

Autovalores e autovetores — além de $n = 2$

Carlos Tomei

Essas notas ¹ acompanham três encontros sobre o cálculo de autovalores e autovetores. O assunto é rico, mas habitualmente tratado de forma anacrônica nos textos e cursos de álgebra linear. É claro que três aulas não bastam para honrar o tema, mas alguns fatos banais levaram à escolha dos tópicos:

1. Geometria, sintética ou analítica, é linda, mas álgebra linear é outra coisa: álgebra linear deve tratar de tudo que é linear.
2. Se não existe uma fórmula explícita para a solução de um problema, devemos procurar uma aproximação da solução.
3. Qualquer assunto ganha quando apresentado com aplicações.

Resumindo o conteúdo dos encontros, vamos mostrar duas aplicações de *teoria espectral* (o estudo de autovalores e autovetores). Uma é combinatória (seções 1 e 2): a contagem de caminhos de comprimento dado em um grafo ligando dois vértices fixos. Obteremos fórmulas para contar, por exemplo, quantos caminhos vão de um vértice de um icosaedro ao vértice oposto, passando por mil arestas intermediárias. A outra aplicação é um modelo demográfico simples (seções 3 e 4): vamos estudar a evolução de uma população dividida em faixas etárias, supondo constantes as taxas de fertilidade e mortalidade de cada faixa. A conveniência da teoria espectral vai ficar evidente quando estudarmos, ainda que superficialmente, o *cálculo funcional de matrizes* (seção 5).

Nas duas aplicações, vão surgir matrizes grandes cujos autovalores interessa calcular. Na situação combinatória, vamos escolher matrizes para os quais os autovalores são de fato calculados, explorando a simetria do problema. Na situação demográfica, o processo se inverte: vamos nos inspirar nela para aprender técnicas de aproximação de autovalores e autovetores que são o ponto de partida de praticamente todos os algoritmos da área.

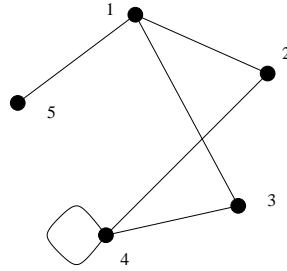
O texto não dá os detalhes de emprego do software, e deve ser pensado como o encarte que acompanha um CD. Na última seção, um kit de sobrevivência de álgebra linear. Parte do material é conteúdo do segundo curso de álgebra linear do Ciclo Básico da PUC-Rio.

1 Matrizes de adjacência

Um *grafo* é simplesmente um conjunto de pontos, os *vértices*, ligados por alguns segmentos, as *arestas*. Dois vértices são *adjacentes* se são ligados por uma aresta. A figura abaixo, por exemplo, é um grafo com cinco vértices, que foram rotulados $1, 2, \dots, 5$, e sete arestas (sem nome). Nele, os vértices 1 e 4 não são adjacentes. Note aliás que existe uma aresta começando e terminando no mesmo vértice 4. Um caminho entre dois vértices é exatamente o que você está pensando: 1243444315 é um caminho entre 1 e 5 de *comprimento* (isto é, o número de arestas percorridas) igual a 9. Um caminho pode ir e voltar sobre si mesmo muitas vezes, ou ficar rodando por uma aresta.

Um outro exemplo é o grafo que tem por vértices e arestas os oito vértices e as doze arestas de um cubo.

¹O autor agradece subsídios da FAPERJ e CNPq.



Grafos são tão simples que é difícil imaginar que dêem origem a perguntas interessantes. A situação é exatamente o contrário: para a maioria das perguntas, não se conhece uma resposta satisfatória. Uma pergunta (que não é tão simples, mas vai ser respondida nesse texto) é a seguinte: quantos caminhos de comprimento 20 existem ligando os vértices 1 e 5 do grafo acima? Outra: quantos caminhos *fechados* (isto é, começando e terminando no mesmo vértice) de comprimento 100 existem no grafo associado ao cubo? Note, aliás, que esses números são muito grandes, e que não são fáceis de escrever usando os termos combinatórios que você aprendeu no colégio (arranjos, combinações...) — na verdade, esses termos são inúteis para esse problema.

Não é óbvio que matrizes possam ajudar a responder a essa pergunta: o procedimento, que descrevemos a seguir, é empregado em problemas de otimização, matemática discreta, teoria de grafos (!), etc.

Dado um grafo G com n vértices, rotule seus vértices com os números $1, 2, \dots, n$ e monte a *matriz de adjacência* de G . Essa matriz, que vamos chamar de A , é $n \times n$, e só tem os números 0 e 1 em suas posições: na posição (i, j) de A (isto é, no cruzamento da linha i com a coluna j de A), colocamos um 0 se os vértices i e j não são adjacentes, e um 1 se o são. Por exemplo, para o grafo G da figura, a matriz A é

$$\begin{pmatrix} 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 \end{pmatrix}.$$

A matriz A é *simétrica* (isto é, suas posições (i, j) e (j, i) são iguais) — afinal, se o vértice i é adjacente a j , o j é adjacente a i .

Estamos prontos para responder à primeira pergunta: o número de caminhos de comprimento 20 ligando os vértices 1 e 5 do grafo da figura é dado pela posição $(1, 5)$ da matriz A^{20} . De maneira mais geral, vale o seguinte teorema, bastante inesperado.

