STRATEGIES IN THE STOCHASTIC ITERATED PRISONER'S DILEMMA

SIWEI LI

ABSTRACT. The stochastic iterated prisoner's dilemma is a kind of iterated prisoner's dilemma game where the strategies of the players are specified in terms of cooperation probabilities. In particular, if both players use finite-memory strategies, then the game can be modeled by a Markov chain. The purpose of this paper is to discuss several important types of strategies related to such a game, especially, zero-determinant strategies. In addition to a description of the results obtained by William H. Press and Freeman J. Dyson (2012) and Ethan Akin (2013), the paper contains more details and verifications. Moreover, while the main concern is memory-one strategies, the author gives some generalizations in the last section.

CONTENTS

1. A BRIEF REVIEW OF MARKOV PROCESS

In this section we shall give a brief review of Markov process mainly based on [1]. In particular, we focus on the discrete-time Markov chain and emphasize its long run behavior, which will serve as a foundation for the following discussion about the stochastic iterated prisoner's dilemma.

Generally speaking, a **Markov process** $\{X_t\}$ is a stochastic process such that given its current state, the future of the process is conditionally independent of the past history of the process. A **Markov chain** is a Markov process whose state space is a finite or countable set, usually labeled by $\{0, 1, 2, ...\}$. More formally, we give the definition of the discrete-time Markov chain:

Date: DEADLINES: Draft AUGUST 18 and Final version AUGUST 29, 2014.

Definition 1.1. A **discrete-time Markov chain** is a stochastic process with a finite or countable set as state space and with $T = \{0, 1, 2, ...\}$ as time index set such that it satisfies the **Markov property**:

$$\Pr(X_{n+1} = j | X_0 = i_0, ..., X_{n-1} = i_{n-1}, X_n = i) = \Pr(X_{n+1} = j | X_n = i)$$

for all time points n and all possible states $i_0, ..., i_{n-1}, i, j$. Here $\Pr(\cdot)$ is the standard notation for probability.

The conditional probability of X_{n+1} being in state j given X_n being in state i, $\Pr(X_{n+1} = j | X_n = i)$, is called a **one-step transition probability**. If all these one-step transition probabilities are independent of the time variable n, such a Markov chain is called **time-homogeneous** and is said to have **stationary transition probabilities** which are denoted by P_{ij}. We restrict our attention to this kind of Markov chains, since the process analyzed in the following sections is of this type.

Usually, these numbers P_{ij} are arranged in a matrix, called the **one-step transition probability matrix**, or just **transition probability matrix**:

$$\mathbf{P} = \begin{bmatrix} P_{00} & P_{01} & P_{02} & ... \\ P_{10} & P_{11} & P_{12} & ... \\ P_{20} & P_{21} & P_{22} & ... \\ ... & ... & ... & ... \end{bmatrix}.$$

A Markov process is completely determined by its transition probability matrix and the initial state/initial distribution.

Similar to the one-step transition probabilities, we also have **n-step transition probabilities**, $\Pr(X_{m+n} = j | X_m = i)$. For stationary Markov chains, these probabilities are denoted by $P_{ij}^{(n)}$. Putting all of them into a matrix, one can get a **n-step transition probability matrix** $\mathbf{P}^{(n)} = [P_{ij}^{(n)}]$. There is an important relation between the one-step transition probability matrix and the n-step transition probability matrix: $\mathbf{P}^{(n)} = \underbrace{\mathbf{P} \times \mathbf{P} \times ... \times \mathbf{P}}_{n\,times} = \mathbf{P}^n$.

Next we shall present some results about the long run behavior of Markov chains.

Definition 1.2. A transition probability matrix, or the corresponding Markov chain, is called **regular** if there is some positive integer k such that the transition matrix $\mathbf{P}$ to the power of k, $\mathbf{P}^k$, has all its elements strictly positive.

Theorem 1.3. *Let* $\mathbf{P}$ *be a regular transition probability matrix on a finite set of states* $\{0, 1, 2, ..., N\}$*. Then there exists a* ***limiting distribution*** $(\pi_0, \pi_1, ..., \pi_N)$ *such that for any states i and j, as* $n \to \infty$*,*

$$P_{ij}^{(n)} \to \pi_j > 0, \text{and, therefore, } \Pr(X_n = j) \to \pi_j > 0,$$

where $(\pi_0, \pi_1, ..., \pi_N)$ *is the unique solution to*

$$\begin{cases} \pi_j = \sum_{k=0}^{N} \pi_k P_{kj}, j = 0, 1, ..., N, \\ \sum_{k=0}^{N} \pi_k = 1. \end{cases}$$

If we let $\boldsymbol{\pi} = (\pi_0, \pi_1, ..., \pi_N)$*, then the system of the first* $N + 1$ *equations above is just equivalent to* $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}$*. Combining with* $\boldsymbol{\pi} \cdot \mathbf{1} = 1$*, where* $\mathbf{1} = (1, 1, ..., 1)$*, one can get the unique* $\boldsymbol{\pi}$*, which has all its entries positive.*

We omit here the proof of the above theorem due to the limited space. There are also other equivalent ways to state the theorem. The associated proofs can be found in many materials for the stochastic process, such as [2], [3] and [4].

In some other situations, there can be no limiting distribution but there exist so-called stationary probability distributions.

Definition 1.4. Let $S = \{0, 1, 2, ...\}$ be the state space of the Markov chain and $\mathbf{P}$ be the corresponding matrix. The row vector $\mathbf{v}$ is called a **stationary probability distribution** if it satisfies

$$\mathbf{v} = \mathbf{v}\mathbf{P}; \quad \mathbf{v} \cdot \mathbf{1} = 1; \quad 0 \leq v_j \leq 1 \quad \forall j \in S.$$

Remark 1.5. Given any initial distribution $\mathbf{v}^{(0)}$ as a row vector, then $\mathbf{v}^{(n+1)} = \mathbf{v}^{(n)}\mathbf{P}$ for $n = 0, 1, 2, ...$, where $\mathbf{v}^{(k)}$ is a row vector standing for the distribution over states after k transitions. The distribution $\mathbf{v}$ in Definition 1.4 is named stationary distribution because if $\Pr(X_0 = i) = v_i$ for all $i \in S$, then $\Pr(X_n = i) = v_i$ for all $n = 0, 1, 2, ...$ and $i \in S$. This can be easily seen by noting that for $n = 1$:

$$\Pr(X_1 = i) = \sum_{k=0}^{\infty} \Pr(X_0 = k) \Pr(X_1 = i | X_0 = k) = \sum_{k=0}^{\infty} v_k P_{ki} = v_i.$$

Moreover, if the limiting distribution exists, then it is the unique stationary distribution. Otherwise there can be more than one stationary distribution. For finite-state Markov chains it turns out that there always exists some stationary distribution $\mathbf{v}$. One proof is given by the Krylov-Bogoliubov Argument, as [4] suggests. The rough idea is as follows. Note that probability simplex is compact and thus any sequence in it has a convergent subsequence, by the Bolzano-Weierstrass property. If we consider the sequence of the so-called Cesaro averages $\{\mathbf{a}_n\}$, where $\mathbf{a}_n = \frac{1}{n}\sum_{k=1}^{n} \mathbf{v}^{(k)}$, then it must have some subsequence $\{\mathbf{a}_{n_k}\}$ that converges to some $\mathbf{v}$. Such $\mathbf{v}$ can be proved to be a stationary distribution.

2. Derivation of zero-determinant strategies

In game theory, there is a typical model, called prisoner's dilemma (PD), which gives a mathematical description of many situations in real life. Prisoner's dilemma is a two-person general-sum game with following rules. Two players, X and Y, choose their actions simultaneously without knowing the other's choice. Two pure strategies, cooperation (C) and defection (D), and any randomized strategies over these two pure ones are the available strategies for both players. If both cooperate, each will earn a reward (R). If one defects and the other cooperates, the defector will get a larger temptation payoff (T) while the cooperator will get a smaller sucker's payoff (S). If both defect, each will gain a punishment payoff (P). For prisoner's dilemma, the values of R,T,S and P should satisfy two conditions: $T > R > P > S$ and $2R > T + S$. The former guarantees that mutual defection is the only Nash Equilibrium for the one-shot prisoner's dilemma and the latter ensures mutual cooperation to be globally optimal. Conventionally, $(T, R, P, S) = (5, 3, 1, 0)$.

If the game is played only once or only a fixed number of times, there is no room for the emergence of cooperation, but this is not the case when the game is played infinitely. This is one reason why (infinitely) iterated prisoner's dilemma (IPD), in which the same two players meet again and again, has drawn so much attention. There are many beautiful results related to this game. Here we introduce a newly revealed type of strategy, called **zero-determinant strategies**, discovered by William H. Press and Freeman J. Dyson [5]. This type of strategy overturns the traditional belief that there is no ultimatum strategy for a player to unilaterally claim an unfair share of rewards. In their original paper [5], they presented several impressive results with brief explanations. Here we shall follow their derivation and supplement more details.

Before going on, we clearly state the following assumptions that we make for the game:

(1) What a player cares about is only the long run expected payoff per round.
(2) Each player can only remember the outcome of the single previous round and thus uses **memory-one strategies**. This implies that the repeated prisoner's dilemma game follows a four-state Markov chain. (In the Appendix A of [5], it is shown that we can make this assumption without loss of generality in some sense. Moreover, the game can be described as a finite-state Markov chain as long as both players choose strategies based on finite memory, as discussed in Section §6.)

In each single round there are four possible outcomes, $xy \in \{CC, CD, DC, DD\}$, where x and y represent X's and Y's choices respectively and C and D denote cooperation and defection. Then we can represent a memory-one strategy by a four tuple specifying the probability of cooperation in the current round given the outcome of the previous move: X's strategy is $\mathbf{p} = (p_1, p_2, p_3, p_4)$ corresponding to the previous outcome $xy \in \{CC, CD, DC, DD\}$; Y's strategy, from his viewpoint, is $\mathbf{q} = (q_1, q_2, q_3, q_4)$ given the previous outcome $yx \in \{CC, CD, DC, DD\}$. Note that p_2 and q_3 do not correspond to the same previous outcome as X and Y look at the game from their own views. (In fact, if we instead treat the payoff that one gets in a round of the game as its outcome, the indices $1, 2, 3, 4$ in one's strategy vector just correspond to his own previous payoffs R, S, T, P, which is possibly a more consistent description.) This way of specifying strategies allows us to treat the game as a stochastic process, and we call it **stochastic iterated prisoner's dilemma**. Furthermore, as already mentioned, this process is Markov. In the following we shall describe the game from X's point of view. The state space of the Markov chain is $\{CC, CD, DC, DD\}$ with the first letter representing X's choice.

With all these at hand and keeping in mind that two players move simultaneously in each round, we can then calculate the 16 transition probabilities and write out the transition probability matrix for the game, denoted by $\mathbf{P}(\mathbf{p}, \mathbf{q})$ since it is fully determined by the two player's strategies, $\mathbf{p}$ and $\mathbf{q}$. For instance, the probability of getting an outcome $xy = CD$ given the previous outcome $xy = CD$ equals $p_2(1 - q_3)$. Then $\mathbf{P}(\mathbf{p}, \mathbf{q})$ equals:

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

Our derivation in this section is for the case where there is a unique limiting distribution for the Markov matrix $\mathbf{P}(\mathbf{p}, \mathbf{q})$ defined above. Then the long run expected payoff per round is just the payoff in the Markov limiting state. When the Markov matrix $\mathbf{P}(\mathbf{p}, \mathbf{q})$ has no unique limiting distribution, we can also give similar results (see Section §5).

According to Theorem 1.3, the limiting distribution row vector $\boldsymbol{\pi}$ of the Markov matrix $\mathbf{P} = \mathbf{P}(\mathbf{p}, \mathbf{q})$ or any nonzero multiple $\mathbf{w}$ of $\boldsymbol{\pi}$ (in the following, we may call $\mathbf{w}$ the limiting vector) should satisfy

$$\mathbf{wP} = \mathbf{w}.$$

If we let $\boldsymbol{\mathcal{L}} := \mathbf{P} - \mathbf{I}_4$ where $\mathbf{I}_n$ is the $n \times n$ identity matrix, then the above equation is equivalent to

$$\mathbf{w}\boldsymbol{\mathcal{L}} = \mathbf{0}, \text{ where } \mathbf{0} = (0, 0, 0, 0).$$

Since $\mathbf{w}\boldsymbol{\mathcal{L}} = \mathbf{0}$ has nonzero solution $\mathbf{w}$, $\boldsymbol{\mathcal{L}}$ is singular. That is, $\boldsymbol{\mathcal{L}}$ has zero determinant: $\det(\boldsymbol{\mathcal{L}}) = 0$.

