Technical Research Report

Systems Research Center

Harvard

Supported by the National Science Foundation Engineering Research Center Program (NSFD CD 8803012), Industry and the University

Risk-Sensitive Control and Dynamic Games for Partially Observed Discrete - Time Nonlinear Systems

by M.R. James, J.S. Baras and R.J. Elliott

TR 92-124

Risk–Sensitive Control and Dynamic Games for Partially Observed Discrete–Time Nonlinear Systems

M.R. James* J.S. Baras† R.J. Elliott‡

Abstract

In this paper we solve a finite–horizon partially observed risk–sensitive stochastic optimal control problem for discrete–time nonlinear systems, and obtain small noise and small risk limits. The small noise limit is interpreted as a deterministic partially observed dynamic game, and new insights into the optimal solution of such game problems are obtained. Both the risk–sensitive stochastic control problem and the deterministic dynamic game problem are solved using information states, dynamic programming, and associated separated policies. A certainty equivalence principle is also discussed. Our results have implications for the nonlinear robust stabilization problem. The small risk limit is a standard partially observed risk–neutral stochastic optimal control problem.

Key words: Nonlinear partially observed stochastic systems, risk–sensitive optimal control, dynamical games, nonlinear filtering, large deviations, output feedback robust control.

1 Introduction

Recent interest in risk–sensitive stochastic control problems is due in part to connections with H_∞ or robust control problems and dynamic games. The solution of a risk–sensitive problem leads to a conservative optimal policy, corresponding to the controller's aversion to risk.

*Department of Systems Engineering and Cooperative Research Centre for Robust and Adaptive Systems, Australian National University, GPO Box 4, Canberra, ACT 2601, Australia.

†Martin Marietta Chair in Systems Engineering, Department of Electrical Engineering and Institute for Systems Research, University of Maryland, College Park, MD 20742, USA. Research supported in part by NSF Grant CDR—88—00108.

‡Department of Statistics and Applied Probability, University of Alberta, Edmunton, Alberta T6G 2G1, Canada.

For linear/quadratic risk–sensitive problems with full state information, Jacobson [15] established the connection with dynamic games. The analogous nonlinear problem was studied recently, and a dynamic game is obtained as a small noise limit (James [16], Fleming–McEneaney [10], Whittle [23], Campi–James [5]). A risk–neutral stochastic control problem obtains as a small risk limit (James [16], Campi–James [5]).

Whittle [22] solved the discrete-time linear/quadratic risk–sensitive stochastic control problem with incomplete state information, and characterized the solution in terms of a certainty equivalence principle. The analogous continuous–time problem was solved by Bensoussan–van Schuppen [4], where the problem was converted to an equivalent one with full state information. A conversion technique has also been used to solve partially observed linear/quadratic H_∞ and dynamic game problems (Basar–Bernhard [3], Doyle *et al* [6], Limebeer *et al* [20], Rhee–Speyer [21], and others). The nonlinear continuous–time partially observed risk–sensitive stochastic control problem was considered by Whittle [25], and an approximate solution was stated using a certainty equivalence principle when the noise is small; these results are not rigorous, but are very insightfull.

In this paper we consider the finite–horizon partially observed risk–sensitive stochastic control problem for discrete–time nonlinear systems. The solution to this problem together with large deviation limits lead to new insights into the optimal solution of partially observed dynamic game problems and related robust control problems.

The risk–sensitive stochastic control problem is solved by defining an information state and an associated value function, and applying dynamic programming (§2). The dynamic programming equation is a nonlinear infinite–dimensional recursion. Our approach is motivated by the method used by Bensoussan–van Schuppen [4], and the well–known separation method for risk–neutral problems. The resulting optimal controller is expressed in terms of a separated policy through the information state.

In §3 we obtain the small noise limit of both the information state and the value function. Logarithmic transformations are employed in each case. The information state limit is similar to large deviations limit results for nonlinear filters (Hijab [13], [14], Baras *et al* [1], James–Baras [18], James [17]), where the limit filter can be used as an observer for the limit deterministic system. The limit of the value function also satisfies a nonlinear infinite–dimensional recursion, which in §4 is interpreted as the dynamic programming equation for a deterministic partially observed dynamic game. As a by–product, we obtain an information state and value function for this game, and a verification theorem. The optimal output feedback controller is given by a separated policy through the information state. The information state, which depends on the output path, is a function of the state variable, and evolves forward in time. The value function is a function of the information state, evolves backwards in time, and determines the optimal control policy. The structure of the controller we obtain is similar to that arising in the solution of the linear/quadratic problem, which involves a pair of Riccati equations, one corresponding to estimation, and one to control (Rhee–Speyer [21]). We identify a certain saddle point condition under which the certainty equivalence policy proposed by Whittle [22], [25] and Basar–Bernhard [3] is optimal, using our verification theorem. This policy involves the forward dynamic programming recursion for the information state, and a backward recursion for the value function for the corresponding dynamic game problem with full state information. This latter value function is (like the information state) a function of the state variable, and consequently easier to compute.

The treatment of the robust nonlinear output feedback control problem using stochastic control formulations leads naturally to the "correct"feedback structure of an "observer"and "controller". Our approach leads directly to this structure through limiting processes which involve large deviation principles. The method clearly establishes the separation of the feedback policy and provides a framework for evaluating practical recipes. This correspondence and application to the discrete–time, nonlinear, robust, output feedback stabilization problem will be described in detail in a different publication [2].

Finally, the small risk limit is evaluated in §5, and shown to be a standard risk–neutral partially observed stochastic control problem. The notion of information state and the use of dynamic programming is well known for risk–neutral problems (e.g. Kumar–Varaiya [19], Elliott–Moore [9]).

2 The Risk–Sensitive Control Problem

2.1 Dynamics

On a probability space $(\Omega, \mathcal{F}, \mathbf{P}^u)$ we consider a risk–sensitive stochastic control problem for the discrete–time system

$$(2.1) \qquad \begin{cases} x_{k+1}^\varepsilon &= b(x_k^\varepsilon, u_k) + w_k^\varepsilon, \\ y_{k+1}^\varepsilon &= h(x_k^\varepsilon) + v_k^\varepsilon, \end{cases}$$

on the finite time interval $k = 0, 1, 2, \ldots M$. The process x^ε represents the state of the system, and is not directly measured. The process y^ε is measured, and is called the observation process. This observation process can be used to select the control actions u_k. We will write $x_{k,l}^\varepsilon$ for the sequence $x_k^\varepsilon, \cdots, x_l^\varepsilon$, etc. $\mathcal{G}_k$ and $\mathcal{Y}_k$ denote the complete filtrations generated by $(x_{0,k}^\varepsilon, y_{0,k}^\varepsilon)$ and $y_{0,k}^\varepsilon$ respectively.

We assume:

(i) x_0^ε has density $\rho(x) = (2\pi)^{-n/2} \exp(-\frac{1}{2}|x|^2)$.

(ii) $\{w_k^\varepsilon\}$ is an $\mathbf{R}^n$-valued iid noise sequence with density $\psi^\varepsilon(w) = (2\pi\varepsilon)^{-\frac{n}{2}} \exp(-\frac{1}{2\varepsilon}|w|^2)$.

(iii) $y_0^\varepsilon = 0$.

(iv) $\{v_k^\varepsilon\}$ is a real-valued iid noise sequence with density $\phi^\varepsilon(v) = (2\pi\varepsilon)^{-\frac{1}{2}} \exp(-\frac{1}{2\varepsilon}|v|^2)$, independent of x_0^ε and $\{w_k^\varepsilon\}$.

(v) $b \in C^1(\mathbf{R}^n \times \mathbf{R}^m, \mathbf{R}^n)$ is bounded and uniformly continuous.

(vi) The controls u_k take values in $U \subset \mathbf{R}^m$, assumed compact, and are $\mathcal{Y}_k$ measurable. We write $\mathcal{U}_{k,l}$ for the set of such control processes defined on the interval $k, \ldots, l$.

(vii) $h \in C(\mathbf{R}^n)$ is bounded and uniformly continuous.

The probability measure $\mathbf{P}^u$ can be defined in terms of an equivalent reference measure $\mathbf{P}^\dagger$ using the discrete analog of Girsanov's Theorem [9]. Under $\mathbf{P}^\dagger$, $\{y_k^\varepsilon\}$ is iid with density ϕ^ε, independent of $\{x_k^\varepsilon\}$, and x^ε satisfies the first equation in (2.1). For $u \in \mathcal{U}_{0,M-1}$,

$$\frac{d\mathbf{P}^u}{d\mathbf{P}^\dagger}|_{\mathcal{G}_k} \;=\; Z_k^\varepsilon \;=\; \Pi_{l=1}^k \Psi^\varepsilon(x_{l-1}^\varepsilon, y_l^\varepsilon),$$

where

$$\Psi^\varepsilon(x,y) \;\triangleq\; \exp\left(-\tfrac{1}{\varepsilon}\left[\tfrac{1}{2}|h(x)|^2 - h(x)y\right]\right).$$

2.2 Cost

The cost function is defined for admissible $u \in \mathcal{U}_{0,M-1}$ by

$$J^{\mu,\varepsilon}(u) \;=\; \mathbf{E}^u\left[\exp\frac{\mu}{\varepsilon}\left(\sum_{l=0}^{M-1} L(x_l^\varepsilon, u_l) + \Phi(x_M^\varepsilon)\right)\right], \tag{2.2}$$

and the *partially observed risk-sensitive stochastic control problem* is to find $u^* \in \mathcal{U}_{0,M-1}$ such that

$$J^{\mu,\varepsilon}(u^*) \;=\; \inf_{u \in \mathcal{U}_{0,M-1}} J^{\mu,\varepsilon}(u).$$

Here, we assume:

(viii) $L \in C(\mathbf{R}^n \times \mathbf{R}^m)$ is non–negative, bounded and uniformly continuous.

(ix) $\Phi \in C(\mathbf{R}^n)$ is non–negative, bounded, and uniformly continuous.

Remark 2.1 The assumptions (i) through (ix) are stronger than necessary. For example, boundedness assumption for b can be replace by a linear growth condition. In addition, a "diffusion"coefficient can be inserted into equation (2.1). Other choices for the initial density ρ are possible; see Remark 4.1.

The parameters $\mu > 0$ and $\varepsilon > 0$ are measures of risk sensitivity and noise variance. In view of our assumptions, the cost function is finite for all $\mu > 0$, $\varepsilon > 0$.

