Perturbed zero-sum games with applications to dynamic games

Eitan ALTMAN
INRIA, B.P. 93
2004 Route des Lucioles
06902 Sophia-Antipolis Cedex, France

Eugene A. FEINBERG
Harriman School for Management and Policy
SUNY at Stony Brook
Stony Brook, NY 11794-3775, U.S.A

Jerzy FILAR
School of Mathematics
University of South Australia
The Levels, Australia, 5095

Vladimir A. GAITSGORY
School of Mathematics
University of South Australia
The Levels, Australia, 5095

Abstract

This paper deals with perturbed matrix games. The main result is that the sets of solutions of perturbed games converge to subsets of solutions of appropriate lexicographic games. We consider applications of these results to dynamic games. In particular, we consider applications to repeated games with weighted discounted criteria and to finite-horizon stochastic games with perturbed transition probabilities and rewards.

1 Introduction

This paper deals with perturbed matrix games and their applications to stochastic and repeated games. For matrix games with continuously perturbed matrices, the values of perturbed games converge to the value of the original game; Tijs and Vrieze [8] or Filar and Vrieze [2], Proposition G.5. For continuous perturbations, the upper limits of the sets of optimal solutions of matrix games belong to the sets of optimal solutions of the original game; see the same references. In the case of a linear perturbation, the value of the game is differentiable in the direction of its perturbation and this derivative is equal to the value of the following lexicographic game: the players play the game with a perturbation matrix on the sets of optimal policies for the original game; Mills [5].

In this paper we study perturbations which are more general that linear perturbations. In particular, we consider perturbations which contain a linear term plus a sum of terms proportional

to the higher powers of the perturbation parameter. These powers may be noninteger. We show that Mill's [5] result also holds for these more general perturbations (Corollary 3.1). In addition, the limits of the sets of optimal policies exist and belong to the sets of solutions of the described lexicographic game (Theorem 3.1). For the simplest case of a linear perturbation, Example 3.1 below shows that the limit of solution sets can be strictly smaller than the set of solutions for the appropriate lexicographic game. For perturbations of higher orders it is possible to consider lexicographic games when the players play sequentially the games with perturbation matrixes of higher orders on the polytopes of optimal solutions for previous lexicographic optimal policies. Example 3.2 shows that, in the case of a quadratic perturbation, the limiting solutions of the perturbed game may have no common points with the optimal solutions of the lexicographic game which the players play on the sets of lexicographic solutions for the linear perturbed games with the payoff matrix being the matrix of the second-order perturbations. Our proofs are based on the analysis of asymptotic solutions of linear programs (see Jeroslow [4]) for perturbed matrix games.

We consider applications of these results to stochastic and repeated games. In particular, we consider two models: (i) finite horizon stochastic games with finite state and action sets and with perturbed transition probabilities and reward functions, (ii) repeated games with objective functions equal to the sums of the total discounted rewards with different discount factors. In particular, for repeated games with several discount factors we show that any limit of optimal solutions is optimal for the following lexicographic matrix game: (a) play the game with the payoffs that correspond to the biggest discount factor, (b) play the game with the payoffs that correspond to the second biggest discount factor on the sets of optimal policies for the first game.

2 The model

Let G_ϵ be a family of matrices of size $m \times n$ and $\epsilon \in [0, \epsilon^*]$, where ϵ^* is a positive number. We denote $G = G_0$.

Player 1 maximizes over the policies U, which are probabilities over the m rows, and player 2 minimizes over the policies V, which are the probability measures over the n columns.

We assume elements of G_ϵ are continuous in ϵ. The zero-sum game with matrix G_ϵ is called a perturbed game, and we are interested in characterizing the limit of the values and optimal policies of G_ϵ.

We consider two types of perturbations:

- (P1) $G_\epsilon = G + \epsilon F$, where G and F be two real matrices of size $m \times n$,

- (P2) $G_\epsilon = G + \epsilon F + \epsilon^2 F_1 + \ldots + \epsilon^{L+1} F_L$, where G, F and $F_l, l = 1, \ldots, L$, are real matrices of size $m \times n$.

Unless otherwise stated, we consider the more general case (P2).