Recall Cramer's rule:

Proposition 2.1. *Let* $\mathbf{A}$ *be an* $n \times n$ *matrix and* $\mathbf{I}_n$ *be the* $n \times n$ *identity matrix. Let* $\mathrm{Adj}(\mathbf{A})$ *denote the adjugate matrix of* $\mathbf{A}$*, defined to be the transpose of the cofactor matrix of* $\mathbf{A}$*. Then* $\mathrm{Adj}(\mathbf{A})\mathbf{A} = \det(\mathbf{A})\mathbf{I}_n$*.*

Applying Cramer's rule to $\boldsymbol{\mathcal{L}}$, along with $\det(\boldsymbol{\mathcal{L}}) = 0$, we get

$$\mathrm{Adj}(\boldsymbol{\mathcal{L}})\boldsymbol{\mathcal{L}} = \det(\boldsymbol{\mathcal{L}})\mathbf{I}_4 = \mathbf{0}_{4,4},$$

where $\mathbf{0_{n,n}}$ is the $n \times n$ zero matrix. If we let $\mathcal{L}_{ij}$ represent the (i, j) cofactor of $\boldsymbol{\mathcal{L}}$, then $\mathrm{Adj}(\boldsymbol{\mathcal{L}})$ can be expressed explicitly:

$$\mathrm{Adj}(\boldsymbol{\mathcal{L}}) = \begin{bmatrix} \mathcal{L}_{11} & \mathcal{L}_{21} & \mathcal{L}_{31} & \mathcal{L}_{41} \\ \mathcal{L}_{12} & \mathcal{L}_{22} & \mathcal{L}_{32} & \mathcal{L}_{42} \\ \mathcal{L}_{13} & \mathcal{L}_{23} & \mathcal{L}_{33} & \mathcal{L}_{43} \\ \mathcal{L}_{14} & \mathcal{L}_{24} & \mathcal{L}_{34} & \mathcal{L}_{44} \end{bmatrix}.$$

For the Markov chain with a unique limiting distribution, the solution to the equation $\mathbf{w}\boldsymbol{\mathcal{L}} = \mathbf{0}$ (or $\mathbf{wP} = \mathbf{w}$) is unique up to a scalar factor. That is to say, the solution space of $\mathbf{w}\boldsymbol{\mathcal{L}} = \mathbf{0}$ is one dimensional. Since $\boldsymbol{\mathcal{L}}$ is a 4×4 matrix, we have $\mathrm{rank}(\boldsymbol{\mathcal{L}}) = 3$. Therefore, $\mathrm{Adj}(\boldsymbol{\mathcal{L}})$ must be a nonzero matrix. Without loss of generality (See Remark 2.9), we may assume that the fourth row $(\mathcal{L}_{14}, \mathcal{L}_{24}, \mathcal{L}_{34}, \mathcal{L}_{44})$ of $\mathrm{Adj}(\boldsymbol{\mathcal{L}})$ is nonzero.

Moreover, since $\mathbf{w}\boldsymbol{\mathcal{L}} = \mathbf{0}$ and $\mathrm{Adj}(\boldsymbol{\mathcal{L}})\boldsymbol{\mathcal{L}} = \mathbf{0}_{4,4}$, noting that $\mathbf{w}\boldsymbol{\mathcal{L}} = \mathbf{0}$ has one-dimensional solution space, we can conclude that every row of $\mathrm{Adj}(\boldsymbol{\mathcal{L}})$ must be proportional to $\mathbf{w}$. Thus, $\mathbf{w} = \eta(\mathcal{L}_{14}, \mathcal{L}_{24}, \mathcal{L}_{34}, \mathcal{L}_{44})$ for some $\eta \neq 0$.

Now we are ready to figure out a formula for the dot product of the limiting vector $\mathbf{w}$ with an arbitrary four-dimensional vector $\mathbf{f} = (f_1, f_2, f_3, f_4)$. By the definition of dot product, we know:

$$\mathbf{w} \cdot \mathbf{f} = \eta(\mathcal{L}_{14}f_1 + \mathcal{L}_{24}f_2 + \mathcal{L}_{34}f_3 + \mathcal{L}_{44}f_4).$$

Now We can thus transform $\boldsymbol{\mathcal{L}}$ to $\boldsymbol{\mathcal{L}'}$ by adding column one to column two and column three respectively:

$$\begin{aligned}\boldsymbol{\mathcal{L}} &:= \mathbf{P} - \mathbf{I}_4 \\ &= \begin{bmatrix} p_1q_1 - 1 & p_1(1-q_1) & (1-p_1)q_1 & (1-p_1)(1-q_1) \\ p_2q_3 & p_2(1-q_3)-1 & (1-p_2)q_3 & (1-p_2)(1-q_3) \\ p_3q_2 & p_3(1-q_2) & (1-p_3)q_2 - 1 & (1-p_3)(1-q_2) \\ p_4q_4 & p_4(1-q_4) & (1-p_4)q_4 & (1-p_4)(1-q_4)-1 \end{bmatrix} \\ &\Rightarrow \begin{bmatrix} p_1q_1 - 1 & p_1 - 1 & (1-p_1)q_1 & (1-p_1)(1-q_1) \\ p_2q_3 & p_2 - 1 & (1-p_2)q_3 & (1-p_2)(1-q_3) \\ p_3q_2 & p_3 & (1-p_3)q_2 - 1 & (1-p_3)(1-q_2) \\ p_4q_4 & p_4 & (1-p_4)q_4 & (1-p_4)(1-q_4)-1 \end{bmatrix} \\ &\Rightarrow \begin{bmatrix} p_1q_1 - 1 & p_1 - 1 & q_1 - 1 & (1-p_1)(1-q_1) \\ p_2q_3 & p_2 - 1 & q_3 & (1-p_2)(1-q_3) \\ p_3q_2 & p_3 & q_2 - 1 & (1-p_3)(1-q_2) \\ p_4q_4 & p_4 & q_4 & (1-p_4)(1-q_4)-1 \end{bmatrix} = \boldsymbol{\mathcal{L}'}.\end{aligned}$$

Let $\mathcal{L}'_{ij}$ represent the (i,j) cofactor of $\boldsymbol{\mathcal{L}'}$. Recall that for any matrix, if a multiple of a column is added to another column, the determinant of the matrix remains the same. Note that only the first three columns of $\boldsymbol{\mathcal{L}}$ have been manipulated, so it is easy to see $\mathcal{L}_{i4} = \mathcal{L}'_{i4}$ for $i = 1,2,3,4$. If we replace the last column of $\boldsymbol{\mathcal{L}'}$ by the transpose of the arbitrary four-dimensional vector $\mathbf{f} = (f_1, f_2, f_3, f_4)$ and then compute the determinant of the corresponding matrix by expanding along the fourth column, a relation between this determinant and the value of $\mathbf{w} \cdot \mathbf{f}$ will show up:

$$\begin{aligned}\det \begin{bmatrix} p_1q_1 - 1 & p_1 - 1 & q_1 - 1 & f_1 \\ p_2q_3 & p_2 - 1 & q_3 & f_2 \\ p_3q_2 & p_3 & q_2 - 1 & f_3 \\ p_4q_4 & p_4 & q_4 & f_4 \end{bmatrix} &= f_1\mathcal{L}'_{14} + f_2\mathcal{L}'_{24} + f_3\mathcal{L}'_{34} + f_2\mathcal{L}'_{44} \\ &= \mathcal{L}_{14}f_1 + \mathcal{L}_{24}f_2 + \mathcal{L}_{34}f_3 + \mathcal{L}_{44}f_4 = \frac{1}{\eta}(\mathbf{w} \cdot \mathbf{f}).\end{aligned}$$

The formula derived above is summarized as a proposition:

Proposition 2.2. *Suppose* $\mathbf{P}$ *is the transition matrix for a Markov chain with limiting distribution. Assume the fourth row of* $\mathrm{Adj}(\boldsymbol{\mathcal{L}})$ *is nonzero. Let* $\mathbf{w}$ *be the limiting row vector, possibly without normalization. That is,* $\mathbf{w}$ *is a vector satisfying* $\mathbf{wP} = \mathbf{w}$*. Then for any four-dimensional vector* $\mathbf{f} = (f_1, f_2, f_3, f_4)$*, there exists some* $\eta \neq 0$ *that depends on* $\mathbf{w}$ *only, such that the following equation holds:*

$$\mathbf{w} \cdot \mathbf{f} = \eta D(\mathbf{p}, \mathbf{q}, \mathbf{f}), \text{where } D(\mathbf{p}, \mathbf{q}, \mathbf{f}) := \det \begin{bmatrix} p_1q_1 - 1 & p_1 - 1 & q_1 - 1 & f_1 \\ p_2q_3 & p_2 - 1 & q_3 & f_2 \\ p_3q_2 & p_3 & q_2 - 1 & f_3 \\ p_4q_4 & p_4 & q_4 & f_4 \end{bmatrix}.$$

This result is very interesting because the determinant $D(\mathbf{p}, \mathbf{q}, \mathbf{f})$ has a peculiar feature: its second column $(p_1 - 1, p_2 - 1, p_3, p_4)^T$ only depends on X's strategy $\mathbf{p} = (p_1, p_2, p_3, p_4)$; its third column $(q_1 - 1, q_3, q_2 - 1, q_4)^T$ only depends on Y's strategy $\mathbf{q} = (q_1, q_2, q_3, q_4)$. For convenience, we denote the transposes of the second column and third column by $\widetilde{\mathbf{p}}$ and $\widehat{\mathbf{q}}$, respectively:

$$\widetilde{\mathbf{p}} \coloneqq (p_1 - 1, p_2 - 1, p_3, p_4), \quad \widehat{\mathbf{q}} \coloneqq (q_1 - 1, q_3, q_2 - 1, q_4).$$

Since the exact limiting distribution vector $\boldsymbol{\pi}$ has its components sum to 1, $\boldsymbol{\pi} = \frac{\mathbf{w}}{\mathbf{w}\cdot\mathbf{1}}$. Recall that we describe the state space of the process as $xy \in \{CC, CD, DC, DD\}$. The corresponding payoff vectors for X and Y are $\mathbf{S_X} = (R, S, T, P)$ and $\mathbf{S_Y} = (R, T, S, P)$, respectively. Therefore the expected payoff per round under the limiting distribution of the Markov chain, s_X and s_Y, are:

$$s_X = \boldsymbol{\pi} \cdot \mathbf{S_X} = \frac{\mathbf{w}}{\mathbf{w}\cdot\mathbf{1}} \cdot \mathbf{S_X} = \frac{\mathbf{w}\cdot\mathbf{S_X}}{\mathbf{w}\cdot\mathbf{1}} = \frac{\eta D(\mathbf{p},\mathbf{q},\mathbf{S_X})}{\eta D(\mathbf{p},\mathbf{q},\mathbf{1})} = \frac{D(\mathbf{p},\mathbf{q},\mathbf{S_X})}{D(\mathbf{p},\mathbf{q},\mathbf{1})}, \tag{2.3}$$

$$s_Y = \boldsymbol{\pi} \cdot \mathbf{S_Y} = \frac{\mathbf{w}}{\mathbf{w}\cdot\mathbf{1}} \cdot \mathbf{S_Y} = \frac{\mathbf{w}\cdot\mathbf{S_Y}}{\mathbf{w}\cdot\mathbf{1}} = \frac{\eta D(\mathbf{p},\mathbf{q},\mathbf{S_Y})}{\eta D(\mathbf{p},\mathbf{q},\mathbf{1})} = \frac{D(\mathbf{p},\mathbf{q},\mathbf{S_Y})}{D(\mathbf{p},\mathbf{q},\mathbf{1})}, \tag{2.4}$$

where $\mathbf{1} = (1,1,1,1)$. Note that the denominator, $D(\mathbf{p},\mathbf{q},\mathbf{1})$, in the above expressions cannot be zero under the assumption that the Markov chain has a unique limiting distribution. This can be seen by recalling that $\mathbf{w}$ is a nonzero multiple of the limiting distribution $\boldsymbol{\pi}$ which has all entries non-negative with sum one and $D(\mathbf{p},\mathbf{q},\mathbf{1})$ equals to $\mathbf{w}\cdot\mathbf{1}$ which is just the sum of all entries of $\mathbf{w}$.

