1. INTRODUCTION

We consider a risk-sensitive optimal control problem for hidden Markov models (HMM), i.e., controlled Markov chains where state information is only available to the controller via an output (message) process. The optimal control of HMM under standard, risk-neutral performance criteria, e.g., discounted and average costs, has received much attention in the past. Many basic results and numerous applications have been reported in the literature in this subject; see [ABFGM], [BE], [KV], and references therein. Controlled Markov chains with full state information and a risk-sensitive performance criterion have also received some attention, dating back at least to the work of Howard and Matheson [HOM]; see also [CSO].

On the other hand, quite the opposite is the situation for HMM under risk-sensitive criteria, e.g., expected value of the exponential of additive costs. Whittle and others (see [WHI]) have extensively studied the risk-sensitive optimal control of partially-observable linear exponential quadratic Gaussian (LEQG) systems. More recently, James, Baras and Elliott [JBE], [BJ], have treated the risk-sensitive partially-observable optimal control problem of discrete-time non-linear systems. An earlier attempt to study HMM with risk-sensitive criteria by Gheorghe [GHE] suffers from a basic conceptual flaw, as explained below.

The paucity of results in this subject area can be mostly attributed to the lack in the past of appropriate sufficient statistics, or information states. As is well known, if the cost criterion being considered is of the type "expected value of additive costs," then the posterior probability density, given all available information up to the present, constitutes a sufficient statistic for control (or information state); see [ABFGM], [BE], [KV]. The latter result was originally proved by Shiryaev in the early sixties, who also proved that this was not the case for non-additive cost criteria; see [SHI] and references therein. In particular, the posterior probability density is not a sufficient statistic for HMM under an "exponential of sum of costs" type of criterion, which is non-additive. This fact was overlooked in [GHE], thus invalidating the claims of optimality for the policies obtained in that paper.

Recently, James, Baras, and Elliott [BJ], [JBE] have derived information states for HMM under an "exponential of additive costs" criterion, and have given dynamic programming equations from which optimal values and controls can be computed, for problems with a finite horizon. Building upon the results by Baras, James and Elliott, we report in this paper initial results of an investigation on the nature and form of risk-sensitive controllers. Essentially, the question that we pose is:

How does risk-sensitivity manifest itself in a controller?

Whittle [WHI] has addressed a similar question for the LEQG problem, and he has shown that much insight can be gained from a comparison of the risk-neutral (i.e., the classical LQG) and risk sensitive equations describing the optimal controller. We initiate a similar investigation for HMM, via an examination of a popular benchmark problem. We examine the dynamic programming equations for both the risk-neutral and risk-sensitive cases and, among other results, compare the structure of optimal controllers thus obtained. Furthermore, we show that indeed the risk-sensitive controller and its corresponding information state converge to the known solutions for the risk-neutral situation, as the risk factor goes to zero.

The paper is organized as follows. In section 2 we recall the main results from [BJ]-[JBE] that will be needed for our developments. Section 3 presents some general results. Section 4 presents our case study.

---

Certainly, one can pose a well defined stochastic optimal control problem given any statistic. However, if the chosen statistic is not sufficient, then one cannot hope to obtain the "overall" optimal policy, except by serendipity.
2. THE CONTROLLED HIDDEN MARKOV MODEL

A controlled hidden Markov model is given by a five-tuple \( (X, Y, \mathbf{U}, \{P(u) : u \in \mathbf{U}\}, \{Q(u) : u \in \mathbf{U}\}) \), here \( X = \{1, 2, \ldots, N_X\} \) is the finite set of (internal) states, \( Y = \{1, 2, \ldots, N_Y\} \) is the set of observations (or messages), \( \mathbf{U} = \{1, 2, \ldots, N_U\} \) is the set of decisions (or controls). In addition, we have that \( P(u) := [p_{ij}(u)] \) is the \( N_X \times N_X \) state transition matrix, and \( Q(u) := [q_{xy}(u)] \) is the \( N_Y \times N_Y \) state/message matrix, i.e., \( q_{xy}(u) \) is the probability of receiving message \( y \) when the state is \( x \) and action \( u \) has been selected. In the operations research literature similar models are called partially observable Markov decision processes [FAM1], [FAM2]. Two types of information patterns are of interest.

