

Proyecciones Journal of Mathematics
Vol. 31, N° 3, pp. 219-233, September 2012.
Universidad Católica del Norte
Antofagasta - Chile

Optimal replacement in a system of n -machines with random horizon

ROCIO ILHUICATZI-ROLDÁN

*Benemérita Universidad Autónoma de Puebla, Mexico
and*

HUGO CRUZ-SUÁREZ

*Benemérita Universidad Autónoma de Puebla, Mexico
Received : May 2011. Accepted : May 2012*

Abstract

This paper considers a system consisting of independently operating n -machines, which follows a deterioration processes with an associated cost function. It is assumed that the system is observed at discrete time and the objective function is the total expected cost. Also, it is considered that the horizon of the problem is random. For this problem, a replacement optimal policy that minimize the operation cost of the system is provided. Besides, a numerical example through a program in Matlab is presented.

Keywords : *Mathematical programming, Optimal stochastic control, Dynamic programming, Markov decision processes.*

1. Introduction

In industrial processes is common the deterioration of electronic components or machines, then it is important to provide replacement strategies for the optimization of these systems. The problem of optimal replacement is modeled in different ways. For example, in [9] the problem is studied considering a single machine which follows a deterioration stochastic process with various quality levels in a continuous time. Also, the optimal replacement has been implemented based on future technological advances, in this case, a non-stationary process is considered and the optimal decision is characterized using a forecast horizon approach (see [6]). On the other hand, in [2] the problem is studied with n -machines considering two heuristics rules of replacement that make possible the search of optimal policies. These rules are as follow: the first suggests that a machine is replaced only if all older machines are replaced and the second one indicates that in any stage all machines of the same age are either kept or replaced.

In this work the model proposed in [1] has been considered, which involves a single machine that follows a Markov process of deterioration with D possible levels, where D is a positive integer, associated with operation costs of the machine at each level. The machine is observed in discrete time and, depending on the deterioration level, the following situations are possible:

1. Leaving operate for a additional period of time.
2. Replaced it with a cost $R > 0$.

Fixing a finite horizon of operation for the system is determined an optimal replacement policy, such that the expected total cost incurred is minimal. Now, in this paper a system consisting of n -machines with independent deterioration processes is studied, assuming that the system is operating in a random horizon. This novel consideration in the model is because it is possible that external factors obligate to conclude the process before expected, for example, bankruptcy of the firm in an economic model (see [8], p. 125). The approach to analyze this problem is through Markov decision processes.

This paper is organized as follows. In Section 2 the basic theory of Markov decision processes is presented. In Section 3, a general study on Markov decision processes with random horizon is provided, which allows to solve the problem that is proposed. Afterwards, the problem of optimal replacement in a system with n -machines and random horizon is described

in Section 4. Then, in the following section, the Markov control model and the dynamic programming equations are structured. For solving numerical cases a program in Matlab is elaborated. Finally, in Section 6 some numerical results are illustrated.

2. Basic theory of Markov decision processes

The theory presented in this section can be consulted in [3].

Consider the following non-homogeneous Markov decision or control model:

$$(X, A, \{A(x) \mid x \in X\}, Q, c_t, t \in \{0, 1, 2, \dots\})$$

where

- a) X is a Borel space, called the state space;
- b) A is a Borel space, called control or action set;
- c) $\{A(x) \mid x \in X\}$ is a nonempty measurable subsets $A(x)$ of A , where $A(x)$ denotes the set of feasible controls or actions when the system is in state $x \in X$, and with the property that the set $\mathbf{K} := \{(x, a) \mid x \in X, a \in A(x)\}$ of feasible state-actions pairs is a measurable subset of $X \times A$.
- d) Q is a stochastic kernel on X given \mathbf{K} called the transition law, i.e. $Q(\cdot \mid \cdot)$ is a function such that $Q(\cdot \mid k)$ is a probability measure on X for each fixed $k \in \mathbf{K}$ and $Q(B \mid \cdot)$ is a measurable function on X for each $B \in B(X)$.
- e) $c_t : \mathbf{K} \rightarrow \mathbf{R}$ is a measurable function called the cost-per-stage or one-stage cost function.

