International Game Theory Review
Vol. 22, No. 2 (2020) 2040002 (27 pages)

DOI: 10.1142/S0219198920400022

On Finding Large Sets of Rewards in Two-Player ETP–ESP Games

Reinoud Joosten*
IEBIS, BMS, University of Twente
POB 217, 7500 AE Enschede, The Netherlands
r.a.m.g.joosten@utwente.nl

Llea Samuel
Coöperatieve Rabobank U.A.
Croeselaan 18, 3521 CB Utrecht, The Netherlands

Received 10 August 2018
Revised 27 April 2019
Accepted 18 January 2020
Published 4 May 2020

Games with endogenous transition probabilities and endogenous stage payoffs (or ETP–ESP games for short) are stochastic games in which both the transition probabilities and the payoffs at any stage are continuous functions of the relative frequencies of all past action combinations chosen.

We present methods to compute large sets of jointly-convergent pure-strategy rewards in two-player ETP–ESP games with communicating states under the limiting average reward criterion. Such sets are useful in determining feasible rewards in a game, and instrumental in obtaining the set of (Nash) equilibrium rewards.

Keywords: Stochastic games; average rewards; endogenous transition probabilities and stage payoffs; (non)homogeneous Markov chains.

Mathematics Subject Classification 2020: 91A15, 91A80, 60J10

1. Introduction

Our main purpose is to present several methods to compute large sets of feasible rewards in games with endogenous transition probabilities and endogenous stage payoffs (ETP–ESP games, see Joosten and Samuel [2017]). Such nonterminating stochastic games combine and generalize ESP games in which the stage payoffs may change as a result of the history of the play until then, and ETP games in which the transition probabilities may change as a result of the history of the play until then. FD games, i.e., games with frequency-dependent stage payoffs, are a

*Corresponding author.

subclass of the former (cf., Brenner and Witt [2003] and Joosten *et al.* [2003]), whereas Joosten and Meijboom [2010] introduced the latter class.

We aim to contribute to filling what we see as a void in the field of stochastic games, namely that there seems little work on establishing their feasible rewards. Computing often seems concerned with finding one or a few feasible rewards with attractive features, such as for Nash equilibria. In many cases, even computing only a limited number of rewards is very involved and various ardent restrictions on strategies or the structure of the games under consideration are required (see e.g., Raghavan and Filar [1991], Filar and Vrieze [1996], Vrieze [1981, 1987] and Vrieze *et al.* [1983]).

In recent years, a series of contributions introducing several types of stochastic games engineered in the sense of Aumann [2008] has appeared. Early contributions in this series were generalizations of repeated games called FD games. Joosten *et al.* [2003] presented a method of analysis for FD games akin to the one for the Folk theorems for repeated and stochastic games (cf., e.g., Aumann [1981], Hart [1985], Forges [1986], Dutta [1995] and Thuijsman and Vrieze [1998]). Such methods yield large sets of rewards corresponding to Nash equilibria involving threats assuming all agents wish to maximize their average[a] payoffs independently over an infinite time horizon.

Central in Joosten *et al.* [2003] were jointly-convergent (pure) strategies, i.e., strategy combinations inducing play such that the relative frequency of each possible action combination converges with probability one as time goes to infinity. The approach led to new developments regarding the analysis of repeated games with vanishing actions [Joosten *et al.*, 1995; Joosten, 2005] providing alternatives to methods of analysis developed independently by Schoenmakers *et al.* [2002], Schoenmakers [2004] and Joosten [1996]. It was crucial in a series of contributions on environmental pollution problems (e.g., Joosten [2019]), competitive advertising combined with Cournot or Bertrand competition (e.g., Joosten [2009, 2015]) and fishery games (e.g., Joosten [2007a, 2007b, 2014, 2016] and Joosten and Samuel [2017]).

Determining the set of jointly-convergent pure-strategy rewards under the limiting average reward criterion is useful to establish which long run average payoffs are feasible in a given game. In a standard repeated game, this set is the convex hull of the rewards corresponding to the entries of its payoff matrix. For FD games however, visualizations provide the insight that this set need not be convex, let alone the convex hull of a finite number of rewards. Joosten *et al.* [2003] prove by construction that the convex hull of the set of jointly-convergent pure-strategy rewards is feasible for the class of games studied there as well.

The set of jointly-convergent pure-strategy rewards is also useful in determining (Nash) equilibrium rewards in the fashion of the Folk Theorem. Rewards giving each player at least the reward he can maximally obtain if all opponents collectively try to

[a]Technically, we take the lim inf of the average payoffs as time goes to infinity.

minimize his long term average payoffs are called individually rational. Joosten *et al.* [2003] showed that the convex hull of all individually-rational jointly-convergent pure-strategy rewards can be obtained by an equilibrium involving threats.

For the purpose of obtaining the set of jointly-convergent pure-strategy rewards, it seems natural to turn to computational methods. Early endeavors imposed a grid on the space of the originals, i.e., a unit simplex, and calculated the associated rewards for each grid point (e.g., Joosten *et al.* [2003]). The ease of use grew when replacing the grid-based approach by randomized methods to obtain originals (e.g., Joosten [2016, 2015]).

A computational procedure supporting Joosten and Meijboom [2018], additionally checks all the so-called balance equations implied in the context of Markov chains (cf., e.g., Kemeny and Snell [1976]). If one of these balance equations is violated, the algorithm discards the original and moves on to generate another candidate. Results were obtained quickly with this innovation for repeated games and certain stochastic games in Joosten and Meijboom [2010]. Finding large sets of rewards took slightly more time for stochastic games in which transition probabilities satisfy the Markov property.

For stochastic games with endogenous transitions, however, the very object of interest and analysis in Joosten and Meijboom [2018], computations, took an excessive amount of time. For instance, it might take a couple of days to obtain a certain (large) number of pairs of rewards, where finding similarly large sets would take a couple of seconds for FD games, to a couple of minutes for certain "standard" stochastic games with an ergodic aperiodic Markov chain encompassing all states. This disappointing finding made chances for further progress look dim for quite some time.

Joosten and Samuel [2017] extended the framework of Joosten and Meijboom [2018] by combining it with the framework of Joosten *et al.* [2003]. For this, it was imperative to improve the efficiency of the computational procedures first. The difference between Joosten and Samuel [2017] and this paper is, that the former limits the scope of the setting as its goal is to introduce a new type of stochastic game first and foremost, and the computational method remains hidden. Here, we provide our methods, improvements and subsequent insights for more general settings.

Next, we introduce our model and give an example. In Sec. 3, we describe the computational procedures for the three different types of stochastic games. Section 4 treats different steps to enhance efficiency. Section 5 discusses generalizations. Section 6 concludes.

2. Application for Stochastic Games

We briefly[b] introduce a stochastic game with frequency-dependent stage payoffs and frequency-dependent transition probabilities from Joosten and Samuel [2017]

[b]For more elaborate definitions, modeling backgrounds and motivations, we gladly refer to Joosten *et al.* [2003], Joosten and Meijboom [2018] and Joosten and Samuel [2017].

for the sake of clarity and discussion. The game has two players A and B, two states ω^1 and ω^2, and in both states, both players have two actions each. Player A chooses top or bottom row, his opponent B chooses left or right column.[c]

The game is played at discrete moments in time called stages, the play continues forever at stages $t = 1, 2, \ldots$ As we have both the stage payoffs and transition probabilities being determined endogenously, i.e., by past play, let us, first, capture this formally. The past play until stage t, $t > 1$, is captured by two relative frequency matrices

$$X^{s,t} = \begin{bmatrix} x^{s,t}_{1,1} & x^{s,t}_{1,2} \\ x^{s,t}_{2,1} & x^{s,t}_{2,2} \end{bmatrix}, \quad s = 1, 2.$$

Here, e.g., $x^{1,t}_{1,1}$ is the relative frequency of action pair top-left in state ω^1 having occurred until stage t and $x^{2,t}_{2,1}$ is the relative frequency of action pair bottom-left in state ω^2 in past play. By logic, $x^t = (x^{1,t}_{1,1}, \ldots, x^{1,t}_{2,2}, x^{2,t}_{1,1}, \ldots, x^{2,t}_{2,2}) \in \Delta^7 = \{y \in \mathbb{R}^8 | y_i \geq 0 \text{ for all } i = 1, \ldots, 8 \text{ and } \sum_{j=1}^{8} y_j = 1\}$. Vector x^t is called a *relative frequency vector*.

The interaction during the play is represented by the following two matrices:

$$\widetilde{\omega}^s(\cdot) = \begin{bmatrix} (\theta^s_{1,1}(\cdot), p^s_{1,1}(\cdot)) & (\theta^s_{1,2}(\cdot), p^s_{1,2}(\cdot)) \\ (\theta^s_{2,1}(\cdot), p^s_{2,1}(\cdot)) & (\theta^s_{2,2}(\cdot), p^s_{2,2}(\cdot)) \end{bmatrix}, \quad s = 1, 2.$$

Here, all functions $p^s_{i,j} : \Delta^7 \to [0, 1]$, $\theta^s_{i,j} : \Delta^7 \to \mathbb{R}_+ \cup \{0\}$ are assumed continuous for all $i, j, s = 1, 2$.

Matrix $\widetilde{\omega}^s$ incorporates everything necessary for the description of the play occurring in and from state ω^s at a certain stage t. Say that the play is in state ω^s at stage t, then the relative frequency vector x^t is known. Then each entry of $\widetilde{\omega}^s$ contains an ordered pair $(\theta^s_{i,j}(x^t), p^s_{i,j}(x^t))$ denoting the immediate payoffs to the players $\theta^s_{i,j}(x^t) = (\theta^{s,A}_{i,j}(x^t), \theta^{s,B}_{i,j}(x^t))$, and the probability vector $p^s_{i,j}(x^t) = (p^{s,1}_{i,j}(x^t), p^{s,2}_{i,j}(x^t)) = (p^{s,1}_{i,j}(x^t), 1 - p^{s,1}_{i,j}(x^t))$ that the system moves to ω^1, ω^2, respectively, at stage $t + 1$, if the corresponding action pair (i, j) is chosen in the current state t.

