Completely Mixed Stochastic Games with Small Unfixed Discount Factor

Konstantin Avrachenkov, Anastasiia Varava

Completely Mixed Stochastic Games with Small Unfixed Discount Factor

Konstantin Avrachenkov and Anastasiia Varava

Inria Sophia Antipolis, France and KTH, Sweden
k.avrachenkov@inria.fr, varava@kth.se

Abstract. Motivated by uncertainty in the value of the interest rate, we study discounted zero-sum stochastic games with unfixed discount factor. Our general goal is to obtain a power series expansion of the value of the game with respect to the discount factor around its nominal value. We consider a specific but important class of stochastic games – completely mixed stochastic games. As an illustrative example we take tax evasion model.
Keywords: discounted stochastic games, completely mixed games, unfixed discount factor, power series, Shapley-Snow kernel, tax evasion model.
AMS 2000 subject classification: Primary 91A15, Secondary 49L20

1 Introduction

The present work is motivated by uncertainty in the value of the interest rate. We consider the discounted stochastic game and pose a question what happens with the value of the game if the interest rate, or equivalently, the discount factor, deviates from its nominal value. In the spirit of the perturbation analysis [2], we try to find efficient algorithms for computation of some initial coefficients of the power series of the value of the game with respect to the discount factor.

The perturbation analysis of stochastic games with respect to the discount factor appears to be very challenging in its full generality (see e.g., [7, 11]). Therefore, in this work we limit ourselves to a specific but important class of stochastic games – completely mixed stochastic games. In particular, in the case of completely mixed stochastic games the value of the game has a Taylor series expansion at the vicinity of zero discount factor.

Our approach is based on generalization of the Shapley value iterations [9] from the field of real numbers to the field of power series. Such an approach was successfully used before for Blackwell optimality in Markov decision processes [6], for singularly perturbed Markov decision processes [1] and for Blackwell equilibrium [3] in stochastic games with perfect information or switching controller games [4]. It is interesting to observe that in the present setting each Shapley value iteration produces an exact new coefficient in the Taylor series of the game value.

The structure of the paper is as follows: In the next section we define the discounted stochastic game and provide necessary background material on stochastic and matrix games. In Section 3 we study the case of completely mixed stochastic games. In Section 4 we provide an illustrative example of the tax evasion model. Finally, in Section 5 we give conclusions and discuss open problems.

2 Background on Discounted Stochastic Games and Matrix Games

The notion of stochastic game was first introduced by Shapley in 1953 [9]. Following [9], we consider two-person zero-sum stochastic games on infinite time horizon and discounted payoff. The game has a finite set of positions, called states. For each state there are two finite sets of actions for the first and the second player, respectively. Each pair of actions corresponding to the same state defines the immediate reward for both players as well as the probabilities of transitions to the other states. At each step the players simultaneously choose actions and receive corresponding rewards. After that the system immediately moves to the next state with respect to the probability distribution defined by chosen pair of actions. Let us next formally define the two-person zero-sum stochastic game (an interested reader can find much more information on stochastic games in the book by Filar and Vrieze [5]).

Definition 1. *A system with the following structure is called two-person zero-sum stochastic game Γ:*

1. *there are two players, P_1 and P_2 (also called "the first player" and "the second player", respectively);*
2. *$S = \{1, 2, .., n\}$ is a finite set of states of the game;*
3. *$A^i(s) = \{1, 2, .., m_i(s)\}$ represent sets of actions of i^{th} player with respect to the current state $s \in S$;*
4. *the function $r(s, i, j)$ represents immediate rewards for player 1, and $-r(s, i, j)$ is the immediate reward for player 2. Here $s \in S$, $i \in A^1(s), j \in A^1(s)$, which means that the game is currently in state s and the players choose actions i and j, respectively. One usually denotes the matrix of immediate rewards in state s by $R(s)$.*
5. *Transition probabilities $p(s'|s, i, j) : s, s' \in S, i \in A^1(s), j \in A^2(s)$ where $p(s'|s, i, j)$ is the probability of transition from state s to state s' given that players 1 and 2 choose actions $i \in A^1(s), j \in A^2(s)$, respectively. It is assumed that the transition probabilities and the immediate rewards are known to both players.*

A strategy for a player is a rule of selecting an action at each step of the game. In general, strategies can depend on complete history of the game until the current stage. Such strategies are called behavioural strategies. We are looking at the simpler class of strategies called stationary strategies which depend only on the current state s and not on how s has been reached.