For a matrix game with a matrix H we denote by U_H and V_H the sets of optimal policies for player 1 and 2 respectively. These sets are convex polytopes whose extreme points can be computed via linear programming. We say that a policy is basic if it corresponds to a basis in the LP. For $H = G_\epsilon$ we sometimes write U_ϵ and V_ϵ instead of U_{G_ϵ} and V_{G_ϵ} respectively.

Since the sets of basic variables for game G_ϵ is finite, there exists some interval $I =]0, \epsilon_0]$ such that for all ϵ in that interval, the same set of basic variables is optimal in the LP (see [4]). The policy u_ϵ corresponding to any basis can be expressed as a rational function of $\epsilon \in]0, \epsilon_0]$, i.e. the ratio between two polynomials in ϵ with real coefficients (see [4]), and therefore it can also be given as a Laurent expansion of the form

$$u_\epsilon = u_0 + \epsilon u_1 + \epsilon^2 u_2 + \ldots . \tag{1}$$

There are no negative powers of ϵ since, clearly, u_ϵ, being bounded, does not have poles. (1) implies

$$u_\epsilon = u_0 + \epsilon u_1 + o(\epsilon). \tag{2}$$

The similar representation also holds for v_ϵ corresponding to a fixed basis in each one of the LPs that are used for computing the optimal policy for the 2nd player.

As observed by Jeroslow [4], there is a finite number of basic sets in the LP and for each basic set (P2) implies that the value of the game and the optimal policies are rational functions of ϵ. Therefore for some positive ϵ_0 there is a finite number, say K, of sets of basic variables in the LP and each of these sets corresponds to optimal policies (u_ϵ, v_ϵ) for all $\epsilon \in I =]0, \epsilon_0]$. In other words, there are numbers $\{(u_0^k, v_0^k) |\ k = 1, 2, \ldots, K\}$ such that u_ϵ, $\epsilon \in]0, \epsilon_0[$, is an optimal basic policy for player 1 in game G_ϵ if and only if for some $k = 1, \ldots, K$

$$u_\epsilon = u_0^k + \epsilon u_1^k + o(\epsilon) \tag{3}$$

The similar representation takes place for player 2.

It follows from the above discussion and from (3) that the value of G_ϵ can be expanded as:

$$\textbf{val}\ G_\epsilon = u_0^* G v_0^* + \epsilon(u_0^* G v_1^* + u_1^* G v_0^* + u_0^* F v_0^*) + o(\epsilon), \qquad \epsilon \in I, \tag{4}$$

where u_i^* and v_i^* are the coefficients in representations (3) when an arbitrary optimal basis is fixed for each player on the interval $]0, \epsilon_0]$.

Let $U_G \subset U$ and $V_G \subset V$ be the compact sets of policies that are optimal for the two players in the game G. Consider the game F over the restricted set of policies U_G and V_G, and denote by U_{GF} and V_{GF} the corresponding sets of optimal policies. We call this a lexicographic game, and these sets - the sets of lexicographic optimal policies. The value of this game is denoted by **val** (GF).

3 Main results

Consider matrix games G and G_ϵ. If $\lim_{\epsilon\to 0} G_\epsilon = G$ then

$$\lim_{\epsilon\to 0} \textbf{val } G_\epsilon = \textbf{val } G,$$

see Theorem 2.1 in Tijs and Vrieze [8] or Proposition G.5 in Filar and Vrieze [2]. If $\epsilon(l) \to 0$, $(u_\epsilon, v_\epsilon) \in (U_\epsilon, V_\epsilon)$, and $(u_{\epsilon(l)}, v_{\epsilon(l)}) \to (u, v)$ then $(u, v) \in (U, V)$; see the same references. Under (P2) the following stronger result holds.

Theorem 3.1 *Consider the perturbed game (P2). There exist* $\lim_{\epsilon\to 0} U_\epsilon$ *and* $\lim_{\epsilon\to 0} V_\epsilon$ *and these limits are polytopes. Let* (u_ϵ, v_ϵ) *be an optimal solution for the perturbed game (P2). Let* $\epsilon(l) \to 0$ *be any sequence along which some limits*

$$u' = \lim_{l\to\infty} u_{\epsilon(l)}, \quad v' = \lim_{l\to\infty} v_{\epsilon(l)}$$

exist. Then $u' \in U_{GF}$*, and* $v' \in V_{GF}$*. Therefore,* $\lim_{\epsilon\to 0} U_\epsilon \subseteq U_{GF}$ *and* $\lim_{\epsilon\to 0} V_\epsilon \subseteq V_{GF}$*.*

We remark that Theorem 3.1 and its proof hold for any perturbation that satisfies (3) and $G_\epsilon = G + \epsilon F + o(\epsilon)$.

