Available online at www.sciencedirect.com

SCIENCE DIRECT®

Operations Research Letters 34 (2006) 412–420

Operations Research Letters

www.elsevier.com/locate/orl

A note on two-person zero-sum communicating stochastic games

Zeynep Müge Avşar[a], Melike Baykal-Gürsoy[b,*]

[a]Department of Industrial Engineering, Middle East Technical University, Ankara, Turkey
[b]Department of Industrial and Systems Engineering, Rutgers, The State University of New Jersey, NJ, USA

Received 10 December 2004; accepted 24 July 2005
Available online 22 September 2005

Abstract

For undiscounted two-person zero-sum communicating stochastic games with finite state and action spaces, a solution procedure is proposed that exploits the communication property, i.e., working with irreducible games over restricted strategy spaces. The proposed procedure gives the value of the communicating game with an arbitrarily small error when the value is independent of the initial state.

Keywords: Stochastic games; Communicating chains; Irreducible games

1. Introduction

Two-person zero-sum communicating stochastic games considered in this note are played sequentially under the long-run average payoff criterion (undiscounted case). A stochastic game is called communicating if it is irreducible under at least one stationary strategy pair of the players (decision makers). The scope of this paper is restricted to solving communicating games under the condition that the game value is independent of the initial state. Note that, communicating stochastic games that are unichain under every stationary strategy pair, form a subclass of games which satisfy the above condition.

In [13,9,19], stochastic games with different ergodic structures are studied from an algorithmic viewpoint under the long-run average payoff criterion. Hoffman and Karp [13] propose an iterative algorithm that finds optimal stationary strategies for an irreducible stochastic game. Algorithms due to Federgruen [9] and Van der Wal [19] give ε-optimal stationary strategies under some assumptions that imply unichain case or independence of the game value from the initial state. For general stochastic games, Filar et al. [10] give a nonlinear programming formulation for finding the best stationary strategies with respect to a measure of distance from optimality.

The communication property is introduced by Bather [4] for Markov decision processes (MDPs) as the existence of a policy for every pair of states that makes one state reachable from the other one. It is

[*] Corresponding author.
E-mail address: gursoy@rci.rutgers.edu (M. Baykal-Gürsoy).

0167-6377/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.orl.2005.07.008

studied further by Ross and Varadarajan [14,15] and Baykal-Gürsoy and Ross [5]. The purpose of investigating the behaviour of communicating MDPs in the above mentioned studies is to solve certain MDP problems (MDPs with constraints or nonlinear objectives) by decomposing the state space into (open or closed) communicating classes in an hierarchical manner. The importance of studying communicating games in this article also arises once the similar hierarchical decomposition procedure proposed in [2] for stochastic games is to be employed. An extension of the communication property is used by Federgruen [8] for stochastic games. Federgruen imposes a condition that for any stationary strategy of one player there is a stationary strategy for the other one that makes the process irreducible. Thus, the difference between a stochastic game that satisfies the conditions in [8] and a communicating stochastic game is that in the latter one when an arbitrary stationary strategy is assigned to one player the other player may not have a strategy that makes the process irreducible whereas in the former case there exists at least one such strategy for the other player. Hence, Federgruen's conditions are more restrictive than a direct adaptation of Bather's definition to stochastic games. Under these more restrictive conditions, Federgruen [8] shows that there exist limiting average optimal stationary strategies for the $\mathscr{N}$-person nonzero-sum games.

To solve undiscounted two-person zero-sum communicating stochastic games, we propose to employ the algorithm in [13] for irreducible games by exploiting the communication property. Concentrating on a restricted set of feasible stationary strategies such that the probability of taking any action pair is strictly positive, a communicating stochastic game behaves as an irreducible one and Hoffman and Karp's algorithm can be applied by incorporating a transformation from the restricted set of strategies to the original set. With these changes, under the condition that the value of the communicating stochastic game is independent of the initial state, it is shown that the proposed procedure gives value of the communicating game with an error of ε for any $\varepsilon > 0$ when the restricted strategy space is sufficiently large. This procedure could also be used to obtain the ε-optimal stationary policy pair for both players.

Independently of our work, Evangelista et al. [7] use a similar approach to show that the stochastic games with additive reward and additive transition structure possess ε-optimal stationary strategies. This is a restrictive condition, since it does not hold in general. It should be pointed out that restricting the strategy space in order to obtain irreducible games corresponds to perturbing the set of mixed actions slightly. Some related analyses on the perturbation of Markov chains with applications to stochastic games are in [16].

Organization of this note is as follows: In Section 2, notation is briefly introduced and some definitions are given. Communicating stochastic game examples are included in Section 3. The restricted game model is presented and convergence is proven for the solution of a sequence of restricted games that approach the original game in Sections 3.1 and 3.2, respectively. Concluding remarks are presented in the final section.

2. Notation and definitions

Let X_n be the random variable denoting the state of the two-person zero-sum stochastic game at epoch n. It takes values from a finite state space $\mathscr{S}=\{1,\ldots,S\}$. After observing the state of the game at epoch n, players I and II simultaneously take actions represented by the random variables A_n and B_n, respectively. The finite sets of actions available in state i for players I and II are $\mathscr{A}_i=\{1,\ldots,M_i\}$ and $\mathscr{B}_i=\{1,\ldots,N_i\}$, respectively. Then, $\{(X_n,A_n,B_n), n=1,2,\ldots\}$ is the underlying stochastic process of the game. As a function of the state visited and the actions taken at epoch n, player II pays player I a payoff $R_n=R(X_n,A_n,B_n)$ instantaneously. It is assumed that the payoffs are finite. Given that the process is in state i and action pair (a,b) is taken by the players at epoch n, the transition probability of being in state j at the next epoch is $P(X_{n+1}=j|X_n=i, A_n=a, B_n=b)$. We consider time-homogeneous processes, i.e., $E\,(R(X_n=i, A_n=a, B_n=b))=r_{iab}$ and $P\,(X_{n+1}=j|X_n=i, A_n=a, B_n=b) = P_{iabj}$ for all $n=1,2,\ldots$.