Information Pattern 1 (IP1):

At decision epoch \( t \), the system is in the (unobserved) 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:

\[
Prob(Y_{t+1} = y | X_{t+1} = x, U_t = u) = q_{xy}(u). \tag{2.1.a}
\]

Hence, based on

\[
I_t^{(1)} := (Y_0, U_0, Y_1, \ldots, U_t, Y_{t+1}),
\]

a new decision \( U_{t+1} \) is selected.

Information Pattern 2 (IP2):

At decision epoch \( t \), the system is in the (unobserved) state \( X_t = i \), a decision \( U_t = u \) is taken, and an observation \( Y_{t+1} \) is gathered, such that:

\[
Prob(Y_{t+1} = y | X_t = i, U_t = u) = q_{xy}(u). \tag{2.1.b}
\]

The state then evolves to \( X_{t+1} = j \) with probability \( p_{ij}(u) \). Hence, based on

\[
I_t^{(2)} := (U_0, Y_1, U_1, Y_2, \ldots, U_t, Y_{t+1}),
\]

a new decision \( U_{t+1} \) is selected.

Hereafter we will simply write \( I_t \) and \( Y_t \) for a generic information pattern and the filtration generated by the available observations, respectively, up to decision epoch \( t \).

Given an expected cost per stage \((i, u) \rightarrow c(i, u)\), then the sum of costs for the finite horizon \( M \) is given by:

\[
C := \sum_{t=0}^{M-1} c(X_t, U_t). \tag{2.2}
\]

The risk-sensitive optimal control problem is that of finding a control policy \( \pi = \{\pi_0, \pi_1, \ldots, \pi_{M-1}\} \), with \( T_t \sim \pi_t(I_t) \in \mathbf{U} \), such that the following criterion is minimized:

\[
J^*(\pi, X_0) := \text{sgn}(\gamma)E^\pi[\exp(\gamma \cdot C)], \tag{2.3}
\]

where \( \gamma \neq 0 \) is the risk-factor, and \( \text{sgn}(\gamma) \) is the sign of \( \gamma \). By computing the Taylor series expansion of \( J^*(\pi, X_0) \) about \( E^\pi[C] \), when \( \gamma \) is sufficiently small, the risk sensitivity of the above criterion becomes evident in that, in addition to the standard expected sum of costs, a second order term in the expansion measures the variance \( \text{var}(C) \); see [WHI] for details. If \( \gamma > 0 \), then the controller is risk-averse or pessimistic, whereas if \( \gamma < 0 \) then the controller is risk-seeking or optimistic.

2.1 INFORMATION STATES

As for the risk-neutral case [ABFGM], [BE], [KV], an equivalent stochastic optimal control problem can be formulated in terms of information states and separated policies. Here we follow the work of Baras, James and Elliott [RBE]-[BI], who derived information states both for the continuous [RBE] and the discrete [BI] variables cases. First, we need to equivalently reformulate the stochastic control problem in terms of a canonical measure, as follows. Let \( \mathcal{F}_t \) be the filtration generated by the available observations up to decision epoch \( t \), and let \( \mathcal{G}_t \) be the filtration generated by the sequence of states and observations up to that time. Then the probability measure induced by a policy \( \pi \) is equivalent to a distribution \( P_t^\pi \), under which \( \{Y_t\} \) is independently and identically distributed (i.i.d), independent of \( \{X_t\} \), and the latter is a controlled Markov chain with transition matrix as above. We have that

\[
\frac{dP_t^\pi}{dP_t}\big|_{\mathcal{G}_t} = \lambda_t^\pi, \tag{2.4}
\]

where

\[
\lambda_t^\pi = \begin{cases} 
N_Y \cdot q_{XY}, & \mathcal{G}_t \text{ generated by (IP1)}; \\
N_Y \cdot q_{XY'}, & \mathcal{G}_t \text{ generated by (IP2)}. 
\end{cases}
\]

Then, the cost incurred by using the policy \( \pi \) is given by

\[
J^*(\pi, X_0) := \text{sgn}(\gamma)E_{X_0}^\pi[\exp(\gamma \cdot C)] = \text{sgn}(\gamma)E_{X_0}^\pi[\lambda_t^\pi \exp(\gamma \cdot C)]. \tag{2.5}
\]