Proof. Formula (3) implies the convergence of U_ϵ and that the limit of U_ϵ is a convex hull of $\{u_0^k : \ k = 1, \ldots, K\}$. The similar statement is true for player 2.

It follows from (4) that

$$\begin{aligned} \textbf{val } G_\epsilon &= u_0^* G v_0^* + \epsilon(u_0^* G v_1^* + u_1^* G v_0^* + u_0^* F v_0^*) + o(\epsilon) \\ &= \inf_v \left[u_0^* G v + \epsilon(u_1^* G + u_0^* F)v + o(\epsilon)\right] \geq \inf_v \left[uGv + \epsilon uFv + o(\epsilon)\right], \end{aligned} \tag{5}$$

for all policies $u \in U$. Since the last inequality holds for all $\epsilon \in I$,

$$\inf_v u_0^* G v \geq \inf_v uGv,$$

for all policies $u \in U$. This establishes directly the fact that $u_0^* \in U_G$ which also follows from Proposition G.5 in [2]. Similarly $v_0^* \in V_G$.

We thus focus on (5) for the case when $u \in U_G$. In this case (5) implies

$$(u_0^* G v_1^* + u_1^* G v_0^* + u_0^* F v_0^*) + o(1) = \inf_{v \in V_G} [(u_1^* G + u_0^* F)v + o(1)] \geq \inf_{v \in V_G} uFv, \tag{6}$$

for all policies $u \in U$.

We observe that

$$(u_0^* G)_j \geq \textbf{val } G, \qquad j = 1, ..., n.$$

(if for some j this were not true, then the optimal response of player 2 to u^* would yield a value strictly smaller than **val** G, which contradicts the fact that u_0^* is optimal for the matrix game G). Similarly,

$$(G v_0^*)_i \leq \textbf{val } G, \qquad i = 1, ..., m.$$

We further note that

$$v_0^*(j) = 0 \text{ for any } j \text{ for which } (u_0^* G)_j > \textbf{val } G,$$

otherwise, player 1 could achieve more than **val** G against v_0^*; this contradicts the fact that $v_0^* \in V_G$. Moreover, if $v_0^*(j) = 0$ then $v_1^*(j) \geq 0$ since v_ϵ^* is nonnegative. It then follows from an argument similar to the one for $v_0^*(j)$ that $v_1^*(j) = 0$. We conclude that $v_1^*(j) \neq 0$ only if $(u_0^* G)_j =$ **val** G. Since $\sum_j v_1^*(j) = 0$, this implies that

$$u_0^* G v_1^* = 0.$$

Similarly, $u_1^* G v_0^* = 0$. It then follows from (6) that

$$u_0^* F v_0^* \geq \inf_{v \in V_G} uFv$$

for any $u \in U_G$, so that $u \in U_{GF}$. We get similarly $v \in V_{GF}$. ∎

The following corollary follows from the proof of Theorem 3.1. Under assumption (P1) this result was proved by Mills [5].

Corollary 3.1 *Consider the perturbed game (P2). Then*

$$\lim_{\epsilon \to 0} \frac{\textbf{val } G_\epsilon - \textbf{val } G}{\epsilon} = \textbf{val } (GF).$$

A natural question is whether $\lim_{\epsilon\to 0} U_\epsilon = U_{GF}$ and $\lim_{\epsilon\to 0} V_\epsilon = V_{GF}$ when $G_\epsilon = G + \epsilon F$. The following example provides the negative answer.