Remark 2.1. *The Markov control model that is considered in this paper is a non-homogeneous system in the cost. A general nonhomogeneous Markov decision model is $(X_t, A_t, \{A_t(x) \mid x \in X_t\}, Q_t, c_t, t \in \{0, 1, 2, \dots\})$, this kind of model can be consulted in [8].*

Let $(X, A, \{A(x) \mid x \in X\}, Q, c_t, t \in \{0, 1, 2, \dots\})$ be a Markov control model and for each $t = 0, 1, \dots$ define the space \mathbf{H}_t of admissible histories up to time t as $\mathbf{H}_0 := X$, $\mathbf{H}_t := \mathbf{K}^t \times X = \mathbf{K} \times \mathbf{H}_{t-1}$, for $t = 1, 2, \dots$

An arbitrary element h_t of \mathbf{H}_t , which is called an admissible t -history or simply t -history, is a vector of the form $h_t = (x_0, a_0, \dots, x_{t-1}, a_{t-1}, x_t)$, with $(x_i, a_i) \in \mathbf{K}$ for $i = 0, 1, \dots, t-1$ and $x_t \in X$.

Then, a control policy is a sequence $\pi = \{\pi_t, t = 0, 1, \dots\}$ of stochastic kernels π_t on the control set A given \mathbf{H}_t such that $\pi_t(A(x_t) | h_t) = 1$, $h_t \in \mathbf{H}_t, t = 0, 1, \dots$

Let Φ be the set of all stochastic kernels φ such that $\varphi(A(x) | x) = 1$ for all $x \in X$, and let \mathbf{F} be the set of all measurable functions $f : X \rightarrow A$ satisfying that $f(x) \in A(x)$ for all $x \in X$. The functions in \mathbf{F} are called selectors of the multifunction $x \mapsto A(x), x \in X$.

The set of all policies is denoted by Π . In this work, deterministic Markov policies are characterized. A policy is called deterministic Markov policy, if there is a sequence $\{f_t\}$ of functions $f \in \mathbf{F}$ such that $\pi_t(\cdot | h_t)$ is concentrated at $f_t(x_t) \in A(x_t)$ for all $h_t \in \mathbf{H}_t, t = 0, 1, 2, \dots$, i.e. $\pi_t(C | h_t) = I_C[f_t(x_t)]$, for all $C \in B(A)$.

Let (Ω, F) be a measurable space consisting of the canonical sample space $\Omega := \mathbf{H}_\infty$ and F is the corresponding product σ -algebra. The elements of Ω are sequences of the form $\omega = (x_0, a_0, x_1, a_1, \dots)$, with $x_t \in X$ and $a_t \in A, t = 0, 1, 2, \dots$. The projections x_t and a_t of Ω to the sets X and A are called state and action variables, respectively. Let $\pi = \{\pi_t\}$ be an arbitrary control policy and μ an arbitrary probability measure on X , referred to as the initial distribution. Then, by the theorem of C. Ionescu-Tulcea (see [3]), there exists a unique probability measure P_μ^π on (Ω, F) which, is concentrated on \mathbf{H}_∞ , i.e., $P_\mu^\pi(\mathbf{H}_\infty) = 1$. The stochastic process $(\Omega, F, P_\mu^\pi, \{x_t\})$ is called a discrete-time Markov control process or Markov decision process.

The expectation operator with respect to P_μ^π is denoted by E_μ^π . If μ is concentrated at the initial state $x \in X$, then we write P_μ^π and E_μ^π as P_x^π and E_x^π , respectively.

Let $\pi \in \Pi$ and $x \in X$. The expected total cost with finite horizon N is defined by

$$J(\pi, x) := E_x^\pi \left[\sum_{t=0}^{N-1} c_t(x_t, a_t) + c_N(x_N) \right],$$

where c_N is a measurable function called the terminal cost function. Consider the Markov control model $(X, A, \{A(x) | x \in X\}, Q, c_t, t \in \{0, 1, 2, \dots\})$ and suppose that wish to minimize $J(\pi, x)$. Define the optimal value function as

$$J(x) := \inf_{\pi \in \Pi} J(\pi, x),$$

$x \in X$. The optimal control problem is to determine a policy $\pi^* \in \Pi$ such that $J(\pi^*, x) = J(x), x \in X$.

