

C I Ê N C I A A B E R T A

Teias Matemáticas

Frentes na Ciência e na Sociedade

M. PAULA SERRA DE OLIVEIRA

Coordenadora

gradiva • Imprensa da Universidade de Coimbra

(Página deixada propositadamente em branco)

MARIA PAULA SERRA DE OLIVEIRA

Coordenadora

TEIAS MATEMÁTICAS

Frentes na Ciência e na Sociedade



gradiva



Imprensa da Universidade de Coimbra

© *Gradiva – Publicações, L.^{da} / Imprensa da Universidade de Coimbra, 2004*

Coordenação editorial: *Maria Paula Serra de Oliveira*

Tradução: *Artur Soares Alves*

Carlota Isabel Leitão Pires Simões

Francisco José Craveiro de Carvalho

João Filipe Cortez Rodrigues Queiró

José Miguel Dordio Martinho de Almeida Urbano

Lia Sandra dos Santos

Mário da Silva Rosa

Paulo Eduardo Aragão Aleixo Neves de Oliveira

Revisão do texto: *Isabel Pedrome*

Capa: *António Barros* [Imprensa da Universidade. Coimbra], com imagem de

E. M. de Melo e Castro, “Fract 010 explod MC”, Dezembro de 2003

[Fractal original gerado no Fractint com tratamento no Photoshop 7.0]

Infografia: *Estúdios Estimulus* [design]

Paginação: *António Resende e Victor Hugo Fernandes*

Impressão e acabamento: *G.C. – Gráfica de Coimbra, L.^{da}*

Reservados os direitos para Portugal por:

Gradiva – Publicações, L.^{da} e Imprensa da Universidade de Coimbra

Gradiva – Publicações, L.^{da}

Rua Almeida e Sousa, 21, r/c, esq. • 1399-041 Lisboa

Telefs. 21 397 40 67/8 • 21 397 13 57 • 21 395 34 70

Fax 21 395 34 71 • Email: gradiva@ip.pt

URL: <http://www.gradiva.pt>

Imprensa da Universidade de Coimbra

Rua Antero de Quental, 195 • 3000-033 Coimbra

Telefs. 351 239 85 31 10

Fax 351 239 85 31 19 • e-mail: fjrpess@ci.uc.pt

URL: <http://www.imp.uc.pt>

ISBN: 972-662-970-5

ISBN Digital: 978-989-26-0485-5

DOI: <http://dx.doi.org/10.14195/978-989-26-0485-5>

1.^a edição: *Maio de 2004*

Depósito legal n.º 210431/04

OBRA PUBLICADA COM O PATROCÍNIO DE:
CENTRO DE MATEMÁTICA DA UNIVERSIDADE DE COIMBRA
DEPARTAMENTO DE MATEMÁTICA DA UNIVERSIDADE DE COIMBRA

FCT Fundação para a Ciência e a Tecnologia

MINISTÉRIO DA CIÊNCIA E DO ENSINO SUPERIOR

Portugal

Sumário

<i>Prefácio</i>	7
Barry Cipra <i>Math as She is Wrote</i>	13
Nuno Crato <i>Algumas experiências de divulgação da matemática na imprensa portuguesa</i>	25
Kevin Paul Scherer e Marco Avellaneda <i>Todos por um ... um por todos? Uma análise de componentes principais da dívida de obrigações Brady da América Latina de 1994 a 2000</i>	35
José M. Bernardo <i>Interpretação de resultados eleitorais: uma análise bayesiana</i>	63
João Sousa Andrade <i>A matemática e a economia</i>	77
Mário F. Lages <i>Sobre a inferência em estudos de opinião</i>	89
David M. Holland <i>Ciência global do ambiente e matemática aplicada moderna</i>	107
Michael F. Goodchild <i>Desafios à matemática na geografia</i>	121
C. Borrego, A. I. Miranda, M. Lopes e A. M. Costa <i>Matemática e ambiente: A redescoberta dos fundamentos básicos</i>	141
João Agria Torres <i>A matemática e a forma de ver o mundo</i>	155

Boris Gutkin, David Pinto e Bard Ermentrout	
<i>Matemática da neurociência: de neurónios a circuitos e de circuitos a sistemas</i>	181
Alfio Quarteroni	
<i>Modelação Matemática: uma síntese entre a teoria e as aplicações...</i>	207
J. J. Pedroso de Lima	
<i>A matemática e a compreensão da medicina</i>	219
L. Trabuco	
<i>A estrutura óssea do fémur</i>	233
Guerino Mazzola	
<i>Em direcção à «grande unificação» de composição, análise e execução musicais</i>	247
A. J. Tromba	
<i>A matemática no melhor mundo possível</i>	269
Rui Mário Gonçalves	
<i>A arte e a ciência no século XX</i>	283
F. J. Craveiro de Carvalho	
<i>Sem título, 2000, técnica mista, dimensões variáveis</i>	295

Durante o ano 2000 — Ano Internacional da Matemática — o Departamento de Matemática da Universidade de Coimbra promoveu múltiplos eventos que culminaram no Congresso *Teias Matemáticas, Frentes na Ciência e na Sociedade*, de cuja comissão organizadora fizeram parte, para além da coordenadora desta edição, os professores do Departamento de Matemática Fernanda Patrício, Carlota Simões, José Carlos Teixeira e Luís Nunes Vicente, reunindo-se neste livro o conjunto de conferências então proferidas.

A sua publicação foi possível graças à generosidade dos investigadores que fizeram a tradução e a revisão de alguns dos textos, originariamente escritos em língua inglesa, e a quem deixamos os nossos agradecimentos. Estes estendem-se também ao Departamento de Matemática da Faculdade de Ciências e Tecnologia da Universidade de Coimbra, ao Centro de Matemática da Universidade de Coimbra e à Fundação para a Ciência e a Tecnologia, cujo apoio financeiro foi determinante na execução desta obra.

O trabalho de preparação desta edição fica a dever-se à preciosa colaboração da Dr.^a Rute Andrade, a quem manifestamos a nossa gratidão.

(Página deixada propositadamente em branco)

Prefácio

Corriam em Babel os primeiros anos da criação do mundo. O Génesis, no seu capítulo XI, conta que todos os povos da Terra falavam a mesma língua e que isso lhes deu uma espantosa energia para tentar o impossível: alcançar o céu. Alcançar o céu com uma torre, a Torre de Babel, empresa esta que não era de homens, mas sim de deuses. E Deus, o Deus do Velho Testamento, reagiu com ira, confundindo a língua dos homens, que deixou de ser única.

Milhões de anos mais tarde, já não no centro da Babilónia, mas muitos quilómetros para ocidente, em Ulm, pequena aldeia da Baviera, René Descartes teve um sonho. Um sonho, um pesadelo ou uma dessas visões que acreditamos só terem acontecido na Idade Média. Imaginou algo com a função da linguagem única dos homens de Babel: o «método da Razão», como então lhe chamou e que tão detalhadamente descreveu no seu *Discurso do Método*. Permitiria o método da razão, segundo Descartes, a unificação de toda a ciência e de todo o conhecimento.

Duas gerações mais tarde, o matemático e filósofo Leibniz falou de uma «característica universalis». O mesmo sonho — a criação de uma linguagem ou de um método com os quais os problemas da ciência, das leis e da política pudessem ser analisados.

Babel e depois Descartes e Leibniz. Em comum, nestas três evocações, a referência a um único e poderoso instrumento — a linguagem (ou o método?) —, omnipresente, flexível e penetrante. Um instrumento capaz

de desencadear forças para chegar aos céus de Babel ou para descobrir o caminho que levava à verdade em todas as áreas de conhecimento.

Foram três sinais de uma visão matematizante do universo. Babel e a metáfora da linguagem única, representando a linguagem matemática. Descartes, Leibniz e a eficácia do método da razão, do método dedutivo.

Desde René Descartes, cerca de doze gerações passaram. Para além das metáforas, dos símbolos e dos sinais, é natural que nos interroguemos sobre o percurso da matematização nascente, que Descartes viveu no século XVIII. Qual o impacto da matemática quando aplicada a um universo situado para além da própria matemática? O que significa hoje a matematização do universo?

Reflectir sobre estes percursos, impactos e finalidades foi o objectivo do Departamento de Matemática da Universidade de Coimbra, que durante o ano 2000 — Ano Internacional da Matemática — promoveu múltiplos eventos que culminaram no Congresso Teias Matemáticas, Frentes na Ciência e na Sociedade, cujas conferências reunimos neste livro.

Quando surgiu a ideia de abordar este tema, com o propósito de divulgar o papel da matemática como sólido motor do desenvolvimento científico, a primeira questão que se colocou foi a dificuldade da comunicação em Matemática. É a complexidade das matérias estudadas, o seu nível de abstracção, a sua linguagem algo esotérica, ou a dificuldade dos seus protagonistas em as transmitir que explica a difícil «mediatização» da matemática e a sua quase total ausência dos meios de comunicação social? Para reflectir sobre esta estranha ausência, convidámos Barry Cipra, escritor e colaborador da Society for Industrial and Applied Mathematics, e também Nuno Crato, Professor do Instituto Superior de Economia e Gestão da Universidade Técnica de Lisboa.

Para além da reflexão sobre as dificuldades da comunicação era preciso definir claramente o nosso propósito ao publicar as Teias Matemáticas: informar, naturalmente, mas também seduzir, desvendar, «desconstruir» um edifício que por vezes parece demasiadamente perfeito e enigmático. Havia assim que optar por zonas do saber matemático que tivessem, em nosso entender, um potencial de sedução mais intenso, por se situarem mais afastadas das aplicações correntes, daquelas que desde sempre acompanharam a matemática na sua marcha de progresso.

A finança e a política não escaparam ao espantoso poder criador e ordenador da matemática. A finança hoje não se limita a uma algoritmia simples; os mercados financeiros, máquinas complexas mas facilmente adaptáveis, têm de matematizar a incerteza e as volatilidades de rápida variação. Marco Avellaneda, professor do Instituto Courant, de Nova Iorque

e também consultor da Bolsa de Wall Street, escreve sobre a modelização matemática em finanças na idade da informação. Tradicionalmente afastada da matemática, regendo-se por códigos quase sempre avessos ao rigor e ao espírito dedutivo, também a máquina política colocou hoje nos bastidores da sua cena alguns aspectos do saber matemático. José M. Bernardo, professor da Universidade de Valência, especialista em análise bayesiana e sua aplicação à teoria das sondagens e profundo conhecedor das recentes vicissitudes do panorama eleitoral espanhol, desvenda-nos nesta obra algumas das complicadas relações que se podem estabelecer entre a matemática e a política. A matemática, a finança e a política são ainda discutidas por Mário Lages, professor da Universidade Católica e Director do Centro de Estudos de Sondagens da Opinião da mesma universidade, e por João Sousa Andrade, professor da Faculdade de Economia da Universidade de Coimbra.

A força ordenadora e criativa da disciplina matemática estendeu-se também, nos últimos anos e talvez inesperadamente, a esse sacrário que é o conhecimento biomédico. De facto, os instrumentos da matemática têm vindo a penetrar em muitas regiões de tal conhecimento. Para além da presença implícita do saber matemático na tecnologia, que está na base dos meios de apoio ao diagnóstico, os modelos matemáticos permitem hoje explicar, prever e decidir num vasto número de situações. Alguns modelos matemáticos que descrevem o comportamento fisiológico e patológico do sistema cardiovascular são discutidos por Alfio Quarteroni, professor do Politécnico de Milão e da École Polytechnique Fédérale de Lausanne, na Suíça. Ainda no âmbito dos problemas biomédicos, um trabalho de Bard Ermentrout e de Boris Gutkin descreve de que forma a organização do córtex visual explica certos aspectos do funcionamento das redes neuronais.

A análise de outros aspectos da utilização da matemática em medicina é ainda objecto de reflexão de Luís Trabucho, professor do Departamento de Matemática da Faculdade de Ciências de Lisboa, e de J. J. Pedroso de Lima, professor da Faculdade de Medicina da Universidade de Coimbra.

O papel dos instrumentos matemáticos em geografia e nas ciências do ambiente é também discutido sob múltiplos aspectos. Michael Goodchild, professor da Universidade da Califórnia em Santa Bárbara, faz uma ampla exposição sobre os desafios que a geografia e a informação geográfica colocam, reflectindo sobre aplicações de inúmeros domínios da matemática. Os complexos problemas colocados pela climatologia, desde a génese dos tufões, como o El Niño, até ao aquecimento global do planeta e a sua interacção com os modernos instrumentos da matemática são focados por David Holland, professor do Instituto Courant,

de Nova Iorque. A matemática, a geografia e o ambiente são ainda amplamente discutidos por João Agria Torres, do Instituto Português de Cartografia e Cadastro, e por Carlos Borrego, professor do Departamento de Engenharia do Ambiente da Universidade de Aveiro.

Também as regras que se vislumbram na criação do universo são passíveis de uma visão matemática. Anthony Tromba, professor da Universidade da Califórnia, em Santa Cruz, torna claras algumas delas, mostrando por que razão a «A matemática está no melhor dos mundos». No melhor dos mundos, pois — apesar da suspeita de eliminar o sonho, imprimindo uma ordem e uma disciplina absolutas em todos os aspectos da vida humana — é possível descobrir paraísos perdidos para além desta aparente frieza. Referimo-nos às surpreendentes relações entre a matemática e a arte. Guerino Mazzola, professor da Universidade de Zurique, apresenta-nos uma teoria matemática para representar de forma unificada os objectos musicais. Outros paraísos é também possível entrever nos textos sobre arte e matemática de Rui Mário Gonçalves, professor da Faculdade de Letras da Universidade de Lisboa, e de F. J. Craveiro de Carvalho, professor do Departamento de Matemática da Universidade de Coimbra.

Demasiado épica poderá ter parecido a descrição destes mundos matematizados que os textos que se seguem vão apresentar.

Surgirá então talvez a questão de saber se há qualquer coisa no universo que não possa ser objecto de uma teoria matemática. Perguntar-se-á se não há crateras, distâncias sustentadas, se não há mistérios. Se todos os grandes fascínios obedecerão a regras. A regras matemáticas. Uma possível resposta terá dois tempos. Um primeiro tempo no mundo físico, onde tudo parece ser «matematizável». Embora complexo hoje, poderá, seguramente, ser formalizado amanhã. Um segundo tempo surge algures para além do mundo físico, onde acreditamos que os comportamentos, as emoções, os credos e os sonhos se não submeterão a nenhuma implacável teia matemática, continuando a desenhar crateras, suspensões e mistérios.

Faremos assim nestas teias matemáticas uma viagem alucinante de ponto de partida remoto: as regras de criação do universo, as galáxias, a história da Terra, com o seu núcleo central, os gelos, a litosfera, os oceanos, o seu espaço envolvente.

E neste percurso cósmico a matemática plasma-se, descreve e explica.

Surge a vida na Terra, as espécies evoluem e diversificam-se. De entre elas uma evolui até um estádio a que se convencionou chamar *Homo sapiens*.

Compacidades e porosidades da estrutura óssea deste *Homo sapiens* organizam-se de modo surpreendentemente otimizado de um ponto de vista matemático, explicando, entre outros aspectos, a sua posição erecta. Penetraremos no seu sistema cardiovascular e simularemos diástoles, sístoles, fluxos. Preveremos falhas. E porque do *Homo* que é *sapiens* estamos a falar, observaremos ainda a complexidade do seu córtex cerebral.

E a disciplina matemática plasma-se, descreve e explica.

Atingido o estágio último do seu desenvolvimento, organiza-se o homem em grupos de complexidade crescente, cria estruturas sociais, económicas e políticas. Organiza o território, mede distâncias, cria sistemas de informação.

Acidentalmente destrói o planeta.

E a disciplina matemática cria modelos, constrói técnicas, plasma-se, descreve e prevê.

Cansado de ser *sapiens* torna-se por fim *ludens*. E é então que faz música, pintura, poesia e literatura. E a matemática, surpreendentemente, plasma-se, organiza, torna-se lúdica, ela também.

Alucinante esta viagem. Alucinante porque percorreremos milhões de anos de história, e de histórias com uma única linguagem. A linguagem matemática.

Recordemos o Génesis. Velho Testamento. Capítulo XI. Os homens tinham uma única língua, o que criou sinergias capazes de produzir uma torre que tocasse os céus. A Torre de Babel.

Olhemos agora, algures, um livro desconhecido. Dir-vos-íamos «o grande livro do universo» se não plagiássemos alguém. Versículo único: os homens esqueceram a ira do Deus do Velho Testamento e criaram uma nova linguagem capaz de construir teias que abarcam o universo.

Maria Paula Serra de Oliveira

(Página deixada propositadamente em branco)

Barry Cipra
Colaborador da Society for Industrial
and Applied Mathematics

Math as She is Wrote¹

Devo talvez começar por explicar o título deste texto. A ideia vem do título de um livro de frases português com tradução em inglês do século XIX, *English as She is Spoke*, de Pedro Carolino. O pobre Pedro não sabia inglês. Só tinha um manual com a versão francesa de expressões portuguesas e um dicionário de francês-inglês. A sua utilização em conjunto deu origem a alguns resultados excelentes, como o seguinte (primeiro em português, depois em inglês):

O sophista Zenon, um dos mais despejados humanos em sustentar paradoxos, negava, em certa ocasião, na presença de Diogenes, a existencia do movimento. O que ouvido pelo cynico, entra logo a cabriolar, e a correr diante dos ouvintes.

The sophist Zenon, the most hardy of all men at to sustain some paradoxes, was denied one day, before Diogenes, the existence of the motion. This was put him self immediately to make two or three turns in the session house.

A cavallo dado não se lhe olha para o dente.

A horse baared don't look him the tooth.

¹ Tradução de João Filipe Queiró, professor do Departamento de Matemática, Universidade de Coimbra.

Causa-me algum embaraço admitir que durante muito tempo o livro de Carolino foi a única obra literária portuguesa que conheci. Só há pouco tempo comecei a ler (em traduções) alguns dos muitos excelentes escritores que Portugal produziu. Estou especialmente enamorado de alguns poetas portugueses dos princípios do século XX, incluindo Alberto Caeiro, Álvaro de Campos e Ricardo Reis².

Normalmente, o livro de Carolino teria desaparecido, mas foi aproveitado como exemplo de humor involuntário pelo grande humorista da América, o escritor do século XIX Samuel Langhorne Clemens, mais conhecido pelo seu nome artístico, Mark Twain. «Ninguém pode acrescentar nada ao absurdo deste livro», escreveu ele numa introdução a *English as She is Spoke*. Graças a Twain, a vergonha de Carolino continuou à venda nas livrarias até aos dias de hoje.

E isto traz-nos a «Math as She is Wrote». A lição de Carolino é que corremos sempre um risco ao escrever com pretensões de autoridade sobre temas de que no fundo somos ignorantes. O que se diz pode facilmente parecer idiota a alguém que realmente sabe do assunto. Como jornalista da matemática, estou sempre consciente das oportunidades de dizer disparates. Lorde Northcliffe, o rei dos tablóides ingleses, apresentou o problema deste modo, ao definir jornalismo como «uma profissão que consiste em explicar aos outros o que ela própria não compreende».

Mais à frente darei alguns exemplos, de vários artigos recentes que escrevi sobre assuntos que pessoalmente não entendo. Mas vou começar com algumas observações gerais sobre a actividade de escrever.

De vez em quando perguntam-me qual o segredo da boa escrita. Quem me dera saber! Aprecio a posição de Fernando Pessoa sobre o assunto, no *Livro do Desassossego*: «Gosto de dizer. Direi melhor: gosto de palavrar. As palavras são para mim corpos tocáveis, sereias visíveis, sensualidades incorporadas.» Contudo, em matéria de conselhos úteis, o melhor que consigo dar-vos é que arranjem um revisor.

Não estou a brincar. Ter alguém que leia o que vocês escreveram e corrija os problemas que o texto traga é uma das melhores maneiras de se tornarem bons escritores. Poucas pessoas o fazem, o que pode explicar

² Para os leitores que não estejam familiarizados com a literatura portuguesa, há aqui uma brincadeira: Caeiro, Campos e Reis são todos o mesmo escritor, Fernando Pessoa, que escreveu também sob outros pseudónimos (chamava-lhes «heterónimos»). Falando aproximadamente, Pessoa é o Walt Whitman da poesia portuguesa.

Outro bom conselho está contido num velho provérbio: «Uma figura vale mais que mil palavras». Na realidade, isso não é inteiramente verdade. Como *freelancer*, muitas vezes sou pago pelo espaço que preencho. Os editores não se importam que o preencha com palavras ou com figuras. De facto, até preferem figuras. Ora uma página numa revista como a *Science* leva cerca de mil palavras. E uma figura típica ocupa cerca de um quarto de página. Portanto, façam as contas: uma figura vale aproximadamente 250 palavras.

Mas voltemos à escrita sobre matemática e à maneira «certa» de fazer as coisas. Um dos truques é evitar o calão técnico — as palavras ou as frases que só têm significado para um pequeno círculo de pessoas. Falando por alto, há três tipos de leitores: os matemáticos que já são especialistas no assunto sobre o qual se está a escrever, outros matemáticos que têm pelo menos uma compreensão geral do modo matemático de dizer as coisas e as massas matematicamente incultas. É para este último e vasto grupo que é preciso apontar se se quer popularizar a matemática. Mas usar esse truque pode ajudar, creio, a atingir também o segundo grupo, e mesmo, até certo ponto, o primeiro.

Vejamos um exemplo. Há uns anos (de facto em 1989), eu estava a escrever um artigo sobre um importante desenvolvimento na teoria das curvas elípticas relacionado com uma coisa chamada o grupo de Tate-Shafarevich. Esse desenvolvimento era uma técnica que permitiu aos especialistas em teoria dos números construir exemplos de curvas elípticas cujo grupo de Tate-Shafarevich é finito; a conjectura era (e continua a ser!) que o grupo de Tate-Shafarevich de *qualquer* curva elíptica é finito, mas, antes deste desenvolvimento, havia só um exemplo para o qual se sabia que a conjectura era verdadeira. Isto é muito interessante. Mas que diabo é um grupo de Tate-Shafarevich?

A minha primeira tentativa não teve muito sucesso:

O grupo de Tate-Shafarevich é uma estrutura algébrica que aparece no estudo de pontos racionais de uma curva elíptica.

Note-se que já tinha explicado antes o que é um «ponto racional». Esta explicação era má, e eu sabia-o. Assim, perguntei a um dos especialistas no assunto se me podia ajudar com uma explicação para leigos. Depois de pensar um momento, ele sugeriu o seguinte: o grupo de Tate-Shafarevich é o análogo para curvas elípticas do grupo das classes de ideais para corpos de números algébricos.

Ora bem, isto pode fazer todo o sentido para vocês. Na realidade, até *fazia* sentido para mim, porque a minha pós-graduação foi sobre teoria

dos números e uma vez frequentei uma disciplina de teoria algébrica dos números. Contudo, penso que concordarão que não é provável que seja muito esclarecedor para as massas incultas — ou mesmo para grande parte das massas cultas.

Eis o que eu acabei por escrever:

O grupo de Tate-Shafarevich é o tapete para debaixo do qual os especialistas em teoria dos números varrem todas as dificuldades do cálculo de pontos racionais de curvas elípticas.

Não há dúvida de que isto não é uma definição, mas transmite a ideia que se pretende. Em particular, ajuda a explicar por que razão o tal desenvolvimento é tão importante: se o tapete é finito, só pode haver uma dificuldade finita no cálculo de pontos racionais.

Mesmo quando se escreve para especialistas, uma abordagem suave e não técnica pode ajudar. Em 1952 foi publicado um artigo intitulado «Dois teoremas sobre funções contínuas». Um dos teoremas dizia qualquer coisa como isto:

Sejam f e g funções contínuas de $[0,1]$ para $[0,1]$ tais que $f(0) = g(0)=0$ e $f(1)=g(1)=1$. Então existem funções contínuas u e v de $[0,1]$ para $[0,1]$ tais que $u(0)=v(0)=0$, $u(1)=v(1)=1$, e $f - u = g - v$.

É um bom teorema; no entanto, rapidamente foi esquecido. Quarenta anos depois foi publicado um artigo intitulado «O problema dos alpinistas». O teorema principal desse artigo foi apresentado de forma expositiva da seguinte maneira:

Dois alpinistas podem coordenar as suas subidas em lados opostos de uma montanha de forma a estarem *sempre* à mesma altitude.

Convenientemente analisado, este teorema é precisamente o mesmo que o outro.

Um último exemplo. Há uns anos escrevi um artigo sobre dinâmica dos fluidos. Nesse artigo precisava de explicar o que são as equações de Navier-Stokes. A minha primeira tentativa foi assim:

As equações de Navier-Stokes afirmam a conservação da massa e do momento para uma substância que se move com

velocidade variável em cada ponto do espaço e é governada pelas leis de Newton sobre o movimento.

Isto veio do revisor com o seguinte comentário:

Pouco claro. O que é que entra nas equações e o que é que sai delas?

Acabámos por trocar alguma correspondência sobre o assunto. Muitas vezes o meu revisor manda a sua própria versão do que eu tentei explicar, e muitas vezes ela está completamente errada — ou é mesmo simplesmente disparatada. Isto pode ser frustrante (especialmente se já estamos em cima do prazo), mas de facto é extremamente útil. A tentação é rejeitar os comentários do revisor como delírios de um lunático ou equívocos de um cretino. Afinal a explicação original é perfeitamente clara, matematicamente clara. Mas o facto é que só é perfeitamente clara *se* o leitor for matemático. O revisor não é matemático e, mais pertinentemente, também o leitor pretendido não é matemático. E o revisor também não é um tolo (normalmente); o que ele é, na verdade, é *o leitor mais atento que eu vou ter*. Portanto, se ele não conseguir perceber o que eu escrevi, qual é a probabilidade de eu conseguir transmitir a ideia ao leitor médio?

Se a versão do revisor fosse directamente para a tipografia, as coisas estariam mal. O que faz o processo funcionar é que nós repetimos: eu tento perceber como é que ele conseguiu transformar a minha prosa claríssima numa coisa sem nexos, e depois tento reformular o texto de forma que faça sentido para um não-matemático (num certo caso, num artigo sobre qualquer coisa de teoria dos grafos, consegui finalmente perceber — porque ele me disse — que o revisor pensava que uma «aresta» de um grafo só podia ser uma das linhas entre vértices na periferia do grafo quando o desenhamos). No caso das equações de Navier-Stokes acabámos por chegar a uma formulação que ambos achámos ser razoavelmente clara:

As equações de Navier-Stokes apresentam as leis do movimento de Newton, que normalmente são aplicadas a objectos sólidos, em termos adequados para um fluido, em que a velocidade e a densidade podem variar de ponto para ponto no espaço.

Uma coisa a observar aqui é o aparecimento da palavra «fluido», que estava totalmente ausente do original! Claro que se pode objectar que a velocidade e a densidade também podem variar de ponto para ponto no

espaço no caso de um sólido, mas é melhor ignorar estas questões mais subtis.

Para leitores com formação matemática, é evidente que para explicar as equações de Navier-Stokes só preciso de as escrever:

$$\frac{\partial u_i}{\partial t} + \sum_{j=1}^n u_j \frac{\partial u_i}{\partial x_j} = \nu \Delta u_i - \frac{\partial p}{\partial x_i}$$

$$\nabla \cdot u = 0$$

ν = viscosidade

Está claro?

A propósito destas equações faço referência à seguinte conjectura: se $u(x,0)$ for «suave», então $u(x,t)$ é suave para *todos* os $t \geq 0$.

Se conseguirem provar a verdade (ou a falsidade) desta conjectura, têm uma possibilidade de chegar à fama e à fortuna. Ela é um dos sete problemas matemáticos fundamentais incluídos em 2000 pelo Clay Mathematics Institute num conjunto cuja solução seria premiada com um milhão de dólares. Os outros referem-se à equação de Yang-Mills (em teoria quântica dos campos), à hipótese de Riemann, à conjectura de Poincaré, à conjectura de Hodge, à conjectura de Birch-Swinnerton-Dyer e ao problema P vs. NP em computação teórica. Podem ver-se pormenores na página do instituto em www.claymath.org.

Vou tentar descrever o último, o problema P vs. NP, porque é um bom exemplo de um problema matemático difícil que surgiu por causa dos computadores.

P denota a classe de problemas que podem ser resolvidos em «tempo polinomial». Isto é uma forma técnica de dizer problemas «fáceis». Um exemplo é o problema da busca. Por exemplo, 101 aparece na lista seguinte?

1011, 10, 1, 1101, 11, 1010, 111, 110, 1, 10101, 1010, 110,
1001, 1011, 10100, 110101, 1011, 10001, 11, 10110, 110,
10, 1011

Este problema é de tempo polinomial porque o tempo que demora resolvê-lo é limitado por uma potência do tamanho do problema, que neste caso é o número de algarismos na lista.

NP, por seu turno, representa a classe de problemas que podem ser resolvidos em «tempo polinomial não determinista». Isto é uma forma técnica de dizer que são problemas «difíceis» (embora a própria palavra *difícil* seja usada num sentido técnico). Mais precisamente, um problema NP é uma pergunta de resposta sim/não em que a resposta «sim» pode ser verificada por um cálculo em «tempo polinomial». Um problema NP famoso é a factorização. Por exemplo, pode mostrar-se facilmente que a pergunta «Será que o número 5 493 216 597 273 416 ... 117 tem um factor primo menor que 427 972 417 ... 271?» tem a resposta «sim», se conhecermos — ou adivinharmos — o factor primo apropriado. É aqui que entra o não-determinismo.

Esta definição é curiosamente assimétrica, porque uma resposta «não» a um problema NP pode ou não ser fácil de verificar. A classe dos problemas em que a resposta «não» é fácil de verificar mas «sim» pode ou não sê-lo chama-se co-NP. A questão da primalidade (por exemplo, o número 5 493 216 597 ... 117 é primo?) pertence a essa classe. Na teoria da complexidade computacional há uma hierarquia de classes de complexidade que incluem P, NP e co-NP. Uma pequena parte da hierarquia está na figura 2.

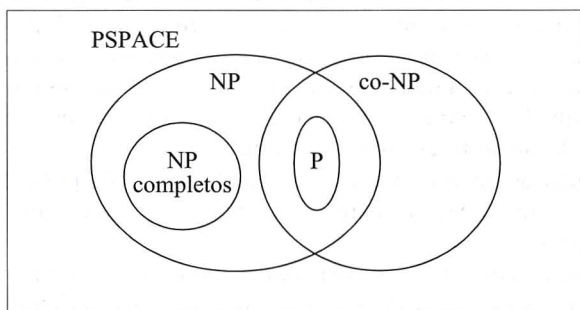


Fig. 2

A classe PSPACE consiste nos problemas computacionais que podem ser resolvidos usando uma quantidade polinomial de memória. A razão por que P está na intersecção de NP com co-NP é óbvia: é *sempre* possível responder a uma pergunta em P (e nem é preciso adivinhar). De facto, foi conjecturado que P é *toda* a intersecção de NP com co-NP. Contudo, a conjectura famosa, uma das sete do Instituto Clay, é que NP é *maior* do que P — isto é, que existem problemas que são mesmo difíceis.

O balão na figura com a legenda «NP completos» precisa de mais algumas explicações. De modo simplificado, trata-se de problemas que podem ser usados como «chave» para todos os outros na classe NP. Um exemplo é o clássico problema do caixeiro-viajante. O que isto significa é que qualquer problema na classe NP — a factorização, por exemplo — pode, com uma quantidade polinomial de trabalho, ser reformulado como um problema do caixeiro-viajante. A consequência disto é que, se se vier a concluir que existe um algoritmo de tempo polinomial para o problema do caixeiro-viajante, *tudo* o que está em NP cai dentro de P. Dito de outra forma, se se acredita (ou se supõe) que é verdadeira a conjectura de que P não é igual a NP, então pertencer à classe dos problemas NP completos é «prova» de que um problema é verdadeiramente difícil.

Há muitas centenas de exemplos de problemas NP completos, alguns dos quais vou referir daqui a pouco. No entanto, começo por mencionar dois problemas que estão em NP mas que não se sabe se são NP completos: a factorização e o problema do isomorfismo de grafos. Na verdade, conjectura-se que estes *não* são NP completos, apesar de se acreditar que ambos estão fora de P.

A lista de problemas NP completos tem vindo a crescer com regularidade desde que o conceito foi definido e os primeiros exemplos dados, no princípio da década de 70. Em 2000, Sorin Istrail, dos Sandia National Laboratories (actualmente na Celera Genomics), acrescentou um exemplo impressionante: o modelo de Ising tridimensional de um vidro de *spin*. Trata-se de um problema em mecânica estatística que surgiu nos anos 20, quando Ernst Ising estudou o modelo «ferromagnético» de dimensão 1. Em geral, o modelo de Ising considera uma rede de átomos i , cada um dos quais pode estar em um de dois estados, $\sigma_i = \pm 1$. Entre átomos adjacentes i e j há um «coeficiente de emparelhamento» J_{ij} para a interacção dos dois átomos. Este coeficiente entra no «hamiltoniano» do sistema,

$$H(\sigma) = -\sum J_{ij} \sigma_i \sigma_j / kT,$$

onde a soma é feita para todos os pares de átomos adjacentes, k é a constante de Boltzmann e T é a temperatura. E temos a importante «função de partição»,

$$Z(T) = \sum_{\sigma} e^{-H(\sigma)} = \sum_E n(E) e^{-E},$$

onde a primeira soma é feita para todas as configurações possíveis do sistema (isto é, atribuições de ± 1 aos átomos individuais) e a segunda soma

é feita para os correspondentes valores do hamiltoniano. Um problema computacional (de facto há vários) é encontrar um «estado básico», uma configuração de energia mínima.

Para o modelo ferromagnético original de Ising isso não é problema: ser ferromagnético significa que todos os J_{ij} são iguais a 1 e o estado básico obtém-se fazendo todos os σ_i tomar o mesmo valor (não importa se 1 se -1). O caso antiferromagnético é igualmente fácil, mas o caso de um vidro de *spin*, em que pode ocorrer uma mistura arbitrária de valores para os coeficientes de emparelhamento, é outra história. Mesmo assim, é trivial para uma rede de dimensão 1, e «fácil» para redes bidimensionais, no sentido de que pertence à classe P (a demonstração para o caso bidimensional, no entanto, não é nada fácil!). Mas mais nada. Istrail mostrou que o problema do vidro de *spin* é NP completo para redes não planares. Isto inclui o problema em dimensão 3, mas também casos bidimensionais em que, além das interações com o vizinho mais próximo, são incluídas interações com os segundos mais próximos.

O modelo de Ising é importante no estudo de transições de fase, que são mudanças bruscas de estado ocorrendo a certa temperatura crítica, como o congelamento da água a 0 graus Celsius. Estranhamente, os especialistas em computação descobriram há pouco tempo alguns fenómenos análogos na sua própria área: certos problemas NP parecem mudar de extremamente fáceis para extremamente difíceis quando o problema atinge um certo tamanho crítico.

O exemplo principal, estudado do ponto de vista das transições de fase pela primeira vez por Scott Kirkpatrick (um físico) e Bart Selman (um especialista em computação), é um problema NP completo conhecido por 3-SAT. Curiosamente, o 3-SAT é o problema NP completo *original*, o primeiro exemplo alguma vez encontrado. O problema diz respeito à «satisfactibilidade» de expressões lógicas como

$$(x_1 \vee x_2 \vee \bar{x}_3) (\bar{x}_1 \vee x_3 \vee x_4) \wedge \dots \wedge (x_{15} \vee \bar{x}_{23} \vee x_{102}),$$

onde há, digamos, M condições, cada uma com três variáveis booleanas, e um total de, digamos, N variáveis booleanas $\{x_1, x_2, \dots, x_N\}$ (os símbolos \wedge e \vee significam «ou» e «e», e \bar{x} significa «não x »). Diz-se que a expressão é satisfazível se alguma atribuição de valores lógicos, Verdadeiro e Falso, às variáveis torna a expressão verdadeira. Mais geralmente, o problema k -SAT diz respeito a expressões em que aparecem k variáveis em cada condição.

Kirkpatrick e Selman estudaram transições de fase no problema k -SAT para $k=3, 4, 5$ e 6 (o caso $k=2$ tem menos interesse porque o problema

2-SAT pertence à classe P). O que eles descobriram foi que, quando o quociente M/N aumenta, a probabilidade de que uma expressão k -SAT gerada aleatoriamente não seja satisfazível passa de perto de 0% para quase 100%, com uma transição assintoticamente brusca num valor que depende de k . Em particular, para $k=3$, o valor crítico é $M/N \approx 4.17$, como indicado na figura 3.

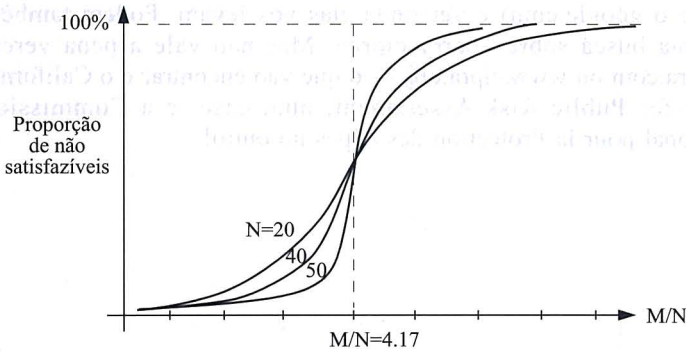


Fig. 3

Além disso, a quantidade de trabalho computacional exigida para dizer se uma expressão é ou não satisfazível parece ter um pico nas vizinhanças do mesmo valor crítico (figura 4).

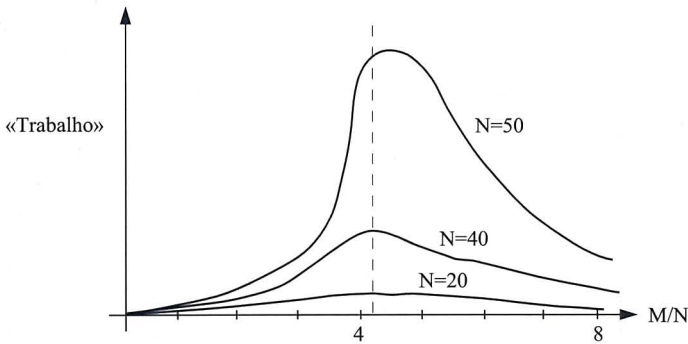


Fig. 4

Resta perceber o que significa exactamente isto, mas muitas pessoas estão actualmente a investigar a «mecânica estatística» de problemas computacionais.

Bom, penso que chega. Disse-vos o que sei sobre a escrita (não muito) e dei-vos alguns exemplos do que acontece quando tentamos explicar assuntos que pessoalmente não entendemos. Espero, todavia, ter-vos proporcionado alguma compreensão de certas coisas e estimulado o vosso interesse nelas. Se quiserem aprender mais, uma boa maneira de o fazer é irem à Internet e introduzirem expressões como «*israil ising np-complete*» ou «*kirkpatrick selman k-sat*» no vosso motor de busca favorito (o meu é o google.com) e ver onde elas vos levam. Podem também tentar uma busca sobre «*barry cipra*». Mas não vale a pena verem www.cipra.com ou www.cipra.org — o que vão encontrar é o California Institute for Public Risk Assessment, num caso, e a Commission International pour la Protection des Alpes no outro!

Nuno Crato
Departamento de Matemática do
Instituto Superior de Economia e Gestão

Algumas experiências de divulgação da matemática na imprensa portuguesa

Os cientistas e os académicos costumam queixar-se da comunicação social e dos jornalistas. Afirmam, e é em grande parte verdade, que a imprensa não cobre as suas actividades, não dedica espaço à divulgação e ao noticiário científico e, quando o faz, comete imprecisões e erros grosseiros. Todos têm histórias para contar. Uns citam disparates crassos que apareceram na imprensa. Outros referem alturas em que foram citados de forma enganadora. A maioria queixa-se da falta de cobertura das actividades de natureza científica, que são preteridas em favor do *fait divers*, dos escândalos do momento ou de temas de importância secundária. É difícil encontrar um cientista ou académico que não critique asperamente os jornalistas.

Do outro lado encontramos jornalistas que se queixam dos académicos. Dizem que estes são incompreensíveis e não fazem um esforço para se tornarem acessíveis. Acusam-nos de serem intransigentes nos pormenores e de não perceberem que a imprensa tem os seus critérios e necessita de transformar o acontecimento em notícia apelativa, relevante e actual, actividade na qual os cientistas são altamente ignorantes.

Esta incompreensão entre os dois campos é velha e já foi discutida até à exaustão. Parece inútil continuar a bater nas mesmas teclas sem

tentar estabelecer pontes entre os dois campos. O que é preciso é procurar ultrapassar incompreensões entre académicos e jornalistas, pois ambos têm a lucrar com uma maior cooperação: os jornalistas com uma informação mais actual, rigorosa e rica e os cientistas com o interesse do público, com a divulgação das suas actividades e com o prestígio da sua profissão.

O QUE OS CIENTISTAS E OS ACADÉMICOS DEVEM SABER SOBRE JORNALISMO

O jornalismo é uma arte com algumas regras e técnicas precisas. O bom jornalismo é uma arte difícil. Para o cientista ver a sua actividade divulgada tem de saber dar ao jornalista informações actualizadas, precisas e aliciantes.

Os critérios que habitualmente se enunciam quando se fala na selecção da notícia são a actualidade, o significado e o interesse público. A actualidade é fácil de entender: ao contrário dos académicos, o público não está interessado em notícias ou artigos intemporais. Lê-se o jornal para saber o que está a acontecer, para pertencer ao grupo de pessoas informadas, para compreender a evolução do mundo, para ter tema de conversa. O segundo critério, do significado, é entendido como a relevância perceptível da peça jornalística, não as suas implicações remotas ou para um grupo restrito. O terceiro critério, do interesse do público, é imperioso: a imprensa não pode sobreviver sem leitores.

É possível escrever uma boa peça jornalística que não satisfaça algum destes critérios, mas não é possível escrever uma que falhe nos três. Ao procurar a actualidade, o cientista ou o divulgador deve fornecer informações recentes, sobre realizações acontecidas num passado muito próximo ou a acontecer num futuro também muito próximo. Muitas vezes, a actualidade não está no cerne da peça jornalística, mas apenas na sua apresentação — o artigo torna-se actual não porque fale de uma descoberta recente, mas porque referencia problemas sentidos na actualidade. Tornar actual uma peça de divulgação científica é uma arte.

O significado pode ser entendido no seu aspecto mais imediato, mas o divulgador deve saber que essa visão é errada. As pessoas não se interessam apenas pelo que as afecta no dia-a-dia, mas também pelo que afecta a sua compreensão do mundo. O público está muitas vezes mais interessado nas novas descobertas sobre a idade do universo do que em novos métodos de tratamento do cancro. Dar significado visível a uma descoberta aparentemente irrelevante é outra arte da divulgação.

O interesse do público é um critério disciplinador do cientista ou do académico que pretende fazer divulgação científica. Não deve ser entendido como uma limitação, mas sim como um desafio. O interesse do público é menos imediatista e limitado do que muitas vezes se supõe. As pessoas estão interessadas em alargar a sua cultura científica, querem saber o que são fractais e o que é o caos, querem compreender a radioactividade, a Internet e as viagens interplanetárias.

O MITO DO DESINTERESSE DO PÚBLICO

Julgamos, muitas vezes, que o público está pouco interessado na actividade científica e cultural e que está mais atento a notícias sensacionalistas, a intrigas e a pormenores secundários da vida. Se isso pode ser verdade em termos gerais, a realidade é que há uma parcela importante do público, talvez sobretudo do público jovem, que está atenta ao noticiário científico e que tenta seguir as descobertas importantes do nosso tempo, assim como as polémicas científicas e culturais.

O que acontece é que há muita gente que se desinteressa daquilo que nós, cientistas e académicos, julgamos importante, pela simples razão de que não sabemos chegar a essas pessoas e não percebemos que a divulgação científica é radicalmente diferente da actividade de ensino académico.

Há sinais bem visíveis do interesse do público. Um bom exemplo é o crescente interesse pela literatura de divulgação científica. Praticamente todas as grandes editoras portuguesas começaram a publicar livros dos melhores autores de divulgação. Publicam-se hoje incomparavelmente mais obras sobre ciência do que há meia dúzia de anos.

Outro factor novo e entusiasmante é a proliferação de planetários, centros de Ciência Viva, museus de ciência e locais similares que aparecem um pouco por todo o país. É certo que esses locais são menos visitados do que muitos dos seus congéneres estrangeiros, nomeadamente norte-americanos. Mas a verdade é que são muito mais frequentados do que muitos museus tradicionais.

Finalmente, a própria imprensa dedica hoje muito mais páginas à ciência do que ainda há poucos anos. Praticamente todos os grandes diários e semanários têm secções de ciência ou de ciência e ambiente.

O que se passa é que o interesse que agora desperta é um interesse novo, muito diferente do que existia há algumas gerações atrás. Quando se organizam conferências e debates sobre reflexões culturais e científicas abstractas, o público é reduzido. Quando se montam telescópios para observar as manchas solares, o público é numeroso. Será que esse facto

é negativo? Será que é triste que as pessoas se desinteressem de debates sobre «a ética da ciência e a ciência da ética», enquanto se interessam por observar as maravilhas do céu através de telescópios?

A resposta a esta questão é decisiva para saber em que ponto nos colocamos quanto ao interesse do público e à divulgação científica que nos interessa fazer. Julgo que o público tem razão e que as discussões vagas de temas fluidos são pouco relevantes, enquanto o conhecimento e a participação directa é educativa e estimulante. É este tipo de divulgação científica que me interessa fazer e por isso julgo que o público nos envia sinais positivos.

O MITO DO DESINTERESSE DA IMPRENSA

É certo que a comunicação social não dá o relevo que julgamos merecido à actividade científica. É certo que a imprensa não relata com rigor muita actividade científica e cultural que gostaríamos de ver mais difundida. É certo que os noticiários dão maior destaque a temas que consideramos menores, em detrimento da actividade académica, científica e de divulgação.

Mas não é menos certo que os jornalistas estão receptivos à informação científica e à ideia de a apresentar. O problema é que não é fácil fazê-lo. Qualquer jornalista precisa de fontes. Ora as fontes nacionais são habitualmente pobres. Basta perguntar o que fazemos nas universidades e laboratórios de investigação para preparar uma informação completa, actualizada, simultaneamente rigorosa e acessível e que a imprensa possa utilizar. O que fazemos é pouco ou nada. Se olharmos para as fontes internacionais, nomeadamente as melhores, veremos como estamos a anos-luz de distância de produzir uma informação que a imprensa possa utilizar. Olhe-se para a página na Internet da NASA, por exemplo, e leiam-se as notícias e comunicados de imprensa que aí aparecem. São documentos completos, informativos, acompanhados de ilustrações magníficas e de diagramas rigorosos. São textos apelativos, bem escritos, com títulos e subtítulos sugestivos, semeados de declarações de cientistas e de informações interessantes. Em anexo vêm habitualmente pormenores complementares, referências de outras fontes e sugestões de pistas para o jornalista alargar a sua notícia.

Parece pois que bastaria à imprensa copiar ou adaptar essas notícias internacionais para produzir um noticiário científico muito mais interessante. Isso é verdade e explica, em parte, a razão por que as páginas dedicadas à ciência têm uma maioria de fontes estrangeiras. Mas os

jornalistas não podem adaptar muitas dessas notícias porque não as sabem traduzir em português. É culpa deles e da sua ignorância? Não, mais uma vez a culpa é nossa, pois os intelectuais e académicos não souberam ainda produzir dicionários completos e actualizados, nomeadamente dos termos científicos. Tentemos traduzir do inglês uma notícia científica. Assim como um matemático teria grandes dificuldades em traduzir termos de biologia e um astrónomo não conseguiria traduzir muita da nomenclatura própria da medicina, também um jornalista receia traduzir certas notícias científicas. Infelizmente, os dicionários que existem são muito incompletos, sobretudo em termos científicos. Será culpa dos jornalistas ou nossa, dos académicos?

A INSUPORTÁVEL VAIDADE DO SER¹

Falemos agora de alguns aspectos importantes a ter em conta pelos que se interessam pela divulgação científica. O primeiro desses aspectos é a clareza da escrita.

Trata-se de um aspecto central e fundamentalíssimo. Os intelectuais literatos conseguem (felizmente cada vez menos!) fazer publicar críticas de cinema ou de literatura que são absolutamente incompreensíveis para o público. Muitas vezes, podem-se ler críticas de cinema que não têm um mínimo de informação e que não conseguem transmitir ao leitor uma única

¹ A expressão «insustentável leveza do ser» é utilizada à exaustão pelos nossos críticos. É uma referência ao título de um livro célebre do escritor checo Milan Kundera, que, em francês, uma das línguas desse autor, e na qual esta obra primeiramente apareceu, em 1984, se intitula *L'insoutenable légèreté de l'être*. Em checo o livro intitula-se *Nesnesitelná lehkost by tí*, o que significa exactamente o mesmo. Em inglês, o título apareceu correctamente traduzido como *The Unbearable Lightness of Being*. Em português, o livro e o filme apareceram com o incompreensível título *A Insustentável Leveza do Ser*. Pois é desse incompreensível título que muitos intelectuais tanto gostam.

O erro de tradução é evidente: «insoutenable», tal como «nesnesitelná» e «unbearable», significa «insuportável» ou «intolerável», e não «insustentável». No português falado e escrito, «insuportável» é aquilo que não se pode aguentar ou a que não se consegue resistir, enquanto «insustentável» significa aquilo que não se pode defender ou não se consegue segurar. A «leveza do ser» pode ser difícil de encarar, mas certamente não é fácil nem difícil de segurar... Quando um intelectual literato reflecte sobre a «insustentável leveza do ser» está provavelmente a falar de... nada.

ideia sobre o filme em causa. O mesmo acontece com muita crítica literária. Por alguma razão estranha — talvez enraizada na ideia absurda de que a cultura se identifica com o inatingível e com as humanidades e não com a ciência —, esses críticos conseguem publicar peças jornalísticas que são autênticos atentados à inteligência do público e ao jornalismo. Mas essas peças incompreensíveis continuam a aparecer.

Mesmo que a imprensa portuguesa nos deixasse utilizar o mesmo estilo hermético nas peças científicas (felizmente, não deixa!), não é isso que queremos com a divulgação científica. É importante perceber que os textos jornalísticos incompreensíveis não denotam cultura nem inteligência por parte dos seus autores, mas apenas sobranceira, pedantismo e ignorância. Quando alguém não sabe explicar um tema em palavras simples, esse alguém quase certamente não o percebe e esconde a sua ignorância e a falta de clareza das suas ideias sob a capa de frases rebuscadas, de uma terminologia hermética e de conceitos abstractos que tão a propósito vêm nesse artigo como em qualquer outro. Basta atentar no jargão de alguns críticos culturais para verificar a sua falta de imaginação, a fluidez dos seus conceitos e a sua ignorância. As frases feitas rodam sempre em torno de uma combinatória de um léxico muito limitado: «o último romântico», «o rigor do olhar», «uma leitura poemática», «os vectores da leitura», «a insustentável leveza do ser», etc.

Devemos rir-nos destas peças herméticas. É saudável rir-nos e prometer a nós mesmos não cair nestes embustes. Não é sinal de cultura ler ou ouvir disparates com ar inteligente. Muito menos é sinal de cultura aproveitar a imprensa para destilar uma pretensa erudição repleta de frases feitas.

COMO SE PODE SER CLARO

Ser claro não é fácil. É difícil, muito difícil. Quando alguém começa conscienciosamente a escrever, é natural que desenvolva uma preocupação extrema pelo rigor do português. Essa preocupação é importante, mas o purismo pode ser um objectivo secundário. Ainda hoje não escrevo «detalhe», pois é um francesismo a que prefiro «pormenor» ou «minúcia». Mas não me choca que esse termo apareça. Utilizo outros termos igualmente impuros, tais como «currículos», pois pouca gente se lembra (ou sabe) que «curricula» é o plural do latim «curriculum». Quer dizer, a clareza parece ser mais importante do que o purismo ortográfico.

O que é verdadeiramente importante é que as ideias sejam claras, apareçam em frases simples, com termos simples e conhecidos de todos,

e que se sigam logicamente de período para período e de parágrafo para parágrafo. No fundo, escrever claramente não é impossível. Trata-se de evitar períodos longos, de incluir apenas uma ideia em cada frase e de escrever com o objectivo de se ser entendido.

Onde o jornalismo científico tem problemas próprios é no uso de jargão e na simplificação das ideias. Não se pode escrever um artigo jornalístico como se escreve um artigo científico especializado. Há, quase sempre, concessões inevitáveis à clareza que chocam o nosso espírito científico. Se é preciso falar em funções contínuas, como se pode traduzir essa ideia num artigo jornalístico? Pode dizer-se que se trata de funções que podem ser desenhadas sem nunca levantar o lápis do papel; no entanto, nós sabemos que essa definição não é correcta. Muitas vezes não é possível incluir definições correctas no texto. Que se pode então fazer?

Não há uma resposta geral, pois o problema deve ser visto caso a caso. O académico tem habitualmente pruridos, pois teme que os seus pares considerem ignorância a não utilização dos termos e das definições mais correctas. Esses receios têm de ser ultrapassados, pois são paralisantes. Não se escreve para se ser bem visto pelos colegas, mas para se ser entendido pelo público. Felizmente, cada vez é maior o número de académicos e cientistas que o compreendem.

Se quisermos estabelecer uma regra geral, deveremos recomendar que se limitem os termos técnicos e que se expliquem os conceitos, que não se apresentem como definições completas o que são apenas intuições e que se deixe entrever a maior complexidade do tema sem defraudar o leitor.

A IMPORTÂNCIA DO ESPÍRITO CIENTÍFICO

Há várias razões pelas quais o jornalismo científico é importante. Uma delas é despertar vocações científicas nos mais jovens. Outra é informar os interessados. Outra, ainda, é despertar o espírito crítico e democrático que são apanágios da ciência.

Há variadas maneiras de escrever peças científicas. Uma das mais interessantes é a que fornece simultaneamente informações factuais actualizadas, desperta interesse por seguir os desenvolvimentos na área, enquadra a notícia no curso histórico da batalha por compreender determinado fenómeno e permite relacionar os factos com vivências do leitor. Fazer tudo isto pode ser muito ambicioso, mas é importante ter em conta estes objectivos.

Ao transmitir informações actualizadas estamos a cumprir um dos requisitos fundamentais do jornalismo, mas estamos também a interessar

o leitor em seguir os desenvolvimentos na área. Ao enquadrar a notícia no curso histórico do desenvolvimento da ciência, estamos a mostrar que a actividade científica não é uma actividade fechada, que as descobertas nunca são definitivas, mas sim passos no avanço do conhecimento científico.

É muito importante que os cidadãos percebam estes aspectos do jornalismo de divulgação científica. Nada é mais irritante do que as notícias de novas descobertas que levam o leitor a concluir que afinal os cientistas apenas se enganam. Este tipo de notícias aparece com grande frequência, sobretudo quando se trata de descobertas clínicas. Alguns jornalistas, estupidamente, adoram dar a entender que os médicos nada sabem. Vêm-se com frequência notícias que contradizem afirmações anteriores: «Agora os médicos descobriram que o açúcar faz bem, quando há tempos diziam que fazia mal. Em quem se pode confiar?» Pois o que o jornalista deveria ser capaz de explicar é que os investigadores tinham razão quando diziam que o açúcar faz mal e voltaram a ter razão quando disseram que o açúcar faz bem. Como é isso possível? Naturalmente, quando se descobriram os malefícios do açúcar, estava-se a contrariar certos hábitos de consumo exagerado. Quando se descobriram os benefícios do açúcar estava-se a dizer que, em certas condições, o incremento desse consumo pode ser benéfico. Quais as condições num e noutro caso? Para que tipo de pessoas? Em que doses? Em que alturas? É isso que cabe aos jornalistas científicos explicar. De outra forma, a sua notícia é irrelevante e mal informada.

Finalmente, ao referir a notícia científica à vivência das pessoas está-se a lançar sementes muito mais fecundas do que as que podem ficar de uma notícia que apenas relata factos distantes. Não se trata apenas de dar exemplos e de transmitir as ideias de forma perceptível, embora isso seja importante². Trata-se de ajudar as pessoas a pensar, no dia-a-dia, com a informação científica que lhes é proporcionada.

Descobriu-se há pouco, por exemplo, que a ilusão visual de a Lua ser maior quando nasce deriva das referências do observador junto ao horizonte. É interessante informar o leitor dessa descoberta. Mas mais

² Numa peça de divulgação científica portuguesa lê-se, por exemplo, «a uma velocidade tal que ligaria Nova Iorque a Los Angeles em menos de quatro minutos». Não seria preferível falar antes de Lisboa e Roma, ou de cidades de localização e distância mais directamente sentida pelos leitores? Só que isso daria trabalho, obrigaria a calcular distâncias e a fazer contas, enquanto aproveitar a informação fornecida por fontes norte-americanas é mais simples.

interessante ainda é sugerir observações que todos podem fazer e que confirmam a origem dessa ilusão óptica. Igualmente, ao provar a conjectura de Kepler sobre o empilhamento de esferas resolveu-se um problema velho de cinco séculos. Isso é interessante, mas ainda mais interessante é mostrar diagramas com o empilhamento óptimo de laranjas e desafiar o leitor a procurar empilhamentos alternativos. Outro exemplo ainda: o Sol passou agora (meados de 2000) por um máximo de actividade. É interessante explicar que esse máximo foi revelado pelas manchas solares. Igualmente interessante é explicar como a descoberta dessas manchas, nos princípios do século XVII, levou a romper com vários conceitos cosmológicos medievais. Também é curioso relatar brevemente o processo de descoberta do ciclo solar. Mas o mais interessante é mostrar fotografias de manchas e explicar ao leitor como pode, ele próprio, observar directamente essas formações solares.

O esforço de divulgação científica pode ajudar o público a compreender melhor o mundo. Poder-se-á desejar mais?

(Página deixada propositadamente em branco)

Kevin Paul Scherer e Marco Avellaneda
Instituto Courant de Ciências Matemáticas
Universidade de Nova Iorque, E.U.A.

Todos por um ... um por todos? Uma análise de componentes principais da dívida de obrigações Brady da América Latina¹ de 1994 a 2000

Utilizaremos a análise de componentes principais (ACP) para estudar as obrigações Brady dos quatro países devedores mais importantes da América Latina: Argentina, Brasil, México e Venezuela. O nosso conjunto de dados abrange um período de cinco anos e meio, com início em Julho de 1994, e consiste nos rendimentos diários (*stripped*) das obrigações ao par e com desconto de cada país. Examinaremos o comportamento dos valores e vectores próprios das matrizes de covariância empíricas calculadas sequencialmente em diferentes períodos. Provaremos que, de um modo geral, existem dois componentes ou factores estatisticamente significativos, que explicam até 90% da variância obtida. O vector próprio associado ao maior valor próprio corresponde à variância atribuída ao risco «regional» (América Latina). A segunda componente sugere

¹ Tradução de Lia Sandra dos Santos. Revisão de João Sousa Andrade, professor da Faculdade de Economia, Universidade de Coimbra.

Esta investigação foi parcialmente financiada pela National Science Foundation.

claramente a existência de um factor de risco de volatilidade associado à dívida venezuelana em relação ao resto da região. Uma análise factorial, dependente do tempo, revela que a importância da variância explicada pelo factor se altera no tempo e que esta variância pode ser interpretada até certo ponto em termos de acontecimentos de mercado. Em particular, analisaremos a relação entre a evolução dos factores ACP e as alterações no mercado que ocorreram durante o período estudado, incluindo o famoso efeito Tequilha, a crise asiática, a desvalorização do rublo e a desvalorização do real.

1. INTRODUÇÃO

Durante a década de 1970 as instituições financeiras ocidentais emprestaram milhares de milhões de dólares aos países em desenvolvimento da América Latina. Os principais devedores, a Argentina, o Brasil, o México e a Venezuela, começaram a faltar ao pagamento do reembolso destas obrigações, desde o início e até cerca de meados da década de 1980. O Departamento do Tesouro dos Estados Unidos, então sob a direcção do secretário Nicholas Brady, apercebeu-se das consequências negativas que o incumprimento generalizado teria no sistema bancário e no próprio desenvolvimento económico da região. No final de 1980 e início de 1990 começaram a ser tomadas medidas destinadas a reestruturar a dívida pendente da América Latina. As instituições financeiras perdoaram parte dos empréstimos pendentes em troca de novas obrigações de longo prazo caucionadas por cupões zero de obrigações *Strip* do Tesouro dos Estados Unidos. Os instrumentos da dívida que surgiram com esta reestruturação são conhecidos, de um modo geral, como «obrigações Brady». Apesar de existirem muitas variantes na dívida Brady da América Latina, incluindo algumas que são caucionadas por outros tipos de garantias, a estrutura da dívida principal, que existe actualmente, tem a forma de obrigações ao par e com desconto. Estruturalmente, as obrigações Brady ao par e com desconto são muito similares, excepto relativamente ao tipo de cupões e ao preço de emissão. As primeiras possuem uma taxa de cupão fixa e foram emitidas ao par, enquanto as últimas têm uma taxa de cupão flutuante LIBOR semianual e foram emitidas com desconto. Os pagamentos dos cupões de umas e outras estão garantidos por um período de 12 a 18 meses até à data de vencimento do pagamento do último cupão. Quando emitidos, ambos são registados com uma data de vencimento final de 30 anos, e, à excepção das obrigações ao par brasileiras, são resgatáveis ao par numa base semianual. As obrigações com desconto são basicamente

obrigações de taxas flutuantes e as obrigações ao par possuem cupões fixos. As suas durações (que decompõe em factores o risco de país e da taxa do juro) são aproximadamente iguais, mas incluiremos na análise feita neste artigo ambos os tipos de obrigações para obtermos uma amostra mais heterogénea.

O objectivo deste trabalho é analisar a volatilidade do mercado da dívida Brady da América Latina, com base em preços históricos, e os *spreads* de mercado. Um assunto fundamental debatido pelos investidores e analistas financeiros — particularmente durante as perturbações do mercado — é até que medida estes países estão correlacionados, ou, dito de outra forma, em que medida os mercados de capital podem «distinguir», em termos de risco de crédito e de risco de liquidez, os diferentes países da região. Será que o valor da dívida «se desloca em conjunto» quando ocorrem choques de mercado? Se a resposta a esta questão fosse dada com base na observação dos dados do rendimento das obrigações Brady ilustrados nas figuras 1a e 1b, abaixo, a resposta seria um sim inequívoco.

No sentido de darmos uma resposta mais precisa e detalhada, faremos uma análise de componentes principais (ACP) (Fluery, 1988) na matriz de covariância empírica construída a partir das alterações diárias nos rendimentos das obrigações ao par e com desconto dos quatro países. De um ponto de vista matemático, a ACP dá dois tipos básicos de informação: as *raízes características* e os *vectores característicos*. Os vectores característicos formam uma base ortogonal que é utilizada para interpretar a relação entre as variáveis (preço e alterações de rendimento, no caso presente). As raízes características medem a contribuição de cada componente para a variância total². A imagem global oferecida pela análise ACP é uma decomposição do risco de mercado Brady em choques não correlacionados, ou factores de volatilidade. Como acontece muitas vezes, a análise de componentes principais é também útil para «reduzir a dimensão», isto é, para identificar alguns dos factores importantes que descrevem as fontes principais de volatilidade do mercado e contribuem para a sua interpretação económica.

² Referimo-nos também aos vectores característicos como componentes.

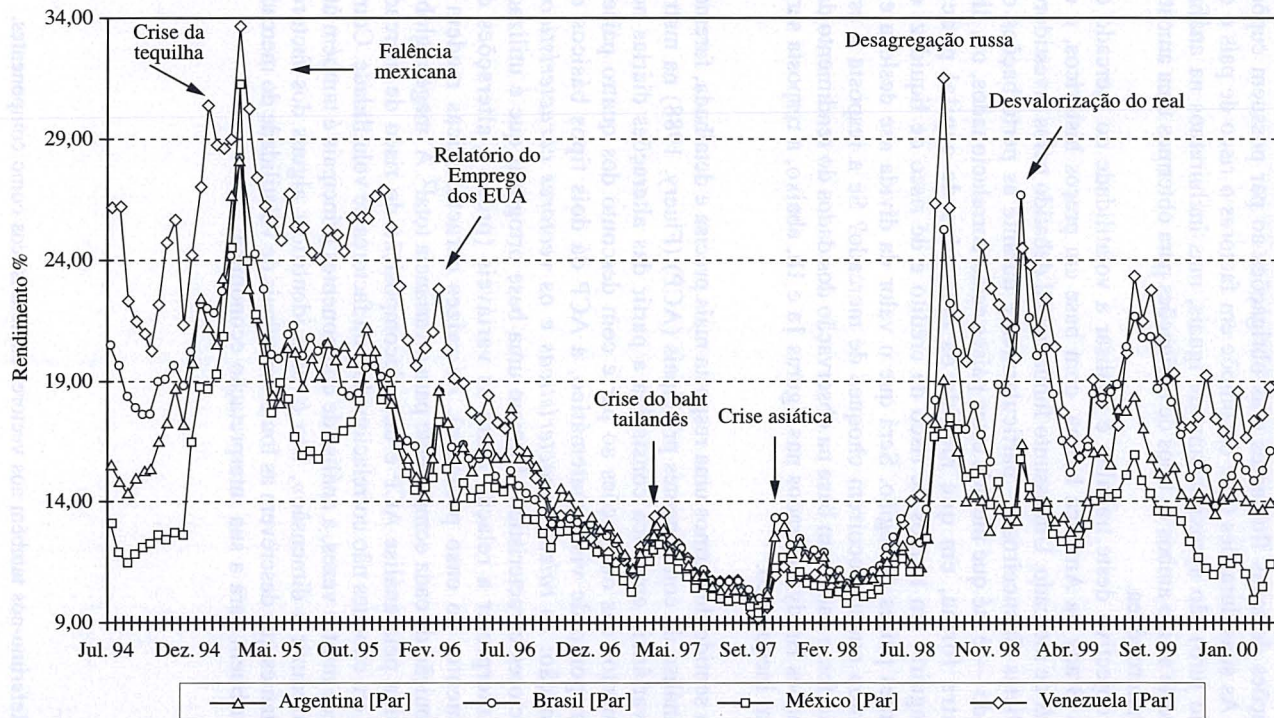


Fig. 1a — Dívida em obrigações Brady ao par: níveis históricos de rendimento por país

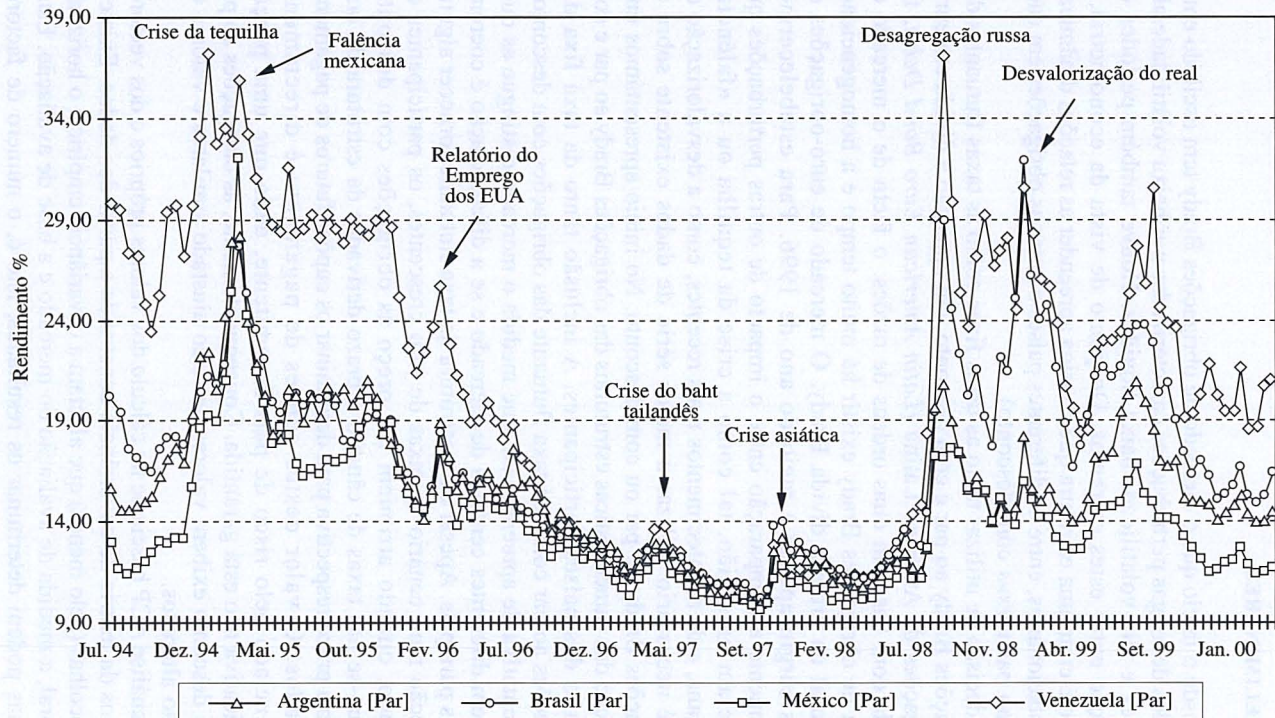


Fig. 1b — Dívida em obrigações Brady com desconto: níveis históricos de rendimento por país

2. PRELIMINARES

Desde o início que o mercado de obrigações Brady tem oscilado entre períodos de longas perturbações caracterizadas por baixa volatilidade/alta liquidez e alta volatilidade/baixa liquidez. Houve também períodos de transição entre estes extremos. Do ponto de vista da econometria, o desafio é criar uma estrutura que consiga apreender as relações dinâmicas, não estacionárias, entre os diferentes países e entre as obrigações em cada país (ao par *versus* com desconto).

A decisão de utilizar tanto as taxas fixas como as taxas flutuantes das obrigações Brady ao par e com desconto, em oposição à dívida em euro-obrigações da América Latina (*Latin American Euro Bond Debt*), foi tomada com base em duas ordens de razões: o facto de o mercado de dívida de obrigações Brady existir há muito tempo e a homogeneidade estrutural da própria dívida Brady. O mercado de euro-obrigações de países surgiu apenas a meio do ano de 1996. Para estabelecermos eficazmente a comparação entre o impacto de outras perturbações que ocorreram na região, tal como a crise da tequilha ou a «falência» mexicana, e de acontecimentos mais recentes, como a desvalorização do real, é necessário utilizar a maior série de dados existente sobre as obrigações Brady ao par ou com desconto. No início apresentámos uma descrição das características estruturais das obrigações Brady ao par e com desconto dos países participantes. A inclusão tanto da taxa fixa das obrigações ao par como da taxa flutuante das obrigações com desconto é uma tentativa de apreender em que medida o mercado distingue as duas taxas em diferentes cenários de mercado e se a diferenciação é coerente de país para país. Apesar de a estrutura da taxa flutuante fornecer alguma protecção num cenário de taxas de juro crescentes, os participantes do mercado, quando atribuem um preço às obrigações com desconto, baseiam-se nas taxas de câmbio a prazo derivadas da estrutura da taxa de juro a prazo respectiva para determinar os cupões futuros de pagamento esperados. O valor destes cupões de pagamento é directamente influenciado pelo risco de país do emitente, dado que uma porção significativa não está garantida. Consequentemente, as obrigações ao par e com desconto exibem valores de risco ajustado similares e valores de duração alterados.

A análise ACP baseia-se no cálculo dos valores próprios e dos vectores próprios da matriz de covariância construída a partir dos dados. Existem três escolhas (pelo menos) que afectam a covariância empírica: o horizonte temporal, a medida de avaliação do mercado e a base de avaliação. Estas escolhas podem determinar os resultados, isto é, o número de factores

explicativos, a estrutura dos vectores próprios e a amplitude dos valores próprios. Dado que a ACP não pode ser determinada de uma forma única, devemos ter em consideração diferentes horizontes de investimento, a duração das situações de perturbação e a nossa própria experiência de mercado para interpretar os dados.

Em particular, a escolha da frequência com que as medições são efectuadas e a duração do período estudado terão um impacto significativo nos resultados obtidos através da ACP. O nosso objectivo é conseguir um certo equilíbrio entre, por um lado, a apreensão de informação suficientemente detalhada e, por outro, a redução do ruído estatístico (Laloux, Bouchaud e Potters, 1999). Limitados a cinco anos e meio de dados, os resultados baseados em observações mensais e semanais mostraram ser dominados pelo ruído e não ter estrutura suficiente. Por esta razão, medimos as alterações de rendimento numa base diária. Após experimentar diferentes tamanhos de «janelas» de negócio, oscilando entre 45 e 180 dias, optámos por utilizar uma «janela» de negócio de 120 dias como o período de observação para cada uma das matrizes de covariância³. A comparação dos resultados das doze matrizes de covariância independentes que abarcam todo o período de observação constitui o ponto fulcral da nossa investigação. Decompondo o mercado pelo procedimento descrito, podemos medir explicitamente e interpretar o impacto dos diferentes acontecimentos económicos que afectaram o mercado de obrigações Brady nos últimos cinco anos.

Relativamente à escolha da base de avaliação, considerámos as variações absolutas do rendimento diário para as obrigações Brady ao par e com desconto da Argentina, Brasil, México e Venezuela (oito variáveis) durante um período de cinco anos e meio, com início em Julho de 1994⁴. Os rendimentos por país ou dos cupões *strip* representam a percepção que o mercado tem da capacidade de um país devedor de cumprir a parte das suas obrigações de dívida não caucionadas, ou não garantidas. No contexto da teoria das opções de crédito, os rendimentos por país medem a probabilidade inferida de não cumprimento e reflectem a percepção do mercado da liquidez da dívida. Verificámos que os resultados obtidos para os dois primeiros componentes da matriz de covariância são essencialmente

³ Os intervalos de tempo são conhecidos informalmente por «janelas». Reflectem o número de dias de transacção activos, não os dias do calendário.

⁴ Os dados relativos ao rendimento por país foram obtidos a partir de uma fonte de dados de mercado privada.

independentes da escolha da base de avaliação. Por outras palavras, a estrutura da covariância é análoga se considerarmos a série das alterações relativas nos rendimentos, no fluxo dos rendimentos ou nos preços.

Com base nesta abordagem da «dinâmica da ACP», pretendemos responder às seguintes questões: Tendo em conta a covariância dos rendimentos, quais são os principais componentes de risco para o mercado? Existirá uma diferença na estrutura da ACP quando comparamos eventos exógenos (por exemplo, a crise asiática e a crise do rublo, em 1998) e eventos endógenos (a «falência» mexicana, em 1995, e a desvalorização do real brasileiro, em 1999)? Em que medida os comportamentos do mercado a longo e a curto prazo estão correlacionados? Qual o impacto do preço do petróleo bruto, se é o que, sobre as obrigações Brady do México e da Venezuela devido à ligação dos *Value Recovery Rights* ou dos *Warrants* ao preço do petróleo? De que forma afectam as diferentes condições de mercado a integridade estrutural da ACP? Poderemos utilizar esta informação para desenvolver um modelo de valor relativo para o mercado Brady?

A técnica ACP tem sido utilizada por diversos autores para analisar dados financeiros. Tanto quanto é do nosso conhecimento, não foi ainda aplicada de forma sistemática à dívida Brady da América Latina. Para uma revisão da literatura sobre a ACP, remetemos o leitor para Litterman e Scheinkman (1992) (estrutura de prazo das taxas de juros nos Estados Unidos), Avellaneda e Zhu (1997) (estrutura de prazo da volatilidade implícita de opções cambiais), Gouriieroux *et al.* (1997), para o Índice CAC 40 e Laloux *et al.* (1999) e para o *Standard and Poor's 500 Index*. Muitos destes estudos utilizam a totalidade de dados disponíveis para calcular a covariância empírica, um procedimento inteiramente compatível com a estacionaridade dos rendimentos⁵. Na nossa opinião, a ACP baseada em janelas de observação alargadas adequa-se melhor à análise de mercados maduros com ciclos económicos estáveis, coerentes com a noção matemática de estacionaridade. Em contraste, o mercado de dívida Brady da América Latina é relativamente jovem; falta-lhe a profundidade dos mercados principais do G7 e conhece períodos esporádicos de alta volatilidade e mudanças de regime. No caso das Bradys da América

⁵ Uma hipótese de estabilidade, ou uma versão «aproximada», podem ser rejeitadas em muitos casos. Por exemplo, Laloux *et al.* rejeitaram a hipótese de estabilidade dos retornos do *S&P 500 Index* baseada nas propriedades espectrais da ACP.

Latina, medir a volatilidade do mercado com base numa única janela de observação alongada iria subestimar a amplitude dos movimentos do mercado e falsear a apreensão das reacções do mercado a eventuais perturbações. Por este motivo, afastamo-nos do paradigma clássico da ACP, considerando janelas de observação sucessivas e as correspondentes matrizes de covariância, e não uma única janela de observação alongada.

3. ANÁLISE DOS RESULTADOS: A ORDEM RESULTANTE DO CAOS

Por razões de comparação, começamos pela análise da ACP «estática», que utiliza um período de observação de cinco anos e meio. A correspondente ACP revela três componentes principais significativos e facilmente identificáveis. A figura 2 representa graficamente a direcção e amplitude para o primeiro, segundo e terceiro componentes, que explicam em média 71%, 13% e 6% do movimento do mercado durante o período em análise. De acordo com esta visão «estática» do mercado, os três componentes principais explicam quase 90% da variância do mercado.

O primeiro componente principal é tal que todos os coeficientes possuem o mesmo sinal. Representa a volatilidade de uma região como um todo. Os coeficientes do vector componente principal medem as alterações relativas típicas no valor ou nos rendimentos das obrigações correspondentes, que surgem em resposta a um choque global na região.

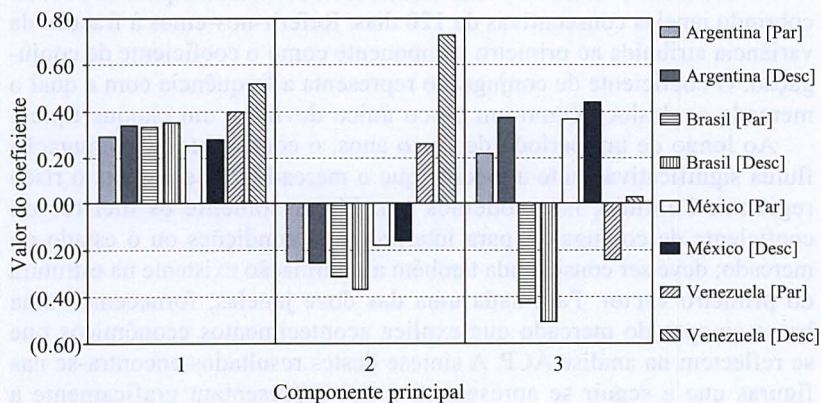


Fig. 2 — Estrutura da componente principal estático

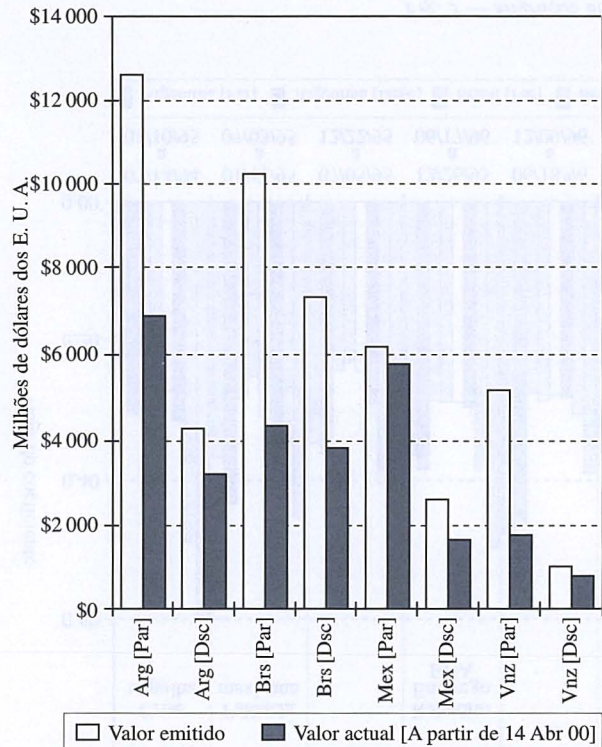
As obrigações ao par e com desconto da Venezuela são as mais voláteis durante movimentos regionais, seguidas pelas do Brasil, Argentina e México. As diferenças de sensibilidades entre as obrigações ao par e com desconto do mesmo país emitente resultam directamente de um prémio de liquidez que o mercado impõe às obrigações com desconto devido à sua dimensão mais reduzida.