Let $\boldsymbol{f}^n$ and $\boldsymbol{h}^n$ be the behaviour strategies (probability distributions over the action spaces given the complete history) of player I and II, respectively, at epoch n. Since a stationary strategy is a behaviour strategy such that the dependence on history is through the current state being visited, let the vectors $\boldsymbol{\alpha}=(\alpha_{11},\alpha_{12},\ldots,\alpha_{1M_1};\ldots;\alpha_{S1},\ldots,\alpha_{SM_S})$ and $\boldsymbol{\beta}=(\beta_{11},\beta_{12},\ldots,\beta_{1N_1};\ldots;\beta_{S1},\ldots,\beta_{SN_S})$ denote the stationary strategies for players I and II, respectively, where $\alpha_{ia}=P(A_n=a|X_n=i)$ and $\beta_{ib}=P(B_n=b|X_n=i)$ for every epoch n when the current state is i.

In case players' strategies are specified, corresponding expected payoff and transition probabilities will be denoted by writing the vectors of these strategies in parentheses. If, on the other hand, an action is fixed for a player, the subscript will be retained. For example, $r_i(\boldsymbol{\alpha},\boldsymbol{\beta})=\sum_{a\in\mathscr{A}_i}\sum_{b\in\mathscr{B}_i}r_{iab}\alpha_{ia}\beta_{ib}$ is the expected payoff in state i given the strategy pair $(\boldsymbol{\alpha},\boldsymbol{\beta})$. Another example could be $P_{iaj}(\boldsymbol{\beta})=\sum_{b\in\mathscr{B}_i}P_{iabj}\ \beta_{ib}$, the transition probability of visiting state j at the next epoch, given that the current state is i and the first player takes action a and the second player's strategy is $\boldsymbol{\beta}$.

When the initial state is i, the long-run average expected payoff to player I under the strategy pair $(\boldsymbol{f},\boldsymbol{h})$ is defined as

$$\phi_i(\boldsymbol{f},\boldsymbol{h})=\liminf_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}E_{\boldsymbol{f},\boldsymbol{h}}(R_n|X_1=i).$$

The payoff to player II is $-\phi_i(\boldsymbol{f},\boldsymbol{h})$. If the long-run average expected payoff is independent of the initial state under the behaviour strategies $\boldsymbol{f}$ and $\boldsymbol{h}$, then the subscript i is dropped in $\phi_i(\boldsymbol{f},\boldsymbol{h})$. Since the objective of player I is to maximize his average expected reward and the objective of player II is to minimize his average expected loss, the strategy pair $(\hat{\boldsymbol{f}},\hat{\boldsymbol{h}})$ is said to be ε-optimal ($\varepsilon>0$) if

$$\phi_i(\boldsymbol{f},\hat{\boldsymbol{h}})-\varepsilon\leqslant\phi_i(\hat{\boldsymbol{f}},\hat{\boldsymbol{h}})\leqslant\phi_i(\hat{\boldsymbol{f}},\boldsymbol{h})+\varepsilon \tag{1}$$

is satisfied for all behaviour strategies $\boldsymbol{f}$ and $\boldsymbol{h}$ and all $i\in\mathscr{S}$. $(\hat{\boldsymbol{f}},\hat{\boldsymbol{h}})$ is called optimal and forms the limiting average equilibrium, and $\phi_i(\hat{\boldsymbol{f}},\hat{\boldsymbol{h}})$ is called value of the game for initial state i, on which both players agree if and only if $\varepsilon=0$ in (1).

Ergodic structure of stochastic games is determined by analyzing the underlying Markov chain $P(\boldsymbol{\alpha},\boldsymbol{\beta})$ of the state process $\{X_n, n=1,2,\ldots\}$ under every pure strategy pair $(\boldsymbol{\alpha},\boldsymbol{\beta})$ (a stationary strategy is called pure if it assigns only one action to each state).

Definition 1. A stochastic game is said to be unichain if the Markov chain induced by every pure strategy pair has one recurrent class and a (possibly empty) set of transient states.

If there are no transient states, then the game is called irreducible. Note that a stochastic game is irreducible when the states are reachable from each other under every stationary strategy pair. Since it is known from [12] that there exist optimal stationary strategies $\boldsymbol{\alpha}^*$, $\boldsymbol{\beta}^*$ for irreducible undiscounted stochastic games, the minimax theorem can be stated as

$$\begin{aligned}\max_{\boldsymbol{\alpha}\in C_0^1}\ \min_{\boldsymbol{\beta}\in C_0^2}\phi(\boldsymbol{\alpha},\boldsymbol{\beta})&=\phi(\boldsymbol{\alpha}^*,\boldsymbol{\beta}^*)\\&=\min_{\boldsymbol{\beta}\in C_0^2}\ \max_{\boldsymbol{\alpha}\in C_0^1}\phi(\boldsymbol{\alpha},\boldsymbol{\beta}),\end{aligned} \tag{2}$$

where $C_0^1=\{\boldsymbol{\alpha}\,|\,\sum_{a\in\mathscr{A}_i}\alpha_{ia}=1,\ i\in\mathscr{S},\text{ and }\boldsymbol{\alpha}\geqslant\boldsymbol{0}\}$, $C_0^2=\{\boldsymbol{\beta}\,|\,\sum_{b\in\mathscr{B}_i}\beta_{ib}=1,\ i\in\mathscr{S},\text{ and }\boldsymbol{\beta}\geqslant\boldsymbol{0}\}$. The space of feasible stationary strategies such that $\boldsymbol{\alpha}\in C_0^1$ and $\boldsymbol{\beta}\in C_0^2$ will be denoted by C_0. Subscript i of ϕ denoting the initial state is dropped in (2) because the game under consideration is irreducible.

