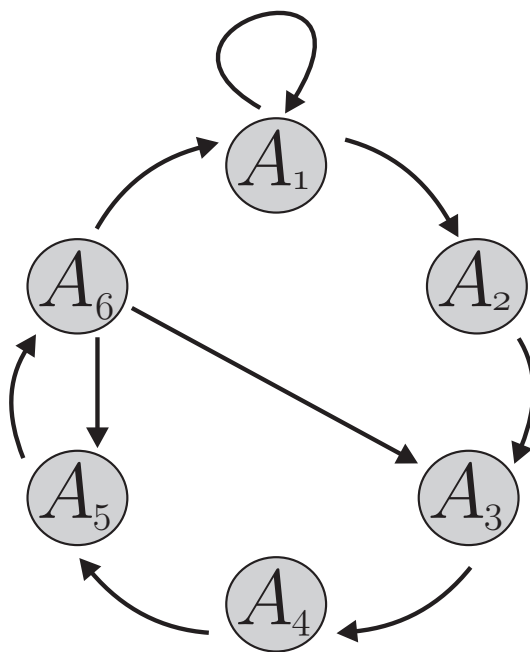


Figura 8: grafo da órbita 7-periódica

2	$A_6 \rightarrow A_5 \rightarrow A_6$
4	$A_6 \rightarrow A_3 \rightarrow A_4 \rightarrow A_5 \rightarrow A_6$
6	$A_1 \rightarrow A_2 \rightarrow \dots \rightarrow A_6 \rightarrow A_1$

- qualquer número $q > 7$: $A_1 \rightarrow \dots \rightarrow A_6 \rightarrow A_1 \rightarrow A_1 \rightarrow \dots \rightarrow A_1$.

Como afirmado pelo teorema de Sharkovskii.

2.5 Atacando o Caso Geral: p é Potência de 2

Este caso é relativamente fácil de tratar, bastando para tal obter os dois seguintes resultados:

Proposição 11. *Se f tem uma órbita p -periódica, com p par, tem também uma órbita 2-periódica.*

Para o provar, toma-se o Lema 10 e ou obtemos directamente uma órbita 2-periódica ou, efectuando uma análise semelhante à do caso $p = 7$, constrói-se o grafo de transição da órbita p -periódica, onde se identifica uma órbita 2-periódica.

Proposição 12. *Se f tem órbita 2^k -periódica, tem órbitas de períodos $2^{k-1}, \dots, 4, 2, 1$*

Como f tem órbita 2^k -periódica, $f^{2^{k-2}}$ tem órbita 4-periódica. Então, pela Proposição 11, $f^{2^{k-2}}$ tem uma órbita 2-periódica, a qual corresponde a um ponto fixo de $f^{2^{k-1}}$. Esse ponto fixo corresponde a uma órbita 2^{k-1} -periódica de f , e estendendo este raciocínio obtemos a proposição pretendida.

2.6 Atacando o Caso Geral: o Caso Intermédio

Omitimos a demonstração deste caso pois não surgem aqui ideias realmente novas, apenas uma conjugação do que já foi feito para os casos em que p é ímpar ou potência de 2. A explicação detalhada de como fazê-lo pode ser encontrada em [3]. Damos então por concluída a demonstração do teorema de Sharkovskii!

3 E Depois do Teorema...

Deve ser referido que o Teorema de Sharkovskii é óptimo, no sentido em que podemos enunciar e provar uma afirmação recíproca:

Teorema 13. *Para qualquer inteiro positivo p existe uma função $f : I \rightarrow I$ contínua, com uma órbita p -periódica e sem órbitas de período q , qualquer que seja $q \prec p$.*

Existe também uma função contínua com órbitas 2^k -periódicas para todo o k , natural, mas sem órbitas de outros períodos.

Uma explicação de como construir estas funções, acompanhada de vários exemplos, pode ser encontrada quer em [3] quer em [4].

Várias tentativas de generalizar o Teorema, ou de encontrar seus análogos noutras áreas, como por exemplo para equações diferenciais, foram levadas a cabo. No entanto, em muitos casos provou-se não ser possível fazer uma generalização directa, e noutros apenas foi possível transpor parte da informação do teorema, obtendo por exemplo ordenações não completas. Muito foi feito, no entanto, sobre os resultados de Sharkovskii e de colaboradores seus, em áreas como as teorias da estabilidade, das oscilações, da bifurcação e do caos.

Para uma introdução a estas áreas recomenda-se a leitura de [1], [3] ou [6]. Em [2] podem-se encontrar resumidos alguns dos avanços nestas áreas efectuados nas últimas décadas, bem como uma versão traduzida do artigo original de Sharkovskii.

Referências

- [1] K. T. Alligood, T. D. Sauer e J. A. Yorke, *Chaos: An Introduction to Dynamical Systems*, Springer-Verlag, 1996.
- [2] L. Alsedà, F. Balibrea, J. Llibre e M. Misiurewicz (Eds.), *Thirty Years After Sharkovskii's Theorem: New Perspectives*, World Scientific Publishing Company, 1996.
- [3] L. S. Block e W. A. Coppel, *Dynamics in One Dimension*, Springer-Verlag Lecture Notes in Mathematics **1513**, 1992.
- [4] R. L. Devaney, *An Introduction to Chaotic Dynamical Systems*, Benjamin/Cummings Publishing, 1986.
- [5] T. Y. Li e J. A. Yorke, Period three implies chaos, *American Mathematical Monthly* **82** (1975), 985-992.
- [6] Yu. L. Maistrenko, E. Yu. Romanenko, A. N. Sharkovsky, *Difference Equations and Their Applications*, Kluwer Academic Publishers, 1993.