Also notice that the expected payoff for a player (Equations (2.3) and (2.4)) depends linearly on his/her own payoff vector. As a consequence, we can perform linear combination of the expected payoffs of the two players:

$$\alpha s_X + \beta s_Y + \gamma = \frac{D(\mathbf{p},\mathbf{q},\alpha\mathbf{S_X} + \beta\mathbf{S_Y} + \gamma\mathbf{1})}{D(\mathbf{p},\mathbf{q},\mathbf{1})}. \tag{2.5}$$

Equation (2.5) is the key to establish the following crucial proposition:

Proposition 2.6. *It is possible for both players to choose some strategies to unilaterally enforce a linear relation between their expected payoffs. More precisely, for some (not all) values of α, β and γ, if X can choose a strategy satisfying $\widetilde{\mathbf{p}} = (p_1 - 1, p_2 - 1, p_3, p_4) = \alpha\mathbf{S_X} + \beta\mathbf{S_Y} + \gamma\mathbf{1}$, or if Y can choose a strategy satisfying $\widehat{\mathbf{q}} = (q_1 - 1, q_3, q_2 - 1, q_4) = \alpha\mathbf{S_X} + \beta\mathbf{S_Y} + \gamma\mathbf{1}$, then regardless of the opponent's strategy, a linear relation between the expected payoffs of the two players in the limiting state will hold:*

$$\alpha s_X + \beta s_Y + \gamma = 0. \tag{2.7}$$

Proof. Based on what we have derived, the proof is straightforward. Recall that if a matrix has two columns identical or proportional, its determinant is zero. Also recall that $D(\mathbf{p},\mathbf{q},\alpha\mathbf{S_X} + \beta\mathbf{S_Y} + \gamma\mathbf{1})$ has its second column fully controlled by X. So if X chooses a satisfying $\widetilde{\mathbf{p}} = (p_1 - 1, p_2 - 1, p_3, p_4) = \alpha\mathbf{S_X} + \beta\mathbf{S_Y} + \gamma\mathbf{1}$, then the second and the fourth column of $D(\mathbf{p},\mathbf{q},\alpha\mathbf{S_X} + \beta\mathbf{S_Y} + \gamma\mathbf{1})$ will be identical and thus $D(\mathbf{p},\mathbf{q},\alpha\mathbf{S_X} + \beta\mathbf{S_Y} + \gamma\mathbf{1}) = 0$, no matter what the values of the other two columns are. This implies $\alpha s_X + \beta s_Y + \gamma = \frac{D(\mathbf{p},\mathbf{q},\alpha\mathbf{S_X} + \beta\mathbf{S_Y} + \gamma\mathbf{1})}{D(\mathbf{p},\mathbf{q},\mathbf{1})} = 0$. X can enforce this relation without considering Y's strategy. A similar argument can be used to prove that Y has such a power as well by choosing a strategy satisfying $\widehat{\mathbf{q}} = (q_1 - 1, q_3, q_2 - 1, q_4) = \alpha\mathbf{S_X} + \beta\mathbf{S_Y} + \gamma\mathbf{1}$. □

Definition 2.8. Zero-determinant (ZD) strategies refer to the strategies described in Proposition 2.6.

Remark 2.9. If we assume some other row, instead of the fourth row, of $\mathrm{Adj}(\boldsymbol{\mathcal{L}})$ is nonzero and therefore use that row to express $\mathbf{w}$, we can get similar results as Proposition 2.2 with the only possible difference being the exact expression of $D(\mathbf{p}, \mathbf{q}, \mathbf{f})$. But we can still have two columns be $\widetilde{\mathbf{p}}^T$ and $\widehat{\mathbf{q}}^T$ defined above in the expression of $D(\mathbf{p}, \mathbf{q}, \mathbf{f})$. Equations (2.3) and (2.4) also hold and we can still define the same zero-determinant strategies as in Proposition 2.6.

Notice that to be feasible, strategies $\mathbf{p}$ and $\mathbf{q}$ of X and Y respectively should have all components between zero and one, which means that it could be the case that for some α, β and γ there is no zero-determinant strategy feasible for X and/or Y. Even so, there indeed exist feasible zero-determinant strategies. For instance, if we take the conventional values $(T, R, P, S) = (5, 3, 1, 0)$, then X can unilaterally enforce a linear relation $\frac{1}{16}s_X - \frac{1}{8}s_Y + \frac{1}{8} = 0$ by adopting the memory-one zero-determinant strategy $\mathbf{p} = (\frac{15}{16}, \frac{1}{2}, \frac{7}{16}, \frac{1}{16})$. Note that the linear relation $\frac{1}{16}s_X - \frac{1}{8}s_Y + \frac{1}{8} = 0$ is equivalent to $s_X - 2s_Y + 2 = 0$.

Moreover, we can use the same argument to derive zero-determinant strategies for any iterated 2×2 games: suppose X has two strategies X_1, X_2 and Y has two strategies Y_1, Y_2. X and Y may have different pure strategies. Then for a single round, there are four possible outcomes $\{X_1Y_1, X_2Y_1, X_2Y_1, X_2Y_2\}$. Suppose the corresponding payoff vectors for X and Y are $\mathbf{S_X} = (x_1, x_2, x_3, x_4)$ and $\mathbf{S_Y} = (y_1, y_2, y_3, y_4)$, respectively, then their expected payoffs in the limiting state will have exactly the same expressions as Equations (2.3) and (2.4). Equation (2.5) and Proposition 2.6 also hold without any changes.

In the following we still limit our attention to iterated prisoner's dilemma.

3. Discussion on special kinds of zero-determinant strategies

In this section we shall study some special kinds of zero-determinant strategies derived in Section §2 more carefully, as [5] did. There are mainly two key applications of zero-determinant strategies: (1) a player using zero-determinant strategies can unilaterally set the opponent's expected payoff to a value within a certain range; (2) a player using zero-determinant strategies can demand a larger share of the total payoffs over the mutual defection value. In this case the zero-determinant strategies are called, according to Press and Dyson [5], **extortionate strategies**.

We assume that X knows the zero-determinant strategies and our analysis is from X's point of view.

3.1. **Determine the opponent's expected payoff unilaterally.** According to Proposition 2.6, a player using zero-determinant strategy can unilaterally enforce a linear relation between the expected payoffs, as shown in Equation (2.7). In particular, we can take $\alpha = 0$ in Equation (2.7). More specifically, if X uses a strategy such that $\widetilde{\mathbf{p}} = \beta\mathbf{S_Y} + \gamma\mathbf{1}$, then $\beta s_Y + \gamma = 0$, i.e. $s_Y = -\frac{\gamma}{\beta}$, which is absolutely determined by X's strategy $\mathbf{p} = (p_1, p_2, p_3, p_4)$ and independent of Y's strategy $\mathbf{q} = (q_1, q_2, q_3, q_4)$. We can also solve the desired strategy explicitly:

$$\begin{aligned}
&\widetilde{\mathbf{p}} \equiv (p_1-1, p_2-1, p_3, p_4) = \beta \mathbf{S_Y} + \gamma \mathbf{1},\\
\Leftrightarrow &(p_1-1, p_2-1, p_3, p_4) = \beta(R,T,S,P) + \gamma(1,1,1,1),\\
\Rightarrow &\beta = \frac{-(1-p_1)-p_4}{R-P}, \quad \gamma = \frac{(1-p_1)P + p_4 R}{R-P},\\
\Rightarrow &p_2 = \frac{p_1(T-P)-(1+p_4)(T-R)}{R-P}, \quad p_3 = \frac{(1-p_1)(P-S)+p_4(R-S)}{R-P}.
\end{aligned}$$

Therefore, the desired strategy is

$$\mathbf{p} = \left(p_1, \frac{p_1(T-P)-(1+p_4)(T-R)}{R-P}, \frac{(1-p_1)(P-S)+p_4(R-S)}{R-P}, p_4\right). \tag{3.1}$$

This strategy is feasible if and only if $p_i \in [0,1]$ for all $i = 1,2,3,4$. By solving four inequalities $0 \le p_i \le 1$ simultaneously, we see that the strategy in (3.1) is feasible if and only if

$$\begin{aligned}
p_1 &\in \left[\max\left\{\frac{T-R}{T-P}, 1-\frac{R-P}{P-S}\right\}, 1\right],\\
p_4 &\in \left[0, \min\left\{\frac{T-P}{T-R}p_1 - 1, 1-\frac{(2-p_1)(P-S)}{R-S}\right\}\right].
\end{aligned}$$

Intuitively, for the strategy in (3.1) to be feasible, p_1 should be large and p_4 should be small and thus p_2 is large and p_3 is small.

Under X's strategy in (3.1), Y's expected payoff is given by

$$s_Y = -\frac{\gamma}{\beta} = \frac{(1-p_1)P + p_4 R}{(1-p_1)+p_4}.$$

In this case, s_Y is actually a weighted average of the mutual-defection payoff (P) and the mutual-cooperation payoff (R). Since $0 \le p_1, p_4 \le 1$, the range of values of Y's expected payoff that can be determined unilaterally by X is $[P, R]$. If $p_1 = 1, p_4 \neq 0$, then $s_Y = R$. If $p_4 = 0, p_1 \neq 1$, then $s_Y = P$. For other values of p_1 and p_4 such that $\mathbf{p}$ is feasible, Y's score will be between P and R. Note that we must have $p_1 \neq 1$ or $p_4 \neq 0$ to calculate s_Y. $p_1 = 1$ and $p_4 = 0$ in (3.1) imply $\mathbf{p} = (1,1,0,0)$ which is not a valid zero-determinant strategy and will be discussed in a moment.

3.2. **Unable to set one's own payoff.** It seems that X should have the ability to unilaterally set his/her own long run payoff as well by using zero-determinant strategies. Indeed, if we set $\beta = 0$ in Equation (2.7), then X's expected payoff is $s_X = -\frac{\gamma}{\alpha}$. The desired strategy should satisfy $\widetilde{\mathbf{p}} = \alpha \mathbf{S_X} + \gamma \mathbf{1}$:

$$\begin{aligned}
&\widetilde{\mathbf{p}} = (p_1-1, p_2-1, p_3, p_4) = \alpha \mathbf{S_X} + \gamma \mathbf{1},\\
\Leftrightarrow &(p_1-1, p_2-1, p_3, p_4) = \alpha(R,S,T,P) + \gamma(1,1,1,1),\\
\Rightarrow &\alpha = \frac{-(1-p_1)-p_4}{R-P}, \quad \gamma = \frac{(1-p_1)P + p_4 R}{R-P},\\
\Rightarrow &p_2 = \frac{(1+p_4)(R-S)-p_1(P-S)}{R-P}, \quad p_3 = \frac{-(1-p_1)(T-P)-p_4(T-R)}{R-P}.
\end{aligned}$$

It is obvious that $p_3 \le 0$. Also, since $(1+p_4)(R-S) - p_1(P-S) - (R-P) = (1-p_1)(P-S) + p_4(R-S) \ge 0$, we know $p_2 \ge 1$. Hence, there is only one feasible solution in this case, i.e. $\mathbf{p} = (p_1, p_2, p_3, p_4) = (1,1,0,0)$. If X uses this strategy, then the second column of $\boldsymbol{\mathcal{L}'}$ has all elements to be zero and therefore $D(\mathbf{p}, \mathbf{q}, \mathbf{f}) = 0$

for any vector $\mathbf{f}$, so the denominators of Equations (2.3) and (2.4) are zero and hence we cannot use Proposition 2.6 directly in this case, In essence, this is because there is no unique limiting distribution for the game if $\mathbf{p} = (p_1, p_2, p_3, p_4) = (1, 1, 0, 0)$. On the other hand, using this strategy means that X will either always cooperate or always defect. As a result, X cannot control his/her long run payoff because it is influenced by Y's strategy.

3.3. **Demand an extortionate share.** By adopting zero-determinant strategies, it is also possible for X to get a pre-specified extortionate share of the total amount above the mutual defection payoffs in the long run, as we now elucidate.