Definition 2. If there exists a pure strategy pair $(\boldsymbol{\alpha},\boldsymbol{\beta})$ such that j is accessible from i under $(\boldsymbol{\alpha},\boldsymbol{\beta})$ for all ordered pairs of states (i,j), then the stochastic game is said to be communicating.

Note that an irreducible stochastic game possesses the communication property. In a communicating stochastic game, $P(\boldsymbol{\alpha},\boldsymbol{\beta})$ is irreducible for every stationary strategy pair $(\boldsymbol{\alpha},\boldsymbol{\beta})$ that satisfies $\alpha_{ia}\beta_{ib}>0$ for all $a\in\mathscr{A}_i$, $b\in\mathscr{B}_i$, $i\in\mathscr{S}$. A communicating stochastic game may have a unichain or multichain structure.

In order to determine how good a stationary strategy pair $(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}})$ is over C_0, Filar et al. [10] introduce the distance function, $\sum_{i\in\mathscr{S}}(\max_{\boldsymbol{\alpha}\in C_0^1}\phi_i(\boldsymbol{\alpha},\hat{\boldsymbol{\beta}})-\min_{\boldsymbol{\beta}\in C_0^2}\phi_i(\hat{\boldsymbol{\alpha}},\boldsymbol{\beta}))$, that quantifies the distance from optimality. In case relation (2) is not satisfied over C_0, best stationary strategies are found by minimizing the distance function. Otherwise, the minimum distance is zero at $(\boldsymbol{\alpha}^*,\boldsymbol{\beta}^*)$.

3. Communicating stochastic games

In this section, first some example games are presented in order to investigate the behaviour of game value and ergodic structure of such games under optimal or best stationary strategies. Then, the proposed procedure is introduced and its convergence is studied.

Matrix notation used for each state i is such that each entry corresponds to an action pair (a, b) of the players. The value in the upper diagonal of each entry is r_{iab} while the lower diagonal gives either the probability distribution over the future states, $(P_{iab1}, P_{iab2}, \ldots, P_{iabS})$, or just the next state to be visited.

In the first two examples, value of the game is not independent of the initial state. Thus, the proposed procedure cannot be used to solve these types of games. In Examples 1 and 3, value of the game is achieved under stationary strategies, while in Examples 2 and 4, it is not.

Example 1. Although this communicating game is multichain (it has more than one recurrent class under a stationary strategy pair), communication property specified in Definition 2 is satisfied. The optimal stationary strategy is $(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*) = ((1; 1, 0), (1, 0; 1))$, under which the game has two disjoint chains. Value of the game is 0 (1) when the initial state is 1 (2).

0 1	1 2

state 1

1 2
0 1

state 2

Example 2. A variation on the Big Match by Blackwell and Ferguson [6] (due to Koos Vrieze). For (initial) state 1, there does not exist an optimal strategy over the space of stationary strategies because $\max_{\boldsymbol{\alpha}\in C_0^1} \min_{\boldsymbol{\beta}\in C_0^2} \phi_1(\boldsymbol{\alpha}, \boldsymbol{\beta}) = 0$, whereas $\min_{\boldsymbol{\beta}\in C_0^2} \max_{\boldsymbol{\alpha}\in C_0^1} \phi_1(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \frac{1}{2}$. The best stationary strategy pairs in state 1 are such that player II chooses each action with a probability of $\frac{1}{2}$ and player I may choose any stationary strategy. Unlike state 1, value of the game in states 2 and 3 (which are 0 and 1, respectively) is achieved under stationary strategies.

1 1	0 1
0 2	1 3

state 1

0 2	1 1

state 2

1 3
0 1

state 3

Example 3. Under the condition that the game value is independent of the initial state, the following question may arise: does there exist at least one optimal stationary strategy pair under which the game is unichain? As this example shows, the answer is not affirmative. The ergodic structure of a communicating stochastic game under every optimal stationary strategy pair may be multichain even though the value of this game is independent of the initial state. The only optimal stationary strategy pair for this game is $(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*) = ((1, 0; 0, 1), (0, 1; 1, 0))$ and value of the game is 1 which is independent of the initial state. $P(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*)$ is multichain such that each of the states defines an ergodic class by itself.

2 1	1 1
0 2	0 1

state 1

0 2	1 1
1 2	2 2

state 2

Example 4 (*due to Koos Vrieze*). Since $\min_{\boldsymbol{\beta}\in C_0^2} \max_{\boldsymbol{\alpha}\in C_0^1} \phi_1(\boldsymbol{\alpha}, \boldsymbol{\beta}) = 1 = \sup_{\boldsymbol{\alpha}\in C_0^1} \inf_{\boldsymbol{\beta}\in C_0^2} \phi_1(\boldsymbol{\alpha}, \boldsymbol{\beta})$, player I does not have an optimal stationary strategy but every stationary strategy is optimal for player II. Note that value of the game does not depend on the initial state, so there exist ε-optimal stationary strategies for player I given any $\varepsilon > 0$.

0 1	1 1
1 2	0 1

state 1

1 2
0 1

state 2

Proposed procedure deals with communicating games that have state-independent value as in the last two examples.

3.1. The algorithm

Proposed procedure requires employment of the iterative algorithm developed by Hoffman and Karp [13] over a restricted strategy space. Therefore, in this subsection, first, Hoffman and Karp's algorithm is reviewed, and then the proposed procedure is presented.

