

UFMG - Universidade Federal de Minas Gerais

**ESTRATÉGIAS DE DETERMINANTE ZERO DE  
PRESS-DYSON E ESTRATÉGIAS BOAS**

Mestranda: Zimar Rejane Mota Vieira Freitas

Orientador: Armando G. M. Neves

# Dissertação de Mestrado

## **ESTRATÉGIAS DE DETERMINANTE ZERO DE PRESS-DYSON E ESTRATÉGIAS BOAS**

Dissertação apresentada ao Programa de Pós-Graduação em Matemática do Instituto de Ciências Exatas da Universidade Federal de Minas Gerais como requisito parcial para a obtenção do grau de Mestre em Matemática.

Mestranda: Zimar Rejane Mota Vieira Freitas

Orientador: Armando G. M. Neves

Belo Horizonte  
Dezembro de 2017

# Resumo

Cooperação pressupõe algum tipo de doação, que pode não ser retribuída. Mesmo assim, diversos modelos matemáticos [[6], [10], [8]] mostram que a cooperação é possível como resultado de um processo biológico evolutivo. O Dilema do Prisioneiro (DP) é um paradigma capaz de representar situações de cooperação. Uma maneira para que a cooperação se torne possível é a repetição da interação. Por este motivo estudaremos uma extensão do DP simples denominada Dilema do Prisioneiro Infinitamente Repetido (DPIR). Nesse contexto, o objetivo deste trabalho é expor uma classe recentemente descoberta de estratégias de memória um para o DPIR, as estratégias de determinante zero de Press e Dyson. Iremos analisá-las do ponto de vista do trabalho de Akin [1]. Para fins de estudos do fenômeno da cooperação, o conceito de estratégias boas introduzido por Akin é bastante relevante. Veremos que existem estratégias boas que não são de determinante zero e estratégias de determinante zero que não são boas. Embora nesta dissertação não introduziremos nenhum tipo de dinâmica evolutiva, espera-se que ao final da leitura, o leitor tenha percebido que as estratégias boas tornam possível o surgimento da cooperação.

**Palavras-chave:** Teoria de Jogos, Cooperação, Dilema do Prisioneiro, estratégias de determinante zero, estratégias boas.

# Abstract

Cooperation assumes a sort of donation, which may not be reciprocated. Notwithstanding, several mathematical models [[6], [10], [8]] show that cooperation is possible as outcome of an evolutionary biological process. The Prisoner Dilemma (PD) is a paradigm able to represent several situations of cooperation and the repetition of an interaction is a way in which cooperation is possible. For this reason, we will study an extension of the simple PD called Infinitely Repetitive Prisoner Dilemma (IRPD). In this context, the aim of this work is to expose a recently discovered class of memory one strategies to the IRPD, the Press and Dysson's zero determinant strategies. We will analyze them from Akin's point of view [1]. For the purpose of cooperation phenomenon studies, the concept of good strategies introduced by Akin is quite relevant. We will see that there exist good strategies which are not zero determinant and zero determinant strategies that are not good. Although we will not introduce any type of evolutionary dynamics in this dissertation, after completing its reading the reader is expected to realize that the good strategies make the emergence of cooperation possible.

**Key-words:** Game Theory, Cooperation, Prisoner Dilemma, zero determinant strategies, good strategies.

# Sumário

<b>1</b>	<b>Ferramentas matemáticas</b>	<b>5</b>
1.1	Cadeias de Markov . . . . .	5
1.1.1	Argumento de convergência da cadeia de Markov . . . . .	12
1.1.2	Classificação de estados em uma cadeia de Markov com número finito de estados . . . . .	18
1.2	O Dilema do Prisioneiro infinitamente repetido (DPIR) . . . . .	21
1.3	Pagamentos médios por jogada no DPR . . . . .	25
<b>2</b>	<b>Estratégias de determinante zero de Press e Dyson</b>	<b>29</b>
2.1	Equação de Press-Dyson . . . . .	29
<b>3</b>	<b>As estratégias de determinante zero (DZ) como subconjunto de todas as estratégias de memória um</b>	<b>35</b>
3.1	Estratégias boas . . . . .	37
3.2	A base de Press-Dyson e as estratégias DZ . . . . .	41
3.3	Geometria do conjunto de estratégias de memória um na representação da base de Press-Dyson . . . . .	43
3.3.1	Estratégias excepcionais . . . . .	44
3.3.2	Estratégias não excepcionais . . . . .	45
3.4	Estratégias não excepcionais DZ . . . . .	50
3.4.1	Equalizadoras . . . . .	50
3.4.2	Extorsivas . . . . .	50
3.4.3	Generosas . . . . .	52
3.4.4	Estratégias vértice . . . . .	54
3.5	Estratégias DZ justas . . . . .	55
3.6	Competição entre jogadores que escolhem estratégias DZ . . . . .	57

# Introdução

Cooperação é um termo que descreve todo comportamento que gera um custo para um agente e um benefício para o recipiente da ação. Segundo Axelrod [2] a cooperação pode surgir até mesmo entre inimigos. Durante a Primeira Guerra Mundial, apesar de horríveis batalhas, o território onde acontecia a guerra foi cenário para o surgimento da cooperação. Em curto espaço de tempo, o melhor era atacar o inimigo para matar, independente se houve ou não ataque por parte do inimigo, mas como a convivência no campo de batalha foi por um longo período de tempo, houve mudança no cenário. Inicialmente, a estratégia de ataque era dominante, mas com o passar do tempo, as estratégias passaram a ser condicionais, ou seja, atacar caso fossem atacados. Os seres humanos cooperam em grande escala, dando origem a cidades, estados e países. Mas a cooperação é sempre vulnerável à exploração por traidores [6]. O comportamento humano está ligado tanto ao egoísmo quanto à cooperação.

Axelrod questiona em seu livro [2]: como desenvolver a cooperação em situações em que cada indivíduo tem um incentivo para ser egoísta? Seu objetivo foi desenvolver uma teoria da cooperação que pudesse ser usada para descobrir o que é necessário para emergir a cooperação. Para ele, o individualismo leva a um resultado ruim para todos. Uma poderosa ferramenta capaz de representar situações de cooperação é o famoso jogo do Dilema do Prisioneiro (DP).

O DP em sua forma tradicional descreve a história de dois indivíduos suspeitos por cometerem um crime e que são presos pela polícia, porém são mantidos em celas separadas. O promotor lhes oferece separadamente um acordo: se um dos prisioneiros delatar, ou seja, trair o outro, e o comparsa ficar calado, ele sairá livre enquanto o comparsa será condenado a 10 anos de prisão. Se ambos delatarem, os dois prisioneiros ficarão presos por 6 anos. Se ambos ficarem em silêncio, isto é, cooperarem um com o outro, a polícia só

poderá mantê-los por 1 ano na cadeia por falta de provas. Cada prisioneiro toma sua decisão sem saber da decisão do outro. Note que a traição é a melhor estratégia, independente do comportamento do outro jogador. Caso o outro seja traidor, traí-lo leva a 6 anos de prisão, enquanto uma cooperação levaria a 10 anos. Se o outro for um cooperador, traí-lo traz a liberdade imediata, enquanto cooperar com ele leva a um ano ainda na cadeia. Jogadores racionais e egoístas, ao jogarem o dilema do prisioneiro apenas uma vez, optarão pela traição, apesar de os jogadores se beneficiarem se ambos cooperarem.

Este simples jogo nos forneceu a base para as análises usadas neste trabalho. Posteriormente, mostraremos que se houver repetição da interação do dilema do prisioneiro, existem estratégias que de alguma forma estimulam o surgimento da cooperação. A mera repetição do jogo acrescenta uma grande quantidade de possíveis estratégias em vez de somente cooperar ou trair. Temos então, de acordo com o número de repetições do jogo, diversas modalidades possíveis para o Dilema do Prisioneiro Repetido (DPR).

Em 2012, Press e Dyson publicaram artigo [7] descrevendo o que chamaram estratégias de determinante zero para o Dilema do Prisioneiro Repetido. O assunto já era bastante explorado, mas as estratégias descobertas por Press e Dyson tinham propriedades surpreendentes, o que causou grande furor na comunidade de especialistas em Teoria de Jogos e sua aplicação em Biologia. Podemos destacar dentre as estratégias de Press e Dyson três classes notáveis: equalizadoras, extorsivas e generosas. Stewart e Plotkin [11] descrevem as estratégias de Press e Dyson de maneira simples e instrutiva e fornecem resultados sobre o sucesso dessas estratégias no contexto do torneio promovido por Axelrod nos anos 1970 [2].

No capítulo 1, trabalhamos com ferramentas matemáticas que são pré requisitos indispensáveis para se alcançar os resultados dos capítulos subsequentes. Introduzimos cadeias de Markov, que são processos estocásticos de tempo discreto em que um acontecimento atual pode afetar o resultado seguinte, mas os resultados futuros não sofrem influências de acontecimentos ocorridos há mais de uma unidade de tempo no passado. Falaremos da convergência de cadeias de Markov para distribuições estacionárias. Uma cadeia não convergente no sentido usual pode convergir no sentido de Cesaro. A convergência de uma cadeia no sentido de Cesaro aparece no Teorema 8 do capítulo 3 um dos principais resultados deste trabalho. Ainda no primeiro capítulo, introduzimos formalmente o dilema do prisioneiro, que visa descrever situações que envolvem cooperação. Definimos combinações lineares

convexas e mostramos que todos os pagamentos possíveis no DP para um par de jogadores estão contidos em um quadrilátero convexo.

No Capítulo 2, falamos das estratégias de determinante zero usando o formalismo original do artigo de Press e Dyson [7]. O objetivo foi mostrar os métodos e resultados para estratégias de Determinante Zero (DZ).

Dentre os diversos artigos publicados após [7] selecionamos [1] que, curiosamente, não foi publicado. Este artigo introduz um formalismo matemático mais completo que o de [7] e permite entender melhor as estratégias de determinante zero como um conjunto dentro do conjunto de todas as estratégias de memória um. No Capítulo 3 apresentamos alguns dos resultados de [1]. Em [1] o autor descreve as estratégias de memória um (ou Markovianas) para o dilema do prisioneiro repetido, enfatizando as estratégias que chama boas, que são uma generalização das estratégias generosas descobertas por Stewart e Plotkin em [11]. Nosso objetivo foi o de expor de maneira pedagógica alguns dos resultados desse artigo.

Concluimos que as estratégias DZ equalizadoras e extorsivas, apesar de permitirem que um jogador controle o resultado do jogo, não devem ter sucesso no contexto evolutivo [3]. No entanto, as estratégias DZ generosas são boas e têm potencial para serem bem sucedidas e dominarem a população, como outras estratégias generosas [8].



# Capítulo 1

## Ferramentas matemáticas

### 1.1 Cadeias de Markov

Uma *variável aleatória* é, grosso modo, uma variável quantitativa cujos valores dependem de fatores aleatórios. Um exemplo clássico de variável aleatória é o lançamento de um dado em que pode resultar qualquer número de 1 a 6. Seja  $\Omega$  o conjunto de todos os resultados possíveis de um experimento aleatório, que chamaremos de *espaço amostral*. Os subconjuntos do espaço amostral recebem o nome de *eventos*. Em teoria de probabilidades atribuímos a cada evento  $A$  em uma  $\sigma$ -álgebra de elementos de  $\Omega$  um número  $P(A)$  que chamaremos *probabilidade de  $A$* .

**Definição 1** *Seja  $\rho(\Omega)$  uma  $\sigma$ -álgebra de subconjuntos de  $\Omega$ . Uma função  $P : \rho(\Omega) \rightarrow \mathbb{R}$  será uma probabilidade se:*

(i)  $P(A) \geq 0, \quad \forall A \in \rho(\Omega)$

(ii)  $P(\emptyset) = 0$

(iii)  $P(\Omega) = 1$

(iv) *Se  $A_1, A_2, \dots \in \rho(\Omega)$  são eventos mutuamente disjuntos, isto é,  $A_i \cap A_j = \emptyset$  se  $i \neq j$ , então  $P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n)$ .*

**Definição 2** *Sejam  $A$  e  $B$  eventos em  $\rho(\Omega)$ . Se  $P(B) \neq 0$ , podemos definir a probabilidade condicional de  $A$  dado  $B$  por*

$$P(A|B) = \frac{P(A \cap B)}{P(B)}. \quad (1.1)$$

Julgamos suficiente esta breve passagem pela Teoria de Probabilidades para chegarmos aos objetivos que propomos neste trabalho. Para aprofundamento a respeito recomendamos, por exemplo, [9].

Um conjunto de variáveis aleatórias  $(X_t)_{t=0}^{\infty}$  representando a evolução de uma série de valores no decorrer do tempo  $t$  é denominado *processo estocástico* de tempo discreto. O matemático russo Andrei Andreyevich Markov iniciou um importante estudo dos processos em que um dado atual pode afetar o resultado no instante seguinte, mas dados de um passado mais remoto são irrelevantes. Uma cadeia de Markov é um caso particular de um processo estocástico e é formalmente definida como segue:

**Definição 3** *Uma sequência de variáveis aleatórias  $X_0, X_1, X_2, \dots$ , todas assumindo valores em um mesmo conjunto  $E$ , é chamada uma cadeia de Markov (CM) se, para qualquer  $t \in \mathbb{N}$  e quaisquer  $i_0, i_1, \dots, i_{t+1} \in E$ ,*

$$P(X_{t+1} = i_{t+1} | X_t = i_t, X_{t-1} = i_{t-1}, \dots, X_1 = i_1, X_0 = i_0) = P(X_1 = i_{t+1} | X_0 = i_t). \quad (1.2)$$

A identidade (1.2) define a *propriedade de Markov*. A propriedade de Markov diz que a função de distribuição de probabilidade no tempo  $t + 1$  depende do estado  $i_t$  no tempo  $t$  sem depender dos estados nos tempos anteriores  $0, 1, \dots, t - 1$  e é independente de  $t$ . Faz sentido portanto definir

$$m_{ij} = P(X_{t+1} = j | X_t = i) \quad (1.3)$$

como a probabilidade de que o sistema passe do estado  $i$  no tempo  $t$  para o estado  $j$  no tempo  $t + 1$ , sendo  $m_{ij}$  independente de  $t$ . Caso  $E$  seja finito, podemos chamar seus elementos de  $1, 2, \dots, s$  e definir a matriz de transição da cadeia de Markov como

$$M = \begin{bmatrix} m_{11} & m_{12} & \cdots & m_{1s} \\ m_{21} & m_{22} & \cdots & m_{2s} \\ \vdots & \vdots & \ddots & \vdots \\ m_{s1} & m_{s2} & \cdots & m_{ss} \end{bmatrix}. \quad (1.4)$$

A matriz de transição (1.4) pode também ser chamada de Matriz de Markov. **Observação:** Alguns autores usam uma definição mais geral de CM em que matriz de transição pode depender do tempo. Nesse caso, a definição adotada nesta dissertação refere-se ao que eles chamam cadeia de Markov homogênea.

**Definição 4** Uma matriz  $M$  é dita não-negativa se todos os elementos  $m_{ij}$  forem não-negativos.

Como os elementos de uma matriz de transição  $M$  de uma cadeia de Markov são probabilidades, então todos devem estar em  $[0,1]$  e, conseqüentemente toda matriz de transição de cadeia de Markov é uma matriz não-negativa. Em  $M$ , a entrada  $m_{12}$  representa a probabilidade de sair do estado 1 e ir para o estado 2,  $m_{11}$  é a probabilidade que o sistema vai continuar no estado 1 depois de ter estado em 1, e assim por diante.

Uma característica importante da matriz de transição associada a uma Cadeia de Markov é que a soma de cada linha deve ser igual a 1:

$$\sum_{j=1}^s P(X_{t+1} = j | X_t = i) = \sum_{j=1}^s m_{ij} = 1 .$$

Uma matriz de elementos não-negativos com essa propriedade é dita *estocástica por linhas*. Podemos chamar um vetor de componentes não-negativas com soma 1 de *vetor de probabilidade*.

**Proposição 1** Se  $M$  é uma matriz  $s \times s$  estocástica por linhas, então 1 é autovalor de  $M$ .

**Prova:** Seja  $\mathbf{1} \in \mathbb{R}^s$  o vetor com todas as componentes iguais a 1. Já que,  $M$  é estocástica por linhas, a soma de cada uma das linhas de  $M$  é igual a um. Daí,  $M\mathbf{1} = \mathbf{1}$ . Portanto, 1 é autovalor de  $M$ . ■

**Definição 5** O autovalor dominante é o maior autovalor em valor absoluto de uma matriz.

Além de mostrar que 1 é autovalor de uma matriz estocástica, necessitaremos de argumentar que 1 é o autovalor dominante. Antes de provarmos esse resultado demonstraremos um resultado geral e pouco conhecido, porém simples de se provar.

**Definição 6** Seja  $M$  uma matriz  $s \times s$  com elementos  $m_{ij} \in \mathbb{C}$ . O  $i$ -ésimo disco de Gershgorin de  $M$  é definido por

$$D_i = \{z \in \mathbb{C} : |z - m_{ii}| \leq \sum_{j=1, j \neq i}^s |m_{ij}|\} ,$$

$i = 1, 2, \dots, s$ .  $D_i$  é portanto o disco fechado no plano complexo centrado no ponto  $m_{ii}$  com raio  $R_i = \sum_{j \neq i} |m_{ij}|$ .

