[J. Math. Systems, Estimation & Control, 7(3), 1997, 371-374]

Robust and Risk-Sensitive Output Feedback Control for Finite State Machines and Hidden Markov Models

J.S. Baras* M.R. James†

November 8, 1995

Abstract

The purpose of this paper is to develop a framework for designing controllers for finite state systems which are robust with respect to uncertainties. A deterministic model for uncertainties is introduced, leading to a dynamic game formulation of the robust control problem. This problem is solved using an appropriate information state. A risk-sensitive stochastic control problem is formulated and solved for Hidden Markov Models, corresponding to situations where the model for the uncertainties is stochastic. The two problems are related using small noise limits.

Key words: Robust control, output feedback dynamic games, finite state machines, output feedback risk-sensitive stochastic optimal control, hidden Markov models.

AMS subject classifications (1991): 93B36, 93C41, 49K35, 93E20.

1 Introduction

A *finite state machine* (FSM) is a discrete–time system defined by the model

$$\begin{cases} x_{k+1} = f(x_k, u_k), \\ y_{k+1} = g(x_k), \quad k = 0, 1, \ldots, M, \end{cases} \tag{1.1}$$

*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 Grant NSFD CDR 8803012 through the Engineering Research Centers Program.

†Department of Engineering, Faculty of Engineering and Information Technology, Australian National University, Canberra, ACT 0200, Australia. Phone: Int + 61 6 249 4889, Fax: Int + 61 6 249 0506, E-mail: Matthew.James@anu.edu.au WWW: http://spigot.anu.edu.au/people/mat/home.html
Research supported in part by the Cooperative Research Centre for Robust and Adaptive Systems.

where the state x_k evolves in a finite set $\mathbf{X}$, and the control u_k and output y_k take values in finite sets $\mathbf{U}$ and $\mathbf{Y}$, respectively. These sets have n, m, and p elements, respectively. The behavior of the FSM is described by a state transition map $f : \mathbf{X} \times \mathbf{U} \to \mathbf{X}$ and an output map $g : \mathbf{X} \to \mathbf{Y}$.

FSM models, together with accompanying optimal control problems, have been used widely in applications. However, it is typically the case that deterministic treatments of such problems do not specifically deal with *disturbances*, e.g., as arising from modelling errors, sensor noise, etc. In this paper we propose and solve a general *robust* control problem for FSMs, paralleling the framework that has been developed for linear systems (e.g. Doyle *et al* [2]). The approach we adopt is motivated by [6], [7]. We thus develop a general framework for robust output feedback control of FSMs which specifically incorporates a deterministic model for disturbances and their effects.

Hidden Markov Models (HMM) are a different but closely related class of models, and numerous filtering, estimation, and control problems for them have been proposed and employed in applications. These models use a *probabilistic* description of disturbances. However, the majority of applications to date use a *risk-neutral* stochastic optimal control formulation. It is clear from the work of Jacobson [4], Whittle [11] and others that a controller more conservative than the risk-neutral one can be very useful. Indeed, it is well known that risk-sensitive controllers are very closely related to robust controllers, see [2], [6]. Here, we formulate and solve such a *risk-sensitive* stochastic optimal control problem for HMMs. Our solution, which is interesting in itself, leads us to the solution of the robust control problem for FSMs mentioned above. This is achieved by using a HMM which is designed to be a small random perturbation of the FSM (1.1), and employing large deviation limits as in [6]. It is possible to solve the robust control problem directly using an appropriate information state, once it is known, as in [7] (the large deviation limit identifies an information state).

The robust control problem for FSMs is formulated in §2; this entails defining a deterministic disturbance mechanism with associated cost functions. In §3, a stochastic disturbance model and a risk-sensitive control problem are defined. The risk-sensitive problem is solved, and a small noise limit is evaluated and used in §4 to solve the robust control problem of §2.

2 Formulation of the Robust Control Problem

2.1 Deterministic Perturbation

The FSM model (1.1) predicts that if the current state is x and a control input u is applied, the next state will be $x' = f(x, u)$. However, a disturbance may affect the actual system and result in a transfer to a state $x'' \neq x'$ instead. Similarly, the model predicts the next output to be $y' = g(x)$, whereas a disturbance may cause an output $y'' \neq y'$ to be observed. Additionally, the initial state x_0 may be unknown, and consequently we shall regard it as a disturbance.

We model the influence of disturbances as follows. Consider the following FSM model with two additional (disturbance) inputs w and v:

$$\begin{cases} x_{k+1} &= b(x_k, u_k, w_k), \\ y_{k+1} &= h(x_k, v_k), \qquad k = 0, 1, \ldots, M, \end{cases} \tag{2.1}$$

where, w_k and v_k take values in finite sets $\mathbf{W}$ and $\mathbf{V}$ respectively, and as in (1.1), $x_k \in \mathbf{X}$, $y_k \in \mathbf{Y}$, $u_k \in \mathbf{U}$. Thus $x'' = b(x, u, w)$ for some $w \in \mathbf{W}$, and $y'' = h(x, v)$ for some $v \in \mathbf{V}$.

The functions $b : \mathbf{X} \times \mathbf{U} \times \mathbf{W} \to \mathbf{X}$ and $h : \mathbf{X} \times \mathbf{V} \to \mathbf{Y}$ are required to satisfy the following consistency conditions:

$$\begin{cases} \text{there exists } w_\emptyset \in \mathbf{W} \text{ such that} \\ b(x, u, w_\emptyset) \;=\; f(x, u) \text{ for all } x \in \mathbf{X}, u \in \mathbf{U}, \end{cases} \tag{2.2}$$

$$\begin{cases} \text{there exists } v_\emptyset \in \mathbf{V} \text{ such that} \\ h(x, v_\emptyset) \;=\; g(x) \text{ for all } x \in \mathbf{X}. \end{cases} \tag{2.3}$$

The symbols $w_\emptyset$ and $v_\emptyset$ [8] play the role of "zero inputs", so that when no disturbances are present (i.e. $w_k \equiv w_\emptyset$, and $v_k \equiv v_\emptyset$), the behavior of (2.1) is the same as (1.1).

We will assume that there exists a null control $u_\emptyset \in \mathbf{U}$ and an equilibrium or rest state $x_\emptyset \in \mathbf{X}$ such that

$$x_\emptyset \;=\; f(x_\emptyset, u_\emptyset).$$

The set of possible initial states is denoted $N_0 \subset \mathbf{X}$, and assumed to contain $x_\emptyset$, while the set of possible future states for the disturbance model (2.1) is

$$N_{\mathbf{X}}(x, u) = \{b(x, u, w) \,:\, w \in \mathbf{W}\} \;\subset\; \mathbf{X},$$

and the corresponding set of possible future outputs is

$$N_{\mathbf{Y}}(x) = \{h(x, v) \,:\, v \in \mathbf{V}\} \;\subset \mathbf{Y}.$$

These sets can be thought of as "neighborhoods" of the nominal future values $f(x, u)$, $g(x)$, and are determined by the maps b and h. These can be designed as appropriate for the application at hand.