Hoffman and Karp's algorithm converges for the class of irreducible undiscounted stochastic games. Their algorithmic construction is based on the use of MDP methodology. In two-person games, when the strategy of one player is fixed, the problem reduces to an MDP problem for the other player. From the MDP literature, it is known that given $\boldsymbol{\beta}$ the following Linear Programming (LP) formulation can be used to solve $\max_{\boldsymbol{\alpha}} \phi(\boldsymbol{\alpha}, \boldsymbol{\beta})$ when the MDP under consideration is irreducible.

*Problem P*1

$$\begin{array}{ll} \text{Min} & g \\ \text{s.t.} & g + v_i \geqslant r_{ia}(\boldsymbol{\beta}) + \sum_{j\in\mathscr{S}} P_{iaj}(\boldsymbol{\beta})\, v_j, \\ & \quad i \in \mathscr{S},\ \ a \in \mathscr{A}_i, \\ & g \text{ unrestricted}, \\ & v_i \text{ unrestricted}, \quad i \in \mathscr{S}, \end{array}$$

where decision variables g and v_i can be interpreted as the long-run average expected payoff and the change in the total payoff when the initial state is i, respectively, under the limiting conditions. So, by definition the value of v_i is determined up to a constant. Using the complementary slackness property, the optimality condition is given as

$$g^* + v_i^* = \max_{a\in\mathscr{A}_i} \left(r_{ia}(\boldsymbol{\beta}) + \sum_{j\in\mathscr{S}} P_{iaj}(\boldsymbol{\beta})\, v_j^* \right), \quad i \in \mathscr{S},$$

in the average payoff MDP.

In the light of this insight into the problem, Hoffman and Karp [13] define the matrix game $\Gamma_i(\boldsymbol{v})$ with entries $(r_{iab} + \sum_{j\in\mathscr{S}} P_{iabj}\, v_j)$ for each action pair (a, b) when the initial state is i and give the optimality condition as follows: $g^* + v_i^* = val\,\Gamma_i(\boldsymbol{v}^*)$, $i \in \mathscr{S}$, where the right-hand side denotes the value of $\Gamma_i(\boldsymbol{v}^*)$. Given v_j values, for each initial state i, the matrix game $\Gamma_i(\boldsymbol{v})$ can be solved using the following LP:

*Problem P*2(i)

$$\begin{array}{ll} \text{Min} & w_i \\ \text{s.t.} & \sum_{b\in\mathscr{B}_i} \beta_{ib} \left(r_{iab} + \sum_{j\in\mathscr{S}} P_{iabj}\, v_j \right) \leqslant w_i, \\ & \quad a \in \mathscr{A}_i, \\ & \boldsymbol{\beta} \in C_0^2, \\ & w_i \text{ unrestricted}. \end{array}$$

Optimal solutions to *Problem P*2(i) give the optimal strategies of the second player and $w_i = val\,\Gamma_i(\boldsymbol{v})$, $i \in \mathscr{S}$, for given $\boldsymbol{v}$ vector. The dual of *Problem P*2(i) is used to find the first player's optimal stationary strategy for given $\boldsymbol{v}$.

By recalling the minimax theorem, one can understand better the use of MDP theory in analyzing the behaviour of stochastic games. Hoffman and Karp's iterative algorithm can be implemented in two ways. One way is to minimize $\max_{\boldsymbol{\alpha}} \phi(\boldsymbol{\alpha}, \boldsymbol{\beta})$ over all β in C_0^2 and the other is to maximize $\min_{\boldsymbol{\beta}} \phi(\boldsymbol{\alpha}, \boldsymbol{\beta})$ over all $\boldsymbol{\alpha}$ in C_0^1 until achieving convergence at $\boldsymbol{\beta}^*$ and $\boldsymbol{\alpha}^*$, respectively. Hoffman and Karp [13] prove the convergence of their algorithm to the value of the irreducible stochastic game. This proof is based on the observation that the sequence of g values obtained through iterations of the algorithm is monotone nonincreasing, and g and $\boldsymbol{v} = (v_1, v_2, \ldots, v_S)$ vary in a compact set. The second observation follows because at every iteration g and $\boldsymbol{v}$ exist, and are unique for $v_S = 0$ and are continuous functions of $\boldsymbol{\beta}$ which vary in a compact set.

By definition, a communicating stochastic game is irreducible if a positive probability is assigned to every action pair. For every sufficiently small $\eta > 0$, define the sets

$$C_\eta^1 = \left\{ \boldsymbol{\alpha} \,\middle|\, \sum_{a\in\mathscr{A}_i} \alpha_{ia} = 1,\ i \in \mathscr{S},\ \text{and } \boldsymbol{\alpha} \geqslant \eta \boldsymbol{e} \right\} \quad \text{and}$$

$$C_\eta^2 = \left\{ \boldsymbol{\beta} \,\middle|\, \sum_{b\in\mathscr{B}_i} \beta_{ib} = 1,\ i \in \mathscr{S},\ \text{and } \boldsymbol{\beta} \geqslant \eta \boldsymbol{e} \right\},$$

where $\boldsymbol{e}$ is the vector with appropriate size such that all of its entries are 1. Let $C_\eta = \{(\boldsymbol{\alpha}, \boldsymbol{\beta}) | \boldsymbol{\alpha} \in C_\eta^1, \boldsymbol{\beta} \in C_\eta^2\}$. In order to have a one-to-one correspondence between the extreme points of C_0 and C_η, every action should be taken with a value strictly greater than all the other actions' smallest value η, i.e., $\eta < 1 - \eta(M_i - 1)$ and $\eta < 1 - \eta(N_i - 1)$, $i \in \mathscr{S}$, have to hold. These two strict inequalities imply that $\eta < \min\{1/M_i, 1/N_i\}$, $i \in \mathscr{S}$.