An approach to analyze the optimal control problem is Dynamic Programming (DP). In the following section, a theorem to guarantee the existence of optimal policies for the optimal control problem is presented.

3. Markov decision processes with random horizon

Let (Ω', F') a measurable space. Consider the Markov control model $(X, A, \{A(x) \mid x \in X\}, Q, c_t, t \in \{0, 1, 2, \dots\})$ and the following performance criterion:

$$J_\tau(\pi, x) := E \sum_{t=0}^{\tau} c_t(x_t, a_t),$$

$\pi \in \Pi$, $x \in X$, where τ is a random variable on (Ω', F') with known probability distribution $P(\tau = t) = \rho_t$, $t = 0, 1, 2, \dots, T$, where T is a positive integer.

Remark 3.1. E is the expected value with respect to the joint distribution of the process $\{x_t, a_t\}$ and τ .

Define the optimal value function as

$$J_\tau(x) := \inf_{\pi \in \Pi} J_\tau(\pi, x),$$

$x \in X$.

Assumption 3.2. τ is independent of $\{(x_t, a_t)\}$.

By Assumption 3.2, it follows that

$$\begin{aligned} J_\tau(\pi, x) &= E \left[E \left(\sum_{t=0}^{\tau} c_t(x_t, a_t) \mid \tau \right) \right] \\ &= \sum_{n=0}^T \left(E_x^\pi \sum_{t=0}^n c_t(x_t, a_t) \right) \rho_n \\ &= \sum_{t=0}^T \sum_{n=t}^T E_x^\pi c_t(x_t, a_t) \rho_n \\ &= E_x^\pi \sum_{t=0}^T P_t c_t(x_t, a_t), \end{aligned}$$

$\pi \in \Pi$ and $x \in X$, where $P_t = \sum_{n=t}^T \rho_n = P(\tau \geq t)$.

Thus, the optimal control problem with a random horizon τ is equivalent to an optimal control problem with a finite horizon $T + 1$, a cost per stage $P_t c_t$ and a null terminal cost (see [5] and [4]).

The following result is motivated by Theorem 3.2.1 in [3], in our case the proof to let a random horizon was adapted.

Theorem 3.3. *Let J_0, J_1, \dots, J_{T+1} be the functions on X defined by $J_{T+1}(x) := 0$ and for $t = T, T - 1, \dots, 0$,*

$$(3.1) \quad J_t(x) := \min_{A(x)} \left[P_t c_t(x, a) + \int_X J_{t+1}(y) Q(dy | x, a) \right].$$

Suppose that J_t is a measurable function for each $t = 0, 1, \dots, T$. Then, there is a selector $f_t \in \mathbf{F}$ such that $f_t(x) \in A(x)$ attains the minimum in (3.1) for all $x \in X$ and $t = 0, 1, \dots, T$; i.e.

$$(3.2) \quad J_t(x) = P_t c_t(x, f_t) + \int_X J_{t+1}(y) Q(dy | x, f_t),$$

$x \in X, t = 0, 1, \dots, T$. Then the deterministic Markov policy $\pi^ = \{f_0, f_1, \dots, f_T\}$ is optimal and the optimal value function J_τ equals J_0 , i.e.,*

$$J_\tau(x) = J_\tau(\pi^*, x) = J_0(x),$$

$x \in X$.

Proof. Let $\pi \in \Pi$ be an arbitrary policy and define

$$C_t(\pi, x) := E^\pi \left[\sum_{n=t}^T P_n c_n(x_n, a_n) \mid x_n = x \right],$$

for $t = 0, 1, \dots, T$, and $C_{T+1}(\pi, x) := 0$. $C_t(\pi, x)$ is called the cost from time t onwards when using the policy π and $x_t = x$.

In particular note that

$$(3.3) \quad J_\tau(\pi, x) = C_0(\pi, x)$$

By backward induction, for $t = T + 1$ $C_{T+1}(\pi, x) = 0 = J_{T+1}(x)$.

