A Stochastic Minimax Optimal Control Problem on Markov Chains with Infinite Horizon

Silvia C. Di Marco, Roberto L.V. González

INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

A stochastic minimax optimal control problem on Markov chains with infinite horizon

Silvia C. Di Marco and Roberto L.V. González

N ° 2946

July 1996

THÈME 4

ISSN 0249-6399

INRIA
ROCQUENCOURT

A stochastic minimax optimal control problem on Markov chains with infinite horizon

Silvia C. Di Marco * and Roberto L.V. González *

Thème 4 — Simulation et optimisation
de systèmes complexes
Projet Promath

Rapport de recherche n ° 2946 — July 1996 — 15 pages

Abstract: We consider here a stochastic discrete minimax optimal control problem defined on a finite state Markov chain in the case of infinite horizon. We prove the existence of an optimal control in terms of a generalized feedback policies. We characterize the optimal cost function and we present iterative methods to compute it numerically.

Key-words: stochastic control, optimal control, minimax optimization, Markov chain, value iteration, policy iteration, feedback.

(Résumé : tsvp)

*CONICET – Inst. Beppo Levi, Dpto. Matemática, Fac. Cs. Ex., Ing. y Agr., Universidad Nacional de Rosario, Rosario, Argentine. This paper is included in the activities developed in the frame of the Cooperation Projet INRIA–Instituto de Matemática Beppo Levi, Coordinators of the projet: E. Rofman–R. González

Unité de recherche INRIA Rocquencourt
Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
Téléphone : (33 1) 39 63 55 11 – Télécopie : (33 1) 39 63 53 30

Un problème de contrôle optimal stochastique de type minimax sur chaînes de Markov avec horizon infini

Résumé : On considère ici un problème stochastique de contrôle optimal de type minimax avec horizon infini. Le système analysé est une chaîne de Markov finie. On étudie l'existence d'une solution définie comme une loi de contrôle en boucle fermée généralisée et on présente deux méthodes itératives pour le calcul approché de la fonction coût optimal.

Mots-clé : contrôle stochastique, contrôle optimal, chaînes de Markov, itération sur les valeurs, itération sur les politiques, contrôle en boucle fermée.

AMS Classification: 93C55, 93E20, 93E25

Contents

1 The stochastic optimization problem

We study a stochastic optimal control problem of minimax type defined on a finite controlled Markov chain.

In the following sections we present the elements of the control problem, we characterize the associated value function and we present two iterative procedures to compute it numerically. Using a suitable auxiliary variable, we define an optimal (generalized) feedback control.

Description of the problem

We consider a discrete time stochastic process Y which at each discrete time $n \in \mathbb{N}_\circ$ is in one of the N states i belonging to $\Omega = \{i : i = 1, \dots N\}$. Moreover, we assume that, at each n an action a can be chosen in a finite set A, with $|A| = N_A$, (controlled finite Markov chain (see [14])). The transition probabilities of the process Y are given by

$$P\left\{Y(n+1) = j \;/\; Y(n) = i, \alpha(n) = a\right\} = p_{i,j}(a)\,. \tag{1}$$

So, $p_{i,j}(a)$ denotes the transition probability from the state i to the state j when the control a is used at i. We denote by $\alpha(\cdot)$ the control policy applied to the chain.

We denote by $\mathcal{A}$ the set of non-anticipative or progressively measurable control policies (see [7, 8, 9, 12]). We say that a non-anticipative policy α is stationary, in case this policy is defined by a map $\Omega \mapsto A$. Since there exists a bijection between the stationary policies (of a N states problem) and the set A^N, we consider $\alpha \in A^N$ when α is stationary.

We are interested in minimizing, with respect to the non-anticipative control policies, the expectation of the maximum of a function f on the chain along the time ($f : \Omega \times A \mapsto \mathbb{R}_\circ^+$). So, the functional to be minimized is given by

$$J(i, \alpha(\cdot)) = \mathbb{E}_\alpha \left\{ \max_{n \in \mathbb{N}_\circ} f(Y_\alpha(n), \alpha(n)) \;/\; Y_\alpha(0) = i \right\}, \tag{2}$$

where $\mathbb{E}_\alpha$ denotes the expectation of the process when the policy α has been chosen. The maximum of the right side of (2) is well defined because it is computed on a finite number of states and actions.