Teorema: Seja G um grafo com n pontos rotulados pelos números $1, 2, \dots, n$, e com matriz de adjacência A . O número de caminhos de comprimento c ligando os vértices i e j é a posição (i, j) da matriz A^c . A própria matriz de adjacência conta os caminhos de comprimento $c = 1$ entre os

vértices i e j : em outras palavras, o teorema para esse caso segue diretamente da definição de matriz de adjacência. Para entender um pouco mais o resultado, vamos considerar o caso $c = 2$: vamos ver porque o número de caminhos ligando i a j com um único vértice intermediário é dado pela posição (i, j) da matriz A^2 . Para facilitar a notação, vamos escrever $M_{i,j}$ para representar a posição (i, j) da matriz arbitrária M . A posição (i, j) de A^2 (isto é, $(A^2)_{i,j}$) é dada pela expressão

$$(A^2)_{i,j} = \sum_{k=1}^n A_{i,k} A_{k,j}$$

(para convencer-se disso, eleve uma matriz pequena ao quadrado explicitamente — note, aliás, que essa expressão seria o comando central de um programa que eleva uma matriz ao quadrado). Só há uma maneira do produto $A_{i,k}A_{k,j}$ ser diferente de zero: as posições $A_{i,k}$ e $A_{k,j}$ têm que ser iguais a 1, o que quer dizer que o grafo original tem uma aresta ligando i a k e outra ligando k a j — mas isto é exatamente dizer que existe um caminho entre i e j de comprimento 2 passando por k . Como k pode ser qualquer ponto do grafo (isto é, k vai de 1 a n no somatório), e cada caminho deste tipo contribui com o número $1 = 1 \times 1$ para o somatório, o termo direito da expressão é o número total de caminhos de i a j de comprimento 2, e o teorema está demonstrado para esse caso. Para comprimentos arbitrários, a demonstração é muito parecida, combinando esse argumento com indução: tente.

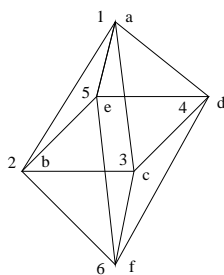
A moral é que um problema combinatório bastante complicado se reduziu a calcular potências de matrizes. Claro, a dificuldade agora é saber calcular potências de forma competente, e para isso temos o cálculo funcional, descrito na última seção.

Exercício: Que multiplicações matriciais você faria para calcular A^{20} ? Procure uma seqüência de poucas multiplicações — use um software para fazer essas contas.

Essas idéias se estendem para outros tipos de grafos. Por exemplo, poderíamos supor que as arestas são orientadas (e representar isso colocando setinhas em cada aresta), de maneira que seja possível ir de i a j , mas não voltar. Ou poderíamos supor que existem várias arestas ligando i a j . Em cada situação dessas, o problema de contar caminhos continua fazendo sentido, e uma maneira de resolvê-lo é considerar a definição adequada de matriz adjacência em cada caso. Mas isso fica para outra.

2 O octaedro e além

Considere o *grafo do octaedro* — seis vértices e doze arestas, com adjacências como no octaedro regular, desenhado abaixo. Nessa seção, vamos calcular os autovalores da sua matriz de adjacência A sem nem calcular o polinômio característico de A (veja a seção 5, para lembrar do vocabulário habitual): existem argumentos muito gerais que se aplicam bem a esse problema.



Os argumentos nos dispensam até de escrever a matriz A , mas não vamos ser obsessivos:

$$A = \begin{pmatrix} 0 & 1 & 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 1 & 0 \end{pmatrix}.$$

A primeira observação é que um vetor de \mathbb{R}^6 sobre o qual age A tem uma interpretação visual natural: ele corresponde a uma *disposição* de seis números nos vértices do octaedro — a rotulação

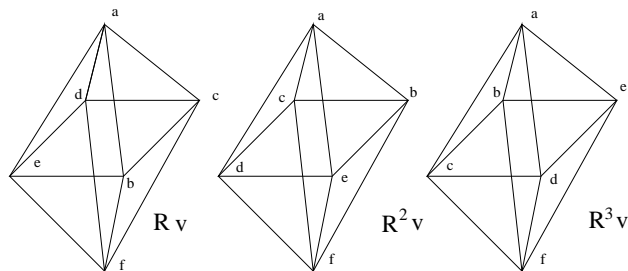
dos vértices usada para construir A é uma bijeção entre as coordenadas do vetor e os vértices do octaedro.

A ação de A sobre o vetor geral $v = (a, b, c, d, e, f)$ obtém

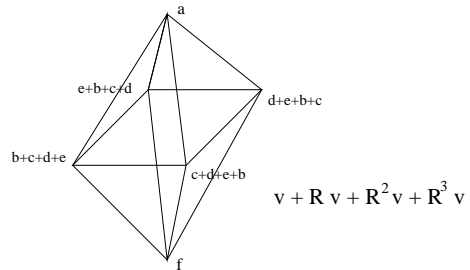
$$Av = (b + c + d + e, a + f + c + e, a + f + b + d, a + f + c + e, a + f + d + a, b + c + d + e).$$

Isso também pode ser visto sem fazer contas: disponha a, b, c, d, e e f pelos vértices do octaedro como na figura acima, e obtenha a nova disposição somando os valores encontrados nos vizinhos de cada vértice. Como exemplo, o vetor $(1, 1, 1, 1, 1, 1)$ é levado por A ao vetor $(4, 4, 4, 4, 4, 4)$ (você usou a matriz ou 'viu' o resultado?). Aliás, acabamos de aprender que $(1, 1, 1, 1, 1, 1)$ é um dos autovetores de A , associado ao autovalor 4. Calcular os outros cinco autovalores de A não vai ser tão fácil: vamos usar dois truques visuais.