É claro, pela observação dos dados, que o segundo componente principal reflecte uma nítida distinção de mercado entre a Venezuela e os outros países da América Latina, a Argentina, o Brasil e o México. Pensamos que este isolamento, assim como a grande sensibilidade da dívida venezuelana aos movimentos regionais, se deve a três factores. Primeiro, o montante inicial e actual da dívida Brady emitida pela Venezuela é consideravelmente menor do que dos outros países. Segundo, tal como a figura 3 indica, o Produto Interno Bruto (PIB) venezuelano é o mais reduzido dos quatro países, e, terceiro, é também o menos diversificado sob o ponto de vista económico, dependendo das receitas do petróleo em cerca de 50%. O terceiro componente principal reflecte sobretudo a relação entre as três maiores economias da América Latina: a Argentina, o Brasil e o México. O Brasil é de longe a maior economia, com um PIB aproximadamente idêntico ao das economias da Argentina e do México em conjunto. Consideramos o segundo e o terceiro componente principal representativos das duas principais relações regionais que existem na América Latina.

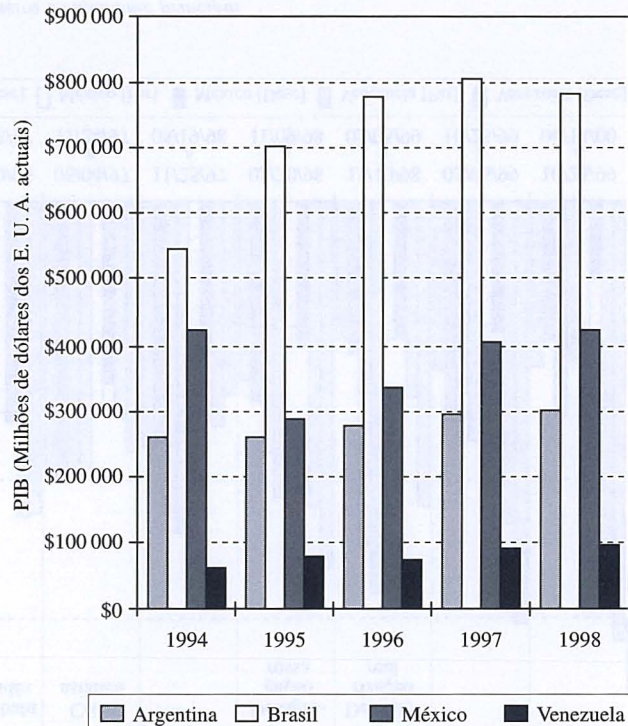
Para fazer face às deficiências inerentes à análise estática, apresentamos em seguida uma perspectiva do mercado de um ponto de vista mais dinâmico.

Para elaborar uma análise dinâmica, efectuamos uma sequência de ACP cobrindo janelas consecutivas de 120 dias. Referir-nos-emos à fracção da variância atribuída ao primeiro componente como o coeficiente de conjugação. O coeficiente de conjugação representa a frequência com a qual o mercado se desloca como um bloco único devido a um choque típico.

Ao longo de um período de cinco anos, o coeficiente de conjugação flutua significativamente à medida que o mercado tem em conta o risco regional. Contudo, não podemos considerar somente os méritos do coeficiente de conjugação para interpretar as condições ou o estado do mercado; deve ser considerada também a informação existente na estrutura do primeiro vector. Para cada uma das doze janelas, fornecemos uma breve sinopse do mercado que explica acontecimentos económicos que se reflectem na análise ACP. A síntese destes resultados encontra-se nas figuras que a seguir se apresentam e que representam graficamente a evolução no tempo das estruturas dos componentes principais e acontecimentos observados durante os cinco anos e meio.



Fonte: Bloomberg



Fonte: Estatística e análise quantitativa IDB com base em estatísticas oficiais dos países membros

Figs. 3 e 4 — Emissão inicial e valor actual em circulação/produto interno bruto

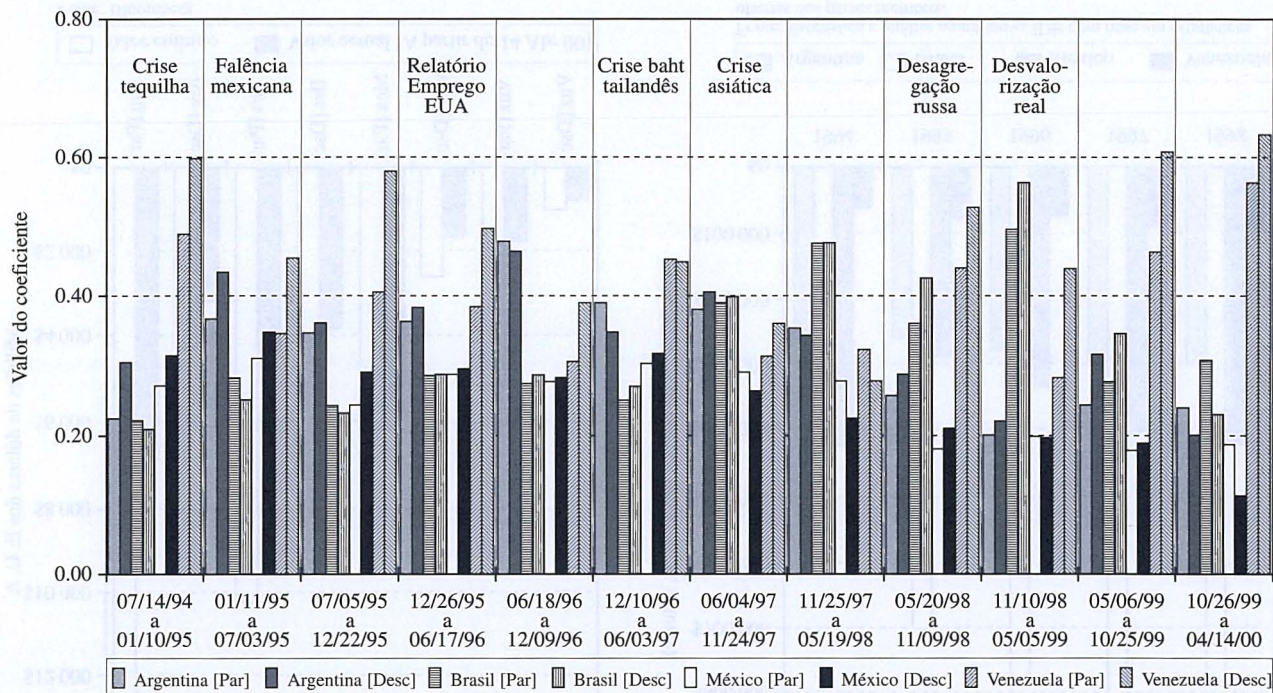


Fig. 5 — Migração do primeiro componente principal

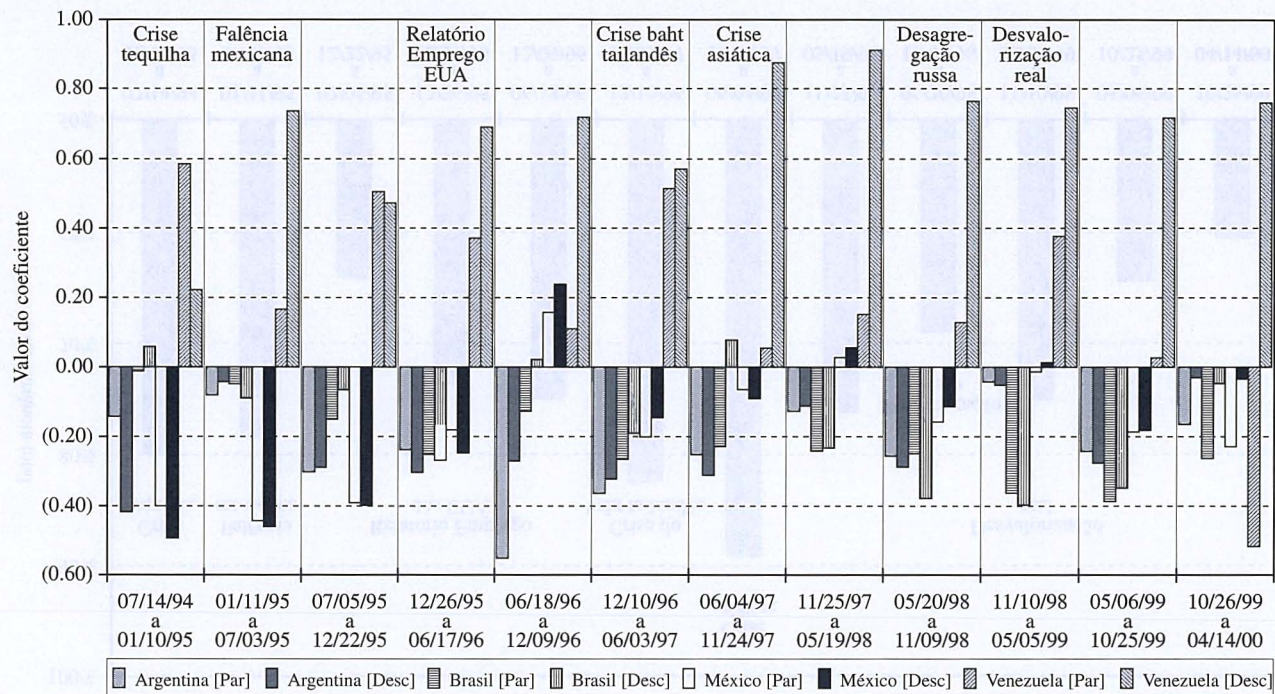


Fig. 6 — Migração do segundo componente principal

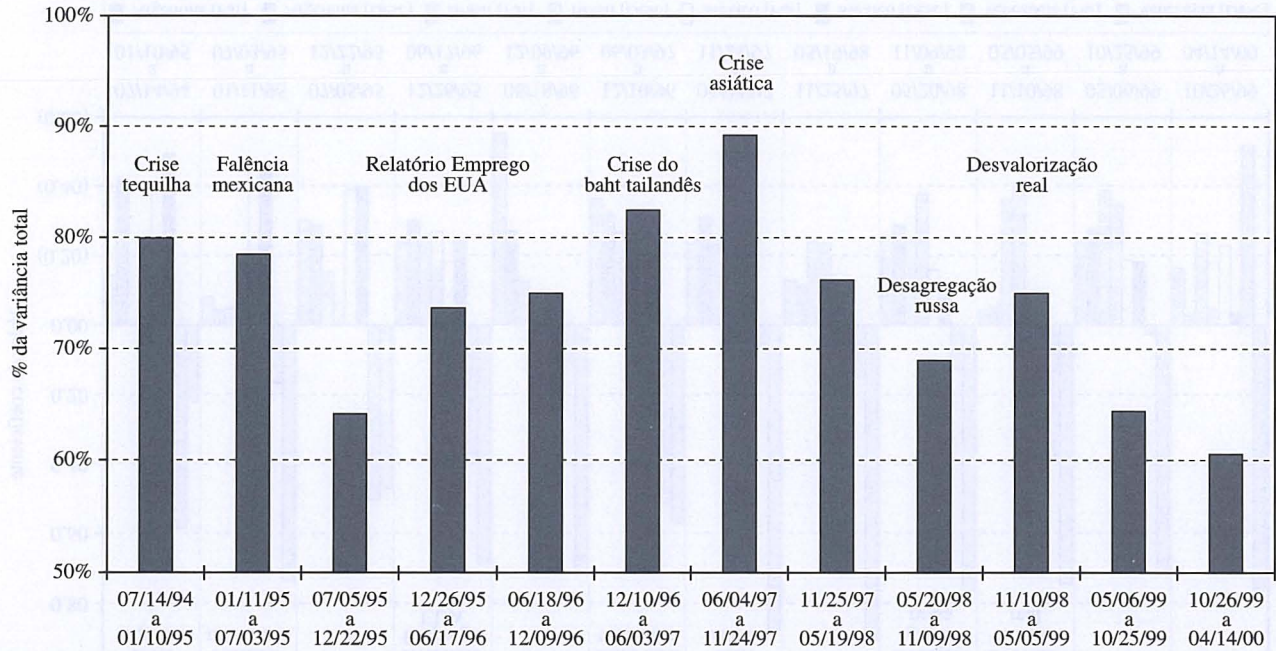


Fig. 7 — Percentagem da variância total do primeiro componente principal [coeficiente de conjugação]

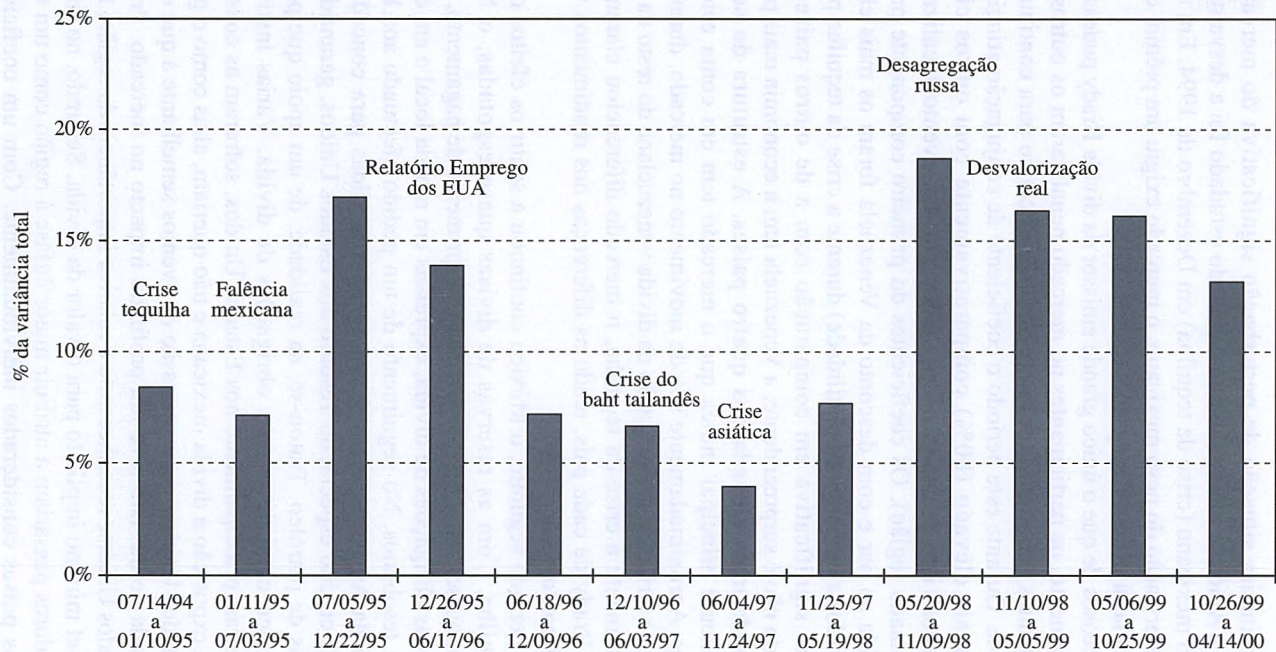


Fig. 8 — Percentagem da variância total do segundo componente principal

A primeira situação de perturbação significativa do mercado que abalou a América Latina durante o período estudado foi a desvalorização do peso mexicano (crise da tequilha) em Dezembro de 1994. Em reacção à desvalorização do peso mexicano, o mercado exigiu um prémio de risco substancial à dívida Brady do México.

Receosos de que o único grande emissor da dívida Brady pudesse faltar ao pagamento, os participantes no mercado penalizaram os outros países participantes, exigindo-lhes um prémio significativo para continuarem a possuí-la. Durante este período o coeficiente de conjugação atingiu o seu valor mais elevado (80%) comparativamente com outros choques endógenos (isto é, choques originados por um evento localizado em determinada região). Os coeficientes do primeiro componente principal da dívida ao par e com desconto da Venezuela foram os mais elevados (associados a uma maior volatilidade) durante a crise da tequilha por uma margem significativa em comparação com a de outros países. Este resultado não é surpreendente; a Venezuela tem a economia mais pequena e a dívida mais elevada dos quatro países. A estrutura do segundo componente principal indica que o mercado tem em conta estes dois factores. Aproximadamente 9% do movimento no mercado, durante este período, é atribuído à segregação da dívida venezuelana do resto da região. Anteriormente à crise da tequilha, o mercado diferenciou claramente a dívida Brady de cada país, tendo as diferenças nos rendimentos variado 75-600 pontos base.

No período seguinte, o México continuou a sentir os efeitos da crise da tequilha. Com as reservas de divisas quase esgotadas, o México confrontou-se com a perspectiva do incumprimento de pagamento, a curto prazo, das obrigações de dívida, expressas em moeda local e em dólares, cetes e tesobonos. No seguimento de um pedido efectuado aos Estados Unidos da América, foram desencadeadas medidas para conceder uma linha de crédito especial do Tesouro dos Estados Unidos, garantida pelas reservas de petróleo. Tratou-se, na realidade, de um apoio que permitiu ao México cumprir as suas obrigações de dívida. Várias instituições financeiras, principalmente nos Estados Unidos, sofreram as consequências da exposição à dívida mexicana e não queriam, aliás como o governo dos Estados Unidos, outra sucessão de eventos semelhante à que ocorreu na década de 80. Isto teve um profundo impacto no mercado. Primeiro, os Estados Unidos, reconhecendo a relativa importância da região, fixaram um nível mínimo implícito para o valor da dívida. Segundo, no futuro os investidores passariam a atribuir maior ênfase à região como um todo do que aos países considerados individualmente. Com um coeficiente de

conjugação quase tão elevado como durante a crise da tequilha, 79%, a dívida Brady deslocou-se para outro nível, à medida que o mercado ia reagindo à crise e às negociações mexicanas de «dólares por petróleo». Estes acontecimentos são responsáveis por uma alteração estrutural significativa no segundo componente principal do período anterior. A relação intra-regional dominante, que explica 7,25% da actividade do mercado, reflecte exclusivamente o movimento da Venezuela *versus* México, enquanto relega a Argentina e o Brasil para a posição de parceiros intra-regionais.

O período de Julho a Dezembro de 1995 foi relativamente benigno na medida em que o mercado continuou a «digerir» os acontecimentos dos seis meses precedentes. O desenvolvimento mais importante neste período foi a emergência do segundo componente principal como factor significativo na explicação do movimento do mercado. É de assinalar que o segundo componente principal sempre existiu, mas apenas com uma influência limitada até este período. Explicando quase 17% da variância do mercado neste período, o segundo componente principal isolou o aumento do *spread* de rendimento relativo da Venezuela *versus* Argentina, México e, em menor grau, obrigações ao par do Brasil.

Durante 1996, período que corresponde à quarta e quinta janelas, assistimos ao início de um período de dois anos e meio de uma procura, sem precedentes históricos, de dívida da América Latina por parte dos investidores. Em Março de 1996 o Relatório do Emprego dos Estados Unidos deu origem ao único choque significativo deste período, dado que o mercado foi apanhado desprevenido quando foram criados 700 000 novos postos de trabalho na economia dos Estados Unidos, alertando para a possibilidade de inflação. A dívida Brady da Argentina teve claramente a reacção mais adversa ao Relatório, devido ao programa de dolarização que de uma forma directa ligava a sua moeda ao dólar americano. Reflectindo quer o choque do emprego quer a confiança geral do investidor na região, o coeficiente de conjugação atingiu 74%, o que representa um aumento de 10% relativamente ao período anterior, e manteve este nível durante a última metade de 1996. Na passagem do quarto para o quinto período assistimos a uma mudança significativa na estrutura do segundo componente principal. Até à segunda metade de 1996, o segundo componente principal incluiu a variância do mercado atribuída à Venezuela *versus* Argentina, Brasil e México. Esta relação alterou-se para passar a reflectir um movimento regional «Norte *versus* Sul» entre, por um lado, a Venezuela e o México e, por outro, a Argentina e o Brasil. Evidentemente, com os primeiros dois países a dependerem

fortemente das receitas do petróleo, torna-se necessário analisar o impacto potencial que os preços do petróleo bruto podem ter na importância desta estrutura. Em nossa opinião, o decréscimo, entre a primeira e a segunda metade de 1996, de 14% para 7% no segundo componente principal pode ser atribuído à alteração estrutural que então ocorreu, isto é, a adopção de uma percepção mais monolítica da região (e dos seus hemisférios).

Em 1997, dois choques exógenos originados na Ásia foram responsáveis pelos dois maiores coeficientes de conjugação observados durante todo o período em análise. Estes factos tornaram-se conhecidos como a crise do baht tailandês e a crise asiática. Caracterizamos estes eventos como fortes choques que tiveram um impacto limitado e negativo no mercado latino das obrigações Brady por um período relativamente curto. A crise do baht tailandês, que ocorreu no final da Primavera de 1997, consolidou ainda mais a importância do coeficiente de conjugação, principalmente devido a uma mudança na variância a partir do terceiro e de outros componentes de ordem superior. Contudo, a proporção da variância do mercado atribuída ao segundo componente principal baixou ligeiramente. A estrutura do segundo componente reverteu para uma volatilidade intra-regional causada, por um lado, pela Venezuela e, por outro lado, pela Argentina, pelo Brasil e pelo México. Após uma calma de seis meses, a crise asiática atingiu a já abalada base de confiança do investidor na região, à medida que as obrigações Brady ao par e com desconto se iam deslocando em conjunto em 90% do período. A presença destes dois choques reforçou a importância do coeficiente de conjugação. Anteriormente à crise, o primeiro componente principal explicava 75% do movimento do mercado, um estado de correlação relativamente elevado, tendo «saltado» então para um estado mais extremo em virtude dos dois choques externos. Isto mostra uma certa «mentalidade de grupo» no investimento da região. Com uma tão significativa proporção do movimento do mercado explicada pelo primeiro componente, o segundo torna-se irrelevante, caindo mesmo abaixo dos 4%.

O primeiro semestre de 1998 será recordado como o da «bonança antes da tempestade». Nunca a participação do investidor no mercado foi tão elevada, tendo os rendimentos das obrigações atingido os níveis mais baixos de sempre. O coeficiente de conjugação regressou a um nível mais modesto, ligeiramente acima dos 75%, e a importância do segundo componente principal duplicou, atingindo aproximadamente 8%. Contudo, o segundo componente principal altera mais uma vez a sua estrutura, isolando a relação da Venezuela *versus* Argentina, sem a participação efectiva do Brasil e do México.

O catalizador que será recordado como o mais dramático acontecimento financeiro da última década foi a falta de pagamento em moeda local da dívida externa por parte da Rússia. O quase colapso da estrutura económica da Rússia foi causado pela desvalorização do rublo em Agosto de 1998, quando a confiança do investidor na capacidade do governo russo de cumprir as suas obrigações foi repentinamente abalada no Verão. No início do Outono a Rússia faltou não só ao pagamento da dívida em moeda local ou GKO⁶, mas também faltou ao da dívida emitida durante a era soviética. Este acontecimento teve repercussões significativas no mundo financeiro e foi conduzindo a um forte «salto para a qualidade», à medida que os investidores se passavam para a dívida do governo dos Estados Unidos e dos governos europeus. A dívida da América Latina, a par e com desconto, que alguns meses antes rendia 8%-9%, negociava-se agora com rendimentos na ordem dos 12%-13%. Particularmente interessante do ponto de vista da ACP é o facto de o coeficiente de conjugação ter baixado no período anterior de 76% para 69% e de o segundo componente principal ter registado um pico de variância de 19%. Por este motivo, apesar do aumento dos *spreads*, o coeficiente de conjugação não atingiu os níveis de 1997. Pensamos que a relação próxima entre o México e os Estados Unidos e a política económica da Argentina, ligada ao dólar dos Estados Unidos, estão subjacentes a este fenómeno. Por exemplo, os dados mostram que os rendimentos para a dívida brasileira e venezuelana aumentaram significativamente, mais do que os rendimentos da Argentina e do México. Durante este período de agitação, a grande importância do segundo componente principal reflectiu-se no prémio de risco pedido para sustentar a dívida venezuelana relativamente ao resto da região. Pensamos que quando o mercado das obrigações Brady esteve sob pressão no Outono de 1998, a inexistência de uma emissão importante de dívida Brady venezuelana e o facto de a sua economia ser mais pequena que a dos outros países foram os factores significativos que conduziram ao aumento da importância do segundo componente.

⁶ *Gosudarstvennyie Kratkosrochnyie Obligatsii*, ou GKO, são obrigações públicas fixadas em rublos russos. Entre Fevereiro de 1998 e Junho de 1998 o governo russo emitiu aproximadamente 65 milhares de milhões (rublos) com uma data de vencimento de 50 anos. O Fundo Monetário Internacional propôs vários planos para reestruturar a dívida por pagar, mas nenhum foi aceite quer pela Rússia quer pelos credores.

A confiança moderada do investidor regressou à região após a desagregação russa, mas foi de curta duração, já que o Brasil sofreu uma desvalorização da sua moeda em Janeiro de 1999. O resultado foi um aumento no coeficiente de conjugação de um valor situado ligeiramente abaixo de 70% para 75%, devido sobretudo a uma mudança da variância do mercado do segundo componente principal, que decresceu de 2,5% para 16%. A dívida brasileira ao par e com desconto registou a maior mudança relativa e absoluta nos coeficientes do primeiro componente principal do período anterior e é a principal responsável pelo aumento da variância do coeficiente de conjugação dos períodos anteriores. O impacto da desvalorização brasileira foi sentido a nível do segundo componente principal. O mercado concentrou-se claramente na relação entre o Brasil e a Venezuela durante este período, tendo percebido, numa base relativa, que o risco inerente a uma tomada de posição na dívida brasileira era maior do que em relação à Venezuela. Isto reflectiu-se no aumento do volume de investimentos no Brasil.

No último semestre de 1999 e no primeiro semestre de 2000 assisteu-se ao decréscimo da importância do coeficiente de conjugação, 64% e 60% respectivamente, níveis que não se tinham registado desde o segundo semestre de 1995. A estabilidade e a confiança do investidor regressaram gradualmente ao mercado após a desvalorização brasileira. Contudo, é evidente que as dívidas dos respectivos países foram sobretudo encaradas a nível individual, em vez de num todo colectivo. O modelo ACP confirma isto com o aparecimento do terceiro componente principal. É difícil, no entanto, tecer conclusões a longo prazo com base na contribuição do terceiro componente principal, uma vez que este não teve um papel significativo anteriormente. Contudo, confirma-se a difusão da variância do mercado desde o primeiro componente principal até ao terceiro e outros componentes de ordem superior. O segundo componente principal regressa à sua forma histórica, mas regista um decréscimo ligeiro durante o segundo semestre de 2000.

Associando os resultados da análise ACP ao comportamento do mercado durante os últimos cinco anos, é possível compreender melhor o mercado de dívida de obrigações Brady da América Latina do que procedendo a uma análise do cenário estático de um só período. Com a informação agora disponível podemos concentrar os nossos esforços na resposta a algumas das questões colocadas anteriormente.

4. ANÁLISE DE COMPONENTES PRINCIPAIS: A ESTRUTURA DESVENDADA

4.1 Coeficiente de conjugação

O coeficiente de conjugação (percentagem da variância explicada pelo primeiro componente) requer uma maior atenção, devido ao seu peso na compreensão do «risco regional». Existem três regimes que medem até que ponto a dívida de obrigações Brady se desloca colectivamente num dado período: fraco, forte e extremo. Para avaliar a força relativa da conjugação, concentramo-nos (i) na direcção ou estado do mercado, (ii) na volatilidade das mudanças no rendimento para a dívida ao par e com desconto (vide figura 9), (iii) no tipo de perturbação e (iv) no nível absoluto do coeficiente de conjugação^{7, 8}.

Um coeficiente de conjugação inferior a 65% indica que a dívida de obrigações Brady da América Latina não está bem correlacionada. Uma conjugação fraca é indicativa do comportamento de «área limite» do mercado, em que se coloca maior ênfase no risco de cada país do que no risco regional. As conjugações fracas surgem normalmente depois de situações de perturbação prolongadas e representam um período durante qual o mercado restabelece o equilíbrio enquanto avalia o impacto do choque.

No extremo oposto do espectro, um regime é considerado extremo quando o coeficiente de conjugação ultrapassa os 80%. Durante regimes de conjugação extremos existe uma relação quase de um para um no movimento direccional da respectiva dívida Brady. Os participantes no mercado subestimam o risco associado ao investimento num país e concentram-se quase exclusivamente no risco agregado da região. Existem dois factores fundamentais que consideramos serem responsáveis pelo facto de o mercado atingir um tão alto nível de correlação: primeiro, um

⁷ A variância ou volatilidade neste contexto é associada às obrigações ao par e com desconto e não deve ser confundida com a variância aplicável a um componente principal específico. A variância neste contexto é usada alternadamente com a volatilidade do mercado. Quanto maior for a variância, maior é a volatilidade nas alterações ou movimentos dos rendimentos dos cupões.

⁸ Existem dois tipos de situações de perturbação ou choques, agudos ou prolongados. Os choques agudos são de curta duração e não têm um impacto significativo no comportamento do mercado a longo prazo. Os choques prolongados significam uma mudança fundamental na direcção do mercado cujo efeito pode influenciá-lo durante um período significativo.

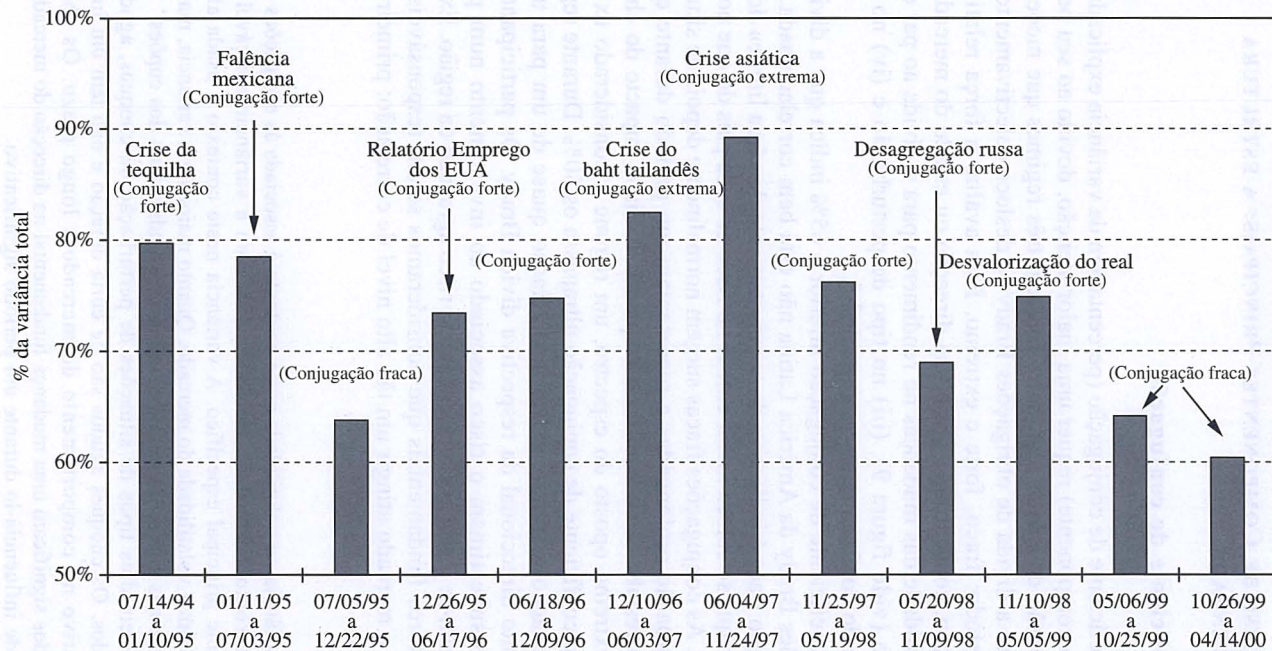


Fig. 9 — Regimes do coeficiente de conjugação

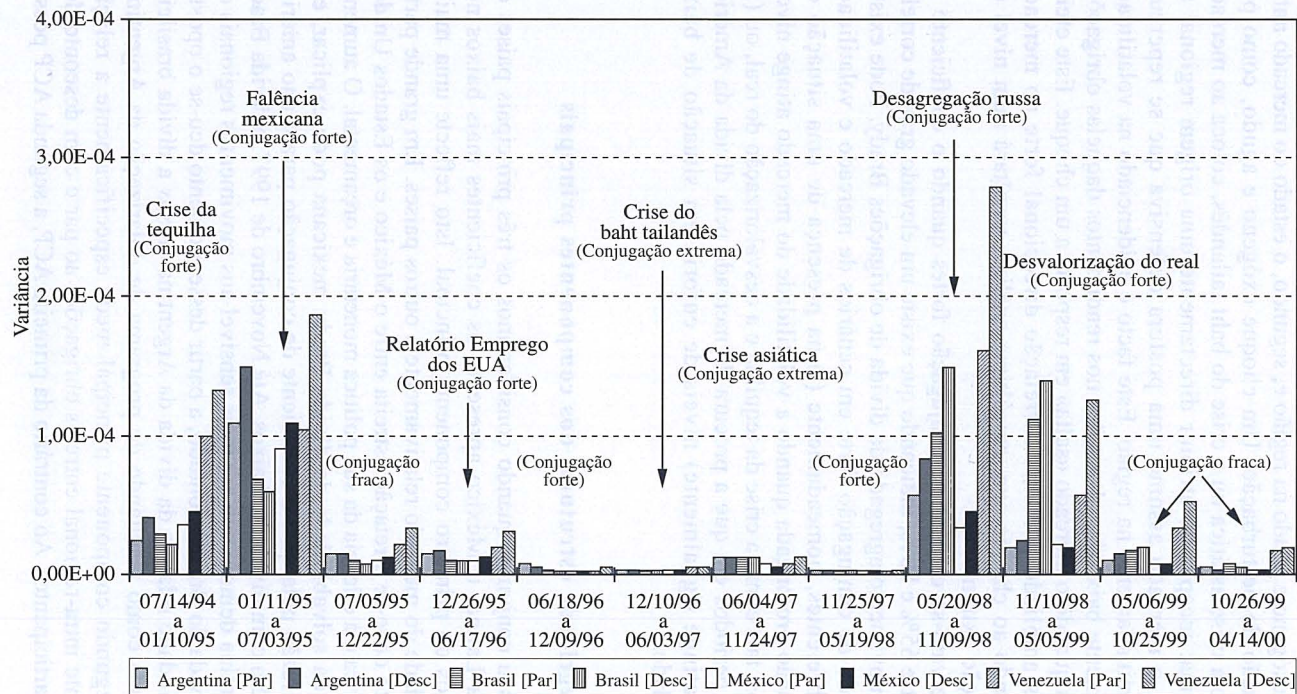


Fig. 10 — Níveis de variância do spread das obrigações ao par e com desconto

choque exógeno e agudo na região e, segundo, o estado do mercado antes da situação de perturbação. Um choque exógeno e agudo, como por exemplo a crise asiática ou a crise do baht tailandês, coloca ao mercado um dilema: incapaz de atribuir directamente uma origem regional ao choque, o mercado assume uma postura defensiva que se repercute indiscriminadamente na região. Este facto é evidenciado na volatilidade relativamente baixa das alterações nos rendimentos daquelas obrigações públicas quando o mercado «salta» em resposta a um choque. Este efeito é depois ampliado por uma correlação direccional forte no mercado conduzindo ao choque que, em última análise, resultará num nível de conjugação extremo.

Observam-se estados de conjugação fortes quando o coeficiente se situa entre 65% e 80%, indicando que existe um elevado grau de correlação no movimento agregado da dívida de obrigações Brady. Pode existir um estado de conjugação forte em cenários de mercado e volatilidade muito diferentes, nomeadamente (i) na presença de uma situação de perturbação prolongada quando a volatilidade do mercado atinge níveis extremos, tais como a crise da tequilha e a desvalorização do real, ou (ii) durante períodos em que a procura de mercado pela dívida da América Latina atinge (igualmente) níveis de euforia em situação de baixa volatilidade.

4.2 Integridade estrutural dos componentes principais

No seu conjunto, quando consideramos os três principais países da América Latina, o México apresenta os coeficientes mais baixos nos elementos do primeiro componente principal. Isto reflecte uma maior estabilidade do mercado relativamente a outros países. Em grande parte, este facto deve-se à relação estreita entre o México e os Estados Unidos da América e à eficácia da sua política monetária e orçamental. O aumento recente da solvabilidade da dívida Brady mexicana pode explicar, em parte, a razão pela qual o coeficiente de conjugação para o ano anterior se mantém com níveis tão baixos. Até Novembro de 1997, a dívida Brady da Argentina demonstrou ser mais sensível aos movimentos regionais do que a dívida do Brasil. Contudo, a partir desse momento deu-se o oposto. A estabilidade relativa da dívida da Argentina *versus* a dívida brasileira é encarada como resultando do programa de dolarização da Argentina.

O segundo componente principal mede especificamente a relação dominante intra-regional entre as obrigações ao par e com desconto dos países participantes. Ao contrário da primeira ACP, a segunda ACP possui

uma estrutura dinâmica e pode variar significativamente de período para período. Teria sido particularmente imprudente que a estrutura do segundo componente principal permanecesse idêntica dados os acontecimentos econômicos que afetaram a região. Como a tabela abaixo indica, as estruturas observadas com mais frequência isolam os movimentos entre a dívida ao par e com desconto da Venezuela relativamente à dívida da Argentina, Brasil e México⁹.

Tabela 1 — Estruturas dinâmicas da segunda ACP

Segunda estrutura da ACP	Número de observações
Venezuela vs. Argentina / México / Brasil	5
Venezuela vs. Argentina / Brasil	2
Venezuela vs. Argentina / México	1
Venezuela vs. Brasil ou México	1 cada

A segunda estrutura da ACP é claramente definida pela Venezuela *versus* Argentina, Brasil e México. Os períodos em que esta relação não se verificou correspondem àquelas janelas de observação em que ocorreram situações de perturbação com origem na América Latina ou durante um período no qual o valor do coeficiente de conjugação é forte ou extremo.

Como salientámos anteriormente, a tabela explica apenas 10 dos 12 períodos. Não incluímos as estruturas da segunda ACP na análise em virtude da instabilidade estatística dos coeficientes. Em particular, os erros assintóticos, que funcionam como estimativas aproximadas da variabilidade dos coeficientes, estiveram muito acima dos limites aceitáveis.

4.3 O contraste entre a ACP dinâmica e estática

Se considerarmos os resultados da ACP estática como representativos da média estatística de várias classes do mercado a longo prazo e os resultados da ACP dinâmica representativos do comportamento do

⁹ O número de observações registado na tabela representa apenas 10 das 12 janelas possíveis de período de negociação. Uma das estruturas dos componentes principais não cumpriu certas exigências de estabilidade e foi, por isso, rejeitada.

mercado a curto prazo, concluímos que existem diferenças significativas entre os dois. O exemplo mais evidente é o comportamento do coeficiente de conjugação. Com base no critério utilizado para determinar o regime do coeficiente de conjugação, descreveríamos o coeficiente de conjugação estático como marginalmente forte, indicando que o mercado se desloca em forte correlação com uma volatilidade moderada. É evidente por esta junção que quer o coeficiente de conjugação quer a volatilidade do mercado, num ambiente dinâmico da ACP, se comportam com uma variabilidade significativa. O primeiro desloca-se entre 60% e 90%, enquanto o último exhibe, em diversas ocasiões, períodos de elevada volatilidade. Nesta base, concluímos que existe um elevado risco se sobrestimarmos ou subestimarmos o comportamento do mercado, no caso de considerarmos apenas a ACP estática. Este facto é ainda reforçado quando estabelecemos comparações envolvendo os coeficientes do primeiro componente principal ou a estrutura da segunda ACP. No que toca a esta última, é evidente que a significância estatística da segunda ACP nem sempre é substancial, nem reflecte de forma coerente uma estrutura estática. Os efeitos das perturbações são abafados em análises estáticas, dado que somos incapazes de avaliar o seu impacto num determinado país ou região como um todo.

Utilizaram-se três factores para modelar o comportamento do mercado numa análise estática; contudo, as limitações de estabilidade ou a falta de coerência estrutural restringiram a ACP dinâmico ao primeiro e segundo componentes principais. O modelo da ACP dinâmico, com dois factores, explicou em média 85% da proporção da variância do mercado, com limites inferiores e superiores de respectivamente 70% e 93%. O valor obtido com uma análise estática ACP, com três factores, é 90%. Consideramos que, apesar de utilizar menos um factor, a ACP dinâmico oferece uma interpretação muito mais realista do comportamento do mercado, com um custo estatístico relativamente baixo.

5. CONCLUSÕES E FUTURA APLICAÇÃO DA ACP

Para procedermos à análise descrita neste artigo baseámo-nos na covariância dos preços/rendimentos de mercado das obrigações Brady da América Latina e analisámos a sua estrutura de forma dinâmico. Foi dada especial atenção às alterações na importância relativa do factor face aos diferentes choques macroeconómicos que afectaram a região nos últimos seis anos. É da nossa opinião que a ACP contribuiu para determinar a variação da importância dos componentes, isto é, do «risco de correlação».

Em especial, a análise levada a cabo mostra claramente a presença constante da Venezuela como fonte principal de risco intra-regional. Por si só, este resultado já deveria ter implicações interessantes em termos das estratégias de afectação de activos a adoptar na região. Seria interessante estudar a dinâmica da futura ACP à medida que novos acontecimentos se vão desenrolando.

A análise de componentes principais encara o mercado em termos da volatilidade e do risco de correlação, em oposição à medida mais convencional de «retorno total». Um resultado estatístico interessante que emerge da análise ACP é o coeficiente de conjugação, ou a percentagem da variância explicada pelo primeiro componente. A observação das variações do coeficiente de conjugação no tempo permite construir uma imagem dinâmica da forma como o mercado encarou o risco «regional» nos últimos seis anos. Uma última aplicação importante da análise factorial, que poderá ser objecto de investigações futuras, é a utilização destes resultados para construir modelos de risco de crédito destinados a avaliar a protecção contra a falta de pagamento, dívidas de obrigações com garantias e outros instrumentos de crédito.

Referências

- Anderson, T. W. (1963) – *Asymptotic theory for principal component analysis. Annals of Mathematical Statistics*, 34, 122-148.
- Avellaneda, M. e Zhu, Y. (1997) – An E-Arch Model for Term Structure of Implied Volatility of FX Options. *International Journal of Theoretical and Applied Finance*.
- Devlin, S. J., Gnanadesikan, R. e Keramidas, E. M. (1987) – A Graphical Procedure for Comparing the Principal Components of Several Covariance Matrices. *Communications in Statistics B-Computation and Simulation*, 16, 161-191.
- Fluery, B. (1988) – *Common Principal Components and Related Multivariate Models*. John Wiley & Sons, Nova Iorque.
- Fuller, W. (1996) – *Introduction to Statistical Time Series*. John Wiley & Sons, Nova Iorque.
- Gourrieroux, Ch., Scaillet, O. e Szafarz, A. (1997) – *Econometrie de la Finance. Ed. Economica*, Paris.
- Krzanowski, W. J. (1984a) – Sensitivity of Principal Components. *Journal of the Royal Statistical Society, série B*, 46, 558-563.
- Laloux, L., Cizeau, C., Bouchaud, J. P., e Potters, M. (1999) – Noise Dressing of Financial Correlation Matrices. *Physical Review Letters* 83, 7, 1467.

- Litterman, R. e Scheinkman, J. (1991) – Common Factors Affecting Bond Returns. *Journal of Fixed Income*, Junho, 55-61.
- Schloerb, F. P., Heyer M. H. (1997) – Application of Principal Component Analysis to Large-Scale Spectral Line Imaging Studies of the Interstellar Medium. *The Astrophysical Journal*, 475, 173-187.
- Strang, Gilbert (1988, 3.^a ed.) – *Linear Algebra and Its Applications*. Saunders College Publishing, Nova Iorque.
- Country Analysis Briefs: Central & South America*.
- <http://www.eia.doe.gov/emeu/cabs/venez.html> (13 de Janeiro de 2000)
- <http://www.eia.doe.gov/emeu/cabs/mexifull.html> (1 de Fevereiro de 2000)

Referências

- Anderson T. W. (1987) – Asymptotic theory for principal component analysis. *Journal of Mathematical Analysis*, 34, 112-132.
- Avramis M e Zou Y (1997) – An E-Asset Model in Latin America and Applied. *Journal of TX Online International Journal of Theoretical and Applied Finance*.
- Devlin, S. J., Gnanadesikan, R., Kruskal, J. B. (1982) – A Graphical Procedure for Comparing the Principal Components of Several Covariance Matrices. *Communication in Statistics-B-Computer and Simulation*, 10, 101-131.
- Fisher, B. (1985) – Common Principal Components and Related Statistics. *Models, John Wiley & Sons, Nova Iorque*.
- Fisher, W. (1986) – Introduction to Statistical Theory. John Wiley & Sons, Nova Iorque.
- Gouriozet, C.J., Scaillet, O. e Sarkar, A. (1997) – *Économique de la Finance*. Ed. Economica, Paris.
- Kruskal, W. J. (1985) – Singularity of Principal Component Analysis of the Royal Statistical Society serie B, 40, 558-565.
- Laloux, L., Garreaud, C., Battisti, J. F. e Pinheiro, M. (1999) – Noise Damping in Financial Correlation Matrices. *Physical Review Letters* 82, 7, 1407.

José M. Bernardo
Universidade de Valência
Espanha

Interpretação de resultados eleitorais: uma análise bayesiana¹

Nos dias que se seguem a actos eleitorais, tanto os meios de comunicação social como os responsáveis políticos concentram parte dos seus discursos num exercício de adivinhação da migração de votos relativamente ao acto eleitoral precedente: quem ganhou votos a quem, quem perdeu votos para quem. Tipicamente, os argumentos apresentados para justificar cada posição são pouco mais do que simples comparações entre os resultados globais de duas eleições, sem qualquer tentativa de efectuar uma análise estatística. Neste artigo apresenta-se uma formalização deste problema, faz-se uma apresentação de resultados básicos do paradigma bayesiano que torna possível resolvê-lo e descreve-se, resumidamente, um método bayesiano de estimação das probabilidades de transição que descrevem a migração de votos, com base numa análise hierárquica dos votos registados em dois actos eleitorais num determinado conjunto de círculos eleitorais. O método será aplicado, a título de ilustração, a resultados eleitorais recentes em Espanha.

¹ Tradução de Paulo A. Oliveira, professor do Departamento de Matemática, Universidade de Coimbra.

1. O PROBLEMA

Quando são anunciados os resultados de uma eleição, tanto os meios de comunicação social como os responsáveis políticos se mostram muito interessados na análise de *modificações* ocorridas nas preferências políticas dos eleitores relativamente ao último acto eleitoral. Para formalizar o problema, representamos por m o número de partidos políticos ou coligações que se apresentaram a votos no acto eleitoral mais recente e por k o número de partidos políticos ou coligações que participaram nas eleições que as precederam (incluindo, em ambos os casos, a abstenção como um partido político). As modificações nas preferências políticas dos eleitores podem ser descritas pelo conjunto de proporções

$$\left\{ p_{i,j}, j=1, \dots, m, \sum_{i=1}^m p_{i,j} = 1 \right\}, \quad i=1, \dots, k,$$

onde cada $P_{i,j}$ representa a proporção de eleitores que votaram, nas últimas eleições, no partido j entre os que haviam votado no partido i na eleição anterior. Isto é, $P_{i,j}$ é a proporção de eleitores que o partido i perdeu em favor do partido j . Em particular, $P_{i,i}$ (a proporção de eleitores do partido i que voltaram a votar no partido i) mede o grau de fidelidade ao partido i por parte do seu eleitorado. O problema em análise consiste na estimação das proporções $P_{i,j}$ a partir dos resultados disponíveis.

Se apenas dispuséssemos dos resultados *globais* de ambos os actos eleitorais, isto é, do número total de votos $a = \{a_1, \dots, a_k\}$ e $b = \{b_1, \dots, b_m\}$ obtidos por cada um dos partidos que se apresentaram a cada eleição, o problema seria naturalmente insolúvel. Mas os resultados eleitorais são *conhecidos para cada um dos círculos eleitorais* em que se divide a área geográfica em análise, mesas de voto, freguesias, concelhos ou distritos, dependendo do nível de detalhe com que se pretende trabalhar. Logo, os dados disponíveis, que representaremos por D , consistem tipicamente no número de votos, $\{a_i, b_i\}$, $i=1, \dots, N$, que cada partido recebeu em cada eleição e em cada um dos N círculos eleitorais. Cada um destes resultados parciais está probabilisticamente relacionado com as proporções desconhecidas, $P_{i,j}$. Portanto, os dados disponíveis D constituem uma amostra de tamanho N de um modelo apropriado, parametrizado pelas proporções $P_{i,j}$ a partir dos quais os parâmetros desconhecidos podem ser estimados. O modelo relevante tem uma estrutura *hierárquica* multinomial complexa e, como consequência, o problema de estimação colocado é não

trivial. Os métodos de análise estatística convencionais (frequentistas) não conseguem fornecer uma resposta adequada para a inferência em modelos hierárquicos, mas o problema pode ser resolvido utilizando uma abordagem bayesiana.

2. ESTATÍSTICA BAYESIANA

Os resultados experimentais ou obtidos por observação consistem geralmente em conjuntos de dados (possivelmente vários) que podem ser descritos formalmente por $D = \{x_1, \dots, x_n\}$, onde as observações x_i gozam de alguma «homogeneidade» (possivelmente multidimensional). Os métodos da estatística permitem obter conclusões acerca da natureza do processo que produziu esses resultados e também acerca do comportamento que esperamos observar em resultados futuros do mesmo processo. Um ponto essencial de *qualquer* análise estatística consiste na especificação de um *modelo probabilístico* que se presume ser capaz de descrever o mecanismo que gerou os dados D em função de um parâmetro (possivelmente multidimensional), $\omega \in \Omega$, por vezes denominado *estado da natureza*, acerca do qual qualquer conhecimento, a existir, é limitado. Todas as conclusões estatísticas subsequentes estão naturalmente condicionadas pelo modelo considerado.

Ao contrário do que é habitual nos vários ramos da matemática, os métodos convencionais da inferência estatística carecem de uma base axiomática; consequentemente, os métodos propostos enfermam, por vezes, de incompatibilidade entre si e a análise de um mesmo conjunto de dados pode levar a resultados incompatíveis quando são aplicados procedimentos diferentes, ainda que intuitivos. A aproximação bayesiana à inferência estatística, em claro contraste com os métodos convencionais, está firmemente apoiada em bases axiomáticas, que fornecem uma estrutura lógica unificadora e garantem a consistência mútua dos diversos procedimentos propostos. Os métodos bayesianos constituem assim um paradigma *completo* para a inferência estatística, uma revolução científica no sentido de Kuhn.

A estatística bayesiana apenas requer a *matemática* da teoria das probabilidades e a *interpretação* da probabilidade que está mais próxima da utilização corrente desta palavra na linguagem diária de não especialistas em matemática: não é por acaso que alguns dos trabalhos mais importantes em estatística bayesiana, como os trabalhos de Laplace, de Finetti ou de Jeffreys, têm o título *Teoria das Probabilidades*. As consequências práticas da adopção do paradigma bayesiano são

bastante vastas. De facto, os métodos bayesianos (i) reduzem a inferência estatística a problemas de teoria das probabilidades, minimizando portanto a necessidade de conceitos novos, e (ii) servem para discernir entre os vários métodos convencionais, indicando uma justificação lógica para alguns (e tornando explícitas as condições em que são válidos) ou provando a inconsistência lógica de outros.

A principal consequência desta fundamentação é a *necessidade* matemática de descrever todas as incertezas num problema à custa de distribuições de probabilidade. Em particular, os parâmetros desconhecidos em modelos probabilísticos *devem* ser descritos por uma distribuição de probabilidade conjunta que permita explicitar a informação disponível acerca dos seus valores; este ponto é, na maioria das vezes, identificado como o princípio mais característico da análise bayesiana. Note-se que, no âmbito do paradigma bayesiano, e em claro contraste com a estatística convencional, *os parâmetros são tratados como variáveis aleatórias*. Este tratamento não é apenas uma descrição da sua variabilidade (os parâmetros são, tipicamente, quantidades fixas mas desconhecidas), mas uma descrição do *grau de incerteza* acerca dos seus valores.

Um caso particular importante é aquele em que não existe nenhuma informação prévia relevante acerca dos parâmetros ou aquele em que a informação prévia existente é subjectiva e se pretende uma análise «objectiva», unicamente baseada em hipóteses acerca do modelo e bem fundamentadas nos dados existentes. Esta questão é resolvida pela *análise de referência*, que utiliza conceitos teóricos de informação para deduzir as distribuições de referência *a posteriori*, definidas para incluir conclusões inferenciais acerca das quantidades em estudo, baseando-se apenas no modelo que foi assumido e nos dados observados.

2.1. A probabilidade como medida de incerteza

A estatística bayesiana utiliza a palavra *probabilidade* no mesmo sentido em que esta palavra é utilizada na linguagem corrente, como uma *medida da incerteza*, condicionada pelo conhecimento corrente, que associamos à ocorrência de um acontecimento particular em função da informação disponível e das hipóteses formuladas no modelo aceite. Assim, a probabilidade de um acontecimento E sob as condições C será denotada por $\Pr(E|C)$ e representa a medida do grau de confiança que depositamos na ocorrência de E sob as condições C . Por vezes a probabilidade de um acontecimento sob determinadas condições pode ser associada à frequência relativa de acontecimentos «similares» sob condições

«similares». No entanto, esta associação nem sempre é possível. É importante notar que uma probabilidade é *sempre* uma função de dois argumentos, o acontecimento E , cujo grau de incerteza está a ser medido, e as condições C , sob as quais a medição do grau de incerteza é efectuada; probabilidades «em termos absolutos» não existem. Nas aplicações mais correntes deste formalismo, estamos interessados na probabilidade de um certo acontecimento E , conhecidos alguns dados D , as hipóteses H que assumimos acerca do mecanismo que gerou os dados, e o *conhecimento* K relevante para o contexto, que possa existir. Assim, $\Pr(E|D,H,K)$ deve ser interpretada como a medida do grau de confiança que depositamos na ocorrência do acontecimento E , tendo em conta os dados D , as hipóteses H e o conhecimento específico K disponível, como uma medida da «verosimilhança» do acontecimento E sob aquelas condições. Para ilustrar os vários conceitos, descreve-se a seguir um exemplo simples.

Estimação de uma proporção

Efectua-se uma sondagem para estimar a proporção θ de indivíduos na população que possuem determinada propriedade. Escolhe-se uma amostra aleatória constituída por n elementos, r dos quais apresentam a referida propriedade. Estaremos então interessados em utilizar estes resultados da amostragem para definir uma região em $[0,1]$ onde seja plausível que esteja o verdadeiro valor (desconhecido) de θ ; esta informação será descrita por *probabilidades* da forma $\Pr(a < \theta < b | r,n,H,K)$, a medida do grau de incerteza, em função da informação disponível, acerca do acontecimento « θ pertence ao intervalo (a,b) » dada a informação que se obtém dos dados (r,n) , das hipóteses H consideradas para o mecanismo que gerou a amostra (uma amostra aleatória de tamanho n de provas de Bernoulli) e ainda de qualquer outra informação relevante K conhecida acerca dos valores θ . Por exemplo, uma sondagem sobre as preferências políticas de 1500 cidadãos indicou 720 respostas favoráveis à implementação de determinada medida legislativa. Se concluíssemos que $\Pr(\theta < 0.5 | 720,1500, H,K) = 0.933$, isso indicaria uma probabilidade de cerca de 93% de perder um referendo para tentar legitimar essa medida legislativa. De forma análoga, se efectuarmos uma pesquisa para encontrar o número de pessoas infectadas com determinada doença, sobre 100 pessoas, e não encontrarmos nenhuma infectada, poderíamos concluir que $\Pr(\theta < 0.01 | 0,100, H,K) = 0,844$, isto é, que a probabilidade de a proporção de infectados na população ser inferior a 1% é de cerca de 84%.

2.2. O paradigma bayesiano

A análise estatística de um conjunto D de dados inicia-se habitualmente por uma avaliação descritiva informal, que é utilizada para sugerir um primeiro *modelo probabilístico* formal $\{p(D|\omega), \omega \in \Omega\}$ que se supõe representar, para algum valor (desconhecido) de ω , o mecanismo probabilístico que gerou os dados observados D . É possível introduzir argumentação axiomática para justificar a necessidade lógica de uma distribuição de probabilidade *a priori* $p(\omega|K)$ sobre o espaço de parâmetros Ω , descrevendo a informação disponível K acerca do valor de ω antes da observação de quaisquer dados (veja-se, por exemplo, Bernardo e Smith, 1994, cap. 2). Deduz-se então, utilizando argumentos da teoria das probabilidades convencionais, que, se o modelo probabilístico está correcto, toda a informação disponível acerca do valor de ω após a observação dos dados D está contida na distribuição de probabilidade *a posteriori* cuja densidade de probabilidade, $p(\omega|D,H,K)$, pode ser obtida de forma imediata a partir do teorema de Bayes como

$$p(\omega|D,H,K) \propto p(D|\omega,A) p(\omega|K)$$

onde H representa o conjunto de hipóteses colocadas pelo modelo de probabilidade assumido. É a utilização sistemática do teorema de Bayes, para incorporar a informação obtida a partir de dados observados, que justifica o adjectivo *bayesiano*, que habitualmente se atribui a este paradigma. É óbvio, a partir do teorema de Bayes, que a qualquer valor de ω para o qual a densidade *a priori* seja nula corresponderá o valor nulo na densidade *a posteriori*. Logo, é correntemente assumido que (restringindo, se necessário, o *espaço dos parâmetros* Ω) as distribuições *a priori* são *estritamente positivas* (como diz Savage, mantenhamos um espírito aberto ou, pelo menos, entreaberto). Para simplificar a apresentação, a indicação explícita a A e K será omitida das notações a utilizar, mas o facto de que *todas* as afirmações acerca de ω são condicionais, *também* relativamente às hipóteses e à informação prévia disponível, deverá estar sempre presente no espírito do leitor.

Do ponto de vista bayesiano, o resultado final de um problema de inferência acerca de *qualquer* quantidade desconhecida é a sua distribuição de probabilidade *a posteriori*. Isto é, conhecendo os dados observados D e as condições C , *tudo* o que se pode dizer acerca de qualquer função ω que dependa dos parâmetros que descrevem o modelo está contido na distribuição *a posteriori* $p(\omega|D,C)$.

Para tornar a interpretação das conclusões mais fácil para o destinatário final da análise estatística, é muitas vezes conveniente resumir a

informação contida na distribuição *a posteriori*, exibindo intervalos aos quais, à luz dos dados observados, é razoável supor que o verdadeiro valor da quantidade em estudo pertence. Este procedimento está relacionado com o conceito frequentista de intervalo de confiança, mas é conceptualmente bastante diferente: um *intervalo de confiança* apenas permite indicar que, *se o procedimento for repetido uma infinidade de vezes*, então o correspondente intervalo conterá o verdadeiro valor do parâmetro numa certa proporção (previamente definida) dos casos. Contudo, do ponto de vista frequentista nada pode ser dito acerca da probabilidade de o verdadeiro valor do parâmetro estar no intervalo apresentado *conhecidos os dados disponíveis*, uma probabilidade que aparece naturalmente a partir da distribuição *a posteriori* quando se utiliza o paradigma bayesiano. Ilustremos estes conceitos com um exemplo.

Estimação de uma proporção (*continuação*). Suponhamos que os dados observados consistem em n observações de Bernoulli com parâmetro θ , das quais r são positivas. Assim, $p(D | \theta, n) = \theta^r (1-\theta)^{n-r}$, e suponhamos que o conhecimento prévio dos valores que θ assume é descrito por uma distribuição beta com parâmetros α e β , $Be(\theta | \alpha, \beta)$. Neste caso obtemos $p(\theta | \alpha, \beta) \propto \theta^{\alpha-1} (1-\theta)^{\beta-1}$. Utilizando agora o teorema de Bayes, a densidade *a posteriori* associada ao parâmetro θ é

$$p(\theta | r, n, \alpha, \beta) \propto \theta^r (1-\theta)^{n-r} \theta^{\alpha-1} (1-\theta)^{\beta-1} \propto \theta^{r+\alpha-1} (1-\theta)^{n-r+\beta-1},$$

que identifica uma distribuição $Be(\theta | r + \alpha, n - r + \beta)$.

Suponhamos que θ representa a proporção de eleitores que votam favoravelmente uma determinada medida legislativa num referendo. A experiência de sondagens anteriores e a informação disponível acerca de θ poderá tornar razoável a suposição de a distribuição *a priori* do parâmetro ser descrita pela distribuição $Be(\theta | 50, 50)$. Esta distribuição prevê que é igualmente provável a aprovação e a reprovação da medida legislativa que está a ser referendada e que a probabilidade de a aprovação ou reprovação da medida legislativa obter menos de 60% dos votos é igual a 0.95. Uma sondagem aleatória é efectuada a 1500 eleitores, registando-se 720 eleitores que aprovam a medida legislativa em causa. A utilização das distribuições descritas atrás indica que a distribuição *a posteriori* de θ é $Be(\theta | 730, 790)$. As distribuições *a priori* e *a posteriori* estão representadas na figura 1. É de realçar que, como se poderia esperar, o efeito dos dados consiste numa redução drástica da incerteza inicial que o modelo atribuía ao valor de θ e, portanto, ao resultado do referendo. Mais precisamente, $\Pr(\theta < 0.5 | 720, 1500, H) = 0.933$ (que corresponde à área a sombreado na figura 1), pelo que, após a inclusão dos resultados

da sondagem no modelo definido, a probabilidade de o referendo ser desfavorável à medida legislativa em causa deverá ser de cerca de 93%.

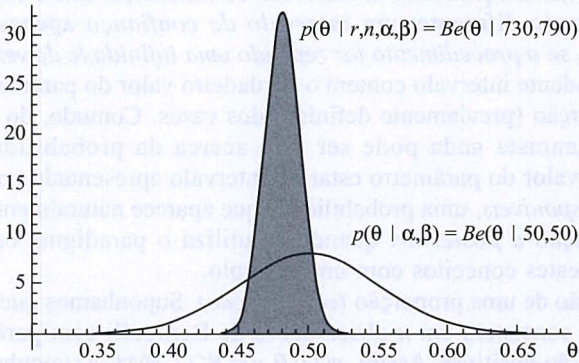


Fig. 1 — Densidades das distribuições a priori e a posteriori para a proporção θ de eleitores que votariam positivamente num referendo

A situação *a priori* para este problema em que não há informação inicial disponível consiste em tomar $\pi(\theta) \propto \theta^{-1/2} (1-\theta)^{-1/2}$, isto é, uma distribuição beta $Be(\theta | 0.5, 0.5)$. A correspondente distribuição *a posteriori* é $Be(\theta | r+0.5, n-r+0.5)$, onde, mais uma vez, r representa o número de respostas positivas.

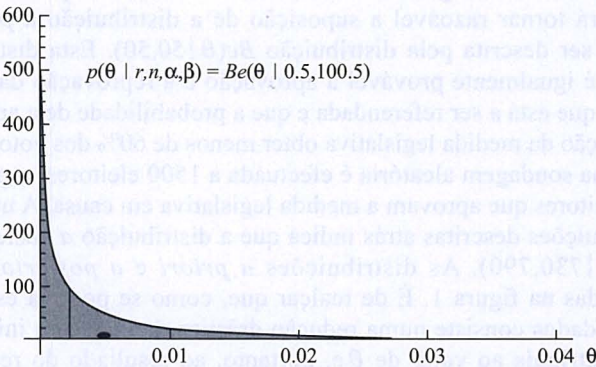


Fig. 2 — Distribuição a posteriori para a proporção de indivíduos infectados numa população, dada a observação de indivíduos dos quais nenhum estava infectado

Suponhamos agora que em 100 indivíduos escolhidos ao acaso foi testada a presença de uma infecção, tendo-se verificado que nenhum dos indivíduos escolhidos estava infectado, isto é, $r = 0$. A figura 2 mostra que a distribuição *a posteriori* para a proporção θ de indivíduos na população inteira que estão infectados é $Be(\theta | 0.5, 100.5)$. Logo, baseando-nos nos resultados observados, podemos concluir que a proporção de indivíduos infectados é seguramente inferior a 5% (pois, para a distribuição *a posteriori*, a probabilidade do acontecimento $\{\theta > 0.05\}$ é 0.001), que θ é inferior a 0.01 com probabilidade 0.844 (a área da região a sombreado na figura 2), que é igualmente provável que θ esteja acima ou esteja abaixo de 0.23% (a mediana da distribuição *a posteriori* está representada por um traço vertical na figura 2) e que um indivíduo escolhido ao acaso naquela população tem uma probabilidade 0.005 (a média da distribuição *a posteriori* está representada por um círculo negro na figura 2) de estar infectado, já que $\Pr(x=1 | r, n) = E[\theta | r, n] = 0,005$. Se quisermos uma aproximação numérica para o valor de θ , deveríamos escolher a mediana 0.0023 ou 0.23%. Note-se que a solução convencional para este problema, baseada no comportamento assintótico do estimador da máxima verosimilhança, é

$$\hat{\theta} = \frac{r}{n} = 0,$$

independentemente do valor de n . Esta conclusão, de a proporção de infectados ser nula, não faz sentido neste cenário.

Nos últimos anos foram publicados vários livros sobre estatística bayesiana, que recomendamos ao leitor mais interessado. Destacamos alguns, referindo-os por ordem cronológica de publicação: Berger (1985), Lee (1989), Bernardo e Smith (1994), O'Hagan (1994) e Gelman *et al.* (1995). Para uma introdução simples à estatística bayesiana veja-se Bernardo (2001), incluído na *Encyclopedia of Life Support Systems*.

A estatística bayesiana permite a análise de modelos probabilísticos complexos, como os de estrutura hierárquica, que estão fora do alcance dos métodos convencionais. O modelo probabilístico apropriado para a análise das transferências de votos entre actos eleitorais é, precisamente, um modelo hierárquico.

3. A SOLUÇÃO PARA O PROBLEMA PROPOSTO

O modelo probabilístico considerado para a estimação da redistribuição de votos entre dois actos eleitorais consecutivos baseia-se na *permutabilidade parcial* dos eleitores que votaram na eleição prévia. Este princípio faz aparecer (veja-se Bernardo e Smith, 1994, cap. 4) um produto de k distribuições multinomiais, onde k é o número de partidos

políticos que se apresentaram à eleição anterior, o que implica que se considerem como parâmetros *simultaneamente* os $P_{i,j}$ que representam a proporção, de entre os votantes que na eleição anterior haviam votado no partido i , de votantes no partido j na presente eleição, e a matriz de variáveis latentes $n_{i,j,u}$ que descreve o número (desconhecido) de eleitores que alteraram o seu voto de i para j em cada um dos círculos eleitorais u , todas sujeitas às condições (conhecidas) impostas pelo número total de votos obtido por cada partido em cada um dos círculos eleitorais u , que são também conhecidos.

Tabela 1 — Cidade de Valência. Resultados por círculo eleitoral nas eleições de 1995

Círculo	PP	PSOE	EU	UV	Bloc	Outros	Abst.
1	11 474	2 732	1 628	1 491	223	166	4 257
2	21 025	4 584	2 835	2 547	298	227	6 823
3	21 591	5 840	3 425	3 282	377	258	8 277
4	8 267	4 579	2 942	1 704	212	210	682
5	13 911	7 549	4 303	2 927	310	313	11 242
6	14 223	2 748	1 649	1 341	232	156	5 284
7	13 148	9 067	4 604	3 126	227	279	11 675
8	14 127	7 411	5 195	3 294	281	323	15 404
9	12 088	8 593	4 862	3 347	236	304	12 981
10	18 010	9 941	5 985	4 907	414	382	18 128
11	11 227	11 500	5 239	4 635	334	304	16 643
12	11 317	8 906	4 559	2 756	280	242	14 328
13	11 000	6 387	3 944	2 061	339	222	10 453
14	7 628	4 346	2 800	1 581	232	162	7 849
15	9 174	8 404	4 330	2 432	221	238	11 736
16	7 432	7 056	3 590	2 013	179	251	12 498
17	1 568	1 019	373	877	70	16	1 306
18	2 337	2 693	1 021	654	44	78	3 496
19	3 754	2 793	1 287	2 593	113	70	5 624

Para aproximar as variáveis latentes utiliza-se programação linear. Obtém-se assim um perfil de verosimilhanças que depende apenas dos parâmetros $P_{i,j}$ que nos interessam. Utilizando uma distribuição *a priori* adequada (Bernardo, 1979; Berger e Bernardo, 1992; Bernardo e Ramón, 1998), é possível obter uma distribuição *a posteriori* para os $P_{i,j}$ e, em particular, os seus valores médios *a posteriori* (que constituem as aproximações pretendidas) e ainda os desvios padrões, também *a posteriori* (que descrevem os erros previsíveis em cada estimativa).

É importante realçar que a metodologia delineada depende *exclusivamente* dos resultados eleitorais registados. Em particular, não requer nenhum conhecimento adicional proveniente, por exemplo, de sondagens. A título de ilustração da metodologia, consideremos as alterações no padrão da votação na cidade de Valência, Espanha, entre as eleições ocorridas em 1995 e em 1999, baseando-nos exclusivamente nos resultados eleitorais nestas duas eleições em cada um dos 19 círculos eleitorais em que a cidade está dividida (tabelas 1 e 2). Os partidos políticos considerados foram: PP (conservadores), PSOE (socialistas), EU (comunistas), UV (direita-nacionalistas), Bloc, (esquerda-nacionalistas), outros (pequenos partidos) e a abstenção (eleitores que não votaram).

Tabela 2 — Cidade de Valência. Resultados por círculo eleitoral nas eleições de 1999

Círculo	PP	PSOE	EU	UV	Bloc	Outros	Abst.
1	9 778	2 319	786	599	643	154	7 692
2	18 261	4 087	1 165	991	993	302	12 540
3	19 329	5 340	1 469	1 260	1 241	317	14 094
4	7 816	4 775	1 168	687	701	196	9 391
5	13 839	7 300	1 970	1 281	998	301	14 866
6	12 416	2 709	657	611	654	188	8 398
7	13 000	8 365	1 811	1 257	825	338	16 530
8	15 107	8 302	2 209	1 505	1 180	362	17 370
9	12 595	8 659	1 947	1 468	931	294	16 517
10	18 570	9 445	2 496	1 933	1 184	734	23 405
11	12 617	10 251	2 460	2 167	1 050	443	20 894
12	12 182	8 482	2 054	1 280	935	291	17 164
13	10 919	6 477	1 696	1 070	1 011	274	12 959
14	7 750	4 699	1 225	693	698	190	9 343
15	9 717	7 862	1 708	1 012	735	317	15 184
16	8 409	7 126	1 509	1 036	689	256	13 994
17	1 839	872	178	461	215	22	1 642
18	2 597	2 599	389	391	194	63	4 090
19	4 911	2 663	591	740	325	984	6 020

Utilizando a metodologia descrita atrás obtemos uma distribuição *a posteriori* para a matriz 7×7 que descreve as probabilidades de transição entre cada par de partidos políticos. Os valores esperados para as probabilidades, expressos em percentagens (isto é, $100P_{i,j}$), são os da matriz de transições indicada na tabela 3. Esta tabela dá-nos uma análise

detalhada da estrutura de transferência de votos. Por exemplo, os dois maiores partidos (PP e PSOE) mantiveram na eleição de 1999 praticamente todos os eleitores que neles haviam votado na eleição de 1995 (95,8% e 95,3%, respectivamente), enquanto os comunistas (EU) apenas mantêm 43% do seu eleitorado, com perto de 50% a decidir-se pela abstenção na eleição de 1999.

Tabela 3 — Estrutura de transição. Distribuição percentual dos votos em 1999 em função do voto em 1995

% Votos 95→99	PP	PSOE	EU	UV	Bloc	Outros	Abst.
PP	95,8	0,0	0,0	0,0	0,7	0,0	3,5
PSOE	0,0	95,3	0,0	0,0	4,0	0,3	0,4
EU	0,2	0,0	43,0	0,0	6,7	0,2	49,9
UV	15,9	3,5	0,0	43,3	0,0	3,0	34,3
Bloc	0,0	0,0	0,0	0,0	100,0	0,0	0,0
Outros	0,0	0,0	0,0	0,0	0,0	98,5	1,5
Abst.	0,0	0,0	0,0	0,0	0,0	0,0	100,0

Como são conhecidos os resultados globais, as aproximações da tabela 3 podem naturalmente ser reescritas em termos absolutos, oferecendo uma estimativa da relevância efectiva da transferência de votos. Assim (veja-se a tabela 4), cerca de 4800 eleitores entre os 116 146 que votaram PSOE em 1995 deverão ter votado em Bloc na eleição de 1999, e, dos 47 569 eleitores que votaram em UV em 1995, estima-se que cerca de 7500 votaram em PP em 1999.

Tabela 4 — Estrutura de transição. Distribuição absoluta dos votos em 1999 em função do voto em 1995

Partido	Votos em 1995	PP	PSOE	EU	UV	Bloc	Outros	Abst.
PP	213 299	204 202	0	0	0	1 542	0	7 555
PSOE	116 146	0	110 592	0	0	4 759	359	436
EU	64 570	151	0	27 711	0	4 314	133	32 261
UV	47 569	7 547	1 654	0	20 516	0	1 391	16 461
Bloc	4 619	0	0	0	0	4 619	0	0
Outros	4 199	0	0	0	0	0	4 136	63
Abst.	184 820	0	0	0	0	0	0	184 820
Votos em 1999	211 900	112 246	27 711	20 516	15 234	6 019	241 596	

Referências

- Berger, J. O. (1995) – *Statistical Decision Theory and Bayesian Analysis*. Berlim: Springer.
- Berger, J. O. e Bernardo, J. M. (1992) – On the development of reference priors. *Bayesian Statistics 4* (J. M. Bernardo, J. O. Berger, A. P. Dawid e A. F. M. Smith, eds.). Oxford: University Press, 35-60 (com discussão).
- Bernardo, J. M. (1979) – Reference posterior distributions for Bayesian inference. *J. Roy. Statist. Soc. B* 41, 113-147 (com discussão). Reimpresso em *Bayesian Inference* (N. G. Polson e G. C. Tiao, eds.). Brookfield, VT: Edward Elgar, 1995, 229-263.
- Bernardo J. M. (2001) – Bayesian Statistics. *Encyclopedia of Life Support Systems* (EOLSS). Paris: UNESCO (em publicação).
- Bernardo, J. M. e Ramón, J. M. (1998) – An introduction to Bayesian reference analysis: inference on the ratio of multinomial parameters. *The Statistician* 47, 101-135.
- Bernardo, J. M. e Smith, A. F. M. (1994) – *Bayesian Theory*. Chichester: Wiley.
- Gelman, A., Carlin, J. B., Stern, H. e Rubin, D. B. (1995). *Bayesian Data Analysis*. Londres: Chapman and Hill.
- Lee, P. M. (1989) – *Bayesian Statistics: an Introduction*. Londres: Edward Arnold.
- O'Hagan, A. (1994) – Bayesian Inference (Kendall's Advanced Theory of Statistics 2B). Londres: Edward Arnold.

(Página deixada propositadamente em branco)

A matemática e a economia

Como digo, nas primeiras aulas, aos estudantes de Introdução à Economia, na Faculdade de Economia, a análise económica, nas suas relações de interdisciplinaridade, apresenta relações privilegiadas com a matemática. Estas relações levam a que os economistas tenham de conhecer e aplicar conhecimentos matemáticos, que, devido ao seu grau de complexidade, têm dificuldade em dominar. Ao longo dos anos, eu e alguns colegas, com esforço e nem sempre com sucesso, procurámos levar essa ideia para a Faculdade de Economia. E hoje, com orgulho, podemos dizer que neste aspecto o futuro está garantido.

Procurarei focar de forma breve os aspectos onde a economia encontrou a matemática e dela necessita. A perspectiva, obviamente pessoal, procura ter em conta o progresso em economia e o apoio que lhe deu a matemática.

Tal como para as restantes ciências sociais, o sucesso da física mecânica no século XIX, teve implicações sobre o que poderia vir a ser uma física social. A procura de uma beleza formal num mundo marcado pela ausência de incerteza vai ser importante no desenvolvimento da economia. A influência da física na forma de pensar dos economistas levou-os a identificar o tempo com o espaço, onde as forças em presença conduziriam aos desejados equilíbrios. O conceito de força acabará por valorizar a análise estática e a estática comparada¹. Ao mesmo tempo, através da ideia

¹ Veja-se sobre o conceito de tempo e a sua utilização em economia Barre (1950), Robinson (1980), Gram (1989) e ainda Burstein (1991).

de lei tendencial — que o conceito de forças e a sua acção justificavam —, imuniza-se a análise à possível refutação.

A revolução marginalista, iniciada nos anos 70 do século XIX, foi fundamental para a utilização da matemática na economia, sobretudo através das correntes do equilíbrio parcial e do equilíbrio geral. Se hoje o maior desenvolvimento de uma economia matematizada vem dos Estados Unidos, não devemos esquecer que os primeiros economistas matemáticos foram franceses: Cournot (1838)² e Dupuit (1844)³, mesmo antes de Léon Walras (1952). O isolamento do Ocidente foi responsável pelo desconhecimento dos trabalhos do russo Dimitriev (passagem do século XIX para o XX) e do japonês Keyshibata (início do século XX).

O primado do individualismo metodológico, tomado como uma *atomic proposition* (Whitehead e Russell, 1962) e imposto pela análise dominante, acabará por ter consequências ao nível do papel da matemática na economia. Graças ao individualismo metodológico, a representação do sujeito da economia, do homem da economia, pôde ser feita com base em princípios mínimos (Machlup, 1978), o que torna esta última mais abstracta e propícia à utilização do método hipotético-dedutivo (Andrade, 2000). Assim, o discurso da economia vai-se distinguir do discurso das outras ciências sociais, a ponto de se afirmar que, se o discurso económico se não funda numa representação matemática, pressupõe essa mesma representação (Kolm, 1986). A formalização avançou para formas bastante complexas, embora numa primeira fase limitada a relações determinísticas (Allen, 1963)⁴.

Não bastasse o individualismo metodológico ser aceite, em geral, pelos economistas, Popper (1985) ainda fortaleceu esse postulado ao apresentá-lo como um princípio mínimo de análise, que assim estaria isento da aplicação do seu critério de demarcação científica baseado no *falsificacionismo* (Nunes, 1996).

A Segunda Guerra Mundial veio determinar a necessidade do planeamento da economia nacional, da produção ao consumo, e os acontecimentos subsequentes, a necessidade de previsão para actuação dos governos dos países capitalistas. O planeamento da produção acabará por levar ao desenvolvimento das técnicas de *input-output*. O nome do economista de origem russa Leontief, Prémio Nobel em 1973, que apresentou e desenvolveu este tipo de análise, ficará para sempre ligado

² Veja-se Shubik (1987).

³ Veja-se Ekelund (1987).

⁴ O livro de Allen, tão marcante, é disso um exemplo.

aos grandes avanços da economia (Leontief, 1941). Os modelos de *input-output* conduzem ao desenvolvimento das aplicações de álgebra linear na economia, assim como ao desenvolvimento da programação linear.

Apesar de os economistas também serem escravos da moda, as aplicações do que também ficou conhecido por modelo de Leontief continuam a fazer-se, por exemplo na análise de recursos escassos não renováveis, onde caem certas formas de energia, e no da sua utilização para ter em conta a dinâmica de interdependência de custos e preços entre os ramos da economia.

A necessidade de previsão está hoje tão arreigada nos nossos comportamentos que envolvem decisões correntes ou futuras, que imaginamos que tenha sido sempre assim. Mais, a ausência de previsão passou mesmo a ser considerada sinónimo de incerteza. A incerteza, inerente à tomada de decisões económicas, passou a ser bastante minimizada graças à elaboração de previsões.