In terms of the reference measure, the cost can be expressed as

$$J^{\mu,\varepsilon}(u) \;=\; \mathbf{E}^\dagger\left[Z_M^\varepsilon \exp\frac{\mu}{\varepsilon}\left(\sum_{l=0}^{M-1} L(x_l^\varepsilon, u_l) + \Phi(x_M^\varepsilon)\right)\right]. \tag{2.3}$$

2.3 Information State

We consider the space $L^\infty(\mathbf{R}^n)$ and its dual $L^{\infty *}(\mathbf{R}^n)$, which includes $L^1(\mathbf{R}^n)$. We will denote the natural bilinear pairing between $L^\infty(\mathbf{R}^n)$ and $L^{\infty *}(\mathbf{R}^n)$ by $\langle \tau, \nu \rangle$ for $\tau \in L^{\infty *}(\mathbf{R}^n)$, $\nu \in L^\infty(\mathbf{R}^n)$. In particular, for $\sigma \in L^1(\mathbf{R}^n)$ and $\nu \in L^\infty(\mathbf{R}^n)$ we have

$$\langle \sigma, \nu \rangle = \int_{\mathbf{R}^n} \sigma(x)\nu(x)\, dx.$$

We now define an information state process $\sigma_k^{\mu,\varepsilon} \in L^{\infty *}(\mathbf{R}^n)$ by

$$\langle \sigma_k^{\mu,\varepsilon}, \eta \rangle = \mathbf{E}^\dagger \left[\eta(x_k^\varepsilon) \exp\left(\frac{\mu}{\varepsilon} \sum_{l=0}^{k-1} L(x_l^\varepsilon, u_l) \right) Z_k^\varepsilon \,|\, \mathcal{Y}_k \right] \tag{2.4}$$

for all test functions η in $L^\infty(\mathbf{R}^n)$, for $k = 1, \ldots, M$ and $\sigma_0^{\mu,\varepsilon} = \rho \in L^1(\mathbf{R}^n)$. We introduce the bounded linear operator $\Sigma^{\mu,\varepsilon} : L^\infty(\mathbf{R}^n) \to L^\infty(\mathbf{R}^n)$ defined by

$$\Sigma^{\mu,\varepsilon}(u,y)\nu\,(\xi) \triangleq \int_{\mathbf{R}^n} \psi^\varepsilon(z - b(\xi,u))\nu(z)\, dz \exp(\frac{\mu}{\varepsilon} L(\xi,u))\Psi^\varepsilon(\xi,y). \tag{2.5}$$

The bounded linear operator $\Sigma^{\mu,\varepsilon *} : L^{\infty *}(\mathbf{R}^n) \to L^{\infty *}(\mathbf{R}^n)$ adjoint to $\Sigma^{\mu,\varepsilon}$ is defined by

$$\langle \Sigma^{\mu,\varepsilon *}\tau, \eta \rangle = \langle \tau, \Sigma^{\mu,\varepsilon}\eta \rangle$$

for all $\tau \in L^{\infty *}(\mathbf{R}^n)$, $\eta \in L^\infty(\mathbf{R}^n)$.

The following theorem establishes that $\sigma_k^{\mu,\varepsilon}$ is in $L^1(\mathbf{R}^n)$ and its evolution is governed by the operator $\Sigma^{\mu,\varepsilon *}$, and for $\sigma \in L^1(\mathbf{R}^n)$, $\eta \in L^\infty(\mathbf{R}^n)$, we have

$$\Sigma^{\mu,\varepsilon *}(u,y)\sigma\,(z) = \int_{\mathbf{R}^n} \psi^\varepsilon(z - b(\xi,u)) \exp(\frac{\mu}{\varepsilon} L(\xi,u))\Psi^\varepsilon(\xi,y)\sigma(\xi)\, d\xi. \tag{2.6}$$

Theorem 2.2 *The information state $\sigma_k^{\mu,\varepsilon}$ satisfies the recursion*

$$\begin{cases} \sigma_k^{\mu,\varepsilon} = \Sigma^{\mu,\varepsilon *}(u_{k-1}, y_k^\varepsilon)\, \sigma_{k-1}^{\mu,\varepsilon} \\ \sigma_0^{\mu,\varepsilon} = \rho. \end{cases} \tag{2.7}$$

Further, $\sigma_k^{\mu,\varepsilon} \in L^1(\mathbf{R}^n)$ since $\rho \in L^1(\mathbf{R}^n)$ and $\Sigma^{\mu,\varepsilon *}$ maps $L^1(\mathbf{R}^n)$ into $L^1(\mathbf{R}^n)$.

PROOF. We follow Elliott–Moore [9]. ¿From (2.4), we have

$$\begin{aligned}
&\langle \sigma_k^{\mu,\varepsilon}, \eta\rangle \\
&= E^{\dagger}\Big[\eta(b(x_{k-1}^{\varepsilon}, u_{k-1}) + w_{k-1}^{\varepsilon}) \exp \tfrac{\mu}{\varepsilon} L(x_{k-1}^{\varepsilon}, u_{k-1}) \Psi^{\varepsilon}(x_{k-1}^{\varepsilon}, y_k^{\varepsilon}) \\
&\qquad \cdot \exp \tfrac{\mu}{\varepsilon} \textstyle\sum_{l=0}^{k-2} L(x_l^{\varepsilon}, u_l) Z_{k-1}^{\varepsilon} | \mathcal{Y}_k\Big] \\
&= E^{\dagger}\Big[\textstyle\int_{\mathbf{R}^n} \eta(b(x_{k-1}^{\varepsilon}, u_{k-1}) + w) \exp \tfrac{\mu}{\varepsilon} L(x_{k-1}^{\varepsilon}, u_{k-1}) \Psi^{\varepsilon}(x_{k-1}^{\varepsilon}, y_k^{\varepsilon}) \\
&\qquad \cdot \exp \tfrac{\mu}{\varepsilon} \textstyle\sum_{l=0}^{k-2} L(x_l^{\varepsilon}, u_l) Z_{k-1}^{\varepsilon} \psi^{\varepsilon}(w)\, dw\, | \mathcal{Y}_k\Big] \\
&= \langle \sigma_{k-1}^{\mu,\varepsilon}, \textstyle\int_{\mathbf{R}^n} \eta(b(\cdot\,, u_{k-1}) + w) \exp \tfrac{\mu}{\varepsilon} L(\cdot\,, u_{k-1}) \Psi^{\varepsilon}(\cdot\,, y_k^{\varepsilon}) \psi^{\varepsilon}(w)\, dw\rangle \\
&= \langle \sigma_{k-1}^{\mu,\varepsilon}, \Sigma^{\mu,\varepsilon}(u_{k-1}, y_k^{\varepsilon})\eta\rangle \\
&= \langle \Sigma^{\mu,\varepsilon\,*}(u_{k-1}, y_k^{\varepsilon})\sigma_{k-1}^{\mu,\varepsilon}, \eta\rangle.
\end{aligned}$$

This holds for all η in $L^{\infty}(\mathbf{R}^n)$; hence (2.7).

The fact that $\Sigma^{\mu,\varepsilon\,*}$ maps $L^1(\mathbf{R}^n)$ into $L^1(\mathbf{R}^n)$ follows easily from (2.6) and the properties of ψ^{ε}, Ψ^{ε}, and L. □

Remark 2.3 When $L \equiv 0$, the recursion (2.7) reduces to the Duncan–Mortensen–Zakai equation for the unnormalized conditional density [19].

The operator $\Sigma^{\mu,\varepsilon}$ actually maps $C_b(\mathbf{R}^n)$ into $C_b(\mathbf{R}^n)$. Then we can define a process $\nu_k^{\mu,\varepsilon} \in C_b(\mathbf{R}^n)$ by

$$(2.8) \qquad \left\{ \begin{array}{rcl} \nu_{k-1}^{\mu,\varepsilon} &=& \Sigma^{\mu,\varepsilon}(u_{k-1}, y_k^{\varepsilon})\, \nu_k^{\mu,\varepsilon} \\ \nu_M^{\mu,\varepsilon} &=& \exp \frac{\mu}{\varepsilon}\Phi. \end{array} \right.$$

It is straightforward to establish the adjoint relationships

$$(2.9) \qquad \begin{array}{rcl} \langle \Sigma^{\mu,\varepsilon\,*}\sigma, \nu\rangle &=& \langle \sigma, \Sigma^{\mu,\varepsilon}\nu\rangle, \\ \langle \sigma_k^{\mu,\varepsilon}, \nu_k^{\mu,\varepsilon}\rangle &=& \langle \sigma_{k-1}^{\mu,\varepsilon}, \nu_{k-1}^{\mu,\varepsilon}\rangle \end{array}$$

for all $\sigma \in L^1(\mathbf{R}^n)$, $\nu \in C_b(\mathbf{R}^n)$, and all k.

2.4 Alternate Representation of the Cost

Following [4], we define for $u \in \mathcal{U}_{0,M-1}$

$$(2.10) \qquad K^{\mu,\varepsilon}(u) = \mathbf{E}^{\dagger}\left[\langle \sigma_M^{\mu,\varepsilon}, \exp \frac{\mu}{\varepsilon}\Phi\rangle\right],$$

a cost function associated with the new "state"process $\sigma_k^{\mu,\varepsilon}$.

Theorem 2.4 *We have for all $u \in \mathcal{U}_{0,M-1}$*

$$J^{\mu,\varepsilon}(u) = K^{\mu,\varepsilon}(u). \tag{2.11}$$

PROOF. By (2.4),

$$\begin{aligned} K^{\mu,\varepsilon}(u) &= \mathbf{E}^{\dagger}\left[\mathbf{E}^{\dagger}\left[\exp \tfrac{\mu}{\varepsilon}\Phi(x^{\varepsilon}_{M}) \exp \tfrac{\mu}{\varepsilon}\left(\textstyle\sum_{l=0}^{M-1} L(x^{\varepsilon}_{l}, u_l)\right) Z^{\varepsilon}_{M} \,|\, \mathcal{Y}_M\right]\right] \\ &= \mathbf{E}^{\dagger}\left[\exp \tfrac{\mu}{\varepsilon}\left(\textstyle\sum_{l=0}^{M-1} L(x^{\varepsilon}_{l}, u_l) + \Phi(x^{\varepsilon}_{M})\right) Z^{\varepsilon}_{M}\right] \\ &= J^{\mu,\varepsilon}(u) \end{aligned}$$

using (2.3). □

We now define an alternate but equivalent stochastic control problem with complete state information. Under the measure $\mathbf{P}^{u}$, consider the state process $\sigma^{\mu,\varepsilon}_{k}$ governed by (2.7) and the cost $K^{\mu,\varepsilon}(u)$ given by (2.10). The new problem is to find $u^{*} \in \mathcal{U}_{0,M-1}$ minimizing $K^{\mu,\varepsilon}$.

Let $\mathcal{U}^{s}_{k,l}$ denote the set of control processes defined on the interval $k, \ldots, l$ which are adapted to $\sigma(\sigma^{\mu,\varepsilon}_{j},\ k \leq j \leq l)$. Such policies are called *separated* [19].

2.5 Dynamic Programming

The alternate stochastic control problem can be solved using dynamic programming. Consider now the state $\sigma^{\mu,\varepsilon}$ on the interval $k, \ldots, M$ with initial condition $\sigma^{\mu,\varepsilon}_{k} = \sigma \in L^{1}(\mathbf{R}^{n})$:

$$\begin{cases} \sigma^{\mu,\varepsilon}_{l} = \Sigma^{\mu,\varepsilon\,*}(u_{l-1}, y^{\varepsilon}_{l})\, \sigma^{\mu,\varepsilon}_{l-1}, & k+1 \leq l \leq M, \\ \sigma^{\mu,\varepsilon}_{k} = \sigma. \end{cases} \tag{2.12}$$

The corresponding value function for this control problem is defined for $\sigma \in L^{1}(\mathbf{R}^{n})$ by

$$S^{\mu,\varepsilon}(\sigma, k) = \inf_{u \in \mathcal{U}_{k,M-1}} \mathbf{E}^{\dagger}\left[\langle \sigma^{\mu,\varepsilon}_{k},\, \nu^{\mu,\varepsilon}_{k}\rangle \,|\, \sigma^{\mu,\varepsilon}_{k} = \sigma\right]. \tag{2.13}$$

Note that this function is expressed in terms of the adjoint process $\nu^{\mu,\varepsilon}_{k}$, given by (2.8).