Example 3.1. Let

$$G = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \qquad F = \begin{pmatrix} 2 & 1 & 2 & 0 \\ 1 & 2 & 2 & 0 \end{pmatrix}.$$

Player 2 has three equivalent policies in game G. It is easy to see that $U_G = \{(0.5, 0.5)^t\}$, **val** $G = 0.5$, and

$$V_G = \{p_1(0.5, 0, 0, 0.5) + p_2(0, 0.5, 0, 0.5) + p_3(0, 0, 0.5, 0.5)|\ p_1 + p_2 + p_3 = 1, p_i \geq 0, i = 1, 2, 3\}.$$

The game GF is equivalent to a 1×3 matrix game with the payoff matrix $(\frac{3}{4}, \frac{3}{4}, 1)$. Therefore,

$$V_{GF} = \{p_1(0.5, 0, 0, 0.5) + p_2(0, 0.5, 0, 0.5)|\ p_1 + p_2 = 1, p_i \geq 0, i = 1, 2\}.$$

For $\epsilon > 0$ we consider a matrix game $G + \epsilon F$. This 2×4 game can be solved explicitly. We have $U_\epsilon = \{(\frac{1-\epsilon}{2+\epsilon}, \frac{1+2\epsilon}{2+\epsilon})^t\}$, **val** $(G + \epsilon F) = \frac{1+2\epsilon}{2+\epsilon}$, and only policies 1 and 4 of player 2 are active. Therefore, policies 2 and 3 of player 2 can be excluded. We delete columns 2 and 3 of matrix G_ϵ and solve the appropriate 2×2 game. From this solution we get

$$V_\epsilon = \{(\frac{1}{2+\epsilon}, 0, 0, \frac{1+\epsilon}{2+\epsilon})\}.$$

We have that $\lim_{\epsilon\to 0} V_\epsilon = \{(0.5, 0, 0, 0.5)\} \neq V_{GF}$. ∎

We have defined lexicographic games for two matrices G and F. However, it is possible to define a lexicographic game for any finite sequence of $m \times n$ matrices $F_1, F_2, \ldots, F_k$. If $k = 1$ then the lexicographic game is F_1 and the sets of optimal solutions for players 1 and 2 are polytopes. If for some $i = 1, \ldots, k-1$, the lexicographic game $F_1F_2 \ldots F_i$ id defined and the sets of optimal solutions for players 1 and 2 are polytopes, the lexicographic game $F_1F_2 \ldots F_iF_{i+1}$ is game the F_{i+1} on these polytopes. Then the set of optimal solutions of this game are polytopes too; see Altman, Feinberg, and Shwartz [1] for details.

Let G_ϵ satisfies (P2). In view of Theorem 3.1, a natural question is whether $\lim_{\epsilon\to 0} U_\epsilon \subseteq U_{GFF_1 \ldots F_L}$. The following example gives the negative answer to this question.

Example 3.2. Let $G_\epsilon = G + \epsilon F + \epsilon^2 F_1$, where matrices G and F have are defined in Example 3.1 and

$$F_1 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

We have from Example 3.1 that $U_{GF} = \{(0.5, 0.5)^t\}$ and V_{GF} is a convex combination of vectors $(0.5, 0, 0, 0.5)$ and $(0, 0.5, 0, 0.5)$. Since the second player minimizes the payoff, we have $V_{GFF_1} = \{(0, 0.5, 0, 0.5)\}$.

Now we solve explicitly the 2 by 4 matrix game G_ϵ. We have that

$$U_\epsilon = \{(\frac{1-\epsilon}{2+\epsilon+\epsilon^2}, \frac{(1+\epsilon)^2}{2+\epsilon+\epsilon^2})\}, \qquad \textbf{val } G_\epsilon = \frac{(1+\epsilon)^2}{2+\epsilon+\epsilon^2}),$$

and only actions 1 and 4 are active for player 2 when ϵ is small. We delete columns 2 and 3 from G_ϵ and solve the (2×2) game. We have

$$V_\epsilon = \{(\frac{1}{2+\epsilon+\epsilon^2}, 0, 0, \frac{1+\epsilon+\epsilon^2}{2+\epsilon+\epsilon^2})\}.$$

Thus,

$$\lim_{\epsilon \to 0} V_\epsilon = \{(0.5, 0, 0, 0.5)\} \neq \{(0, 0.5, 0, 0.5)\} = V_{GFF_1}.$$

4 Extensions

Our first extension is to a perturbation of the form:

- (P3) All entries of G_ϵ are rational functions of ϵ (the division of two polynomials of ϵ with real coefficients). We assume that G_ϵ has no poles in 0.