To give further insights, we provide the same functions as in Joosten and Samuel [2017] and explain briefly the following:

$$\theta^1_{1,1}(x) = 4\theta^2_{1,1}(x) = (16, 16)c(x), \quad \theta^1_{1,2}(x) = 4\theta^2_{1,2}(x) = (14, 28)c(x),$$

$$\theta^1_{2,1}(x) = 4\theta^2_{1,3}(x) = (28, 14)c(x), \quad \theta^1_{2,2}(x) = 4\theta^2_{1,4}(x) = (24, 24)c(x),$$

$$c(x) = 1 - \frac{x^1_{1,2} + x^1_{2,1} + 2(x^2_{1,2} + x^2_{2,1})}{4} - \frac{x^1_{2,2} + 2x^2_{2,2}}{3}.$$

Note that the stage payoffs in ω^1 (or High) are four times the payoffs in ω^2 (or Low) *ceteris paribus*; $c(x)$ is common to all stage payoff functions. So, the stage

[c]We may also refer to actions by numbers 1, 2 indicating top or bottom, respectively, for Player A and left or right, respectively, for Player B.

payoffs at stage t, if $x^t = x$, depend on the eight elements of the relative frequency vector x. Clearly, $c(x) = 1$ if and only if $x^1_{1,1} + x^2_{1,1} = 1$, $c(x) = \frac{1}{3}$ if and only if $x^2_{2,2} = 1$, so, $\frac{1}{3} \leq c(x) \leq 1$. Let

$$P = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.35 & 0.3 & 0.3 & 0.25 & 0.2 & 0.15 & 0.15 & 0.05 \\ 0.35 & 0.3 & 0.3 & 0.25 & 0.2 & 0.15 & 0.15 & 0.5 \\ 0.7 & 0.6 & 0.6 & 0.5 & 0.4 & 0.3 & 0.3 & 0.1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.35 & 0.3 & 0.3 & 0.25 & 0.2 & 0.15 & 0.15 & 0.05 \\ 0.35 & 0.3 & 0.3 & 0.25 & 0.2 & 0.15 & 0.15 & 0.5 \\ 0.7 & 0.6 & 0.6 & 0.5 & 0.4 & 0.3 & 0.3 & 0.1 \end{bmatrix}.$$

With respect to the transition probabilities, let furthermore

$$\begin{aligned} p^{s,1}(x) &= \pi^{s,1} - x \cdot P = (p^{1,1}_{1,1}(x), \ldots, p^{1,1}_{2,2}(x), p^{2,1}_{1,1}(x), \ldots, p^{2,1}_{2,2}(x)), \\ p^{s,2}(x) &= (1 - p^{1,1}_{1,1}(x), \ldots, 1 - p^{1,1}_{2,2}(x), 1 - p^{2,1}_{1,1}(x), \ldots, 1 - p^{2,1}_{2,2}(x)), \\ \text{with } \pi^{s,1} &= (0.8\ 0.7\ 0.7\ 0.6\ 0.5\ 0.4\ 0.4\ 0.15). \end{aligned}$$

The continued use of action pair $(1, 1)$ in both states does not impair the transition probabilities to go to ω^1. All other action pairs decrease the transition probabilities to move to ω^1.

The stage payoffs and the transition probabilities depend on the vector of relative frequencies capturing the play of the agents until the present stage. These rather simple linear functions $p^{k,l}(\cdot)$ can be generalized considerably, but the results depend crucially on continuity (cf., e.g., Joosten *et al.* [2003] and Joosten and Meijboom [2018]).

2.1. *Notations for the general case*

In the example, we used the three twos: two players, two states and two actions per player. The following pertains to the general two-player case and this has some implications for notations.
State space: $\Omega = \{\omega^1, \ldots, \omega^N\}$.
Actions in state ω^k: $i \in I^k = \{1, \ldots, m^k\}$, $j \in J^k = \{1, \ldots, n^k\}$.
Entry/action pair in state ω^k: $(i, j) \in I^k \times J^k$.
Number of entries/action pairs in state $\omega^k \in \Omega : m^k \cdot n^k$.
Total number of entries: $\#E = \sum_{k=1}^{N} m^k \cdot n^k$.
Relative frequency of action pair (i, j) in state ω^k: $x^k_{i,j}$.
For state ω^k: $(x^k) = (x^k_{1,1}, \ldots, x^k_{1,m^k}, \ldots, x^k_{n^k,1}, \ldots, x^k_{n^k,m^k})$.

Relative frequency vector: $x = (x^1, \ldots, x^N) \in \Delta^{\#E-1}$.
Relative state frequency for ω^k given x: $X^k(x) = \sum_{i \in I^k} \sum_{j \in J^k} x^k_{i,j}$.
Relative state frequency vector given x: $X(x) = (X^1(x), \ldots, X^N(x)) \in \Delta^{N-1}$.
Normalized relative frequency vector given x: $y(x) = (\frac{1}{X^1(x)} x^1, \ldots, \frac{1}{X^N(x)} x^N)$.

Transition probability from state ω^k to state ω^l if action pair $(i,j) \in I^k \times J^k$ is played in ω^k given x : $p^{k,l}_{i,j}(x)$.
Payoffs of action pair (i,j) in state ω^k given x: $\theta^k_{i,j}(x)$.
Vector of payoffs in state ω^k given x:

$$(\theta^k(x)) = (\theta^k_{1,1}(x), \ldots, \theta^k_{1,m^k}(x), \ldots, \theta^k_{n^k,1}(x), \ldots, \theta^k_{n^k,m^k}(x)).$$

Transition probability from state ω^k to ω^l given x:

$$P^{k,l}(x) = \sum_{i=1}^{m^k} \sum_{j=1}^{n^k} x^k_{i,j} \cdot p^{k,l}_{i,j}(x).$$

Normalized transition probability from state ω^k to ω^l given x:

$$F^{k,l}(x) = \sum_{i=1}^{m^k} \sum_{j=1}^{n^k} y^k_{i,j}(x) \cdot p^{k,l}_{i,j}(x).$$

Balance equation for state k given x:

$$\sum_{l \neq k} P^{k,l}(x) = \sum_{l \neq k} P^{l,k}(x) \Leftrightarrow X^k \sum_{l \neq k} F^{k,l}(x) = \sum_{l \neq k} X^l F^{l,k}(x).$$

Vector of transition probabilities in state ω^k given x:

$$(p^k(x)) = (p^{k,1}_{1,1}(x), \ldots, p^{k,N}_{1,m^k}(x), \ldots, p^{k,1}_{n^k,1}(x), \ldots, p^{k,N}_{n^k,m^k}(x)).$$

Limiting average rewards given x: $\gamma(x) = \sum_{k=1}^{N} \sum_{i=1}^{m^k} \sum_{j=1}^{n^k} x^k_{i,j} \cdot \theta^k_{i,j}(x)$.
Vector of transition probabilities given x: $p(x) = (p^1(x), \ldots, p^N(x))$.
Vector of payoffs given x: $\theta(x) = (\theta^1(x), \ldots, \theta^N(x))$.
System of balance equations given x:

$$\left.\begin{array}{c} \sum_{l \neq 1} P^{1,l}(x) = \sum_{l \neq 1} P^{l,1}(x) \\ \vdots \\ \sum_{l \neq N} P^{N,l}(x) = \sum_{l \neq N} P^{l,N}(x) \end{array}\right\} \Leftrightarrow \left\{\begin{array}{c} X^1 \sum_{l \neq 1} F^{1,l}(x) = \sum_{l \neq 1} X^l F^{l,1}(x), \\ \vdots \\ X^N \sum_{l \neq N} F^{N,l}(x) = \sum_{l \neq N} X^l F^{l,N}(x). \end{array}\right.$$

Remark 1. We assume[d] that an $\varepsilon > 0$ and number Z exist such that for all $x \in \Delta^{\#E-1}$, it holds that the transition matrix

$$F(x) = \begin{bmatrix} F^{1,1}(x) & \cdots & F^{1,N}(x) \\ \vdots & \ddots & \vdots \\ F^{N,1}(x) & \cdots & F^{N,N}(x) \end{bmatrix},$$

satisfies $[F(x)^z]_{k,l} > \varepsilon$ for all integers $z \geq Z$ and for all $k, l \in N$. Therefore, all components of the matrix $F(x)^z$, the zth power of $F(x)$, are bounded away from zero for all z large enough.

Remark 2. Let $N \times N$-matrix $A(x)$ for given x, be defined by

$$A_{i,j}(x) = \begin{cases} 1 & \text{if } i = N, \\ 1 - F^{i,i}(x) & \text{if } i \neq N,\ j = i, \\ -F^{j,i}(x) & \text{otherwise.} \end{cases}$$

Suppose $Q(x) = (Q^1(x), \ldots, Q^N(x))$ satisfies

$$A(x) \cdot Q(x)^\top = \begin{bmatrix} 0^{N-1} \\ 1 \end{bmatrix}, \tag{1}$$

where 0^{N-1} is an $(N-1)$-vector of zeros. Then $Q(x)$ is the solution to the system of balance equations, as well as $\sum_{k=1}^{N} Q^k(x) = 1$.

Remark 3. Remark 1 guarantees that the Markov chain at hand is irreducible and aperiodic, or stated differently, the set of all states is ergodic. $Q(x)$ is called the *stationary distribution*. The stationary distribution is unique and positive in all components (cf., e.g., Kemeny and Snell [1976]). Remark 1 also implies that $A^{-1}(x)$ exists, $Q(x)^\top$ is its last column by (1) and $Q(x)$ equals any row of $[F(x)]^\infty \equiv \lim_{t\to\infty}[F(x)]^t$.