Os modelos econométricos⁵ desenvolveram-se primeiro para apoiar decisões dos governos e depois como uma importante actividade privada altamente lucrativa⁶. Não deixa de ser irónico que tenha sido a Cowles Commission (for Research in Economics) (CFRE, 2000b), fundada em 1932, a ser responsável pelo desenvolvimento dos métodos quantitativos na economia, procurando tornar a ciência económica uma ciência mais rigorosa. O interesse central do criador da fundação residia no conhecimento mais preciso do comportamento do mercado financeiro (CFRE, 2000a)⁷. Newton já havia desejado e ensaiado o mesmo.

A previsão criou o seu próprio vício, ou seja, a dependência de modelos econométricos. Por esta razão, assistimos ao desenvolvimento da análise estatística e econométrica, ao aparecimento da necessidade de encontrar aplicações informáticas que auxiliassem o economista.

A moda dos grandes modelos, pesados, de inspiração sobretudo keynesiana, foi abalada com a crise do início dos anos 70. Tratou-se de um fenómeno temporário, ou não fosse a imunização a primeira atitude dos investigadores. Adaptaram-se os modelos tendo em conta um maior peso das decisões de oferta e novas metodologias econométricas, que se desenvolveram. A maior crítica partira de Lucas: os parâmetros dos

⁵ Para uma história da econometria na análise económica, Hendry e Morgan (1995).

⁶ Para uma indicação elucidativa, mas já desactualizada, veja-se Whitley (1994).

⁷ Cowles perseguiu esse objectivo mesmo antes da criação da fundação.

modelos não poderiam ser independentes das alterações de política económica (Lucas, 1972). O progresso em econometria não se fez esperar. Mas, como o provou Sargent, com a hipótese de *observationally equivalent* (Sargent, 1976), os modelos para serem eficazes necessitam do conhecimento da teoria económica (Walsh, 1998)⁸.

A segunda revolução do individualismo metodológico correspondeu à mera consideração de que o indivíduo racional não devia basear as suas decisões em expectativas sistematicamente enviesadas. Os indivíduos devem agir de forma racional, usando a informação de que dispõem quando tomam as suas decisões. A revolução das expectativas racionais levou a que até para o indivíduo tenhamos de formalizar o comportamento da economia, para que este possa ter em conta as acções de todos os outros. O conhecimento e o desenvolvimento do cálculo recursivo tornou-se uma necessidade⁹. A macroeconomia dinâmica é assim herdeira de Wald, Bellman, e também de Kalman¹⁰. Convenhamos que para este desenvolvimento muito contribuiu a evolução dos meios de cálculo à disposição dos economistas. Para além dos modelos e das suas soluções, tornou-se igualmente importante a realização de simulações e a análise de sensibilidade das soluções a alterações nos parâmetros. Tratando-se de modelos não-lineares, não é difícil perceber as dificuldades matemáticas e estatísticas que o economista passou a enfrentar¹¹.

O ramo *gémeo* da economia, o *finance*, veio a ter uma importância visível nos anos 80 e 90. Os pais e impulsionadores da nova disciplina serão homens vindos da física. O progresso em *finance* só foi possível graças à competência desses investigadores. A atribuição em 1990 do Prémio Nobel a Markowitz, Miller e Sharpe foi o primeiro reconhecimento académico do novo ramo. Em 1997, Merton e Scholes, sucedem àqueles economistas.

A complexidade da análise pode facilmente ser apercebida se dissermos que o modelo primeiro da análise financeira consiste num problema de programação quadrática. A análise do valor de mercado dos activos financeiros, e assim do risco, gerou a necessidade de um apoio

⁸ Pp. 20-2.

⁹ São disso exemplo Lucas e Stockey (1989) e Ljungqvist e Sargent (2000).

¹⁰ Na clara formulação de Shubik, «The search for an explicit economic dynamics is a search for the Philosopher's stone», Shubik (2000), p. 2. De referir também Turnovsky (1995).

¹¹ Vejam-se Sargent (1987) e Azariadis (1998).

estatístico e matemático enorme. Não é pois de estranhar que nos últimos anos os intermediários financeiros tenham sido os principais compradores dos grandes computadores, a par dos sistemas de defesa, nas principais economias.

O papel que as aplicações informáticas poderiam ter na desestabilização do mercado financeiro levou, no final dos anos 80, a que se sugerisse a possibilidade de proibição do seu uso. A imaginação e a competência dos que elaboraram e actualizaram esses programas foram suficientes para não gerar opiniões unânimes sobre a situação do mercado financeiro. Esses *programas de negócio* tornaram-se assim um elemento importante da actividade financeira.

Mas integrar o risco não é o mesmo que integrar a incerteza (Knight, 1921). Esta última, não sendo probabilizável, não poderá integrar aqueles modelos (Nunes, 1998)¹². E, bem perto dos nossos dias, assistimos a verdadeiros colapsos financeiros no Japão e nos Estados Unidos, tendo estado envolvidos nesses acontecimentos Nobéis da economia, como conselheiros ou mesmo como fundadores e executivos desses intermediários financeiros. O mais recente foi o caso do LTCM (*Long Term Capital Management*) (Dunbar, 2000). Este fundo entrou em colapso definitivo em Setembro de 1998. Para evitar a sua falência, que seria desastrosa para os Estados Unidos e para a economia mundial, recebeu uma ajuda de 834 milhões de contos (escudos actuais) em 28 de Setembro. Para reanimar o mercado afectado pela impressionante desvalorização do capital da LTCM, a taxa de juro da autoridade monetária norte-americana, do Fed¹³, será ainda diminuída em 0,25% em 29 de Setembro, 14 de Outubro e 17 de Novembro. As crises financeiras são de tal forma perigosas, e temidas pelos governos, que estes se dispõem a gastar fortunas para as evitar: a crise dos anos 80 das «Saving and Loan Associations» custou ao governo norte-americano mais do que a Guerra do Vietname.

Apesar das «certezas» e das avaliações dos «riscos» dos economistas, a incerteza continua presente e, de vez em quando, teima em pregar partidas, uma forma de nos lembrar que o mundo não se reduz à representação que dele podemos fazer, por mais formal e bonita que ela seja. Mas os mercados financeiros não seriam o que são hoje sem a formalização matemática dos economistas.

¹² Sobretudo cap. IV, pp. 141-179, e anexo 3, pp. 201-214.

¹³ Federal Reserve Board. Instituição equivalente ao Banco Central Europeu.

Se a matemática tivesse a magia de resolver todos os nossos problemas, de formalizar, compreender e prever, num mundo incerto, o Dr. Dulcamara daria explicações de matemática, Nemorino perderia Adina e Fausto teria tido Mefistófeles a ensinar-lhe matemática esquecendo, entretanto, Margarida.

Se até aqui apenas me referi ao que se passa com o que podemos chamar ramo ortodoxo da análise económica, tem também interesse ver o que se passou com um ramo heterodoxo: o da economia marxista. Não a economia marxista ensinada nos países ex-socialistas, mas a que se estudava fora deles, embora a corrente ortodoxa marxista continuasse agarrada ao materialismo histórico e dialéctico, que posteriormente também degenera em «teoria da regulação», encontrando assim o (falso) apoio positivo de que necessitava. Falso, porque não passava de *story telling*. Uma outra corrente de economia marxista, positiva, vai desenvolver-se no seguimento da obra de Sraffa (1960). Esta corrente irá basear-se no uso de modelos lineares de produção (Duarte, 1994). A álgebra linear foi usada até aos seus limites. Estes economistas levantarão o problema da medida do capital e aproximam-se dos economistas keynesianos. A teoria do capital interessou os grandes economistas matemáticos até meados dos anos 80¹⁴. Nas faculdades de Economia que optaram por ensinar este ramo de economia marxista, o estudante entrava nas aulas à espera da *revolução* e encontrava uma matemática complicada para ele e mesmo complexa para a maioria dos economistas.

O que se passou com a economia marxista é ilustrativo do uso da liberdade de escolha com que nos defrontamos em economia. O economista com conhecimentos de matemática poderia escolher a escola que entendesse mais correcta e prometedora. O economista sem conhecimentos de matemática não tinha liberdade de escolha ...

Falei atrás da segunda revolução do individualismo metodológico e das suas consequências ao nível da formalização. A utilização do conceito de *bounded rationality* (Simon, 1976 e 1978) não terá destronado aquele outro conceito? Em primeiro lugar devemos insistir no facto de este último conceito ser definido em relação ao primeiro. Também o conceito de racionalidade individual pode ser tomado de forma que o meio em que ela é exercida aja sobre ela. Neste caso, perdemos o mais geral, o aspecto otimizador do comportamento, mas não o seu aspecto maximizador (Sen, 1996).

¹⁴ Kurz (1987) e Pasinetti e Scazzieri (1987).

A simplicidade de representação do indivíduo é agora substituída pela complexidade. As interações sucedem-se. A inteligência artificial, em geral, e as redes neuronais, em particular, têm aqui a sua porta grande de entrada na economia. Em aspectos particulares já haviam entrado. Como disse Sargent (1993), o economista das expectativas racionais era um super Deus; criava indivíduos que conheciam melhor do que ele (economista) o funcionamento da economia, e afinal deve criar indivíduos mais à sua semelhança, que erram e aprendem, que voltam a errar e a aprender ...

Em suma, um comportamento menos simples do homem da economia acarreta uma maior complexidade da análise económica, e por essa via a complexidade dos meios de análise matemática. No mesmo sentido, acaba por apontar a introdução das emoções nas escolhas racionais dos indivíduos (Damásio, 1994).

Há muito que os economistas abandonaram a escravatura do realismo das hipóteses (Friedman, 1953) como base de partida para a investigação. Por isso, a introdução de formas não ilimitadas de racionalidade deixa o economista num dilema. A introdução de um *realismo* no comportamento do homem da economia tem de ser avaliada do ponto de vista dos seus resultados. Não adianta bater no nosso *homem*. Ele não protesta. Como disse Espinosa, o conceito de cão não ladra! Para já, sabemos que os resultados daqueles modelos estão demasiado dependentes de hipóteses *ad hoc* sobre comportamentos.

Dilema idêntico se coloca ao problema do *agente representativo* na macroeconomia dinâmica. Ninguém duvida que tomar um agente que representa todos os outros pode ser uma hipótese demasiado *irrealista*. Mas a contrapartida é a complexidade matemática que resulta de tomarmos agentes heterodoxos. Para além, mais uma vez, de aumentarem as hipóteses *ad hoc* sobre comportamentos que resultam do acréscimo de parâmetros do modelo.

O tratamento da informação assimétrica é muito interessante para o economista. No entanto, as exigências de conhecimentos matemáticos são bastante grandes, pelo que não são muitos os economistas com capacidade para formalizar tais comportamentos, que são importantes sobretudo em economia financeira.

Já houve tempos em que os economistas pensaram a economia como a ciência social abrangente das restantes¹⁵. A questão deixou de fazer sentido. O que hoje temos é um método usado em economia que, em

¹⁵ Osório (1996) tinha essa posição. Sobre este autor veja-se nesta obra o prefácio do professor Jacinto Nunes.

algumas situações, se impõe nas outras ciências sociais. O método de investigação começou a ser conhecido por outras disciplinas através de um sociólogo e de um economista. O rigor de análise e a elegância de apresentação têm com certeza a sua importância neste reconhecimento (Coleman, 1990 e Becker, 1981).

Uma corrente da economia, designada por «economia dos ciclos reais», introduziu uma nova metodologia de investigação, de que os chamados modelos de equilíbrio geral também fazem uso. A formalização bastante avançada e, em geral, a utilização de programação dinâmica, permite a representação da economia baseada em comportamentos definidos *a priori*. A sua construção depende da opinião do investigador sobre o que deve ou não ser incluído. Se é assim que se investiga em qualquer ciência, aqui o problema é diferente devido ao tipo de não refutação que é usado. Em geral o número de parâmetros é elevado e a escolha dos seus valores feita de forma a que o conjunto das equações do modelo produza soluções próximas dos valores efectivos. Se a reprodução dos valores das variáveis, para as quais o modelo foi construído, for boa, então o modelo é bom. Se a reprodução for má, voltamos à calibragem do modelo, ou seja, à escolha de novos valores para os parâmetros (King, 1995, e King e Rebelo, 2000). Tal como num treino de Fórmula Um. Só que ninguém, até hoje, atribuiu o título de cientista a Schumacher.

Esta metodologia é um retrocesso ao *verificacionismo* do início do século XX (Nunes, 1996, e Jacob, 1980). Podemos classificar esta prática como *jogos de economia* que apenas os que possuem elevados conhecimentos de matemática podem jogar, mas não como construção científica.

Em suma, a matemática surge-nos como um instrumento de análise do qual é impossível prescindir e sem o qual o progresso não será possível. Mas o seu uso de forma abusiva pode afastar a análise do seu objecto de estudo e não respeitar condições metodológicas de investigação¹⁶. Afinal, a economia pode ser tomada como aquilo que os economistas fazem, e nisso, estes, fazem um uso crescente da matemática.

Feynman (1989) disse que lamentava quem não tinha conhecimentos de matemática porque não podia apreciar a natureza. Por sua vez, Kolm (1986)¹⁷ disse que o mundo, hoje, seria bem diferente se Marx, em vez de na sua juventude ter estudado filosofia, tivesse estudado matemática.

¹⁶ Ainda que se aceite a posição de Hanson (1961).

¹⁷ P. 106.

Referências

- Allen, R. G. D. (1963) – *Mathematical Economics*. Londres: Macmillan.
- Andrade, J. S. (2000) – Algumas Observações sobre a Metodologia da Economia. *Episteme*, 5-6, pp. 39-76.
- Azariadis, C. (1998) – *Intertemporal Macroeconomics*. Oxford: Blackwell Publishers.
- Barre, R. (1950) – *La Période dans l'Analyse Économique, une approche à l'étude du temps*. Paris: Société d'Édition d'Enseignement Supérieur.
- Becker, G. (1981) – *A Treatise on the Family*. Cambridge, Ma.: Harvard University Press.
- Burstein, M. (1991) – History versus Equilibrium, Joan Robinson and Time in Economics, in *The Joan Robinson Legacy*. I. Rima ed. Armonk: M. E. Sharpe, pp. 49-61.
- CFRE (2000a) – Alfred Cowles, 3.^a ed., Founder and President, Setembro 1932-Setembro 1934. <http://cowles.econ.yale.edu>, Yale University.
- CFRE (2000b) – A Short History. <http://cowles.econ.yale.edu>, Yale University.
- Coleman, J. (1990) – *Foundations of Social Theory*. Cambridge, Ma.: Harvard University Press.
- Cournot, A. A. (1838) – *Recherches sur les Principes Mathématiques de la Théorie des Richesses*. Paris: Livraria Hachette.
- Damáσιο, A. (1994) – *O Erro de Descartes, Emoção, Razão e Cérebro Humano*. Lisboa: Publicações Europa-América.
- Duarte, A. D. (1994) – *Valeur, Prix, Dynamique non-Proportionnelle. Une mise en perspective des apports critiques et constructifs du courant neocambridgien*, 550 pp.; Coimbra: Universidade de Coimbra.
- Dunbar, N. (2000) – *Inventing Money, the Story of Long-Term Capital Management and the Legends Behind It*. Chichester: John Wiley & Sons.
- Dupuit, A. J. E. J. (1844) – De la Méasure de l'Utilité des Travaux Publics. *Annales des Ponts et Chaussées*, 13.
- Ekelund, R. (1987) – Dupuit, Arsène-Jules-Emile Juvenal, in *The new Palgrave: A dictionary of economics*. J. Eatwell, M. Milgate and P. Newman eds. Londres: Macmillan Press, pp. 943-4.
- Feynman, R. (1989) – *O que é uma Lei Física*. Lisboa: Gradiva.
- Friedman, M. (1953) – The Methodology of Positive Economics, in *Essays in Positive Economics*. M. Friedman ed., Chicago: The University of Chicago Press, pp. 3-43.
- Gram, H. (1989) – Ideology and Time: criticisms of General Equilibrium, in *Joan Robinson and Modern Economic Theory*. G. Feiwel ed., Londres: Macmillan Press, pp. 285-302.

- Hanson, N. (1961) – Is There a Logic of Discovery?, in *Current Issues in the Philosophy of Science*. H. Feigl and G. Maxwell, eds., Nova Iorque: Reinehart and Winston, pp. 20-35.
- Hendry, D., M. Morgan eds. (1995) – *The Foundations of Econometric Analysis*. Cambridge: Cambridge University Press.
- Jacob, P. (1980) – Comment peut-on ne pas être Empiriste?, in *De Vienne à Cambridge. L'héritage du Positivisme Logique*. P. Jacob, ed., Paris: Gallimard, pp. 7-52.
- King, R. (1995) – Quantitative Theory and Econometrics. *Federal Reserve Bank of Richmond Economic Review*, 81, pp. 53-105.
- King, R., S. Rebelo (2000) – Resuscitating Real Business Cycles. *Rochester Center for Economic Research*, Working Paper No. 467.
- Knight, F. (1921) – *Risk, Uncertainty and Profit*. Chicago: University of Chicago Press.
- Kolm, S. C. (1986) – *Philosophie de l'Économie*. Paris: Seuil.
- Kurz, H. (1987) – Capital Theory: Debates, in *The New Palgrave: A Dictionary of Economics*. J. Eatwell, M. Milgate and P. Newman, eds., Londres: Macmillan Press, pp. 357-62.
- Leontief, W. (1941) – *The Structure of the American Economy, 1919-1929*. Cambridge, Ma.: Harvard University Press.
- Ljungqvist, L., T. Sargent (2000) – *Recursive Macroeconomic Theory*. Cambridge, Ma.: The MIT Press.
- Lucas, R. (1972) – Expectations and the Neutrality of Money. *Journal of Economic Theory*, 63, pp. 103-124.
- Lucas, R., N. Stokey (1989) – *Recursive Methods in Economic Dynamics*. Cambridge, Ma.: Harvard University Press.
- Machlup, F. (1978) – *Methodology of Economics and Other Social Sciences*. Nova Iorque: Academic Press.
- Nunes, J. (1996) – *Introdução à Epistemologia e Metodologia Económicas – O Pós-Positivismo*. Coimbra: Boletim de Ciências Económicas.
- Nunes, J. (1998) – *O Pensamento de Keynes*. Lisboa: Imprensa Nacional – Casa da Moeda.
- Osório, A. H. (1996) – *A Matemática na Economia Pura: a Troca*. Lisboa: Banco de Portugal.
- Pasinetti, L., R. Scazzieri (1987) – Capital Theory: Paradoxes, in *The New Palgrave: A Dictionary of Economics*. J. Eatwell, M. Milgate e P. Newman, eds., Londres: Macmillan Press, pp. 363-68.
- Popper, K. (1985) – The Rationality Principle, in *Popper Selections*. D. Miller, ed., Nova Iorque: Princeton University Press, pp. 357-365.
- Robinson, J. (1980) – Time in Economic Theory, in *Further Contributions to Modern Economics*. J. Robinson, ed., Oxford: Basil Blackwell, pp. 86-95.

- Sargent, T. (1976) – The Observational Equivalence of Natural and Unnatural Rate Theories of Macroeconomics. *Journal of Political Economy*, 84:3, pp. 631-640.
- Sargent, T. (1987) – *Dynamic Macroeconomic Theory*. Cambridge, Ma.: Harvard University Press.
- Sargent, T. (1993) – *Bounded Rationality in Macroeconomics*. Oxford: Clarendon Press.
- Sen, A. (1996) – Maximization and the Act of Choice. *Temi di Discussione, Banca de Italia*, 270, Maio.
- Shubik, M. (1987) – Cournot, Antoine Augustin, in *The New Palgrave: A Dictionary of Economics*. J. Eatwell, M. Milgate e P. Newman, eds., Londres: Macmillan Press, pp. 708-712.
- Shubik, M. (2000) – The Theory of Money. *Cowles Foundation Discussion Papers*, 1253.
- Simon, H. A. (1976) – From Substantive to Procedural Rationality, in *Method and Appraisal in Economics*. S. Latsis, ed., Cambridge: Cambridge University Press, pp. 129-148.
- Simon, H. A. (1978) – Rationality as Process and as Product of Thought. *American Economic Review*, 68:2, pp. 1-16.
- Sraffa, P. (1960) – *Production of Commodities by Means of Commodities*. Cambridge: Cambridge University Press.
- Turnovsky, S. (1995) – *Methods of Macroeconomic Dynamics*. Cambridge, Ma.: The MIT Press.
- Walras, L. (1952) – *Éléments d'Économie Politique Pure, ou théorie de la richesse sociale*. Paris: Librairie Générale de Droit et de Jurisprudence (1874).
- Walsh, C. (1998) – *Monetary Theory and Policy*. Cambridge, Ma.: MIT Press.
- Whitehead, A., B. Russell (1962) – *Principia Mathematica*, 1927, *Paper Edition to *56*. Cambridge: Cambridge University Press.
- Whitley, J. (1994) – *A Course in Macroeconomic Modelling and Forecasting*. Londres: Harvester Wheatsheaf.

(Página deixada propositadamente em branco)

Mário F. Lages
Director do Centro de Estudos
e Sondagens de Opinião da
Faculdade de Ciências Humanas
da Universidade Católica

Sobre a inferência em estudos de opinião

INTRODUÇÃO

Os problemas relacionados com a compreensão dos comportamentos humanos, designadamente os referentes à opinião pública, não são de solução simples, sendo a matemática apresentada, por vezes, como a vara de condão capaz de transformar o pensamento impreciso em conhecimento seguro. A nível básico, certamente o será, a acreditarmos em Bertrand Russell, que afirma serem a «matemática e a lógica [...] a mesma coisa» (Russel 1972 [1903]: v). Possivelmente por isso é que todas as ciências humanas a têm vindo a utilizar para dizer com rigor o que a linguagem vulgar não faz de maneira convincente. Os casos mais claros são os da demografia e da economia, esta tratada noutro texto desta colectânea. Quanto à demografia, pode dizer-se que teve um desenvolvimento paradigmático, sendo uma das ciências humanas que mais têm utilizado a matemática. Basta recordar os trabalhos sobre a mortalidade de Breslau, por Edmund Halley, nos já longínquos idos de 1693, a lei do crescimento geométrico da população de R. Malthus (1798) e as suas várias correcções, designadamente logísticas, desde Verhulst (1838), os contributos de A. Lotka (1998) para a teoria das populações estáveis, os logitos de sobreviventes de Brass (1974), ou a modelização da fecundidade de Coale e Trussell (1978).

O argumento poderia ser desenvolvido quase indefinidamente, sem qualquer vantagem, porém, já que a questão nele subjacente está nos

antípodas do que nos interessa aqui. De facto, os modelos globais dos comportamentos, entre os quais os demográficos, levam a esquecer não só a pluralidade e complexidade das interacções comportamentais, mas também a sua não linearidade. Assim, os pressupostos de uma teoria determinista da história e dos comportamentos — de que são exemplos os estádios de Vico, Comte e Frazer — ao estabelecer sequências obrigatórias e universais, reduzem essa complexidade a uma evolução pré-determinada.

Tais perspectivas, fundamentalmente macroteóricas, não podem ter em conta que os fenómenos colectivos dependem de decisões nem sempre racionalizadas ou automáticas. O jogo das interacções, diferencialmente definido pelos actores sociais e teorizado pela sociologia compreensiva, leva a dizer que a conclusão não está sempre contida nas premissas. Por isso é que, se o bater das asas da célebre borboleta de Pequim provoca uma tempestade em Nova Iorque, a potenciação do efeito de eco é notória nos fenómenos sociais, como podemos deduzir de um exemplo recente: a visita de Ariel Sharon à zona das mesquitas de Jerusalém levou a uma nova Intifada, que levou às balas israelitas, que levou à morte de uma criança diante das câmaras da televisão e perante a protecção impotente do pai, que levou ao linchamento de dois soldados israelitas em Ramalah, que levou à quase-guerra hoje em curso no Médio Oriente.

Esta sequência perversa e não previsível dos actos humanos, exacerbados no seu significado pelos sentimentos mais diversos, muitas vezes vividos à flor da pele das sensibilidades colectivas, manifesta, pela sua natureza excessiva e extrema, quanto a percepção das acções alheias condiciona, numa perspectiva weberiana, toda a interacção humana e social. Por isso se insiste cada vez mais em que a indeterminação dos actos humanos tem a sua correspondência nos próprios fenómenos naturais, sobretudo a nível micro. A incerteza não pode deixar de ser introduzida na própria definição da realidade e nas técnicas de análise a que ela pode ser submetida.

Este conjunto de considerações permite centrar a intenção deste trabalho: ilustrar os pressupostos da análise das atitudes expressas em sondagens por questionário, o instrumento que os sociólogos têm vindo a privilegiar para a captação das circunstâncias definidoras dos comportamentos. Tentarei, por isso, apresentar alguns contornos da produção científica em sociologia, pondo em destaque alguns enviesamentos que a estatística pode produzir. No âmago do tema está uma teia de concepções e de métodos, de não fácil destrinça, tal como a realidade sobre a qual o sociólogo se debruça. Na sequência destas reflexões farei a sua aplicação aos resultados de uma sondagem.

NOTAS EPISTEMOLÓGICAS

E porque de teia conceptual, que não só matemática, se trata, coloco algumas questões sobre os processos de produção dos factos que constituem a matéria dos estudos de opinião. Sendo eles o primeiro objecto de construção teórica, convirá reflectir sobre os instrumentos que os originam e sobre o seu efeito na produção da realidade.

Na base de grande parte das empiria sociológica moderna está o questionário, um instrumento não inocente que resulta de várias escolhas, todas elas sujeitas a imprecisões, em parte decorrentes do tipo de variáveis que o estruturam. O investigador tem muitas vezes de se contentar com variáveis nominais, ou, quanto muito, ordinais, o que naturalmente se reflecte no poder das estatísticas utilizáveis. Por outro lado, a matemática tem vindo a produzir métodos específicos de análise de dados. Contudo, as teorias que com eles podem ser construídas nem sempre resistem a uma crítica rigorosa. De resto, muitas vezes, os métodos afastam-se do que o cientista considera intuitivo, dificultando a sua utilização inteligível. Os resultados são, assim, inseguros e pouco sustentados teoricamente. Acontece mesmo que a teoria não é suficiente para evitar desvios. De facto, nem todos os olhares conseguem ver correctamente a realidade. O olhar teoricamente cego certamente o não faz.

Relacionado com estas observações está o facto de que, em sociologia, como se deduz do que foi dito acerca da perspectiva compreensiva, não é fácil observar o preceito da ruptura epistemológica que Bachelard recomendava como condição da produção científica. A interpretação do actor acerca da acção, própria e alheia, faz parte integrante da situação e da interacção nela produzida. Além disso, em qualquer inquérito de opinião, por mais clara e inequívoca que seja a pergunta, a resposta é sempre dada em função do entendimento circunstancial que dela faz o inquirido. Esta circunstância incorpora contextos vivenciais que constituem o pano de fundo em que os significados são entrevistados a partir de uma experiência pessoal sobredeterminada culturalmente. Por isso é pouco provável que todos os entrevistados entendam todas as perguntas da mesma maneira. Os comportamentos têm uma forte componente de ambiguidade. E o próprio actor social pode ter duas interpretações diferentes dos mesmos factos em dois momentos da sua vida, sejam eles separados por dias, horas ou mesmo minutos. Não é invulgar alguém mudar de opinião no decurso de uma entrevista. O questionário e a relação entrevistador-entrevistado podem contribuir para isso.

O distanciamento entre a empiria e a teoria que a ruptura epistemológica supõe não é, pois, facilmente atingido. A teoria, por mais implícita

que seja, do actor acerca da sua própria acção faz parte integrante desta. Desde W. Thomas que se sabe, segundo a interpretação de R. Merton (1957: 421), que quando alguém toma uma definição da situação como real, ela se torna real nas suas consequências. Conquanto, na primitiva formulação de Thomas, essa definição, de natureza individual, pudesse admitir contradições (cf. Thomas e Znaniecki 1998: 12), não deixa de ser importante para o nosso argumento dizer que toda a definição da situação faz parte integrante do sistema de acções e reacções que determinam o seu desenvolvimento, com inclusão destas teorias implícitas. Por outro lado, a realidade sociológica é feita de imagens implicadas, transmitidas por quem está interessado na sua definição. Os factos sociais não são constituídos apenas por expressões de sentimentos, atitudes e comportamentos, mas também pelas configurações dos meios em que emergem. E tanto assim é que tomamos, às vezes, por realidades, as imagens que se formam e deformam no ecrã das identificações pessoais e colectivas.

Assim, as mediações expressivas são como artes de engano, não porque consciente, obrigatória e propositadamente o queiram ser, mas porque nem sempre está realizado, entre outras coisas, o processo de auto-identificação prévia do respondente, sendo certo que só ele poderia induzir a resposta «verdadeira». A rapidez do inquérito, a relação interpessoal que o produz, a invasão da privacidade que representa, são circunstâncias que dificultam a expressão sincera, mesmo por quem não está na defensiva. Por outro lado, o inquérito pode produzir no respondente uma clara consciência de realidades e atitudes desconhecidas, sendo esta tomada de consciência acompanhada de conflitos interiores que o podem levar a produzir uma imagem ideal de si — do que é, faz, sente, pensa e sabe —, tanto para salvar a face como para melhorá-la. Pode mesmo acontecer que se convença de que a realidade assim definida é exacta.

Sendo, além disso, o inquérito um momento de afirmação da conformidade com o que é correcto do ponto de vista político, ético ou comunitário, a definição das atitudes nem sempre é lógica e coerente. O espaço de formação e expressão da opinião é, de facto, multidimensional. E as linhas de força que percorrem este espaço não são facilmente destrinchadas por quem está sob o império da pergunta incontornável, a que é necessário responder imediatamente.

Tudo isto é relevante para o nosso propósito, designadamente no caso em que a principal preocupação do cientista é definir o que é o «homem médio», como acontece em certas metodologias tipológicas — à maneira do que já se propunha A. Quételet — por mais diversificado que este homem seja, a partir de uma multiplicidade de critérios analíticos.

Ora as identidades modernas ou pós-modernas são cada vez mais circunstanciais e contextuais. Num momento o actor age segundo a norma, noutro está na moda, no seguinte faz parte da excepção; aqui está em unísono com alguém, acolá com outrem. Desta sorte, há cada vez menos conformidades absolutas e permanentes. A busca da originalidade e da diferença — ou melhor, de uma identidade diferencial — constitui uma das características dominantes dos comportamentos actuais, não só das vanguardas, mas também das classes médias, ou mesmo baixas, que anseiam pelo reconhecimento e pela notoriedade, ao contrário do que acontecia com as personalidades tradicionais e endonómicas dominantes entre nós, respectivamente nas comunidades rurais e nas classes médias urbanas, até ao último quartel do século XX. Nem a simplicidade estrutural das primeiras, nem os ditames conservadores das segundas proporcionavam o debate de ideias e as mudanças rápidas que hoje acontecem em quase todos os meios sociais.

Nesses contextos, em grande parte ultrapassados, as instituições, com os seus amplos reflexos de ditames e sanções, formavam um complexo concêntrico, quase concentracionário, fortemente condicionador dos comportamentos. Por isso eles eram relativamente previsíveis. Hoje, porém, a cultura dominante exprime-se em fragmentações aptitudinais e comportamentais, em conflitualidades ideológicas e religiosas, em contradições éticas e doutrinárias, em impasses vários nas relações sociais. E, por mais que se tente congruar as oposições e as diferenças, a acção humana é simultaneamente imbricada e contrastante, tanto ao nível das sociedades, das comunidades e dos grupos, como do próprio indivíduo. Daí a dificuldade na manutenção consistente de comportamentos e opiniões, o que apela para modelos interpretativos necessariamente complexos.

IMPRECISÃO, IMPLICAÇÕES FRACAS E PARADOXOS

Tudo isto tem implicações óbvias nos modelos de análise. Mas a questão fundamental é saber se os comportamentos podem ser matematizados, o que poderá parecer supérfluo, já que, de uma maneira ou de outra, isso tem sido feito. Não se deixará, no entanto, de notar que há quem encontre dificuldades no argumento. Coleman, por exemplo, na sua já clássica *Introduction to Mathematical Sociology* (1964: 3), diz que o isomorfismo entre as estruturas matemáticas e as ciências da natureza não se verifica em relação à «ciência social». De facto, «as teorias verbais

e os resultados da investigação [...] são definidos de maneira tão vaga ou tão fraca que é difícil traduzi-los em linguagem matemática, e, quando traduzidos, frequentemente não mostram o seu isomorfismo com partes fortes da matemática». O problema estaria em que, «[e]m geral, a construção teórica em sociologia tem-se confinado à riqueza — e à ambiguidade — da linguagem ordinária» (Coleman 1964: 3).

Estando a polissemia no âmago da realidade social e sendo a regularidade, em parte, uma abstracção, é correcto dizer que os comportamentos humanos são cada vez mais refractários a lógicas deterministas (pelo menos numa parte importante da sua expressão) ou a teorias de causalidade absolutas e unidireccionadas. A teoria dos jogos, por exemplo, veio fazer encarar numa perspectiva diferente alguns desses processos. Mas a ilógica das decisões, sobretudo ao nível das interacções micro-sociais, necessita porventura de ser trabalhada com outros instrumentos. De qualquer maneira, Raymond Boudon (1971: 4), ao debruçar-se sobre a incerteza das associações causais entre fenómenos sociais, defende que a sociologia apenas pode chegar a uma «teoria das implicações fracas». Por mais que nela se pretenda formular um corpo de proposições integradas e irrefutáveis, em termos de lógica dedutiva, tal objectivo não é facilmente atingido. Para ele é claro que «[a] natureza do social nunca permite observar implicações estritas» (Boudon 1971: 18).

Estas questões, e outras semelhantes, têm vindo a ser reflectidas por matemáticos e cientistas sociais. Hoje fala-se, por exemplo, de conjuntos imprecisos, tema importante para ambas as ciências. Mas o próprio conceito de imprecisão (*fuzziness*) é complexo, pois nem se identifica com probabilidade, no sentido de frequência, nem se distingue sempre de probabilidade subjectiva ou se confunde com a ambiguidade ou falta de intersubjectividade. Por outro lado, é distinto de algumas espécies de generalidade (cf. Smithson 1987: 13).

O desenvolvimento de alguns métodos e algoritmos específicos relacionados com esta temática — designadamente em *A Ciência do Impreciso*, de Abraham Moles (1995) e em *Fuzzy Sets for the Behavioral and Social Sciences*, de Michael Smithson (1987) — tem em conta a natureza imprecisa de muitos factos sociais, cuja análise está longe de poder cumprir a regra da clareza e distinção recomendada pela metodologia cartesiana, ou de levar a uma teoria da causalidade estrita. Mas estes métodos permitem esperar uma definição mais adequada das coisas.

No mesmo contexto se pode falar de uma outra teoria emergente, a das facetas, ou seja, os aspectos ou características de uma situação, em que é tomada em linha de conta, por exemplo, «a positividade dos eventos,

o seu número, a sua variabilidade e a dificuldade em os tratar» (Borg e Shye 1995: 2). De acordo com esta teoria, seria importante a noção de perfis ou estructuplos, ou seja, a «combinação dos elementos das facetas, individualmente tomados em cada uma» (Borg e Shye 1995: 2).

Não é obviamente este o lugar apropriado para o desenvolvimento das questões levantadas por este conjunto, antigo e recente, de instrumentações matemáticas. A sua importância para o nosso tema pode, no entanto, ser entrevista através de duas impropriedades, significadas pelo efeito de Condorcet e pelo paradoxo de Simpson. O primeiro, posto em evidência por um dos «primeiros utilizadores da matemática nas ciências sociais», mostra que, quando pedimos a um grupo de pessoas que ordene três objectos por ordem de preferência, frequentemente acontece que as suas opções são incompatíveis entre si, de sorte que «preferências individuais coerentes tornam-se incoerentes quando se juntam» (Boudon 1971: 4). Por outro lado, Simpson notou que pode haver «uma associação positiva marginal entre duas variáveis binárias B e C, embora em função de uma terceira variável, A, a associação entre C e B possa ser negativa» (cf. Cox 1992: 291).

Tanto o efeito de Condorcet como o paradoxo de Simpson mostram que a associação e a causalidade podem conduzir a erro, sendo frequentemente necessário decompor os fenómenos por forma a compreender as suas verdadeiras interacções, reforços e debilitações mútuas. A noção de causalidade supõe que uma variável está associada a uma ou mais características de um fenómeno. Mas esta relação depende de muitas circunstâncias em que o efeito pode ser anulado. Por outro lado, o importante não é estabelecer os termos extremos dos efeitos produzidos, mas definir as intermediações existentes, para já não falar dos efeitos retroactivos, entendidos em termos sistémicos e não temporais.

Nenhuma destas clarificações metodológicas poderá ser levada a cabo sem uma ordenação teórica dos diversos condicionamentos em causa, o que implica a definição prévia da teia dos efeitos causais, a sua reciprocidade e retroactividade eventuais. Para além disso, é necessário verificar em que sentido a causalidade se verifica, podendo acontecer, como refere Rosenblatt (cf. Cox 1992: 293-294), que determinados sistemas não lineares sejam determinísticos numa direcção e estocásticos noutra, caso se inverta a variável tempo.

O papel do cientista está assim na definição da cadeia teórica das relações entre os fenómenos, num modelo em que a realidade é considerada interagente, ou, numa linguagem mais antropológica, intersignificativa. Neste mesmo sentido se podem formular modelos que podem ser considerados «potencialmente causais» (Cox 1992: 294), para o que será

necessário distinguir variáveis imediatas, intermédias e envolventes, cada uma delas com uma distância específica em relação aos efeitos produzidos. Assim, a definição das mediações de causalidade não pode ser substituída pelo «ver se dá» em «corridas de programas» não suportadas por relações teóricas claras entre as variáveis.

TIPOS E ATRIBUTOS

Algumas das dificuldades anteriormente assinaladas são particularmente visíveis nas análises tipológicas, teorizadas, entre outros, por Max Weber, as quais têm vindo a ser utilizadas de uma forma extensiva nos métodos de análise factorial e de «clusters». Na perspectiva weberiana, tal procedimento recomenda-se porque os tipos ideais correspondem a um «complexo que é entendido como um sistema internamente consistente». Para ele, «este constructo é em si mesmo como uma utopia a que se chegou pela acentuação analítica de certos elementos da realidade» (Weber 1949: 90), procedendo-se na sua definição a uma espécie de polarização das características observadas. Daqui resulta que o tipo ideal é uma pura construção teórica que não tem outra intenção que não seja a clarificação dos conceitos.

Nos métodos factoriais, a tipologia tem a ver com a descrição interpretativa da realidade. Do conjunto, necessariamente compósito e circunstancial — pois depende das variáveis disponíveis para tal composição —, pretende-se tirar conclusões acerca das relações últimas entre as características. O método é sedutor. Através da matemática e da correlação entre os atributos constrói-se uma realidade teórica sobre os factores que lhes estão subjacentes. Não deixo, no entanto, de notar que nestas análises pode facilmente imiscuir-se o efeito de Condorcet, sobretudo se se usam modelos micro para construir explicações macro. Tomo, com efeito, como ajustada a advertência de Miller, num contexto um pouco diferente: «uma análise quantitativa dos indivíduos só pode conduzir a conclusões precisas acerca dos indivíduos; uma análise de lugares, só a conclusões precisas acerca dos lugares; e uma análise de época, a conclusões acerca das épocas» (Miller 1995: 155-156). E se é certo que a «interpretação [de um resultado estatístico] não faz frequentemente mais do que transformar noutro, com coordenadas diferentes, o espaço de atributos de um sistema tipológico, [acontece que] com isso se muda simultaneamente uma redução noutra». Ora, como diz Lazarsfeld (1993: 166), deve-se atender a que «[a]s classificações tipológicas usadas na investigação científica actual são um tanto vagas e portanto para elas pode ser fornecida mais

do que uma substrução¹ lógica. Mas os diferentes espaços de atributos provindos destas diferentes substruções podem ser transformados uns nos outros» (Lazarsfeld 1993: 166).

A construção das tipologias está, pois, marcada por dificuldades que não podem ser supridas por técnicas matemáticas que não sejam suportadas teoricamente. Lazarsfeld (1993: 166) recomenda que «quando estão em causa diversos atributos, será melhor [de preferência a escolhê-los, à partida] que o estudioso se familiarize em profundidade com o material, ordenando-o em seguida em alguns tipos principais numa base inteiramente impressionista. Só depois deverá reconsiderar a matéria, subestruturar as suas próprias intuições tipológicas num espaço de atributos adequado e pôr em evidência a redução que usou implicitamente. Os melhores resultados serão provavelmente obtidos justamente na combinação de um inquérito genérico e de uma análise sistemática subsequente.» A questão fundamental está em que o espaço de atributos que permitem uma tipologização é por natureza multidimensional. Nenhuma redução a que se queira proceder, por mais automática que seja, elimina a necessidade de ser liderada teoricamente, não permitindo que os casos postos de lado, por não concretizarem algumas das características, venha a inviabilizar as relações descobertas.

FELICIDADE E RELIGIÃO

Estes elementos relativos à análise de dados de opinião levam-me a concretizar melhor o que pretendo dizer através de algumas considerações sobre as associações supostas num trabalho do Instituto de Ciências Sociais (Cabral *et al.* 2000), cujos principais resultados foram divulgados numa conferência de imprensa no dia 17 de Março de 2000². O facto de se tratar de uma investigação sociológica de vulto e de, nessa apresentação, terem sido formuladas algumas conclusões surpreendentes, suportadas pela utilização de uma metodologia tipológica, justificam estas reflexões. Com efeito, nas notícias das televisões e dos jornais sobre o evento foi referido que o estudo revelava que os católicos praticantes eram

¹ O conceito de substrução, sugerido por Hempel e Oppenheim, é o seguinte: «O procedimento para encontrar num determinado sistema de tipos o espaço de atributos a que pertence a redução que implicitamente foi usada» (Lazarsfeld 1993: 162). A redução é «um agrupamento de combinações de atributos que envolve mais de um atributo» (Lazarsfeld 1993: 161).

² Agradeço ao professor José Machado Pais o ter-me facultado este estudo (Pais 2000).

infelizes. A notícia baseava-se não só no que foi então afirmado, mas também no já referido trabalho de Machado Pais. Entre as características por ele referidas na figura 1, intitulada «As teias da religiosidade conservadora: católicos ritualistas, moralistas e tradicionais», é dito serem eles, entre outras coisas, «nada felizes», pensarem que o «mundo está cheio de maldade e de pecado» e que «o destino não se altera» (Pais 2000: 24). Por oposição, os laicos «urbanos e elitistas» (figura 3) são «bastante felizes», pensam «que a vida faz sentido sem Deus» e têm «expectativas elitistas» (Pais 2000: 24). O autor afirma ainda que «o dado mais saliente se prende com o facto de 58% dos Portugueses que se acham ‘nada felizes’ com a sua vida se encontrarem no grupo dos católicos ritualistas» (Pais 2000: 11). O contrário acontece com os laicos urbanos e elitistas: «Ao pensarem em todos os aspectos da sua vida actual, 62% consideram-se bastante felizes» (Pais 2000: 17).

O que parece estar contido em tal leitura, sobretudo nas figuras mencionadas, é que há uma relação negativa entre religião e felicidade. Só quem não acredita em Deus e não reza é feliz. Assim, rapidamente se passa à ideia de que o catolicismo é causa de infelicidade. O que não é dito é que só 8,2% de todos os católicos, ritualistas ou não, se considera «nada feliz». Por outro lado, tomando os resultados do inquérito pelo seu valor facial, chegaríamos à conclusão de que os seguidores de outras religiões não cristãs são menos felizes do que os católicos: 11,1% confessam-se «nada felizes». Acima de todos em termos de infelicidade estariam os protestantes: 33,3% têm esse mesmo sentimento de ausência total de felicidade³.

Se, por outro lado, agregarmos as categorias dos que se dizem «pouco» ou «nada» felizes e as cruzarmos pela sua posição religiosa, verificamos que os católicos não têm as mais altas percentagens. Contra 43,8% de católicos pouco ou nada felizes, há 61,9% de protestantes e outros cristãos e 55,5% de sequazes de outras religiões na mesma situação. Apenas os que não têm nenhuma religião têm menor percentagem (28,3%) do que os católicos⁴. Em contraponto com isto, temos as pessoas que se

³ Os cálculos aqui apresentados utilizam todos os casos referidos na sondagem ICS (Cabral *et al.* 2000) sem os ponderar. A ponderação apenas ocultaria problemas de representatividade que a amostra efectivamente tem.

⁴ Os valores dos testes estatísticos para a tabela de contingência das quatro categorias de religião («protestante/outra cristã», «católica», «outra não cristã» e «nenhuma») pelos quatro graus de felicidade («muito feliz», «razoavelmente feliz», «pouco feliz» e «nada feliz») são os seguintes: $\chi^2 = 33,807$; g.l. = 9; $p = 0.000$; phi: 0,169; contingência: 0,167; Gamma: -0,227; Spearman: -0,090.

consideram «muito felizes»: protestantes e outros cristãos: 14,3%; católicos: 19,4%; sem religião: 19,6%; outros não cristãos: 33,3%. Neste aspecto, as pessoas sem religião não estão melhor do que os católicos e são menos felizes do que os seguidores de outras religiões. Por outro lado, no que se refere a esta última categoria, o resultado — 33,3% de pessoas felizes — introduz uma contradição com os 55,5% «pouco» e «nada» felizes do mesmo grupo.

Seja como for, estes dados não sustentam a associação estabelecida entre catolicismo e infelicidade. O valor facial dos dados diz que o catolicismo está associado a menor infelicidade do que as religiões que não o catolicismo e que, se os urbanos elitistas atingem maiores percentagens de «razoavelmente felizes» (52,2%), os católicos têm 36,8% na mesma categoria, contra os 23,8% dos protestantes e os 33,3% das religiões não cristãs. No entanto, a estatística conjunta dos «muito» e «razoavelmente» felizes é favorável aos urbanos irreligiosos e elitistas, já que 71,8% estão nestas duas categorias, contra 56,2% de católicos, 44,4% de outras religiões e 38,1% de protestantes. Estes resultados alertam-nos para algumas contradições que interessará aprofundar, designadamente no sentido de verificar se não há outras variáveis mais fortemente definidoras das variações nos sentimentos de felicidade.

A IRRELEVÂNCIA DAS PRÁTICAS RELIGIOSAS NA FELICIDADE

Convém insistir que não se pode desvincular a interpretação de qualquer sondagem das considerações metodológicas feitas acima. As relações entre os elementos constituintes de uma teoria dependem do estatuto teórico das variáveis, querendo isto dizer que as causas ou os condicionamentos dos fenómenos podem ser ou não mediados. Não se pode fazer uma teoria da implicação, da associação ou da causalidade se não se tiver uma ideia acerca de quão mediata ou imediatamente os sentimentos, as atitudes e os comportamentos estão relacionados com variáveis de natureza mais ou menos objectiva. Por outras palavras, uma boa estratégia de análise obriga a clarificar pelo menos três questões: 1. o modelo utilizado: causalidade, implicação, ou associação; 2. o estatuto teórico das variáveis e a cadeia de relações supostas; 3. a retroacção dos efeitos.

Não poderei obviamente desenvolver aqui tudo o que está implicado nestas questões. Mas não tenho dúvidas em dizer que as variáveis materiais, designadamente aquelas que têm a ver com as condições de vida, podem aparecer aos actores sociais como mais directamente

associadas aos seus sentimentos de felicidade ou infelicidade do que as relacionadas com constructos culturais abstractos. Nas coisas sociais, o que parece é. Os factos do dia-a-dia são governados por teorias implícitas elaboradas por cada um dos actores sociais. O que se não toca dificilmente se interioriza e teoriza. As apreciações e avaliações da situação decorrem de comparações que os actores estabelecem com o que têm à sua volta.

Aplicando estes pressupostos, dir-se-ia, de forma simplificada, que uma situação penosa não favorece a apreciação da felicidade, muita ou pouca, que porventura se tenha. Ao teórico implícito que é cada inquirido — a matéria prima de análise é o que pensa, sabe e diz de si mesmo — as variáveis de natureza cultural, entre as quais as religiosas, são demasiado longínquas para que reconheça a sua importância estratégica na definição da sua felicidade, sendo, como são, mediadas pelas condições materiais de vida. Ou seja, as características psicológicas dos actores e as suas atitudes têm a sobredeterminá-las o sistema social, que delas constitui a referência imediata. A cultura, por seu lado, interfere na diferenciação dos comportamentos ao permear as estruturas com as legitimações e justificações que levam à aceitação do sistema.

Se assim é, compreende-se que as concepções e atitudes religiosas, constitutivas da referência última do sistema cultural, tenham importância própria, mas não muito manifesta ou aparente. Em meu entender, só há dois casos em que a religião poderia ser claramente determinante do sentimento de felicidade: (i) quando a vida social fosse determinada por uma moral anti-humana, imposta por instituições sociais totalitárias, que conduzissem ao total domínio do indivíduo; (ii) ou quando, no outro extremo, a união mística determinasse a vida das pessoas, o que configuraria a suma felicidade. Fora destas situações, a relação entre felicidade e religião é sempre equívoca.

Estas afirmações levam-me a considerar que a clarificação dos conceitos utilizados é tanto mais necessária quanto os dados são imprecisos e relacionados entre si de forma não inteiramente perceptível ou mensurável. Mas esta tarefa nem sempre é possível, dada a natureza difusa de alguns comportamentos.

Segundo o desígnio metodológico implicado nestes considerandos, procedeu-se ao estudo da significância das variações do sentimento de felicidade, tal como foi expresso pelos respondentes ao inquérito em causa, em função de algumas variáveis que concretizam a teoria aqui desenvolvida. Os resultados obtidos confirmam as dúvidas levantadas no ponto anterior, onde no entanto verificámos haver significância na distribuição dos sentimentos de felicidade em função da religião professada.

Um dos aspectos em que a pertença religiosa pode ser mais claramente detectada, em termos das questões do inquérito, são as idas à igreja, que admitiam as seguintes categorias: «1 ou mais vezes na semana», «2 ou 3 vezes na semana», «1 vez por mês, algumas vezes no ano», «menos de algumas vezes por ano» e «nunca». No que se refere aos que se sentem «muito felizes», segundo tais categorias, temos, por ordem, as seguintes percentagens: 21,5%, 17,6%, 10,3%, 20,0 e 15,6%. As pessoas do primeiro grupo são, pois, as mais felizes, seguidas de perto apenas pelas que vão à igreja menos de uma vez por ano. Por outro lado, as pessoas que nunca vão à igreja não são as mais felizes. Paralelamente, as pessoas que se dizem «nada felizes» atingem, pela ordem referida, os seguintes valores: 9,8%, 8,4%, 10,3%, 7,3%, 11,2% e 3,1%. Por estes dados se vê serem as pessoas que vão menos de uma vez por ano à igreja as menos felizes⁵.

O que ressalta destes dados, e dos não comentados relativos às categorias intermédias (razoavelmente e pouco feliz), é que as diferenças são demasiado pequenas para serem significativas. Nenhum dos testes aplicáveis às variáveis ordinais tem significância. A probabilidade dos testes *d* de Somer, *Tau b* e *Tau c* de Kendall, Gamma e Spearman é de 40%. Os valores Gamma e Spearman são respectivamente -0.027 e -0.024, não sustentando, por isso, nenhuma conclusão sobre qualquer associação ou sobre o seu sentido.

FACTORES NÃO RELIGIOSOS DA FELICIDADE

A hipótese formulada anteriormente a respeito da influência das condições de vida nos sentimentos de felicidade pode ser avaliada através da associação entre estes sentimentos, por um lado, e a classe social e o rendimento familiar, por outro, sendo estas duas variáveis tomadas como expressões das condições de vida, a primeira mais mediata do que a segunda. Segundo a hipótese, o rendimento familiar estaria mais fortemente associado à felicidade do que a classe social, o que efectivamente se verifica, como passo a mostrar, anotando antes, para o bom entendimento dos resultados referentes à classe social, que ela não foi definida *a posteriori* pelos investigadores, mas resulta da autotransclassificação

⁵ Tomando esta distribuição, juntamente com a das categorias «razoavelmente feliz» e «pouco feliz», obtemos os valores: $\chi^2 = 19.835$; g.l. = 15; $p = 0.178$. Note-se ainda que as duas primeiras categorias estão invertidas, em termos de ordem lógica, o que para o caso não é irrelevante.

dos inquiridos. Este facto introduz, por si só, alguma imprecisão nos dados. Mas isso não é específico daquela variável. Também os sentimentos de felicidade ou infelicidade que os inquiridos dizem ter resultam de classificações fluidas.

Quadro 1 – Percentagens de sentimentos de felicidade segundo a classe social subjectiva

Grau de felicidade	Classe baixa	Classe trabalhadora	Classe média baixa	Classe média	Classe média alta	Classe alta	Total
Muito feliz	5,6	22,1	21,9	21,4	34,6	3,8	19,7
Razoavelmente feliz	24,3	32,0	32,1	50,3	50,0	50,0	37,3
Pouco feliz	50,7	38,5	39,4	23,7	15,4	26,9	35,0
Nada feliz	19,4	7,4	6,6	4,6	0,0	19,2	8,1

Fonte: Cálculos sobre Cabral 2000

O certo é que a percentagem de pessoas da classe média alta muito felizes é cerca de 7 vezes maior que a da classe baixa. Por outro lado, a percentagem das classes trabalhadora, média baixa e média, muito felizes é muito semelhante, possivelmente porque as diferenças de classe também não são importantes. O reverso da medalha, relativo aos pouco felizes, também é verdadeiro, o que confirma a hipótese: as pessoas da classe baixa são confessadamente muito mais infelizes do que as das outras classes, excepção feita à classe alta, fundamentalmente constituída pelos urbanos elitistas referidos acima, o que parece contraditório. Todos os testes mostram que as variações são estatisticamente significativas⁶.

O segundo indicador das condições de vida é o rendimento familiar. Em termos hipotéticos, é de admitir uma associação positiva entre felicidade e rendimento e que esta associação é mais importante que a existente entre felicidade e classe social, dado que, por um lado, a definição da classe é mais fluida do que o rendimento e, por outro, incorpora muito mais componentes, o que se pode traduzir em efeitos contraditórios em relação aos sentimentos de felicidade. No conjunto, estas propostas interpretativas confirmam-se.

⁶ As estatísticas são as seguintes: $\chi^2 = 115.965$; g.l.= 15; p = 0.000; Gamma: -0.252; Spearman: -0.209.

Quadro 2 – Percentagens de sentimentos de felicidade segundo o rendimento familiar (em contos)

Grau de felicidade	< 31 c.	31-60 c.	61-100 c.	101-160 c.	161-300 c.	> 300 c.	Total
Muito feliz	5,1	13,8	18,5	19,5	23,9	29,9	19,7
Razoavelmente feliz	25,6	22,2	29,2	42,5	44,5	51,5	36,8
Pouco feliz	48,7	44,3	41,3	35,5	27,3	17,5	35,4
Nada feliz	20,5	19,8	11,0	2,6	4,2	1,0	8,0

Fonte: Cálculos sobre Cabral 2000

O quadro 2 mostra, com efeito, que a percentagem de pessoas que se consideram muito felizes cresce monotonicamente com o rendimento familiar. A percentagem de muito felizes com mais de 300 contos de rendimento é cerca de 6 vezes maior do que a de pessoas com menos de 31 contos. Exactamente o contrário aparece na categoria «nada feliz». Usando tal indicador dir-se-ia, forçando um pouco os termos, que as pessoas com menos de 31 contos de rendimento são vinte vezes menos felizes do que as que têm 300 contos ou mais de rendimento⁷.

É, por outro lado, de assinalar o facto de os inquiridos, na sua globalidade, se confessarem mais felizes do que infelizes: a percentagem daqueles que se dizem nada felizes (8,0%) é mais de duas vezes inferior à daqueles que se dizem muito felizes (19,7%). Regista-se, por outro lado, que as variações em causa não decorrem de erros intrínsecos à amostra.

Este conjunto de resultados relativos à significância das distribuições e referentes às três variáveis estudadas (frequência das idas à igreja, classe social e rendimento) levou a verificar que tanto o valor de χ^2 como o das estatísticas Gamma e Spearman crescem em função da incorporação das condições materiais de existência nas variáveis utilizadas.

Assim, o rendimento seria a variável com mais força explicativa, como se pode verificar pela síntese das estatísticas apresentadas no quadro 3. A seguir viria a classe social. No pólo da insignificância estaria a frequência da igreja, a qual não diferenciaria suficiente e significativamente as pessoas em termos de sentimentos de felicidade.

⁷ As estatísticas são as seguintes: $\chi^2 = 125.452$; g.l.= 15; p = 0.000; Gamma: -0.313; Spearman: -0.277.

Quadro 3 – Testes de associação entre variáveis da base de dados ICS

Associação	χ^2	g.l.	p	Gamma	Spearman
Felicidade/frequência da igreja	19,84	15	0,18	-0,027	-0,024
Felicidade/classe social subjectiva	115,97	15	0,00	-0,252	-0,209
Felicidade/rendimento familiar	125,45	15	0,00	-0,313	-0,277

CONCLUSÃO

Constituindo estes breves apontamentos apenas o início de um possível estudo sobre as condições da felicidade, não poderemos tirar conclusões definitivas sobre os méritos dos processos utilizados. Mas os resultados obtidos permitem chamar a atenção para o facto de os pressupostos e os métodos de análise terem óbvias consequências nas teorias desenvolvidas, sobretudo em matérias onde o efeito de Condorcet ou o paradoxo de Simpson espreitam continuamente a desatenção do analista. Por outro lado, o exemplo estudado parece mostrar que, em termos da utilização da matemática na construção das teorias sociais, as teias do significado só podem ser tecidas por Penélopes incansáveis que, no fazer e refazer constante das teorias, vão encontrando novos sentidos — o mais das vezes ocultos, como em Delfos — que subjazem aos comportamentos sociais, em seus enquadramentos institucionais e sobredeterminações culturais.

O percurso feito — da crítica da construção do objecto à sua interpretação — revelou alguns dos escolhos que se encontram em todas as investigações relativas às relações entre ideias, sentimentos, atitudes e suas condicionantes sociais. Por outro lado, mostrou que as relações entre os factos podem ser obnubiladas por variáveis compósitas e de imprecisa definição, sendo a análise perturbada pela incompatibilidade das opções e pela contaminação contraditória de dependências culturais múltiplas.

O eventual mérito da análise realizada estará no estabelecimento de relações entre variáveis singelas, onde as associações podem ser mais claramente entrevistas. A partir delas, outras mais complexas poderão ser estudadas, devendo no entanto atender-se sempre a que os resultados não podem ser contraditórios nem afastar-se muito do senso comum, que nunca é tão erróneo como uma análise superficial pode fazer crer.

Por outro lado, as associações descobertas parecem encaminhar a interpretação para formulações conformes com a teoria disponível. Além disso, estas aproximações simples parecem indicar que a complexidade

dos métodos não garante a correção do que não pode ser sustentado pela teoria.

Mas a questão de fundo, aqui tocada muito ao de leve, é que o conhecimento científico das causalidades, implicações ou associações entre ideias, sentimentos, atitudes, comportamentos, determinações institucionais, traços culturais, valores e símbolos sociais — a teia sócio-cultural — não se compadece com leituras apressadas. É fruto de uma longa paciência. E o seu principal instrumento é a matemática, onde a lógica é levada às últimas consequências.

Referências

- Alker, JR., Hayward R., Karl W. Deutsch, Antoine H. Stoetzel, eds. (1973) – *Mathematical Approaches to Politics*, Amsterdão, Elsevier.
- Berk, Richard A. (1988) – Causal Inference for Sociological Data, in Neil J. Smelser, ed., *Handbook of Sociology*, Londres, Sage, pp. 155-172.
- Borg, Igwer, Samuel Shye (1995) – *Facet Theory, Form and Content*, Londres, Sage.
- Boudon, Raymond (1971) – *Les mathématiques en sociologie*, Paris, PUF.
- Brass, W. (1974) – Perspectives in Population Prediction, Illustrated by the Statistics of England and Wales, *Journal of the Royal Statistical Society, Series A*, 137 (4), pp. 532-570.
- Cabral, Manuel Villaverde, Jorge Vala, José Machado Pais e Alice Ramos (2000) – *Atitudes e Práticas Religiosas dos Portugueses, Base de dados 2*, Lisboa, Imprensa de Ciências Sociais.
- Coale, Ansley J., T. James Trussell (1978) – Technical Note: Finding the Two Parameters that Specify a Model Schedule of Marital Fertility, *Population Index*, 44, pp. 203-213.
- Coleman, James S. (1964) – *Introduction to Mathematical Sociology*, Glencoe, The Free Press.
- Cox, D. R. (1992) – Causality: Some Statistical Aspects, *Journal of the Royal Statistical Society, Series A*, 155 (2), pp. 291-301.
- Halley, Edmund (1693) – An estimate of the degree of the mortality of mankind, in David Smith, Nathan Keyfitz, eds., *Mathematical Demography, Selected Papers*, Berlim, Springer-Verlag, 1977, pp. 21-26.
- Lazarsfeld, Paul F. (1972) – Some Remarks on Typological Procedures in Social Research, in suo *On Social Research and its Language*, Chicago, The University of Chicago Press, pp.158-172.
- Lotka, Alfred J. (1998) – *Analytical Theory of Biological Populations*, Nova Iorque, Plenum Press, 220 p. [1.^a ed.: 1934].

- Merton, Robert K. (1957) – *Social Theory and Social Structure*, Glencoe, The Free Press.
- Miller, W. L. (1995) – Quantitative Methods, in David Marsh & Gerry Stoker, eds., *Theory and Methods in Political Science*, Londres, MacMillan, pp. 154-172.
- Moles, Abraham (1995) – *As ciências do impreciso*, Porto, Afrontamento.
- Pais, José Machado (2000) – *O que explica a religiosidade dos portugueses, Um ensaio de análise tipológica*, manuscrito, 48 p.
- Russell, Bertrand (1972) (1903) – *The Principles of Mathematics*, Londres, G. Allen & Unwin.
- Sanders, David (1995) – Behavioral Analysis, in David Marsh & Gerry Stoker, eds., *Theory and Methods in Political Science*, Londres, MacMillan, pp. 58-75.
- Smithson, Michael (1987) – *Fuzzy Set Analysis for Behavioral and Social Sciences*, Nova Iorque, Springer-Verlag.
- Thomas, W. I., & F. Znaniecki (1998) – *Le paysan polonais en Europe et en Amérique, Recit de vie d'un migrant (Chicago 1919)*, trad. de Ives Gaudillat, Paris, Nathan.
- Weber, Max (1949) – *The Methodology of the Social Sciences*, trad. E. Shils & H. Finch, Nova Iorque, The Free Press.

David M. Holland
Instituto Courant de Ciências Matemáticas
Universidade de Nova Iorque, EUA

Ciência global do ambiente e matemática aplicada moderna¹

RESUMO

A ciência global do ambiente é uma actividade interdisciplinar baseada em conhecimentos de física, química, biologia e outras disciplinas. Utiliza a matemática aplicada moderna e as ciências da computação para dispor de linguagem e ferramentas através das quais as teorias científicas se desenvolvem e são confrontadas com as observações obtidas directamente na própria natureza. Este artigo examina alguns dos complexos e delicados equilíbrios responsáveis pelo actual estado climático do ambiente físico do planeta Terra, bem como o papel da matemática aplicada moderna como auxiliar na formulação e resolução adequadas de modelos realistas do sistema Terra. Apresentam-se sucintamente os últimos avanços da ciência global do ambiente, bem como as possíveis direcções de desenvolvimento futuro desta ciência e as crescentes oportunidades de interacção com a área da matemática aplicada moderna. A história da Terra fornece indicações de que ocorrem grandes mudanças nas componentes do sistema climático, nomeadamente na Terra sólida, nos oceanos, nas calotes de

¹ Tradução de José Miguel Urbano, professor do Departamento de Matemática, Universidade de Coimbra. Revisão de Manuel Maria Godinho, professor do Departamento de Ciências da Terra, Universidade de Coimbra.

gelo, nos mantos de gelo que cobrem alguns mares, na atmosfera e na própria vida. Começando pelo processo de formação do planeta, descrevem-se diversos fenómenos que ocorrem num largo espectro de escalas temporais, desde a lenta deriva dos continentes às rápidas mudanças diárias no estado do tempo. É importante estudar as componentes individuais do sistema climático da Terra, mas é também de enfatizar o facto de existirem interações complexas entre essas componentes. Estas interações podem controlar, de modos subtils, o que à primeira vista poderiam parecer fenómenos inexplicáveis — alguns exemplos notáveis são as glaciações, o aquecimento global e o El Niño. A investigação científica destes ou doutros fenómenos correlativos deu origem a algumas das mais complexas equações com derivadas parciais encontradas até hoje, relativamente às quais as técnicas de resolução e as propriedades das soluções ainda não são completamente conhecidas.

1. INTRODUÇÃO

Neste artigo faremos uma revisão pormenorizada das propriedades das cinco «esferas» que fundamentalmente constituem o nosso ambiente global, destacando em simultâneo a relação entre a compreensão dos processos que ocorrem nessas esferas e a aplicação das poderosas ferramentas da matemática aplicada moderna. Começando pelo interior da Terra, e deslocando-nos para o exterior, essas esferas são: a litosfera, a hidrosfera, a criosfera, a atmosfera e a biosfera, cuja composição essencial é, respectivamente, de rocha, água, gelo, ar e matéria viva. Em termos técnicos, a designação *litosfera* corresponde apenas à parte mais externa da Terra sólida, mas neste artigo usá-la-emos também para referir indirectamente os processos que ocorrem a um nível mais profundo.

Ao tratar matematicamente problemas das ciências do ambiente que envolvam estas cinco esferas, parte da sua complexidade deriva do largo espectro de escalas de tempo e de espaço envolvidas. Quanto às escalas temporais, esboça-se na figura 1 um espectro que se estende de dias até éons. Nesta figura indicam-se alguns exemplos em que se associam certos fenómenos a determinadas escalas de tempo. Sendo geralmente verdade que os processos na litosfera ocorrem ao longo de grandes intervalos de tempo, e na atmosfera durante curtos intervalos, tal verdade não é rígida e há muitas excepções a esta classificação simplista dos processos. No entanto, ela permite evidenciar que em qualquer tentativa para construir um modelo matemático que abarque a totalidade do sistema Terra é necessário ter em conta o problema do espectro incrivelmente largo de escalas de tempo em que os processos ocorrem.

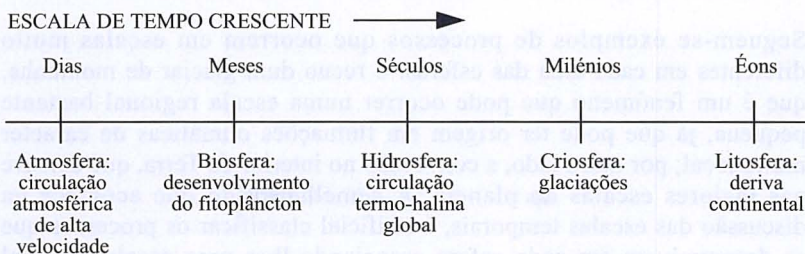


Fig. 1 — Esquema do espectro de escalas de tempo tipicamente encontradas em problemas de ciências da Terra

Para além deste aspecto, os processos físicos que ocorrem nas várias esferas interagem uns com os outros, tornando o problema, em certa medida, menos «separável» em termos de escalas de tempo do que seria de supor à primeira vista. Por exemplo, considere-se a circulação do carbono através do sistema Terra, um elemento que está actualmente no centro da polémica em torno do aquecimento causado pelo efeito de estufa. A ciclagem do carbono envolve o seu transporte, relativamente rápido, sob a forma de dióxido de carbono através da atmosfera, influenciando claramente o balanço da radiação infravermelha da Terra, a sua captação pela vida vegetal, tanto nos continentes como nos oceanos, a sua subsequente acumulação em sedimentos nos fundos oceânicos, a sua subducção para o interior do manto por efeito de forças tectónicas, e, finalmente, a sua libertação para a atmosfera através de erupções vulcânicas, repetindo-se este ciclo a partir da atmosfera. Resumindo, o carbono circula através de quase todas as componentes do sistema Terra e em quase todas as escalas de tempo concebíveis.

Apesar do que acaba de se expor, há situações em que a separação por escalas de tempo é de facto bastante apropriada, dependendo, como é óbvio, dos pormenores do problema a tratar. Esta simplificação torna o estudo de alguns problemas em ciências do ambiente passíveis de um tratamento matemático. O exemplo mais conhecido é, provavelmente, o da previsão meteorológica, em que podem ser feitas, com alguma precisão, previsões sobre a evolução da atmosfera em escalas de tempo de alguns dias, ainda que se não considerem as variações no estado das outras esferas que compõem o sistema Terra.

A consideração de escalas espaciais assume a mesma importância que a de escalas temporais. Na figura 2 mostra-se um espectro de escalas que abrange a escala relativamente pequena de uma dada «região», a da superfície dos continentes e do planeta e, por fim, a do volume deste.

Seguem-se exemplos de processos que ocorrem em escalas muito diferentes em cada uma das esferas: o recuo dum glaciador de montanha, que é um fenómeno que pode ocorrer numa escala regional bastante pequena, já que pode ter origem em flutuações climáticas de carácter muito local; por outro lado, a convecção no interior da Terra, que decorre nas maiores escalas do planeta. À semelhança do que acontece na discussão das escalas temporais, é artificial classificar os processos que se desenvolvem em cada esfera associando-lhes uma escala espacial rígida; só é útil no sentido em que permite evidenciar a larga gama de escalas espaciais em que os processos ocorrem.

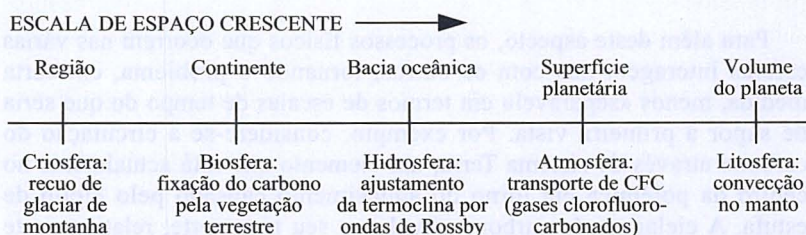


Fig. 2 — Esquema de espectro de escalas de espaço tipicamente encontradas em problemas de ciências da Terra

No que segue discutiremos alguns fenómenos interessantes que ocorrem em cada uma das cinco esferas (secção 2) e posteriormente apresentaremos algumas das ferramentas matemáticas (secção 3) desenvolvidas e aplicadas na resolução deste tipo de problemas. Na secção 4 serão apresentadas algumas conclusões.

2. A TERRA

O sistema Terra pode, num sentido muito lato, ser descrito como uma grande bola de rocha (a litosfera) que possui uma fina camada de água a cobrir parcialmente a sua superfície (a hidrosfera) e uma fina camada de gelo a cobrir, também parcialmente, a sua superfície (a criosfera), estando todas estas esferas envolvidas por ar (a atmosfera). A vida (a biosfera) existe em todas elas; provavelmente merecerá menos atenção a vida que existe em condições extremas — desde as colónias de bactérias que vivem em profundidade a temperaturas relativamente elevadas, até às algas que prosperam sob o manto de gelo de alguns mares. Analisaremos agora com

mais pormenor os processos físicos e biológicos que ocorrem nestas cinco esferas; como motivação de fundo, encaramos estes fenómenos únicos como problemas altamente relevantes do ponto de vista da aplicação das ferramentas da matemática aplicada moderna.

2.1 A litosfera

Tecnicamente, a palavra «litosfera» refere-se à camada externa rígida da Terra, situada acima da astenosfera; é constituída pelas crostas continental e oceânica e pelo manto superior, estendendo-se até uma profundidade que varia entre 100 km e 200 km. Para mais pormenores sobre os processos que ocorrem na Terra sólida e respectiva nomenclatura remetemos o leitor para a excelente monografia de Press e Siever (1978). Tendo em vista o nosso objectivo, salientamos o facto de a Terra abaixo de nós ser constituída essencialmente por rocha sólida, e de esta rocha sofrer transformações complexas em escalas de tempo geológico, as quais provocam alterações de fundo a nível da existência de água, gelo, ar e vida neste planeta. O inverso também é verdadeiro, já que, em certa medida, a água, o gelo, o ar e a vida podem erodir ou alterar as rochas expostas à superfície.

O melhor conhecimento da estrutura da Terra sólida (ver figura 3) revelou a existência de um núcleo metálico interno rodeado por um manto

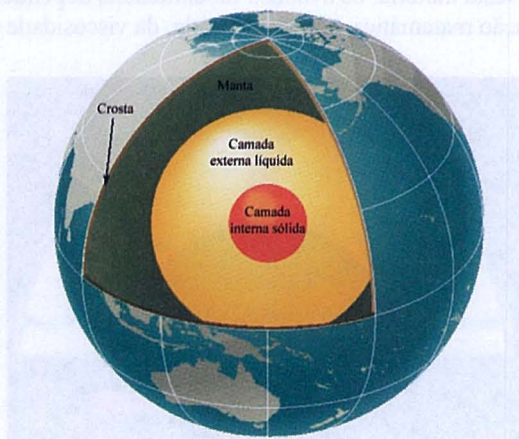


Fig. 3 — Esquema tridimensional, em corte, da Terra sólida (segundo Song e Richards, 1996). Observa-se o núcleo metálico (com uma camada interna sólida e uma externa líquida), o manto e a crosta

de rocha, a que se segue, no exterior, uma crosta, onde residem as bacias oceânicas e as massas continentais. Concluiu-se que os processos que se desenvolvem em cada uma destas três camadas da Terra sólida desempenham um papel fundamental na evolução de todo o sistema Terra. Sem dúvida que a revolução científica que anunciou a introdução e a aceitação da teoria das placas litosféricas oferece o exemplo mais evidente da importância destes processos. O facto de a crosta oceânica estar constantemente a ser consumida e regenerada e, em consequência, de a superfície dos continentes estar a mover-se, tem como resultado directo o afundamento de rochas (que são cristalinas) até uma profundidade de centenas de quilómetros. A figura 4 mostra um exemplo de convecção no manto; a cor vermelha simboliza material quente a subir em direcção à superfície, a azul representa material superficial mais frio a descer para o interior.

A escolha deste exemplo deve-se ao facto de ele representar adequadamente uma ampla classe de problemas do âmbito da ciência do ambiente, na resolução dos quais a metodologia da matemática aplicada moderna pode ser útil. No sentido mais geral, a classe de problemas a que nos referimos é aquela que envolve o movimento convectivo tridimensional de fluidos diversos. No caso específico da convecção no manto, o fluido é tratado como não-newtoniano, sendo, neste caso particular, apropriado considerá-lo com as propriedades reológicas da olivina. Há um tema que, pela sua importância, é comum a quase todas as subdisciplinas da ciência do ambiente: é o estudo dos movimentos convectivos. Nesta matéria, os avanços fundamentais dependem de se fazer uma interpretação matemática, com significado, da viscosidade de um fluido.

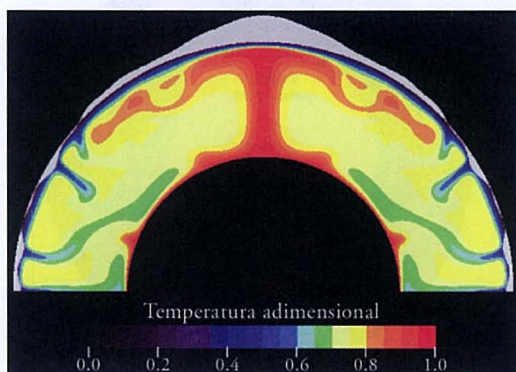


Fig. 4 — Representação parcial (pressuposta simétrica) dum corte no manto sólido, com plumas convectivas ascendentes (vermelho) e descendentes (azul) (segundo Kiefer e Kellog, 1998)

2.2 A hidrosfera

A hidrosfera é composta não só pela água das bacias oceânicas que constituem o oceano mundial, mas também pela dos numerosos lagos e rios e pelos recursos hídricos subterrâneos que existem nos continentes. Para o nosso objectivo, o termo é tomado apenas na acepção de água no estado líquido. O maior volume individual de água é, de longe, o reservatório do oceano mundial, não obstante tratar-se duma mistura salina. À superfície desse reservatório o fluido move-se por influências atmosféricas, quer mecanicamente, devido à força exercida pelo vento, quer termodinamicamente, por efeito directo do calor e da transferência de massa devida a variações de densidade. Ao contrário da convecção no manto, que é provocada por um aquecimento a partir de baixo, a convecção oceânica de grande escala é originada pelo arrefecimento a partir de cima, especialmente em regiões de latitude elevada. Um dos últimos grandes avanços em oceanografia foi a identificação de um padrão convectivo cíclico de escala global, actualmente designado por circulação termo-halina global.

Com base em simulações feitas através de modelos computacionais e em dados de sondagens efectuadas nas calotes geladas, crê-se que o padrão convectivo actual da circulação oceânica global terá sido muito diferente no passado, o que deixa antever que haverá múltiplos equilíbrios neste sistema de origem convectiva. Estamos, pois, em presença dum sistema dinâmico muito rico no sentido matemático.

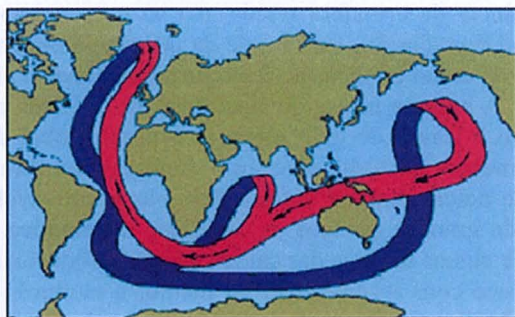


Fig. 5 — Esquema da circulação termo-halina global (segundo Broecker, 1991). A circulação convectiva é originada em vários locais do Atlântico Norte, onde a água fria e salina da superfície se afunda, em virtude da sua densidade, movimentando-se lentamente para sul em profundidades abissais e entrando no Índico e no Pacífico (percurso a azul); nestes oceanos ascende, aquece, e retorna por níveis superficiais ao Atlântico Norte (percurso a vermelho)

Não obstante serem de salientar as semelhanças entre as esferas que compõem o sistema Terra, são também de referir óbvias e profundas diferenças. Um dos caracteres que marcam a diferença entre as descrições matemáticas da litosfera e as da hidrosfera é a existência de turbulência nesta última. Além disso, essa turbulência existe num ambiente altamente rotacional, que possui a capacidade surpreendente de gerar estruturas coerentes em larga escala a partir de escoamentos aparentemente desorganizados e turbulentos. Esta propriedade de auto-organização é, de facto, confirmada pelas equações utilizadas para descrever os escoamentos oceânicos e atmosféricos, mas o conjunto completo de soluções para o escoamento permanece por explorar devido ao largo espectro de regimes de escoamento que podem ocorrer nos oceanos e na atmosfera.

2.3 A criosfera

O abaixamento da temperatura à superfície do planeta com o aumento da altitude ou da latitude conduz à congelação de grandes reservatórios de água, formando o gelo que constitui a nossa criosfera. São subcomponentes da criosfera as enormes calotes de gelo da Antártida e da Gronelândia; a cobertura de gelo do Ártico e dos mares do Sul, com flutuações sazonais; os glaciares de montanha que fluem por encostas a grande altitude; o *permafrost* que existe em grandes extensões continentais; e o manto de neve, de espessura e extensão variáveis com as estações, que cobre vastas áreas continentais no hemisfério norte.

A importância da criosfera reside não só no seu impacto sobre o balanço global da radiação, em virtude da influência directa que tem no albedo da Terra, como também no facto de condicionar o nível do mar à escala global. A enorme massa de água doce armazenada no continente antártico sob a forma de gelo elevaria o nível global do mar umas dezenas de metros se todo esse gelo fundisse subitamente — um acontecimento potencialmente catastrófico mas improvável.

Tem havido grandes avanços na descrição matemática do manto de gelo que cobre alguns mares e das calotes geladas. Todavia, à semelhança do que acontece com a convecção no manto, a evolução temporal da criosfera está necessariamente dependente da descrição precisa da viscosidade do fluido, neste caso gelo sujeito a uma gama de pressões e temperaturas moderadas. Esta viscosidade é ainda mais difícil de descrever quando se trata dum fluido tão descontínuo como este. O manto de gelo de alguns mares apresenta à superfície fendas mais ou menos extensas; o mesmo acontece com o gelo que constitui a componente terrestre da

criosfera, as calotes geladas. Têm sido realizados vários estudos que consideram o gelo um material granular (ver figura 6), por forma a lidar com a natureza invulgarmente descontínua deste meio fluido (Hopkins, 1996).