It then follows that G_ϵ can be expressed as

$$G_\epsilon = G + \epsilon F + \sum_{l=1}^{\infty} \epsilon^{l+1} F_l, \tag{7}$$

where G, F and $F_l, l = 1, ...,$ are real matrices of size $m \times n$.

It follows from [4] that the set of basic coordinates that are optimal for the LP that correspond to the matrix game is fixed for some interval $I' = (0, \epsilon']$. A policy u_ϵ corresponding to any fixed set of basic coordinates can also be expressed as a rational function in ϵ; see Jeroslow [4].

Note that G_ϵ is uniformly bounded on I'. Hence both u_ϵ, v_ϵ corresponding to any basis, as well as the value of the game can be expressed as Taylor series in ϵ. We may thus repeat the steps of Sections 2, 3 and obtain the results of Theorem 3.1 and of Corollary 3.1.

Next, we extend the spaces of policies to infinite ones. Let U and V be some spaces and consider the function $G_\epsilon : U \times V \to \mathbb{R}$. Assume that G_ϵ has a value for all ϵ sufficiently small. Consider the following form of perturbation:

Assume that there exists optimal policies of the form

$$u_\epsilon^* = u_0 + \epsilon u_1 + o_1(\epsilon), \qquad v_\epsilon^* = v_0 + \epsilon v_1 + o_2(\epsilon).$$

The statements of Theorem 3.1 and of Corollary 3.1 still hold. The proofs extend with minor modifications.

Our second extension is to a perturbation of the form:

- (P4) $G_\epsilon = G + \epsilon F + \epsilon^{\alpha_1} F_1 + \ldots + \epsilon^{\alpha_L} F_L$, where F, G and $F_l, l = 1, \ldots, L$ are real matrices of size $m \times n$ and $1 < \alpha_1 < \ldots \alpha_L < \infty$.

We say that a function $P(\epsilon)$ is an irrational polynomial if for some integer L

$$P(\epsilon) = a + b\epsilon + c_1\epsilon^{\alpha_1} + \ldots c_L\epsilon^{\alpha_L}, \tag{8}$$

where $\alpha_l > 1$, $l = 1, \ldots, L$. Without loss of generality we can assume that $\alpha_1 < \alpha_2 < \ldots < \alpha_L$.

We observe that if an irrational polynomial P is not identical to 0 then $P(\epsilon) = \epsilon^\alpha(d + o(1))$, where $\alpha \geq 0$ and $d \neq 0$. Therefore,

$$P(\epsilon) \neq 0, \qquad \epsilon \in]0, \epsilon_0], \tag{9}$$

for some positive ϵ_0.

The sum and the products of two irrational polynomials are irrational polynomials. Let P and Q be irrational polynomials and assume that Q is not identical to 0. We consider a ratio P/Q. Let $\mathcal{M}$ be the set of all ratios of irrational polynomials. If $R_1, R_2 \in \mathcal{M}$ then $R_1 + R_2 \in \mathcal{M}$ $R_1 R_2 \in \mathcal{M}$, and, if R_2 is not identical to 0, $\frac{R_1}{R_2} \in \mathcal{M}$.

We consider a perturbed game with matrix G_ϵ defined in (P4). Without loss of generality we assume that **val** $G > 0$. For game G_ϵ we consider a standard LP that computes an optimal policy for player 1; see for example [2]. The optimal variables of this LP are equal to $u_\epsilon(\textbf{val}\ G(\epsilon))^{-1}$, where u_ϵ is an optimal policy for player 1 in game G_ϵ. We apply Jeroslow's [4] approach to these LPs. Each feasible basic solution x of an LP can be represented as $x = (x_b, x_N)$, where x_b is a vector of basic variables. Therefore, the set of all coordinates of x is partitioned into two subsets: a subset of basic coordinates and a subset of nonbasic coordinates. This partition defines vector x.

Let x_ϵ be the feasible basic solution of the LP for game G_ϵ when a set of basic coordinates is fixed. Then the elements of vector x_ϵ belong to $\mathcal{M}$ and $f_\epsilon(x_\epsilon) \in \mathcal{M}$, where f_ϵ is the objective function of the LP for game G_ϵ. If x_ϵ is an optimal solution of this LP then $f_\epsilon(x_\epsilon) = (\mathbf{val}\ G_\epsilon)^{-1}$.