Now, suppose that for some $t = T, T - 1, \dots, 0$, $C_{t+1}(\pi, x) \geq J_{t+1}(x)$, $x \in X$.

Then

$$\begin{aligned} C_t(\pi, x) &= E^\pi \left[P_t c_t(x_t, a_t) + \sum_{n=t+1}^T P_n c_n(x_n, a_n) \mid x_t = x \right] \\ &= \int_A \left[P_t c_t(x, a) + \int_X C_{t+1}(y) Q(dy | x, a) \right] \pi_t(da | x), \end{aligned}$$

and by the induction hypothesis

$$\begin{aligned} C_t(\pi, x) &\geq \int_A \left[P_t c_t(x, a) + \int_X J_{t+1}(y) Q(dy | x, a) \right] \pi_t(da | x) \\ &\geq \min_{A(x)} \left[P_t c_t(x, a) + \int_X J_{t+1}(y) Q(dy | x, a) \pi_t(da | x) \right] \end{aligned}$$

hence $C_t(\pi, x) \geq J_t(x)$, $x \in X$ and $t = 0, 1, \dots, T + 1$.

On other hand, if $C_{t+1}(\pi, x) = J_{t+1}(x)$ for all $x \in X$ with $\pi = \pi^*$, $\pi_t(\cdot | h_t)$ is the measure of Dirac concentrated at $f_t(x_t)$, then the equality holds throughout the previous calculations obtaining $C_t(\pi, x) = J_t(x)$. Then, if $C_t(\pi, x) \geq J_t(x)$, in particular for $t = 0$ and of (3.3), $J_\tau(\pi, x) \geq J_0(x)$ and for $\pi = \pi^*$, $J_\tau(\pi^*, x) = J_0(x)$.□

4. Description of the problem

Consider a system consisting of n machines, each with an independent stochastic process of deterioration, whose possible levels of deterioration are denoted by $1, 2, 3, \dots, D$, where D is a positive integer. Level one denotes that the machine is in perfect condition. Suppose that deterioration level is increasing, i.e., that a machine operating at level i is better than $i + 1$, $i = 1, 2, 3, \dots, D - 1$.

Let $P = (p_{i,j})_{D \times D}$ be the matrix of transition probabilities for going from level i to level j (identical for the n -machines). Since a machine can not move to better level of deterioration, $p_{i,j} = 0$ if $j < i$. Let $g : \{1, 2, 3, \dots, D\} \rightarrow \mathbf{R}$ be a known function, which will measure the cost of operation of a machine. Suppose that g is nondecreasing, i.e. $g(1) \leq g(2) \leq \dots \leq g(D)$, and in the beginning of each period of time can be taken the following options.

- a) Operate the machine k , $k = 1, 2, \dots, n$ in a level of deterioration for this time period, or
- b) replace by a new one with a fixed cost $R > 0$.

Also consider that the system can operate for τ time periods, where τ is a random variable independent of the process followed by the system with probability distribution $P(\tau = t) = \rho_t$, $t = 0, 1, 2, \dots, T$, where T is a positive integer.

The problem consists on determining optimal replacement policies that minimize the expected total cost of operation of the system.

5. Modeling the problem

The problem is solved through the theory of Markov decision processes. This requires building the corresponding Markov control model. At the beginning of an arbitrary time period, the state of the system can be registered as (d_1, d_2, \dots, d_n) where d_k , $k = 1, 2, \dots, n$, is the level of deterioration in which the machine is operating, therefore the state space is defined by:

$$(5.1) \quad X = \{(d_1, d_2, \dots, d_n) | d_k \in \{1, 2, \dots, D\}, k = 1, 2, \dots, n\},$$

where $\text{card}(X) = D^n$ states. A replacement action can be represented by (a_1, a_2, \dots, a_n) with $a_k = 0$ or $a_k = 1$, where $a_k = 0$ means let that the machine k operate on the level d_k and $a_k = 1$ means replace it. At this way

$$(5.2) \quad A = A(x) = \{(a_1, a_2, \dots, a_n) | a_k \in \{0, 1\}, k = 1, 2, \dots, n\},$$