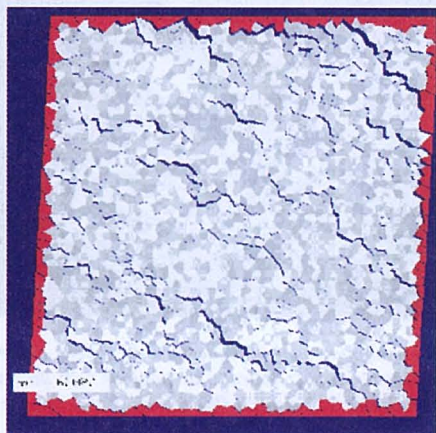


Fig. 6 — Tratamento, como material granular, do gelo de cobertura de um mar. Instantâneo duma simulação computacional em que um bloco de gelo tomado como modelo é sujeito a forças de cizalhamento. As áreas a azul mostram o aparecimento de fracturas de grande escala. Segundo Hopkins (1996)

2.4 A atmosfera

O fino véu de gases que envolve a superfície do nosso planeta e constitui a atmosfera pode considerar-se a componente mais bem estudada do sistema Terra. Isto resulta, em larga medida, do facto de haver uma rede global de observações que forneceu aos investigadores um conjunto relativamente rico de dados, os quais possibilitaram a construção de teorias matemáticas explicativas dos movimentos da atmosfera. Os modelos computacionais actuais, construídos após décadas de investigação cuidada, revelam uma notável capacidade para simular o comportamento real da atmosfera; a capacidade para simular a cobertura de nuvens é disto um claro exemplo (ver figura 7). No entanto, continua a colocar-se o problema do tratamento matemático de fenómenos de escoamento que ocorrem em escalas espaciais e temporais abaixo das admitidas por um modelo computacional, continuando, assim, por resolver a questão do tratamento de fenómenos que ocorrem numa escala abaixo de certo limite.

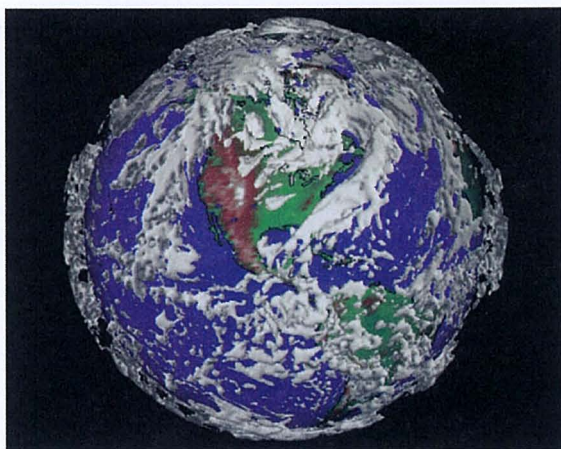


Fig. 7 — Simulação computacional da cobertura de nuvens a partir de um modelo atmosférico de circulação global. Retirada de páginas web do National Center for Atmospheric Research (<http://www.scd.ucar.edu>)

A atmosfera comporta uma grande variedade de instabilidades, que têm sido amplamente analisadas de um ponto de vista matemático relativamente rigoroso. A mais conhecida destas instabilidades é a chamada instabilidade baroclínica, um processo responsável pelas variações na localização das correntes atmosféricas de alta velocidade² que se movimentam sobre os hemisférios norte e sul. Há outras instabilidades ainda não completamente compreendidas, nomeadamente a que respeita ao processo explosivo que conduz à formação de tornados e tufões. A compreensão destes fenómenos numa perspectiva puramente matemática constitui um enorme desafio.

2.5 A biosfera

A biosfera é claramente a esfera mais distinta do sistema Terra, já que a complexidade dos mecanismos que governam o seu funcionamento é muito mais intrincada do que a observada em qualquer das outras esferas, caracterizadas essencialmente por mecanismos físicos. A compreensão da evolução e das flutuações da biosfera requer a compreensão e o

² Em língua inglesa: *jet stream*.

conhecimento profundos dos mecanismos químicos e biológicos, para além do arsenal físico e matemático correntemente usado em ciência do ambiente. A biosfera existe nos oceanos (ver figura 8) e nos continentes, e a sua descrição matemática tem progredido nos últimos anos. No entanto, um modelo matemático completo desta esfera e das suas interações com as outras componentes do sistema Terra está ainda por construir. A capacidade para formular um modelo matemático apropriado da biosfera e para prever a sua evolução continua a ser um desafio central na ciência do sistema Terra.

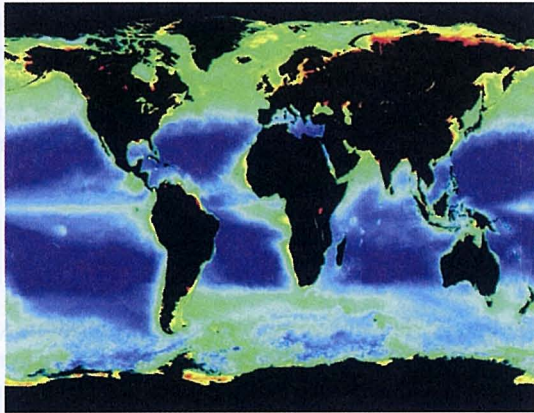


Fig. 8 — Distribuição da concentração de clorofila na biosfera oceânica, baseada em dados obtidos por satélite. As concentrações mais elevadas estão representadas por áreas a vermelho, as mais baixas a azul. Retirada de páginas da web da NASA (<http://earthobservatory.nasa.gov/>)

3. O QUADRO MATEMÁTICO

Uma discussão exaustiva do quadro matemático no qual a ciência do ambiente se desenvolveu ao longo das últimas décadas constitui matéria vasta que está muito para além do âmbito deste artigo. Podemos, contudo, descrever algumas características comuns aos diversos tipos de abordagem que foram bem sucedidos ao trazerem novas ideias e ao ajudarem a compreender o funcionamento do sistema Terra. Pode considerar-se estas características comuns dizem respeito a (i) modelos matemáticos, (ii) técnicas de solução, e (iii) propriedades das soluções. É o que a seguir se discutirá sucintamente.

3.1 Modelos matemáticos

Ao construir modelos matemáticos das várias componentes do sistema Terra, a maioria dos investigadores optou pela formulação dos problemas físicos recorrendo a equações com derivadas parciais (EDP). Esta abordagem resulta naturalmente da descrição das leis de conservação de massa, momento e energia, as quais são bastante semelhantes, não obstante se expressem de forma ligeiramente diferente nas várias subdisciplinas das ciências da Terra, como é o caso das equações da convecção no manto e na atmosfera. Recentemente, as equações diferenciais estocásticas (EDE) tornaram-se uma ferramenta corrente e têm sido aplicadas a problemas em que a acção sobre um sistema é descrita de modo mais preciso como processo aleatório. A título de exemplo, as previsões climáticas de longo prazo podem ser analisadas nesta perspectiva, considerando-se as flutuações de curto prazo no estado do tempo como essencialmente ligadas ao ruído aleatório do sistema.

3.2 Técnicas de solução

O desenvolvimento de técnicas analíticas de que resultem soluções para algumas das principais EDP da ciência do ambiente conheceu recentemente uma espécie de renascimento com os avanços significativos do *software* de álgebra computacional simbólica. Concretamente, estes avanços permitem hoje em dia obter expansões assintóticas e em série de potências de ordem muito mais elevada do que era anteriormente possível.

Nos últimos anos, os métodos de eleição utilizados para obter soluções de EDP na ciência do ambiente têm sido sobretudo os métodos numéricos. O crescimento explosivo das capacidades computacionais colocou os métodos espectrais e os das diferenças finitas e dos elementos finitos em lugar de destaque entre os muitos avanços que recentemente se verificaram na aplicação de técnicas matemáticas à ciência do ambiente. Alguns foram claramente ilustrados na secção 2. Por exemplo, a simulação da cobertura global de nuvens apresentada na figura 7 ilustra bem estes novos avanços.

3.3 Propriedades das soluções

Uma vez formulados os modelos matemáticos apropriados, e desenvolvidas as técnicas de solução adequadas, resta ainda um terceiro e muito importante aspecto do quadro matemático — a compreensão das

propriedades e do comportamento das soluções obtidas. Um exemplo elucidativo foi a descoberta de que mesmo um simples sistema de convecção térmica pode exibir as propriedades do caos. Outro exemplo é o de sistemas turbulentos capazes de se auto-organizarem, criando estruturas coerentes de escoamento em larga escala. Na verdade, é neste terceiro aspecto do quadro matemático que as poderosas ferramentas analíticas da matemática aplicada moderna podem ter um impacto bastante significativo no desenvolvimento de uma profunda compreensão do funcionamento do sistema Terra.

Outra ilustração mais específica deste facto, que não foi abordada na secção 2.1 a propósito dos processos litosféricos, é o fenómeno do campo magnético da Terra (o geodínamo), que resulta de correntes de convecção no núcleo exterior, composto de metal líquido; este caso constitui um problema matematicamente rico, que permanece, em larga medida, sem solução. Apesar de as equações que governam a magneto-hidrodinâmica num ambiente fluido convectivo serem conhecidas, a forma como as soluções poderão exibir oscilações auto-excitadas, compatíveis com a conhecida inversão não periódica do campo magnético da Terra, permanece algo misteriosa. É, sem dúvida, um exemplo, e há muitos outros, de um problema matemático extremamente rico e atractivo, que poderá, provavelmente, ser mais bem compreendido recorrendo explicitamente, por exemplo, às técnicas da teoria dos sistemas dinâmicos.

4. OBSERVAÇÕES FINAIS

Neste artigo foi apresentada uma visão geral de alguns dos principais avanços registados no campo da ciência do ambiente ao longo das últimas décadas. A ciência do ambiente foi definida de modo a abranger o incrivelmente largo espectro de processos físicos e biológicos, os quais ocorrem numa gama de escalas espaciais e temporais ainda mais larga. Como se referiu na abordagem desses processos, o espectro pode estender-se do lento processo convectivo que ocorre no manto à escala planetária e em intervalos de tempo de éons, ao rápido processo convectivo de pequena escala associado a uma trovoadas de fim de tarde. No entanto, apesar das diferenças de escalas, quer espaciais quer temporais, tais processos, aparentemente díspares, partilham muitos elementos. Será sobretudo através do desenvolvimento de um quadro matemático consistente para o estudo de propriedades mecânicas e termodinâmicas dos diversos fluidos que constituem o nosso planeta que melhor compreenderemos muitos dos processos responsáveis pela evolução do

ambiente global e suas interações. Há um futuro promissor para a aplicação das poderosas técnicas analíticas da matemática aplicada moderna aos problemas mais importantes das ciências da Terra, alguns dos quais foram descritos ao longo deste artigo. Novos avanços nas técnicas numéricas e analíticas, bem como o aperfeiçoamento de técnicas já existentes e bem desenvolvidas, importadas de outras áreas de aplicação irão provavelmente criar o ímpeto para grandes avanços e sucessos nas ciências do ambiente no decurso deste século.

AGRADECIMENTOS

O autor agradece reconhecido o apoio das bolsas OPP-9901039, OPP-9984966 e OPP-0084286 do Office of Polar Programs of the United States National Science Foundation e da bolsa NAG-5-8475 do Polar Research Program of the National Aeronautical Space Administration.

Referências

- Broecker, W. S. (1991) – The great ocean conveyor. *Oceanography*, 4, 79-89.
- Kiefer, W., and L. Kellogg (1998) – Geoid Anomalies and Dynamics Topography from time-dependent, Spherical Axisymmetric Mantle Convection. *Physics of the Earth and Planetary Interiors*, 106, 237-256.
- Hopkins, M. A. (1996) – On the mesoscale interaction of lead ice and floes. *Journal of Geophysical Research-Oceans*, 101, 18315-18325.
- Press, F., and R. Siever (1978) – Earth. W. H. Freeman and Company, São Francisco.
- Song, X., and P. G. Richards (1996) – Seismological evidence for differential rotation of the Earth's inner core. *Nature*, 382, 221-224.

Michael F. Goodchild
Universidade da Califórnia
Santa Bárbara, EUA

Desafios à matemática na geografia¹

INTRODUÇÃO

A ciência matemática e a geografia têm usufruído de uma longa e frutuosa associação, desde que despertou o interesse inicial da humanidade por medir a Terra até à actual e crescente dependência dos sistemas informáticos para o estudo da geometria, da topologia e das propriedades estatísticas da informação sobre a superfície da Terra. A matemática ajudou a resolver no passado, no âmbito da geografia, muitos problemas importantes, como, por exemplo, o da descrição eficiente da forma complexa da Terra como pré-requisito da cartografia exacta, continuando a ser um instrumento imprescindível para os geógrafos na sua busca contínua de um melhor conhecimento da Terra como o lar da humanidade.

Neste artigo será feita uma breve sùmula da associação entre a matemática e a geografia, desde o tempo das primitivas comunidade recolectoras até à complexa sociedade actual, baseada na informação. Em seguida serão discutidos quatro exemplos específicos do modo como a matemática constitui um instrumento fundamental na geografia moderna. Ao fazê-lo, espero contribuir com indicações inegáveis de que a interacção entre a matemática e a geografia tem um passado ilustre, um presente interessante e um futuro excitante.

¹ Tradução de Artur Soares Alves, professor do Departamento de Matemática, Universidade de Coimbra.

RESUMO HISTÓRICO

A ideia de criar uma representação abstracta do meio circundante é mais antiga do que a própria humanidade. Sabe-se que as abelhas comunicam informações sobre a localização dos recursos alimentares dançando umas para as outras; uma comunicação muito mais sofisticada tornou-se possível quando os humanos desenvolveram a linguagem e as aptidões necessárias para traçar representações sob a forma de mapas primitivos nas paredes das cavernas e na poeira do chão. Os primeiros navegadores do Pacífico utilizavam composições adequadas de paus para trocar informações e para transmitir esse conhecimento às gerações mais novas. Porém, a emergência da agricultura sedentária, há cerca de dez mil anos, e as sociedades complexas que com ela evoluíram criaram a necessidade não somente de representar a Terra mas também de a medir. A geometria (literalmente *medida da Terra*) evoluiu, de certo modo, como resposta à necessidade de registar a propriedade e a produtividade da terra e também para permitir a construção de estruturas permanentes de abrigo, de armazenamento do grão e para satisfazer outros fins comunais.

A agricultura e a construção são, comparativamente, actividades com um âmbito local, mas o comércio entre comunidades humanas situadas a longas distâncias, por mar ou por terra, estimulou o interesse pela navegação e pela relação entre a Terra e os outros corpos celestes. Desenvolveram-se métodos para registar e medir as superfícies curvas e, por volta de 200 a.C., Eratóstenes deduziu o raio de curvatura da Terra comparando a elevação do Sol acima do horizonte ao meio-dia em dois lugares situados a uma distância conhecida e sobre o mesmo meridiano.

O avanço do cristianismo e da sua doutrina ajudou ou prejudicou, dependendo do ponto de vista, a evolução da descrição matemática da Terra. A crença de que Deus criara uma Terra perfeita conduziu à representação de um universo geocêntrico e aos antigos mapas que representavam a região mediterrânica com uma forma geométrica perfeita, com Jerusalém, centro da cristandade, em posição de destaque. Os conhecimentos resultantes dos descobrimentos portugueses, no início do século XV, eram claramente incompatíveis com estas crenças, porém, foi só no século XVII que o peso da evidência empírica se sobrepôs finalmente à doutrina cristã, abrindo o caminho para a Idade das Luzes (só recentemente a Igreja admitiu o seu erro em perseguir Galileu por causa da sua oposição, cientificamente fundamentada, ao universo geocêntrico).

A invenção da imprensa criou um dos mais antigos desafios matemáticos propostos pela geografia. Todos os sistemas que se desenvolveram em torno da imprensa para disseminar o conhecimento

humano — papel, livros, atlas, cartas, envelopes — são planos, mas a superfície da Terra é curva. Por isso, qualquer representação da Terra à escala, a duas dimensões, como por exemplo um mapa, terá de ser distorcida, aumentando o grau de distorção com a proporção da superfície da Terra representada. Ainda hoje, a ciência das projecções cartográficas continua a levantar uma série de problemas estimulantes para os matemáticos. São necessárias novas projecções para se obter um conhecimento e uma representação cada vez mais precisos da forma da Terra e para servir de suporte às necessidades particulares de novas aplicações.

O século XVIII trouxe consigo um novo interesse pela representação precisa da forma detalhada da superfície da Terra. A construção de canais tornava necessário fazer nivelamentos precisos e a artilharia requeria o conhecimento exacto da altura relativa. A noção do uso de contornos, ou curvas de nível, para mostrar a forma em pormenor surgiu nesta altura, associada em primeiro lugar à batimetria ou representação da profundidade das águas. A elevação acima do nível do mar não é medida com a mesma facilidade que a profundidade abaixo do nível do mar e, assim, decorreu algum tempo até que esses métodos pudessem ser aplicados na elaboração de mapas rigorosos de curvas de nível para a elevação do terreno.

Um dos primeiros mapas deste tipo foi elaborado em meados do século XVIII com o objectivo de levar a cabo uma experiência, proposta por Sir Isaac Newton certa de 100 anos antes, para estimar a densidade da Terra. Schiehallion é uma montanha relativamente isolada, aproximadamente cónica, na Escócia. Em 1774 uma equipa de topógrafos criou um mapa exacto de curvas de nível e, a partir daí, calculou uma estimativa do volume da montanha. Medições precisas da direcção e grandeza da força gravitacional da Terra foram então efectuadas para obter a densidade desconhecida.

As Descobertas iniciadas por Portugal no século XV criaram a necessidade de uma cartografia exacta, a que os matemáticos responderam com métodos que permitiam determinar a posição com extrema precisão. Todos esses métodos requeriam um modelo matemático básico da forma da Terra e, embora uma esfera seja suficientemente exacta para muitos objectivos, um elipsóide (obtido por rotação de uma elipse em torno do eixo menor) constitui uma representação muito mais precisa da Terra porque esta é ligeiramente achatada ao longo do seu eixo, por um factor de cerca de 1/300. Em meados do século XX, várias medições rigorosas em diferentes partes da Terra deram origem a uma grande variedade de elipsóides, cada um deles destinado a ajustar-se melhor a um determinado

sector geográfico (ou administrativo). A tabela 1 mostra alguns dos mais importantes elipsóides e os respectivos parâmetros associados. Mesmo actualmente, os mapas produzidos por países adjacentes não se ajustam necessariamente ao longo da sua fronteira comum, a menos que estes países utilizem o mesmo elipsóide e a mesma projecção. A Guerra Fria e a necessidade de guiar com precisão os mísseis intercontinentais levou finalmente à adopção, na segunda metade do século XX, de um elipsóide padrão internacional.

Tabela 1 — Alguns elipsóides importantes de referência (o WGS84 tornou-se o elipsóide mais importante usado internacionalmente).

Elipsóide	Semi-eixo maior (m)	1/achatamento
Airy 1830	6377563.396	299.3249646
Modified Airy	6377340.189	299.3249646
Australian National	6378160	298.25
Bessel 1841 (Namibia)	6377483.865	299.1528128
Bessel 1841	6377397.155	299.1528128
Clarke 1866	6378206.4	294.9786982
Clarke 1880	6378249.145	293.465
Everest (India 1830)	6377276.345	300.8017
Everest (Sabah and Sarawak)	6377298.556	300.8017
Everest (India 1956)	6377301.243	300.8017
Everest (Malaysia 1969)	6377295.664	300.8017
Everest (Malaysia and Singapore)	6377304.063	300.8017
Everest (Pakistan)	6377309.613	300.8017
Modified Fischer 1960	6378155	298.3
Helmert 1906	6378200	298.3
Hough 1960	6378270	297
Indonesian 1974	6378160	298.247
International 1924	6378388	297
Krassovsky 1940	6378245	298.3
GRS 80	6378137	298.257222101
South American 1969	6378160	298.25
WGS 72	6378135	298.26
WGS 84	6378137	298.257223563

Nos últimos trezentos anos, a geodesia, ou medição exacta da forma da Terra e respectivas propriedades, juntamente com a cartografia, concluíram praticamente o projecto que os estudiosos iniciaram há mais de dois milénios. Hoje em dia conhecemos a forma e as propriedades gravitacionais da Terra com exactidão suficiente para medir mudanças milimétricas na posição dos continentes; possuímos métodos bastante desenvolvidos para cartografar a superfície de outros planetas, incluindo asteróides de forma não-esférica e luas longínquas; existe ainda o Sistema

de Posicionamento Global (GPS), uma constelação de satélites artificiais da Terra que, pela primeira vez na história da Humanidade, permite a medição exacta da posição à superfície terrestre relativamente a um sistema de coordenadas geocêntricas. As imagens obtidas pelos satélites em torno da órbita da Terra podem ser processadas utilizando sofisticados algoritmos informáticos para uma rápida actualização dos mapas. As bases de dados geográficos contêm quantidades enormes de informação sobre as características da superfície terrestre e são, em muitos casos, acessíveis de qualquer lugar, via Internet. Os modelos matemáticos são utilizados rotineiramente para prever os movimentos da atmosfera e dos oceanos, para analisar fluxos comerciais, tráfego rodoviário, chamadas telefónicas, idas às compras e para examinar padrões de doenças. A matemática e a geografia estão agora intimamente ligadas e mais do que nunca através da tecnologia da informação.

As próximas quatro secções abordam áreas onde a matemática e a geografia interagem de um modo particularmente estimulante e interessante. O conjunto é de facto muito pessoal, reflectindo os meus próprios interesses e gostos, assim como experiências acumuladas nas últimas três décadas.

PROPRIEDADES TOPOLÓGICAS DAS BACIAS HIDROGRÁFICAS E DOS POVOAMENTOS

A superfície da Terra é inacreditavelmente complexa. Suponhamos que se quer examiná-la globalmente e analisar características tão pequenas como as que medem cerca de um metro. Há aproximadamente 500 000 000 quilómetros quadrados de superfície e, se atribuíssemos um único carácter (1 *byte*) à representação de cada metro quadrado, a base de dados resultante teria um tamanho da ordem de um *petabyte* (1 *petabyte* = 10^{15} *bytes*; esta estimativa pressupõe a não compressão dos dados). Se se digitalizarem estes dados, preenchendo um ecrã de computador com dados novos a cada segundo que passa (cerca de um quilómetro quadrado por segundo para a resolução de um ecrã típico), digitalizar toda a Terra demoraria a vida inteira de um ser humano. É assim extremamente importante encontrar representações compactas e eficientes das características significativas para aplicações específicas.

Em 1945, o geomorfologista Horton ficou intrigado com os complexos padrões formados na superfície terrestre pelas redes de cursos de água ou bacias hidrográficas. Seriam essas redes de canais tão complexas como pareciam ou apresentariam algumas simetrias e padrões básicos? Seriam

determinadas representações dos cursos de água mais adequadas do que outras para revelar essas simetrias e padrões? Os detalhes geométricos dos cursos de água pareciam relativamente pouco importantes, mas seria provável que as propriedades topológicas — noutras palavras, as propriedades que se mantêm quando as propriedades geométricas são alteradas por estiramento ou distorção — pudessem revelar padrões interessantes. Horton concebeu um método de numeração dos cursos de água de uma bacia hidrográfica para identificar as suas propriedades topológicas baseando-se na suposição de que uma rede de cursos de água é um grafo em árvore (não contém circuitos). O método, na forma mais tarde modificada por Strahler (1952), está indicado na figura 1. Os cursos de água mais pequenos foram assinalados primeiro com um «1». Qualquer curso de água resultante da junção de dois «1» é assinalado com um «2». De um modo semelhante, juntando-se dois «2» forma-se um novo «3» e, geralmente, a junção de dois cursos de água de ordem i produz um novo curso de ordem $i+1$ (a junção de um curso de ordem i e de um curso de água de ordem menor do que i não produz um novo curso de água).

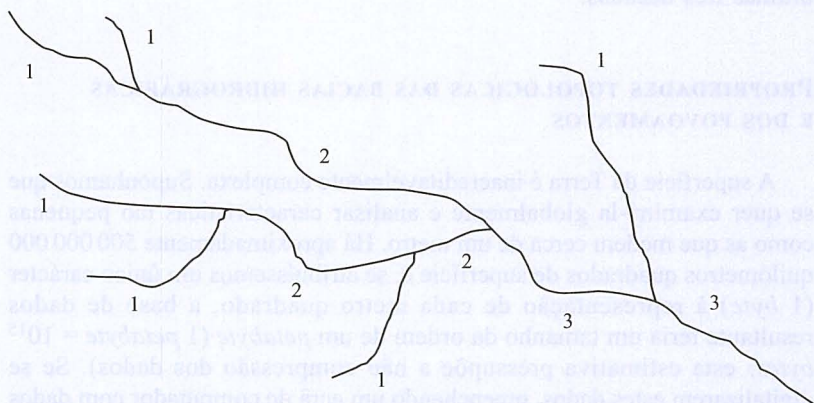


Fig. 1 — O sistema de Horton para numerar as características topológicas de redes de água, modificado por Strahler

Ao simplificar as bacias hidrográficas deste modo, Horton descobriu algumas regularidades empíricas notáveis (Horton, 1945). A *Law of Stream Numbers* foi constatada contando os números de cursos de cada ordem e assinalando-os num gráfico de dispersão e colocando o logaritmo do número em contraste com o número de ordem. Na maioria dos casos, o gráfico mostrou-se, de um modo notável, ajustado a uma linha recta.

Por outras palavras, a relação entre o número de cursos de uma dada ordem i , $N(i)$, e o número de cursos da ordem mais alta seguinte, $N(i+1)$, revelou-se uma constante. Em geral, $N(i) / N(i+1) = B$, sendo B uma constante designada por *taxa de bifurcação* e apresentando um valor de aproximadamente 4,5 para muitas redes de cursos de água. Horton tinha encontrado o que parecia ser uma forte regularidade, facto este que suscitou a busca de uma teoria explicativa apropriada.

Quando duas variáveis são consideradas num gráfico de dispersão é frequentemente utilizada uma linha recta como indicador de uma forte relação, ao passo que duas variáveis que não estão relacionadas produzem uma nuvem dispersa de pontos. Mas os gráficos de Horton não são gráficos de dispersão normais. Tendo em conta o modo como as duas variáveis são definidas, o gráfico nunca poderá resultar numa nuvem de pontos, pois nenhum $N(i+1)$ é superior a $N(i)/2$. Num artigo publicado em 1966 (Shreve, 1966; ver também Shreve, 1967; Abrahams, 1984), cerca de vinte anos depois do trabalho original de Horton, Shreve formulou a pergunta óbvia: como seria a topologia de uma rede «aleatória» de cursos de água que não sofresse qualquer influência das leis físicas e não apresentasse nenhum dos padrões observados por Horton?

Shreve sugeriu que uma definição razoável de «aleatório», neste caso, seria todas as possíveis topologias de redes, com um dado número de cursos de água de primeira ordem, terem a mesma probabilidade. A figura 2 mostra as 14 redes teoricamente possíveis e topologicamente distintas sendo $N(1) = 5$; presumivelmente as forças da natureza tornam algumas delas impossíveis ou menos prováveis que outras, e este facto conduz à aparente regularidade de Horton. No entanto, através da análise matemática, Shreve conseguiu mostrar que a lei de Horton seria também verdadeira para estas hipotéticas redes de canais se estas ocorressem com igual probabilidade — por outras palavras, a lei era uma simples consequência do acaso estatístico, e não de um processo físico influenciando as probabilidades. Os efeitos dos processos físicos revelavam-se melhor em certos desvios da linha recta de Horton do que na própria linha.

As bacias hidrográficas são um exemplo de um fenómeno físico e os cursos de água que Horton estudou não apresentavam praticamente nenhuma modificação introduzida pelo homem. Mas os geógrafos também estudam os padrões formados na superfície terrestre pela actividade humana e questionam-se do mesmo modo sobre a existência de padrões e regularidades. Um dos padrões mais impressionantes é o formado pelas povoações — aldeias, vilas e cidades. A sua localização é muitas vezes determinada pelas características do terreno, tal como as linhas costeiras, as margens e os rios, mas, em alguns casos, principalmente nas planícies

da América do Norte, Europa ou Ásia, as povoações podem localizar-se praticamente em qualquer lugar.

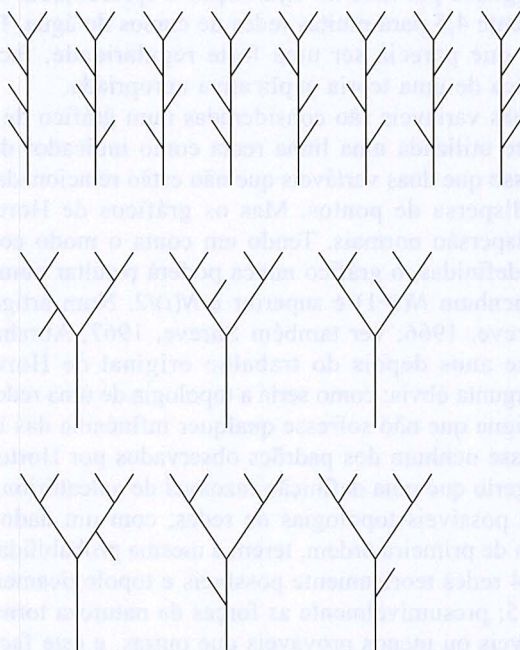


Fig. 2 — As catorze redes de canais topologicamente distintas que se podem formar a partir de cinco cursos de água de primeira ordem. Oito têm um curso de água de segunda ordem; as restantes seis possuem dois de segunda ordem e um de terceira ordem

Em 1933, o geógrafo Christaller propôs uma teoria das povoações, ou locais centrais, baseada em princípios simples de comportamento económico (Christaller, 1966). O nível mais baixo de povoamento, uma aldeia, fornece os serviços mais simples, como por exemplo, a mercearia; os serviços mais complexos, como os centros comerciais, só estão disponíveis nas povoações maiores. Os investidores de grandes armazéns procuram maximizar o rendimento ou o investimento e tentam estabelecer-se nos sítios mais lucrativos. Sob certos pressupostos, a competição resultará em agrupamentos regulares de povoações na paisagem, em que cada ordem, desde a mais baixa, irá ocupar um nó numa grelha hexagonal. As povoações de ordem mais alta também irão formar uma grelha

hexagonal com uma menor densidade, em co-localização com os centros de ordem mais baixa. Os números de centros em cada nível estarão fixados numa razão fixa igual a 3, 4, 7 ou a produtos destes números inteiros.

A lei de Horton surgiu a partir da observação, tendo-se depois partido para o estabelecimento da teoria. Christaller desenvolveu primeiro a teoria, procurando posteriormente a sua validação empírica. Seria possível comprovar, particularmente em áreas de povoamento denso em planície, os padrões teóricos previstos por esta teoria? E, mesmo que não se encontrassem padrões perfeitamente hexagonais, haveria, ainda assim, vestígios de perfeição? Um vasto número de estudos acerca deste problema surgiu então nos anos 50 e 60.

Um dos mais interessantes dependia da possibilidade de identificar a área de influência de cada povoação. Suponha-se que cada povoação formava o centro de uma unidade administrativa, tal como um distrito (para comunidades maiores) ou uma freguesia. A unidade poderia não ser um hexágono geométrico, mas teria talvez algumas das propriedades topológicas dos hexágonos, por exemplo, talvez houvesse uma tendência para que cada unidade tivesse seis vizinhos. Identificaram-se áreas adequadas, onde as condições favoreciam a teoria de Christaller e onde as unidades administrativas foram devidamente definidas e cartografadas, tendo os estudos nestas áreas encontrado um número médio de vizinhos muito próximo de seis.

Mas embora a intuição nos diga que uma média de 5,91 é uma notável aproximação de 6, tal como no caso de Horton, falta-nos uma base racional para a avaliação. Que média se obteria a partir de uma configuração aleatória de unidades que não tivessem sido influenciadas pelas forças económicas admitidas por Christaller? Neste caso uma simples taça de poliuretano forneceria a resposta. As bolhas de poliuretano são geradas por processos que estão longe de ser análogos aos processos económicos que operam na paisagem humana, mas, curiosamente, uma simples análise mostra que também elas têm um número médio de vizinhos muito próximo dos 6.

A teoria subjacente a esta problemática é uma lei descoberta por Euler. Se se tomar uma pavimentação (divisão de um plano em áreas), os números de áreas A , de arestas entre elas E , e de vértices onde se tocam as arestas N , devem satisfazer a expressão $A - E + N = 1$. Definindo as regras apropriadas de modo a identificar ou contar objectos, a lei de Euler citada conduz directamente ao resultado observado, notando-se, embora, que o número médio de vizinhos é inferior a 6 porque as arestas partilhadas pelas áreas adjacentes são contadas duas vezes, ao passo que as arestas que circundam o mosaico são contadas uma vez.

Estes dois exemplos demonstram um dos princípios básicos da análise geográfica. Sabemos que os olhos e o cérebro podem ser facilmente enganados pela aparência do padrão, e muitas ilusões ópticas atestam isto mesmo. Têm-se encontrado padrões e formas na disposição aleatória das estrelas no céu nocturno ou nas variações da superfície de Marte. Mas uma análise matemática cuidadosa mostrou, em muitos destes casos, que o que parecem ser padrões sistemáticos não são mais do que acontecimentos ao acaso no espaço, uma vez que se encontrem e implementem definições apropriadas de acaso. O ser humano parece não estar, de um modo geral, suficientemente equipado para discriminar os padrões das características aleatórias no espaço.

GEOESTATÍSTICA, FRACTAIS E DEPENDÊNCIA ESPACIAL

Qualquer estudante de Estatística conhece o conceito de independência — a noção de que acontecimentos sucessivos não têm memória de outros acontecimentos passados. Uma moeda, por exemplo, não tem memória do seu comportamento passado e o resultado de um lançamento não influencia o seguinte. A independência é um pressuposto razoável em muitas experiências e é a base de muitos testes típicos de estatística.

Mas tudo é diferente na superfície terrestre. O geógrafo Waldo Tobler exprimiu esta diferença do seguinte modo: «Tudo está relacionado, mas as coisas próximas estão mais relacionadas do que as distantes». Por outras palavras, a independência quase nunca é um pressuposto válido para acontecimentos que ocorrem à superfície da Terra, excepto quando esses acontecimentos são muito longínquos entre si. Por exemplo, considere-se o fenómeno da temperatura do ar, que varia com o local como uma função $T(x)$, onde T representa a temperatura e x a posição. Suponhamos que se conhece a temperatura em x e que temos de estimá-la noutra local, y (talvez porque se pretenda viajar até lá). Quando x e y estão muito próximos, $T(x)$ é uma estimativa muito segura de $T(y)$ porque a temperatura na superfície da Terra só se altera rapidamente em circunstâncias especiais, como, por exemplo, durante a passagem de frentes climáticas ou na proximidade de águas frias. Quanto mais afastado y estiver de x , menos exacto será $T(x)$ como estimativa de $T(y)$.

A observação de Tobler é suficientemente genérica para lhe poder ser conferido o estatuto de lei. Têm sido concebidas várias medidas de dependência espacial e o termo geral para as designar é *autocorrelação espacial*. A lei de Tobler estabelece, na sua essência, que a autocorrelação

espacial é fortemente positiva para distâncias curtas e enfraquece com a distância. Existem vários termos para designar a distância à qual a autocorrelação espacial decresce para zero, como por exemplo *range* e *grain*. A autocorrelação espacial negativa ocorre quando a diferença entre fenómenos a uma dada distância é maior do que a variação global (um tabuleiro de xadrez é um bom exemplo), mas a autocorrelação espacial negativa entre algumas distâncias deve implicar a autocorrelação espacial positiva entre outras distâncias. De facto, um mundo sem autocorrelação espacial seria quase impossível de imaginar ou habitar. Cada alteração microscópica de posição produziria tanta variação como a que existe no mundo como um todo e a vida seria como estar a cair constantemente de uma falésia.

Um dos modos mais eficazes de representar a dependência espacial é através da função que os geoestatísticos designam por *variograma*. Na figura 3, cada ponto denota um par de localizações. O valor marcado no eixo horizontal é a distância entre elas e o valor no eixo vertical é o quadrado da diferença dos valores em causa. Neste exemplo, no eixo vertical assinalam-se os quadrados das diferenças das temperaturas. As barras verticais denotam intervalos convenientes do eixo horizontal. Calcula-se a média dos pontos dentro de cada intervalo para se marcarem as pintas. Finalmente ajusta-se a estas pintas uma curva regular que representa uma função pertencente a um conjunto padrão de funções matemáticas. Esta curva é o variograma, denotado pela função $2\gamma(d)$, onde d representa a distância. Geralmente constata-se que os variogramas, para fenómenos na superfície da Terra, têm uma forma notavelmente consistente, como se ilustra na figura caracterizada por um crescimento monótono seguido por uma atenuação no domínio onde a variância já não aumenta.

Esta observação tem sido muito útil no tratamento da informação geográfica. Uma aplicação frequente ocorre quando um fenómeno foi sujeito a amostragem num determinado número de locais, como acontece, por exemplo, na amostragem de fenómenos climáticos, como a temperatura, a precipitação ou a pressão atmosférica. Razões monetárias impedem que estes fenómenos possam ser medidos em qualquer local. Assim, em locais onde não são efectuadas medições procede-se a estimativas explorando a lei de Tobler e as propriedades específicas do variograma. O método espacial de interpolação designado por Kriging foi inventado por um geólogo, nos anos 50, para implementar este conceito. Noutras disciplinas foram inventados métodos muito semelhantes. A área da *geoestatística* baseia-se nestes métodos.

Metade da média do quadrado das
diferenças (semivariância)

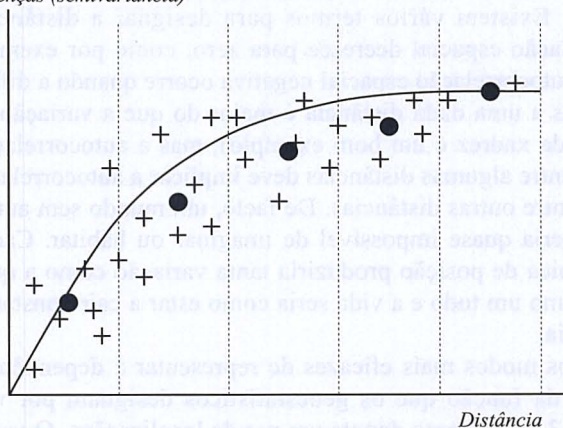


Fig. 3 — Um semivariograma típico. Os símbolos são explicados no texto

Mas a correlação espacial não é a única propriedade matemática dominante dos dados geográficos. Quem já trabalhou exaustivamente com dados geográficos refere os efeitos da complexidade e, particularmente, da relação entre complexidade e resolução espacial: quanto mais de perto se olha a geografia mais pormenores se vêem. As linhas costeiras mostram claramente esta propriedade, tornando-se cada vez mais sinuosas num exame próximo e, conseqüentemente, também mais extensas. Nenhum geógrafo leva a sério os debates sobre a extensão da orla costeira ou sobre a extensão das fronteiras políticas, sabendo estes que é possível produzir uma larga variedade de extensões dependendo da escala do exame a efectuar. Por exemplo, em 18 de Novembro de 2000, o *Seattle Times* noticiou um novo cálculo do comprimento da costa da Noruega. A nova estimativa, calculada a partir de uma representação digital detalhada, tinha mais 13 760 milhas que a extensão anteriormente estimada em 35 580 milhas. O objectivo desta notícia era corrigir um erro, mas para os geógrafos tratava-se do simples resultado de um exame mais pormenorizado da linha costeira.

Benôit Mandelbrot foi o primeiro a dar a estas observações um enquadramento matemático formal, a que deu o nome de *fractais* ou fenómenos geométricos que exibem dimensões fraccionárias (Mandelbrot, 1983). As propriedades fractais podem ser associadas a pontos, linhas, áreas ou superfícies, e muitos dos exemplos que Mandelbrot utilizou nos

seus livros eram de natureza geográfica. Os fractais mostram propriedades de *auto-semelhança*, o que significa que uma parte do objecto tem propriedades, tais como a forma geométrica, que são verdadeiras para o objecto considerado como um todo. Um corolário interessante desta propriedade é a impossibilidade de determinar a escala de um objecto fractal, dado que se pode estar a olhar para uma parte de um tamanho qualquer. As propriedades dos fractais comportam-se sistematicamente como mudanças de escala, fornecendo um dos poucos enquadramentos teóricos que permitem estudar o efeito da escala. Por exemplo, o comprimento de uma linha fractal altera-se consoante o nível de detalhe a que é mostrada e medida, numa relação simples do tipo potência. Este comportamento tinha sido observado anteriormente por Richardson no seu estudo sobre as extensões das fronteiras políticas (Richardson, 1961), mas os fractais de Mandelbrot são os primeiros a fornecer uma explicação teórica da questão. Goodchild e Mark (1987) apresentam uma visão geral da importância dos fractais na geografia.

Os fractais têm variogramas bem comportados, não admitindo, no entanto, tangentes — comportamento que conduziu à utilização do termo *monstro matemático* para descrever os fractais, o que aliás acontece na representação gráfica de muitos fenómenos geográficos. O relevo da superfície da Terra é muitas vezes concebido como uma função matemática de posição, $z = z(x)$, e em muitos estudos utilizaram-se funções típicas, tais como polinómios ou séries de Fourier, para descrever partes da superfície. Contudo, também é verdade que a superfície da Terra mostra frequentemente falésias, cumeeiras, canais e outros fenómenos que representam descontinuidades na elevação ou no declive.

Por outras palavras, a função z não é contínua e diferenciável em todo o lado e, logo, as suas tangentes não estão bem definidas em todo o lado. Embora recorramos com frequência a metáforas de *terreno* quando se fala de funções matemáticas (por exemplo, *superfície*, *elevação*, *gradiente*), há sem dúvida fortes razões para o facto de ser falacioso pensar no terreno como tendo uma forma matemática perfeita. É interessante especular sobre a importância da metáfora *terreno* para os iniciadores do cálculo (Newton e Leibniz) e interrogarmo-nos sobre se os terrenos que lhes eram familiares eram particularmente livres de descontinuidades.

Uma consequência deste problema é que, à semelhança do que acontece no exemplo do comprimento da linha costeira discutido anteriormente, o declive da superfície da Terra não pode ser definido independentemente da resolução espacial. Na prática, o declive é definido através

de uma aproximação de diferenças finitas, utilizando tipicamente as seguintes equações:

$$b = (z_3 + 2z_6 + z_9 - z_1 - 2z_4 - z_7) / 8D$$

$$c = (z_1 + 2z_2 + z_3 - z_7 - 2z_8 - z_9) / 8D$$

onde b e c representam declives das tangentes respectivamente nas direcções x e y , D é o espaçamento da grelha e z_i denota a cota do ponto i , como se mostra na figura 4. A ausência de tangentes bem definidas à superfície da Terra obriga a que as estimativas de b e c dependam explicitamente de D . Por exemplo, os declives calculados a partir de uma grelha com espaçamento de 30 metros são diferentes dos declives na mesma localização calculados a partir de uma grelha com espaçamento de 100 metros (existem vários conjuntos de dados com aproximadamente estes dois espaçamentos para a maior parte dos EUA).

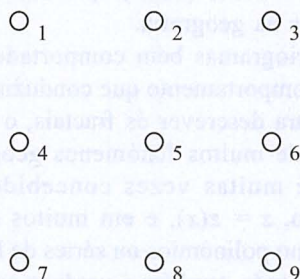


Fig. 4 — Notação utilizada na estimativa do declive

Os fractais mostraram-se muito úteis na ciência da informação geográfica. Primeiro, como foi dito anteriormente, fornecem um enquadramento teórico para o estudo dos efeitos da escala, permitindo-nos antever até que ponto se ganham ou perdem detalhes quando a resolução espacial se altera e determinar os benefícios e os custos do aumento da resolução quando se projectam satélites para obter imagens da superfície da Terra. As curvas fractais descritas por Peano e Hilbert (figura 5) constituem também uma base para estruturas de dados eficientes e para a concepção de ordenações unidimensionais do espaço que procuram preservar relações espaciais bidimensionais. Estas ideias são implementadas em muitos dos actuais sistemas de informação geográfica. Nos últimos anos os investigadores têm vindo a estudar a forma como estes conceitos se podem aplicar à superfície curva da Terra.

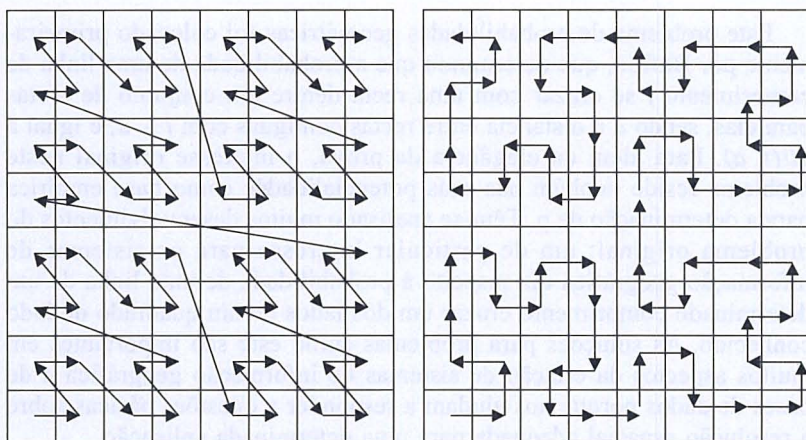


Fig. 5 — As curvas de Peano e Hilbert que preenchem o espaço têm sido implementadas em sistemas de informação geográfica como base de técnicas de compressão de dados e esquemas de indexação

PROBABILIDADE GEOMÉTRICA

A probabilidade geométrica é um ramo da matemática que se dedica ao estudo de inferências estatísticas dos acontecimentos geométricos. Tem sido uma fonte de problemas estimulantes e mesmo de diversão para os matemáticos. Recentemente, surgiu uma nova série de desafios relacionados com o interesse em prever o desempenho dos sistemas de informação geográfica. Muitos dos processos actualmente aplicados aos dados geográficos têm resultados variáveis, dependendo da frequência com que ocorrem determinadas condições extremas ou patológicas. Por exemplo, considere-se o problema do cálculo da distância entre dois pontos na superfície terrestre. A projecção *Universal Transverse Mercator* (UTM) é vulgarmente utilizada como base das coordenadas em sistemas de informação geográfica e foi adoptada como padrão militar internacional pela NATO. A projecção UTM divide a Terra em faixas ou zonas com amplitude de seis graus de longitude. Se dois pontos se situarem na mesma zona, então a distância entre eles pode ser calculada a partir das suas coordenadas UTM e de uma expressão simples do teorema de Pitágoras, com um erro máximo de 0,04%. Mas, se os dois pontos se situarem em zonas UTM diferentes, têm de ser efectuados cálculos muito mais complexos e morosos. O desempenho depende assim directamente da probabilidade de qualquer par de pontos se situar na mesma zona.

Este problema de probabilidades geométricas foi colocado primeiramente por Buffon, que determinou que a probabilidade de uma linha de comprimento l se cruzar com uma recta dentre um conjunto de rectas paralelas, sendo a a distância entre rectas contíguas com $l < a$, é igual a $2l/(\pi a)$. Para além da elegância da prova, o interesse original neste problema reside também nas suas potencialidades como base empírica para a determinação de π . Têm-se analisado muitos desenvolvimentos do problema original; um de particular interesse para os sistemas de informação geográfica diz respeito à probabilidade de uma linha de um determinado comprimento cruzar um dos lados de um quadrado de lado conhecido. As soluções para problemas como este são importantes em muitos aspectos da criação de sistemas de informação geográfica e de bases de dados porque nos ajudam a responder a questões básicas sobre a resolução espacial adequada para uma determinada aplicação.

OPTIMIZAÇÃO ESPACIAL

A última destas quatro secções aborda uma classe de problemas que desde sempre me intrigaram, em parte por causa da sua complexidade e em parte por causa da natureza não intuitiva de muitas das suas soluções. Trata-se de problemas de optimização espacial ou de busca de soluções óptimas para problemas colocados no espaço tridimensional. Um dos mais conhecidos é o problema do caixeiro viajante (*traveling salesman problem* – TSP), ou seja, o problema de seleccionar o melhor caminho para visitar um determinado número de destinos numa só viagem, a partir de um dado ponto. Se houver n destinos, incluindo o ponto de partida, o número possível de viagens é $(n-1)!/2$ (o factor 2 entra porque qualquer viagem pode ser executada nos dois sentidos). Este número cresce muito rapidamente com n e, conseqüentemente, o problema TSP não pode ser resolvido em tempo razoável a não ser para valores pequenos de n .

Apesar da sua complexidade inerente, estes problemas são agora rotineiramente resolvidos através de métodos heurísticos, métodos de solução que se espera possam vir a oferecer soluções satisfatoriamente próximas da óptima. Dado um par de pontos identificados pelas suas direcções, qualquer utilizador da Internet pode contactar um certo número de *sites* (incluindo www.mapquest.com) e obter a solução para o problema do caminho mais curto, ou o trajecto óptimo entre os dois pontos através da teia de ruas. Milhões de pessoas acedem a estes *sites* todos os dias para planearem as férias ou viagens de negócios em áreas que lhes são desconhecidas.

Considere-se um conjunto de pontos distribuídos num plano, cada um com coordenadas (x_i, y_i) e com pesos w_i . Se estes pesos representarem contagem de pessoas, será útil saber a localização (X, Y) que minimiza a distância total percorrida, por outras palavras, o ponto que minimiza:

$$Z = \sum_i w_i [(x_i - X)^2 + (y_i - Y)^2]^{1/2}$$

Este ponto tem sido designado de várias formas como, por exemplo, o *minimizante de percurso agregado* (MAT – *point of minimum aggregate travel*) e as suas propriedades estudadas extensivamente. Uma das mais interessantes é que a função $Z(X, Y)$ não tem derivadas definidas nos pontos (x_i, y_i) . Webber (1928) utilizou o MAT para elaborar a sua teoria de localização industrial, propondo que as fábricas se localizassem em sítios em que os custos totais de transporte de matérias-primas para a empresa e a distribuição dos produtos pelo mercado fossem minimizados.

Há um dispositivo físico que permite obter uma solução para o problema MAT que se designa por *quadro de Varignon*. Numa placa rígida na horizontal, fazem-se furos em cada uma das posições (x_i, y_i) . Em cada furo pendura-se uma corda e suspende-se um peso w_i . As pontas livres das cordas juntam-se e atam-se por cima da placa, soltando-se o nó (ver figura 6). Ignorando os efeitos do atrito, a posição encontrada pelo nó será o ponto MAT.

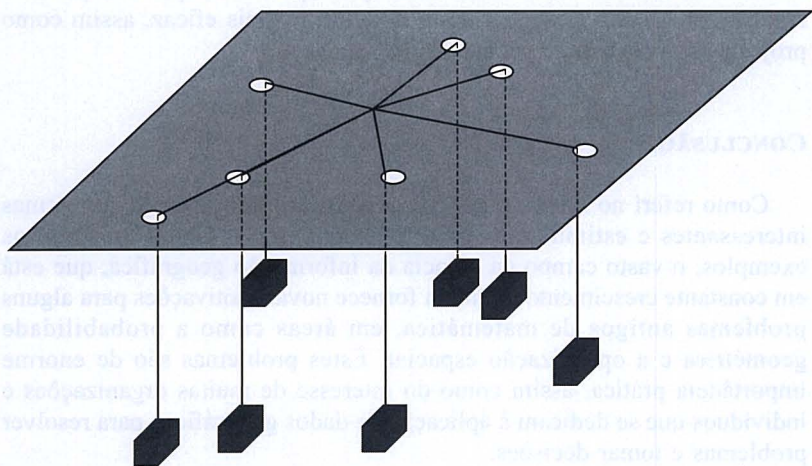


Fig. 6 — A experiência do quadro de Varignon. Ver a explicação no texto

Considere-se agora a seguinte experiência mental. Suponhamos que um peso representa um fornecimento para a fábrica que está localizada de modo ótimo no nó e suponhamos que o abastecimento se esgota. Encontra-se uma nova fonte de abastecimento mais afastada, mas na mesma direção. Faz-se um novo furo neste local e repete-se a experiência. Será que a localização da fábrica se move? A intuição diz que sim, dado que o ponto MAT minimiza as distâncias e a distância mudou. Mas o quadro de Varignon diz que não, dado que a força que puxa a corda é a mesma independentemente da localização dos furos e todas as outras forças permanecem imutáveis.

A utilização do cálculo diferencial para obter as derivadas de Z a respeito de X e Y mostra que a analogia física está correcta e a intuição errada. A intuição parece mais compatível com as propriedades do centróide, que minimiza a soma pesada dos quadrados das distâncias e se moverá nesta experiência mental, do que com as propriedades do ponto MAT. Propriedades não intuitivas, de algum modo semelhantes, podem ser facilmente encontradas para o caso unidimensional de pontos localizados ao longo de uma linha e para muitas generalizações do problema MAT básico. De maneira geral, parece que a intuição humana é particularmente limitada no que diz respeito a encontrar soluções para problemas de optimização espacial, fazendo assim eco de um ponto levantado anteriormente relativo a problemas estatísticos. À semelhança do que então se referiu, a lição a retirar deste exemplo é que os instrumentos matemáticos são uma ajuda preciosa, que nos permite resolver problemas geográficos de uma forma mais eficaz, assim como projectar soluções mais adequadas.

CONCLUSÃO

Como referi no início, a geografia constitui uma fonte de problemas interessantes e estimulantes para os matemáticos. Como mostram os exemplos, o vasto campo da ciência da informação geográfica, que está em constante crescimento, também fornece novas motivações para alguns problemas antigos de matemática, em áreas como a probabilidade geométrica e a optimização espacial. Estes problemas são de enorme importância prática, assim como do interesse de muitas organizações e indivíduos que se dedicam à aplicação de dados geográficos para resolver problemas e tomar decisões.

Nesta análise tentei evidenciar o que costumo considerar o conflito frequente entre a intuição humana e a lógica matemática relativamente a

tópicos geográficos. Nós, humanos, trabalhamos constantemente com o espaço. Acreditamos que «uma fotografia vale mais que mil palavras» e que as crianças aprendem a ter capacidades espaciais numa idade tenra. Mas a familiaridade também parece ser uma fonte de falsa segurança porque o que aprendemos sobre o espaço não está necessariamente correcto. Devido a ilusões ópticas e efeitos relacionados, vemos muitas vezes padrões onde não existem; a nossa intuição no que respeita à optimização espacial é nitidamente errónea; falta-nos, muitas vezes, o sentido correcto do que esperar em propriedades topológicas, como, por exemplo, o número médio de arestas por face num mosaico. Actualmente há a noção generalizada da crescente importância da localização geográfica, assim como do pensamento espacial na sociedade moderna. Espero ter conseguido mostrar claramente que esta importância crescente justifica que se dê maior atenção à matemática da geografia.

Referências

- Abrahams, A. D. (1984) – Channel networks: A geomorphological perspective. *Water Resources Research* 20: 161-188.
- Christaller, W. (1966) – *The Central Places of Southern Germany*, translated by C. W. Baskin. Englewood Cliffs, New Jersey: Prentice-Hall.
- Goodchild, M. F., e D. M. Mark (1987) – The fractal nature of geographic phenomena. *Annals of the Association of American Geographers* 77: 265-278.
- Horton, R. E. (1945) – Erosional development of streams and their drainage basins: Hydrophysical approach to quantitative morphology. *Geological Society of America Bulletin* 56: 275-370.
- Mandelbrot, B. B. (1983) – *The Fractal Geometry of Nature*. São Francisco: Freeman.
- Richardson, L. F. (1961) – The problem of contiguity: An appendix to the statistics of deadly quarrels. *General Systems Yearbook* 6: 139-187.
- Shreve, R. L. (1966) – Statistical law of stream numbers. *Journal of Geology* 74: 17-37.
- Shreve, R. L. (1967) – Infinite topologically random channel networks. *Journal of Geology* 75: 178-186.
- Strahler, A. N. (1952) – Hypsometric analysis of erosional topography. *Geological Society of America Bulletin* 63: 1117-1142.
- Weber, A. (1928) – *Alfred Weber's Theory of the Location of Industries*. Translated by C.J. Friedrich. Chicago: University of Chicago Press.

(Página deixada propositadamente em branco)

Matemática e ambiente: a redescoberta dos fundamentos básicos

INTRODUÇÃO

A constatação de que a humanidade faz parte da biosfera do seu planeta contribuiu significativamente para o reconhecimento crescente da interdependência e da necessidade de proteger o ambiente, não só para a civilização actual, mas também para as gerações futuras. Um dos aspectos mais importantes, associado à complexidade dos processos ambientais, que os cientistas, os académicos, os engenheiros e mesmo os políticos são actualmente instados a reconhecer é a natureza interdisciplinar das ciências do ambiente. O rápido desaparecimento da divisão tradicional entre experimentalistas e teóricos e a colaboração activa entre as disciplinas (classicamente separadas) da física, da química, da biologia, da matemática, da informática, etc., procurando as inter-relações existentes, constitui a base do sucesso dos actuais especialistas do ambiente.

Os sistemas naturais são função de diversos factores não lineares, sendo o seu estudo mais bem sucedido quando ocorre em sistemas experimentais e numéricos, o que implica forçosamente simplificações. Apesar destas, o modelo mental que cada engenheiro do ambiente tem de um sistema necessita de ser formalizado matematicamente para se tornar operacional e útil: enquanto a física analisa o ecossistema como um conjunto de partículas e a química se preocupa com a sua composição, a

engenharia do ambiente não só as divide em partículas vivas e mortas, como também descreve as suas principais interacções funcionais no sistema ambiental.

A procura do suporte matemático para compreender o ambiente está associada aos instrumentos disponíveis para a resolução de modelos, isto é, quer aos métodos numéricos quer aos computadores modernos. Por exemplo, os sofisticados modelos de qualidade do ar são resolvidos em potentes computadores, obtendo os engenheiros do ambiente a previsão dos episódios de poluição do ar, com uma elevada resolução espacial e temporal. No entanto, a complexidade dos processos atmosféricos ultrapassa a capacidade dos computadores actuais. Presentemente, os dados simulados resultam da aplicação de métodos numéricos na resolução de sistemas de equações às derivadas parciais, que representam os processos dinâmicos na atmosfera, aos quais se juntam alguns milhares de reacções químicas que integram os fenómenos da qualidade do ar.

Estes complexos sistemas de equações necessitam de condições iniciais e de fronteira que têm de ser fornecidas ao modelo, para que os resultados da simulação sejam tão precisos quanto possível. Estes dados podem ser o resultado de outras simulações ou podem provir de observações. Às incertezas numéricas relacionadas com os métodos numéricos e com as capacidades computacionais dever-se-ão associar as incertezas das observações referentes à precisão dos equipamentos, tempos de amostragem, médias temporais, etc. O valor final, o denominado valor «real», resulta do uso de muitas e diferentes técnicas matemáticas, tais como métodos variacionais, técnicas analíticas, estatística aplicada, etc.

A redescoberta desta ligação entre as ciências básicas, de onde emergem as ciências aplicadas — de que a engenharia é um bom exemplo —, leva a confirmar a necessidade imperiosa de dar uma formação forte e alargada em ciências ditas fundamentais aos alunos universitários. Apenas esta base científica garante tranquilidade num mundo em constante mutação e onde a interdisciplinaridade se tornou incontornável.

Este trabalho pretende dar uma perspectiva geral da utilização da matemática na engenharia do ambiente, fornecendo exemplos ilustrativos deste contributo, que, apesar de focarem essencialmente os processos atmosféricos, são também aplicáveis a qualquer outro processo ambiental.

1. A INTERDISCIPLINARIDADE DAS CIÊNCIAS DO AMBIENTE

O nascimento do pensamento científico, traduzido no desenvolvimento de métodos (experimental, dedutivo, demonstração pelo absurdo, etc.) que permitiram estabelecer as leis da natureza, foi um marco na racionalidade.

A matemática constituiu desde sempre a linguagem de base e o suporte à formulação dos problemas e das suas soluções.

Dada a complexidade da natureza e dos fenómenos que a regem, o homem não tem capacidade de controlar todas as variáveis envolvidas. Este facto, aliado à tentativa de manipulação dos recursos naturais para proveito próprio, tem conduzido à degradação desses recursos e do ambiente natural. Para fazer face a esta problemática surgiu a engenharia do ambiente, com vocação para a remediação dos desequilíbrios causados nos sistemas naturais, distinguindo-se das outras engenharias pela sua capacidade integradora. De facto, os vários componentes do ambiente não são estanques; pelo contrário, são interdependentes, pelo que, irremediavelmente, qualquer alteração com impacto sobre um determinado meio terá reflexos nos outros meios: ar, água, solo e biota.

A compreensão do ambiente como um todo e dos seus componentes em particular requer um conjunto de conhecimentos que não se limitam a um ramo específico da ciência. É necessário conhecer os fenómenos físicos, os processos de transformação química e as relações tróficas que ocorrem nos ecossistemas. A figura 1 ilustra este tipo de filosofia e abordagem (Borrego *et al.*, 2000).

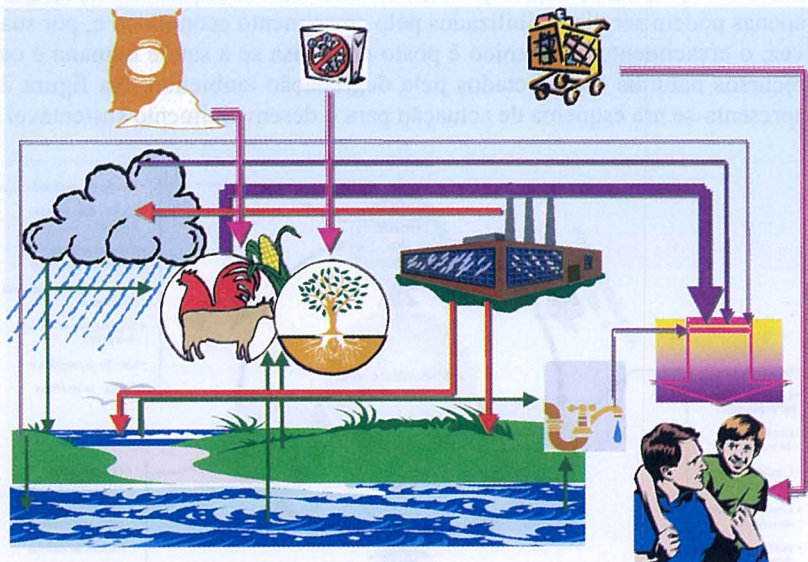


Fig. 1 — Esquema ilustrativo da interação estabelecida entre o ar, a água, o solo, o biota (animais e plantas) e o homem (último elemento da cadeia alimentar) (in Borrego *et al.*, 2000)

Através de vários processos, tais como a precipitação, a infiltração, o escoamento subterrâneo e a evapotranspiração, estabelecem-se interações entre o ar e a superfície (animais e plantas). Quando ocorre um episódio de poluição, todo o ciclo é afectado. Os poluentes libertados para a atmosfera poderão ser transportados e vir a ser depositados a grandes distâncias das fontes emissoras. Os efluentes líquidos contaminarão águas e solos, que por sua vez estão associados à cadeia alimentar animal.

O engenheiro do ambiente deverá ter capacidade para criar, modificar, valorizar e simultaneamente proteger o meio em que o homem vive, a fim de satisfazer as necessidades deste. O ensino de engenharia do ambiente deve fazer a síntese científico-técnica, mas também social e humanística, para a resolução de problemas e desenvolver a criatividade e a imaginação. A interdisciplinaridade das ciências do ambiente e da formação de um engenheiro do ambiente veio a ser confirmada, na década de 90, quando o desenvolvimento sustentável foi estabelecido como meta prioritária a nível planetário. Este implica uma perspectiva integrada para políticos e decisores, na qual a protecção ambiental e o crescimento económico a longo prazo são vistos como complementares e não como incompatíveis, em última análise mutuamente dependentes (Borrego *et al.*, 1999a): a solução dos problemas ambientais requer recursos que apenas podem ser disponibilizados pelo crescimento económico e, por sua vez, o crescimento económico é posto em causa se a saúde humana e os recursos naturais são afectados pela degradação ambiental. Na figura 2 apresenta-se um esquema de actuação para o desenvolvimento sustentável.

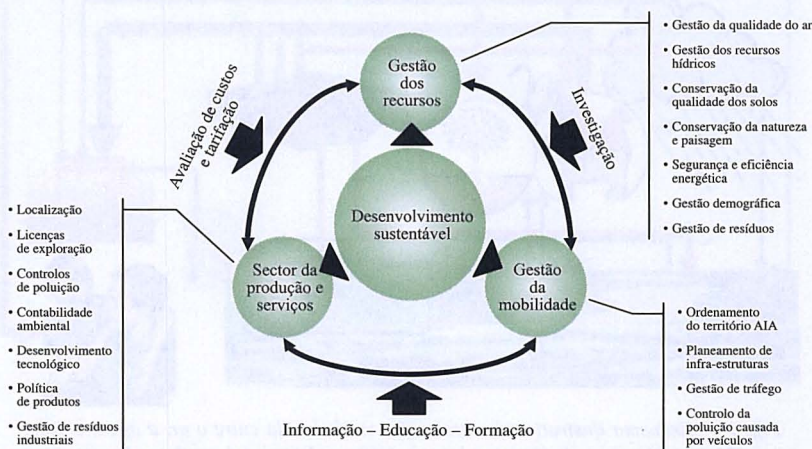


Fig. 2 — Esquema de actuação do desenvolvimento sustentável (Borrego *et al.*, 1999a)

O compromisso assumido recentemente, em 1997, em Quioto, pelos países industrializados, relativamente à redução das emissões para a atmosfera de gases com efeito de estufa (GEE), ilustra com clareza a forte relação entre a componente ambiente e a componente desenvolvimento sócio-económico. De facto, a tendência de evolução das emissões de GEE depende de inúmeras variáveis, tais como o desenvolvimento económico, o crescimento populacional e os padrões de consumo energético. O desenvolvimento de sectores económicos como o dos transportes, a agricultura e a gestão de resíduos tem também um papel de relevo neste processo (Borrego *et al.*, 1999b).

Portugal assumiu como objectivo nacional o crescimento das suas emissões dos 6 GEE em 27% e de dióxido de carbono (CO_2) em 40% em relação ao ano de 1990. A verificação do compromisso assumido em Quioto implica a comparação de cenários de emissões com o inventário de base elaborado para o ano de 1990, que constitui o instrumento de análise para avaliação da eficácia da aplicação de medidas de controlo e/ou redução das emissões de GEE. Na figura 3 estão representadas as curvas evolutivas das emissões de GEE, construídas a partir dos inventários nacionais de emissões de gases para a atmosfera de 1990 até 1995 e extrapoladas até 2010 com base em diferentes cenários de crescimento médio (Borrego *et al.*, 1999c). Nos gráficos estão ainda representadas as metas acordadas em Quioto. Estes cenários foram determinados considerando as emissões de base multiplicadas por factores de crescimento específicos para as diferentes fontes.

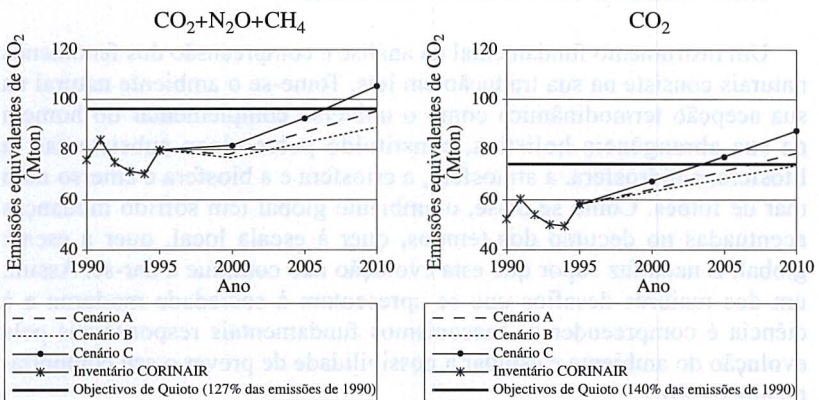


Fig. 3 — Cenários evolutivos possíveis das emissões equivalentes de GEE (Borrego *et al.*, 1999c)

A partir da figura 3 é possível verificar que, se o desenvolvimento do país ocorrer num cenário «business as usual» (cenário B), sem que ocorra qualquer intervenção mitigadora, é de esperar que em 2010 os limites estabelecidos para Portugal sejam atingidos ou mesmo ultrapassados. A situação torna-se mais gravosa no caso de se verificar um crescimento acima do esperado (cenário C), das actividades económicas, da população, dos requisitos energéticos do país ou mesmo uma alteração significativa do uso do solo. Nesta circunstâncias as emissões cresceriam, relativamente a 1990, 64% para o CO₂ e 38% para os 3 GEE, o que comprometeria significativamente o cumprimento do protocolo de Quioto.

Nos bastidores do debate político «economia e ambiente» aparece a matemática, a jogar nos dois lados deste jogo de «puxa e empurra». A matemática dá argumentos aos economistas e à sua análise financeira sobre a competitividade dos mercados, a maximização de lucros e minimização de custos, não dando no entanto resposta à questão de quantificação do custo real dos recursos naturais tendo em consideração as externalidades. A matemática dá também argumentos aos ecologistas, contribuindo para o estabelecimentos de relações de causa-efeito. Inconscientemente, toda a argumentação de base assenta em números, expressões numéricas e relações matemáticas, que contribuem para a formulação dos problemas, permitindo um conhecimento mais profundo e extenso da realidade.

2. MODELOS MATEMÁTICOS: EQUACIONAR A NATUREZA PARA MELHOR CONHECER A REALIDADE

Um instrumento fundamental da análise e compreensão dos fenómenos naturais consiste na sua tradução em leis. Tome-se o ambiente natural na sua aceção termodinâmica como o universo complementar do homem na sua abrangência holística, constituído pelos cinco subsistemas: a litosfera, a hidrosfera, a atmosfera, a criosfera e a biosfera e emerso num mar de fotões. Como se disse, o ambiente global tem sofrido mudanças acentuadas no decurso dos tempos, quer à escala local, quer à escala global. E nada faz supor que esta evolução não continue a dar-se. Assim, um dos maiores desafios que se apresentam à sociedade moderna e à ciência é compreender os mecanismos fundamentais responsáveis pela evolução do ambiente e estudar a possibilidade de prever o seu comportamento futuro.

A ciência clássica assenta no determinismo, na reversibilidade e na causalidade estrita: o futuro de um sistema está «contido» no seu presente.

A distinção entre passado, presente e futuro seria apenas uma ilusão. Hoje experimenta-se uma mudança radical na nossa atitude. Por toda a parte se descobre a multiplicidade, a temporalidade e a complexidade.

A ciência clássica preocupava-se principalmente com a regularidade, a ordem e a estruturação dos sistemas. Ora a ordem é a excepção no universo. E a desordem é incomparavelmente mais vasta do que a coesão estrutural. Mas tem-se verificado, de facto, que a irreversibilidade pode ser uma fonte de ordem. Longe do equilíbrio podem gerar-se estruturas coerentes e ordenadas. Por outro lado, existe hoje uma compreensão muito mais profunda do comportamento dos sistemas dinâmicos, não lineares, especialmente de sistemas sujeitos a condições fronteiras de não equilíbrio.

Sabe-se que a concepção clássica, baseada na estabilidade, na regularidade e na permanência, não cobre aspectos, alguns essenciais, do comportamento dinâmico dos sistemas. De acordo com Lorenz, Saltzman, Prigogine e outros, o comportamento do ambiente só se pode compreender em termos desta nova linguagem da dinâmica dos sistemas instáveis. A sensibilidade para as condições iniciais é afinal uma manifestação de falta de previsibilidade. Esta pode ser quantificada a partir dos expoentes de Liapunov, que medem a taxa média da divergência exponencial e que permitem, portanto, avaliar os limites de previsibilidade de um sistema natural.

Por exemplo, parece que há vários regimes estacionários acessíveis à dinâmica da atmosfera. Lorenz mostrou que nos sistemas meteorológicos o inverso do expoente de Liapunov é da ordem de 2,5 dias, o que significa que o número de informações meteorológicas para a previsão do tempo aumenta exponencialmente com um factor de 2,5 dias. Por isso, a previsão na atmosfera, seja do que for, a médio prazo não será nunca possível usando métodos deterministas! O conhecimento imperfeito das condições iniciais não permite formular previsões, fora dum horizonte temporal limitado, que define a previsibilidade do sistema.

As equações matemáticas que traduzem um determinado modelo conceptual de uma determinada realidade são em muitos casos também complexas e de difícil resolução. Apesar dos progressos tecnológicos no campo da informática e da computação, a capacidade de cálculo ainda é limitada, pelo que na maioria dos casos há necessidade de recorrer a simplificações dos modelos matemáticos, com base em condições iniciais, condições limite e parametrizações. Estes dados podem ser resultado da simulação de outros modelos ou podem provir de observações. Por exemplo, os dados meteorológicos, necessários nos modelos de dispersão de poluentes na atmosfera, podem ser resultado da simulação de modelos

meteorológicos de prognóstico ou de diagnóstico, que geram campos bi ou tridimensionais dos parâmetros de interesse (temperatura, velocidade e direcção do vento, etc.), ou, nas aplicações mais simplistas, provir de observações meteorológicas.

Um modelo simplificado, apesar das limitações da sua aplicação, as quais nunca devem ser negligenciadas pelo operador ou por quem analisa os resultados da simulação, não deixa de ser um poderoso instrumento de diagnóstico e de previsão dos fenómenos naturais.

O modelo de qualidade do ar, representado na figura 4, é um modelo de dispersão cujo objectivo é determinar o campo de concentrações $C(x,y,z,t)$ de um determinado poluente atmosférico provocado pela sua emissão para a atmosfera. O modelo simula as interacções das emissões de poluentes na atmosfera e o impacto resultante na qualidade do ar, utilizando como informação de base as emissões, a meteorologia, a orografia e a química da atmosfera, entre outros dados.

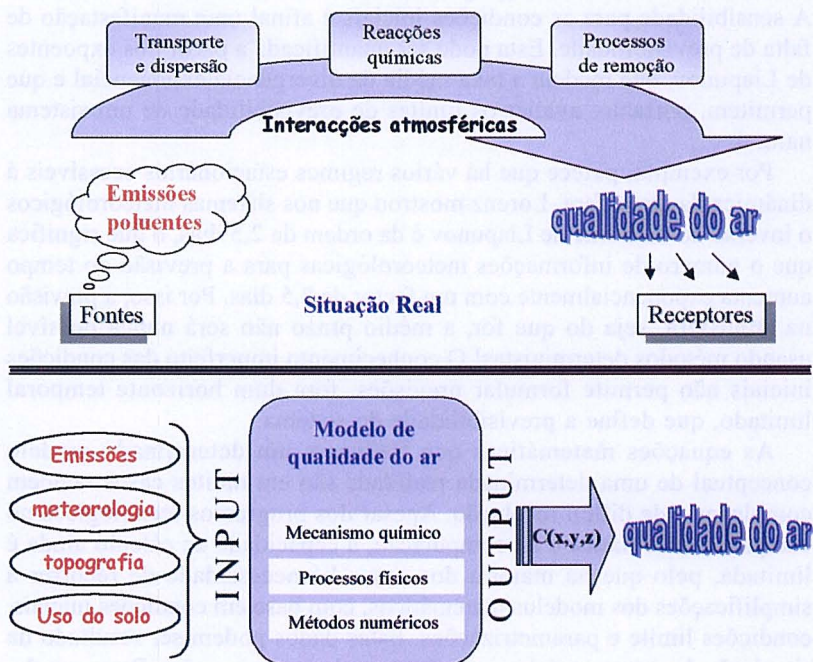


Fig. 4 — Transposição da situação real para um modelo de qualidade do ar (Borrego, 1994)

Um outro factor a ter em conta relaciona-se com o conhecimento dessa mesma realidade. A primeira etapa de apreensão da realidade é estabelecida pelos sentidos (visão, audição, tacto, paladar e cheiro); no entanto dada a sua capacidade limitada e baixa fiabilidade, os cientistas foram desenvolvendo técnicas que permitem apreender a realidade através da medição de parâmetros característicos. Este método também é falível devido às técnicas e equipamento envolvidos, também eles limitados e falíveis. A aplicação das técnicas de análise matemática e estatística é fundamental neste processo de conhecimento, na medida em que permite estabelecer intervalos de confiança nas medições, aferir equipamentos e técnicas, tornar os dados reproduzíveis e definir os limites de aplicação dessas técnicas. A determinação de parâmetros estatísticos é um dos critérios de avaliação da qualidade dos sistemas ambientais e do cumprimento de normas legisladas.

3. APLICAÇÃO DE MODELOS MATEMÁTICOS NA COMPREENSÃO DOS FENÓMENOS ATMOSFÉRICOS

A poluição atmosférica, uma importante área da engenharia do ambiente, estuda os fenómenos físicos e químicos que ocorrem no meio ar, analisando as perturbações induzidas neste receptor pela introdução de gases e partículas (poluentes). Estes agentes perturbadores podem ser originados por processos naturais (vulcões, erosão, floresta e outras actividades biológicas) ou por actividades humanas (fontes antropogénicas), designando-se «poluentes primários», ou resultar da transformação química ou física destes «poluentes secundários». Dependendo da intensidade da fonte, das características dos poluentes e do grau de desequilíbrio induzido pela perturbação, a atmosfera pode ter capacidade de regeneração, repondo o equilíbrio inicial, ou pelo contrário tender para um novo estado de equilíbrio.

As perturbações induzidas pelos poluentes atmosféricos, temporárias ou definitivas, têm efeitos ou impacto noutros meios (solo, água, seres vivos). A gestão da qualidade do ar, uma área de aplicação da poluição atmosférica, engloba o estudo das relações causa-efeito dos poluentes no ambiente, em particular nos seres vivos, o estabelecimento de indicadores de qualidade e a definição de medidas mitigadoras da poluição e do seu impacto ambiental. Numa perspectiva política, a gestão da qualidade do ar passa pela definição de estratégias para o recurso ar, através da escolha simultânea dos objectivos e dos meios que permitam pôr em prática essa política.

A qualidade do ar de uma dada região, traduzida normalmente em termos de concentrações dos poluentes, depende, para além das emissões de poluentes, dos fenómenos físicos que regulam a dinâmica da atmosfera, nomeadamente o transporte ou advecção e a dispersão, das transformações químicas sofridas pelos poluentes (por exemplo processos fotoquímicos), bem como dos mecanismos de remoção desses poluentes da atmosfera (deposição seca e húmida). A qualidade do ar pode ser avaliada directamente através da medição das concentrações dos poluentes, a partir das quais se pode inferir dos efeitos agudos, crónicos ou cumulativos sobre os receptores (seres vivos, património construído, etc.). As medições directas e o seu tratamento estatístico constituem um pilar fundamental da gestão da qualidade do ar.

A modelação é um instrumento importante nos estudos de qualidade do ar. Podem distinguir-se basicamente dois tipos de modelação: física e matemática. A primeira técnica baseia-se na construção de réplicas de elementos representativos da realidade (relevo, edifícios), com base em relações de semelhança (geométrica, cinemática, dinâmica), sendo estas réplicas inseridas num túnel de vento que simula condições de escoamento semelhantes às verificadas na atmosfera. Contrastando com esta técnica, os modelos matemáticos utilizam técnicas de análise (álgebra, cálculo) para resolver directamente parte ou o conjunto de equações que traduzem o escoamento e interações atmosféricas.

Existem vários tipos de modelos numéricos, consoante as escalas do problema que se pretende analisar e o tipo de abordagem ou de formulação do problema. Atendendo à escala espacial pode-se falar em modelos globais ou planetários, de macroscale, de mesoscale, locais ou de microscale. Em termos temporais, os modelos podem ser classificados como episódicos ou de longo termo.

A formulação matemática por detrás dos modelos é mais ou menos complexa, consoante o tipo de simplificações aplicáveis na situação para a qual o modelo foi desenvolvido. Por exemplo, a força de Coriolis, uma força existente devido à rotação da terra em torno do seu eixo, é um elemento importante no estudo das circulações de grande escala (modelos planetários e de macroscale). No entanto, o seu efeito pode ser desprezado no estudo do escoamento atmosférico de microscale pelo que as equações do movimento podem ser simplificadas por eliminação do termo correspondente. Pelo contrário, o efeito da rugosidade do solo (e das forças de atrito associadas) é um fenómeno bastante relevante em escoamentos de escala local, mas negligenciável nos de grande escala, nos quais se torna mais relevante a topografia do terreno ou o tipo de cobertura do solo.