We also observe that (9) implies that for any function R in $\mathcal{M}$ there is $\epsilon_0 > 0$ such that $R(\epsilon) > 0$ for all $\epsilon \in]0, \epsilon_0]$. In particular, if we fix the set of basic coordinates, it is true for $R(\epsilon) = f_\epsilon(x_\epsilon)$. We also observe that the difference of two functions in $\mathcal{M}$ belongs to $\mathcal{M}$. The set of all possible partitions into basis and nonbasic coordinates is finite. Therefore, there is a nonempty interval $I =]0, \epsilon_0[$ and a finite number, say K, of sets of basic coordinates such that for all $\epsilon \in I$ each of these sets define optimal solutions for these LPs for all $\epsilon \in I$. We fix one of these sets of basic coordinates. Let x_ϵ be the appropriate basic solution of the LP. Then $\mathbf{val}\ G_\epsilon = 1/f(x_\epsilon) \in \mathcal{M}$ and the elements of the vector x_ϵ are in $\mathcal{M}$. For the corresponding basic optimal policy u_ϵ of the game G_ϵ we have that $u_\epsilon = x_\epsilon/\mathbf{val}\ G_\epsilon$ and therefore $u_\epsilon \in \mathcal{M}$. In addition, u_ϵ are probability distributions and therefore they are bounded. Therefore $u_\epsilon = u_0 + \epsilon u_1 + o(\epsilon)$, $\epsilon \in I$, for some u_i, $i = 1, 2$. Each of these K sets of basic coordinates defines optimal basic policies u_ϵ^k for the games G_ϵ, $\epsilon \in I$, $k = 1, \ldots, K$. Therefore u_ϵ an optimal basic policy for player 1 if and only if (3) holds for some $k = 1, \ldots, K$. This implies that Theorem 3.1 and corollary 3.1 hold for the perturbed game (P4). We also remark that Theorem 3.1 holds for a more general perturbation $G_\epsilon = G + \epsilon^{\alpha_0} F + \sum_{l=1}^{L} \epsilon^{\alpha_L} F_l$, where $0 < \alpha_0 < \alpha_1 \ldots < \alpha_L < \infty$. In this case, we can substitute ϵ^{α_0} with a new variable and apply Theorem 3.1 for (P4).

5 Application to perturbed stochastic games

Consider a perturbed Markov game with a finite state space $\mathbf{X}$ and finite *action spaces* $\mathbf{A}$ and $\mathbf{B}$. We assume without loss of generality that $\mathbf{A}$ and $\mathbf{B}$ are the same for all states. We assume that the transitions are controlled only by player 2, i.e. the probability to go from state x to y in one step is only a function p^ϵ_{xby} of the action b of player 2. It is given by the transition probability

$$p^\epsilon_{xby} := \sum_{l=0}^{L} \epsilon^l p_{xby}(l), \tag{10}$$

where L is some integer. We allow for subprobability measures, i.e. $\sum_{y \in \mathbf{X}} p^\epsilon_{xby} < 1$.

A behavioral policy u in the *policy space* U is described as $u = \{u_1, u_2, ...\}$, where the decision rule u_t, applied at time epoch t, is a probability measure over $\mathbf{A}$ conditioned on the whole history

of actions and states prior to t, as well as on the state at time t. A behavioral policy for which all measures u_t are concentrated on a single action are called pure behavioral policies.

We shall consider the infinite horizon case. A mixed stationary policy u is a probability measure of the set of pure stationary policies.

Given an initial distribution β on $\mathbf{X}$, each policy pair u induces a probability measure denoted by $P^{u,v}_{\beta,\epsilon}$ on the space of sample paths of states and actions (which serves as the canonical sample space Ω). The corresponding expectation operator is denoted by $E^{u,v}_{\beta,\epsilon}$. On this probability space the state and action processes, x_t, a_t, b_t, $t = 1, 2, ..., N$ are defined, as well as the history process $h_t = (x_1, a_1, b_1, ..., x_t)$.