**Teorema 1** (Gershgorin). *Seja  $M_{s \times s}$  uma matriz complexa qualquer. Suponha que  $v = (v_1, \dots, v_s) \in \mathbb{C}^s$  seja autovetor de  $M$  com autovalor  $\lambda \in \mathbb{C}$  e que para um certo  $i$  tem-se  $|v_j| \leq |v_i|$ ,  $j = 1, 2, \dots, s$ . Então  $\lambda \in D_i$ . Em particular, todos os autovalores de  $M$  se encontram na união  $\bigcup_{i=1}^s D_i$  dos discos de Gershgorin de  $M$ .*

**Prova:** Seja  $\lambda$  autovalor de  $M$  e  $v$  um autovetor correspondente. Suponha que  $i$  seja a posição da maior componente de  $v$  em módulo. Assim, a  $i$ -ésima componente da equação  $Mv = \lambda v$  é  $\sum_{j=1}^s m_{ij}v_j = \lambda v_i$ . Separando o  $i$ -ésimo termo da soma, obtemos

$$\sum_{j \neq i} m_{ij}v_j = (\lambda - m_{ii})v_i .$$

Passando o módulo em ambos os lados

$$\left| \sum_{j \neq i} m_{ij}v_j \right| = |\lambda - m_{ii}||v_i| ,$$

aplicando a desigualdade triangular e o fato de que a  $i$ -ésima componente de  $v$  é a maior em módulo, temos

$$\left| \sum_{j \neq i} m_{ij}v_j \right| \leq \sum_{j \neq i} |m_{ij}||v_j| \leq \sum_{j \neq i} |m_{ij}||v_i| .$$

Assim

$$|\lambda - m_{ii}||v_i| \leq |v_i|R_i .$$

Cancelando  $|v_i|$  obtemos finalmente que  $\lambda \in D_i$ . Em geral não temos controle sobre qual das componentes de  $v$  será a maior, mas podemos garantir que  $\lambda \in \bigcup_{i=1}^s D_i$ . O mesmo argumento vale para qualquer dos autovalores e portanto todos os autovalores de  $M$  estão em  $\bigcup_{i=1}^s D_i$ . ■

**Corolário 1** *Se  $M$  é uma matriz  $s \times s$  estocástica por linha, e  $\lambda$  é autovalor, então  $|\lambda| \leq 1$ .*

**Prova:** Os discos de Gershgorin são centrados em  $m_{ii}$  e possuem raio  $R_i = \sum_{j \neq i} m_{ij}$ . Temos que  $m_{ii} \in [0, 1]$ , ou seja, os discos possuem centro no eixo real entre 0 e 1 e como  $m_{ii} + R_i = 1$ , todos possuem 1 em sua fronteira. O

fato de  $M$  ser estocástica por linhas nos fornece  $R_i \leq 1$ , observe Figura 1.1. Seja  $\lambda$  autovalor qualquer de  $M$ . Se  $\lambda \in D_i$ , temos

$$\lambda = \lambda - m_{ii} + m_{ii}$$

o que implica

$$|\lambda| \leq |\lambda - m_{ii}| + |m_{ii}| \leq \sum_{j \neq i} |m_{ij}| + |m_{ii}| = R_i + m_{ii} = 1 .$$

■

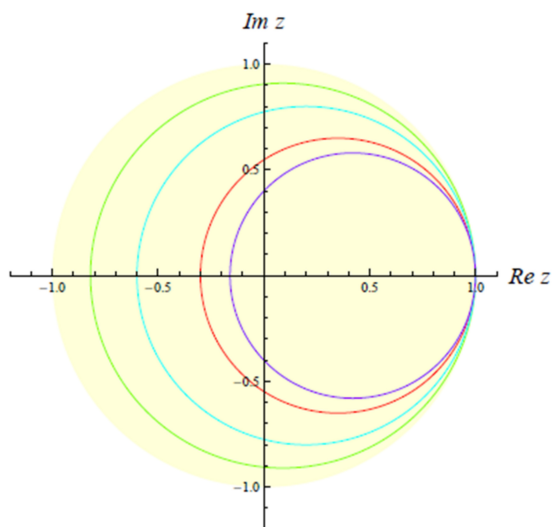


Figura 1.1: Ilustração dos discos de Gershgorin para demonstração do Corolário 1.

Pode-se associar a uma matriz de transição um grafo orientado segundo a definição seguinte:

**Definição 7** O grafo orientado de adjacência de uma matriz  $s \times s$  é o grafo orientado, ou digrafo, onde os vértices são os estados  $1, 2, \dots, s$  e existe uma aresta orientada partindo do estado  $i$  e chegando ao estado  $j$  se, e somente se,  $m_{ij} \neq 0$ .

**Definição 8** Um caminho em um grafo orientado partindo do estado  $i$  e chegando ao estado  $j$ , com possíveis arestas intermediárias, é um conjunto de arestas do grafo que saia do estado  $i$  e chegue até o estado  $j$ . O comprimento de um caminho que liga  $i$  a  $j$  é o número de arestas no caminho.

**Definição 9** Um grafo orientado é dito fortemente conexo, se para quaisquer dois vértices  $i$  e  $j$  do grafo existir um caminho que parta do estado  $i$  e chegue até  $j$ .

**Exemplo 1** O grafo orientado da Figura 1.2 é fortemente conexo. Por exemplo, o caminho partindo do vértice 1, chega ao vértice 3 passando pelos vértices 2 e 4. Note que a partir de qualquer estado, existe algum caminho para alcançar qualquer outro estado. Por outro lado, o grafo orientado da

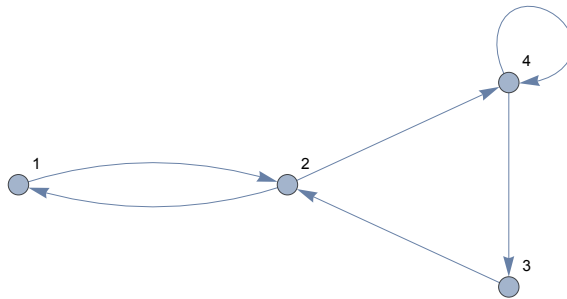


Figura 1.2: Exemplo de grafo orientado fortemente conexo.

Figura 1.3 não é fortemente conexo, já que não existe caminho que parta do estado 1 e chegue ao estado 4.

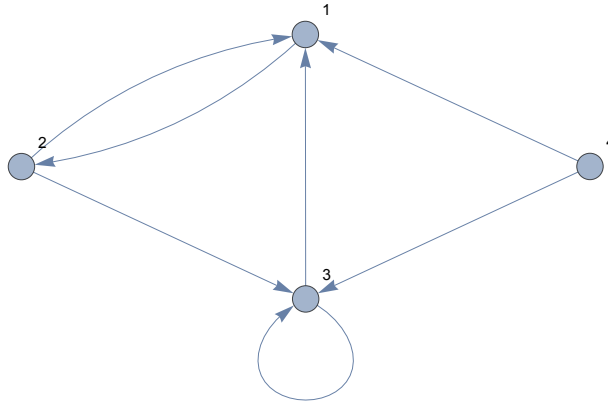


Figura 1.3: Exemplo de grafo orientado não fortemente conexo, o estado 4 não é acessível.

Adiante, veremos resultados importantes da teoria de Perron-Frobenius que dizem respeito a matrizes irredutíveis. Por simplicidade, faremos uma definição geométrica de irredutibilidade a partir dos grafos orientados de adjacência. Existe uma outra definição puramente algébrica e equivalente para o conceito, veja [4].

**Definição 10** *Uma matriz não-negativa é dita irredutível se o seu grafo orientado de adjacências for fortemente conexo. Se a matriz de transição de uma cadeia de Markov for irredutível, diremos também que a cadeia de Markov é irredutível.*

Esta definição diz que é possível passar de algum estado para qualquer outro por um caminho sobre o grafo de adjacências, não necessariamente em um passo. A Figura 1.2 ilustra o grafo orientado de adjacências para uma cadeia irredutível, enquanto a Figura 1.3 ilustra o mesmo para uma cadeia redutível.

O próximo teorema trata de autovalores de matrizes não-negativas. Quando aplicado às matrizes de transição de cadeias de Markov, o mesmo garante a existência e unicidade de distribuição de probabilidade invariante para cadeias de Markov irredutíveis. Aqui vamos nos ater somente ao enunciado. Para uma demonstração, veja, por exemplo, [4].

**Teorema 2 (Perron – Frobenius).** *Suponha que  $M$  seja uma matriz não-negativa irredutível. Então:*

- (1) O autovalor dominante  $r$  de  $M$  é estritamente positivo e é raiz simples do polinômio característico de  $M$ .
- (2) O autoespaço de  $M$  relacionado ao autovalor dominante possui dimensão 1.
- (3) Existe um autovetor  $v$  correspondente ao autovalor dominante  $r$  com todas as componentes estritamente positivas.

Apesar de o autoespaço relacionado ao autovalor dominante ser unidimensional, a hipótese de irredutibilidade de  $M$  ainda não é suficiente para excluir que há possibilidade de que  $M$  possua algum outro autovalor com módulo igual ao do autovalor dominante. Uma condição que garante a unicidade do autovalor dominante é a chamada primitividade.

**Definição 11** *Seja  $M_{s \times s}$  uma matriz não-negativa irredutível com autovalor dominante  $r$ . Suponha que  $M$  possua exatamente  $h$  autovalores de módulo  $r$ . O número  $h$  é chamado o índice de imprimitividade de  $M$ . Se  $h = 1$ , então a matriz  $M$  é dita primitiva, caso contrário, é imprimitiva.*

**Teorema 3** *O espectro de uma matriz irredutível de índice de imprimitividade  $h$  é invariante sob uma rotação do plano complexo de  $\frac{2\pi}{h}$ , mas não de uma rotação de ângulo positivo menor do que  $\frac{2\pi}{h}$ .*

A demonstração deste resultado e do seguinte pode ser encontrada em [4].

**Teorema 4** *Uma matriz irredutível que possui traço positivo é primitiva.*

### 1.1.1 Argumento de convergência da cadeia de Markov

Frequentemente as situações modeladas por cadeias de Markov atingem estados estacionários. No caso do DPIP, a convergência para uma distribuição estacionária será o instrumento que nos permite definir pagamentos médios para jogadores que repetem o jogo indefinidamente. No intuito de verificarmos o fenômeno da convergência, iremos predispor de mais algumas ferramentas matemáticas. O próximo resultado mostra a utilidade de se colocar as probabilidades de transição em forma de uma matriz.

**Proposição 2** *Seja  $(X_t)_{t=0}^{\infty}$  cadeia de Markov com espaços de estados  $E$  finito e matriz de transição  $M$ .*



- (i) Se  $w$  é vetor de probabilidade, então  $v = wM$  também é um vetor de probabilidade.
- (ii) Se a  $i$ -ésima componente  $w_i$  de  $w$  denota a probabilidade de que  $X_t = i$ ,  $i \in E$ , então  $\forall j \in E$  a probabilidade de que  $X_{t+1} = j$  vale  $(wM)_j$ .

**Prova:** (i)

$$v_j = \sum_{i=1}^s w_i m_{ij} .$$

Como os elementos  $m_{ij}$  e as componentes de  $w_i$  são todos não-negativos, claramente as componentes de  $v_j$  também o serão. Falta provar que a soma das componentes  $v_j$  também é igual a 1.

$$\begin{aligned} \sum_{j=1}^s v_j &= \sum_{j=1}^s \sum_{i=1}^s w_i m_{ij} \\ &= \sum_{i=1}^s w_i \sum_{j=1}^s m_{ij} \\ &= \sum_{i=1}^s w_i \\ &= 1 . \end{aligned}$$

Como  $M$  é estocástica por linhas e  $w$  é um vetor de probabilidade segue (i).

(ii) Note que

$$\{X_{t+1} = j\} = \bigcup_{i \in E} (\{X_{t+1} = j\} \cap \{X_t = i\}) ,$$

pois  $\bigcup_{i \in E} \{X_t = i\} = \Omega$ . Como os  $\{X_t = i\}$  são mutuamente disjuntos, ver Definição 1 (iii), usando a definição de probabilidade condicional em (1.1), temos

$$\begin{aligned} P(X_{t+1} = j) &= \sum_{i \in E} P(\{X_{t+1} = j\} \cap \{X_t = i\}) \\ &= \sum_{i \in E} P(X_t = i) P(X_{t+1} = j | X_t = i) \\ &= \sum_{i \in E} (w)_i m_{ij} = (wM)_j . \end{aligned}$$

Na penúltima para última linha usamos (1.3). ■

Dada uma cadeia de Markov no estado  $i$  no tempo  $t$ , queremos saber como determinar a probabilidade de que a cadeia de Markov estará no estado  $j$  após  $r$  unidades de tempo.

**Corolário 2** *Seja  $Q$  a matriz de transição de  $r$  passos de uma cadeia de Markov, isto é, a matriz cujos elementos são*

$$q_{ij} = P(X_{t+r} = j | X_t = i) = P(X_r = j | X_0 = i) .$$

*Então  $Q = M^r$ . Em outras palavras, a matriz de transição a  $r$  passos é a  $r$ -ésima potência da matriz de transição.*

Além dos resultados vistos até agora, devemos também observar que para um vetor de probabilidade  $v$  ser estacionário deverá acontecer que se a distribuição de probabilidade dos estados for dada no tempo  $t$  pelo vetor  $v$ , então nos tempos  $t + 1, t + 2, \dots$  a distribuição de probabilidade continuará a ser  $v$ . Portanto, pela Proposição 2 só teremos uma distribuição estacionária  $v$  caso  $v$  seja um autovetor *esquerdo* da matriz de transição  $M$  com autovalor 1, ou seja,

$$vM = v . \tag{1.5}$$

Apesar de ser pouco usual trabalhar com autovetores esquerdos, observe que  $v$  é autovetor esquerdo de  $M$  se, e somente se, for autovetor da transposta de  $M$ .

Para sumarizar toda discussão sobre a convergência de uma cadeia de Markov para uma distribuição estacionária, usaremos o próximo teorema.

**Teorema 5** *Considere uma cadeia de Markov com  $s$  estados e matriz de transição  $M$  irredutível e primitiva. Então, para qualquer vetor de probabilidade  $x_0 \in \mathbb{R}^s$ ,*

$$\lim_{r \rightarrow \infty} x_0 M^r = v ,$$

*onde  $v$  é a distribuição estacionária da cadeia de Markov. Em outras palavras, tomando qualquer vetor de probabilidade  $x_0$  para o estado inicial, a distribuição de probabilidade da cadeia convergirá para a distribuição estacionária.*

**Prova:** Vamos supor como simplificação adicional que a matriz  $M$  seja diagonalizável. Caso essa hipótese não se verifique, as mesmas ideias da prova

seguinte podem ser usadas trocando-se a base de  $\mathbb{C}^s$  constituída por autovetores de  $M$  por uma base construída a partir da forma canônica de Jordan da mesma matriz. Seja  $B = \{v, w_2, \dots, w_s\}$  uma base de  $\mathbb{C}^s$  composta por autovalores de  $M$ . Pelo enunciado do teorema, temos que  $v$  é a distribuição estacionária da cadeia de Markov. Pelo Teorema de Perron-Frobenius,  $v$  é o único elemento da base  $B$  associado ao autoespaço com autovalor 1. A hipótese da primitividade de  $M$ , veja Teorema 4, implica que os autovalores associados aos demais elementos de  $B$  possuem módulo estritamente menor que 1. Se  $w = x_0$  é o vetor de probabilidade que descreve a cadeia no instante inicial, então pela Proposição 2 e Corolário 2, temos que a distribuição de probabilidade dos estados da cadeia no tempo  $s$  é

$$x_s = x_0 M^s .$$

Escrevendo a condição inicial  $x_0$  na base  $B$ ,

$$x_0 = c_1 v + c_2 w_2 + \dots + c_s w_s ,$$

obtemos

$$x_s = c_1 v + c_2 w_2 \lambda_2^s + \dots + c_s w_s \lambda_s^s .$$

Como  $|\lambda_j| < 1$ ,  $j = 2, \dots, s$ , então  $x_s \rightarrow c_1 v$ , quando  $s \rightarrow \infty$ . Como  $x_0$  é vetor de probabilidade,  $x_s$  também o é, veja Proposição 2. Então só podemos ter  $c_1 = 1$ . ■

Com este argumento, podemos concluir que, independente da distribuição inicial  $w$ , cadeias de Markov com número finito de estados irredutíveis e primitivas sempre acabam convergindo para uma distribuição estacionária  $v$ , que é o autovetor esquerdo de  $M$ , com autovalor 1.

Para ilustrar o que afirmamos acima, vejamos um exemplo simples com dados fictícios.

**Exemplo 2** *Imagine uma situação em que há indivíduos que preferem uma comunicação através do uso de redes sociais, que denotaremos por produto 1, e outros que preferem ligações telefônicas, denotado por produto 2. Suponha que um indivíduo está consumindo o produto 1 e volte a consumir o produto 1 com uma probabilidade de 90%. Se o indivíduo consome o produto 2, a probabilidade de voltar a consumir o produto 2 é 80%. O grafo orientado que representa esta situação pode ser visto na figura 1.4. A matriz de transição é*

$$M = \begin{bmatrix} 0,9 & 0,1 \\ 0,2 & 0,8 \end{bmatrix} .$$

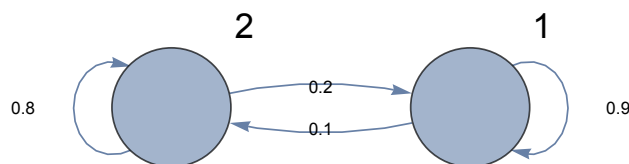


Figura 1.4: Grafo orientado que representa o consumo dos produtos 1 e 2.

Note que o grafo orientado é fortemente conexo (de qualquer estado consegue-se chegar ao outro). Para encontrar o estado estacionário, basta fazer  $v = vM$ . Assim,

$$[v_1 \ v_2] = [v_1 \ v_2] \begin{bmatrix} 0,9 & 0,1 \\ 0,2 & 0,8 \end{bmatrix},$$

obtendo o sistema

$$\begin{cases} v_1 = 0,9v_1 + 0,2v_2 \\ v_2 = 0,1v_1 + 0,8v_2 \end{cases}.$$

Resolvendo o sistema, encontramos  $v_1 = \frac{2}{3}$  e  $v_2 = \frac{1}{3}$ . Isto significa que o estado estacionário é  $v = [\frac{2}{3} \ \frac{1}{3}]$ . Em outras palavras, não importa o estado inicial, a chance de um indivíduo consumir o produto 1 é de aproximadamente 67% enquanto o consumo do produto 2 é cerca de 33%.

Este exemplo ilustrou o caso em que o grafo de adjacências de  $M$  é fortemente conexo, logo  $M$  é irredutível. Podemos concluir que  $M$  é convergente já que o limite  $v$  é a única distribuição estacionária e não importa o estado inicial. A irredutibilidade e a primitividade são a situação mais típica das cadeias de Markov, mas para desenvolver o restante dessa dissertação precisaremos entender o comportamento de cadeias de Markov em que uma ou ambas essas hipóteses falham.

Para exemplificar o que acontece no caso de cadeias de Markov irredutíveis, mas não primitivas, começamos por falar em convergência no sentido de Cesaro. Uma generalização do conceito de convergência de uma sequência é o de *convergência de Cesaro*.

**Definição 12** Seja  $(a_n)$  sequência infinita de elementos de  $\mathbb{R}^d$ ,  $m_n = \frac{a_1 + a_2 + \dots + a_n}{n}$  a média dos primeiros elementos de  $(a_n)$ . Dizemos que a sequência  $(a_n)$  é convergente no sentido de Cesaro para o limite  $m$  se  $\lim_{n \rightarrow \infty} m_n = m$ .

Se a sequência converge para um valor  $m$  no sentido usual, então pode-se provar facilmente que também converge no sentido de Cesaro para o mesmo valor  $m$ . A característica importante da convergência de Cesaro é que uma sequência divergente pode ter um limite de Cesaro bem definido.

**Exemplo 3** Seja  $a_n = \frac{1}{2}(1 - (-1)^{n+1})$  para  $n \geq 1$ . Obtemos  $a_1 = 0, a_2 = 1, a_3 = 0, a_4 = 1, \dots$ . Como

$$\frac{a_1 + a_2 + \dots + a_n}{n} = \begin{cases} \frac{1}{2}, & \text{se } n \text{ é par} \\ \frac{n+1}{2n}, & \text{se } n \text{ é ímpar} \end{cases},$$

daí

$$\lim_{n \rightarrow \infty} \frac{a_1 + a_2 + \dots + a_n}{n} = \frac{1}{2}.$$

Concluimos que a sequência  $(a_n)$ , embora divergente no sentido usual, converge para  $\frac{1}{2}$  no sentido de Cesaro.

Ilustraremos agora, através de um exemplo, o comportamento periódico de uma cadeia, que ocorre quando a matriz de transição é irredutível, mas imprimitiva. Neste caso não há uma convergência no sentido usual, mas a cadeia ainda converge no sentido de Cesaro para a distribuição estacionária.