Definition 2. *A stationary strategy of a player is a function from the state space to the set of probability distributions on player's action set. A strategy is called pure or deterministic when for each state of the game the player deterministically choose exactly one action with probability 1.*

In his paper [9], Shapley has shown that it is enough to consider only stationary strategies for the discounted stochastic games, so we restrict ourselves to them.

Discounted payments are accumulating throughout the game. The first player aims to maximize the β-discounted payoff, whereas the second player aims to minimize it.

Definition 3. *Given an initial state s_0, a pair of stationary strategies (f, g) of players 1 and 2, resp., and a discount factor $\beta \in [0, 1)$, we define β-discounted payoffs as follows:*

$$[J_\beta(f, g)](s_0) = \sum_{t=0}^{\infty} \beta^t E_{s_0}^{f,g}[r_t],$$

where t corresponds to discrete moments of time and r_t is an immediate payoff on a corresponding (t^{th}) step of the game with respect to the initial state s_0 and strategies of the players.

Under this payoff one can define an equilibrium pair of strategies and the value vector of the game.

Definition 4. *A pair (f^*, g^*) such that*

$$[J_\beta(f, g^*)](s) \leq [J_\beta(f^*, g^*)](s) \leq [J_\beta(f^*, g)](s),$$

for all f and g and for all $s \in S$, is called a pair of equilibrium strategies. The vector $J_\beta(f^*, g^*)$ is called equilibrium value vector, or game value vector and is denoted by $\mathbf{v}$.

One can view a zero-sum discounted stochastic game $\Gamma(\beta)$ as a generalization of static matrix game to a multistate and multistage situation. Indeed, assume that $|S| = N$ and $m_1(s) = |A^1(s)|, m_2(s) = |A^2(s)|$ for each $s \in S$. Then we can naturally define N matrix games that are in one-to-one correspondence with the states of $\Gamma(\beta)$:

$$R(s) = [r(s, a_1, a_2)]_{a_1=1, a_2=1}^{m_1(s), m_2(s)}$$

Now we can think of each state of a game as of a simple matrix game $R(s)$ with corresponding action sets $A^1(s), A^2(s)$. Notice that actions of players determine not only their rewards (as in the classical matrix game), but also probability transitions $p(s'|s, a_1, a_2)$ to the matrix game $R(s')$ that can be played at the next step.

In his original paper [9], Shapley has shown that under the discounted payoff criterion, there always exist an equilibrium pair of strategies, and the equilibrium value vector $\mathbf{v}(f^*, g^*)$ is unique. More precisely, the following theorem takes place:

Theorem 1. (*Shapley, 1953*) *The discounted, zero-sum, stochastic game $\Gamma(\beta)$ possesses the value vector* $\mathbf{v}$ *that is unique solution of the equations*

$$v(s) = \mathrm{val}[R(s, \mathbf{v})]$$

for all $s \in \mathbf{S}$*, where* $\mathbf{v}^T = (v(1), v(2), ..., v(N))^T$ *and*

$$R(s, \mathbf{v}) = \left[r(s, a_1, a_2) + \beta \sum_{s' \in \mathbf{S}} p(s'|s, a_1, a_2) v(s') \right]$$

We call the following set of equations N (one equation per state of the game) "Shapley equations":

$$v(s) = \mathrm{val} \left[r(s, a_1, a_2) + \beta \sum_{s' \in \mathbf{S}} p(s'|s, a_1, a_2) v(s') \right], \ s \in S$$

