Analysis of a Risk Sensitive Control Problem for Hidden Markov Chains[1]

Daniel Hernández-Hernández[2], Steven I. Marcus[3] and Pedram J. Fard[4]

§1 Introduction

Recently considerable attention has been given to the study of risk-sensitive stochasic control problems [1]-[13]. Risk-sensitive control is a generalization of the risk-neutral approach, in which we seek to minimize an exponential of the sum of costs. In particular, there has been work on risk-sensitive Markov Decision Processes (MDPs) – i.e., discrete-time problems with a finite or countable state space; this research is reviewed in [12]. We focus here on risk-sensitive control of partially observable MDPs, also known as Hidden Markov Models (HMMs). For such problems, an information state and dynamic programming equations are introduced in [1]. Structural results for the value function in the partially observed setting are provided in [7]. The average cost partially observed problem has been studied in [8].

As in [7], we study the structure of risk-sensitive controllers (policies), to attempt to understand the implications of the risk-sensitive cost on the behavior of the controller, and to compare the behavior of risk-sensitive and risk-neutral controllers. We will also study the behavior of the controller in the "large risk" (or "worst case") limit (see also [4]). In order to study these questions, we analyze a replacement problem involving the control of failure-prone units in production/manufacturing systems. This system is analyzed in [5] for the average cost risk-neutral criterion and in [7] for risk-sensitive finite horizon cost.

[1]Supported in part by the National Science Foundation under grant EEC 9402384

[2]Department of Mathematics, CINVESTAV-IPN, MEXICO, dher@math.cinvestav.mx

[3]Electrical Engineering Department and Institute for Systems Research, University of Maryland, College Park, Maryland 20742, marcus@isr.umd.edu

[4]Electrical Engineering Department and Institute for Systems Research, University of Maryland, College Park, Maryland 20742, pedram@isr.umd.edu

The remainder of the paper is organized as follows. In Section 2 we introduce the partially observed risk sensitive control problem and a verification theorem is presented. Section 3 presents the replacement problem and proves existence of a solution to the dynamic programming equation. The structure of the optimal policy is analyzed in Section 4. Finally, in Section 4 we present simulation results and compare the behavior of optimal risk-sensitive policies to the risk-neutral and worst case behaviors.

§2 Partially Observed Controlled Markov Model

Let the partially observed controlled Markov model be given by a five-tuple $(X, Y, U, \{P(u) : u \in U\}, \{Q(u) : u \in U\})$; here $X = \{1, 2, ..., N_X\}$ is the finite set of states, $Y = \{1, 2, ..., N_Y\}$ is the set of observations (or outputs), $U = \{1, 2, ..., N_U\}$ is the set of decisions (or controls). In addition, let $P(u) := [p_{ij}(u)]$ be the $N_X \times N_X$ state transition matrix, and $Q(u) := [q_{xy}(u)]$ is the $N_X \times N_Y$ state/output matrix, i.e., $q_{xy}(u)$ is the probability of receiving message y when the state is x and action u has been taken. We adopt the following information pattern:

Information Pattern: At decision epoch t, the system is in the (unobservable) state $x_t = i$, a decision $u_t = u$ is taken, and the state evolves to $x_{t+1} = j$ with probability $p_{ij}(u)$. Once the state has evolved to x_{t+1}, an observation y_{t+1} is gathered, such that:

$$Prob\{y_{t+1} = y | x_{t+1} = i, u_t = u\} = q_{x,y}(u).$$

Hence, based on $\mathcal{I}_t := (y_0, u_0, y_1, ..., u_t, y_{t+1})$, a new decision u_{t+1} is selected. Let $\mathcal{Y}_t$ be the σ-field generated by the available observations up to t.

Given an expected cost per stage $(i, u) \mapsto c(i, u)$, the sum of costs for the finite horizon T is given by

$$\mathcal{C}_T := \sum_{t=0}^{T-1} c(x_t, u_t).$$

The *risk-sensitive control* problem is that of finding a control policy $\pi = \{\pi_0, \pi_1, ..., \pi_{T-1}\}$, with $\mathcal{I}_t \mapsto \pi_t(\mathcal{I}_t) \in U$, such that

$$J^\gamma(p_0, \pi) = \limsup_{T\to\infty} \frac{\gamma}{T} \log \mathbb{E}^\pi_{p_0} \exp\{\frac{1}{\gamma}\mathcal{C}_T\},$$

is minimized, where $\gamma^{-1} > 0$ is the *risk factor*; here $\mathbb{E}^\pi_{p_0}$ denotes the expectation induced by policy π and, implicitly, the initial distribution of the state.