We can express the linear relation (2.7) using another form:

$$s_X - P = \chi(s_Y - P), \tag{3.2}$$

where $\chi \geq 1$ is called **extortion factor**. We may also multiply both sides by another nonzero parameter, ϕ, which is used to ensure the feasibility of the strategy, and move the right hand side to the left:

$$\phi[s_X - P - \chi(s_Y - P)] = 0.$$

To enforce such a linear relation, by Proposition 2.6, what X needs to do is just choose a strategy satisfying

$$\widetilde{\mathbf{p}} = \phi[\mathbf{S_X} - P\mathbf{1} - \chi(\mathbf{S_Y} - P\mathbf{1})]. \tag{3.3}$$

That is,

$$p_1 = 1 - \phi(\chi - 1)(R - P), \quad p_2 = 1 - \phi[(P - S) + \chi(T - P)], \tag{3.4}$$

$$p_3 = \phi[(T - P) + \chi(P - S)],\ p_4 = 0. \tag{3.5}$$

Such a strategy is feasible for any χ and sufficiently small ϕ. Given the value of χ,

$$\begin{aligned} & 0 \leq p_i \leq 1, \quad i = 1, 2, 3, 4, \\ \Leftrightarrow \quad & 0 < \phi \leq \min\left\{ \frac{1}{(P - S) + \chi(T - P)}, \frac{1}{\chi(P - S) + (T - P)} \right\}. \end{aligned}$$

Using this strategy, called **extortionate strategy**, X is able to get an extortionate share, but the absolute amount of payoff for X still depends on Y's strategy. Since the payoffs for X and Y are subject to the linear relation (3.2), it is clear that X's and Y's payoff will be maximized at the same time. If Y always cooperates, that is, if Y uses the strategy $\mathbf{q} = (1, 1, 1, 1)$, then in each round of the game X can get more, compared with the case where Y has a possibility to defect.(In fact, no matter what strategy X uses, he can get the maximum payoff achievable by his strategy if he is facing a "silly" player who cooperates all the time unconditionally.) In other words, if X uses the extortionate strategy, then Y can maximize his own payoff by cooperating all the time, which also maximizes X's payoff and even gives X more than Y himself. To get an expression for this maximum payoff of X, we can use the formula obtained in Section §2. Note that with $\widetilde{\mathbf{p}} = \phi[\mathbf{S_X} - P\mathbf{1} - \chi(\mathbf{S_Y} - P\mathbf{1})]$ and $\mathbf{q} = (1, 1, 1, 1)$, we have:

$$\begin{aligned} D(\mathbf{p}, \mathbf{q}, \mathbf{S_X}) &= -T\phi(\chi - 1)(R - P) - R\phi[(T - P) + \chi(P - S)], \\ D(\mathbf{p}, \mathbf{q}, \mathbf{S_Y}) &= -S\phi(\chi - 1)(R - P) - R\phi[(T - P) + \chi(P - S)], \\ D(\mathbf{p}, \mathbf{q}, \mathbf{1}) &= -\phi(\chi - 1)(R - P) - \phi[(T - P) + \chi(P - S)]. \end{aligned}$$

Therefore, X's maximum payoff by using the extortionate strategy and and Y's maximum payoff when facing an extortionate player are given by

$$s_X = \frac{D(\mathbf{p},\mathbf{q},\mathbf{S_X})}{D(\mathbf{p},\mathbf{q},\mathbf{1})} = \frac{P(T-R)+\chi[R(P-S)+T(R-P)]}{(T-R)+\chi(R-S)},$$
$$s_Y = \frac{D(\mathbf{p},\mathbf{q},\mathbf{S_Y})}{D(\mathbf{p},\mathbf{q},\mathbf{1})} = \frac{R(T-S)+(\chi-1)P(R-S)}{(T-R)+\chi(R-S)}.$$

A special case is $\chi = 1$, which means that X chooses a strategy to make both players have the same expected payoff that has maximum R. $\chi = 1$ corresponds to the strategy $\mathbf{p} = (1, 0, 1, 0)$, the famous **Tit-for-tat** strategy.

Remark 3.6. A natural question here is what will happen if both players use extortionate strategies, that is, what will happen if both players are trying to earn a larger share of the total amount above the mutual defection payoffs. It is obvious that it is impossible for both players to get a larger share simultaneously, but it is certainly allowed for both players to choose any strategies that they want. Suppose X chooses an extortionate strategy $\mathbf{p}$ such that $s_X - P = \chi_1(s_Y - P)$ and Y chooses an extortionate strategy $\mathbf{q}$ such that $s_Y - P = \chi_2(s_X - P)$, where $\chi_1 \geq 1$ and $\chi_2 \geq 1$. Then, $\widetilde{\mathbf{p}} = \phi_1[\mathbf{S_X} - P\mathbf{1} - \chi_1(\mathbf{S_Y} - P\mathbf{1})]$; $\widehat{\mathbf{q}} = \phi_2[\mathbf{S_Y} - P\mathbf{1} - \chi_2(\mathbf{S_X} - P\mathbf{1})]$; ϕ_1 and ϕ_2 are nonzero parameters such that the strategies are feasible. More explicitly, as given by (3.4) and (3.5),

$$\mathbf{p} = (1-\phi_1(\chi_1-1)(R-P), 1-\phi_1[(P-S)+\chi_1(T-P)], \phi_1[(T-P)+\chi_1(P-S)], 0),$$
$$\mathbf{q} = (1-\phi_2(\chi_2-1)(R-P), 1-\phi_2[(P-S)+\chi_2(T-P)], \phi_2[(T-P)+\chi_2(P-S)], 0).$$

Under the two strategies, the payoffs for both players satisfy

$$\begin{cases} s_X - P = \chi_1(s_Y - P), \\ s_Y - P = \chi_2(s_X - P). \end{cases}$$

When χ_1 and χ_2 are not equal to 1 at the same time, the only solution is $s_X = s_Y = P$. A reasonable explanation comes from the observation that both p_4 and q_4 are zero. This means that once the two players defect at the same time in a round of the game, both will defect forever, resulting in the long run expected payoff for each of them to be P. This result implies that if both players want to extort the other, then neither of them can get more than P on average in the long run.

When $\chi_1 = 1$ as well as $\chi_2 = 1$, then certainly we have $s_X = s_Y$.

4. Further discussion on games with zero-determinant players

Now we shall present more properties of zero-determinant strategies derived in Section §2 and also discuss the situation where both players use zero-determinant strategies, based on the results in [6]. For convenience, we shall normalize the payoffs T, R, P, S. It does no harm if they are added by a same number or multiplied by a same positive number. As a consequence, without changing the structure of the game, we can assume that (i) $T = 1, S = 0$; (ii) $0 < P < R, \frac{1}{2} < R < 1$. Then the payoff vectors in a single game become $\mathbf{S_X} = (R, 0, 1, P)$ and $\mathbf{S_Y} = (R, 1, 0, P)$.

We begin by considering that X is a player using a zero-determinant strategy such that $\widetilde{\mathbf{p}} = (p_1 - 1, p_2 - 1, p_3, p_4) = \alpha\mathbf{S_X} + \beta\mathbf{S_Y} + \gamma\mathbf{1} = \alpha(R, 0, 1, P) + \beta(R, 1, 0, P) + \gamma(1, 1, 1, 1)$. For the strategy to be feasible, the values of α, β and γ should satisfy that so-called sign constraints:

$$p_1 - 1 = (\alpha+\beta)R + \gamma \le 0, \quad p_2 - 1 = \beta + \gamma \le 0,$$
$$p_3 = \alpha + \gamma \ge 0, \quad p_4 = (\alpha+\beta)P + \gamma \ge 0,$$

and the so-called size constraints that the absolute value of each entry of $\widetilde{\mathbf{p}}$ is at most 1.

Lemma 4.1. *If $\widetilde{\mathbf{p}} = \alpha \mathbf{S_X} + \beta \mathbf{S_Y} + \gamma \mathbf{1}$ satisfies the sign constraints, then $\alpha+\beta \le 0$ and $\gamma \ge 0$. Moreover, $\alpha + \beta = 0$ if and only if $\gamma = 0$.*

Proof. Consider the four inequalities under the sign constraints. Subtracting the fourth from the first gives $(\alpha+\beta)(R-P) \le 0$. Since $R > P$, $\alpha+\beta \le 0$. The fourth inequality then implies $\gamma \ge 0$ since $P > 0$. Using the first and the fourth again gives that $\alpha + \beta = 0$ if and only if $\gamma = 0$. □

According to this Lemma, if $\alpha+\beta = 0$, then $\gamma = 0$, and vice versa. $\alpha+\beta = \gamma = 0$ gives a strategy $\mathbf{p} = (1, 1-\alpha, \alpha, 0)$, which is a mixture of two common strategies *Tit-for-tat*, $\mathbf{p} = (1,0,1,0)$, and *Repeat*, $\mathbf{p} = (1,1,0,0)$. This strategy will also be mentioned in the next section. Excluding this, we can assume that $\gamma > 0$ and thus $\alpha + \beta < 0$. Since now $\gamma \ne 0$, we can define

$$\overline{\alpha} = \frac{\alpha}{\gamma}, \quad \overline{\beta} = \frac{\beta}{\gamma}.$$

Then $\widetilde{\mathbf{p}} = \gamma(\overline{\alpha}\mathbf{S_X} + \overline{\beta}\mathbf{S_Y} + \mathbf{1})$, and the sign constraints become

$$-P^{-1} \le \overline{\alpha} + \overline{\beta} \le -R^{-1}, \quad \overline{\beta} \le -1 \le \overline{\alpha}.$$

Definition 4.2. Define **ZDSstrip** to be the set $\{(x,y) : -P^{-1} \le x + y \le -R^{-1}, y \le -1 \le x\}$.

Any $(\overline{\alpha}, \overline{\beta}) \in ZDSstrip$ and sufficiently small $\gamma > 0$ give a feasible zero-determinant strategy $\mathbf{p}$.

Lemma 4.3. *Suppose $(\overline{\alpha}, \overline{\beta}) \in ZDSstrip$ and $\overline{\alpha} + \overline{\beta} = -Z^{-1}$. Then we have $-\overline{\beta} \ge max(1, |\overline{\alpha}|)$, and furthermore, $-\overline{\beta} = |\overline{\alpha}|$ if and only if $\overline{\alpha} = \overline{\beta} = -1$. Suppose $(\overline{a}, \overline{b}) \in ZDSstrip$. Then $D \coloneqq \overline{\beta}\overline{b} - \overline{\alpha}\overline{a} \ge 0$ with $D = 0$ if and only if $\overline{\alpha} = \overline{\beta} = \overline{a} = \overline{b} = -1$. Note that $(-1,-1) \in ZDSstrip$ if and only if $0 < P \le \frac{1}{2}$.*

Proof. $(\overline{\alpha}, \overline{\beta}) \in ZDSstrip$ means that $(\overline{\alpha}, \overline{\beta})$ satisfies the sign constraints $-P^{-1} \le \overline{\alpha} + \overline{\beta} \le -R^{-1}$ and $\overline{\beta} \le -1 \le \overline{\alpha}$. Since $\overline{\alpha} + \overline{\beta} = -Z^{-1}$, Z satisfies $P \le Z \le R$. Then $\overline{\alpha} + \overline{\beta} = -Z^{-1}$ and $Z \ge P > 0$ imply $-\overline{\beta} = \overline{\alpha} + Z^{-1} > \overline{\alpha}$. On the other hand, by the sign constraints, $-\overline{\beta} \ge 1 \ge -\overline{\alpha}$ and thus $-\overline{\beta} = -\overline{\alpha}$ if and only if $\overline{\alpha} = \overline{\beta} = -1$. Combining these we prove the first half of the lemma. For $(\overline{a}, \overline{b}) \in ZDSstrip$, we also have $-b \ge |\overline{a}| \ge 0$, which means that $D \coloneqq \overline{\beta}\overline{b} - \overline{\alpha}\overline{a} \ge (-\overline{\beta})(-\overline{b}) - |\overline{\alpha}||\overline{a}| \ge 0$ and the equality holds if and only if $-\overline{\beta} = |\overline{\alpha}|$ and $-\overline{b} = |\overline{a}|$, that is, using the first half of the lemma, if and only if $\overline{\alpha} = \overline{\beta} = \overline{a} = \overline{b} = -1$. Note that $(-1,-1) \in ZDSstrip$ if and only if $-P^{-1} \le -2 \le -R^{-1}$. Our condition for R, $\frac{1}{2} < R < 1$, makes $-R^{-1} \ge -2$ always hold. Hence, $(-1,-1) \in ZDSstrip$ if and only if $-P^{-1} \le -2$, namely, $0 < P \le \frac{1}{2}$. □

Proposition 4.4. *Suppose X uses a feasible zero-determinant strategy* $\mathbf{p}$ *such that* $\widetilde{\mathbf{p}} = \gamma(\overline{\alpha}\mathbf{S_X} + \overline{\beta}\mathbf{S_Y} + \mathbf{1})$. *Let* $\overline{\alpha} + \overline{\beta} = -Z^{-1} \in [P, R]$. *Then for any strategy* $\mathbf{q}$ *used by Y, the long run expected payoffs to X and Y,* s_X *and* s_Y*, satisfy*

$$\overline{\alpha}Z(s_X - s_Y) = s_Y - Z. \tag{4.5}$$

Let $\kappa = \frac{\overline{\alpha}Z}{1+\overline{\alpha}Z}$*, then for any strategy* $\mathbf{q}$ *used by Y,*

$$\kappa(s_X - Z) = s_Y - Z. \tag{4.6}$$

Also, $\kappa < 1$ *and* κ *has the same sign as* $\overline{\alpha}$.