Chame de R uma rotação de 90 graus em torno de um eixo do octaedro. Se você dispõe o autovetor v sobre os vértices do octaedro, e gira o octaedro por R fazendo com que os números sobre os vértices acompanhem a rotação, você encontra outra disposição que por sua vez, quando interpretada como um vetor Rv em \mathbb{R}^6 , é outro autovetor de A , associado ao mesmo autovalor λ . De fato, *as propriedades que caracterizam um autovetor são descritas em termos de adjacências de vértices, e não se alteram por uma rotação*. Nas figura abaixo, a disposição associada ao vetor v deu origem a três outros autovetores, por rotação sucessiva em torno do eixo vertical.



Mas em que isso simplifica o cálculo dos autovalores? É que se você soma autovetores associados a um mesmo autovalor, você encontra ainda outros autovetores associados a esse autovalor — em palavras mais técnicas, o conjunto de autovetores associado a um autovalor fixo é um subespaço. Agora, tome um autovetor v cuja coordenada a é não nula (afinal, todo autovetor tem alguma coordenada não nula: batize-a de a) e considere a média $(v + Rv + R^2v + R^3v)/4$ (que também tem alguma coordenada não nula — qual?), cuja disposição está apresentada na figura abaixo: você está vendo outro autovetor de A , associado ao mesmo autovalor de v , Rv , R^2v e R^3v .



Nos vértices mais alto e mais baixo, a disposição continua tendo a e f , mas nos quatro vértices equatoriais agora aparece mesmo número $g = (b + c + d + e)/4$. Moral: todo autovalor de A tem um autovetor *equatorial* $w = (a, g, g, g, g, f)$.

Se você aplica A a um vetor equatorial, você encontra outro — num vocabulário mais elegante, o subespaço de vetores equatoriais forma uma *subespaço invariante* por A . De fato, $Aw = (4g, a +$

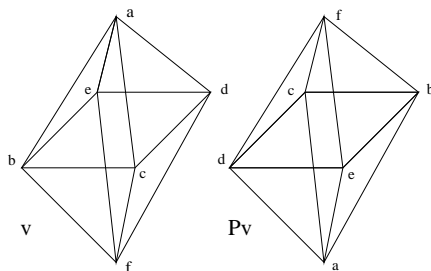
$f + 2g, a + f + 2g, a + f + 2g, a + f + 2g, 4g)$, claramente equatorial. De forma mais compacta, um vetor de três coordenadas (s, h, i) , que guarda os três números relevantes de um vetor equatorial é levado ao vetor $(4h, s + i + 2h, 4h)$. Moral ao quadrado: a matriz A tem os mesmos autovalores que a matriz

$$\tilde{A} = \begin{pmatrix} 0 & 4 & 0 \\ 1 & 2 & 1 \\ 0 & 4 & 0 \end{pmatrix}.$$

A matriz A tem seis autovalores e \tilde{A} só tem três! Isso não é paradoxal: cada autovalor de A é autovalor de \tilde{A} , mas as multiplicidades de cada autovalor podem baixar, e isso é o que acontece. Pelo teorema espectral (seção 5), A tem seis autovalores reais, e seis autovetores independentes. A matriz \tilde{A} , por sua vez, só tem autovetores que representam os autovetores equatoriais de A : autovetores desse tipo são raros.

Bom, ótimas notícias: o polinômio característico de \tilde{A} tem grau três, bem menor do que o polinômio característico de A . De quebra, já sabemos um de seus autovalores: 4. A essa altura, o cálculo dos autovalores de A se resume a resolver uma equação de segundo grau. Mas em vez de seguir essa trilha, vamos partir para mais um truque muito conveniente em outras situações.

Vamos levar em conta mais uma simetria evidente do problema. Quem tem um autovetor $v = (a, b, c, d, e, f)$ tem outro autovetor $Pv = (f, d, e, b, c, a)$ associado ao mesmo autovalor ('veja': P é a simetria pelo centro do octaedro).



Agora, como ao trabalharmos com a rotação R , note que $v + Pv = (a + f, b + d, c + e, b + d, c + e, a + f)$ e $v - Pv = (a - f, b - d, c - e, d - b, e - c, f - a)$, ambos vetores com formas especiais — vamos chamá-los de vetores *pares* e *ímpares*, em consonância com a definição habitual de funções pares e ímpares (se você não sabe o que é isso, esqueça e simplesmente use as palavras). Para começar, vetores pares (e ímpares) só tem três graus de liberdade: quem conhece as três primeiras coordenadas, conhece as três outras. Além disso, tanto os vetores pares quanto os ímpares constituem subespaços vetoriais, também invariantes por A , e vetores equatoriais são levados por P a outros vetores equatoriais. Finalmente, todo vetor de \mathbb{R}^6 é a soma única de um vetor par e um ímpar: a identidade óbvia

$$v = \frac{v + Pv}{2} + \frac{v - Pv}{2}$$

pode ser escrita em coordenadas, tanto para vetores em \mathbb{R}^6 quanto para os vetores em \mathbb{R}^3 , sobre os quais age \tilde{A} :

$$(s, e, i) = \left(\frac{s+i}{2}, e, \frac{s+i}{2}\right) + \left(\frac{s-i}{2}, 0, \frac{-s+i}{2}\right).$$

Dessa última identidade, resulta que \tilde{A} age sobre um vetor par (s, e, s) e sobre um vetor ímpar $(s, 0, -s)$ obtendo os vetores $(4e, 2s + 2e, 4e)$ e $(0, 0, 0)$. Assim, as versões compactas da ação de \tilde{A} sobre os subespaços de vetores equatoriais pares e equatoriais ímpares (descritos por coordenadas

(s, e) no primeiro caso e s no segundo) correspondem a duas matrizes, uma de dimensão dois e outra de dimensão um,

$$\tilde{A}_p = \begin{pmatrix} 0 & 4 \\ 2 & 2 \end{pmatrix}, \tilde{A}_i = (0),$$

com autovalores 4, -2 e 0 — esses são *todos* os autovalores de A , a matriz de adjacência do ocatedro. Como você calcularia suas multiplicidades? Note que elas não são necessárias, se estivermos apenas interessados em calcular potências de A usando o cálculo funcional, como descrito na seção 5.