Remark 4. For $x \in \Delta^{\#E-1}$, let $Q(x)$ satisfy (1), then let the associated *rescaled* relative frequency vector $\widetilde{x} \equiv (\frac{Q(x)^1}{X^1(x)}x^1, \ldots, \frac{Q(x)^N}{X^N(x)}x^N)$.

2.2. *Strategies, rewards and equilibria*

Since a strategy is a game plan for the entire infinite time horizon, allowing it to depend on any condition makes a comprehensive analysis of infinitely repeated games quite impossible. Most restrictions in the literature put requirements on what aspects the strategies are conditional upon. For instance, a *history-dependent* strategy prescribes a possibly mixed action to be played at each stage conditional on the current stage and state, as well as on the full history until then, i.e., all states

[d]A referee suggested this elegant formulation, the current form implies our earlier one by compactness of $\Delta^{\#E-1}$ and continuity of the functions involved.

visited and all action combinations already realized. Less general strategies are for instance, *action independent* ones conditioning on all states having been visited before, but *not* on the action combinations chosen (cf., Herings and Predtetchinski [2012]); *Markov strategies* condition on the current state and stage; *stationary strategies* condition on the present state (cf., e.g., Filar and Vrieze [1996]). Let $\mathcal{X}^A$ and $\mathcal{X}^B$ denote the set of history-dependent strategies of player A, B, respectively.

A strategy is *pure*, if at *each* stage a *pure action* is chosen. The set of pure strategies for player $A(B)$ is $\mathcal{P}^A(\mathcal{P}^B)$ and $\mathcal{P} \equiv \mathcal{P}^A \times \mathcal{P}^B$.

The strategy pair $(\pi, \sigma) \in \mathcal{X}^A \times \mathcal{X}^B$ is *jointly convergent* if and only if $x \in \Delta^{\#E-1}$ exists such that for all $\varepsilon > 0$, $i \in \{1, 2, \ldots, \#E\}$:

$$\limsup_{t\to\infty} \Pr_{\pi,\sigma}[|x_i^t - x_i| \geq \varepsilon] = 0. \tag{14}$$

$\Pr_{\pi,\sigma}$ denotes the probability under strategy pair (π, σ). $\mathcal{JC}$ denotes the set of jointly-convergent strategy pairs. Under such a pair of strategies, the relative frequency of each action pair in both states as play goes to infinity, converges to a fixed number with probability 1 in the terminology of Billingsley [1986, p. 274]). Obviously, $\mathcal{JC} \subset \mathcal{X}^A \times \mathcal{X}^B$.

The players receive an infinite stream of stage payoffs, they are assumed to wish to maximize their average rewards. For a given pair of strategies (π, σ), $R_t^A(\pi, \sigma)$ $(R_t^B(\pi, \sigma))$ is the expected payoff to player $A(B)$ at stage t under strategy combination (π, σ). Player A's *average reward*, is $\gamma^A(\pi, \sigma) = \liminf_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} R_t^A(\pi, \sigma)$, his opponent gets average rewards $\gamma^B(\pi, \sigma)$ defined similarly, *mutatis mutandis*, and $\gamma(\pi, \sigma) \equiv (\gamma^A(\pi, \sigma), \gamma^B(\pi, \sigma))$.[e]

The *set of jointly-convergent pure-strategy rewards* $P^{\mathcal{JC}}$ is, then, the set of pairs of rewards obtained by using a pair of jointly-convergent pure strategies. For vector $x \in \Delta^{\#E-1}$, the *x-averaged payoffs* $(a, b)_x$ are given by

$$(a, b)_x = \sum_{i=1}^{8} x_i \theta_i(x).$$

The next result connects notions introduced in the the preceding parts of this section.

Proposition 1 (Joosten and Samuel [2017]). *Let strategy pair* $(\pi, \sigma) \in \mathcal{JC}$ *and let* $x \in \Delta^{\#E-1}$ *for which* (2) *is satisfied, then the average payoffs are given by* $\gamma(\pi, \sigma) = (a, b)_x$.

The strategy pair $(\pi^*, \sigma^*) \in \mathcal{X}^A \times \mathcal{X}^B$ is *a* $(Nash)$ *equilibrium* if and only if

$$\gamma^A(\pi^*, \sigma^*) \geq \gamma^A(\pi, \sigma^*) \text{ for all } \pi \in \mathcal{X}^A,$$
$$\gamma^B(\pi^*, \sigma^*) \geq \gamma^B(\pi^*, \sigma) \text{ for all } \sigma \in \mathcal{X}^B.$$

[e]An anonymous referee recommended to add that these rewards do not depend on where the play starts, whereas in general stochastic games they very well might.

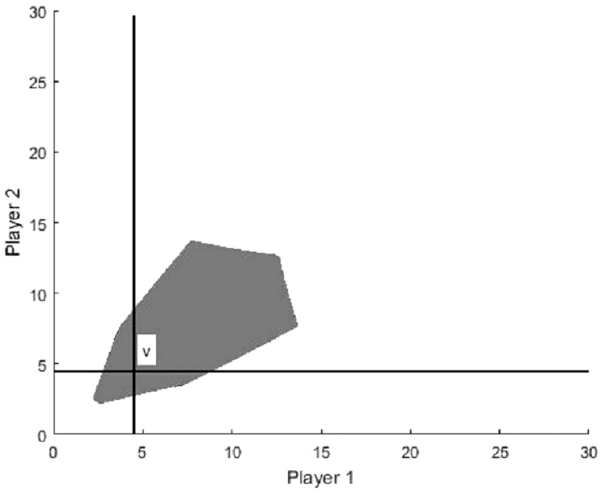

Fig. 1. The set $P^{\mathcal{JC}}$ in Joosten and Samuel [2017]. Each reward to the North East of the threat point, v, the intersection of the two lines, can be supported by a Nash equilibrium involving threats.

We call $v = (v^A, v^B)$ the *threat point*, where $v^A = \min_{\sigma \in \mathcal{X}^B} \max_{\pi \in \mathcal{X}^A} \gamma^A(\pi, \sigma)$ and $v^B = \min_{\pi \in \mathcal{X}^A} \max_{\sigma \in \mathcal{X}^B} \gamma^B(\pi, \sigma)$. So, v^A is the highest amount A can get if B tries to minimize A's average payoffs. A pair of *individually-rational* (feasible) rewards attributes each player at least the threat-point reward. Let E be defined as follows:

$$E = \{\gamma(\pi, \sigma) \in P^{\mathcal{JC}} \,|\, \gamma^A(\pi, \sigma) > v^A,\ \gamma^B(\pi, \sigma) > v^B\}.$$

Then the following result has been established.

Theorem 1 (Joosten and Samuel [2017]). *Each pair of rewards in the convex hull of E can be supported by an equilibrium.*

The latter two formal results are illustrated in Fig. 1.

3. Description of the Computational Setup

Before we describe the ideas behind the computations, we make some reservations. For a large part, we relied on Matlab. Our description therefore will abstain from detailed computational procedures hidden to (the average) us(er). We are just modelers facing challenges requiring computational solutions, we leave the development of more efficient (Matlab) codes to more knowledgeable programmers.

3.1. *A computation-inspired ordering*

In Sec. 2, we presented a stochastic game in which both the transition probabilities and the stage payoffs are determined endogenously as these may depend on the relative frequencies of all possible action combinations in all states realized. Based

Table 1. A computation-inspired ordering among stochastic games.

	CTP		ETP
	Type I (p_0)	Type II (x, p_0)	Type III ($x, p(x)$)
CSP (θ_0)	AIT games	Stochastic games*	CAIT games ETP games*
ESP ($\theta(x)$)	FD games* AIT FD games	Stochastic FD games*	CAIT–ESP games ETP–ESP games

on our computational experience, we came up with an ordering of stochastic games shown in Table 1.

In Table 1, subscript zero indicates that the object at hand is constant over time. CTP (respectively, ETP) on top denotes constant (respectively, endogenous) transition probabilities; CSP (respectively, ESP) on the left denotes constant (respectively, endogenous) stage payoffs. The term endogenous with regards to both functions means dependent on history of the play as demonstrated in the subsection 2.1. CTP (respectively, CSP) games can be seen as invariant special cases of ETP (respectively, ESP) games. Both Type-I and Type-II classes belong to the CTP part. Type-III models have as a special feature that transition probabilities depend on the relative frequency vectors representing past play.

Type-I models have transition probabilities from state to state which only depend on the state the play is in, i.e., $p^{k,l}_{i,j}(x) = p^{k,l}$ for all $i, j \in m^k \times n^k$, $k, l \in N$, $x \in \Delta^{\#E-1}$, hence, independent from the actions chosen and independent from the history of the play. Type-II models have transition probabilities which not only depend on the current state but also on the actions chosen by the agents in that state, i.e., $p^{k,l}_{i,j}(x) = p^{k,l}_{i,j}$ for all $i, j \in m^k \times n^k$, $k, l \in N$, $x \in \Delta^{\#E-1}$, yet independent from the history of the play. Type-III models have transition probabilities $p^{k,l}_{i,j}(x)$ depending on the current state, i.e., ω^k, $k \in N$, the actions $i, j \in m^k \times n^k$ chosen in ω^k as well as the history of the play condensed in $x \in \Delta^{\#E-1}$. Every Type-I model is a special case of a Type-II model and every Type-II model is a special case of a Type-III model.

The asterisks denote names already taken: stochastic games (in the narrow sense) were introduced by Shapley [1953] and the nonterminating version by Gillette [1957], FD games by Brenner and Witt [2003], ETP games by Joosten and Meijboom [2010, 2018], ETP–ESP games by Joosten and Samuel [2017]. Mahohoma [2014] introduced the term stochastic FD games.