The optimal cost function is given by

$$u(i) = \inf \{J(i, \alpha(\cdot)) : \alpha(\cdot) \in \mathcal{A}\}. \tag{3}$$

Our objective is to find an optimal policy that realizes u and, if it is possible, to find an optimal feedback control policy.

Remark 1.1 Even when the chosen policy is stationary, the stochastic process

$$\zeta_n = \left\{ \max_{\nu \leq n} f(Y_\alpha(\nu), \alpha(\nu)) \right\}$$

is not a Markov one. At time n, it depends not only on the state $Y_\alpha(n)$ but also on the previous ones. To apply dynamical programming techniques to the problem (3), we transform it to a Markov process through the introduction of an auxiliary variable.

2 Conversion to a Markov problem

2.1 Introduction of an auxiliary state

We consider $i_\circ$ the initial state of the process Y_α, where $\alpha(\cdot)$ is an arbitrary non-anticipative policy. The auxiliary process Z is given by a sequence of random variables z_n with values in F, which evolves according to the equation

$$z_{n+1} = z_n + (f(Y_\alpha(n), \alpha(n)) - z_n)^+ = \max \{f(Y_\alpha(n), \alpha(n)),\ z_n\},$$

with $z_\circ \in F$. F is the set of admissible states of the process Z,

$$F = \{0\} \bigcup \{f(i, a) : i \in \Omega, a \in A\}, \quad |F| = N \times N_A + 1. \tag{4}$$

Remark 2.1 It is easy to see that

$$z_n = \max \left\{ z_\circ, \max_{k=1,\dots,n} f(Y_\alpha(k), \alpha(k)) \right\}. \tag{5}$$

In consequence it is immediate to check that the variable z_n "stores" the maximum of f up to time n, on the trajectory Y associated to the employed policy α.

For each $a \in A$, we define the operator $T_a : F \times \Omega \to F$ such that

$$T_a(p,i) = p + (f(i,a) - p)^+ . \tag{6}$$

We consider now the extended stochastic process (Z,Y) whose state at time $n \in \mathbb{N}_\circ$ is given by

$$(z_n, Y(n)) \in F \times \Omega. \tag{7}$$

The transition probabilities of the stochastic process (Z,Y) are

$$P\{\,(z_{n+1}, Y_\alpha(n+1)) = (q,j) \;/\; (z_n, Y_\alpha(n)) = (p,i), \alpha(n) = a\,\} = p_{i,j}(a)\,\mathcal{X}_{q=T_a(p,i)}\,.$$

In consequence, (Z,Y) is a Markov chain.

Definition 2.1 Let α be a non-anticipative policy. We say that α is a *generalized feedback* if it is given by a monovalued function $\alpha : F \times \Omega \mapsto A$.

Definition 2.2 We give the following definition of recurrent state for the chain (Z,Y); (see [1, 13, 14]). We say that (p,i) is a *recurrent state* under the action of the feedback control policy α, if there exists $C_p(i) \subset \Omega$, such that $\forall\, j \in C_p(i)$,

- $T_a(p,j) = p$;
- there exists $m, k \in \mathbb{N}$ such that $p^m_{i,j} > 0$ and $p^k_{j,i} > 0$;
- $p_{j,l} = 0$, $\forall\, l \notin C_p(i)$.

2.2 The auxiliary problem

We consider in the chain (Z,Y) the stochastic optimal control problem with infinite horizon, where the cost functional is defined by

$$v_\alpha(p,i) = \mathbb{E}_\alpha \left\{ \sum_{n=0}^{\infty} (f(Y_\alpha(n), \alpha(n)) - z_n)^+ \;/\; Y_\alpha(0) = i, z_\circ = p \right\}, \tag{8}$$

and the associated optimal cost function is

$$v(p,i) = \min\{v_\alpha(p,i) : \alpha(\cdot) \in \mathcal{A}\}\,. \tag{9}$$

Proposition 2.1 *The following properties hold*

1. $v(p,i) \geq 0,\ \forall\ (p,i) \in F \times \Omega$.
2. *Let* $\tilde{F} = \max\{f(i,a) : i \in \Omega, a \in A\}$, *then* $v(\tilde{F},i) = 0,\ \forall\ i \in \Omega$.