**Exemplo 4** Seja  $n$  um número par e positivo. Considere  $n$  pedras em um lago colocadas sobre uma circunferência e numeradas em ordem crescente no sentido horário. A cada unidade de tempo um sapo pula com probabilidade  $p \in (0, 1)$  para a pedra mais próxima no sentido horário e  $1 - p$  para a pedra mais próxima no sentido anti-horário. Suponha que no instante 0 o sapo esteja na pedra 1, isto é,  $v_0 = (1, 0, 0, \dots, 0)$ . Qual a probabilidade de o sapo estar na posição  $j$  no instante  $k$ ?

A matriz de transição é

$$M = \begin{bmatrix} 0 & p & 0 & 0 & 0 & \dots & 0 & 0 & 1-p \\ 1-p & 0 & p & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 1-p & 0 & p & 0 & \dots & 0 & 0 & 0 \\ \vdots & \dots & \vdots & \dots & \vdots & \dots & \vdots & \dots & \vdots \\ \vdots & \dots & \vdots & \dots & \vdots & \dots & \vdots & \dots & \vdots \\ \vdots & \dots & \vdots & \dots & \vdots & \dots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \dots & 0 & p & 0 \\ 0 & 0 & 0 & 0 & 0 & \dots & 1-p & 0 & p \\ p & 0 & 0 & 0 & 0 & \dots & 0 & 1-p & 0 \end{bmatrix}.$$

Note que  $M$  é irredutível e que a distribuição invariante é  $v = (\frac{1}{n}, \dots, \frac{1}{n})$ .  $M$  possui autovalores 1 e -1 e outros autovalores de módulo menor que 1. Por possuir dois autovalores de módulo 1,  $M$  é imprimitiva com índice  $h = 2$ . Há duas distribuições de probabilidade  $v_1 = (0, \frac{2}{n}, \dots, 0, \frac{2}{n})$  e  $v_2 = (\frac{2}{n}, 0, \dots, \frac{2}{n}, 0)$  concentradas uma nas pedras pares, outra nas pedras ímpares, que são distribuições estacionárias da cadeia com matriz  $M^2$  e tais que  $v_2 = v_1 M$ ,  $v_1 = v_2 M$ . O estado inicial  $v_0$  converge no sentido usual para um ciclo de período 2 e no sentido de Cesaro para a distribuição invariante  $v = \frac{1}{2}(v_1 + v_2)$ .

Pode-se provar que se a cadeia é irredutível, mas imprimitiva de índice  $h$  haverá um fenômeno de periodicidade da cadeia de Markov com período  $h$ . Ou seja, existirão vetores de probabilidade  $v_1, v_2, \dots, v_h$ , tais que  $v_2 = v_1 M$ ,  $v_3 = v_2 M$ ,  $\dots$ ,  $v_h = v_{h-1} M$ ,  $v_1 = v_h M$ . A distribuição da cadeia convergirá no sentido de Cesaro para a distribuição estacionária  $v = \frac{1}{h}(v_1 + \dots + v_h)$ .

### 1.1.2 Classificação de estados em uma cadeia de Markov com número finito de estados

Os resultados a seguir esclarecem os outros fenômenos que podem ocorrer no limite de tempo infinito para qualquer cadeia de Markov com número finito de estados. Daqui em diante considere o conjunto de estados  $E = \{1, 2, \dots, s\}$  e uma cadeia de Markov com espaço de estados  $E$ . Seja  $G$  o grafo orientado de adjacências da matriz de transição da cadeia.

**Definição 13** Dados  $i, j \in E$  dizemos que  $j$  é alcançável a partir de  $i$ , denotado por,  $i \rightarrow j$ , se existe algum caminho em  $G$  que parta de  $i$  e chegue a  $j$ .

**Definição 14** Dados  $i, j \in E$ , dizemos que  $i$  e  $j$  são comunicantes se  $i \rightarrow j$  e  $j \rightarrow i$ . Denotamos a comunicação de  $i$  e  $j$  por  $i \leftrightarrow j$ .

Como pode acontecer que  $i \leftrightarrow i$ , então  $\leftrightarrow$  não é, em geral, uma relação de equivalência. Mas se definirmos  $E' \subset E$  como sendo o conjunto dos estados que comunicam consigo mesmos, então a relação  $\leftrightarrow$  restrita a  $E'$  passa a ser uma relação de equivalência.

Observe que  $E' \neq \emptyset$ . De fato, considere uma realização qualquer da cadeia de Markov. Sejam  $i_1, i_2, i_3, \dots, i_s$  os estados sucessivos ocupados pela realização da cadeia em  $s$  unidades de tempo. Como  $E$  é finito, algum estado terá que se repetir, ou seja, existe  $k \in E$  e instantes  $l, m$  tais que  $i_l = i_m = k$ . Logo  $k \in E'$ .

**Definição 15** (i) Um conjunto de estados não-vazio  $F \subset E$  é chamado um conjunto fechado se todo caminho começando em  $F$  nunca sai de  $F$ . Em outras palavras, se  $i \in F$  e  $j \notin F$ , então  $i \not\rightarrow j$ .

(ii) Um conjunto terminal é um conjunto  $T \subset E'$  que é uma classe de equivalência pela relação de comunicação e é fechado.

(iii) Um estado  $i \in E$  é recorrente se pertence a algum conjunto terminal.

(iv) Um estado  $i \in E$  é transiente se não pertence a nenhum conjunto terminal. Ou ainda, um estado é transiente se, e somente se, não é recorrente.

**Exemplo 5** Considere uma cadeia cujo grafo orientado de adjacências é o da Figura 1.5. Observe neste exemplo que  $5 \rightarrow 5$  e que  $\{1, 2, 3, 4, 5, 6\}$ ,

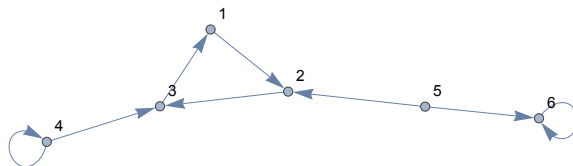


Figura 1.5: Grafo orientado do Exemplo 5.

$\{1, 2, 3, 6\}$ ,  $\{1, 2, 3\}$  e  $\{6\}$  são exemplos de conjuntos fechados. Temos que  $E' = \{1, 2, 3, 4, 6\}$  e as classes de comunicação são  $\{1, 2, 3\}$ ,  $\{4\}$  e  $\{6\}$ . Como  $\{4\}$  não é fechado, os conjuntos terminais desse exemplo são  $\{1, 2, 3\}$  e  $\{6\}$ . Os estados 1, 2, 3 e 6 são recorrentes e 4 e 5 são transientes.

**Teorema 6** (i) Conjuntos terminais distintos são disjuntos.

(ii) Qualquer conjunto fechado contém pelo menos um conjunto terminal.

(iii) De qualquer estado transiente existe um caminho para algum conjunto terminal.

**Prova:** (i) Pela definição de conjunto terminal como uma classe de equivalência, é claro que conjuntos terminais distintos são disjuntos.

(ii) Seja  $F$  conjunto fechado. Tomando um caminho de  $s$  passos em  $F$ , pela finitude de  $E$  é necessário que este caminho passe pelo menos duas vezes

em pelo menos um estado  $j_1$ . Seja  $C_1$  a classe de comunicação de  $j_1$ . Caso  $C_1$  seja fechado, então é terminal e acaba a demonstração. Caso  $C_1$  não seja fechado, então existe algum caminho começando em  $C_1$  que não retorna a  $C_1$ . Tomando tal caminho e requerendo que este tenha pelo menos  $s$  passos após sair de  $C_1$ , encontraremos outro estado  $j_2 \notin C_1$  em que o caminho passou pelo menos duas vezes. Defina  $C_2$  como a classe de  $j_2$ . Caso  $C_2$  não seja fechado, é só continuar o procedimento com  $j_3$  em  $C_3$ , etc. Como o número de estados é finito, em algum momento encontraremos um  $C_n$  que é terminal.

(iii) Dado  $i \in E$ , o conjunto  $A_i = \{j \in E, i \rightarrow j\}$  é claramente fechado. Se  $i$  é transiente, por (ii) existe um caminho que liga  $i$  a um conjunto terminal.

■

Da teoria de Perron-Frobenius, se  $M$  é irredutível, então todo o conjunto  $E$  é um único conjunto terminal e a distribuição estacionária única prevê probabilidade positiva para todos os estados.

Tudo que foi feito até agora nesta primeira seção teve o propósito de nos fazer entender a questão da convergência de cadeias de Markov com um número finito de estados. A cadeia pode convergir no sentido usual, independentemente do estado inicial, para a única distribuição estacionária, quando se tem a matriz de transição irredutível e primitiva. No caso em que a cadeia é irredutível e imprimitiva, isto é, existe mais de um autovalor de módulo 1, não há convergência usual, mas há convergência no sentido de Cesaro para a única distribuição estacionária.

Caso uma cadeia de Markov com um número finito de estados seja re-  
 dutível, não se pode aplicar diretamente o Teorema de Perron-Frobenius. Neste caso, usando que conjuntos terminais são fechados, podemos definir a restrição da cadeia somente aos estados recorrentes. Basta tomar uma distribuição de probabilidade inicial com probabilidade nula nos estados transientes. Retirando da matriz de transição as linhas e colunas correspondentes aos estados transientes, obtemos uma nova matriz de transição estocástica por linhas correspondente à cadeia restrita.

Uma das possibilidades é que a matriz de transição da cadeia restrita seja irredutível, podendo ser primitiva ou não. Neste caso, podemos aplicar à cadeia restrita a análise já feita para cadeias irredutíveis. Resulta que para a cadeia completa, sem restrição, existe um único conjunto terminal e que os estados fora desse conjunto terminal são transientes. Neste caso, pelo item (iii) do Teorema 6, para qualquer distribuição de probabilidade inicial teremos convergência (possivelmente só no sentido de Cesaro) para a única distribuição estacionária  $v$ . Essa distribuição será suportada sobre os estados



recorrentes, ou seja, o vetor  $v$  possuirá componente  $v_i > 0$  se e somente se o estado  $i$  for recorrente.

A outra possibilidade é que a matriz de transição para a cadeia restrita aos estados recorrentes ainda seja redutível. Neste caso haverá mais de um conjunto terminal para a cadeia restrita. Podemos então dividir a cadeia restrita em novas restrições suas a cada um dos conjuntos terminais. As restrições da restrição serão agora cadeias irredutíveis pois, pela definição de conjunto terminal como uma classe de comunicação fechada, a restrição a um conjunto terminal possui grafo orientado fortemente conexo. Para a cadeia completa, sem restrição, haverá uma distribuição estacionária para cada conjunto terminal, e essa será suportada sobre os estados desse conjunto terminal. Dada uma distribuição de probabilidade inicial, a cadeia convergirá no sentido de Cesaro para uma combinação linear das distribuições estacionárias. Diferentes realizações da cadeia convergirão em geral para distribuições estacionárias distintas.

Exemplos de cadeias de Markov redutíveis serão encontradas no Capítulo 3 quando falaremos sobre a competição entre jogadores que usam estratégias de determinante zero.

## 1.2 O Dilema do Prisioneiro infinitamente repetido (DPIR)

Uma situação comum em Teoria de Jogos é quando os indivíduos interagem dois a dois. Cada indivíduo tem à sua disposição  $n$  estratégias e recebe um pagamento por sua escolha. Tais pagamentos são dados por uma matriz de pagamento  $A$ ,  $n \times n$ , em que o elemento  $a_{ij}$  denota o pagamento de um indivíduo jogando a estratégia  $i$  contra um outro que joga a estratégia  $j$ . O Dilema do Prisioneiro é um paradigma para situações que envolvem cooperação e as estratégias possíveis são duas: cooperação (C) e traição (D). Considerando C como estratégia 1 e D como estratégia 2, a matriz de pagamento é

$$A = \begin{bmatrix} R & S \\ T & P \end{bmatrix}, \quad (1.6)$$

onde o DP é caracterizado pela ordem

$$T > R > P > S \quad (1.7)$$

e

$$P < \frac{S+T}{2} < R. \quad (1.8)$$

A condição  $\frac{S+T}{2} < R$ , tradicionalmente incluída como parte da definição do DP, significa que é mais vantajoso para um par de jogadores cooperarem sempre em lugar de alternarem entre cooperação e traição, caso em que receberiam pagamento médio  $\frac{1}{2}(T+S)$ . Sem essa condição não se consegue, no caso em que a interação é repetida (DPR), que a cooperação seja suficientemente favorecida de modo a permitir sua evolução. A desigualdade  $\frac{S+T}{2} > P$ , incluída também aqui como parte da definição do DP, nem sempre aparece. Em [8] ela apareceu também como necessária para a evolução da cooperação, pois um período de alternância entre cooperação e traição é frequentemente necessário antes que jogadores comecem a cooperar. A desigualdade implica então que nesse período os jogadores recebam recompensas maiores que o  $P$  da traição mútua. Na seção 1.3 veremos uma interpretação geométrica das desigualdades (1.8): elas garantem que o conjunto de todos os pagamentos para um par de jogadores esteja contido em um quadrilátero convexo. Caso alguma delas seja violada, a região do plano que contém todos os pares possíveis de pagamentos passa a ser um triângulo.

Se ambos os jogadores cooperam, cada um deles recebe o bom pagamento de cooperação mútua  $R$  (recompensa). Se ambos os jogadores traem, cada um deles recebe a punição por traição, o mau pagamento  $P$ . No entanto, se um jogador coopera e o outro não, então o traidor recebe o maior pagamento  $T$  (tentação), enquanto o cooperador obtém o menor pagamento,  $S$  (sucker's pay-off). Por causa da hierarquia (1.7), vemos que independentemente da estratégia do adversário, a traição é sempre a melhor escolha.

Ao considerarmos, por exemplo, o DP em sua forma tradicional onde as recompensas são dadas em anos de cadeia, se ambos os prisioneiros cooperam um com o outro, são condenados a 1 ano de prisão, enquanto se ambos traem, ficam presos por 6 anos, mas se um trai e o outro coopera, o traidor sai livre enquanto o cooperador fica condenado a 10 anos de prisão. A matriz de pagamento que representa esta situação obedece (1.7) e (1.8) com  $T = 0$ ,  $R = -1$ ,  $P = -6$  e  $S = -10$ . As recompensas são negativas, pois são anos de prisão.

**Definição 16** *Seja  $E$ ,  $n \times n$ , matriz de pagamento em um jogo entre dois*

jogadores e  $n$  estratégias.

$$E = \begin{bmatrix} e_{11} & \dots & e_{1j} & \dots & e_{1n} \\ \vdots & \dots & \vdots & \dots & \vdots \\ e_{j1} & \dots & e_{jj} & \dots & e_{jn} \\ \vdots & \dots & \vdots & \dots & \vdots \\ e_{n1} & \dots & e_{nj} & \dots & e_{nn} \end{bmatrix} .$$

A estratégia  $j$  é um equilíbrio de Nash (EN) se  $e_{jj} \geq e_{ij}$  para todo  $i \in 1, 2, \dots, n$ . Se a desigualdade acima for estrita, a estratégia  $j$  é dita um EN estrito.

Em outras palavras, quando ambos os jogadores jogam o equilíbrio de Nash, nenhum deles se arrepende de sua escolha caso o outro jogador mude de estratégia.

Voltando ao jogo DP, a estratégia D é um equilíbrio de Nash estrito, enquanto a estratégia C não é um equilíbrio de Nash. Uma consequência é que jogadores racionais irão sempre escolher a traição.

Para que a cooperação se torne possível, o jogo deve ser repetido. O cientista político Robert Axelrod [2] estudou uma extensão do DP simples denominada Dilema do Prisioneiro Repetido (DPR). Axelrod [2] optou por realizar sorteio, ao fim de cada rodada, com uma probabilidade  $w > 0$  de que o jogo continuasse por mais uma rodada. Desta forma, os jogadores não tinham conhecimento em que momento terminaria o jogo.

Para explorar o espaço das possíveis estratégias, Axelrod convidou especialistas em teoria de jogos para apresentar estratégias codificadas como programas de computador e realizou nos anos 1970 um campeonato do DPR [2], onde as estratégias automatizadas competiam umas com as outras. A estratégia que ganhou o torneio - dentre as submetidas - era a de “olho por olho, dente por dente” ou Tit-for-Tat (TFT). Um jogador TFT começa cooperando e em seguida faz o mesmo que o seu adversário fez na rodada anterior, ou seja, retalia com traição quando recebeu traição e coopera após receber cooperação. A estratégia TFT é um exemplo daquilo que definiremos como estratégia de memória um. Em particular, é uma estratégia reativa, que baseia sua resposta somente sobre o resultado da rodada imediatamente anterior. Estratégias reativas aparecem por exemplo em [10], que encontrou resultados interessantes em um experimento computacional para evolução de uma população de indivíduos adotando 100 estratégias reativas. Neste experimento, após um longo período de tempo, venceu a estratégia GTFT

(Generous Tit For Tat), uma estratégia leal que, ao receber traição, é capaz de perdoar o adversário. Sugerimos também [8], que estudou a evolução de uma população no Dilema do Prisioneiro Infinitamente Repetido (DPIR), provando rigorosamente que resultados semelhantes aos de [10] são obtidos quando a população é composta por indivíduos adotando as três estratégias mais importantes dentre as do experimento computacional.

Neste trabalho, a versão que estudaremos do DP é o Dilema do Prisioneiro Infinitamente Repetido (DPIR), caracterizado por  $w = 1$ , ou seja, os dois jogadores têm a certeza de encontrar-se novamente para outra rodada do jogo. Em cada rodada do DPIR o estado do par de jogadores X e Y pode ser um dos quatro seguintes: CC, CD, DC ou DD. A primeira das duas letras representa a jogada de um jogador X (cooperou ou traiu) e a segunda representa a jogada de um segundo jogador Y. Uma estratégia de memória um é descrita por um vetor  $p = (p_1, p_2, p_3, p_4)$  em que as componentes representam as probabilidades de cooperação do jogador para cada um dos quatro estados possíveis para a rodada anterior:

- $p_1$  é a probabilidade de o jogador jogar C quando ambos os jogadores cooperaram na rodada anterior (CC);
- $p_2$  é a probabilidade de o jogador jogar C quando cooperou na rodada anterior, mas o outro jogador o traiu (CD);
- Similarmente,  $p_3$  e  $p_4$  representam respectivamente as probabilidades de cooperação do jogador quando ele trai e o outro coopera e quando ambos traem.

Quando o estado é CC, cada um dos dois jogadores ganha a recompensa  $R$ . No estado CD, o jogador X recebe a menor recompensa  $S$ , enquanto Y recebe a grande recompensa  $T$ . Quando o estado é DC, X ganha  $T$  e Y ganha  $S$ . Quando ambos traem, o estado é DD e cada um recebe a punição por traição  $P$ . Podemos definir os vetores de recompensa para cada jogador por

$$S_X = (R, S, T, P) \quad \text{e} \quad S_Y = (R, T, S, P) . \quad (1.9)$$

$S_X$  e  $S_Y$  denotam portanto os pagamentos respectivos para os jogadores X e Y em cada um dos 4 estados para o conjunto de dois jogadores.