AIT games[f] are stochastic games with action-independent transition probabilities, i.e., the latter do not depend on the actions chosen by the agents, neither

[f]The name was inspired by a similar term in Herings and Predtetchinski [2012].

present actions, nor past ones. They may however, depend on the current state. Rather trivial AIT games are repeated games as there is only one state for a such a game. A less trivial example is a game in which the play follows a random walk over N states, i.e., if the play is in state k at stage t, then at stage $t+1$ the play will continue in state $k-1$ (modulo N) with probability p and with the complementary probability in state $k+1$ (modulo N) (cf., e.g., Kemeny and Snell [1976]).

Another example of an AIT game would be a two-player game in which the stage payoffs at stage $t = 1, 2, \ldots$ of the play are given by a bi-matrix $(\cos\frac{2\pi(t-1)}{k}A, \cos\frac{2\pi(t-1)}{k}B)$ for some fixed integer k and a pair of fixed matrices A and B. In that case, the number of distinct states is (at most) k, transitions are deterministic and independent from the current action choices. So, the play returns cyclically[g] to the different states.

CAIT games are stochastic games with current-action-independent transition probabilities, i.e., the current transition probabilities do not depend on the current actions chosen by the agents. They may however depend on the current state, and on the actions played by the agents in the past.

In Type-I and Type-II models, determining the stationary distribution may require some computational efforts, but the solution found is exact. For Type-III models, the stationary distribution depends on the relative frequency vector in combination with the transition probabilities which in turn are determined by the relative frequency vector. So, determining a stationary distribution is like shooting at a moving target. We find a stationary distribution solving system (1), rescale the relative frequency vector using this stationary distribution inducing new transition probabilities, find another stationary distribution solving the new system (1) and so on.

In our opinion, *all games above* are stochastic games in the broad sense, e.g., all FD games and all AIT games are stochastic games, too. The term coined by Mahohoma [2014] is a tautology in the broad interpretation. The question whether the stage payoffs depend on the relative frequency vector, turned out to be of minor consequence for the efficiency and complexity of the computational procedures. Hence, it seems sufficient to just have three instead of six types of games in our ordering of stochastic games.

3.2. *Computational procedures*

The computations for the general Type-III model will also give the desired answers to the Type-I and Type-II models. However, the efficiency of computations for these other types can be improved significantly, so we present these computations as well.

First, we have to decide how many pairs of jointly-convergent pure strategy rewards we wish to generate. Let us call this integer V. The transition-probabilities

[g]Uyttendaele *et al.* [2012] introduced evolutionary games in which the underlying fitness matrices are periodic in continuous time. Discrete-time versions of their underlying games with time-dependent payoff (fitness) bi-matrices could be interpreted as AIT games.

function $p(\cdot)$ and the stage-payoff functions $\theta(\cdot)$ are known, i.e., they form the primitives for the calculations.

3.2.1. *Type*-I *models*

Draw $x(1) = (x^1(1), \ldots, x^N(1)) \in \Delta^{\#E-1}$ at random and compute

$$y(x(1)) := \left(\frac{1}{X^1(x(1))} x^1(1), \ldots, \frac{1}{X^N(x(1))} x^N(1) \right),$$
$$F^{k,l}(x(1)) \ := p^{k,l}.$$

Solve the following system to find stationary distribution $Q(x(1))^\top$,

$$A(x(1)) \cdot Q(x(1)) = \begin{bmatrix} 0^{N-1} \\ 1 \end{bmatrix}.$$

Set

$$x(1) := (Q^1(x(1))y^1(x(1)), \ldots, Q^N(x(1))y^N(x(1))),$$
$$\theta(x(1)) := (\theta^1(x(1)), \ldots, \theta^N(x(1))),$$
$$\gamma(x(1)) := \sum_{k=1}^{N} \sum_{i=1}^{m^k} \sum_{j=1}^{n^k} x_{i,j}^k(1) \cdot \theta_{i,j}^k(x(1)).$$

Next, save $Q := Q(x(1))$, $x(1)$ and $\gamma(x(1))$. Enter Loop I with $v := 2$.

Loop I. Draw $x(v)$ at random, and set/compute

$$x(v) := \left(\frac{Q^1}{X^1(x(v))} x^1(v), \ldots, \frac{Q^N}{X^N(x(v))} x^N(v) \right),$$
$$\theta(x(v)) := (\theta^1(x(v)), \ldots, \theta^N(x(v))),$$
$$\gamma(x(v)) := \sum_{k=1}^{N} \sum_{i=1}^{m^k} \sum_{j=1}^{n^k} x_{i,j}^k(v) \cdot \theta_{i,j}^k(x(v)).$$

Save $x(v)$ and $\gamma(x(v))$ for further use. If $v = V$, then *stop*. Otherwise, proceed with Loop I with $v := v + 1$.

Comment. Note that for AIT games, it holds that $\sum_{i=1}^{m^k} \sum_{j=1}^{n^k} y_{i,j}^k(x(1)) \cdot p_{i,j}^{k,l}(x(1)) = p^{k,l}$ as $p_{i,j}^{k,l}(\cdot) = p_{i',j'}^{k,l}(\cdot) \equiv p^{k,l}$ for all $(i,j),(i',j') \in I^k \times J^k$ and any state $k = 1, \ldots, N$. We compute the stationary distribution *once*, as the transition probabilities do not depend on the relative frequency vector selected randomly. All randomly drawn relative frequency vectors are rescaled with one and the same stationary distribution.

3.2.2. *Type*-II *models*

Start Loop II with $v := 1$.

Loop II. Draw $x(v) = (x^1(v), \ldots, x^N(v)) \in \Delta^{\#E-1}$ at random and compute

$$y(x(v)) := \left(\frac{1}{X^1(x(v))}x^1(v), \ldots, \frac{1}{X^N(x(v))}x^N(v)\right),$$

$$p(x(v)) := (p^1, \ldots, p^N),$$

$$F^{k,l}(x(v)) := \sum_{i=1}^{m^k}\sum_{j=1}^{n^k} y^k_{i,j}(x(v)) \cdot p^{k,l}_{i,j}.$$

Solve the following system to find stationary distribution $Q(x(1))^\top$.

$$A(x(v)) \cdot Q(x(v)) = \begin{bmatrix} 0^{N-1} \\ 1 \end{bmatrix}.$$

Set

$$x(v) := (Q^1(x(v))y^1, \ldots, Q^N(x(v))y^N).$$

$$\theta(x(v)) := (\theta^1(x(v)), \ldots, \theta^N(x(v))),$$

$$\gamma(x(v)) := \sum_{k=1}^{N}\sum_{i=1}^{m^k}\sum_{j=1}^{n^k} x^k_{i,j}(v) \cdot \theta^k_{i,j}(x(v)).$$

Then $x(v)$ and $\gamma(x(v))$ are saved for further use. If $v = V$, then *stop*. Otherwise, go to Loop II with $v := v + 1$.

Comment. Here, we used the fact that $p(x(v)) = p(x'(v))$ for all $x(v), x'(v)$. We compute *one* stationary distribution *for each random draw* of a relative frequency vector by solving a linear system of equations. Then we rescale the original draw using the stationary distribution to obtain the rewards corresponding to the rescaled relative frequency vector.

3.2.3. *Type*-III *models*

Perform Loop IIIa starting with $v := 1$.

Loop IIIa. Draw $x(v) = (x^1(v), \ldots, x^N(v)) \in \Delta^{\#E-1}$ at random and compute

$$y(x(v)) := \left(\frac{1}{X^1(x(v))}x^1(v), \ldots, \frac{1}{X^N(x(v))}x^N(v)\right),$$

$$p(x(v)) \ := \ = (p^1(x(v)), \ldots, p^N(x(v))),$$

$$F^{k,l}(x(v)) := \sum_{i=1}^{m^k}\sum_{j=1}^{n^k} y^k_{i,j}(x(v)) \cdot p^{k,l}_{i,j}(x(v)).$$

Then solve the system

$$A(x(v)) \cdot \widetilde{X}(x(v)) = \begin{bmatrix} 0^{N-1} \\ 1 \end{bmatrix}.$$

Set

$$\begin{aligned} x(v,0) &:= x(v), \\ Q(x(v,0)) &:= (X^1(x(v)), \ldots, X^N(x(v))), \\ x(v,1) &:= (\widetilde{X}^1(x(v))y^1, \ldots, \widetilde{X}^N(x(v))y^N), \\ Q(x(v,1)) &:= \widetilde{X}(x(k)). \end{aligned}$$

Then *enter Loop* III*b*, *starting with* $w := 1$.

Loop IIIb. If $||Q(x(v,w)) - Q(x(v,w-1))||_\infty < \varepsilon$, then compute

$$\theta(x(v,w)) := (\theta^1(x(v,w)), \ldots, \theta^N(x(v,w))),$$

$$\gamma(x(v,w)) := \sum_{k=1}^{N} \sum_{i=1}^{m^k} \sum_{j=1}^{n^k} x_{i,j}^k(v,w) \cdot \theta_{i,j}^k(x(v,w)).$$

Set

$$\begin{aligned} x(v) &:= x(v,w), \\ \theta(x(v)) &:= \theta(x(v,w)), \\ \gamma(x(v)) &:= \gamma(x(v,w)). \end{aligned}$$

Then $x(v)$ and $\gamma(x(v))$ are saved for further use. If $v = V$, then *stop*.
Otherwise proceed with *Loop* III*a*, *setting* $v := v + 1$.
Otherwise, if $w = W$, *proceed* with *Loop* III*a*, with $v := v$.
Otherwise, compute

$$p(x(v,w)) = (p^1(x(v,w)), \ldots, p^N(x(v,w))),$$

$$F^{k,l}(x(v,w)) = \sum_{i=1}^{m^k} \sum_{j=1}^{n^k} y_{i,j}^k(x(v,w)) \cdot p_{i,j}^{k,l}(x(v,w)).$$

Then solve the system

$$A(x(v,w)) \cdot \widetilde{X}(x(v,w)) = \begin{bmatrix} 0^{N-1} \\ 1 \end{bmatrix}.$$

Set

$$\begin{aligned} Q(x(v,w+1)) &:= \widetilde{X}(x(v,w))^\top, \\ x(v,w+1) &:= (\widetilde{X}^1(x(v,w))y^1, \ldots, \widetilde{X}^N(x(v,w))y^N), \end{aligned}$$

Set $w := w + 1$, and go back to *Loop* III*b*.