Proof. The first property comes from the definition of v. To prove the second one, from the definition of the process Z, we have that $\forall\, n \in \mathbb{N}_\circ,\ z_n = \tilde{F}$. Then, every term in the sum (8) is equal to zero.

□

Proposition 2.2 *Let g be a generalized feedback control policy and (p,i) be a recurrent state of the chain determined by g. Then $v_g(p,i) = 0$.*

Proof. Let $C_p(i)$ denote the recurrent class to which (p,i) belongs. Then,

$$T_g(p,j) = p + (f(j,g(p,j)) - p)^+ = p,\ \forall\ (p,j) \in C_p(i),$$

which implies that

$$(f(j,g(p,j)) - p)^+ = 0,\ \forall\ (p,j) \in C_p(i).$$

Since $\{Y_g(n) : n \in \mathbb{N}\}$ is a stochastic process which remains forever in $C_p(i)$, replacing in (8) we have that

$$v_g(p,i) = \mathbb{E}_g\left\{\sum_{n=0}^{\infty} (f(Y_g(n), g(z_n, Y_g(n))) - z_n)^+ \ / \ Y_g(0) = i, z_\circ = p\right\} = 0.$$

□

Now we give a relation concerning the minimax problem (3) and the stochastic control problem with cumulative cost (9).

Proposition 2.3 $u(i) = v(0,i) = \max\{v(p,i) : p \in F\}$.

Proof. It is easy to check from its definition that $v(p,i)$ is non-increasing with respect to $p \in F$, so

$$v(0,i) = \max\{v(p,i) : p \geq 0\}.$$

Besides,

$$\begin{aligned} v(0,i) &= \min_{\alpha \in \mathcal{A}}\left\{\mathbb{E}_\alpha\left\{\sum_{n=0}^{\infty} (f(Y_\alpha(n), \alpha(n)) - z_n)^+ \ / \ Y_\alpha(0) = i, z_\circ = 0\right\}\right\} \\ &= \min_{\alpha \in \mathcal{A}}\left\{\mathbb{E}_\alpha\left\{\max_{n \in \mathbb{N}_\circ} f(Y_\alpha(n), \alpha(n)) \ / \ Y_\alpha(0) = i\right\}\right\} \\ &= u(i). \end{aligned}$$

□

2.3 Dynamical programming

2.3.1 The dynamical programming principle

The stochastic control problem (9) has been widely studied in the literature of this subject (see e.g. [14, 15]). We refer to chapter 6 of [14] for the proofs of the following Proposition and Corollary.

Proposition 2.4 *The function v satisfies the following dynamical programming equation:*

$$v(p,i) = \min_{a \in A} \left\{ (f(i,a) - p)^+ + \mathbb{E}_a \left(v(T_a(p,i), Y_\alpha(1)) \right) \right\}. \tag{10}$$

Corollary 2.1 *Let π be a generalized feedback policy such that $\pi(p,i)$ produces the minimum of the right side of* (10)*. Then, π is an optimal policy.*

Corollary 2.2 *Let π be an optimal generalized feedback policy. For any (p,i) recurrent state of the chain associated to π, it results $v(p,i) = 0$.*

Proof. It follows from the Proposition 2.2 and the Corollary 2.1.

□

2.3.2 The operator M and its properties

Definition 2.3 We define the operator $M : \mathbb{R}^{(N \times N_A + 1) \times N} \mapsto \mathbb{R}^{(N \times N_A + 1) \times N}$ such that

$$(Mw)(p,i) = \min_{a \in A} \left\{ (f(i,a) - p)^+ + \mathbb{E}_a \left(w(T_a(p,i), Y_\alpha(1)) \right) \right\}. \tag{11}$$

Remark 2.2 From the dynamical programming principle proved in the Proposition 2.4, we have that

$$M\, v = v. \tag{12}$$

Definition 2.4 For each generalized feedback control policy g, we define the operator
$M_g : \mathbb{R}^{(N \times N_A + 1) \times N} \mapsto \mathbb{R}^{(N \times N_A + 1) \times N}$ such that

$$(M_g\, w)\,(p,i) = (f(i, g(p,i)) - p)^+ + \mathbb{E}_g \left(w(T_{g(p,i)}(p,i), Y_g(1)) \right). \tag{13}$$

Remark 2.3 If g is a generalized feedback, it is easy to see that v_g is a fixed point of M_g .