Then the risk sensitive optimal control problem is to find a policy $\pi^* \in \Pi$ (set of admissible controls) that minimizes $J^\gamma(p_0, \pi)$.

Remark 2.1 Under suitable conditions, a solution to this problem has been given in [8] using dynamic programming techniques. There, the problem is transformed into a completely observed problem by introducing a sufficient statistic, and proving that a separated optimal control can be obtained from the corresponding dynamic programming equation. However, those conditions are not satisfied for some problems, e.g. the two state replacement problem. In particular, the conditions of [8] are essentially equivalent to the primitivity of $P(u)$ for all u, which is not the case for the problem considered in Section 3 below. Thus another approach must be followed.

Next we transform the above partially observed problem into one that is completely observed. For full details and justifications consult [1,2,8,12].

Let $\pi \in \Pi$ be given, and let $\mathcal{G}_t = \sigma(x_0, \ldots, x_t; y_1, \ldots, y_t)$ be the σ-field generated by the state and output processes up to time t. Let $\mathcal{P}^\pi$ be the probability measure induced by policy π. Then there exists a probability measure $\mathcal{P}^\dagger$ (with corresponding expectation $\mathbb{E}^\dagger$) on $\mathcal{G}_t$, under which $\{y_t\}$ are independent and identically distributed random variables, uniformly distributed, independent of $\{x_{t'}, t' \leq t\}$, and $\{x_t\}$ is a Markov chain with the same distribution as above. Further, $\mathcal{P}^\dagger$ is equivalent to $\mathcal{P}^\pi$ on $\mathcal{G}_t$, and

$$\frac{d\mathcal{P}^\pi}{d\mathcal{P}^\dagger}|_{\mathcal{G}_t} =: \lambda_t^\pi = \prod_{k=1}^{t} N_Y q_{x_k y_k}(u_{k-1}).$$

In [1] (see also [2]), Baras and James defined the information state

$$\sigma_t(x) := \mathbb{E}^\dagger[I_{\{x_t = x\}} \exp\{\frac{1}{\gamma}\mathcal{C}_t\}\lambda_t^\pi | \mathcal{Y}_t],$$

which is computable through the recursive formula

$$\begin{aligned} \sigma_{t+1} &= M(y,u)\sigma_t \\ \sigma_0 &= p_0, \end{aligned}$$

where $M(y,u)$ has entries

$$M(y,u)_{xx'} = N_Y p_{x'x}(u) q_{x'y}(u) \exp\{\frac{1}{\gamma}c(x',u)\}. \tag{2.1}$$

Then σ_t is a sufficient statistic for the finite horizon risk sensitive output feedback control problem. However, in order to consider the infinite horizon case, it is necessary to introduce a normalized statistic $\rho_t \in S$ (see [8]), with $S = \{\rho \in \mathbb{R}^{N_X} | \rho^i \geq 0$ and $\sum_{i=1}^{N_X} \rho^i = 1\}$, which is defined by $\rho_0 = p_0$ and $\rho_t = \sigma_t / |\sigma_t|$, where $|\sigma| = \sum \sigma^i$.

Let $F : S \times Y \times U \to S$ and $G : S \times Y \times U \to \mathbb{R}$ be defined by

$$F(\rho, y, u) := \frac{M(y,u)\rho}{|M(y,u)\rho|} \quad \text{and} \quad G(\rho, y, u) := \log |M(y,u)\rho|, \tag{2.2}$$

respectively. Then for each $\pi \in \Pi$ and initial distribution p_0, the cost functional $J(p_0, \pi)$ can be expressed in terms of the information state ρ (see [8] Section 4), as

$$J(p_0, \pi) = \limsup_{T \to \infty} \frac{\gamma}{T} \log \mathbb{E}^\dagger \exp\{\sum_{t=0}^{T-1} G(\rho_t, y_{t+1}, u_t)\}, \tag{2.3}$$

and the corresponding dynamic programming equation is

$$\exp\{\gamma\Lambda + W(\rho)\} = \inf_{u \in U} \mathbb{E}^\dagger \exp\{W(F(\rho, y, u)) + G(\rho, y, u)\} \tag{2.4}$$

Theorem 2.2 (Verification). Assume that there exist a number Λ and a bounded function $W : S \to \mathbb{R}$ such that (2.4) holds. Then, the stationary policy π^*, with $\pi^*(\rho)$ achieving the minimum on the right hand side of (2.4) is optimal and Λ is the optimal risk-sensitive average cost.