2.2 Cost Functions

To quantify the effect of the disturbances, a measure of their "sizes" is required. To this end, one specifies functions

$$\phi_w : \mathbf{W} \times \mathbf{X} \times \mathbf{U} \to \mathbf{R}, \qquad \phi_v : \mathbf{V} \times \mathbf{X} \to \mathbf{R}, \qquad \beta : \mathbf{X} \to \mathbf{R},$$

with the following properties:

$$\begin{cases} \phi_w(w_\emptyset; x, u) = 0 \;\text{ for all } x \in \mathbf{X}, u \in \mathbf{U}, \\ \phi_w(w; x, u) > 0 \;\text{ for all } w \neq w_\emptyset \in \mathbf{W}, x \in \mathbf{X}, u \in \mathbf{U}, \end{cases} \tag{2.4}$$

$$\begin{cases} \phi_v(v_\emptyset; x) = 0 \text{ for all } x \in \mathbf{X}, \\ \\ \phi_v(v; x) > 0 \text{ for all } v \neq v_\emptyset \in \mathbf{V}, x \in \mathbf{X}, u \in \mathbf{U}, \end{cases} \tag{2.5}$$

and

$$\begin{cases} \beta(x_\emptyset) = 0, \\ \\ +\infty > \beta(x_0) \geq 0 \text{ for all } x_0 \neq x_\emptyset \in N_0, \\ \\ \beta(x_0) = +\infty \text{ for all } x_0 \notin N_0, \end{cases} \tag{2.6}$$

We think of $\phi_w(w; x, u)$ as the magnitude of the disturbance w as it affects the system when it is in state x with control u applied, and $\phi_v(v; x)$ as the magnitude of the disturbance v when in state x. Of course, null disturbances are assigned zero cost. The cost function β specifies the amount of uncertainty regarding the initial state. The two extreme choices are (i) $\beta(x_0) = 0$ for all $x_0 \in \mathbf{X}$ (high uncertainty), and (ii) $\beta(x_0) = -\delta_\emptyset$ (certainty), where

$$\delta_{x'}(x) \;=\; \begin{cases} 0 & \text{if } x = x', \\ \\ -\infty & \text{if } x \neq x'. \end{cases}$$

Associated with these cost functions are quantities which define the optimal cost of transfering from x to x'' and the optimal cost of producing the output y''. These quantities will be used in the solution of the robust control problem below. They are defined by

$$\begin{aligned} U(x, x''; u) &\stackrel{\triangle}{=} \min\nolimits_{w \in \mathbf{W}} \left\{ \phi_w(w; x, u) \,:\, x'' = b(x, u, w) \right\}, \\ V(x, y'') &\stackrel{\triangle}{=} \min\nolimits_{v \in \mathbf{V}} \left\{ \phi_v(v; x) \,:\, y'' = h(x, v) \right\}. \end{aligned} \tag{2.7}$$

We adopt the convention that the minimum over an empty set equals $+\infty$. Thus U and V are extended real valued functions. Note that

$$\begin{aligned} U(x, f(x, u); u) &= 0 \quad \text{for all } x \in \mathbf{X}, u \in \mathbf{U}, \\ U(x, b(x, u, w); u) &> 0 \quad \text{for all } w \neq w_\emptyset \in \mathbf{W}, x \in \mathbf{X}, u \in \mathbf{U}, \\ U(x, x''; u) &= +\infty \text{ if } x'' \notin N_{\mathbf{X}}(x, u). \end{aligned}$$

and

$$\begin{aligned} V(x, g(x)) &= 0 \quad \text{for all } x \in \mathbf{X}, \\ V(x, h(x, v)) &> 0 \quad \text{for all } v \neq v_\emptyset \in \mathbf{V}, x \in \mathbf{X}, \\ V(x, y'') &= +\infty \text{ if } y'' \notin N_{\mathbf{Y}}(x). \end{aligned}$$

2.3 Robust Control

As part of the problem specification, one defines an additional output quantity

$$z_{k+1} \;=\; \ell(x_k, u_k), \tag{2.8}$$

where z_k takes values in a finite set $\mathbf{Z}$, and $\ell : \mathbf{X} \times \mathbf{U} \to \mathbf{Z}$. We assume there exists a specific null element $z_\emptyset \in \mathbf{Z}$ such that

$$\ell(x_\emptyset, u_\emptyset) \;=\; z_\emptyset. \tag{2.9}$$

A cost function for this output is also specified, with the properties

$$\begin{cases} \phi_z(z_\emptyset) \;=\; 0, \\ \\ \phi_z(z) \;\geq\; 0 \;\; \text{for all } z \in \mathbf{Z}. \end{cases} \tag{2.10}$$

The output quantity z and its associated cost function ϕ_z encode the performance objective of the problem at hand. To summarize, the complete system is described by the equations

$$\begin{cases} x_{k+1} \;=\; b(x_k, u_k, w_k), \\ \\ z_{k+1} \;=\; \ell(x_k, u_k), \\ \\ y_{k+1} \;=\; h(x_k, v_k), \qquad k = 0, 1, \ldots, M. \end{cases} \tag{2.11}$$

The state variable x_k is not measured directly, and so the controller must make use of information available in the output signal $y_{0,k}$; i.e., the controller must be an *output feedback* controller. We denote by $\mathcal{O}_{k,l}$ the set of non–anticipating control policies defined on the interval $[k, l]$; i.e., those controls for which there exist functions $\bar{u}_j : \mathbf{Y}^{j-k+1} \to \mathbf{U}$ such that $u_j = \bar{u}_j(y_{k+1,j})$ for each $j \in [k, l]$.

The *output feedback robust control problem* we wish to solve is the following: given $\gamma > 0$ and a finite time interval $[0, M]$ find an output feedback controller $u \in \mathcal{O}_{0,M-1}$ such that

$$\sum_{k=0}^{M-1} \phi_z(z_{k+1}) \;\leq\; \beta(x_0) + \gamma \sum_{k=0}^{M-1} \left(\phi_w(w_k; x_k, u_k) + \phi_v(v_k; x_k)\right) \tag{2.12}$$

for all $(w, v) \in \mathbf{W}^M \times \mathbf{V}^M$, $x_0 \in \mathbf{X}$.