Comment. We generate (the first part of) a sequence of (approximating) stationary distributions $\{Q(x(v,w))\}_{w=1}^{\infty}$ for each original randomly drawn relative frequency vector $x(v)$. In each step of the algorithm, we compute, having determined $x(v,w-1)$, $p(x(v,w-1))$ and $A(x(v,w-1))$ already, a new approximating stationary distribution $Q(x(v,w))$ solving Eq. (1), i.e., such that it satisfies $A(x(v,w-1))\cdot Q(x(v,w)) = \begin{bmatrix} 0^{N-1} \\ 1 \end{bmatrix}$. Then we check whether $||Q(x(v,w)) - Q(x(v,w-1))||_{\infty} < \varepsilon$, i.e., whether $Q(x(v,w))$ is close enough to the previous approximating stationary distribution $Q(x(v,w-1))$. If not, $x(v,w)$ is used to obtain new transition probabilities $p(x(v,w))$ determining, then solving Eq. (1) anew with $A(x(v,w))$ to get a new candidate $Q(x(v,w+1))$. If the difference between two consecutive candidate stationary distributions is sufficiently small, we rescale the originally drawn $x(v)$ with the last stationary distribution found and we obtain the corresponding rewards $\gamma(x(v))$ setting them equal to $\gamma(x(v,w))$. If we cannot find a sufficiently close pair of consecutive approximating stationary distributions in the manner described until $w = W$, then the algorithm terminates the search for the current original randomly drawn relative frequency vector $x(v)$ and it is discarded. A new relative frequency vector $x(v)$ is drawn randomly and the whole procedure is repeated.

4. Speed and Efficiency

In the previous section, we presented the basics to obtain a large set of feasible rewards. Here, we present a so-called accelerator with a great potential of speeding up our computations considerably, measured in number of iterations or in total computing time. We discuss the possibility that the computations lead to cycling and our precautions taken in that respect. We finally propose to split up the calculations into parts in which starting relative frequency vectors are generated randomly but contain a certain number of zeros by design.

4.1. *Aitken's delta squared method*

The computations for the Type-III models may be sped up by making use of Aitken's delta squared method (cf., e.g., Burden and Faires [2010]). This method applied to the stationary distributions generated in Loop IIIb boils down to the following. For the sequence $\{Q^k(x(v,w))\}_{w\in\mathbb{N}}$, for fixed $k = 1,\ldots,N$, we generate a parallel sequence as follows, where we write $\Delta Q^k(w)$ for $Q^k(x(v,w)) - Q^k(x(v,w-1))$:

$$\Lambda^k(x(v,w)) = Q^k(x(v,w+2)) - \frac{(\Delta Q^k(w+2))^2}{\Delta Q^k(w+2) - \Delta Q^k(w+1)}.$$

Let $\lim_{w\to\infty} Q^k(x(v,w)) = \lim_{w\to\infty} \Lambda^k(x(v,w)) = \widetilde{\Lambda}^k(x(v))$, $k = 1,\ldots,N$.

Aitken's method is not geared to dimensions higher than one. However, it treats the computations guaranteeing the convergence of any original sequence, it

improves upon in speed of convergence, as a black box. So, we treat each component-sequence as a converging sequence upon which to apply the accelerator. What remains to be done is to obtain a limit for the sequence of stationary distributions $\{Q(x(v,w))\}_{w\in\mathbb{N}}$, i.e., we normalize the results as follows:

$$\widetilde{Q}(x(v)) = \frac{1}{\sum_{k=1}^{N} \widetilde{\Lambda}^k(x(v))}(\widetilde{\Lambda}^1(x(v)), \ldots, \widetilde{\Lambda}^N(x(v))).$$

If $\{Q^k(x(v,w))\}_{w\in\mathbb{N}}$ converges, then $\{\Lambda^k(x(v,w))\}_{w\in\mathbb{N}}$ converges much faster in general. If the accelerator does not help enough, one can apply it to the accelerator successively (see also Fig. 2).

4.2. *Safeguards against indefinite computing times*

We interpret the computations as an approximation of a fixed point of the function $\Gamma : \Delta^{N-1} \to \Delta^{N-1}$, where the input is a candidate stationary distribution and the output is a new candidate, i.e., $Q_{n+1} = \Gamma(Q_n)$. As Γ is continuous,[h] Brouwer's fixed point theorem applies, i.e., $\exists Q^* \in \Delta^{N-1} : Q^* = \Gamma(Q^*)$.

Cycling or very slow convergence might be a problem in our computations.[i] We set a fixed number W and discard the associated relative frequency vector

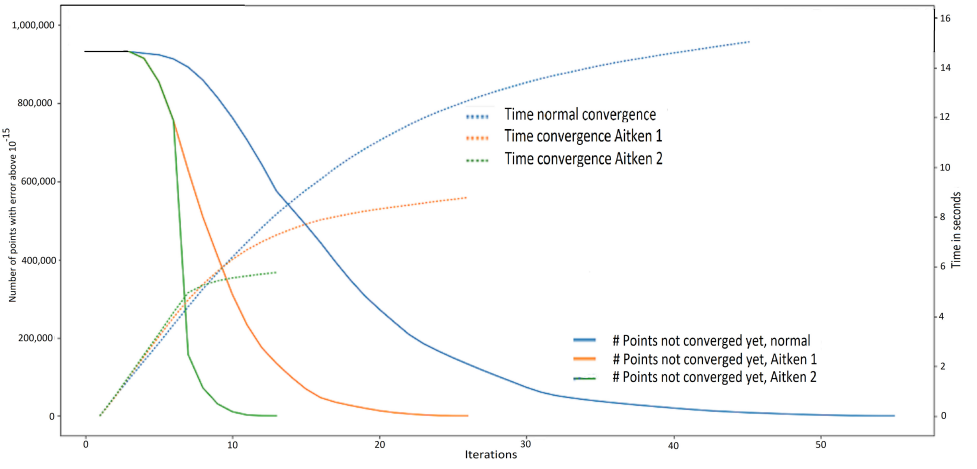

Fig. 2. The three downward (upward) sloping curves denote the number of vectors requiring at least a certain number of iterations before converging (computing times). The "normal" procedure requires less than 60 iterations (16 s). Accelerators need only 28 (10 s), respectively, 13 (6 s) iterations.

[h]Continuity of this function follows from Remark 1.

[i]For standard autonomous, irreducible, a-cyclic Markov chains, a stationary distribution can be determined by methods having contraction-like properties converging exponentially fast (e.g., Häggström [2002] and Levin *et al.* [2009]). However, if the absolute value of the second-largest eigenvalue is close to one, "exponentially fast" can be quite slow (e.g., Häggström [2002], Rosenthal [1995] and Doob [1953].

if the number of iterations in Loop IIIb exceeds W, and move on with another relative frequency vector in Loop IIIa. The down side of this measure is that "many" candidates may be thrown away. We propose to run a test with rather low V and W, keeping track of the proportion of computations exceeding W iterations in Loop IIIb, followed by running tests with increasing W until a reasonable proportion of computations is terminated. If that occurs, the number V can be increased considerably.

Alternatively, a fixed point algorithm could be added such as the generalized Newton method and quasi-Newton variants thereof (see Harker and Pang [1990] for an overview). However, Newton-like methods must start sufficiently close to the solution Q^* in order to guarantee convergence. For algorithms which do not suffer from this restriction, we refer to e.g., Laan and Talman [1979, 1981], Kojima and Yamamoto [1984] and Todd [1976].

4.3. *Split calculations interior and boundary starting vectors*

U-shaped beta distributions in our early computational efforts for Type-I games (cf., Joosten [2016]) yielded sufficient insights in the geometric properties of the reward sets already for relatively low V, i.e., the number of rewards generated. What came back to haunt us, is the tendency of randomized approaches to yield clustered feasible rewards (cf., Fig. 3).

Here, the density of rewards near the outskirts, especially on the upper right hand, is lower than in the lower left-hand interior part of the cluster of points. Simply increasing the number of rewards to be computed will fill up the low density parts eventually. However, for every "interesting" pair of rewards to be gained,

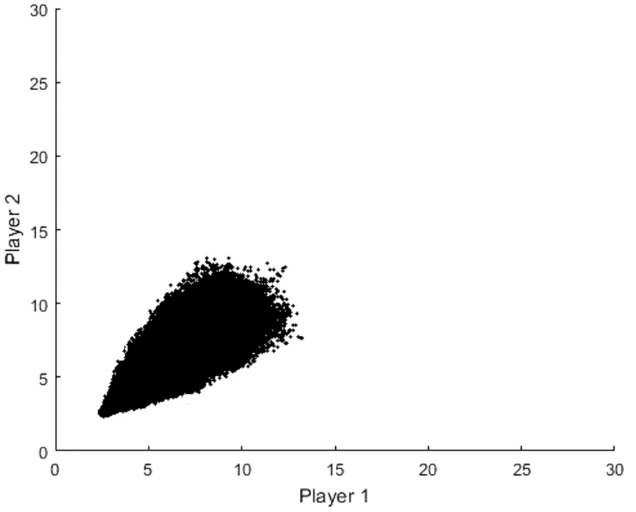

Fig. 3. The 50,000 rewards from the Type-III game presented in Sec. 2, from randomly generated starting relative frequency vectors.

a huge number of points will be generated in the very dense parts. Hence, for every useful bit of information, a huge deadweight of rather useless information is generated.