For the proof of this theorem, see [8, Theorem 4.1].

§3 The replacement problem

This section contains the main results of the paper. We consider the risk sensitive control of a replacement problem, which has been studied in the risk neutral case [5,6]. This is a two-state replacement problem which models failure-prone units in manufacturing systems, communication systems, etc. Let x_t be the state of the unit, which can either be working ($x_t = 0$) or degraded/failed ($x_t = 1$), and the control actions are to operate the system in its current condition ($u_t = 0$), or replace the unit with a new one ($u_t = 1$). The transition matrices $P(u)$ for the state x_t are given by

$$P(0) = \begin{bmatrix} 1-\theta & \theta \\ 0 & 1 \end{bmatrix}; \quad P(1) = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix},$$

with $0 < \theta < 1$. The observation process $\{y_t\}$ gives the correct information about the state x_t with probability q ($q \in (\frac{1}{2}, 1)$), that is,

$$Q = \begin{bmatrix} q & 1-q \\ 1-q & q \end{bmatrix}$$

.

Let $(x,u) \mapsto c(x,u)$ be the cost function defined by $c(0,0) = 0,\ c(1,0) = C,\ c(x,1) = R$, with $0 < C < R$. With the above definitions, the matrices $M(y,u)$ defined in (2.1) are given by

$$(3.1)\quad M(y,0) = 2\begin{bmatrix} q_y(1-\theta) & 0 \\ (1-q_y)\theta & (1-q_y)e^{\frac{1}{\gamma}C} \end{bmatrix};\ M(y,1) = 2\begin{bmatrix} q_y e^{\frac{1}{\gamma}R} & q_y e^{\frac{1}{\gamma}R} \\ 0 & 0 \end{bmatrix}$$

where $q_y := q(1-y) + (1-q)y,\ y = 0,1$.

In this case, the information state ρ_t has the form

$$\rho_t = \begin{bmatrix} 1-\alpha_t \\ \alpha_t \end{bmatrix},$$

with $\alpha_t \in [0,1]$.

Throughout we consider α_t as our information state, with dynamics (see (2.2)) given by

$$(3.2)\qquad \alpha_{t+1} = f(\alpha_t, y_{t+1}, u_t),$$

where

$$\begin{cases} f(\alpha,y,0) &= \dfrac{\alpha[(1-q_y)(e^{\frac{1}{\gamma}C}-\theta)] + \theta(1-q_y)}{\alpha[(1-q_y)(e^{\frac{1}{\gamma}C}-\theta) - (1-\theta)q_y] + \theta(1-q_y) + (1-\theta)q_y} \\ f(\alpha,y,1) &= 0. \end{cases}$$

Remark 3.1 The control action $u = 1$ plays the role of a reset action.

The next proposition summarizes several properties of $f(\cdot,y,0)$, which will be used later in this section. Its proof follows in a straightforward manner by algebraic manipulations, and we omit it.

Proposition 3.1 (a) $f(\cdot,y,0)$ is strictly increasing for $y = 0,1$.
(b) $f(\cdot,1,0)$ is strictly concave.
(c) $f(\cdot,0,0)$ is either convex or concave depending on whether the sign of $(1-q)(e^{\frac{1}{\gamma}C}-\theta) - (1-\theta)q$ is negative or positive.
(d) $\alpha < f(\alpha,1,0)$ and $f(\alpha,0,0) < f(\alpha,1,0)$ for all $\alpha \in [0,1)$, with equality for $\alpha = 1$.
(e) Assume that $1 < \frac{q(1-\theta)}{1-q}$, and define $\Gamma := (0, \frac{1}{C}\log\frac{q(1-\theta)}{1-q})$. If $\gamma \in \Gamma$, then there exists $\alpha_0 \in (0,1)$ such that $f(\alpha_0,0,0) = \alpha_0$.

Now, in order to prove some structural results for the optimal policy, we shall prove the existence of solution to the Equation (2.4), which in this case can be written as

$$e^{\gamma\Lambda+W(\alpha)} = \min\left\{\mathbb{E}^{\dagger} e^{W(f(\alpha,y,0))+G(\alpha,y,0)},\ e^{\frac{1}{\gamma}R}\right\}. \tag{3.3}$$

In the risk neutral case [5] this equation was solved using the vanishing discount approach. However, for the risk-averse case, we approximate the value function of the average cost risk sensitive control problem through the value function of a discounted cost dynamic game satisfying

$$e^{W_\beta(\alpha)} = \min\left\{\mathbb{E}^{\dagger} e^{\beta W_\beta(f(\alpha,y,0))+G(\alpha,y,0)},\ e^{\frac{1}{\gamma}R+\beta W_\beta(0)}\right\} \tag{3.4}$$

We will not describe this dynamic game here. Interested readers can find the details in [8,11,12].

