

A MATEMÁTICA POR TRÁS DO NASCIMENTO DO *GOOGLE*

BRUNO SANTIAGO

1. A MANEIRA INTELIGENTE DE PROJETER UM BUSCADOR

Se você é jovem demais para lembrar como era navegar na internet no fim dos anos 90 possivelmente não tem memórias do impacto causado pelo *Google* nesse mercado. A empresa foi fundada em 1998, por dois amigos Lawrence Page e Sergey Brin. Eles eram colegas de doutorado no Departamento de Ciência da Computação da Universidade de Stanford.

O problema de tese de Page era justamente entender a estrutura *matemática* subjacente a internet, seguindo uma ideia proposta pelo seu orientador Terry Winograd: estudar a estrutura de links inerente a internet. Esse tipo de estrutura de links já havia sido objeto de estudo por matemáticos e cientistas da computação e aplicados em muitas situações, notadamente a dinâmica das colaborações científicas. Terry propôs a Page de colocar todo esse arcabouço que já estava a disposição para entender a internet, que começava e explodir naquele momento (meados dos anos 90).

Essas pesquisas deram frutos e em 1996, Page e Brin lançaram o algoritmo *PageRank* que permite ranquear as páginas da internet em termos da “importância” de cada uma delas. Dois anos depois eles reportaram a nova ideia, já patenteada, numa revista científica [2].

Naquele momento, já existiam buscadores da *Internet* (uma ferramenta que permite ao usuário buscar páginas que citem uma determinada palavra ou frase) como *Yahoo!*, *MSN*, *etc...*. Todos funcionavam essencialmente como contadores de palavras: o buscador te reporta em primeiro lugar a página com maior número de ocorrências da palavra ou frase que você digitou. Assim, por exemplo, se você escrevesse uma página que só contém a sigla UFF repetida um milhão de vezes, essa página aparecia em primeiro lugar ao buscar “UFF”, ao invés do site da Universidade Federal Fluminense.

Brin e Page tiveram a ideia de combinar o ranqueamento que tinham inventado com a contagem de palavras para criar um novo tipo de buscador. O princípio subjacente é que em geral o usuário vai preferir a página com maior importância na qual a palavra ou frase digitada aparece. Assim nasceu o *Google*, uma das maiores empresas do planeta nos dias de hoje.

E o que permitiu ao *Google* se destacar tão rapidamente dos seus concorrentes? Matemática! A chave para o sucesso da empresa foi justamente o algoritmo matemático de ranqueamento das páginas da internet.

Neste artigo vou tentar explicar como funciona esse algoritmo e toda a matemática por trás dele. Desse modo, começo contanto o problema do ranqueamento da internet e a solução proposta por Page e Brin. Em seguida, vou falar de cada tópico matemático que fundamenta essa solução. Procurei escrever o artigo de forma que a primeira seção não tenha pré-requisitos matemáticos e de modo que

eles vão aumentando **gradual e gentilmente** ao longo das seções seguintes. No entanto, fiz um esforço para que o texto seja o mais auto-contido possível, o que justifica o tamanho do artigo. Mas antes de começarmos, uma pausa para

1.0.1. *Anekdotes matemáticas.* De fato, a história do *Google* é tão recheada de matemática que o próprio nome é inspirado num livro de matemática, “Matemática e Imaginação”, escrito por Edward Kasner e James Newman [5]. Nesse livro delicioso de divulgação¹ os autores começam falando sobre os mistérios dos números muito grandes. Eles inclusive inventaram um número tão grande que nada que existe no universo pode chegar a essa quantidade: 10^{100} , ou 1 seguido de 100 zeros. Kesner queria dar um nome a esse número fabuloso e passou a tarefa ao seu filho de 6 anos. A criança decidiu nomear o número de *googol*. Ou seja

$$\text{googol} \stackrel{\text{def}}{=} 10^{100}.$$

Um exemplo de quão grande é esse número: desde o surgimento do universo, há 14 bilhões de anos, passaram-se “apenas” $4,7 \times 10^{17}$ segundos. Nem perto de um googol... Para enfatizar que o conjunto dos números naturais é infinito, e por isso contém números de tamanhos além da nossa imaginação, Kesner e Newman também inventaram o googolplex: 1 seguido de um googol de zeros

$$\text{googolplex} \stackrel{\text{def}}{=} 10^{\text{googol}}.$$

A sede atual do *Google* se chama Googleplex. :D

1.1. **O PageRank.** Nessa seção vamos descrever a solução do problema do ranqueamento das páginas da internet. A pergunta que queremos responder é a seguinte: como elencar as páginas da internet, da mais importante (dentre todas as páginas) a menos importante?

Esse é um problema onde parece muito natural pensar em usar matemática para atacá-lo, afinal os números possuem um ranqueamento intrínseco entre si: sabemos que $0.5 > 0.489!$ Então, se pudermos associar um número a cada página que reflita a magnitude da sua importância, teremos resolvido o problema.

O primeiro passo para se resolver um problema usando matemática é tentar fazer um *modelo matemático* da situação: descrever a internet (nesse caso particular que estamos interessados) através de algum objeto matemático. Mas a rainha das ciências é tão rica em objetos interessantes que isso é quase de graça, uma vez que se olha para o problema com um certo olhar.

Com efeito, toda modelagem matemática requer algum tipo de hipótese simplificadora, ou descritiva. Como falei na introdução, Page estava estudando a estrutura de links da internet. A hipótese que ele e seu colega Brin fizeram foi de que

*a importância de uma página depende apenas
dos links que ela recebe.*

Page descreveu a sugestão do seu orientador para estudar a estrutura de links da internet como “a melhor ideia que recebeu na vida”. O momento onde fazemos hipóteses para permitir a modelagem é sempre subjetivo, depende de como enxergamos a situação, depende do *feeling*. Você pode olhar o mesmo problema e achar mais razoável outro tipo de hipóteses.

¹Por sorte, havia um exemplar dele na maravilhosa biblioteca do CEFET/RJ, onde fiz o ensino médio, e posso te garantir, caro leitor, isso mudou minha vida.

A partir daí, se estamos interessados apenas nos links que as páginas fazem entre si, podemos imaginar cada página da internet como um ponto no plano e representar um link como uma seta da página A para a página B, se a página A faz link para a página B.²

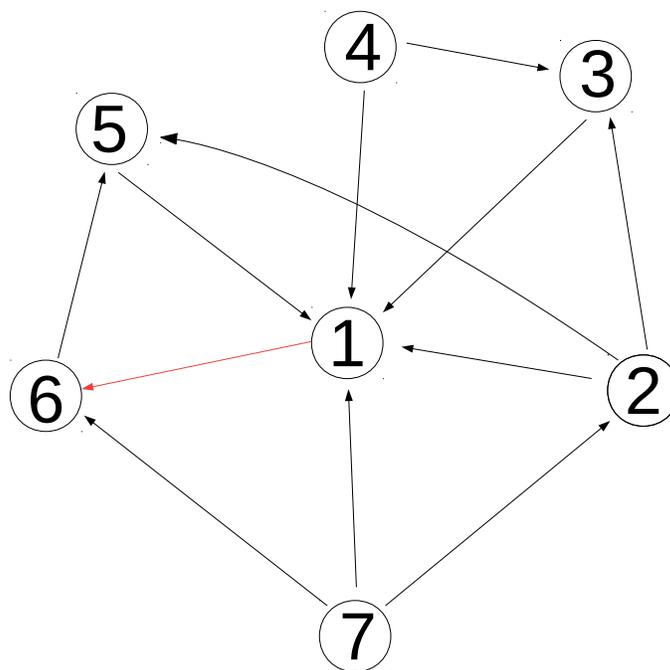


FIGURA 1. A estrutura de *links* da internet pode ser descrita através de um grafo.

Uma estratégia inicial pode ser ranquear como mais importante aquela página que recebe mais links. Repare na Figura 1. A página (1) claramente recebe mais links do que todas as outras, e indiscutivelmente deve ser a mais importante. Agora repare que as páginas (6), (5) e (3) recebem ambas o mesmo número de links. No entanto, a página (6) recebe um link (em vermelho na Figura 1) de uma página muito importante. Isso deveria torná-la mais importante do que as demais!

Isso indica que devemos procurar uma maneira mais sofisticada de atribuir importância as páginas, e trás uma outra ideia: um link vindo de uma página que recebe muitos links tem mais *qualidade* e por isso carrega muita importância. Por exemplo, como a página (6) tem bastante importância e aponta para a página (5), a página (5) deve ser considerada mais importante do que a página (3). Ambas recebem 2 links, mas a página (3) recebe links de páginas pouco importantes.

²Esse tipo de figura com pontos e links entre pontos é que chamamos em matemática de um *grafo*, e o ramo da matemática que estuda isso é a Teoria de Grafos.

No entanto, seguindo essa linha parece um processo complicado e sem fim atribuir a importância das páginas: você vê a quantidade de links que cada página recebe, depois você atualiza a importância das páginas que recebem links de páginas com muitos links e assim sucessivamente. Mas como isso acaba?

Essa é a beleza de atacar um problema usando matemática. Você cria um modelo abstrato da situação real, estuda apenas as relações entre os objetos que aparecem no seu modelo abstrato e... bum! De repente uma coisa salta aos olhos, uma coisa que estava lá o tempo todo, mas estava escondida, inacessível aos olhos porque a única chave para abrir as portas do esconderijo é a matemática.

De fato, você pode pensar que cada página doa uma fração da sua importância às páginas para as quais ela aponta. Assim toda página doa toda a sua importância mas recebe frações de importância vindas das páginas que para ela apontam. Como supomos que a única coisa que influencia na importância de uma página é a quantidade e a qualidade dos links que ela recebe, podemos assumir que a importância de uma página é igual a soma das frações de importância que ela recebe.

Vamos tentar traduzir isso numa equação.

Começamos enumerando as páginas como P_1, P_2, \dots, P_n , onde n representa o número total de páginas da internet. Vamos chamar de x_1, x_2, \dots, x_n , respectivamente o valor numérico da importância que cada página tem. Observe que faz sentido que o valor da importância seja um número maior do que ou igual a zero e menor do que ou igual a 1, ou seja, represente uma porcentagem, uma probabilidade. Assim, uma importância próxima de 1 significa que aquela página vai ser frequentemente visitada, enquanto uma importância próxima de zero significa que a página vai ser raramente visitada. Observe que os números x_1, \dots, x_n são justamente as incógnitas cujo valor queremos descobrir. Para isso, vamos fazer como sempre fazemos: tentar montar uma equação, uma igualdade numérica, envolvendo as incógnitas e números conhecidos e tentar resolver essa equação.

Além disso, como alguma página sempre vai ser visitada, a importância “total” da internet deve ser 100%. Por isso, podemos supor que a soma das importâncias de todas as páginas sempre dá $1 = 100\%$, ou seja

$$(1.1) \quad \sum_{\ell=1}^n x_{\ell} \stackrel{\text{def}}{=} x_1 + x_2 + x_3 + \dots + x_n = 1.$$

Temos assim a nossa primeira equação envolvendo os números x_1, \dots, x_n .

Agora, sejam d_1, d_2, \dots, d_n respectivamente o número de páginas para as quais a página 1 aponta, a página 2 aponta e assim por diante. Assim, por exemplo, na Figura 1, $d_2 = 3$. No nosso modelo, cada página j doa uma fração igual a $1/d_j$ da sua importância x_j à cada página para a qual aponta. Por exemplo, a página 2 doa $1/3$ da sua importância x_2 à página 3, $1/3$ à página 1 e $1/3$ à página 5. E qual vai ser a importância da página 2? Ora, ela recebe link apenas da página 7, então sua importância vai ser

$$x_2 = \frac{x_7}{d_7}.$$

Já a página 1 recebe links de quase todo mundo. Sua importância vai ser

$$x_1 = \frac{x_2}{d_2} + \frac{x_3}{d_3} + \frac{x_4}{d_4} + \frac{x_5}{d_5} + \frac{x_7}{d_7}.$$

Para a nossa internet de 7 páginas, vamos ter um sistema de 7 equações e 7 incógnitas como as duas acima. Em geral, podemos denotar por β_{ij} a fração da

importância da página j doada a página i . Assim, $\beta_{ij} = 1/d_j$ se $j \rightarrow i$ e $\beta_{ij} = 0$ caso contrário. A importância x_i deve satisfazer a equação

$$(1.2) \quad x_i = \sum_{j=1}^n \beta_{ij} x_j.$$

1.1.1. *Como resolver esse sistema de equações?* Se você estudou sistemas lineares no ensino médio, ou já fez um curso de álgebra linear na universidade, vai lembrar que a equação (1.2) nada mais é que um sistema linear! Então, vamos tentar resolver o sistema com 7 equações proveniente da internet fictícia da Figura 1. Olhando para ela podemos ver que $d_1 = 1$, $d_2 = 3$, $d_3 = 1$, $d_4 = 2$, $d_5 = 1$, $d_6 = 1$ e $d_7 = 3$. Logo, o sistema (1.2) se escreve explicitamente assim:

$$\begin{aligned} x_1 &= \frac{x_2}{3} + x_3 + \frac{x_4}{2} + x_5 + \frac{x_7}{3} \\ x_2 &= \frac{x_7}{3} \\ x_3 &= \frac{x_2}{3} + \frac{x_4}{2} \\ x_4 &= 0 \\ x_5 &= \frac{x_2}{3} + x_6 \\ x_6 &= x_1 + \frac{x_7}{3} \\ x_7 &= 0 \end{aligned}$$

Da quarta e da última equação deduzimos que $x_2 = x_3 = 0$ e portanto colocando essas informações na quinta, na sexta e na primeira equação vemos que $x_1 = x_5 = x_6$. Por fim, como as importâncias x_1, \dots, x_n devem satisfazer também a equação (1.1) temos que

$$\begin{aligned} 1 = x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7 &= 3x_1 \\ \therefore x_1 &= \frac{1}{3} \end{aligned}$$

Logo, o vetor de importância da nossa internet fictícia é $v = (1/3, 0, 0, 0, 1/3, 1/3, 0)$.