Suponha que X use a estratégia  $p = (p_1, p_2, p_3, p_4)$  e Y use a estratégia  $q = (q_1, q_2, q_3, q_4)$ . Os estados sucessivos deste jogo são determinados por probabilidades dependentes unicamente da rodada anterior, ou seja, seguem

uma cadeia de Markov com quatro estados, onde a matriz de transição é dada por

$$M = \begin{bmatrix} p_1q_1 & p_1(1 - q_1) & (1 - p_1)q_1 & (1 - p_1)(1 - q_1) \\ p_2q_3 & p_2(1 - q_3) & (1 - p_2)q_3 & (1 - p_2)(1 - q_3) \\ p_3q_2 & p_3(1 - q_2) & (1 - p_3)q_2 & (1 - p_3)(1 - q_2) \\ p_4q_4 & p_4(1 - q_4) & (1 - p_4)q_4 & (1 - p_4)(1 - q_4) \end{bmatrix}. \quad (1.10)$$

Para entendermos as entradas da matriz de transição  $M$ , seja  $m_{ij}$  entrada da matriz de transição, onde  $i, j = 1, \dots, 4$ , representam os estados (CC, CD, DC, DD). Para calcular, por exemplo, a entrada  $m_{23}$  que significa a probabilidade de passar do estado 2 para o estado 3, temos

$$\begin{aligned} m_{23} &= P(CD \rightarrow DC) \\ &= (1 - p_2)q_3. \end{aligned} \quad (1.11)$$

Para passar da primeira para a segunda igualdade, fizemos o produto da probabilidade de X escolher D com a probabilidade de Y escolher C. Sendo  $p_2$  a probabilidade de X escolher C se o estado anterior era CD assim  $1 - p_2$  é a probabilidade complementar para X, ou seja, representa a probabilidade de X escolher D já que a rodada anterior foi CD. De forma análoga  $q_3$  representa a probabilidade de Y escolher C, já que na rodada anterior o estado foi CD. Repare que o estado CD, é visto do ponto de vista de Y como se fosse DC, o estado 3.

Se dois jogadores adotam estratégias de memória um, acontece frequentemente, como visto na seção anterior, que existe uma única distribuição limite  $v$  no sentido de Cesaro da cadeia de Markov. Na seção seguinte veremos que tal distribuição define um pagamento médio por rodada para os jogadores.

### 1.3 Pagamentos médios por jogada no DPR

Suponha que X e Y sejam jogadores arbitrários que jogam uma sequência finita ou infinita de jogadas no DPR. Suponha que nessa sequência de jogadas o vetor  $v = (v_1, v_2, v_3, v_4)$  denote a distribuição de frequência com que o par de jogadores esteve nos estados 1, 2, 3 e 4. Portanto, para  $i = 1, 2, 3, 4$ ,  $v_i \in [0, 1]$  com  $\sum_{i=1}^4 v_i = 1$ . Os pagamentos médios  $s_X$  e  $s_Y$  dos jogadores nessa sequência de jogadas obedecem portanto

$$(s_X, s_Y) = (\langle v, S_X \rangle, \langle v, S_Y \rangle). \quad (1.12)$$

Caso X e Y adotem estratégias de memória um e  $M$  é irredutível,  $v$  se aproxima da distribuição limite no sentido de Cesaro da cadeia de Markov com matriz de transição (1.10). Este resultado motiva as seguintes definições:

**Definição 17** *Seja  $V$  um espaço vetorial e  $v_1, v_2, \dots, v_n \in V$ . Uma combinação linear convexa de  $v_1, v_2, \dots, v_n$  é uma combinação linear  $\alpha_1 v_1 + \alpha_2 v_2 + \dots + \alpha_n v_n$ , onde  $\alpha_i \in [0, 1]$  e  $\sum_{i=1}^n \alpha_i = 1$ .*

**Definição 18** *A envoltória convexa de um conjunto de pontos  $v_1, v_2, \dots, v_n$  em  $V$  é o conjunto de todas as combinações lineares convexas de  $v_1, v_2, \dots, v_n$ .*

Se  $V = \mathbb{R}^2$ , é possível ter uma visão intuitiva da envoltória convexa de um conjunto de  $n$  pontos como a região do plano delimitada por uma tira elástica esticada ao redor dos pontos. Vejamos alguns exemplos para objetos restritos ao plano:

- Se  $n = 2$ , a envoltória convexa é o segmento de reta que tem os dois pontos como extremos.
- Se  $n = 3$ , a envoltória convexa é o triângulo que tem os três pontos como vértices, ou, o maior dentre os segmentos de reta com extremos nos três pontos dados caso os pontos estejam alinhados.

**Definição 19** *Um conjunto  $C \subset V$  é convexo se para todo  $v_1, v_2 \in C$ , todas as combinações lineares convexas  $\alpha_1 v_1 + (1 - \alpha_1) v_2$  estão em  $C$ .*

A envoltória convexa de qualquer conjunto é sempre um conjunto convexo.

**Observações:**

- (i) Nem todo quadrilátero é convexo, embora todo triângulo o seja. Na Figura 1.6, os pontos  $v$  e  $w$  do quadrilátero  $ABCD$  são tais que as combinações lineares convexas suas não estão todas contidas no quadrilátero  $ABCD$ . Este quadrilátero, portanto, não é convexo.
- (ii) O ponto  $(s_X, s_Y)$  em  $\mathbb{R}^2$  de (1.12) é uma combinação linear convexa dos pontos  $(S, T)$ ,  $(T, S)$ ,  $(P, P)$  e  $(R, R)$ . De fato,

$$\begin{aligned} (s_X, s_Y) &= (\langle v, S_X \rangle, \langle v, S_Y \rangle) \\ &= v_1 (R, R) + v_2 (S, T) + v_3 (T, S) + v_4 (P, P). \end{aligned}$$

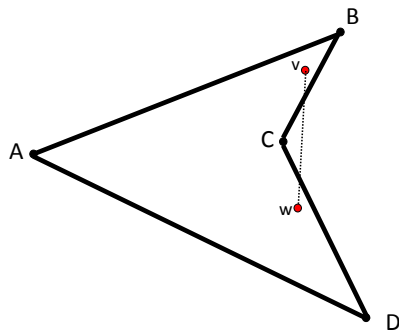


Figura 1.6: Exemplo de quadrilátero não convexo.

Assim, o conjunto de todos os pares de pagamentos médios possíveis no DPR está contido na envoltória convexa de  $(S, T)$ ,  $(T, S)$ ,  $(P, P)$  e  $(R, R)$ . O teorema a seguir mostra que essa envoltória convexa será um quadrilátero se, e somente se, as desigualdades (1.8) forem satisfeitas. Caso contrário, será um triângulo.

**Teorema 7** *Os pontos  $(S, T)$ ,  $(T, S)$ ,  $(P, P)$  e  $(R, R)$  são vértices de um quadrilátero convexo se, e somente se,  $P < \frac{S+T}{2} < R$ .*

Prova:

(a) Se  $P = \frac{S+T}{2}$ , os pontos  $(P, P)$ ,  $(S, T)$  e  $(T, S)$  estarão alinhados, de forma que os quatro pontos não são vértices de um quadrilátero. Se  $P > \frac{S+T}{2}$ , o segmento que liga  $(S, T)$  e  $(T, S)$  estará fora do quadrilátero com vértice nos quatro pontos. O quadrilátero será, portanto, não convexo.

(b) Análoga.

■

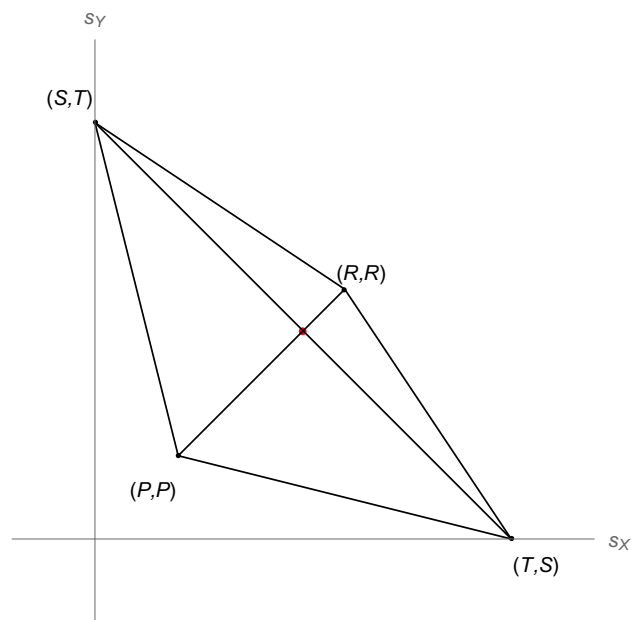


Figura 1.7: Quadrilátero do DPR. O ponto vermelho é o ponto médio do segmento que liga os vértices  $(S, T)$  e  $(T, S)$ .



## Capítulo 2

# Estratégias de determinante zero de Press e Dyson

Press e Dyson descobriram recentemente [7] uma nova classe de estratégias de memória um para o Dilema do Prisioneiro Infinitamente Repetido (DPIR), que chamaram estratégias de determinante zero (DZ). Tais estratégias serviram de inspiração para trabalhos de outros autores como, por exemplo, Stewart e Plotkin [11], Hilbe, Nowak e Sigmund [3] e Akin [1].

Mostraremos neste capítulo os métodos e resultados utilizados por Press e Dyson em seu artigo [7].

### 2.1 Equação de Press-Dyson

Suponha que X e Y sejam jogadores de memória um com estratégias dadas respectivamente por vetores  $p$  e  $q$ . Seja  $M$  a matriz de transição (1.10) da cadeia de Markov que descreve a mudança temporal do estado do par de jogadores. Suponha que  $M$  é irredutível.

A matriz  $N \equiv M - I$  é singular, já que  $M$  tem um autovalor unitário. Da irredutibilidade de  $M$  segue que o espaço nulo de  $N$  tem dimensão 1. Seja  $v$  a distribuição estacionária da cadeia de Markov. Observe que  $v$  satisfaz

$$v M = v \text{ ou } v N = 0. \quad (2.1)$$

Seja  $C_{ij} = (-1)^{i+j} \det(N_{ij})$  o cofator da entrada  $n_{ij}$  onde  $N_{ij}$  é o menor

da matriz  $N$  e  $i, j = 1, \dots, 4$ . Aplicando a extensão em cofatores  $N$ , obtemos

$$\det(N) = \sum_{i=1}^4 C_{ij} n_{ij} .$$

Estes cofatores não dependem da coluna  $j$  de  $N$ . Seja

$$L_j = (C_{1j} \ C_{2j} \ C_{3j} \ C_{4j}) .$$

Assim

$$\det(N) = L_j N_j ,$$

onde  $N_j$  é a  $j$ -ésima coluna de  $N$ ,  $j = 1, \dots, 4$ . Se  $j \neq k$ , então  $L_j N_k$  é o determinante da matriz obtida trocando a  $j$ -ésima coluna de  $N$  por  $N_k$ , sendo portanto zero. Logo

$$L_j N_k = \delta_{jk} \det(N) , \tag{2.2}$$

onde  $\delta_{jk}$  é o delta de Kronecker definido por

$$\delta_{jk} = \begin{cases} 0, & j \neq k \\ 1, & j = k \end{cases} .$$

Seja  $L = \text{adj}(N)$  a adjunta clássica de  $N$ , ou seja a transposta da matriz dos cofatores de  $N$ . A  $j$ -ésima linha de  $L$  é portanto formada pelos cofatores da  $j$ -ésima coluna de  $N$ . De (2.2) temos então

$$L N_k = \text{adj}(N) N_k = \det(N) e_k$$

com  $e_k$  representando a  $k$ -ésima coluna da matriz identidade,  $k = 1, \dots, 4$ . Juntando as colunas obtemos

$$\text{adj}(N) N = \det(N) I ,$$

que é uma forma da conhecida Regra de Cramer. Já que  $N$  é singular, isto implica

$$\text{adj}(N) N = 0 . \tag{2.3}$$

Já que o espaço nulo de  $N$  é unidimensional, então (2.3) implica que cada

linha de  $\text{adj}(N)$  é proporcional a  $v$ . Em particular, se  $l_4$  é a 4ª linha de  $\text{adj}(N)$ , temos que existe  $\alpha \in \mathbb{R}$  tal que

$$l_4 = \alpha v \quad (2.4)$$

ou, caso  $\alpha \neq 0$ ,

$$v = \frac{1}{\alpha} l_4. \quad (2.5)$$

Deve ser verdade que  $\alpha \neq 0$ , mas o artigo [7] não o prova. Embora tenhamos tentado um pouco, também não conseguimos uma prova desse fato. Por outro lado, o formalismo rigoroso do capítulo seguinte para reconstruir os resultados de [7] dispensa tal prova. Em [7] a equação acima nem mesmo aparece explicitamente. Aparentemente os autores tomaram  $\alpha = 1$  e consertaram o erro mais adiante.

Seja  $N_{i_4}$  a matriz  $3 \times 3$  obtida de  $N$  retirando a 4ª coluna e a  $i$ -ésima linha. Sabemos que somar múltiplos de outras colunas a uma coluna de uma matriz não altera seu determinante. Somando a primeira coluna de  $N_{14}$  à segunda e à terceira colunas da mesma, temos

$$\det(N_{14}) = \begin{vmatrix} p_2 q_3 & p_2 - 1 & q_3 \\ p_3 q_2 & p_3 & q_2 - 1 \\ p_4 q_4 & p_4 & q_4 \end{vmatrix}. \quad (2.6)$$

Analogamente, somando a primeira coluna à segunda e à terceira de  $N_{24}$ , obtemos

$$\det(N_{24}) = \begin{vmatrix} p_1 q_1 - 1 & p_1 - 1 & q_1 - 1 \\ p_3 q_2 & p_3 & q_2 - 1 \\ p_4 q_4 & p_4 & q_4 \end{vmatrix}. \quad (2.7)$$

Finalmente, se  $f = (f_1, f_2, f_3, f_4)$  é um vetor arbitrário, definimos

$$N(p, q, f) = \begin{bmatrix} p_1 q_1 - 1 & p_1 - 1 & q_1 - 1 & f_1 \\ p_2 q_3 & p_2 - 1 & q_3 & f_2 \\ p_3 q_2 & p_3 & q_2 - 1 & f_3 \\ p_4 q_4 & \underbrace{p_4}_{\tilde{p}} & \underbrace{q_4}_{\tilde{q}} & f_4 \end{bmatrix}. \quad (2.8)$$

Note que a segunda coluna da matriz no lado direito de (2.8), designada  $\tilde{p} = (p_1 - 1, p_2 - 1, p_3, p_4)$ , depende apenas do vetor estratégia  $p$  de  $X$ . Podemos dizer que  $\tilde{p}$  está sob controle exclusivo de  $X$ . O mesmo ocorre com a terceira coluna de (2.8) nomeada  $\tilde{q} = (q_1 - 1, q_3, q_2 - 1, q_4)$ , que está sob controle exclusivo de  $Y$ .

Observe que, pelas expressões anteriores para os determinantes das submatrizes  $N_{i4}$ , então, se definirmos  $D(p, q, f) = \det[N(p, q, f)]$ , usando a expansão em cofatores da 4ª coluna, este pode ser escrito como

$$\begin{aligned} D(p, q, f) &= \sum_{i=1}^4 f_i (-1)^{i+4} \det(N_{i4}) \\ &= \langle f, l_4 \rangle, \end{aligned}$$

onde usamos que o vetor dos cofatores da 4ª coluna de  $N(p, q, f)$  coincide com a 4ª linha  $l_4$  de  $\text{adj}(N)$ . Usando (2.5) obtemos

$$\begin{aligned} D(p, q, f) &= \langle f, \alpha v \rangle \\ &= \alpha \langle f, v \rangle \end{aligned}$$

Se  $\mathbf{1}$  designa o vetor com todas as componentes iguais a 1, então, usando a fórmula acima e a normalização de  $v$ , temos que

$$\alpha = D(p, q, \mathbf{1}).$$

Assim

$$\langle f, v \rangle = \frac{D(p, q, f)}{D(p, q, \mathbf{1})}. \quad (2.9)$$

A fórmula (2.9) calcula o produto escalar de um vetor arbitrário  $f$  com o vetor estacionário  $v$  da matriz de transição (1.10) como uma razão de determinantes em que a segunda e terceira colunas dependem respectivamente somente das estratégias de  $X$  e  $Y$ . Esta fórmula foi obtida originalmente em

[7]. Sabendo que a matriz de recompensa de X é  $S_X = (R, S, T, P)$ , enquanto de Y é  $S_Y = (R, T, S, P)$ , no estado estacionário seus respectivos pagamentos são

$$s_X = \langle S_X, v \rangle = \frac{D(p, q, S_X)}{D(p, q, \mathbf{1})} \quad (2.10)$$

$$s_Y = \langle S_Y, v \rangle = \frac{D(p, q, S_Y)}{D(p, q, \mathbf{1})} . \quad (2.11)$$

Por propriedades elementares de determinantes,

$$\alpha s_X + \beta s_Y + \gamma = \frac{D(p, q, \alpha S_X + \beta S_Y + \gamma \mathbf{1})}{D(p, q, \mathbf{1})} . \quad (2.12)$$

A equação (2.12) mostra que se X escolher uma estratégia que satisfaça  $\tilde{p} = \alpha S_X + \beta S_Y + \gamma \mathbf{1}$ , ou se Y escolher uma estratégia tal que  $\tilde{q} = \alpha S_X + \beta S_Y + \gamma \mathbf{1}$ , então o determinante no numerador se anula já que colunas iguais resultam em determinante nulo. Assim, será satisfeita uma relação polinomial de grau 1 entre os dois pagamentos

$$\alpha s_X + \beta s_Y + \gamma = 0 . \quad (2.13)$$

Press e Dyson [7] mostraram que de fato existem estratégias que anulam o determinante no numerador de (2.12) e nomearam-nas estratégias de *determinante zero* (DZ). No Capítulo 3 mostraremos a existência de uma infinidade de estratégias de determinante zero, entenderemos a relação entre elas e as demais estratégias de memória um, assim como estudaremos algumas classes especiais com interessantes propriedades, já conhecidas por Press e Dyson, dentre as estratégias de determinante zero.

Na Figura 2.1, extraída de [3], todos os quadros ilustram o pagamento dos jogadores X e Y se ambos usam estratégias memória um no DPR (com  $T = 3$ ,  $R = 2$ ,  $P = 0$  e  $S = -1$ ). Em cada um dos quadros, a estratégia do jogador X foi fixada. O quadro A representa os pares de pagamento dos dois jogadores quando o jogador X adota  $p = (1, 0, 0, 1)$ , estratégia conhecida na literatura [6] por Win-Stay-Lose-Shift (ou WSLS). Deste quadro, vê-se que a estratégia WSLS não é DZ, já que claramente os pontos não estão todos sobre uma reta, como em (2.13). Nos gráficos B, C e D o jogador X adota estratégias DZ e, conseqüentemente, os pontos que representam os pagamentos do par de jogadores estão todos alinhados. A estratégia adotada por X em C está na classe das chamadas estratégias equalizadoras. Em D a

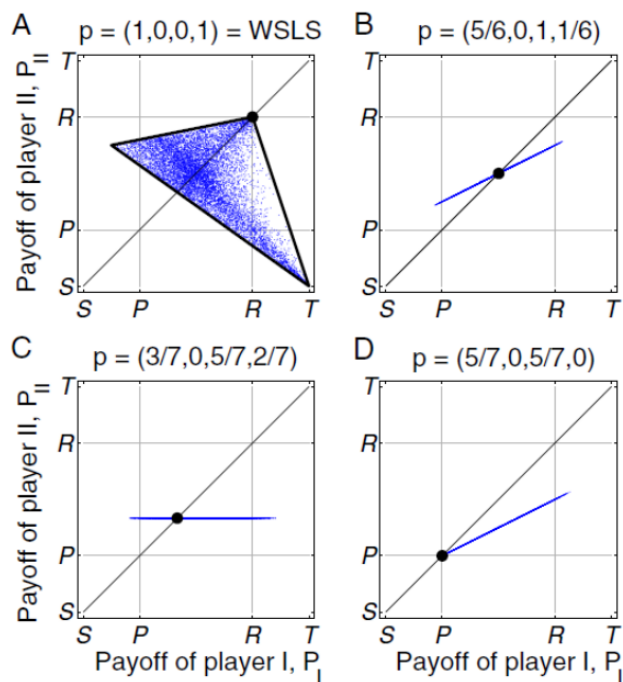


Figura 2.1: Figura extraída de [3] que representa os pares de pagamentos de dois jogadores que usaram estratégias de memória um no jogo DPR.

estratégia de X está na classe das estratégias de extorsão. Essas duas classes, já individualizadas no artigo do Press e Dyson, merecerão uma certa atenção no Capítulo 3.