Você tem todo o direito de se perguntar para que considerar o segundo truque — afinal, acabamos caindo numa equação de segundo grau de qualquer forma para calcular os autovalores de \tilde{A}_p . É que agora você está pronto para calcular autovalores de matrizes de adjacência bem maiores. Considere, por exemplo, o icosaedro com uma matriz de adjacência A de dimensão 12. O argumento envolvendo uma rotação mostra que A tem os mesmos autovalores de uma matriz 4×4 \tilde{A} correspondendo à ação de A sobre autovetores que só tem quatro valores diferentes possíveis: um para o vértice superior, um para cada um dos trópicos, e um para o vértice inferior. Agora, usando a simetria central P , o cálculo dos autovalores de \tilde{A} se reduz ao cálculo dos autovalores de duas matrizes 2×2 \tilde{A}_p e \tilde{A}_i , agindo sobre vetores tropicais pares e tropicais ímpares.

Você já tem 'tecnologia' para calcular os autovalores de todas as matrizes de adjacência dos poliedros regulares em três dimensões, e de vários semi-regulares (que são poliedros que admitem faces de dois tipos diferentes dispostos com certa regularidade — a bola de futebol costurada com hexágonos e pentágonos é o poliedro semi-regular mais importante de nossa cultura). Os interessados podem ir além lendo [ST1] e [ST2]. Só para provocar, existe um poliedro regular (melhor, um politopo) em quatro dimensões que tem 600 vértices, logo dotado de uma matriz de adjacência A de dimensão 600. O truque das rotações leva ao estudo de uma matriz \tilde{A} com dimensão 'apenas' 46 ('viu' em $\mathbb{R}^{4?}$), e a simetria central por sua vez quebra essa matriz em duas matrizes de dimensão 23. Os deuses essa vez são generosos: os dois polinômios característicos de grau 23 se fatoram sobre os números racionais obtendo apenas fatores de graus 1 ou 2!

Os dois truques acima são exemplos muito simples de um conjunto de técnicas maravilhosas para lidar com simetrias — a teoria das representações de grupos. Tomara que você reencontre o assunto com mais profundidade um dia.

[ST1] Nicolau Saldanha e Carlos Tomei, *Spectra of Regular Polytopes*, Discr.Comp. Geom. 7, 403-414 (1992).

[ST2] Nicolau Saldanha e Carlos Tomei, *Spectra of Semi-regular Polytopes*, Boletim Soc. Bras. Mat. 29, 25-51 (1998).

3 Um modelo populacional

Esse é um exemplo de uma situação em que álgebra linear tem muito a dizer. Pelo seu curso afora, você vai perceber a enorme aplicabilidade do assunto — álgebra linear serve para tratar de *modelos lineares* (!) e, francamente, quase tudo que se modela na vida real tem que ser feito assim: a teoria não linear em geral é muito mais difícil.

Queremos fazer um modelo de crescimento populacional que tenha alguma sofisticação. Vamos começar com um primeiro, simples demais. Imagine que você sabe a *taxa de fertilidade* de uma população, usualmente dada por uma porcentagem por unidade de tempo. Assim, por exemplo, se a taxa é de 2% ao ano, e a população em um ano é 500, no ano seguinte ela será 510 (isto é, 1.02×500 — aliás, e dois anos depois?). Em geral, modelos desse tipo fazem uso de informação obtida através de um recenseamento realizado em um certo momento, como é o caso da população e da taxa de

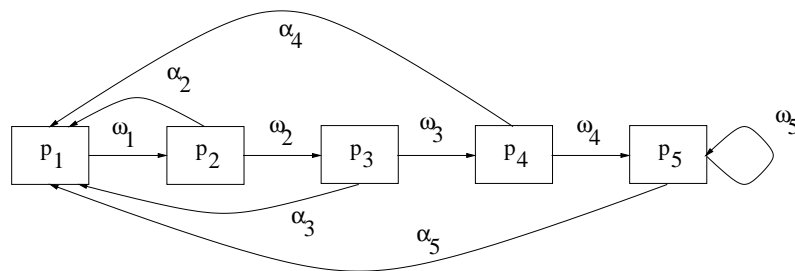
fertilidade, e a partir delas esses modelos especulam sobre o comportamento populacional um pouco antes, ou um pouco depois desse momento.

O modelo que queremos estudar é mais complexo. Vamos supor que estamos interessados em prever a *distribuição por idades* de uma população — isto é, gostaríamos de saber como muda a quantidade de indivíduos de uma certa faixa etária a medida que o tempo passa. Por exemplo, vamos considerar uma população dividida em cinco faixas etárias, dadas em anos por intervalos temporais $[0, 1)$, $[1, 2)$, $[2, 3)$, $[3, 4)$ e $[4, \infty)$. A última faixa inclui todos os indivíduos com idade maior ou igual a quatro anos — isso nos dispensa de fixar um prazo de vida terminal para os indivíduos dessa população. Vamos também supor que um recenseamento no tempo $t = 0$ obteve as populações $p_1(0), p_2(0), p_3(0), p_4(0)$ e $p_5(0)$, e vamos pensar como podemos fazer uso dessas informações para aproximar as várias populações alguns anos adiante. Note que chamar o tempo inicial de 0 deve ser irrelevante: o crescimento da população não tem nada a ver com o nome que damos a esse momento. As observações de que necessitamos são quase óbvias:

1. Depois de um ano, cada indivíduo passa de uma faixa etária para a seguinte, a não ser que ele pertença à última faixa. Para ser um pouco mais realista, vamos empregar *taxas de mortalidade* ω_i para indicar que fração da população de uma faixa i sobrevive até o ano seguinte.
2. A cada ano, nascem novos indivíduos, que são acrescentados à primeira faixa etária. Mais uma vez, vamos supor que cada faixa i contribui com uma fração α_i , a *taxa de fertilidade*, para a primeira faixa.