Theorem 2.5 (Dynamic programming equation.) *The value function $S^{\mu,\varepsilon}$ satisfies the recursion*

$$\begin{cases} S^{\mu,\varepsilon}(\sigma, k) = \inf_{u \in U} \mathbf{E}^{\dagger}\left[S^{\mu,\varepsilon}(\Sigma^{\mu,\varepsilon\,*}(u, y^{\varepsilon}_{k+1})\sigma, k+1)\right] \\ S^{\mu,\varepsilon}(\sigma, M) = \langle \sigma, \exp \tfrac{\mu}{\varepsilon}\Phi\rangle. \end{cases} \tag{2.14}$$

PROOF. The proof is similar to Elliott–Moore [9], Theorem 4.5.

$$\begin{aligned}
S^{\mu,\varepsilon}(\sigma,k) &= \inf_{u\in\mathcal{U}_{k,k}} \inf_{v\in\mathcal{U}_{k+1,M-1}} \mathbf{E}^{\dagger}\left[\langle \sigma_k^{\mu,\varepsilon},\, \Sigma^{\mu,\varepsilon}(u_k, y_{k+1}^{\varepsilon})\nu_{k+1}^{\mu,\varepsilon}\rangle \mid \sigma_k^{\mu,\varepsilon}=\sigma\right] \\
&= \inf_{u\in\mathcal{U}_{k,k}} \inf_{v\in\mathcal{U}_{k+1,M-1}} \mathbf{E}^{\dagger}\left[\mathbf{E}^{\dagger}\left[\langle \Sigma^{\mu,\varepsilon\,*}(u_k, y_{k+1}^{\varepsilon})\sigma_k^{\mu,\varepsilon},\, \nu_{k+1}^{\mu,\varepsilon}\rangle \mid \mathcal{Y}_{k+1}\right] \mid \sigma_k^{\mu,\varepsilon}=\sigma\right] \\
&= \inf_{u\in\mathcal{U}_{k,k}} \mathbf{E}^{\dagger}\left[\inf_{v\in\mathcal{U}_{k+1,M-1}} \mathbf{E}^{\dagger}\left[\langle \Sigma^{\mu,\varepsilon\,*}(u_k, y_{k+1}^{\varepsilon})\sigma_k^{\mu,\varepsilon},\, \nu_{k+1}^{\mu,\varepsilon}\rangle \mid \mathcal{Y}_{k+1}\right] \mid \sigma_k^{\mu,\varepsilon}=\sigma\right] \\
&= \inf_{u\in\mathcal{U}_{k,k}} \mathbf{E}^{\dagger}\left[\inf_{v\in\mathcal{U}_{k+1,M-1}} \mathbf{E}^{\dagger}\left[\langle \sigma_{k+1}^{\mu,\varepsilon},\, \nu_{k+1}^{\mu,\varepsilon}\rangle \mid \sigma_{k+1}^{\mu,\varepsilon}=\Sigma^{\mu,\varepsilon\,*}(u_k, y_{k+1}^{\varepsilon})\sigma\right]\right] \\
&= \inf_{u\in\mathcal{U}_{k,k}} \mathbf{E}^{\dagger}\left[S^{\mu,\varepsilon}(\Sigma^{\mu,\varepsilon\,*}(u_k, y_{k+1}^{\varepsilon})\sigma, k+1)\right]
\end{aligned}$$

The interchange of minimization and conditional expectation is justified because of the lattice property of the set of controls [7], Chapter 16. □

Theorem 2.6 (Verification.) *Suppose that $u^* \in \mathcal{U}^s_{0,M-1}$ is a policy such that, for each $k=0,\ldots,M-1$, $u_k^* = \bar{u}_k^*(\sigma_k^{\mu,\varepsilon})$, where $\bar{u}_k^*(\sigma)$ achieves the minimum in (2.14). Then $u^*\in\mathcal{U}_{0,M-1}$ and is an optimal policy for the partially observed risk–sensitive stochastic control problem (§2.2).*

PROOF. We follow the proof of Elliott–Moore [9], Theorem 4.7. Define

$$\bar{S}^{\mu,\varepsilon}(\sigma,k;u) \;=\; \mathbf{E}^{\dagger}\left[\langle \sigma_k^{\mu,\varepsilon},\, \nu_k^{\mu,\varepsilon}\rangle \mid \sigma_k^{\mu,\varepsilon}=\sigma\right].$$

We claim that

$$S^{\mu,\varepsilon}(\sigma,k) \;=\; \bar{S}^{\mu,\varepsilon}(\sigma,k;u^*) \tag{2.15}$$

for each $k=0,1,\ldots,M$.

For $k=M$, (2.15) is clearly satisfied. Assume now that (2.15) holds for $k+1,\ldots,M$. Then

$$\begin{aligned}
\bar{S}^{\mu,\varepsilon}(\sigma,k;u^*) &= \mathbf{E}^{\dagger}\left[\mathbf{E}^{\dagger}\left[\langle \Sigma^{\mu,\varepsilon\,*}(u_k^*, y_{k+1}^{\varepsilon})\sigma_k^{\mu,\varepsilon},\, \nu_{k+1}^{\mu,\varepsilon}(u_{k+1,M-1}^*)\rangle \mid \mathcal{Y}_{k+1}\right] \mid \sigma_k^{\mu,\varepsilon}=\sigma\right] \\
&= \mathbf{E}^{\dagger}\left[\bar{S}^{\mu,\varepsilon}(\Sigma^{\mu,\varepsilon\,*}(u_k^*, y_{k+1}^{\varepsilon})\sigma, k+1; u_{k+1,M-1}^*)\right] \\
&= \mathbf{E}^{\dagger}\left[S^{\mu,\varepsilon}(\Sigma^{\mu,\varepsilon\,*}(u_k^*, y_{k+1}^{\varepsilon})\sigma, k+1)\right] \\
&= S^{\mu,\varepsilon}(\sigma,k).
\end{aligned}$$

from (2.14). This proves (2.15).

¿From (2.15), setting $k=0$ and $\sigma=\rho\in L^1(\mathbf{R}^n)$ we obtain

$$\bar{S}^{\mu,\varepsilon}(\rho,0;u^*) \;=\; S^{\mu,\varepsilon}(\rho,0) \;\le\; \bar{S}^{\mu,\varepsilon}(\rho,0;u)$$

for any $u \in \mathcal{U}_{0,M-1}$. Comparing (2.10) and the definitions of $S^{\mu,\varepsilon}$, $\bar{S}^{\mu,\varepsilon}$ above this implies

$$K^{\mu,\varepsilon}(u^*) \leq K^{\mu,\varepsilon}(u)$$

for all $u \in \mathcal{U}_{0,M-1}$. Using Theorem 2.4, we complete the proof. $\square$

Remark 2.7 The significance of Theorem 2.6 is that it establishes the optimal policy of the risk–sensitive stochastic control problem as a *separated* policy through the process $\sigma_k^{\mu,\varepsilon}$, defined in §2.3, which serves as an "information state"[19].

3 Small Noise Limit

In order to obtain a limit variational problem as $\varepsilon \to 0$, we must first obtain limit results for the information state (and its dual).

3.1 Information State

For $\gamma \in G \triangleq \{\gamma \in \mathbf{R}^2 : \gamma_1 > 0, \gamma_2 \geq 0\}$ define

$$\mathcal{D}^\gamma \triangleq \left\{p \in C(\mathbf{R}^n) : p(x) \leq -\gamma_1|x|^2 + \gamma_2\right\},$$

$$\mathcal{D} \triangleq \left\{p \in C(\mathbf{R}^n) : p(x) \leq -\gamma_1|x|^2 + \gamma_2 \text{ for some } \gamma \in G\right\},$$

and write

$$C_b(\mathbf{R}^n) \triangleq \{q \in C(\mathbf{R}^n) : |q(x)| \leq C, \quad \text{for some} \quad C \geq 0\}.$$

We equip these spaces with the topology of uniform convergence on compact subsets. In the sequel, $B(x, \alpha) \subset \mathbf{R}^p$ denotes the open ball centered at $x \in \mathbf{R}^p$ of radius $\alpha > 0$.

The "sup pairing"

$$(p,\, q) \triangleq \sup_{x \in \mathbf{R}^n} \{p(x) + q(x)\}, \tag{3.1}$$

is defined for $p \in \mathcal{D}$, $q \in C_b(\mathbf{R}^n)$, and arises naturally in view of the Varadhan–Laplace lemma (see the Appendix):

$$\lim_{\varepsilon \to 0} \frac{\varepsilon}{\mu} \log\langle e^{\frac{\mu}{\varepsilon}p},\, e^{\frac{\mu}{\varepsilon}q}\rangle = (p,\, q) \tag{3.2}$$

(uniformly on compact subsets of $\mathcal{D}^\gamma \times C_b(\mathbf{R}^n)$, for each $\gamma \in G$).

Define operators $\Lambda^{\mu *} : \mathcal{D} \to \mathcal{D}$, and $\Lambda^\mu : C_b(\mathbf{R}^n) \to C_b(\mathbf{R}^n)$ by

$$\begin{aligned} \Lambda^{\mu *}(u,y)p(z) &\triangleq \sup\nolimits_{\xi \in \mathbf{R}^n} \left\{L(\xi,u) - \tfrac{1}{2\mu}|z - b(\xi,u)|^2 - \tfrac{1}{\mu}\left[\tfrac{1}{2}|h(\xi)|^2 - h(\xi)y\right] + p(\xi)\right\}, \\ \Lambda^{\mu}(u,y)q(\xi) &\triangleq \sup\nolimits_{z \in \mathbf{R}^n} \left\{-\tfrac{1}{2\mu}|z - b(\xi,u)|^2 + q(z)\right\} + L(\xi,u) - \tfrac{1}{\mu}\left[\tfrac{1}{2}|h(\xi)|^2 - h(\xi)y\right]. \end{aligned} \tag{3.3}$$

With respect to the "sup pairing"$(\cdot\,,\,\cdot)$, these operators satisfy:

$$(\Lambda^{\mu *} p,\, q) \;=\; (p,\, \Lambda^{\mu} q)\,. \tag{3.4}$$

Also, $\Lambda^{\mu *}(u,y) \,:\, \mathcal{D}^{\gamma} \to \mathcal{D}$ is continuous for each $\gamma \in G$; in fact, the map $(u,y,p) \mapsto \Lambda^{\mu *}(u,y)p$, $U \times \mathbf{R} \times \mathcal{D}^{\gamma} \to \mathcal{D}$ is continuous.

The next theorem is a logarithmic limit result for the information state and its dual, stated in terms of operators (i.e. semigroups).

Theorem 3.1 *We have*

$$\begin{aligned} \lim_{\varepsilon \to 0} \tfrac{\varepsilon}{\mu} \log \Sigma^{\mu,\varepsilon *}(u,y) e^{\frac{\mu}{\varepsilon} p} &= \Lambda^{\mu *}(u,y)p, \\ \lim_{\varepsilon \to 0} \tfrac{\varepsilon}{\mu} \log \Sigma^{\mu,\varepsilon}(u,y) e^{\frac{\mu}{\varepsilon} q} &= \Lambda^{\mu}(u,y)q \end{aligned} \tag{3.5}$$

in $\mathcal{D}$ uniformly on compact subsets of $U \times \mathbf{R} \times \mathcal{D}^{\gamma}$ for each $\gamma \in G$, respectively in $C_b(\mathbf{R}^n)$ uniformly on compact subsets of $U \times \mathbf{R} \times C_b(\mathbf{R}^n)$.

PROOF. From (2.6), we have

$$\begin{aligned} \tfrac{\varepsilon}{\mu} \log \Sigma^{\mu,\varepsilon *}(u,y) e^{\frac{\mu}{\varepsilon} p}(z) \;=\; & \tfrac{\varepsilon}{\mu} \log \textstyle\int_{\mathbf{R}^n} \exp \tfrac{\mu}{\varepsilon} \Big(-\tfrac{1}{2\mu} |z - b(\xi,u)|^2 - \tfrac{n\varepsilon}{2\mu} \log(2\pi\varepsilon) \\ & -\tfrac{1}{\mu} \left[\tfrac{1}{2} |h(\xi)|^2 - h(\xi)y \right] + L(\xi,u) + p(\xi) \Big)\, d\xi \end{aligned}$$

Therefore,

$$\begin{aligned} & \lim_{\varepsilon \to 0} \tfrac{\varepsilon}{\mu} \log \Sigma^{\mu,\varepsilon *}(u,y) e^{\frac{\mu}{\varepsilon} p}(z) \\ &= \sup_{\xi \in \mathbf{R}^n} \left\{ L(\xi,u) - \tfrac{1}{2\mu} |z - b(\xi,u)|^2 - \tfrac{1}{\mu} \left[\tfrac{1}{2} |h(\xi)|^2 - h(\xi)y \right] + p(\xi) \right\} \\ &= \Lambda^{\mu *}(u,y)p(z) \end{aligned}$$

uniformly in $a = (z,u,y,p) \in A$, by Lemma 6.1; where $A = B(0,R) \times U \times B(0,R) \times K$, and $K \subset \mathcal{D}^{\gamma}$ is compact. This proves the first part of (3.5). The second part is proven similarly. □

3.2 Risk-Sensitive Value Function

The next theorem evaluates the small noise limit of the risk-sensitive value function $S^{\mu,\varepsilon}$. It involves two large deviations type limits, one corresponding to estimation and one to control.