1.1.2. *Testando a técnica num sistema ainda mais fácil.* Vamos agora considerar uma outra internet fictícia, como na Figura 2. Temos apenas três incógnitas a descobrir nesse sistema. Como $d_1 = 2$, $d_2 = 2$ e $d_3 = 0$. Assim, nesse caso particular o sistema de equações (1.2) se escreve

$$\begin{aligned} x_1 &= \frac{x_2}{2} \\ x_2 &= \frac{x_1}{2} \\ x_3 &= \frac{x_1}{2} + \frac{x_2}{2} \end{aligned}$$

Substituindo a segunda equação na primeira obtemos $x_1 = x_1/4$ o que implica que $x_1 = 0$. Mas isso implica então que $x_2 = x_3 = 0$. Ou seja, todas as páginas têm importância zero! No entanto, como $x_1 + x_2 + x_3 = 1$, temos uma situação absurda! Ou seja, as equações que montamos para modelar o problema do ranqueamento da internet não tem solução no caso da Figura 1.2.

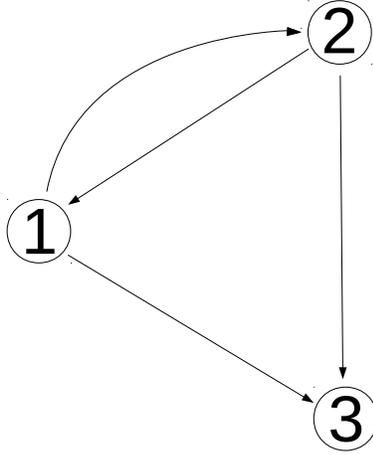


FIGURA 2. A única solução possível para essa internet é $x_1 = x_2 = x_3 = 0$, mas isso é incompatível com a equação (1.1).

1.1.3. *Corrigindo o modelo, ou: se aproximando mais da realidade.* Se você observar, nosso modelo negligenciou uma característica fundamental da internet: na verdade todas as páginas estão ligadas entre si! Eu posso ir da página P_i a página P_j apenas digitando a URL da página P_j no navegador.

Assim, por exemplo, na Figura 2 a página 3 não cita nenhuma outra e, no modelo que fizemos antes, uma vez que um usuário “aterriza” nela não pode mais sair. Só que isso não condiz nem um pouco com a realidade! Ainda que essa página não cite nenhuma outra, podemos sair dela para qualquer outra! E a equação (1.2) negligenciou completamente isso.

Mas não há motivo para pânico, sempre vamos errar e o que importa é aprender com nossos erros. Agora vamos tentar escrever uma nova equação, que leve em conta o que desprezamos até aqui; portanto, vamos tentar fazer uma nova equação, um modelo, mais fiel à realidade do nosso problema.

Para tanto, temos que fazer uma nova hipótese, de natureza estatística, empírica, baseada na observação, no “chute”:

na maior parte das vezes o usuário segue os links entre as páginas, mas algumas vezes ele digita a URL e se “teletransporta” no grafo da página P_i à página P_j .

Assim, existe uma chance de o usuário digitar a URL. Brin e Page colocaram essa chance como sendo de 15% [2]. Podemos traduzir isso para as nossas equações da seguinte forma: uma parte da importância da página P_i (digamos, como Brin e Page fizeram, 85%) vem das frações de importância que ela recebe das páginas que apontam para ela. Uma outra parte, complementar a anterior, vem da chance de um usuário teleportar-se diretamente da página P_j à página P_i digitando a URL no navegador. Essa chance é diretamente proporcional à importância da página P_i .

Então, consideramos um número $0 < p < 1$, que representa a chance de o usuário digitar uma URL diretamente. Brin e Page deram a esse número o nome de fator

de amortecimento, e como dissemos, consideraram $p \sim 15\%$. Para escrever o novo sistema de equações envolvendo as importâncias, precisamos ainda *reconsiderar* a nossa definição dos números β_{ij} . Esse número representava a fração de importância que a página j doa para a página i , quando há um link $j \rightarrow i$. Em particular, se uma página não cita nenhuma outra, como a página 3 na figura 2, $\beta_{ij} = 0$, para todo $i = 1, \dots, n$. À luz do que propusemos acima para remodelar o problema, toda vez que uma página j não citar nenhuma outra vamos colocar $\beta_{ij} = 1/n$.

Essa é uma escolha que não tem justificativa matemática, apenas estamos tentando trazer as nossas equações para mais perto da realidade. Intuitivamente, significa que quando um usuário entra numa página que não cita ninguém ele pode ir dela para qualquer outra página. É uma mudança drástica de ponto de vista: quando uma página não cita ninguém na verdade ela cita todo mundo (inclusive ela própria)! Porém, isso parece mais próximo da nossa experiência com a internet: é muito mais frequente um usuário sair de uma página que não cita nada, e continuar surfando, do que ficar ancorado muito tempo no mesmo lugar. Embora ambos possam acontecer, estamos dando pesos diferentes para cada possibilidade.

Vamos então recapitular a nova definição dos números β_{ij} : se $d_j = 0$ (o que significa que a página j não cita ninguém) colocamos $\beta_{ij} = 1/n$, para todo $i = 1, \dots, n$. Se $d_j > 0$ colocamos a definição antiga, isto é, $\beta_{ij} = 1/d_j$, se $j \rightarrow i$ e $\beta_{ij} = 0$ caso contrário.

Resumindo: um surfista navegando aleatoriamente pela internet, na maior parte do tempo, vai seguir links entre páginas ou pular de uma página que não cita ninguém para qualquer outra. Com uma frequência menor (dada pelo fator de amortecimento $0 < p < 1$ de Brin e Page) ele vai digitar uma URL e “teleportar-se” no grafo.

Com isso, toda página j doa uma fração p/n de sua importância para **todas** as páginas da internet. Além disso, para cada página i que aponta ela doa $(1-p)\beta_{ij}$ de sua importância (com a convenção de que se ela não aponta para ninguém, $\beta_{ij} = 1/n$ para todo i).

Com isso, o sistema de equações envolvendo as importâncias x_i vai ser

$$(1.3) \quad x_i = \sum_{j=1}^n \left(\frac{p}{n} + (1-p)\beta_{ij} \right) x_j.$$

1.1.4. *Vamos ver se isso resolve nosso problema.* Vamos escrever a equação (1.3) no caso da Figura 2. Vamos considerar $p = 0,15$. Assim, $1-p = 0,85$. Como nesse caso $n = 3$, é útil observar que $p/n = 0,05$.

Temos que $d_1 = d_2 = 2$ e $d_3 = 0$. Assim, temos que $\beta_{11} = \beta_{22} = 0$ e $\beta_{13} = \beta_{23} = \beta_{33} = 0,85/3$, ao passo que $\beta_{12} = \beta_{21} = \beta_{31} = \beta_{32} = 0,5$. Portanto, o sistema (1.3) fica

$$\begin{aligned} x_1 &= (0,05 + 0,85 \times 0)x_1 + (0,05 + 0,85 \times 0,5)x_2 + (0,05 + 0,85 \times \frac{1}{3})x_3 \\ x_2 &= (0,05 + 0,85 \times 0,5)x_1 + (0,05 + 0,85 \times 0)x_2 + (0,05 + 0,85 \times \frac{1}{3})x_3 \\ x_3 &= (0,05 + 0,85 \times 0,5)x_1 + (0,05 + 0,85 \times 0,5)x_2 + (0,05 + 0,85 \times \frac{1}{3})x_3. \end{aligned}$$

Fazendo as contas dentro dos parêntesis chegamos a forma

$$\begin{aligned}x_1 &= 0,05x_1 + 0,475x_2 + \frac{1}{3}x_3 \\x_2 &= 0,475x_1 + 0,05x_2 + \frac{1}{3}x_3 \\x_3 &= 0,475x_1 + 0,475x_2 + \frac{1}{3}x_3.\end{aligned}$$

Se você já fez um curso de álgebra linear, você pode resolver esse sistema pelo método do escalonamento. No entanto, como acredito que mais importante do que saber fazer as contas que resolvem um problema é saber *quais* contas resolvem esse problema, você pode fazer como eu fiz: entrei no *Google* e digitei “calculadora de matrizes” e entrei no site <https://matrixcalc.org/pt/>. Clicando no botão “solução pelo método de Gauss” vemos que a solução geral do sistema em função de x_3 é

$$(x_1, x_2, x_3) = \left(\frac{40}{57}x_3, \frac{40}{57}x_3, x_3\right).$$

Como as importâncias devem satisfazer ainda a equação (1.1), ou seja,

$$x_1 + x_2 + x_3 = 1,$$

devemos ter

$$\begin{aligned}\frac{80}{57}x_3 + x_3 &= 1 \\ \therefore x_3 &= \frac{57}{137}\end{aligned}$$

Portanto, o vetor *PageRank* da internet 2 é

$$\left(\frac{2280}{7809}, \frac{2280}{7809}, \frac{57}{137}\right) \approx (0.29, 0.29, 0.42).$$

1.2. Mas e o *PageRank* real? Com o modelo que fizemos, determinar o ranqueamento das páginas da internet equivale a resolver o sistema de equações (1.3). Como a internet já em 1998 possuía algo da ordem de 300 milhões de páginas, isso significa uma quantidade monstruosa de cálculos, desafiadora até para os melhores computadores.

Ou seja, modelamos o problema do ranqueamento da internet num problema (impossível?) de resolver um sistema bizarramente monstruoso de equações.

O que é incrível é que uma vez montado o sistema de equações (1.3), podemos olhar para ele com o ponto de vista da aleatoriedade, do caos e aí... bum! Um novo mundo se abre diante dos nossos olhos, e tudo se resolve, como num passe de mágica o caos traz a certeza, a determinação. Vamos explicar essa poesia dos antagonicos a seguir.

2. CADEIAS DE MARKOV

Vamos olhar Figura 1 de novo e com outra visão agora. Vamos imaginar um usuário que entra na internet num determinado momento e começa a “surfar” no grafo. A cada momento o surfista pula de uma página a outra. Vamos supor como antes, que na maioria das vezes ele segue um link de uma página a outra, mas de vez em quando ele pula de uma página a outra qualquer (digitando a URL de destino). Nesse modelo, a página com maior importância será a página para a qual há uma maior probabilidade do surfista aleatório parar em um dado momento.

Com essa interpretação, podemos colocar o problema do ranqueamento das páginas da internet como um caso particular da seguinte situação geral: suponha agora um sistema qualquer com apenas um número finito de estados possíveis, digamos

$$E_1, \dots, E_n$$

Imagine que a gente não conheça a lei de evolução do sistema

$$f : \{E_1, \dots, E_n\} \rightarrow \{E_1, \dots, E_n\},$$

que diz como ele passa de um estado a outro no futuro. Como entender esse sistema? Vamos analisar essa pergunta a partir de um exemplo específico, simples. Com isso, vamos sair completamente do contexto da internet e falar de outras coisas. Na verdade, ao abrir a mente para novas ideias poderemos voltar muito mais fortes para o nosso problema.

Exemplo 2.1 (Modelos Markovianos de Desemprego). Vamos propor um modelo para descrever a evolução no tempo da taxa de desemprego. Vamos fazer o acordo de que o modelo será muito simples, o bastante para que possamos dar uma descrição matemática completa.

Podemos pensar da seguinte forma: para cada pessoa temos dois estados possíveis em cada instante de tempo. Ou a pessoa encontra-se empregada, ou desempregada. Além disso, no próximo instante de tempo cada pessoa pode mudar de estado ou permanecer no estado anterior.

Então, vamos supor que a economia encontra-se numa situação tal que existe uma certa chance, uma *probabilidade* no jargão matemático, de uma pessoa empregada tornar-se desempregada no mês seguinte, e, reciprocamente, existe uma chance de uma pessoa desempregada conseguir um emprego daqui a um mês. A hipótese crucial que simplifica o nosso modelo, mas que possibilita uma análise matemática *reveladora*, é de que **essas chances permanecem constantes ao longo do tempo**.

Assim, matematicamente podemos dizer que no nosso modelo uma existe uma probabilidade $0 < p < 1$ de uma pessoa inicialmente empregada continuar empregada, e uma chance complementar $1 - p$ dessa mesma pessoa tornar-se desempregada.

Analogamente, existe uma probabilidade $0 < q < 1$ de uma pessoa inicialmente desempregada continuar assim, e uma chance complementar $1 - q$ dela mudar de situação e conseguir um emprego.

Os parâmetros p e q devem refletir a situação econômica: se p e q têm valores muito próximos de 1, isso pode ser lido como um cenário de relativa estabilidade econômica ao passo que se são muito pequenos temos altas chances de aumento da taxa de desemprego.

Podemos interpretar isso como um sistema de dois estados, e que evolui no tempo de acordo com probabilidades que refletem as chances de mudança (ou permanência) de estado.

Suponhamos que $0 < \alpha_0 < 1$ seja a taxa de desemprego do mês atual. Se a economia segue o nosso modelo, qual será a taxa de desemprego do mês seguinte?

Com efeito, seja $\beta_0 = 1 - \alpha_0$. Esse número representa a porcentagem de pessoas empregadas na economia. Além disso, a população é repartida entre empregados e desempregados. Sabemos que dentre os primeiros, uma porcentagem p vai continuar empregada, ao passo que uma porcentagem $1 - p$ deles vai mudar de situação e

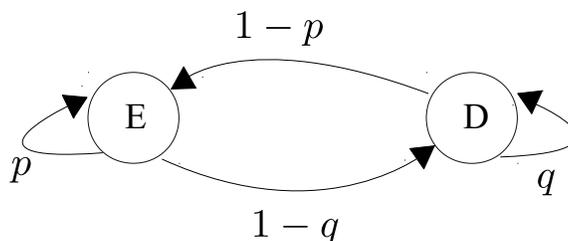


FIGURA 3. O jogo do emprego: a cada mês uma proporção fixa de empregados muda de situação, e uma proporção complementar a essa permanece na mesma.

tornar-se desempregada. Por outro lado, das pessoas que estão desempregadas, uma parcela q tende a assim continuar no próximo mês, ao passo que uma parcela $1 - q$ tende a perder o emprego.