Lemma 3.2 There exists a unique nondecreasing concave function $W_\beta : [0,1] \to \mathbb{R}$ such that (3.4) holds.

Proof. Let $C[0,1]$ be the set of real valued continuous functions defined on $[0,1]$, and define the operator $T : C[0,1] \to C[0,1]$ by

$$T\phi(\alpha) = \log\min\left\{\mathbb{E}^{\dagger} e^{\beta\phi(f(\alpha,y,0))+G(\alpha,y,0)}, e^{\frac{1}{\gamma}R+\beta\phi(0)}\right\}, \quad \phi \in C[0,1].$$

Then straightforward calculations show that T is a nondecreasing operator, i.e., if $\phi(\cdot) \le \varphi(\cdot)$, then $T\phi(\cdot) \le T\varphi(\cdot)$. Let $W_\beta^0 := 0$ and define $W_\beta^{n+1} := T^n W_\beta^0$. Then $W_\beta^n \nearrow W_\beta$, where W_β is the unique fixed point of T. Moreover, using induction, it can be easily seen that $W_\beta^n(\cdot)$ is nondecreasing and concave for all $n = 0, 1 \ldots$, and therefore so is W_β. □

From (3.4), we have that, for each $\alpha \in$ [0,1],

$$e^{W_\beta(\alpha)} \le e^{\frac{1}{\gamma}R+\beta W_\beta(0)}$$

Then, using the fact that $W_\beta(\cdot)$ is nondecreasing in [0,1], we have that

$$0 \le W_\beta(\alpha) - W_\beta(0) \le \frac{1}{\gamma}R \ \text{ for all } \ \alpha \in [0,1]. \tag{3.5}$$

Theorem 3.3 There exist a number Λ and a bounded function $W : [0,1] \to \mathbb{R}$ such that (3.3) is satisfied.

Proof. In [5] the conditions (C') and (UBGT) were introduced; those imply the existence of a solution to the risk neutral average cost dynamic programming equation. Here, both conditions are satisfied, because of Proposition 3.1(a) and (3.3). Following the same steps as in [5], for the risk sensitive case, these conditions also imply the existence of solution to the equation (3.3). $\square$

§4 Structure of optimal policies

In this section we show that the optimal policy is of the "control limit" or "threshold" type, as is the case in the risk neutral average cost case [5].

Lemma 4.1 If every average-cost optimal policy replaces at $\alpha \in [0,1]$, then every average-cost optimal policy replaces in the interval $[\alpha, 1]$.

Proof. The argument used in the proof of Theorem 6.1(i) in [5] works here, noting that the solution of Equation (3.4) (see Lemma 3.2) is concave. $\square$

Lemma 4.2 It is average-cost optimal to produce in the interval $[0, f(0,0,0)]$.

Proof. We prove first that it is average-cost optimal to produce at $\alpha = 0$. From (3.4), we have that

$$e^{W_\beta(0)} = \min\left\{\mathbb{E}^\dagger e^{\beta W_\beta(f(0,y,0))+G(0,y,0)},\ e^{\frac{1}{\gamma}R+\beta W_\beta(0)}\right\}. \tag{4.1}$$

Suppose that

$$\mathbb{E}^\dagger e^{\beta W_\beta(f(0,y,0))+G(0,y,0)} > e^{\frac{1}{\gamma}R+\beta W_\beta(0)}.$$

Then

$$e^{W_\beta(0)} = e^{\frac{1}{\gamma}R+\beta W_\beta(0)},$$

that is,

$$(1-\beta)W_\beta(0) = \frac{1}{\gamma}R. \tag{4.2}$$

On the other hand,

$$e^{W_\beta(1)} \le e^{\beta W_\beta(1)+\frac{1}{\gamma}C},$$

i.e.

$$(1-\beta)W_\beta(1) \le \frac{1}{\gamma}C.$$

Then, using the monotonicity of $W_\beta(\cdot)$, we have

$$\frac{1}{\gamma}R = (1-\beta)W_\beta(0) \le (1-\beta)W_\beta(1) \le \frac{1}{\gamma}C,$$

which is a contradiction. Thus, the minimum on the r.h.s. of (4.1) is achieved at $u = 0$, and using Lemma 4.5 in [5], it follows that it is average-cost optimal to produce at $\alpha = 0$.