Following [EM] and [JBE], the information state for our problem is given by

\[
\sigma^*_t(i) := E_{X_0}^\pi[1[X_t = i] \exp(\gamma \cdot C) | Y_t], \tag{2.6}
\]

where \( 1[A] \) is the indicator function of the event \( A \), and \( \sigma^*_t(i) = 1[X_0 = i] \). Notice that \( \sigma^*_t \in \mathbb{R}_{+}^{N_X} := \)
\{ \sigma \in \mathbb{R}^{N_X} \mid \sigma_i \geq 0, \forall i \} \). With this definition of information state, similar results as in the risk-neutral case can be obtained. In particular, one obtains a recursive updating formula for \( \sigma^T \), and dynamic programming equations for value functions giving necessary and sufficient optimality conditions for separated policies, i.e., maps \( \sigma^T \mapsto h(\sigma^T) \in U \); see [JBE], [BJ] for details.

3. SOME GENERAL RESULTS

As derived by James, Baras and Elliott [JBE], [BJ], recursive equations are obtained for the information state process \( \{ \sigma^T \} \) and the cost-to-go value functions \( J^T(\cdot, M - k) \), \( k = 0, 1, \ldots, M \), where \( J^T(\cdot, M - k) : \mathbb{R}^{N_X} \rightarrow \mathbb{R} \), and

\[
\min_{\pi} \{ J^T(\pi, X_0) \} = E^T \{ J^T(\sigma^T_M, M) \}. \tag{3.1}
\]

The information state process is driven by the output (observations) path, and evolves forward in time. The dynamic programming recursion for the value functions evolves backward in time, and determines the optimal (separated) control policy.

As in the completely observed case [HOM], define the 

**disutility contribution matrix** as:

\[
[D(u)_{i,j}]_i = p_{i,j}(u) \cdot \exp(\gamma(\gamma(i, u))). \tag{3.2}
\]

Then, the following lemma gives the recursions that govern the evolution of the information state [BJ], [JBE].

**Lemma 3.1:** The information state process \( \{ \sigma^T \} \) is recursively computable as:

\[
\sigma^T_{i+1} = \begin{cases} 
N_X \cdot \text{Q}(Y_{i+1}, U_i)D^T(U_i)\sigma^T_i & \text{for (IP1)}; \\
N_X \cdot D^T(U_i)\bar{Q}(Y_{i+1}, U_i)\sigma^T_i & \text{for (IP2)};
\end{cases} \tag{3.3}
\]

where \( \bar{Q}(y, u) := \text{diag}(q_{i,y})(u) \).

**Remark 3.1:** For risk-sensitive completely observed Markov chains with finite state and control sets, it is well known that the disutility matrix \( D(u) \) governs the evolution of the disutility [HOM]. On the other hand, for risk-neutral HMM models, the information state used is the conditional probability distribution of the unobservable state, given the available observations [ABFGM], [BE], [KV]. The unnormalized form of this conditional probability distribution is given by similar recursions as in (3.3), with \( D(u) \) replaced by \( P(u) \). Therefore, we see that (3.3) is the “natural” extrapolation of the standard risk-neutral information state.

For ease of presentation, we will hereafter consider exclusively (IP1). Simple modifications to the results to be presented would give the corresponding analogues for (IP2). Furthermore, we will denote by \( T(u, y) \) the matrix

\[
T(u, y) := N_Y \cdot \bar{Q}(y, u)D^T(u). \tag{3.4}
\]

The following result follows directly from [JBE]; see also [BJ].

**Lemma 3.2:** The dynamic programming equations for the value functions in this problem are given as:

\[
\begin{align*}
J^T(\sigma, M) &= \sum_{i=1}^{N_X} \sigma_i; \\
J^T(\sigma, M - k) &= \min_{u \in X} \left\{ E^T \left[ J^T(T(u, Y_{M-k+1})\sigma) \right] \right\}.
\end{align*} \tag{3.5}
\]

Furthermore, a separated policy \( \pi^* = \{ \pi^*_0, \ldots, \pi^*_{M-1} \} \) that attains the minimum in (3.5) is risk-sensitive optimal.

Recall that \( E^T[\cdot] \) is the expectation with respect to the canonical measure \( \mathcal{P}^T \), and thus for a given function \( f : Y \rightarrow \mathbb{R} \),

\[
E^T[f(Y_i)] = \frac{1}{N_Y} \sum_{y=1}^{N_Y} f(y). \tag{3.6}
\]

Next, we present some results for the risk sensitive case that have similar counterparts in the standard risk-neutral case; see [ABFGM], [BE], [FAM1], [KV], [SSO].