Talvez queiramos impor que poucas faixas etárias sejam férteis: por exemplo, poderíamos imaginar que, depois de uma certa idade, os indivíduos não se reproduzissem. Isso não é problema — basta anular as taxas de fertilidade de algumas faixas etárias. Da mesma forma, o modelo é suficientemente versátil para levar em conta que infantes e idosos morrem mais: é questão de ajustar as taxas de mortalidade associadas. Outra possibilidade seria tratar separadamente do crescimento populacional por idade e por sexo (supondo, por exemplo, que a espécie tenha, digamos, dois sexos), mas vamos deixar isso de lado. O importante é que estamos supondo que, assim como dispomos das populações em tempo $t = 0$, somos capazes de obter suas taxas de fertilidade e mortalidade, que, pelo menos por alguns anos, são supostas constantes.

Os itens acima descrevem o modelo, mas existem três outras formas de transmitir a mesma informação. A primeira é o gráfico abaixo. As caixas podem conter os valores das populações, e as setas indicam visualmente como indivíduos contribuem para as populações no ano seguinte. Certifique-se de que o gráfico, devidamente interpretado, contém exatamente a informação descrita nos itens.



A segunda maneira de descrever o modelo é através das equações abaixo:

$$\begin{aligned}
p_1(t+1) &= \alpha_2 p_2(t) + \alpha_3 p_3(t) + \alpha_4 p_4(t) + \alpha_5 p_5(t), \\
p_2(t+1) &= \omega_1 p_1(t), \\
p_3(t+1) &= \omega_2 p_2(t), \\
p_4(t+1) &= \omega_3 p_3(t), \\
p_5(t+1) &= \omega_4 p_4(t) + \omega_5 p_5(t).
\end{aligned}$$

Mais uma vez, convença-se de que essas equações dizem a mesma coisa que os itens ou o gráfico acima. Na terceira maneira, escrevemos as equações acima em notação matricial:

$$\begin{pmatrix} p_1(t+1) \\ p_2(t+1) \\ p_3(t+1) \\ p_4(t+1) \\ p_5(t+1) \end{pmatrix} = \begin{pmatrix} 0 & \alpha_2 & \alpha_3 & \alpha_4 & \alpha_5 \\ \omega_1 & 0 & 0 & 0 & 0 \\ 0 & \omega_2 & 0 & 0 & 0 \\ 0 & 0 & \omega_3 & 0 & 0 \\ 0 & 0 & 0 & \omega_4 & \omega_5 \end{pmatrix} \begin{pmatrix} p_1(t) \\ p_2(t) \\ p_3(t) \\ p_4(t) \\ p_5(t) \end{pmatrix}.$$

Para facilitar a notação, defina um vetor com cinco coordenadas,

$$p(t) = (p_1(t), p_2(t), p_3(t), p_4(t), p_5(t))^T,$$

onde o T indica que o vetor está sendo pensado na vertical, e seja M a matriz 5×5 acima. O modelo então é equivalente á equação

$$p(t+1) = M p(t).$$

É claro, aliás, que $p(t+2) = M^2 p(t)$, ou, se quisermos avançar n anos, $p(t+n) = M^n p(t)$.

As duas últimas representações do modelo são fáceis de implementar em máquina. Desta forma, podemos estudar computacionalmente a evolução temporal de uma população a partir dos dados iniciais. O modelo só é satisfatório enquanto não houver alteração substancial das taxas de fertilidade e mortalidade. Numa situação de super-população, por exemplo, esses parâmetros mudam drasticamente. Um modelo desse tipo, aliás, foi empregado por alguns anos para estabelecer uma largura conveniente dos buracos das redes de pesca no Mediterrâneo — pense como.

Se todas as descrições acima são equivalentes, por quê interessar-se pela representação matricial? A primeira razão é simples de entender: vetores e matrizes são fáceis de programar. A outra razão é que, a partir de álgebra linear, podemos deduzir fatos sobre a evolução da população que não são evidentes — vamos dar um exemplo. Em vez de considerarmos a população em cada faixa etária, vamos estudar como muda a fração correspondente a cada faixa etária na população total: isso é equivalente a, dada uma população (vetorial!) $p(t)$, dividirmos todas as coordenadas de $p(t)$ pelo mesmo número de modo a obter um novo vetor, digamos $q(t)$, a *distribuição etária* da população, cujas coordenadas somam 1: as coordenadas de $q(t)$ são números entre 0 e 1, denotando as frações que nos interessam. O fato surpreendente é que, para hipóteses muito gerais, a medida que t aumenta, o vetor $q(t)$ converge de um vetor fixo que não tem nada a ver com a distribuição etária inicial! Isto quer dizer, por exemplo, que distribuições etárias iniciais muito diferentes vão ficar cada vez mais parecidas a longo prazo. A demonstração disso usa um resultado mais difícil de álgebra linear, o teorema de Perron-Frobenius, que está delineado na seção 5.