Proposition 2.5 *Let g be a generalized feedback. Then, the following properties hold:*

1. *M_g and M are monotone.*
2. $\left(M_g^{(n)}0\right)\ (p,i) \to v_g(p,i)$ *when* $n \to \infty$.

Proof.

1. The monotony of operators M_g and M results from the same property for the expectation.
2. It is easy to prove by induction that $\forall\, n \in \mathbb{N}$

$$\left(M_g^{(n)}0\right)\ (p,i) = \mathbb{E}_g\left\{\sum_{k=0}^{n}\left(f(Y_g(k),g(z(k),k)) - z_k\right)^+ \ / \ Y_g(0) = i, z_\circ = p\right\}. \tag{14}$$

Passing to the limit, by Lebesgue's Theorem, it is valid that

$$\begin{aligned}
&\lim_{n\to\infty}\left(M_g^{(n)}0\right)\ (p,i)\\
&= \lim_{n\to\infty}\mathbb{E}_g\left\{\sum_{k=0}^{n}\left(f(Y_g(k),g(z(k),k)) - z_k\right)^+ \ / \ Y_g(0) = i, z_\circ = p\right\}\\
&= \mathbb{E}_g\left\{\lim_{n\to\infty}\sum_{k=0}^{n}\left(f(Y_g(k),g(z(k),k)) - z_k\right)^+ \ / \ Y_g(0) = i, z_\circ = p\right\}\\
&= \mathbb{E}_g\left\{\sum_{k=0}^{\infty}\left(f(Y_g(k),g(z(k),k)) - z_k\right)^+ \ / \ Y_g(0) = i, z_\circ = p\right\} = v_g(p,i).
\end{aligned}$$

□

2.3.3 Characterization of the optimal cost v

Definition 2.5 In relation to the equation (12), we define the associated sets of supersolutions and subsolutions:

$$S = \{s : F \times \Omega \mapsto \mathbb{R}_\circ^+ \ / \ M\, s \le s\}\,,$$
$$W = \{w : F \times \Omega \mapsto \mathbb{R} \ / \ M\, w \ge w\}\,.$$

It is easy to see that $v \in S \bigcap W$.

In the following theorems the value function v is characterized as the minimum element of the set of supersolutions, and as the limit of a particular sequence of subsolutions.

Theorem 2.1 $v = \min\{s : s \in S\}$.

Proof. Let $s \in S$, $(p,i) \in F \times \Omega$. Let π be a generalized feedback that realizes the maximum appearing in the definition of $(M\,s)(p,i)$, i.e.

$$(M\,s)(p,i) = (M_\pi\,s)\,(p,i)\,.$$

By definition, it is valid that $s \geq 0$. From Proposition 2.5, it results $M_\pi s \geq M_\pi 0 \geq 0$. Then, $\forall n \in \mathbb{N}$

$$s(p,i) \geq (M_\pi^n 0)\,(p,i)\,.$$

Again from Proposition 2.5, it follows that

$$(M_\pi^n 0)\,(p,i) \to v_\pi(p,i)\,.$$

Consequently, $s(p,i) \geq v_\pi(p,i) \geq v(p,i)$. Since this property is valid $\forall\ s \in S$, it also holds for $\underline{s} = \min\{s : s \in S\}$. Then,

$$\underline{s}(p,i) \geq v(p,i)\,.$$

□

Now we can prove the following theorem.

Theorem 2.2 $\left(M^{(n)}0\right)\,(p,i) \to v(p,i)$ *when* $n \to \infty$.