Remark 2.1 The formulation of this problem is similar to the H_∞ problem in the time domain for nonlinear systems [7], motivated by the formulation for linear systems [2].

2.4 Dynamic Game

The robust control problem formulated above can be recast as a dynamic game problem, see, e.g., [7] and the references contained therein. The payoff function for a controller $u \in \mathcal{O}_{0,M-1}$ (player 1) and disturbances $(w, v, x_0) \in \mathbf{W}^M \times \mathbf{V}^M \times \mathbf{X}$ (player 2) is given by

$$J^\gamma(u, w, v, x_0) \;\stackrel{\triangle}{=}\; -\beta(x_0) + \sum_{k=0}^{M-1} \phi_z(z_{k+1}) - \gamma \left(\phi_w(w_k; x_k, u_k) + \phi_v(v_k; x_k)\right).$$

We consider the upper game for this payoff given the dynamics (2.1). Define

$$J^\gamma(u) \stackrel{\triangle}{=} \max_{(w,v)\in\mathbf{W}^M\times\mathbf{V}^M} \max_{x_0\in\mathbf{X}} \left\{J^\gamma(u,w,v,x_0)\right\}.$$

The bound

$$0 \le J^\gamma(u) \le M \max_{z\in\mathbf{Z}} \phi_z(z) \tag{2.13}$$

is readily verified. The dynamic game problem is to find an output feedback controller $u^* \in \mathcal{O}_{0,M-1}$ such that

$$J^\gamma(u^*) = \min_{u\in\mathcal{O}_{0,M-1}} J^\gamma(u). \tag{2.14}$$

Then if

$$J^\gamma(u^*) = 0, \tag{2.15}$$

the robust control objective (2.12) is achieved.

We will solve this dynamic game problem in §4.

2.5 State Feedback Robust Control

For completeness and clarity, we consider now the solution of the robust control problem in the special case where complete state information is available. In this case, $g(x) \equiv x$ and $h(x,v) \equiv x$, so that $y_{k+1} = x_k$ for all k, and $\mathcal{O}_{k,l} = \mathcal{S}_{k,l}$, the class of $\mathbf{U}$–valued non–anticipating functions of the state $x_{k,l}$.

The *state feedback robust control problem* is to find a controller $u \in \mathcal{S}_{0,M-1}$ such that

$$\sum_{k=0}^{M-1} \phi_z(z_{k+1}) \le \gamma \sum_{k=0}^{M-1} \phi_w(w_k; x_k, u_k) \tag{2.16}$$

for all $w \in \mathbf{W}^M$, given $x_0 = x_\emptyset$, and $\gamma > 0$.

The equivalent dynamic game problem is to find $u^* \in \mathcal{S}_{0,M-1}$ such that

$$J^\gamma(u^*) = \min_{u\in\mathcal{S}_{0,M-1}} J^\gamma(u) = 0, \tag{2.17}$$

where now

$$J^\gamma(u) \stackrel{\triangle}{=} \max_{w\in\mathbf{W}^M} \left\{ \sum_{k=0}^{M-1} \phi_z(z_{k+1}) - \gamma\phi_w(w_k; x_k, u_k) \,:\, x_0 = x_\emptyset \right\}.$$

Note that

$$0 \le J^\gamma(u) \le M \max_{z\in\mathbf{Z}} \phi_z(z). \tag{2.18}$$

The upper value function for this game is defined by

$$\bar{f}_k^\gamma(x) \stackrel{\triangle}{=} \min_{u\in\mathcal{S}_{k,M-1}} \max_{w\in\mathbf{W}^{M-k}} \left\{ \sum_{l=k}^{M-1} \phi_z(z_{l+1}) - \gamma\phi_w(w_l; x_l, u_l) \,:\, x_k = x \right\}, \tag{2.19}$$

and the corresponding dynamic programming equation is

$$\begin{cases} \bar{f}_k^\gamma(x) = \min_{u\in\mathbf{U}} \max_{w\in\mathbf{W}} \{ \bar{f}_{k+1}^\gamma(b(x,u,w)) + \phi_z(\ell(x,u)) \\ \qquad\qquad -\gamma\phi_w(w;x,u)\} \\ \bar{f}_M^\gamma(x) = 0. \end{cases} \tag{2.20}$$

Theorem 2.2 (Necessity) *Assume that $u^s \in \mathcal{S}_{0,M-1}$ solves the state feedback robust control problem. Then there exists a solution $\bar{f}^\gamma$ to the dynamic programming equation (2.20) such that $\bar{f}_k^\gamma(x) \geq 0$, $\bar{f}_0^\gamma(x_\emptyset) = 0$.* (Sufficiency) *Assume that there exists a solution $\bar{f}^\gamma$ of the dynamic programming equation (2.20) such that $\bar{f}_k^\gamma(x) \geq 0$, $\bar{f}_0^\gamma(x_\emptyset) = 0$. Let $\tilde{u}_k^*(x)$ be the control value achieving the minimum in (2.20). Then $\tilde{u}_k^*(x)$ is a state feedback controller which solves the state feedback robust control problem (2.16).*

Proof. For $x \in \mathbf{X}$, $k \in [0, M]$ define $\bar{f}_k^\gamma(x)$ by (2.19). Then we have

$$0 \leq \bar{f}_k^\gamma(x) \leq \beta(x),$$

for some constant $\beta(x)$ depending on M. By dynamic programming $\bar{f}^\gamma$ solves equation (2.20). By definition,

$$\bar{f}_0^\gamma(x_\emptyset) \leq J^\gamma(u^s) \leq 0,$$

and so $\bar{f}_0^\gamma(x_\emptyset) = 0$. This proves the necessity part.

As for sufficiency, standard dynamic programming arguments and the hypotheses imply that

$$0 = \bar{f}_0^\gamma(x_\emptyset) = J^\gamma(u^*).$$

Therefore by (2.17) we see that u^* solves the state feedback robust control problem. □

3 A Risk–Sensitive Stochastic Control Problem

In this section we formulate and solve a risk-sensitive stochastic control problem for Hidden Markov Models (HMM). While the particular HMM treated is a random perturbation of the FSM (1.1), [5], the method applies more generally.

3.1 Random Perturbation

The random perturbation defined below is a stochastic analog of the deterministic perturbation introduced in §2, and indeed much of the notation from §2 will be used here.