Theorem 3.2 *The function $W^\mu(p,k)$ defined for $p \in \mathcal{D}$ by*

$$W^\mu(p,k) \ \triangleq\ \lim_{\varepsilon\to 0} \frac{\varepsilon}{\mu} \log S^{\mu,\varepsilon}(e^{\frac{\mu}{\varepsilon}p}, k) \tag{3.6}$$

exists (i.e. the sequence converges uniformly on compact subsets of $\mathcal{D}^\gamma$ ($\gamma \in G$)), is continuous on $\mathcal{D}^\gamma$ ($\gamma \in G$), and satisfies the recursion

$$\left\{ \begin{array}{rcl} W^\mu(p,k) &=& \inf_{u\in U} \sup_{y\in \mathbf{R}} \left\{ W^\mu(\Lambda^{\mu *}(u,y)p, k+1) - \frac{1}{2\mu}|y|^2 \right\} \\ W^\mu(p,M) &=& (p,\,\Phi)\,. \end{array} \right. \tag{3.7}$$

Proof. The result is clearly true for $k = M$ because of the second of (2.14), (3.2), and the continuity of $p \mapsto (p,\,\Phi)$ on each $\mathcal{D}^\gamma$.

Assume the conclusions hold for $k+1, \ldots, M$. Select $\gamma \in G$ and $K \subset \mathcal{D}^\gamma$ compact. In what follows $C > 0$, etc, will denote a universal constant. ¿From Theorem 2.5 and (3.6), we need to compute

$$\begin{array}{rcl} \lim_{\varepsilon\to 0} \frac{\varepsilon}{\mu} \log S^{\mu,\varepsilon}(e^{\frac{\mu}{\varepsilon}p}, k) &=& \lim_{\varepsilon\to 0} \frac{\varepsilon}{\mu} \log \inf_{u\in U} \mathbf{E}^\dagger \left[S^{\mu,\varepsilon}(\Sigma^{\mu,\varepsilon *}(u, y^\varepsilon_{k+1}) e^{\frac{\mu}{\varepsilon}p}, k+1) \right] \\ &=& \lim_{\varepsilon\to 0} \inf_{u\in U} \frac{\varepsilon}{\mu} \log \mathbf{E}^\dagger \left[S^{\mu,\varepsilon}(\Sigma^{\mu,\varepsilon *}(u, y^\varepsilon_{k+1}) e^{\frac{\mu}{\varepsilon}p}, k+1) \right]. \end{array} \tag{3.8}$$

The last equality is due to the monotonicity of the logarithm function.

Direct calculation verifies the estimate

$$\frac{\varepsilon}{\mu} \log S^{\mu,\varepsilon}(\Sigma^{\mu,\varepsilon *}(u,y) e^{\frac{\mu}{\varepsilon}p}, k+1) \ \le\ C(1+|y|)$$

for all $u \in U$, $y \in \mathbf{R}$, $p \in K$, $\varepsilon < \varepsilon'$ for some $\varepsilon' > 0$, and the inclusion

$$\frac{\varepsilon}{\mu} \log \Sigma^{\mu,\varepsilon *}(u,y) e^{\frac{\mu}{\varepsilon}p} \ \in \mathcal{D}^{\gamma(|y|)}$$

for all $u \in U$, $y \in \mathbf{R}$, $p \in K$, $\varepsilon < \varepsilon'$, for some $\gamma(|y|) \in G$. The fact that $\gamma(|y|)$ depends on $|y|$ complicates matters a little. If we select $R > 0$ and consider those y for which $|y| \le R$, then there exists $\gamma_R \in G$ such that

$$\frac{\varepsilon}{\mu} \log \Sigma^{\mu,\varepsilon *}(u,y) e^{\frac{\mu}{\varepsilon}p} \ \in \mathcal{D}^{\gamma_R}$$

for all $u \in U$, $|y| \leq R$, $p \in K$, for all $\varepsilon > 0$ sufficiently small. Considering the right hand side of (3.8) we have

$$
\begin{aligned}
V^{\mu,\varepsilon}(p,k;u) &\triangleq \tfrac{\varepsilon}{\mu} \log \mathbf{E}^{\dagger} \left[S^{\mu,\varepsilon}(\Sigma^{\mu,\varepsilon *}(u, y^{\varepsilon}_{k+1}) e^{\frac{\mu}{\varepsilon}p}, k+1) \right] \\
&= \tfrac{\varepsilon}{\mu} \log \textstyle\int_{\mathbf{R}} \phi^{\varepsilon}(y) S^{\mu,\varepsilon}(\Sigma^{\mu,\varepsilon *}(u,y) e^{\frac{\mu}{\varepsilon}p}, k+1)\, dy \\
&= \tfrac{\varepsilon}{\mu} \log \left\{ \textstyle\int_{|y| \leq R} S^{\mu,\varepsilon}(\Sigma^{\mu,\varepsilon *}(u,y) e^{\frac{\mu}{\varepsilon}p}, k+1)\, dy \right. \qquad (3.9) \\
&\quad + \left. \textstyle\int_{|y| \geq R} \phi^{\varepsilon}(y) S^{\mu,\varepsilon}(\Sigma^{\mu,\varepsilon *}(u,y) e^{\frac{\mu}{\varepsilon}p}, k+1)\, dy \right\} \\
&\triangleq \tfrac{\varepsilon}{\mu} \log \{A + B\}.
\end{aligned}
$$

Now using the bounds above we can write

$$
\frac{\varepsilon}{\mu} \log B \;\leq\; \frac{\varepsilon}{\mu} \log \int_{|y| \geq R} \exp \frac{\mu}{\varepsilon} \left(C(1 + |y|) - \tfrac{1}{2\mu}|y|^2 - \frac{\varepsilon}{2\mu} \log(2\pi\varepsilon) \right) dy,
$$

and using a standard estimate for gaussian integrals we obtain

$$
\begin{aligned}
\tfrac{\varepsilon}{\mu} \log B &\leq C_1 - C_2 R^2 \qquad (3.10) \\
&\leq -C'
\end{aligned}
$$

as $\varepsilon \to 0$, uniformly in $u \in U$, $p \in K$, where $C' > 0$ if $R > 0$ is chosen sufficiently large.

By the induction hypothesis $W^{\mu}(p, k+1)$ exists and

$$
W^{\mu}(p,k+1) \;=\; \lim_{\varepsilon \to 0} \frac{\varepsilon}{\mu} \log S^{\mu,\varepsilon}(e^{\frac{\mu}{\varepsilon}p}, k+1)
$$

uniformly on $\mathcal{D}^{\gamma_R}$. We also have from Theorem 3.1 that

$$
\Lambda^{\mu *}(u,y)p \;=\; \lim_{\varepsilon \to 0} \frac{\varepsilon}{\mu} \log \Sigma^{\mu,\varepsilon *}(u,y) e^{\frac{\mu}{\varepsilon}p}
$$

uniformly on $U \times B(0,R) \times K$, and

$$
\Lambda^{\mu *}(u,y)p \in \mathcal{D}^{\gamma_R}
$$

for all $(u,y,p) \in U \times B(0,R) \times K$.

Consider the function $W^{\mu}(\Lambda^{\mu *}(u,y)p, k+1) - \frac{1}{2\mu}|y|^2$. Due to the induction hypothesis and the properties of $\Lambda^{\mu,\varepsilon}$, it is continuous in p, y, u, and bounded in y; all properties uniformly in $(u,p) \in U \times K$. Therefore we can choose R large enough so that both (3.10) is satisfied and

$$
\underset{y \in \mathbf{R}}{\operatorname{argmax}} \{ W^{\mu}(\Lambda^{\mu *}(u,y)p, k+1) - \tfrac{1}{2\mu}|y|^2 \} \subset B(0,R).
$$

We keep R fixed from now on.

Combining the above limits for $S^{\mu,\varepsilon}$ and $\Sigma^{\mu,\varepsilon}$ we have that

$$\begin{aligned}
&\lim_{\varepsilon\to 0} \tfrac{\varepsilon}{\mu} \log \left\{ S^{\mu,\varepsilon}(\Sigma^{\mu,\varepsilon *}(u,y)e^{\frac{\mu}{\varepsilon}p}, k+1) \exp \tfrac{\mu}{\varepsilon}\left(-\tfrac{1}{2\mu}|y|^2 - \tfrac{\varepsilon}{2\mu}\log(2\pi\varepsilon)\right)\right\} \\
&= W^{\mu}(\Lambda^{\mu *}(u,y)p, k+1) - \tfrac{1}{2\mu}|y|^2
\end{aligned} \tag{3.11}$$

uniformly in $u \in U$, $p \in K$, $y \in B(0,R)$.

We can now proceed with the further computation of (3.9). Indeed, we follow the proof of Lemma 6.1 with

$$S^{\mu,\varepsilon}(\Sigma^{\mu,\varepsilon *}(u,y)e^{\frac{\mu}{\varepsilon}p}, k+1) \exp \frac{\mu}{\varepsilon}\left(-\tfrac{1}{2\mu}|y|^2 - \frac{\varepsilon}{2\mu}\log(2\pi\varepsilon)\right)$$

replacing $\exp(F_a^{\varepsilon}/\varepsilon)$, and

$$W^{\mu}(\Lambda^{\mu *}(u,y)p, k+1) - \tfrac{1}{2\mu}|y|^2$$

replacing F_a, and $a = (u,p) \in A = U \times K$. Then

$$\lim_{\varepsilon\to 0} \frac{\varepsilon}{\mu} \log A \;=\; \sup_{y\in R} \left\{ W^{\mu}(\Lambda^{\mu *}(u,y)p, k+1) - \tfrac{1}{2\mu}|y|^2 \right\} \tag{3.12}$$

uniformly in $U \times K$. It is crucial that R is chosen as above so that the uniform bound required by condition (iii) of Lemma 6.1 is satisfied. Consequently

$$\begin{aligned}
\lim_{\varepsilon\to 0} V^{\mu,\varepsilon}(p,k;u) &= \lim_{\varepsilon\to 0} \tfrac{\varepsilon}{\mu} \log (A+B) \\
&= \lim_{\varepsilon\to 0} \tfrac{\varepsilon}{\mu} \log A\,(1 + B/A) \\
&= \lim_{\varepsilon\to 0} \tfrac{\varepsilon}{\mu} \log A \\
&= \sup_{y\in\mathbf{R}} \left\{ W^{\mu}(\Lambda^{\mu *}(u,y)p, k+1) - \tfrac{1}{2\mu}|y|^2 \right\}
\end{aligned}$$

uniformly on $U \times K$, since $\lim_{\varepsilon\to 0} \frac{\varepsilon}{\mu}\log(1 + B/A) = 0$ from (3.10), and (3.12).