O longo parágrafo acima pode ser traduzido em duas equações simples. Se chamarmos de α_1 a taxa de desemprego do mês seguinte, vimos acima que o conjunto das pessoas desempregadas do próximo mês é a soma das pessoas que continuaram desempregadas (uma proporção q da taxa de desemprego) com aquelas que tornaram-se desempregadas de um mês para o outro (uma proporção $1 - p$ da taxa de emprego). Matematicamente,

$$\alpha_1 = q\alpha_0 + (1 - p)\beta_0.$$

Similarmente, a taxa de emprego no mês seguinte será (veja a figura 3)

$$\beta_1 = p\beta_0 + (1 - q)\alpha_0.$$

Essas duas equações podem ser colocadas numa linguagem ainda mais elegante: a da Álgebra Linear! De fato, em notação matricial as duas equações acima se tornam

$$\begin{bmatrix} \alpha_1 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} q & 1 - p \\ 1 - q & p \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \beta_0 \end{bmatrix}.$$

Assim, a evolução da taxa de desemprego pode ser modelada por uma simples (e eficaz!!!) multiplicação matricial. Como veremos adiante, a *estrutura* específica do tipo de matriz que aparece nesse exemplo permite a dedução de propriedades extremamente ricas e que explicam muitos fenômenos que não saltam aos olhos à primeira vista.

No exemplo acima temos dois estados possíveis e probabilidades de transitar de um estado a outro. Além disso, a soma das probabilidades associadas a um determinado estado é sempre 1: afinal, uma pessoa empregada no mês seguinte com 100% de chance vai estar ou empregada ou desempregada.

Em geral, uma *Cadeia de Markov* refere-se a um sistema com um certo número finito n de estados possíveis em que de um instante de tempo ao próximo o sistema pode transitar de um estado $i \in \{1, \dots, n\}$ a um estado $j \in \{1, \dots, n\}$. Numa Cadeia de Markov, não sabemos exatamente para qual estado vamos se estivermos no estado 3. Basicamente, podemos ir para qualquer estado. A única coisa que

sabemos são as probabilidades de transição $0 < p_{ij} < 1$ do estado j para o estado i .

Então a evolução de uma coisa pode ser descrita adequadamente através da interpretação probabilística. Se inicialmente temos chance x_i de estar em cada estado i , quais serão as chances de estarmos em cada estado i uma unidade de tempo no futuro?

Isso pode ser dito mais formalmente da seguinte maneira. Dada uma distribuição inicial (x_1^0, \dots, x_n^0) de probabilidades para cada estado $1, \dots, n$ qual será a distribuição futura (x_1^1, \dots, x_n^1) ?

Ora, se para cada j temos uma chance x_j de estar no estado j , e ali estando temos uma chance p_{ij} de sair para o estado i , a chance de acabarmos no estado i na próxima etapa será

$$x_i^1 = \sum_{j=1}^n p_{ij} x_j.$$

Como antes podemos escrever a mesma coisa em notação matricial. Seja $P = [p_{ij}]$ a *matriz de transição*. Se $x^0 = (x_1^0, \dots, x_n^0)$ é um vetor de distribuição de probabilidades inicial, o vetor $x^1 = (x_1^1, \dots, x_n^1)$ será dado pela equação matricial

$$x^1 = P(x^0),$$

que explicitamente se escreve

$$\begin{bmatrix} x_1^1 \\ \vdots \\ x_n^1 \end{bmatrix} = \begin{bmatrix} p_{11} & \cdots & p_{1n} \\ \vdots & \ddots & \vdots \\ p_{n1} & \cdots & p_{nn} \end{bmatrix} \begin{bmatrix} x_1^0 \\ \vdots \\ x_n^0 \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^n p_{1j} x_j^0 \\ \vdots \\ \sum_{j=1}^n p_{nj} x_j^0 \end{bmatrix}.$$

Note que como cada número x_i^0 é uma probabilidade deve portanto satisfazer $0 \leq x_i^0 \leq 1$. Além disso, como o sistema está sempre em *algum* estado, devemos ter $\sum_{i=1}^n x_i^0 = 1$. Similarmente, como o sistema sempre transita para algum estado (que pode inclusive ser o mesmo de agora) devemos ter

$$\sum_{i=1}^n p_{ij} = 1,$$

para cada $j = 1, \dots, n$. Esse tipo de matriz é tão especial que merece um nome pomposo.

Definição 2.2. Uma matriz $P = [p_{ij}]$ é dita *estocástica* se todas as suas entradas são números entre 0 e 1 (ou seja $0 < p_{ij} < 1$, para todo par i, j) cujas colunas somam 1 (ou seja $\sum_{i=1}^n p_{ij} = 1$).

Assim, a evolução de uma dada distribuição de probabilidades inicial numa Cadeia de Markov consiste em aplicar sucessivamente uma matriz estocástica a esse vetor de probabilidade inicial.

Ou seja, para cada vetor de distribuição de probabilidades $v = (x_1, \dots, x_n)$, com cada $x_i \in [0, 1]$ representando a probabilidade do sistema estar no estado i no instante inicial, a probabilidade do sistema estar no estado i em 50 unidades de tempo no futuro será a i -ésima entrada do vetor $P^{50}(v)$.

Agora, façamos a seguinte pergunta maliciosa: será que existe um vetor de probabilidade tal que essa distribuição de chances de o sistema estar em cada estado

não muda nunca ao longo do tempo? Seria possível que a distribuição de probabilidades de sistema evolui ao longo do tempo para uma espécie de distribuição de equilíbrio, estacionária, que não muda mais?

A resposta a essa pergunta será a resposta final para o problema do ranqueamento da internet. Com efeito, podemos imaginar temos uma distribuição inicial de probabilidade, cada página P_i tem uma chance $1/n$ de ser visitada no primeiro momento pelo surfista aleatório. Colocamos essas probabilidades no lugar das importâncias x_i na equação (1.3).

Observe que a chance de um usuário estando na página P_j ir para a página P_i vai ser dada por

$$p_{ij} \stackrel{\text{def}}{=} \left(\frac{p}{n} + (1-p)\beta_{ij} \right),$$

pois temos a chance dele se teletransportar p/n somada com a chance dele seguir um possível link $j \rightarrow i$. A matriz $\mathcal{G} = [p_{ij}]$ é chamada matriz Google. Como as vezes a natureza se comporta de modo matemático, temos o seguinte

Lema 2.3. \mathcal{G} é uma matriz estocástica.

Demonstração. Primeiro, observe que $\beta_{ij} = 1/d_j$ se $j \rightarrow i$ e $\beta_{ij} = 0$ caso contrário. Lembre que d_j é o número de páginas para as quais P_j aponta. Veja a figura 1, nela $d_7 = 1/3$, $d_1 = 1$, $d_4 = 1/2$ e assim por diante. Assim, por exemplo a página 7 doa $1/3$ de sua importância à cada página que aponta (P_1 , P_2 e P_6). Isso ilustra o fato geral de que toda página doa toda sua importância. Em outras palavras, fixado j , se $d_j > 0$ então a soma dos β_{ij} é igual a $d_j \times 1/d_j = 1$, e se $d_j = 0$ então a soma dos β_{ij} é $n \times 1/n = 1$. Com a notação de somatório isso fica

$$\sum_{i=1}^n \beta_{ij} = 1.$$

Depois, note que quando somarmos todos os p_{ij} com j fixado e i variando de 1 a n , vamos somar n vezes a parcela p/n (o que vai dar p) e vamos somar

$$(1-p)\beta_{1j} + (1-p)\beta_{2j} + \dots + (1-p)\beta_{nj}.$$

Nessa outra soma podemos colocar $1-p$ em evidência, que vai multiplicar a soma dos β_{ij} . Resumindo, temos

$$\begin{aligned} \sum_{i=1}^n p_{ij} &= \sum_{i=1}^n \left(\frac{p}{n} + (1-p)\beta_{ij} \right) \\ &= n \times \frac{p}{n} + \sum_{i=1}^n (1-p)\beta_{ij} \\ &= p + (1-p) \sum_{i=1}^n \beta_{ij} = p + (1-p) = 1. \end{aligned}$$

Isso prova que a soma das entradas em cada coluna da matriz \mathcal{G} é igual a 1, e demonstra que \mathcal{G} é estocástica. \square

Esse lema diz que para compreender o movimento ao longo do tempo de um surfista aleatório na internet devemos compreender a evolução no tempo de uma Cadeia de Markov especial, para a qual a matriz de transição é \mathcal{G} .

2.0.1. *E no que isso nos ajuda?* Temos um problema que ainda não dissemos como resolver: a equação (1.3) do vetor *PageRank* possui **sempre** solução? Corrigimos o modelo justamente para evitar *patologias*, exemplos sem solução como o grafo da Figura 2. Além disso, no “mundo real” o sistema (1.3) é composto por milhões de equações. Assim, ainda que a solução exista, como calculá-la?

Para responder essas perguntas, mais uma vez, vamos abrir ainda mais nossos horizontes e olhar para um conceito matemático ainda mais geral do que Cadeias de Markov.

Isso é como tirar férias e fazer uma viagem. Saímos do lugar onde estamos acostumados a viver para conhecer novos horizontes, climas diferentes, comidas diferentes, mas depois voltamos para casa e nosso trabalho de todos os dias. Mas nunca voltamos iguais, e sempre voltamos com novas ideias.

Então vamos viajar um pouco para voltarmos com novas ideias sobre como estudar o comportamento assintótico de cadeias de Markov.

3. SISTEMAS DINÂMICOS

Um *sistema dinâmico* é um modelo matemático que descreve a evolução no tempo de algum fenômeno de interesse. Pode ser por exemplo um sistema mecânico, um sistema econômico, um sistema social e por aí vai! As cadeias de Markov são apenas um exemplo particular de sistema dinâmico.

Se o tempo evolui de forma discreta, o modelo matemático se reduz ao dado de uma função $f : X \rightarrow X$, que a cada configuração inicial $x_0 \in X$ associa a configuração 1 unidade de tempo adiante no futuro $f(x_0)$.

Exemplo 3.1 (Sinuca circular). Você já jogou sinuca numa mesa circular? Pois bem, imagine uma mesa circular e uma bolinha posicionada em algum lugar dessa mesa. Ao atirar a bolinha em alguma direção, suponha que ela vai se deslocar na mesa sem atrito e que não há perda de energia ao bater com o bordo da mesa. Nesse caso a bolinha vai ricochetear cada vez que se encontrar com o bordo da mesa. Temos um problema concreto: tentar dizer qualitativamente o que vai acontecer com a bolinha. Fazemos hipóteses simplificadoras (nada de atrito nem de perda de energia) para tentar obter um modelo matemático que descreva a trajetória da bolinha.

Com essas hipóteses, podemos descrever matematicamente a situação da seguinte forma. Consideramos o bordo da mesa como o círculo \mathbb{S}^1 , parametrizado pelo intervalo $[0, 2\pi)$. Assim, cada ponto do círculo corresponde a um ângulo $\theta \in [0, 2\pi)$ e quando damos uma volta completa no círculo voltamos ao ângulo zero.³

Usando que o ângulo de um raio com a tangente ao círculo vale sempre 90 graus, podemos mostrar facilmente (veja a Figura 4) sempre que $x_0 \in \mathbb{S}^1$ e $x_1 \in \mathbb{S}^1$ são batidas sucessivas então $x_1 - x_0 = \theta$, onde θ depende apenas do ângulo α que a primeira batida faz com a tangente do círculo.

Ou seja, a trajetória que a bolinha vai fazer é completamente determinada pela função $f : \mathbb{S}^1 \rightarrow \mathbb{S}^1$ que a cada ponto $x_0 \in \mathbb{S}^1$ associa o ponto $f(x_0)$ obtido pela rotação por um ângulo θ do ponto x_0 .

Temos então um sistema dinâmico $f : \mathbb{S}^1 \rightarrow \mathbb{S}^1$, um objeto matemático abstrato, que provém da nossa análise de um problema concreto!

³Se você já viu o conceito de relação de equivalência, isso equivale a descrever \mathbb{S}^1 como o quociente \mathbb{R} pela relação de equivalência $x \sim y \iff x - y \in 2\pi\mathbb{Z}$.

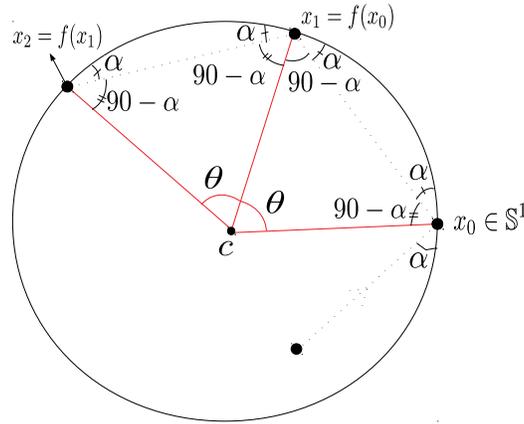


FIGURA 4. Rotação do círculo.

Em outras palavras, no nosso modelo matemático o jogo de bilhar na mesa circular pode ser entendido através da aplicação de rotação f . O comportamento do bilhar fica determinado pela sequência $x_n = f^n(x_0)$ que corresponde a aplicarmos sucessivamente por n vezes a função f ao ponto x_0 .

A rotação do círculo, embora extremamente simples, é um objeto matemático maravilhoso sobre o qual podemos falar muitas coisas surpreendentes. Não vamos explorá-la nesse texto, contudo. Deixo apenas o exercício abaixo para o entretenimento do leitor.

- Exercício 3.2.** (a) Se $\theta \in \mathbb{Q}$ então a sequência $x_n = f^n(x_0)$ é finita.
 (b) Se $\theta \in \mathbb{R} \setminus \mathbb{Q}$ então para todo $x \in \mathbb{S}^1$ existe uma subsequência $n_k \rightarrow \infty$ tal que $x_{n_k} \rightarrow x$. Em particular, x_n é uma sequência infinita que se espalha densamente pelo círculo.

3.1. Espaços Métricos. O palco onde acontece um sistema dinâmico é, em geral, um espaço onde podemos medir distâncias entre pontos. Parece uma coisa tão corriqueira, medir distâncias, que nem percebemos que para cada problema pode haver uma maneira de medir distâncias mais adequada.

Por exemplo, imagine um taxista que se desloca do ponto A ao ponto B (veja a Figura 5). O custo desse deslocamento depende da distância entre os pontos A e B. Mas qual seria a medida que realmente interessa ao taxista? O comprimento da linha reta que liga A e B (o caminho mais curto, que é como pensamos em medir distâncias usualmente) ou a distância que ele vai percorrer efetivamente?