In addition, we call the following map "Shapley operator":

$$\mathrm{T} : \mathbb{R}^N \to \mathbb{R}^N, \quad \mathrm{T}(\mathbf{x}_s) = \mathrm{val}\, R(s, \mathbf{x}_s)$$

Shapley [9] has shown that this operator is a contraction with coefficient β. In the above mentioned paper he has also shown how one can reduce the process of solving stochastic game to solving several static matrix games. The idea of this algorithm is based on the fact that the operator T is a contraction. The value of the game is the unique fixed point of this operator, or, in other words, the unique solution of the equation

$$\mathbf{x} = \mathrm{T}\, \mathbf{x}.$$

By simple iterations one can find the approximated value of the game. Let $\mathbf{v}_0$ be an arbitrary initial approximation. Then, by Banach fixed point theorem, the sequence of approximations $\mathbf{v}_k$ converges to the exact solution $\mathbf{v}$ of the game:

$$\mathbf{v}_{k+1} = \mathrm{T}\, \mathbf{v}_k, \quad k = 0, 1, ...$$

Notice that at each iteration we have to solve N matrix games. As we mentioned in the previous section, the problem of solving a matrix game is of polynomial complexity (e.g., is solvable by linear programming). We will call this algorithm Shapley value iteration.

Let us also recall some useful facts about static matrix games. Each static matrix game can be presented as a matrix M by identifying rows with pure strategies of player 1 and columns with pure strategies of player 2. The element $M[a_1, a_2]$ of a matrix represents the reward $r(a_1, a_2)$. Clearly, the first player aims to maximize his payoff, whereas the second player aims to minimize his cost. We assume that both players are rational. It is known that there always exists a pair of equilibrium strategies (f^*, g^*) such that:

$$\forall f, g : \ r(f, g^*) \leq r(f^*, g^*) \leq r(f^*, g).$$

The value $v = r(f^*, g^*)$ is called the value of the game and is known to be unique. By solving a game one usually means finding its value (and, possibly, optimal strategies). Solving a matrix game is not a trivial problem. There are several approaches to it, and one of them is due to Shapley and Snow [10].

Theorem 2. (*Shapley and Snow, 1950*) *If A is a matrix game and* $\operatorname{val} A \neq 0$, *then A has a square invertible submatrix $\hat{A}$, called a Shapley-Snow kernel, such that:*

- $\operatorname{val} A = \operatorname{val} \hat{A} = \dfrac{\det \hat{A}}{\sum_{ij} \operatorname{adj} \hat{A}[i,j]}$;
- *There is a pair of equilibrium strategies $(\hat{x}, \hat{y})$ for $\hat{A}$ which are also equilibrium strategies for A (after inserting zeroes at corresponding entries) that satisfy*

$$(\hat{x})^T = (\operatorname{val} A)\mathbf{1}^T \hat{A}^{-1}, \quad \hat{y} = (\operatorname{val} A)\hat{A}^{-1}\mathbf{1},$$

where $\mathbf{1} = (1, .., 1)^T$.

Without loss of generality, we assume that all rewards of the game are strictly positive and hence the value is positive as well, and so Shapley-Snow kernels are always defined.

Following [11], let us call Shapley-Snow kernels, which are completely mixed, *cmv-kernels*. It has been shown in [11] that they always exist in arbitrary matrix game. From now we will consider only these kernels.

Despite the fact that this theorem has a theoretical value, in practice matrix games are usually solved in other ways, e.g., by linear programming. In particular, this implies that a matrix game can be solved in polynomial time.

3 Completely Mixed Stochastic Games

Consider a special class of stochastic games – completely mixed stochastic games.

Definition 5. *Stochastic game is called completely mixed, if for each state $s \in S$, the Shapley matrix*

$$R(s, \mathbf{v}(\beta), \beta) = \left[r_{s,i,j} + \beta \textstyle\sum_{l=1..N} p^l_{s,i,j} v_l(\beta)\right]_{i,j=1}^{|A^1(s)|\times|A^2(s)|}, \quad \forall s \in S$$

is completely mixed.