Let $r_l : \mathbf{X} \times \mathbf{A} \times \mathbf{B} \to \mathbb{R}$, be (real valued) reward functions, $l = 0, \ldots, L$, and consider the total expected reward function:

$$R^{\epsilon}_{\beta}(u, v) = E^{u,v}_{\beta,\epsilon} \sum_{s=1}^{N+1} r^{\epsilon}(x_s, a_s, b_s), \tag{11}$$

where $\epsilon > 0$ and

$$r^{\epsilon}(x, a, b) = \sum_{l=0}^{L} \epsilon^l r_l(x, a, b). \tag{12}$$

The set of pure stationary policies is finite. Thus when considering the game over the set of mixed pure stationary policies, we are in the framework of (finite) matrix games described in the Section 4. We note that this game has indeed a saddle point within the mixed pure stationary policies, as well as among the set of stationary randmoized policies, and there exists a simple equivalence between these two classes of policies (see [6] Thm. 3.1).

Clearly the game is of the form of (P2), so we may apply the main results of Theorem 3.1 and Corollary 3.1:

Theorem 5.1 *Consider the perturbed stochastic game. Then*
(i) The values of R^{ϵ}_{β} converge to that of R^{0}_{β}, and

$$\lim_{\epsilon \to 0} \frac{\mathbf{val}\ R^{\epsilon}_{\beta} - \mathbf{val}\ R^{0}_{\beta}}{\epsilon} = \mathbf{val}\ (GF).$$

(ii) Let (u_ϵ, v_ϵ) be optimal (randomized stationary or mixed pure stationary) for the ϵ stochastic game. Let $\epsilon(l) \to 0$ be any sequence along which some limits

$$u' = \lim_{l \to \infty} u_{\epsilon(l)}, \quad v' = \lim_{l \to \infty} v_{\epsilon(l)}$$

exist. Then $u' \in U_{GF}$, *and* $v' \in V_{GF}$.

Note: it is only for the mixed policies that we have the representation of the stochastic game as a matrix game. In order to obtain the results for the behavioral policies, we have to use the equivalence between the behavioral and mixed policies, as well as the fact that we can select a continuous mapping between them. If $\{u_{\epsilon(l)}\}_l$ and u are behavioral, then the corresponding mixed policies also converge. Since the limit of the mixed policies is in U_{GF}, it follows that so is u'.

Theorem 5.1 extends results in [7] where it was shown that $\lim_{\epsilon\to 0} \mathbf{val}\ R^\epsilon_\beta - \mathbf{val}\ R^0_\beta \epsilon = \mathbf{val}\ A$, and that $u' \in U_G$, and $v' \in V_G$. We also remark that Theorem 5.1 holds if instead of perturbations of type (P2) of the transition probabilities and reward functions we consider a more general perturbation of type (P4) of these objects.

Consider finally the more general case where the transition probabilities p^ϵ_{xaby} may depend on the acations of both players, and consider a finite horizon of N steps. One can now study the convergence of perturbed stochastic games with finite horizon through the dynamic programming equation

$$v_t(x) = \mathbf{val}_{ab}\left[r(x,a,b) + \sum_y p^\epsilon_{xaby} v_{t-1}(y)\right], \quad t > 0, \tag{13}$$

$$v_0(x) = \mathbf{val}_{ab} r_0(x,a,b).$$

A saddle point (u, v) can be obtained for both players in Markov policies by using at time $N - t$ the argument that achieves the value in the tth equation above (as function of x). At each stage t we are faced with a perturbed matrix game. Note that the value $v_0(x)$ may be expressed as a an infinite sum $v_0(x) = \sum_{s=0}^\infty \epsilon^s v_0^s(x)$ and need not be polynomial, so we are in the framework of Section 4 for the next stage. However, we can show by induction that the perturbation at each stage is of the form (P3) presented in Section 4, so the results there hold.

Indeed, since $r_0(x)$ is polynomial in ϵ, its value $v_0(x)$ is analytic functions in ϵ in a neighborhood of 0, as it follows from the argument above (3) (note that we are in the framework of Problem (P2)). Now, making the inductive hypothesis that $v_{t-1}(x)$ is analytic in ϵ in a neighborhood of 0 for all x, it follows that the term in square brackets in (13) is analytic in ϵ in some neighbrohood of 0. We are thus in the framework of Problem (P3), and by the argument below (7) we conclude that $v_t(x)$ is analytic in ϵ in a neighborhood of 0 for all x. Hence for any t there is a neighborhood of 0 such that $v_t(x)$ is analytic in ϵ in that neighborhood.