**Lemma 3.3:** The value functions given by (3.5) are concave functions of \( \sigma \in \mathbb{R}^{N_X} \).

**Proof:**

We proceed by induction in \( k \), with the case \( k = 0 \) being trivially verified from (3.5). Assume that the claim holds true for \( 0 \leq k - 1 < M \). Let \( 0 \leq \lambda \leq 1 \) and \( \sigma_1, \sigma_2 \in \mathbb{R}^{N_X} \), and define \( \tilde{\sigma} := \lambda \sigma_1 + (1 - \lambda)\sigma_2 \). Then we have that:

\[
J^T(\tilde{\sigma}, M - k)
\]

\[
= \min_{u \in X} \left\{ \frac{1}{N_Y} \sum_{y=1}^{N_Y} J^T(T(u, y) \cdot \tilde{\sigma}, M - k + 1) \right\}
\]

\[
\geq \min_{u \in X} \left\{ \frac{1}{N_Y} \sum_{y=1}^{N_Y} \left[ \lambda J^T(T(u, y) \cdot \sigma_1, M - k + 1) \right. \\
+ (1 - \lambda) J^T(T(u, y) \cdot \sigma_2, M - k + 1) \right. \\
\left. \left. + \lambda J^T(\sigma_1, M - k) + (1 - \lambda) J^T(\sigma_2, M - k) \right] \right\}
\]

(3.7)

and the result follows.

**Remark 3.2:** For the risk-neutral case, a similar result was initially pointed out by Shiryaev (see [SHII] and references therein), and shown in detail by Aström (see [BE] and references therein). This result
has played a key role in showing optimality of structured policies in the risk-neutral case; see [FAM1] and references therein.

Next, define recursively sets of vectors in \( \mathbb{R}^{N_U}_+ \) as follows:
\[
A_0 := \{ 1 = (1, 1, \ldots, 1)^T \},
\]
\[
A_k := \left\{ \frac{1}{N_Y} \sum_{y=1}^{N_Y} a_y^T \cdot T(u, y) : a_y \in A_{k-1}, u \in U \right\}.
\]

(3.8)

We have the following result, which has important computational implications.

**Lemma 3.4:** The value functions given by (3.5) are piecewise linear functions in \( \sigma \in \mathbb{R}^{N_U}_+ \), such that:
\[
J^*(\sigma, M - k) = \min_{\sigma \in A_k} \{ \sigma^T \cdot \sigma \}.
\]

(3.9)

**Proof:**

We proceed by induction in \( k \), with the case \( k = 0 \) being trivially verified from (3.5). Assume that the claim holds true for \( 0 \leq k - 1 < M \), then:
\[
J^*(\sigma, M - k) = \min_{\sigma \in A_k} \left\{ \frac{1}{N_Y} \sum_{y=1}^{N_Y} \min_{\alpha \in A_{k-1}} \{ \alpha^T \cdot T(u, y) \sigma \} \right\}
\]
\[
= \min_{\sigma \in A_k} \left\{ \frac{1}{N_Y} \sum_{y=1}^{N_Y} \alpha^T \cdot T(u, y) \sigma \right\}
\]
\[
= \min_{\alpha \in A_k} \{ \alpha^T \cdot \sigma \}
\]

(3.10)

where \( \alpha^T(u, y, \sigma) \) denotes a minimizer in the expression on the right of the first equality above. \( \square \)

4. A CASE STUDY

As a first step in our investigation, we have considered a popular benchmark problem for which much is known in the risk-neutral case. This is a two-state-replacement problem which models failure-prone units in production/manufacturing systems, communication systems, etc. The underlying state of the unit can either be working \( (X_t = 0) \) or failed \( (X_t = 1) \), and the available actions are to keep \( (U_t = 0) \) the current unit or replace \( (U_t = 1) \) the unit by a new one. The messages received have probability \( 1/2 < q < 1 \) of coinciding with the true state of the unit. The state transition matrices are given as:
\[
P(0) = \begin{bmatrix} 1 - \theta & 0 \\ \theta & 1 \end{bmatrix}, \quad P(1) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix};
\]