In other words, for each state s the cmv-kernel of the Shapley matrix $R(s, \mathbf{v}(\beta), \beta)$ is the whole matrix per se. Clearly, in this case for each $s \in S$ it is required that $|A^1(s)| = |A^2(s)|$.

We can propose an easily verifiable condition to check if a stochastic game is completely mixed for small values of the discount factor.

Assumption 1. *Assume that we are given a game $\Gamma(\beta)$ in which the static matrix game at each state $s \in S$ is completely mixed.*

The above assumption implies useful structural properties of the stochastic game. Namely, we have

Lemma 1. *Let Assumption 1 hold. Then, the stochastic game $\Gamma(\beta)$ is completely mixed for the values of the discount factor β in some interval $[0, \delta)$ and the value vector of the game possesses a Taylor series expansion in the vicinity of $\beta_0 = 0$ with the convergence radius $R \le \delta$.*

Proof. Since the set of completely mixed matrix games is open in the space of all matrix games of the corresponding dimension [11], there exist a neighborhood $\beta \in \omega = [0, \delta)$, such that for all $\beta \in \omega$ and all states $s \in S$, the games defined by $R(s, \mathbf{v}(\beta), \beta)$ are completely mixed. It then implies that for all $\beta \in \omega$ the stochastic game $\Gamma(\beta)$ is completely mixed as well.

It follows from the arguments in the proof of Lemma 4.1 in [11] that in this case the value function $\mathbf{v}(\beta)$ of the stochastic game $\Gamma(\beta)$ is analytic on $\beta \in \omega$. More precisely, the fact that the cmv-kernels of Shapley matrices do not change in some neigbourhood of $\beta = 0$ is crucial here. In its turn, it is a consequence of the completely-mixed assumption on the payoff matrices.

Thus, the value of the game can be represented as a Taylor series expansion around zero for each state s, with the convergence radius $0 < R \le \delta$. □

Our goal is to find an approximation of the value function $\mathbf{v}(\beta)$ given by the first m terms of its power series expansion. We take a constant vector as an initial approximation, e.g., $\mathbf{v}_0(\beta) := 0$. By $\mathbf{v_k}(\beta)$ we denote the k^{th} approximation of $\mathbf{v}(\beta)$. Then the next approximation can be obtained as follows:

$$v_{k+1,s}(\beta) = \frac{\det\left(R(s, \mathbf{v_k}(\beta), \beta)\right)}{\sum_{i,j=1}^{r} \operatorname{adj}\left(R(s, \mathbf{v_k}(\beta), \beta)\right)[i,j]}, \qquad \forall s \in S. \tag{1}$$

Here $\operatorname{adj}\left(R(s, \mathbf{v_k}(\beta), \beta)\right)$ is the adjugate matrix of $R(s, \mathbf{v_k}(\beta), \beta)$. Recall that each entry of $R(s, \mathbf{v_k}(\beta), \beta)$ looks as follows:

$$R(s, \mathbf{v_k}(\beta), \beta)[i,j] = r_{s,i,j} + \beta \sum_{l=1..N} p^l_{s,i,j} v_{k,l}(\beta).$$

Each $v_{k,l}(\beta)$ is a rational function, continuous (and analytical) for $\beta \in \omega$. Thus, we can present it as a power series expansion

$$v_{k,l}(\beta) = \sum_{i=0}^{\infty} a^i_{k,l} \beta^i.$$

Substituting this representation into the Shapley matrix, we obtain

$$\begin{aligned} R(s, \mathbf{v_k}(\beta), \beta)[i,j] = r_{s,i,j} + \beta \sum_{l=1..N} p^l_{s,i,j} \left(\sum_{i=0}^{\infty} a^i_{k,l} \beta^i \right) = \\ r_{s,i,j} + \beta \sum_{l=1..N} p^l_{s,i,j} \left(a^0_{k,l} + a^1_{k,l}\beta + a^2_{k,l}\beta^2 + ... \right) = \\ r_{s,i,j} + \beta \sum_{l=1..N} p^l_{s,i,j} a^0_{k,l} + \beta^2 \sum_{l=1..N} p^l_{s,i,j} a^1_{k,l} + \beta^3 \sum_{l=1..N} p^l_{s,i,j} a^2_{k,l} + ... \end{aligned} \tag{2}$$