Our heuristic was to split up V into parts reserved for restricted calculations. We start with relative frequency vectors with one nonzero component per state, which belongs to an $(N-1)$-dimensional facet of the $(\#E-1)$-dimensional unit simplex. Next, we allow at most two nonzero components yielding rewards in higher-dimensional facets of the unit simplex. Gradually the number of zeros per starting state is decreased.

Example. Suppose that in the example of Sec. 2, we let the algorithm generate random relative frequency vectors with one nonzero component per state. Then to cover every possibility we need only few iterations as there are 16 such vectors. For the case that the algorithm generates one nonzero component in the first state and two nonzero components in the second, let us discuss what may occur. Suppose the algorithm draws

$$x = [0.8 \quad 0 \quad 0 \quad 0 \quad 0.1 \quad 0 \quad 0.1 \quad 0]$$

then the following computations are made

$$y = \left[1 \quad 0 \quad 0 \quad 0 \quad \frac{1}{2} \quad 0 \quad \frac{1}{2} \quad 0\right],$$

$$p^{1,1}_{1,1}(x) = 0.8 - 0.175 + 0.175Q = 0.625 + 0.175Q,$$

$$p^{2,1}_{1,1}(x) = 0.5 - 0.1 + 0.1Q = 0.4 + 0.1Q,$$

$$p^{2,1}_{2,1}(x) = 0.4 - 0.075 + 0.075Q = 0.325 + 0.075Q,$$

$$F^{1,2}(x) = 1 - (0.625 + 0.175Q) = 0.375 - 0.175Q,$$

$$F^{2,1}(x) = \frac{1}{2}(0.4 + 0.1Q) + \frac{1}{2}(0.325 + 0.075Q) = 0.0875Q + 0.3625.$$

Then $Q = 0.60734813$ solves the balance equations

$$(1 - p^{1,1}_{1,1}(x))Q = (1-Q)\frac{1}{2}(p^{2,1}_{1,1}(x) + p^{2,1}_{2,1}(x)),$$

$$Q(0.375 - 0.175Q) = (1-Q)(0.0875Q + 0.3625).$$

The rescaled relative frequency vector x' becomes

$$x' = \left[0.60734813 \quad 0 \quad 0 \quad 0 \quad \frac{1}{2}(1 - 0.60734813) \quad 0 \quad \frac{1}{2}(1 - 0.60734813) \quad 0\right].$$

Then $c(x') = 0.901\,84$ and $\gamma(x') = (10.711, 10.092)$. Similarly, for vectors

$$x_{p,q} = [q \quad 0 \quad 0 \quad 0 \quad p(1-q) \quad 0 \quad (1-p)(1-q) \quad 0]$$

satisfying $p, q \in [0, 1]$, the method induces

$$y_p = [1 \quad 0 \quad 0 \quad 0 \quad p \quad 0 \quad (1-p) \quad 0].$$

From this, similar equations lead to a reward $\gamma(x_{p,q})$. So, $\{\gamma(x_{p,q})\}_{p,q\in[0,1]}$ is a one-dimensional set, generically. The number of iterations necessary to obtain clear insights on the geometric properties of this set, is quite low in general. Similarly, we can deal with all starting vectors restricted to three nonzero components, one corresponding to one state and two to the other. Note that there are 48 such one-dimensional sets of rewards.

Next, we turn to starting vectors with four nonzero components, e.g.,

$$x_{p,q,r} = [qr \quad 0 \quad 0 \quad (1-q)r \quad p(1-r) \quad 0 \quad (1-p)(1-r) \quad 0]$$

satisfying $p, q, r \in [0, 1]$, the method induces

$$y_{p,q} = [q \quad 0 \quad 0 \quad (1-q) \quad p \quad 0 \quad (1-p) \quad 0].$$

So, $\{\gamma(x_{p,q,r})\}_{p,q,r\in[0,1]}$ is a two-dimensional set of rewards, generically. The necessary insights for these too, are obtained by a low number of iterations, but usually by generating more rewards than for the one-dimensional ones.

Comparing Figs. 3 and 4, we consider the latter to be more informative. By "strategically" keeping a couple of components equal to zero, more information at much lower numbers of rewards generated may be obtained.

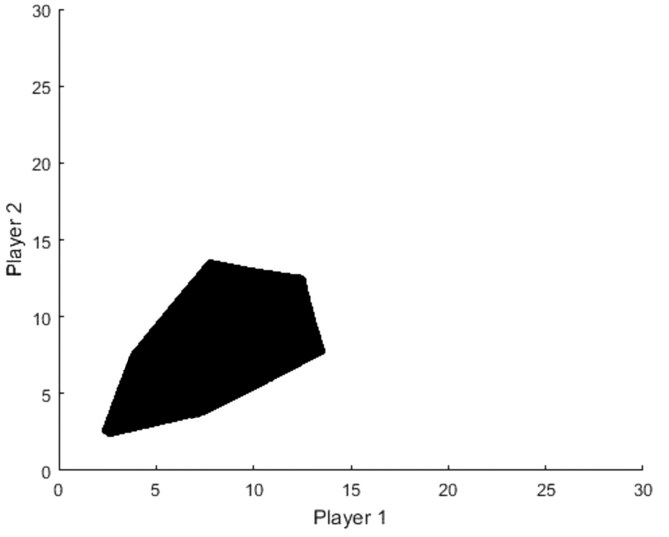

Fig. 4. The 50,000 rewards from the Type-III game presented in Sec. 2, from starting relative frequency vectors with four zeros.

5. Extending the Range of Applicability

In obtaining the first results, we leaned on the following classification of Markov chains by Kemeny and Snell [1976, p. 35]. Next, we investigate, using this classification, where the computational procedures can be applied fruitfully. We will see the implications of Remark 1, and we are interested in relaxing the restrictions implied.

I Chains without transient sets

- I-0 Several ergodic sets: these are unrelated, hence, each ergodic set can be investigated separately.
- — Unique ergodic set.
- I-A Regular: no matter where the process starts, after sufficiently much time has elapsed, the system can be in any state (*regular Markov chain*).
- I-B Cyclic: for any given starting position, the system will move through the states in a definite order, returning to the starting position after a fixed number of steps (*cyclic Markov chain*).

II Chains with transient sets.

- II-A All ergodic sets are singletons (*absorbing chain*).
- II-B All ergodic sets are regular, but not all are unit sets.
- II-C All ergodic sets are cyclic.
- II-D Cyclic and regular ergodic sets coexist.

Remark 1 reduces the applicability of the computational method to games in which all states communicate, i.e., there is a unique ergodic set. Moreover, the Markov chain induced is irreducible, so all states are in the ergodic set and therefore, visited infinitely often as time goes to infinity, and it is a periodic, i.e., the process can be in any state after a sufficient amount of time has elapsed (Class I-A). This means technically that the inverse matrix of $A(x)$ in (1) exists, and there exist several ways to compute the stationary distribution, one is of course to use $A^{-1}(x)$ directly.

Checking whether the ergodic set is cyclic/periodic using the formula in Remark 1 may be a bit cumbersome, but a periodic ergodic set has all diagonal transition probabilities equal to zero (cf., e.g., Kemeny and Snell [1976]). Only then, such a check is necessary.

Kemeny and Snell [1976] used a visualization pertaining to the transition matrix $F(x)$ to aid in the classification. If this matrix is irreducible we are in a Class I-A or Class I-B setting. The computational methods do not work in the Class I-B framework.

If $F(x)$ is not irreducible, we can rewrite it as follows (this example's dimensionality is without loss of generality):

$$F(x) = \begin{bmatrix} A_1 & 0_{1,2} & 0_{1,3} & 0_{1,4} & 0_{1,5} \\ B_{2,1} & A_2 & 0_{2,3} & 0_{2,4} & 0_{2,5} \\ B_{3,1} & B_{3,2} & A_3 & 0_{3,4} & 0_{3,5} \\ B_{4,1} & B_{4,2} & B_{4,3} & A_4 & 0_{4,5} \\ B_{5,1} & B_{5,2} & B_{5,3} & B_{5,4} & A_5 \end{bmatrix},$$

where $A_1, \ldots, A_5$ are the square irreducible matrices corresponding to the equivalence classes of states.[j] We denote the cardinality of the equivalence class $i = 1, \ldots, 5$ by $|A_i|$, i.e., matrix A_i has $|A_i|$ rows and columns. Each above-diagonal matrix $0_{k,l}$ is a zero $|A_k| \times |A_l|$-matrix containing transition probabilities from states in equivalence class k to states in equivalence class $l > k$, and each matrix $B_{k,l}$ is a $|A_l| \times |A_k|$-matrix of transition probabilities from states in equivalence class k to states in equivalence class l, $k > l$. The latter matrices are used in the classification.

If all $B_{k,l}$, $k > l$, are zero matrices, the system has unrelated ergodic sets and we are in Class I-0. The following matrix pertains to that situation:

$$F(x) = \begin{bmatrix} A_1 & 0_{1,2} & 0_{1,3} & 0_{1,4} & 0_{1,5} \\ 0_{2,1} & A_2 & 0_{2,3} & 0_{2,4} & 0_{2,5} \\ 0_{3,1} & 0_{3,2} & A_3 & 0_{3,4} & 0_{3,5} \\ 0_{4,1} & 0_{4,2} & 0_{4,3} & A_4 & 0_{4,5} \\ 0_{5,1} & 0_{5,2} & 0_{5,3} & 0_{5,4} & A_5 \end{bmatrix}.$$

Then the analysis can be restricted to the independent subsystems (cf., e.g., Kemeny and Snell [1976]). The isolated ergodic set of a subsystem can be either Class I-A or Class I-B. The algorithms proposed can be applied here to each of these unrelated Class I-A ergodic sets. To predict the long run behavior of the system, it must be known where the system starts.