Claramente para o taxista é impossível seguir a linha reta que vai de A até B e por isso ele vai percorrer o caminho em vermelho na Figura 5. Portanto, para os objetivos dele, a distância entre A e B é o comprimento da curva vermelha.

Esse exemplo simples deixa entrever que a noção de “distância entre A e B” não é universal. Depende do contexto. Ou em outras palavras, para cada problema que

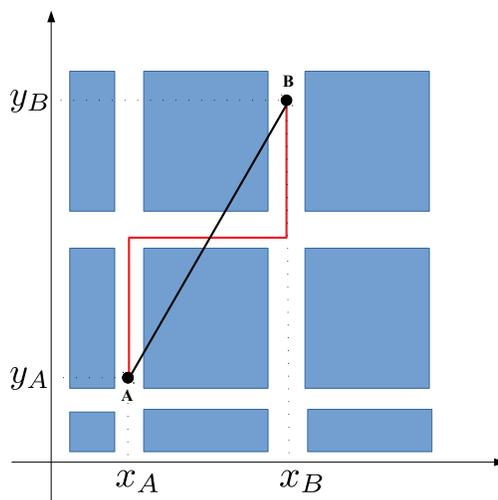


FIGURA 5. Distância efetivamente percorrida pelo taxista vs comprimento do segmento AB.

pretendemos resolver, deve haver uma maneira medir essa distância que seja mais adequada.

Em matemática, toda vez que percebemos que um determinado conceito é na verdade mais maleável do que imaginávamos a priori tentamos *generalizá-lo*. O conceito que vamos introduzir a seguir surge exatamente da flexibilização da ideia de medir distâncias.

Definição 3.3. Seja X um conjunto. Uma função $d : X \times X \rightarrow [0, \infty)$ é chamada uma *função distância* ou simplesmente uma *métrica* se satisfaz as condições a seguir

- (1) $d(x, y) = 0 \iff x = y$
- (2) $d(x, y) = d(y, x)$
- (3) $d(x, z) \leq d(x, y) + d(y, z)$ (Desigualdade Triangular)

Um conjunto X munido de uma métrica d é chamado um *espaço métrico*. Denotamos um espaço métrico por (X, d) .

Exemplo 3.4. (a) Considere X como sendo o plano cartesiano \mathbb{R}^2 , como na Figura 5. Podemos medir a distância entre dois pontos A e B através do comprimento do segmento de reta que liga A e B . Esse segmento é a hipotenusa de um triângulo retângulo cujos catetos medem, respectivamente $|x_B - x_A|$ e $|y_B - y_A|$. Aplicando então o Teorema de Pitágoras temos

$$d_E(A, B) = \sqrt{|x_B - x_A|^2 + |y_B - y_A|^2}.$$

A esse número damos o nome de *distância euclidiana* entre A e B .

(b) A distância do taxista, por sua vez, é o comprimento da curva vermelha. Analisando a Figura 5 vemos que

$$d_T(A, B) = |x_B - x_A| + |y_B - y_A|.$$

Esse número fornece uma função distância diferente para o mesmo conjunto $X = \mathbb{R}^2$.

- (c) Seja $X = \mathbb{R}^d$ o espaço euclidiano de dimensão d , ou seja \mathbb{R}^d é apenas o conjunto das listas ordenadas de números com d entradas. A métrica euclidiana em \mathbb{R}^d , que generaliza a métrica do item (a), é a função que a cada par $x = (x_1, \dots, x_d), y = (y_1, \dots, y_d) \in \mathbb{R}^d$ associa o número

$$d_E(x, y) = \sqrt{\sum_{j=1}^d |x_j - y_j|^2}.$$

Nesse texto, um espaço métrico serve principalmente de palco para um sistema dinâmico atuar. Cada ponto no espaço métrico representa uma configuração do sistema, e dada uma configuração inicial, o sistema evolui no tempo assumindo a cada instante uma nova configuração. Dessa forma, a evolução do sistema pode ser traduzida como uma sucessão de pontos no espaço. O que estamos interessados é em saber para onde vai essa sucessão de pontos. Será que ela converge para algum lugar, ou se comporta erraticamente no espaço?

Uma característica fundamental da matemática é o rigor. Sempre que nos debruçamos em um problema, procuramos formular o pouco que entendemos de forma precisa, para que possamos esmiuçar os mínimos detalhes. Por isso, para estudar o comportamento de um sistema dinâmico um dos conceitos fundamentais é a ideia de convergência.

Definição 3.5. Uma sequência $\{x_n\}$ num espaço métrico X é dita *convergente* a um ponto $x \in X$ se, para todo $\varepsilon > 0$, existe $n_0 \in \mathbb{N}$ tal que

$$n \geq n_0 \implies d(x_n, x) < \varepsilon.$$

A sequência de números $x_n = 1/n$ converge à 0 monotonamente, pois conforme n aumenta, $1/n$ diminui. Por outro lado, a sequência de pontos no plano $x_n = (n, n^2 - n)$ não converge nem se acumula em nenhum lugar: ela “sobe” pelo gráfico da função $y = x^2 - x$ em direção ao infinito! Veja a figura 6

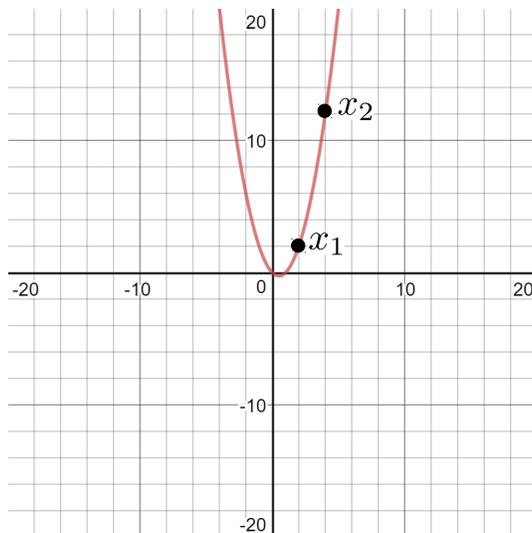


FIGURA 6. A curva vermelha é o gráfico de $y = x^2 - x$.

A maioria das seqüências que podemos inventar com fórmulas algébricas assim vai ter um comportamento razoavelmente monótono, chato. Ou vai convergir a um ponto, ou vai se acumular hora aqui, hora acolá (como a seqüência $x_n = (-1)^n$, que hora vale 1, hora vale -1).

Por outro lado, quando olhamos um número, como 3,1415, podemos decompô-lo em uma *parte inteira* 3 e uma *parte fracionária* 0,1415. Com essa operação, a seqüência definida por

$$x_n = \text{parte fracionária de } n\pi$$

é de outro tipo. Primeiro, ela não diverge para o infinito pois ao tomarmos a parte fracionária estamos olhando apenas números entre 0 e 1. Depois ela não converge a um ponto nenhum; ela se acumula densamente em todos os pontos entre 0 e 1! (compare com o Exercício 3.2).

Mais uma vez, acabamos de passar por um conceito importante. Vamos formalizar essa ideia de uma seqüência que se acumula em lugares diferentes.

Definição 3.6. Uma *subseqüência* de uma seqüência $\{x_n\}$ é qualquer subconjunto $\{x_{n_k}\} \subset \{x_n\}$ tal que os índices n_k da subseqüência formem um subconjunto ilimitado dos naturais.

Por exemplo, a subseqüência dos termos x_n com n par, com n primo, com n múltiplo de um googolplex etc...

Quando uma seqüência se acumula em algum lugar é porque ela possui uma subseqüência que converge a um ponto. Isso nos permite distinguir dois tipos de espaços métricos: aqueles onde seqüências são permitidas de escapar ao infinito e aqueles onde isso não é possível. Do ponto de vista de um sistema dinâmico, uma trajetória que escapa para o infinito pode ser pensada como tendo um comportamento trivial. Isso diz que o tipo de palco mais adequado para um sistema dinâmico atuar de ver o seguinte:

Definição 3.7. Um espaço métrico X é dito *compacto* se toda seqüência admite subseqüência convergente a um ponto de X .

A seguir vamos ver um exemplo de espaço métrico compacto.

Definição 3.8. (1) Um subconjunto $F \subset X$ de um espaço métrico X é dito *fechado* se toda seqüência de pontos em F que converge a algum ponto, converge na verdade a um ponto de F , i.e. se $x_n \in F$ e $x_n \rightarrow x$ então $x \in F$.

(2) Um subconjunto $F \subset \mathbb{R}^d$ é dito *limitado* se a distância entre pontos de F e a origem é limitada, i.e se existe um número $c > 0$ tal que para todo $x \in F$, vale $d_E(x, 0) < c$.

Num espaço euclideano quando uma seqüência não escapa ao infinito ela obrigatoriamente se acumula em algum lugar.⁴

Teorema 3.9. *Se $F \subset \mathbb{R}^d$ é fechado e limitado então F é compacto.*

Demonstração. Apesar de um dos pilares da matemática ser o rigor, nesse artigo eu gostaria que o rigor fosse crescendo aos poucos. Além disso, as ideias constituem o

⁴Esse é um fenômeno exclusivo dos espaços de dimensão finita. Em espaços de dimensão infinita uma seqüência pode simplesmente pular de uma dimensão a outra, sem escapar para o infinito e sem se acumular em nenhum lugar.

outro pilar da matemática. Por isso, para essa demonstração vou apresentar apenas um esboço contendo as ideias principais. Vamos falar da demonstração como se $d = 2$ (ou seja, como se estivéssemos no plano cartesiano). Isso traz apenas um conforto visual, mas tecnicamente os argumentos são os mesmos.

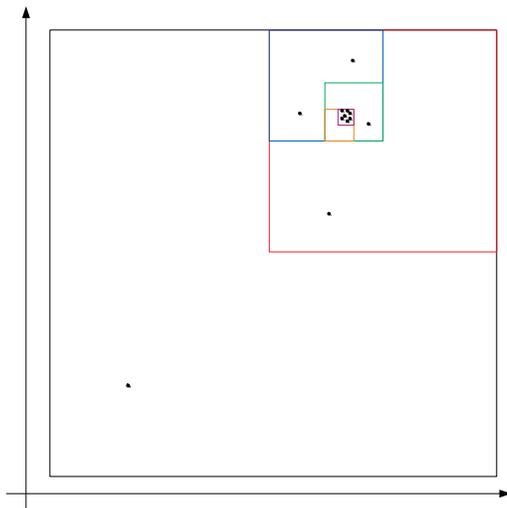


FIGURA 7. Os quadradinhos cada vez menores contém sempre infinitos pontos da sequência.

Inicialmente vamos ver que toda sequência $\{x_n\}$ que é limitada possui subsequência convergente. Primeiro, por ser limitada podemos assumir que todos os pontos estão contidos num quadrado grande. Dividimos esse quadrado em quatro quadrados com metade do tamanho cada um. Como a sequência tem infinitos pontos, em pelo menos um desses quadradinhos menores vamos ter infinitos pontos da sequência. Agora, dividimos esse quadradinho menor em quatro quadradinhos ainda menores, com metade do tamanho. Mais uma vez, como a sequência tem infinitos pontos, em um desses quadradinhos temos infinitos pontos da sequência. Prosseguimos assim, tomando quadradinhos cada vez menores sempre contendo infinitos termos da sequência. Veja a Figura 7. Como os quadradinhos são cada vez menores eles se transformam em um ponto no limite. Esse ponto é o limite de uma subsequência porque em cada etapa a sequência tem infinitos termos.

Agora isso diz que toda sequência em um conjunto limitado possui subsequência convergente. Se esse conjunto for também fechado, o limite da subsequência será um elemento do conjunto. Isso completa o esboço da prova. \square

O teorema que acabamos de ver é um dos resultados fundamentais daquilo que chamamos em matemática de *análise*. A análise lida com processos infinitos, questões de continuidade e aplicações desses conceitos. Uma das aplicações mais importantes da análise é o que veremos a seguir.

4. TEOREMAS DE PONTO FIXO

Quando temos um sistema dinâmico $f : X \rightarrow X$, as vezes existem pontos cuja trajetória pelo sistema é a mais trivial possível: ela não muda nunca. A esses

pontos, damos o nome de *pontos fixos*. Em outras palavras, um ponto $x_0 \in X$ do espaço de configurações é dito um ponto fixo se $f(x_0) = x_0$.

Nem todo sistema dinâmico possui pontos fixos. Por exemplo, a função $f : \mathbb{R} \rightarrow \mathbb{R}$ dada por $f(x) = x + 1$ não fixa nenhum ponto. No entanto, as vezes, propriedades particulares do espaço de configurações e do sistema podem implicar a existência de pontos fixos.

Por outro lado, garantir a existência de pontos fixos é um problema na fronteira entre dinâmica e análise. Afinal, como bem me ensinou o mestre Jedi Alexander Arbieto, a pergunta dinâmica

$$\text{Existe } x_0 \in X \text{ tal que } f(x_0) = x_0?$$

pode ser interpretada como uma pergunta de análise

$$\text{Existe solução, dentro do espaço } X, \text{ para a equação } f(x_0) = x_0?$$

Assim é por exemplo que muitas equações são resolvidas por intermédio de teoremas de ponto fixo e que estes por sua vez são provados via técnicas analíticas. Vejamos abaixo alguns exemplos de teoremas de ponto fixo.

Exemplo 4.1. (1) Seja $f : [0, 1] \rightarrow [0, 1]$ contínua.⁵ Então, f possui pontos fixos. De fato, se nem 0 nem 1 são fixos por f , caso em que o teorema estaria verificado, consideramos $g(x) = f(x) - x$. Então, como $f(0) > 0$ e $f(1) < 1$ temos que $g : [0, 1] \rightarrow \mathbb{R}$ satisfaz $g(0) > 0$ e $g(1) < 0$. Como f é contínua, g também o é, e portanto, pelo teorema do valor intermediário, (veja a Figura 9) existe $x_0 \in (0, 1)$ tal que $g(x_0) = 0$, logo x_0 é um ponto fixo para f .

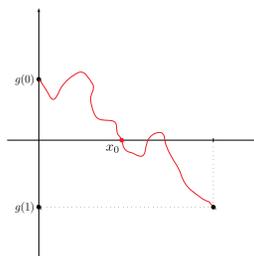


FIGURA 8. Teorema do Valor Intermediário: o gráfico de g tem que descer continuamente do ponto $g(0)$ até o ponto $g(1)$. Para fazer isso, ele é obrigado a cruzar o eixo horizontal.