Now, define $\tilde{\boldsymbol{\alpha}}_i^a = (\tilde{\alpha}_{i1}^a, \tilde{\alpha}_{i2}^a, \ldots, \tilde{\alpha}_{iM_i}^a)$ for all $i \in \mathscr{S}$ and $a \in \mathscr{A}_i$, and define $\tilde{\boldsymbol{\beta}}_i^b = (\tilde{\beta}_{i1}^b, \tilde{\beta}_{i2}^b, \ldots, \tilde{\beta}_{iN_i}^b)$for all $i \in \mathscr{S}$ and $b \in \mathscr{B}_i$ as

$$\tilde{\alpha}_{ic}^a = \begin{cases} 1-\eta(M_i-1) & \text{if } c=a, \\ \eta & \text{otherwise,} \end{cases} \quad \text{and}$$

$$\tilde{\beta}_{id}^b = \begin{cases} 1-\eta(N_i-1) & \text{if } d=b, \\ \eta & \text{otherwise,} \end{cases}$$

respectively. Define *Problem P*3 by replacing $r_{ia}(\boldsymbol{\beta})$ and $P_{iaj}(\boldsymbol{\beta})$ in *Problem P*1 with $r_i(\tilde{\boldsymbol{\alpha}}_i^a, \boldsymbol{\beta})$ and $P_{ij}(\tilde{\boldsymbol{\alpha}}_i^a, \boldsymbol{\beta})$, respectively. Similarly, let *Problem P*4(i) be such that r_{iab} and P_{iabj} in *Problem P*2(i) are replaced with $r_{ib}(\tilde{\boldsymbol{\alpha}}_i^a)$ and $P_{ibj}(\tilde{\boldsymbol{\alpha}}_i^a)$, respectively, and $\boldsymbol{\beta} \in C_\eta^2$.

Since a communicating stochastic game is irreducible under every stationary strategy pair in C_η, there exists an optimal stationary strategy pair $(\boldsymbol{\alpha}^\eta, \boldsymbol{\beta}^\eta)$ such that the minimax condition in (2) holds over C_η. Denote the value of the stochastic game over C_η, i.e., $\phi(\boldsymbol{\alpha}^\eta, \boldsymbol{\beta}^\eta)$, by g^η. Next, we present the proposed procedure that requires the implementation of Hoffman and Karp's algorithm over C_η with $0 < \eta < \min_{i \in \mathscr{S}} \{\min\{1/M_i, 1/N_i\}\}$.

Step 0: Choose a stationary strategy $\boldsymbol{\beta}(1) \in C_\eta^2$ for player II. Let $n = 1$.

Step 1: Increment n by one. Given $\boldsymbol{\beta}(n-1)$, solve *Problem P*3 to find $g(n)$, $\boldsymbol{v}(n)$ letting $v_S = 0$.

Step 2: Given $\boldsymbol{v}(n)$, solve *Problem P*4(i) for all $i \in \mathscr{S}$ to find $w_i(n)$, $i \in \mathscr{S}$, and $\boldsymbol{\beta}(n)$.

If $\boldsymbol{\beta}(n) = \boldsymbol{\beta}(n-1)$, then $\boldsymbol{\beta}(n)$ is an optimal strategy for player II and $g(n)$ is the value of the stochastic game on C_η, stop.

Otherwise, go to step 1.

To illustrate the implementation of this procedure (both the one given above to find $\boldsymbol{\beta}^\eta$ and its dual version to find $\boldsymbol{\alpha}^\eta$), Examples 3 and 4 are studied. In Example 3, let η be 0.001. The procedure stops in the second iteration at $(\boldsymbol{\alpha}, \boldsymbol{\beta}) = ((1-\eta, \eta; \eta, 1-\eta), (\eta, 1-\eta; 1-\eta, \eta))$ with corresponding game value of 0.9999995, when the initial $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ chosen at step 0 is $((\eta, 1-\eta; 1-\eta, \eta), (1-\eta, \eta; \eta, 1-\eta))$. On the other hand, for Example 4 convergence of the procedure is slow at least for the initial $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ tried, i.e., $((1-\eta, \eta; 1-\eta, \eta), (1-\eta, \eta; 1))$, with $\eta = 0.001$. In this case, the procedure is stuck around $((0.968, 0.032; 0.999, 0.001), (0.03, 0.97; 1))$ with a game value around 0.97. As η gets smaller, it is observed that convergence is to $((1-\eta, \eta; 1-\eta, \eta), (\eta, 1-\eta; 1))$ with the corresponding game value approaching 1.

3.2. Convergence

Since a communicating stochastic game is irreducible over C_η, convergence of the Hoffman and Karp's algorithm (which is also outlined in the previous subsection) for given η follows from [13]. To see the convergence, one can also look at the following equivalent representation of the restricted game.