4 Aproximando autovalores

O exemplo da evolução de população dividida em faixas etárias dá o caminho das pedras para aproximar autovalores de uma matriz M . Comece com um vetor v ao acaso: no exemplo, uma

pirâmide etária. Agora, calcule vetores $Mv, M^2v, \dots, M^k v$, e normalize de vez em quando. Para o exemplo, a normalização é importante porque o conceito de pirâmide etária exige que as componentes somem 1. Para a situação mais, geral, é bom normalizar para evitar que os vetores da seqüência calculada não cresçam ou encolham demais — pense no que aconteceria se M fosse um múltiplo da identidade. Agora, com um pouco de sorte (isto é, sob hipóteses que serão explicitadas em breve), a seqüência $(M^k v)/\|M^k v\|$ alinha-se com um autovetor de M . Quem tem boas aproximações de um autovetor tem também boas aproximações do autovalor associado.

Vamos tornar essa idéia mais precisa: os argumentos vão ao mesmo tempo demonstrar o comportamento do modelo demográfico e justificar porque o procedimento numérico funciona, sob hipóteses adequadas.

Para simplificar, vamos supor que M seja uma matriz com autovalores distintos, todos reais, e que λ_1 seja o autovalor de maior módulo. Denote por λ_i os autovalores de M , associados a autovetores v_i : $Mv_i = \lambda_i v_i$. Seja v um vetor qualquer, que se expande em $v = \sum_i c_i v_i$ na base de autovetores. As contas a seguir falam por si:

$$\begin{aligned} v &= c_1 v_1 + c_2 v_2 + \dots + c_n v_n, \\ Mv &= c_1 \lambda_1 v_1 + c_2 \lambda_2 v_2 + \dots + c_n \lambda_n v_n, \\ M^k v &= c_1 \lambda_1^k v_1 + c_2 \lambda_2^k v_2 + \dots + c_n \lambda_n^k v_n, \\ M^k v &\sim c_1 v_1 + c_2 \left(\frac{\lambda_2}{\lambda_1}\right)^k v_2 + \dots + c_n \left(\frac{\lambda_n}{\lambda_1}\right)^k v_n, \end{aligned}$$

onde o sinal \sim quer dizer que os dois lados da equação são vetores alinhados, isto é, diferem de um escalar multiplicativo. As frações $(\lambda_i/\lambda_1)^k$ vão a zero, quando k vai a infinito, já que λ_1 é o autovalor de maior módulo. Se c_1 não é nulo, fica então claro que $M^k v$ se alinha com v_1 quando $k \rightarrow \infty$. Em palavras imprecisas, a longo prazo ganha a componente do vetor inicial na direção mais espichada pela matriz M . Não interessa se esse vetor inicial tem uma componente pequena para começar (o valor de c_1), desde que ele tenha algo ($c_1 \neq 0$). Como as contas na vida real sempre têm erros de arredondamento, é quase impossível perceber se c_1 de fato é zero: ele vai deixar de sê-lo ao longo do algoritmo. Mas isso pode fazer com que o processo leve mais tempo (k grande) para caracterizar convergência.

Agora, vamos enfraquecer um pouco as hipóteses. Note que não é necessário pedir para que os autovalores sejam reais, já que estamos comparando módulos de números. Em princípio, podemos pedir apenas para que o autovalor de maior módulo seja real. Aliás, se dois autovalores tiverem o mesmo maior módulo, uma pequena alteração nas contas acima mostra que $M^k v$ tende a se aproximar do subespaço gerado pelos autovetores associados aos autovalores de módulo máximo. É possível também mostrar que o efeito de autovalores repetidos (mesmo sem autovetores independentes associados) é inócua, e que autovalores complexos de módulo máximo também são caracterizados por um padrão de convergência geométrico.

Em vez de descer a detalhes sobre esses aspectos, vamos nos restringir a duas situações freqüentes. Na primeira, a matriz M é real, simétrica, e pelo teorema espectral (seção 5), as contas acima só necessitam de extensão para o caso de termos mais de um autovalor de módulo máximo, o que não é difícil de fazer. Na segunda situação, vamos supor que a matriz M satisfaz as hipóteses do teorema de Perron ou de Frobenius (seção 5), com um único autovalor de maior módulo. Nesse caso, o teorema de Perron-Frobenius assegura que o autovalor de maior módulo é um número real, positivo, e que ele é o único com esse módulo, o que é bastante também para assegurar convergência do método. Aliás, isso explica o comportamento da distribuição etária a longo prazo: ela converge para o autovetor (de coordenadas positivas, como toda quebra de população em frações deve ter) associado ao autovalor de maior módulo. Cada iteração, para k grande, essencialmente multiplica a distribuição pelo autovalor, e o destino da população como um todo — se cresce ou se extingue —

depende desse autovalor ser maior ou menor do que um. A distribuição, por ser normalizada, não percebe o tamanho do autovalor.

O processo descrito acima é o *método das potências* para cálculo do autovalor de maior módulo de uma matriz, ponto de partida para tantos outros algoritmos numéricos. Note que a velocidade de convergência depende do tamanho de c_1 e do quociente λ_2/λ_1 : quanto maior o espaçamento entre os dois autovalores, mais rápida a convergência. O método de potências só funciona se estivermos interessados no cálculo de λ_1 . Essa limitação, ironicamente, é consequência de uma das propriedades fundamentais de álgebra linear: o comportamento a longo prazo de um sistema modelado por uma transição $v_{n+1} = Mv_n$ depende essencialmente do autovalor de maior módulo de M , e é sentido especialmente na direção do autovetor associado. A matriz M pode ser enorme, mas para entender o comportamento a longo prazo da iteração só precisamos saber muito pouco sobre ela. Nas aplicações, às vezes são estudados uns poucos autovalores de módulo um pouco menor do que λ_1 , justamente porque eles podem estar muito próximos, atrasando o desempenho do algoritmo e simultaneamente influenciando por mais tempo o comportamento assintótico do modelo.