O fato que acabamos de demonstrar é o caso em dimensão 1 do famoso Teorema do Ponto Fixo de Brower, que diz que toda aplicação contínua da uma bola d -dimensional em si mesma possui pelo menos um ponto fixo. Podemos interpretar esse enunciado da seguinte forma: se você tentar embaralhar os pontos de uma bola continuamente, pelo menos um deles vai ficar parado. Nesse caso, é como se a forma (uma bola) do espaço de configurações, bem como o fato de f respeitar essa forma (sendo contínua), *impusessem* a existência do ponto fixo.

⁵Um sistema dinâmico $f : X \rightarrow X$ é dito uma *função contínua* se, para cada $x_0 \in X$ e cada $\varepsilon > 0$ existe um $\delta > 0$ tais que $d(x, x_0) < \delta \implies d(f(x), f(x_0)) < \varepsilon$. Em palavras, a função f não dá saltos bruscos.

- (2) Seja (X, d) um espaço métrico completo⁶ e $f : X \rightarrow X$ uma *contração*, i.e. existe $0 < c < 1$ tal que $d(f(x), f(y)) \leq cd(x, y)$, para todos $x, y \in X$. Então f possui um único ponto fixo. Esse é o também famoso teorema do ponto fixo de Banach. Aqui, a dinâmica especial que contrai o espaço todo força a existência e a unicidade do ponto fixo. Este é um dos teoremas de pontos fixos mais utilizados, com inúmeras aplicações pela matemática afora.⁷

4.1. O Teorema de Edelstein. Vamos agora efetivamente colocar as mãos a trabalhar e *demonstrar* um teorema de ponto fixo. Eu escolhi apresentar um enunciado que se assemelha muito ao teorema de Banach, mas que é muito menos conhecido do que o seu primo famoso. Ei-lo:

Teorema 4.2 (Edelstein [4]). *Seja (X, d) um espaço métrico compacto e seja $f : X \rightarrow X$ tal que $d(f(x), f(y)) < d(x, y)$, para todos $x, y \in X$, com $x \neq y$. Então, f possui um único ponto fixo e além disso $f^n(x)$ converge a esse ponto fixo, qualquer que seja $x \in X$.*

Alguns leitores, habituados à filosofia típica da análise de que compacidade implica em uniformidade, poderiam se perguntar se o teorema de Edelstein não é na realidade um rele Corolário do seu primo famoso, Banach.

Isso porque, podemos ser induzidos a pensar que como

$$\frac{d(f(x), f(y))}{d(x, y)} < 1,$$

para todo x, y , e como f é contínua, por compacidade teríamos um máximo que deve ser estritamente menor do que 1. No entanto, a desigualdade mencionada só vale quando $x \neq y$. Ou seja, ela vale em $X \times X \setminus \Delta$, onde $\Delta = \{(x, y) \in X \times X; x = y\}$. E este é um subconjunto não compacto do produto cartesiano!

E na verdade podemos ter funções que satisfazem a hipótese de Edelstein, são contrações fracas, mas não são contrações. Por exemplo, a função $f : [0, 1] \rightarrow [0, 1]$ dada por $f(x) = \frac{1}{1+x}$. Observe que

$$|f(x) - f(y)| = \left| \frac{1}{1+x} - \frac{1}{1+y} \right| = \left| \frac{x-y}{(1+x)(1+y)} \right|.$$

Logo, se $x \neq y$, então

$$\frac{|f(x) - f(y)|}{|x - y|} = \frac{1}{(1+x)(1+y)} < 1,$$

No entanto, quando $x, y \rightarrow 0$, temos $\frac{|f(x) - f(y)|}{|x - y|} \rightarrow 1$.

Exercício 4.3. A diferença entre os dois teoremas pode ficar ainda mais dramática: considere $f : [0, 1] \rightarrow [0, 1]$, dada por $f(x) = \frac{x}{1+x}$. Mostre que f^n não é uma contração em nenhuma vizinhança de 0, para todo $n > 0$.

⁶Se você nunca foi apresentado a essa definição, pense numa aplicação $f : \mathbb{R}^d \rightarrow \mathbb{R}^d$ que contrai as distâncias por um fator uniforme $c \in (0, 1)$. A demonstração inclusive não vai mudar nada desse caso particular para o caso geral.

⁷Essas e muitos outros tópicos de análise o leitor pode encontrar no livro [7]

Prova do Teorema de Edelstein. ⁸ Antes de começar a demonstração efetivamente, deixo um **exercício** para o leitor: prove que se $f : X \rightarrow X$ satisfaz $d(f(x), f(y)) < d(x, y)$, para todos $x \neq y$ em X , então f é contínua. Intuitivamente, como f contrai distâncias, não podem haver saltos nos valores que assume.

Agora, passamos a demonstração.

Seja $g : X \rightarrow [0, +\infty)$, dada por $g(x) = d(x, f(x))$. Como f é contínua e a função “medir distâncias” também é contínua, segue que g é contínua, ao compormos funções contínuas continuamos tendo uma função contínua. Aqui usamos a compacidade do espaço de configurações X . Ela implica que g possui um mínimo.⁹ Seja $a \in X$ um ponto no qual esse mínimo é assumido. Ou seja, $g(x) \geq g(a)$, para todo $x \in X$.

Isso implica que a é obrigatoriamente um ponto fixo de f . De fato, se não fosse teríamos $a \neq f(a)$, e portanto

$$g(f(a)) = d(f(a), f^2(a)) < d(a, f(a)) = g(a),$$

o que é absurdo. Isso prova a existência de um ponto fixo para f . A unicidade segue com um raciocínio análogo. Com efeito, se $b \neq a$ fosse outro ponto fixo de f então teríamos

$$d(a, b) = d(f(a), f(b)) < d(a, b),$$

um absurdo.

Passamos agora à demonstrar que o único ponto fixo é também um *atrator global para a dinâmica*: se aplicamos sucessivamente a função f a qualquer ponto do espaço X , nos aproximamos cada vez mais do ponto fixo.

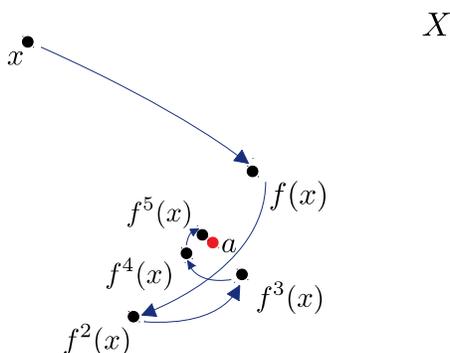


FIGURA 9. A trajetória de $x \in X$ se aproxima assintoticamente do ponto fixo a .

⁸Essa bela demonstração foi retirada de um artigo disponível (em inglês) na página pessoal de Keith Conrad. Aliás, aos amantes da matemática, recomendo visitar a página dele.

⁹Esse é um teorema clássico da análise, cuja demonstração deixo a cargo do leitor: toda função contínua num espaço compacto assume um máximo e um mínimo. Você também pode achar a prova em [7] se preferir.

Tomemos um ponto $x_0 \in X$ qualquer. Nosso objetivo é demonstrar que a sequência $x_n := f^n(x_0)$ converge ao ponto a . Se supomos por contradição que isso não acontece é porque infinitos termos da sequência passam longe do ponto a , mas como X é compacto esses infinitos termos tem que se acumular em algum lugar, que não pode ser o ponto a . Logo, existe uma subsequência x_{n_i} que converge a outro ponto $y \in X$ ($y \neq a$). Para mais detalhes quanto a esse passo, veja o exercício que segue essa demonstração.

Em particular, como a é o único ponto fixo de f , devemos ter $f(y) \neq y$. Considere a sequência de números reais $\mu_n = d(x_n, a)$. Temos que

$$\begin{aligned}\mu_{n+1} &= d(x_{n+1}, a) = d(f(x_n), f(a)) \\ &< d(x_n, a) = \mu_n.\end{aligned}$$

Isso mostra que $\{\mu_n\}$ é uma sequência monótona decrescente, portanto converge. Seja $\mu = \lim \mu_n$.

Agora observe que, pela continuidade de f , se $x_{n_i} \rightarrow y$ então $x_{n_i+1} \rightarrow f(y)$. Segue daí que

$$\begin{aligned}\mu &= \lim \mu_{n_i+1} = \lim d(x_{n_i+1}, a) \\ &= d(f(y), a) = d(f(y), f(a)) < d(y, a) \\ &= \lim d(x_{n_i}, a) = \lim \mu_{n_i} = \mu,\end{aligned}$$

o que é absurdo. Isso completa a prova. \square

Exercício 4.4. Prove a afirmação esboçada no texto: seja x_n uma sequência no espaço métrico compacto X que **não** converge ao ponto $a \in X$. Então, existe $y \in X$, com $y \neq a$, e uma subsequência x_{n_i} que converge a y .

5. O TEOREMA DE PERRON-FROBENIUS

Agora que estamos de volta da nossa viagem para o mundo dos sistemas dinâmicos e seus pontos fixos, vamos voltar nossa atenção ao problema que motivou este artigo. Inspirados pela nossa viagem, vamos adotar um ponto de vista dinâmico para o problema do ranqueamento da internet.

Primeiro, decidimos que a importância de uma página será medida pela probabilidade de um usuário eventualmente clicar nela. Páginas muito importantes têm altas chances de serem clicadas. Parece razoável, não?

No instante inicial em que o usuário entra na internet cada página tem a mesma chance de ser visitada ($1/n$, onde n é a quantidade de páginas da internet). Na linguagem de cadeias de Markov, cada página representa um estado do sistema e a página onde o usuário estiver num dado instante de tempo é o estado do sistema naquele instante.

A partir daí, as probabilidades de transição são atualizadas, como numa cadeia de Markov, usando-se a matriz \mathcal{G} , que é estocástica.

Agora observe a equação (1.3). Se $x = (x_1, \dots, x_n)$ satisfaz (1.3) então $x = \mathcal{G}(x)$. Ou seja, a pergunta sobre como se distribuem assintoticamente (i.e. quando deixamos o tempo correr indeterminadamente) as probabilidades de transição da cadeia de Markov do Google, ou de uma cadeia de Markov em geral, se traduz matematicamente da seguinte maneira: existe $v = (x_1, \dots, x_n)$, com $x_i \in [0, 1]$ e $\sum_{i=1}^n x_i = 1$ (ou seja, v é um *vetor de probabilidades* da Cadeia de Markov) tal que a matriz de transição $P = [p_{ij}]$ aplicada a v não faz nada com ele (isto é, $P(v) = v$)?

Essa pergunta pode ser pensada como um **problema de ponto fixo** para a aplicação $P : \mathbb{R}^n \rightarrow \mathbb{R}^n$. Ou ainda, na linguagem da álgebra linear, como o problema de demonstrar a existência de um vetor de probabilidades que é um autovetor da matriz P , associado ao autovalor 1.

Eu, pessoalmente adoro esse problema! Ele tem todas as características dos melhores problemas de matemática.

- (1) É fácil de enunciar. Se você acabou de sair do ensino médio, em cinco linhas te digo o que é uma matriz estocástica e a pergunta é apenas saber se existe uma lista de números em $[0, 1]$ que quando multiplicada por essa matriz permanece inalterada.
- (2) É difícil de provar! Se você tem medo de coisas com dimensão maior do que 3, mesmo em dimensão 3 é um problema desafiador.
- (3) A solução do problema *explica* muitas coisas e de fato aprendemos coisas novas ao resolver esse problema.

O último item acima é especialmente verdade nesse caso. A solução do problema é um dos resultados mais importantes da matemática do século XX.

Teorema 5.1 (Perron-Frobenius). *Seja $P = [p_{ij}]_{n \times n}$ uma matriz estocástica. Então existe um único vetor $v = (a_1, \dots, a_n) \in \mathbb{R}^n$, com $0 < a_i < 1$, para todo $i = 1, \dots, n$ e com $\sum_{i=1}^n a_i = 1$ tal que $P(v) = v$. Além disso, se $v_0 = (x_1, \dots, x_n)$ é qualquer outro vetor com $0 \leq x_i \leq 1$ e $\sum_{i=1}^n x_i = 1$ então*

$$P^n(v_0) \rightarrow v.$$

O Teorema de Perron-Frobenius pode ser lido, no contexto de Cadeias de Markov, da seguinte forma. *Existe uma única distribuição de probabilidades que é estacionária para o sistema. Além disso, ao longo do tempo, qualquer outra distribuição converge para esta distribuição de equilíbrio.* Essa última frase inclusive mostra como o teorema pode ser aplicado, em caso de necessidade, para efetivamente encontrar a configuração de equilíbrio: basta deixar o sistema fluir, ou seja aplicar a matriz algumas vezes, para obtermos uma configuração próxima da configuração estacionária.

Esse é o algoritmo *PageRank*: aplicar a matriz \mathcal{G} sucessivamente a um vetor de probabilidades inicial para obter aproximações cada vez melhores da configuração estacionária.

Nas linhas que seguem vamos ver a demonstração do Teorema de Perron-Frobenius. Como eu gosto de um olhar problemas bonitos por todos os ângulos possíveis, vou apresentar primeiro a solução em dimensão 3.

5.1. Prova da existência da configuração estacionária em dimensão três.

¹⁰ Nosso primeiro objetivo é demonstrar que existe $v \in \mathbb{R}^3$ tal que $P(v) = v$. Note que isto é equivalente a mostrar que $0 = P(v) - v = (P - I)v$. Portanto, devemos mostrar que existe pelo menos um vetor com todas as entradas positivas e que esteja no *núcleo* da transformação linear $P - I$, onde I , como sempre, representa a matriz identidade.

¹⁰Nessa demonstração vou fazer referência a alguns resultados básicos de álgebra linear, que podem ser encontrados em [6] e [8]

Vamos mostrar que o núcleo de $I - P$ é não-trivial. Note que

$$I - P = \begin{bmatrix} 1 - p_{11} & -p_{12} & -p_{13} \\ -p_{21} & 1 - p_{22} & -p_{23} \\ -p_{31} & -p_{32} & 1 - p_{33} \end{bmatrix}.$$

Como $\sum_{i=1}^3 p_{ij} = 1$, para todo $j = 1, 2, 3$, vemos que a soma de cada coluna de $I - P$ é igual a zero. Como a coluna j é a imagem $(I - P)e_j$, vemos que a imagem de $I - P$ está contida no núcleo do funcional linear $f : \mathbb{R}^3 \rightarrow \mathbb{R}$, dado por $f(x, y, z) = x + y + z$. Como o núcleo de f é um plano e como, pelo que acabamos de ver, a imagem de $I - P$ está contida nesse plano, concluímos que a imagem de $I - P$ possui dimensão no máximo 2. Pelo Teorema do Núcleo e da Imagem, concluímos que o núcleo de $I - P$ possui dimensão pelo menos 1 e, portanto, é não-trivial.