(4.1)

see [WCC], [FAM1], [FAM2] for more details. With the above definitions, the matrices used to update the information state vector are given by:
\[
T(0, 0) = \begin{bmatrix} q(1 - \theta) & 0 \\ (1 - q)\theta & q \end{bmatrix},
\]
\[
T(0, 1) = \begin{bmatrix} 1 - \theta & 0 \\ \theta & q \end{bmatrix},
\]
\[
T(1, y) = \begin{bmatrix} q e^{\gamma R} & q e^{\gamma R} \\ 0 & 0 \end{bmatrix}, \quad y = 0, 1.
\]

(4.2)

To gain insight into the problem, we proceed to perform a few steps of the dynamic programming recursions of (3.5), and then a more general result will be given. For ease of presentation, we consider the risk-averse case only \( (\gamma > 0) \); the risk-seeking case is treated similarly.

**STAGE \( k = 0 \).**

\[
J^*(\sigma, M) = \sigma_1 + \sigma_2.
\]

(4.3)

**STAGE \( k = 1 \).**

\[
J^*(\sigma, M - 1) = \min \left\{ \mathbb{E}^1 \left[ J^*_0 (\sigma, M - 1) \right] \right\};
\]

(4.4)

where \( J^*_0 \) and \( J^*_1 \) denote the terms corresponding to \( u = 0 \) and \( u = 1 \), respectively, i.e.:
\[
\mathbb{E}^1 \left[ J^*_0 (\sigma, M - 1) \right] = \sigma_1 + \sigma_2 + e^{\gamma C} \sigma_2;
\]
\[
\mathbb{E}^1 \left[ J^*_1 (\sigma, M - 1) \right] = e^{\gamma R} (\sigma_1 + \sigma_2) = e^{\gamma R} J^*(\sigma, M).
\]

(4.5)

Therefore, the region in \( \mathbb{R}^2_+ \) where it is optimal to keep the unit is defined by the condition:
\[
\mathbb{E}^1 \left[ J^*_0 (\sigma, M - 1) \right] \leq \mathbb{E}^1 \left[ J^*_1 (\sigma, M - 1) \right]
\]
\[
\iff \frac{1 - e^{\gamma R}}{e^{\gamma R} - e^{\gamma C}} \sigma_1 \leq \sigma_2.
\]

(4.6)

Thus, since \( \gamma > 0, R > C > 0 \) and \( \sigma_1 \geq 0 \), \( i = 1, 2 \), then (4.6) implies that it is always optimal to keep the unit, i.e., \( \sigma^*_M (\sigma) = 0, \forall \sigma \in \mathbb{R}^2_+ \).

**STAGE \( k = 2 \).**

We have that \( J^*(\cdot, M - 1) = J^*_0 (\cdot, M - 1) \). Therefore:
\[
\mathbb{E}^1 \left[ J^*_0 (\sigma, M - 2) \right] = \left[ 1 - e^{\gamma C} \right] \mathbb{E}^1 \left[ J^*_1 (\sigma, M - 2) \right]
\]
\[
= (1 - \theta) + e^{\gamma C} \mathbb{E}^1 \left[ J^*_0 (\sigma, M - 2) \right] + e^{2\gamma C} \sigma_2;
\]

(4.7.a)

\[
\mathbb{E}^1 \left[ J^*_1 (\sigma, M - 2) \right] = e^{\gamma R} (\sigma_1 + \sigma_2).
\]

(4.7.b)

Then the condition for it to be optimal to keep the unit is:
\[
\mathbb{E}^1 \left[ J^*_0 (\sigma, M - 2) \right] \leq \mathbb{E}^1 \left[ J^*_1 (\sigma, M - 2) \right]
\]

(4.8.a)
\[ (1 - \theta)(1 - e^{\gamma R}) + \theta(e^{\gamma C} - e^{\gamma R}) \sigma_1 \leq e^{\gamma R} - e^{\gamma C} \sigma_2. \]  
(4.8.b)

Therefore, it is always optimal to keep the unit, i.e., \( \pi^*_M = 0, \forall \theta \in \mathbb{R}_+^2 \), and only if \( R \geq 2C \). Otherwise, \( \mathbb{R}_+^2 \) is partitioned into two regions by a line, as described by:
\[ \frac{(1 - \theta)(e^{\gamma R} - 1) + \theta(e^{\gamma R} - e^{\gamma C})}{e^{\gamma R} - e^{\gamma C}} \sigma_1 \geq \sigma_2. \]  
(4.9)

Thus, the optimal policy is of the threshold type, such that it is optimal to keep the unit for \( \sigma \) values below and on the line determined by (4.9), and to replace the unit otherwise. We will refer to the regions in \( \mathbb{R}_+^2 \) thus defined as the keep and replace regions.