Mas o que fazer para calcular outros autovalores? Em certas aplicações, é importante saber se existem autovalores dentro de um intervalo dos reais, ou dentro de uma região complexa. O método das potências admite uma extensão simplíssima e poderosa. Para começar, seja p um polinômio. É fácil ver (comece com polinômios simples) que $p(M)v_i = p(\lambda_i)v_i$, onde, como sempre, $Mv_i = \lambda_i v_i$. Agora, refaça as contas fundamentais da seção:

$$\begin{aligned} v &= c_1 v_1 + c_2 v_2 + \dots + c_n v_n, \\ p(M)v &= c_1 p(\lambda_1)v_1 + c_2 p(\lambda_2)v_2 + \dots + c_n p(\lambda_n)v_n, \\ p(M)^k v &= c_1 p(\lambda_1)^k v_1 + c_2 p(\lambda_2)^k v_2 + \dots + c_n p(\lambda_n)^k v_n, \end{aligned}$$

mas agora talvez p seja maior em outro autovalor: em outras palavras, a escolha de p determina em que parte do plano complexo você quer procurar autovalores! Ao normalizar, a ênfase vai automaticamente privilegiar as componentes dos autovetores desta região. Assim, por exemplo, se M é uma matriz real, simétrica com autovalores entre 0 e 1 e você quer procurar autovalores perto de $1/2$, você podia escolher o polinômio $p(x) = x(1-x)$. Claro que, agora, a velocidade de convergência é determinada por um quociente da forma $p(\lambda_i)/p(\lambda_j)$.

Vamos estender essa idéia só mais um pouquinho. Em vez de considerar um polinômio p , podemos considerar iterações de uma função f . Claro que a dificuldade está em poder fazer as contas necessárias, tanto na teoria quanto na prática. Vamos nos limitar a funções racionais (isto é, quocientes de dois polinômios). A parte teórica não é difícil: de novo, o ingrediente fundamental é que $f(M)v_i = f(\lambda_i)v_i$ e o estudo da iteração $f^k(M)v$ procede como para polinômios. O quociente $p(M)/q(M)$ é na verdade a matriz $p(M)(q(M))^{-1} = (q(M))^{-1}p(M)$, já que todas as funções de uma matriz comutam entre si. Na prática, é raro alguém calcular explicitamente a matriz inversa. Note que, para efeitos de iteração precisamos apenas calcular $f(M)v$. Assim, por exemplo, se quisermos calcular $M^2/(M-5I)v$, podemos fazer as contas

$$v, Mv, M(Mv), (M-5I)^{-1}(M(Mv)).$$

Duas observações são oportunas. Não é vantagem calcular M^2 para depois aplicar em v : a seqüência $v, Mv, M(Mv)$ custa menos. Para calcular $x = (M-5I)^{-1}(M(Mv))$, resolva $(M-5I)x = (M(Mv))$: resolver um sistema linear é muito mais barato do que inverter uma matriz.

Agora o pulo do gato: a escolha apropriada de uma função racional na iteração é um grande acelerador de convergência. Note que a função $f(x) = 1/(x-s)$ dá muita ênfase a números próximos de s . Mais até, se x_1 e x_2 estão próximos de s , os valores $f(x_1)$ e $f(x_2)$ freqüentemente são tais que o quociente $f(x_2)/f(x_1)$ é substancialmente maior do que x_2/x_1 , o que mostra que a iteração com

um s adequado converge mais rápido (para ver isso, desenhe o gráfico de $f!$). Essa escolha de F na literatura é chamada de *iteração inversa com translação* (em inglês, *inverse iteration with shift*, daí o s). A situação se realimenta: se queremos aproximar um autovetor v associado a um autovalor λ , podemos começar com aproximações v_0 e λ_0 , e, a medida que avançamos a iteração, usamos as aproximações v_k do autovetor e λ_k do autovalor, esse para atualizar o shift s . A convergência eventual desse método é espetacular, como se vê em máquina.

A teoria associada a estes refinamentos é apenas uma parte de uma enorme área de atividade em análise numérica. Os algoritmos disponíveis já dão conta de matrizes bastante grandes, mas a demanda hoje em dia dirige-se ao desenvolvimento de algoritmos para classes especiais de matrizes, com dimensões na ordem de milhares, dezenas de milhares.

5 Álgebra linear básica e o cálculo funcional

Vamos listar os fatos da vida: consulte seu livro de álgebra linear preferido para os detalhes. Dada uma matriz M de dimensão n (isto é, $n \times n$), um vetor não nulo v e um número λ para o qual valem $Mv = \lambda v$ são o *autovetor* v associado ao autovalor λ . Os autovalores de M são as raízes do polinômio característico $\det(M - \lambda I)$, onde I é a matriz identidade $n \times n$. Assim, matrizes $n \times n$ sempre têm n autovalores, mas nem sempre têm n autovetores independentes. Quando isso acontece, a matriz M é *diagonalizável*, e pode ser escrita como

$$M = PDP^{-1},$$

onde D é uma matriz diagonal contendo os autovalores de M e as colunas de P são autovetores apropriados. Duas matrizes que diferem de uma conjugação (isto é, A e B tais que $A = PBP^{-1}$) têm os mesmos autovalores. É isso que garante, por exemplo, que os autovalores das matrizes de adjacência de um grafo (definidas na primeira seção) não dependem da rotulação dos vértices do grafo.

Existem duas classes importantes de matrizes reais diagonalizáveis: as matrizes cujos autovalores são todos diferentes (as matrizes de *espectro simples*) e as matrizes simétricas. O teorema espectral diz até mais: uma matriz simétrica admite uma base ortonormal de autovetores, o que permite tomar P na sua diagonalização como sendo uma matriz ortogonal, mas não usaremos esses fatos nesse texto.