Proof. If $\widetilde{\mathbf{p}} = \gamma(\overline{\alpha}\mathbf{S_X} + \overline{\beta}\mathbf{S_Y} + \mathbf{1})$, then according to Proposition 2.6, s_X and s_Y are subject to the linear relation

$$\overline{\alpha}s_X + \overline{\beta}s_Y + 1 = 0. \tag{4.7}$$

So $\overline{\alpha}Zs_X + \overline{\beta}Zs_Y + Z = 0$. Then $\overline{\alpha}Zs_X + (-Z^{-1} - \overline{\alpha})Zs_Y + Z = 0$ because $\overline{\beta} = -Z^{-1} - \overline{\alpha}$. After rearranging the terms, we get Equation (4.5).

For a feasible zero-determinant strategy $\mathbf{p}$, $\overline{\beta} \le -1 < 0$, which implies $\overline{\alpha} = -Z^{-1} - \overline{\beta} > -Z^{-1}$ and thus $1 + \overline{\alpha}Z > 1 - Z^{-1}Z = 0$ since $Z \ge P > 0$. So $\kappa = \frac{\overline{\alpha}Z}{1+\overline{\alpha}Z}$ has the same sign as $\overline{\alpha}$ and $\kappa = \frac{\overline{\alpha}Z}{1+\overline{\alpha}Z} = 1 - \frac{1}{1+\overline{\alpha}Z} < 1$. Dividing both sides of Equation (4.5) by $(1 + \overline{\alpha}Z)$ and rearranging the terms, we get Equation (4.6). □

Next we consider what will happen if both players use zero-determinant strategies. We begin by looking at Y's zero-determinant strategy $\mathbf{q}$. Proposition 2.6 shows that $\mathbf{q}$ should be of the form such that $\widehat{\mathbf{q}} = (q_1 - 1, q_3, q_2 - 1, q_4) = b\mathbf{S_X} + a\mathbf{S_Y} + g\mathbf{1}$ for some numbers a, b and g. Then a, b and g are subject to sign constraints and size constraints as well, similar to those for X's strategy $\mathbf{p}$. Y's version of Lemma 4.1 is:

Lemma 4.8. *If* $\widehat{\mathbf{q}} = (q_1 - 1, q_2, q_3 - 1, q_4) = b\mathbf{S_X} + a\mathbf{S_Y} + g\mathbf{1}$ *satisfies the sign constraints, then* $a + b \le 0$ *and* $g \ge 0$. *Moreover,* $a + b = 0$ *if and only if* $g = 0$.

We exclude the case $g = 0$ as before and focus on the case $g > 0$ and thus $a + b < 0$. Then we can define $\overline{b} = \frac{b}{g}$ and $\overline{a} = \frac{a}{g}$. Then $\widehat{\mathbf{q}} = g(\overline{b}\mathbf{S_X} + \overline{a}\mathbf{S_Y} + \mathbf{1})$, where $\overline{b}$ and $\overline{a}$ satisfy the sign constraints: $-P^{-1} \le \overline{a} + \overline{b} \le -R^{-1}$ and $\overline{b} \le -1 \le \overline{a}$. This means that $(\overline{a}, \overline{b})$ should lie in exactly the same feasible set as X's $(\overline{\alpha}, \overline{\beta})$, that is, $(\overline{a}, \overline{b}) \in ZDSstrip$. Any $(\overline{a}, \overline{b}) \in ZDSstrip$ and sufficiently small $g > 0$ also give a feasible zero-determinant strategy $\mathbf{q}$.

Y's version of Proposition 4.4 is:

Proposition 4.9. *Suppose Y uses a feasible zero-determinant strategy* $\mathbf{q}$ *such that* $\widehat{\mathbf{q}} = g(\overline{b}\mathbf{S_X} + \overline{a}\mathbf{S_Y} + \mathbf{1})$. *Let* $\overline{a} + \overline{b} = -Z^{-1} \in [P, R]$. *Then for any strategy* $\mathbf{p}$ *used by X, the long run expected payoffs to X and Y,* s_X *and* s_Y*, satisfy* $\overline{a}Z(s_Y - s_X) = s_X - Z$. *Let* $\kappa = \frac{\overline{a}Z}{1+\overline{a}Z}$*, then for any strategy* $\mathbf{p}$ *used by X,* $\kappa(s_Y - Z) = s_X - Z$. *Also,* $\kappa < 1$ *and* κ *has the same sign as* $\overline{a}$.

The next proposition serves as a summary of what will happen in a game of two zero-determinant players.

Proposition 4.10. *Suppose X use a zero-determinant strategy with* $\widetilde{\mathbf{p}} = \gamma(\overline{\alpha}\mathbf{S_X} + \overline{\beta}\mathbf{S_Y} + \mathbf{1})$ *and Y uses a zero-determinant strategy with* $\widehat{\mathbf{q}} = g(\overline{b}\mathbf{S_X} + \overline{a}\mathbf{S_Y} + \mathbf{1})$. *Let* $\overline{\alpha} + \overline{\beta} = -Z_X^{-1}$ *and* $\overline{a} + \overline{b} = -Z_Y^{-1}$; $Z_X, Z_Y \in [P, R]$. s_X *and* s_Y *are the long* *run expected payoffs to X and Y, respectively. Then, (i) $Z_X = Z_Y$ if and only if $s_X = s_Y$, and in this case $Z_X = Z_Y = s_X = s_Y$; (ii) $s_Y > s_X$ if and only if $Z_X > Z_Y$; (iii) If $Z_X > Z_Y$, we have the following:*

$$\begin{cases} \overline{\alpha} > 0 \Rightarrow Z_X > s_Y > s_X; \\ \overline{\alpha} = 0 \Rightarrow Z_X = s_Y > s_X; \\ \overline{\alpha} < 0 \Rightarrow s_Y > Z_X > s_X; \end{cases} \quad \begin{cases} \overline{a} > 0 \Rightarrow s_Y > s_X > Z_Y; \\ \overline{a} = 0 \Rightarrow s_Y > s_X = Z_Y; \\ \overline{a} < 0 \Rightarrow s_Y > Z_Y > s_X. \end{cases}$$

Proof. Proposition 2.6 shows that X's strategy with $\widetilde{\mathbf{p}} = \gamma(\overline{\alpha}\mathbf{S_X} + \overline{\beta}\mathbf{S_Y} + \mathbf{1})$ leads to $\overline{\alpha}s_X + \overline{\beta}s_Y + 1 = 0$ while Y's strategy with $\widehat{\mathbf{q}} = g(\overline{b}\mathbf{S_X} + \overline{a}\mathbf{S_Y} + \mathbf{1})$ leads to $\overline{b}s_X + \overline{a}s_Y + 1 = 0$. If they both use such strategies, then s_X and s_Y will satisfy the two equations, $\overline{\alpha}s_X + \overline{\beta}s_Y + 1 = 0$ and $\overline{b}s_X + \overline{a}s_Y + 1 = 0$, simultaneously. One singular case is that $D := \overline{\beta b} - \overline{\alpha a} = 0$, which, by Lemma 4.3, happens if and only if $\overline{\alpha} = \overline{\beta} = \overline{a} = \overline{b} = -1$. In this case, both strategies are the same and thus both have the same expected payoff: $s_X = s_Y = \frac{1}{2}$, and $Z_X = Z_Y = s_X = s_Y = \frac{1}{2}$, which is consistent with part (i). Otherwise $D = \overline{\beta b} - \overline{\alpha a} > 0$. For this general case,

$$s_X = D^{-1}(\overline{a} - \overline{\beta}), \quad s_Y = D^{-1}(\overline{\alpha} - \overline{b}), \tag{4.11}$$

$$\Rightarrow s_Y - s_X = D^{-1}[(\overline{\alpha} + \overline{\beta}) - (\overline{a} + \overline{b})] = D^{-1}(Z_Y^{-1} - Z_X^{-1}). \tag{4.12}$$

We can further verify that $s_X, s_Y \in [0,1]$ as it should be, by the following argument. For feasible strategies $\widetilde{\mathbf{p}}$ and $\widehat{\mathbf{q}}$, $(\overline{\alpha}, \overline{\beta})$ and $(\overline{a}, \overline{b})$ are in the $ZDSstrip$. So $\overline{a} \geq -1 \geq \overline{\beta}$ and thus $s_X = D^{-1}(\overline{a} - \overline{\beta}) \geq 0$. On the other hand, since $0 < P \leq Z_X, Z_Y \leq R < 1$, we have $-\overline{\beta} = \overline{\alpha} + Z_X^{-1} > \overline{\alpha} + 1 \geq 0$ and $-\overline{b} = \overline{a} + Z_Y^{-1} > \overline{a} + 1 \geq 0$. Note that $-\overline{b} - 1 \geq 0$. So $\overline{a}(\overline{\alpha}+1) \leq (-\overline{b}-1)(-\overline{\beta}) = (\overline{b}+1)\overline{\beta}$ and thus $0 \leq \overline{a} - \overline{\beta} \leq \overline{\beta b} - \overline{\alpha a} = D$. This implies $s_X \leq 1$. Hence $s_X \in [0,1]$. Similarly $s_Y \in [0,1]$.

(a) On the one hand, if $Z_X = Z_Y$, then (4.12) implies $s_X = s_Y$. On the other hand, according to Proposition 4.4 and Proposition 4.9, we see that s_X and s_Y satisfy $\overline{\alpha}Z_X(s_X - s_Y) = s_Y - Z_X$ and $\overline{a}Z_Y(s_Y - s_X) = s_X - Z_Y$ simultaneously. So if $s_X = s_Y$, then $s_X = s_Y = Z_X = Z_Y$. (i) is true.

(b) Observing that $D > 0$ and $s_Y - s_X = D^{-1}(Z_Y^{-1} - Z_X^{-1})$, (ii) is true.

(c) From (ii), assuming $Z_X > Z_Y$ implies $s_Y > s_X$, i.e. $s_X - s_Y < 0$. Since $\overline{\alpha}Z_X(s_X - s_Y) = s_Y - Z_X$ and $Z_X \geq P > 0$, $s_Y - Z_X$ and $\overline{\alpha}$ have opposite signs. Hence, $\overline{\alpha} > 0 \Rightarrow Z_X > s_Y > s_X$ and $\overline{\alpha} = 0 \Rightarrow Z_X = s_Y > s_X$. When $\overline{\alpha} < 0$, then $s_Y > Z_X$. To compare s_X and Z_X in this case, we use Equation (4.6) in Proposition 4.4. s_X and s_Y are linked by $\kappa_X(s_X - Z_X) = s_Y - Z_X$. $\overline{\alpha} < 0$ implies $\kappa_X < 0$, so $s_X - Z_X$ and $s_Y - Z_X$ have opposite signs. Since $s_Y > Z_X$, $s_X < Z_X$. This completes the first half of (iii). A similar argument applies to the second half if we use the relation $\overline{a}Z_Y(s_Y - s_X) = s_X - Z_Y$ and $\kappa_Y(s_Y - Z_Y) = s_X - Z_Y$ instead.