Let us define a reduction $\mod \beta^m$ on the power series:

$$a(\beta) = \sum_{i=0}^{\infty} a_i \beta^i;$$

$$a(\beta) \mod \beta^m = \sum_{i=0}^{\infty} a_i \beta^i \mod \beta^m = \sum_{i=0}^{m} a_i \beta^i.$$

It can be easily seen that if

$$a(\beta) \mod \beta^m = b(\beta) \mod \beta^m, \quad c(\beta) \mod \beta^m = d(\beta) \mod \beta^m$$

then

$$(a(\beta) \pm c(\beta)) \mod \beta^m = (b(\beta) \pm d(\beta)) \mod \beta^m, \tag{3}$$

$$(a(\beta) \cdot c(\beta)) \mod \beta^m = (b(\beta) \cdot d(\beta)) \mod \beta^m. \tag{4}$$

Now let us look at the formula of the value of a completely mixed matrix game $R(s, \mathbf{v_k}(\beta), \beta)$:

$$\operatorname{val}(R(s, \mathbf{v_k}(\beta), \beta)) = \frac{\det(R(s, \mathbf{v_k}(\beta), \beta))}{\sum_{i,j=1}^{r} \operatorname{adj}(R(s, \mathbf{v_k}(\beta), \beta))[i,j]}. \tag{5}$$

Firstly consider the numerator. Let us use the series representation of the entries of the Shapley matrix. Then from the way of computing the determinant it is easy to see that the numerator can be written as follows:

$$\det(R(s, \mathbf{v_k}(\beta), \beta)) = c^0_{k,s} + c^1_{k,s}\beta + c^2_{k,s}\beta^2 + c^3_{k,s}\beta^3 + ...$$

So, it can be represented as a power series expansion.

Now consider the denominator. Similarly, the sum of the elements of the adjugate matrix is the sum of cofactors of the initial matrix, which are (up to sign) determinants of its $r-1$ submatrices. So the denominator can be written as

$$\sum_{i,j=1}^{r} \operatorname{adj}(R(s, \mathbf{v_k}(\beta), \beta))[i,j] = d^0_{k,s} + d^1_{k,s}\beta + d^2_{k,s}\beta^2 + d^3_{k,s}\beta^3 + ...$$

Note that from the definition of a completely-mixed game we have $c^0_{k,s} \neq 0$, $d^0_{k,s} \neq 0$. Now $v^s_{k+1}(\beta)$ can be expressed as a ratio of two power series:

$$v^s_{k+1}(\beta) = \operatorname{val}(R(s, \mathbf{v_k}(\beta), \beta)) = \frac{c^0_{k,s} + c^1_{k,s}\beta + c^2_{k,s}\beta^2 + c^3_{k,s}\beta^3 + ...}{d^0_{k,s} + d^1_{k,s}\beta + d^2_{k,s}\beta^2 + d^3_{k,s}\beta^3 + ...} \tag{6}$$

This is one of the key observations in our method. From now we will use the fact that $v^s_{k+1}(\beta)$ can be represented as

$$v_{k+1,s}(\beta) = \frac{P_{k,s}(\beta)}{Q_{k,s}(\beta)},$$

where $P_{k,s}(\beta)$ and $Q_{k,s}(\beta)$ are power series of β:

$$P_{k,s}(\beta) = c^0_{k,s} + c^1_{k,s}\beta + c^2_{k,s}\beta^2 + c^3_{k,s}\beta^3 + ...,$$

and

$$Q_{k,s}(\beta) = d^0_{k,s} + d^1_{k,s}\beta + d^2_{k,s}\beta^2 + d^3_{k,s}\beta^3 + ...$$

We shall show that if at some moment K we have calculated the value function expansion at $\beta = 0$ up to the m-th term, then at the next iteration we get the next $(m+1)$-st term of the expansion. Formally, we can formulate the following statement.