Let $0 < \alpha \leq f(0,0,0)$ and assume that it is average-cost optimal to repair at α. Then it is average-cost optimal to repair in $[\alpha, 1]$ by Lemma 4.1. The average cost incurred by this policy is R. In view of Proposition 3.1, also a policy that produces in $[0, f(0,0,0)]$ and replaces in $[(f(0,0,0), 1]$ incurs an average cost of R, and thus it is optimal to produce at α. $\square$

Now we introduce the following condition.

Condition C1. For each $\beta \in (0,1)$

$$\frac{1}{\gamma}C \leq (1-\beta)[\frac{1}{\gamma}R + \beta W_\beta(0)].$$

Theorem 4.3 (a) If (C1) holds, then it is optimal to produce for all $\alpha \in [0,1]$.
(b) If (C1) does not hold, then there exist $\alpha_\gamma \in [0,1)$ such that it is optimal to produce in $[0, \alpha_\gamma)$ and repair in $[\alpha_\gamma, 1]$.

The proof requires the following lemma.

Lemma 4.4 For each $\alpha \in [0,1]$ the function $K(\beta) := (1-\beta)[\frac{1}{\gamma}R + \beta W_\beta(\alpha)]$ is nonincreasing.

Proof From the proof of Lemma 3.2, it follows that there exists a sequence $\{W_\beta^n\}$ such that $W_\beta^0 = 0$, and $W_\beta^n \nearrow W_\beta$. Then, in order to prove the lemma, it is enough to prove that for each $\alpha \in [0,1]$ the function

$$K^n(\beta) = (1-\beta)[\frac{1}{\gamma}R + \beta W_\beta^n(\alpha)] \tag{4.3}$$

is nonincreasing for $n = 0, 1, \ldots$. We first prove that for each $\alpha \in [0,1]$

$$K_1^n(\beta) = (1-\beta)W_\beta^n(\alpha) \ \text{ is not decreasing for all } \ n = 0, 1 \ldots. \tag{4.4}$$

We proceed by induction. For $n = 0$ (4.4) holds since $W_\beta^0 = 0$. Now we assume (4.4) for $n = k$. Let $\alpha \in [0,1]$, and let $0 < \beta' < \beta < 1$. Then,

$$\begin{aligned} &(1-\beta')W_{\beta'}^{k+1}(\alpha) - (1-\beta)W_\beta^{k+1}(\alpha) \qquad (4.5)\\ = \;& (1-\beta') \sup_{\mu \in P(Y)} \left\{ \int [\beta' W_{\beta'}^k(f(\alpha,y,u)) - \log(\frac{\mu(y)}{g(\alpha,y,u)})]\mu(dy) \right\} \end{aligned}$$

$$
\begin{aligned}
& -(1-\beta)\sup_{\mu\in P(Y)}\left\{\int[\beta W_\beta^k(f(\alpha,y,u))-\log(\frac{\mu(y)}{g(\alpha,y,u)})]\mu(dy)\right\} \\
\leq\ & (1-\beta')\beta'\int W_\beta^k(f(\alpha,y,u))\mu_{\beta'}(dy)-(1-\beta')\int\log(\frac{\mu_{\beta'}(y)}{g(\alpha,y,u)})\mu_{\beta'}(dy) \\
& -(1-\beta)\beta\int W_\beta^k(f(\alpha,y,u))\mu_{\beta'}(dy)+(1-\beta)\int\log(\frac{\mu_{\beta'}(y)}{g(\alpha,y,u)})\mu_{\beta'}(dy) \\
\leq\ & (\beta'-\beta)[\int[(1-\beta)W_\beta^k(f(\alpha,y,u))+\log(\frac{\mu_{\beta'}(y)}{g(\alpha,y,u)})]\mu_{\beta'}(dy)] \\
\leq\ & 0,
\end{aligned}
$$

where we have used the fact that W_β^k is nonnegative.