where its cardinality is 2^n actions. For an arbitrary machine k , let $P^0 = (p_{i,j}^0)_{D \times D} = P$ be the transition matrix of the process of deterioration when the machine is not replaced. Let $P^1 = (p_{i,j}^1)_{D \times D}$ be the transition matrix when the machine is replaced, where $p_{i,j}^1 = 1$, if $j = 1$ and $p_{i,j}^1 = 0$ in otherwise (safely when machine k was replaced, the machine goes to level one). Let $Q^a = (q_{i,j}^a)_{D^n \times D^n}$ be the transition matrix of the state i at state j of the system, when the action $a \in A$ is taken, $i, j \in X$. For the independence of the deterioration processes of the machines, it is obtained that

$$(5.3) \quad q_{(i_1, i_2, \dots, i_n), (j_1, j_2, \dots, j_n)}^{(a_1, a_2, \dots, a_n)} = p_{i_1, j_1}^{a_1} \cdot p_{i_2, j_2}^{a_2} \cdot \dots \cdot p_{i_n, j_n}^{a_n}.$$

Also at any time period t

$$c(x_t, a_t) = \sum_{k=1}^n \gamma(x_{k,t}, a_{k,t}),$$

where $\gamma(x_{k,t}, a_{k,t}) = g(x_{k,t})$, if $a_{k,t} = 0$ and $\gamma(x_{k,t}, a_{k,t}) = g(1) + R$, if $a_{k,t} = 1$. In this case $x_{k,t}$ and $a_{k,t}$ represent the state and the action at time t of the machine k , respectively. Therefore the performance criterion for this problem is

$$J_\tau(\pi, x) := E \sum_{t=0}^{\tau} c(x_t, a_t),$$

$\pi \in \Pi$ and $x \in X$, where τ is the planning horizon of the problem.

Then, using Theorem 3.3, it is obtained that the dynamic programming equation is given by $J_{T+1}(x) = 0$, and

$$(5.4) \quad J_t(x) = \min_{a \in A(x)} \left[P_t c(x, a) + \sum_{y \in X} J_{t+1}(y) q_{x,y}^a \right],$$

$t = 0, 1, \dots, T$.

Remark 5.1. *The assumption of measurability and existence of selectors in Theorem 3.3 holds, due to X and A are finites (see [8], p. 90).*

The amount of numerical calculations involving the equation (5.4) depends on the value of n , D and T . For this reason a program in Matlab that carry out numerical calculations was implemented.

Below, the algorithm followed for the elaboration of the program is presented.

1. Read of the following data:
 - n , the number of machines.
 - D , the number of deterioration levels.
 - P , the transition matrix of deterioration process.
 - $g(i)$, the cost for operating a machine at level of deterioration i , $i = 1, 2, 3, \dots, D$.
 - R , the cost per unit replaced.
 - T , the support of the horizon.
 - $\{\rho_t, t = 1, 2, 3, \dots, T\}$ probability distribution of the horizon.
2. Construct the state space X , using (5.1).
3. Construct the action space A , using (5.2).
4. Calculate the transition matrices Q^a for each $a \in A$, through (5.3).
5. Obtain the optimal policy and optimal value function, as follow
 - 5.1. do $t = T + 1$ and $J_t(x) = 0$ for each $x \in X$.
 - 5.2. if $t = 0$, stop, then $J_t(x)$ is the optimal value function and $\{f_0, f_1, \dots, f_{T+1}\}$ is the optimal policy. Else, go to 5.3.

- 5.3. replace t by $t - 1$ and calculate $J_t(x)$ for $x \in X$ by means of the equation $J_t(x) = \min_{a \in A(x)} \left[P_t c(x, a) + \sum_{y \in X} J_{t+1}(y) q_{x,y}^a \right]$, doing $f_t(x) = a$, for some $a \in \operatorname{argmin}_{a \in A(x)} \left[P_t c(x, a) + \sum_{y \in X} J_{t+1}(y) q_{x,y}^a \right]$. Return to 5.2.

6. Example

Consider the optimal replacement problem with a random horizon and the following numerical values: $n = 3$, $D = 3$,

$$P = \begin{pmatrix} 0.4 & 0.3 & 0.3 \\ 0 & 0.3 & 0.7 \\ 0 & 0 & 1 \end{pmatrix},$$

$g = (5, 7, 29)$, $R = 4$, $T = 6$ and $\rho = (0.1, 0.1, 0.3, 0.2, 0.15, 0.15)$.