6 Application to repeated zero-sum games with weighted discounted payoffs

Let us consider a zero-sum repeated game with finite action sets $\mathbf{A}$ and $\mathbf{B}$ for players 1 and 2 respectively. There are L payoff matrices $F_1, F_2, \ldots, F_L$ and there are L discount factors $\beta_1, \beta_2, \ldots, \beta_L$ where $1 > \beta_1 > \beta_2 \ldots \beta_L > 0$. The payoff function is

$$G(u,v) = E^{u,v} \sum_{l=1}^{L} \sum_{n=0}^{\infty} \beta_l^n F_l(a_n, b_n).$$

Finding an optimal randomized Markov policy for this game is equivalent to finding optimal policies for the sequence of games

$$G_{\epsilon_n} = F_1 + \epsilon_n F_2 + \sum_{l=3}^{L} (\epsilon_n)^{\alpha_l} F_l$$

with $\epsilon_n = (\frac{\beta_2}{\beta_1})^n$, $n = 0, 1, \ldots$, and $\alpha_l = \log_{\frac{\beta_2}{\beta_1}} \frac{\beta_l}{\beta_1}$.

Filar and Vrieze [2] proved the existence of ϵ-optimal ultimately stationary policies for zero-sum stochastic games with weighted discounted payoffs and with finite sets of states and actions. Altman, Feinberg, and Shwartz [1] proved the existence of optimal policies, which are Markov and ultimately stationary, for such games with perfect information. Example 6.1 in [1] shows that optimal ultimately stationary policies may not exist for general Markov games when there is no perfect information assumption. The statement in the previous paragraph implies that, for repeated games with several different discount factors, all limits of optimal actions as the time parameter tends to ∞ are optimal for the lexicographic game defined by payoff matrices corresponding to two largest discount factors. We conjecture that this result holds for stochastic games with weighted discounted payoffs and with finite state and action sets; see Altman, Feinberg, and Shwartz [1].

We remark that for stochastic games with perfect information the sets of optimal solutions coincide from some step N onward with the set of optimal solutions for the lexicographic game; Altman, Feinberg, and Shwartz [1]. Example 3.1 implies that, without the perfect information assumption, the limiting sets of optimal solutions can be strictly smaller than the sets of appropriate solutions for the lexicographic game even in the case of two discount factors. Example 3.2 demonstrates that, in the case of three or more different discount factors, the limiting sets of optimal solutions may have no common points with the sets of optimal solutions of the lexicographic game defined by the payoff matrices corresponding to three largest discount factors.

Acknowledgment. The research of the second author was partially supported by NSF Grant DMI-9500746.

References

[1] E. Altman, E.A. Feinberg, and A. Shwartz, "Weighted discounted stochastic games with perfect information," Proc. of the 7th International Symposium on Dynamic Games and Applications, Vol. 1, 18-31, 1996 (eds. J.A. Filar, V. Gaitsgory, F. Imado). To appear in *Annals of the International Society of Dynamic Games.*

[2] J.A. Filar and O.J. Vrieze, "Weighted reward criteria in competitive Markov decision processes," *ZOR*, **36**, 343- 358, 1992

[3] J. Filar and K. Vrieze, *Competitive Markov Decision Processes,* Springer, NY, 1997.

[4] R. G. Jeroslow, "Asymptotic linear programming" *Oper. Res.*, **21**, 1128-1141, 1973.

[5] H.D. Mills, "Marginal values of matrix games and linear programs," *Ann. Math. Stud.* **38**, 183-193, 1956.

[6] A. S. Nowak and T. E. S. Raghavan, "A finite step algorithm via a bimatrix game to a single controller non-zero sum stochastic game", *Mathematical Programming* 59, 249-259, 1993.

[7] M. Tidball and E. Altman, "Approximations in dynamic zero-sum games, I," *SIAM J. Control and Optimization*, **34**, 311-328, 1996.

[8] S.H. Tijs and O.J. Vrieze, "Perturbation Theory for Games in Normal Form and Stochastic Games," *JOTA*, **30**, 549-567, 1980.