Now we shall prove that for each $\alpha\in[0,1]$, the function

$$
K_2^n(\beta):=\beta W_\beta^n(\alpha)\ \text{ is nonincreasing for all }\quad n=0,1\dots. \tag{4.6}
$$

Again we proceed by induction. For $n=0$ (4.6) obviously holds, and assume that it remain true for $n=k$. then, for each $\alpha\in[0,1]$, and $0<\beta'<\beta<1$,

$$
\begin{aligned}
W_\beta^{k+1}(\alpha) &= \sup_{\mu\in P(Y)}\left\{\int[\beta W_\beta^k(f(\alpha,y,u))-\log(\frac{\mu(y)}{g(\alpha,y,u)})]\mu(dy)\right\} \\
&\leq \sup_{\mu\in P(Y)}\left\{\int[\beta' W_{\beta'}^k(f(\alpha,y,u))-\log(\frac{\mu(y)}{g(\alpha,y,u)})]\mu(dy)\right\} \\
&= W_{\beta'}^{k+1}(\alpha). \qquad (4.7)
\end{aligned}
$$

Then, from (4.5) and (4.7), it follows that (4.6) holds for $n=k+1$. Finally, (4.3) follows in a straightforward way from (4.6). □

Proof of Theorem 4.3. (a) From Lemma 4.4 it is sufficient to prove that if Condition C1 holds, then for $\alpha=1$ the minimum on the r.h.s. of (4.1) is achieved at $u=0$, for each $\beta\in(0,1)$. From (3.2) we have that

$$
W_\beta(1)=\min\{\beta W_\beta(1)+\frac{1}{\gamma}C,\beta W_\beta(0)+\frac{1}{\gamma}R\}
$$

Then, from C1,

$$
\begin{aligned}
\beta W_\beta(1)+\frac{1}{\gamma}C &\leq \beta W_\beta(1)+(1-\beta)[\frac{1}{\gamma}R+\beta W_\beta(0)] \\
&\leq \frac{1}{\gamma}R+\beta W_\beta(0).
\end{aligned}
$$

If C1 does not hold, the same arguments as in part (a) give that it is optimal to repair at $\alpha = 1$. Taking the limit when $\beta \nearrow 1$, and using Lemma 4.4, we get $C > \Lambda$. Assume that the statement of the theorem is not true, i.e., that there is not $\alpha_R \in (0,1)$ such that it is optimal to repair in $[\alpha_R, 1]$. Then it will be optimal to produce in $[0,1)$. The cost of this policy is

$$\mathbb{E}^\dagger \exp\{\sum_{t=0}^{T-1} log|M(\alpha_t, y_{t+1}, 0)|\} = \mathbb{E}^\dagger \prod_{t=0}^{T+1}(1 + (e^{\frac{1}{\gamma}C} - 1)\alpha_t)$$

Now, by the continuity of $\alpha \to 1 + (e^{\frac{1}{\gamma}C} - 1)\alpha$, we have that for all $\epsilon > 0$, there exists $\delta > 0$ such that if $1-\delta < \alpha < 1$ then $|1+(e^{\frac{1}{\gamma}C}-1)\alpha - e^{\frac{1}{\gamma}C}| < \epsilon$. Then, given $\epsilon > 0$, arbitrary but fixed, and α such that $\alpha \in (1-\delta, 1)$,

$$\Lambda = \limsup_{T\to\infty} \frac{\gamma}{T}\mathbb{E}^\dagger \exp\{\sum_{t=0}^{T-1} log|M(\alpha_t, y_{t+1}, 0)|\} > \gamma \log(e^{\frac{1}{\gamma}C} - \epsilon).$$

Since ϵ was chosen arbitrarily, we get that $\Lambda \geq C$, a contradiction. $\square$

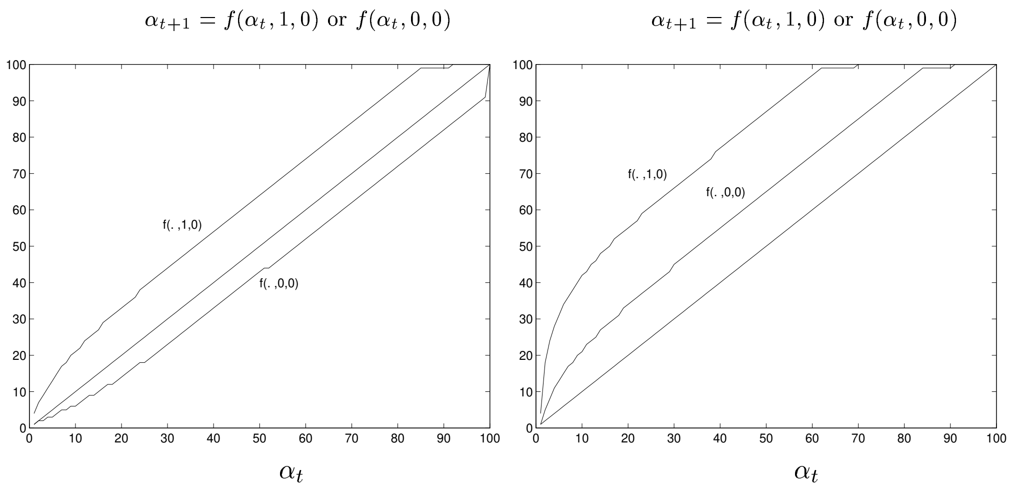
§5 Simulation and comparison of results for different parameters