In the table 6.1, the optimal expected total cost is presented for each initial state.

In the table 6.2, the optimal action depending on the stage and state of the system is reported.

For a given number of machines n , consider the state of perfect conditions of the system as the initial state, i.e. $x_0 = (1, 1, \dots, 1)_{n \times 1}$. Let $J_\tau^n(x_0)$ be the optimal expected value. In table 6.3, the optimal value is illustrated for different values of n , whose graph is shown in figure 6.1.

The numerical calculations show that there is a linear relation between the number of machines and the optimal expected value, with initial state x_0 . Moreover, it can be seen that in this particular case, it is possible to write the following relation: $J_\tau^n(x_0) = nJ_\tau^1(x_0)$. In a general case, with initial state, $x_0 = (d_1, d_2, \dots, d_n)$, computationally it has been observed that $J_\tau^n(x_0) = J_\tau^1(d_1) + J_\tau^1(d_2) + \dots + J_\tau^1(d_n)$, which would divide the problem with n machines in n problems with a single machine, this would give many computational benefits.

initial state x	optimal value $J_T(x)$	initial state x	optimal value $J_T(x)$
1 1 1	67.8226	1 1 2	70.4790
1 1 3	70.4790	1 2 1	70.4790
1 2 2	73.1353	1 2 3	73.1353
1 3 1	70.4790	1 3 2	73.1353
1 3 3	73.1353	2 1 1	70.4790
2 1 2	73.1353	2 1 3	73.1353
2 2 1	73.1353	2 2 2	75.7917
2 2 3	75.7917	2 3 1	73.1353
2 3 2	75.7917	2 3 3	75.7917
3 1 1	70.4790	3 1 2	73.1353
3 1 3	73.1353	3 2 1	73.1353
3 2 2	75.7917	3 2 3	75.7917
3 3 1	73.1353	3 3 2	75.7917
3 3 3	75.7917		

Table 6.1: Optimal value for each initial state

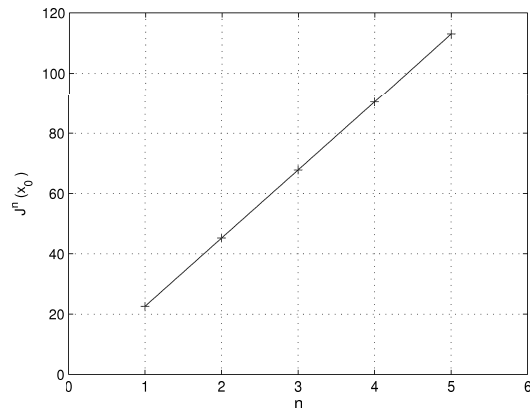


Figure 6.1 : Optimal value with initial state $x_0 = (1, 1, 1)$ in function of n .

7. Conclusion

In this paper a problem of optimal replacement in a system of n -machines considering random horizon has been analyzed. The problem is modeled as a Markov decision process, using this approach it is possible to characterize the optimal replacement policy through of dynamic programming equation. Moreover, a program in Matlab that helps to solve numerical cases was developed. Results obtained for special cases suggest that we can divide the problem with n -machines in n problems with a single machine, which reduce the course of dimensionality (see [7]).

References

- [1] Bertsekas D.; Programming and Optimal Control, Athena Scientific, Massachusetts, 1987.
- [2] Childress S., Durango-Cohen P.; On Parallel Machine Replacement Problems with General Replacement Functions and Stochastic Deterioration, Vol. 52, Naval Research Logistics, 2005.
- [3] Hernández-Lerma O., Lasserre J. B.; Discrete-Time Markov Control Processes: Basic Optimality Criteria, Springer-Verlag, New York, 1996.
- [4] Iida T., Mori M.; Markov Decision Processes with Random Horizon, Journal of the Operations Research, Vol. 39, No. 4, 1996.
- [5] Levhari D., Mirman L. J.; Savings and Consumption with an Uncertain Horizon, Journal of Political Economy, Vol. 85, No. 2, 1977.
- [6] Nair S. K., Hopp W. J.; A model for equipment replacement due to technological obsolescence, European Journal of Operational Research, Vol. 63, 1992.
- [7] Powell W. B.; Approximate Dynamic Programming: Solving the Curses of Dimensionality, John Wiley & Sons, New Jersey, 2007.