Let G be a communicating game with the law of motion P_{iabj} in which the value is independent of the initial state. Let G^η be its restricted game with the law of motion P_{iabj}^η. To obtain P^η explicitly, note that, extreme points of C_η are all possible combinations of the vectors $\tilde{\boldsymbol{\alpha}}_i^1, \ldots, \tilde{\boldsymbol{\alpha}}_i^{M_i}$ and $\tilde{\boldsymbol{\beta}}_i^1, \ldots, \tilde{\boldsymbol{\beta}}_i^{N_i}$. A pure strategy pair $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ in C_0, i.e., an extreme point of C_0 such that (a_i, b_i) is the action pair taken in state i, corresponds to the extreme point of C_η such that $(\tilde{\boldsymbol{\alpha}}_i^{a_i}, \tilde{\boldsymbol{\beta}}_i^{b_i})$ is taken in state i. The law of motion for the restricted game defined over C_η is

$$\begin{aligned} P_{iabj}^\eta &\equiv P_{ij}(\tilde{\boldsymbol{\alpha}}_i^a, \tilde{\boldsymbol{\beta}}_i^b) \\ &= (1-\eta(M_i-1))\,(1-\eta(N_i-1))\,P_{iabj} \\ &\quad + (1-\eta(M_i-1))\,\eta \sum_{d \neq b} P_{iadj} \\ &\quad + \eta\,(1-\eta(N_i-1)) \sum_{c \neq a} P_{icbj} \\ &\quad + \eta^2 \sum_{c \neq a} \sum_{d \neq b} P_{icdj} \end{aligned}$$

in terms of P_{iabj}. Hoffman and Karp's algorithm applied to this perturbed game will converge to the optimal stationary strategy pair $(\boldsymbol{\alpha}^\eta, \boldsymbol{\beta}^\eta)$ and the optimal value of the game g^η for given η.

To prove the convergence of the proposed procedure to the value of the communicating game as η gets smaller, note that when the value of a stochastic game is independent of the initial state there exist ε-optimal stationary strategies for any $\varepsilon > 0$ [11]. Let $(\boldsymbol{\alpha}_\varepsilon, \boldsymbol{\beta}_\varepsilon)$ denote an ε-optimal stationary strategy pair for a given communicating stochastic game whose value is state-independent, and $\left\{(\boldsymbol{\alpha}^{(n)}, \boldsymbol{\beta}^{(n)})\right\}_{n=1}^{\infty}$ be a sequence such that $(\boldsymbol{\alpha}^{(n)}, \boldsymbol{\beta}^{(n)}) \in C_{\eta_n}$, $\eta_n > 0$ for all n, $\eta_n \to 0$ and $(\boldsymbol{\alpha}^{(n)}, \boldsymbol{\beta}^{(n)}) \to (\boldsymbol{\alpha}_\varepsilon, \boldsymbol{\beta}_\varepsilon)$ as $n \to \infty$. The convergence proof of Proposition 1 results from the observation that $\left(\boldsymbol{\alpha}^{(n)}, \boldsymbol{\beta}^{(n)}\right)$ is ε-optimal for any $\varepsilon > 0$ for both the original game over C_0 and the restricted game over C_{η_n} when n is sufficiently large. The following lemma shows that $\left(\boldsymbol{\alpha}^{(n)}, \boldsymbol{\beta}^{(n)}\right)$ is ε-optimal for the original game over C_0.

Lemma 1. *Consider a communicating stochastic game whose value is independent of the initial state. Let $(\boldsymbol{\alpha}_\varepsilon, \boldsymbol{\beta}_\varepsilon)$ be an ε-optimal stationary strategy pair. For any $\varepsilon > 0$, consider $\left(\boldsymbol{\alpha}^{(n)}, \boldsymbol{\beta}^{(n)}\right) \in C_{\eta_n}$ such that $\eta_n > 0$ for all n, $\eta_n \to 0$ and $\left(\boldsymbol{\alpha}^{(n)}, \boldsymbol{\beta}^{(n)}\right) \to (\boldsymbol{\alpha}_\varepsilon, \boldsymbol{\beta}_\varepsilon)$ as $n \to \infty$. Then, $\left(\boldsymbol{\alpha}^{(n)}, \boldsymbol{\beta}^{(n)}\right)$ is ε-optimal for the original game if n is large enough.*

The proof of Lemma 1, which can be found in the longer version of this paper [3], is given by investigating the behaviour of $\min_{\boldsymbol{\beta} \in C_0^2} \phi(\boldsymbol{\alpha}, \boldsymbol{\beta})$ $(\max_{\boldsymbol{\alpha} \in C_0^1} \phi(\boldsymbol{\alpha}, \boldsymbol{\beta}))$ around $\boldsymbol{\alpha}_\varepsilon$ $(\boldsymbol{\beta}_\varepsilon)$ within $C_{\eta_n}^1$ $(C_{\eta_n}^2)$ letting n go to infinity because ϕ is not necessarily continuous over C_0. Note that when $\boldsymbol{\alpha}$ in $C_{\eta_n}^1$ ($\boldsymbol{\beta}$ in $C_{\eta_n}^2$) is fixed for player I (II) in a communicating game, the resulting MDP problem for player II (I) is a communicating MDP, which means that ϕ turns out to be state-independent (subscript i is dropped) with respect to the minimizing (maximizing) strategy of player II (I).

Remark. Let $(\boldsymbol{\alpha}_\varepsilon, \boldsymbol{\beta}_\varepsilon)$ denote an ε-optimal stationary strategy pair for a given communicating stochastic game whose value is state-independent, and $\left\{(\boldsymbol{\alpha}^{(n)}, \boldsymbol{\beta}^{(n)})\right\}_{n=1}^{\infty}$ be a sequence such that $\left(\boldsymbol{\alpha}^{(n)}, \boldsymbol{\beta}^{(n)}\right) \in C_{\eta_n}$, $\eta_n > 0$ for all n, $\eta_n \to 0$ and $\left(\boldsymbol{\alpha}^{(n)}, \boldsymbol{\beta}^{(n)}\right) \to (\boldsymbol{\alpha}_{\varepsilon/2}, \boldsymbol{\beta}_{\varepsilon/2})$ as $n \to \infty$. An immediate discussion for ε-optimality of $\left(\boldsymbol{\alpha}^{(n)}, \boldsymbol{\beta}^{(n)}\right)$ over C_0 would be as follows: For any $\varepsilon > 0$, there exists an N_ε such that by choosing strategy $\boldsymbol{\beta}_{\varepsilon/2}$ player II guarantees to have an expected average payoff equal to at most $\phi^* + \varepsilon/2$ over N_ε stages for any $\boldsymbol{\alpha}$ taken by player I and any initial state. Then, for a sufficiently large n, with strategy $\boldsymbol{\beta}^{(n)}$ for player II the expected average payoff over N_ε stages would be at most $\phi^* + \varepsilon$ for any strategy of player I and any initial state. Since the game value is independent of the initial state, the same arguments hold for the next N_ε stages. Using the law of large numbers it can be concluded that $\phi_i\left(\boldsymbol{\alpha}, \boldsymbol{\beta}^{(n)}\right)$ for any i would be at most $\phi^* + \varepsilon$ for sufficiently large n. Thus, repeating similar arguments for $\boldsymbol{\alpha}^{(n)}$ also, it is shown that $\left(\boldsymbol{\alpha}^{(n)}, \boldsymbol{\beta}^{(n)}\right)$ is ε-optimal over C_0.