□

5. Good strategies and strategies of Nash type

In this section we are going to introduce some other strategy concepts in the stochastic iterated prisoner's dilemma, including agreeable strategies, firm strategies, good strategies and strategies of Nash type. These concepts are proposed by Ethan Akin in [6]. As is the case for the stochastic iterated prisoner's dilemma, both players, X and Y, specify their strategies in terms of cooperation probabilities, as a result of which there will be a probability distribution $\mathbf{v} = (v_1, v_2, v_3, v_4)$ over the set of outcomes $\{CC, CD, DC, DD\}$ in a round of the game. Again, the first letter represents X's choice and we still index these four states by $1, 2, 3, 4$. $\mathbf{v}$ satisfies $\mathbf{v} \cdot \mathbf{1} = 1$ and $v_i \in [0, 1]$ for $i = 1, 2, 3, 4$. With X's payoff vector $\mathbf{S_X} = (R, S, T, P)$ and Y's payoff vector $\mathbf{S_Y} = (R, T, S, P)$ corresponding to the four outcomes, the expected payoffs to X and Y with respect to the distribution $\mathbf{v}$, denoted by s_X and s_Y, are given by the dot products:

$$s_X = \mathbf{v} \cdot \mathbf{S_X}, \quad s_Y = \mathbf{v} \cdot \mathbf{S_Y}.$$

Remember that R,T,S and P satisfy two conditions: $T > R > P > S$ and $2R > T + S$.

Proposition 5.1. *Suppose* $\mathbf{v} = (v_1, v_2, v_3, v_4)$ *is a distribution. Then,* $s_Y - s_X = (T-S)(v_2-v_3)$*, which implies* $s_Y = s_X$ *if and only if* $v_2 = v_3$*. Moreover,* $s_X+s_Y \leq 2R$ *and the following statements are equivalent: (i)* $s_X + s_Y = 2R$*; (ii)* $v_1 = 1$*; (iii)* $s_X = s_Y = R$*.*

Proof. The proposition can be proved easily by using $s_Y - s_X = \mathbf{v} \cdot \mathbf{S_Y} - \mathbf{v} \cdot \mathbf{S_X} = (v_2 - v_3)(T - S)$, $s_X + s_Y = \mathbf{v} \cdot \mathbf{S_Y} + \mathbf{v} \cdot \mathbf{S_X} = 2v_1R + (v_2 + v_3)(T + S) + 2v_4P$, $T > R > P > S$ and $2R > T + S$. □

In the following our discussion is still limited to memory-one strategies. Then exactly the same as in the previous sections, X's strategy is $\mathbf{p} = (p_1, p_2, p_3, p_4)$ and Y's strategy, from his viewpoint, is $\mathbf{q} = (q_1, q_2, q_3, q_4)$. Notice that the same strategy for X and Y are described by the same probability vector. We still look at the game from X's perspective. Let $\mathbf{v}$ be a stationary vector associated with the Markov process of the game. $\mathbf{v}$ always exists, according to Remark 1.5. s_X and s_Y are the expected payoffs to X and Y respectively under the stationary distribution $\mathbf{v}$.

Definition 5.2. A strategy $\mathbf{p}$ is **agreeable** if $p_1 = 1$ and is **firm** if $p_4 = 0$.

Example 5.3. The strategies, *Tit-for-tat* $\mathbf{p} = (1, 0, 1, 0)$ and *Repeat* $\mathbf{p} = (1, 1, 0, 0)$, are both agreeable and firm. The extortionate strategy in Section §3.3 is firm.

Definition 5.4. A memory-one strategy for X is **good** if (i) it is agreeable; (ii) for any strategy of Y and any associated stationary distribution, $s_Y \geq R$ implies $s_X = s_Y = R$. A memory-one strategy is of **Nash type** if (i) it is agreeable; (ii) for any strategy of Y and any associated stationary distribution, $s_Y \geq R$ implies $s_Y = R$.

By definition, a good strategy is of Nash type but the converse is not true.

Example 5.5. *Repeat* $= (1, 1, 0, 0)$ is agreeable but not of Nash type and not good: if both players use *Repeat* and the outcome of the first round is $xy = CD$, then $s_Y = T$ and $s_X = S$ because the same outcome will always occur.

Definition 5.6. Let $\mathbf{e}_{12} = (1, 1, 0, 0)$. The **X Press-Dyson vector** $\widetilde{\mathbf{p}} = (\widetilde{p_1}, \widetilde{p_2}, \widetilde{p_3}, \widetilde{p_4})$ of a strategy $\mathbf{p}$ of X is defined to be $\widetilde{\mathbf{p}} := \mathbf{p} - \mathbf{e}_{12}$, which is just the transpose of the second column of the matrix $\boldsymbol{\mathcal{L}'}$ in Section §2.

Since the determinant of the matrix whose four columns are $\mathbf{S_X} = (R, S, T, P)$, $\mathbf{S_Y} = (R, T, S, P)$, $\mathbf{1} = (1, 1, 1, 1)$ and $\mathbf{e}_{23} = (0, 1, 1, 0)$ respectively is equal to $-2(R - P)(T - S)$ that is nonzero, $\mathbf{S_X} = (R, S, T, P), \mathbf{S_Y} = (R, T, S, P), \mathbf{1} = (1, 1, 1, 1)$ and $\mathbf{e}_{23} = (0, 1, 1, 0)$ form a basis for $\mathbb{R}^4$. Hence we can write $\widetilde{\mathbf{p}}$ as a linear combination of $\mathbf{S_X} = (R, S, T, P), \mathbf{S_Y} = (R, T, S, P), \mathbf{1} = (1, 1, 1, 1)$ and $\mathbf{e}_{23} = (0, 1, 1, 0)$: $\widetilde{\mathbf{p}} = \alpha\mathbf{S_X} + \beta\mathbf{S_Y} + \gamma\mathbf{1} + \delta\mathbf{e}_{23}$ for some $\alpha, \beta, \gamma, \delta$.

Proposition 5.7. *Suppose X uses a strategy* $\mathbf{p}$ *with X Press-Dyson vector* $\widetilde{\mathbf{p}}$ *and Y uses a strategy that leads to a sequence of distributions* $\{\mathbf{v}^{(n)}, n = 1, 2, ...\}$ *with* $\mathbf{v}^{(k)}$ *representing the distribution over the states in the* k^{th} *round of the game. Let* $\mathbf{v}$ *be an associated stationary distribution. Then,*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} (\mathbf{v}^{(k)} \cdot \widetilde{\mathbf{p}}) = 0, \text{ and therefore } \mathbf{v} \cdot \widetilde{\mathbf{p}} = 0.$$

Proof. The probability that X cooperates in the n^{th} round, denoted by $v_{12}^{(n)}$, is $v_{12}^{(n)} = v_1^{(n)} + v_2^{(n)} = \mathbf{v}^{(n)} \cdot \mathbf{e}_{12}$. The probability that X cooperates in the $(n+1)^{th}$ round, denoted by $v_{12}^{(n+1)}$, is $v_{12}^{(n+1)} = \mathbf{v}^{(n)} \cdot \mathbf{p}$. Thus,

$$v_{12}^{(n+1)} - v_{12}^{(n)} = \mathbf{v}^{(n)} \cdot \mathbf{p} - \mathbf{v}^{(n)} \cdot \mathbf{e}_{12} = \mathbf{v}^{(n)} \cdot (\mathbf{p} - \mathbf{e}_{12}) = \mathbf{v}^{(n)} \cdot \widetilde{\mathbf{p}}.$$

This implies $v_{12}^{(n+1)} - v_{12}^{(1)} = \sum_{k=1}^{n} (v_{12}^{(k+1)} - v_{12}^{(k)}) = \sum_{k=1}^{n} (\mathbf{v}^{(k)} \cdot \widetilde{\mathbf{p}})$.

Since $0 \le v_{12}^{(k)} \le 1$ for any k,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} (\mathbf{v}^{(k)} \cdot \widetilde{\mathbf{p}}) = \lim_{n\to\infty} \frac{1}{n} (v_{12}^{(n+1)} - v_{12}^{(1)}) = 0.$$

For the stationary distribution $\mathbf{v}$ that is the limit of some subsequence of the Cesaro averages $\{\frac{1}{n}\sum_{k=1}^{n} \mathbf{v}^{(k)}\}$, the continuity of the dot product implies $\mathbf{v} \cdot \widetilde{\mathbf{p}} = 0$. □

Corollary 5.8. *Suppose X uses a strategy* $\mathbf{p}$ *having X Press-Dyson vector* $\widetilde{\mathbf{p}} = \alpha\mathbf{S_X} + \beta\mathbf{S_Y} + \gamma\mathbf{1} + \delta\mathbf{e}_{23}$ *and Y uses any strategy* $\mathbf{q}$*. Let* $\mathbf{v}$ *be a stationary distribution of the Markov process of the game. Denote* $v_{23} = \mathbf{v} \cdot \mathbf{e}_{23} = v_2 + v_3$*. Then the long run expected payoffs per round,* s_X *and* s_Y*, satisfy the following linear equation, called* ***Press-Dyson Equation****,*

$$\alpha s_X + \beta s_Y + \gamma + \delta v_{23} = 0. \tag{5.9}$$

Proof. According to Proposition 5.7, we have $0 = \mathbf{v} \cdot \widetilde{\mathbf{p}} = \alpha\mathbf{v} \cdot \mathbf{S_X} + \beta\mathbf{v} \cdot \mathbf{S_Y} + \gamma\mathbf{v} \cdot \mathbf{1} + \delta\mathbf{v} \cdot \mathbf{e}_{23} = \alpha s_X + \beta s_Y + \gamma + \delta v_{23}$. □

In particular, if we take $\delta = 0$, then under the stationary state, using strategy $\widetilde{\mathbf{p}} = \alpha\mathbf{S_X} + \beta\mathbf{S_Y} + \gamma\mathbf{1}$ can unilaterally enforce a linear relation $\alpha s_X + \beta s_Y + \gamma = 0$ between the expected payoffs of the opponent and oneself. Such a strategy is just the same as the zero-determinant strategy obtained in Section §2 using Press and Dyson's argument, but here $\mathbf{v}$ is a stationary distribution and is not necessarily to be the unique limiting distribution that may not exist. Hence, zero-determinant strategies exist in a general Markov process of the game that has a stationary distribution $\mathbf{v}$.

The next theorem is about the the necessary and sufficient condition for an agreeable strategy to be of Nash type or to be good.

Theorem 5.10. *Let* $\mathbf{p} = (p_1, p_2, p_3, p_4)$ *be an agreeable strategy but not Repeat, i.e.* $p_1 = 1$ *but* $\mathbf{p} \neq (1, 1, 0, 0)$*. The non-Repeat agreeable strategy* $\mathbf{p}$ *is of Nash type if* *and only if it satisfies the following two inequalities:*

$$\frac{T-R}{R-S}p_3 \le 1-p_2 \quad \textit{and} \quad \frac{T-R}{R-P}p_4 \le 1-p_2.$$

The non-Repeat agreeable strategy $\mathbf{p}$ *is good if and only if both inequalities above are strict.*

For convenience, we still normalize the payoffs T, R, P, S from now on. We assume that (i) $T = 1, S = 0$; (ii) $0 < P < R, \frac{1}{2} < R < 1$. The payoff vectors in a single round are $\mathbf{S_X} = (R, 0, 1, P)$ and $\mathbf{S_Y} = (R, 1, 0, P)$. Then Theorem 5.10 is equivalent to the following:

Theorem 5.11. *Consider the normalized stochastic iterated prisoner's dilemma. Let* $\mathbf{p} = (p_1, p_2, p_3, p_4)$ *be an agreeable strategy but not Repeat, i.e.* $p_1 = 1$ *but* $\mathbf{p} \ne (1, 1, 0, 0)$*. The non-Repeat agreeable strategy* $\mathbf{p}$ *is of Nash type if and only if it satisfies the following two inequalities:*

$$\frac{1-R}{R}p_3 \le 1-p_2 \quad \textit{and} \quad \frac{1-R}{R-P}p_4 \le 1-p_2. \tag{5.12}$$

The non-Repeat agreeable strategy $\mathbf{p}$ *is good if and only if both inequalities above are strict.*

Proof. In the normalized game, $\mathbf{S_X} = (R, 0, 1, P)$ and $\mathbf{S_Y} = (R, 1, 0, P)$. Under stationary state $\mathbf{v} = (v_1, v_2, v_3, v_4)$, $s_Y = \mathbf{v} \cdot \mathbf{S_Y} = v_1 R + v_2 + v_4 P$. Also recall that by definition, if a strategy is good, then it is of Nash type, but converse is not true. If a strategy is not of Nash type, then it is not good.

First of all, if $p_2 = 1$, then $\mathbf{p} = (1, 1, p_3, p_4)$. If Y uses strategy $\mathbf{q} = (0, 0, 0, 1)$, then $\mathbf{v} = (0, 1, 0, 0)$ is a stationary vector associated with this game. Under this stationary state, $s_X = 0$ and $s_Y = 1 > R$, so no matter what p_3 and p_4 are, $\mathbf{p} = (1, 1, p_3, p_4)$ is not of Nash type and thus not good. On the other hand, $p_2 = 1$ gives no solution to the inequalities (5.12) except $p_3 = p_4 = 0$, i.e. $\mathbf{p} = (1, 1, 0, 0)$, which has been excluded by our assumptions in the theorem.