Os sistemas mais complexos utilizados no estudo da qualidade do ar incluem um conjunto de modelos (meteorológico, de dispersão, de emissões) e módulos de cálculo (mecanismo de reacções fotoquímicas, cálculo de deposição, cálculo de sobre-elevação de penachos, cálculo de forçamento radiativo, etc.), pré-processadores (que convertem os dados de entrada para o formato específico do modelo) e pós-processadores, que tratam os dados, por exemplo estatisticamente, de modo a facilitar a sua análise e a sua comparação como valores legislados ou com valores de referência.

O sistema de modelos MAR (Barros, 1999) é um dos sistemas complexos que permitem simular o transporte, a dispersão e a produção fotoquímica num domínio de mesoscala. A sua formulação baseia-se na resolução numérica das equações base da atmosfera, assumindo algumas simplificações que estão associadas ao tipo de escoamento simulado (mesoscala) e parametrizações (parametrização da turbulência atmosférica, por exemplo). Na figura 5 apresentam-se os campos horizontais superficiais de ventos e de concentrações de ozono obtidos através da aplicação do sistema MAR à região da Grande Lisboa, considerando um dia típico de Verão e todas as fontes emissoras localizadas nesse domínio (Borrego *et al.*, 1999d). É interessante verificar a forma como o penacho gerado, principalmente em Lisboa e na península de Setúbal, acompanha o escoamento de brisa de mar que se desenvolve na região durante a tarde, verificando-se o transporte dos poluentes para sudeste do domínio.

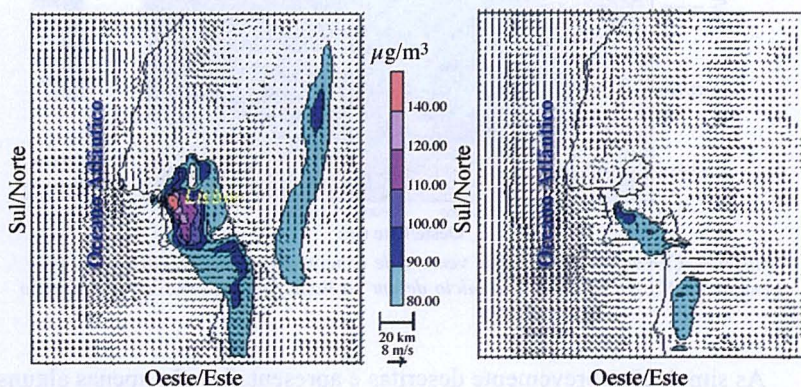


Fig. 5 — Campos de ventos e de concentração de ozono sobre a região de Lisboa, calculados pelo sistema de modelos MAR, para as 14 e 18 horas de um dia típico de Verão (Borrego *et al.*, 1999d)

Na figura 6 representam-se os campos de ventos e de concentração de um poluente considerado passivo, o monóxido de carbono (CO), obtidos recorrendo a um modelo local de qualidade do ar, o modelo VADIS (Martins, 1999). Neste exercício de simulação pretendia-se analisar o efeito na qualidade do ar e na saúde humana das emissões para a atmosfera de um incêndio florestal ocorrido na serra do Marvão.

Os níveis estimados de concentração de CO apontam para um impacto significativo na qualidade do ar. Algumas das povoações existentes no domínio de simulação poderão ser atingidas por valores de concentração desse poluente superiores aos aconselhados pela Organização Mundial de Saúde ($30\,000\ \mu\text{g}\cdot\text{m}^{-3}$), pelo que deverão ser considerados cuidados especiais com grupos mais sensíveis de pessoas, nomeadamente doentes, crianças e idosos.

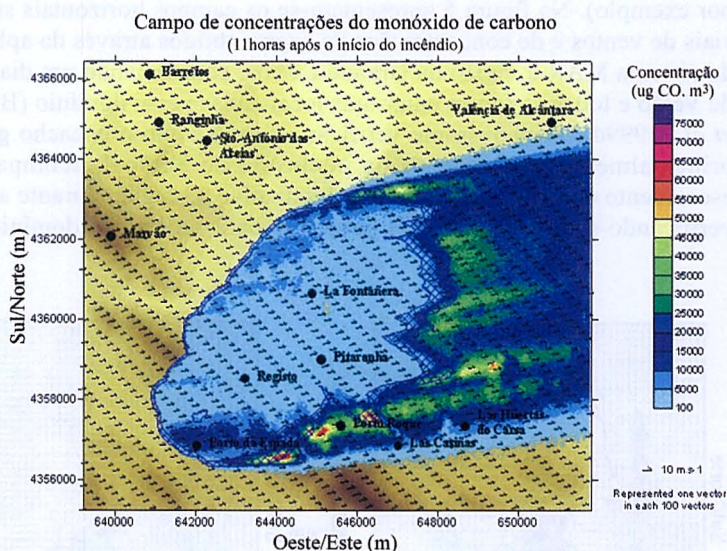


Fig. 6 — Campos superficiais de vento e de concentração de CO, calculados pelo modelo VADIS, 11 horas após o início de um incêndio florestal na serra do Marvão

As simulações brevemente descritas e apresentadas são apenas alguns exemplos de aplicação da modelação numérica ao estudo de situações reais, e ilustram bem a importância desta ferramenta como instrumento de apoio à análise, à decisão e no desenvolvimento de cenários.

6. COMENTÁRIOS FINAIS

O universo como hoje o conhecemos é um conjunto de modelos conceptuais que resultam da combinação de percepção, intuição, desenvolvimento teórico e apreensão da realidade através da medição de parâmetros característicos. Uma parte significativa desse conhecimento provém do raciocínio matemático que suporta a componente teórica e experimental da aprendizagem.

Para compreender e resolver os problemas ambientais e as suas consequências é indispensável o suporte matemático, quer na qualificação dos dados resultantes da experimentação, quer na sua descrição científica.

A aplicação de modelos matemáticos na simulação dos fenómenos atmosféricos exemplifica claramente o forte suporte da matemática para a compreensão e resolução de questões na área do ambiente.

Referências

- Barros, N. (1999) – Poluição Atmosférica por Foto-Oxidantes: O Ozono Troposférico na Região de Lisboa, Dissertação apresentada na Universidade de Aveiro para obtenção do grau de Doutor em Ciências Aplicadas ao Ambiente.
- Borrego, C. (1994) – Poluição Atmosférica I. Departamento de Ambiente e Ordenamento, Universidade de Aveiro.
- Borrego, C., Martins, J. M. e Lopes, M. (1999a) – Instrumentos Técnico Científicos de Gestão Ambiental. 4.º Encontro Nacional de Ecologia, 2-4 de Dezembro, Portugal.
- Borrego, C., Miranda, A. I., Conceição, M., Carvalho, A. C., Lopes, M. e Tchepel O. (1999b) – Climate Change and Air Quality Management: Emissions and Consequences. 6th International Conference on Environmental Science & Technology, 30 de Agosto-2 de Setembro, Samos, Grécia, vol. A, pp. 438-446.
- Borrego, C., Nunes, M., Barros, N., Portas, M. J., Monteiro, A., Lacasta, N. e Barata, P. (1999c) – A Indústria e as Alterações Climáticas. Contexto político, sócio-económico e medidas de mitigação da poluição atmosférica. AMB-QA-08/99. Relatório Final do estudo elaborado no âmbito do projecto: Universidade do Norte – Horizonte 2000. Departamento de Ambiente e Ordenamento, Universidade de Aveiro, Aveiro, Portugal.
- Borrego *et al.* (1999d) – AMAZOC, Ambiente atmosférico em zonas costeiras: avaliação da capacidade de carga do ecossistema. Relatório científico final. Contrato PRAXIS 3/3.2/AMB/38/94.

- Borrego, C., Miranda, A. I., Coutinho, M., Ferreira, J. e Carvalho, A. C. (2000) – Air Quality Management in Portugal: Example of Needs and Available Tools. 2nd International Workshop on Biomonitoring of Atmospheric Pollution, 28 de Agosto-3 de Setembro, Açores, Portugal.
- Martins, J. M. (1999) – Dispersão de poluentes atmosféricos em condições de vento fraco. Dissertação apresentada na Universidade de Aveiro para obtenção do grau de Doutor em Ciências Aplicadas ao Ambiente.

Referências

- Borrego, C. (1999) – Poluição Atmosférica por Foto-Oxidantes. O Óxono. *Introdução ao Registo de Labor. Dissertação apresentada na Universidade de Aveiro para obtenção do grau de Doutor em Ciências Aplicadas ao Ambiente.*
- Borrego, C. (1991) – Poluição Atmosférica. I. Departamento de Ambiente e Ordenamento, Universidade de Aveiro.
- Borrego, C., Martins, J. M. e Lopes, M. (1999) – Instrumentos Técnicos. Comissão de Gestão Ambiental. Fórum Nacional de Poluição, 2-4 de Dezembro, Portugal.
- Borrego, C., Miranda, A. I., Coutinho, M., Carvalho, A. C., Lopes, M. e Tóthfalvi, G. (1999) – Climate Change and Air Quality Management. *Environm. Sci. & Technology. 33rd International Conference on Environmental Science & Technology. 30 de Agosto-2 de Setembro. Szekes, Orosz. Vol. A, pp. 418-420.*
- Borrego, C., Pinho, M., Borrego, R., Borrego, M. L., Monteiro, A. I. e Costa, N. S. (1999) – A Indústria e as Alterações Climáticas. *Conferência ANB-2000-Comunidade e medidas de mitigação de poluição atmosférica. ANB-2000-99. Relatório final do estudo elaborado no âmbito do projecto. Universidade do Minho - Horizonte 2000. Departamento de Ambiente e Ordenamento, Universidade de Aveiro, Aveiro, Portugal.*
- Borrego, C. et al. (1995) – AMASOL - Ambiente associado em zonas costeiras. *Relatório de actividades de campo de monitorização. Laboratório científico final. Centro PEXIS 333 ZAMARRIM.*

A matemática e a forma de ver o mundo

RESUMO

«Medir e representar a Terra» é uma expressão que pode descrever resumidamente a actividade do engenheiro geógrafo.

A medição e representação do espaço que nos rodeia, e do mundo em geral, sempre foi uma necessidade do homem, com aplicação em diferentes actividades: transportes, estratégia militar, agricultura, caracterização da propriedade, planeamento, etc., sem excluir as de natureza predominantemente especulativa e científica.

Os processos de medição (geodesia) e representação (cartografia) da Terra foram evoluindo ao longo dos tempos, com recurso a técnicas e tecnologias cada vez mais sofisticadas. Para além dos aspectos puramente tecnológicos, a conceptualização e modelação da realidade carecem de formulações matemáticas adequadas.

Para modelar matematicamente o espaço apresentam-se sempre várias possibilidades, que diferem entre si de diversas formas, por exemplo, do tipo de função e variáveis usadas, das técnicas numéricas associadas à capacidade e eficácia computacional, etc. Deste modo, a representação do espaço depende da escolha dos números e das formulações matemáticas utilizadas.

Essa representação, sob a forma de cartografia tradicional ou, mais modernamente, de visualização apoiada em sistemas de informação, condiciona a nossa forma de ver o mundo. Em tudo isto, obviamente, a matemática não é inocente.

Na presente comunicação faz-se uma retrospectiva da evolução dos processos de medição e representação da Terra e aponta-se, com alguns exemplos, de que modo a nossa percepção do espaço e dos factos pode ser condicionada pelos números e formulações matemáticas. Em suma, procura-se descobrir algumas das teias com que a matemática enreda a realidade.

1. INTRODUÇÃO

A engenharia geográfica actual é herdeira de um conjunto de actividades que acompanham praticamente todo o percurso da humanidade, e que se podem resumir com a expressão «medir e representar a Terra».

De facto, o ser humano sentiu desde sempre a necessidade de conhecer o espaço que o rodeava, de forma a poder guardar e transmitir o conhecimento que ia adquirindo dos diferentes fenómenos espaciais (zonas de caça, rotas de navegação, territórios ocupados, etc.) que influenciavam a sua vida.

Obviamente que para representar é necessário um conhecimento prévio do tamanho e forma do espaço a representar, ou, dito de outra maneira, da posição tridimensional dos objectos e das ocorrências espaciais num determinado sistema de referência. Numa primeira fase da história do homem, esta actividade não tinha quaisquer características científicas, isto é, não se apoiava em princípios que permitissem aferir o rigor com que a representação era feita. À medida que o conhecimento no âmbito das ciências (como a matemática e a física) foi progredindo, foi sendo possível transportar para as actividades de medição e representação do espaço os conceitos que permitiram a criação das ciências associadas à medição e representação da Terra, a geodesia e a cartografia.

Desta forma, a geodesia e a cartografia autonomizaram-se enquanto áreas científicas, tendo sempre como base os princípios da matemática e da física, essencialmente, e mais recentemente das ciências da informação. As formulações da matemática são particularmente determinantes para a modelação de fenómenos. As diferentes possibilidades de criar os modelos que representam a realidade espacial oferecidas por esta ciência condicionam claramente a representação associada a esses modelos.

Quando alguém pretende ir para um local onde nunca tenha estado, a melhor maneira de ter um conhecimento prévio desse lugar é a consulta de um mapa, elaborado segundo um conjunto de leis matemáticas que

traduzem um conjunto de características que o seu autor pretendeu evidenciar, e concebido a partir da utilização de um determinado modelo matemático e físico da Terra. Pode dizer-se assim que o mapa é um modelo da realidade e que a percepção da realidade é feita não a partir da sua observação directa, mas essencialmente do modo como essa realidade foi modelada e transmitida.

Este facto tem actualmente uma expressão mais vincada, decorrente da utilização de sistemas de informação, que também introduzem factores que podem modificar a forma como vemos o mundo que nos rodeia.

Nos dois capítulos seguintes far-se-á uma abordagem breve à evolução dos processos de medição e representação da Terra e do modo como as formulações matemáticas condicionaram e condicionam a nossa forma de ver o mundo, sendo o quarto capítulo dedicado ao posicionamento e navegação, com recurso às técnicas espaciais e às tecnologias da informação.

2. A FORMA E AS DIMENSÕES DA TERRA

2.1 Primeiras abordagens

Como foi referido anteriormente, as primeiras noções sobre a forma e as dimensões da Terra careciam de qualquer fundamentação científica. Os filósofos tinham a sua interpretação muito própria do mundo, baseada essencialmente no conhecimento transmitido pelos viajantes que percorriam terras e mares, necessariamente em zonas bastante restritas.

Assim, a escola de Mileto, através de Tales, cerca de 600 a.C., concebia o mundo como um disco achatado envolto por um oceano infinito (figura 1). Esta ideia já tinha sido expressa por Homero na *Iliada*, cerca de 800 a.C. Esta obra designava o oceano circundante por *Oceanus*. A concepção de Anaximandro, da mesma escola, era semelhante, apenas diferindo na orientação Norte-Sul do disco.



Fig. 1 — Concepção de Tales (cerca de 600 a.C.)

Cerca de 250 anos mais tarde, Aristóteles concebe a Terra como uma esfera e faz a primeira alusão à existência de um campo gravítico, obviamente com outra designação, introduzindo simultaneamente a primeira abordagem científica. A argumentação incluía o facto de a forma da sombra da Terra na Lua durante os eclipses ter sempre um contorno circular, independentemente da posição relativa dos dois corpos no espaço; ora tal só seria possível se a Terra fosse uma esfera (Jones, T., 1967). Concomitantemente, foi utilizado um outro argumento: quando um barco se afastava no horizonte, o que desaparecia em último lugar eram os mastros e as velas.

Esta evolução do conhecimento da forma da Terra, associada ao desenvolvimento da astronomia, permitiu a Eratóstenes, cerca de 230 a.C., realizar as primeiras operações que conduziram à determinação do raio da esfera terrestre e, conseqüentemente, das dimensões do nosso planeta; é por isso considerado por vários autores (Torge, W., 1991) o fundador científico da geodesia.

Possuindo o conhecimento de que num determinado dia do ano os raios solares incidiam perpendicularmente num poço em Assuão quando o Sol atingia a altura máxima (isto é, cruzava o meridiano do lugar), e de que noutra lugar (Alexandria), situado mais a norte sobre o mesmo meridiano, no mesmo dia, os raios solares provocavam sombra nos objectos verticais, foi possível a Eratóstenes determinar o valor angular do arco de meridiano (Φ) entre os dois lugares. Bastava-lhe agora medir o comprimento desse arco de meridiano (S) para determinar o raio (R) da esfera (figura 2) usando a conhecida relação matemática

$$R = S / \Phi,$$

com Φ expresso em radianos.

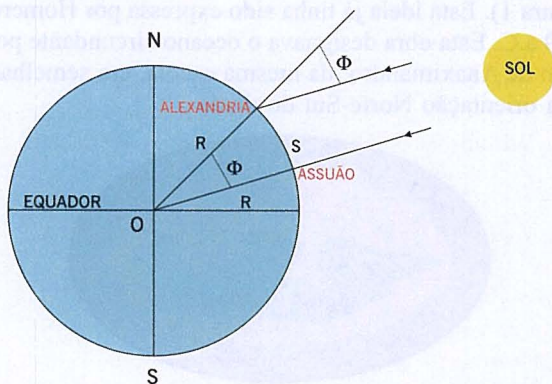


Fig. 2 — Determinação do raio da Terra por Eratóstenes

Reza a história que Eratóstenes mediu o comprimento do arco de meridiano a partir das distâncias patentes na informação cadastral egípcia, as quais eram obtidas através da contagem de passos de camelo, conhecendo o equivalente linear de cada passo.

O valor encontrado por Eratóstenes para a distância foi de cerca 5000 estádias egípcias, e o valor angular do arco de meridiano foi de cerca $1/50$ de uma circunferência. Transportando para unidades métricas (1 estádia = $157,5$ m), encontra-se um valor de 6267 km para o raio da esfera, o que é cerca de 2% inferior ao actualmente estabelecido (Torge, W., 1991).

Esta determinação, pelos resultados obtidos, é notável. Deve no entanto ter-se em consideração que as operações realizadas por Eratóstenes continham várias fontes de erro. Por um lado, sabe-se hoje que Assuão e Alexandria não se encontram no mesmo meridiano, e é óbvio que a determinação da distância sofreu as vicissitudes decorrentes da forma como foi realizada. Encontramos aqui um bom exemplo de como os erros de alguma forma se compensaram, ao invés do que as leis de Murphy preconizam.

Durante todo o período da história que se seguiu, até ao século XVII, foram realizadas várias determinações do raio da Terra baseadas no modelo esférico, em especial a partir do momento em que passaram a estar disponíveis instrumentos para observação astronómica. O procedimento seguia em geral a essência do método de Eratóstenes: medir os valores angulares e lineares de um arco de meridiano.

Os procedimentos operacionais conduziram à concepção do que hoje se designa por *rede geodésica*. O valor angular do arco de meridiano era determinado por observação astronómica em dois pontos situados sobre (ou quase) o mesmo meridiano. Quanto à medição da distância entre esses pontos, era conseguida com recurso à expansão de uma distância medida sobre um ou mais lados de um triângulo através de uma cadeia de triângulos. São de destacar as operações realizadas por Tycho Brahe, na Dinamarca, e por Willebrord Snellius, na Holanda, nos finais do século XVI.

2.2 O campo gravítico e a figura da Terra

As considerações sobre o campo gravítico terrestre iriam desempenhar um papel fundamental na etapa seguinte da determinação da forma e dimensões da Terra, que consistiu essencialmente no reconhecimento de que a forma do nosso planeta se assemelharia à de um elipsóide achatado nos pólos.

O primeiro sinal deste facto foi obtido por Cassini, em 1691, quando observou Júpiter com um telescópio e verificou que este planeta tinha um achatamento polar, quantificado pelo próprio Cassini: o diâmetro equatorial seria 1/15 maior que o eixo polar.

O segundo sinal proveio do francês Richer. Este astrónomo tinha sido enviado pelo seu governo para a ilha de Cayenne, situada na zona equatorial, com o objectivo de determinar as paralaxes de vários planetas do sistema solar, os movimentos de Mercúrio, as coordenadas astronómicas de algumas estrelas não visíveis em Paris, etc., e, o mais importante de tudo para a nossa história, o comprimento do pêndulo de segundos em Cayenne.

Ora Richer verificou que teria de o encurtar relativamente ao valor que trazia de Paris, onde tinha sido perfeitamente ajustado, por ter verificado que perdia diariamente 2 minutos e 28 segundos nos seus relógios de pêndulo. Quando regressou a Paris verificou que tinha de realizar a operação inversa, por comparação com o pêndulo de Paris, para obter novamente oscilações de 1 segundo. Este mesmo fenómeno foi verificado por Halley, em 1677, quando realizou observações em Santa Helena relativamente ao comprimento do pêndulo que tinha transportado de Inglaterra, e confirmado por experiências posteriores.

É neste cenário que surge Isaac Newton. Conjugando estes dois sinais, Newton estipulou, na *Proposição XVIII, Teorema XVI*, da sua obra *Philosophiæ naturalis principia mathematica* (1678), que «os eixos dos planetas são menores que os diâmetros perpendiculares aos eixos». O achatamento de Júpiter observado foi explicado considerando que a matéria está em estado fluido e que o seu comportamento num corpo em rotação seria descer ao longo do eixo para o equador, alargando desta forma o diâmetro equatorial e encurtando o eixo.

A consequência seguinte é que a força da gravidade é mais intensa nos pólos do que no equador, já que a distância é ali menor e a força centrífuga é variável com a latitude, exercendo-se numa direcção oposta à força resultante da atracção (newtoniana) da massa da Terra sobre uma massa à sua superfície. A figura 3 documenta este facto relativamente às acelerações (aceleração centrífuga a_c , aceleração newtoniana a_N e aceleração da gravidade a_g).

As normas destes vectores são dadas por

$$|\overline{a_N}| = GM / R^2 \quad \text{e} \quad |\overline{a_c}| = \omega^2 r,$$

em que G , M , R , ω e r representam, respectivamente, a constante de gravitação universal, a massa da Terra, a distância de um ponto da superfície ao centro da Terra, a velocidade angular de rotação da Terra e o raio do paralelo onde o ponto se localiza.

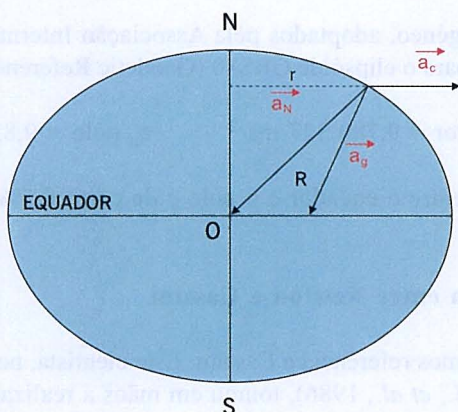


Fig. 3 — A aceleração da gravidade como resultante das acelerações centrífuga e newtoniana

Foi possível a Newton, a partir das diferenças de gravidade encontradas, estabelecer o elipsóide rotacional como figura de equilíbrio de uma Terra homogénea, fluida e em rotação. Mais ainda, determinou que o achatamento do elipsóide teria o valor de $1/230$, com base na teoria de uma Terra homogénea; Huygens determinou o valor de $1/576$, com base numa grande concentração de massas no centro. Registe-se que o valor actual é de cerca $1/298$.

O achatamento geométrico f pode obter-se a partir dos semi-eixos maior (a) e menor (b) da elipse meridiana, de acordo com a relação

$$f = (a - b) / a.$$

Os valores dados por Newton e Huygens podem ser obtidos a partir das equações estabelecidas por Clairaut em 1743:

$$\begin{aligned} f &= 3(C-A) / (2 M a^2) + m / 2; \\ f_g &= 2 f - 9(C-A) / (2 M a^2) + m; \\ f + f_g &= 5 / 2 m, \end{aligned}$$

em que A e C são os momentos de inércia, m é a relação entre as acelerações centrífuga e newtoniana equatorial ($m = 1/288$) e f_g é o achatamento gravítico.

Na hipótese de Newton, $f = f_g = 5/4 m$, e na de Huygens $f = 1/2 m$. Note-se que os valores da gravidade teórica equatorial e polar para um

elipsóide homogéneo, adoptados pela Associação Internacional de Geodesia em 1980 para o elipsóide GRS80 (Geodetic Reference System), são:

$$a_g \text{ equador} = 9,780\,327 \text{ ms}^{-2} \quad a_g \text{ pólo} = 9,832\,129 \text{ ms}^{-2}.$$

A variação entre o equador e o pólo é de cerca 0,5%.

2.3 A polémica entre Newton e Cassini

Já antes fizemos referência a Cassini. Este cientista, no final do século XVII (Kakkuri, J., *et al.*, 1986), tomou em mãos a realização do projecto de medição do arco de meridiano entre Dunquerque e Perpignan, da forma como o havia feito Picard. Das suas medições concluiu que «[...] parece bastante evidente que os graus de um meridiano são maiores quanto mais perto do equador e, contrariamente, diminuem à medida que se aproximam do pólo».

Ora esta afirmação era exactamente contrária à teoria de Newton, e em redor desta circunstância estabeleceu-se uma enorme polémica. A forma de resolver a questão foi encontrada pela Academia Real das Ciências francesa, que em 1734 decidiu realizar uma experiência controlada.

Assim, foram organizadas duas expedições a locais afastados em latitude. Em 1736 Maupertuis partiu para a Lapónia, de onde regressou em Agosto de 1737, com as observações realizadas. Da sua equipa faziam parte Celsius e Clairaut, entre outros. A outra expedição teve como destino o Peru e era chefiada por Bouguer. Enquanto a primeira operação se desenrolou sem percalços de maior, a segunda sofreu várias vicissitudes, o que fez que só em 1744 fosse possível a Bouguer apresentar os seus resultados.

Os cálculos realizados permitiram, por um lado, confirmar a teoria de Newton em detrimento da de Cassini e, por outro, calcular o valor de $1/207$ para o achatamento polar. Os resultados discrepantes de Cassini foram explicados por um lado por alguns erros de cálculo e por outro pela pouca precisão dos instrumentos utilizados.

Esta operação e o conjunto de acontecimentos que a rodearam, envolvendo os mais notáveis cientistas da época e de todos os tempos, constituíram um dos marcos mais notáveis da ciência, não só pelos meios que foram utilizados como pelo conjunto de desenvolvimentos científicos

resultantes. Revelou também o enorme poder da matemática e da física, que levou Voltaire a afirmar (Kakkuri, J., *et al.*, 1986):

*Vous avez confirmé dans ces lieux pleins d'ennui
Ce que Newton connut sans sortir de chez lui.*

2.4 A evolução do conhecimento até à nossa época

Depois do reconhecimento de que o modelo matemático que mais se adequava à forma da Terra era um elipsóide de revolução em torno do eixo menor, assistiu-se a um grande desenvolvimento científico e à generalização das operações de determinação de arcos de meridiano por forma a obter medidas cada vez mais rigorosas do nosso planeta. As operações de medição foram também aproveitadas para criar redes geodésicas nacionais e regionais; a este propósito, mostra-se na figura 4 a Triangulação Geodésica de 1ª Ordem de Portugal, concebida por Filipe Folque em 1876. No entanto, as técnicas utilizadas eram semelhantes, apenas diferindo numa cada vez maior precisão dos equipamentos e no refinamento das técnicas de cálculo.

A partir dos anos 60 do século XX, a utilização do espaço exterior veio trazer um novo impulso a este domínio do conhecimento, contribuindo para uma espectacular melhoria da precisão na determinação dos parâmetros necessários ao conhecimento da forma e dimensões da Terra.

Portugal foi palco de uma das primeiras operações de medição baseadas na utilização de satélites artificiais, que consistiu na ligação do arquipélago dos Açores ao continente através da observação dos satélites passivos Echo 1 e Echo 2 (Levallois, J. J., 1988), realizada em conjunto pelo Instituto Geográfico Nacional de França e pelo Instituto Geográfico e Cadastral de Portugal. A técnica utilizada foi a triangulação espacial, que tinha os dias contados, já que pouco depois, nos anos 70, surge o sistema Transit do Departamento de Defesa dos Estados Unidos, e nos anos 80 o GPS.

Mas não foi só a utilização de satélites artificiais que deu a sua contribuição para o melhor conhecimento da forma e dimensões da Terra. Outras técnicas são usadas, como o VLBI (Very Long Baseline Interferometry), o SLR (Satellite Laser Ranging) e o LLR (Lunar Laser Ranging).

Para se ter uma percepção global da evolução do conhecimento nesta matéria, apresentam-se na tabela seguinte algumas das determinações do semi-eixo maior e achatamento do elipsóide ao longo dos tempos. Registe-se que alguns dos valores apresentados são aproximados, e que a precisão associada ao último é de cerca de 1 metro.

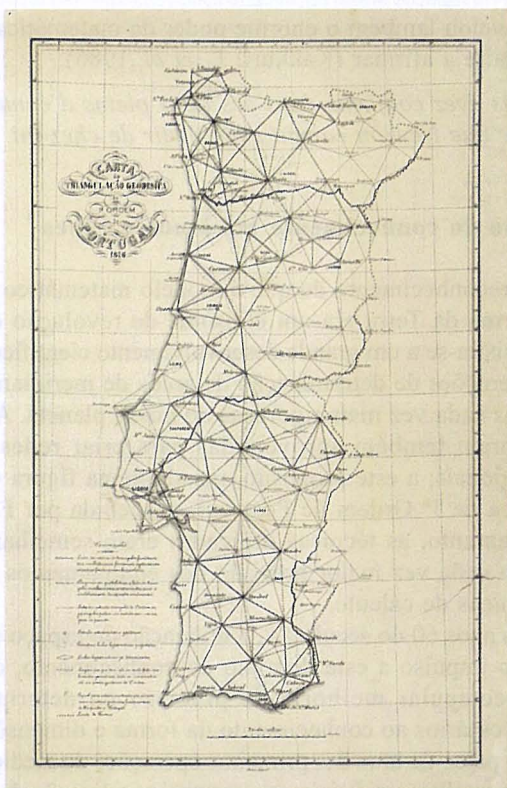


Fig. 4 — Carta da Triangulação Geodésica de 1.ª Ordem de Filipe Folque, 1876.
(Mapoteca do Instituto Português de Cartografia e Cadastro)

	Ano	a (m)	1/f
Eratóstenes	230 a.C.	6 267 000	∞
Bouguer, Maupertuis	1744	6 397 300	207
Delambre	1810	6 376 985	308.64
Everest	1830	6 377 276	300.8
Airy	1830	6 377 563	299.3
Bessel	1840	6 377 397	299.15
Clarke	1880	6 378 245	293.5
Hayford	1910	6 378 388	297
Jeffreys	1948	6 378 097	297.28
GRS80	1980	6 378 137	298.26

3. A REPRESENTAÇÃO DA TERRA

3.1 Os modelos matemáticos da cartografia

As aventuras da medição e representação da Terra estão claramente associadas. As primeiras tentativas cartográficas não tinham qualquer fundamentação científica e as representações da realidade eram condicionadas pelo conhecimento que se ia adquirindo do espaço envolvente. O mapa surge como necessidade de transmitir esse conhecimento, isto é, constitui-se como um meio de comunicação, com uma semântica própria.

Sem pretensões de fazer uma história da cartografia, que não cabe obviamente num escrito desta natureza, vale a pena referir que o cadastro da propriedade egípcio constituiu talvez a primeira forma de cartografia com alguns fundamentos científicos e, como se viu antes, esteve na base da primeira determinação do valor do raio terrestre.

Esta história vai ser retomada apenas no século XVI, quando Gerardo Kramer concebe, para navegação, a primeira representação da Terra baseada numa formulação matemática. Atendendo às finalidades dessa carta, a linha de rumo (e azimute, obviamente) constante, a loxodrómia, é representada por uma recta, bem como os meridianos (Guedes, J., 2000). Esta representação designa-se por «projecção de Mercator» (figura 5).

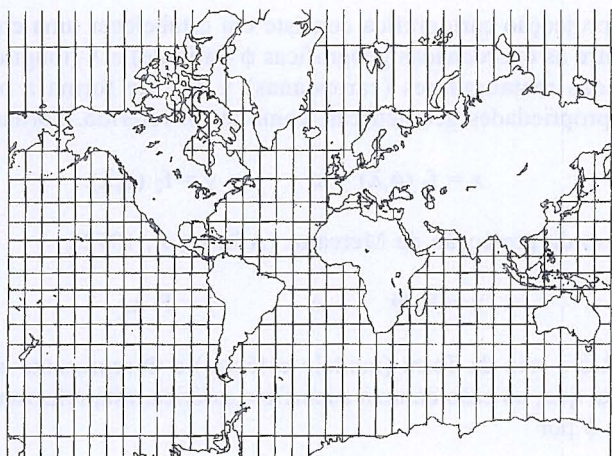


Fig. 5 — Projecção de Mercator

A projecção de Mercator é uma transformação matemática inspirada numa projecção cilíndrica normal (projecção geométrica pura, em que a superfície terrestre é envolvida por uma superfície cilíndrica com o eixo coincidente com o eixo de rotação da Terra) e tem a propriedade de manter os ângulos infinitesimais (e consequentemente as formas), designando-se por isso de «conforme».

O problema da cartografia é realmente complicado, já que não é possível planificar uma esfera ou um elipsóide. A figura 6 mostra as analogias geométricas que são normalmente utilizadas para as representações cartográficas, bem como a impossibilidade de representar a superfície terrestre num plano sem introduzir alguma espécie de deformação.

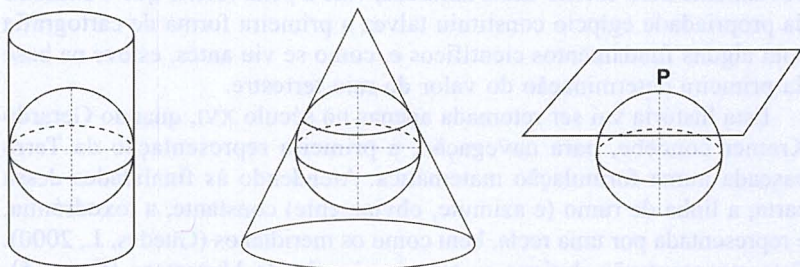


Fig. 6 — Analogias geométricas para as projecções cilíndricas, cónicas e azimutais

Uma projecção cartográfica consiste em estabelecer uma correspondência entre as coordenadas geográficas ϕ (latitude) e λ (longitude) e as coordenadas rectangulares (cartesianas) x e y , de forma a preservar algumas propriedades geométricas, como já foi referido. Formalmente,

$$x = f_1(\phi, \lambda) \quad \text{e} \quad y = f_2(\phi, \lambda).$$

No caso da projecção de Mercator (Afonso, J., 1972),

$$x = R \Phi \quad \text{e} \quad y = R \lambda,$$

em que R é o raio da Terra (modelo esférico) e Φ representa a latitude isométrica, que, no caso de uma superfície esférica, está relacionada com a latitude ϕ por

$$\Phi = \ln (\operatorname{tg} (\pi/4 + \phi/2)).$$

Desta formulação resulta que a distância entre os paralelos vai aumentando do equador para os pólos, e é a forma de conseguir representar por uma recta a linha de rumo constante, com óbvias vantagens para a navegação, e manter a propriedade da conformidade.

A necessidade de cartografar países em diferentes localizações geográficas, com determinadas configurações geométricas, e regiões mais ou menos extensas do globo, fez que fossem sendo criadas projecções cartográficas cujas propriedades geométricas se adequassem melhor às finalidades pretendidas.

Em Portugal utiliza-se actualmente a projecção de Gauss (1820), com a formulação introduzida posteriormente por Krüger (1912), pelo que também se designa por Gauss-Krüger. É uma projecção conforme, e pode dizer-se, como analogia geométrica, que corresponde, grosso modo, à projecção da superfície do elipsóide sobre uma superfície cilíndrica cujo eixo é perpendicular ao eixo de rotação da Terra; por este motivo também é por vezes designada por «transversa de Mercator». O meridiano central representa-se por uma recta. A correspondência entre as coordenadas geográficas e as rectangulares é estabelecida por (Afonso, J., 1972)

$$x = \lambda N \cos \phi + (\lambda^3/3!) N A \cos^3 \phi + (\lambda^5/5!) N B \cos^5 \phi;$$

$$y = \sigma + (\lambda^2/2!) N \sin \phi \cos \phi + (\lambda^4/4!) N C \sin \phi \cos^3 \phi,$$

em que σ representa o comprimento de um arco de meridiano (dependente da latitude) e as outras variáveis são igualmente dependentes da latitude do ponto:

$$\sigma = g_1(\phi); \quad N = g_2(\phi);$$

$$A = g_3(\phi); \quad B = g_4(\phi); \quad C = g_5(\phi).$$

Refira-se a propósito que o cálculo de s não é directo no caso do modelo elipsóidico, já que somos confrontados com um integral elíptico de 3.ª espécie, pelo que se recorre a métodos de integração numérica. Novamente, evidencia-se aqui a importância da matemática através de uma das suas disciplinas.

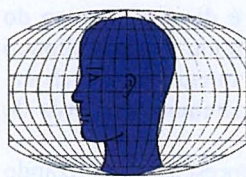
A figura 7 ilustra a maneira como Portugal é representado utilizando esta projecção, particularmente adequada a uma região com um desenvolvimento Norte-Sul. O meridiano central da projecção encontra-se no centro de Portugal.



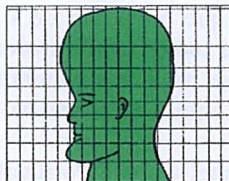
Fig. 7 — Projecção de Gauss utilizada em Portugal

3.2 Como se adequam os modelos à realidade?

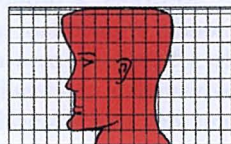
A teoria das projecções cartográficas tem uma ligação directa com a matemática, em particular com a geometria diferencial. Existem inúmeras, concebidas de forma a satisfazer determinados requisitos, que podem ser técnicos (manutenção de ângulos infinitesimais ou áreas finitas, por exemplo) ou ainda políticos ou estratégicos (Black, J., 1997). De facto, a história da cartografia é preche de situações em que a escolha de determinada projecção é condicionada pela visão que se pretende dar da realidade. A matemática, enquanto ciência de base para a cartografia, aparece-nos assim como um instrumento para modelar essa realidade. A figura 8 (Guedes, J., 2000) mostra o modo como ocorrem as deformações resultantes de três modelações diferentes.



Mollweide



Mercator



Cilíndrica equivalente

Fig. 8 — Deformações produzidas por três projecções cartográficas distintas

«Os mapas são um modelo selectivo da realidade» (Black, J., 1997). Esta frase resume bem o que antes foi expresso. Além do exemplo mostrado na figura 8, e para se ter uma ideia mais clara destas teias tecidas pela matemática, apresentam-se na figura 9 várias representações da Gronelândia, onde é patente a enorme diferença de áreas e de forma resultantes de cada representação. Para a generalidade das pessoas, esta ilha apresenta-se como possuindo enormes dimensões, quando, de facto, a sua área é de cerca quatro vezes a área da França.

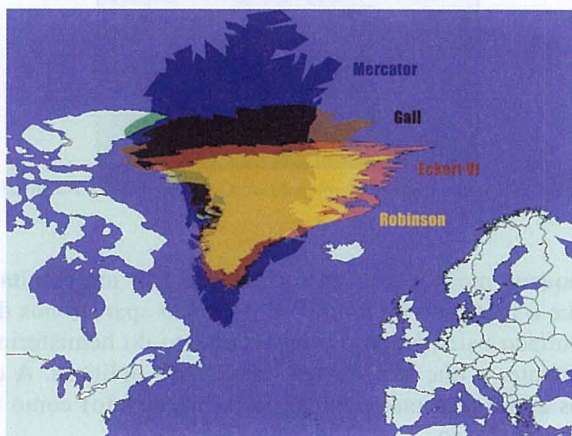


Fig. 9 — A Gronelândia representada por quatro projecções cartográficas distintas

Posto isto, uma questão se coloca. Qual o modelo que deve ser utilizado? A resposta já foi sendo dada antes; a escolha do modelo matemático é condicionada pelas características que se pretende inculir à representação, de acordo com a finalidade. Exemplificando, o sistema de projecção utilizado em Portugal não deve ser alargado globalmente, pois provoca distorsões que não são razoáveis e modifica completamente a ideia que temos sobre a posição relativa dos continentes. O resultado da extensão da projecção é patente na figura 10.

Como nota final neste tema, vejamos o que acontece se for feita uma pequena alteração a um dos modelos matemáticos antes referidos. Mudando o sinal nas fórmulas que caracterizam a projecção de Mercator, obtém-se

$$x = -R \Phi \quad e \quad y = -R \lambda.$$



Fig. 10 — Extensão da projecção de Gauss usada em Portugal

Esta pequena modificação na formulação tem um resultado prático drástico, que é mostrado na figura 11: o mundo aparece-nos de pés para o ar. O resultado psicológico é o aparecimento do hemisfério sul como parte dominante, o que tem óbvias incidências políticas. A cartografia aparece-nos assim (e a matemática, por arrastamento) como um instrumento de manipulação.



Fig. 11 — O mundo é dominado pelo hemisfério Sul

3.3 A intervenção das ciências da informação

É difícil encontrar uma actividade do mundo actual em que não haja intervenção das ciências da informação, em níveis mais ou menos complexos. A cartografia, e em geral as actividades que lidam com informação geo-referenciada, foi das primeiras a procurar adaptar os seus paradigmas aos métodos tornados possíveis pela informática, cuja expressão mais visível são os sistemas de informação geográfica.

Um dos paradigmas que tiveram de ser modificados foi o que diz respeito à forma de representação dos objectos cartográficos, de modo a tornar possível a sua percepção pelos sistemas computacionais e a construção de bases de dados gráficas. A cartografia digital criou os seus formatos próprios, que são basicamente o matricial e vectorial (Molenaar, M., 1998). Novamente, a matemática foi fundamental neste processo, já que no primeiro caso é claramente visível uma intervenção da álgebra e no segundo as noções de topologia são fundamentais para criar estruturas de dados manejáveis.

Com este novo meio de produção, processamento, armazenagem e manipulação de dados espaciais tornou-se também possível a sua rápida integração, permitindo criar uma mais-valia na sua exploração, que conduz por sua vez à obtenção de mais informações (Korte, G., 1997). É o caso, por exemplo, da associação de bases de dados gráficas e alfanuméricas e de imagem, o que permite uma percepção mais correcta da realidade.

A figura 12 apresenta um exemplo desta associação. Nela estão representados um ortofotomapa digital (matricial) da região de Avis, vendo-se parte da albufeira do Maranhão, e a cartografia digital (vectorial) codificada de forma a poder ser explorada pelo sistema de informação (IPCC, 2000). No lado esquerdo da figura encontra-se a legenda. Este sistema permite activar ou desactivar níveis de informação e fazer diferentes tipos de análise.

Seria monótono, e nem esse é o propósito deste artigo, referir todas as áreas da informação geográfica a que a informática trouxe uma mais-valia. Refira-se apenas, como mais um exemplo, a construção de modelos numéricos de elevação (Fonseca, A., *et al.*, 2000), que permitem também a exploração por sistemas de informação geográfica e possibilitam vários tipos de análise que, sem estes meios, seriam bastante mais morosas e complicadas.

É o caso do cálculo de intervisibilidades, importante para o planeamento em telecomunicações, e de pendentes de terreno e exposição solar, com aplicações óbvias nas áreas agrícola e hidrológica, entre outras. Também aqui os modelos matemáticos associados são importantes, pois

a sua maior ou menor capacidade de interpretar a realidade a partir de um conjunto finito e discreto de dados condiciona a forma como é representado o relevo e os resultados das análises espaciais.

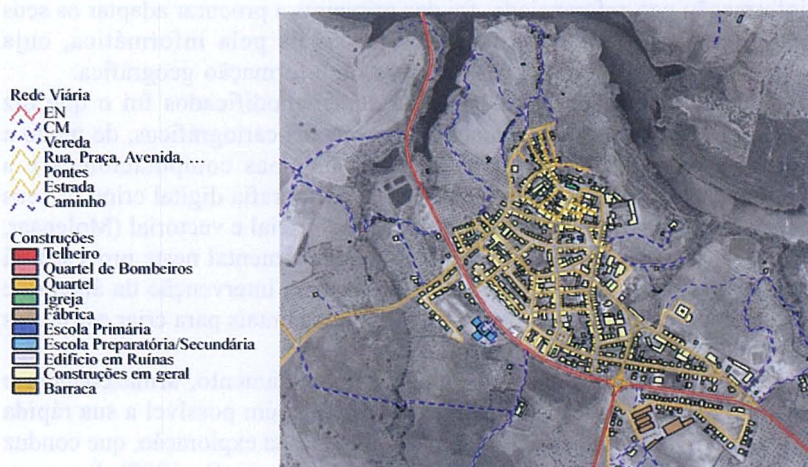


Fig. 12 — Sistema de informação geográfica

4. O POSICIONAMENTO E A NAVEGAÇÃO

4.1 Como nos localizamos?

Na actividade humana actual e na sociedade da mobilidade em que vivemos é cada vez mais fundamental conhecermos a nossa posição no espaço e sabermos qual a direcção a tomar. Por outras palavras, o posicionamento e a navegação tornaram-se actividades de tal forma importantes que já é comum encontrar viaturas terrestres com estes sistemas.

É clara a ligação à temática que estamos a tratar. No fim de contas, está-se a lidar com cartografia digital e com sistemas de posicionamento por satélite, na maioria dos casos o Sistema de Posicionamento Global (GPS).

Será que aqui também a matemática influi na nossa forma de ver o mundo? É o que vamos ver a seguir.

O GPS é um instrumento fundamental para a geodesia actual, como antes foi referido. As aplicações nesta área vão do posicionamento de alta precisão até à determinação dos parâmetros de orientação da Terra, passando pelo controlo de deformações e pela geodinâmica. Embora o sistema tenha iniciado o seu funcionamento há cerca de vinte anos, apenas na última década passou a estar acessível de forma consistente. Nos últimos cinco anos assistiu-se a um proliferar de aplicações e utilizações, decorrentes não só do conhecimento público cada vez maior das suas capacidades e potencialidades como também do baixo custo dos receptores, acessíveis à bolsa do cidadão médio. É evidente que estamos a falar agora de receptores que apenas utilizam a informação contida no código C/A e destinados a aplicações fundamentalmente na área do posicionamento de baixa precisão e navegação, e não dos equipamentos utilizados em geodesia (Torres, J., *et al.*, 1996).

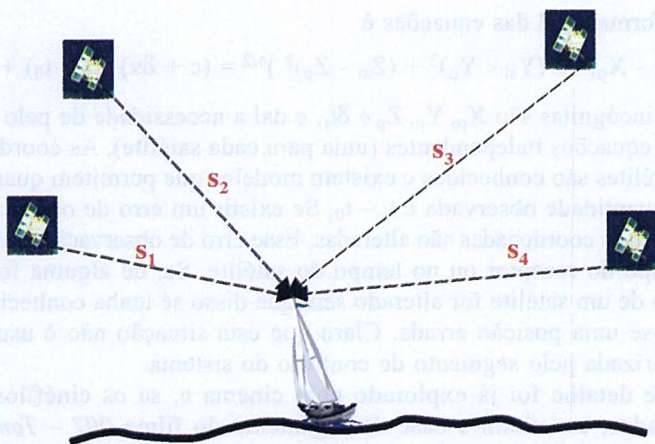


Fig. 13 — Posicionamento e navegação com GPS

O posicionamento baseado na recepção e processamento da informação contida nos códigos necessita da observação simultânea de quatro satélites (figura 13). As equações (Bastos, L., 1986) que permitem o cálculo da posição do receptor têm a seguinte forma

$$S_i = ((X_{si} - X_p)^2 + (Y_{si} - Y_p)^2 + (Z_{si} - Z_p)^2)^{1/2},$$

em que

X_p, Y_p, Z_p — coordenadas cartesianas tridimensionais do centro de fase da antena;

X_{si}, Y_{si}, Z_{si} — coordenadas cartesianas tridimensionais do satélite i ;

s_i — distância verdadeira entre o centro de fase da antena e o satélite i , dada por $s_i = d_i + v \delta t_r$;

d_i — distância observada entre o centro de fase da antena e o satélite i , dada por $d_i = v (t_i - t_0)$;

v — velocidade de propagação do sinal dada por $v = c + \delta v$;

δt_r — erro de sincronização dos relógios do receptor (t_r) e dos satélites (t_s) dado por $\delta t_r = t_s - t_r$ (supõe-se que os relógios de cada satélite estão sincronizados entre si);

t_i — instante em que o sinal do satélite i é recebido;

t_0 — instante em que o sinal é emitido;

c — velocidade de propagação da luz no vácuo;

δv — atraso na propagação do sinal.

A forma final das equações é

$$((X_{si} - X_p)^2 + (Y_{si} - Y_p)^2 + (Z_{si} - Z_p)^2)^{1/2} = (c + \delta v) ((t_i - t_0) + \delta t_r).$$

As incógnitas são X_p, Y_p, Z_p e δt_r , e daí a necessidade de pelo menos quatro equações independentes (uma para cada satélite). As coordenadas dos satélites são conhecidas e existem modelos que permitem quantificar δv ; a quantidade observada é $t_i - t_0$. Se existir um erro de observação, é claro que as coordenadas são alteradas. Esse erro de observação pode estar no tempo do receptor ou no tempo do satélite. Se, de alguma forma, o relógio de um satélite for alterado sem que disso se tenha conhecimento, obtém-se uma posição errada. Claro que esta situação não é usual, e é monitorizada pelo segmento de controlo do sistema.

Este detalhe foi já explorado pelo cinema e, se os cinéfilos estão recordados, constituiu a base do argumento do filme *007 – Tomorrow Never Dies*, onde uma intriga internacional era provocada pela introdução de um erro de tempo num dos satélites da constelação GPS, que originou a modificação das coordenadas de um barco. É então legítimo perguntar: *será que sabemos mesmo onde estamos?*

4.2 As trajectórias reais e a sua representação

Para a navegação é necessário conhecer não só a posição do objecto que se está a deslocar como também a trajectória a seguir. Já antes se viu

que a projecção de Mercator contribuiu para a representação das trajectórias marítimas com base na loxodrómia (linha de azimute constante, representada na carta por uma linha recta). Essa linha facilitava obviamente o cálculo, mas não representa a distância mais curta entre dois pontos, no caso de uma superfície esférica ou elipsóidica. A distância mais curta é a geodésica (ou ortodrómia), que se pode definir da seguinte forma: uma linha traçada sobre uma superfície é geodésica se, em cada ponto, a normal principal à linha for normal à superfície. A sua equação é (Torres, J., 1980)

$$r \operatorname{sen} \alpha = C,$$

em que r representa o raio do paralelo num ponto, α o azimute da geodésica nesse ponto e C uma constante própria de cada geodésica. Esta equação, aparentemente simples, é extremamente complicada de resolver sobre o elipsóide de revolução, e só se consegue fazê-lo através de processos de integração numérica. A sua resolução permite conhecer a distância mais curta entre dois pontos e a configuração da linha geodésica que une esses pontos. No caso da superfície esférica, a geodésica é um arco de círculo máximo, e a sua representação na projecção de Mercator não é uma linha recta, mas uma curva (figura 14).

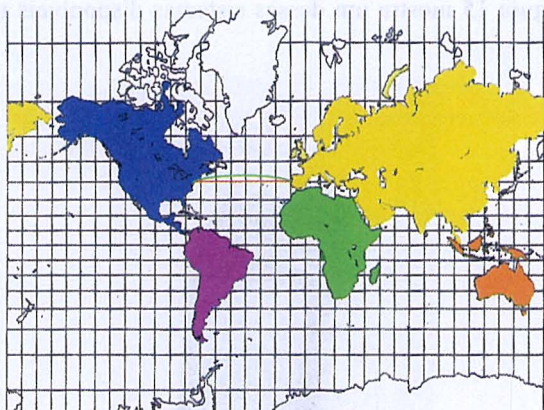


Fig. 14 — Traçado de uma geodésica e de um arco de paralelo (loxodrómia)

Na figura 14 estão traçadas a geodésica (arco) e o arco de paralelo (segmento de recta) entre Lisboa e um outro ponto situado à mesma latitude no continente americano, na projecção de Mercator; neste caso o

arco de paralelo é uma loxodromia, já que intersecta todos os meridianos segundo um ângulo de 90° . Aparentemente, a distância mais curta é a da linha recta. Não é esse o caso, pois se for calculada a distância entre Lisboa e Nova Iorque para cada uma das trajectórias obtém-se

Geodésica: 5 421 874,731 m;

Loxodromia: 5 544 339,170 m,

o que representa uma diferença de cerca de 120 km.

Quem já viajou em aviões que vão mostrando a trajectória seguida pôde constatar esta ilusão originada pela representação cartográfica, e não é raro ouvirem-se comentários a este propósito entre os passageiros. Pode fazer-se então a pergunta: *Será que as companhias aéreas estão a usar a projecção cartográfica mais adequada às finalidades?*

4.3 O admirável mundo novo

Os actuais sistemas de localização e navegação, já disponíveis em automóveis e a que brevemente teremos acesso através dos telemóveis, utilizam as tecnologias e os princípios da cartografia digital e dos GNSS (Global Navigation Satellite Systems), de que o GPS é um dos componentes. A figura 15 mostra um desses sistemas disponíveis no mercado.



Fig. 15 — Equipamento para posicionamento e navegação

Este é o admirável mundo novo em que vivemos, e para o qual a contribuição da geodesia e da cartografia tem sido e continuará a ser fundamental, embora em termos diferentes daqui em diante. Os requisitos dos utilizadores obrigam a uma mudança de paradigma sobre a forma como a informação é posta ao serviço das tecnologias, conduzindo à procura de novos modelos para a informação cartográfica e ao desenvolvimento de novos sistemas de navegação global, como é o caso do sistema Galileo, a ser implantado brevemente na Europa, e ao aperfeiçoamento dos existentes, no caso do GPS.

5. NOTAS FINAIS

Pretendeu-se dar, ao longo do texto e tendo em conta o tema que serviu de base à sua feitura (*Teias Matemáticas*), uma ideia da importância da matemática nos problemas relacionados com a forma, dimensões e representação da Terra, através de uma abordagem que pretendeu realçar os momentos mais marcantes e decisivos desta longa história, bem como a influência desta ciência nas várias etapas deste percurso notável.

Não se deixou também de focar os desenvolvimentos actuais e o que se pode esperar num futuro mais ou menos próximo no que respeita à informação espacial.

Apontou-se, com alguns exemplos, de que modo a nossa percepção do espaço pode ser condicionada pelos números e formulações matemáticas: a matemática não é inocente. Em suma, descobrimos algumas das teias com que a matemática enreda a realidade.

A propósito do tema, e para terminar, transcreve-se um texto a todos os títulos delicioso da obra *A Caça ao Snark*, (Carrol, L., 1986) da autoria de Lewis Carrol, que, embora matemático, deixou o seu nome para a posteridade por via dos seus dotes literários.

*Quanto ao sineiro, toda a gente o adorava!
Que figura, que graça e que classe!
E como aquela face solene brilhava!
Que saber espelhava aquela face!*

*Trouxera para bordo uma carta do mar
Onde de terra nem cheiro havia.
Era um papel facilimo de interpretar
E os homens pularam de alegria.*

«*Abaixo Mercator, Pólo Norte e Equador,
Eixos, meridianos, linhas tropicais.*»
*Bradou o capitão. E a tripulação:
«Simples símbolos são convencionais.»*

«*Mapas são o diabo! Tanta ilha e cabo!
Mas o nosso Sineiro, seja Deus louvado,
Com tal nada perfeito, faz-nos grande jeito!*»
Clamou o barco à uma, entusiasmado.

Agradecimentos

Agradeço ao Instituto Português de Cartografia e Cadastro (IPCC) a disponibilização da informação e aos colegas que contribuíram para a realização do trabalho e sua revisão final: José Barreiro Guedes, Teresa Cardoso e Henrique Silva, engenheiros geógrafos, e António Pereira, operador de sistemas.

Um agradecimento é também devido à Maria de Deus, professora de Matemática, pelo apoio intelectual e moral e pela paciência com que aguentou as noites e fins-de-semana que foram precisos para levar esta nau a bom porto.

Referências

- Afonso, J. (1972) – *Cartografia Matemática*, Caderno Técnico n.º 31 do Instituto Geográfico e Cadastral, Lisboa.
- Bastos, L. (1986) – *O Sistema Global de Posicionamento*, Revista n.º 6 do Instituto Geográfico e Cadastral, IGC, Lisboa.
- Black, J. (1997) – *Maps and Politics*, Reaktion Books, Londres.
- Carrol, L. (1986) – *A Caça ao Snark*, Edições Afrontamento, Lisboa.
- Fonseca, A., Fernandes, J., Bica, V. (2000) – *Carta Imagem de Portugal à Escala 1:100 000, Relatório 2: Geração do Modelo Numérico Altimétrico*, DT 13:00 do Instituto Português de Cartografia e Cadastro, Lisboa.
- Guedes, J. (2000) – *Notas pessoais*.
- IPCC (2000) – *Elementos de trabalho do Instituto Português de Cartografia e Cadastro*, Lisboa.
- Jones, T. (1967) – *The Figure of the Earth*, Coronado Press, Lawrence.
- Kakkuri, J., Kukkamaki, T., Levallois, J., Moritz, H. (1986) – *Le 250^e Anniversaire de la Mesure de l'Arc du Meridien en Laponie*, Publicação n.º 103 do Finnish Geodetic Institute, Helsínquia.

- Korte, G. (1997) – *The GIS Book, 4th Edition*, OnWord Press, Santa Fé.
- Levallois, J. (1988) – *Mesurer la Terre*, Presse de l'École Nationale des Points et Chaussées, Paris.
- Molenaar, M. (1998) – *An Introduction to the Theory of Spatial Object Modelling for GIS*, Taylor & Francis, Londres.
- Torge, W. (1991) – *Geodesy*, Walter de Gruyter, Berlim, Nova Iorque.
- Torres, J. (1980) – *Grande Geodésica*, Caderno Técnico n.º 36 do Instituto Geográfico e Cadastral, Lisboa.
- Torres, J., Lima, J. (1996) – *O Sistema Global de Posicionamento (GPS)*, Ingenium (Revista da Ordem dos Engenheiros) n.º 12, II série, Lisboa.

Outras fontes

- Boletins informativos da American Airlines e da Iberia
<http://www.007.com>
<http://igsbc.jpl.nasa.gov/index.html>

(Página deixada propositadamente em branco)

Boris Gutkin, David Pinto
Unité de Neurosciences Integratives et Computationelles,
Institute Alfred Fessard, França

e Bard Ermentrout
Universidade de Pittsburgh, EUA

Matemática da neurociência: de neurónios a circuitos e de circuitos a sistemas¹

1. INTRODUÇÃO

Desde há muitos anos que a matemática e os métodos computacionais têm vindo a desempenhar um papel importante no nosso conhecimento do sistema nervoso. O objectivo deste capítulo consiste em apresentar alguns exemplos que mostram a forma como certas técnicas matemáticas podem ser aplicadas, a vários níveis, para aumentar o nosso conhecimento sobre os sistemas neuronais. Começaremos com uma descrição dos princípios biofísicos subjacentes à formação do potencial de acção em neurónios individuais. A ideia fundamental da modelação é representar as propriedades eléctricas de membranas biológicas através de um circuito equivalente formado por condensadores e resistências ligados em paralelo. Utilizaremos depois métodos do domínio das equações diferenciais para estudar um modelo simplificado de um neurónio individual e mostrar como se pode reduzir a dinâmica no plano a um sistema dinâmico escalar

¹ Tradução de Mário da Silva Rosa, professor do Departamento de Matemática, Universidade de Coimbra. Revisão de F. Geraldes, professor do Departamento de Bioquímica, Universidade de Coimbra, e E. Quinta Ferreira, professora do Departamento de Física, Universidade de Coimbra.

num círculo. Utilizam-se simulações do modelo reduzido para explicar a estatística da formação de pontas (*spikes*) de neurónios individuais accionados por estímulos ruidosos. Seguidamente, analisaremos circuitos neuronais simples, envolvendo redes de neurónios excitatórios e inibitórios. Uma aproximação de campo médio reduz estas redes a um sistema planar e a análise do plano de fases permite-nos explicar resultados experimentais do sistema somato-sensorial (tacto) do rato. Finalmente, examinaremos redes alargadas, organizadas espacialmente. Aplicaremos a teoria da bifurcação a estas redes e, com base nos resultados, explicaremos os padrões observados durante alucinações visuais.

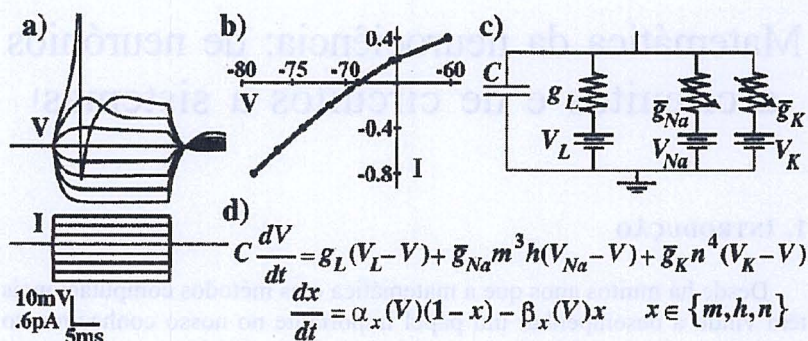


Fig. 1 — Modelo de circuito equivalente de actividade neuronal a) Alterações do potencial de membrana neuronal (V) em resposta a diferentes níveis de corrente injectada (I). b) Relação I - V de respostas neuronais. c) Diagrama do circuito que representa correntes capacitivas e resistivas, activas e passivas. d) Equações de Hodgkin Huxley do equilíbrio de corrente e da cinética dos canais com «portas» (gating kinetics)

2. MODELOS DO CIRCUITO EQUIVALENTE

Para representar a dinâmica da actividade eléctrica observada em neurónios individuais tornou-se corrente a utilização do modelo de circuito equivalente. Este modelo é baseado na ideia de que a actividade neuronal pode ser completamente descrita pelo fluxo de diferentes correntes associadas à membrana neuronal. As correntes dividem-se em dois tipos: aquelas que podem ser representadas por elementos lineares do circuito (correntes passivas) e as que são dependentes do potencial do tempo e

requerem uma dinâmica mais complexa (correntes activas). Ambos os conjuntos podem ser entendidos em termos das observações experimentais apresentadas na figura 1a) (ver o trabalho de revisão de Johnston e Wu, 1997).

2.1 Correntes passivas

A primeira observação é que os neurónios mantêm uma diferença de potencial constante através das suas membranas, o que se designou por «potencial de repouso», ($V_m \approx -65mV$). Este potencial deve-se: 1) ao facto de as membranas neuronais serem semipermeáveis, permitindo que apenas certos iões passem de um lado para o outro (maioritariamente sódio, Na^+ , potássio, K^+ , e cálcio, Ca^{++}); 2) ao facto de os neurónios manterem activamente um gradiente de concentração daqueles mesmos iões, através da membrana; 3) ao facto de os iões em causa possuírem uma carga eléctrica. Por exemplo, o potássio, K^+ , tem uma concentração maior no interior da célula e uma carga positiva. Forças de difusão tendem a fazer sair K^+ da célula. A perda subsequente de iões positivos conduz ao aparecimento de uma carga negativa no interior da membrana. A força eléctrica resultante atrai iões positivos, incluindo os ligados a K^+ , novamente para dentro da célula. O potencial de repouso, também designado por potencial de equilíbrio, é o nível de potencial para o qual as forças eléctricas e de difusão, devidas a todos os iões permeáveis, estão em equilíbrio. O elemento de circuito equivalente, utilizado para representar este potencial, é uma força electromotriz (FEM) ou bateria.

A segunda observação é que a injeção directa de corrente num neurónio induz uma variação incremental no potencial. Para uma certa gama de intensidades de corrente, a variação induzida no potencial é linear (figura 1b), o que, de acordo com a lei de Ohm ($V = IR$), sugere que a membrana está, em parte, a funcionar como uma resistência linear (as não linearidades na relação serão analisadas posteriormente). Experimentalmente, a conductância da membrana é medida a partir do declive da linha obtida através da representação gráfica da variação do potencial induzida por intensidades diferentes de corrente injectada. Ainda que muitos iões diferentes contribuam para a alteração do potencial, utiliza-se frequentemente uma única conductância de fuga (*leak conductance*), g_L , para contemplar todas as correntes iónicas passivas.

A terceira observação é que as variações do potencial, devido à injeção de corrente, se comportam exponencialmente no tempo. Isto sugere que as membranas neuronais operam como um circuito-RC, isto é, uma resistência linear em paralelo com um condensador.

As membranas neuronais são compostas por uma dupla camada de moléculas lipídicas na qual estão imersas proteínas. As proteínas possuem resistividades específicas baixas ($\sim 1\Omega\text{cm}$) e formam os canais proteicos através dos quais o fluxo de iões carregados atravessa a membrana. Por outro lado, os lípidos têm resistividades específicas elevadas ($\sim 10^9\Omega\text{cm}$), formando um isolamento eficiente entre as soluções celulares interna e externa, altamente electrolíticas, o que constitui exactamente o arranjo característico de um condensador de placas paralelas. A corrente que é injectada na membrana divide-se numa corrente capacitiva que carrega a bicamada lipídica e numa corrente iónica que passa através dos canais proteicos. Para um circuito-RC, a constante tempo da variação do potencial é $\tau = RC$. Portanto, experimentalmente, a capacidade de um neurónio é obtida como $C = \tau/R$, onde τ é a constante de tempo observada correspondente à taxa de variação exponencial do potencial de membrana e R é a resistência medida do modo acima descrito.

2.2 Correntes activas

A maior parte das correntes observadas exibem uma condutância constante apenas numa certa gama de potenciais. De facto, é de esperar alguma não linearidade na relação I-V, mesmo no caso de correntes totalmente passivas. Por exemplo, quando o potencial é tal que ambas as forças, de difusão e eléctrica, produzem movimentos de iões na mesma direcção, a condutância observada pode ser maior do que aquela que uma simples relação linear faria prever. Em muitos casos, contudo, a não linearidade, ou *rectificação*, numa condutância de canal não pode ser explicada apenas por propriedades passivas. Além disso, observa-se frequentemente que a condutância depende não só do potencial como também do tempo (ver figura 1a).

A fim de explicar a *rectificação anómala* observada em muitas correntes neuronais, Hodgkin e Huxley (1952) propuseram o modelo de condutância do canal com portas (*gating model*). A ideia é que alguns canais de proteínas, que permitem a passagem de iões através da membrana, possuem também partículas carregadas que abrem ou fecham o canal (*gating particles*) de uma forma dependente do potencial e do tempo, alterando assim a condutância. Por exemplo, a condutância de correntes de K^+ *activas* é representada por uma condutância máxima g_K vezes n^4 , onde n é a probabilidade de que uma partícula-porta (*gating particle*) esteja na posição que contribui para a abertura do canal e o expoente 4 indica o número destas partículas por canal.

A dinâmica das partículas-porta (*gating particles*) é modelada usando cinéticas de primeira ordem, como se mostra na figura 1d, onde $\alpha(V)$ é a taxa, dependente do potencial, à qual os canais abertos se fecham e $\beta(V)$ é a velocidade à qual os canais fechados se abrem. No caso de correntes de Na^+ activas, as experiências sugerem que há dois tipos de partículas-porta independentes envolvidas, três de activação, m , e uma de inactivação, h . A dependência do potencial e do tempo de cada «porta» podem determinar-se experimentalmente examinando as alterações de condutância associadas a cada corrente individual. O elemento de circuito equivalente utilizado para representar a condutância neuronal activa é uma resistência variável.

2.3 Equações de equilíbrio de corrente

Consideradas conjuntamente, as correntes passiva e activa acima descritas caracterizam o circuito equivalente ou modelo de condutância em paralelo de actividade neuronal, mostrado na figura 1c. A equação que descreve a dinâmica do circuito é uma consequência da primeira lei de Kirchoff: é nula a soma de todas as correntes que convergem num nodo. Em particular, uma corrente aplicada divide-se numa corrente capacitiva que carrega a membrana e em correntes resistivas que passam através dos canais activos e passivos. Normalmente, a equação é escrita como se indica na figura 1d, em que a corrente capacitiva ($C \cdot dV/dt$) é igual à soma das várias correntes iónicas.

O grande mérito do modelo de Hodgkin e Huxley foi a sua capacidade de explicar muito mais do que a sensibilidade das várias condutâncias neuronais ao potencial e ao tempo. Uma vez correctamente caracterizada cada uma das correntes, constatou-se que os mesmos mecanismos não lineares das «portas» eram suficientes para explicar a formação de potenciais de acção, isto é, pontas (*spikes*) de potencial «tudo ou nada» explosivos que dão início à sinalização sináptica entre neurónios. Utilizando o mesmo conjunto de equações de equilíbrio de corrente, Hodgkin e Huxley demonstraram como os potenciais de acção resultam duma sequência estereotipada de variações de condutâncias entre as várias correntes iónicas.

Além de Na^+ e K^+ , foram também descritas outras condutâncias que dependem do potencial e do tempo. Considera-se que os canais de Ca^{++} estão na base da formação de trens de potenciais de acção que caracterizam a actividade nalguns neurónios do tálamo e do córtex cerebral (McCormick e Huguenard, 1992). O decréscimo no tempo da taxa de

formação de potenciais de acção (*spike adaptation*) depende de canais de K^+ que são sensíveis não apenas ao potencial, mas também à concentração de Ca^{++} no interior do neurónio (Lancaster e Adams, 1986). Os canais de Na^+ que funcionam numa escala de tempo mais lenta e não têm porta de inactivação podem ser responsáveis pelo aumento de sinais que chegam de outros neurónios (Stafstrom *et al.*, 1982). Cada uma destas correntes pode ser incorporada no modelo de equilíbrio de corrente, com a adição de condutâncias e cinéticas apropriadas, ao conjunto de equações.

2.4 Controlo dinâmico (*Dynamic Clamp*)

Uma aplicação recente e muito interessante das equações de equilíbrio de corrente está relacionada com a modificação da função de neurónios reais. Isto é feito com modelos computacionais em tempo real e o método designa-se por «dynamic clamp» ou «modelo híbrido» (Sharp *et al.*, 1993). A ideia subjacente é ligar neurónios reais a correntes iónicas simuladas, ou mesmo a membranas completamente simuladas. Utiliza-se um eléctrodo para medir o potencial de membrana de um neurónio real. Esta medida é usada para calcular correntes iónicas simuladas. Estas são depois novamente injectadas no neurónio através do mesmo eléctrodo. As aplicações mais comuns da técnica consistem em «inserir» novos canais eléctricos na membrana de um neurónio ou em «subtrair» um canal existente e estudar o respectivo efeito no comportamento do neurónio. Inicialmente o método foi aplicado ao estudo dos neurónios no sistema digestivo da lagosta (Sharp *et al.* 1993), enquanto em estudos mais recentes se tem examinado o comportamento de neurónios individuais no cérebro de mamíferos (Jaeger e Brower, 1999). Uma outra aplicação que está actualmente a ser desenvolvida consiste em utilizar o método do modelo híbrido na construção de circuitos de neurónios simulados, modelados em «microchips» analógicos programáveis (Le Masson *et al.*, 1995). Colaborações recentes têm explorado esta técnica para estudar como controlam os circuitos neuronais os ritmos cerebrais durante o sono (Le Masson, comunicação pessoal).

3. PLANOS DE FASES E «SPIKING»

O modelo de equilíbrio de corrente de Hodgkin e Huxley acima descrito é um sistema dinâmico a quatro dimensões, o que torna difícil a sua análise matemática rigorosa. Morris e Lecar (ver Rinzel e Ermentrout,

1998) construíram um modelo de neurónio muito simples baseado apenas em três condutâncias: um canal de cálcio rápido, um canal de potássio lento e um canal de fuga passivo. Com base na formulação do circuito equivalente acima descrita, as equações a considerar são:

$$C \frac{dV}{dt} = I + g_L(V_L - V) + g_{Ca}m_\infty(V) (V_{Ca} - V) + g_K\omega(V_K - V)$$

$$\frac{d\omega}{dt} = \frac{\omega_\infty(V) - \omega}{\tau_\omega(V)}$$

Sendo suficientemente rápidas, pode considerar-se que as cinéticas das «portas» do canal de cálcio, $m_\infty(V)$, actuam instantaneamente. Assim, o sistema requer apenas duas variáveis dinâmicas e pode ser analisado, de uma forma rigorosa e completa, através de métodos associados ao plano de fases.

Estes métodos são dos mais úteis que surgiram na neurociência teórica oriundos da teoria qualitativa de equações diferenciais. O objectivo é compreender a dinâmica de duas variáveis, representando graficamente uma como uma função da outra. A designação de plano de fases deriva do facto de, através da representação gráfica de duas variáveis dinâmicas, em particular das que têm alguma espécie de comportamento periódico ou estereotipado, podermos estudar a «fase» relativa de uma das variáveis face à outra. Por exemplo, consideremos um neurónio formando um potencial de acção. O potencial de membrana afasta-se de um nível de repouso até atingir o nível máximo do potencial, voltando a descer através de um processo de queda ou *pós-hiperpolarização*, regressando então à posição de repouso. A variável de recuperação, ω , segue uma trajectória correspondente em resposta ao potencial de membrana. Se representarmos num gráfico o potencial face à variável de recuperação ω (ver figura 2a e b), o resultado é uma curva fechada, ou *trajectória*, no plano de fases.

O plano de fases é caracterizado por várias marcas ou pontos especiais que ajudam a definir e a visualizar as dinâmicas. Por exemplo, as linhas de inclinação nula do sistema são definidas como as curvas ao longo das quais uma ou a outra variável permanece constante. A curva ao longo da qual $dV/dt = 0$, define a linha de inclinação nula V , enquanto $d\omega/dt = 0$ define a linha de inclinação nula ω . Nos pontos onde aquelas duas linhas se intersectam, nenhuma das variáveis se altera; os pontos de intersecção definem os estados de equilíbrio do sistema. Um tal ponto, marcado por R, corresponde ao potencial de repouso do neurónio. Trata-se de um ponto fixo atractivo: pequenas perturbações transitórias no potencial de

membrana, por intermédio, por exemplo, da aplicação de uma corrente externa ou por activação de conductâncias que dependem do potencial, serão amortecidas com o tempo e o potencial voltará ao ponto de repouso. Um outro ponto do estado de equilíbrio é o assinalado com T – potencial de limiar do neurónio. Este é um ponto fixo repulsivo, pelo que as trajectórias que se aproximem de T retornam à posição de repouso ou continuam, após terem produzido um potencial de acção.

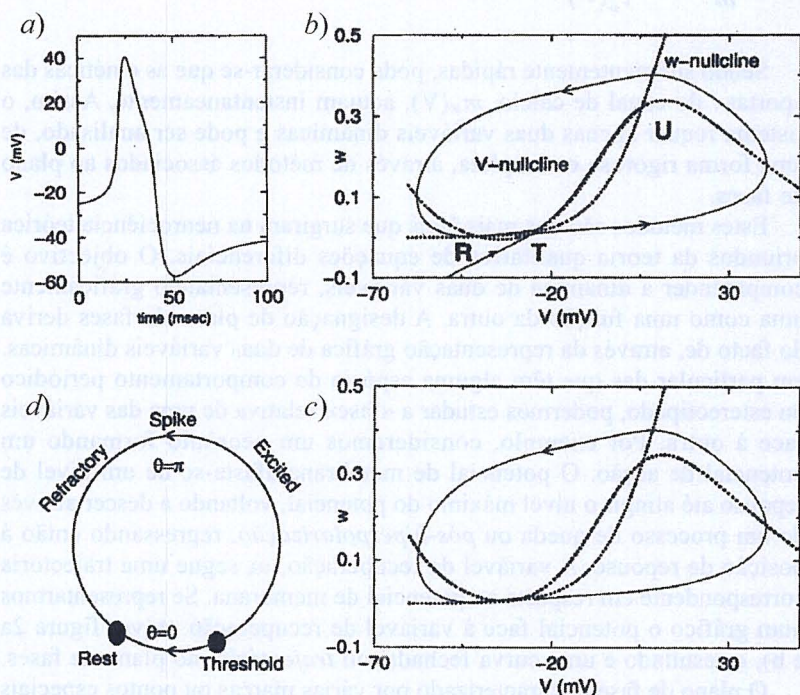


Fig. 2 — a) Um potencial de acção, ou ponta, no potencial de membrana representado graficamente no tempo; b) Diagrama de fases do mesmo potencial de acção. Nele estão representadas as linhas de inclinação nula V e w . As suas intersecções correspondem a pontos fixos: R é o ponto atractivo, T é o limiar que corresponde a um ponto-sela (atractivo numa direcção e repulsivo noutras direcções), U é um ponto repulsivo; c) Após o aumento da corrente injectada, a linha de inclinação nula V cresce e não existem pontos de repouso estáveis, mas sim um ciclo-limite atractivo correspondente ao estado de explosão; d) Diagramas de estado equivalentes para o neurónio-teta. A cheio marcam-se os pontos críticos do modelo — R e T. O modelo inclui também o potencial de acção e o comportamento refractário do modelo completo

O plano de fases permite visualizar o efeito da injeção de correntes num neurónio. Matematicamente, injectar uma corrente estável positiva corresponde a adicionar uma constante positiva ao segundo membro da equação do potencial. No plano de fases, o potencial da linha de inclinação nula V é aumentada, o que altera a posição dos pontos fixos. Quando o potencial daquela linha cresce suficientemente, os estados estacionários R e T convergem e depois desaparecem. Isto revela uma solução periódica ou solução de *ciclo-limite* no plano de fases correspondente à formação repetitiva de potenciais de acção, ou disparo, pelo neurónio. À semelhança dos estados de equilíbrio estável, os ciclos-limite podem ser atractivos ou repulsivos. Apenas as oscilações estáveis podem ser realizadas fisiologicamente (excepto quando se aplicam controlos especiais), mas as soluções periódicas instáveis podem actuar como separatrizes, dividindo o plano de fases em bases de atracção para múltiplos atractores. Neste modelo particular, notemos que as linhas de inclinação nula podem estar arbitrariamente próximas uma da outra (figura 2c), pelo que a frequência das oscilações pode ser arbitrariamente baixa. O aumento de corrente aplicada conduz a frequências mais elevadas na medida em que o «canal» entre as linhas de inclinação nula se alarga.

Um pormenor importante e um pouco surpreendente é que a trajectória do ciclo-limite permanece constante na sua forma, independentemente de quão rápidas são as explosões da célula (ou atravessa a curva fechada). Do ponto de vista experimental, isto corresponde à forma invariante dos potenciais de acção gerados. Em termos matemáticos, permite-nos considerar aquela linha fechada como uma estrutura invariante que não se altera com a dinâmica dos sistemas. Por este motivo, se conseguirmos um processo para marcar apenas as quantidades que se alteram, deveremos ser capazes de simplificar o modelo.

Formalmente, isto é feito através do método de formas normais da teoria da bifurcação. Descreveremos apenas as ideias heurísticas que lhe estão subjacentes.