To complete the proof, we use the continuity of the map $(p,u) \mapsto V^{\mu,\varepsilon}(p,k;u)$, $\mathcal{D}^{\gamma} \times U \to \mathbf{R}$ to obtain

$$\begin{aligned}
\lim_{\varepsilon\to 0} \tfrac{\varepsilon}{\mu} \log S^{\mu,\varepsilon}(e^{\frac{\mu}{\varepsilon}p}, k) &= \lim_{\varepsilon\to 0} \inf_{u\in U} V^{\mu,\varepsilon}(p,k;u) \\
&= \inf_{u\in U} \lim_{\varepsilon\to 0} V^{\mu,\varepsilon}(p,k;u) \\
&= \inf_{u\in U} \sup_{y\in\mathbf{R}} \left\{ W^{\mu}(\Lambda^{\mu *}(u,y)p, k+1) - \tfrac{1}{2\mu}|y|^2 \right\} \\
&= W^{\mu}(p,k)
\end{aligned} \tag{3.13}$$

uniformly on K. The last equality holds by the definition (3.6).

The sequence converges uniformly on K, and as a result $W^\mu(p,k)$ is continuous on $\mathcal{D}^\gamma$. This completes the proof. □

Remark 3.3 In §4, equation (3.7) will be interpreted as the optimal cost function (upper value) for a *dynamic game problem.*

Remark 3.4 Note that there are two large deviation type limits involved in the result of Theorem 3.2. The first is expressed in the use of (3.5) in (3.11) and corresponds to "state estimation"or "observers". The second is embodied in (3.6) and corresponds to the relationship of the stochastic risk–sensitive optimal control problem to the deterministic game that (3.7), (3.13) imply.

4 A Dynamic Game Problem

4.1 Dynamics

We consider a two–player deterministic partially observed dynamic game for the discrete-time system

$$(4.1)\qquad \begin{cases} x_{k+1} &= b(x_k,u_k)+w_k, \\ y_{k+1} &= h(x_k)+v_k, \end{cases}$$

on the finite time interval $k=0,1,2,\dots M$, where

(i) $x(0)=x_0$ is the unknown initial condition, and $y(0)=0$.

(ii) Player 1 selects the U–valued control u_k, which is required to be a non–anticipating functional of the observation path y. We write $\mathcal{U}_{k,l}$ for the set of such controls defined on the interval $k,\dots,l$, and note that $u\in\mathcal{U}_{k,l}$ if and only if for each $j\in[k,l]$ there exists a function $\bar{u}_j:\mathbf{R}^{(j-k+1)n}\to U$ such that $u_j=\bar{u}_j(y_{k+1,j})$.

(iii) Player 2 selects the $\mathbf{R}^n\times\mathbf{R}$–valued disturbance (w_k,v_k), which is a square summable open loop sequence. We let $\ell_2([k,l],\mathbf{R}^p)$ denote the set of square summable $\mathbf{R}^p$–valued sequences defined on the interval $k,\dots,l$ $(p=n,1,n+1)$.

4.2 Cost

The payoff function for the game is defined for admissible $u\in\mathcal{U}_{0,M-1}$ and $(w,v)\in\ell_2([0,M-1],\mathbf{R}^{n+1})$ by

$$(4.2)\quad J^\mu(u,w,v) = \sup_{x_0\in\mathbf{R}^n}\left\{\alpha(x_0)+\sum_{l=0}^{M-1}L(x_l,u_l)+\Phi(x_M)-\frac{1}{\mu}\sum_{l=0}^{M-1}\frac{1}{2}\left(|w_l|^2+|v_l|^2\right)\right\},$$

where $\alpha\in\mathcal{D}$.

Remark 4.1 This formulation treats x_0 as part of the uncertainty (to be chosen by nature). *A priori* knowledge of x_0 is incorporated in the cost function $\alpha \in \mathcal{D}$. Our theory also applies to the case $\alpha = 0$, which corresponds to no *a priori* information, and note $\lim_{\varepsilon \to 0} \varepsilon \log \rho = 0$, where ρ is the initial density for the risk–sensitive stochastic control problem. One can alternatively select $\alpha \in \mathcal{D}$ and define the initial density by $\rho(x) = c_\varepsilon \exp(-\alpha(x)/\varepsilon)$, where c_ε is a normalizing constant.

The (upper) game is defined as follows. Let

$$J^\mu(u) \;=\; \sup_{(w,v)\in\ell_2([0,M-1],\mathbf{R}^{n+1})} J^\mu(u,w,v),$$

and the *(upper) partially observed dynamic game problem* is to find $u^* \in \mathcal{U}_{0,M-1}$ such that

$$J^\mu(u^*) \;=\; \inf_{u\in\mathcal{U}_{0,M-1}} J^\mu(u).$$

The cost function $J^\mu(u)$ can be rewritten in the form

$$(4.3)\qquad \begin{aligned} J^\mu(u) \;=\;& \sup_{x\in\ell_2([0,M],\mathbf{R}^n)} \sup_{y\in\ell_2([1,M],\mathbf{R})} \Big\{\alpha(x_0) + \textstyle\sum_{l=0}^{M-1} L(x_l,u_l) + \Phi(x_M) \\ & -\tfrac{1}{\mu} \textstyle\sum_{l=0}^{M-1} \tfrac{1}{2}\left(|x_{l+1} - b(x_l,u_l)|^2 + |y_{l+1} - h(x_l)|^2\right)\Big\}. \end{aligned}$$

This cost function is finite for all $\mu > 0$.

4.3 Information State

Motivated by the use of an information state, and associated separation policies, in solving the risk–sensitive stochastic control problem, given a control policy $u \in \mathcal{U}_{0,M-1}$ and an observation path $y \in \ell_2([0,M],\mathbf{R})$, we define an "information state" $p_k^\mu \in \mathcal{D}$ and its "dual" $q_k^\mu \in C_b(\mathbf{R}^n)$ for the game problem by the recursions

$$(4.4)\qquad \begin{cases} p_k^\mu \;=\; \Lambda^{\mu *}(u_{k-1},y_k)p_{k-1}^\mu \\[2ex] p_0^\mu \;=\; \alpha, \end{cases}$$

$$(4.5)\qquad \begin{cases} q_{k-1}^\mu \;=\; \Lambda^{\mu}(u_{k-1},y_k)q_k^\mu \\[2ex] q_M^\mu \;=\; \Phi, \end{cases}$$

where Λ^μ and $\Lambda^{\mu *}$ are as defined in (3.3). Note that, in the "sup pairing" notation of (3.1),

$$(4.6)\qquad (p_k^\mu,\, q_k^\mu) \;=\; \left(p_{k-1}^\mu,\, q_{k-1}^\mu\right)$$

for all k, and

$$(4.7)\qquad \sigma_k^{\mu,\varepsilon} \;\sim\; \exp\frac{\mu}{\varepsilon}p_k^\mu, \qquad \nu_k^{\mu,\varepsilon} \;\sim\; \exp\frac{\mu}{\varepsilon}q_k^\mu$$

in probability as $\varepsilon \to 0$.

Remark 4.2 The asymptotic formulae (4.7) are similar to the large deviation limit for nonlinear filtering ([13], [14], [18], [17]). In addition, if $L \equiv 0$ the recursions (4.4) reduce to the equations for Mortensen's method of minimum energy estimation.

4.4 Alternate Representation of the Cost

Define for $u \in \mathcal{U}_{0,M-1}$

$$K^\mu(u) \;=\; \sup_{y \in \ell_2([1,M],\mathbf{R})} \left\{ (p_M^\mu,\, \Phi) - \tfrac{1}{2\mu} \sum_{l=0}^{M-1} |y_{l+1}|^2 \right\}, \tag{4.8}$$

a cost function associated with the new "state"process p_k^μ.

Theorem 4.3 *We have for all* $u \in \mathcal{U}_{0,M-1}$

$$J^\mu(u) \;=\; K^\mu(u). \tag{4.9}$$

PROOF. Iterating (4.4), we see that

$$\begin{aligned} p_M^\mu(z) \;=\; \sup\nolimits_{x \in \ell_2([0,M],\mathbf{R}^n)} \Big\{ & \alpha(x_0) + \textstyle\sum_{l=0}^{M-1} L(x_l, u_l) - \frac{1}{\mu} \sum_{l=0}^{M-1} \frac{1}{2} |x_{l+1} - b(x_l, u_l)|^2 \\ & - \tfrac{1}{\mu} \textstyle\sum_{l=0}^{M-1} \left[\frac{1}{2} |h(x_l)|^2 - h(x_l) y_{l+1} \right] \;:\; x_M = z \Big\}. \end{aligned}$$

Substitution of this equality into (4.8) yields (4.9). □

We now define an alternate deterministic dynamic game problem with complete state information. Consider the state sequence p_k^μ with dynamics (4.4) and the cost $K^\mu(u)$: find $u^* \in \mathcal{U}_{0,M-1}^s$ minimizing K^μ. Here, $\mathcal{U}_{k,l}^s$ denotes the set of separated control policies, through the information state p_k^μ, defined on the interval $k, \ldots, l$, i.e. those which are non–anticipating functionals of $(p_j^\mu,\ k \le j \le l)$. Note that $\mathcal{U}_{k,l}^s \subset \mathcal{U}_{k,l}$.

4.5 Dynamic Programming

Consider now the state p^μ on the interval $k, \ldots, M$ with initial condition $p_k^\mu = p \in \mathcal{D}$:

$$\begin{cases} p_l^\mu \;=\; \Lambda^{\mu *}(u_{l-1}, y_l) p_{l-1}^\mu, & k+1 \le l \le M, \\ p_k^\mu \;=\; p. \end{cases} \tag{4.10}$$

The (upper) value function is defined for $p \in \mathcal{D}$ by

$$W^\mu(p,k) \;=\; \inf_{u \in \mathcal{U}_{k,M-1}} \; \sup_{y \in \ell_2([k+1,M],\mathbf{R})} \left\{ (p_k^\mu,\, q_k^\mu) - \tfrac{1}{2\mu} \sum_{l=k}^{M-1} |y_{l+1}|^2 \;:\; p_k^\mu = p \right\}. \tag{4.11}$$

Theorem 4.4 *The value function $W^\mu(p,k)$ defined by (4.11) is the unique solution of the dynamic programming equation (3.7).*

Proof. From (4.11), (4.4), (4.5), and (4.6), we have

$$
\begin{aligned}
& W^\mu(p,k) \\
&= \inf_{u_k\in\mathcal{U}_{k,k}} \inf_{v\in\mathcal{U}_{k+1,M-1}} \sup_{y_{k+1}\in\mathbf{R}} \sup_{y\in\ell_2([k+2,M],\mathbf{R})} \\
&\qquad \Big\{ \Big(p_k^\mu,\ \Lambda^\mu(u_k,y_{k+1})q_{k+1}^\mu(v_{k+1,M-1})\Big) - \tfrac{1}{2\mu}\textstyle\sum_{l=k+1}^{M-1}|y_{l+1}|^2 - \tfrac{1}{2\mu}|y_{k+1}|^2 \ :\ p_k^\mu = p \Big\} \\
&= \inf_{u_k\in\mathcal{U}_{k,k}} \sup_{y_{k+1}\in\mathbf{R}} \inf_{v\in\mathcal{U}_{k+1,M-1}} \sup_{y\in\ell_2([k+2,M],\mathbf{R})} \\
&\qquad \Big\{ \Big(\Lambda^{\mu *}(u_k,y_{k+1})p_k^\mu,\ q_{k+1}^\mu(v_{k+1,M-1})\Big) - \tfrac{1}{2\mu}\textstyle\sum_{l=k+1}^{M-1}|y_{l+1}|^2 - \tfrac{1}{2\mu}|y_{k+1}|^2 \ :\ p_k^\mu = p \Big\} \\
&= \inf_{u_k\in\mathcal{U}_{k,k}} \sup_{y_{k+1}\in\mathbf{R}} \Big\{ W^\mu(\Lambda^{\mu *}(u_k,y_{k+1})p, k+1) - \tfrac{1}{2\mu}|y_{k+1}|^2 \Big\}
\end{aligned}
$$

which is the same as (3.7). Here, the interchange of the minimization over $v_{k+1,M-1}$ and maximization over y_{k+1} is justified because these terms are not coupled in the expression being optimized. □

Remark 4.5 We conclude from Theorems 3.2 and 4.4 that the small noise limit of the partially observed stochastic risk–sensitive problem is a partially observed deterministic dynamic game problem.