Therefore, we just need to consider $p_2 \in [0, 1)$. Notice that because $p_1 = 1$, using Proposition 5.7 and $\widetilde{p_1} = p_1 - 1 = 0$ gives $v_2\widetilde{p_2} + v_3\widetilde{p_3} + v_4\widetilde{p_4} = 0$, that is,

$$(1-p_2)v_2 = v_3 p_3 + v_4 p_4. \tag{5.13}$$

Since $s_Y - R = (v_1 R + v_2 + v_4 P) - (v_1 + v_2 + v_3 + v_4)R = v_2(1-R) - v_3 R - v_4(R-P)$ and $1 - p_2 > 0$,

$$\begin{aligned} s_Y \ge R &\Leftrightarrow v_2(1-R) \ge v_3 R + v_4(R-P), \\ &\Leftrightarrow (1-p_2)v_2(1-R) \ge (1-p_2)v_3 R + (1-p_2)v_4(R-P), \\ &\Leftrightarrow (v_3 p_3 + v_4 p_4)(1-R) \ge (1-p_2)v_3 R + (1-p_2)v_4(R-P), \\ &\Leftrightarrow [p_3(1-R) - (1-p_2)R]v_3 \ge [(1-p_2)(R-P) - p_4(1-R)]v_4. \end{aligned}$$

Let $A = p_3(1-R) - (1-p_2)R$ and $B = (1-p_2)(R-P) - p_4(1-R)$, then $s_Y \ge R$ if and only if $Av_3 \ge Bv_4$ (and obviously $s_Y = R$ if and only if $Av_3 = Bv_4$). The inequalities (5.12) are the same as $A \le 0$ and $B \ge 0$, so what we need to prove now is: (i) the necessary and sufficient condition for a non-Repeat agreeable strategy to be of Nash type is $A \le 0$ and $B \ge 0$; (ii) the necessary and sufficient condition for a non-Repeat agreeable strategy to be good is $A < 0$ and $B > 0$. By considering the signs of A and B and discussing them case by case, we can prove the two claims simultaneously.

(i) $A > 0$: Suppose against X's strategy $\mathbf{p} = (1, p_2, p_3, p_4)$, Y uses strategy $\mathbf{q} = (0, 1, 1, 1)$. Then (in the long run) it is impossible to get the outcome $xy = DD$, so $v_4 = 0$. Also, $q_1 = 0$ implies $v_1 < 1$; $p_2 < 1$ and $A = p_3(1-R) - (1-p_2)R > 0$ imply $p_3 > 0$. Using Equation (5.13) here gives $(1-p_2)v_2 = v_3 p_3$, which implies $v_2, v_3 > 0$ (otherwise $v_2 = v_3 = 0 \Rightarrow v_1 = 1$, contradiction!). So $Av_3 > 0 = Bv_4$, which means $s_Y > R$. Therefore, $\mathbf{p}$ is not Nash and thus not good.

(ii) $A = 0, B > 0$: Then $Av_3 \geq Bv_4$ if and only if $v_4 = 0$. When $v_4 = 0$, $Av_3 = 0 = Bv_4$ and thus $s_Y = R$. So $\mathbf{p}$ is Nash. On the other hand, if Y uses strategy $\mathbf{q} = (0, 1, 1, 1)$, then similar to case (i), we have $v_4 = 0$ and $v_1 < 1$. By Proposition 5.1, $s_X + s_Y < 2R$, and since $s_Y = R$, $s_X < R$. Hence, though $\mathbf{p}$ is Nash, $\mathbf{p}$ is not good.

(iii) $A = 0, B = 0$: Then $\forall \mathbf{q}$, $Av_3 = 0 = Bv_4$. So $\forall \mathbf{q}$, $s_Y = R$, which means $\mathbf{p}$ is Nash. On the other hand, if Y uses some strategy with $q_1 < 1$, then $v_1 < 1$. By Proposition 5.1, $s_X + s_Y < 2R$ and since $s_Y = R$, $s_X < R$. Hence, though $\mathbf{p}$ is Nash, $\mathbf{p}$ is not good.

(iv) $A \leq 0, B < 0$: Suppose against X's strategy $\mathbf{p} = (1, p_2, p_3, p_4)$, Y uses strategy $\mathbf{q} = (0, 0, 0, 0)$. Then (in the long run) it is impossible to get the outcome $xy = DC$, so $v_3 = 0$. Also, $q_1 = 0$ implies $v_1 < 1$; $p_2 < 1$ and $B = (1-p_2)(R-P) - p_4(1-R) < 0$ imply $p_4 > 0$. Using Equation (5.13) here gives $(1-p_2)v_2 = v_4 p_4$, which implies $v_2, v_4 > 0$ (otherwise $v_2 = v_4 = 0 \Rightarrow v_1 = 1$, contradiction!). So $Av_3 = 0 > Bv_4$, which means $s_Y > R$. Therefore, $\mathbf{p}$ is not Nash and thus not good.

(v) $A < 0, B = 0$: Then $Av_3 \geq Bv_4$ if and only if $v_3 = 0$. When $v_3 = 0$, $Av_3 = 0 = Bv_4$ and thus $s_Y = R$. So $\mathbf{p}$ is Nash. On the other hand, if Y uses strategy $\mathbf{q} = (0, 0, 0, 0)$, then similar to case (iv), we have $v_3 = 0$ and $v_1 < 1$. By Proposition 5.1, $s_X + s_Y < 2R$, and since $s_Y = R$, $s_X < R$. Hence, though $\mathbf{p}$ is Nash, $\mathbf{p}$ is not good.

(vi) $A < 0, B > 0$: Then $Av_3 \geq Bv_4$ if and only if $v_3 = v_4 = 0$. Using Equation (5.13) here gives $(1-p_2)v_2 = v_3 p_3 + v_4 p_4 = 0$. Since $p_2 < 1$, $v_2 = 0$ and thus $v_1 = 1 - v_2 - v_3 - v_4 = 1$. So $s_Y \geq R$ implies $v_1 = 1$, which, by Proposition 5.1, implies $s_X = s_Y = R$. This means that $\mathbf{p}$ is good and thus Nash.

All possible cases have been considered above and therefore we complete the proof. □

Corollary 5.14. *Consider the normalized stochastic iterated prisoner's dilemma. Let* $\mathbf{p} = (p_1, p_2, p_3, p_4)$ *be an agreeable strategy but not Repeat. Suppose* $\mathbf{p}$ *has X Press-Dyson vector* $\widetilde{\mathbf{p}} = \alpha \mathbf{S_X} + \beta \mathbf{S_Y} + \gamma \mathbf{1} + \delta \mathbf{e}_{23}$, $(\alpha, \beta, \gamma, \delta) \neq (0, 0, 0, 0)$. *The strategy* $\mathbf{p}$ *is of Nash type if and only if it satisfies the following:*

$$\max(\delta, \frac{\delta}{2R-1}) \leq \alpha.$$

The strategy $\mathbf{p}$ *is good if and only if the inequality above is strict.*

Proof. In the normalized game, $\mathbf{S_X} = (R, 0, 1, P)$ and $\mathbf{S_Y} = (R, 1, 0, P)$. Thus, $\widetilde{\mathbf{p}} = (\widetilde{p_1}, \widetilde{p_2}, \widetilde{p_3}, \widetilde{p_4}) = \alpha \mathbf{S_X} + \beta \mathbf{S_Y} + \gamma \mathbf{1} + \delta \mathbf{e}_{23} = ((\alpha+\beta)R + \gamma, \beta + \gamma + \delta, \alpha + \gamma + \delta, (\alpha+\beta)P + \gamma)$. For an agreeable strategy, $\widetilde{p_1} = p_1 - 1 = 0$, i.e. $(\alpha+\beta)R + \gamma = 0$, so $\beta = -\alpha - \frac{\gamma}{R}$. Then,

$$(1-p_2) = -\widetilde{p_2} = -(\beta+\gamma+\delta) = \alpha + \frac{1-R}{R}\gamma - \delta,$$
$$p_3 = \widetilde{p_3} = \alpha+\gamma+\delta, \quad p_4 = \widetilde{p_4} = (\alpha+\beta)P + \gamma = \frac{R-P}{R}\gamma.$$

Hence, the inequalities in (5.12) are equivalent to $\frac{1-R}{R}(\alpha+\gamma+\delta) \le \alpha + \frac{1-R}{R}\gamma - \delta$ and $\frac{1-R}{R-P}\frac{R-P}{R}\gamma \le \alpha + \frac{1-R}{R}\gamma - \delta$. After simplification they become $\frac{\delta}{2R-1} \le \alpha$ and $\delta \le \alpha$. Then it is clear that this corollary is just a special version of Theorem 5.11. □

For completeness, we also state the general version of the above corollary:

Corollary 5.15. *Consider the prisoner's dilemma without normalizing the payoffs. Let* $\mathbf{p} = (p_1, p_2, p_3, p_4)$ *be an agreeable strategy but not Repeat. Suppose* $\mathbf{p}$ *has X Press-Dyson vector* $\widetilde{\mathbf{p}} = \alpha\mathbf{S_X} + \beta\mathbf{S_Y} + \gamma\mathbf{1} + \delta\mathbf{e}_{23}$*,* $(\alpha,\beta,\gamma,\delta) \ne (0,0,0,0)$*. The strategy* $\mathbf{p}$ *is of Nash type if and only if it satisfies the following:*

$$\max\left(\frac{\delta}{T-S}, \frac{\delta}{2R-(T+S)}\right) \le \alpha.$$

The strategy $\mathbf{p}$ *is good if and only if the inequality above is strict.*

6. Generalization to memory-n strategies

We have so far focused on memory-one strategies that make the game a Markov process. Though in the Appendix A of [5] it is shown that assuming that players use memory-one strategies will not lose generality to a certain extent, it would be better if we can figure out what will happen if players use strategies based on longer memory. In fact, in a stochastic iterated prisoner's dilemma where, by definition, the strategies are specified in terms of cooperation probabilities, as long as both players use finite-memory strategies, then the game can be modeled by a Markov process and the idea of zero-determinant strategies applies as well.

To illustrate this point, we consider the case where both players are memory-two. That is, their strategies are based on the outcomes of the previous two rounds. For the general cases, the state space can be very large but we can still manipulate them in a similar way.

Denote the sequence of outcomes in the game by $\{U_n, n = 1, 2, ...\}$ where $U_n \in \{CC, CD, DC, DD\}$ with the first letter representing X's choice. It seems that using a memory-two strategy means that given the current outcome, the next outcome still depends on the previous outcome, so the process alone is not Markov. This is true, but there is a way to make it Markovian. The trick is to consider instead a new process, denoted by $\{V_n, n = 1, 2, ...\}$, such that each state of it describes the outcome of the previous round and the outcome of the current round. Given the current state which is specified by the outcome of the previous round and the outcome of the current round, then the future states will be conditionally independent of the past states because the strategies are memory-two, so this is a Markov process. Since for a single round there are 4 possible outcomes, for two rounds there will be $4 \times 4 = 16$ possible situations. and thus the state space of $\{V_n\}$ contains 16 elements. If we use the symbol E_XE_Y/F_XF_Y to represent a state where the outcome of the previous round is E_XE_Y and the outcome of the current round is F_XF_Y and $E_X, E_Y, F_X, F_Y \in \{C, D\}$, then the state space is