Como os dois pontos fixos R e T convergem e desaparecem e o ciclo-limite emerge, o sistema aproxima-se de um ponto de bifurcação nas soluções dinâmicas. Quando examinamos os valores próprios do modelo linearizado próximo desta bifurcação constatamos que um único valor próprio real vai para 0. Isto significa que a bifurcação é a de um *nodo-sela num círculo invariante* («saddle-node»), o que fornece a base matemática para tratar a trajectória circular como um invariante. Para encontrar o «centro» matemático desse círculo, descrevemos a variação do potencial à sua volta, utilizando uma variável angular θ . Consideramos este ângulo igual a zero ($\theta = 0$) quando o nível do potencial é o de

repouso, no ponto R, e consideramos que θ cresce à volta do círculo na direcção contrária à do movimento dos ponteiros do relógio. O potencial de acção é então uma revolução do ângulo relativo ω/v ou «fase» de 0 até π , voltando a 2π . O artifício está em encontrar uma transformação matemática das variáveis do modelo original do circuito equivalente, tal que uma das novas variáveis percorra o ciclo-limite invariante e as restantes variáveis sejam ortogonais àquela. Para um sistema dinâmico com uma bifurcação nodo-sela, uma tal transformação pode sempre ser encontrada e conduz à equação:

$$\frac{dx}{dt} = qx^2 + pI, \quad (1)$$

onde a variável x descreve a dinâmica, p e I são parâmetros que incluem, por exemplo, a influência de «inputs» externos, e $q > 0$ é um parâmetro que depende dos detalhes do modelo original completo. Esta equação é um tanto incómoda, uma vez que a sua solução se torna infinita em tempo finito. Tratamos a singularidade envolvendo x num círculo utilizando a mudança de variáveis $x = \tan \theta/2$, chegando assim matematicamente à variável de fases acima descrita. A equação final para o modelo em θ vem:

$$\frac{d\theta}{dt} = q(1 - \cos\theta) + (1 + \cos\theta)pI. \quad (2)$$

Este modelo é muito simples; contudo, reproduz o comportamento do modelo mais complexo do circuito equivalente. De facto, o neurónio-teta é um modelo canónico para uma vasta classe de modelos baseados em condutâncias que exibem dinâmicas de bifurcação com nodo-sela. Mais especificamente, qualquer modelo de um neurónio baseado em condutâncias que exiba uma bifurcação nodo-sela pode ser descrito pelo neurónio-teta. Por isso, se conseguirmos compreender o comportamento do neurónio-teta, saberemos mais acerca da função de muitos tipos diferentes de neurónios. Por exemplo, eis algumas das propriedades do neurónio-teta que reflectem o comportamento geral de neurónios piramidais encontrados no neocortex:

- Os potenciais de acção são acontecimentos do tipo tudo ou nada, ou seja, quando estimulada, a actividade abaixo do limiar T regressa ao repouso. Mas, uma vez passado o ponto T, ocorre uma ponta (*spike*) completa.
- O modelo e os neurónios geram sequências contínuas de potenciais de acção em resposta a injeções de correntes constantes.
- Surgem explosões repetitivas com frequências arbitrariamente baixas.

- A curva (IF) da relação corrente de entrada/frequência de saída pode ser facilmente ajustada por uma raiz quadrada (IF instantânea) ou por uma função linear (IF estacionária), ambas observadas em neurónios reais.

Todas estas propriedades são básicas e nos mostram que o modelo reduzido contém «informação» dinâmica sobre os neurónios reais. A seguir utiliza-se este modelo reduzido para explorar a dinâmica subjacente a comportamentos mais complexos.

4. ESTATÍSTICA DE ACTIVIDADE NEURONAL DO CÓRTEX «IN VIVO»

Um debate de longa data em neurociência prende-se com a questão de saber se os neurónios codificam a informação acerca do mundo em termos da frequência média de formação de potenciais de acção ou através da distribuição precisa no tempo de potenciais de acção individuais. A segunda hipótese tem sido considerada improvável, uma vez que o disparo de neurónios no cérebro vivo é muito irregular (ver, por exemplo, Shadlen e Newsome, 1998). Por outro lado, várias experiências mostraram que os neurónios são capazes de produzir de um modo fidedigno potenciais de acção definidos no tempo. Numa experiência, em particular, foi demonstrado que injectando níveis constantes de corrente os neurónios formam potenciais de acção a uma velocidade contida num certo conjunto de valores, mas com distribuição temporal irregular. Contudo, após repetidas injeções de corrente com ruído, os tempos particulares de disparo foram muito mais precisos relativamente ao sinal de entrada (Mainen e Sejnowski, 1995). Os resultados do modelo demonstram que o neurónio-teta apresenta respostas semelhantes para sinais de entrada constantes *versus* sinais de entrada com ruídos (figura 3) (Gutkin e Ermentrout, 1998). Examinaremos aqui o modo como o modelo permite captar e explicar estas conclusões não intuitivas.

O neurónio-teta com potencial de repouso com ruído mas com corrente de entrada constante é basicamente um oscilador intrínseco perturbado por um efeito de potencial aleatório ou, em linguagem matemática, por um processo de renovação não-linear. Assim, o tempo de partida de cada ciclo sucessivo de potenciais de acção depende do tempo do potencial de acção anterior. Uma vez que estes tempos são perturbados pela variação do «potencial», a incerteza aumenta em cada potencial de acção sucessivo. De facto, pode mostrar-se em termos analíticos (Gutkin e Ermentrout, 1998) e numéricos (figura 3) que esta incerteza, medida como a variância

de tempos de potenciais de acção em cada experiência, é representada pela raiz quadrada do número de potenciais de acção multiplicado pela variância do tempo do primeiro potencial de acção.

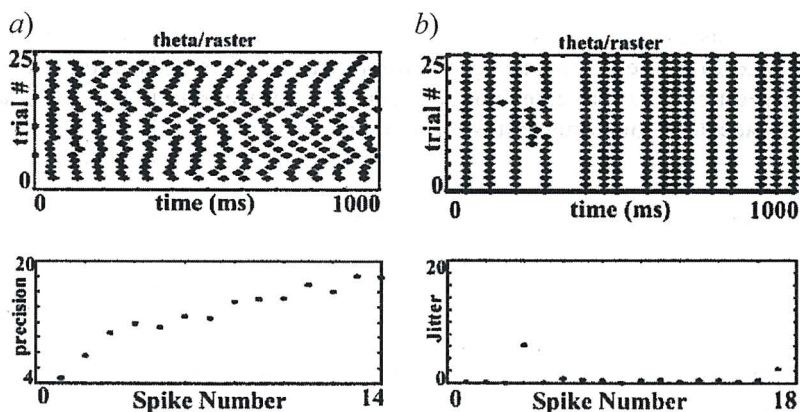


Fig. 3 — a) Resposta do modelo de neurónio-teta (com potencial de repouso com ruído) quando estimulado por uma corrente constante. O gráfico mostra os tempos de formação de potenciais de acção para ensaios repetidos com o mesmo estímulo constante. Os primeiros potenciais de acção aparecem com a mesma distribuição enquanto os últimos se deslocam aleatoriamente. Em baixo mostra-se o gráfico da incerteza, nos tempos dos potenciais de acção, calculados como o desvio padrão dos tempos obtidos nos ensaios. Como se pode constatar, a incerteza aumenta em cada potencial de acção sucessivo. b) Resposta do mesmo modelo de neurónio-teta quando estimulado repetidamente por uma corrente perturbada (variada). Os potenciais de acção permanecem consistentes durante cada ensaio (gráfico superior) e a incerteza mantém-se baixa de potencial de acção para potencial de acção (gráfico inferior)

Por outro lado, quando o mesmo neurónio-teta é estimulado por correntes repetidas com ruído, os tempos dos potenciais de acção não dependem uns dos outros, mas da dinâmica dos estímulos. Se o estímulo for de tal forma que conduza rapidamente o potencial para além de T, então forma-se um potencial de acção para o mesmo ponto de entrada em todos os ensaios. Desta forma, a incerteza é baixa e não aumenta em cada potencial de acção.

Isto significa que, para que os neurónios codifiquem informação com o tempo de formação do potencial de acção preciso, deverão fazê-lo em resposta a entradas de corrente do tipo choque. Ao mesmo tempo, entradas lentas serão codificadas a uma velocidade de disparo neuronal média.

A análise em curso do neurónio-teta sugere que estes dois esquemas de codificação poderão não ser mutuamente exclusivos, mas que ambos os modos podem contribuir para a saída de potenciais de acção de um neurónio individual (Gutkin e Ermentrout, em preparação).

5. MODELOS DE ACTIVIDADE

As secções anteriores mostraram como reduzir as equações de equilíbrio de corrente a um simples modelo escalar para compreender a dinâmica de neurónios individuais. Em muitos casos, relevantes a nível experimental, é desejável ir mais além na simplificação da dinâmica, utilizando os chamados modelos de actividade ou modelos de velocidade de disparo. Nestes modelos, a quantidade relevante não é o tempo de formação do potencial de acção ou o potencial de um neurónio individual, mas antes o nível de actividade generalizada ou velocidade de disparo em neurónios individuais ou em populações neuronais.

Em contraste com a representação biofísica fornecida pelas equações de equilíbrio de corrente, os modelos de actividade baseiam-se numa descrição mais funcional da actividade neuronal. Estes modelos são sobretudo utilizados para estudar interações entre vastas populações neuronais, ou, mais geralmente, nos casos em que um modelo biofísico se torna impraticável ou muito complexo. Nesta secção apresentaremos uma derivação funcional do modelo de actividade e descreveremos um exemplo em que é habitualmente utilizado para examinar o processamento de sinais de entrada sensoriais nos bigodes dos roedores.

5.1 Dedução das equações

A equação central dos modelos de actividade descreve a relação entre três medidas distintas de actividade neuronal: potencial, velocidade de disparo e impulso sináptico (descrito em baixo). A formação de um potencial de acção ou o disparo de um neurónio activa ligações sinápticas transmitidas a outros neurónios. As sinapses activadas induzem alterações de potencial ou potenciais pós-sinápticos (PPS) nos neurónios receptores. Os PPS são muitas vezes representados como seguindo um curso temporal característico, descrito através de uma função «alfa», $\alpha(t)$ (Rall, 1998). Os PPS podem ser positivos ou negativos, dependendo de a sinapse ser excitatória (EPPS) ou inibitória (IPPS). Para uma série de potenciais de

acção, as funções alfa somam-se e o potencial, V , no neurónio pós-sináptico é dado por

$$V(t) = \sum_i \alpha(t - t_i),$$

em que i corresponde ao número de potenciais de acção e t_i é o tempo de chegada para cada um deles. Numa população de neurónios, a distribuição de potenciais de acção no tempo é descrita frequentemente através de uma função contínua que representa a velocidade média de disparo dessa população, $F(t)$. Esta função depende do nível médio do potencial, pelo que $F(t) = F(V(t))$. Se separar a população em neurónios excitatórios *versus* neurónios inibitórios, a equação para o potencial excitatório médio é dada por

$$V_e(t) = ee \int_{-\infty}^t \alpha_e(t-t') F_e(V_e(t')) dt' - ie \int_{-\infty}^t \alpha_i(t-t') F_i(V_i(t')) dt'. \quad (3)$$

Nesta equação as duas populações têm distintas funções α , F de taxas de disparo e ie (por exemplo) representa a eficácia relativa de sinapses da população inibitória para a excitatória. O potencial médio da população inibitória é definido de modo semelhante.

Pode exprimir-se uma formulação alternativa do modelo de actividade, se definirmos o *impulso sináptico* de cada população por

$$S_e = \int_{-\infty}^t \alpha_e(t-t') F_e(V_e(t')) dt'$$

$$S_i = \int_{-\infty}^t \alpha_i(t-t') F_i(V_i(t')) dt'$$

pelo que, tomando novamente a população excitatória como exemplo,

$$S_e = \int_{-\infty}^t \alpha_e(t-t') F_e(eeS_e - ieS_i) dt'. \quad (4)$$

Além disso, se admitirmos que as funções alfa são descritas por simples funções exponenciais decrescentes (e.g. $\alpha_e = e^{-t/\tau_e}$), então podemos diferenciar para obter,

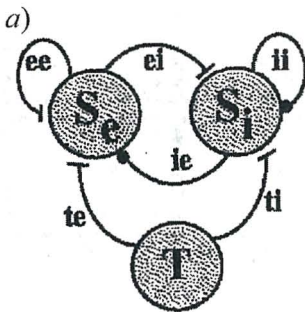
$$\tau_e \frac{dS_e}{dt} + S_e = F_e(eeS_e - ieS_i).$$

A equação (3) é designada formulação do potencial e a equação (4) é a formulação da actividade (Ermentrout, 1997). Intuitivamente, o termo *impulso sináptico* deriva das unidades de eeS_e , descrevendo a alteração de potencial induzida por sinapses excitatórias. Na maior parte dos modelos, ee mede-se em unidades de volts \times sinapse, a eficácia da ligação

vezes o número de sinapses. Assim, S_e mede-se na unidade $1/\text{sinapse}$, uma quantidade adimensional por sinapse, i.e., impulso sináptico (Pinto *et al.*, 1996). Como se descreve a seguir, a formulação da actividade permite, em particular, a incorporação directa de dados biológicos no sistema modelo.

5.2 Bigodes e barris

Ao deflectir-se um bigode na face de um roedor, activa-se uma cadeia de neurónios desde a periferia, passando pelo eixo do cérebro, o tálamo e seguindo até um feixe de neurónios definido anatomicamente no neocortex, normalmente designado por «barril de bigodes» (Woolsey e van der Loos, 1970). Cada barril contém duas populações principais de células, neurónios excitatórios e neurónios inibitórios. Ambas as populações recebem do tálamo sinais sinápticos excitatórios. Além disso, ambas as populações estabelecem ligações recíprocas com neurónios do outro tipo e contactos recorrentes a neurónios do mesmo tipo (figura 4a) (Simons, 1978). Os neurónios excitatórios no barril também formam contactos sinápticos com neurónios situados noutras regiões do cérebro (White, 1989). Portanto, utilizando quer o modelo de actividade quer os dados experimentais, o objectivo é compreender como é que interacções as sinápticas no interior de um barril operam nos sinais de entrada provenientes do tálamo e os transformam em sinais de saída que emanam da população de neurónios excitatórios.



$$\begin{aligned}
 b) \quad & \tau_e \frac{dS_e}{dt} + S_e = F_e (ee S_e - ie S_i + te T) \\
 & \tau_i \frac{dS_i}{dt} + S_i = F_i (ei S_e - ii S_i + ti T) \\
 & \tau_e \frac{dT}{dt} + T = \text{THAL}(t).
 \end{aligned}$$

Fig. 4 — Modelo de actividade do barril-cortex de um roedor. a) Diagrama esquemático de ligações sinápticas entre populações excitatórias (S_e) e inibitórias (S_i) do barril e do tálamo (T). As sinapses excitatórias são representadas por linhas e as sinapses inibitórias por pontos; b) Formulação da actividade do barril incorporando a entrada do tálamo

Experiências laboratoriais medem o número e o tempo de potenciais de acção gerados por neurónios do tálamo e do barril em resposta a uma deflecção. Geralmente, reúnem-se os dados de muitos neurónios e calcula-se a velocidade média de disparo para cada população. A utilidade do modelo de actividade deriva do facto de representar a actividade neuronal de uma forma compatível com estes dados experimentais. Uma vez que tanto os dados das populações de neurónios do tálamo como dos neurónios do barril são medidos em termos de velocidade média de disparo, os dados talâmicos podem ser utilizados directamente como sinais de entrada para o modelo (THAL(t)), tal como os dados da população de neurónios excitatórios do barril o podem ser para avaliar a saída excitatória do modelo (F_e , figura 4b).

Na prática, a forma precisa das funções de velocidade de disparo (F_e e F_i) é determinada experimentalmente através da análise das velocidades de disparo dos neurónios do tálamo e do barril em resposta a um conjunto de deflecções bem seleccionadas (Pinto *et al.*, 1996; Kyriazi, 1993). Quando as respostas simuladas estiverem de acordo com os dados experimentais, o modelo permitirá prever respostas a novos estímulos e compreender os mecanismos através dos quais aquelas respostas são geradas.

Uma previsão do modelo de actividade do barril é que a intensidade da resposta excitatória é sensível à distribuição dos sinais de entrada da população talâmica.

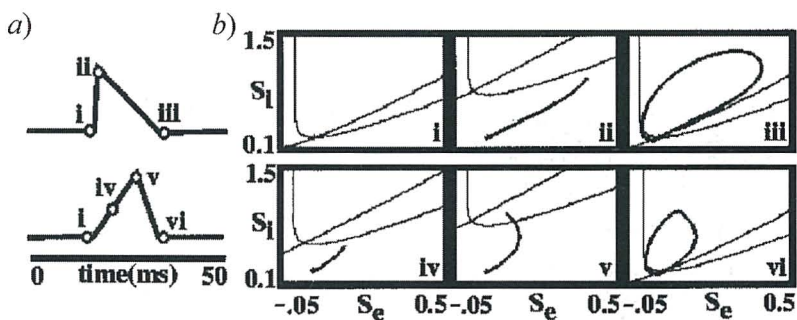


Fig. 5 — Análise do plano de fases de resposta simulada do barril a distribuições de entrada talâmicas. a) Duas distribuições de entrada talâmicas, diferindo no curso temporal, mas com a mesma actividade total; b) Representação no plano de fases de respostas da população do barril. Planos sucessivos mostram as respostas progressivas nos instantes de tempo indicados em numeração romana

Para compreender este facto, examinaremos o plano de fases obtido a partir do modelo, em resposta a entradas de duas populações talâmicas (Pinto *et al.*, 2001). No plano de fases, representa-se graficamente no eixo dos x e no eixo dos y o nível de actividade das populações excitatórias e o das inibitórias, respectivamente. Quando ocorre um sinal de entrada são traçados níveis de alteração da actividade pela trajectória de resposta nos sucessivos planos, nos instantes de tempo indicados pelos números romanos (figura 5a e 5b). A linha curva a fino (recta) é a linha de inclinação nula excitatória (inibitória), constituída por pontos nos quais $dS_e/dt = 0$ ($dS_i/dt = 0$).

Para o sinal de entrada talâmico de crescimento rápido, as linhas de inclinação nula sobem rapidamente (figura 5b, i a ii). Para o sinal de entrada de crescimento lento, as linhas de inclinação nula sobem mais lentamente. A trajectória-resposta permanece próxima da linha de inclinação nula onde as alterações na actividade são relativamente menores (figura 5b, i a iv a v). Assim, apesar de ambos os sinais de entrada representarem a mesma quantidade total de actividade talâmica (figura 5a), a distribuição de entrada de crescimento rápido gera uma resposta muito maior da rede de trajectórias do que a da distribuição de entrada de crescimento lento (figura 5b, iii *versus* vi). Esta previsão do modelo tem sido verificada experimentalmente através da análise da resposta de neurónios reais a diferentes distribuições de entrada talâmicas (Pinto *et al.*, 2000).

6. DAS PINTURAS RUPESTRES À ESTRUTURA DO CÓRTEX

Durante anos as pessoas têm sentido um grande fascínio pelas pinturas rupestres do Paleolítico, tal como as que foram encontradas nas famosas cavernas de Lascaux, em França, e nas paredes de arenito no Sudoeste da América. Estas gravuras descrevem muitas vezes animais e seres humanos. Contudo, há também desenhos mais abstractos, nomeadamente espirais, raios solares e mandalas (círculos). Uma das questões que os antropólogos se colocam é saber qual é o significado destes símbolos abstractos. Um certo número de investigadores sugeriu que estas pinturas foram desenhadas por «xamãs» («homens-medicina») durante estados de transe e que os desenhos representam imagens visuais que resultam de um estado alterado de consciência (Hedges, 1981; Lewis-Williams, 1984; Hudson e Lee, 1984). Em particular, alguns antropólogos estabelecem uma analogia entre as formas icónicas da arte rupestre e as configurações descritas em fases iniciais de alucinação (Kluver, 1996), bem como as

configurações que surgem em reacção a certos estímulos visuais, por exemplo, a pressão do globo ocular e a luz trémula (Tyler, 1978). Estas últimas configurações, chamadas fotópsias, e os seus análogos induzidos por estupefacientes foram classificadas por Kluver em quatro tipos diferentes, que designou «constantes de forma»:

- túneis e funis
- espirais
- reticulados, tais como favos de mel e tabuleiros de damas
- teias de aranha

Na figura 6 são apresentadas imagens representativas destas formas. Assim, de acordo com os antropólogos, os desenhos geométricos abstractos encontrados nas cavernas e nas pinturas em pedra são descrições de imagens vislumbradas por xamãs que estariam em estados mentais induzidos por transe ou por efeitos de «drogas». De facto, como iremos ver, a maioria dos desenhos mais comuns podem ser previstos através da análise da estrutura de configurações da actividade neuronal no córtex visual.

Vimos como se passa de modelos biofísicos pormenorizados para modelos reduzidos que permitem estudar circuitos sensoriais locais. Que acontece quando muitos desses circuitos são acoplados de um modo espacialmente estruturado? O córtex visual estriado é notável devido à sua organização, quase cristalina, de células que respondem à orientação, cor e posição no campo visual.

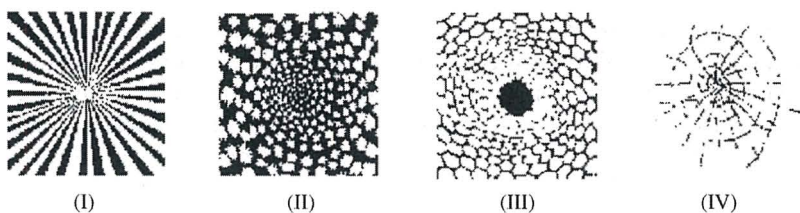


Fig. 6 — Constantes de forma (I) funil, (II) espiral, (III) reticulado, (IV) teia de aranha

Ou seja, o córtex visual está organizado em colunas (à semelhança dos barris) e no interior de cada coluna estão neurónios que respondem a (i) posições espaciais na retina; (ii) sinais de entrada dos olhos esquerdo e direito; (iii) linhas orientadas; outras modalidades visuais. As ligações

entre as células não são aleatórias. Por exemplo, as células numa coluna que respondem a linhas verticais estão mais fortemente ligadas a células da mesma orientação em colunas vizinhas. As consequências desta ligação altamente estruturada têm sido objecto de análise num vasto número de artigos. Ermentrout e Cowan (1979) mostraram que as partições simétricas espontâneas em redes corticais espacialmente organizadas poderão justificar as configurações grosseiras da actividade espacial registadas durante alucinações. Mais recentemente, Bressloff *et al.* (2000) consideraram o acoplamento entre células sensíveis à orientação para explicar outros tipos de configurações alucinatórias mais complexas. No final desta secção, recorreremos a estas noções para compreender o contexto em que se poderão inserir os desenhos das cavernas acima descritos.

Começemos por considerar uma arquitectura muito geral da rede neuronal descrita por:

$$\tau_i \frac{du_i}{dt} = -u_i + F_i \left(\sum_j W_{ij} u_j \right) \quad (5)$$

que representa as actividades de muitos neurónios acoplados. W_{ij} representa a intensidade das ligações entre neurónios, pelo que a equação (5) é uma generalização da equação (4). Esta rede pode ser separada em camadas. Suponhamos que eliminamos completamente a actividade de «fundo» («background») da rede, pelo que $u_i(t)$ representa o desvio do «fundo» e $F_i(0) = 0$. Para simplificar pomos $\tau_i = 1$. Poderemos perguntar se este estado de «fundo» é ou não estável. Para esse efeito, precisamos de linearizar em torno do estado de repouso $u = 0$. A estabilidade é determinada através dos valores próprios da matriz:

$$M = -I + Q$$

onde $Q_{ij} = F_i'(0) W_{ij}$. Se Q for bastante pequeno (ou seja, há apenas uma fraca ligação entre as unidades ou $F_i'(0)$ é pequeno), então os valores próprios de M têm partes reais negativas e todas as perturbações decaem para o estado de «fundo». Contudo, suponhamos que alguma influência externa (e.g., luz trémula, alucinogéneos, etc.) «excita» a rede tornando, por exemplo, F_i' elevadas. Então é possível que alguns dos valores próprios de M venham positivos, o que conduz a instabilidades. A «forma» das actividades resultantes desta instabilidade é uma combinação de todos os vectores próprios de Q cujos valores próprios se tornaram simultanea-

mente positivos. Genericamente, para uma rede simétrica aleatória, espera-se uma bifurcação «transcrítica» ou em «forquilha» conduzindo a uma forma excitada única que reflecte a organização aleatória da rede. Se o efeito dos alucinogêneos ou de outros estímulos externos é produzir um aumento geral de excitabilidade da rede, então é possível que o estado de «fundo» possa tornar-se instável conduzindo a uma configuração espontânea de actividade que reflecte a organização associativa da rede recorrente. Se existem muitos valores próprios praticamente idênticos, então as configurações resultantes podem ser misturas complexas dos vectores próprios e das dinâmicas que regem as configurações.

Consideremos, por exemplo, uma matriz associativa de pesos,

$$W_{ij} = \sum_{k=1}^m \xi_i^k \xi_j^k$$

onde $\xi^1, \xi^2, \dots, \xi^m$ são m vectores a memorizar. Se cada um deles tem amplitude 1 e são quase ortogonais, então a matriz peso, W , tem um valor próprio $\lambda = 1$ de multiplicidade m , sendo estes m vectores os vectores próprios associados. Os métodos padrão de bifurcação podem ser aplicados para deduzir um conjunto de equações para os m vectores próprios com a forma:

$$\dot{A}_i = \nu A_i + p_2(A_1, \dots, A_m) + p_3(A_1, \dots, A_m) \quad (6)$$

onde ν é um parâmetro que controla a estabilidade do estado de «fundo» e p_2, p_3 são termos quadráticos e cúbicos, respectivamente. A configuração da actividade que surge passado o ponto de bifurcação tem a forma:

$$u(t) \approx A_1(t)\xi^1 + \dots + A_m(t)\xi^m.$$

Portanto, para uma rede associativamente ligada, as configurações de actividade resultantes, que ocorrem quando o estado de «fundo» se torna instável, são misturas dinamicamente variadas de várias «memórias formadas».

Que tem isto a ver com «constantes de forma»? A análise acima esquematizada assume que a matriz-peso, W_{ij} , não tem estrutura ou tem uma estrutura proveniente de memórias formadas. Em áreas associativas do cérebro, é razoável considerar esta hipótese. Contudo, como mencionámos acima, o córtex visual possui uma certa estrutura. As ligações estão longe de ser aleatórias e têm numerosas simetrias intrínsecas. Por exemplo, se a matriz dos pesos, W , é a associada a uma malha espacial bidimensional, com ligações que dependem apenas da distância, então os vectores próprios correspondentes a W têm uma estrutura periódica dupla

(e.g., tiras, hexágonos ou pontos). As configurações resultantes que surgem na malha bidimensional são combinações de padrões formados por tiras e pontos.

Como poderá ser apreendida uma configuração de tiras no córtex? A resposta é a chave para a compreensão das constantes de forma de Kluver. Considere-se uma configuração de luz imposta na retina. Uma vez que há um «mapa» topográfico da retina para o córtex visual, isto resultará numa configuração no córtex. Reciprocamente, cada configuração espacial de activação no córtex está associada a uma configuração virtual na retina. Em particular, Schwartz (1977) mostrou que cada ponto (r, θ) da retina, em coordenadas, está associado a um ponto (x, y) no córtex. A transformação dada por:

$$x = \frac{\alpha}{\varepsilon} \ln \left(1 + \frac{\varepsilon}{\omega_0} r \right), \quad y = \frac{\beta r \theta}{\omega_0 + \varepsilon r},$$

e a inversa desta,

$$r = \frac{\omega_0}{\varepsilon} \left(e^{\frac{x\varepsilon}{\alpha}} - 1 \right), \quad \theta = \frac{\varepsilon y}{\beta \left(1 - e^{-\frac{x\varepsilon}{\alpha}} \right)},$$

permite-nos transformar uma configuração cortical na sua análoga retinal. Perto do ponto central (*fovea*), a transformação é a identidade em escala. Longe desse ponto central (*fovea*), a transformação é o logaritmo complexo e a sua inversa é a exponencial complexa. A figura 7 mostra a acção desta transformação e o respectivo efeito numa constante de forma de tipo funil e de tipo espiral.

Esta transformação conjugada com a noção de que as alucinações visuais resultam da perda de estabilidade, primeiro no córtex estriado e posteriormente em áreas de associação mais elevada, permite-nos explicar as constantes de forma de Kluver e explicar indirectamente os desenhos abstractos das gravuras xamanistas. Ermentrout e Cowan (1979) estudaram a simetria de ligações com base na posição espacial e utilizaram a teoria de grupos para mostrar que as configurações espaciais que bifurcam a partir do estado de «fundo» não são mais que tiras e pontos. A forma normal das equações (6) possui uma estrutura muito simplificada. Por exemplo, surge a seguinte forma:

$$A'_1 = vA_1 - A_1(aA_1^2 + bA_2^2)$$

$$\dot{A}_2 = vA_2 - A_2(aA_2^2 + bA_1^2)$$

onde a, b são constantes dependentes do modelo e A_1 é a amplitude correspondente às tiras horizontais de actividade, enquanto A_2 é a amplitude correspondente às tiras verticais. Se $A_1 \neq 0, A_2 = 0$ ($A_1 = 0, A_2 \neq 0$), então ocorrem tiras horizontais (verticais), enquanto se $A_1 = A_2$, então ocorrem pontos. A estabilidade destas configurações de bifurcação depende apenas das constantes a, b . A figura 8 mostra exemplos de tiras horizontais e pontos de actividade em coordenadas corticais (em cima à esquerda), bem como a imagem observada em coordenadas retinais (em cima à direita).

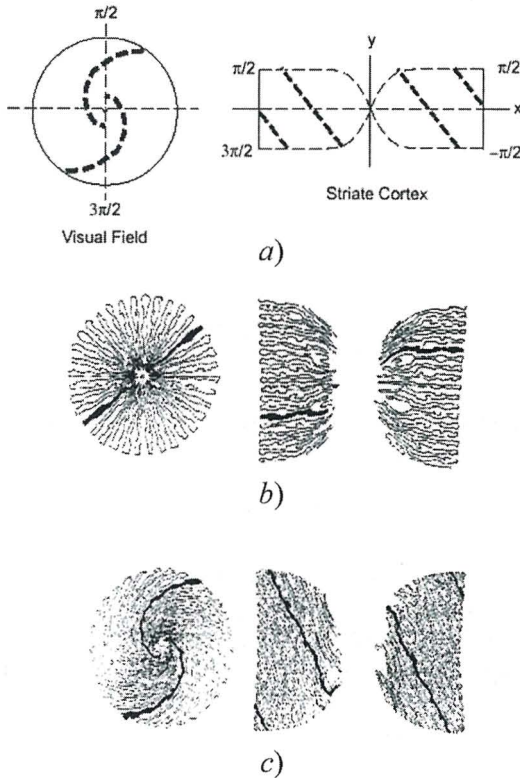


Fig. 7 — (a) A transformação retinal \rightarrow cortical (retina \rightarrow córtex) e sua acção, a (b) a constante de forma «funil» e (c) a constante de forma «espiral»

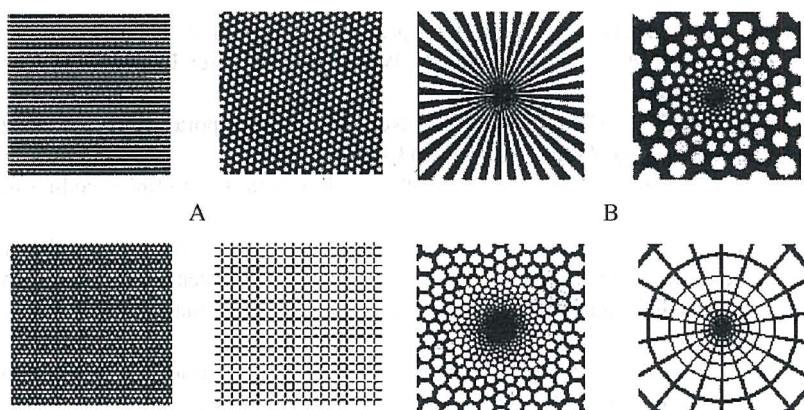


Fig. 8 — Configurações de bifurcação em coordenadas corticais (à esquerda) e em coordenadas retiniais (à direita). As imagens superiores correspondem a configurações originadas por ligações dependentes da distância e as inferiores correspondem a ligações dependentes da distância e da orientação (segundo Bressloff, et al.)

Se a estrutura da matriz de conexão incorpora a orientação específica de neurónios corticais, então os vectores próprios correspondentes são mais complicados, mas são interpretados pelo sistema visual como segmentos de recta dispostos em filas periódicas. Bressloff *et al.* (2000) analisaram este caso e deduziram as equações das formas normais. As configurações que correspondem aos pontos fixos destas equações são mais complexas do que simples tiras e pontos. Contudo, completam a enumeração das constantes de forma. As imagens inferiores da figura 8 mostram dois exemplos que correspondem a constantes de forma do tipo reticulado e do tipo teia de aranha.

Referências

- Bressloff, P. C., Cowan, J. D., Golubitsky, M., Thomas, P. J., Wiener, M. (2000) – Geometric visual hallucinations, Euclidean symmetry, and the functional architecture of striate cortex, Proc. Royal Society (a publicar).
- Del Negro, C. A., Hsiao, C. F., Chandler, S. H., Garfinkel, S. (1998) – Evidence for a novel bursting mechanism in rodent trigeminal neurons. *Biophys. J.* 75, 174-182.
- Ermentrout, G. B., e Cowan, J. D., (1979) – A Mathematical Theory of Visual Hallucinations, *Kybernetik* 34: 137-150.

- Ermentrout, G. B., (1985) – A model for premigrainous auras. In *The Neurobiology of Pain*. ed. A. H. Holden e W. Winlow. Manchester University Press: Manchester.
- Ermentrout, G. B., (1998) – Neuronal Nets as spatio-temporal pattern forming systems. *Rep. Prog. Phys.* 61: 353-430.
- Gutkin, B. S., e Ermentrout, G. B., (2001) – How neurons multiplex coding of fast and slow stimuli (em preparação).
- Gutkin, B. S., e Ermentrout, G. B., (1998) – Dynamics of membrane excitability determine interspike interval variability: a link between spike generation mechanisms and cortical spike train statistics. *Neuronal Comput.* 10(5): 1047-1065.
- Hedges, K., (1981) – Phosphenes in the context of Native American rock art, in *American Indian Rock Art* vol VII, VIII, ed. FG Bock.
- Hudson, T., e Lee, G. (1984) – Function and symbolism in Chumash rock art. *J. New World Archaeology* 6: 26-47.
- Hodgkin, A. L., e Huxley, A. F. (1952) – A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Phys.* (Londres) 117: 500-544.
- Jaeger, D., e Bower, J. M. (1999) – Synaptic control of spiking in cerebellar Purkinje cells: dynamic current clamp based on model conductances. *J. Neurosci.* 19(14): 6090-101.
- Johnston, D., e Wu, S., (1997) – Foundations of cellular neurophysiology. MIT Press.
- Kluver, H., (1996) – *Mescal and the Mechanisms of Hallucination*. University of Chicago Press: Chicago
- Kyriazi, H. T., e Simons, D. J., (1993) – Thalamocortical response transformations in simulated whisker barrels. *J. Neurosci.* 13: 1601-1615.
- Lancaster, B., e Adams, P. R., (1986) – Calcium-dependent current generating the after-hyperpolarization of hippocampal neurons. *J. Neurophys.* 55(6): 1268-82.
- Le Masson, G., Le Masson, S., Moulins, M., (1995) – From conductances to neuronal network properties: analysis of simple circuits using hybrid network method. *Prog. Biophys. Mol. Biol.* 64(2-3): 201-20.
- Lewis-Williams, J. D. e Dowson, T. A. (1988) – The signs of all times: entoptic phenomena in upper Paleolithic art. *Current Antropology* 29: 201-245.
- Mainen, Z. F. e Sejnowski, T. S. (1995) – Reliability of spike timing in neocortical neurons. *Science* 268: 1503-1506.
- McCormick, D. A., e Huguenard, J. R., (1992) – A Model of the Electrophysiological Properties of Thalamocortical Relay Neurons. *J. Neurophys.* 68: 1384-400.
- Pinto, D. J., Brumberg, J. C., Simons D. J., e Ermentrout, G. B., (1996) – A quantitative population model of whisker barrels: Re-examining the Wilson-Cowan equations. *J. Comput. Neurosci.* 3: 247-264.

- Pinto, D. J., Brumberg, J. C., Simons, D. J., (2000) – Circuit dynamics and coding strategies in rodent somatosensory cortex. *J. Neurophysiol.* 83: 1158-1166.
- Pinto, D. J., Hartings, J. D., Brumberg, J. C., Ermentrout, G. B., Simons, D. J., (2001) – Recurrent excitation in damping cortical circuits: An alternative mechanism for response tuning. *J. Neurosci* (submetido).
- Rall, W., e Agmon-Snir, A. (1998) – Cable theory for dendritic neurons. In *Methods in Neuronal Modeling*. C Koch, I Segev, eds. MIT Press Cambridge: MA. pp. 27-93.
- Rinzel, J., e Ermentrout, G. B., (1998) – Analysis of neuronal excitability and oscillations. In *Methods in Neuronal Modeling*. C Koch, I Segev, eds., MIT Press Cambridge: MA.
- Swartz, E., (1977) – Spatial mapping in the primary sensory projection: analytic structure and relevance to projection. *Biol. Cybernetics* 25: 181-194.
- Shadlen, W. M., e Newsome, W. T., (1998) – The variable discharge of cortical neurons: implications for connectivity, computation and information coding. *J. Neurosci.* 18: 3870-96.
- Sharp, A. A., O’Neil, M. B., Abbott, L. F., e Marder, E., (1993) – Dynamic clamp: computer-generated conductances in real neurons. *J. Neurophysiol.* 69 (3): 992-5.
- Simons, D. J. (1978) – Response properties of vibrissa units in rat SI somatosensory neocortex. *J. Neurophysiol.* 41: 798-820.
- Stafstrom, C. E., Schwindt, P. C., e Crill, W. E., (1982) – Negative slope conductance due to a persistent subthreshold sodium current in cat neocortical neurons in vitro. *Brain Res.* 236: 221-226
- Tyler, C. (1978) – Some new entoptic phenomena. *Vision Research* 18: 1633-1639.
- White, E. L., (1989) – *Cortical Circuits: Synaptic organization of the cerebral cortex*. Birkhauser: Boston, MA.
- Woolsey, T. A., van der Loos, H., (1970) – The structural organization of layer IV in the somatosensory region (SI) of mouse cerebral cortex. *Brain Res.* 17: 205-242.

(Página deixada propositadamente em branco)

Alfio Quarteroni

École Polytechnique Fédérale de Lausanne, Suíça

Politécnico de Milão, Itália

Modelação matemática: uma síntese entre a teoria e as aplicações¹

1. INTRODUÇÃO

O objectivo deste texto é analisar a modelação matemática, o seu papel na ciência da engenharia bem como em ambientes não académicos.

Nas últimas décadas testemunhámos um enorme aumento da utilização da matemática no desenvolvimento teórico de diversos campos científicos bem como em aplicações nos mais variados aspectos da vida real.

Há cerca de meio século, salvo raras excepções, a expressão «matemática aplicada» significava essencialmente a aplicação da matemática à mecânica. Hoje em dia a matemática pode ser considerada um elemento da aprendizagem e do processo descritivo de inúmeras disciplinas como, por exemplo, a física, a química, a biologia, a engenharia, a medicina e a economia.

A modelação matemática é um processo que pretende resolver problemas provenientes destas disciplinas, ou seja, a sua representação por um conjunto de equações matemáticas, a análise das suas propriedades qualitativas, o estabelecimento de métodos numéricos para a sua

¹ Tradução de Lia Sandra dos Santos.

aproximação e, finalmente, a sua implementação em computador através de algoritmos apropriados (figura 1).

Podemos questionar-nos sobre o que motiva a utilização da modelação matemática, que benefícios se podem obter com a sua utilização. No tocante às ciências experimentais, elas recorrem cada vez mais à utilização de ferramentas matemáticas sofisticadas, o que lhes permite evoluir de um estado empírico inicial para uma estrutura logicamente organizada. As motivações de outros tipos de ciências podem, no entanto, ser de natureza diferente.

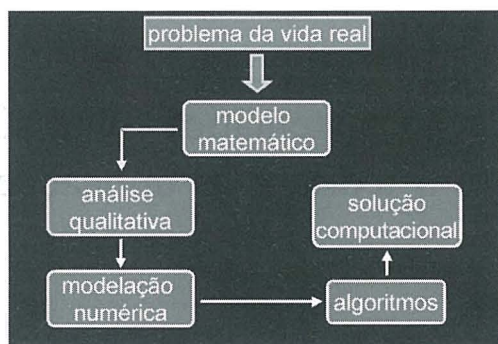


Fig. 1 — Modelação matemática

Podem citar-se inúmeros casos de utilização de modelos matemáticos para simular ou mesmo controlar dinâmicas sociais ou demográficas. Em muitos dos problemas encontrados na economia, os modelos matemáticos fornecem informações quantitativas, operando com um número de variáveis desconhecidas consideravelmente maior do que aquele que poderia ser tido em consideração no contexto de uma análise qualitativa. Isto acontece, por exemplo, sempre que desejamos prever a reacção de sistemas fortemente interactivos quando as condições de referência (i.e., o mercado) se alteram muito rapidamente. Por último, deve ser referido que foi dado um forte impulso à utilização de modelos matemáticos em grande escala, pela análise sistémica, através da qual o campo de observação foi alargado para uma escala global. Um exemplo significativo é o modelo, criado durante o período da Guerra Fria, para a previsão do comportamento da atmosfera após a utilização de armas nucleares, perante a terrível perspectiva do inverno nuclear. Outro modelo é aquele que procura prever a alteração atmosférica global, envolvendo a interacção

entre os oceanos, a terra e a atmosfera, e cujo objectivo é a simulação precisa das variações climáticas devidas ao efeito de estufa.

A MODELAÇÃO MATEMÁTICA NA CULTURA DA ENGENHARIA

Através da modelação matemática, um problema proveniente do mundo real é transferido do seu universo próprio para um *habitat* onde pode ser correctamente analisado e resolvido numericamente, sendo depois transferido de novo para o seu contexto de origem para visualização, análise e validação dos resultados obtidos.

Um modelo não expressa necessariamente a verdadeira natureza do problema nem a sua complexidade, pois a realidade é muitas vezes demasiado complexa, não sendo possível a sua rigorosa representação em termos de fórmulas matemáticas. Contudo, o modelo tem de fornecer uma síntese útil do problema.

Uma característica da modelação é a abstracção, isto é, a capacidade de identificar características comuns em campos diferentes, permitindo a elaboração, *a priori*, de ideias muito gerais, e depois a sua aplicação em contextos muito diferentes uns dos outros.

Muitos ramos da engenharia, tais como a aeronáutica, a mecânica, a engenharia civil, a química e a engenharia nuclear têm tradicionalmente beneficiado com a utilização de modelos matemáticos para o planeamento, controlo, optimização e gestão de processos industriais e tecnológicos. Mais recentemente, os modelos matemáticos foram também introduzidos noutros campos, de entre os quais mencionamos, sem ser exaustivos, a microelectrónica, o ambiente e as finanças (vide figura 2).

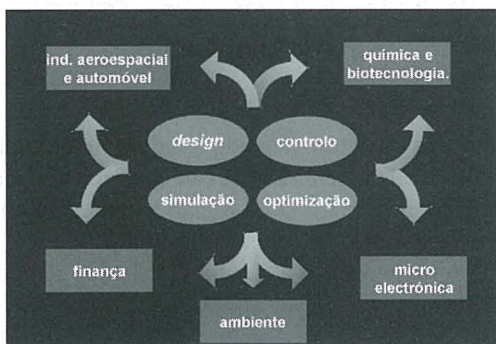


Fig. 2 — Objectivos dos modelos matemáticos e campos de aplicação

Os engenheiros utilizam a análise experimental e a simulação numérica de forma complementar. A última é mais flexível no que diz respeito à variação de parâmetros físicos ou à mudança das condições de limite e às formas geométricas. A simulação numérica pode fornecer uma descrição pontual de todas as variáveis que intervêm no processo em questão.

A modelação matemática é por isso um elemento de ligação entre a análise experimental e o planeamento da produção (*vide* figura 3). No topo deste processo, os modelos matemáticos são sempre inspirados por princípios físicos fundamentais, tais como as equações do equilíbrio estático, ou os princípios da massa, energia e o momento de conservação da dinâmica do meio contínuo, sendo a descrição da reologia dos materiais, a determinação de condições de limite e a dimensionalização e medições dos coeficientes do problema e parâmetros descritivos fornecidos pela análise experimental.

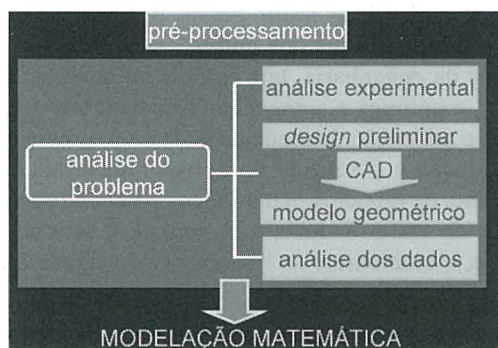


Fig. 3 — Análise preliminar

Muitas vezes, uma característica adicional do modelo matemático é a apresentação de um modelo geométrico, isto é, a representação por sólidos modeladores ou ferramentas CAD da região tridimensional complexa na qual as equações do modelo devem ser resolvidas.

Na sequência deste processo, a complexidade dos campos de vectores tridimensionais obtidos tem de ser analisada de uma forma logicamente organizada e validada à luz dos dados experimentais disponíveis, e principalmente da intuição da engenharia. Esta análise retrospectiva, por

sua vez, conduz a um processo iterativo que permite modificar o modelo, as suas equações ou os seus parâmetros, até que os resultados numéricos obtidos estejam de acordo com as verificações experimentais para uma grande família de casos de teste (*vide* figura 4).

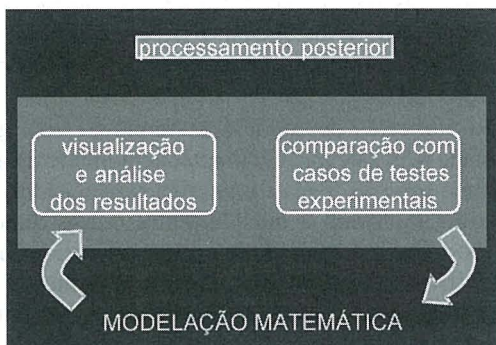


Fig. 4 — Análise a posteriori

O exemplo ilustrado na figura 5 refere-se ao projecto de um veículo (um carro ou um avião) e mostra a interação entre os testes no túnel de vento e a simulação numérica, antes de se obter o projecto final.

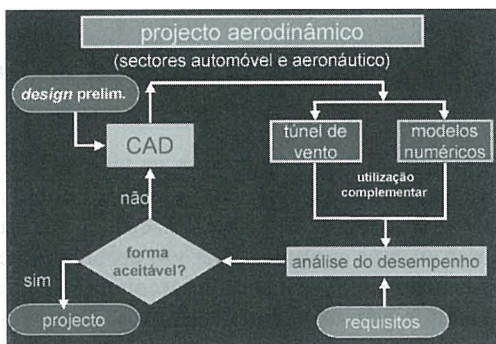


Fig. 5 — Da concepção inicial à final

A SIMULAÇÃO NUMÉRICA

Para um especialista de matemática aplicada, o principal objectivo é a resolução efectiva do problema expresso pelo modelo matemático. Muitas vezes, estes problemas são determinados por complexas equações de derivadas parciais, cuja solução explícita não é conhecida ou cuja representação não é passível de construção. Os algoritmos numéricos têm então de ser desenvolvidos de forma a permitir sempre a solução do modelo em questão, desde que este esteja bem formulado do ponto de vista matemático.

A simulação numérica do modelo matemático visa garantir que a solução aproximada seja calculada computacionalmente de modo eficiente e que o erro entre a solução física e numérica seja pequeno e controlável.

A eficiência do algoritmo é medida, entre outros factores, pela sua complexidade, isto é, a quantidade de recursos necessários para implementar o algoritmo no computador, nomeadamente o tempo de computação e a memória requerida. A este propósito, note-se que os computadores disponíveis hoje em dia não permitem a solução de problemas de complexidade arbitrária.

Considerem-se alguns exemplos provenientes de sistemas determinísticos cujo comportamento é descrito por equações de derivadas parciais da dinâmica dos fluidos. Esta disciplina é de um interesse primordial, não só porque se encontra numa larga variedade de aplicações, mas também porque dá origem a um conjunto de problemas que estão entre os mais difíceis de resolver de toda a matemática aplicada.

UM EXEMPLO INTERESSANTE: MODELAÇÃO NA DINÂMICA DOS FLUIDOS

Os fluidos têm uma enorme importância na vida quotidiana. De facto, a dinâmica da atmosfera, a dispersão dos poluentes no ar, a formação de correntes e a circulação de sedimentos em cursos de água, a corrente sanguínea no nosso sistema cardiovascular são problemas de dinâmica dos fluidos. Outros processos de natureza aparentemente diferente podem também ser descritos por modelos de fluidos. Por exemplo, a nível microscópico, os electrões num meio semiconductor comportam-se como um fluido condutor de corrente.

Todos os exemplos precedentes, bem como muitos outros que se poderiam citar, podem ser modelados por um sistema de equações de

derivadas parciais introduzido pelo engenheiro francês Louis Marie Henri Navier e pelo físico irlandês Sir George Gabriel Stokes. Note-se que apesar de estas equações serem conhecidas há mais de um século, muitos aspectos do comportamento dos fluidos escapam ainda à nossa compreensão.

Os métodos analíticos podem proporcionar uma forma explícita de solução para as equações de Navier-Stokes, no caso de serem consideradas hipóteses tão restritivas que o problema simplificado tenha pouco interesse prático.

A tendência dos fluidos para mostrar um comportamento complexo ou, mais precisamente, turbulento é a razão para tal dificuldade. A propósito, a turbulência não é uma propriedade constitutiva (ou intrínseca) do fluido; é, sim, um regime específico do escoamento do fluido que surge sempre que um grande número de graus de liberdade intervém na dinâmica dos fluidos. Tratando-se de um modelo não-linear, todas as escalas de movimento possível interagem. Tal interacção é responsável pela turbulência.

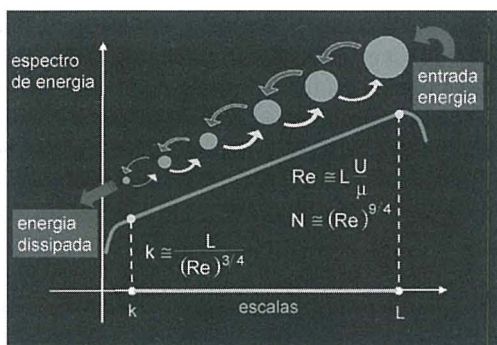


Fig. 6 — Escalas de turbulência e espectro de energia

Um indicador do nível de turbulência é o chamado número de Reynolds, directamente proporcional à velocidade média do fluido e inversamente proporcional à sua viscosidade molecular. O número de Reynolds expressa a importância da convecção que ocorre em escalas macroscópicas, relativamente à dissipação, que ocorre em vez disso a nível molecular.

A análise das equações de Navier-Stokes permite afirmar que o número de graus de liberdade de uma corrente turbulenta é da ordem de $N=Re^{9/4}$

(por uma questão de simplicidade, podemos pensar que estamos a trabalhar em cada nível temporal com uma série de Fourier com N modos). Dado que o número de Reynolds de muitas correntes, por exemplo as que circundam num avião, é da ordem de 10^6 , o número de graus de liberdade correspondente pode ser maior que 10^{13} , isto é, dez biliões.

A energia do fluido em grandes escalas, devido às condições de fronteira e às forças volúmicas, é transferida para escalas cada vez menores com interacção não-linear dos graus de liberdade activos no fluido, através de uma espécie de cascata de energia, baixando para uma escala tão pequena que a energia se dissipa no calor. A escala activa mais pequena, que indicaremos por K , é conhecida por escala de *Kolmogorov* e é directamente proporcional à escala maior e inversamente proporcional a $Re^{3/4}$ (vide figura 6).

Por exemplo, na turbulência atmosférica, a que tem de ser simulada nas previsões meteorológicas, o número de Reynolds é da ordem de 10^8 . Isto implica que a entrada de energia, na escala de um quilómetro, afecta até as escalas na ordem de um milímetro. Uma característica geral da turbulência do fluido é a presença das regiões de «escoamento coerente», com turbulência que se auto-organiza dando origem a padrões de escoamento regular e não caótico, espirais normalmente designadas por vórtices. As estruturas coerentes estão organizadas de forma hierárquica, no sentido em que os últimos vórtices geram vórtices ainda mais pequenos, e assim sucessivamente até um diâmetro K da ordem da escala de Kolmogorov, como ilustra a figura 6. Reciprocamente, todas as escalas são influenciadas por uma escala mais pequena num processo em cadeia interminável.

Quando utilizamos a simulação numérica quase nunca podemos ter em consideração todo o conjunto das escalas activas, desde a macroescala L até à escala K de Kolmogorov.

Por exemplo, na simulação da troca de energia até à escala de Kolmogorov, na turbulência atmosférica, devíamos utilizar um passo de discretização espacial da ordem de um milímetro. Isto implicaria resolver um sistema dinâmico com 10^{18} variáveis, sistema este que nenhum computador existente consegue tratar. É por isso necessário desistir da simulação numérica directa e utilizar métodos de redução que aproximam apenas um número limitado de escalas, mas fornecem um modelo conveniente da troca de energia de pequenas escalas e de grandes escalas, através de processos apropriados de cálculo da média e de normalização (vide figura 7 para uma representação qualitativa desta abordagem).

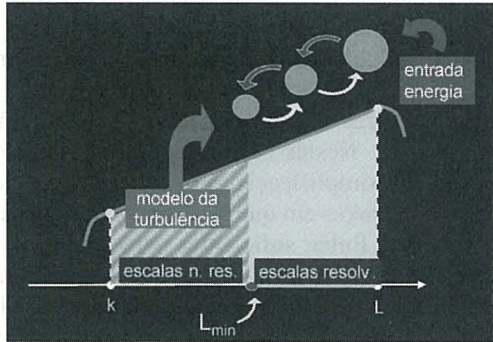


Fig. 7 — Escalas na simulação de escoamentos turbulentos

Uma outra dificuldade reside na modelação e simulação dos fluxos em presença de reacções químicas, quer no caso de problemas de uma só fase quer no caso de problemas multifase. Por limitações de tempo não desenvolveremos este aspecto, apesar do interesse que assume em muitas aplicações. Na indústria, a utilização da experimentação constituiu durante muito tempo o único instrumento disponível para analisar os fenómenos relacionados com os escoamentos dos fluidos. Em alguns casos, no entanto, a temperatura e a velocidade envolvidas são tão elevadas que a experimentação em túneis de vento é mesmo impossível, como acontece, por exemplo, na reentrada na atmosfera de um veículo espacial. Noutras situações, como por exemplo o estudo dos efeitos fisiopatológicos induzidos pela dinâmica do sangue, a experimentação *in vivo*, além de ser pouco precisa, é obviamente crítica para o paciente.

A modelação numérica do movimento dos fluidos, se for realizada com grande rigor matemático, pode fornecer resultados precisos para a compreensão de fenómenos complexos e ser um instrumento de apoio na análise experimental e na realização de projectos industriais. Naturalmente, é preciso desenvolver métodos numéricos adequados e utilizar algoritmos que explorem de modo óptimo as modernas arquitecturas de cálculo vectorial e paralelo.

SIMPLIFICAR PARA RESOLVER

A complexidade dos problemas a resolver pode ser, no entanto, demasiadamente elevada relativamente ao papel que a simulação numérica pode assumir. Por exemplo, na fase do projecto e optimização de um

veículo na indústria automobilística, recorre-se à utilização interactiva de diversos códigos de cálculo para analisar as várias componentes do projecto. Os tempos de elaboração necessários à análise de alguns elementos são muitas vezes demasiadamente elevados, não permitindo várias simulações diárias. Nestes casos impõe-se uma reformulação do modelo e a sua posterior simplificação. Por exemplo, o modelo reduzido das equações de Navier-Stokes em que se elimina a viscosidade dá origem às chamadas equações de Euler, suficientes para prever de modo preciso a pressão na aerodinâmica externa e também as trocas de energia interna entre fluido e máquina. Uma simplificação posterior conduz à equação do potencial não-linear, uma única equação e não um sistema de equações como anteriormente, que é adequada para descrever fluxos irrotacionais e isotrópicos, sendo por este motivo frequentemente utilizada para a simulação de regimes não ultra-sónicos.

É claro que a utilização de modelos simplificados diminui drasticamente a complexidade do problema, tornando assim possível efectuar simulações computacionais, que de outro modo não seriam realizáveis. Esta simplificação deve ser, no entanto, justificada: do ponto de vista físico, não deve alterar o significado do problema em análise, e do ponto de vista matemático deve conservar as propriedades teóricas fundamentais do modelo original. A síntese destas duas exigências nem sempre é fácil e exige um esforço conjunto de matemáticos e engenheiros. Num outro âmbito, a redução da complexidade pode também obter-se recorrendo a partições geométricas do problema, o que o torna particularmente adaptado à utilização do cálculo paralelo. Neste caso, transforma-se o problema numérico original numa sucessão de problemas de dimensão reduzida que podem ser resolvidos simultaneamente num ambiente de cálculo com multiprocessadores. Na figura 8 ilustra-se de modo esquemático o processo de simulação do fluxo que se gera nas vizinhanças da asa de um avião durante a aterragem.

O desenvolvimento de arquitecturas paralelas incentivou os matemáticos a estudar novos métodos de cálculo, baseando-se muitas vezes na reformulação do próprio modelo matemático.

Na figura 9 mostra-se que a diminuição do tempo de cálculo, devido a vinte anos de evolução no *hardware*, é inferior à diminuição resultante do progressivo aperfeiçoamento dos métodos numéricos, incluindo a sua adaptação às arquitecturas vectoriais.

O desenvolvimento em simbiose do *hardware* e do *software* é, sem dúvida, uma das condições determinantes para resolver com sucesso os modelos matemáticos de complexidade crescente que emergem nos desafios científicos contemporâneos.

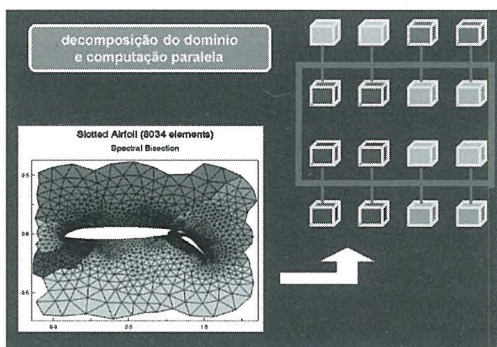


Fig. 8 — Decomposição do domínio e cálculo paralelo

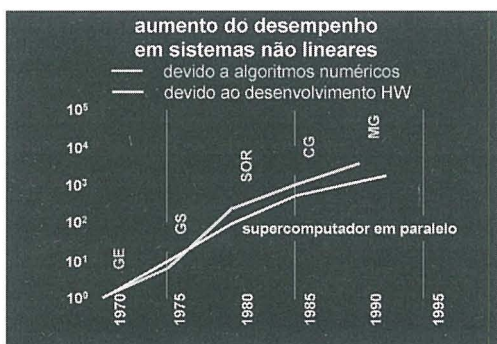


Fig. 9 — Diminuição do tempo de cálculo na resolução de sistemas lineares

(Página deixada propositadamente em branco)

A matemática e a compreensão da medicina

O envolvimento das ciências exactas na evolução da ciência médica tem sido um processo lento mas irreversível. Nas últimas décadas, as ciências biomédicas progrediram, em grande medida, graças à forte contribuição da física, da matemática, da informática e da engenharia. Num número considerável de áreas da medicina, são reconhecidos os esforços no sentido de associar conteúdo científico e precisão matemática às novas abordagens dos mecanismos fisiológicos, ao desenvolvimento e avaliação das novas técnicas de diagnóstico e à optimização das metodologias instaladas. Por outro lado, em contextos restritos, os métodos matemáticos são utilizados, desde há muito, na biologia e na medicina. Certas relações entre a biologia e a matemática são até tradicionais, nomeadamente na genética, na epidemiologia e na investigação do crescimento das populações. Se existem ainda dúvidas em considerar a matemática biológica como uma ciência autónoma, já não restam dúvidas de que a genética matemática é, desde há muito tempo, uma disciplina estabelecida.

Além disso, actualmente, a forte contribuição da estatística e de diversos métodos matemáticos de processamento de dados é uma constante em áreas específicas da medicina e da biologia.

Existem ramos, tais como a medicina nuclear e especializações da imagiologia como a ressonância magnética, que, pela sua própria natureza, envolveram desde o início fortes componentes matemáticas, físicas e

tecnológicas e cujo progresso tem dependido de forma considerável de um conhecimento altamente especializado nestes campos.

Conceitos físicos e métodos matemáticos, que na maioria das outras áreas da medicina são utilizados apenas em investigação e desenvolvimento, têm sido, na prática, ferramentas de rotina em medicina nuclear. Por exemplo, a determinação de espaços de distribuição, os ajustamentos lineares e não-lineares, as abordagens de modelos estocásticos, as aproximações polinomiais e a análise compartimental. Podemos provavelmente afirmar que, no que se refere à utilização do método científico em abordagens clínicas, a medicina nuclear está muito mais avançada do que outras áreas da medicina.

Desde a década de 60 o computador tem tido um impacto gradual em praticamente todos os aspectos da medicina e, particularmente, na aplicação dos métodos matemáticos. A utilização corrente de computadores em linha está na base da maior parte dos últimos avanços da medicina. Contudo, o computador tem criado exigências cada vez maiores, não só em medicina como noutras disciplinas com problemas comuns. Para responder a estas exigências, desenvolveram-se novas subespecialidades com uma forte componente matemática e com objectivos bem definidos, tais como o processamento de imagem, a manipulação de dados dinâmicos, o reconhecimento de padrões, a simulação e a visualização 3-D.

A teoria dos sistemas é mais um exemplo das potencialidades dos métodos matemáticos em biomedicina. Esta teoria tem permitido o desenvolvimento de métodos elaborados para abordar eficazmente processos fisiológicos com interesse do ponto de vista do diagnóstico. Em medicina, recorre-se frequentemente a abordagens funcionais com base na teoria dos sistemas, utilizando modelos e análogos como ferramentas de partida.

Por exemplo, em medicina nuclear os modelos permitem-nos prever matematicamente a história, no tempo, das moléculas radioactivas e avaliar a sua concentração, bem como a dos seus metabolitos activos, em vários compartimentos. Num sentido lato, cada metodologia da medicina nuclear é um procedimento da teoria dos sistemas: utiliza-se uma função de entrada para estudar o sistema através de uma resposta induzida. Em medicina nuclear o sistema é um compartimento, ou um conjunto de compartimentos, com propriedades metabólicas específicas. A função de entrada está intimamente relacionada com os radiofármacos utilizados e a resposta detectada depende do transporte e do destino biológico do traçador utilizado.

A medicina nuclear tem progredido devido aos avanços registados em cada uma das fases deste processo: melhoria nas funções de entrada devido à introdução de novos radiofármacos, actualizações da tecnologia e fiabilidade dos detectores, avanços na capacidade de modelação dos processos metabólicos sofridos pelos produtos aplicados à entrada dos sistemas e, finalmente, aumento da capacidade de extracção e processamento de informação a partir dos resultados. Um compartimento é um espaço de distribuição teórica de uma espécie molecular tal como é dado pelo método de diluição, utilizando um traçador. Pode significar qualquer coisa anatomicamente bem definida, como o plasma, ou uma distribuição teórica do volume fisiológico sem limites materiais.

Existem muitos compartimentos nos organismos vivos, quer fisicamente identificáveis quer só teoricamente definíveis. De facto, um organismo pode, fundamentalmente, ser considerado um conjunto de compartimentos interrelacionados de forma dinâmica. Ocorre uma troca contínua de matéria e energia entre estes compartimentos e também com o ambiente. Estas trocas são os processos físico-químicos relacionados com a absorção, distribuição, síntese, degradação e excreção. A doença é uma alteração de uma ou de algumas destas trocas ou dos seus mecanismos reguladores.

Geralmente, os modelos utilizados em medicina nuclear podem envolver sistemas compartimentais ou não compartimentais e conduzir a sistemas de equações diferenciais ou simplesmente a equações diferenciais ou integrais. As soluções destas equações conduzem a expressões, que deverão reproduzir os dados experimentais e o comportamento do sistema.

A convolução é uma operação importante no contexto dos problemas de entrada-saída dos sistemas. Representa-se pela equação integral (Eq. 1) onde $h(t)$ é um operador que converte um objecto ou uma função de entrada $I(t)$ numa função de saída $R(t)$ mensurável. A sua aplicabilidade baseia-se na suposição de linearidade do sistema em estudo, bem como na invariância do tempo (comportamento constante no tempo) das funções envolvidas.

$$R(t) = \int_0^t h(t - \tau) I(\tau) d\tau \quad (1)$$

A convolução das funções $h(t)$ e $I(t)$ é frequentemente representada sob a forma

$$R(t) = h(t) * I(t) \quad (2)$$

onde o símbolo * significa produto de convolução. Pode ser provado que a ordem das funções na Eq. 2 é intermutável. A função $h(t)$ é a função de resposta do sistema a um impulso, isto é, a resposta do sistema a uma função delta de intensidade unitária, no instante zero, ou, em termos de medicina nuclear, uma injeção muito rápida de actividade unitária no início da experiência.

A convolução pode ser visualizada se decomposermos a função $I(t)$ num número infinito de funções delta deslocadas no eixo dos tempos, tendo como amplitude o valor da função em cada instante e considerando como resultado final a soma das respostas do sistema a todas estas funções delta.

A maioria das situações práticas em medicina nuclear obriga a uma aproximação discreta da convolução. Quando as funções envolvidas são funções amostradas, a convolução integral assume a forma

$$R(k) = \sum_{n=0}^k h(k-n)I(n) \quad (3)$$

onde $R(k)$, $h(k-n)$ e $I(n)$ são os valores amostrados das funções. A Eq. 3 pode também ser escrita em forma matricial como

$$R = I \times H \quad (4)$$

onde I é uma matriz triangular de ordem $(k+1)$, em que a linha $n+1$ consiste em termos $I(n)$, $I(n-1)$, $I(n-2)$, ..., $I(0)$, seguidos de $(k-n)$ zeros. H e R são matrizes coluna com $k+1$ elementos.

Contudo, na maioria dos casos, o problema importante é o inverso da convolução, isto é, pretende-se obter $I(t)$ quando $R(t)$ e $h(t)$ são conhecidos ou pretende-se obter $h(t)$ quando $I(t)$ e $R(t)$ são conhecidos. Nestes casos, é teoricamente possível desconvoluir a função integranda, desconhecida, a partir das funções conhecidas e os procedimentos matemáticos necessários designam-se por desconvolução. A desconvolução é, assim, uma operação definida como sendo o inverso da convolução (De Lima J. J. P., 1984, 1995 e 1996).

As aplicações da desconvolução unidimensional e bidimensional são referidas desde a década de 50 em estudos dinâmicos e metabólicos na medicina nuclear, nomeadamente em estudos cardiovasculares, renográficos e gastroenterológicos, bem como nos procedimentos de processamento de imagem, como, por exemplo, nos algoritmos de restauração e de reconstrução de imagem.

A curva de diluição da actividade de um órgão após a injeção do traçador depende do órgão, do respectivo caudal sanguíneo e também da

forma do bolus de traçador aplicado. Em princípio, a desconvolução das curvas actividade/tempo anula os efeitos da forma do bolus mas, em muitas situações práticas, os resultados têm-se mostrado instáveis para uma utilização segura. A desconvolução contínua, após ajuste a funções com expressão analítica conhecida, leva muitas vezes a equações integrais com soluções gerais utilizáveis, mas, na prática, isto só acontece em condições restritas ou em casos muito especiais. Frequentemente, nos processos metabólicos o órgão é modelado como um sistema, a entrada é uma função actividade-tempo, por exemplo, do sangue e a função de resposta a actividade-tempo no próprio órgão (ou função de retenção). Utilizando a desconvolução, a função resíduo pode ser recuperada.

A desconvolução discreta é levada a cabo através da divisão polinomial, da divisão de matrizes, dos algoritmos de transformação rápida de Fourier e dos métodos de minimização numérica (Nimon C. C. *et al.*, 1980). Quando o ruído afecta os dados, os procedimentos de desconvolução são fortemente perturbados. A desconvolução é um processo condicionado para a autodestruição, pois mesmo pequenos erros nos dados obtidos, para as duas funções medidas, podem originar grandes desvios na solução. A perturbação nestas técnicas, causada pelo ruído, pode ser estudada acrescentando conteúdo de ruído conhecido às funções que estão a ser alvo de desconvolução e utilizando estes dados para desenvolver métodos eficientes de filtragem do ruído.

A detecção e a quantificação dos *shunts* intracardíacos da esquerda para a direita é uma aplicação bem sucedida da desconvolução em medicina nuclear (Alderson P. O. *et al.* 1979; Brendel A. J., 1983 e Nimon C. C. *et al.*, 1980). A curva de actividade-tempo da região do pulmão após uma injeção intravenosa de um traçador radioactivo não difusível atinge um pico devido à passagem de actividade proveniente da parte direita do coração. Se houver um *shunt* esquerdo-direito, a curva registará um segundo pico por causa da recirculação anterior originada pelo *shunt*. O tamanho relativo destes picos permite quantificar a dimensão do *shunt*, mas, muitas vezes, tal não é possível directamente, dado que os picos têm uma fracção considerável de sobreposição. Se, na imagem exibida, estiverem desenhadas regiões por cima da veia cava superior e do pulmão e se se obtiverem curvas de actividade-tempo, respectivamente $i(t)$ e $g(t)$, poderemos escrever

$$g(t) = a(t) * i(t) \quad (5)$$

onde $a(t)$ é a função de resíduo para o pulmão ROI (Região de Interesse). A função $a(t)$ é calculada pela desconvolução após filtragem. Isto torna-

-se necessário, pois o ruído estatístico encontra-se presente. A função obtida pela desconvolução assemelha-se à verdadeira função de resíduo e permite uma avaliação mais adequada do caudal da derivação.

Uma situação particularmente favorável para a aplicação directa da desconvolução, pois o ruído é pequeno e os dados são precisos, é a detecção/quantificação dos *shunts* arterio-venosos na circulação pulmonar utilizando microesferas de albumina humana marcadas com ^{99m}Tc . Este estudo pode ser realizado efectuando uma aquisição rápida de primeira passagem em vista posterior (0,5-0,8 s/matriz durante 80 s) em pacientes durante um estudo de rotina de perfusão do pulmão, sem nenhuma aplicação adicional de radiação ao paciente. A curva actividade/tempo na ROI pulmonar para a primeira passagem é obtida a partir da sequência de imagens adquirida. Presume-se que a actividade de colóide, que é transportada através da circulação pulmonar, será retida pelos capilares na primeira passagem e contribuirá com uma função em escada para a curva de actividade/tempo do pulmão, enquanto a actividade que segue pelos *shunts* intrapulmonares fornecerá uma curva de diluição que se inicia nos primeiros instantes. A curva de actividade/tempo de primeira passagem para a área do pulmão, ou seja, a curva de detecção de resíduo é, assim, a sobreposição destas duas curvas. A normalização do resultado para uma injeção instantânea de entrada (função delta) na entrada arterial do pulmão é obtida por desconvolução. As funções que intervêm são a função resíduo e a função de entrada, isto é, a curva actividade/tempo numa ROI na subclávia (ou ventrículo direito), após subtracção do fundo. O alisamento e a correcção do início da função de entrada são geralmente necessários para obter os melhores resultados.

A razão entre o caudal dos *shunts* e o caudal efectivo, ou seja, o caudal colateral pulmonar relativo, é dada por $d = A/(aT)$, onde A é a área subtendida pela curva de diluição, T é o tempo médio de passagem correspondente e a é a amplitude da função em escada, isto é, o valor constante da curva actividade-tempo desconvoluída após a anulação da curva de diluição (De Lima J. J. P., 1995).

A análise de desconvolução tem sido utilizada em diversos estudos clínicos, nomeadamente em estudos sobre a circulação cerebral Britton K. E. *et al.*, 1981), estudos cardíacos (Stritzke P., 1990) e estudos hepáticos (Brown H. P. *et al.*, 1988), mas, aparentemente, os estudos funcionais renais são o domínio onde se realizou a maior parte do trabalho (Britton K. E. *et al.*, 1980).

No nosso departamento (Floyd C. E. *et al.*, 1985) desenvolveu-se um método que descrevemos sucintamente a seguir e que se baseia nas

aplicações das técnicas de desconvolução ao modelo clássico de compartimentos para o estudo quantitativo da função hepática a partir da cintigrafia hepatobiliar utilizando os dados da ^{99m}Tc -mebrofenin.

A ^{99m}Tc -mebrofenin possui uma elevada especificidade para os hepatócitos, com uma cinética a processar-se num curto espaço de tempo, já que a remoção do sangue é eficaz (98%), o metabolismo intra-hepático é rápido e não ocorre captação entérica (Floyd C. E. *et al.*, 1985). Uma hora após a administração intravenosa de ^{99m}Tc -mebrofenin em indivíduos normais, o espaço extravascular retém cerca de 0,6% e a urina menos de 1%.

Consideremos um modelo aberto do sistema hepatobiliar, com três compartimentos, em estado estacionário e sem refluxo (figura 1). Injecta-se uma actividade A_0 de ^{99m}Tc -mebrofenin no compartimento sanguíneo para $t = 0$ com homogeneização imediata. Todas as reacções entre os compartimentos são de primeira ordem.

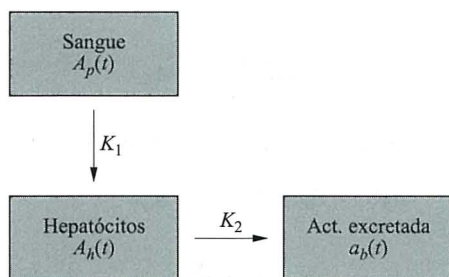


Fig. 1 — Modelo com três compartimentos para a quantificação de cintigrafia hepatobiliar

As taxas k_1 e k_2 são constantes, as funções $a_p(t)$, $a_h(t)$ e $a_b(t)$ são respectivamente valores da actividade nos compartimentos de sangue total, de parênquima hepático total e de excreção total, como mostra a figura 1. As equações diferenciais que representam as variações instantâneas da actividade no tempo nos três compartimentos são:

$$\frac{da_p}{dt} = -k_1 a_p, \quad (6)$$

para o compartimento de sangue,

$$\frac{da_h}{dt} = k_1 a_p - k_2 a_h, \quad (7)$$

para o compartimento do parênquima hepático e

$$\frac{da_b}{dt} = k_2 a_h \quad (8)$$

para o compartimento de excreção total.

A integração destas equações conduz a

$$a_p(t) = A_0 e^{-k_1 t} \quad (9)$$

para o compartimento de actividade sanguínea.

Para a actividade no compartimento do parênquima hepático temos

$$a_h(t) = A_p e^{-k_1 t} + A_h e^{-k_2 t}, \quad (10)$$

com

$$A_p = \frac{k_1 A_0}{k_2 - k_1} = -A_h. \quad (11)$$

A actividade no compartimento de excreção total é:

$$a_b(t) = A_0 + A_b e^{-k_1 t} + A_p e^{-k_2 t}, \quad (12)$$

onde

$$A_b = -\frac{k_2 A_0}{k_2 - k_1}. \quad (13)$$

Estamos interessados principalmente nas equações 9 e 10. Para efectuarmos o estudo são necessárias as curvas de actividade-tempo do sangue e do fígado, após a injeção de $^{99m}\text{Tc-mebrofenin}$. Numa região de interesse traçada na imagem cardíaca (ROI), obtida com a câmara gama, detecta-se uma fracção α do total de actividade do sangue e regista-se uma função $a'_p(t)$, formalmente idêntica à apresentada na equação 9, mas com um valor no instante zero dado por

$$A'_0 = \alpha A_0. \quad (14)$$

A actividade detectada externamente na ROI hepática é uma soma da actividade existente no parênquima hepático, no sangue hepático e

também na bÍlis em trÁnsito para o intestino. A actividade no espaço intersticial nÁo foi considerada neste modelo. Esta decisÁo é realista, dado que a percentagem de traçador no espaço extravascular para a Última geraçÁo de radiofÁrmacos hepatobiliares é mÍnima (~0,6%, 60 minutos apÓs a injeçÁo).

No nosso modelo, a actividade da bÍlis em trÁnsito, detectada na ROI hepÁtica, é considerada insignificante, o que é verdade nos primeiros instantes. Posteriormente, se forem detectadas as fracçÓes β e γ , respectivamente, da actividade hepÁtica total e da actividade total do sangue e tendo em consideraçÁo as equaçÓes 9 e 10, a actividade detectada na ROI é

$$a'_h(t) = \beta A_p e^{-k_1 t} + \beta A_h e^{-k_2 t} + \gamma A_0 e^{-k_1 t} . \quad (15)$$

Isto é,

$$a'_h(t) = A'_p e^{-k_1 t} + A'_h e^{-k_2 t} \quad (16)$$

onde

$$A'_p = \beta \frac{k_1 A_0}{k_2 - k_1} + \gamma A_0 , \quad (17)$$

e

$$A'_h = -\beta \frac{k_1 A_0}{k_2 - k_1} . \quad (18)$$

Admitindo que a teoria dos sistemas é aplicÁvel, a funçÁo $a_p(t)$ pode ser considerada como uma funçÁo de entrada no sistema e a funçÁo $a'_h(t)$ a resposta do sistema a esta funçÁo. A funçÁo de resposta $a'_h(t)$ pode também ser considerada como o resultado da convoluçÁo de $a'_p(t)$ com uma funçÁo de retençÁo, $h(t)$, isto é, a funçÁo que representa a curva hepÁtica teÓrica de actividade/tempo apÓs a injeçÁo instantÁnea de um radiofÁrmaco na artéria hepÁtica (ou veia porta), sem recirculaçÁo. Desta forma,

$$a'_h(t) = a'_p(t) * h(t) . \quad (19)$$

A funçÁo de retençÁo $h(t)$ é caracterÍstica do sistema e independente de $a'_p(t)$. Para conhecer $h(t)$ tem de ser realizada a desconvoluçÁo da equaçÁo 19. Aplicando transformadas de Laplace a esta equaçÁo chegamos a:

$$A_h(s) = A_p(s) H(s) , \quad (20)$$

onde $A_h(s)$, $A_p(s)$ e $H(s)$ sÁo as transformadas de Laplace de $a'_h(t)$, $a'_p(t)$ e $h(t)$ respectivamente.

Por conseguinte,

$$H(s) = Ah(s) / Ap(s) . \quad (21)$$

Os cálculos das transformadas de Laplace de $A_p(s)$ e $A_h(s)$ e a sua substituição na equação 21 conduzem a:

$$H(s) = \frac{A'_p}{A'_0} + \frac{A'_h}{A'_0} k_1 \frac{1}{s+k_2} + \frac{A'_h}{A'_0} \frac{s}{s+k_2} \quad (22)$$

A função $h(t)$ é obtida pela transformação inversa de $H(s)$, isto é

$$h(t) = H^{-1}(s) , \quad (23)$$

o que leva a

$$h(t) = \left[\frac{A'_p}{A'_0} + \frac{A'_h}{A'_0} \right] \delta(t) + \left[-\frac{A'_h}{A'_0} (k_2 - k_1) \right] e^{-k_2 t} . \quad (24)$$

Após a substituição de A'_p , A'_h e A'_0 pelas suas próprias expressões e fazendo

$$A = \frac{\gamma}{\alpha}$$

e

$$B = \frac{\beta k_1}{\alpha} , \quad (25)$$

estabelece-se que

$$h(t) = A\delta(t) + Be^{-k_2 t} . \quad (26)$$

A equação 26 é a soma de uma função exponencial com uma curva de diluição de área A , situada na origem. A representação gráfica desta função é apresentada na figura 2.

A relação « r » entre a área da função delta e o valor da função $Be^{-k_2 t}$ para $t = 0$ é

$$r = \frac{\gamma}{\beta k_1} \quad (27)$$

O valor de γ que corresponde à fracção do total da actividade do sangue detectada simultaneamente com a actividade nos hepatócitos na ROI hepática é dado por

$$\gamma = r \beta k_1 . \quad (28)$$

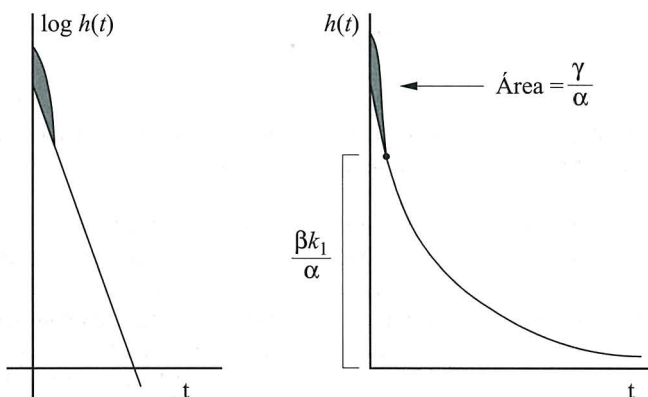


Fig. 2 — Representação gráfica da função $h(t)$ em forma semi-logaritmica e linear

A constante β representa a eficiência da detecção de radioactividade no parênquima do fígado, que se considera unitária (α e γ são valores relativos), vindo então,

$$\gamma = r k_1 \quad (29)$$

Na prática, o valor de γ resulta quase exclusivamente da actividade do sangue hepático, dado que, no campo de visão do detector, a contribuição da actividade do sangue nas estruturas localizadas fora do fígado é pequena. Então, γ (e também « r ») depende da capacidade dos hepatócitos de captarem os radiofármacos. Este método pode ser uma ferramenta útil para a quantificação da insuficiência hepática.