The optimal policies for this problem under the risk-sensitive and risk-neutral cost criteria are of threshold type; however, for each value of the risk factor γ^{-1}, this policy is a function of a *different* information state. Thus the policies cannot be meaningfully compared simply by comparing the thresholds. In order to compare policies for different values of the risk factor, we have compared by Monte Carlo simulation the percentage of repair actions induced by each policy.

Two sets of parameters were chosen and the effect of the risk factor on other variables in the machine replacement problem were investigated. The interval [0,1] was partitioned into 100 points, so the information state α is limited to 100 values. Information state evolution curves are shown in Figure 1(a,b) for two different values of risk factor. As the risk factor γ^{-1} increases, there is a faster increase in the information state. Note that in Figure 1(b) (larger risk factor), even when the observation indicates that the machine is in good condition, the information state increases.

Figure 2(a,b) show how the threshold value changes with γ and Figures 3(a,b) show the replacement percentages for two different sets of parameters.

When an increase of the risk factor results in a decrease of the threshold ($\frac{R}{C} < 2$), the replacement percentage increases almost monotonically

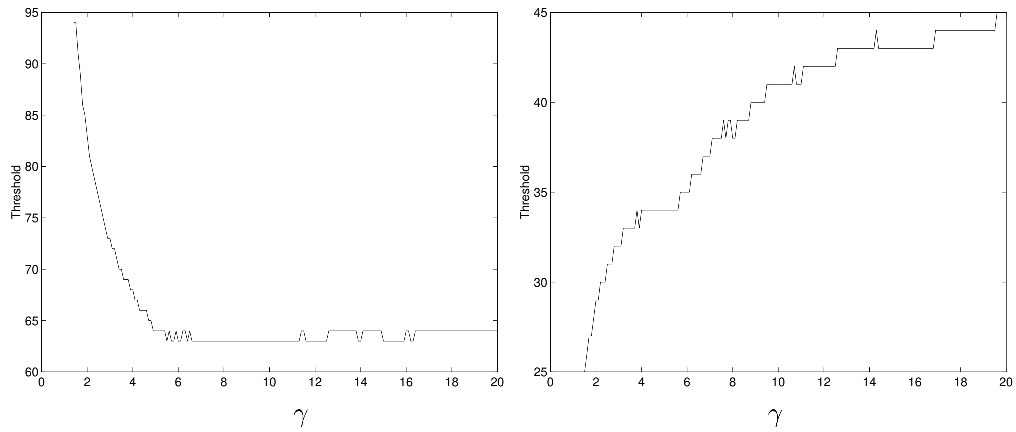


(a) $\gamma^{-1} = 0.05$ (b) $\gamma^{-1} = 0.5$

Figure 1: One step transition curves for information state (α_{t+1} vs. α_t) for different values of γ. The identity map (the line $\alpha_{t+1} = \alpha_t$) is given for comparison.

(the only downward jumps appear when there is is an upward jump in the threshold). But when the threshold increases with the risk factor, there is no monotonicity in the trend of change in the replacement percentage. There are discontinuities in the replacement percentage and threshold curves which are a result of discontinuous change of information state.

The ratio $\frac{R}{C}$ is an important factor in determining the characteristics of the system. When $\frac{R}{C} < 2$ (i.e., $\frac{R}{2} < C$) the replacement percentage approaches 50% for large values of risk factor. This can be interpreted as follows: as $\gamma^{-1} \to \infty$, the risk-sensitive problem approaches the worst-case (or minimax) problem [4]. Since in the worst case, the machine fails as often as possible (one step after a repair), the optimal policy is to replace (i.e. reset the state to 0) every other step, with an average cost of $R/2$. When $\frac{R}{C} > 2$ (i.e., $\frac{R}{2} > C$) if we increase the risk factor, replacing every other step is more costly than not replacing at all and incurring an average cost of C . So when the risk factor gets large enough the replacement percentage suddenly drops to zero.