A *controlled Hidden Markov Model* consists of an $\mathbf{X}$ valued controlled Markov chain x_k^ε together with a $\mathbf{Y}$ valued output process y_k^ε whose behavior is determined by a state transition matrix $A^\varepsilon(u)$ and an output probability matrix $B^\varepsilon(x)$ [9]. These matrices are defined by their components

$$\begin{aligned} A^\varepsilon(u)_{x,x''} &\stackrel{\triangle}{=} \tfrac{1}{Z^\varepsilon_{x,u}} \exp\left(-\tfrac{1}{\varepsilon}U(x,x'';u)\right), \\ B^\varepsilon(x)_{y''} &\stackrel{\triangle}{=} \tfrac{1}{Z^\varepsilon_{x}} \exp\left(-\tfrac{1}{\varepsilon}V(x,y'')\right), \end{aligned} \tag{3.1}$$

where the functions U and V are defined by (2.7), and the normalizing constants $Z^\varepsilon_{x,u}$ and Z^ε_x are chosen so that

$$\sum_{x''\in\mathbf{X}} A^\varepsilon(u)_{x,x''} = 1, \qquad \sum_{y''\in\mathbf{Y}} B^\varepsilon(x)_{y''} = 1.$$

Similarly, an initial distribution can be defined by

$$\rho^\varepsilon(x_0) = \tfrac{1}{Z^\varepsilon_{x_0}} \exp\left(-\tfrac{1}{\varepsilon}\beta(x_0)\right). \tag{3.2}$$

Thus

$$\mathbf{P}^u\left(x^\varepsilon_{k+1} = x'' \,|\, x_k = x, x_{0,k-1}, u_k = u, u_{0,k-1}\right) = A^\varepsilon(u)_{x,x''},$$

$$\text{and} \quad \mathbf{P}^u\left(y^\varepsilon_{k+1} = y'' \,|\, x_k = x\right) = B^\varepsilon(x)_{y''},$$

where $\mathbf{P}^u$ is the probability distribution on $\mathbf{X}^{M+1}\times\mathbf{Y}^M$ defined by a control policy $u \in \mathcal{O}_{0,M-1}$:

$$\mathbf{P}^u(x_{0,M}, y_{1,M}) = \Pi_{k=0}^{M-1} A^\varepsilon(u_k)_{x_k,x_{k+1}} B^\varepsilon(x_k)_{y_{k+1}} \rho^\varepsilon(x_0)$$

The HMM (3.1) satisfies the consistency conditions

$$\lim_{\varepsilon\to 0} A^\varepsilon(u) = A(u), \qquad \lim_{\varepsilon\to 0} B^\varepsilon(x) = B(x),$$

where

$$A(u)_{x,x''} = \begin{cases} 1 & \text{if } x'' = f(x,u), \\ 0 & \text{if } x'' \neq f(x,u), \end{cases}$$

$$B(x)_{y''} = \begin{cases} 1 & \text{if } y'' = g(x), \\ 0 & \text{if } y'' \neq g(x). \end{cases}$$

If β has a unique minimum at $x_\emptyset$, then

$$\lim_{\varepsilon\to 0} \rho^\varepsilon(x_0) = \rho(x_0),$$

where

$$\rho(x_0) = \begin{cases} 1 & \text{if } x_0 = x_\emptyset, \\ 0 & \text{if } x_0 \neq x_\emptyset. \end{cases}$$

Thus if $x^\varepsilon_0 \to x_\emptyset$, we have

$$(x^\varepsilon_{0,M}, y^\varepsilon_{1,M}) \to (x_{0,M}, y_{1,M})$$

in probability as $\varepsilon \to 0$, where $(x_{0,M}, y_{1,M})$ are the state and output paths for the FSM with the same control policy and initial condition. Therefore the HMM (3.1) is indeed a random perturbation of the FSM (1.1).

The probability distribution $\mathbf{P}^u$ is equivalent to a distribution $\mathbf{P}^\dagger$ under which $\{y_k^\varepsilon\}$ is iid uniformly distributed on $\mathbf{Y}$, independent of $\{x_k^\varepsilon\}$, and $\{x_k^\varepsilon\}$ is a controlled Markov chain as above:

$$\mathbf{P}^\dagger(x_{0,M}, y_{1,M}) \;=\; \Pi_{k=0}^{M-1}\left(A^\varepsilon(u_k)_{x_k,x_{k+1}}\frac{1}{p}\right)\rho^\varepsilon(x_0),$$

where p here denotes the number of outputs. Note that

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

where

$$\Psi^\varepsilon(x, y'') \;\stackrel{\triangle}{=}\; pB^\varepsilon(x)_{y''},$$

and $\mathcal{G}_k$ is the filtration generated by $(x_{0,k}^\varepsilon,\, y_{1,k}^\varepsilon)$.

Remark 3.1 λ_k^ε is the Radon-Nikodym derivative of $\mathbf{P}^u$ with respect to $\mathbf{P}^\dagger$ on the measurable space $(\mathbf{X}^{M+1}\times\mathbf{Y}^M, \mathcal{G}_k)$, and has the explcit form given. As a consequence, for a $\mathcal{G}_M$-measurable random variable Ξ,

$$\mathbf{E}^u[\Xi] \stackrel{\triangle}{=} \int \Xi d\mathbf{P}^u = \int \Xi\lambda_M^\varepsilon d\mathbf{P}^\dagger \stackrel{\triangle}{=} \mathbf{E}^\dagger[\lambda_M^\varepsilon \Xi].$$

3.2 Cost

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

$$J^{\gamma,\varepsilon}(u) \;=\; \mathbf{E}^u\left[\exp\frac{1}{\gamma\varepsilon}\sum_{l=0}^{M-1}\phi_z\left(\ell(x_l^\varepsilon, u_l)\right)\right] \tag{3.3}$$

and the *output feedback risk-sensitive stochastic control problem* for the HMM (3.1) is to find $u^* \in \mathcal{O}_{0,M-1}$ such that

$$J^{\gamma,\varepsilon}(u^*) \;=\; \min_{u\in\mathcal{O}_{0,M-1}} J^{\gamma,\varepsilon}(u).$$

Remark 3.2 The risk-sensitive cost (3.3) was introduced by Jacobson [4] in the context of linear systems and quadratic cost functions, and was refered to as LEQG. This cost measure was further studied by Whittle [10], [11] and others. The exponential appearing in (3.3) does not commute with the expectation, and so the risk-sensitive cost is fundamentally different to the more familiar risk-neutral cost

$$\mathbf{E}^u\left[\sum_{l=0}^{M-1}\phi_z\left(\ell(x_l^\varepsilon, u_l)\right)\right].$$

The exponential has the effect of heavily penalizing large values of the sum, leading to a controller which is more conservative, or risk-sensitive. This latter terminology comes from economics.