O método em análise tem algumas semelhanças com um método não paramétrico para a quantificação de cintigrafias hepatobiliares proposto por Juni (Juni J. E. *et al.*, 1979; Lima J. M. C. P., 1997 e Tagge E. *et al.* 1987): utiliza-se também a técnica matemática da desconvolução e obtém-se uma função de retenção semelhante. Este método não se baseia, contudo, na modelação por compartimentos.

A aplicação do método proposto para a quantificação de cintigrafias hepatobiliares pode fornecer informações acerca do fluxo do sangue hepático e da capacidade funcional dos hepatócitos. Apesar de a situação real ser por vezes diferente das condições ideais consideradas no modelo teórico e de diversos factores restringirem a utilização clínica do método, a sua aplicabilidade como ferramenta de diagnóstico na medicina nuclear tem sido comprovada com sucesso.

Referências

- Alderson, P. O., Kenneth, H. D., Mendenhall, K. G., Guadiani, V. A., Watson, D. C., Links, J. M., Wagner, Jr. H. N., (1979) – Deconvolution analysis in radio-nuclide quantitation of left-to-right cardiac shunts. *Eur. J. Nucl. Med.*, 20: 502-506.
- Brendel, A. J., Commenges, D., Salamon, R., Ducassou, D., Blanquet, P., (1983) – Deconvolution analysis of radionuclide angiography curves: problems arising from fragmented bolus injections. *Eur. J. Nucl. Med.*, 8: 93-98
- Britton, K. E., Granowska, M., Nimon, C. C., (1980) – Total and regional cerebral blood flow — a new quantitative non-invasive method for cerebrovascular disease. In: *Medical Radionuclide imaging*; IAEA Vienna. 1981; II: p. 315.
- Britton, K. E., Nimon, C. C., Whitfield, H. N., Fry, I. K., Hendry, W. F., Wickham, J. E. A., (1980) – The evaluation of obstructive nephropathy by means of parenchymal retention functions. In: *Radionuclide Nephrology*. George Thieme Verlag.: 164.
- Brown, H. P., Juni, J. J., Lieberman, D. A., Krishnamurthy, G. T., (1988) – Hepatocyte versus biliary disease: A distinction by deconvolution analysis of ^{99m}Tc IDA time-activity curves. *Eur. J. Nucl. Med.*; 29: 623-630.
- De Lima, J. J. P., (1995) – Image processing in nuclear pneumology. *Computer methods and programs in biomedicine*; 48: 7-14
- De Lima, J. J. P., (1984) – Applications of convolution and deconvolution methods in nuclear medicine. In *Technical advances in biomedical physics* (ed.: Dendy P. P., Ernest D. W. and Sengun A.). NATO ASI Series nº 77. Pp. 337-344.
- De Lima, J. J. P., (1988) – New possibilities of computer studies in nuclear medicine in *Progress in computer-assisted function analysis* (ed.: Willems J. L., van Bommel and Michel J.). Elsevier. Amsterdão. Pp. 409-414.
- De Lima, J. J. P., (1996) – Nuclear medicine and mathematics. *Eur. J. Nucl. Med.* 23;6: 705-719, 14.
- Floyd, C. E., Jaszczak, R. J., Greer, K. J., Coleman, R. E., (1985) – Deconvolution of Compton scatter in SPECT. *Eur. J. Nucl. Med.*; 26: 403-408.
- Juni, J. E., (1994) – Scintigraphic measurement of hepatic function in *Nuclear medicine in clinical diagnosis and treatment* (Eds. Murray I. P., Ell P. J.). Churchill Livingstone. Nova Iorque. Pp. 439-444.
- Juni, J. E., Thrall, J. H., Froelich, J. W., *et al.* (1988) – The appended curve technique for deconvolutional analysis – method and validation. *Eur. J. Nucl. Med.* 14: 403-407
- Levin, S. G., (1979) – Understanding and using statistics in nuclear medicine. *Eur. J. Nucl. Med.*; 20: 550-558.
- Lima, J. M. C. P., (1997) – Quantificação em hepatologia nuclear. Dissertação de doutoramento. Univ. de Coimbra.

- Nimon, C. C., Lee, T. Y., Britton, K. E., Granowska, M., Gruenewald, S., (1981)
– Practical applications of deconvolution techniques in dynamic studies.
In: *Medical Radionuclide imaging*. IAEA Vienna; I: pp. 367, 1980.
- Stritzke, P., (1990) – Deconvolution using orthogonal polynomials in Nuclear
Medicine. *IEEE Trans. on Med Imag*; 9: 11-23
- Tagge, E., Campbell, Jr. D. A., Reichelle, R., *et al.* (1987) – Quantitative
scintigraphy with deconvolutional analysis for the dynamic measurement
of hepatic function. *J. of Surgical Research* 42: 605-612

(Página deixada propositadamente em branco)

A estrutura óssea do fémur

1. RESUMO

O esqueleto dos vertebrados é essencialmente constituído por dois tipos de estrutura óssea: osso cortical — denso e compacto; osso trabecular — poroso e esponjoso.

Em 1892, o fisiologista alemão Julius Wolff propôs uma explicação para a distribuição destes dois tipos de estrutura designada, actualmente, por *lei de Wolff*.

A ideia subjacente consiste numa visão dinâmica da estrutura óssea como consequência da sua adaptabilidade às diversas solicitações externas. Num local onde as tensões mecânicas passem a ser mais elevadas existirá deposição de matéria óssea, enquanto num outro, onde a partir de determinado momento as tensões diminuem substancialmente passará a existir absorção de matéria óssea. A este processo de absorção/deposição de matéria óssea dá-se o nome de remodelação óssea.

Possuir um modelo fiável de remodelação óssea é da maior importância no caso dos implantes ortopédicos, do tratamento de fracturas, da biomecânica desportiva e da prevenção da osteoporose, do tratamento de assimetrias ósseas durante o crescimento, etc.

A lei de Wolff afirma ainda que, *perante uma mudança de estímulos exteriores, a remodelação se dá segundo direcções privilegiadas associadas às direcções de maior tensão mecânica*. Esta afirmação tem conduzido à elaboração dos mais variados modelos analíticos e empíricos nos últimos cem anos.

Com o advento dos grandes meios de computação e o desenvolvimento de conceitos matemáticos associados à otimização de estruturas foi possível começar a ter uma maior compreensão do processo mecânico de remodelação óssea e, simultaneamente, generalizar a maioria dos modelos propostos neste último século. Nestes novos modelos, a lei de Wolff surge, naturalmente, associada a condições necessárias de estacionaridade de determinados funcionais de energia.

A discretização destes modelos matemáticos tem conduzido a simulações numéricas que, além de permitirem uma melhor compreensão do fenómeno da remodelação óssea, começam a desempenhar uma enorme ajuda na prática clínica.

Neste artigo ilustrar-se-ão estes conceitos aplicando-os ao estudo da estrutura óssea do fémur.

2. OPTIMIZAÇÃO DE ESTRUTURAS. CASO UNIDIMENSIONAL

Considere-se uma barra ocupando o intervalo $]0, L[\subset \mathbb{R}$ de secção de área A fixa num dos extremos, por exemplo em $x = 0$, e solicitada por uma força aplicada F , no outro extremo, localizado em $x = L$.

Considere-se que o material de que a barra é feita possa possuir uma microestrutura formada pela combinação de dois materiais-base, homogêneos e isótropos, de módulo de elasticidade E^+ e E^- , nas proporções γ e $(1 - \gamma)$, respectivamente, com

$$0 \leq \gamma \leq 1 . \quad (1)$$

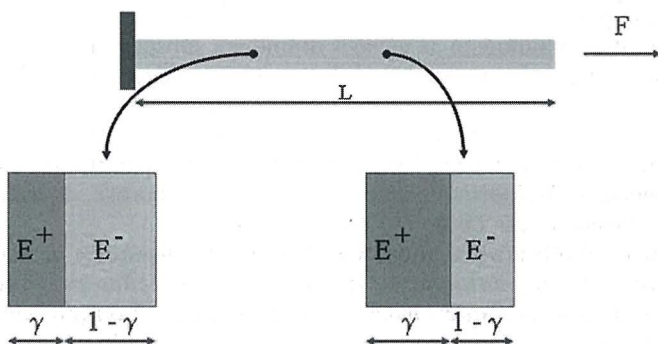


Fig. 1 — Exemplificação do tipo de microestrutura da barra

A massa específica de cada um destes materiais será designada por ρ^+ e ρ^- , respectivamente. Por simplicidade consideram-se estes dois materiais *bem ordenados*, isto é:

$$E^+ > E^-, \quad \rho^+ > \rho^- . \quad (2)$$

Nos problemas de optimização de topologia é usual considerar a minimização do trabalho das forças aplicadas, isto é, calcular a distribuição da fracção volúmica γ , de modo a que a grandeza $Fu(L)$ seja o menor possível, onde u designa deslocamento axial da barra (*cf.* Bendsøe, M. P. e Kikuchi, N., 1988 e Bendsøe, M. P., 1995). Sem mais restrições, a solução para este problema é dada por $\gamma = 1$. Consequentemente, toda a barra seria constituída pelo material mais forte, E^+ . Porém, se houver um custo em adicionar material E^+ , poderá acontecer que este seja proibitivo e que, em algumas partes, ou no seu todo, a barra tenha de ser constituída também por material E^- , numa proporção a determinar. Além disso, não há razão para esperar que o valor de γ numa determinada secção seja o mesmo que noutra secção, isto é, poderá ter-se $\gamma = \gamma(x)$.

Considere-se então o problema de minimizar o funcional

$$I(\gamma) = F u(L) + k \int_0^L \rho^H A , \quad (3)$$

onde as unidades da constante k são trabalho/massa representando, fisicamente, o trabalho necessário realizar para adicionar ao sistema uma unidade de massa. A grandeza ρ^H representa a densidade macroscópica (homogeneizada) da barra formada pelos dois materiais em questão e será uma função de γ .

Para que a estrutura esteja em equilíbrio, o campo de deslocamentos u tem de verificar as equações de equilíbrio, que escreveremos na forma variacional, isto é, $u \in V = \{v \in H^1(0, L) : v(0) = 0\}$:

$$\int_0^L E^H A u' v' dx - F v(L) = 0, \quad \forall v \in V, \quad (4)$$

onde E^H representa o módulo de elasticidade macroscópico (homogeneizado) do material de que é feita a barra e o símbolo «'» designam a derivada em ordem à variável x .

A utilização da teoria da homogeneização é uma das várias possibilidades para representar a influência da microestrutura nas equações de equilíbrio, obtidas a partir da mecânica dos meios contínuos. Trata-se de uma das teorias mais apropriadas a uma análise matemática do problema (*cf.* Bendsøe, M. P., e Kikuchi, N., 1988; Bendsøe, M. P., 1995; Guedes, J. M., e Kikuchi, N., 1990; e Murat, F., e Tartar, L., 1997).

Resumindo, o problema de otimização de topologia é constituído pela minimização do funcional (3), sujeito às equações (1) e (4) e, no caso presente, para materiais verificando (2). Para resolver este problema pode considerar-se o lagrangiano:

$$L = F u(L) + k \int_0^L \rho^H A dx + \lambda \left[\int_0^L E^H A u v dx - F v(L) \right] dx + \int_0^L \tau^+ (\gamma - 1) dx + \int_0^L \tau^- \gamma dx \quad (5)$$

onde λ , τ^+ e τ^- são multiplicadores de Lagrange associados aos constrangimentos (4), $\gamma \leq 1$ e $\gamma \geq 0$, respectivamente.

Das condições necessárias de estacionaridade do lagrangiano L obtém-se, para além de (4) e para $x \in]0, L[$:

$$v = -\frac{1}{\lambda} u, \quad (6)$$

$$k \frac{dp^H}{d\gamma} A - \frac{dE^H}{d\gamma} A |u|^2 + \tau^+ - \tau^- = 0 \quad (7)$$

$$\tau^+ \geq 0, \quad (8)$$

$$\tau^- \geq 0, \quad (9)$$

$$\tau^+ (\gamma - 1) = 0 \quad (10)$$

$$\tau^- \gamma = 0 \quad (11)$$

Até este momento não se especificou o tipo de microestrutura da barra. Considere-se, sem perda de generalidade, o caso mais simples de uma microestrutura laminada, formada pelos materiais E^+ e E^- orientados segundo um sistema de eixos paralelo a Ox_1 e a Ox_2 , nas proporções γ e $(1-\gamma)$, respectivamente (cf. figura 1).

Da teoria da homogeneização tem-se (cf. Bendsøe, M. P., 1995; Guedes, J. M., e Kikuchi, N., 1990; e Murat, F., e Tartar, L., 1997):

$$p^H = \gamma p^+ + (1-\gamma) p^-, \quad \frac{1}{E^H} = \frac{\gamma}{E^+} + \frac{1-\gamma}{E^-}, \quad (12)$$

Assim, a equação (7) toma a forma:

$$k(\rho^+ - \rho^-) A - \frac{E^+ E^- (E^+ - E^-)}{[\gamma E^- + (1-\gamma) E^+]^2} |u|^2 A + \tau^+ - \tau^- = 0.$$

Da equação de equilíbrio obtém-se $(E^H A u')^2 = F^2$, o que, substituindo na equação anterior, conduz a

$$\underbrace{k(\rho^+ - \rho^-)A - F^2}_{x} \frac{(E^+ - E^-)}{AE^+E^-} = -\tau^+ + \tau^- \quad (13)$$

Juntamente com (8)-(11), esta é a equação que conduz à solução do problema. De facto, se o custo (k) for muito baixo, tem-se $\chi < 0$ e, consequentemente, $\tau^- = 0$, $\tau^+ > 0$, o que implica, de (10), que $\gamma = 1$. Por outras palavras, se o custo de adicionar material for muito baixo, a barra será formada pelo material mais rígido (geralmente mais oneroso), de módulo de elasticidade E^+ . Se o custo de adicionar material for elevado, tem-se $\chi > 0$ e, consequentemente, $\tau^- > 0$, $\tau^+ = 0$, o que implica, de (11), que $\gamma = 0$. Isto é, a barra será formada pelo material mais fraco, de módulo de elasticidade E^- .

Resumindo, tem-se:

$$\chi > 0 \Rightarrow \tau^+ = 0, \tau^- > 0 \Rightarrow \gamma = 0 \Rightarrow E^H = E^-, \quad (14)$$

$$\chi < 0 \Rightarrow \tau^+ > 0, \tau^- = 0 \Rightarrow \gamma = 1 \Rightarrow E^H = E^+ \quad (15)$$

$$\chi = 0 \Rightarrow \tau^+ = 0, \tau^- = 0 \Rightarrow 0 \leq \gamma \leq 1 \Rightarrow E^H = \frac{1}{\frac{\gamma}{E^+} + \frac{1-\gamma}{E^-}} \quad (16)$$

ou, de forma equivalente:

$$F^2 < k(\rho^+ - \rho^-) A^2 \frac{E^+E^-}{E^+ - E^-} \Rightarrow \gamma = 0 \Rightarrow E^H = E^-, \quad (17)$$

$$F^2 > k(\rho^+ - \rho^-) A^2 \frac{E^+E^-}{E^+ - E^-} \Rightarrow \gamma = 1 \Rightarrow E^H = E^+, \quad (18)$$

$$F^2 = k(\rho^+ - \rho^-) A^2 \frac{E^+E^-}{E^+ - E^-} \Rightarrow 0 \leq \gamma \leq 1 \Rightarrow E^H = \frac{1}{\frac{\gamma}{E^+} + \frac{1-\gamma}{E^-}}, \quad (19)$$

Verifica-se assim que, quando a força aplicada é tal que

$$F^2 = k(\rho^+ - \rho^-) A^2 \frac{E^+E^-}{E^+ - E^-}, \quad (20)$$

a solução não é única, podendo a barra ser formada por qualquer combinação dos dois materiais.

Em conclusão, à medida que se aumenta a força, mantendo o custo (k) fixo, (ou à medida que se diminui o custo mantendo a força F fixa), a solução do problema é dada por uma barra constituída pelo material mais fraco de propriedades (E^- , ρ^-) até que se atinja a relação (19), onde existirá uma multiplicidade de soluções caracterizada por $0 \leq \gamma \leq 1$, para se passar a ter uma barra constituída apenas pelo material de propriedades (E^+ , ρ^+) quando a força aumentar ainda mais.

De um modo empírico, pode dizer-se que à medida que as solicitações exteriores variam, o material de que a barra é feita se adapta para suportar as novas solicitações. É este mesmo tipo de análise, envolvendo a minimização do trabalho das forças aplicadas juntamente com o custo de adicionar material e utilizando a teoria da homogeneização para representar a influência da microestrutura nas equações de equilíbrio, obtidas através da mecânica dos meios contínuos, que nos dará uma compreensão da estrutura óssea do fémur.

2. A ESTRUTURA ÓSSEA DO FÉMUR

A procura de modelos capazes de reproduzir e explicar, de um ponto de vista mecânico, o processo de remodelação óssea começou com o trabalho pioneiro do fisiologista alemão Julius Wolff. De um ponto de vista estrutural, o osso é extremamente complexo. No fémur podem distinguir-se essencialmente três zonas: uma formada por *osso cortical*, denso e compacto (cerca de 85% da matéria óssea); outra caracterizada pela ausência de material ósseo e onde se inserem vários sistemas e tecidos biológicos que não contribuem, de forma substancial, para a resistência mecânica do fémur e uma terceira zona formada por material esponjoso e poroso designada por *zona trabecular* (cf. figura 2).

Em alguns pontos desta região trabecular poderá o leitor observar que a orientação das trabéculas parece ser bem determinada. Foi esta observação que permitiu a Julius Wolff elaborar o que ficou conhecido como lei de Wolff e que afirma que, *perante uma mudança de estímulos exteriores, a remodelação se dá segundo direcções privilegiadas associadas às direcções de maior tensão mecânica* (cf. figuras 2 e 3).

Durante o século XX vários foram os modelos que surgiram para explicar os fenómenos de remodelação óssea. Alguns baseiam-se em factos experimentais, outros em considerações empíricas, mas quase todos consistem numa equação diferencial ordinária cuja variável independente é a fracção volúmica (ou a massa específica) e a variável dependente o tempo. Porém, a completa compreensão e descrição do fenómeno da remodelação óssea é ainda um problema em aberto.

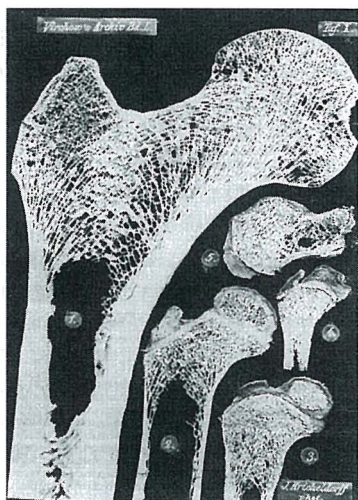


Fig. 2 — Fotografia da estrutura do fêmur apresentada no artigo de J. Wolff de 1870

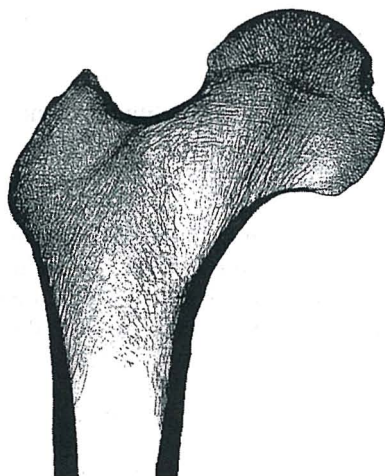


Fig. 3 — Radiografia da cabeça do fêmur

Nos últimos anos pensou-se que, do ponto de vista macroscópico, o fenômeno de remodelação óssea poderia ser visto como um processo de otimização de uma estrutura. O funcional a minimizar seria o trabalho

das forças aplicadas, com um termo de penalização para a adição de massa, e as equações de estado seriam as equações de equilíbrio fornecidas pela teoria da elasticidade linearizada. A influência da microestrutura seria modelada utilizando a teoria da homogeneização. Isto é, trata-se de uma generalização directa, ao caso tridimensional, do anteriormente exposto.

Mais especificamente, seja Ω um subconjunto de R^3 , aberto, limitado, conexo. Seja $x=(x_i)$, $1 \leq i \leq 3$, um ponto genérico de Ω . Seja $\partial\Omega = \overline{\Gamma_0} \cup \overline{\Gamma_1}$ com $\overline{\Gamma_0} \cap \overline{\Gamma_1} = \emptyset$ sua fronteira. Considere-se um corpo que ocupa o volume Ω e está submetido a um sistema de forças interiores, por unidade de volume, dadas por $f = (f_i(x))$, $1 \leq i \leq 3$; e a um sistema de forças de superfície, $F = (F_i(x))$, $1 \leq i \leq 3$, por unidade de área, em Γ_1 . Designe $n = (n_i)$, $1 \leq i \leq 3$, a normal unitária exterior a $\partial\Omega$ e $u = (u_i)$, $1 \leq i \leq 3$, o campo de deslocamentos. É possível demonstrar que, neste caso, o estado de tensão num ponto $x \in \Omega$ fica completamente caracterizado pelo tensor das tensões de Cauchy, $\sigma(u) = (\sigma_{ij}(u))$, $1 \leq i, j \leq 3$; o estado de extensão num ponto $x \in \Omega$ fica completamente caracterizado pelo tensor das extensões linearizado, $e(u) = (e_{ij}(u))$, $1 \leq i, j \leq 3$, onde:

$$e_{ij}(u) = \frac{1}{2} \left(\frac{\partial u_j}{\partial x_i} + \frac{\partial u_i}{\partial x_j} \right), \text{ para todo } 1 \leq i, j \leq 3. \quad (21)$$

Neste caso a lei de Hooke (lei constitutiva) toma a forma geral:

$$\sigma_{ij}(u) = a_{ijkm} e_{km}(u), \quad (22)$$

onde $a_{ijkm} = (a_{ijkm}(x))$ designam as componentes de um tensor de quarta ordem e onde se utilizou a convenção de soma no índice repetido.

Considerando que na parte da fronteira Γ_0 o corpo está fixo, a teoria da elasticidade linearizada diz-nos que o equilíbrio do corpo é regido pelo sistema seguinte:

$$-\partial_j \sigma_{ij} = f_i, \text{ em } \Omega, \quad (23)$$

$$u_i = 0, \text{ sobre } \Gamma_1, \quad (24)$$

$$\sigma_{ij} n_j = F_i, \text{ sobre } \Gamma_2, \quad (25)$$

onde ∂_j designa a derivada parcial em relação à variável x_j e onde, novamente, se utilizou a convenção de soma no índice repetido. Trata-se de um sistema de três equações diferenciais com derivadas parciais, do

tipo elíptico, para as três componentes do campo de deslocamentos $u_i(x)$, $1 \leq i \leq 3$. Para o leitor ter uma ideia da sua complexidade escrevem-se as equações por extenso. As equações correspondentes a (23) têm a forma:

$$-\frac{\partial \sigma_{11}}{\partial x_1} - \frac{\partial \sigma_{12}}{\partial x_2} - \frac{\partial \sigma_{13}}{\partial x_3} = f_1 \quad (26)$$

$$-\frac{\partial \sigma_{21}}{\partial x_1} - \frac{\partial \sigma_{22}}{\partial x_2} - \frac{\partial \sigma_{23}}{\partial x_3} = f_2 \quad (27)$$

$$-\frac{\partial \sigma_{31}}{\partial x_1} - \frac{\partial \sigma_{32}}{\partial x_2} - \frac{\partial \sigma_{33}}{\partial x_3} = f_3 \quad (28)$$

e as correspondentes a (25):

$$\sigma_{11}n_1 + \sigma_{12}n_2 + \sigma_{13}n_3 = F_1, \quad (29)$$

$$\sigma_{21}n_1 + \sigma_{22}n_2 + \sigma_{23}n_3 = F_2, \quad (30)$$

$$\sigma_{31}n_1 + \sigma_{32}n_2 + \sigma_{33}n_3 = F_3, \quad (31)$$

onde se deve substituir a lei de Hooke generalizada (22). No caso unidimensional estas equações reduzem-se a

$$-(E^H Au')' = f = 0, \quad \text{em }]0, L[, \quad (32)$$

$$u(0) = 0, \quad (33)$$

$$(E^H Au') = F, \quad \text{no extremo } x = L, \quad (34)$$

respectivamente, cuja formulação variacional é dada por (4). Neste mesmo caso, a lei de Hooke tomou a forma particular seguinte:

$$\sigma_{11} = E^H e_{11}(u) = E^H u', \quad (35)$$

onde se considerou que a direcção axial da barra coincidia com a do eixo das abcissas Ox_1 .

Assim, no caso tridimensional, o funcional a minimizar toma a forma:

$$\int_{\Omega} f_i u_i dx + \int_{\Gamma_1} F_i u_i ds + k \int_{\Omega} \rho^H, \quad (36)$$

sujeito à equação de estado seguinte:

$$\int_{\Omega} E_{ijkm}^H e_{km}(u) e_{ij}(v) dx = \int_{\Omega} f_i v_i dx + \int_{\Gamma_1} F_i v_i ds, \quad \forall v \in V, \quad (37)$$

onde agora $V = \left\{ v \in [H^1(\Omega)]^3 : v = 0 \text{ sobre } \Gamma_0 \right\}$, e que não é mais do que a versão variacional das equações de equilíbrio (23)-(25).

O lagrangiano toma agora a forma:

$$\begin{aligned} L = & \int_{\Omega} f_i u_i dx + \int_{\Gamma_1} F_i u_i ds + k \int_{\Omega} \rho^H + \\ & + \lambda \left[\int_{\Omega} E_{ijkm}^H e_{km}(u) e_{ij}(v) dx = \int_{\Omega} f_i v_i dx + \int_{\Gamma_1} F_i v_i ds \right] + \\ & + \int_{\Omega} \tau^+ (\gamma - 1) dx + \int_{\Omega} \tau^- \gamma dx, \end{aligned} \quad (38)$$

Das condições necessárias de estacionaridade obtêm-se as equações (6), (8)-(11), (37), sendo a equação (7) substituída por:

$$k \frac{d\rho^H}{d\gamma} - \frac{dE_{ijkm}^H}{d\gamma} e_{km}(u) e_{ij}(u) + \tau^+ - \tau^- = 0, \quad (39)$$

Torna-se agora necessário calcular as derivadas em ordem à fracção volúmica γ da densidade e coeficientes de elasticidade homogeneizados, ρ^H e E_{ijkm}^H , respectivamente. Para laminados simples, formados por materiais com propriedades elásticas E_{ijkm}^+ e E_{ijkm}^- , nas proporções γ e $1 - \gamma$ respectivamente, tem-se, para o caso bidimensional:

$$E_{1111}^H = \frac{E_{1111}^+ E_{1111}^-}{\gamma E_{1111}^- + (1 - \gamma) E_{1111}^+},$$

$$E_{1212}^H = \frac{E_{1212}^+ E_{1212}^-}{\gamma E_{1212}^- + (1 - \gamma) E_{1212}^+},$$

$$E_{1122}^H = \left[\gamma \frac{E_{1122}^+}{E_{1111}^+} + (1 - \gamma) \frac{E_{1122}^-}{E_{1111}^-} \right] \frac{E_{1111}^+ E_{1111}^-}{\gamma E_{1111}^- + (1 - \gamma) E_{1111}^+},$$

$$\begin{aligned} E_{2222}^H = & \gamma E_{2222}^+ + (1 - \gamma) E_{2222}^- - \left[\gamma \frac{(E_{1122}^+)^2}{E_{1111}^+} + (1 - \gamma) \frac{(E_{1122}^-)^2}{E_{1111}^-} \right] + \\ & + \left[\gamma \frac{E_{1122}^+}{E_{1111}^+} + (1 - \gamma) \frac{E_{1122}^-}{E_{1111}^-} \right]^2 \frac{E_{1111}^+ E_{1111}^-}{\gamma E_{1111}^- + (1 - \gamma) E_{1111}^+}. \end{aligned}$$

No caso de tensão plana, estas expressões simplificam-se e tem-se:

$$E_{1111}^H = \frac{1}{1-\nu^2} I_1,$$

$$E_{1212}^H = \frac{1}{2(1+\nu)} I_1,$$

$$E_{1122}^H = \frac{\nu}{1-\nu^2} I_1,$$

$$E_{2222}^H = I_2 \frac{\nu^2}{1-\nu^2} I_1,$$

onde

$$I_1 = \frac{E^+ E^-}{\gamma E^- + (1-\gamma) E^+}, \quad I_2 = \gamma E^+ + (1-\gamma) E^-,$$

e onde ν designa o coeficiente de Poisson, que se supõe o mesmo para ambos os materiais. Trata-se da generalização das equações (12), onde ρ^H mantém a mesma forma.

A dependência destas expressões em γ é simples, pelo que as derivadas em (39) são de cálculo imediato. No caso de se optar por uma representação da microestrutura mais complexa, permitindo, por exemplo, a rotação da microestrutura, o que se torna necessário para análise do fémur, há que recorrer às técnicas de derivação numérica, mas as ideias fundamentais são as expostas (*cf.* Bendsøe, M. P., e Kikuchi, N., 1988; Bendsøe, M. P., 1995; Murat, F., e Tartar, L., 1997; e Fernandes, P. R., Folgado, J., e Miranda, P. S., 2000).

3. EXEMPLOS NUMÉRICOS APLICADOS À ANÁLISE DA ESTRUTURA ÓSSEA DO FÉMUR

Utilizando o método dos elementos finitos para a resolução numérica do problema da elasticidade e um algoritmo de optimização para resolver a condição necessária de estacionaridade (39), Fernandes, P., Rodrigues, H., e Jacobs, C., (1999) obtiveram o resultado ilustrado na figura seguinte:

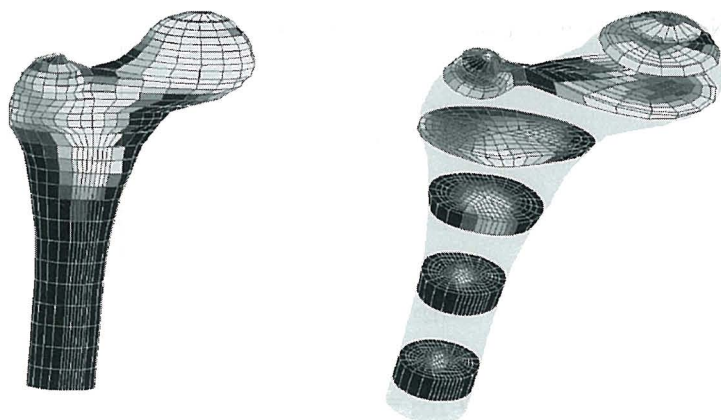


Fig. 4 — Distribuição de densidade no fêmur

A comparação da distribuição, macroscópica, da densidade com uma radiografia está feita na figura seguinte:

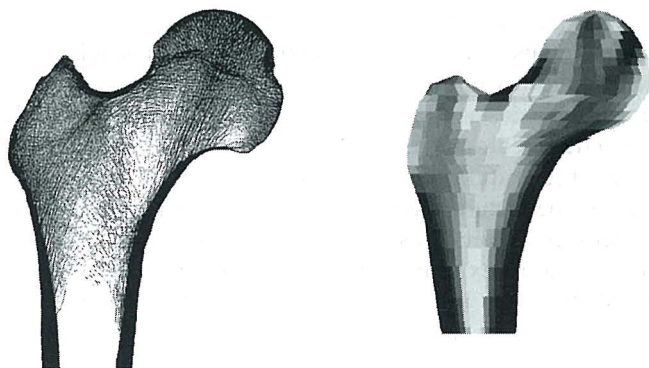


Fig. 5 — Comparação entre o resultado numérico e uma radiografia

Calculando em cada ponto os valores e vectores próprios do tensor das tensões de Cauchy, σ_{ij} , obtêm-se os valores e as direcções de maior tensão mecânica. Representando estas grandezas graficamente, observa-se uma boa correlação com a lei de Wolff, notando-se nitidamente, o desenvolvimento do osso trabecular ao longo das direcções (vectores

próprios) de tensões mais elevadas (valores próprios) do tensor das tensões de Cauchy. Este aspecto teórico e numérico da lei de Wolff encontra-se representado na figura seguinte:

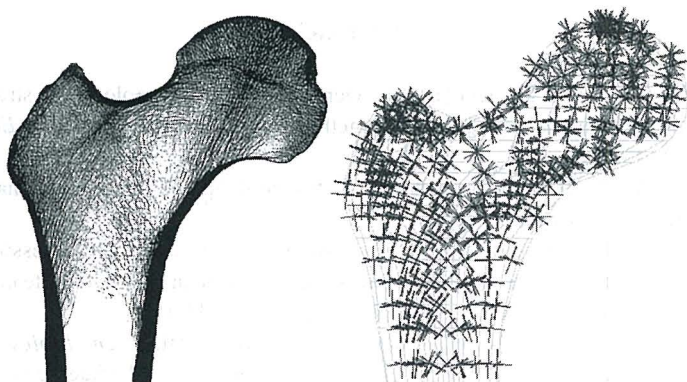


Fig. 6 — Verificação numérica da lei de Wolff

Possuindo um algoritmo com estas características, nada impede que se faça o estudo simultâneo do conjunto osso-prótese. O resultado numérico obtido por Fernandes, Folgado e Miranda, 2000, encontra-se representado na figura seguinte:



Fig. 7 — Distribuição numérica da densidade com e sem prótese

Resultado este de grande utilidade no projecto e desenvolvimento de próteses, bem como no eventual acompanhamento, no futuro, da prática clínica.

Referências

- Bendsøe, M. P., e N. Kikuchi (1988) – Generating optimal topologies in structural design using a homogenization method, *Comp. Meth. Appl. Mech. Engng.*, 71, 192-224.
- Bendsøe, M. P., (1995) – Optimization of structural topology, shape, and material, Berlim, Springer Verlag.
- Guedes, J. M., e N. Kikuchi (1990) – preprocessing and postprocessing for materials based on the homogenization method with adaptive finite element methods, *Comp. Meth. Appl. Sci. Engrn.*, 83, 143-198.
- Murat, F., e L. Tartar (1997) – consultar os vários artigos em *Topics in the Mathematical Modelling of Composite Materials*, A. Cherkaev e B. Khon, eds.; Birhäuser.
- P. Fernandes, H. Rodrigues e C. Jacobs (1999) – *A Model of Bone Adaptation Using a Global Optimisation Criterion Based On the Trajectorial Theory of Wolff*; Computer Methods in Biomechanics and Biomedical Engineering, vol. 2, p. 125-138.
- P. R. Fernandes, J. Folgado e P. S. Miranda (2000) – *Modelos Computacionais para o Projecto de Próteses Ortopédicas*, VI Congresso Nacional de Mecânica Aplicada e Computacional, vol. 1, pp. 135-144.

Guerino Mazzola
University and ETH Zurich, Suíça
Internet Institute for Music Science

Em direcção à «grande unificação» de composição, análise e execução musicais¹

A matemática do século XX deu origem a um processo dramático de reescrita de teorias musicais e correspondente tecnologia da informação. Apresentamos uma visão geral dos modelos de harmonia e contraponto que foram concebidos neste espírito. Baseadas numa teoria matemática da representação unificada de objectos musicais, estas teorias foram implementadas em programas informáticos e aplicadas à análise, composição e execução musicais. Esta representação unificada baseia-se em disciplinas da matemática moderna tais como a teoria dos topos, a lógica e a geometria, bem como em paradigmas de programação orientada para objectos. Os métodos matemáticos utilizados nesta descrição de fenómenos musicais incluem a teoria de módulos, a combinatória, a geometria algébrica, a estatística, os campos de vectores, as equações diferenciais ordinárias, as equações às derivadas parciais e a topologia algébrica. A unificação não só tem em conta a representação de objectos musicais e o tratamento unificado dos processos específicos, como em última análise aponta para uma fusão de verdade e beleza.

¹ Tradução de Carlota Simões, professora do Departamento de Matemática, Universidade de Coimbra.

1. INTRODUÇÃO

«Grande unificação» é uma expressão proveniente da física das partículas e representa a ideia de que três das quatro forças básicas: forte, fraca e interacção electromagnética são descritas como instâncias especiais de uma única super força subjacente². A alusão a este paradigma tem como objectivo estabelecer uma relação entre tendências e resultados até agora considerados campos dispersos e não relacionados da música e da musicologia, tais como

- a representação de objectos musicais de natureza muito diferente, tais como notas de uma partitura, acontecimentos de sons físicos, acordes, motivos, ritmos, formas de natureza mista, agrupamentos de objectos musicais tal como eles intervêm em símbolos de agrupamento utilizados em partituras (ligaduras, por exemplo), e estruturas de execução tais como hierarquias de tempo;
- a teoria de funções de Riemann³ em harmonia e a teoria de contraponto de Fux;
- a análise musical de melodias e ritmos;
- a teoria da execução e da análise musical;
- os espaços de afinação para representação de notas.

Isto apenas para mencionar alguns dos aspectos de unificação mais importantes. Neste contexto, não podemos dar mais que um breve esboço destes desenvolvimentos e remeter para o livro *The Topos of Music – Geometric Logic of Concepts, Theory and Performance* (Mazzola, G., et al., 2001) para uma descrição detalhada.

O objectivo principal do presente artigo é apresentar uma teoria matemática para uma parte profunda da natureza humana interior, nomeadamente a música, em contraste com a descrição matemática habitual da natureza exterior em ciências naturais. Mas esta transferência de métodos clássicos para as humanidades não se restringe à construção de modelos matemáticos; também inclui a metodologia da verificação e falsificação experimentais dos modelos. Isto significa que o processo científico tradicional aplicado nas ciências naturais é reconstruído nas humanidades: experiências de natureza mental tornaram-se uma «alternativa» possível para a física (ver Mazzola, G., et al., 1998, para uma discussão alargada desta extensão metodológica).

² A «teoria do todo» acrescenta ainda a gravidade à grande unificação.

³ Não confundir o musicólogo Hugo Riemann com o matemático Bernhard Riemann!

2. QUANTA PARA MODULAÇÃO TONAL

O modelo de modulação a seguir apresentado é uma primeira construção muito simples que, no entanto, conduz a resultados muito precisos e possui a propriedade típica dos modelos matemáticos em música: permitir uma generalização quase automática de situações para as quais a teoria de música clássica para a qual o modelo foi construído não tem qualquer resposta. No caso da modulação que originalmente foi modelada para escalas maiores, a generalização estende-se a escalas de sete notas arbitrárias. Esta é uma propriedade que tem consequências profundas e que discutiremos para o modelo de contraponto abordado no parágrafo seguinte.

2.1. O princípio antrópico

A propriedade de extensibilidade de um modelo matemático volta a colocar a teoria da música existente — que ele modeliza — num campo de potenciais teorias fictícias. Este facto situa a factualidade histórica numa relação com os potenciais «mundos da música». A justificação puramente histórica das regras de modulação existentes, por exemplo, não constitui por si só uma fundamentação para esta escolha, o que torna esta abordagem puramente histórica uma base de conhecimento limitada: sabemos que é assim, mas não sabemos porquê, nem porque é que outras possibilidades não são exequíveis. Em contraste, a abordagem matemática fornece um campo de teorias potenciais nas quais a actual teoria da música pode ser questionada em relação às suas possíveis propriedades especiais relativamente a variantes não existentes.

Isto evoca a ideia de Leibniz segundo a qual o mundo existente é o melhor de todos os mundos possíveis: será a teoria da música existente a melhor escolha possível? Ou existirá, pelo menos, uma outra distinta? Em cosmologia, esta ideia foi reformulada sob a designação de «princípio antrópico» (Barrow, J., e Tripler, F., 1986). Este princípio afirma que as leis da física são as melhores possíveis para a existência humana. Mais precisamente (e menos radicalmente), o princípio antrópico consiste em estabelecer um teorema que afirma que uma pequena variação nas constantes fundamentais, tais como a constante gravitacional ou a carga eléctrica de prótons e electrões, tornaria impossível qualquer complexidade molecular de ordem superior, tal como a que é necessária para a bioquímica do carbono.

2.2. Esquema de modulação de Schönberg

No seu texto clássico sobre harmonia, Arnold Schönberg (1911) descreveu a modulação tonal como um processo de três fases: a neutralização, o ponto de viragem para a nova tonalidade e a confirmação cadencial (ver figura 1).

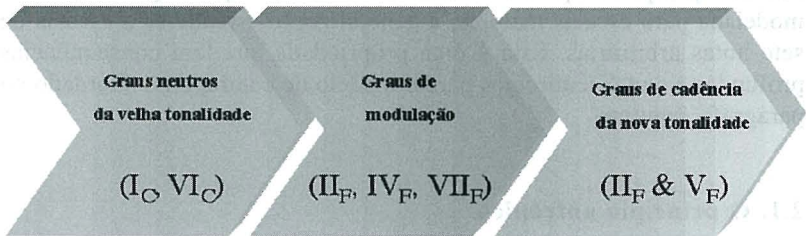


Fig. 1 — As três fases do esquema de modulação de Schönberg no caso de uma modulação de dó maior (C) para fá maior (F)

A afirmação central refere-se à lista de graus de modulação na fase intermédia ou fase do ponto de viragem. Para criar um modelo matemático devem primeiro definir-se os termos musicológicos e os factos de acordo com uma terminologia matemática rigorosa. Deste modo, temos de responder às seguintes questões:

- Qual é o conjunto de tonalidades admitidas?
- O que é um grau?
- O que é uma cadência?
- Qual é o mecanismo de modulação subjacente?
- Como podem os graus de modulação ser deduzidos a partir de uma resposta às questões anteriores?

Começaremos por trabalhar no contexto da afinação bem temperada, e só iremos considerar notas módulo oitava, ou seja, vamos considerar as doze classes de notas cromáticas, fazendo $dó = 0$, $dó \# = ré b = 1$, $ré = 2$, ..., $si = 11$. Deste modo, o conjunto das classes de notas identifica-se com o grupo cíclico Z_{12} . Consideramos doze escalas diatónicas possíveis, $dó = \{0, 2, 4, 5, 7, 9, 11\}$ (ver figura 2) e as respectivas onze escalas transpostas, $fá, si b, mi b, \dots, lá, ré, sol$, no ciclo das quartas (que se obtém por transposições sucessivas por intervalos de quarta, ou seja, cinco unidades).

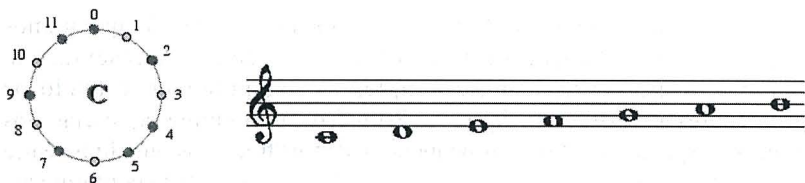


Fig. 2 — A escala de dó maior (C) como um subconjunto de sete elementos do grupo cíclico das doze classes de notas em afinção bem temperada

Para definir uma estrutura de tonalidade numa escala S , consideramos as sete tríades $I_S, II_S, III_S, IV_S, V_S, VI_S, VII_S$, que são conjuntos de três elementos da escala dada, tal como se mostra para a escala de dó maior (C) na figura 3.

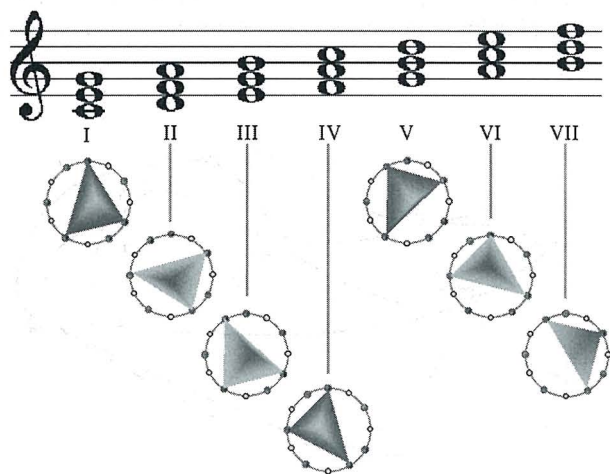


Fig. 3 — As tríades formadas sobre os sete graus da escala de dó maior definem uma cobertura quase geográfica da escala por cartas

Esta cobertura pode ser comparada a um atlas geográfico cujas cartas são os sete graus. Para construir uma representação englobante da sua configuração em sobreposição mútua, a topologia algébrica introduz o

conceito de *nervo*. O nervo desta cobertura é a representação que iremos utilizar na visualização de tonalidade que a seguir apresentamos. O conceito de nervo é muito simples de compreender, estando os pormenores descritos na figura 5. Primeiro distribuímos as cartas dos graus no espaço (os triângulos pequenos), depois traçamos uma linha entre duas cartas quaisquer que tenham intersecção não vazia. Esta configuração gráfica visualiza o que Schönberg chama a «banda harmónica» no seu livro *Harmony* (Schönberg, 1911). Infelizmente, Schönberg não deu um passo em frente que seria decisivo: considerar tríades dos diferentes graus com intersecção não vazia. Desta vez, desenhamos uma superfície triangular entre quaisquer desses três graus. Verificamos então que o objecto gráfico gerado por este processo é uma bem conhecida fita de Moebius, motivo pela qual se designa por «banda harmónica». Será nesta geometria surpreendente da geografia de tríades que iremos basear a definição de tonalidade⁴.

Sem apresentar mais detalhes, convém notar que a ausência de orientação na banda harmónica (não há de facto distinção entre lado da frente e lado de trás!) é responsável por algumas das maiores desvantagens da teoria de funções harmónicas de Hugo Riemann (Mazzola, G., 1990b).

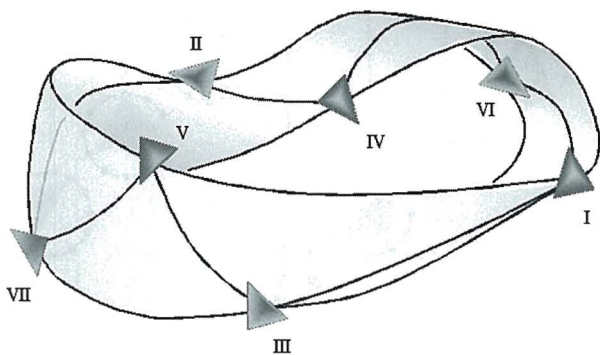


Fig. 4 — A banda harmónica é uma fita de Moebius. Matematicamente, é o nervo da cobertura da escala pelos seus sete graus. O grafo das linhas de ligação (sem as superfícies triangulares) entre os graus é a banda harmónica no sentido apontado por Schönberg (1911)

⁴ Notemos que intersecções não vazias de quatro graus diferentes não podem ocorrer, logo estamos no extremo desta construção de nervo.

Tendo em conta estas definições e um conceito de cadência, relacionado com conjuntos minimais de graus que definem de modo único a tonalidade envolvente, o mecanismo de modulação da tonalidade S para a tonalidade T é definido por uma simetria que transforma S em T. À semelhança do que acontece na física das partículas, materializaremos esta força de transformação por um «*quantum* de modulação» M. Por definição, este é um conjunto de classes de notas, juntamente com uma cobertura por tríades de tal modo que a simetria de modulação é uma simetria interior a M, com algumas propriedades técnicas adicionais. A importância deste *quantum* de modulação reside no facto de ele constituir «*matéria musical*» — tal como os *quanta* da modulação física são forças materializadas —, e por isso preconiza um conjunto de graus que devem ser tocados na parte central do processo de Schoenberg apresentado na figura 1.

O pressuposto central deste modelo de modulação é o seguinte: *os quanta de modulação existem para todos os pares de tonalidades e conduzem exactamente aos graus de modulação descritos por Schönberg (1911)*. A figura 5 mostra um *quantum* M para a modulação de dó para mi b. Reconhecemos as duas fitas harmónicas, a claro, bem como o *quantum* de ligação, a escuro, na representação de nervo. Os graus de modulação são exactamente os vértices de M que se encontram na banda harmónica (quatro graus no exemplo da figura 5) da tonalidade pretendida.

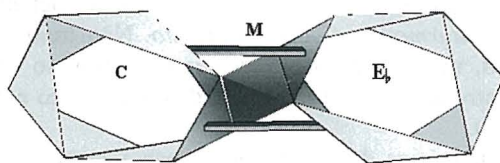


Fig. 5 — Visualização de um quantum de modulação M para a transição de dó para mi b. A representação de nervo de M mostra a estrutura de conexão do meio em cor escura, em particular contém dois simplexes de dimensão 5 (simbolizados por barras horizontais) acima e abaixo do simplex a 3 dimensões (o tetraedro)

Os pressupostos deste modelo e o mecanismo subjacente, nomeadamente a transformação da simetria inerente, podem ser também testados em peças do repertório da música clássica, como, por exemplo, na *Opus 106* de Beethoven. Foi possível obter uma compreensão mais aprofundada dos seus conhecidos planos de modulação, complexos e intrincados,

oscilando alternadamente entre um «mundo» em torno da tonalidade dominante de si *b* maior, e um «anti-mundo» em torno da tonalidade de si menor (teses de Erwin Ratz e Jurgen Uhde, Mazzola, G., 1990b).

O modelo de modulação acima descrito ilustra perfeitamente o «princípio antrópico», um assunto que deriva da modelação matemática. O modelo pode ser alargado a qualquer escala de sete sons e respectivas bandas harmónicas (Muzzolini, D., 1995), o que permite localizar a escala maior num mundo de potenciais variantes. O resultado principal neste contexto variacional é que *a escala maior ocupa uma posição singular na medida em que o seu arsenal de modulação apresenta o mínimo de variedades possíveis entre todas as escalas que admitem modulações entre duas tonalidades quaisquer*. A dominação histórica da escala maior encontra assim uma justificação suplementar num quadro perfeitamente sistemático, justificação essa que a abordagem puramente histórica não pode *a priori* fornecer!

3. CONTRAPONTO VIA DEFORMAÇÕES DO ESPAÇO TANGENTE

O segundo modelo desta análise, o contraponto clássico, é muito importante por três razões. Em primeiro lugar, ilustra uma vez mais o princípio antrópico; em segundo lugar, torna plausível que seja necessário um quadro conceptual poderoso para enquadrar as sistematizações formais heterogéneas no contexto de uma teoria musical matemática, e em terceiro lugar representa um dos mais interessantes assuntos da actual investigação, *a unificação da harmonia e do contraponto*. Não iremos abordar esta última questão, mas deve, contudo, ser mencionado que existe um isomorfismo entre uma estrutura central de contraponto e harmonia (ver Noll, Th., 1995).

A teoria matemática do contraponto baseia-se numa representação alternativa das doze classes de notas do modelo de modulação anterior. Mais precisamente, referimo-nos aos doze intervalos musicais, medidos em múltiplos de passos de meio-tom: unísono = 0, segunda menor = 1, segunda maior = 2, terceira menor = 3, terceira maior = 4, quarta = 5, tritono = 6, quinta = 7, sexta menor = 8, sexta maior = 9, sétima menor = 10 e sétima maior = 11. A oitava é identificada com o unísono módulo oitava, os intervalos maiores que a oitava são identificados com o seu resto módulo oitava, tal como usualmente se faz em contraponto clássico. Assim, o nosso modelo de intervalos é construído a partir da redução módulo oitava. Este conjunto é também visto como o grupo cíclico Z_{12} . No entanto, a sua representação prévia num círculo é um tanto

primitiva, já que não tem em conta a estrutura algébrica relevante. De facto, a chamada decomposição de Sylow (neste caso o teorema de classificação para grupos abelianos finitos) conduz-nos ao produto directo $Z_{12} = Z_3 \times Z_4$. Em termos geométricos, isto significa que temos de pensar num toro (discreto) em vez de num círculo (ver figura 6). Nesta representação, obtêm-se três círculos de terceiras menores, cada um deles a uma terceira maior do seguinte, ou de forma equivalente, quatro círculos de terceiras maiores, cada um deles a uma terceira menor do seguinte. Esta representação algébrica está evidentemente cheia de significado musical, isto é, a decomposição Sylow parece ser a mais adequada para representar classes de intervalos!

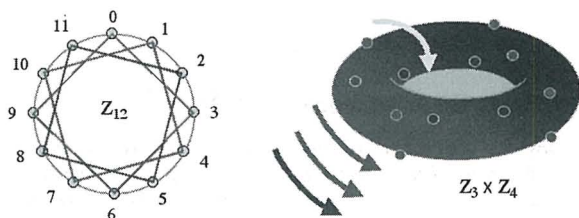


Fig. 6 — O conjunto de intervalos Z_{12} na sua representação Sylow como um produto directo do ciclo de terceiras maiores Z_3 e do ciclo das terceiras menores Z_4

Pode verificar-se que a dicotomia contrapontística dos seis intervalos consonantes $K = \{0, 3, 4, 7, 8, 9\}$ e dos seis intervalos dissonantes $D = \{1, 2, 5, 6, 10, 11\}$ tem uma posição relevante no conjunto de todas as 924^5 dicotomias seis por seis de intervalos (Mazzola, G., 1990b). Note-se que a quarta (intervalo de cinco meios-tons) é dissonante na teoria contrapontística básica, e não consonante, como preconizam os modelos físicos de intervalos. Trata-se de uma diferença importante: uma abordagem genuína da teoria da música não é apenas uma perspectiva física; é um mundo autónomo de entidades simbólicas que privilegia a composição e a construção musicais em vez das restrições acústicas! As propriedades características mais evidentes da dicotomia K/D são apresentadas na figura 7. Nesta figura, as duas metades K e D estão

⁵ Combinações de doze elementos seis a seis.

distribuídas no toro de tal modo que se mantêm separadas uma da outra a uma distância máxima, sob condição de que existe uma única simetria afim, nomeadamente $d= 5c +2$, que transforma K em D e vice-versa.

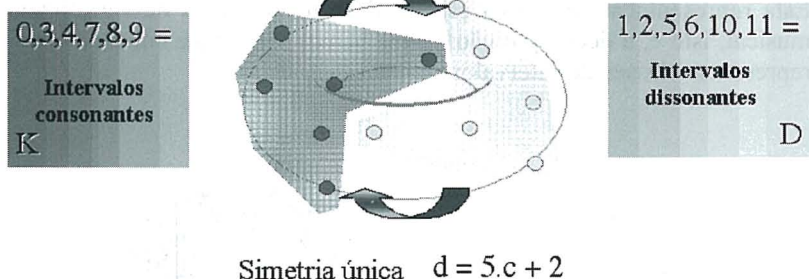


Fig. 7 — No toro dos intervalos, as consonâncias (K) e as dissonâncias (D) estão distribuídas de modo característico e distinto: mantêm-se separadas entre si a uma distância máxima, sob a condição de que exista uma única simetria afim, nomeadamente $d= 5c+2$, que transforma K em D e vice-versa. Entre todas as possíveis 924 dicotomias seis por seis, esta dicotomia K/D pode ser caracterizada por estas propriedades e pela propriedade da parte consonante $7k$ (isto é, quando considerada um conjunto de múltiplos da quinta) ser um monóide multiplicativo

Para além desta representação e caracterização geométricas da medida dos intervalos, a abordagem contrapontística necessita de novos objectos matemáticos: os intervalos vistos como pares, que consistem numa nota de *cantus firmus* a mais uma quantidade de intervalo b , apontando na nota do *discantus*. Formalizemos estes dados numa representação do intervalo pela expressão algébrica $i = a + \epsilon b$. Estritamente falando, referimo-nos ao anel $Z_{12}[\epsilon]$ dos polinómios $f(\epsilon)$ com coeficientes no anel Z_{12} . Assim, intuitivamente, um intervalo $i = a + \epsilon b$ representa o *cantus firmus* de «ponto base» a , juntamente com uma «tangente» infinitesimal b (ver figura 8 para esta geometria tangente).

Os intervalos consonantes definem o subconjunto de 72 elementos $K=Z_{12}+\epsilon K$, enquanto o seu complementar $D=Z_{12}+\epsilon D$ define o conjunto dos intervalos dissonantes. Mais uma vez temos essencialmente apenas uma simetria autocomplementar A: $K \rightarrow D$.

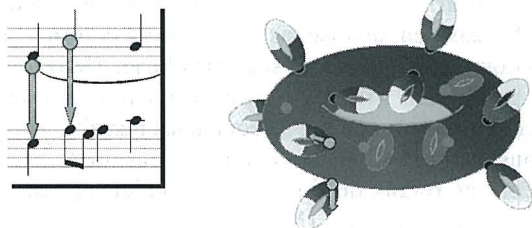


Fig. 8 — Os intervalos contrapontísticos são dados por uma nota base de um cantus firmus, juntamente com uma quantidade intervalar direccionada na nota do discantus. Em termos geométricos, podemos interpretar este facto como sendo um vector tangente no toro das classes de notas, sendo a quantidade intervalar a pequena flecha aplicada no ponto base do cantus firmus (esta representação baseia-se na representação de Sylow dos elementos de Z_{12} , sejam eles classes de notas ou quantidades intervalares). As pontas das setas dos intervalos consonantes estão na metade K do intervalo consonante (metade escura), enquanto as dos intervalos dissonantes estão na metade clara dos pequenos toros

O modelo de contraponto baseia-se na seguinte ideia: em contraponto elementar é apenas permitido passar de um intervalo consonante para outro. Assim, por exemplo, uma sequência $2+\epsilon 3, 4+\epsilon 5$ não é permitida porque termina numa quarta dissonante. A teoria elementar, tal como foi concebida por Johann Joseph Fux no seu famoso *Gradus ad Parnassum* (Fux, J. J., 1725), considera um conjunto de regras que definem restrições adicionais a possíveis sequências de intervalos consonantes, sendo a mais importante a proibição de quintas paralelas, ou seja, sequências do tipo $a+\epsilon 7, b+\epsilon 7$. De acordo com a perspectiva contrapontística de definir sequências de intervalos, devemos tentar estabelecer uma espécie de tensão entre intervalos consonantes e dissonantes quando passamos de um intervalo para o seguinte. Mas, assumindo um ponto de vista contraditório, devemos permanecer simultaneamente na metade consonante! Este conflito pode ser resolvido introduzindo *simetrias de deformação* no espaço tangente $Z_{12}[\epsilon]$. Esta deformação permite-nos «saltar» de uma consonância para outra, mas sob a condição de que a primeira viva na metade consonante deformada, enquanto a outra vive na metade dissonante deformada! Desta forma, *modelamos a tensão contrapontística por uma deformação do espaço tangente*. Tal como para o modelo de modulação, esta ideia é proveniente da física, já que podem surgir forças físicas quando se aplicam simetrias locais a um sistema físico. Isto pode ocorrer, por exemplo, se se distorcer um balão (basta segurá-lo em ambas as mãos, rodá-lo no sentido dos ponteiros do relógio com a mão direita

e no sentido contrário com a mão esquerda), o que produz forças de deformação. No caso em análise, as simetrias de deformação podem ser reconhecidas como sendo de facto simetrias locais (Mazzola, G., 1990b) e, portanto, produzem «forças de deformação». No sentido acima indicado, esta abordagem conduz a sequências permitidas e proibidas de intervalos consonantes (tangentes $a+\epsilon b$) e o resultado é extremamente congruente com as regras de Fux, em particular, aplica-se a regra da proibição das quintas paralelas (Mazzola, G., 1990b).

Mais uma vez, podemos incluir este modelo em generalizações canónicas. As propriedades características de dicotomia consonância/dissonância podem ser variadas e conduzem a outras cinco classes de dicotomias distintas. Em cada uma delas, o contraponto pode ser executado, o que permite afirmar que a *dicotomia clássica consonância/dissonância está situada numa posição globalmente singular* no mundo dos contrapontos potenciais! Mais ainda, e isto cria de certo modo um ambiente mágico, existe uma dicotomia polar I/J contra a dicotomia consonância/dissonância que é correctamente definida pelos seis intervalos próprios a partir da tónica da escala maior. Enquanto a escala maior é a mais adequada para executar o contraponto Fux (K/D), uma escala cujos intervalos estão relacionados com K é a melhor para executar o contraponto I/J. Por estranho que pareça, esta escala está fortemente relacionada com as escalas usadas nas ragas indianas (Hichert, J., 1993). Assim, concluímos que os modelos matemáticos induzem interessantes variações interculturais e extensões do «melhor de todos os possíveis mundos» da música europeia.

Convém acrescentar que o presente modelo de contraponto foi também sujeito a extensos testes empíricos por electroencefalograma profundo EEG (Mazzola, G., *et al.*, 1989), que revelaram uma resposta significativa das regiões límbica e auditiva do cérebro para a simetria autocomplementar.

4. ESPAÇOS CONCEPTUAIS UNIFICADOS E SOFTWARE PARA OBJECTOS MUSICAIS

A variedade de objectos musicais e musicológicos é extremamente rica e diversa. Vimos que diferentes estruturas algébricas e geométricas podem intervir ao nível da harmonia e do contraponto. Quando abandonamos o contexto bem temperado, há ainda outros tipos de módulos que intervêm, e, se tivermos de modelar objectos entre partituras, teoria musical e investigação de execução, a variedade acaba por exceder qualquer modelo

pré-fixado de dados como os que encontramos em sistemas de organização de bases de dados. Para implementar a teoria musical matemática no *software* RUBATO® (Mazzola, G., e Zahorka, O., 1996), desenvolvido no Laboratório Multimédia da Universidade de Zurique, este problema teve de ser abordado de modo genérico. Sem entrar em detalhes, apresentamos aqui as ideias principais. Uma vez que a tarefa de construir um modelo de dados «universal» era de facto enciclopédica, os princípios básicos da ciência enciclopédica (Mazzola, G., *et al.*, 1998), em particular a partir da *Encyclopédie* de Diderot (Auroux, S., 1979), foram considerados e reformulados no contexto do paradigma da programação orientada para objectos. Estes princípios são a unidade, a completude e o discurso. O modelo de dados para um espaço de conceito musical universal implementou (e desenvolveu a um nível matemático teórico, incluindo a teoria de topos, o derradeiro casamento entre a lógica e a geometria (Mac Lane, S., e Moerdijk, I., 1994)) estes princípios da seguinte forma (ver figura 9): a unidade foi traduzida num conceito baseado em recursão, incluindo construção conceptual auto-referencial circular. A condição de completude foi modelada seguindo um arsenal universal de tipos de ramificação (incluindo tudo o que a matemática moderna oferece como construção de objectos universais, ou seja, limites, co-limites e objectos poderosos da teoria do topos). Finalmente, o princípio do discurso foi implementado através da extensibilidade dinâmica e ilimitada do quadro conceptual dado por novas construções, incluindo ordenações canónicas totais entre os objectos existentes.

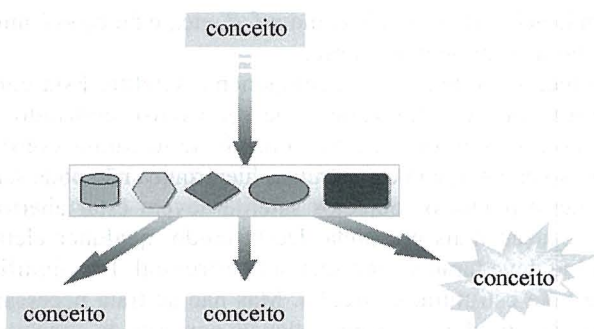


Fig. 9 — No design e na teoria de modelos de dados para conceitos musicais, os princípios enciclopédicos de unidade, completude e discurso são implementados através da recursividade conceptual (1), de tipos de ramificação universal (2) e da extensibilidade dinâmica sem restrições (3) e da ordenação do quadro conceptual

Em vez de abordarmos definições formais, preferimos ilustrar a construção conceptual para um conceito circular musicalmente importante: a nota estrutural (ver figura 10). O problema reside no facto de muitas vezes, em composição musical, análise e representação, as notas não serem meros objectos, mas permanecerem agrupadas a uma encapsulação hierárquica de objectos delegados de *status* hierarquicamente inferior, tal como acontece por exemplo, em ornamentos (como os trilos) ou na análise schenkeriana. Num trilo, é-nos dada uma nota que transporta um grupo de notas ornamentais de carácter secundário e que são apenas uma pequena variação da nota âncora (ver figura 10, em cima à esquerda). Na análise schenkeriana — também corroborada pela abordagem de Jackendoff-Lerdahl (1983) — de estruturas de agrupamento global em partituras musicais, certas partes da partitura são consideradas uma espécie de desdobramento estrutural construído em torno de notas pivotais de modo recursivo.

Para formalizar esta configuração, o conceito de nota estrutural abarca dois constituintes, uma nota âncora e uma componente satélite. Esta primeira ramificação do conceito é uma espécie de produto cartesiano (mais geralmente, um limite) ou, logicamente falando, uma conjunção de atributos. A componente da nota âncora é a parte inofensiva, uma vez que descreve uma estrutura de nota de piano comum. Uma nota âncora é dada por um produto cartesiano e ramifica-se em intensidade, altura, timbre e duração. Cada um destes atributos é uma construção conceptual de um tipo básico ou simples: timbre refere-se a um número real, altura refere-se a um inteiro (passos de meio-tom), intensidade refere-se a palavras (assentando na álgebra monóide sobre o alfabeto e os inteiros) para descrever indicações do som tais como *mf*, *ff*, etc., e duração é um número real, tal como a componente timbre.

A parte menos inofensiva é a componente satélite. Esta contém um conjunto finito de objectos (que pode ser vazio), cobrindo objectos delegados como as notas ornamentais do trilo numa configuração hierárquica. No entanto, como a estrutura hierárquica não pode ser forçada a terminar neste ponto, os objectos satélite devem estar abertos a uma eventual delegação mais profunda. Deste modo, qualquer elemento no «saco» satélite deve também ser uma nota estrutural. Isto significa que o conceito de nota estrutural é circular. Mas não se trata necessariamente de uma situação circular para uma instância concreta do conceito de nota estrutural. De facto, a colecção de satélites pode tornar-se o conjunto vazio, e então, não aparecendo indeterminações, a instância do conceito é uma construção finita. No entanto, a existência de um conceito circular de nota estrutural não é um assunto trivial. Podíamos considerar esta

estrutura como uma *teoria de Galois de conceitos*, uma vez que estes conceitos podem ser definidos por «equações conceptuais» como números algébricos (tal como $x^2+1=0$ para a famosa unidade imaginária). Isto introduz um novo ramo da teoria do topos: a teoria de Galois conceptual (ver Mazzola, G., *et al.*, 2001, para pormenores técnicos).

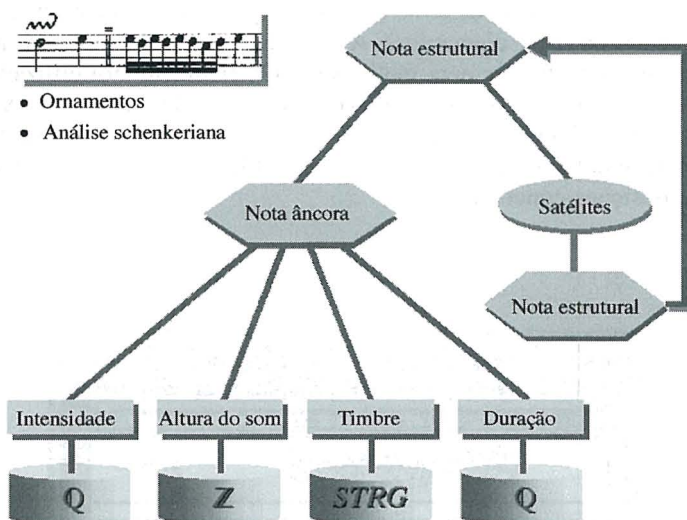


Fig. 10 — Um exemplo notável de uma construção conceptual em música é definido pelo conceito de nota estrutural. Essencialmente, este conceito formaliza estruturas onde temos um objecto âncora rodeado por um agrupamento de «objectos delegados» de hierarquia inferior. Estas construções são necessárias quando se trata de ornamentos gerais (tais como os trilos) ou de análise schenkeriana global de partituras musicais. Neste exemplo, a construção do conceito recursivo cristaliza na componente de satélite do conceito, um atributo que contém objectos da mesma arquitectura a priori, à semelhança da estrutura da nota estrutural inicial

Para além da grande vantagem de poder ser implementada em ambientes de *software* para computadores, esta linguagem conceptual tem a grande desvantagem de a classificação dos seus objectos se tornar uma tarefa extremamente difícil no campo da matemática abstracta, desde a enumeração simples de todos os casos possíveis até à mais sofisticada determinação de todas as classes de isomorfismos de determinados tipos (Mazzola, G., *et al.*, 2001).

Mesmo em situações elementares, a variedade de objectos torna-se incrivelmente grande. Vejamos, por exemplo, conceitos tais como os motivos. Um motivo pode ser encarado como um subconjunto do produto cartesiano $(Z_{12})^2$, no sentido de a primeira cordenada ser a duração, um inteiro módulo 12 unidades de tempo, enquanto o segundo corresponde à classe de altura do som. Dizemos que dois motivos são equivalentes (isomórficos) se e só se existir um isomorfismo afim do espaço ambiente $(Z_{12})^2$ que tranforma um deles no outro (o grupo desses isomorfismos afins consiste em 663 552 elementos). A lista de representativos para as 26 classes de isomorfismos para motivos de três elementos foi utilizada em composição (Mazzola, G., 1990a) e análise (Mazzola, G., 1990b) musicais. Isto significa que a classificação é não apenas interessante em termos matemáticos, mas pode também ser utilizada para síntese e análise de composições concretas.

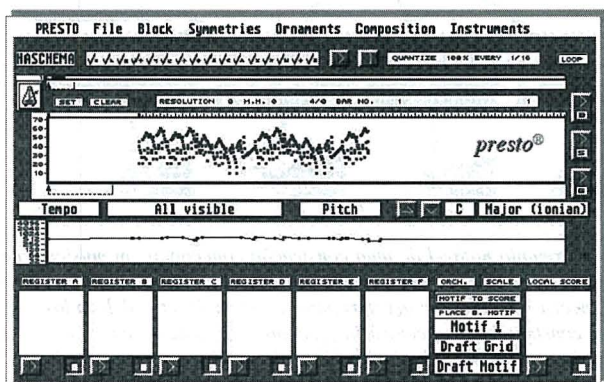


Fig. 11 — A janela principal do software de composição presto®. Este software foi desenvolvido em 1986 segundo uma recomendação de Herbert von Karajan. Fizeram-se vários CD de estilos diferentes (jazz e música contemporânea) utilizando este programa. Foi concebido de modo a permitir a aplicação de ferramentas de construção paradigmáticas da teoria matemática da música, como, por exemplo, transformações de simetria e deformações

A riqueza exorbitante de motivos musicais torna-se evidente quando tentamos contar o número de classes de isomorfismo de motivos de 72 elementos em $(Z_{12})^2$ (metade da cardinalidade do espaço ambiente).

Com base na teoria da enumeração de Polya e de de Bruijn, Harald Friperntinger (1993) conseguiu contá-los, tendo obtido o número

$$2\ 230\ 741\ 522\ 540\ 743\ 033\ 415\ 296\ 821\ 609\ 381\ 912 \approx 2.23 \times 10^{36}$$

enquanto o número de estrelas numa galáxia média não excede 10^{11} . Deste modo, temos ainda algumas opções composicionais.

A figura 11 mostra uma interface principal do *software* comercial para composição *presto*[®] (Mazzola, G., 1990b), que foi implementado e aperfeiçoado no período de 1986 a 1994 em computadores Atari sob recomendação de Herbert von Karajan. Utilizando este *software*, fizeram-se vários CD, entre os quais um CD de *jazz* no qual se utilizou a classificação de motivos de três elementos. Os críticos de *jazz* não reconheceram a composição para piano, percussão e guitarra eléctrica como sendo música digitalmente gerada e executada.

O *software* que executa o quadro conceptual acima referido é uma versão aperfeiçoada do *software* RUBATO[®]. Actualmente, está a ser desenvolvido por dois grupos de investigação, um na Universidade Técnica de Berlim (www.mamuth.de) e o outro no Laboratório Multimédia da Universidade de Zurique (www.encycloSPACE.org). A nova versão do RUBATO[®] funciona na tecnologia Mac OS X e utiliza o sistema UNIX, bem como componentes de *software* baseados em JAVA (estes últimos concebidos para navegação visual na Internet). Inclui componentes analíticas (análise rítmica, melódica, harmónica), uma componente composicional (de facto, um melhoramento do antigo *presto*[®]), um módulo de execução, módulos de filtragem para diferentes formatos de importação/exportação e análise lógica de construções conceptuais (ver www.mamuth.de, e www.encycloSPACE.org). A figura 12 apresenta a arquitectura global do RUBATO[®]. Mostra-se aqui uma janela de aplicação principal do DBMS (*Data Base Management System*: sistema de gestão de bases de dados) deduzida a partir da presente estrutura conceptual; a terminologia técnica é «forma» para espaços ambientes que incluem itens conceptuais, enquanto os últimos se designam por «denotadores» e representam uma espécie de ponto no seu espaço de formas ambiente. Adiciona-se um número de módulos dinamicamente carregáveis, os chamados RUBETTES[®], à aplicação básica. A natureza universal dos modelos de dados permite a transferência sem restrições de dados entre quaisquer dois desses RUBETTES[®].

Deve salientar-se que uma tal plataforma de *software* avançado é a base de um novo estilo de investigação interactiva. O RUBATO[®] parece, de facto, um grande edifício onde trabalham diferentes especialistas e onde

global. Não podemos abordar em pormenor esta matéria, por isso restringimo-nos a uma apresentação geral qualitativa. A ideia que está na base deste módulo é a de que a execução de uma determinada partitura pode ser dirigida pelo uso de resultados analíticos sobre essa partitura; esta é a abordagem clássica adoptada pelo músico e filósofo Theodor W. Adorno (1956). Isto não significa que a motivação emocional ou gestual não seja importante, mas actualmente a análise é a mais desenvolvida componente crítica da execução. Para compreender a natureza da execução, temos de ter consciência de que se trata de uma transformação do nível simbólico de objectos da partitura para o nível físico de eventos sonoros (ver figura 13).

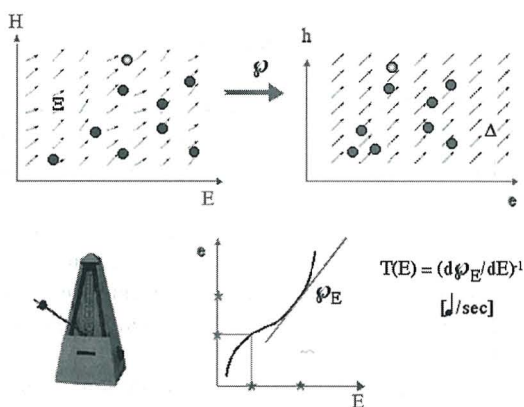


Fig. 13 — A execução é uma transformação ϕ de uma realidade simbólica da partitura para a realidade física dos eventos acústicos. A transformação pode ser descrita pela sua derivada, como se ilustra na metade inferior desta figura. No caso do tempo de duração, a derivada é aquilo de que precisamos para definir tempo musical. A duração física é obtida do tempo por uma integração. Quando consideramos vários parâmetros de tom, tem de se generalizar a curva tempo a um campo de execução X (em cima à esquerda) cujo integral conduz à transformação execução

A transformação execução ϕ pode ser descrita pelo campo de imagem inversa Ξ do campo diagonal $\Delta = (1, \dots, 1)$ no espaço físico (em cima à direita na figura 13). No caso de uma execução conduzida por uma análise A , este campo Ξ é deformado de acordo com operadores específicos Ω , conduzindo a um novo campo $\Xi^* = \Omega(\Xi, A)$. Por exemplo, operadores de

tempo actuam na componente de tempo de Ξ , enquanto operadores de afinação actuam na componente altura⁶ de Ξ .

O conceito de análise no *software* RUBATO[®] foi especificado de modo que a análise seja representada por uma função de ponderação numérica das notas, das barras de compasso, das pausas, etc., da partitura dada. Esta exigência garante que os operadores de execução podem ser construídos a partir de dados numéricos e não a partir de textos difusos tradicionais em teoria musical. Cada RUBETTE[®] analítico conduz a tais funções de ponderação. Esta metodologia proporcionou-nos execuções muito interessantes e convincentes de diversas peças, em particular do Contrapuntus III da *Arte da Fuga*, de Johann Sebastian Bach (Stange-Elbe, J., 2000).

Referências

- Adorno, Th. W., (1956) – Fragment über Musik und Sprache. Stuttgart, Jahresring
- Auroux, S., (1979) – *La sémiotique des encyclopédistes*. Payot, Paris
- Barrow, J., e Tipler, F., (1986) – *The Anthropic Cosmological Principle*: Oxford University Press, Nova Iorque
- Friepertinger, H., (1993) – Endliche Gruppenaktionen in Funktionenmengen — Das Lemma von Burnside — Repräsentantenkonstruktionen — Anwendungen in der Musiktheorie. Dissertação de doutoramento, Univ. Graz
- Fux, J. J., (1742) – *Gradus ad Parnassum* (1725). Dt. und kommentiert von L. Mitzler, Leipzig
- Hichert, J., (1993) – Verallgemeinerung des Kontrapunkttheorems für die Hierarchie aller starken Dichotomien in temperierter Stimmung. Dissertação de licenciatura, TU Ilmenau
- Jackendoff, R., e Lerdahl, F., (1983) – *A Generative Theory of Tonal Music*. MIT Press, Cambridge MA
- Mac Lane, S., e Moerdijk, I., (1994) – *Sheaves in Geometry and Logic*. Springer, Nova Iorque
- Mazzola, G., (1990a) – SYNTHESIS. StoA music, Zurique
- Mazzola, G., (1990b) – *Geometrie der Töne*. Birkhäuser, Basileia
- Mazzola, G., et al. (1998) – Humanities@EncycloSPACE. Schweizerischer Wissenschaftsrat, www.encycloSPACE.org, Berna
- Mazzola, G., et al. (2001) – *The Topos of Music*. To appear, Birkhäuser, Basileia

⁶ Notemos que, em geral, campos de execução não são produtos cartesianos de factores unidimensionais.

- Mazzola, G., *et al.* (1989) – A Symmetry-Oriented Mathematical Model of Classical Counterpoint and Related Neurophysiological Investigations by Depth-EEG. In: Hargittai I (ed.): Symmetry II, CAMWA, Pergamon, Nova Iorque
- Mazzola, G., e Zahorka, O., (1996) – RUBATO on the Internet (www.rubato.org). Univ. Zúriq
- Muzzolini, D., (1995) – Musical Modulation by Symmetries. J. for Music Theory
- Noll, Th., (1995) – Morphologische Grundlagen der abendländischen Harmonik. Dissertação de doutoramento, TU Berlim
- Schönberg, A., (1966) – Harmonielehre (1911). Universal Edition, Viena
- Stange-Elbe, J., (2000) – Analyse- und Interpretationsaspekte zu J. S. Bachs Kunst der Fuge mit Werkzeugen der objektorientierten Informationstechnologie. U Osnabrück, Osnabrück
- Ver www.mamuth.de (site of the KiTMaMuTh research group for mathematical music theory at TU Berlin, director: Thomas Noll)
- Ver link em www.encyclospace.org to the Multimedia Lab of U Zúriq

(Página deixada propositadamente em branco)

A. J. Tromba
Universidade da Califórnia
Santa Cruz, EUA

A matemática no melhor mundo possível¹

De acordo com os estudos de ADN mitocondrial, a nossa espécie terá começado há uns 200 000 anos. Contudo, alguns estudiosos crêem que a verdadeira consciência começou muito mais tarde.

Quando a consciência de si e do mundo se instalaram, o planeta deve ter sido um lugar terrivelmente assustador e inexplicável. O homem estava cercado por acontecimentos muito para além da sua compreensão e do seu controlo; de tremores de terra, cheias, tufões e secas a cometas e eclipses solares e lunares.