Isso prova que existe $v \in \mathbb{R}^3$ diferente do vetor nulo, tal que $P(v) = v$. Vamos agora mostrar que podemos escolher um tal v com todas as entradas positivas. Para isso, supomos por contradição, que $v = (x, y, z)$ possui entradas negativas. Note que se todas as entradas de v forem negativas, basta tomarmos $-v$. Então, podemos supor que apenas duas entradas podem ser negativas (digamos, no máximo x e y são negativos; os outros casos são análogos).

Como $(I - P)v = 0$, temos que

$$\begin{bmatrix} 1 - p_{11} & -p_{12} & -p_{13} \\ -p_{21} & 1 - p_{22} & -p_{23} \\ -p_{31} & -p_{32} & 1 - p_{33} \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

Se $x < 0$, $y, z \geq 0$ como $1 - p_{11} > 0$, teremos que

$$0 = (1 - p_{11})x - p_{12}y - p_{13}z < 0,$$

o que é absurdo. Se $x, y < 0$ e $z \geq 0$, teremos que

$$0 = -p_{31}x - p_{32}y + (1 - p_{33})z > 0,$$

o que também é absurdo. Vemos assim que, em qualquer caso, assumir que v possui entradas < 0 nos leva a um absurdo. Portanto todas as entradas de v são não negativas. Para requerer que a soma das entradas de v dá 1, usamos um argumento geométrico que veremos a seguir.¹¹ \square

Exercício 5.2. Releia a demonstração acima e com lápis e papel do lado refaça-a em dimensão dois.

5.2. A pseudo métrica projetiva de Hilbert. Uma das coisas lindas sobre a prova do Perron-Frobenius, que já foi citada acima mas que vou repetir porque isso me emociona, é que realmente aprendemos algo novo ao resolver o problema. *Spoiler alert:* a ideia da demonstração é baseada na observação de que como as entradas de $P = [p_{ij}]$ são todas positivas o quadrante positivo

$$\mathbb{R}_+^n := \{x = (x_1, \dots, x_n) \in \mathbb{R}^n; x_i \geq 0\}$$

é estritamente invariante pela ação de P . Se você olhar bem, o conjunto \mathbb{R}_+^n se assemelha a um cone (meio quadrado, mas vale a analogia, não?). Assim, podemos tentar mostrar que a matriz P vai encolhendo esse cone cada vez mais e mais até que no final ele é transformado numa reta. Veja a Figura 10

¹¹A reta gerada por um vetor com todas as entradas positivas cruza uma única vez o simplexo probabilístico. Veja a Figura 10.

Essa imagem é muito sedutora, mas a dimensão n , que pode ser arbitrariamente grande (e é muito importante para as aplicações que o teorema seja verdade mesmo em dimensões incrivelmente gigantescas). Seguindo uma ideia do matemático americano G.Birkhoff vamos usar uma abordagem conceitual astuta para vencer essa dificuldade.

Então, sem mais delongas, vamos direto ao primeiro capítulo dessa história.

5.2.1. O *simplexo probabilístico*.

Definição 5.3. O conjunto $\mathbb{P}_+^n = \{x \in \mathbb{R}_+^n; 0 \leq x_i \leq 1 \text{ e } \sum_{i=1}^n x_i = 1\}$ é chamado o *simplexo probabilístico*.

Vamos considerar também o conjunto $\mathbb{P}_{++}^n = \{x \in \mathbb{P}_+^n; 0 < x_i < 1\}$. Note que \mathbb{P}_+^n é compacto (veja o Lema 5.9 mais adiante) mas \mathbb{P}_{++}^n não! Por outro lado, lembre-se que uma matriz $P = [p_{ij}]$ é uma aplicação contínua $P : \mathbb{R}^n \rightarrow \mathbb{R}^n$. Em particular, $X := P(\mathbb{P}_+^n) \subset \mathbb{R}^n$ é compacto, pois é a imagem de um compacto por uma aplicação contínua.¹²

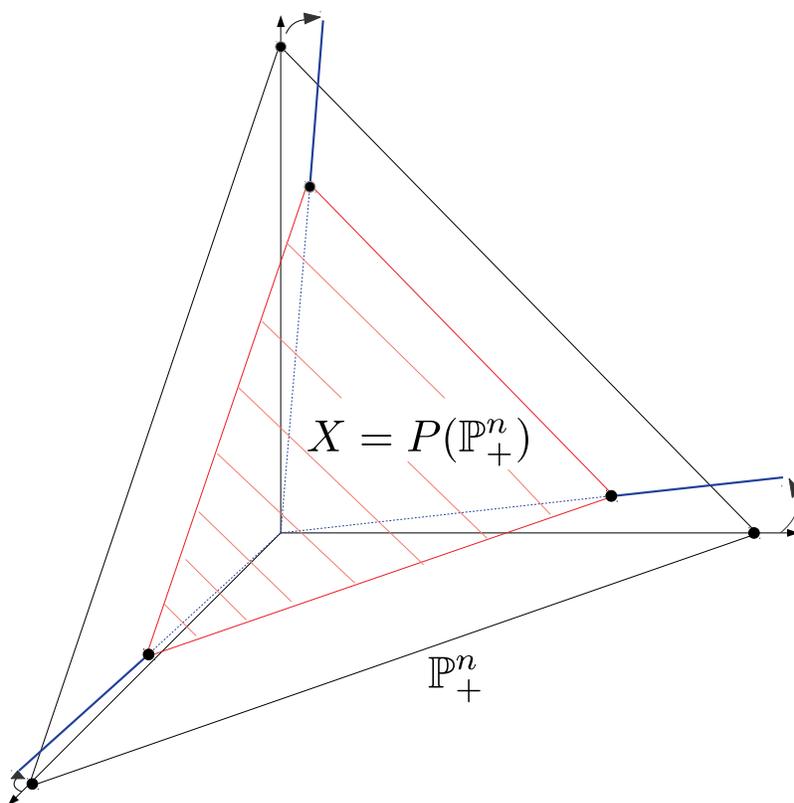


FIGURA 10. P contrai o cone positivo \mathbb{R}_+^n e podemos detectar isso no simplexo probabilístico.

¹²Esse é um bom exercício, caro leitor, todaa função contínua transforma conjuntos compactos em conjuntos compactos.

Vamos formalizar a observação descrita acima de que o cone \mathbb{R}_+^n é invariante pela ação de uma matriz estocástica. Se você olhar bem para provar isso basta olhar para o simplexo probabilístico, pois \mathbb{R}_+^n é uma união de semi-retas e cada uma delas cruza o simplexo probabilístico uma única vez. Por, isso vamos provar o seguinte.

Lema 5.4. *Se $P = [p_{ij}]$ é uma matriz estocástica então $X = P(\mathbb{P}_+^n) \subset \mathbb{P}_{++}^n$.*

Demonstração. Observe que \mathbb{P}_{++}^n é convexo: se $v^1, \dots, v^k \in \mathbb{P}_{++}^n$ e $\lambda_1, \dots, \lambda_k \in [0, 1]$ satisfazem $\sum_{l=1}^k \lambda_l = 1$ então

$$v = \sum_{l=1}^k \lambda_l v^l \in \mathbb{P}_{++}^n.$$

Com efeito, se $v^i = (x_1^i, \dots, x_n^i)$ então temos por hipótese que $\sum_{l=1}^n x_l^i = 1$, para todo $i = 1, \dots, k$. Agora observe que somar todas as entradas do vetor $v = \sum_{l=1}^k \lambda_l v^l$ equivale a somar todas as entradas da matriz $k \times n$ abaixo, cujas entradas somam 1 (veja a Figura 11).

$$\begin{array}{c} \lambda_1 v^1 \left[\begin{array}{ccc} \lambda_1 x_1^1 & \dots & \lambda_1 x_n^1 \end{array} \right] \cdot \lambda_1 \\ \vdots \\ \lambda_k v^k \left[\begin{array}{ccc} \lambda_k x_1^k & \dots & \lambda_k x_n^k \end{array} \right] \cdot \lambda_k \\ \hline = 1 \end{array}$$

FIGURA 11. Como $\sum_{l=1}^n \lambda_l x_l^i = \lambda_i$, cada linha da matriz soma λ_i , e como a soma dos λ_i dá 1, a soma de todas as entradas dá 1.

Isso prova que $v \in \mathbb{P}_{++}^n$ e portanto que \mathbb{P}_{++}^n é convexo. Agora note que os vetores $e_1 = (1, 0, \dots, 0)$, $e_2 = (0, 1, 0, \dots, 0), \dots, e_n = (0, \dots, 0, 1)$ da base canônica de \mathbb{R}^n são precisamente os vértices do simplexo probabilístico (veja Figura 10 mais uma vez). Além disso, as colunas de $P = [p_{ij}]$ são as imagens da base canônica: Como $0 < p_{ij} < 1$, para todo $i, j = 1, \dots, n$, e como a soma das entradas nas colunas

$$P = \begin{bmatrix} p_{11} & \dots & p_{1n} \\ \vdots & \dots & \vdots \\ p_{n1} & \dots & p_{nn} \end{bmatrix}$$

$\begin{array}{ccc} \uparrow & & \uparrow \\ P(e_1) & & P(e_n) \end{array}$

de P dá 1, ou seja $\sum_{i=1}^n p_{ij} = 1$, temos que $P(e_j) \in \mathbb{P}_{++}^n$. Em particular, toda combinação convexa dos vetores $\{P(e_j)\}_{j=1}^n$ pertence ainda a \mathbb{P}_{++}^n .

Logo, se $x = (x_1, \dots, x_n) \in \mathbb{P}_+^n$ então

$$\begin{aligned} P(x) &= \begin{bmatrix} p_{11} & \cdots & p_{1n} \\ \vdots & \cdots & \vdots \\ p_{n1} & \cdots & p_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = x_1 \begin{bmatrix} p_{11} \\ \vdots \\ p_{n1} \end{bmatrix} + \cdots + x_n \begin{bmatrix} p_{1n} \\ \vdots \\ p_{nn} \end{bmatrix} \\ &= x_1 P(e_1) + \cdots + x_n P(e_n) \in \mathbb{P}_{++}^n, \end{aligned}$$

e portanto $P(\mathbb{P}_+^n) \subset \mathbb{P}_{++}^n$, como queria demonstrar. \square

5.2.2. *Construção da métrica.* O Lema 5.4 mostra que de fato a Figura 10 é acurada: por uma iteração a matriz P de fato contrai o simplexo probabilístico. No entanto, a nossa intuição era de que aplicações sucessivas de P contraem cada vez mais. Para poder efetivamente “enxergar” essa contração precisamos de uma maneira de medir distâncias, uma métrica adequada. É tendo isso em mente que vamos introduzir os conceitos a seguir.

O conjunto $\mathbb{R}_{++}^n = \{(x_1, \dots, x_n) \in \mathbb{R}^n; x_i > 0, \forall i = 1, \dots, n\}$ é um *grupo multiplicativo*¹³ com a operação de multiplicação coordenada a coordenada:

$$xy = (x_1y_1, \dots, x_ny_n).$$

A identidade nesse conjunto é o ponto $(1, \dots, 1)$. E o inverso de (x_1, \dots, x_n) é $(x_1^{-1}, \dots, x_n^{-1})$.

Considere a função $\|\cdot\| : \mathbb{R}_{++}^n \rightarrow [1, +\infty)$ dada por

$$\|x\| = \max_{i,j} \left\{ \frac{x_i}{x_j} \right\} = \frac{\max x_i}{\min x_j}.$$

Por exemplo, no espaço \mathbb{R}^3 podemos calcular $\|(3, 3, 3)\| = 1$, $\|(1, 2, 3)\| = 3$ e assim por diante. Agora, preste bastante atenção na definição dessa função. Você está dividindo a maior entrada do vetor pela menor. Assim, em vetores com todas as entradas iguais vai dar sempre 1. Ou seja, essa função é pensada para “medir o quão distante um vetor está de ter todas as suas entradas iguais”.

Ela joga o papel de uma norma, uma maneira de medir comprimento de vetores. No entanto, as propriedades usuais de uma norma são aditivas ao passo que essa “falsa norma” que temos aqui tem propriedades multiplicativas. Vejamos:

Proposição 5.5. (1) $\|x\| = 1$ se, e somente se, $x = (c, \dots, c)$, para algum número real $c > 0$.

(2) $\|x\| = \|x^{-1}\|$

(3) $\|xy\| \leq \|x\|\|y\|$

(4) $\|cx\| = \|x\|$.

Demonstração. (1) Se $x = (c, \dots, c)$, para algum $c > 0$, então a menor entrada de x é c e a maior também. Portanto,

$$\|x\| = \frac{c}{c} = 1.$$

¹³Se expressão *grupo multiplicativo* pareceu bizarra para você provavelmente você ainda não foi apresentado ao maravilhoso mundo da *teoria de grupos*. Em matemática, muitas vezes a gente se esquece da forma específica que certos objetos possuem e se concentra apenas nas relações que os objetos guardam entre si. Isso permite, por exemplo, elucidar simetrias surpreendentes do espaço, assunto que é tratado por essa tal teoria de grupos. Quando falamos de um grupo estamos nos referindo a coleções de objetos que podem se combinar multiplicativa (ou aditivamente) mais ou menos como os números, satisfazendo mais ou menos as mesmas propriedades que os números satisfazem.

Por outro lado, se $\|x\| = 1$ então

$$\max x_i = \min x_j.$$

Ou seja, a menor e a maior entrada de x são iguais. Isso só pode acontecer se todas as entradas de x forem iguais, exatamente como prometido.