A descoberta das estratégias DZ, inspirou respostas de autores renomados como Akin [1], Stewart e Plotkin [11], e também Hilbe, Nowak e Sigmund [3] que produziram um texto sobre a evolução da extorsão no DPR.

## Capítulo 3

# As estratégias de determinante zero (DZ) como subconjunto de todas as estratégias de memória um

A surpreendente descoberta em 2012 das estratégias de determinante zero [7] despertou ideias importantes que estudaremos neste capítulo. Como vimos no capítulo anterior, Press e Dyson [7] usaram na descoberta dessas estratégias um formalismo arbitrário baseado nas propriedades de determinantes e na regra de Cramer. Mostraremos outros resultados interessantes e mais gerais para descrever as estratégias DZ, apresentados por Akin em [1].

Suponha que um par de jogadores X e Y interaja através do DPIR. Suponhamos que X adota uma estratégia de memória um com vetor  $p$ . Na maior parte dos casos, embora faremos algumas exceções pontuais, suponhamos que Y também adota uma estratégia de memória um com vetor  $q$ . Neste caso, a evolução temporal dos estados do par de jogadores é descrita por uma cadeia de Markov cuja matriz de transição é dada por (1.10).

Conforme vimos na Seção 1.1, haverá sempre pelo menos um vetor distribuição de probabilidade  $v$  estacionário para a cadeia. Na maior parte dos casos, a cadeia é irredutível,  $v$  é único e a distribuição da cadeia converge para  $v$  no sentido de Cesaro, independente da condição inicial. Caso a cadeia seja redutível, a distribuição de probabilidade limite pode não ser única e a distribuição limite poderá depender do estado inicial.

Em qualquer dos casos, seja  $v$  único ou não, podemos sempre afirmar

que se  $X$  e  $Y$  adotam estratégias de memória um, haverá uma distribuição estacionária  $v$  e, como visto na Seção 1.3, os pagamentos médios por jogada  $s_X$  e  $s_Y$  estarão bem definidos.

Iremos supor que os elementos  $T$ ,  $R$ ,  $P$  e  $S$  da matriz obedecem à hierarquia (1.7) e às desigualdades (1.8).

**Proposição 3** *Seja  $v$  a distribuição de probabilidade limite para a evolução temporal dos estados do par de jogadores  $X$  e  $Y$ . Então os pagamentos médios para o par de jogadores no DPIR satisfazem*

$$s_Y - s_X = (v_2 - v_3)(T - S) .$$

*Em particular,  $s_Y = s_X$  se e somente se  $v_2 = v_3$ . Além disso:*

(a) *Se  $\frac{S+T}{2} < R$ , então*

$$s_Y + s_X \leq 2R , \tag{3.1}$$

*com igualdade se, e somente se,  $v = (1, 0, 0, 0)$  se, e somente se,  $s_Y = s_X = R$ .*

(b) *Se  $P < \frac{S+T}{2}$ , então*

$$s_Y + s_X \geq 2P , \tag{3.2}$$

*com igualdade se, e somente se,  $v = (0, 0, 0, 1)$  se, e somente se,  $s_Y = s_X = P$ .*

**Prova:** Se  $v = (v_1, v_2, v_3, v_4)$ , de (1.12) temos que

$$\begin{aligned} s_Y - s_X &= \langle v, S_Y \rangle - \langle v, S_X \rangle \\ &= \langle (v_1, v_2, v_3, v_4), (R, T, S, P) \rangle - \langle (v_1, v_2, v_3, v_4), (R, S, T, P) \rangle \\ &= (v_2 - v_3)(T - S) . \end{aligned}$$

Portanto  $s_Y = s_X$  se e somente se  $v_2 = v_3$ . Além disso,

$$\begin{aligned} \frac{1}{2}(s_Y + s_X) &= \frac{1}{2}(\langle v, S_Y \rangle + \langle v, S_X \rangle) \\ &= \frac{1}{2}[\langle (v_1, v_2, v_3, v_4), (R, T, S, P) \rangle + \langle (v_1, v_2, v_3, v_4), (R, S, T, P) \rangle] \\ &= \frac{1}{2}(v_1R + v_2T + v_3S + v_4P + v_1R + v_2S + v_3T + v_4P) \\ &= \frac{1}{2}[2v_1R + (v_2 + v_3)(T + S) + 2v_4P] . \end{aligned}$$



Caso  $(S + T) < 2R$ , usando (1.7),

$$\frac{1}{2}(s_Y + s_X) \leq R(v_1 + v_2 + v_3 + v_4) = R.$$

Vale a igualdade  $\iff v = (1, 0, 0, 0) \iff s_Y = s_X = R$ , já que  $v_1 = 1$  indica que o pagamento médio para o par de jogadores é  $R$  (cooperação mútua). Analogamente, caso  $2P < (S + T)$ ,

$$\frac{1}{2}(s_Y + s_X) \geq \frac{1}{2}[2v_1P + 2(v_2 + v_3)P + 2v_4P] = P$$

com igualdade  $\iff v = (0, 0, 0, 1) \iff s_Y = s_X = P$ , já que  $v_4 = 1$  indica que o pagamento médio para o par de jogadores é  $P$ , que é a não-cooperação mútua. ■

### 3.1 Estratégias boas

O vetor  $\tilde{p}$  que aparece na segunda coluna da matriz (2.8) tem um papel especial no formalismo que desenvolvemos a partir daqui.

**Definição 20** *O vetor de Press-Dyson  $\tilde{p}$  associado à estratégia  $p$  de memória um é  $\tilde{p} = p - e_{12}$  onde  $e_{12} = (1, 1, 0, 0)$ .*

Como dissemos anteriormente, nem sempre consideraremos que o jogador  $Y$  joga uma estratégia de memória um. Em alguns resultados importantes é suficiente que exista uma distribuição de probabilidade que possibilite definir pagamentos médios por jogada para os jogadores  $X$  e  $Y$ .

**Definição 21** *Um padrão estratégico é uma estratégia para o jogador  $Y$ , não necessariamente de memória um, tal que exista uma distribuição de probabilidade para a ocupação dos estados  $CC$ ,  $CD$ ,  $DC$  e  $DD$  pelo par de jogadores  $X$  e  $Y$ .*

**Definição 22** *Uma estratégia de memória um escolhida por  $X$  é chamada **boa** se  $p_1 = 1$  e se, para qualquer padrão estratégico escolhido pelo adversário  $Y$  e qualquer distribuição limite associada, os pagamentos médios satisfazem*

$$s_Y \geq R \implies s_Y = s_X = R. \quad (3.3)$$

Apesar de sua simplicidade, o resultado a seguir é um dos mais importantes no que segue. A formulação presente foi apresentada originalmente por Hilbe, Nowak e Sigmund [3] e repetida por Akin [1]. Observe o papel proeminente do vetor  $\tilde{p}$  associado à estratégia  $p$ .

**Teorema 8** *Suponha que  $X$  use a estratégia cujo vetor de Press-Dyson é  $\tilde{p}$ . Se o adversário  $Y$  usa um padrão estratégico tal que  $v^n$  é a distribuição de probabilidade no tempo  $n$ , então  $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \langle v^k, \tilde{p} \rangle = 0$ . Se  $v^k \rightarrow v$  no sentido de Cesaro, então*

$$\langle v, \tilde{p} \rangle = 0. \quad (3.4)$$

**Prova:** Levando em conta que  $v_1^n$  e  $v_2^n$  são as probabilidades dos estados 1 (CC) e 2 (CD), então o produto escalar  $v_{12}^n \equiv \langle v^n, e_{12} \rangle = v_1^n + v_2^n$  é a probabilidade de  $X$  jogar  $C$  na  $n$ -ésima rodada do jogo. Por outro lado, pela definição do vetor de estratégia  $p$  como sendo o vetor cujas componentes são probabilidades de cooperar em cada um dos quatro estados do par de jogadores, temos que a probabilidade de  $X$  jogar  $C$  na  $(n+1)$ -ésima rodada é  $\langle v^n, p \rangle$ . Portanto,  $v_{12}^{n+1} = \langle v^n, p \rangle$ . Assim,

$$v_{12}^{n+1} - v_{12}^n = \langle v^n, p \rangle - \langle v^n, e_{12} \rangle = \langle v^n, p - e_{12} \rangle = \langle v^n, \tilde{p} \rangle.$$

Usando soma telescópica, obtemos

$$\begin{aligned} \sum_{k=1}^n \langle v^k, \tilde{p} \rangle &= (v_{12}^2 - v_{12}^1) + (v_{12}^3 - v_{12}^2) + \dots + (v_{12}^{n+1} - v_{12}^n) \\ &= (v_{12}^{n+1} - v_{12}^1). \end{aligned}$$

Multiplicando por  $\frac{1}{n}$  em ambos os lados, temos

$$\frac{1}{n} \sum_{k=1}^n \langle v^k, \tilde{p} \rangle = \frac{1}{n} (v_{12}^{n+1} - v_{12}^1).$$

Como o módulo da diferença  $v_{12}^{n+1} - v_{12}^1$  é limitado superiormente por 1, então o lado direito da última expressão tende a zero quando  $n$  tende ao infinito. Logo, provamos que

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \langle v^k, \tilde{p} \rangle = 0.$$

Agora, se  $v^k \rightarrow v$  no sentido de Cesaro, façamos

$$W_n = \frac{1}{n} \sum_{k=1}^n v^k .$$

Pela Definição 12,  $W_n \rightarrow v$  no sentido usual. Assim,

$$\lim_{n \rightarrow \infty} (\langle W_n, \tilde{p} \rangle) = 0 .$$

Como

$$\begin{aligned} | \langle W_n, \tilde{p} \rangle - \langle v, \tilde{p} \rangle | &= | \langle (W_n - v), \tilde{p} \rangle | \\ &\leq \|W_n - v\| \|\tilde{p}\| \rightarrow 0 , \end{aligned}$$

obtemos (3.4). ■

**Teorema 9** *Suponha  $p = (p_1, p_2, p_3, p_4)$  um vetor estratégia. Sejam*

$$A = (T - R)p_3 - (1 - p_2)(R - S)$$

e

$$B = (T - R)p_4 - (1 - p_2)(R - P) .$$

*Então  $p$  é uma estratégia boa se, e somente se,  $p_2 < 1$ ,  $A < 0$  e  $B < 0$ .*

**Prova:** Seja  $v$  o vetor de probabilidade que descreve a distribuição estacionária para o par de jogadores. Primeiramente, eliminamos a possibilidade de  $p_2 = 1$ . Suponha então que X use a estratégia  $p = (1, 1, p_3, p_4)$ . Vamos mostrar que se a estratégia de Y é  $q = (0, 0, 0, 0)$ , então teremos  $s_Y > R$  com  $s_X < R$ . O grafo orientado de adjacências da cadeia de Markov para essa situação está na Figura 3.1. Note que  $CC$ ,  $DC$  e  $DD$  são estados transientes e assim  $v_1 = v_3 = v_4 = 0$ . Portanto a única componente não nula de  $v$  é  $v_2 = 1$ . Assim, por (1.12) e pela hierarquia (1.7), temos  $s_X = S < P$  e  $s_Y = T > R$ . Portanto, quando  $p_2 = 1$ , a estratégia não é boa.

De agora em diante podemos então supor que  $p_2 < 1$ . Podemos calcular

$$\begin{aligned} s_Y - R &= \langle S_Y, v \rangle - R(v_1 + v_2 + v_3 + v_4) \\ &= \langle (R, T, S, P), (v_1, v_2, v_3, v_4) \rangle - R(v_1 + v_2 + v_3 + v_4) \\ &= (T - R)v_2 + (S - R)v_3 + (P - R)v_4 . \end{aligned} \tag{3.5}$$

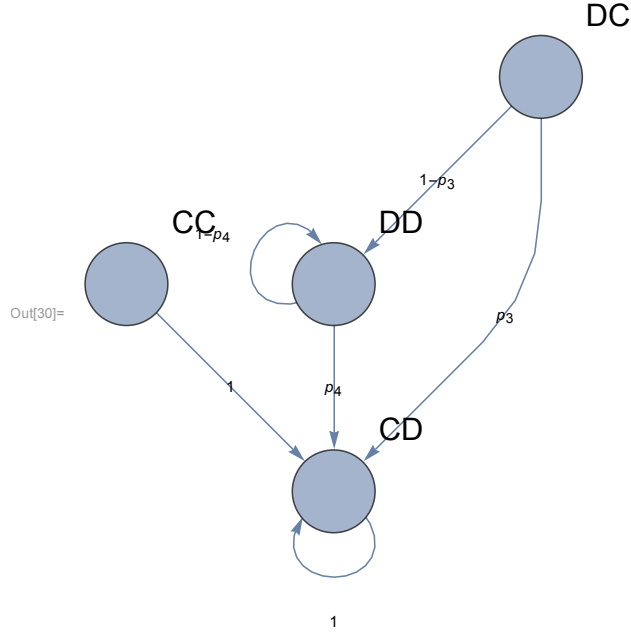


Figura 3.1: Grafo orientado que representa a demonstração do Teorema 9 quando  $p_2 = 1$ .

Levando em conta que  $p_1 = 1$ , então  $\tilde{p} = (0, p_2 - 1, p_3, p_4)$  e segue de (3.4) que

$$v_2(p_2 - 1) + v_3p_3 + v_4p_4 = 0 \quad (3.6)$$

que permite escrever o lado direito de (3.5) somente em termos de  $v_3$  e  $v_4$ . Multiplicando ambos os lados de (3.5) por  $(1 - p_2)$ , obtemos finalmente

$$(1 - p_2)(s_Y - R) = Av_3 + Bv_4, \quad (3.7)$$

onde  $A$  e  $B$  estão definidos no enunciado deste teorema.

Suponha  $A < 0$  e  $B < 0$ . Se  $s_Y \geq R$ ,  $v_3 = v_4 = 0$ . De (3.6) e  $p_2 < 1$  temos que  $v_2 = 0$  e isto implica  $v = (1, 0, 0, 0)$ . Assim, pela Proposição 3 temos  $s_X = s_Y = R$ , mostrando que  $p$  é uma estratégia boa.

Mostraremos agora que caso não valham  $A < 0$  e  $B < 0$ , a estratégia  $p$  não pode ser boa. Suponha  $A \geq 0$ . Suponha que  $Y$  use uma estratégia  $q = (0, 1, 1, 1)$ . O grafo orientado que representa as transições na cadeia de Markov para o par de jogadores  $X$  e  $Y$  está na Figura 3.2. Note que o estado

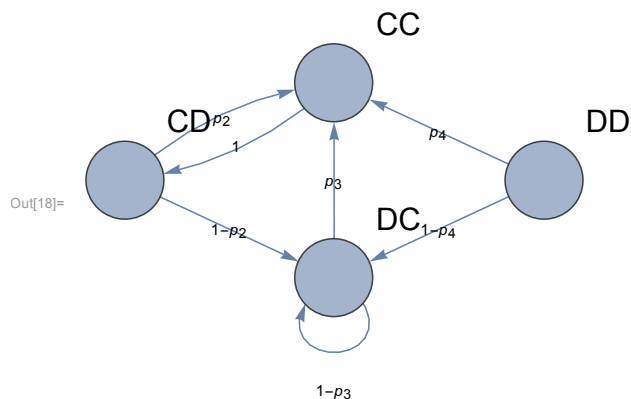


Figura 3.2: Grafo orientado que representa a demonstração do Teorema 9 quando  $p_2 < 1$ ,  $A \geq 0$ .

$DD$  é transitente, portanto  $v_4 = 0$ . O estado  $CC$  não é absorvente, logo  $v_1 < 1$ . Como  $v_4 = 0$  e  $A \geq 0$  então (3.7) implica que

$$s_Y \geq R .$$

Ao mesmo tempo, ver Proposição 3, como  $v_1 < 1$ , então  $s_X + s_Y < 2R$ , que implica  $s_X < R$ , provando que  $p$  não é boa.

Suponha agora  $B \geq 0$ . Se  $Y$  usa estratégia ALLD, ou seja,  $q = (0, 0, 0, 0)$  então as transições da cadeia de Markov estão ilustradas pelo grafo da Figura 3.3, em que os estados  $CC$  e  $DC$  são transientes, logo  $v_1 = 0$  e  $v_3 = 0$ . De (3.7) e já que  $v_3 = 0$  e  $B \geq 0$ , então

$$s_Y \geq R .$$

Novamente, usando um argumento análogo ao do caso  $A \geq 0$ , vemos que  $p$  não é uma estratégia boa. ■

### 3.2 A base de Press-Dyson e as estratégias DZ

Uma outra ideia muito importante de Press e Dyson foi de escrever  $\tilde{p}$  usando uma boa base para  $\mathbb{R}^4$ .

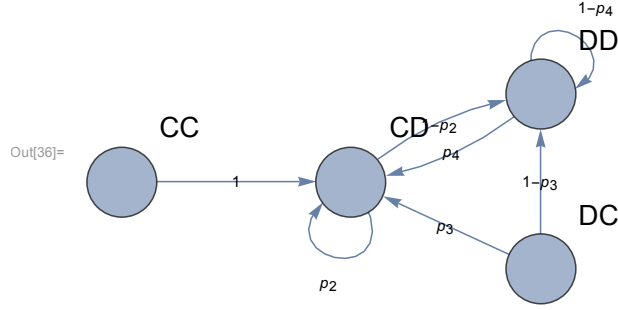


Figura 3.3: Grafo orientado que representa a demonstração do Teorema 9 quando  $p_1 = 1$ ,  $p_2 < 1$ ,  $B \geq 0$ .

**Proposição 4** Os vetores  $S_X = (R, S, T, P)$ ,  $S_Y = (R, T, S, P)$ ,  $\mathbf{1} = (1, 1, 1, 1)$  e  $e_{23} = (0, 1, 1, 0)$  constituem uma base  $\mathbb{R}^4$ .

**Prova:** Basta verificar que

$$\det \begin{bmatrix} R & R & 1 & 0 \\ S & T & 1 & 1 \\ T & S & 1 & 1 \\ P & P & 1 & 0 \end{bmatrix} = -2(R - P)(T - S) \neq 0.$$

■

Chamaremos  $\{S_X, S_Y, \mathbf{1}, e_{23}\}$  de *base de Press-Dyson*. O próximo resultado, consequência direta do Teorema 8, nos leva bem próximos aos resultados obtidos por Press e Dyson em [7].