Para além de ter dificuldade em perceber estes acontecimentos naturais terríveis, muito provavelmente a humanidade tinha dificuldade em tomar conta dos seus haveres, da passagem do tempo ou mesmo do planeamento do futuro. Sabemos que os processos de contagem começaram há pelo menos 35 000 anos. Marcas em fragmentos ósseos indicam que, mesmo nessa altura, se podem encontrar rudimentos do conceito de base aritmética (por exemplo, as bases 5 e 10).

A história dos números, isto é, o desenvolvimento do sistema aritmético actual, notação aritmética e operações, é fascinante. É uma

¹ Tradução de F. J. Craveiro de Carvalho, professor do Departamento de Matemática, Universidade de Coimbra.

história que põe a claro a dificuldade que muitas grandes civilizações não foram capazes de vencer para obter um sistema de cálculo eficaz. Mas os números, em si e por si, não são matemática. A palavra grega *mathema* significa saber, conhecimento, compreensão, percepção — sugerindo que a história da matemática começou com tentativas para compreender o mundo.

Havia quatro assuntos que os antigos Gregos consideravam fundamentais: a aritmética, a geometria, a astronomia e a música. Descobriram também que os princípios astronómicos podiam ser descritos em termos de conceitos geométricos e os princípios musicais em termos de ideias matemáticas. Ficam então dois assuntos fundamentais: os números e a geometria.

Mas os Gregos haviam estabelecido também uma ligação entre os números e a geometria, precisando, que um número é o comprimento de um segmento. Esta ligação foi subsequentemente destruída pela descoberta, no século V, antes de Cristo, de que $\sqrt{2}$ era incomensurável, isto é, não podia ser expresso como a razão de dois inteiros. Foi apenas, basicamente, reestabelecida no século XVII com a descoberta das coordenadas geométricas por Fermat e Descartes.

Os antigos Gregos puseram-nos no longo caminho que vai do terror e da inexplicabilidade constantes à idade científica moderna, onde temos conseguido uma compreensão extraordinária dos fenómenos físicos e biológicos. O primeiro passo brilhante neste caminho foi, simplesmente, a hipótese de que o mundo era racional e «podia» ser compreendido e, mais, podia ser compreendido através da linguagem da matemática; isto é, pelos números e pela geometria. Platão afirmava que as leis matemáticas, eternas e invariáveis, são a essência da realidade e que a realidade objectiva e universalmente válida pertence apenas ao campo das ideias. Do ponto de vista do século XX podemos apenas maravilhar-nos com a audácia deste ponto de vista.

Os Gregos estudaram figuras geométricas planas bem como curvas (a circunferência, a elipse, a parábola e a hipérbole; isto é, as secções cónicas) e as superfícies obtidas pela rotação destas curvas em torno de um eixo central. Imaginaram circunferências a rolar em circunferências, que por sua vez podiam rolar noutras circunferências, criando assim classes de curvas inteiramente novas.

Os cientistas da Antiguidade usaram estas curvas para criarem um modelo matemático do movimento dos planetas e das estrelas no céu. Pitágoras supunha que as estrelas estavam ligadas a uma esfera de cristal que rodava diariamente em torno de um eixo passando pela Terra.

Analogamente, os outros sete corpos celestes antigos — o Sol, a Lua, Mercúrio, Marte, Júpiter, Vénus e Saturno — estavam cada um deles ligados a uma esfera móvel própria.

Mais tarde, Platão, no seu *Timeu*, ligou a criação do tempo à do céu e dos planetas. O Sol, no seu curso, define o dia e a noite bem como o ano, enquanto a Lua descreve o mês quando, na sua órbita circular, ultrapassa o Sol outra vez, «e os ciclos dos outros planetas os homens não os compreenderam, excepto alguns entre muitos, e não deram nem nomes a estes ciclos, nem mediram as suas órbitas usando números, fazendo observações, de modo que não sabem verdadeiramente que a diversidade intrincada e a alteração maravilhosa do vaguear daqueles planetas é tempo também». O conceito pitagórico dos planetas e dos céus tornou-se mais tarde uma teoria do movimento dos corpos celestes, que formou a base da astronomia até ao século XVI. O que aconteceu foi o seguinte:

As secções cónicas foram estudadas durante muitos anos, mas os resultados principais foram obtidos por Apolónio de Perga (262 a.C. – 190 a.C.). São descritos nos seus oito livros, *Secções Cónicas*, dos quais os primeiros quatro são revisões do trabalho feito por Euclides, trabalho que posteriormente se perdeu para sempre. Também o oitavo livro do tratado de Apolónio parece ter-se perdido.

A descoberta das secções cónicas é atribuída a Menecmo, membro da escola de matemática de Platão. Esta escola floresceu na cidade-estado de Atenas no século IV, logo a seguir ao século de Péricles, o período clássico da arte, arquitectura e filosofia gregas, em que foi construída a Acrópole.

Platão, um estudante de Sócrates, fundou a sua escola, a Academia, numa zona sagrada de Atenas chamada Akademeia (em honra do herói Akademos). O nome de todas as academias posteriores derivou do nome desta instituição, que existiu ininterruptamente durante cerca de um milhar de anos, até ser dissolvida pelo imperador Justiniano, em 529 d.C. A escola de Platão era como uma pequena universidade onde o filósofo e os amigos leccionavam. Dois dos grandes matemáticos da antiguidade, Eudoxo de Cnidos (408 a.C. – 355 a.C.) e Teeteto (420 a.C. – 367 a.C.), foram membros da Academia. Embora Platão não fosse um matemático, apreciava tanto esta ciência que exigia que cada estudante dedicasse dez anos às ciências matemáticas e mais cinco à filosofia. A inscrição à entrada da Academia diria «*Que não entre aqui ninguém que não seja familiar com a geometria*».

Conta-se que Platão sugeriu o seguinte problema aos seus estudantes: explicar o movimento dos corpos celestes combinando vários movimentos

circulares e esféricos. Platão não apreciava especialmente a astronomia como ciência prática, igualmente útil para os lavradores e para os homens do mar. Na sua opinião, só merecia atenção como recreio para os géometras, para quem era uma fonte de problemas interessantes.

Que problemas apresentavam os movimentos dos planetas? Observados da Terra, estes movimentos parecem muito complicados. Os movimentos do Sol e da Lua podem ser grosseiramente descritos como circulares e de velocidade constante, mas os desvios da órbita circular eram uma fonte de problemas para os Gregos, que se sentiram desafiados a encontrar algumas explicações para aquelas irregularidades. As órbitas observadas dos planetas são ainda mais complicadas porque, à medida que os planetas fazem uma revolução, invertem o sentido por algum tempo, voltam a invertê-lo para avançar, tudo com velocidade variável. Os Gregos procuraram compreender este movimento aparentemente descontrolado por meio de um modelo geométrico.

Eudoxo estudou o problema e propôs uma teoria elegante e puramente geométrica dos movimentos celestes, que, contudo, tinha o defeito grave de não explicar muitos dos dados empíricos. Assim, no século III a.C., Apolónio de Perga sugeriu que as órbitas celestes deviam ser explicadas pela combinação de movimentos circulares. Nasceu assim a teoria astronómica mais importante dos dois milénios seguintes. Foi desenvolvida detalhadamente por Hiparco, o maior astrónomo da Antiguidade, no século II, a.C. O seu trabalho chegou até nós através da célebre *Colecção Matemática*, escrita pelo astrónomo grego Ptolomeu de Alexandria, no século II d.C., e chamada *al-magest*, «a maior», pelos astrónomos do mundo árabe da Idade Média. Este livro completou o *sistema geocêntrico* da astronomia, conhecido mais tarde por *sistema ptolomaico*. A teoria de Ptolomeu foi tão bem aceite que na Idade Média se pensava que havia sido dada directamente por Deus ao homem. Todos estes resultados confirmaram o ponto de vista platónico da concepção matemática do universo e estabeleceram a sua ligação ao Criador.

Antes da era grega, eram os deuses que eram responsável pelas atribuições dos homens, não sendo assim surpreendente que, com a ascensão da interpretação racionalista da natureza e com o êxito na descrição matemática da mesma natureza, Deus tenha sido ligado à matemática.

Já em 100 d.C., Nicómaco escreveu: «A aritmética é a mãe de tudo o mais, existindo naturalmente antes do começo, no espírito do Deus criador. Faz parte da concepção universal e extraordinária que o Criador do universo iniciou».

Os Gregos puseram, outras questões fundamentais relativas ao mundo em seu redor e em alguns casos respondera-lhes. Afirmaram que a luz viajava com velocidade finita e em linha recta. Compreenderam a lei da reflexão da luz e perceberam que esta lei era uma consequência de uma lei mais fundamental ainda: que a luz segue o caminho mais curto.

Puseram questões acerca de formas matemáticas ideais e pensaram nas suas ligações com o mundo natural. Platão previu o desenvolvimento da física moderna, porque compreendeu que o mundo objectivo só podia ser interpretado, ou mesmo compreendido, através de uma lente matemática. Na sua busca de formas ideais, classificaram os cinco poliedros regulares: cubo, tetraedro, octaedro, icosaedro e dodecaedro.

Seguindo a linha de pensamento originada no princípio mínimo que descreve os raios luminosos, Pappus interrogou-se sobre se as abelhas construía os seus favos de modo a usarem a menor quantidade de cera, uma questão que não foi respondida até à segunda metade do século XX.

Os Gregos antigos foram também os primeiros a aplicar a matemática na guerra e a tornar possível o uso moderno da matemática, tanto na guerra como nos serviços secretos. Desenvolveram o conceito de código para a entrega de mensagens secretas entre governos e comandantes no campo de batalha. Na verdade, a maior razão da fama de Arquimedes foi a defesa notável da cidade de Siracusa contra os Romanos, em 214 d.C., durante a segunda das três Guerras Púnicas. Chegou até nós a seguinte descrição da batalha:

Quando os Romanos cercaram a cidade depararam com um inimigo a quem o matemático de 73 anos havia fornecido armas inesperadas e poderosas. As legiões romanas, ao aproximarem-se, foram desbaratadas por uma artilharia muito eficaz, que disparava mísseis de pedra de longo alcance, e por uma quantidade enorme de projecteis de curto alcance. Os barcos da frota romana foram afundados por guindastes enormes, que ou deixavam cair pedras sobre eles ou levantavam as proas com uma garra de ferro e subitamente os esmagavam contra a água. Plutarco (na sua *Vida de Marcelo*) relata como o comandante Marcelo se dirigiu ao seu exército:

Não vamos pôr fim à luta contra este Briareus geométrico [um gigante de cem braços de lenda grega] que usa os nossos barcos para tirar água do mar, que infamemente fustigou e obrigou à retirada da sambuca [a escada de assalto romana, usada para escalar as muralhas de uma cidade cercada] e que pela quantidade de mísseis que nos atira ao mesmo tempo supera os gigantes de cem braços da mitologia?

Os soldados romanos estavam aterrorizados. Diz Plutarco,

Bastava que vissem uma corda ou um pedaço de madeira estendendo-se para além das muralhas para fugirem, exclamando que Arquimedes tinha mais uma vez inventado uma nova máquina para os destruir.

Mas para os soldados romanos o pior ainda estava para vir: a redução a cinzas dos seus barcos pelos «espelhos incendiários» de Arquimedes. Em 1200 o historiador bizantino Tzetzes, infelizmente um escritor não particularmente digno de crédito, fez o seguinte relato do que Arquimedes teria conseguido:

Quando os barcos romanos estavam ao alcance do tiro da proa, Arquimedes fez uma espécie de espelho hexagonal, e outros pequenos com 24 lados, que colocou a uma distância apropriada e que ele podia mover com a ajuda de dobradiças e fios de metal. Dispôs o espelho hexagonal de modo a ser bissectado pelo meridiano do Inverno e do Verão e os raios solares recebidos por este espelho, depois de reflectidos, provocaram um grande incêndio que reduziu os barcos romanos a cinzas, ainda que se afastassem do alcance da proa.

Um tal conjunto de pratos reflectores podia ter criado o mesmo efeito que um espelho parabólico.

O cerco de Siracusa durou dois anos e a cidade caiu por traição. A seguir ao assalto, o velho cientista foi morto violentamente por um soldado romano, embora o comandante Marcelo tivesse pedido aos seus homens que poupassem a vida de Arquimedes. Diz a lenda que Arquimedes estava sentado à frente da sua casa a estudar algumas figuras geométricas desenhadas na areia. Quando um soldado romano se aproximou, Arquimedes gritou: «Não mexa nas minhas figuras». Sentindo-se insultado, o soldado matou o grande homem.

Como por compensação, Marcelo erigiu um túmulo a Arquimedes onde estava representada uma esfera circunscrita por um cilindro, simbolizando, de acordo com os desejos de Arquimedes, o seu teorema favorito sobre os volumes do cone, cilindro e esfera. Quando Cícero (100 a.C. – 43 a.C.) visitou a Sicília, conseguiu localizar o monumento coberto de arbustos e espinhos.

Aparentemente Arquimedes considerava as suas contribuições para a engenharia como feitos menores, que não valeria a pena lembrar mais tarde. Segundo Plutarco,

Ele não se preocupou em deixar qualquer trabalho escrito sobre tais assuntos; para ele a construção de instrumentos era uma coisa sórdida e ignóbil, mais geralmente, qualquer arte dirigida para a utilização e o lucro, e só se empenhava naquelas coisas que, pela sua beleza e excelência, estão para além das necessidades habituais da vida.

Embora ocupado com a arte de fazer a guerra, como Leonardo da Vinci dois milénios mais tarde, o verdadeiro compromisso de Arquimedes, como o de muitos matemáticos depois dele, era com a *beleza* e o prazer que fluem do estudo da mais pura de todas as formas de arte, a matemática.

Depois da queda da cidade-estado de Atenas, em 338 a.C., a cidade de Alexandria na foz do Nilo tornou-se o centro da cultura grega, até ser destruída um milénio mais tarde. Por volta de 285 a.C., foi fundado o Museu (ou Templo das Musas). Era uma instituição cultural onde os estudiosos podiam trabalhar e estudar. No seu momento mais alto, a Biblioteca do Museu conteria 750 000 manuscritos. Foi em Alexandria que Euclides escreveu os seus *Elementos* (de matemática), que viriam a tornar-se o livro mais divulgado no mundo ocidental.

A ascensão do cristianismo trouxe maus ventos à cultura grega. Em 392, o imperador romano Teodósio ordenou a destruição dos templos gregos. Hipácia, uma conhecida matemática de Alexandria, recusou renunciar à sua fé grega, foi capturada por fanáticos e morta. Ao mesmo tempo, os livros gregos foram queimados aos milhares. O golpe final para Alexandria e a ciência grega veio com a conquista do Egipto pelos muçulmanos, em 640 d.C.. Os livros que restavam foram destruídos.

A maioria dos estudiosos fugiu para Constantinopla, levando consigo vários manuscritos gregos clássicos. Embora não houvesse ciência ou matemática activas na nova cidade, pelo menos algum do saber clássico foi preservado. A ciência grega tinha sido completamente destruída, mas uma boa parte do seu saber e da sua visão do mundo esperou calmamente em Constantinopla para ser descoberta uns oito séculos mais tarde.

Não quer dizer que todo o saber tenha desaparecido do mundo árabe. Em 830 o astrónomo *Mohammed ibn Musa al-Khowarizmi* baseou-se nos Gregos, hindus e Babilónios para escrever um tratado clássico, *Al-jabr W'al Muqâbala*, dando assim a palavra Al-jabr (álgebra), significando restabelecer. Este trabalho resumia o que na altura se sabia sobre equações.

O alvorecer da era moderna não começou até à chegada de Ricardo, Coração de Leão, às muralhas de Jerusalém. O conhecimento grego foi gradualmente trazido para a Europa e absorvido. Com ele veio a profunda e extremamente interessante perspectiva filosófica de que o universo foi concebido matematicamente. Apesar de sinais de que podia realmente ser assim, entrava em conflito com os ensinamentos da Igreja, nomeadamente que Deus concebeu o universo.

A resolução óbvia foi inferir que Deus usou matemática na concepção do universo ou que *Deus é um matemático*. Por pretensioso que possa parecer, tem sido um princípio que guiou muitos cientistas através dos séculos. Mais, aceitando esta premissa pode inferir-se que, descobrindo as Suas leis matemáticas, se O descobre, ao nosso Criador.

É surpreendente talvez para alguns que esta perspectiva tenha vingado como vingou e, mais surpreendente ainda, que sobreviva até aos dias de hoje. Quando Kepler, em 1619, descobriu as suas três belas leis do movimento em termos das secções cónicas gregas, escreveu na sua *Harmonia do Mundo*:

A sabedoria do Senhor é infinita; são-no também a Sua glória e o Seu poder. Vós, céus, cantai os Seus louvores! Sol, Lua e planetas, glorifiquem-n'O na vossa linguagem inefável! Harmonias celestiais, vós todas que compreendeis os Seus trabalhos maravilhosos, louvai-O. E tu, minha alma, louva o teu Criador! É por Ele e n'Ele que tudo existe. O que conhecemos melhor está n'Ele, bem como na nossa ciência vã. Louvor, honra e glória, a Ele por toda a eternidade.

Na segunda parte deste século, o prémio Nobel da física Paul Dirac expressou, da forma seguinte, esta opinião.

Parece ser uma das características fundamentais da natureza que as leis físicas fundamentais sejam descritas em termos de uma teoria matemática de grande beleza e poder, necessitando-se de matemática de alto nível para a compreender. Pode perguntar-se: porque está a natureza construída segundo estas linhas? Pode apenas responder-se que o nosso conhecimento actual parece mostrar que a natureza é constituída assim. Temos simplesmente de o aceitar. Podia talvez descrever-se a situação dizendo que Deus é um matemático de grande categoria e que usou matemática muito avançada na construção do universo.

As nossas fracas tentativas matemáticas permitem-nos compreender um pouco do universo e, à medida que formos desenvolvendo matemática cada vez mais avançada, podemos esperar compreendê-lo melhor.

Assim, se há Deus, o matemático supremo, talvez haja uma lei matemática universal da qual tudo o mais decorre.

A primeira pedra filosófica e matemática de uma tal lei foi lançada por Gottfried Wilhelm Leibniz, diplomata, filósofo, matemático, cientista e estudioso universal. As suas ideias filosóficas foram apresentadas no seu *Essais de Théodicée sur la Bonté de Dieu, la Liberté de l'Homme et l'Origine du Mal*, que apareceu em 1710. Nele desenvolve a ideia filosófica de que o nosso mundo está organizado para ser *o melhor mundo possível*. Esta filosofia, bem conhecida em círculos eruditos na primeira metade do século XVIII, tinha muitos apoiantes. O mais influente era provavelmente o filósofo alemão Christian Wolff (1679-1754). Leibniz considerou a questão de como a hipótese de um Criador todo-poderoso, onisciente e infinitamente bom podia ser compatível com as imperfeições da Sua criação.

Simplificando muito, a teoria de Leibniz sustentava que Deus não interfere, como um relojoeiro desastrado, de tempos a tempos com as questões do mundo para regular os ponteiros do Seu relógio. Pelo contrário, Deus criou o Seu mundo numa harmonia pré-estabelecida. Como um relojoeiro habilidoso cuidando do seu relógio, Deus fez a natureza de cada coisa concordar com a natureza de todas as outras por toda a eternidade; portanto, todas as partes estão em completa harmonia com cada uma para sempre. Isto, apenas, é digno de Deus, o ser mais inteligente, e todo-poderoso.

E Deus existe, segundo Leibniz, pela razão seguinte: Deus é o ser que possui todas as propriedades em grau mais elevado, tanto quanto estas propriedades possam coexistir. Esta condição é necessária porque algumas propriedades, como a santidade e a onnipotência, não podem coexistir completamente. A santidade de Deus restringe a Sua onnipotência, uma vez que é incompatível com a Sua santidade fazer o mal que Ele podia fazer sendo onnipotente.

Segue-se pois que Deus pode, na verdade, pensar em todos os mundos possíveis, mas pode querer, e portanto pode criar, só o melhor de entre eles: o melhor mundo possível.

Portanto, a existência deste mundo como o melhor possível é uma consequência da existência de Deus. Assim, qualquer outro mundo seria necessariamente menos completo que o nosso; contudo, a existência de

muitos males e desigualdades não é negada: o pecado e o mal existem. Mais, Leibniz prova que um mundo sem pecado e sem mal seria impossível.

Matematicamente, Leibniz diz que Deus usa princípios de otimização na concepção do mundo, dos quais o princípio grego de que a luz segue o trajecto mais curto seria um exemplo. Contudo, não foi proposto nenhum princípio geral abrangendo todos os fenómenos até 1744, quando o cientista francês Pierre-Louis Moreau de Maupertuis propôs o seu grande plano do universo, que se tornou conhecido como a *lei da menor acção*. Maupertuis publicou uma versão elaborada das suas ideias em 1746, «As Leis do Movimento e do Repouso Deduzidas de um Princípio Metafísico». O «princípio metafísico» de Maupertuis é a suposição de que a natureza funciona sempre com a maior economia possível.

Maupertuis definiu o conceito de *acção* como a energia a actuar sobre o tempo ou, formalmente,

$$\text{Acção} = \text{Energia} \times \text{Tempo}$$

Na verdade, esta mesma definição de *acção* foi tirada de uma das cartas de Leibniz. Muito provavelmente Maupertuis soube deste princípio por um membro da grande família Bernoulli, uma dinastia de matemáticos suíços que, a partir de 1690, faz da cidade de Basileia um centro de matemática mundial.

Maupertuis formulou então o seu princípio metafísico como «A Natureza minimiza sempre a *Acção*». Maupertuis viu neste princípio uma expressão da sabedoria do Ser Supremo, de Deus, de acordo com a qual tudo na natureza é executado do modo mais económico possível. Escreveu:

Que satisfação para o espírito humano, ao contemplar estas leis que contêm o princípio do movimento e do repouso para todos os corpos no universo, encontrar a prova da existência d'Aquele que governa o mundo.

Desde a descoberta por Maupertuis do seu princípio, tem-se visto que ele é uma lei básica por detrás de todas as teorias físicas modernas, da mecânica newtoniana às equações de Maxwell do electromagnetismo, da relatividade geral à mecânica quântica e mesmo à física das cordas, que procura explicar o *Big Bang* e o começo do tempo.

Contudo, a ideia de uma razão providente e a compreensão do Criador através das suas leis (matemáticas) perderam importância com a aceleração do progresso científico. Mesmo assim, alguns cientistas modernos ainda sustentaram estes pontos de vista. Entre eles, neste século, o mais proeminente foi Max Planck, o pai da revolução quântica.

No «dia de Leibniz», 29 de Junho de 1922, Max Planck falou na Academia das Ciências de Berlim e fez as observações seguintes sobre a filosofia de Leibniz e o princípio da menor acção.

... A *Teodiceia* ... culmina com a afirmação de que o que quer que ocorra no nosso mundo, grande ou pequeno, na natureza ou na vida espiritual, é definitivamente regulado pela razão divina, e de tal forma que o nosso mundo é o melhor mundo possível.

Manteria Leibniz esta afirmação hoje em dia, tendo em conta a miséria dos tempos actuais, o falhanço amargo de muitos esforços não dirigidos de forma imediata para o ganho material, o facto inegável de a imaginada harmonia geral entre as pessoas estar hoje mais longe da sua concretização do que nunca? Sem dúvida a nossa resposta a esta pergunta teria de ser afirmativa, mesmo se não soubéssemos que Leibniz nunca deixou de se ocupar intensamente até aos seus últimos anos, apesar de um destino adverso e de muitos desapontamentos de todo o tipo, e não erraremos muito se admitirmos que foi exactamente a *Teodiceia* que o apoiou e confortou nos dias de maior sofrimento da sua vida. Mais uma vez, este é um exemplo comovente da velha verdade de que os nossos princípios mais profundos e sagrados estão firmemente enraizados na nossa alma, não dependendo das experiências no mundo exterior.

A ciência moderna, em particular sob a influência do desenvolvimento da noção de casualidade, afastou-se do ponto de vista teleológico de Leibniz. A ciência abandonou a hipótese de uma razão providente, especial, e considera cada acontecimento no mundo natural e no espiritual, pelo menos em princípio, como redutível a estados anteriores. Observamos contudo um facto, em especial na ciência mais exacta, que, neste contexto, pelo menos, é surpreendente. A física de hoje, quanto à sua organização teórica, é completamente governada por um sistema de equações diferenciais espaço-tempo que expressam que cada processo na natureza é totalmente determinado pelos acontecimentos que ocorrem na sua vizinhança temporal e espacial imediata. Todo este rico sistema de equações diferenciais, embora difiram no pormenor, pois referem-se a processos mecânicos, eléctricos, magnéticos e térmicos, está hoje inteiramente

contido num único teorema, no princípio da menor acção. Abreviadamente, este diz que, de todos os processos possíveis, os que de facto ocorrem são os que envolvem um dispêndio mínimo de acção. Como podemos ver, é necessário apenas um pequeno passo para reconhecer na preferência pela menor quantidade de acção a intervenção da razão divina e descobrir portanto uma parte da ordenação teleológica de Leibniz do universo. A diferença de ponto de vista não é, como vemos, de modo nenhum uma diferença essencial, mas apenas uma diferença de interpretação, uma vez que o primeiro se restringe ao muito pequeno, enquanto o outro abrange grandes distâncias no espaço e no tempo. Todos têm de decidir por si qual o ponto de vista básico e perguntar-se que abordagem acabará por ter mais êxito.

Na física de hoje, o princípio da menor acção tem um papel relativamente menor. Não se enquadra completamente no quadro das teorias actuais. Claro que é um enunciado correcto; contudo, habitualmente, não serve como fundamento da teoria mas como um apêndice, verdadeiro mas dispensável, porque a física teórica de hoje está feita inteiramente à medida do princípio dos efeitos locais infinitesimais e considera as extensões a espaços e tempos maiores como uma complicação desnecessária e não económica do método de tratamento.

Portanto a física inclina-se para considerar o princípio da menor acção mais como uma curiosidade acidental e formal do que como um pilar do conhecimento físico.

Assim, foi ainda mais surpreendente que tenha sido descoberto que este princípio, originalmente considerado por Leibniz e Maupertuis como um teorema mecânico, era verdadeiro, sem qualquer restrição, para toda a física do seu tempo, por Hermann von Helmholtz. Recentemente, David Hilbert, usando a versão de Hamilton do teorema, estabeleceu-o na teoria da relatividade geral de Einstein. Quanto mais complicadas se tornam as circunstâncias, menos provável é que o domínio de uma tal abordagem seja incompleto por não poder explicar uma relação geral, simples e geralmente aceite. O nosso desejo de compreensão só será satisfeito se toda a lei, que tenha sido descoberto ser verdadeira, tiver sido compreendida nos seus significado e relevância e integrada na estrutura teórica global.

Numa conferência num dia de Leibniz posterior, 27 de Junho de 1935, Planck voltou a reflectir sobre a frase principal de *Teodiceia*, afirmando que, entre os mundos possíveis, o nosso é o melhor.

Hoje, quando parece ser cada vez mais difícil encontrar um sentido razoável para o desenvolvimento da humanidade, esta frase pode parecer grotesca na sua audácia. Mantê-la-ia Leibniz face aos acontecimentos presentes, aos erros da Grande Guerra, à desconfiança mútua generalizada e ao armamento crescente dos povos? Teríamos de responder afirmativamente a esta pergunta, mesmo se não soubéssemos que Leibniz não abandonou a sua filosofia por causa dos horrores da Guerra dos Trinta Anos, que atingiu os seus dois primeiros anos de vida, nem pela profunda e sub-sequente humilhação do seu país. Mais precisamente, é um facto que a filosofia da *Teodiceia* não pode ser provada nem negada pela história. Não é resultado da experiência, mas a suposição e o ponto de partida da atitude de Leibniz perante o mundo. Segundo o seu ponto de vista filosófico, qualquer contradição aparente, cada sofrimento, cada injustiça no mundo podem ser justificados pelo facto de a limitada mente humana ser incapaz de reconhecer a ligação completa de um acontecimento com a concepção global da criação. Da mesma forma que pela destruição de uma parte, o todo pode avançar, assim um mal local pode ajudar a melhorar e a aperfeiçoar o geral.

Recorde-se que o filho mais velho de Planck, Karl, foi morto em Verdun em 1916; ambas as filhas morreram ao dar à luz os seus primeiros filhos; o filho mais novo, Erwin, foi condenado à morte pelo juiz Freisler, de Hitler, e executado em Janeiro de 1945, pelo seu envolvimento na tentativa de assassinato von Stauffenberg. Pouco depois, Planck perdeu todos os seus haveres pessoais e, com quase 87 anos de idade, fez parte da grande migração de refugiados para o Ocidente, como milhões de outros antes e depois dele².

Universidade da Califórnia
Santa Cruz

² Para uma descrição mais pormenorizada do princípio da menor acção, ver *The Parsimonious Universe*, de Stefan Hildebrandt e Anthony Tromba, Springer-Verlag, Nova Iorque (1996).

Nota do Tradutor: Algumas pessoas concederam-me um pouco do seu tempo e apresentaram-me sugestões sobre a tradução de algumas passagens do texto original. Seguidas ou não, todas se revelaram úteis e justificam o meu agradecimento, que vai para os professores Onésimo Teotónio Almeida, José Ribeiro Ferreira, J. A. Sampaio Martins, Manuel Portela e, em especial, Alan Freeland.

A arte e a ciência no século XX

Dirijo-me aos géometras, cuja fecunda imaginação leva à construção de muitas geometrias. Qual delas escolher? Os físicos são talvez aqueles que mais depressa e melhor juntam os dados da experiência à matemática. Para os físicos, não basta que se fale em n dimensões. Os físicos querem saber como é que, concretamente, as medidas são efectuadas. E os artistas querem saber como é que elas são *sentidas*.

A arte do desenho e suas derivadas, como a pintura e a escultura, precisa também de adoptar uma geometria para ser coerente no seu modo de *comunicação*, visual ou táctil. O artista descobre as geometrias na sua necessidade de *expressão*.

Evocarei a expressividade mais imediata, nas formas mais elementares, para ser mais claro. Essas formas servem porém as construções mais complexas do espírito, assim como os registos dos impulsos vitais.

As palavras «deformação» e «informal» são frequentes no actual estado efémero da crítica de arte. Vamos também utilizá-las, embora saibamos que a arte procura a boa forma. Esteticamente incorrectas, o uso frequente das palavras «deformação» e «informal» revelam que estamos a passar de uns critérios formais a outros: novos critérios que se definirão claramente quando houver melhor encontro entre géometras, físicos e artistas. O que é desejável e, felizmente, possível.

A arte e a ciência dos nossos dias têm algumas origens comuns. São ambas filhas da experiência e do livre pensamento. Porém, a arte e a ciência são frequentemente referidas, mesmo pelos seus praticantes,

como actividades absolutamente distintas. E a sociedade actual, que se desenvolve através da especialização dos seus componentes, deixa cavar o fosso existente entre artistas e cientistas, colocando-os em funções sociais que tendem, cada uma delas, a menorizar as contribuições das outras.

Mas se, por um momento, considerarmos as contribuições admiráveis quer da arte, quer da ciência, poderemos reparar que *ambas* (cada uma a seu modo, e às vezes, sem o saberem, de modo análogo) ajudam a modificar a consciência humana, através do simples exercício da liberdade de pensamento e da difusão da experiência.

Algumas pessoas pensarão que é utópico propor a conjugação da arte e da ciência. Para um crítico de arte é, concretamente, um projecto viável.

Hoje vou limitar-me a apresentar algumas situações análogas na visão que o cientista e o artista têm do mundo enquanto agentes livres.

Desde logo, chamo a atenção para a diferença entre os *modelos* paradigmáticos da pintura no século XIX e no século XX, sendo o primeiro estático e o outro fluido. Um separava o espaço do tempo; o actual não separa. Isto significa que as artes do desenho passaram a introduzir o tempo nas suas sugestões de espaço. Assim, o século XIX adoptou Ingres como modelo, enquanto o século XX encontrou em Paul Klee um dos seus artistas-teóricos mais fecundos, cuja arte rivaliza com a *música*, a arte do tempo interior. O próprio Paul Klee declarou:

«Diz-se que Ingres introduziu a ordem no repouso. Eu gostaria de introduzir, para além do *pathos*, a ordem no movimento.»



Ingres, La grande Baigneuse



Paul Klee, Croissance, 1921

O estetólogo Henri Focillon, que, como filósofo e historiador da arte, poderia satisfazer-se com a lógica dos conceitos e com a contemplação dos objectos artísticos, faz o elogio da mão, mostrando que ela não é apenas a serva obediente do cérebro, mas, sim, uma colaboradora arrojada na investigação da ordem universal, elaborando dados da observação e da expressão, que não existiriam sem ela. Henri Focillon escreveu no seu mais famoso livro, intitulado *A vida das formas*:

«Sem as mãos nunca teria havido geometria, pois são necessários traços e círculos para especular sobre as propriedades da extensão. Antes de poder reconhecer pirâmides, cones, espirais nos cristais e nas conchas, foi ou não necessário que as formas regulares tivessem sido previamente «simuladas» pelos homens, no ar ou sobre a areia?»

Qualquer quadro de Paul Klee é feito de uma maneira voluntariamente elementar, para que nós possamos adivinhar a sua feitura e, nessa contemplação activa, partilhar a génese da visão encantatória do artista. No século XX, Klee e muitos outros artistas não mostram o mundo feito, mas em devir.

Mas, já no final do século XX, a mão e o cérebro pediram auxílio às máquinas para compreender as formas. E alguns bailarinos dançaram com *robots*.

Porém, nem sempre é em dias de paz que as conquistas da arte e da ciência chegam ao grande público.

A técnica, bruscamente desenvolvida durante a Segunda Guerra Mundial, acelerou o tempo exterior. Atingiu-se então uma situação de ruptura entre o mundo mecanizado e a natureza. Passou-se a exigir dos homens um esforço excessivo de adaptação. Tornou-se necessário encontrar revigorantes modos de compensação e de auto-identificação. Verificou-se então, no domínio pictórico, uma intensificação da arte imediata, manifestação do «tempo interior», como o gestualismo de Hartung e outros abstraccionismos que recusam as formas da geometria euclidiana que, entretanto, predominava na técnica. O que até então parecia surgir unicamente da experiência individual de um Miró, de um Henry Moore ou de um Hans Hartung, passou a constituir a urgência expressiva colectiva. Hartung, escreveu:

«Quanto mais longe entrarmos em nós mesmos, mais clara e mais imperativa será a imagem que podemos dar das nossas sedimentações interiores, e também mais *universal* será a nossa expressão.»



H. Hartung, *Composition*

O gestualismo e outras tendências abstractas que rejeitam as formas da geometria euclidiana desenvolveram-se por toda a parte, na Europa de Hartung, na América de Pollock, no Japão de Inoue, etc. O caso japonês é particularmente interessante, porque esta vanguarda reactivou anti-quíssimas tradições da arte da caligrafia.

Estes informalismos constituíram, de facto e imprevisivelmente, o primeiro movimento vanguardista *planetário*, sucedendo às intenções universalistas dos abstraccionismos geométrico-euclidianos. Os «informalisms» puderam então corresponder aos expressionismos iniciados por Van Gogh e outros no final do século XIX, época essa em que também se sentiu com extrema agudeza, tal como agora, o mal e o bem da aceleração técnica e industrial; os «informalisms» puderam também corresponder à tábua rasa *dada* e às premonições surrealistas. Os «informalisms», que radicalizaram práticas abstraccionistas e engendraram novos figurativismos, com Dubuffet, tiveram inicialmente importância como *protesto* (era o NÃO que era audível ruidosamente); mas, activando as capacidades formativas instintivas, os «informalisms» despertaram intuições de geometrias não-euclidianas (era um SIM que se balbuciava).

Comparemos uma pintura euclidiana de Herbin com uma informalista de Jean Dubuffet.



Herbin



Dubuffet

Na passagem de Herbin para Dubuffet, reparamos que a pintura foi deixando de ser um plano onde se projectam formas concebidas previamente, para se afirmar como superfície sensível, texturada, lugar da acção táctil das mãos ou do corpo inteiro. Enquanto os abstracto-geométricos (euclidianos) permaneceram essencialmente relacionados com o exclusivamente *visual*, os não-euclidianos passaram a relacionar-se com o *palpável*; uns procuravam a *harmonia*, outros manifestavam a *vitalidade*.

Geralmente, os euclidianos herdaram do Renascimento a noção de forma fechada concebida num espaço estático puramente visual. Mas no século XX, desde o cubismo, esse espaço estático e monocular foi substituído pela *imagem de ritmo*, pela plurifocalidade e pela polissensorialidade.

No caso de Vasarely, as formas euclidianas, em si mesmas estáticas, são associadas de um modo regrado que intensifica a interacção das cores e das formas, transmutando-as em *energia* e movimentando virtualmente os volumes sugeridos.



Vasarely

Do lado dos informalistas, também o informalismo matérico de Dubuffet é estático. Mas o informalismo torna-se dinâmico no gestualismo de Georges Mathieu. E neste gestualista pode reparar-se na semelhança grafológica entre as pinceladas das suas pinturas e a rápida rubrica fortemente personalizada do próprio artista. Uma ordenação exclusivamente baseada no tempo, e, neste caso, no tempo interior, manifestado através dos gestos espontâneos, substituiu a ordenação exclusivamente visual dos euclidianos, tanto a ordenação estática como a ordenação dinâmica.

De certo modo, a nível teórico, esta distinção entre euclidianos e informalistas seguiu a tradicional oposição da forma e da cor, do contorno e da mancha livre, do intelecto e do sensorial.

Uma forte sugestão de elasticidade constitui a expressividade e o sentido de ordenação a partir da cor e do traço impulsivo de Van Gogh e de Matisse. Repare-se nas modificações formais de Van Gogh exigidas pelos acertos das áreas de cor. Nestes acertos de Van Gogh, como nos de Cézanne, pode começar a detectar-se uma «geometria da cor». Paul Cézanne estava consciente desta nova apreensão das formas e do espaço quando disse a Emile Bernard: «Ao mesmo tempo que se pinta, desenha-se. Quanto mais harmónica se torna a cor, mais nítido é o desenho. Quando a cor atinge a sua riqueza máxima, a forma atinge a máxima plenitude. No contraste e harmonia dos tons reside o segredo do desenho e daquilo que é modelado» (*Souvenirs sur Paul Cézanne*, Cartas e entrevistas, Paris, 1912).

É claro que os contemporâneos de Van Gogh e de Cézanne achavam que eles não sabiam desenhar, que eram desajeitados. Na realidade, eles estavam a adivinhar uma nova maneira de compreender as formas espaciais e simultaneamente inventavam um tipo mais alargado de registo, profundamente expressivo tanto dos dados emocionais como intelectuais. Esse tipo de registo tende a transformar a arte do desenho numa espécie de escrita. Repare-se na analogia entre desenho e caligrafia de Van Gogh, que revela uma noção de forma que se aproxima da da arte japonesa. Reveja-se a deformação topológica de Matisse nas suas figuras aderentes ao plano. Algumas parecem ter sido desenhadas numa superfície elástica, como um balão que depois fosse inchado pelo sopro, mostrando a sua alteração contínua, sem rupturas.

Na problemática da pesquisa de estruturas formais coerentes, é oportuno (e saboroso) lembrar algumas palavras do cientista Henri Poincaré. «Suponhamos um modelo qualquer e a cópia desse modelo recopiado por um desenhador desajeitado; as proporções são alteradas; as rectas traçadas por uma mão trémula sofrem desvios desagradáveis e apresentam

curvaturas desastreadas. Do ponto de vista da geometria métrica, mesmo da geometria projectiva, as duas figuras não são equivalentes; elas são-no, pelo contrário, do ponto de vista da *Analysis situs*».

A escolha das geometrias é importante no conjunto das convenções em que terá de apoiar-se a nova visão, em correlação com as concepções de espaço, tempo e matéria que forem usadas na vida prática e nas estruturas científicas.

O homem já não se considera o centro do universo; mas, na sua vida quotidiana, a substituição da natureza pela artificialidade dá-lhe uma responsabilidade temporal inédita e muito grande. Basta pensar nos actuais problemas ecológicos e no *stress* das grandes cidades.

Sem dúvida, é entretanto mais fácil verificar a relação entre a pintura ou escultura de execução *controlada* e a técnica, na medida em que ambas utilizam disciplinas comuns emanadas da ciência *feita*. É, por exemplo, o caso de Gabo, que corporiza superfícies geradas matematicamente. É também o caso de uma escultura-móvel do inglês Kenneth Martin. É uma escultura pendurada que gira incessantemente. É, portanto, uma forma geométrica, regrada, que integra o tempo. E há também esculturas deste género, como as de Max Bill, que representam modelos que serviram algumas reflexões de ordem topológica, como a superfície de Mobius.



Gabo



Max Bill

Em novo tipo de *trompe-l'œil*, Vasarely constrói volumes e perspectivas contraditórias, que, aniquilando-se mutuamente, reintegram o plano das duas dimensões. Essas perspectivas contraditórias, com a sua força *gestáltica*, conduzem a uma vibração, nos limites euclidianos. Leis estatísticas organizam estes quadros, e as formas volumétricas sugeridas precisam de tempo para vibrarem.

A vibração óptica é rapidamente alcançada nas pinturas do americano Larry Poons. Numa das suas pinturas *op*, as pintas azuis são fortemente contrastantes com o fundo cor de laranja. Agridem a retina. O cérebro reage, e a percepção «saltita», entre as diversas estruturas de arrumação das pintas azuis. Estas diversas subestruturas estão dissimuladas, porque estão sobrepostas.

Num quadro de Sofia Tauber-Arp, Max Bill descobriu várias organizações simples simultâneas. Algumas destas subestruturas são percebidas com muita frequência, outras com pouca frequência, conforme o momento da contemplação. Sofia Tauber-Arp e outros artistas da mesma tendência vanguardista têm uma intuição muito viva destas leis estatísticas da percepção.

Às tensões provocadas por estas rebuscadas contradições do acto da percepção juntam-se muitos outros dados psicofísicos, criando-se um movimento virtual nesta arte de efeitos ópticos violentos. Estes efeitos são universais. «A obra age sobre o aparelho psico-físico perceptivo, e não sobre a base psicológica cultural do espectador», diz Mestrovic, que caracteriza ainda estas pesquisas como apresentadoras de «formação e não formas». Sugere-se a desmaterialização do suporte para o reencontrar como campo de energia. Há quadros de Vasarely que se nos apresentam como campos de pura energia luminocromáticas. Vasarely levou ao máximo de energia as descobertas ópticas dos cientistas e dos pintores do século XIX, nomeadamente Chevreul e os impressionistas.

Ao referir-se às quatro dimensões, é significativo que Vasarely prefira falar do espaço a partir do tempo, ou seja, para ele, o tempo não é a quarta dimensão, mas a primeira. Diz ele: «O espaço é o lugar de um fenómeno que se realiza ao longo do tempo». Também os físicos, actualmente, preferem definir metro-padrão a partir de um período de tempo: o espaço será medido em função da constância da velocidade da luz no vazio. O tempo mede o espaço. A matéria fornece o relógio atómico, com seus rigorosos períodos de desintegração. Repito: com o comportamento da matéria mede-se o tempo, e com o tempo mede-se o espaço. Instaura-se deste modo uma nova correlação dos enquadramentos mentais com que os homens encaram o mundo.

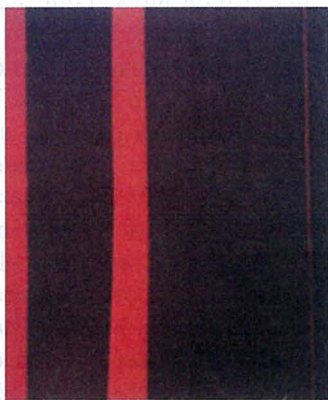
Na pintura, desde o início da sua moderna evolução cromática, com o impressionismo, instaurou-se o processo de produção de efeitos de luz devido à interacção das cores puras, *produção* que exige um certo tempo de percepção. E o espaço transformou-se em campo de energia luminocromática.

Onde os não-euclidianos são mais radicais é na vontade de revelação do próprio processo pictórico que inclui o *sujeito*, pelo que a arte deles

assume talvez a consciência mais aguda da necessidade de manter na nossa época tecnicista a *intencionalidade estética*. Na pintura de Pollock estão registados os signos tal como nasceram da acção do corpo inteiro, numa prática irreductível à repetição mecânica, desencadeadora porém das capacidades expressivas espontâneas que permitem renovar perpetuamente a criação e a contemplação estéticas, subjectivamente dinâmicas. É por aqui que a *arte de hoje* estará talvez na raiz de *ciências futuras*, assim como a arte de ontem não deixou de dar a sua contribuição para o aparecimento das ciências actuais. Eis a razão pela qual é já muito significativa a relação desta arte com a topologia, e mesmo com geometrias mais recentes. Por vezes, a arte é uma presciência. Almada Negreiros tinha a convicção de que «a arte precede a ciência, a perfeição precede a exactidão».

De qualquer modo, através da arte actual, está-se longe do espaço estático do Renascimento, está-se longe da concepção do quadro concebido como janela, está-se longe da redução do observador a um olho fixo. Está-se também longe das pinturas e esculturas concebidas como ilustrações.

O aparecimento de novos objectos artísticos com movimento real não reaproxima a arte de qualquer tipo de ilustração. Alguns artistas utilizam forças electromagnéticas (Takis) e as suas obras familiarizam o público com a multiplicidade dos *campos de forças*. Van Thienen utiliza o vento e o som. Também no âmbito da *Colour Field Painting*, Barnett Newman afirma o valor plano, *flat*, da pintura. Os quadros de Barnett Newman têm grandes dimensões, para que ocupem muito espaço real e as cores envolvam o espectador.



Barnett Newman, Adam

Neste campo de energia cromática, o espectador não tem apenas uma reacção fisiológica, mas também espiritual. A verticalidade exprime o sublime. Barnett Newman negou a estabilidade da pirâmide numa escultura alongada em aço inoxidável. É um obelisco invertido, que se torna indefinido na sua vertical ascendente. O «aqui e agora», onde o observador se encontra, está em contacto imaginário com o infinito da abóbada celeste.

Com as viagens interplanetárias, à medida que o homem deixa de sentir como constante a força gravítica, também deixa de ter necessidade de fazer obedecer os sinais espaciais à vertical (força gravítica na vida habitual) e à horizontal (base).

Precedendo Newman, David Smith tinha já concebido esculturas que, intensificando o dinamismo dos seus signos, ou subindo-os muito, contrariaram o tipo de concepção escultórica que parte do necessário equilíbrio de massas em relação à vertical e à base de sustentação. Como se sabe, a existência de um pedestal tem mantido o cariz antropomórfico da escultura moderna, dificultando o seu desenvolvimento abstraccionista. Depois de Smith, o inglês Anthony Caro e outros libertaram-se mais desse antropomorfismo, realizando esculturas sem pedestal, articulando planos ao longo do chão, recorrendo a chapas e redes metálicas fortemente coloridas.

A cor artificial da indústria é francamente assumida na escultura actual. É o caso de Anthony Caro e de Isaac Witkin. Os materiais de plástico também facilitaram novas concepções escultóricas, em especial o poliuretano, que se expande. Isto leva alguns escultores, como o francês César, a adoptarem uma atitude nova, tanto na execução como na concepção.

Uma topologia intuitiva ressurgue nestas «expansões», mas desdramatizada, ao contrário do que acontecia no tempo de Van Gogh. O mesmo acontece com os desenhos realizados pela arte cibernética, que pode alterar as proporções de uma figura, mecanicamente, até a tornar irreconhecível. Mas, por outro lado, o desenho cibernético reintroduz algo que a modernidade tinha abandonado: a minúcia. Uma espécie de confiança no traçado cibernético faz mergulhar o olhar do espectador em zonas de texturas inabituais, que, em desenhos não mecânicos, ou sem regras, o espectador não levaria em consideração. Isto mostra que o conhecimento das regras auxilia a percepção. Mas o que é fascinante é a rapidez com que nas obras de arte o espectador adivinha o regramento, mesmo que não saiba verbalizá-lo.



Computer Art

A minúcia pode ser abolida, por redundância ou pela introdução de factores aleatórios, na *copy-art* e, em geral, na *random-kunst*, ou arte mecânica que aproveita e regista os dados do acaso.

Nestas artes que usam e abusam da tecnologia electromagnética, o artista pode parecer cada vez mais distante da obra. Na utilização das novas técnicas, a comunicação intersubjectiva, que se espera da arte, parece não existir. Todavia, alguns artistas utilizam o vídeo para ajudar a redescobrir o corpo e outros começam já a divertir-se com os raios *laser*, desenvolvendo a holografia.

Noutros casos, o próprio artista está presente, lembrando-nos drasticamente, na *body art* (ou arte do próprio corpo) e na *performance*, que a arte simboliza pela exemplificação. É através da exemplificação que a arte simboliza.



Body Art

Através de uma documentação fotográfica, pode ver-se a mão do artista italiano Penone apertando uma árvore. A mão do artista é depois substituída por uma mão de ferro, que não deixará crescer aquele bocado de vegetal.

Presença e exemplo, através do próprio corpo, ou simplesmente através da mão, a arte é um *fazer* que conduz a *novas ideias*.

F. J. Craveiro de Carvalho
Departamento de Matemática
Universidade de Coimbra

Sem título, 2000, técnica mista, dimensões variáveis

*Ah, there's no excitement to beat the
excitement of proving a theorem! Until
you find out the next day that it's wrong!*

Cathleen Synge Morawetz

Barbara Hepworth, Carlos de Oliveira, Wallace Stevens são figuras marcantes em vários universos culturais do século XX. Procurarei usar alguns aspectos do seu trabalho como pontos de apoio possíveis para estabelecer algumas pontes entre a arte e a matemática. A abordagem é ingénuo, naturalmente, e deverá ser lida, basicamente, como indicativa da importância que concedo à arte como espaço de fruição do mundo.

1. SALTO EM ALTURA

«Fernando Pessoa dizia:

Aconteceu-me um poema.

A minha maneira de escrever fundamental é muito próxima deste 'acontecer'. O poema aparece feito, emerge, dado (ou como se fosse dado). Como um ditado que escuto e noto.»

São palavras de Sophia de Mello Breyner Andresen (1972), que, mais adiante, continua:

«Deixar que o poema se diga por si, sem intervenção minha (ou sem intervenção que eu veja), como quem segue um ditado (que ora é mais nítido, ora mais confuso), é a minha maneira de escrever.»

Não creio que posto perante uma questão semelhante houvesse algum matemático que pudesse dar uma resposta próxima da de Sophia. Tive no Verão de 1997 a oportunidade de entrevistar o matemático norte-americano Robert Osserman (F. J. Craveiro de Carvalho *et al.*, 1999). Osserman é o autor do livro de divulgação *Poesia do Universo* (Robert Osserman, 1995), onde, a certa altura, refere uma afirmação de David Hilbert sobre um estudante que decidira abandonar a matemática:

Bom, ele não tinha imaginação suficiente para ser um matemático.

Perante a minha pergunta sobre o que é necessário para se «fazer», se ter um matemático, Osserman referiu características como a intuição e a visualização geométricas, o perfeito domínio das diferentes técnicas, a capacidade de cálculo, o conhecimento enciclopédico, existindo em conjunto ou separadamente, mas concluiu de uma forma definitiva:

A única qualidade que nunca vi ausente é a disposição para trabalhar duramente.

Num vértice distante da posição de Sophia temos um poeta como Carlos de Oliveira. O excerto seguinte é formado por parte do penúltimo e pelo segmento final do poema «Salto em Altura», publicado inicialmente no volume *Entre duas Memórias* (Carlos de Oliveira, 1971) (onde, a despropósito, diga-se, aparece um dos poemas que tornam maior a poesia portuguesa. Refiro-me à «Descrição da Guerra em Guernica»).

[...]
*condensa-se o rumor das primeiras
palavras: afinal, são elas;
e logo que os seus voos;
anteriores à escrita; as precipitam
no papel, começa-se a escrever.*

*O saltador em altura
conseguiu transpor
os dois metros e vinte;
músculos a ascenderem
só por si; o treino, a obsessão: à neve,
no estádio sem ninguém;
este filme analisa,
ao retardador, cada um dos seus saltos:
o sonho a decompor-se;
a refazer-se; em fotogramas
sucessivos; como disse,
a primeira forma é ainda
elástica; as outras endurecem
no ar, mais angulosas;
mas todas pesam,
elaborando as leis da queda:
e caem; graves; reduzidas
ao espaço do seu peso.*

Carlos de Oliveira referir-se-á ao seu trabalho poético. No entanto, é-me possível, apoiado na minha experiência de investigador matemático, isolar algumas das características do meu próprio trabalho: a leveza e a alegria das primeiras ideias (*a primeira forma é ainda // elástica*), pequenas construções mentais (*os seus voos; // anteriores à escrita*), o esforço e a solidão (*músculos a ascenderem só por si; o treino... // no estádio sem ninguém*), a verificação exaustiva e o desalento (*este filme analisa, // ao retardador, cada um dos seus saltos: // o sonho a decompor-se; a refazer-se*), os resultados que acabam por se impor como se de forma natural e evidente (*e caem; graves; reduzidas // ao espaço do seu peso*) e, finalmente, o conseguimento (*o saltador em altura // conseguiu transpor // os dois metros e vinte*) e o processo a recomençar.

Há, sem dúvida, na história dessas coisas, registos de exemplos matemáticos de revelação que, aparentemente, contrariam o que acabo de referir. Digo aparentemente porque esses instantes perfeitos apenas ocorrem depois de trabalho árduo e continuado, que estará ausente em Sophia. Citarei dois, dos mais emblemáticos, possivelmente.

O primeiro exemplo passa-se com W. R. Hamilton e tem a ver com a sua tentativa de obter em dimensões mais altas, começando com a dimensão 3, um produto análogo ao dos números complexos. Era o chamado problema da «multiplicação dos ternos». A história ocorreu em 16 de Outubro de 1843 (S. A. Robertson, 1974).

Hamilton caminhava com a esposa ao longo do Royal Canal a caminho de uma sessão da academia em Dublin. Embora fossem falando de vez em quando, a sua mente estava ocupada com outros pensamentos que, finalmente, deram resultado, culminando o trabalho intenso de muitos anos. Hamilton não resistiu e, ao passarem pela Brougham Bridge, gravou na pedra com uma faca as igualdades fundamentais do produto dos quaterniões:

$$i^2 = j^2 = k^2 = i j k = -1.$$

O outro caso passou-se com Henri Poincaré, que conta (1947):

Nessa altura deixei Caen, onde então vivia, para participar numa conferência geológica organizada pela Escola de Minas. As peripécias da viagem fizeram-me esquecer os meus trabalhos matemáticos. Quando chegámos a Coutances subimos para um autocarro para ir não sei aonde. No momento em que punha o pé no degrau veio-me a ideia, sem que aparentemente nada nos meus pensamentos anteriores me tivesse preparado para isso, de que as transformações usadas para definir as funções fuchsianas eram idênticas às da geometria não euclidiana. Não o verifiquei, não tive tempo para o fazer, uma vez que reatei a conversa logo que me sentei, mas senti logo uma certeza absoluta. De volta a Caen verifiquei o resultado para tranquilizar o meu espírito.

Sem o dramatismo destes dois exemplos, todo o investigador matemático poderá contar pequenas histórias semelhantes. Mas o que há aqui a realçar é que a fluência e a facilidade matemáticas, tão apreciadas, são sempre muito mais conseguidas do que espontâneas. Como de resto acontece ainda com alguma poesia. O exemplo seguinte chega-me pela mão acolhedora de Eugénio de Andrade (1995).

Há um belíssimo poema de Antonio Machado que me acompanha há mais de vinte anos. É um poema breve, um dos raros poemas que sei de cor:

La plaza tiene una torre,
la torre tiene un balcón,
el balcón tiene una dama,

la dama una blanca flor.
Ha pasado un caballero
— quién sabe por qué pasó! —
y se ha llevado la plaza
con su torre y balcón,
com su balcón y su dama,
su dama y su blanca flor.

São só dez versos, todos de arte menor, como se diz nos compêndios, e não há nada aqui que não seja maior. Nada falta, nada sobra: tudo é simultaneamente leve e denso, límpido e misterioso, simples e complexo... Pois bem, o poema de Machado, de tão exemplar espontaneidade, teve uma primeira versão de nada menos nada mais que vinte e oito versos e, desses vinte e oito versos, só dois passaram intactos, e um deles com diferença de pontuação, à versão definitiva...

Não menos extremas são as palavras de João Cabral de Melo Neto (Arnaldo Saraiva, 2000):

Tem poemas meus que levaram anos a escrever, como aquele «Tecendo a manhã» de A Educação pela Pedra: passei praticamente dez anos para o concluir¹.

Para as palavras de João Cabral encontramos paralelo na afirmação seguinte de E. C. Zeeman (2001):

No mês passado² consegui provar um teorema que conjecturei há vinte e cinco anos sobre a teoria da proporção de Eudoxus.

2. UMA SERPENTE ∞

Há outros sinais mais visíveis que denunciam a proximidade entre a poesia e a matemática. Um olhar, que nem precisa de ser muito atento, sobre a poesia portuguesa do século XX permite detectar, com facilidade,

¹ Trata-se de um poema com dezasseis versos.

² Janeiro de 2000.

a utilização de palavras conotadas com ideias matemáticas, desviadas ou não da sua área semântica. Eis alguns exemplos significativos:

A ESFERICIDADE OU FEROCIDADE

*qualidade de que gozam alguns
sólidos quando se viram
para nós³*

(Luiza Neto Jorge, 1966)

CATEDRAL DE BURGOS

*A catedral de Burgos tem trinta metros de altura
e as pupilas dos meus olhos dois milímetros de abertura.*

Olha a catedral de Burgos com trinta metros de altura!

(António Gedeão, 1971)

AS COORDENADAS LÍRICAS

Desviou-se o paralelo um quase nada

...

*A geométrica forma dos meus passos
procura um mar redondo.*

...

*Em cada esfera desenho o meu refúgio
— as minhas coordenadas.*

(Fernando Botelho, 1951)

³ Há uma outra versão, posterior, e que surge, por exemplo, em *Os Sítios Sitiados*, deste poema:

*qualidade perigosa a de alguns
sólidos quando perdidos se viram
para nós*

Optei pela versão original, por, em minha opinião, a segunda versão ser significativamente mais fraca.

E poderíamos não ficar por aqui. Melhor do que eu, sobre este assunto falou Manuel Portela na conferência de abertura da exposição «A Matemática na Poesia Portuguesa», realizada no Departamento de Matemática da Universidade de Coimbra em 1999.

Por vezes o poema revela, da matemática, uma proximidade mais subtil e mais profunda. Um exemplo é «A história do frasco»⁴ do poeta americano Wallace Stevens. O início é agressivo e insólito:

*Pousei um frasco no Tennessee,
Um frasco redondo, sobre um outeiro.
E o matagal bravio veio
Cercar aquele outeiro.*

*O matagal ergueu-se, numa vénia,
e espraiou-se em seu redor, alinhado.
Redondo o frasco no chão,
E alto no seu porte aéreo.*

*Dominou tudo ali à volta.
Um frasco cinzento e vazio.
Sem dar por ave ou moita,
Como nada mais no Tennessee⁵.*

Como diz W. M. Priestley (1990), o frasco de Stevens talvez fosse um poema. Ou um momento de paixão, acrescento eu. Contudo o poema tem-se prestado a uma leitura matemática, baseada na ideia da ordenação que um sistema de referência e as coordenadas que lhe estão associadas permitem introduzir. Aqui o sistema seria tridimensional, com a terceira dimensão dada pela verticalidade do frasco e, indo talvez um pouco longe de mais, pode associar-se, como faz Jonathan Holden (1990), a letra *O*, de origem do sistema de referência, à boca redonda do frasco.

Mais perto de nós temos uma das «Tisanas» de Ana Hatherly (1997):

*Era uma vez uma serpente infinita. como era infinita não
havia maneira de se saber onde estava a sua cabeça. de
cada vez que se lhe tirava uma vértebra não fazia falta
nenhuma. podia-se mesmo parti-la deslocá-la emendá-la.
ficava sempre infinita. quem quisesse levar-lhe um bocado
para casa podia pô-lo na parede e contemplar um
fragmento da serpente infinita.*

⁴ «Anecdote of the jar».

⁵ Este poema encontra-se também traduzido em *Ficção Suprema*, Wallace Stevens, publicado pela editora *Assírio e Alvim*.

Quem tiver alguma familiaridade com a teoria de conjuntos reconhecerá tratar-se de uma variação sobre o facto de existirem conjuntos com o mesmo número de elementos que algumas das suas partes. Em particular, o número de elementos de um conjunto infinito não se altera se se lhe retirar um subconjunto finito (Klaus Jänich, 1984).

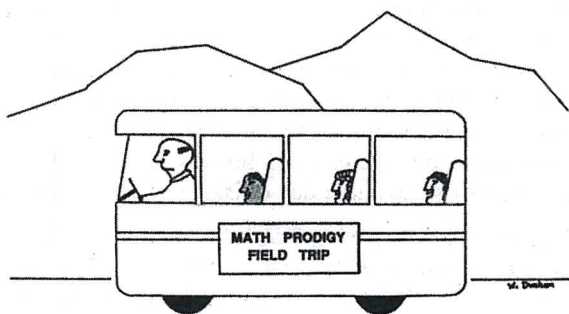
A mesma ideia está subjacente no *cartoon* que se reproduz a seguir, que se refere a uma viagem de estudo de pequenos prodígios matemáticos e se baseia numa *nursery rhyme* muito cantada nesse tipo de saídas escolares em países de língua inglesa:

*Ten green bottles, hanging on a wall,
(repete)
And if one green bottle should accidentally fall
There'll be nine green bottles, hanging on the wall.*

E continua até chegar a:

*One green bottle, hanging on a wall
(repete)
And if one green bottle should accidentally fall,
There'll be no green bottles, hanging on the wall.*

O *cartoon* é tirado da revista *College Mathematics Journal*⁶ (William Dunham, 1992).



"No bottles of beer on the wall,
No bottles of beer.
If one of the bottles should happen to fall,
No bottles of beer on the wall."

Contributed by William Dunham, Hanover College, Hanover, IN

⁶ Reprodução autorizada pela Mathematical Association of America. Reservados todos os direitos.

3. AH, MAS É UM TORO!

Aproveitemos o *cartoon* como pausa de passagem para uma outra forma de arte, a que sou particularmente sensível como géometra e topologista.

Barbara Hepworth nasceu em 1903 no Yorkshire, Reino Unido, tendo morrido de forma trágica em 1975 em St. Ives, na Cornualha. O seu percurso como escultora acompanhou o de outras figuras famosas: Henry Moore, de quem foi colega e de cuja sombra terá tido alguma dificuldade em se libertar, Naum Gabo e Ben Nicholson, este a quem esteve ligada afectivamente.

Duas características essenciais no trabalho de Hepworth são o uso de buracos enormes, que ocupam por vezes toda a espessura do material em utilização, e o uso de fios tensos. Não se discutirão aqui questões de prioridade, mas refira-se que estas duas características se encontram também no trabalho de Henry Moore, assim como o uso de fios aparece no trabalho de Naum Gabo.

Os fios tensos lembram imediatamente as rectas numa superfície regrada ou, no caso dos elementos equivalentes em desenhos, o campo de rectas tangentes a uma dada curva. Tal não é acidental, já que Hepworth na sua juventude estudou modelos matemáticos em Londres e Paris, mas «deixou a ideia ficar adormecida no seu espírito até ser capaz de a usar emocional e não matematicamente» (A. M. Hammacher, 1993). De facto, essas famílias de rectas surgem em desenhos cujo motivo dominante é o *crystal*, enquanto nas esculturas a transparência possibilitada pelo uso de fios surge como contraste às ideias de limite ou cerco que materiais como a madeira, a pedra implicam (A. M. Hammacher, 1993). O uso do buraco ou furo confere ao trabalho de Barbara Hepworth um peso topológico enorme. Um matemático conhecedor de alguns aspectos da teoria das superfícies (W. S. Massey, 1967) exclamará:

Ah, mas é um toro! (simples ou múltiplo conforme as circunstâncias).

ou

[...] *mas topologicamente é uma esfera.*

Não é só essa atracção topológica, para usar palavras de E. C. Zeeman (2001) ao referir-se a Moore e Hepworth, que me interessa aqui. Toca-me ainda mais a permuta entre matemática e figuração, patente nalguns trabalhos que procurarei evidenciar com exemplos retirados de A. M. Hammacher, 1993.

Começemos com as esculturas das figuras 41, 42, 46 ou 51, por exemplo. Trata-se de esculturas compostas por elementos de uma geometria muito despojada e, no entanto, na sua minimalidade, remetem, todas elas, para o intimismo doméstico das naturezas mortas.

Já a uma escultura como a da figura 39 não surpreenderia que em vez de *Two Forms* a autora tivesse dado um título mais referencial. A ideia do par *mãe e filho*, que ocorre natural e imediatamente, não impede a mente treinada matematicamente, e conhecedora da igualdade que relaciona os pontos críticos de uma função com o *número de Euler* (D. R. J. Chillingworth, 1976), de ser instintivamente levada a tentar verificar essa igualdade, no caso da superfície maior, para as *funções altura* determinadas pelos vectores unitários normais ao plano de apoio. Associando 1 (resp. — 1) a cada ponto da superfície em que o plano tangente é paralelo ao plano de apoio e, localmente, a superfície está inteiramente para um dos lados desse plano (resp. — não está) e somando obter-se-á 2, *número de Euler* da esfera.

4. COLAGENS, CITAÇÕES...

Limitei-me aqui a destacar apenas algumas áreas de contacto entre a arte e a matemática, não sendo este texto mais que uma colagem de apontamentos breves. Que seja uma colagem poderá não ser inteiramente desapropriado em face do contexto. A *referência*, a *citação*, a *colagem* são instrumentos usados em arte e não o são menos em matemática. Como diz, autorizadamente, W. T. Gowers na sua notável contribuição para *Mathematics: Frontiers and Perspectives* (V. Arnold *et al.* 2000), «o progresso é frequentemente o resultado de combinações inteligentes de uma larga variedade de resultados existentes».

AGRADECIMENTOS

A Carlota Simões, pela preparação das transparências relativas ao trabalho de Barbara Hepworth apresentadas na sessão de *Teias Matemáticas*; a Manuel Portela, pela tradução, excelente, do poema de Wallace Stevens.

Referências

- Eugénio de Andrade (1995) – *Rosto Precário*, Fundação Eugénio de Andrade.
- Sophia de Mello Breyner Andresen (1972) – *Dual*, Moraes Editores.
- V. Arnold *et al.* (2000) – *Mathematics: Frontiers and Perspectives*, American Mathematical Society.
- Fernando Botelho (1951) – *As Coordenadas Líricas*, Távola Redonda – 1.^a série.
- D. R. J. Chillingworth (1976) – *Differential Topology with a View to Applications*, Pitman Publishing.
- F. J. Craveiro de Carvalho *et al.* (1999) – *Our Speaker Today*, Departamento de Matemática da Universidade de Coimbra.
- William Dunham (1992) – *Cartoon*, College Mathematics Journal, vol. 23, #1.
- António Gedeão (1971) – *Poesias Completas*, Portugalíia.
- A. M. Hammacher (1993) – *Barbara Hepworth*, Thames and Hudson.
- Ana Hatherly (1997) – *351 Tisanas*, Quimera.
- Jonathan Holden (1990) – The Poetry of Wallace Stevens, *Math. Intelligencer*, vol. 12, #1.
- Klaus Jänich (1984) – *Topology*, Springer-Verlag.
- Luiza Neto Jorge (1966) – *O Seu a Seu Tempo*, Ulisseia.
- W. S. Massey (1967) – *Algebraic Topology: an Introduction*, Harcourt, Brace & World, Inc.
- Carlos de Oliveira (1971) – *Entre Duas Memórias*, Publicações Dom Quixote.
- Robert Osserman (1995) – *Poesia do Universo*, Difusão Cultural.
- Henri Poincaré (1947) – *Science et Méthode*, Flammarion.
- Manuel Portela (1999) – *A Matemática na Poesia Portuguesa*, Departamento de Matemática da Universidade de Coimbra.
- W. M. Priestley (1990) – Mathematics and Poetry: How Wide the Gap?, *Math. Intelligencer*, vol. 12, #1.
- S. A. Robertson (1974) – *Topological Surprises (Unforeseen Phenomena in High Dimensional Spaces)*, uma lição inaugural, Universidade de Southampton.
- Arnaldo Saraiva (2000) – *Conversas com Escritores Brasileiros*, edição do Congresso Portugal – Brasil Ano 2000, Porto.
- Wallace Stevens (1991) – *Ficção Suprema*, tradução de Luísa Maria Lucas Queiroz de Campos, Assírio & Alvim.
- E. C. Zeeman (2001) – entrevista em *A Jar in Tennessee*, F. J. Craveiro de Carvalho e Jorge Picado, Departamento de Matemática da Universidade de Coimbra.

(Página deixada propositadamente em branco)

CIÊNCIA ABERTA

1. O JOGO DOS POSSÍVEIS
François Jacob
2. UM POUCO MAIS DE AZUL
H. Reeves
3. O NASCIMENTO DO HOMEM
Robert Clarke
4. A PRODIGIOSA AVENTURA
DAS PLANTAS
Jean-Marie Pelt/Jean-Pierre Cuny
5. COSMOS
Carl Sagan
6. A MEDUSA E O CARACOL
Lewis Thomas
7. O MACACO, A ÁFRICA
E O HOMEM
Yves Coppens
8. OS DRAGÕES DO ÉDEN
Carl Sagan
9. UM MUNDO IMAGINADO
June Goodfield
10. O CÓDIGO CÓSMICO
Heinz R. Pagels
11. CIÊNCIA: CURIOSIDADE
E MALDIÇÃO
Jorge Dias de Deus
12. O POLEGAR DO PANDA
Stephen Jay Gould
13. A HORA DO
DESLUMBRAMENTO
H. Reeves
14. A NOVA ALIANÇA
Ilya Prigogine/Isabelle Stengers
15. PONTES PARA O INFINITO
Michael Guillen
16. O FOGO DE PROMETEU
Charles Lumsden/Edward O.
Wilson
17. O CÉREBRO DE BROCA
Carl Sagan
18. ORIGENS
Robert Shapiro
19. A DUPLA HÉLICE
James Watson
20. OS TRÊS PRIMEIROS
MINUTOS
Steven Weinberg
21. «ESTÁ A BRINCAR,
SR. FEYNMAN!»
Richard P. Feynman
22. NOS BASTIDORES
DA CIÊNCIA
Sebastião J. Formosinho
23. VIDA
Francis Crick
24. SUPERFORÇA
Paul Davies
25. QED — A ESTRANHA TEORIA
DA LUZ E DA MATÉRIA
Richard P. Feynman
26. A ESPUMA DA TERRA
Claude Allègre
27. BREVE HISTÓRIA
DO TEMPO
Stephen W. Hawking
28. O JOGO
Manfred Eigen/Ruthild Winkler
29. EINSTEIN TINHA RAZÃO?
Clifford M. Will
30. PARA UMA NOVA CIÊNCIA
Steven Rose/Lisa Appignanesi
31. A MÃO ESQUERDA
DA CRIAÇÃO
Jonh D. Barrow/Joseph Silk
32. O GENE EGOÍSTA
Richard Dawkins
33. HISTÓRIA CONCISA
DAS MATEMÁTICAS
Dirk J. Struik
34. CIÊNCIA, ORDEM
E CRIATIVIDADE
David Bohm/F. David Peat
35. O QUE É UMA LEI FÍSICA
Richard P. Feynman
36. QUANDO AS GALINHAS
TIVEREM DENTES
Stephen Jay Gould
37. «NEM SEMPRE A BRINCAR,
SR. FEYNMAN!»
Richard P. Feynman
38. CAOS — A CONSTRUÇÃO
DE UMA NOVA CIÊNCIA
James Gleick
39. SIMETRIA PERFEITA
Heinz R. Pagels
40. ENTRE O TEMPO
E A ETERNIDADE
Ilya Prigogine/Isabelle Stengers
41. OS SONHOS DA RAZÃO
Heinz R. Pagels
42. VIAGEM ÀS ESTRELAS
Robert Jastrow
43. MALICORNE
Hubert Reeves
44. INFINITO EM TODAS
AS DIRECÇÕES
Freeman J. Dyson
45. O ÁTOMO ASSOMBRADO
P. C. W. Davies/J. R. Brown
46. MATÉRIA PENSANTE
Jean-Pierre Changeux/Alain
Connes
47. A NATUREZA
REENCONTRADA
Jean-Marie Pelt
48. O CAMINHO QUE NENHUM
HOMEM TRILHOU
Carl Sagan/Richard Turco
49. O SORRISO DO FLAMINGO
Stephen Jay Gould
50. EM BUSCA DA UNIFICAÇÃO
Abdus Salam/Paul Dirac/
/Werner Heisenberg
51. OBJECTOS FRACTAIS
Benoît Mandelbrot
52. A QUARTA DIMENSÃO
Rudy Rucker
53. DEUS JOGA AOS DADOS?
Ian Stewart
54. OS PRÓXIMOS CEM ANOS
Jonathan Weiner
55. IDEIAS E INFORMAÇÃO
Arno Penzias
56. UMA NOVA CONCEPÇÃO
DA TERRA
Seiya Uyeda
57. HOMENS E ROBOTS
Hans Moravec
58. A MATEMÁTICA
E O IMPREVISTO
Ivar Ekeland
59. SUBTIL É O SENHOR
Abraham Pais
60. FLATLAND — O PAÍS PLANO
Edwin A. Abbott
61. FEYNMAN — A NATUREZA
DO GÉNIO
James Gleick
62. COMIDA INTELIGENTE
Jean-Marie Bourre

63. O FIM DA FÍSICA
Stephen Hawking
64. UNIVERSO, COMPUTADORES
E TUDO O RESTO
Carlos Fiolhais
65. OS HOMENS
André Langaney
66. OS PROBLEMAS
DA BIOLOGIA
John Maynard Smith
67. A CRIAÇÃO DO UNIVERSO
Fang Li Zhi/Li Shu Xian
68. A MÁQUINA MÁGICA
A. K. Dewdney
69. O MELHOR DE FEYNMAN
Organização de Laurie M. Brown
e John S. Rigden
70. ÚLTIMAS NOTÍCIAS
DO COSMOS
Hubert Reeves
71. A VIDA É BELA
Stephen Jay Gould
72. OS PROBLEMAS
DA MATEMÁTICA
Ian Stewart
73. POEIRAS DE ESTRELAS
Hubert Reeves
74. A PALAVRA DAS COISAS
Pierre Laszlo
75. A EXPERIÊNCIA
MATEMÁTICA
Philip J. Davis/Reuben Hersh
76. EINSTEIN VIVEU AQUI
Abraham Pais
77. SOMBRAS DE
ANTEPASSADOS
ESQUECIDOS
Carl Sagan/Ann Druyan
78. O PRIMEIRO SEGUNDO
Hubert Reeves
79. A COMUNIDADE VIRTUAL
Howard Rheingold
80. UM MODO DE SER
João Lobo Antunes
81. SONHOS DE UMA TEORIA
FINAL
Steven Weinberg
82. A MAIS BELA HISTÓRIA
DO MUNDO
Hubert Reeves/Joël de Rosnay/
Yves Coppens/Dominique
Simonnet
83. O SÉCULO DOS QUANTA
João Varela
84. O FIM DAS CERTEZAS
Ilya Prigogine
85. A PRIMEIRA IDADE
DA CIÊNCIA
António Manuel Baptista
86. O QUARK E O JAGUAR
Murray Gell-Mann
87. A DIVERSIDADE DA VIDA
Edward O. Wilson
88. A LIÇÃO ESQUECIDA
DE FEYNMAN
David L. Goodstein/Judith
R. Goodstein
89. ORDEM OCULTA
John H. Holland
90. UM MUNDO INFESTADO
DE DEMÓNIOS
Carl Sagan
91. O RATINHO, A MOSCA
E O HOMEM
François Jacob
92. O ÚLTIMO TEOREMA
DE FERMAT
Amir D. Aczel
93. A MENTE VIRTUAL
Roger Penrose
94. SOBRE O FERRO
NOS ESPINAFRES
Jean-François Bouvet (org.)
95. BILÍÕES E BILÍÕES
Carl Sagan
96. CINCO EQUAÇÕES QUE
MUDARAM O MUNDO
Michael Guillen
97. A CIÊNCIA NO GRANDE
TEATRO DO MUNDO
António Manuel Baptista
98. CONCEITOS FUNDAMENTAIS
DA MATEMÁTICA
Bento de Jesus Caraça
99. O MUNDO DENTRO
DO MUNDO
John D. Barrow
100. A CULTURA CIENTÍFICA
E OS SEUS INIMIGOS
O LEGADO DE EINSTEIN
Gerald Holton
101. VIAGENS NO ESPAÇO-TEMPO
Jorge Dias de Deus
102. IMPOSTURAS INTELLECTUAIS
Alan Sokal/Jean Bricmont
103. O ESTRANHO CASO DO GATO
DA SR.^a HUDSON
Colin Bruce
104. AVES, MARAVILHOSAS AVES
Hubert Reeves
105. O HOMEM QUE SÓ GOSTAVA
DE NÚMEROS
Paul Hoffman
106. DECOMPONDO O ARCO-ÍRIS
Richard Dawkins
107. FULL HOUSE
Stephen Jay Gould
108. O UNIVERSO ELEGANTE
Brian Greene
109. GÖDEL, ESCHER, BACH
Douglas R. Hofstadter
110. O SIGNIFICADO DE TUDO
Richard P. Feynman
111. GENOMA
Matt Ridley
112. ZERO
Charles Seife
113. O MISTÉRIO DO BILHETE
DE IDENTIDADE E OUTRAS
HISTÓRIAS
Jorge Buescu
114. $E = mc^2$
David Bodanis
115. AS LIGAÇÕES CÓSMICAS
Carl Sagan
116. O DISCURSO PÓS-MODERNO
CONTRA A CIÊNCIA
António Manuel Baptista
117. O NOSSO HABITAT CÓSMICO
Martin Rees
118. OS GÊNIO DA CIÊNCIA
Abraham Pais
119. NOVE IDEIAS MALUCAS
EM CIÊNCIA
Robert Ehrlich
120. A COISA MAIS PRECIOSA
QUE TEMOS
Carlos Fiolhais
121. FEITICEIROS E CIENTISTAS
Georges Charpak/Henri Broch
122. A ESPÉCIE DAS ORIGENS
António Amorim
123. COMO CONSTRUIR UMA
MÁQUINA DO TEMPO
Paul Davies
124. O GRANDE, O PEQUENO
E A MENTE HUMANA
Roger Penrose
125. COMO RESOLVER
PROBLEMAS
G. Polya
126. DA FALSIFICAÇÃO DE EUROS
AOS PEQUENOS MUNDOS
Jorge Buescu
127. MAIS RÁPIDO QUE A LUZ
João Magueijo
128. O SIGNIFICADO
DA RELATIVIDADE
Albert Einstein
129. FRONTEIRAS DA CIÊNCIA
Rui Fausto, Carlos Fiolhais
e João Filipe Queiró
130. DA CRÍTICA DA CIÊNCIA
À NEGAÇÃO DA CIÊNCIA
Jorge Dias de Deus
131. CONVERSAS COM
UM MATEMÁTICO
Gregory J. Chaitin
132. Y: A DESCENDÊNCIA
DO HOMEM
Steve Jones
133. CRÍTICA DA RAZÃO
AUSENTE
António Manuel Baptista
134. TELAS MATEMÁTICAS
Mária Paula Serra de Oliveira

(Página deixada propositadamente em branco)

Instrumentos matemáticos complexos permitiram realizar com sucesso tarefas tão distintas como a programação de um voo a Marte, a previsão de resultados eleitorais, a explicação do funcionamento de alguns mecanismos do sistema nervoso, ou a abordagem crítica de obras de arte e de textos literários. Da ciência à sociedade, dos grandes avanços técnicos à solidez de uma argumentação lógica, a Matemática constrói teias de uma imensa flexibilidade resultante do carácter universal da sua linguagem.

Neste livro, personalidades de diferentes universos dão o seu testemunho sobre a forma como usam as teias matemáticas para tecer a sua própria visão do mundo.

MARIA PAULA SERRA DE OLIVEIRA é professora de Matemática na Faculdade de Ciências e Tecnologia da Universidade de Coimbra.

ISBN 972-662-970-5



9 789726 629702



gradiva



Imprensa da Universidade de Coimbra