Theorem 4.6 (Verification.) *Suppose that $u^* \in \mathcal{U}_{0,M-1}^s$ is a policy such that, for each $k = 0,\ldots,M-1$, $u_k^* = \bar{u}_k^*(p_k^\mu)$, where $\bar{u}_k^*(p)$ achieves the minimum in (3.7). Then $u^* \in \mathcal{U}_{0,M-1}$ and is an optimal policy for the partially observed dynamic game problem (§§4.1, 4.2).*

Proof. Define

$$
W^\mu(p,k;u) \;=\; \sup_{y\in\ell_2([k+1,M],\mathbf{R})} \left\{ (p_k^\mu,\, q_k^\mu) - \frac{1}{2\mu}\sum_{l=k}^{M-1}|y_{l+1}|^2 \ :\ p_k^\mu = p \right\}.
$$

We claim that

$$
W^\mu(p,k) \;=\; W^\mu(p,k;u^*) \tag{4.12}
$$

for each $k = 0,1,\ldots,M$.

For $k = M$, (4.12) is true. Assume now that (4.12) holds for $k+1, \ldots, M$. Then

$$
\begin{aligned}
& W^\mu(p, k; u^*) \\
&= \sup_{y_{k+1} \in \mathbf{R}} \sup_{y \in \ell_2([k+2, M], \mathbf{R})} \\
&\qquad \left\{ \left(\Lambda^{\mu *}(u_k^*, y_{k+1}) p_k^\mu,\ q_{k+1}^\mu(u_{k+1,M-1}^*) \right) - \tfrac{1}{2\mu} \textstyle\sum_{l=k+1}^{M-1} |y_{l+1}|^2 - \tfrac{1}{2\mu} |y_{k+1}|^2 \ :\ p_k^\mu = p \right\} \\
&= \sup_{y_{k+1} \in \mathbf{R}} \left\{ W^\mu(\Lambda^{\mu *}(u_k^*, y_{k+1}) p_k^\mu, k+1; u_{k+1,M-1}^*) - \tfrac{1}{2\mu} |y_{k+1}|^2 \ :\ p_k^\mu = p \right\} \\
&= \sup_{y_{k+1} \in \mathbf{R}} \left\{ W^\mu(\Lambda^{\mu *}(u_k^*, y_{k+1}) p, k+1) - \tfrac{1}{2\mu} |y_{k+1}|^2 \right\} \\
&= W^\mu(p, k),
\end{aligned}
$$

which proves (4.12).

Next, from (4.12) and setting $k = 0$ and $p = \alpha$ we obtain

$$W^\mu(\alpha, 0; u^*) \ = \ W^\mu(\alpha, 0) \ \le \ W^\mu(\alpha, 0; u)$$

for all $u \in \mathcal{U}_{0,M-1}$, which implies

$$K^\mu(u^*) \ \le \ K^\mu(u)$$

for all $u \in \mathcal{U}_{0,M-1}$. This together with Theorem 4.3 completes the proof. □

4.6 Certainty Equivalence Principle

Remark 4.7 Theorem 4.6 is the "appropriate" separation theorem for the partially observed dynamic game described in §§4.1, 4.2, in that it establishes that the optimal feedback policy is a separated one [19] through the information state p_k^μ which carries all the information from the observations $y_{0,k}$ relevant for control. It is important to note that the solution of this partially observed dynamic game problem involves two infinite–dimensional recursions. One is (4.4), which describes the dynamics of the information state, evolves forward in time, and is a dynamic programming equation in view of (3.3). This equation plays the role of an "observer" in the resulting controller, and is determined by the control problem at hand, and not prescribed *a priori*. The information state $p_k^\mu(x)$ is a function of the state variable $x \in \mathbf{R}^n$. The other recursion is (3.7), which describes the computation of the feedback control as a function of the information state, evolves backward in time, and is a dynamic programming equation. The value function $W^\mu(p, k)$ is a function of the information state variable p, which takes values in the infinite–dimensional space $\mathcal{D}$. An important aspect of our work is to point out this essential difficulty of the nonlinear robust control problem. This is not surprising given that this difficulty is well known in stochastic control. For practical applications, one can try to find suboptimal finite–dimensional schemes that provide performance close to the one predicted by the optimal results obtained here. We are also pursuing the development of numerical schemes that can compute the required recursions (4.4) and (3.7), as well as the incorporation of such schemes into CAD control systems design software based on optimization.

We now relate the above analysis to certainty equivalence principles suggested by Whittle [22], [24], [25] for the risk–sensitive problem, and by Basar–Bernhard [3] for the game problem.

Consider a completely observed dynamic game problem with dynamics

$$x_{k+1} = b(x_k, u_k) + w_k \tag{4.13}$$

and payoff function

$$J^\mu(u, w) = \sum_{l=0}^{M-1} L(x_l, u_l) + \Phi(x_M) - \frac{1}{\mu} \sum_{l=0}^{M-1} \frac{1}{2}|w_l|^2, \tag{4.14}$$

where the initial state x_0 is known, player 1 selects $u \in \mathcal{U}^c_{0,M-1}$ to minimize J^μ, and player 2 selects $w \in \ell_2([0, M-1), \mathbf{R}^n)$ to maximize J^μ. Here, $\mathcal{U}^c_{k,l}$ is the set of U–valued controls which are non–anticipating functionals of the state x defined on the interval $k, \ldots, l$.

Define the upper value (see, e.g. [8]) for this dynamic game by

$$\begin{aligned} \bar{f}^\mu_k(x) = \inf_{u \in \mathcal{U}^c_{k,M-1}} \sup_{w \in \ell_2([k,M-1],\mathbf{R}^n)} \Big\{ & \textstyle\sum_{l=k}^{M-1} L(x_l, u_l) + \Phi(x_M) \\ & -\frac{1}{\mu} \textstyle\sum_{l=k}^{M-1} \frac{1}{2}|w_l|^2 \; : \; x_k = x \Big\}. \end{aligned} \tag{4.15}$$

This function satisfies the dynamic programming equation

$$\begin{cases} \bar{f}^\mu_k(x) = \inf_{u \in U} \sup_{w \in \mathbf{R}^n} \left\{ \bar{f}^\mu_{k+1}(b(x, u) + w) + L(x, u) - \frac{1}{2\mu}|w|^2 \right\} \\ \bar{f}^\mu_M(x) = \Phi(x), \end{cases} \tag{4.16}$$

and if $\bar{v}^*_k(x) \in U$ achieves the minimum in (4.16), then $u^*_k = \bar{v}^*_k(x_k)$ is an optimal feedback policy for this completely observed game.

Whittle [22], [24], [25] solves the partially observed risk–sensitive stochastic control problem by using the solution to the completely observed game (4.16), and the modified filter or "observer"(4.4). He refers to $\bar{f}^\mu_k$ as the *future stress*, and to p^μ_k as the *past stress*, and defines the *minimum stress estimate* $\bar{x}_k$ of x_k by

$$\bar{x}_k \in \operatorname*{argmax}_{x \in \mathbf{R}^n} \left\{ p^\mu_k(x) + \bar{f}^\mu_k(x) \right\} \triangleq \hat{x}^\mu_k, \tag{4.17}$$

where $\hat{x}^\mu_k$ is a *set–valued* variable. For linear systems with quadratic cost, the *certainty equivalence principle* asserts that if $J^\mu(u, w, v)$ is negative definite in (w, v) (and positive definite in u), then $u_k^* = \bar{v}_k^*(\bar{x}_k)$ is an optimal control for the partially observed risk–sensitive problem [22]. For nonlinear systems, Whittle's assertion is that this recipe gives a policy which is approximately optimal for the risk–sensitive stochastic control problem [25].

Remark 4.8 The variable $\hat{x}_k^\mu$ is set–valued and is closely related to the finite–time observer results of James ([17], equation (3.8)), and our earlier observer design methodology [1], [18]. Indeed our construction brings out another essential difficulty of the nonlinear problem, which has to do with multivalued (or set–valued) variables for state estimation and control. We will have more to say about this issue in a forthcoming paper.

We now state a certainty equivalence principle for the partially observed deterministic game problem described in §§4.1, 4.2 (c.f. Basar–Bernhard [3], Chapters 5 & 6).

Theorem 4.9 (Certainty equivalence.) *Let $\bar{f}_k^\mu(x)$ be the upper value for the full state information game (4.13), (4.14). If for all $k = 0, \ldots, M$ and $p \in \mathcal{D}$ we have*

$$W^\mu(p,k) \;=\; \left(p,\, \bar{f}_k^\mu\right), \tag{4.18}$$

then the policy $u^* \in \mathcal{U}_{0,M-1}^s$ defined by

$$u_k^* \;=\; \bar{v}_k^*(\bar{x}_k) \tag{4.19}$$

is an optimal policy for the partially observed game problem (§§4.1, 4.2).

PROOF. Let

$$x_k^*(p) \;\in\; \operatorname*{argmin}_{\xi\in\mathbf{R}^n} \left(p(\xi) + \bar{f}_k^\mu(\xi)\right).$$

Then the minimum stress estimate and the candidate policy defined by (4.19) can be written as

$$\bar{x}_k \;=\; x_k^*(p_k^\mu) \quad \text{and} \quad u_k^* \;=\; \bar{v}_k^*(x_k^*(p_k^\mu)). \tag{4.20}$$

Therefore u_k^* is a separated policy. To check the optimality of u_k^*, we apply the verification Theorem 4.6. We must show that for each k, $v_k^* \triangleq \bar{v}_k^*(x_k^*(p))$ achieves the minimum in (3.7). To prove this, using the hypothesis (4.18) at time $k+1$, we have

$$\begin{aligned}
& W^\mu(p,k) \\
&= \inf\nolimits_{u\in U} \sup\nolimits_{y\in\mathbf{R}} \left\{ W^\mu(\Lambda^{\mu*}(u,y)p, k+1) - \tfrac{1}{2\mu}|y|^2 \right\} \\
&= \inf\nolimits_{u\in U} \sup\nolimits_{y\in\mathbf{R}} \left\{ \left(\Lambda^{\mu*}(u,y)p,\, \bar{f}_{k+1}^\mu\right) - \tfrac{1}{2\mu}|y|^2 \right\} \\
&= \inf\nolimits_{u\in U} \sup\nolimits_{y\in\mathbf{R}} \sup\nolimits_{z\in\mathbf{R}^n} \sup\nolimits_{\xi\in\mathbf{R}^n} \\
&\qquad \left\{ p(\xi) + L(\xi,u) - \tfrac{1}{2\mu}|z - b(\xi,u)|^2 - \tfrac{1}{\mu}\left[\tfrac{1}{2}|h(\xi)|^2 - h(\xi)y\right] - \tfrac{1}{2\mu}|y|^2 + \bar{f}_{k+1}^\mu(z) \right\} \\
&= \inf\nolimits_{u\in U} \sup\nolimits_{\xi\in\mathbf{R}^n} \left(p(\xi) + L(\xi,u) + \sup\nolimits_{z\in\mathbf{R}^n} \left\{ -\tfrac{1}{2\mu}|z - b(\xi,u)|^2 + \bar{f}_{k+1}^\mu(z) \right\} \right).
\end{aligned} \tag{4.21}$$

On the other hand, (4.18) at time k implies

$$
\begin{aligned}
W^\mu(p,k) &= \sup_{\xi\in\mathbf{R}^n}\left(p(\xi)+\bar{f}^\mu_k(\xi)\right) \\
&= \sup_{\xi\in\mathbf{R}^n}\inf_{u\in U}\left(p(\xi)+L(\xi,u)+\sup_{z\in\mathbf{R}^n}\left\{-\tfrac{1}{2\mu}|z-b(\xi,u)|^2+\bar{f}^\mu_{k+1}(z)\right\}\right).
\end{aligned} \tag{4.22}
$$

Thus $W^\mu(p,k)$ is a saddle value for a static game, with saddle point $\xi = x^*_k(p)$, $u = \bar{v}^*_k(x^*_k(p))$. Therefore $\bar{v}^*_k(x^*_k(p))$ achieves the minimum in (3.7). □

Remark 4.10 As described in Remark 4.7 above, the partially observed dynamic game (and the related robust control problem [2]) involve two infinite–dimensional recursions. The significance of the "certainty equivalence"theorem is that if valid, the recursion (3.7) involving the value $W^\mu(p,k)$, $p\in\mathcal{D}$ can be replaced by a simpler recursion (4.15) involving the upper value $\bar{f}^\mu_k(x)$, $x\in\mathbf{R}^n$. This has obvious computational implications.