(a) $\frac{R}{C} > 2$ (b) $\frac{R}{C} < 2$

Figure 2: Threshold values

§6 References

1. J.S. Baras and M.R. James, Robust and risk sensitive output feedback control for finite state machines and hidden Markov models, to appear in *J. Math. Systems, Estimation & Control.*

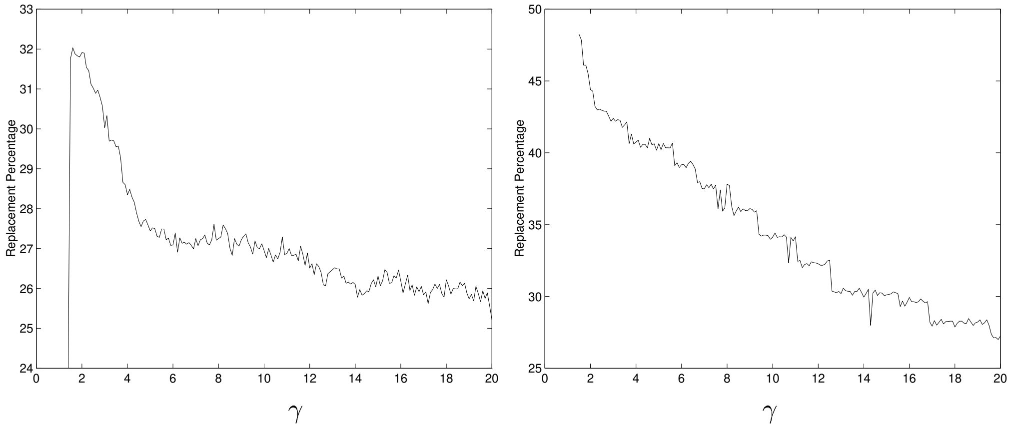
2. M. R. James, J. S. Baras and R. J. Elliott, Risk-sensitive control and dynamic games for partially observed discrete-time nonlinear systems, *IEEE Trans. Aut. Control* **39**, 1994, 780-792.

3. C. D. Charalambous and R. J. Elliott, Certain nonlinear partially observable stochastic optimal control problems with explicit control laws equivalent to LEQG/LQG problems, *IEEE Trans. Aut. Control* **42**, 1997, 482-497.

4. S.P. Coraluppi and S.I. Marcus, Robust control of Markov decision processes and connection to risk sensitive control, *Proc. 34th Allerton Conf. Communication, Control and Computing*, October 1-4, 1996, 353-362.

5. E. Fernandez-Gaucherand, A. Arapostathis and S.I. Marcus, On the average cost optimality equation and the structure of optimal policies for partially observable Markov decision processes, *Annals of Operation Research* **29**, 1991, 471-512.

6. E. Fernández-Gaucherand, A. Arapostathis, and S.I. Marcus, "Analysis of an Adaptive Control Scheme for a Partially Observed Controlled Markov Chain," *IEEE Trans. Automatic Control* **38**, June 1993, 987-993.



(a) $\frac{R}{C} > 2$ (b) $\frac{R}{C} < 2$

Figure 3: Replacement percentage

7. E. Fernández-Gaucherand and S. I. Marcus, Risk-sensitive optimal control of hidden Markov models: structural results, *IEEE Transactions on Automatic Control* (to appear).

8. W. H. Fleming and D. Hernández-Hernández, Risk sensitive control of finite state machines on an infinite horizon II, Technical Report, Division of Applied Mathematics, Brown University.

9. W. H. Fleming and W. M. McEneaney, Risk-sensitive control and differential games, Springer Lecture Notes in Control and Info. Sci. No. 184, 1992, 185-197.

10. W. H. Fleming and W. M. McEneaney, Risk-sensitive control on an infinite horizon, *SIAM J. Control and Optim.* **33**, 1995, 1881-1915.

11. D. Hernández-Hernández and S.I. Marcus, "Risk-Sensitive Control of Markov Processes in Countable State Space, *Systems and Control Letters* **29**, November 1996, 147-155.

12. S.I. Marcus, E. Fernandez-Gaucherand, D. Hernandez-Hernandez, S. Coraluppi, and P. Fard, Risk Sensitive Markov Decision Processes, in *Systems and Control in the Twenty-First Century*, C. I. Byrnes, et. al. (eds.). Boston: Birkhauser, 1997, 263-279.

13. T. Runolfsson, The equivalence between infinite horizon control of stochastic systems with exponential-of-integral performance index and stochastic differential games, *IEEE Trans. Aut. Cont.* **39**, 1994, 1551-1563.