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

$$J^{\gamma,\varepsilon}(u) \;=\; \mathbf{E}^\dagger\left[\lambda_M^\varepsilon\exp\frac{1}{\gamma\varepsilon}\sum_{l=0}^{M-1}\phi_z\left(\ell(x_l^\varepsilon, u_l)\right)\right]. \tag{3.4}$$

3.3 Information State

Following [6], we define an information state process $\sigma_k^{\gamma,\varepsilon} \in \mathbf{R}^n$ by the relation

$$\sigma_k^{\gamma,\varepsilon}(x) \;=\; \mathbf{E}^{\dagger}\left[I_{\{x_k^\varepsilon = x\}} \exp \frac{1}{\gamma\varepsilon} \sum_{l=0}^{k-1} \phi_z\left(\ell(x_l^\varepsilon, u_l)\right) \lambda_k^\varepsilon \,|\, \mathcal{Y}_k \right], \tag{3.5}$$

where $\mathcal{Y}_k$ is the filtration generated by the observation process $y_{1,k}^\varepsilon$, and $\sigma_0^{\gamma,\varepsilon}(x) = I_{\{x = x_\emptyset\}}$.

The evolution of this process is determined by a matrix $\Sigma^{\gamma,\varepsilon}(u, y'')$ whose entries are defined by

$$\Sigma^{\gamma,\varepsilon}(u, y'')_{x,x''} \;\stackrel{\triangle}{=}\; A^\varepsilon(u)_{x,x''}\, \Psi^\varepsilon(x, y'') \,\exp \frac{1}{\gamma\varepsilon}\phi_z\left(\ell(x,u)\right). \tag{3.6}$$

Indeed, the information state is the solution of the recursion

$$\begin{cases} \sigma_k^{\gamma,\varepsilon} \;=\; \Sigma^{\gamma,\varepsilon\,*}(u_{k-1}, y_k^\varepsilon)\, \sigma_{k-1}^{\gamma,\varepsilon} \\ \sigma_0^{\gamma,\varepsilon} \;=\; \rho^\varepsilon, \end{cases} \tag{3.7}$$

where the $*$ denotes matrix transpose.

Remark 3.3 The recursion (3.7) is similar to the recursion formula for the unnormalized conditional distribution in nonlinear filtering, see, e.g. [9]. The derivation of (3.7) appears in [6].

We can also define an adjoint process $\nu_k^{\gamma,\varepsilon} \in \mathbf{R}^n$ by the recursion

$$\begin{cases} \nu_{k-1}^{\gamma,\varepsilon} \;=\; \Sigma^{\gamma,\varepsilon}(u_{k-1}, y_k^\varepsilon)\, \nu_k^{\gamma,\varepsilon} \\ \nu_M^{\gamma,\varepsilon} \;=\; 1. \end{cases} \tag{3.8}$$

Then relative to the inner product

$$\langle \sigma, \nu \rangle \;=\; \sum_{x \in \mathbf{X}} \sigma(x)\nu(x)$$

for $\sigma, \nu \in \mathbf{R}^n$, it is straightforward to establish the adjoint relationships

$$\begin{aligned} \langle \Sigma^{\gamma,\varepsilon\,*}\sigma, \nu \rangle \;&=\; \langle \sigma, \Sigma^{\gamma,\varepsilon}\nu \rangle, \\ \langle \sigma_k^{\gamma,\varepsilon}, \nu_k^{\gamma,\varepsilon} \rangle \;&=\; \langle \sigma_{k-1}^{\gamma,\varepsilon}, \nu_{k-1}^{\gamma,\varepsilon} \rangle \end{aligned} \tag{3.9}$$

for all $\sigma \in \mathbf{R}^n$, $\nu \in \mathbf{R}^n$, and all k.

Remark 3.4 The reason for introducing the information state $\sigma_k^{\gamma,\varepsilon}$ is to replace the original output feedback risk-sensitive stochastic control problem with an equivalent stochastic control problem with a state variable $\sigma_k^{\gamma,\varepsilon}$ which is completely observed, and to solve this new problem using dynamic programming. This will yield a state feedback controller for the new problem, or equivalently, an output feedback controller for the original problem which is *separated* through the information state [9], [1], [6].

As in [6], the cost function can be expressed purely in terms of the information state:

$$J^{\gamma,\varepsilon}(u) \;=\; \mathbf{E}^{\dagger}\left[\langle \sigma_M^{\gamma,\varepsilon},\, 1\rangle\right]. \tag{3.10}$$

To see this, observe that for $u \in \mathcal{O}_{0,M-1}$

$$\begin{aligned}
&\mathbf{E}^{\dagger}\left[\langle \sigma_M^{\gamma,\varepsilon},\, 1\rangle\right]\\
&= \mathbf{E}^{\dagger}\left[\textstyle\sum_{x\in\mathbf{X}} \mathbf{E}^{\dagger}\left[I_{\{x_M^{\varepsilon}=x\}} \exp \frac{1}{\gamma\varepsilon} \sum_{l=0}^{M-1} \phi_z\left(\ell(x_l^{\varepsilon}, u_l)\right) \lambda_M^{\varepsilon} \,|\, \mathcal{Y}_M\right]\right]\\
&= \mathbf{E}^{\dagger}\left[\exp \tfrac{1}{\gamma\varepsilon} \textstyle\sum_{l=0}^{M-1} \phi_z\left(\ell(x_l^{\varepsilon}, u_l)\right) \lambda_M^{\varepsilon}\right]\\
&= J^{\gamma,\varepsilon}(u).
\end{aligned}$$

3.4 Dynamic Programming

Consider the state $\sigma^{\gamma,\varepsilon}$ on the interval $k, \ldots, M$ with initial condition $\sigma_k^{\gamma,\varepsilon} = \sigma \in \mathbf{R}^n$:

$$\begin{cases} \sigma_l^{\gamma,\varepsilon} \;=\; \Sigma^{\gamma,\varepsilon\,*}(u_{l-1}, y_l^{\varepsilon})\, \sigma_{l-1}^{\gamma,\varepsilon}, \quad k+1 \le l \le M, \\ \sigma_k^{\gamma,\varepsilon} \;=\; \sigma. \end{cases} \tag{3.11}$$

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

$$S^{\gamma,\varepsilon}(\sigma, k) \;=\; \min_{u\in\mathcal{O}_{k,M-1}} \mathbf{E}^{\dagger}\left[\langle \sigma_M^{\gamma,\varepsilon},\, 1\rangle \,|\, \sigma_k^{\gamma,\varepsilon} = \sigma\right]. \tag{3.12}$$

The dynamic programming equation for this problem is as follows [6]:

$$\begin{cases} S^{\gamma,\varepsilon}(\sigma, k) \;=\; \min_{u\in U} \mathbf{E}^{\dagger}\left[S^{\gamma,\varepsilon}(\Sigma^{\gamma,\varepsilon\,*}(u, y_{k+1}^{\varepsilon})\sigma, k+1)\right] \\ S^{\gamma,\varepsilon}(\sigma, M) \;=\; \langle \sigma,\, 1\rangle. \end{cases} \tag{3.13}$$

The next theorem is a statement of the dynamic programming solution to the output feedback risk-sensitive stochastic control problem.