**Teorema 10** Se  $p$  é uma estratégia com vetor de Press-Dyson  $\tilde{p} = \alpha S_X + \beta S_Y + \gamma \mathbf{1} + \delta e_{23}$  e se  $Y$  joga um padrão estratégico qualquer tal que exista uma distribuição de probabilidade  $v$  no sentido de Cesaro, então os pagamentos médios  $s_X$  e  $s_Y$  satisfazem

$$\alpha s_X + \beta s_Y + \gamma + \delta v_{23} = 0, \quad (3.8)$$

onde  $v_{23} = \langle e_{23}, v \rangle$ .

**Prova:** Temos que  $\tilde{p} = \alpha S_X + \beta S_Y + \gamma \mathbf{1} + \delta e_{23}$ . Fazendo o produto interno de  $v$  com  $\tilde{p}$ , obtemos

$$\begin{aligned} \langle v, \tilde{p} \rangle &= \alpha \langle S_X, v \rangle + \beta \langle S_Y, v \rangle + \gamma \langle \mathbf{1}, v \rangle + \delta \langle e_{23}, v \rangle \\ &= \alpha s_X + \beta s_Y + \gamma + \delta v_{23} \end{aligned}$$

e já que  $\langle v, \tilde{p} \rangle = 0$  pelo Teorema 8, provamos o resultado (3.8). ■

**Definição 23** *Uma estratégia de determinante zero (DZ) é uma estratégia  $p$  cujo vetor  $\tilde{p}$ , quando escrito na base de Press-Dyson, possui  $\delta = 0$ .*

A grande descoberta de Press e Dyson vem do fato de terem percebido que existem de fato estratégias  $\tilde{p}$  com  $\delta = 0$ . Quando se substitui  $\delta = 0$  em (3.8) recupera-se o resultado (2.13) sem a necessidade de supor que Y adote uma estratégia de memória um. Basta supor que Y adota um padrão estratégico tal que exista uma distribuição de probabilidade estacionária  $v$ . Tampouco precisamos supor que a cadeia de Markov seja irredutível, assim como nos livramos da dificuldade de dividir por  $\alpha$  em (2.5) sem termos provado que  $\alpha \neq 0$ .

O mais surpreendente sobre as estratégias DZ é que enquanto  $\delta \neq 0$ , Y tem algum controle sobre o pagamento de X e o seu próprio através da sua capacidade de influenciar a distribuição estacionária  $v$ . Se X não joga uma estratégia DZ, os pares de pagamentos ficam espalhados, conforme visto no primeiro quadro da Figura 2.1. Assim, se  $\delta \neq 0$ , Y tem uma maior liberdade de influenciar seu próprio pagamento e o de X. Mas quando X escolhe uma estratégia DZ, com  $\delta = 0$ , os pares de pagamentos ficam concentrados em uma linha reta e nenhuma escolha de Y pode fazer com que o par de pagamentos  $(s_X, s_Y)$  saia fora de uma reta.

### 3.3 Geometria do conjunto de estratégias de memória um na representação da base de Press-Dyson

Sabe-se que  $\tilde{p} = p - e_{12} = (p_1 - 1, p_2 - 1, p_3, p_4)$  e  $0 \leq p_1, p_2, p_3, p_4 \leq 1$ , logo  $-1 \leq \tilde{p}_1, \tilde{p}_2 \leq 0$  e  $0 \leq \tilde{p}_3, \tilde{p}_4 \leq 1$ . As restrições  $\tilde{p}_1 \leq 0$ ,  $\tilde{p}_2 \leq 0$ ,  $\tilde{p}_3 \geq 0$  e  $\tilde{p}_4 \geq 0$  são as chamadas *restrições de sinal*. Já que

$$\begin{aligned} \tilde{p}_1 &= (\alpha + \beta)R + \gamma \\ \tilde{p}_2 &= \alpha S + \beta T + \gamma + \delta \\ \tilde{p}_3 &= \alpha T + \beta S + \gamma + \delta \\ \tilde{p}_4 &= (\alpha + \beta)P + \gamma \end{aligned} \quad ,$$

as restrições de sinal se tornam

$$\begin{aligned} (\alpha + \beta)R + \gamma &\leq 0 \\ \alpha S + \beta T + \gamma + \delta &\leq 0 \\ \alpha T + \beta S + \gamma + \delta &\geq 0 \\ (\alpha + \beta)P + \gamma &\geq 0 \end{aligned} \quad (3.9)$$

**Lema 1** *Se  $\tilde{p} = \alpha S_X + \beta S_Y + \gamma \mathbf{1} + \delta e_{23}$  satisfaz as restrições de sinal, então*

$$\alpha + \beta \leq 0 \quad \text{e} \quad \gamma \geq 0.$$

*Além do mais*

$$\alpha + \beta = 0 \Leftrightarrow \gamma = 0.$$

**Prova:** Subtraindo a 4ª desigualdade da 1ª em (3.9), obtemos

$$(\alpha + \beta)(R - P) \leq 0.$$

Como  $(R - P) > 0$ , então  $(\alpha + \beta) \leq 0$ . Da 4ª desigualdade e do fato  $P > 0$  resulta que  $\gamma \geq 0$ . Finalmente, a 1ª e a 4ª desigualdades implicam que  $\alpha + \beta = 0 \Leftrightarrow \gamma = 0$ . ■

Observe que, ambos  $\tilde{p}_1$  e  $\tilde{p}_4$  valerão zero se, e somente se,  $\alpha + \beta = \gamma = 0$ . Isto acontece quando  $p_1 = 1$  e  $p_4 = 0$ .

Um outro tipo de restrição importante para um vetor  $\tilde{p}$  de Press-Dyson é que suas entradas possuem valor absoluto no máximo 1. Estas são as *restrições de tamanho*. Mais adiante voltaremos a falar das restrições de tamanho.

### 3.3.1 Estratégias excepcionais

**Definição 24** *Estratégias excepcionais são aquelas cujos vetores de Press-Dyson possuem  $\gamma = \alpha + \beta = 0$ .*

Decorre da definição acima que  $\tilde{p}_1 = \tilde{p}_4 = 0$ . Pelas restrições de sinal e tamanho  $-1 \leq \tilde{p}_2 \leq 0$  e  $0 \leq \tilde{p}_3 \leq 1$ , assim as estratégias excepcionais são representadas como pontos em um quadrado no plano  $\tilde{p}_2\tilde{p}_3$  com vértices  $(0, 0)$ ,  $(-1, 0)$ ,  $(-1, 1)$ ,  $(0, 1)$ , ver Figura 3.4.

Dois dos vértices do quadrado das estratégias excepcionais são estratégias já mencionadas na literatura e de alguma importância. O vértice  $(-1, 0)$  é conhecido como GRIM: coopera com cooperadores, mas, uma vez traído não



coopera nunca mais. O vértice  $(-1, 1)$  é a famosa estratégia TFT vencedora dos torneios de Axelrod.

Os demais vértices não são muito notáveis.  $(0, 0)$  é chamado REPETE, pois simplesmente repete sua primeira jogada eternamente, independente do que jogar o adversário.  $(0, 1)$  foi chamada por Akin [1] de LAME.

As estratégias excepcionais possuem vetor  $\tilde{p} = (0, \delta - \alpha(T - S), \delta + \alpha(T - S), 0)$  fornecido pelas restrições de sinal (3.9) e o fato de  $\alpha = -\beta$ . Colocando  $\delta = 0$  nas equações para  $\tilde{p}_2$  e  $\tilde{p}_3$ , as estratégias DZ excepcionais são as da diagonal do quadrado.

Para identificar as estratégias boas do quadrado, deve-se usar o Teorema 9. A condição  $B < 0$  só fornece  $\tilde{p}_2 < 0$ . A condição  $A < 0$  leva a  $\tilde{p}_3 < -\frac{R-S}{T-R}\tilde{p}_2$ . A reta  $\tilde{p}_3 = -\frac{R-S}{T-R}\tilde{p}_2$  passa pela origem e possui inclinação menor que -1, pois, por (1.8),  $R - S > \frac{T-S}{2}$  enquanto  $T - R < \frac{T-S}{2}$ . Portanto, essa reta intercepta o segmento TFT-LAME em algum ponto entre estas estratégias. A região das estratégias excepcionais e boas é então a parte do quadrado à esquerda dessa reta. Observe que, com exceção de REPETE, todas as estratégias excepcionais DZ são boas. GRIM é boa, mas não é DZ. TFT é excepcional, DZ é boa.

### 3.3.2 Estratégias não excepcionais

Vamos agora tratar das estratégias mais típicas, que são as não excepcionais, ou seja, aquelas em que temos  $\alpha + \beta < 0$  e  $\gamma > 0$ . Defina  $\bar{\alpha} = \frac{\alpha}{\gamma}$ ,  $\bar{\beta} = \frac{\beta}{\gamma}$ ,  $\bar{\delta} = \frac{\delta}{\gamma}$ . A primeira e a quarta das desigualdades em (3.9) fornecem

$$-P^{-1} \leq \bar{\alpha} + \bar{\beta} \leq -R^{-1}, \quad (3.10)$$

enquanto a segunda e terceira levam a

$$\bar{\alpha}S + \bar{\beta}T \leq -1 - \bar{\delta} \leq \bar{\alpha}T + \bar{\beta}S. \quad (3.11)$$

Em termos de  $\bar{\alpha}$ ,  $\bar{\beta}$  e  $\bar{\delta}$ , podemos reescrever a equação (3.8) como

$$\bar{\alpha}s_X + \bar{\beta}s_Y + \bar{\delta}v_{23} + 1 = 0. \quad (3.12)$$

As restrições de sinal (3.10) identificam as estratégias não excepcionais com pontos no plano  $\bar{\alpha}\bar{\beta}$  localizados entre as retas paralelas  $\bar{\alpha} + \bar{\beta} = -R^{-1}$  e  $\bar{\alpha} + \bar{\beta} = -P^{-1}$ . Já as restrições de sinal (3.11) identificam as estratégias não excepcionais com pontos do plano  $\bar{\alpha}\bar{\beta}$  em um quadrante determinado

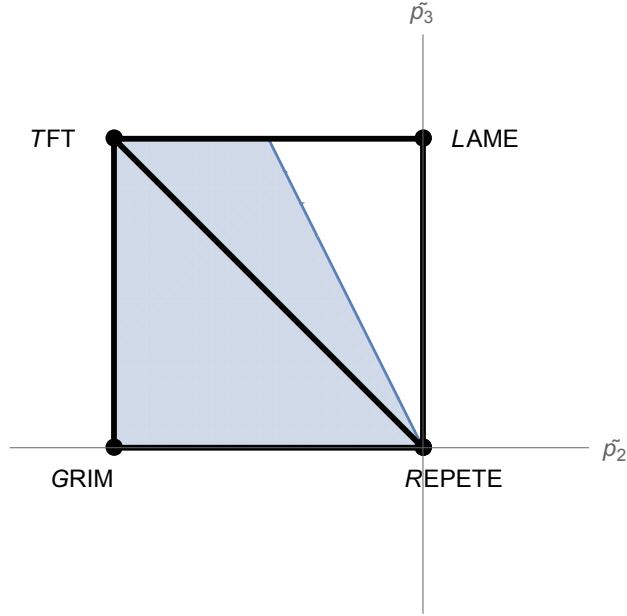


Figura 3.4: As estratégias excepcionais são representadas pelos pontos do quadrado no plano  $\tilde{p}_2\tilde{p}_3$  aqui ilustrado. A diagonal mostrada na figura são as estratégias DZ excepcionais. A região sombreada corresponde às estratégias boas e excepcionais.

por duas retas concorrentes. Para cada  $\bar{\delta} \in \mathbb{R}$ , as estratégias estão portanto associadas aos pontos da interseção entre um quadrante e uma faixa. Esta interseção está ilustrada na Figura 3.5 para o caso  $\bar{\delta} = 0$ .

Os pontos localizados sobre a reta  $\bar{\alpha} + \bar{\beta} = -P^{-1}$  correspondem às estratégias não excepcionais com  $p_4 = 0$ . E os pontos localizados sobre a reta  $\bar{\alpha} + \bar{\beta} = -R^{-1}$  correspondem às estratégias não excepcionais com  $p_1 = 1$ .

**Definição 25** *Seja  $\tilde{p} = \gamma(\bar{\alpha}S_X + \bar{\beta}S_Y + 1)$  o vetor de Press-Dyson de uma estratégia DZ não excepcional. A região EDZ é definida por*

$$EDZ := \{(\bar{\alpha}, \bar{\beta}) / \bar{\alpha}S + \bar{\beta}T \leq -1 \leq \bar{\alpha}T + \bar{\beta}S \text{ e } -P^{-1} \leq \bar{\alpha} + \bar{\beta} \leq -R^{-1}\} .$$

Quando o ponto  $(\bar{\alpha}, \bar{\beta})$  estiver na região EDZ temos que as restrições de sinal são satisfeitas. Desde que  $\gamma > 0$  seja pequeno o suficiente, as restrições de tamanho serão também respeitadas. Observe que o ponto de interseção das retas  $\bar{\alpha}S + \bar{\beta}T + 1 = 0$  e  $\bar{\alpha}T + \bar{\beta}S + 1 = 0$  que definem o quadrante da região

EDZ possui coordenadas  $(-\frac{1}{s+T}, -\frac{1}{s+T})$ . Das desigualdades (1.8) vê-se que este ponto está contido na faixa entre as retas  $\bar{\alpha} + \bar{\beta} = -P^{-1}$  e  $\bar{\alpha} + \bar{\beta} = -R^{-1}$ . Este ponto de interseção está portanto associado às estratégias DZ que chamaremos *estratégias vértice*. Embora as estratégias vértice estejam associadas a um único ponto no plano  $\bar{\alpha}\bar{\beta}$ , é bom lembrar que há infinitas estratégias, pois há infinitas escolhas de  $\gamma$  tais que as restrições de tamanho são satisfeitas. Mais em geral, pelo mesmo motivo, a cada ponto  $(\bar{\alpha}, \bar{\beta})$  na região EDZ estão associadas a infinitas estratégias DZ.

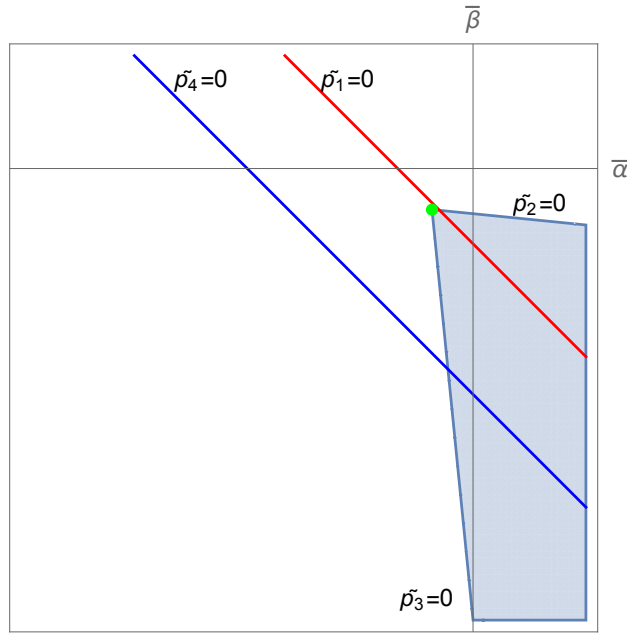


Figura 3.5: A interseção do quadrante em cinza com a faixa entre as duas retas paralelas vermelha e azul é a região EDZ. O ponto verde é a interseção  $(-\frac{1}{s+T}, -\frac{1}{s+T})$  das retas  $\tilde{p}_2 = 0$  e  $\tilde{p}_3 = 0$  e representa as chamadas estratégias vértice.

**Teorema 11** *Suponha que  $X$  utilize uma estratégia  $p$  DZ com  $\tilde{p} = \gamma(\bar{\alpha}S_X + \bar{\beta}S_Y + 1)$ ,  $\gamma > 0$ . Seja  $Z = -(\bar{\alpha} + \bar{\beta})^{-1}$ , de modo que  $P \leq Z \leq R$ . Para qualquer padrão estratégico escolhido por  $Y$  os pares de pagamentos estão contidos em uma reta que passa pelo ponto  $(Z, Z)$ :*

$$s_Y - Z = \chi(s_X - Z) \quad (3.13)$$

onde

$$\chi = \frac{\bar{\alpha}Z}{1 + \bar{\alpha}Z} \quad (3.14)$$

é o coeficiente angular.

**Prova:**

$$\bar{\alpha} + \bar{\beta} = -Z^{-1} \quad \Rightarrow \quad -\bar{\beta}Z = 1 + \bar{\alpha}Z .$$

Já que  $\bar{\delta} = 0$  e multiplicando (3.12) por  $Z$ , obtemos

$$\bar{\alpha}Zs_X + \bar{\beta}Zs_Y + Z = 0 .$$

Substituindo  $\bar{\beta}Z$  encontramos a equação

$$\bar{\alpha}Zs_X - (1 + \bar{\alpha}Z)s_Y + Z = 0 . \quad (3.15)$$

Definindo  $\chi = \frac{\bar{\alpha}Z}{1 + \bar{\alpha}Z}$ , obtemos

$$\chi(s_X + \frac{1}{\bar{\alpha}}) = s_Y ,$$

que se torna (3.13) após rearranjo dos termos. ■

**Corolário 3** *Suponha que  $X$  utilize um vetor estratégia  $p$  com vetor de Press-Dyson  $\tilde{p} = \gamma(\bar{\alpha}S_X + \bar{\beta}S_Y + 1)$ ,  $\gamma > 0$ . Seja  $\bar{\alpha} + \bar{\beta} = -Z^{-1}$ . Suponha que  $Y$  use um padrão estratégico arbitrário.*

(a) *Se  $\bar{\alpha} = 0$ , então  $s_Y = Z$ . Se  $\bar{\alpha} \neq 0$ , então as seguintes afirmações são equivalentes:*

- (i)  $s_Y = s_X$
- (ii)  $s_Y = Z$
- (iii)  $s_X = Z$

(b) *Se  $s_Y > s_X$ , então*

- $\bar{\alpha} > 0 \quad \Rightarrow \quad Z > s_Y > s_X$
- $\bar{\alpha} = 0 \quad \Rightarrow \quad Z = s_Y > s_X$
- $\bar{\alpha} < 0 \quad \Rightarrow \quad s_Y > Z > s_X$

(c) *Se  $s_X > s_Y$ , então*

- $\bar{\alpha} > 0 \Rightarrow s_X > s_Y > Z$
- $\bar{\alpha} = 0 \Rightarrow s_X > s_Y = Z$
- $\bar{\alpha} < 0 \Rightarrow s_X > Z > s_Y$

**Prova:** Veja a Figura 3.6. Os pontos vermelhos são para análise do item (b) e os pontos azuis são para análise do item (c). ■

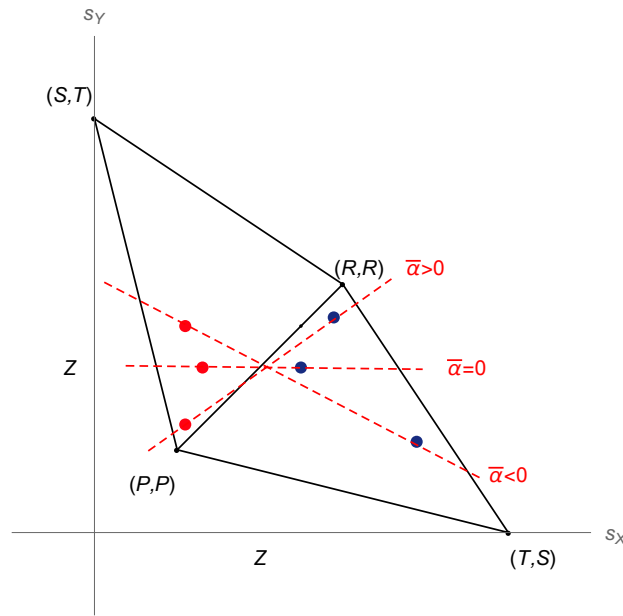


Figura 3.6: Ilustração das situações do Corolário 3.