- [8] Puterman M. L.; Markov Decision Process: Discrete Stochastic Dynamic Programming, John Wiley & Sons, New York, 1994.
- [9] Sheti S. P., Sorger G., Zhou X. Y.; Stability of Real-Time Lot-Scheduling and Machine Replacement Policies with Quality Levels, IEEE Transactions on Automatic Control, Vol. 45, No. 11, 2000.

Maria del Rocio Ilhuicatzí Roldán

Facultad de Ciencias Físico-Matemáticas
Benemérita Universidad Autónoma de Puebla
México
e-mail: rroldan@alumnos.fcfm.buap.mx

and

Hugo Cruz Suárez

Facultad de Ciencias Físico-Matemáticas
Benemérita Universidad Autónoma de Puebla
México
e-mail: hcs@fcfm.buap.mx

stage state	1	2	3	4	5	6
1 1 1	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0
1 1 2	0 0 0	0 0 0	0 0 0	0 0 0	0 0 1	0 0 1
1 1 3	0 0 1	0 0 1	0 0 1	0 0 1	0 0 1	0 0 1
1 2 1	0 0 0	0 0 0	0 0 0	0 0 0	0 1 0	0 1 0
1 2 2	0 0 0	0 0 0	0 0 0	0 0 0	0 1 1	0 1 1
1 2 3	0 0 1	0 0 1	0 0 1	0 0 1	0 1 1	0 1 1
1 3 1	0 1 0	0 1 0	0 1 0	0 1 0	0 1 0	0 1 0
1 3 2	0 1 0	0 1 0	0 1 0	0 1 0	0 1 1	0 1 1
1 3 3	0 1 1	0 1 1	0 1 1	0 1 1	0 1 1	0 1 1
2 1 1	0 0 0	0 0 0	0 0 0	0 0 0	1 0 0	1 0 0
2 1 2	0 0 0	0 0 0	0 0 0	0 0 0	1 0 1	1 0 1
2 1 3	0 0 1	0 0 1	0 0 1	0 0 1	1 0 1	1 0 1
2 2 1	0 0 0	0 0 0	0 0 0	0 0 0	1 1 0	1 1 0
2 2 2	0 0 0	0 0 0	0 0 0	0 0 0	1 1 1	1 1 1
2 2 3	0 0 1	0 0 1	0 0 1	0 0 1	1 1 1	1 1 1
2 3 1	0 1 0	0 1 0	0 1 0	0 1 0	1 1 0	1 1 0
2 3 2	0 1 0	0 1 0	0 1 0	0 1 0	1 1 1	1 1 1
2 3 3	0 1 1	0 1 1	0 1 1	0 1 1	1 1 1	1 1 1
3 1 1	1 0 0	1 0 0	1 0 0	1 0 0	1 0 0	1 0 0
3 1 2	1 0 0	1 0 0	1 0 0	1 0 0	1 0 1	1 0 1
3 1 3	1 0 1	1 0 1	1 0 1	1 0 1	1 0 1	1 0 1
3 2 1	1 0 0	1 0 0	1 0 0	1 0 0	1 1 0	1 1 0
3 2 2	1 0 0	1 0 0	1 0 0	1 0 0	1 1 1	1 1 1
3 2 3	1 0 1	1 0 1	1 0 1	1 0 1	1 1 1	1 1 1
3 3 1	1 1 0	1 1 0	1 1 0	1 1 0	1 1 0	1 1 0
3 3 2	1 1 0	1 1 0	1 1 0	1 1 0	1 1 1	1 1 1
3 3 3	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1

Table 6.2: Optimal policies

n	$J_{\tau}^n(x_0)$
1	22.6075
2	45.2151
3	67.8226
4	90.4301
5	113.0377

Table 6.3: Optimal value with initial state $x_0 = (1, 1, 1)$ for different values of n