- (2) Para cada par $i, j \in \{1, \dots, n\}$,

$$\frac{x_j}{x_i} = \frac{1/x_i}{1/x_j},$$

logo

$$\max_{i,j} \frac{x_j}{x_i} = \max_{i,j} \frac{x_i^{-1}}{x_j^{-1}}.$$

- (3) Sejam $x = (x_1, \dots, x_n)$ e $y = (y_1, \dots, y_n)$ vetores em \mathbb{R}_+^n . Essa demonstração é baseada no fato simples de que o maior valor de uma lista finita de números é um dos números da lista, o maior deles! As vezes coisas profundas em matemática são consequências de fatos estúpidos assim.¹⁴

Por isso, existe um índice $k \in \{1, \dots, n\}$ tal que $\max x_i y_i = x_k y_k$. Contudo,

$$x_k y_k \leq (\max x_i) y_k \leq (\max x_i)(\max y_i).$$

Similarmente, vemos que $\min x_i y_i \geq (\min x_i)(\min y_i)$. Juntas, essas duas desigualdades provam que

$$\frac{\max x_i y_i}{\min x_i y_i} \leq \frac{(\max x_i)(\max y_i)}{(\min x_i)(\min y_i)},$$

que é o mesmo que dizer $\|xy\| \leq \|x\| \|y\|$.

- (4) Se $x = (x_1, \dots, x_n) \in \mathbb{R}_+^n$ e $c > 0$ então $cx = (cx_1, \dots, cx_n)$. Para cada i , temos $cx_i/cx_j = x_i/x_j$, logo

$$\max_{i,j} \left\{ \frac{cx_i}{cx_j} \right\} = \max_{i,j} \left\{ \frac{x_i}{x_j} \right\}. \quad \square$$

Uma norma mede comprimento de vetores, mas uma *métrica* deve medir distâncias entre pontos. Por isso, vamos considerar a função $\Delta : \mathbb{R}_+^n \times \mathbb{R}_+^n \rightarrow [1, \infty)$ dada por

$$\Delta(x, y) = \|xy^{-1}\|.$$

Mais explicitamente, se $x = (x_1, \dots, x_n)$ e $y = (y_1, \dots, y_n)$ são pontos de \mathbb{R}_+^n então

$$\Delta(x, y) = \frac{\max \frac{x_i}{y_i}}{\min \frac{x_j}{y_j}}.$$

Definição 5.6. A *pseudo-métrica projetiva de Hilbert* é a função $d : \mathbb{R}_+^n \times \mathbb{R}_+^n \rightarrow [0, \infty)$ definida por $d(x, y) = \log \Delta(x, y)$.

¹⁴E isso é uma das coisas mais difíceis de se aprender na formação de um matemático.

5.2.3. *Propriedades.* Vamos explicar um pouco esse nome tão *fancy*, ou suntuoso para usar um português a altura. Primeiro, d não é uma métrica¹⁵ pois uma métrica genuína dá zero somente se mede distância de um ponto a ele mesmo. Mas, como veremos em detalhes logo adiante, a função d mede o quanto dois pontos x e y estão longe de serem colineares. Ou seja, de certa forma d mede distâncias entre retas.

Por outro lado, olhe de novo a Figura 10. Perceba que cada reta que parte da origem em \mathbb{R}_+^n cruza o simplexo probabilístico uma única vez. O mesmo vale para $X = P(\mathbb{P}_+^n)$. Dessa forma, cada reta deixa um rastro único ali. Assim, uma métrica que mede distância entre retas mede distância entre pontos do simplexo probabilístico!

Vamos ver nesse parágrafo justamente que a *pseudo-métrica projetiva de Hilbert* é uma métrica se a restringirmos somente ao simplexo probabilístico.

Então, vamos verificar que d tem tudo o que é necessário para isso. Como d está em função de Δ , todas as suas propriedades são derivadas a partir daquelas de Δ .

Lema 5.7. *Sejam $x = (x_1, \dots, x_n)$ e $y = (y_1, \dots, y_n)$ pontos em \mathbb{R}_+^n . Então, valem as seguintes afirmações.*

- (1) $\Delta(x, y) = 1$ se, e somente se, existe $c > 0$ tal que $x = cy$.
- (2) $\Delta(x, y) = \Delta(y, x)$
- (3) $\Delta(x, z) \leq \Delta(x, y)\Delta(y, z)$.

Demonstração. (1) Pela Proposição 5.5 $\Delta(x, y) = \|xy^{-1}\| = 1$ se, e somente se, existe $c > 0$ tal que $xy^{-1} = (c, \dots, c)$. Ou seja, para todas as coordenadas de x e y teremos

$$\frac{x_l}{y_l} = c \iff x_l = cy_l,$$

o que equivale a $x = cy$, provando a afirmação.

- (2) Nessa demonstração vamos usar um fato simples mas muito útil: se a e b são números positivos e $a > b$ então $1/a < 1/b$. Isso tem como consequência o seguinte. Dada qualquer lista finita de números positivos a_1, \dots, a_n , se a_k é o maior deles e a_j é o menor deles então, com relação a lista dos inversos $1/a_1, \dots, 1/a_n$, o número $1/a_k$ é o menor e $1/a_j$ é o maior. Sejam $k, j \in \{1, \dots, n\}$ tais que $\max_l x_l/y_l = x_k/y_k$ e $\min_l x_l/y_l = x_j/y_j$. Em particular, pelo fato que acabamos de mencionar na frase anterior,

$$\max_l y_l/x_l = y_j/x_j \text{ e } \min_l y_l/x_l = y_k/x_k.$$

Então,

$$\begin{aligned} \Delta(x, y) &= \frac{\max_l x_l/y_l}{\min_l x_l/y_l} = \frac{x_k/y_k}{x_j/y_j} \\ &= \frac{x_k}{y_k} \times \frac{y_j}{x_j} = \frac{y_j/x_j}{y_k/x_k} \\ &= \frac{\max_l y_l/x_l}{\min_l y_l/x_l} = \Delta(y, x) \end{aligned}$$

- (3) Aqui usamos o item (3) da Proposição 5.5.

$$\begin{aligned} \Delta(x, z) &= \|xz^{-1}\| = \|xy^{-1}yz^{-1}\| \\ &\leq \|xy^{-1}\| \|yz^{-1}\| = \Delta(x, y)\Delta(y, z). \end{aligned}$$

¹⁵por isso é uma pseudo-métrica!

□

5.2.4. *A promoção da pseudo-métrica.* É comum em matemática nos depararmos com objetos estranhos, que não proporcionam o prazer imediato. Mas como tudo na vida, a essência das coisas está nos detalhes escondidos, que não saltam aos olhos no primeiro encontro. Quando percebemos isso, esses objetos esquisitos ganham uma nova reputação e as vezes podem até ser promovidos na empresa e receber um cargo melhor. É exatamente disso que trata o teorema abaixo.

Proposição 5.8. *A restrição da pseudo-métrica projetiva de Hilbert ao simplexo probabilístico estrito \mathbb{P}_{++}^n é uma métrica*

Essa proposição é a ferramenta que nos permite efetivamente medir a distância entre retas olhando apenas o rastro que elas deixam ao tocar \mathbb{P}_{++}^n . Por isso, ela é um elemento crucial na prova do Teorema de Perron-Frobenius.

Prova da Proposição 5.8. Queremos demonstrar que a função $d : \mathbb{P}_{++}^n \times \mathbb{P}_{++}^n \rightarrow [0, \infty)$ dada por $d(x, y) = \log \Delta(x, y)$ satisfaz todos os requisitos que uma métrica deve satisfazer. Vamos checar esses requisitos um por um. Primeiro, d não pode se importar com a ordem dos pontos. Ou seja, queremos que $d(x, y)$ seja igual a $d(y, x)$. Mas como $\Delta(x, y) = \Delta(y, x)$ (Lema 5.7) isso de fato é verdade. Em seguida, queremos a desigualdade triangular $d(x, z) \leq d(x, y) + d(y, z)$. Novamente pelo Lema 5.7, como

$$\Delta(x, z) \leq \Delta(x, y)\Delta(y, z)$$

e como o logaritmo transforma produtos em somas, segue que

$$\log \Delta(x, z) \leq \log \Delta(x, y) + \log \Delta(y, z),$$

ou seja

$$d(x, z) \leq d(x, y) + d(y, z).$$

Finalmente, precisamos verificar que $d(x, y) = 0$ se, e somente se, $x = y$. Primeiro, se $x = y$ então você pode verificar que $\Delta(x, y) = 1$, o que dá $d(x, y) = 0$.

Para provar a recíproca, vamos precisar argumentar um pouquinho mais. Com efeito, se $d(x, y) = 0$ isso implica, similarmente, que $\Delta(x, y) = 1$. Pelo Lema 5.7 isso diz que existe $c > 0$ tal que $x = cy$. No entanto, como $x, y \in \mathbb{P}_{++}^n$, se $x = (x_1, \dots, x_n)$ e $y = (y_1, \dots, y_n)$ então temos (pela definição de \mathbb{P}_{++}^n que

$$1 = \sum_{l=1}^n x_l = \sum_{l=1}^n y_l.$$

Logo, se $x = cy$ devemos ter $x_l = cy_l$ para todo $l = 1, \dots, n$ e portanto

$$1 = \sum_{l=1}^n x_l = c \sum_{l=1}^n y_l = c \cdot 1 = c,$$

o que prova que $x = y$. □

5.2.5. *Compacidade.* Certamente a propriedade mais amadas pelos matemáticos¹⁶ é a compacidade.

Ela é uma espécie de magia que faz os pontos em movimento num espaço sempre voltarem perto do ponto inicial, faz o vida ter recorrência; evita que tudo mude

¹⁶Ou vamos dizer assim: quanto mais perto você está da Análise, mais você ama compacidade.
Exercício: o quão longe um algebrista está da Análise?

eternamente e nunca vejamos de novo algo parecido com o que já vimos antes (essa é uma interpretação *dinamicista* da compacidade)¹⁷

Vamos ver nessa seção a compacidade entrando em cena na nossa jornada pela demonstração do Teorema de Perron-Frobenius.

Lema 5.9. *O simplexo probabilístico \mathbb{P}_+^n é compacto na métrica euclideana.*

Demonstração. Vamos verificar que o simplexo probabilístico é fechado e limitado com relação a métrica euclideana. Com efeito, ele é limitado porque se $x_i \in \mathbb{R}$ é tal que $0 \leq x_i \leq 1$ então $x_i^2 \leq 1$, o que implica imediatamente que se $x = (x_1, \dots, x_n) \in \mathbb{P}_+^n$ então

$$\|x\|^2 = \sum_{i=1}^n x_i^2 \leq 1,$$

e portanto $\|x\| \leq 1$. Isso mostra que $\mathbb{P}_+^n \subset \overline{B}_1(0)$, e portanto é limitado.

Por outro lado, \mathbb{P}_+^n é fechado porque as condições que o definem são condições fechadas. Com efeito, $x = (x_1, \dots, x_n) \in \mathbb{P}_+^n$ se, e só se, $0 \leq x_i \leq 1$ e $\sum_{i=1}^n x_i = 1$. Suponhamos então que $y_l = (y_1^l, \dots, y_n^l) \in \mathbb{P}_+^n$ convirja a um ponto $x \in \mathbb{R}^d$. Então $y_i^l \rightarrow x_i$ e e as condições seguintes

$$0 \leq y_i^l \leq 1 \text{ e } \sum_{i=1}^n y_i^l = 1$$

são verdadeiras, para todo $l \in \mathbb{N}$. Com isso, usando que a soma dos limites é o limite da soma, temos que

$$\sum_{i=1}^n x_i = \sum_{i=1}^n \lim_{l \rightarrow \infty} y_i^l = \lim_{l \rightarrow \infty} \sum_{i=1}^n y_i^l = 1.$$

Com um raciocínio parecido mostramos que

$$0 \leq x_i \leq 1, \forall i = 1, \dots, n,$$

e portanto $x \in \mathbb{P}_+^n$. Isso prova que o simplexo probabilístico é fechado e completa a demonstração do lema. \square

Como já aprendemos, uma função contínua transforma conjuntos compactos em conjuntos compactos. Como toda transformação linear é contínua, temos o seguinte

Corolário 5.10. *Se $P = [p_{ij}]_{n \times n}$ é uma matriz estocástica então $X = P(\mathbb{P}_+^n)$ é compacto na métrica euclideana.*

Nesse momento estamos entrando num detalhe delicado dessa história. Apresentamos o simplexo probabilístico. Ele é o palco onde vamos buscar o ponto fixo de uma matriz estocástica. Você já deve ter se ligado nessa altura que vamos precisar de compacidade, como quase sempre na lida diária de um matemático.

No entanto, se você lembrar bem da ideia para provar o Teorema de Perron-Frobenius, a gente pensava em medir distância usando a métrica projetiva de Hilbert, e não a métrica euclideana.

Ou seja, temos agora duas maneiras distintas de medir distâncias no simplexo probabilístico. A métrica euclideana que mede o comprimento do segmento de reta ligando os dois pontos (calculado através do Teorema de Pitágoras) e métrica

¹⁷Naturalmente, há quem goste dessa vida selvagem e sem raízes.

projetiva de Hilbert, que mede a distância entre as retas geradas pelos pontos e que “explode” quando uma das coordenadas se aproxima de zero.

Justamente por essa característica, \mathbb{P}_+^n não é compacto com respeito a métrica de Hilbert. No entanto, temos a

Proposição 5.11. *X é compacto na métrica projetiva de Hilbert.*

Antes de demonstrarmos essa proposição, vamos ver um lema que torna precisa uma imagem na Figura 10. Nela, quando aplicamos a matriz estocástica ao simplexo probabilístico obtemos um conjunto onde todas as entradas são positivas e além disso, todas as entradas são “um pouco” maiores do que zero. A versão precisa dessa frase é o seguinte

Lema 5.12. *Existe $\gamma > 0$ tal que todas as entradas dos pontos de X são maiores ou iguais a γ .*

Demonstração. Seja $P = [p_{ij}]$ a nossa matriz estocástica. Por definição, $X = P(\mathbb{P}_+^n)$ é o conjunto de todos os vetores da forma

$$P(x) = \begin{bmatrix} \sum_{j=1}^n p_{1j}x_j \\ \vdots \\ \sum_{j=1}^n p_{nj}x_j \end{bmatrix},$$

onde $x = (x_1, \dots, x_n) \in \mathbb{P}_+^n$, ou seja $0 \leq x_i \leq 1$ e $\sum_{i=1}^n x_i = 1$.