Se  $\bar{\alpha} = 0$ , o coeficiente angular  $\chi$  será nulo e (3.13) implica que  $s_Y = Z$ . As estratégias DZ correspondentes são chamadas *equalizadoras* e foram descritas por Press e Dyson em [7]. Os pagamentos de X e Y estão concentrados em uma reta horizontal. Se  $\bar{\alpha} > 0$ , e  $Z = R$  as estratégias DZ são chamadas *generosas* e foram descritas primeiramente em [11]. As estratégias DZ com  $\bar{\alpha} > 0$  e  $Z = P$  foram reconhecidas já em [7] e chamadas *extorsivas*, ver também [3]. Estes três tipos particulares e interessantes de estratégias DZ serão mais analisados mais adiante nas Subseções 3.4.1 a 3.4.3.

## 3.4 Estratégias não excepcionais DZ

As fórmulas (2.13) de Press e Dyson e (3.12) com  $\bar{\delta} = 0$  impõem naturalmente uma relação afim entre os pagamentos dos dois jogadores do DPR. A existência de tais estratégias tem consequências em longo prazo. Se um jogador X usa estratégia DZ, ele pode impor fortes restrições ao pagamento de seu adversário. Já Y não consegue variar com muita liberdade sua recompensa, mesmo que suas escolhas possam afetar a recompensa de X [11].

No resto desta seção estudaremos classes interessantes de estratégias DZ.

### 3.4.1 Equalizadoras

Para obter uma estratégia equalizadora, basta tomar  $\bar{\alpha} = 0$ , o que resulta em  $\tilde{p} = \gamma(\bar{\beta}S_Y + 1)$ . O que caracteriza uma estratégia equalizadora é que a equação para os pagamentos (3.12) resulta em  $s_Y = -\frac{1}{\bar{\beta}}$ . Ao usar uma estratégia DZ, um jogador fixa o pagamento do seu adversário Y de forma unilateral. Como  $\bar{\alpha} + \bar{\beta} = -Z^{-1}$ , então  $\bar{\alpha} = 0$ , implica que  $\bar{\beta} = -Z^{-1}$ . Escrevendo  $p$  em função  $\tilde{p}$ , encontramos as 4 equações que definem as componentes do vetor estratégia para uma estratégia equalizadora:

$$\begin{aligned} p_1 &= \gamma(-Z^{-1}R + 1) + 1 \\ p_2 &= \gamma(-Z^{-1}T + 1) + 1 \\ p_3 &= \gamma(-Z^{-1}S + 1) \\ p_4 &= \gamma(-Z^{-1}P + 1) \end{aligned} \quad , \quad (3.16)$$

onde  $\gamma > 0$  é pequeno o suficiente de modo a que todas componentes situem-se em  $[0, 1]$ . Na Figura 3.7 ilustramos um exemplo de pagamentos para uma estratégia equalizadora.

**Observação:** Já que as estratégias equalizadoras permitem que X escolha um pagamento fixo para Y, é natural perguntar-se se X poderia escolher um pagamento fixo para si próprio. Press e Dyson [7] já haviam respondido negativamente a essa pergunta. Do ponto de vista de Akin [1], isto decorre simplesmente do fato de que não existem pontos da região EDZ com  $\bar{\beta} = 0$ .

### 3.4.2 Extorsivas

Uma estratégia de extorsão é uma estratégia em que o jogador X ganha mais do que Y e este, no intuito de maximizar seu ganho, maximiza também o de X. Press e Dyson mostraram que existem estratégias de extorsão dentre as

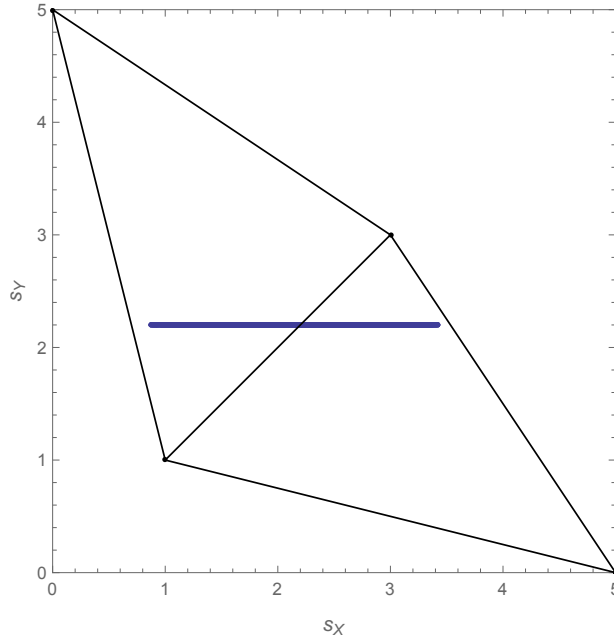


Figura 3.7: Pagamentos para os jogadores X e Y quando X usa a estratégia equalizadora  $p = (\frac{4}{5}, \frac{3}{10}, \frac{11}{20}, \frac{3}{10})$ , obtida com  $\gamma = 0.55$ ,  $Z = 2.2$  e a escolha  $T = 5$ ,  $R = 3$ ,  $P = 1$  e  $S = 0$  contra uma amostra de 10.000 estratégias  $q$  de Y.

DZ. Do ponto de vista de Akin, basta tomar  $\bar{\alpha} > 0$  e  $Z = P$ . A equação (3.13) torna-se

$$s_X - P = \frac{1}{\chi}(s_Y - P),$$

onde, por (3.14),  $0 < \chi < 1$ . Se o jogador X escolhe uma estratégia de extorsão, o adversário Y é forçado a cooperar com o extorsor porque essa opção lhe oferece melhor retorno. Ademais, o jogador X escolhe o fator de extorsão  $\frac{1}{\chi} > 1$  que lhe diz quantas vezes o excesso de  $s_Y$  com relação a  $P$  será o seu próprio excesso com relação a  $P$ . Hilbe, Nowak e Sigmund [3] analisaram o desempenho evolutivo das estratégias de extorsão mostrando que em populações razoavelmente grandes elas podem agir como catalisadores para a evolução da cooperação, mas não são em geral o resultado final do processo evolutivo.

Para adotar uma estratégia extorsiva, basta escolher  $Z = P$ ,  $\bar{\alpha} > 0$ ,

$\bar{\beta} = -P^{-1} - \bar{\alpha}$  em (3.13). Obtemos

$$\begin{aligned} p_1 &= \gamma(1 - P^{-1}R) + 1 \\ p_2 &= \gamma(\bar{\alpha}S - T(P^{-1} + \bar{\alpha}) + 1) + 1 \\ p_3 &= \gamma(\bar{\alpha}T - S(P^{-1} + \bar{\alpha}) + 1) \\ p_4 &= 0 \end{aligned} \quad , \quad (3.17)$$

onde, novamente,  $\gamma > 0$  é pequeno o suficiente para que as componentes de  $p$  estejam em  $[0, 1]$ . Note que as estratégias de extorsão sempre possuem  $p_4 = 0$ .

Suponha que X escolhe uma estratégia de extorsão. O seu pagamento depende da estratégia  $q$  de Y. Se Y quiser aumentar seu ganho, automaticamente aumentará o ganho de X. Por outro lado, Y pode também punir X por sua extorsão e, caso adote a estratégia ALLD, ambos os jogadores ganharão  $P$ .

**Exemplo 6** *X escolhe uma estratégia de extorsão com  $p = (p_1, p_2, p_3, 0)$ . Y decide cooperar plenamente na intenção de maximizar seu ganho com  $q = (1, 1, q_3, q_4)$ . Para encontrar o pagamento de X sob esta estratégia, será necessário encontrar a distribuição estacionária  $v$  e substituí-la em  $s_X = \langle v, S_X \rangle$ . Pode-se checar que  $v = (\frac{p_3}{p_3+1-p_1}, 0, \frac{1-p_1}{p_3+1-p_1}, 0)$ . Logo*

$$\begin{aligned} s_X &= \langle v, S_X \rangle \\ &= \langle (\frac{p_3}{p_3+1-p_1}, 0, \frac{1-p_1}{p_3+1-p_1}, 0), (R, S, T, P) \rangle \\ &= \frac{p_3}{p_3+1-p_1}R + \frac{1-p_1}{p_3+1-p_1}T \end{aligned}$$

Observe que quando Y coopera plenamente, o pagamento de X é maior que  $R$ , uma vez que é uma combinação linear convexa de  $R$  e  $T$ . Por outro lado, se Y jogar ALLD, teremos  $v = (0, 0, 0, 1)$  e ambos os jogadores recebem  $P$ . Para ilustrar os pagamentos das estratégias de extorsão veja Figura 3.8 que ilustra os pagamentos para a estratégia extorsiva  $p = (\frac{8}{9}, \frac{23}{36}, \frac{7}{36}, 0)$  obtida com  $T = 5$ ,  $R = 3$ ,  $P = 1$ ,  $S = 0$  escolhendo fator de extorsão  $\frac{1}{\chi} = 3$  e  $\gamma = \frac{1}{18}$ .

### 3.4.3 Generosas

Em [7] Press e Dyson descrevem apenas as estratégias DZ equalizadoras e extorsivas, mas Stewart e Plotkin [11] descobriram mais uma classe de



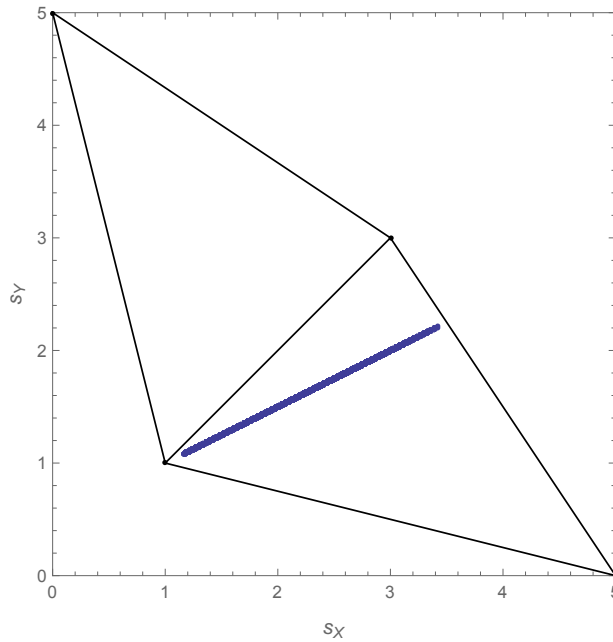


Figura 3.8: Pagamentos para os jogadores X e Y quando X usa a estratégia extorsiva  $p = (\frac{8}{9}, \frac{23}{36}, \frac{7}{36}, 0)$  obtida com  $\frac{1}{\chi} = 3$ ,  $\gamma = \frac{1}{18}$  e a escolha de  $T = 5$ ,  $R = 3$ ,  $P = 1$ ,  $S = 0$  contra uma amostra de 10.000 estratégias  $q$ . Observe que os pagamentos estão sobre um segmento de reta na região  $s_Y < s_X$ .

estratégias DZ, a qual nomearam *estratégias DZ generosas*. Em geral, uma estratégia  $p$  é generosa quando  $p_2 > 0$  ou  $p_4 > 0$ , isto é quando há uma probabilidade positiva de que X irá cooperar se Y trair. As estratégias DZ generosas são definidas tomando  $\bar{\alpha} > 0$  e  $Z = R$ . A equação (3.13) torna-se então

$$s_X - R = \frac{1}{\chi}(s_Y - R),$$

onde, por (3.14),  $0 < \chi < 1$ .

Os pares de pagamentos situam-se em uma reta de inclinação menor que 1 passando por  $(R, R)$  e portanto o pagamento de X é sempre menor que o de Y. Em particular, ambos os pagamentos são menores ou iguais a  $R$  e quando  $s_Y = R$  tem-se também  $s_X = R$  e portanto as estratégias DZ generosas são, pela Definição 22, boas. Podemos definir  $\frac{1}{\chi} > 1$  como o fator de generosidade: tudo que Y ganhar a menos que R, X receberá multiplicado por  $\frac{1}{\chi}$ .

Com as contas análogas às da extorsão obtemos o vetor para as estratégias DZ generosas  $p = (1, \gamma(\bar{\alpha}S - T(R^{-1} + \bar{\alpha}) + 1) + 1, \gamma(\bar{\alpha}T - S(R^{-1} + \bar{\alpha}) + 1), \gamma(1 - R^{-1}P))$ . Note que  $p_1$  é sempre 1.

Quando X escolhe uma estratégia DZ generosa, está incentivando a cooperação de seus adversários, que precisam cooperar para maximizar seus ganhos e, ao fazerem-no, aumentam também os ganhos de X. Este fato explica basicamente o motivo porque estratégias generosas em geral podem ser o resultado final de um processo em que a cooperação emerge de uma população em que a maioria não coopera. O fenômeno foi descoberto por Sigmund e Nowak [10] muito antes da descoberta das estratégias DZ.

Logo após o descobrimento das estratégias DZ, Stewart e Plotkin [11] simularam um torneio com as mesmas estratégias participantes do torneio de Axelrod [2] e acrescentaram a elas uma estratégia DZ generosa. Essa estratégia generosa foi a vencedora do torneio que, originalmente, tinha sido vencido por TFT.

**Exemplo 7** Para valores  $T = 5$ ,  $R = 3$ ,  $P = 1$  e  $S = 0$  e escolhendo  $\gamma = \frac{1}{16}$  e  $\frac{1}{\chi} = 2$  uma estratégia generosa fixa  $p = (1, \frac{31}{48}, \frac{3}{8}, \frac{1}{24})$ . A Figura 3.9 ilustra os pagamentos para essa estratégia.

### 3.4.4 Estratégias vértice

É interessante entendermos melhor as estratégias vértice, que foram mencionadas na Subseção 3.3.2 e definidas como as estratégias associadas ao ponto  $(-\frac{1}{S+T}, -\frac{1}{S+T})$  no vértice da região EDZ. As estratégias vértice não parecem particularmente interessantes do ponto de vista da evolução da cooperação, mas a sua posição de destaque na fronteira da região EDZ chama a atenção. Vamos descobrir que elas ilustram para estratégias DZ fenômenos interessantes de cadeias de Markov. Já que,  $\bar{\alpha} + \bar{\beta} = -Z^{-1}$ , então  $Z = \frac{S+T}{2}$ . Substituindo os valores de  $\bar{\alpha}$  e  $Z$  em (3.14) resulta em  $\chi = -1$ . Assim, os pagamentos de X e Y quando X adota uma estratégia vértice estão exatamente sobre a diagonal do quadrilátero da Figura 1.7 que liga os pontos  $(S, T)$  e  $(T, S)$ . Para obter o vetor  $p$  de uma estratégia vértice, deve-se substituir  $\bar{\alpha}$  e  $\bar{\beta}$  em  $\tilde{p} = \gamma(\bar{\alpha}S_X + \bar{\beta}S_Y + 1)$  resultando

$$p = (\gamma + 1 - \frac{2\gamma R}{S+T}, 1, 0, \gamma - \frac{2\gamma P}{S+T}), \quad (3.18)$$

onde novamente  $\gamma > 0$  é pequeno o suficiente. Observe que, ao adotar estratégias vértice sempre obtemos  $p_2 = 1$  e  $p_3 = 0$ . A Figura 3.10 ilustra os

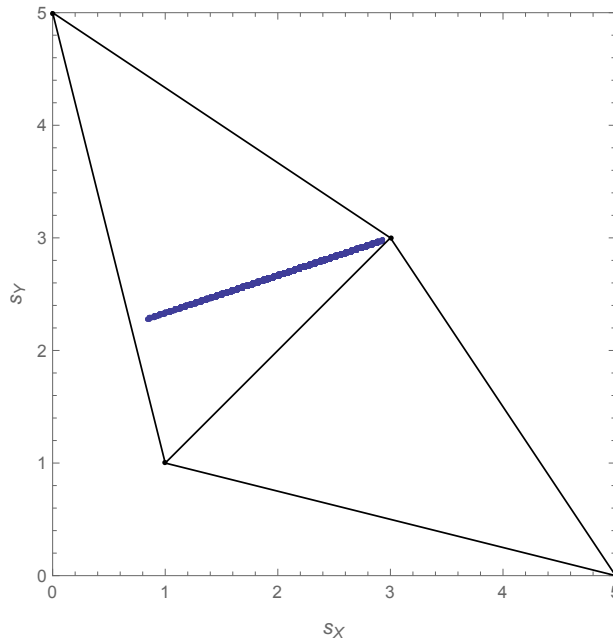


Figura 3.9: Pagamentos para os jogadores X e Y quando X usa a estratégia generosa  $p = (1, \frac{31}{48}, \frac{3}{8}, \frac{1}{24})$ , obtida com  $\frac{1}{\chi} = 2$ ,  $\gamma = \frac{1}{16}$  e a escolha  $T = 5$ ,  $R = 3$ ,  $P = 1$  e  $S = 0$  contra 10.000 estratégias  $q$ . Observe que os pagamentos estão sobre um segmento de reta na região  $s_Y > s_X$ .

pagamentos dessa estratégia para  $\gamma > 0$  pequeno o suficiente e valores  $T = 1$ ,  $R = \frac{3}{5}$ ,  $P = \frac{1}{5}$ ,  $S = 0$ .