If any of the matrices $B_{k,l}$, $k > l$, is nonzero, we are in Class II. If additionally $|A_i| = 1$ for all i, then we are in Class II-A, and the algorithms will not work except in very restricted cases.

If all matrices $B_{k,l}$, $k > l$, are nonzero, then the first equivalence class is the unique ergodic set. We are in a subcase of Class II-B or Class II-C, namely a chain with transient equivalent classes of states but a unique ergodic set. The algorithms work if the first equivalence class is regular, they will not work if the first equivalence class is cyclic.

[j]A class is called an equivalence class if for every state there exists a path to every other state in the class. An equivalence class is called *transient* if the probability of returning to it infinitely often as time goes to infinity is zero, if the process started in it. An equivalence class is called *ergodic* if the probability of returning to it infinitely often as time goes to infinity is one, if the process started in it (cf., e.g., Kemeny and Snell [1976]).

If all matrices $B_{k,l}$ in a row l are zero matrices, equivalence class l is unrelated to lower-indexed equivalence classes. We have unrelated components such as in the following matrix:

$$F(x) = \begin{bmatrix} A_1 & 0_{1,2} & 0_{1,3} & 0_{1,4} & 0_{1,5} \\ B_{2,1} & A_2 & 0_{2,3} & 0_{2,4} & 0_{2,5} \\ B_{3,1} & B_{3,2} & A_3 & 0_{3,4} & 0_{3,5} \\ 0_{4,1} & 0_{4,2} & 0_{4,3} & A_4 & 0_{4,5} \\ 0_{5,1} & 0_{5,2} & 0_{5,3} & B_{5,4} & A_5 \end{bmatrix}.$$

Here, the first three equivalence classes are unrelated to the latter two. If $B_{2,1}, B_{3,1}, B_{3,2}$ and $B_{5,4}$ are nonzero each, then the first and fourth equivalence classes are ergodic sets each. The second, third and fifth equivalence classes are transient and eventually the play ends up in the first equivalence class if the initial play occurs in one of the first three equivalence classes, and play ends up in the fourth equivalence class whenever the play starts in the last two equivalence classes.

Summarizing, the first and the fourth equivalence classes are ergodic sets, which might be regular both (Class II-B), or cyclic both (Class II-C), or one regular the other cyclic (Class II-D). The unrelated ergodic sets can be analyzed separately, and if they are Class II-B with a unique ergodic set the algorithms are applicable to them. The algorithms will fail for Class II-C, and for Class II-D, they will work for the separate unrelated sets if the ergodic set is aperiodic.

Remark 5. If we translate the statements on Markov chains made above to our framework, it is useful to think in terms of x and $F(x)$. For a given x, we can determine $F(x)$ which leads to a Markov chain that should be compatible with the same long term relative frequency vector x. Then if the pair is mutually compatible, the implied Markov chain falls into this categorization.

In Joosten and Samuel [2017], the entire range of such relative frequency vectors was examined. However, precautions were taken to guarantee that the qualitative features of the Markov chains over all possible x remained the same, i.e., we always had a I-A setting, i.e., for each relative frequency vector, the ergodic set encompasses all states.

The above statements regarding other Markov chains in the classification remain valid, if the qualitative features of the Markov chain remain identical over all possible relative frequency vectors. To put this notion of "identical qualitative features" into mathematical terms, x and x' are said to induce Markov chains with identical qualitative features if the following property holds $\mathrm{Car}(F(x)) = \mathrm{Car}(F(x'))$ where for matrix A : $\mathrm{Car}(A) \equiv \{(i,j) \mid [A]_{i,j} > 0\}$. We call this property *strong carrier identity* as it pertains to the carrier of the transition matrix and this implies *weak carrier identity* which is given by $\mathrm{Car}(Q(x)) = \mathrm{Car}(Q(x'))$.

Remark 6. In Joosten and Meijboom [2018], the qualitative features of the Markov chains changed over the range of all relative frequency vectors. There, a subset of

Table 2. Overview of the applicability of the algorithms presented.

Class	Applicability	Comments
I-0	**Yes**	For isolated I-A components
I-A	**Yes**	General
I-B	No	General
II-A	No **Yes**	General Unique ergodic set
II-B	No **Yes**	General Unique ergodic set
II-C	No	General
II-D	**Yes**	For isolated I-A components

the originally accessible states becomes temporarily inaccessible as the play evolves due to transition probabilities to this subset becoming zero as the play continues. In the terminology of this section, the Markov chain may change from Class I-A to a Class II-A or Class II-B (with a unique ergodic set) for different relative frequency vectors. More formally, what must hold in the framework of the model and the applicability of our computational methods is

$$||x - x'|| \to 0 \Longrightarrow \text{car}(F(x)) \subseteq \text{car}(F(x'))$$

(which seems guaranteed by continuity of F) and furthermore $||x - x'|| \to 0 \wedge \text{car}(F(x)) \neq \text{car}(F(x'))$ implies

$$\exists_{\delta>0} : [\text{car}(F(x)) \neq \text{car}(F(x'')) \neq \text{car}(F(x'))] \Longrightarrow ||x - x''|| \geq \delta.$$

Here, $||\cdot||$ is some norm. The algorithm of Joosten and Samuel [2017] allowed the re-analysis of Joosten and Meijboom [2018] and it worked much faster in this framework where the above is satisfied.

We have examined other models, for instance a Class II-A model changing into a Class I-A model changing into a Class II-A model where the ergodic classes in the two Class II-A processes are disjoint. What seems to be the all-decisive property making the algorithms work, is that it involved models such that for each x, *there is a unique ergodic set*. However, in the extensions examined, it always holds that if $\gamma, \gamma' \in P^{\mathcal{JC}}$, then $(\pi, \sigma) \in JC$ exists such that for any $\varepsilon > 0$, integers T_1, T_2 exist satisfying

$$\left| \frac{\sum_{t=1}^{T_1} (R_t^A(\pi,\sigma), R_t^B(\pi,\sigma))}{T_1} - \gamma \right| < \varepsilon, \quad \left| \frac{\sum_{t=1}^{T_2} (R_t^A(\pi,\sigma), R_t^B(\pi,\sigma))}{T_2} - \gamma' \right| < \varepsilon.$$

Note that this is not possible in stochastic games in general.

6. Conclusions and Discussion

We presented algorithms to compute large sets of limiting average rewards for various types of stochastic games in which the stage payoffs or the stage transition

probabilities are endogenously determined as the play evolves. Not all past action choices matter, just the relative frequencies in which all action combinations were chosen before.

We presented a classification of stochastic games based on the degree to which the stationary distribution of the game at hand is determined by present and historic action choices (see Table 1). In a Type-I model, the stationary distribution does not depend on the relative frequency vector, only on the vector of transition probabilities. Hence, it is easily determined and exact. For Type-II models, the stationary distribution depends on the relative frequency vector in combination with the transition probabilities, but the latter are independent of the former. To determine this stationary distribution, it requires some computational efforts, but the solution found is exact. For Type-III models, the stationary distribution depends on the relative frequency vector in combination with the transition probabilities which in turn are determined by the relative frequency vector. Here, determining a stationary distribution is an iterative process, terminated if consecutive stationary distributions are sufficiently close, or if the number of iterations becomes excessive.

The algorithms were built for stochastic games in which the entire set of states is *ergodic*, i.e., the process visits each state infinitely often as time goes to infinity, and *aperiodic*, i.e., after a sufficiently long period of time, the process can be in any state among the ergodic set. In such games jointly-convergent pure-strategies exist, hence, also the associated rewards.

Extending the range of the algorithms beyond this framework is possible by taking several precautions. We gave an overview of classes of the Markov chains generated to which the algorithms can and cannot be applied. In some cases, the algorithms are bound to fail, in others they do not make sense (see Table 2). Each jointly-convergent strategy pair can induce one and only one ergodic set. This does not preclude however, that two jointly-convergent strategy combinations induce two different ergodic sets. Thus, the most obvious extension is to analyze games in which there exists one and only one ergodic set for all jointly-convergent strategy combinations, not necessarily encompassing all states.

The latter has some connection of course to the topic of irreducible stochastic games and unichain stochastic games. Such games are however, always defined in terms of the set of all stationary strategies inducing an irreducible Markov chain or a single ergodic set (see e.g., Rogers [1969], Sobel [1971], Federgruen [1978] and Thuijsman [1992]). It can be confirmed that in such games, jointly-convergent pure-strategy pairs not only exist, but also yield larger sets of limiting average rewards than that of the stationary strategies.

Here, we have shown that the algorithms can be applied to games other than unichain stochastic games. If applied to multichain games, some strict extra demands on the system are required to guarantee a valid framework for the computations (see Remark 6). In Joosten and Meijboom [2018], the relative frequency vector associated with a pair of jointly-convergent pure strategies will either

induce a two- or one-state ergodic set (Markov chain). We have already, fruitfully, used the algorithms to the model of Joosten and Meijboom [2018] and even generalizations.

Game theorists have focussed on the particular stochastic games highlighting problems with existence of equilibrium such as the Big Match and variations [Blackwell and Ferguson, 1968; Filar, 1979] and the Paris Match [Thuijsman, 2003]. Only a few classes of games have been designed with the opposite objective, i.e., guaranteeing several convenient properties regarding existence and computability, such as single-controller games [Parthasarathy and Raghavan, 1981], switching-control games [Filar, 1979, 1981], ARAT games [Raghavan *et al.*, 1985] and SER-SIT games [Parthasarathy *et al.*, 1984].