Proof. From the property 1 of the Proposition 2.1, $v(p,i) \geq 0$, $\forall\ (p,i) \in F \times \Omega$. Moreover, $M\,v = v$ and since $M\,0 \geq 0$, from the monotony of the operator M, we have

$$0 \leq M^{(n)}0 \leq M^{(n+1)}0 \leq v,\ \forall\ n \in \mathbb{N}.$$

In consequence, $\exists\ \underline{v}$ such that

$$0 \leq M^{(n)}0 \to \underline{v} \leq v. \tag{15}$$

From (15) and the continuity of the operator M it follows that $M\underline{v} = \underline{v}$, that implies $\underline{v} \in S$. Since v is the minimum supersolution, we have that $v \leq \underline{v}$. Then, $\underline{v} = v$ and therefore

$$\lim_{n\to\infty} M^{(n)}0 = v.$$

□

3 Iterative computation of the function u

3.1 Value iteration

From the previous theorem we have that the optimal cost can be computed as the limit of the sequence of subsolutions $M^\nu 0$. This property suggests the following iterative scheme.

Algorithm 1

Step 0: $w^\circ = 0$, $\nu = 0$.

Step 1: Compute $w^{\nu+1} = Mw^\nu$.

Step 2: If $w^{\nu+1} = w^\nu$, stop.

Step 3: Set $\nu = \nu + 1$ and go to Step 1.

From Theorem 2.2, the sequence w^ν generated by the algorithm converges to the solution v. The convergence may be dismally slow as it is shown in the following example.

Example 3.1 We consider a chain with N nodes, with $|A| = 1$ and the instantaneous cost f given by

$$f_i = \begin{cases} 0 & i = 1, \dots, N-1, \\ 1 & i = N. \end{cases} \tag{16}$$

The transition probabilities are given by

$$p_{i,j} = \begin{cases} 1-2\rho & j = i, \\ \rho & j = i-1, \\ \rho & j = i+1, \\ 0 & \text{otherwise}, \end{cases} \tag{17}$$

where the equality considered in (17) is the module N equality.

The error $v - w^\nu$ converges to 0 verifying in addition the relation

$$\lim_{\nu\to\infty} \frac{\|v - w^\nu\|}{\lambda^\nu} = C > 0, \tag{18}$$

where $\lambda = 1 - 4\,\rho\,\sin^2(\pi/2N)$. Then, $\lambda \to 1$ when $\rho \to 0$. From (18) we see that the convergence is very slow when $\rho \sim 0$.

3.2 Policy iteration

To avoid the slowness of the convergence of the value iteration algorithm, we try to apply the Newton method. Since the fixed point problem $w = M\,w$ is not a contraction, we do not use the Newton method directly (which is known as the policy iteration method in this type of problems). Instead, we apply it on a sequence of perturbed problems that converges to the original problem and whose associated fixed point operators are contractions.

We consider the following discounted optimal control problem, with λ the discount coefficient, $0 < \lambda < 1$.

$$V_\lambda(p,i) = \min\left\{V_{\pi,\lambda}(p,i) : \pi(\cdot) \in \mathcal{A}\right\}, \tag{19}$$

where

$$V_{\pi,\lambda}(p,i) = \mathbb{E}_\pi\left(\sum_{n=0}^{\infty} \lambda^n \left(f(Y_\pi(n),\pi(n)) - z_n\right)^+ \;/\; Y_\pi(0) = i, z_o = p\right). \tag{20}$$

It is easy to see that V_λ verifies the following dynamical programming equation

$$V_\lambda(p,i) = \min_{a\in A}\left\{(f(i,a) - p)^+ + \lambda \sum_{j=1}^{N} p_{i,j}(a)\;\cdot\; V_\lambda(T_a(p,i),j)\right\}. \tag{21}$$

Definition 3.1 Given $\lambda \in (0,1)$ we define $Q_\lambda : \mathbb{R}^{(N\times N_A+1)\times N} \mapsto \mathbb{R}^{(N\times N_A+1)\times N}$ such that

$$(Q_\lambda\, w)(p,i) = \min_{a\in A}\left\{(f(i,a) - p)^+ + \lambda \sum_{j\in\Omega} p_{i,j}(a)\;\cdot\; w(T_a(p,i),j)\right\}. \tag{22}$$

Remark 3.1 The operator Q_λ is a contraction and V_λ is its unique fixed point. Besides, $V_\lambda \to v$, when $\lambda \to 1$.