Theorem 3.5 *The value function $S^{\gamma,\varepsilon}$ defined by (3.12) is the unique solution to the dynamic programming equation (3.13). Conversely, assume that $S^{\gamma,\varepsilon}$ is the solution of the dynamic programming equation (3.13). Suppose that $u^* \in \mathcal{O}_{0,M-1}$ is a policy such that, for each $k = 0, \ldots, M-1$, $u_k^* = \bar{u}_k^*(\sigma_k^{\gamma,\varepsilon})$, where $\bar{u}_k^*(\sigma)$ achieves the minimum in (3.13). Then u^* is an optimal output feedback controller for the risk–sensitive stochastic control problem (§3.2).*

Proof. The proof is similar to those of Theorems 2.5 and 2.6 in [6] (see also [9]).

Let $S^{\gamma,\varepsilon}(\sigma, k)$ denote the solution to the dynamic programming equation (3.13). We will show that $S^{\gamma,\varepsilon}(\sigma, k)$ equals the RHS of (3.12) and that the policy u^* is optimal. To this end, define

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

We claim that, for all $u \in \mathcal{O}_{k,M-1}$,

$$S^{\gamma,\varepsilon}(\sigma,k) \;\leq\; \bar{S}^{\gamma,\varepsilon}(\sigma,k;u) \tag{3.14}$$

for each $k = 0, 1, \ldots, M$, with equality if $u = u^*$.

For $k = M$, (3.14) is clearly satisfied. Assume now that (3.14) holds for $k+1, \ldots, M$. Then

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

using the induction hypothesis and (3.13). If $u = u^*$ the above inequalities are replaced by equalities. This proves (3.14), and hence that $S^{\gamma,\varepsilon}(\sigma,k)$ equals the RHS of (3.12).

From (3.14), setting $k = 0$ and $\sigma = \rho^\varepsilon$ we obtain

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

for any $u \in \mathcal{O}_{0,M-1}$. This implies

$$J^{\gamma,\varepsilon}(u^*) \;\leq\; J^{\gamma,\varepsilon}(u)$$

for all $u \in \mathcal{O}_{0,M-1}$. This completes the proof. □

Remark 3.6 Note that the controller u_k^* is defined as a function of the information state $\sigma_k^{\gamma,\varepsilon}$, and since $\sigma_k^{\gamma,\varepsilon}$ is a non-anticipating function of $y^\varepsilon_{0,k}$, u_k^* is an output feedback controller for the risk-sensitive stochastic control problem; indeed, u^* is an information state feedback controller.

3.5 Small Noise Limit

In [6] it was shown that a deterministic dynamic game problem is obtained as a small noise limit of a risk–sensitive stochastic control problem. In this subsection, we carry out this limit procedure for the risk-sensitive stochastic control problem defined above. We first obtain a limit for the information state, and use this to evaluate the appropriate limit for the value function. This yields an information state and value function for the dynamic game problem of §2.4. These results will be used in §4 in the solution of the output feedback robust control problem of §2.

Define the matrix $\Lambda^\gamma(u, y'')$ by its entries

$$\Lambda^\gamma(u,y'')_{x,x''} \;\stackrel{\triangle}{=}\; \phi_z(\ell(x,u)) - \gamma\,(U(x,x'';u) + V(x,y''))\,. \tag{3.15}$$

Then we have

$$\lim_{\varepsilon \to 0} \gamma\varepsilon \log \Sigma^{\gamma,\varepsilon}(u,y'')_{x,x''} \;=\; \Lambda^\gamma(u,y'')_{x,x''}. \tag{3.16}$$

The action of the matrix $\Sigma^{\gamma,\varepsilon}$ and its adjoint (transpose) on vectors σ, ν in $\mathbf{R}^n$ is given by the usual matrix multiplication, i.e., sums of products of entries. The action of the matrix $\Lambda^\gamma(u, y'')$ and its adjoint on vectors p, q in $\mathbf{R}^n$ is instead defined in terms of maximization operations as follows:

$$\begin{aligned} \Lambda^{\gamma\,*}(u, y'')p(x'') &\stackrel{\triangle}{=} \max\nolimits_{x\in\mathbf{X}} \left\{\Lambda^\gamma(u, y'')_{x,x''} + p(x)\right\}, \\ \Lambda^\gamma(u, y'')q(x) &\stackrel{\triangle}{=} \max\nolimits_{x''\in\mathbf{X}} \left\{\Lambda^\gamma(u, y'')_{x,x''} + q(x'')\right\}. \end{aligned} \tag{3.17}$$

The inner product $\langle\,\cdot\,,\,\cdot\,\rangle$ is replaced by the "sup-pairing"

$$(p,\, q) \stackrel{\triangle}{=} \max_{x\in\mathbf{X}} \left\{p(x) + q(x)\right\}, \tag{3.18}$$

and in fact we have

$$\lim_{\varepsilon\to 0} \gamma\varepsilon \log\langle e^{\frac{1}{\gamma\varepsilon}p},\, e^{\frac{1}{\gamma\varepsilon}q}\rangle \;=\; (p,\, q). \tag{3.19}$$

The actions corresponding to the matrix $\Lambda^\gamma(u, y'')$ are "adjoint" in the sense that

$$(\Lambda^{\gamma\,*}p,\, q) \;=\; (p,\, \Lambda^\gamma q)\,. \tag{3.20}$$

The limit result for the information state is the following:

Theorem 3.7 *We have*

$$\begin{aligned} \lim\nolimits_{\varepsilon\to 0} \gamma\varepsilon \log \Sigma^{\gamma,\varepsilon\,*}(u, y) e^{\frac{1}{\gamma\varepsilon}p} &= \Lambda^{\gamma\,*}(u, y)p, \\ \lim\nolimits_{\varepsilon\to 0} \gamma\varepsilon \log \Sigma^{\gamma,\varepsilon}(u, y) e^{\frac{1}{\gamma\varepsilon}q} &= \Lambda^\gamma(u, y)q \end{aligned} \tag{3.21}$$

in $\mathbf{R}^n$ *uniformly on* $\mathbf{U}\times\mathbf{Y}\times\mathbf{R}^n$*.*

Proof. First note that, using Lemma A.1,

$$\lim_{\varepsilon\to 0} \varepsilon \log Z^\varepsilon_{x,u} = 0, \quad \text{and} \quad \lim_{\varepsilon\to 0} \varepsilon \log Z^\varepsilon_x = 0.$$