Theorem 3. *Let $\Gamma(\beta)$ be a discounted stochastic game with unfixed discount parameter $\beta \in [0, 1)$ and let $\mathbf{v}(\beta)$ be its value function. Let Assumption 1 hold. Then starting from $\mathbf{v_0} = \underline{0}$, we can obtain K first terms of the Taylor series expansion of $\mathbf{v}(\beta)$ at $\beta_0 = 0$ after K iterations given by* (1).

Proof. Notice that when $\beta = 0$ we have

$$v_{k+1,s}(0) = \frac{c^0_{k,s}}{d^0_{k,s}}.$$

In particular, if $k = 0$ and $\mathbf{v_0} = \underline{0}$, we have that

$$v_{1,s} = \mathrm{val}(R(s, \underline{0}, \beta)) = \mathrm{val}(R(s)) = v_s(0). \tag{7}$$

Let us now consider the derivative of order m of the value function approximation

$$v_{k+1,s}(\beta) = \frac{P_{k,s}(\beta)}{Q_{k,s}(\beta)}.$$

We can think of $v_{k+1,s}(\beta)$ as a rational function of $P_{k,s}(\beta)$ and $Q_{k,s}(\beta)$. Similarly, the derivative $v^{(m)}_{k+1,s}(0)$ is a rational function of $P_{k,s}(0)$, $P^{(1)}_{k,s}(0)$, ..., $P^{(m)}_{k,s}(0)$ and $Q_{k,s}(0)$, $Q^{(1)}_{k,s}(0)$, ..., $Q^{(m)}_{k,s}(0)$.

On the other hand, notice that $P^{(q)}_{k,s}(0) = q! \cdot c^q_{k,s}$ and $Q^{(q)}_{k,s}(0) = q! \cdot d^q_{k,s}$. So, we can say that ${v_{k+1,s}}^{(m)}(0)$ is a rational function of $c^q_{k,s}$ and $d^q_{k,s}$, for $q \in \{0, .., m\}$. Therefore, the first $m+1$ terms of the Taylor series expansion of $v_{k+1,s}(\beta)$ are completely defined by $c^q_{k,s}$ and $d^q_{k,s}$, for $q \in \{0, .., m\}$.

The latter means that if for some $K \in \mathbb{N}$:

$$\begin{aligned}\forall s \in S,\ \forall o, p > K :& \\ & P_{o,s}(\beta) \mod \beta^m = P_{p,s}(\beta) \mod \beta^m, \\ & Q_{o,s}(\beta) \mod \beta^m = Q_{p,s}(\beta) \mod \beta^m,\end{aligned} \tag{8}$$

then the first $m+1$ terms of the value function series expansion will not change at all on the subsequent iterations:

$$\forall s \in S,\ \forall o, p > K + 1 :\quad v_{o,s}(\beta) \mod \beta^m = v_{p,s}(\beta) \mod \beta^m \tag{9}$$

On the other hand, recall that $c^q_{k,s}$ and $d^q_{k,s}$ for $q \in \{0,..,m\}$ are defined by the corresponding Shapley matrix approximation, $R(s, \mathbf{v}_k(\beta), \beta)[i,j]$. Thus, if K terms are comuted exactly, the $(K+1)$-st term will be exactly derived at the next iteration. Invoking the principle of mathematical induction together with the base (7), we conclude the proof of the theorem. □

The complexity of the proposed approach depends on the number of states, on the number of possible actions at each state and, obviously, on the number of terms of the expansion that we are looking for. Assume that we have N states, at each state $s \in \{1,..,N\}$ both of the players have r_s possible actions, and we want to compute K terms of the series expansion of each of the value component. Observe that for each state s at each iteration we have to compute the determinant and the adjugate of an $r_s \times r_s$ matrix. We can do this performing $O(r_s^3)$ operations per iteration using Gaussian elimination. This means that in general our algorithm requires $O(K \cdot \sum_{s=1}^{N} r_s^3)$ operations.