Algorithm 2

Step 0: Let $\{\lambda_\nu\}_{\nu\in\mathbb{N}}$ be a increasing positive sequence, which converges to 1,

$\varepsilon > 0$, $\nu = 1$.

Step 1: Compute $V_{\lambda_\nu} = Q_{\lambda_\nu} V_{\lambda_\nu}$.

Find a stationary policy $\hat{\pi}$ whose cost is V_{λ_ν} .

Step 2: Solve the linear system

$$\left|\begin{array}{ll} w(p,i) = 0 & \forall\ (p,i) \text{ recurrent state,} \\ w(p,i) = (M_{\hat{\pi}} w)\,(p,i) & \forall\ (p,i) \text{ non recurrent state.} \end{array}\right.$$

Step 3: If $w - V_{\lambda_\nu} \leq \varepsilon$, stop.

Set $\nu = \nu + 1$ and go to Step 1.

Remark 3.2 The problem $V_{\lambda_\nu} = Q_{\lambda_\nu} V_{\lambda_\nu}$ at Step 1 can be solved using the algorithms presented in [10]. They converge to the solution of the discounted problem in a finite number of steps.

Convergence of the algorithm

Remark 3.3 Let us see that the stopping rule at Step 3 is verified after a finite number of iterations. In effect, let us assume that $\forall\ \nu \in \mathbb{N}$ we have that

$$w - V_{\lambda_\nu} > \varepsilon\,. \tag{23}$$

It is easy to see that there exists $K > 0$ – dependent only on the data of the problem but independent on the policy used in (24) – such that

$$w - V_{\lambda_\nu} \leq K(1 - \lambda_\nu)\,. \tag{24}$$

Then, from (23) and (24), it results that $\forall\,\nu \in \mathbb{N}$,

$$\varepsilon < K(1 - \lambda_\nu)\,, \tag{25}$$

which implies that $\varepsilon = 0$ because $\lambda_\nu \to 1$. Then, (23) cannot hold $\forall\ \nu \in \mathbb{N}$ and in consequence, the stopping rule is verified for a finite value of ν.

Theorem 3.1 *When the algorithm finishes, a suboptimal policy is available. The cost of this policy is w and it is valid that*

$$v \leq w \leq v + \varepsilon\,. \tag{26}$$

Proof. Since v is the optimal cost of the problem without discount and w is the cost of $\hat{\pi}$, the first inequality is trivial. To prove the second one, from the stopping rule we have that there exists some $\nu \in \mathbb{N}$ such that

$$w - V_{\lambda_\nu} \leq \varepsilon\,. \tag{27}$$

Let $\overline{\pi}$ be any optimal policy, i.e $V_{\overline{\pi},1} = v$, by definition of V_{λ_ν} it results

$$V_{\lambda_\nu} \leq V_{\overline{\pi},\lambda_\nu}\,. \tag{28}$$

Besides $V_{\overline{\pi},1} \geq V_{\overline{\pi},\lambda_\nu}$ because $\lambda_\nu < 1$. From these inequalities, we get

$$w - v \leq \varepsilon\,. \tag{29}$$

□

4 Numerical examples

Example 4.1 We consider $\Omega = \{1,2,3,4\}$ the set of states and $A = \{1,2,3,4\}$ the admissible actions. The instantaneous cost and the transition matrix have been generated randomly. We use $\varepsilon = 10^{-7}$ as the maximum error of approximation in Algorithm 2.

The cost w associated to the suboptimal policy given by the Algorithm 2, is shown in the following table as well as the error $w - V_\lambda$.

$$w = \begin{pmatrix} 0.267 & 0.267 & 0.267 & 0.267 \\ 0.252 & 0.252 & 0.252 & 0.252 \\ 0.116 & 0.116 & 0.116 & 0.116 \\ 0.02 & 0.02 & 0.02 & 0.02 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad w - V_\lambda = 10^{-9} \times \begin{pmatrix} 0.429 & 0 & 0.233 & 0.063 \\ 0.429 & 0 & 0.233 & 0.063 \\ 0.246 & 0 & 0.233 & 0.063 \\ 0.078 & 0 & 0.069 & 0.063 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Example 4.2 Again, let us consider the problem described in Example 3.1 to compare both algorithms. In the value iteration method we have used the stopping test $\|Mw - w\| \leq 10^{-6}$.