Por definição de matriz estocástica, todas as entradas de $P = [p_{ij}]_{n \times n}$ são estritamente positivas. Seja $\gamma = \min_{i,j} p_{ij}$ a menor entrada da matriz p . Ou seja, γ é uma das entradas de P e $p_{ij} \geq \gamma$, para todo $i, j = 1, \dots, n$.

Assim, se $P(x) \in X$ todas as suas entradas satisfazem

$$\sum_{j=1}^n p_{ij}x_j \geq \sum_{j=1}^n \gamma x_j = \gamma \sum_{j=1}^n x_j = \gamma.$$

Portanto, todas as entradas dos pontos de X são maiores ou iguais a γ . □

Para provar a Proposição 5.11, vamos mostrar que X é compacto usando a função identidade! Como X é compacto na métrica euclideana, se a função identidade

$$\text{Id} : (X, d_E) \rightarrow (X, d),$$

$\text{Id}(x) = x$, for contínua, teremos provado que X é compacto na métrica d de Hilbert, já que funções contínuas entre espaços métricos levam conjuntos compactos em conjuntos compactos.

Como falei acima, nesse ponto nossa demonstração entra num detalhe delicado: estamos olhando o mesmo conjunto X e colocando nele duas métricas diferentes, e a de Euclides e a de Hilbert. Então, como *espaços métricos*, temos dois espaços métricos diferentes. Nesse contexto, a função identidade se torna uma função entre dois espaços diferentes. Mas, se ela for contínua, então a sua imagem, o conjunto X , tem que ser compacto na métrica d .

Prova da Proposição 5.11. Passamos então a verificar a continuidade da função identidade. Fixe $x \in X$. Suponha que uma sequência de pontos $x_\ell \in X$, de entradas $(x_\ell^1, \dots, x_\ell^n)$ convirja a $x = (x_1, \dots, x_n)$ com respeito a métrica euclideana:

$$d_E(x_\ell, x) \rightarrow 0, \text{ quando } \ell \rightarrow +\infty.$$

Uma consequência disso é que as seqüências das entradas x_ℓ^i , com $i = 1, \dots, n$, dos pontos x_n , convergem as respectivas entradas x_i do ponto x . Em outras palavras,

$$(5.1) \quad \lim_{\ell \rightarrow +\infty} x_\ell^i = x_i, \quad \forall i = 1, \dots, n.$$

Nosso objetivo é demonstrar que isso implica que $d(x_n, x) \rightarrow 0$, onde d é a distância de Hilbert.

Ora, como $d(x, y) = \log \Delta(x, y)$ isso equivale a demonstrar que $\Delta(x_\ell, x) \rightarrow 1$. Como

$$\Delta(x_\ell, x) = \frac{\max x_\ell^i / x_i}{\min x_\ell^i / x_i}$$

se mostrarmos que

$$\lim_{\ell \rightarrow +\infty} \frac{x_i}{x_\ell^i} = 1, \quad \forall i = 1, \dots, n,$$

teremos resolvido o problema.

No entanto, a dificuldade de lidar com limites de frações é que se ambos, numerador e denominador tendem a zero, não controlamos o limite: temos uma indeterminação do tipo $\frac{0}{0}$. Observe que a indeterminação $\frac{\infty}{\infty}$ não nos preocupa pois, como $x_\ell, x \in X \subset \mathbb{P}_n^+$ todas as coordenadas envolvidas são menores do que 1.

E qual a chave para abrir nossos caminhos então? O Lema 5.12. Eles dizem que para os pontos de X as coordenadas nunca tendem a zero, ou ainda: $x_\ell \geq \gamma > 0$, para todo $\ell \in \mathbb{N}$.

Por isso, podemos dizer tranquilamente que $\Delta(x_\ell, x)$ converge para 1, o que prova que

$$d(x_n, x) \rightarrow 0,$$

como gostaríamos. □

Observação 5.13. O grande obstáculo à compacidade que temos é que a métrica de Hilbert não está definida na fronteira do simplexo probabilístico (pontos que tem alguma entrada nula). A aplicação $\text{Id} : (\mathbb{P}_{++}^n, d_E) \rightarrow (\mathbb{P}_{++}^n, d)$ está definida nessa fronteira, mas o espaço métrico (\mathbb{P}_+^n, d) não está definido (observe que \mathbb{P}_+^n é a união de \mathbb{P}_{++}^n com sua fronteira).

Exercício 5.14. O leitor familiarizado com argumentos típicos de análise pode pensar no seguinte problema: demonstre a Proposição 5.11 sem usar o Lema 5.12.

5.3. Demonstração do Teorema de Perron-Frobenius. Após a construção da métrica de Hilbert no subconjunto \mathbb{P}_{++}^n simplexo probabilístico, e da compacidade da imagem $X = P(\mathbb{P}_+^n)$ do simplexo probabilístico pela matriz P , temos (quase) todos os ingredientes para demonstrar o teorema.

De fato, *spoiler alert*: nossa estratégia será usar o Teorema de Edelstein (Teorema 4.2). Então, vamos querer provar que uma matriz estocástica reduz distâncias na métrica de Hilbert.

Talvez você esteja se perguntando, caro leitor, porque usar a métrica de Hilbert? Com efeito, a Figura 10 parece indicar que de fato uma matriz estocástica é uma contração para a métrica euclídeana. A dificuldade com essa ideia é que a métrica euclídeana não é “adaptada” ao contexto de matrizes estocásticas ao passo que a métrica de Hilbert funciona muito bem nesse cenário.

Exercício 5.15. Tente demonstrar que uma matriz estocástica é uma contração na métrica euclídeana.

5.3.1. *Um lema puramente numérico.* A chave para abrir nossos caminhos agora será o lema puramente numérico com demonstração puramente geométrica (e sem contas) abaixo.

Lema 5.16 (Lema Puramente Numérico). *Sejam $\xi^1, \dots, \xi^n > 0$ e $x, y \in \mathbb{P}_{++}^n$, com $x \neq y$. Então*

$$(5.2) \quad \min \left\{ \frac{x_i}{y_i} \right\} < \frac{\sum_{\ell=1}^n \xi^\ell x_\ell}{\sum_{\ell=1}^n \xi^\ell y_\ell} < \max \left\{ \frac{x_i}{y_i} \right\}$$

Demonstração puramente geométrica. Marque no plano cartesiano xy os pontos x_1, \dots, x_n no eixo horizontal e y_1, \dots, y_n no eixo vertical. Assim, cada número x_i/y_i representa a inclinação¹⁸ da reta r_i que liga a origem ao ponto (x_i, y_i) . Vamos usar índices i_1, \dots, i_n para escrever essas inclinações da menor para a maior¹⁹

$$\frac{x_{i_1}}{y_{i_1}} \leq \dots \leq \frac{x_{i_n}}{y_{i_n}}.$$

Como $x \neq y$ devem existir i_k, i_ℓ tais que

$$\frac{x_{i_k}}{y_{i_k}} > \frac{x_{i_\ell}}{y_{i_\ell}}.$$

Em particular, $x_{i_1}/y_{i_1} > x_{i_n}/y_{i_n}$.

Por outro lado, para cada $\ell \in \{1, \dots, n\}$ o ponto $(\xi^\ell x_\ell, \xi^\ell y_\ell)$ está sob a reta r_ℓ . No entanto, em geral, quando temos duas retas r, s no plano, com a inclinação de r maior do que aquela de s , a soma de um ponto em r com um ponto em s produz um vetor com inclinação entre a inclinação de r e a inclinação de s . Usando esse raciocínio repetidamente com as retas r_{i_ℓ} (veja Figura 5.3.1) vemos que o ponto $(\sum_{\ell=1}^n \xi^\ell x_\ell, \sum_{\ell=1}^n \xi^\ell y_\ell)$ pertence a uma reta cuja inclinação é um número dentro do intervalo

$$\left(\frac{x_{i_1}}{y_{i_1}}, \frac{x_{i_n}}{y_{i_n}} \right).$$

Isso demonstra a estimativa (5.2) e prova o lema. \square

5.3.2. *Prova do Teorema 5.1.* Seja $P = [p_{ij}]$ uma matriz estocástica. Vamos considerar (X, d) como espaço métrico compacto, onde X é a imagem por P do simplexo probabilístico e d é a métrica de Hilbert. Sejam $x, y \in X$, com $x \neq y$. Denote por (x_1, \dots, x_n) e (y_1, \dots, y_n) as coordenadas de x e y , respectivamente. Então,

$$P(x) = \left(\sum_{\ell=1}^n p_{1\ell} x_\ell, \dots, \sum_{\ell=1}^n p_{n\ell} x_\ell \right),$$

e, analogamente,

$$P(y) = \left(\sum_{\ell=1}^n p_{1\ell} y_\ell, \dots, \sum_{\ell=1}^n p_{n\ell} y_\ell \right).$$

¹⁸Esse número dá a cotangente do ângulo que a reta r_i faz com a horizontal, que é igual à tangente do ângulo que a reta faz com a vertical. Quanto maior esse número, mais afastada do eixo vertical a reta estará.

¹⁹Observe que nessa demonstração, para manter cada x_i no eixo horizontal e cada y_i no eixo vertical, optamos por medir a inclinação em relação ao eixo vertical, e não ao eixo horizontal como é de costume.

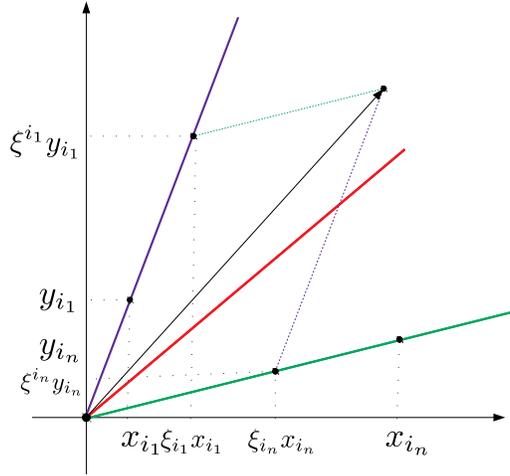


FIGURA 12. Cada ξ^ℓ determina um ponto em uma das retas. A soma de todos esses pontos dá um vetor com inclinação estritamente entre a reta de maior e a de menor inclinação.

Pela definição da métrica de Hilbert. Existem índices $k, j \in \{1, \dots, n\}$ tais que

$$\Delta(P(x), P(y)) = \frac{\sum_{\ell=1}^n p_{k\ell} x_\ell}{\sum_{\ell=1}^n p_{k\ell} y_\ell} \cdot \frac{\sum_{\ell=1}^n p_{j\ell} x_\ell}{\sum_{\ell=1}^n p_{j\ell} y_\ell}.$$

Aplicando o Lema 5.16 vemos que

$$\begin{aligned} \Delta(P(x), P(y)) &= \frac{\sum_{\ell=1}^k p_{k\ell} x_\ell}{\sum_{\ell=1}^n p_{k\ell} y_\ell} \cdot \frac{\sum_{\ell=1}^n p_{j\ell} x_\ell}{\sum_{\ell=1}^n p_{j\ell} y_\ell} \\ &< \max \frac{x_\ell}{y_\ell} \\ &= \Delta(x, y). \end{aligned}$$

Como $d(x, y) = \log \Delta(x, y)$, concluímos que a transformação $P : (X, d) \rightarrow (X, d)$, definida num espaço métrico compacto, é uma contração fraca:

$$x \neq y \in X \implies d(P(x), P(y)) < d(x, y).$$

O Teorema de Perron-Frobenius decorre agora do Teorema de Edelstein (4.2). \square

COMENTÁRIOS SOBRE A BIBLIOGRAFIA

O Teorema de Perron-Frobenius é um clássico e sua prova pode ser encontrada em muitos livros de álgebra linear ou de cadeias de Markov. Peter Lax, por exemplo, no seu excelente compêndio de álgebra linear [6] prova um resultado muito mais geral do que o que foi apresentado aqui e usa ideias de otimização na demonstração. A prova que eu apresentei aqui foi inspirada no artigo [3] que apresenta também uma versão mais geral e muito precisa do teorema. De fato, ele também usa ideias

de otimização para calcular o fator de contração exato de uma matriz estocástica na pseudo-métrica projetiva de Hilbert.

Para informações completas sobre a matemática da internet, o leitor pode consultar o livro [1].

No mais, se você tem críticas, dúvidas ou sugestões, não hesite em me escrever!

AGRADECIMENTOS

Agradeço ao Professor Ralph Teixeira (com quem tenho o privilégio de dividir uma sala na UFF) e seu aluno Matheus, pelas conversas que tivemos sobre o Teorema de Perron-Frobenius e que foram o pontapé inicial deste artigo. Agradeço aos meus alunos de iniciação científica Pedro Cavalcante e Cristiano Bastos pela leitura crítica e por comentários valiosos. Eu agradeço também ao bravo leitor que chegou até aqui.

REFERÊNCIAS

- [1] Anthony Bonato. *A course on the web graph*, volume 89. American Mathematical Soc., 2008.
- [2] Sergey Brin and Lawrence Page. The anatomy of a large-scale hypertextual web search engine. *Computer networks and ISDN systems*, 30(1-7):107–117, 1998.
- [3] Joseph E Carroll. Birkhoff's contraction coefficient. *Linear algebra and its applications*, 389:227–234, 2004.
- [4] Michael Edelstein. On fixed and periodic points under contractive mappings. *Journal of the London Mathematical Society*, 1(1):74–79, 1962.
- [5] Edward Kasner and James Newman. *Matemática e imaginação*. Zahar, 1968.
- [6] Peter D Lax. *Linear algebra. pure and applied mathematics*. 1996.
- [7] Elon Lages Lima. *Espaços métricos*, volume 4. Instituto de Matemática Pura e Aplicada, CNPq Rio de Janeiro, 1983.
- [8] Elon Lages Lima. *Algebra linear*, 2a. edição. IMPA, Rio de Janeiro, 1996.

INSTITUTO DE MATEMÁTICA E ESTATÍSTICA, UNIVERSIDADE FEDERAL FLUMINENSE
Current address: Rua Professor Marcos Waldemar de Freitas Reis, s/n, Bloco H - Campus do Gragoatá São Domingos- Niterói - RJ - Brazil CEP 24.210-201
E-mail address: brunosantiago@id.uff.br