Remark 4.11 A crucial contribution here is that we have identified precisely the condition that one needs to establish the certainty equivalence principle suggested in [22], [25], [3]. The condition is (4.18), i.e., the saddle point condition in (4.21), (4.22). One may ask to what extent this condition can be obviated. Or alternatively, can we show under certain assumptions that (4.18) is satisfied? The proof of Theorem 4.9 can be used as a basis for answering this question. In [22], [25], [3], it is apparently assumed that the full state game problem has a value (equivalent to the existence of a saddle point for this game), and other assumptions are made. The full state game problem has a value $f^\mu_k(x)$ if the upper value $\bar{f}^\mu_k(x)$ coincides with the lower value $\underline{f}^\mu_k(x)$, in which case $f^\mu_k(x)$ is defined to equal this common number. The lower value is defined by

$$
\begin{aligned}
\underline{f}^\mu_k(x) &= \sup_{w\in\ell_2([k,M-1],\mathbf{R}^n)}\inf_{u\in\mathcal{U}^c_{k,M-1}}\left\{\sum_{l=k}^{M-1}L(x_l,u_l)+\Phi(x_M)\right. \\
&\qquad\left. -\tfrac{1}{\mu}\sum_{l=k}^{M-1}\tfrac{1}{2}|w_l|^2 \ :\ x_k=x\right\},
\end{aligned} \tag{4.23}
$$

and solves the dynamic programming equation

$$
\left\{
\begin{aligned}
\underline{f}^\mu_k(x) &= \sup_{w\in\mathbf{R}^n}\inf_{u\in U}\left\{\underline{f}^\mu_{k+1}(b(x,u)+w)+L(x,u)-\tfrac{1}{2\mu}|w|^2\right\} \\
\underline{f}^\mu_M(x) &= \Phi(x).
\end{aligned}
\right. \tag{4.24}
$$

In continuous–time, the Isaacs condition [7], [3] implies the equivalence of the two values.

Remark 4.12 Since the partially observed game is the limit of the partially observed risk–sensitive problem, the policy (4.19) is an approximate optimal policy for the partially observed risk–sensitive problem for small $\varepsilon>0$ (c.f. Whittle [25]).

5 Small Risk Limit

In this section we show that a risk–neutral stochastic control problem is obtained if in the risk–sensitive stochastic control problem the risk–sensitivity parameter μ tends to zero.

5.1 Information State

Define the bounded linear operator $\Sigma^{\varepsilon *} : L^1(\mathbf{R}^n) \to L^1(\mathbf{R}^n)$ by

$$(5.1) \qquad \Sigma^{\varepsilon *}(u,y)\sigma(z) \triangleq \int_{\mathbf{R}^n} \psi^\varepsilon(z - b(\xi,u))\Psi^\varepsilon(\xi,y)\sigma(\xi)\, d\xi.$$

Theorem 5.1 *We have*

$$(5.2) \qquad \lim_{\mu\to 0} \Sigma^{\mu,\varepsilon *}(u,y)\sigma \;=\; \Sigma^{\varepsilon *}(u,y)\sigma,$$

uniformly on bounded subsets of $U \times \mathbf{R} \times L^1(\mathbf{R}^n)$.

PROOF. This result follows simply from the definitions (2.6), (5.1). □

Next, we define a process $\sigma_k^\varepsilon \in L^1(\mathbf{R}^n)$ by the recursion

$$(5.3) \qquad \begin{cases} \sigma_k^\varepsilon \;=\; \Sigma^{\varepsilon *}(u_{k-1}, y_k^\varepsilon)\, \sigma_{k-1}^\varepsilon \\[2ex] \sigma_0^\varepsilon \;=\rho. \end{cases}$$

Remark 5.2 The process σ_k^ε is an unnormalized conditional density of x_k^ε given $\mathcal{Y}_k$, and equation (5.3) is known as the Zakai equation [9], [19].

5.2 A Risk–Neutral Control Problem

We again consider the discrete–time stochastic system (2.1), and formulate a *partially observed risk–neutral stochastic control problem* with cost

$$(5.4) \qquad J^\varepsilon(u) \;=\; \mathbf{E}^u \left[\sum_{l=0}^{M-1} L(x_l^\varepsilon, u_l) + \Phi(x_M^\varepsilon) \right]$$

defined for $u \in \mathcal{U}_{0,M-1}$, where $\mathcal{U}_{0,M-1}$, etc, are as defined in §2. This cost function is finite for all $\varepsilon > 0$.

We quote the following result from [9], [19], which establishes that the optimal policy is separated through the information state σ_k^ε satisfying (5.3).

Theorem 5.3 *The unnormalized conditional density σ_k^ε is an information state for the risk–neutral problem, and the value function defined for $\sigma \in L^1(\mathbf{R}^n)$ by*

$$(5.5) \qquad W^\varepsilon(\sigma,k) = \inf_{u\in\mathcal{U}^s_{k,M-1}} \mathbf{E}^\dagger \left[\sum_{l=k}^{M-1} \langle \sigma_l^\varepsilon, L(\cdot,u)\rangle + \langle \sigma_M^\varepsilon, \Phi\rangle \mid \sigma_k^\varepsilon = \sigma \right]$$

satisfies the dynamic programming equation

$$(5.6) \qquad \begin{cases} W^\varepsilon(\sigma,k) = \inf_{u\in U} \mathbf{E}^\dagger \left[\langle \sigma, L(\cdot,u)\rangle + W^\varepsilon(\Sigma^{\varepsilon *}(u, y^\varepsilon_{k+1})\sigma, k+1)\right] \\ W^\varepsilon(\sigma,M) = \langle \sigma, \Phi\rangle. \end{cases}$$

If $u^* \in \mathcal{U}^s_{0,M-1}$ is a policy such that, for each $k = 0,\dots,M-1$, $u_k^* = \bar{u}_k^*(\sigma_k^\varepsilon)$, where $\bar{u}_k^*(\sigma)$ achieves the minimum in (5.6), then $u^* \in \mathcal{U}_{0,M-1}$ and is an optimal policy for the partially observed risk–neutral problem.

Remark 5.4 The function $W^\varepsilon(\sigma,k)$ depends continuously on $\sigma \in L^1(\mathbf{R}^n)$.

5.3 Risk–Sensitive Value Function

The next theorem evaluates the small risk limit of the risk–sensitive stochastic control problem. Note that normalization of the information state is required.

Theorem 5.5 *We have*

$$(5.7) \qquad \lim_{\mu\to 0} \frac{\varepsilon}{\mu} \log \frac{S^{\mu,\varepsilon}(\sigma,k)}{\langle \sigma, 1\rangle} = \frac{W^\varepsilon(\sigma,k)}{\langle\sigma, 1\rangle}$$

uniformly on bounded subsets of $L^1(\mathbf{R}^n)$.

PROOF. 1. We claim that

$$(5.8) \qquad S^{\mu,\varepsilon}(\sigma,k) = \langle\sigma, 1\rangle + \frac{\mu}{\varepsilon} W^\varepsilon(\sigma,k) + o(\mu)$$

as $\mu \to 0$ uniformly on bounded subsets of $L^1(\mathbf{R}^n)$.

For $k = M$,

$$\begin{aligned} S^{\mu,\varepsilon}(\sigma,M) &= \langle \sigma, e^{\frac{\mu}{\varepsilon}\Phi}\rangle \\ &= \langle\sigma, 1\rangle + \tfrac{\mu}{\varepsilon}\langle\sigma, \Phi\rangle + o(\mu) \\ &= \langle\sigma, 1\rangle + \tfrac{\mu}{\varepsilon} W^\varepsilon(\sigma, M) + o(\mu) \end{aligned}$$

as $\mu \to 0$, uniformly on bounded subsets of $L^1(\mathbf{R}^n)$.

Assume now that (5.8) is true for $k+1, \ldots, M$. Then

$$\begin{aligned}
V^{\mu,\varepsilon}(\sigma, k; u) &\triangleq E^{\dagger}\left[S^{\mu,\varepsilon}(\Sigma^{\mu,\varepsilon *}(u, y^{\varepsilon}_{k+1})\sigma, k+1)\right] \\
&= \textstyle\int_{\mathbf{R}} \phi^{\varepsilon}(y) S^{\mu,\varepsilon}(\Sigma^{\mu,\varepsilon *}(u, y)\sigma, k+1)\, dy \\
&= \textstyle\int_{\mathbf{R}^n} \int_{\mathbf{R}} \phi^{\varepsilon}(y - h(\xi)) \exp \frac{\mu}{\varepsilon} L(\xi, u) S^{\mu,\varepsilon}(\frac{\Sigma^{\mu,\varepsilon *}(u,y)\sigma}{\langle \Sigma^{\mu,\varepsilon *}(u,y)\sigma,\, 1\rangle}, k+1)\sigma(\xi)\, d\xi\, dy \\
&= \textstyle\int_{\mathbf{R}^n} \int_{\mathbf{R}} \phi^{\varepsilon}(y - h(\xi))\{1 + \frac{\mu}{\varepsilon} L(\xi, u) + o(\mu)\} \\
&\qquad \{1 + \tfrac{\mu}{\varepsilon} W^{\varepsilon}(\tfrac{\Sigma^{\mu,\varepsilon *}(u,y)\sigma}{\langle \Sigma^{\mu,\varepsilon *}(u,y)\sigma,\, 1\rangle}, k+1) + o(\mu)\}\sigma(\xi)\, d\xi\, dy \\
&= \langle \sigma,\, 1\rangle + \tfrac{\mu}{\varepsilon}\{\langle \sigma,\, L(\,\cdot\,, u)\rangle \\
&\qquad + \textstyle\int_{\mathbf{R}^n} \int_{\mathbf{R}} \phi^{\varepsilon}(y - h(\xi)) W^{\varepsilon}(\frac{\Sigma^{\varepsilon *}(u,y)\sigma}{\langle \Sigma^{\varepsilon *}(u,y)\sigma,\, 1\rangle}, k+1)\sigma(\xi)\, d\xi\, dy\Big\} + o(\mu) \\
&= \langle \sigma,\, 1\rangle + \tfrac{\mu}{\varepsilon}\{\langle \sigma,\, L(\,\cdot\,, u)\rangle + \textstyle\int_{\mathbf{R}} \phi^{\varepsilon}(y) W^{\varepsilon}(\Sigma^{\varepsilon *}(u, y)\sigma, k+1)\, dy\} + o(\mu)
\end{aligned}$$

as $\mu \to 0$ uniformly on bounded subsets of $U \times L^1(\mathbf{R}^n)$. Thus, using the continuity of $(\sigma, u) \mapsto V^{\mu,\varepsilon}(\sigma, k; u)$,

$$\begin{aligned}
S^{\mu,\varepsilon}(\sigma, k) &= \inf_{u \in U} V^{\mu,\varepsilon}(\sigma, k; u) \\
&= \langle \sigma,\, 1\rangle + \tfrac{\mu}{\varepsilon} \inf_{u \in U}\{\langle \sigma,\, L(\,\cdot\,, u) + \textstyle\int_{\mathbf{R}} \phi^{\varepsilon}(y) W^{\varepsilon}(\Sigma^{\varepsilon *}(u, y)\sigma, k+1)\, dy\} + o(\mu) \\
&= \langle \sigma,\, 1\rangle + \tfrac{\mu}{\varepsilon} W^{\varepsilon}(\sigma, k) + o(\mu),
\end{aligned}$$

uniformly on bounded subsets of $L^1(\mathbf{R}^n)$, proving (5.8).