### 3.5 Estratégias DZ justas

Poder-se-ia perguntar se não existiriam estratégias DZ simultaneamente extorsivas e generosas. Como as estratégias não excepcionais são todas caracterizadas por  $Z = -\frac{1}{\alpha+\beta}$  e estratégias extorsivas e generosas possuem valores diferentes de  $Z$ , então isto só poderia ser possível dentre as estratégias DZ excepcionais. Estas são exatamente as do segmento que liga REPETE a TFT na Figura 3.4. Como elas possuem  $\gamma = \delta = 0$  e  $\beta = -\alpha$ , então, se  $\alpha \neq 0$ , a equação (3.8) torna-se  $s_X = s_Y$ , que é tanto a equação (3.13) para os pagamentos das estratégias de extorsão, quanto (3.13) para os pagamentos

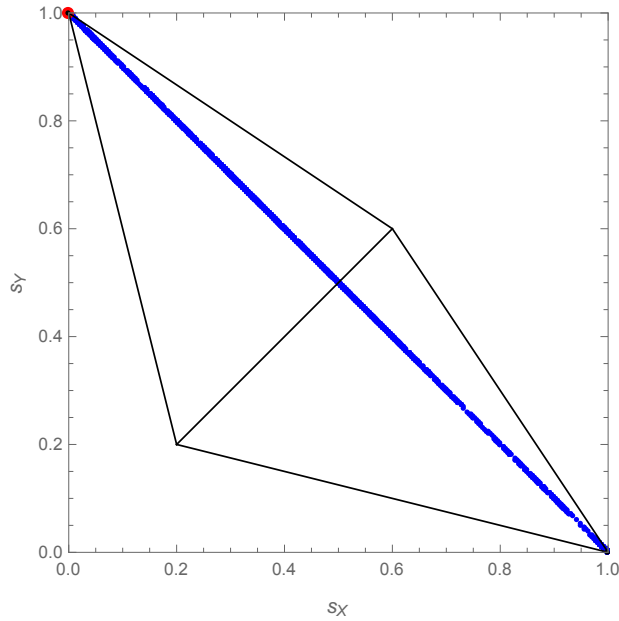


Figura 3.10: Pagamentos para os jogadores X e Y quando X usa a estratégia vértice  $p = (1 - \frac{\gamma}{5}, 1, 0, \frac{3\gamma}{5})$ , obtida com  $Z = \frac{1}{2}$ ,  $\bar{\alpha} = \bar{\beta} = -1$ ,  $\gamma > 0$  e a escolha  $T = 1$ ,  $R = \frac{3}{5}$ ,  $P = \frac{1}{5}$ ,  $S = 0$ . Observe que os pagamentos estão sobre o segmento de reta que liga  $(S, T)$  a  $(T, S)$ .

das estratégias DZ generosas, desde que se tome em ambos os casos  $\chi = 1$  (e portanto  $Z$  é cancelado).

As estratégias DZ justas são aquelas tais que  $s_X = s_Y$  para qualquer escolha de plano estratégico de Y. São portanto as estratégias com  $\tilde{p} = \alpha(S_X - S_Y) = (0, -\alpha(T - S), \alpha(T - S), 0)$ ,  $\alpha \in (0, (T - S)^{-1})$ .

Para uma ilustração dos pagamentos de uma estratégia DZ justa veja Figura 3.11.

O exemplo mais notável de estratégia DZ justa é TFT, de grande interesse histórico tanto por ter sido a vencedora dos torneios de Axelrod, quanto por seu papel no experimento computacional de Sigmund e Nowak [10] de evolução da cooperação. Assim como TFT, as outras estratégias DZ justas possivelmente também podem catalisar a evolução da cooperação, pois são capazes de combaterem as estratégias mais traidoras e se beneficiarem da cooperação com as mais cooperativas.

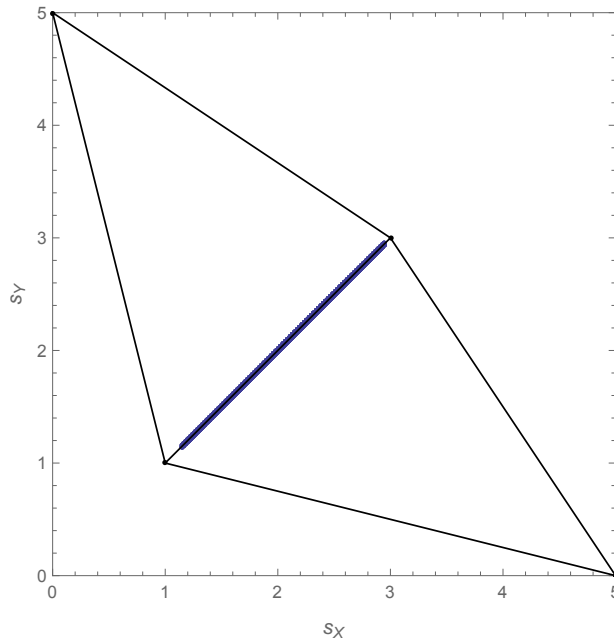


Figura 3.11: Para um valor fixo  $p$  de uma estratégia justa contra uma amostra de 10.000 estratégias  $q$ , os pagamentos dos jogadores X e Y estão sobre a diagonal  $s_Y = s_X$ .

Por completeza convém mencionar que a estratégia na fronteira do segmento das estratégias justas, ou seja, aquela com  $\alpha = 0$  em  $\tilde{p} = \alpha(S_X - S_Y)$  é a estratégia REPETE. Os pagamentos para esta não obedecem à equação  $s_X = s_Y$ . Como a estratégia repete indefinidamente sua primeira jogada, os pagamentos  $s_X$  e  $s_Y$  dependem da primeira jogada de X e da estratégia de Y. A cadeia de Markov possui pelo menos dois estados terminais, podendo ter quatro estados terminais, caso Y também adote REPETE.

### 3.6 Competição entre jogadores que escolhem estratégias DZ

Nas competições vistas em seções anteriores, as estratégias foram escolhidas por X com vetor  $p$  contra um padrão estratégico arbitrário escolhido pelo adversário Y. Suponha que ambos os jogadores escolham estratégias DZ.

Vamos calcular o que acontece quando X e Y usam estratégias DZ associadas, respectivamente, com os pontos  $(\bar{\alpha}, \bar{\beta})$  e  $(\bar{a}, \bar{b})$  na região EDZ. A situação está ilustrada na Figura 3.12. A reta  $E_1$  é o conjunto de todos os possíveis pares de pagamentos quando X usa a estratégia DZ associada ao ponto  $(\bar{\alpha}, \bar{\beta})$  na região EDZ e Y usa uma estratégia qualquer. Se X usa uma estratégia DZ associada ao ponto  $(\bar{a}, \bar{b})$ , então, pelo Teorema 10, os pagamentos médios de X e Y satisfazem a  $\bar{\alpha}s_X + \bar{\beta}s_Y = -1$ , que são os pontos da reta denotada por  $E_2$  na Figura 3.12. Mas como quem adota a estratégia DZ associada ao ponto  $(\bar{a}, \bar{b})$  é Y, então o conjunto de todos os pagamentos para todas as possíveis escolhas para X são os pontos da reta  $\bar{b}s_X + \bar{a}s_Y = -1$ , que é a reta  $E'_2$  da mesma figura.

Encontramos portanto os pagamentos para X e Y resolvendo o sistema de equações lineares

$$\begin{cases} \bar{\alpha}s_X + \bar{\beta}s_Y = -1 \\ \bar{b}s_X + \bar{a}s_Y = -1 \end{cases} \quad (3.19)$$

Neste momento convém que nos preocupemos mais com a posição geométrica das retas cujas equações formam o sistema acima. Para isto, vamos nomear os lados do quadrilátero de pagamentos, ver Figura 3.12. Sejam  $\ell_1$  o lado que liga os vértices  $(P, P)$  e  $(S, T)$ ,  $\ell_2$  o que liga  $(S, T)$  a  $(R, R)$ ,  $\ell_3$  o que liga  $(R, R)$  a  $(T, S)$  e, finalmente  $\ell_4$  o que liga  $(T, S)$  a  $(P, P)$ .

Sabemos que se Y joga ALLD, então o ponto correspondente ao par de pagamentos  $(s_X, s_Y)$  deve necessariamente estar sobre o lado  $\ell_1$ . Se Y porém joga ALLC, o ponto correspondente aos pagamentos de X e Y deve estar no lado oposto  $\ell_3$ . Portanto a reta  $E_1$  de equação  $\bar{\alpha}s_X + \bar{\beta}s_Y = -1$  deve ligar um ponto de  $\ell_1$  a um ponto de  $\ell_3$ .

Por um raciocínio análogo, trocando os papéis de X e Y, a reta  $E'_2$  de equação  $\bar{b}s_X + \bar{a}s_Y = -1$  liga  $\ell_2$  a  $\ell_4$ .

Daí, não é difícil enxergar, a menos de dois casos especiais, que as duas retas mencionadas se interceptam em um ponto do interior do quadrilátero, o sistema (3.19) possui solução única e os pagamentos de dois jogadores DZ estão bem definidos por esse sistema.

Os dois casos especiais são quando ambos os jogadores jogam estratégias justas, ou quando ambos jogam estratégias vértice. Em ambos os casos, o sistema (3.19) possui infinitas soluções. Como calcular os pagamentos dos jogadores em tais casos será examinado nos exemplos a seguir.

Resumindo, o que irá acontecer em ambos os casos em que o sistema (3.19) é indeterminado é que a cadeia de Markov para os estados do par X e

Y será redutível e haverá mais de um conjunto terminal. O vetor distribuição estacionária  $v$  para a cadeia será dependente do estado inicial do par de jogadores. Os pagamentos  $s_X$  e  $s_Y$  estarão bem definidos, como veremos, mas dependem da condição inicial do par.

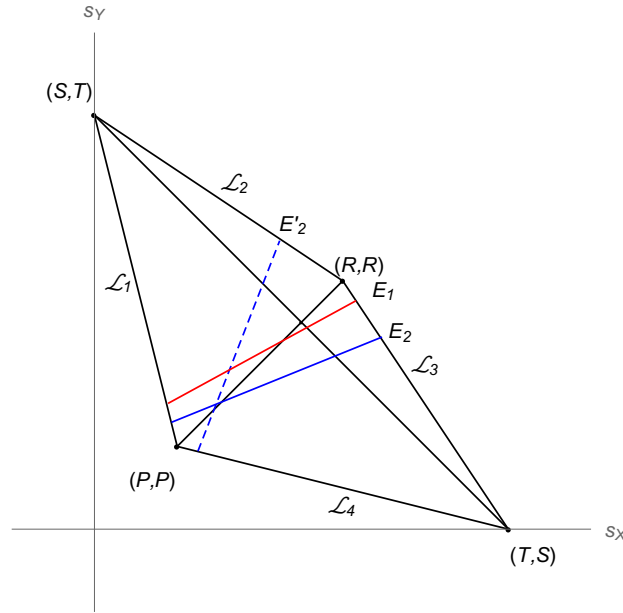


Figura 3.12: Os pagamentos de dois jogadores X e Y que adotem estratégias DZ são as coordenadas do ponto de interseção das retas  $E_1$  e  $E_2'$  explicadas no texto.

**Exemplo 8** *Competição entre dois jogadores que adotam estratégias justas. Pode-se ver pela Figura 3.13 que  $\{CC\}$  e  $\{DD\}$  são conjuntos terminais e, para valores genéricos dos vetores estratégia, são os únicos conjuntos terminais. A estratégia TFT é uma das estratégias justas. Caso X e Y joguem TFT,  $\{CD, DC\}$  também é um conjunto terminal. Como não há unicidade do conjunto terminal, a distribuição estacionária da cadeia de Markov depende da condição inicial. Se o estado inicial for CC ou DD, o par de jogadores permanece no estado inicial por todo o tempo. Os pagamentos  $(s_X, s_Y)$  serão respectivamente  $(R, R)$  e  $(P, P)$ . Agora, para dois jogadores que escolhem TFT, há três comportamentos possíveis: ou ficam em CC todo*

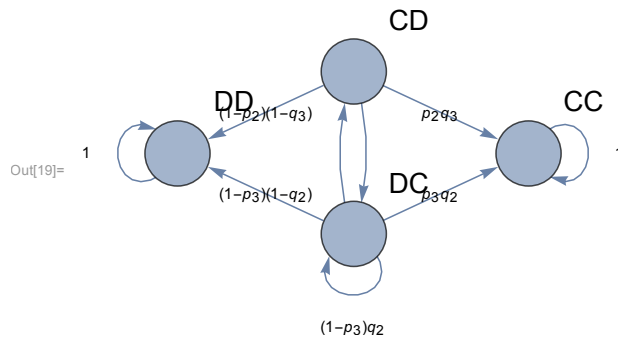


Figura 3.13: Grafo orientado da competição entre jogadores que escolhem estratégia justa.

o tempo, ou ficam em  $DD$  todo o tempo, ou alternam entre  $CD$  e  $DC$  todo o tempo.

**Exemplo 9** *Competição entre dois jogadores que adotam estratégias vértice.* Os vetores são  $p = (p_1, 1, 0, p_4)$  e  $q = (q_1, 1, 0, q_4)$ . Pode-se ver pela Figura 3.14 que os únicos estados recorrentes são  $CD$  e  $DC$ . Existem neste caso, dois conjuntos terminais:  $\{CD\}$  e  $\{DC\}$ . A matriz de transição é redutível e o pagamento médio de  $X$  e  $Y$  depende da condição inicial. Se o estado final for  $CD$ , o pagamento médio será  $(s_X, s_Y) = (S, T)$  e se o estado final for  $DC$  resulta no pagamento médio  $(s_X, s_Y) = (T, S)$ . Este exemplo mostra que os pagamentos dos dois jogadores nunca serão idênticos, ao contrário do que afirma Akin em [1].



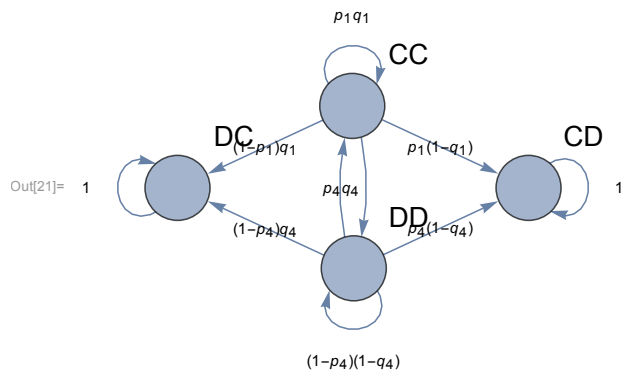


Figura 3.14: Grafo orientado da competição entre jogadores que escolhem estratégia vértice.

# Conclusão

Neste trabalho estudamos fenômenos que envolvem cooperação para entender o surgimento da mesma, apesar de não termos introduzido nenhum tipo de dinâmica evolutiva. A ferramenta da Teoria de Jogos utilizada foi o jogo Dilema do Prisioneiro, no qual acontece uma interação entre dois indivíduos. A cada rodada do jogo os jogadores possuem duas estratégias possíveis: cooperar ou trair. Cada um dos participantes recebe um pagamento por sua escolha. Os vetores de recompensa são descritos por  $S_X = (R, S, T, P)$  e  $S_Y = (R, T, S, P)$ . Pela hierarquia  $T > R > P > S$ , conseguimos perceber que a traição é a melhor escolha quando se joga o DP apenas uma vez.

Para tornar possível a cooperação, o jogo deve ser repetido. Aqui utilizamos uma extensão do DP simples denominada Dilema do Prisioneiro Infinitamente Repetido (DPIR). No DPIR, os dois jogadores têm a certeza de que haverá uma nova rodada do jogo. Em cada rodada do DPIR, o par de jogadores ocupa um dos quatro estados seguintes:  $CC, CD, DC, DD$ . Supomos na maior parte das situações que os jogadores X e Y adotam estratégias de memória um. Os resultados sucessivos deste jogo são então determinados por probabilidades que só dependem da rodada imediatamente anterior, isto é, seguem uma cadeia de Markov. Supomos  $v = (v_1, v_2, v_3, v_4)$  a distribuição limite no sentido de Cesaro da cadeia de Markov. Tal distribuição define um pagamento médio por rodada para os jogadores  $(s_X, s_Y) = (\langle v, S_X \rangle, \langle v, S_Y \rangle)$ .

Uma condição imposta foi  $P < \frac{S+T}{2} < R$ , a qual garante que todos os pares de pagamento possíveis estão contidos em um quadrilátero convexo com vértices  $(S, T), (T, S), (P, P), (R, R)$ .

Uma nova classe de estratégias de memória um para o DPIR, chamadas estratégias de determinante zero (DZ), foram descobertas em 2012 por Press e Dyson [7]. Eles usaram um formalismo arbitrário baseado nas propriedades de determinantes e regra de Cramer. Estas estratégias permitem que o

jogador que as adote tenha um grande controle sobre o resultado do jogo.

Diversos autores renomados ampliaram a análise de Press e Dyson. Por exemplo, Stewart e Plotkin descobriram em [11] as estratégias DZ generosas e forneceram resultados sobre o sucesso dessas estratégias ao repetir o torneio promovido por Axelrod incluindo também estratégias DZ.

Escolhemos estudar [1] e mostrar seus resultados mais interessantes, que situam as estratégias DZ no conjunto de todas as estratégias de memória um. Definimos estratégias boas e concluímos que as estratégias DZ generosas são boas. Pela Figura 3.9, conseguimos perceber que, quando  $Y$  aumenta seu ganho, ele aumenta também o ganho de  $X$ . Mesmo que  $Y$  tenha somente o interesse em aumentar o seu ganho, quando  $X$  adota uma estratégia boa, os jogadores  $X$  e  $Y$  recebem o mesmo pagamento  $R$  de cooperação mútua. Concluímos também que as estratégias justas são boas. Na Figura 3.11, em todos os pontos,  $s_X = s_Y$  e quando  $Y$  maximiza seu pagamento, ele maximiza também o pagamento de  $X$ . Assim,  $X$  e  $Y$  também receberão o pagamento  $R$ . Uma estratégia DZ extorsiva não é boa. Podemos ver na Figura 3.8 que, quando  $Y$  maximiza seu ganho, ele maximiza também o ganho de  $X$  sendo que, em geral,  $X$  ganhará mais do que  $Y$ . O único ponto em que  $s_X = s_Y$  é quando ambos recebem o mesmo pagamento da traição mútua  $P$ . Isto acontece quando  $Y$  minimiza seu ganho. Descrevemos também as estratégias DZ vértice, que não são boas já que  $p_2 = 1$ . O interesse nas estratégias vértice se justifica por sua posição peculiar na fronteira da região das estratégias DZ não excepcionais, ver Figura 3.5. Também porque o confronto de dois jogadores que adotam estratégias vértice exemplifica um caso em que a cadeia de Markov que define os pagamentos dos jogadores possui mais de um conjunto terminal e, conseqüentemente, os pagamentos dependem da condição inicial adotada pelos jogadores.

# Referências Bibliográficas

- [1] Akin Ethan. The Iterated Prisoner's Dilemma: Good Strategies and their Dynamics, arxiv:1211.0969v3 (2017)
- [2] Axelrod Robert. The Evolution of Cooperation, Basic Books (1984)
- [3] Hilbe C, Nowak M, Sigmund K. (2013) The evolution of extortion in iterated Prisoner's Dilemma games, PNAS, 110 no. 17, 6913-6918.
- [4] Minc. H. Y. John Wiley and Sons. Nonnegative Matrices. 1988.
- [5] Neves A. G. M. Aplicações Biológicas de Cadeias de Markov. Notas de minicurso ministrado no III Colóquio de Matemática da Região Norte, Rio de Janeiro, RJ, Brasil, 2014, 93p.
- [6] Nowak Martin A. Evolutionary Dynamics: exploring the equations of life, (2006)
- [7] Press William H; Dyson Freeman J. Iterated Prisoner's Dilemma contains strategies that dominate any evolutionary opponent, P. Natl. Acad. Sci. USA 109, 10409-10413, (2012)
- [8] Rodríguez I. Núñez, Neves A. G. M. Evolution of cooperation in a particular case of the infinitely repeated prisoner's dilemma with three strategies, J. Math. Biol. (2016) 73:1665-1690.
- [9] Ross Sheldon. Probabilidade: um curso moderno com aplicações, 8.ed. Porto Alegre: Bookman, 2010. 608p.
- [10] Sigmund K, Nowak M. Tit for tat in heterogeneous populations. Nature, (355):250-253, (1992).

- [11] Stewart Alexander J; Plotkin Joshua B. Extortion and cooperation in the Prisoner's Dilemma, PNAS, June 26, 2012