**Remark 4.1:** The above result, (4.9), is similar to the results for the risk-neutral case, where there is a threshold value \( 0 < \rho^*_M < 1 \) such that it is optimal to keep the unit if the conditional probability of the unit being failed is not greater than \( \rho^*_M \), and to replace the unit otherwise; see [FAM1], [WCC], and references therein. Furthermore, we see that as \( \gamma \to \infty \), i.e., as the DM becomes infinitely risk averse, the keep region tends to vanish. Thus, the DM tends to replace independently of the value of \( \sigma^*_M = 2 \). This can be explained in that if the decision to replace the unit is taken, then there is complete certainty of the outcome. Following the jargon of Whittle [WHI], it could be said that as \( \gamma \to \infty \) the DM turns "neurotic", since the optimal control policy becomes of the bang bang type: if \( R \geq 2C \), then \( \pi^*_M = 0 \), and otherwise \( \pi^*_M = 1 \).

**STAGE k = 3.**

First, notice that \( T(0, y) \sigma \) will lie on different rays through the origin, depending on the value of \( y \). Thus, if \( 2C > R \) one would need to consider four cases when evaluating the dynamic programming recursion: whether both \( T(0, 0) \sigma \) and \( T(0, 1) \sigma \) lie in the same control region, or in different ones. We omit this analysis here, and instead assume that \( R \geq 2C \), and therefore \( J_T(\sigma, M - 2) = J_T^1(\sigma, M - 2) \). Thus, we have:
\[
E^T[J_T^1(\sigma, M - 2)] = \left[ (1 - \theta) + \theta e^{\gamma C} \right] (1 - e^{\gamma R}) \sigma_1 + e^{\gamma R} \theta e^{\gamma C} \sigma_2
\]
\[= \left[ (1 - \theta)^2 + \theta(1 - \theta) e^{\gamma C} + \theta e^{\gamma C} \right] \sigma_1 + e^{\gamma R} \theta e^{\gamma C} \sigma_2. \]  
(4.10.a)

\[ E^T[J_T^2(\sigma, M - 3)] = e^{\gamma R} \left[ (1 - \theta) + \theta e^{\gamma C} \right] (1 - \theta) \sigma_1 + \theta e^{\gamma C} \sigma_2. \]  
(4.10.b)

Then, the condition for it to be optimal to keep the unit is:
\[ E^T[J_T^2(\sigma, M - 3)] \leq E^T[J_T^1(\sigma, M - 3)]. \]  
(4.11.a)

\[ [(1 - \theta)^2(1 - e^{\gamma R}) + \theta(1 - \theta)(e^{\gamma C} - e^{\gamma R}) + \theta(e^{\gamma C} - e^{\gamma (R + C)})] \sigma_1 \leq e^{\gamma R}(1 - \theta) + \theta e^{\gamma (R + C)} - e^{\gamma C} \sigma_2. \]  
(4.11.b)

Thus, the necessary and sufficient condition for it to be optimal to keep the unit for all \( \sigma \in \mathbb{R}_+^2 \) is:
\[ e^{\gamma C} - e^{\gamma R}(1 - \theta) + \theta e^{\gamma (R + C)} \sigma_1 \geq \sigma_2, \]  
(4.12)

and a sufficient condition for this is that \( R \geq 3C \). If (4.12) is not satisfied, then the optimal policy will be of the threshold type, with \( \mathbb{R}_+^2 \) being partitioned into keep and replace regions by a line, as defined by:
\[ NUM \left[ e^{\gamma C} - e^{\gamma R}(1 - \theta) + \theta e^{\gamma (R + C)} \right] \sigma_1 \geq \sigma_2, \]  
(4.13)

where \( NUM = (1 - \theta)^2(1 - e^{\gamma R}) + \theta(1 - \theta)(e^{\gamma C} - e^{\gamma R}) + \theta(e^{\gamma C} - e^{\gamma (R + C)} - e^{\gamma C}). \)

Studying closely the equations for the above cases, a recursive equation is found for the coefficients in (4.6), (4.8), (4.11), and also for the corresponding equations for subsequent cases. Define:
\[
\alpha_0 := 1; \]
\[
\alpha_1 := (1 - \theta)\alpha_0 + \theta e^{\gamma C}; \]
\[
\vdots \]
\[
\alpha_{k+1} := (1 - \theta)\alpha_k + \theta e^{\gamma C}; \quad k = 0, 1, \ldots, M. \]  
(4.14)

We will need the following simple result.