Em vez disso, usaremos os teoremas de Perron e Frobenius ([G]). Uma matriz é *positiva* se todas as suas entradas forem maiores do que zero. Pelo *teorema de Perron*, uma matriz positiva tem um único autovalor de maior módulo, que é automaticamente real e positivo, e que está associado a um autovetor cujas coordenadas podem ser tomadas todas positivas. O teorema de Frobenius é uma extensão desse resultado. Uma matriz é *não negativa* se todas as suas entradas forem maiores ou iguais a zero e é *irredutível* se alguma potência sua só tem entradas não nulas. Pelo *teorema de Frobenius*, para uma matriz não negativa e irredutível, os autovalores de maior módulo são as raízes de um número real positivo. A possibilidade de encontrarmos mais de um autovalor de maior módulo é muito rara: para saber exatamente quando isso acontece, consulte [G]. No caso de um único autovalor de maior módulo, mais uma vez, temos que o autovetor associado pode ser tomado com todas as coordenadas positivas.

Agora, vamos mostrar que potências de matrizes diagonalizáveis são calculadas com facilidade se conhecemos seus autovalores. De forma mais geral, vamos delinear um procedimento para calcular funções $f(A)$ de uma matriz A . Uma referência muito detalhada para este *cálculo funcional de matrizes* é [B]: lá você vai encontrar um procedimento geral, aplicável a qualquer matriz.

Uma maneira possível de obter $f(A)$ seria calcular autovalores e autovetores de A , para diagonalizá-la,

$$A = PDP^{-1},$$

onde P tem os autovetores de A em suas colunas, e D é uma matriz diagonal, com os autovalores de A em sua diagonal. Então

$$\begin{aligned} A^2 &= (PDP^{-1})^2 = PDP^{-1}PDP^{-1} = PD^2P^{-1}, \\ A^3 &= PD^3P^{-1} \dots \end{aligned}$$

ou de forma mais geral, $f(A) = Pf(D)P^{-1}$. Isso é fácil de ver se f é um polinômio, e precisa de uma demonstração mais cuidadosa quando f é uma série de Taylor (como no caso da exponencial), mas nem por isso deixa de ser verdade. Finalmente, note que $f(D)$, uma matriz diagonal como D , é muito fácil de calcular: basta aplicar f a cada posição diagonal de D .

Para o procedimento que nos interessa, não é nem necessário calcular os autovetores de A . Recapitulando, suponha que você quer calcular $f(A)$, onde A é uma matriz diagonalizável. Comece calculando seus autovalores $\lambda_1, \lambda_2, \dots, \lambda_k$, descartando os autovalores repetidos. A seguir, procure um polinômio p tal que

$$p(\lambda_1) = f(\lambda_1), \dots, p(\lambda_k) = f(\lambda_k).$$

A matriz procurada, $f(A)$, é simplesmente $p(A)$.

Exemplo: Para calcular quantos caminhos de comprimento 1000 vão do vértice 1 ao vértice 6 do octaedro (veja a primeira figura da seção 3), você quer calcular a posição $(1, 6)$ da matriz A^{1000} , onde A é a matriz de adjacência do octaedro. Como A é simétrica, ela é diagonalizável, e já calculamos seus autovalores: 4, -2 e 0. Precisamos então encontrar um polinômio que leva 0 a 0, -2 a $(-2)^{1000}$ e 4 a 4^{1000} . O polinômio $p(x) = ax^2 + bx$ com

$$a = \frac{4^{1000} + 2 \cdot 2^{1000}}{24}, \quad b = \frac{4^{1000} - 4 \cdot 2^{1000}}{12}$$

faz isso. Assim, $A^{1000} = aA^2 + bA$: somando as matrizes, você obtém a resposta.

Note que um pequeno milagre aconteceu no exemplo acima: para calcular uma matriz à milésima potência, bastou calcular um polinômio de grau dois dela.

Exercício ótimo para depois de ler a primeira seção: Mostre que os autovalores da matriz de adjacência do grafo associado ao cubo são os números $-3, -1, 1$ e 3 , aliás, com multiplicidade $1, 3, 3$ e 1 . Calcule o número de caminhos fechados de comprimento 1000 nesse grafo.

Por quê esse procedimento funciona? No caso do cálculo de uma potência de uma matriz diagonalizável, é muito fácil explicar. Suponha que você quer calcular A^n . Diagonalizando $A = PDP^{-1}$, temos que $A^n = PD^nP^{-1}$, onde D^n é a matriz diagonal obtida elevando a n as posições diagonais de D , que são aliás os autovalores de A . Agora, se p é um polinômio que no autovalor λ_i de A toma o valor λ_i^n , temos que $p(A) = Pp(D)P^{-1}$, e $p(D) = D^n$, já que as duas matrizes são matrizes diagonais cujas posições diagonais são da forma $p(\lambda_i) = \lambda_i^n$. Em uma frase, o argumento é quase banal: se A é diagonalizável, duas funções f e g que tomam os mesmos valores nos autovalores de A vão ser tais que $f(A) = g(A)$! Essa idéia elementar é uma fonte de simplificações em cálculos envolvendo matrizes, e por alguma razão misteriosa, não tem a divulgação merecida.

[B] Hamilton Bueno, *Function of Matrices*, Cubo 05/03, Outubro 2003, ou *Funções de Matrices*, I Biental da SBM.

[G] F. Gantmacher, *Matrix theory*, Chelsea, NY, 1974.

Departamento de Matemática, PUC-Rio tomei@mat.puc-rio.br