Next, for $a = (x'', u, y, p)$ define

$$\begin{aligned} F^\varepsilon_a(x) &= F_a(x) - \varepsilon \log Z^\varepsilon_{x,u} - \varepsilon \log Z^\varepsilon_x, \\ F_a(x) &= (\Lambda^\gamma(u, y)_{x,x''} + p(x))/\gamma. \end{aligned}$$

Then

$$\Sigma^{\gamma,\varepsilon\,*}(u, y) e^{\frac{1}{\gamma\varepsilon}p}(x'') = \sum_{x\in\mathbf{X}} e^{F^\varepsilon_a(x)/\varepsilon},$$

and the first limit result follows using Lemma A.1. The second limit is proven similarly. □

In view of this theorem, we define a limit information state and its adjoint by the recursions

$$\begin{cases} p^\gamma_k = \Lambda^{\gamma\,*}(u_{k-1}, y_k) p^\gamma_{k-1} \\ p^\gamma_0 = -\beta, \end{cases} \tag{3.22}$$

and

$$\begin{cases} q^\gamma_{k-1} = \Lambda^\gamma(u_{k-1}, y_k) q^\gamma_k \\ q^\gamma_M = 0. \end{cases} \tag{3.23}$$

Note that

$$(p^\gamma_k,\, q^\gamma_k) \;=\; (p^\gamma_{k-1},\, q^\gamma_{k-1})$$

for all k.

Turning now to the value function, we have:

Theorem 3.8 *The function $W^\gamma(p,k)$ defined for $p \in \mathbf{R}^n$ by*

$$W^\gamma(p,k) \;\stackrel{\triangle}{=}\; \lim_{\varepsilon\to 0} \gamma\varepsilon \log S^{\gamma,\varepsilon}(e^{\frac{1}{\gamma\varepsilon}p}, k) \tag{3.24}$$

exists (i.e. the sequence converges uniformly on $\mathbf{R}^n$), is continuous, and satisfies the recursion

$$\begin{cases} W^\gamma(p,k) = \min_{u\in\mathbf{U}} \max_{y\in\mathbf{Y}} \{W^\gamma(\Lambda^{\gamma\,*}(u,y)p, k+1)\} \\ W^\gamma(p,M) = (p,\, 0)\,. \end{cases} \tag{3.25}$$

Proof. First note that the solution of the recursion (3.25) is a continuous function on $\mathbf{R}^n$ for each k (this is readily verified by induction).

Assume the conclusions hold for $k+1, \ldots, M$. From Theorem 3.5 and (3.24), we need to compute

$$\lim_{\varepsilon\to 0} \gamma\varepsilon \log S^{\gamma,\varepsilon}(e^{\frac{1}{\gamma\varepsilon}p}, k)$$

$$= \lim_{\varepsilon\to 0} \gamma\varepsilon \log \min_{u\in U} \mathbf{E}^\dagger \left[S^{\gamma,\varepsilon}(\Sigma^{\gamma,\varepsilon\,*}(u, y^\varepsilon_{k+1}) e^{\frac{1}{\gamma\varepsilon}p}, k+1) \right] \tag{3.26}$$

$$= \lim_{\varepsilon\to 0} \min_{u\in U} \gamma\varepsilon \log \mathbf{E}^\dagger \left[S^{\gamma,\varepsilon}(\Sigma^{\gamma,\varepsilon\,*}(u, y^\varepsilon_{k+1}) e^{\frac{1}{\gamma\varepsilon}p}, k+1) \right].$$

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

By Theorem 3.7,

$$p^\varepsilon_a \stackrel{\triangle}{=} \gamma\varepsilon \log \Sigma^{\gamma,\varepsilon,*}(u, y'') e^{\frac{1}{\gamma\varepsilon}p} \to p_a \stackrel{\triangle}{=} \Lambda^\gamma(u, y'')p$$

as $\varepsilon \to 0$ uniformly in $a = (u, y'', p)$. It then follows from the induction hypothesis and continuity that

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

uniformly in a. Next, applying Lemma A.1, we obtain

$$\lim_{\varepsilon\to 0} \gamma\varepsilon \log \textstyle\sum_{y''\in\mathbf{Y}} \frac{1}{p} S^{\gamma,\varepsilon}(\Sigma^{\gamma,\varepsilon\,*}(u, y'') e^{\frac{1}{\gamma\varepsilon}p}, k+1)$$

$$= \max_{y''\in\mathbf{Y}} W^\gamma(\Lambda^{\gamma,*}(u, y'')p, k+1)$$

uniformly in (u,p). Therefore the RHS of (3.26) equals

$$\min_{u\in\mathbf{U}} \max_{y\in\mathbf{Y}} \{W^\gamma(\Lambda^{\gamma\,*}(u,y)p, k+1)\}\,,$$

and the limit is uniform in p. □

4 Solution to the Robust Control Problem

4.1 Equivalent Game Problem

We now replace the deterministic output feedback game problem (§2) with an equivalent deterministic game problem with p_k^γ, defined in §3.5, as a completely observed state variable. The solution of this new problem will result in an information state feedback controller, and thus an output feedback controller for the original game problem which is *separated* through the information state.

The next theorem shows that the cost function can be expressed in terms of the information state [6], [1].

Theorem 4.1 *We have for all* $u \in \mathcal{O}_{0,M-1}$

$$J^\gamma(u) \;=\; \max_{y \in \mathbf{Y}^M} \{(p_M^\gamma, 0)\}. \tag{4.1}$$

PROOF. We have, for all $u \in \mathcal{O}_{0,M-1}$,

$$\begin{aligned}
&\max\nolimits_{y\in\mathbf{Y}^M}\{(p_M^\gamma, 0)\}\\
&= \max\nolimits_{y\in\mathbf{Y}^M}\max\nolimits_{x''\in\mathbf{X}}\{p_M^\gamma(x'')\}\\
&= \max\nolimits_{y\in\mathbf{Y}^M}\max\nolimits_{\xi\in\mathbf{X}^{M+1}}\{-\beta(\xi_0)+\textstyle\sum_{l=0}^{M-1}\phi_x(\ell(\xi_l,u_l))\\
&\qquad -\gamma\,(U(\xi_l,\xi_{l+1};u_l)+V(\xi_l,y_{l+1}))\}\\
&= \max\nolimits_{w\in\mathbf{W}^M}\max\nolimits_{v\in\mathbf{V}^M}\max\nolimits_{x_0\in\mathbf{X}}\{J^\gamma(u,w,v,x_0)\}\\
&= J^\gamma(u),
\end{aligned}$$

where we have made use of the definitions for the cost functions U and V (§2.2). □

4.2 Dynamic Programming

Consider now the state p^γ on the interval $k, \ldots, M$ with initial condition $p_k^\gamma = p \in \mathbf{R}^n$:

$$\begin{cases} p_l^\gamma \;=\; \Lambda^{\gamma *}(u_{l-1}, y_l)p_{l-1}^\gamma, \quad k+1 \le l \le M, \\ p_k^\gamma \;=\; p. \end{cases} \tag{4.2}$$

The value function is defined for $p \in \mathbf{R}^n$ by

$$W^\gamma(p,k) \;=\; \min_{u\in\mathcal{O}_{k,M-1}} \max_{y\in\mathbf{Y}^{M-k}} \{(p_M^\gamma, 0) \;:\; p_k^\gamma = p\}. \tag{4.3}$$

The solution of the game problem is expressed as follows.