**Lemma 4.1:** \( \alpha_{k+1} > \alpha_k \), and \( e^{\gamma R}\alpha_k > \alpha_{k+1}; \quad k = 1, 2, \ldots, M. \)

**Proof:**

Note that \( e^{\gamma C} > 1 = \alpha_1 \). Proceeding by induction, suppose that \( e^{\gamma C} > \alpha_k \). Then, from (4.14), we have:
\[ e^{\gamma (k+1)R} = (1 - \theta)e^{\gamma (k+1)C} + \theta e^{\gamma (k+1)C} \]
\[> (1 - \theta)e^{\gamma C} + \theta e^{\gamma C} \]
\[> (1 - \theta)\alpha_k + \theta e^{\gamma C}, \]  
(4.15)

Then the first result follows since:
\[ \alpha_{k+1} - \alpha_k = \theta[e^{\gamma C} - \alpha_k]. \]  
(4.16)

On the other hand, we have that
\[ e^{\gamma R}\alpha_k = (1 - \theta)e^{\gamma R}\alpha_k + \theta e^{\gamma R}\alpha_k \]
\[> (1 - \theta)\alpha_k + \theta e^{\gamma C} = \alpha_{k+1}. \]  
(4.17)

\[ \square \]
Using the above results, we have the following.

**Theorem 4.1**: Let $0 < \overline{K} \leq M$ be given. The necessary and sufficient condition for the policy with $\pi_M() = \ldots = \pi_{M-\overline{K}}() = 0$ to be optimal is that:

$$\frac{e^{\gamma R}}{\sigma_{\overline{K}-1}} \leq e^{\gamma R} R \geq \overline{K} C - \frac{\ln(\alpha_K - 1)}{\gamma}. \quad (4.18)$$

Furthermore, a sufficient condition for the above is that $R > \overline{K} C$. If $1 < \overline{K} \leq M$ is the smallest integer for which (4.18) fails, then the optimal policy $\pi_{M-\overline{K}}$ is of the threshold type, with $R^2_\pi$ being partitioned by a line into *keep* and *replace* regions, as defined by:

$$\frac{e^{\gamma R} \alpha_{\overline{K}} - \alpha_{\overline{K}+1}}{e^{\gamma R} \alpha_{\overline{K}} - e^{\gamma R} \alpha_{\overline{K}}} \geq \sigma_2. \quad (4.19)$$

**Proof**: The proof is by induction, and follows similarly as the cases $k = 2, 3$ done previously. $\square$

**Remark 4.2**: How do the results in Theorem 4.1 compare to known results for the risk-neutral case? Similarly as in [WCC], the dynamic programming equations for this case can be written, with the conditional probability distribution of the state as the information state. Then, it can be shown that the optimal risk-neutral controller has a structure similar to the risk-sensitive controller given in Theorem 4.1. Furthermore, it can be shown that the necessary and sufficient condition in the risk-neutral case for the separated policy $\pi_M() = \ldots = \pi_{M-\overline{K}}() = 0$ to be optimal is:

$$R > \overline{K} C - \sigma_{\overline{K}-1}. \quad (4.20)$$

where $\sigma_{\overline{K}-1}$ is obtained as the derivative with respect to $\gamma$, evaluated as $\gamma \rightarrow 0$, of (4.14). As can be easily verified, the above is nothing but the small risk limit (i.e., as $\gamma \rightarrow 0$) of (4.18). Hence the risk-sensitive controller obtained here has as its small risk limit the known risk-neutral controller, and both controllers have in general a similar structure.

**ACKNOWLEDGEMENTS**

The research of the first author was partially supported by The Engineering Foundation under grant RI-A-93-10, and in part by the National Science Foundation under grant NSF-INT 9201430. The research of the second author was partially supported by the Air Force Office of Scientific Research under Grant F49620-92-J-0045, and in part by the National Science Foundation under Grant CDR-8803012.

**REFERENCES**