Algorithm	Time
Value iteration	29,65 s
Policy iteration	0,27 s

Table 1: Comparative table, $N = 5, N_A = 1, \rho = 0.1$

Remark 4.1 The computational times shown above correspond to a PC 486 (50 Mz) using the programming system MATLAB.

5 Conclusion

We have studied here a discrete minimax stochastic problem defined on a finite Markov chain and we have devised two algorithms to compute numerically the value function of the problem. This function is a natural approximation of the optimal cost corresponding to the continuous time minimax problem (with infinite horizon) studied in [4, 5]. In that case, the Markov chain is obtained via Kushner's procedure to discretize the dynamics of the system appearing in the continuous problem. We want to remark that it seems necessary to use additional hypotheses in order to obtain the convergence of the discrete solutions towards the continuous solution. Among these ones,

- conditions which assure that the finite horizon problems converges to the infinite horizon problem,
- existence of an attractive state or of an attractive cycle.

The analysis of these issues is the subject of [6].

Acknowledgements

The authors would like to thank:

- Elina M. Mancinelli for their careful revision of the manuscript.
- CONICET for support given through the grant PID BID-CONICET N°213.
- The authorities of INRIA for the support given through the Cooperation Project INRIA-Instituto de Matemática Beppo Levi. This work was completed when the first author visited the INRIA Rocquencourt in June 1996.

References

[1] Borkar V.S., *Topics in controlled Markov chains,* Longman–Pitman, London, 1991.

[2] Di Marco S.C., González R.L.V., *A numerical procedure for minimizing the maximum cost*, Rapport de Recherche N°2454, INRIA, 1995.

[3] Di Marco S.C., González R.L.V., *Une procédure numérique pour la minimisation du coût maximum,* Comptes Rendus Acad. Sc. Paris, Série I, Tome 321, pp. 869-874, 1995.

[4] Di Marco S.C., *Sobre la optimización minimax y tiempos de detención óptimos,* Thesis, University of Rosario, Argentine, 1996.

[5] Di Marco S.C., González R.L.V., *Minimax optimal control problem – Infinite horizon case,* work in progress.

[6] Di Marco S.C., *On the numerical solution of minimax optimal control problem with infinite horizon,* work in progress.

[7] Fleming W.H., Rishel R.W., *Deterministic and stochastic optimal control,* Springer–Verlag, New York, 1975.

[8] Fleming W.H., Soner H.M., *Controlled Markov processes and viscosity solutions,* Springer–Verlag, New York, 1993.

[9] Friedman A., *Stochastic differential equations and applications, Vol. I and II*, Academic Press, New York, 1975, 1976.

[10] González R.L.V., Sagastizábal C.A, *Un algorithme pour la résolution rapide d'équations discrètes de Hamilton–Jacobi–Bellman,* Comptes Rendus Acad. Sc. Paris, Série I, Tome 311, pp. 45-50, 1990.

[11] Kushner H.J., *Probability methods for approximations in stochastic control and for elliptic equations*, Academic Press, New York, 1977.

[12] Kushner H.J., Dupuis P.G., *Numerical methods for stochastic control problems in continuous time*, Springer–Verlag, New York, 1992.

[13] Romanovsky V.I., *Discrete Markov chains*, Walters–Noordhoff, Groningen, 1970.

[14] Ross S.M., *Applied probability models with optimization applications,* Holden–Day, San Francisco, 1970.

[15] Ross S.M., *Introduction to stochastic dynamic programming*, Academic Press, New York, 1983.

Unité de recherche INRIA Lorraine, Technopôle de Nancy-Brabois, Campus scientifique,
615 rue du Jardin Botanique, BP 101, 54600 VILLERS LÈS NANCY
Unité de recherche INRIA Rennes, Irisa, Campus universitaire de Beaulieu, 35042 RENNES Cedex
Unité de recherche INRIA Rhône-Alpes, 655, avenue de l'Europe, 38330 MONTBONNOT ST MARTIN
Unité de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex
Unité de recherche INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex

Éditeur
INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
ISSN 0249-6399