4 Example: Tax Evasion Model

As it has been already mentioned, stochastic games is a powerful tool for modeling different real-life situations. They have applications in economics, evolutionary biology, computer networks etc. In this section we consider a simple model of tax evasion as a simple example of a stochastic game application. This model is inspired by [8], but we slightly simplify it in our work, since this particular application is not the main objective of the present work. We propose a numerical example, which is a two-states completely mixed game, and we find an approximation in terms of power series using our technique described above.

We consider a situation where there are two agents with the opposite interests: the taxpayer and the auditor. The objective of the former is to pay as less as possible, while the latter aims to collect as much money as possible. Each time slot, say, each month, the taxpayer has to decide whether declare his income honestly or not. In his turn, each month the auditor can decide whether to trust the taxpayer or to audit. Hence both of the agents have two possible behaviors. To motivate taxpayers to be honest, the auditor introduces a system of penalties and rewards. Furthermore, there are two different possible states. Normally, the taxpayer is assumed to be honest. In such a situation penalties are lower, and rewards are higher. However, if the auditor suspects (based on the experience of previous months) that the taxpayer is a cheater, then rewards are lower and penalties are higher. Hence we have two different states: when the taxpayer is presumed to be honest and when he is assumed to be suspicious. Since the taxpayer can be “short-sighted” and the auditor can be oblivious or subject to staff mobility, we feel that our setting of small unfixed discount factor is particularly relevant in this model. Formally, we can describe the model as the following stochastic game:

Player 1: auditor;

Player 2: taxpayer;

Pure strategies of Player 1: to audit or not to audit (to trust);

Pure strategies of Player 2: to declare the income honestly or to cheat;

States of the game: State 1 (Good) – the taxpayer is presumed to be honest; State 2 (Bad) – the taxpayer is presumed to cheat;

Payoffs: the amount of money that the taxpayer pays; it is defined by 3 constants: n – normal tax; p – penalty for cheating; r – reward for being honest.

The payoff is accumulated throughout the game with a discount parameter $\beta \in [0, 1)$.

We assume that in a "good" state the taxpayer receives a reward for being honest in case if the auditor decides to check his income. He pays the penalty, if he is found to be guilty. Moreover, in this case the auditor becomes more severe and the game moves to the "bad" state, where the penalty is higher and the reward for being honest is lower. However, if after moving to the "bad" state the taxpayer becomes honest and the auditor notices this, the game moves back to the "good" state. The payoffs and the transition probabilities can be expressed with the help of the following tables:

Table 1. State 1 (Good)

	Be honest	**Cheat**
Audit	$n-r$ (1, 0)	$n+p$ (0, 1)
Trust	n (1, 0)	0 (1, 0)

Let us now consider concrete numerical example. Let $n = 5$, $p = 2$, $r = 2$. Then we have the following Shapley equations:

$$v_1(\beta) = \text{val} \begin{bmatrix} 3 + \beta \cdot v_1(\beta) & 7 + \beta \cdot v_2(\beta) \\ 5 + \beta \cdot v_1(\beta) & 0 + \beta \cdot v_1(\beta) \end{bmatrix},$$

$$v_2(\beta) = \text{val} \begin{bmatrix} 4 + \beta \cdot v_1(\beta) & 9 + \beta \cdot v_2(\beta) \\ 5 + \beta \cdot v_2(\beta) & 0 + \beta \cdot v_2(\beta) \end{bmatrix}$$

It is easy to check that when β is close to zero, both of the games at the right-hand side are completely-mixed. On the one hand, the exact solution of this game is given by the following system of equations:

$$v_1(\beta) = \frac{(3 + \beta v_1(\beta))\beta v_1(\beta) - (7 + \beta v_2(\beta))(5 + \beta v_1(\beta))}{\beta v_1(\beta) - \beta v_2(\beta) - 9},$$

Table 2. State 2 (Bad)

	Be honest	**Cheat**
Audit	$n - r/2$ (1, 0)	$n + 2 \cdot p$ (0, 1)
Trust	n (0, 1)	0 (0, 1)

$$v_2(\beta) = \frac{(4 + \beta v_2(\beta))\beta v_2(\beta) - (9 + \beta v_2(\beta))(5 + \beta v_2(\beta))}{\beta v_1(\beta) - \beta v_2(\beta) - 10}. \tag{10}$$

We see that even in this simple example the exact solution is not easy to compute and to represent analytically.