Existence of equilibria is hardly ever problematic in FD games or generalizations thereof, quite the opposite. Folk theorem-like results guarantee usually rather large sets of equilibrium rewards. We do not suffer from difficulties such as that it matters, crucially, in which state the play starts, or that tricky transitions to absorbing sets or other ergodic sets exist. Hence, the algorithms can enjoy a low degree of complexity.

Acknowledgments

We thank two anonymous referees for useful comments and criticism. We are grateful to Joost Muis for double checking our programming and allowing us to use Fig. 2, a result of his own research.

References

Aumann, R. [1981] Survey of repeated games, in *Essays in Game Theory and Mathematical Economics in Honor of Oskar Morgenstern*, Vol. 4 of Gesellschaft, Recht, Wirtschaft (Wissenschaftsverlag, Bibliographishes Institut, Mannheim), pp. 11–42.

Aumann, R. [2008] Game engineering, in *Mathematical Programming and Game Theory for Decision Making*, eds. Neogy, S. K., Bapat, R. B., Das, A. K. & Parthasarathy, T. (World Scientific), pp. 279–285.

Billingsley, P. [1986] *Probability and Measure* (John Wiley & Sons).

Blackwell, D. and Ferguson, T. [1968] The big match, *Ann. Math. Stat.* **39**, 159–163.

Brenner, T. and Witt, U. [2003] Melioration learning in games with constant and frequency-dependent payoffs, *J. Econ. Behav. Organ.* **50**, 429–448.

Burden, R. L. and Faires, J. D. [2010] *Numerical Analysis*, 9th edition (Cengage Learning).

Doob, J. I. [1953] *Stochastic Processes* (Wiley).

Dutta, P. [1995] A folk theorem for stochastic games, *J. Econ. Theory* **66**, 1–32.

Federgruen, A. [1978] On N-person stochastic games with denumerable state space, *Adv. Appl. Probabil.* **10**, 452–471.

Filar, J. [1979] Algorithms for solving some undiscounted stochastic games, Ph.D. Dissertation, University of Illinois.

Filar, J. [1981] Ordered field property for stochastic games when the player who controls transitions changes from state to state, *J. Optim. Theory Appl.* **34**, 503–515.

Filar, J. and Vrieze, O. J. [1996] *Competitive Markov Decision Processes* (Springer).

Forges, F. [1986] An approach to communication equilibria, *Econometrica* **54**, 1375–1385.

Gillette, D. [1957] Stochastic games with zero stop probabilities, in *Contributions to the Theory of Games* III eds. M. Dresher *et al.*, Annals of Mathematical Studies, Vol. 39 (Princeton University Press), pp. 179–187.

Häggström, O. [2002] *Finite Markov Chains and Algorithmic Applications* (Cambridge University Press).

Harker, P. T. and Pang, J. S. [1990] Finite-dimensional variational inequality and nonlinear complementarity problems: A survey of theory, algorithms and applications, *Math. Program.* **48**, 161–220.

Hart, S. [1985] Nonzero-sum two-person repeated games with incomplete information, *Math. Oper. Res.* **10**, 117–153.

Herings, P. J. J. and Predtetchinski, A. [2012] Voting in collective stopping games, GSBE Research Memorandum No. 014, Maastricht University.

Joosten, R. [1996] Dynamics, equilibria, and values, Ph.D. thesis no. 96-37, Faculty of Economics & Business Administration, Maastricht University.

Joosten, R. [2005] A note on repeated games with vanishing actions, *Int. Game Theory Rev.* **7**, 107–115.

Joosten, R. [2007a] Small fish wars: A new class of dynamic fishery-management games, *ICFAI J. Manag. Econ.* **5**, 17–30.

Joosten, R. [2007b] Small fish wars and an authority, in *The Rules of the Game*: *Institutions*, *Law*, *and Economics*, eds. Prinz, A., Steenge, A. E. and Schmidt, J., (LIT-Verlag), pp. 131–162.

Joosten, R. [2009] Strategic advertisement with externalities: A new dynamic approach, in *Modeling, Computation and Optimization*, eds. Neogy, S. K., Das, A. K. and Bapat, R. B. (World Scientific), pp. 21–43.

Joosten, R. [2014] Social dilemmas, time preferences and technology adoption in a commons problem, *J. Bioeconomics* **16**, 239–258.

Joosten, R. [2015] Long-run strategic advertisement and short-run Bertrand competition, *Int. Game Theory Rev.* **17**, 1540014.

Joosten, R. [2016] Strong and weak rarity value: Resource games with complex price-scarcity relationships, *Dyn. Games Appl.* **6**, 97–111.

Joosten, R. [2019] Strategic interaction and externalities: FD-games and pollution, in *Understanding Economic Change: Contributions to an Evolutionary Paradigm in Economics*, eds. Witt, U. and Chai, A. (Cambridge University Press), pp. 288–308.

Joosten, R., Brenner, T. and Witt, U. [2003] Games with frequency-dependent stage payoffs, *Int. J. Game Theory* **31**, 609–620.

Joosten, R. and Meijboom, R. [2010][k] Stochastic games with endogenous transitions, papers on economics and evolution #1024, Max Planck Institute of Economics, Jena, This version of the 2018 paper contains relevant examples lost in the refereeing process.

Joosten, R. and Meijboom, R. [2018] Stochastic games with endogenous transitions, in *Mathematical Programming and Game Theory*, eds. Neogy, S. K., Das, A. K. and Dubey, D. (Springer), pp. 205–226.

Joosten, R. and Samuel, L. [2017] On stochastic fishery games with endogenous stage payoffs and transition probabilities, *Game Theory and Applications*, eds. Li, D.-F., Yang, X. G., Uetz, M. & Xu, G. J., CCIS, Vol. 758 (Springer), pp. 115–133.

Joosten, R., Thuijsman, F. and Peters, H. [1995] Unlearning by not doing: Repeated games with vanishing actions, *Games Econ. Behav.* **9**, 1–7.

Kemeny, J. G. and Snell, J. L. [1976] *Finite Markov Chains* (Springer).

Kojima, M. and Yamamoto, Y. [1984] A unified approach to the implementation of several restart fixed point algorithms and a new variable dimension algorithm, *Math. Program.* **28**, 288–328.

Laan, G. V. D. and Talman, A. J. J. [1979] A restart algorithm for computing fixed points without an extra dimension, *Math. Program.* **17**, 74–84.

Laan, G. V. D. and Talman, A. J. J. [1981] A class of simplicial restart fixed point algorithms without an extra dimension, *Math. Program.* **20**, 33–48.

Levin, D. A., Peres, Y. and Wilmer, E. L. [2009] *Markov Chains and Mixing Times* (AMS Society).

Mahohoma, W. [2014] Stochastic games with frequency dependent stage payoffs, Master thesis DKE 14-21, Department of Knowledge Engineering, Maastricht University.

Parthasarathy, T. and Raghavan, T. E. S. [1981] An orderfield property for stochastic games when one player controls transition probabilities, *J. Optim. Theory Appl.* **33**, 375–392.

Parthasarathy, T., Tijs, S. H. and Vrieze, O. J. [1984] Stochastic games with state independent transitions and separable rewards, in *Selected Topics in Operations Research and Mathematical Economics*, eds. Hammer, G. & Pallaschke, D. (Springer), pp. 262–271.

Raghavan, T. E. S. and Filar, J. [1991] Algorithms for stochastic games — A survey, *Z. Oper. Res.* **35**, 437–472.

Raghavan, T. E. S., Tijs, S. H. and Vrieze, O. J. [1985] On stochastic games with additive reward and transition structure, *J. Optim. Theory Appl.* **47**, 451–464.

Rogers, P. D. [1969] Nonzero-sum stochastic games, Report ORC 69-8, Operations Research Center, University of California, Berkeley.

Rosenthal, J. S. [1995] Convergence rates for Markov chains, *SIAM Rev.* **37**, 387–405.

Schoenmakers, G. M. [2004] The profit of skills in repeated and stochastic games, Ph.D. thesis, Maastricht University.

Schoenmakers, G. M., Flesch, J. and Thuijsman, F. [2002] Coordination games with vanishing actions, *Int. Game Theory Rev.* **4**, 119–126.

Shapley, L. [1953] Stochastic games, *Proc. Natl. Acad. Sci. USA* **39**, 1095–1100.

Sobel, M. J. [1971] Noncooperative stochastic games, *Ann. Math. Stat.* **42**, 1930–1935.

Thuijsman, F. [1992] Optimality and equilibria in stochastic games, CWI-tract 82, Center for Mathematics and Computer Science, Amsterdam.

Thuijsman, F. [2003] The big match and the Paris match, in *Stochastic Games and Applications*, eds. Neyman, A. & Sorin, S. (Kluwer), pp. 195–204.

Thuijsman, F. and Vrieze, O. J. [1998] The power of threats in stochastic games, in *Stochastic and Differential Games, Theory and Numerical Solutions*, eds. Bardi, M., Raghavan, T. E. S. & Parthasarathy, T. (Birkhauser), pp. 343–358.

Todd, M. J. [1976] *The Computation of Fixed Points and Applications*, Lecture Notes in Economics and Mathematical Systems (Springer).

Uyttendaele, P., Thuijsman, F., Collins, P., Peeters, R., Schoenmakers, G. and Westra, R. [2012] Evolutionary games and periodic fitness, *Dyn. Games Appl.* **2**, 335–345.

Vrieze, O. J. [1981] Linear programming and undiscounted stochastic games, *OR Spectrum* **3**, 29–35.

Vrieze, O. J. [1987] Stochastic games with finite state and action spaces, CWI-tract 33, Center for Mathematics and Computer Science, Amsterdam.

Vrieze, O. J., Tijs, S. H., Raghavan, T. E. S. and Filar, J. A. [1983] A finite algorithm for the switching control stochastic game, *OR Spectrum* **5**, 15–24.