Theorem 4.2 *The value function $W^\gamma(p,k)$ defined by (4.3) is the unique solution to the dynamic programming equation (3.25). Further, if $W^\gamma(p,k)$ is the solution of (3.25), and if $u^* \in \mathcal{O}_{0,M-1}$ is a policy such that, for each $k = 0,\dots,M-1$, $u_k^* = \bar{u}_k^*(p_k^\gamma)$, where $\bar{u}_k^*(p)$ achieves the minimum in (3.25), then u^* is an optimal policy for the output feedback dynamic game problem (§2.4).*

PROOF. Standard dynamic programming arguments, similar to those employed in the proof of Theorem 3.5, show that the function $W^\gamma(p,k)$ defined by (4.3) is the solution to the dynamic programming equation (3.25), and

$$W^\gamma(-\beta,0) \;=\; J^\gamma(u^*) \;=\; \min_{u\in\mathcal{O}_{0,M-1}} J^\gamma(u).$$

Therefore u^* is optimal. □

4.3 Robust Control

The solution to the state feedback robust control problem (§2.5) was expressed in terms of the solution $\bar{f}_k^\gamma(x)$ of a dynamic programming equation, and a state feedback controller $\tilde{u}_k^*(x)$ was obtained. The framework we have developed in this paper allows us to characterize the solution of the output feedback robust control problem in terms of the solution $W^\gamma(p,k)$ of a dynamic programming equation, and obtain an output feedback controller $\bar{u}_k^*(p_k^\gamma(\cdot\,;y_{1,k}))$. Note that the information state p_k^γ is also the solution of a dynamic programming equation (3.22).

Theorem 4.3 (Necessity) *Assume that there exists an output feedback controller $u^o \in \mathcal{O}_{0,M-1}$ solving the output feedback robust control problem. Then there exists a solution $W^\gamma(p,k)$ of the dynamic programming equation (3.25) such that $W^\gamma(-\beta,0)=0$.* (Sufficiency) *Assume that there exits a solution $W^\gamma(p,k)$ of the dynamic programming equation (3.25) such that $W^\gamma(-\beta,0)=0$, and let $\bar{u}_k^*(p)$ be a control value achieving the minimum in (3.25). Then $\bar{u}_k^*(p_k^\gamma(\cdot\,;y_{1,k}))$ is an output feedback controller which solves the output feedback robust control problem.*

PROOF. Define W^γ by (4.3). Then by Theorem 4.2 we know that W^γ is the solution of the dynamic programming equation (3.25). Next, we have

$$0 \;\le\; W^\gamma(-\beta,0) \;=\; \min_{u\in\mathcal{O}_{0,M-1}} J^\gamma(u) \;\le\; J^\gamma(u^o) \;\le\; 0,$$

and so $W^\gamma(-\beta,0)=0$.

To prove sufficiency, by Theorem 4.2 we have

$$0 \;=\; W^\gamma(-\beta,0) \;=\; J^\gamma(u^*),$$

which is the same as (2.15). Therefore u^* solves the output feedback robust control problem. □

A Appendix

The following theorem is a version of the Varadhan-Laplace lemma [3].

Lemma A.1 *Let A be a subset of $\mathbf{R}^m$, and F_a^ε, F_a be real valued functions defined on a finite set $\mathbf{X}$ such that*

$$\lim_{\varepsilon\to 0}\sup_{a\in A}\max_{x\in\mathbf{X}}|F_a^\varepsilon(x)-F_a(x)| \;=\; 0.$$

Then

$$\lim_{\varepsilon\to 0}\sup_{a\in A}|\varepsilon\log\sum_{x\in\mathbf{X}}e^{F_a^\varepsilon(x)/\varepsilon} \;-\; \max_{x\in\mathbf{X}}F_a(x)| \;=\; 0. \tag{A.1}$$

Proof. Write $\bar{F}_a^\varepsilon = \max_{x\in\mathbf{X}}F_a^\varepsilon(x)$, $\bar{F}_a = \max_{x\in\mathbf{X}}F_a(x)$. Our assumptions ensure that

$$\lim_{\varepsilon\to 0}\sup_{a\in A}|\bar{F}_a^\varepsilon - \bar{F}_a| = 0. \tag{A.2}$$

Since the maximums are achieved, we deduce

$$e^{\bar{F}_a^\varepsilon/\varepsilon} \;\le\; \sum_{x\in\mathbf{X}}e^{F_a^\varepsilon(x)/\varepsilon} \;\le\; ne^{\bar{F}_a^\varepsilon/\varepsilon}.$$

Take logarithms and multiply by ε to get

$$\bar{F}_a^\varepsilon \le \varepsilon\log\sum_{x\in\mathbf{X}}e^{F_a^\varepsilon(x)/\varepsilon} \le \varepsilon\log n + \bar{F}_a^\varepsilon.$$

Then using (A.2) we deduce (A.1) as desired. □

References

[1] 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.

[2] 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.

[3] M.I. Friedlin and A.D. Wentzell, “Random Perturbations of Dynamical Systems”, Springer-Verlag, New York, 1984.

[4] 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.

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

[6] M.R. James, J.S. Baras and R.J. Elliott, *Risk–Sensitive Control and Dynamic Games for Partially Observed Discrete-Time Nonlinear Systems*, IEEE Trans. Aut. Control, **AC-39-4** (1994) 780—792.

[7] M.R. James and J.S. Baras, *Robust H_∞ Output Feedback Control for Nonlinear Systems*, IEEE Trans. Aut. Control, **40**(6) (1995) 1007—1017.

[8] R.E. Kalman, P.L. Falb and M.A. Arbib, "Topics in Mathematical System Theory", McGraw-Hill, New York, 1969.

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

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

[11] P. Whittle, "Risk-Sensitive Optimal Control", Wiley, New York, 1990.