2. To complete the proof, note that (5.8) implies

$$\frac{S^{\mu,\varepsilon}(\sigma, k)}{\langle \sigma,\, 1\rangle} = 1 + \frac{\mu}{\varepsilon} \frac{W^{\varepsilon}(\sigma, k)}{\langle \sigma,\, 1\rangle} + o(\mu)$$

and hence

$$\frac{\varepsilon}{\mu} \log \frac{S^{\mu,\varepsilon}(\sigma, k)}{\langle \sigma,\, 1\rangle} = \frac{W^{\varepsilon}(\sigma, k)}{\langle \sigma,\, 1\rangle} + o(1)$$

as $\mu \to 0$, uniformly on bounded subsets of $L^1(\mathbf{R}^n)$. □

Remark 5.6 We conclude from Theorems 5.3 and 5.5 that the small risk limit of the partially observed stochastic risk–sensitive problem is a partially observed stochastic risk–neutral problem.

6 Appendix

The following theorem is a version of the Varadhan–Laplace lemma [11]. Below ϱ denotes a metric on $C(\mathbf{R}^m)$ corresponding to uniform convergence on compact subsets, and $B(x,\alpha)$ denotes the open ball centered at x of radius α.

Lemma 6.1 *Let A be a compact space, F_a^ε, $F_a \in C(\mathbf{R}^m)$, and assume*

(i)

$$\lim_{\varepsilon\to 0} \sup_{a\in A} \varrho(F_a^\varepsilon, F_a) \;=\; 0.$$

(ii) The function F_a is uniformly continuous on each set $B(0,R)$, $R>0$, uniformly in $a\in A$.

(iii) There exist $\gamma_1>0$, $\gamma_2\geq 0$ such that

$$F_a^\varepsilon(x),\, F_a(x) \;\leq\; -\gamma_1|x|^2+\gamma_2$$

for all $x\in\mathbf{R}^m$, $a\in A$, $\varepsilon>0$.

Then

(6.1) $$\lim_{\varepsilon\to 0}\sup_{a\in A} |\varepsilon \log \int_{\mathbf{R}^n} e^{F_a^\varepsilon(x)/\varepsilon}\,dx \;-\; \sup_{x\in\mathbf{R}^n} F_a(x)| \;=\; 0.$$

Proof. Write $\bar F_a^\varepsilon = \sup_{x\in\mathbf{R}^m} F_a^\varepsilon(x)$, $\bar F_a = \sup_{x\in\mathbf{R}^m} F_a(x)$. Our assumptions ensure that

$$\lim_{\varepsilon\to 0}\sup_{a\in A} \bar F_a^\varepsilon \;=\; \bar F_a.$$

For $\delta>0$ define

$$B_\delta^{a,\varepsilon} \;=\; \left\{x\in\mathbf{R}^m \;:\; F_a^\varepsilon(x) > \bar F_a^\varepsilon - \delta\right\}.$$

Then the uniform coercivity hypothesis *(iii)* ensures there exists $R>0$ such that $B_\delta^{a,\varepsilon}\subset B(0,R)$.

By hypothesis *(ii)* on $B(0,R)$ and using the uniform convergence on $B(0,R)$, given $\delta>0$ there exists $r>0$ such that

$$|x-x'|<r \;\text{ implies }\; |F_a^\varepsilon(x)-F_a^\varepsilon(x')|<\delta$$

for all $x,x'\in B(0,R)$, $a\in A$, and $\varepsilon>0$ sufficiently small.

Let $x_a^\varepsilon\in \operatorname{argmax} F_a^\varepsilon$. Then $|x-x_a^\varepsilon|<r$ implies $|F_a^\varepsilon(x)-\bar F_a^\varepsilon|<\delta$ for all $a\in A$, and $\varepsilon>0$ sufficiently small. Hence

$$B(x_a^\varepsilon, r)\subset B_\delta^{a,\varepsilon}$$

for all $a \in A$, and $\varepsilon > 0$ sufficiently small. Then

$$\begin{aligned} \alpha_a^\varepsilon &\triangleq \int_{\mathbf{R}^m} \exp\left(F_a^\varepsilon(x)/\varepsilon\right)\, dx \\ &\geq \int_{B_\delta^{a,\varepsilon}} \exp\left(F_a^\varepsilon(x)/\varepsilon\right)\, dx \\ &\geq C_m r^m \exp\left(\frac{\bar{F}_a^\varepsilon - \delta}{\varepsilon}\right). \end{aligned}$$

Thus

$$\begin{aligned} \varepsilon \log \alpha_a^\varepsilon &\geq \varepsilon \log C_m r^m + \bar{F}_a^\varepsilon - \delta \\ &\geq \bar{F}_a - 3\delta \end{aligned} \tag{6.2}$$

for all $\varepsilon > 0$ sufficiently small and all $a \in A$.

Next, for $R > 0$ write

$$\begin{aligned} \alpha_a^\varepsilon &= \int_{|x| \leq R} \exp\left(F_a^\varepsilon(x)/\varepsilon\right)\, dx + \int_{|x| \geq R} \exp\left(F_a^\varepsilon(x)/\varepsilon\right)\, dx \\ &= A + B, \end{aligned}$$

and note that

$$\varepsilon \log \alpha_a^\varepsilon \;=\; \varepsilon \log A \;+\; O(B/A).$$

Now

$$\begin{aligned} B &\leq \int_{|x| \geq R} \exp\left(\frac{-\gamma_1 |x|^2 + \gamma_2}{\varepsilon}\right)\, dx \\ &\leq C_R \exp\left(\frac{C_1 - C_2 R^2}{\varepsilon}\right) \\ &\leq C_R \exp\left(-C'/\varepsilon\right) \end{aligned}$$

where $C_R, C_1, C_2 > 0$, and $C' > 0$ if R is chosen sufficiently large. Also

$$\begin{aligned} \varepsilon \log A &\leq \varepsilon \log \int_{|x| \leq R} \exp\left(\bar{F}_a^\varepsilon / \varepsilon\right)\, dx \\ &\leq \varepsilon \log C_m R^m + \bar{F}_a^\varepsilon \end{aligned}$$

where R is chosen large enough to ensure $\operatorname{argmax}_{x \in \mathbf{R}^m} F_a^\varepsilon(x) \subset B(0, R)$ for all $a \in A$ and all sufficiently small ε. Thus

$$\varepsilon \log \alpha_a^\varepsilon \;\leq\; \bar{F}_a + 3\delta. \tag{6.3}$$

for all $\varepsilon > 0$ sufficiently small and all $a \in A$.

Combining (6.2) and (6.3) we obtain

$$\sup_{a \in A} \left| \varepsilon \log \alpha_a^\varepsilon - \bar{F}_a \right| \;<\; 3\delta$$

for all $\varepsilon > 0$ sufficiently small. This proves the Lemma. $\square$

References

[1] J.S. Baras, A. Bensoussan and M.R. James, *Dynamic Observers as Asymptotic Limits of Recursive Filters: Special Cases*, SIAM J. Applied Math., **48**(5) (1988) 1147—1158.

[2] J.S. Baras and M.R. James, *Robust Output Feedback Control for Nonlinear Discrete–Time Systems*, in preparation.

[3] T. Basar and P. Bernhard, "H^∞–Optimal Control and Related Minimax Design Problems: A Dynamic Game Approach", Birkhauser, Boston, 1991.

[4] A. Bensoussan and J.H. van Schuppen, *Optimal Control of Partially Observable Stochastic Systems with an Exponential-of-Integral Performance Index*, SIAM J. Control Optim., **23** (1985) 599—613.

[5] M. Campi and M.R. James, *Discrete–Time Nonlinear Risk–Sensitive Control* preprint, 1992.

[6] J.C. Doyle, K. Glover, P.P. Khargonekar and B.A. Francis, *State–Space Solutions to Standard H_2 and H_∞ Control Problems*, IEEE Trans. Aut. Control, **AC–34** (8) (1989) 831—847.

[7] R.J. Elliott, "Stochastic Calculus and Applications", Springer Verlag, New York, 1982.

[8] R.J. Elliott, "Viscosity Solutions and Optimal Control", Longman Scientific and Technical, Pitman Research Notes in Mathematics Series, London, 1987.

[9] R.J. Elliott and J.B. Moore, *Discrete Time Partially Observed Control*, preprint, 1992.

[10] W.H. Fleming and W.M. McEneaney, *Risk Sensitive Optimal Control and Differential Games*, Proc. Conf. on Adaptive and Stochastic Control, Univ. of Kansas, Springer Verlag, 1991.

[11] M.I. Friedlin and A.D. Wentzell, "Random Perturbations of Dynamical Systems", Springer–Verlag, New York, 1984.

[12] K. Glover and J.C. Doyle, *State Space Formulae for all Stabilizing Controllers that Satisfy an H_∞ Norm Bound and Relations to Risk Sensitivity*, Systems and Control Letters, **11** (1988) 167—172.

[13] O. Hijab, "Minimum Energy Estimation", Ph.D. Dissertation, Univ. of Calif., Berkeley, 1980.

[14] O. Hijab, *Asymptotic Bayesian Estimation of a First Order Equation with a Small Diffusion*, Annals of Prob., **12** (1984) 890—902.

[15] D.H. Jacobson, *Optimal Stochastic Linear Systems with Exponential Performance Criteria and Their Relation to Deterministic Differential Games*, IEEE Trans. Aut. Control, **AC–18** (2) (1973) 124—131.

[16] M.R. James, *Asymptotic Analysis of Nonlinear Stochastic Risk–Sensitive Control and Differential Games*, Math. Control, Signals, and Systems, **5**(4) (1992) 401—417.

[17] M.R. James, *Finite Time Observer Design by Probabilistic–Variational Methods*, SIAM J. Control Optim., **29**(4) (1991) 954—967.

[18] M.R. James and J.S. Baras, *Nonlinear Filtering and Large Deviations: A PDE-Control Theoretic Approach*, Stochastics, **23**(3) (1988) 391—412.

[19] P.R. Kumar and P. Varaiya, "Stochastic Systems: Estimation, Identification, and Adaptive Control", Prentice-Hall, Englewood Cliffs, 1986.

[20] D.J.N. Limebeer, B.D.O. Anderson, P.P. Khargonekar and M. Green, *A Game Theoretic Approach to H_∞ Control for Time Varying Systems*, SIAM J. Control Optim., **30** (1992) 262—283.

[21] I. Rhee and J.L. Speyer, *A Game Theoretic Approach to a Finite–Time Disturbance Attenuation Problem*, IEEE Trans. Aut. Control, **36** (1991) 1021—1032.

[22] P. Whittle, *Risk–Sensitive Linear/Quadratic/Gaussian Control*, Adv. Appl. Prob., **13** (1981) 764—777.

[23] P. Whittle, *A Risk–Sensitive Maximum Principle*, Systems and Control Letters, **15** (1990) 183—192.

[24] P. Whittle, "Risk–Sensitive Optimal Control", Wiley, New York, 1990.

[25] P. Whittle, *A Risk–Sensitive Maximum Principle: The Case of Imperfect State Observation*, IEEE Trans. Aut. Control, **36** (1991) 793—801.