Although the arguments in the remark might seem immediate, the proof of Lemma 1 is referred to since the analysis of functions $\min_{\boldsymbol{\beta} \in C_0^2} \phi(\boldsymbol{\alpha}, \boldsymbol{\beta})$ and $\max_{\boldsymbol{\alpha} \in C_0^1} \phi(\boldsymbol{\alpha}, \boldsymbol{\beta})$, and its implications might be useful in devising alternative solution procedures (even under some other less restrictive conditions) and would clarify any question about the short cut arguments (especially, to see how $\boldsymbol{\beta}^{(n)}$ guarantees $\phi^* + \varepsilon$) in the remark. Furthermore, the analysis in Lemma 1 (and Lemma 2 in [1,3], see also [20]) given as the prerequisite of Lemma 1) is related to the research stream, an overview of which is given in [18,17].

Finally, we present our result on the convergence of the value over the restricted spaces, i.e., the value of G^η, to the value of G within some ε distance.

Proposition. *For a communicating stochastic game with a value independent of the initial state, the solution obtained by the proposed procedure over C_η, $\eta > 0$, gives the original game value within ε, for any $\varepsilon > 0$, if η is small enough.*

Proof. Given any $\varepsilon > 0$, consider $\left(\boldsymbol{\alpha}^{(n)}, \boldsymbol{\beta}^{(n)}\right) \in C_{\eta_{(n)}}$ such that $\eta_n > 0$ for all n, $\eta_n \to 0$ and $\left(\boldsymbol{\alpha}^{(n)}, \boldsymbol{\beta}^{(n)}\right) \to (\boldsymbol{\alpha}_\varepsilon, \boldsymbol{\beta}_\varepsilon)$ as $n \to \infty$. From Lemma 1, $\left(\boldsymbol{\alpha}^{(n)}, \boldsymbol{\beta}^{(n)}\right)$ is ε-optimal over C_0 for n sufficiently large. Since $C_{\eta_n} \subset C_0$, given any $\varepsilon > 0$, $\left(\boldsymbol{\alpha}^{(n)}, \boldsymbol{\beta}^{(n)}\right)$ is ε-optimal also over C_{η_n} for sufficiently large n. On the other hand, by definition of $(\boldsymbol{\alpha}^{\eta_n}, \boldsymbol{\beta}^{\eta_n})$,

$$\begin{aligned}\min_{\boldsymbol{\beta}\in C^2_{\eta_n}} \phi(\boldsymbol{\alpha}^{(n)}, \boldsymbol{\beta}) &\leqslant \min_{\boldsymbol{\beta}\in C^2_{\eta_n}} \phi(\boldsymbol{\alpha}^{\eta_n}, \boldsymbol{\beta}) = \phi(\boldsymbol{\alpha}^{\eta_n}, \boldsymbol{\beta}^{\eta_n}) \\ &= \max_{\boldsymbol{\alpha}\in C^1_{\eta_n}} \phi(\boldsymbol{\alpha}, \boldsymbol{\beta}^{\eta_n}) \leqslant \max_{\boldsymbol{\alpha}\in C^1_{\eta_n}} \phi(\boldsymbol{\alpha}, \boldsymbol{\beta}^{(n)})\end{aligned} \tag{3}$$

which shows (relaxing lower and upper bounds in (3) by $\phi^* - \varepsilon$ and $\phi^* + \varepsilon$, respectively, due to ε-optimality of $\left(\boldsymbol{\alpha}^{(n)}, \boldsymbol{\beta}^{(n)}\right)$ over C_{η_n} for sufficiently large n) that $\lim_{n\to\infty} \phi(\boldsymbol{\alpha}^{\eta_n}, \boldsymbol{\beta}^{\eta_n}) = \phi^*$. □

The question to be addressed next would be what the distance (as defined in [10]) from optimality is under the strategies found by the proposed procedure. Note that this distance can be computed for any strategy pair by just solving two MDPs. But the challenge would be to answer how the distance from optimality for $(\boldsymbol{\alpha}^{\eta}, \boldsymbol{\beta}^{\eta})$ changes as η varies. Particularly, when η goes to zero, does it decrease? Although the analysis in this note is not enough to claim the convergence of $(\boldsymbol{\alpha}^{\eta}, \boldsymbol{\beta}^{\eta})$ to an ε-optimal policy pair, we have not encountered a counter example in our experiments.

4. Conclusion

The procedure proposed here is based on the use of Hoffman and Karp's algorithm. Obtaining the game value with an error of ε for any $\varepsilon > 0$ is guaranteed under the condition that the game value is independent of the initial state. It is conjectured that ε-optimal stationary strategies could also be evaluated.