$\{CC/CC, CC/CD, CC/DC, CC/DD, CD/CC, CD/CD, CD/DC, CD/DD, DC/CC, DC/CD, DC/DC, DC/DD, DD/CC, DD/CD, DD/DC, DD/DD\}$. X's strategy is $\mathbf{p} = (p_1, p_2, ..., p_{16})$ corresponding to the previous state $E_XE_Y/F_XF_Y \in \{CC/CC, CC/CD, CC/DC, CC/DD, ..., DD/CC, DD/CD, DD/DC, DD/DD\}$. Y considers how to play the game from his own view point, so his strategy is $\mathbf{q} = (q_1, q_2, ..., q_{16})$ corresponding to the previous state $E_YE_X/F_YF_X \in \{CC/CC, CC/CD, CC/DC, CC/DD, ..., DD/CC, DD/CD, DD/DC, DD/DD\}$. In this way the same vector gives the same strategy to X and Y. As before, we shall analyze the game from X's perspective and use the state space of the form E_XE_Y/F_XF_Y. Corresponding to this space, X's strategy is $\mathbf{p} = (p_1, p_2, ..., p_{16})$ and Y's strategy is $\mathbf{q} = (q_1, q_3, q_2, q_4, q_9, q_{11}, q_{10}, q_{12}, q_5, q_7, q_6, q_8, q_{13}, q_{15}, q_{14}, q_{16})$. We can then write the transition matrix $\mathbf{P} = \mathbf{P}(\mathbf{p}, \mathbf{q})$. Note that $\Pr(V_{n+1} = G_XG_Y/H_XH_Y | V_n = E_XE_Y/F_XF_Y) = 0$ if $G_XG_Y \neq F_XF_Y$, so in each row of the matrix there will be at most four nonzero elements. The transition matrix $\mathbf{P}$ is:

$$\begin{bmatrix}
p_1q_1 & p_1(1-q_1) & (1-p_1)q_1 & (1-p_1)(1-q_1) & ... & 0 \\
0 & 0 & 0 & 0 & ... & 0 \\
0 & 0 & 0 & 0 & ... & 0 \\
0 & 0 & 0 & 0 & ... & (1-p_4)(1-q_4) \\
p_5q_9 & p_5(1-q_9) & (1-p_5)q_9 & (1-p_5)(1-q_9) & ... & 0 \\
0 & 0 & 0 & 0 & ... & 0 \\
0 & 0 & 0 & 0 & ... & 0 \\
0 & 0 & 0 & 0 & ... & (1-p_8)(1-q_{12}) \\
p_9q_5 & p_9(1-q_5) & (1-p_9)q_5 & (1-p_9)(1-q_5) & ... & 0 \\
0 & 0 & 0 & 0 & ... & 0 \\
0 & 0 & 0 & 0 & ... & 0 \\
0 & 0 & 0 & 0 & ... & (1-p_{12})(1-q_8) \\
p_{13}q_{13} & p_{13}(1-q_{13}) & (1-p_{13})q_{13} & (1-p_{13})(1-q_{13}) & ... & 0 \\
0 & 0 & 0 & 0 & ... & 0 \\
0 & 0 & 0 & 0 & ... & 0 \\
0 & 0 & 0 & 0 & ... & (1-p_{16})(1-q_{16})
\end{bmatrix}.$$

As in Section §2, we consider the case where the process has a unique limiting distribution $\boldsymbol{\pi} = (\pi_1, \pi_2, ..., \pi_{16})$. Then,

$$\begin{aligned}
\lim_{n\to\infty} \Pr(U_n = CC) &= \pi_1 + \pi_5 + \pi_9 + \pi_{13}, \\
\lim_{n\to\infty} \Pr(U_n = CD) &= \pi_2 + \pi_6 + \pi_{10} + \pi_{14}, \\
\lim_{n\to\infty} \Pr(U_n = DC) &= \pi_3 + \pi_7 + \pi_{11} + \pi_{15}, \\
\lim_{n\to\infty} \Pr(U_n = DD) &= \pi_4 + \pi_8 + \pi_{12} + \pi_{16}.
\end{aligned}$$

Hence, the long run expected payoffs to both players are given by

$$\begin{aligned}
s_X =& R(\pi_1 + \pi_5 + \pi_9 + \pi_{13}) + S(\pi_2 + \pi_6 + \pi_{10} + \pi_{14}) \\
&+ T(\pi_3 + \pi_7 + \pi_{11} + \pi_{15}) + P(\pi_4 + \pi_8 + \pi_{12} + \pi_{16}) = \boldsymbol{\pi} \cdot \mathbf{S_X}, \\
s_X =& R(\pi_1 + \pi_5 + \pi_9 + \pi_{13}) + T(\pi_2 + \pi_6 + \pi_{10} + \pi_{14}) \\
&+ S(\pi_3 + \pi_7 + \pi_{11} + \pi_{15}) + P(\pi_4 + \pi_8 + \pi_{12} + \pi_{16}) = \boldsymbol{\pi} \cdot \mathbf{S_Y},
\end{aligned}$$

$$\begin{aligned}
\text{where} \quad \mathbf{S_X} &= (R, S, T, P, R, S, T, P, R, S, T, P, R, S, T, P), \\
\mathbf{S_Y} &= (R, T, S, P, R, T, S, P, R, T, S, P, R, T, S, P).
\end{aligned}$$

Similar to the memory-one case, we can manipulate $\boldsymbol{\mathcal{L}} := \mathbf{P} - \mathbf{I}_{16}$ by adding one column to another two respectively to get a matrix with some columns solely depending on $\mathbf{p}$ and some columns solely depending on $\mathbf{q}$. For instance, if the nonzero row of Adj($\boldsymbol{\mathcal{L}}$) that we use to express the limiting vector $\mathbf{w}$ is one of the Row $4, 8, 12$ and 16, then to get the similar key determinant as $D(\mathbf{p}, \mathbf{q}, \mathbf{f})$ in Proposition 2.2, we can add the Column k of $\boldsymbol{\mathcal{L}}$ to Column $k+1$ and $k+2$ respectively for $k = 1, 5, 9, 13$, without changing the the value of the (i, j) cofactor of $\boldsymbol{\mathcal{L}}$ for $i = 1, 2, ..., 16$ and $j = 4, 8, 12, 16$. Then the transposes of the related columns become:

$$\begin{aligned}
&\text{Column } 2: \widetilde{\mathbf{p_2}} = (p_1 - 1, -1, 0, 0, p_5, 0, 0, 0, p_9, 0, 0, 0, p_{13}, 0, 0, 0);\\
&\text{Column } 6: \widetilde{\mathbf{p_6}} = (0, p_2, 0, 0, -1, p_6 - 1, 0, 0, 0, p_{10}, 0, 0, 0, p_{14}, 0, 0);\\
&\text{Column } 10: \widetilde{\mathbf{p_{10}}} = (0, 0, p_3, 0, 0, 0, p_7, 0, -1, -1, p_{11}, 0, 0, 0, p_{15}, 0);\\
&\text{Column } 14: \widetilde{\mathbf{p_{14}}} = (0, 0, 0, p_4, 0, 0, 0, p_8, 0, 0, 0, p_{12}, -1, -1, 0, p_{16});\\
&\text{Column } 3: \widehat{\mathbf{q_3}} = (q_1 - 1, 0, -1, 0, q_9, 0, 0, 0, q_5, 0, 0, 0, q_{13}, 0, 0, 0);\\
&\text{Column } 7: \widehat{\mathbf{q_7}} = (0, q_3, 0, 0, -1, q_{11}, -1, 0, 0, q_7, 0, 0, 0, q_{15}, 0, 0);\\
&\text{Column } 11: \widehat{\mathbf{q_{11}}} = (0, 0, q_2, 0, 0, 0, q_{10}, 0, -1, 0, q_6 - 1, 0, 0, 0, q_{14}, 0);\\
&\text{Column } 15: \widehat{\mathbf{q_{15}}} = (0, 0, 0, q_4, 0, 0, 0, q_{12}, 0, 0, 0, q_8, -1, 0, -1, q_{16}).
\end{aligned}$$

Then the desired $D(\mathbf{p}, \mathbf{q}, \mathbf{f})$ that derived from $\boldsymbol{\mathcal{L}}$ is a just a matrix (i) composed of these columns; (ii) with another column of $\boldsymbol{\mathcal{L}}$ replaced by the transpose of an arbitrary vector $\mathbf{f} = (f_1, f_2, ..., f_{16})$; (iii) with the remaining seven columns being the same as $\boldsymbol{\mathcal{L}}$. Then similar to Proposition 2.2 and Equations (2.3) and (2.4), if $\mathbf{w}$ is a nonzero multiple of $\boldsymbol{\pi}$, we have $\mathbf{w} \cdot \mathbf{f} = \eta D(\mathbf{p}, \mathbf{q}, \mathbf{f})$ for some $\eta \neq 0$ only depending on $\mathbf{w}$, $s_X = \frac{D(\mathbf{p}, \mathbf{q}, \mathbf{S_X})}{D(\mathbf{p}, \mathbf{q}, \mathbf{1})}$ and $s_Y = \frac{D(\mathbf{p}, \mathbf{q}, \mathbf{S_Y})}{D(\mathbf{p}, \mathbf{q}, \mathbf{1})}$. Hence, Equation (2.5) hold as well. To enforce this linear relation $\alpha s_x + \beta s_Y + \gamma = 0$, one needs to choose a strategy such that $D(\mathbf{p}, \mathbf{q}, \alpha \mathbf{S_X} + \beta \mathbf{S_Y} + \gamma \mathbf{1}) = 0$. Such a strategy, if feasible, is a generalized version of the zero-determinant strategy discussed in the previous sections.

To see whether the feasible generalized zero-determinant strategies indeed exist, we try to find it for X. To make $D(\mathbf{p}, \mathbf{q}, \alpha \mathbf{S_X} + \beta \mathbf{S_Y} + \gamma \mathbf{1}) = 0$ equal to zero, X should choose $\mathbf{p}$ such that $a\widetilde{\mathbf{p_2}} + b\widetilde{\mathbf{p_6}} + c\widetilde{\mathbf{p_{10}}} + d\widetilde{\mathbf{p_{14}}} = \alpha \mathbf{S_X} + \beta \mathbf{S_Y} + \gamma \mathbf{1}$ for some a, b, c, d. That is,

$$\begin{aligned}
&(\alpha + \beta)R + \gamma = a(p_1 - 1) = ap_5 - b = ap_9 - c = ap_{13} - d,\\
&\alpha S + \beta T + \gamma = -a + bp_2 = b(p_6 - 1) = bp_{10} - c = bp_{14} - d,\\
&\alpha T + \beta S + \gamma = cp_3 = cp_7 = cp_{11} = cp_{15},\\
&(\alpha + \beta)P + \gamma = dp_4 = dp_8 = dp_{12} = dp_{16}.
\end{aligned}$$

The above equations can give a solution of $\mathbf{p}$, expressed in terms of a, b, c, d in addition to $R, S, T, P, \alpha, \beta$ and γ. Then it is possible to choose some suitable values for a, d, c, d such that $p_i \in [0, 1]$, $\forall i = 1, 2, ...16$. Then this is a feasible strategy. As an example, if we take the conventional values $(T, R, P, S) = (5, 3, 1, 0)$, then X can unilaterally enforce a linear relation $s_X - 2s_Y + 2 = 0$ by adopting the memory-two strategy $\mathbf{p} = (\frac{11}{12}, \frac{4}{11}, \frac{7}{9}, \frac{1}{10}, \frac{5}{6}, \frac{3}{11}, \frac{7}{9}, \frac{1}{10}, \frac{2}{3}, \frac{1}{11}, \frac{7}{9}, \frac{1}{10}, \frac{3}{4}, \frac{2}{11}, \frac{7}{9}, \frac{1}{10})$ Nevertheless, similar to the memory-one case, there may not be any feasible zero-determinant strategies for some values of α, β and γ. In other words, a player can enforce some but not all linear relations of the long run expected payoffs.

Acknowledgments. It is a pleasure to thank my mentor, Jacob Perlman, for his guidance and help with this paper throughout the REU program. I would also like to thank Prof. Peter May for organizing this program and for his great encouragement. In addition, I express gratitude to Prof. Shiu-Yuen Cheng for recommending me this topic.

References

[1] Mark A. Pinsky, Samuel Karlin. An Introduction to Stochastic Modeling, Fourth Edition. Academic Press, 2010.
[2] David A. Levin, Yuval Peres, Elizabeth L. Wilmer. Markov Chains and Mixing Times. American Mathematical Society, 2008.
[3] E. Seneta. Non-negative Matrices and Markov Chains, Second Edition. Springer-Verlag New York Inc., 2006.
[4] Steven P. Lalley. Markov Chains: Basic Theory. (Lecture notes, The University of Chicago, accessed on Aug. 1, 2014.) http://galton.uchicago.edu/∼lalley/Courses/312/MarkovChains.pdf
[5] William H. Press, Freeman J. Dyson. Iterated prisoner's dilemma contains strategies that dominate any evolutionary opponent. PNAS **109**(26)(2012), 10409-10413.
[6] Ethan Akin. Good strategies for the iterated prisoner's dilemma. arXiv:1211.0969v2 [math.DS], 2013, 42 pages.