On the other hand, we can apply our technique to easily find the solution as a series expansion in the neighbourhood of zero. Using the system for symbolic computations *Maple*, we obtain the following segments of the series:

$$v_1(\beta) = 35/9 + 2890/729\beta + 940969/236196\beta^2 + 95458552/23914845\beta^3 + \mathrm{O}(\beta^4),$$

$$v_2(\beta) = 9/2 + 1499/360\beta + 593543/145800\beta^2 + 762448589/188956800\beta^3 + \mathrm{O}(\beta^4).$$

5 Conclusion and Discussion

Motivated by uncertainty in the value of the interest rate, we study discounted zero-sum stochastic games with unfixed discount factor. Our general goal is to obtain a power series expansion of the value of the game with respect to the discount factor around its nominal value. Even though we could not solve the problem in its full generality, we considered a specific but important class of stochastic games – completely mixed stochastic games. In this class of games we show that the value vector can be expanded as a Taylor series near zero discount factor and provide a generalization of the Shapley iterations to compute an initial segment of the Taylor series. It is very interesting that iterations subsequently produce exact values of the Taylor series coefficients. We illustrate our technique on tax evasion model.

There is a number of very interesting open research questions. In the case of completely mixed games, can the generalized Shapley iterations be adapted to some nominal values of the discount factor different from zero? Our numerical experiments indicate that such generalization is likely to be possible but there is no any more nice term-by-term convergence. Then, how to estimate the radius of convergence of the obtained power series? If the game is not completely mixed, how one can compute the Puiseux series expansion of the value of the game? In the general case, one needs to deal with the fractional power Puiseux series expansions instead of the Taylor series.

Acknowledgement

This work is partly supported by EU Project Congas FP7-ICT-2011-8-317672.

References

1. Altman, E. Avrachenkov K.E. and Filar, J.A.: Asymptotic linear programming and policy improvement for singularly perturbed Markov decision processes. *Math. meth. oper. res.* **49(1)** (1999) 97–109.
2. Avrachenkov, K.E. Filar J.A. and Howlett, P.G. *Analytic Perturbation Theory and its Applications.* SIAM, 2013.
3. Avrachenkov, K. Cottatellucci, L. and Maggi, L.: Algorithms for uniform optimal strategies in two-player zero-sum stochastic games with perfect information. *Operations Research Letters*, **40(1)** (2012) 56–60.
4. Filar, J.A.: Order field property of stochastic games when the player who controls the transition changes from state to state. *J. Optim. Theory Appl.*, **34** (1981) 505–517.
5. Filar, J.A. and Vrieze, K. *Competitive Markov Decision Processes*, Springer, 1997.
6. Hordijk, A. Dekker, R. and Kallenberg, L.C.M.: Sensitivity-analysis in discounted Markovian decision problems. *Operations-Research-Spektrum*, **7(3)** (1985) 143–151.
7. Neyman, A.: Real algebraic tools in stochastic games. *Stochastic games and applications*, NATO Science Series C, Mathematical and Physical Sciences, **570** (2003) 57–75.
8. Raghavan, T.E.S.: A stochastic model of tax evasion. *Annals of the International Society of Dynamic Games*, **8** (2006) 397–420.
9. Shapley, L.S.: Stochastic games. *Proceedings of the National Academy of Sciences*, **39** (1953) 1095–1100.
10. Shapley, L.S. and Snow, R.N.: Basic solutions of discrete games. *Annals of Mathematics Studies*, **24** (1950) 27–35.
11. Szczechla, W. Connell, S. Filar, J. and Vrieze, K.: (1994) On the Puisuex series expansion of the limit discount equation of stochastic games. *SIAM J. Control Optim.* **35(3)** (1994) 860–875.