The previous work that could be compared to the proposed procedure is due to Federgruen [9] and Van der Wal [19]. Hoffman and Karp's algorithm is an adaptation of the policy-iteration method devised for solving MDP problems. On the other hand, the method employed in both [9,19] is successive approximation (value-iteration) method, optimal long-run average payoff of a stochastic game and for finding stationary ε-optimal strategies. The first one is based on selecting a sequence of discount factors that approaches 1 and solving a discounted game for this sequence, and converges under the condition that there exist optimal stationary strategies with a value independent of the initial state. The second algorithm due to Federgruen [9] is a special case of his first algorithm where interest rate (discountfactor) is 0 (1). Even in MDPs, the convergence of the value-iteration algorithm depends on the aperiodicity property. This problem arises also for stochastic games. Federgruen suggests the use of Schweitzer's data-transformation to obtain strong aperiodicity, which requires $P_{iabj} > 0$ for all i, j, a, b, and guarantees the convergence when the game is unichain under every pure strategy pair. This recurrence condition implies the conditions given for the convergence of Federgruen's first algorithm (see [9]). Van der Wal [19] proved that if the game is unichain under every pure strategy pair or if the functional equation $g\boldsymbol{e} + \boldsymbol{v} = \max_{\boldsymbol{\alpha}} \min_{\boldsymbol{\beta}} \{\boldsymbol{r}(\boldsymbol{\alpha}, \boldsymbol{\beta}) + P(\boldsymbol{\alpha}, \boldsymbol{\beta})\boldsymbol{v}\}$ has a solution, then his algorithm gives ε-optimal stationary strategies. He also uses Schweitzer's data-transformation if the strong aperiodicity property is not satisfied. To summarize, the differences of the proposed approach from [9,19] are (i) the use of policy-iteration method (instead of value-iteration), (ii) the requirement for communication property (instead of unichain ergodic structure in the second algorithm of Federgruen) and (iii) the concentration on a restricted strategy space over which a communicating stochastic game is irreducible (instead of the use of Schweitzer's data-transformation) to guarantee convergence. In all these approaches, the state-independent game value condition is either imposed directly or implied by the assumption on the ergodic structure (as in the case of unichain game assumption).

Future research topics include proving the convergence of $(\boldsymbol{\alpha}^{\eta}, \boldsymbol{\beta}^{\eta})$ to an ε-optimal policy pair, and devising solution procedures for communicating games with the state-independent value condition being relaxed and incorporating such solution procedures into the algorithms based on the hierarchical decomposition of the state space into communicating classes as in [2,16].

Acknowledgements

The authors are grateful to anonymous referees whose suggestions greatly improved the paper.

References

[1] Z.M. Avşar, M. Baykal-Gursoy, Two-person zero-sum communicating stochastic games, Technical Report, ISE Working Series 97-122, Industrial and Systems Engineering Department, Rutgers University, 1997.

[2] Z.M. Avşar, M. Baykal-Gursoy, A decomposition approach for undiscounted two-person zero-sum stochastic games, Math. Method. Oper. Res. 49 (1999) 483–500.

[3] Z.M. Avşar, M. Baykal-Gursoy, A note on two-person zero-sum communicating stochastic games, Technical Report, ISE Working Series 05-001, Industrial and Systems Engineering Department, Rutgers University, 2005.

[4] J. Bather, Optimal decision procedures in finite Markov chains. Part II: communicating systems, Adv. Appl. Prob. 5 (1973) 541–553.

[5] M. Baykal-Gursoy, K.W. Ross, Variability sensitive Markov decision processes, Math. Oper. Res. 17 (1992) 558–571.

[6] D. Blackwell, T.S. Ferguson, The big match, Ann. Math. Stat. 39 (1968) 159–163.

[7] F. Evangelista, T.E.S. Raghavan, O.J. Vrieze, Repeated ARAT games, Department of Mathematics, University of Wisconsin Center-Marathon, Wausau, January 1995.

[8] A. Federgruen, On *N*-person stochastic games with denumerable state space, Adv. Appl. Probab. 10 (1978) 452–472.

[9] A. Federgruen, Successive approximation methods in undiscounted stochastic games, Oper. Res. 28 (1980) 794–809.

[10] J.A. Filar, T.A. Schultz, F. Thuijsman, D.J. Vrieze, Nonlinear programming and stationary equilibria in stochastic games, Math. Programming 50 (1991) 227–238.

[11] J. Filar, K. Vrieze, Competitive Markov Decision Processes, Springer, New York, 1997.

[12] D. Gillette, Stochastic games with zero stop probabilities, Ann. Math. Stud. 39 (1957).

[13] A.J. Hoffman, R.M. Karp, On nonterminating stochastic games, Manage. Sci. 12 (1966) 359–370.

[14] K.W. Ross, R. Varadarajan, Markov decision processes with a sample path constraints: the communicating case, Oper. Res. 37 (1989) 780–790.

[15] K.W. Ross, R. Varadarajan, Multichain Markov decision processes with a sample path constraint: a decomposition approach, Math. Oper. Res. (1991) 195–207.

[16] E. Solan, Perturbations of Markov Chains with Applications to Stochastic Games, Kellogg School of Management, Northwestern University, 2000.

[17] E. Solan, Continuity of the value of competitive Markov decision processes, J. Theor. Probab. 16 (4) (2003) 831–845.

[18] E. Solan, N. Vieille, Perturbed Markov chains, J. Appl. Probab. 40 (1) (2003) 107–122.

[19] J. Van der Wal, Successive approximations for average reward Markov games, Int. J. Game Theory 9 (1980) 13–24.

[20] G. Yin, Q. Zhang, Singularly perturbed discrete-time Markov chains, SIAM J. Appl. Math. 61 (2000) 834–854.