93-08
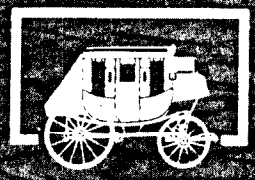
PROCEEDINGS OF THE
32nd IEEE CONFERENCE ON DECISION AND CONTROL

DECEMBER 15-17, 1993
MARRIOTT RIVERCENTER
SAN ANTONIO, TEXAS, USA

VOLUME 4 OF 4

93CH3307-6

IEEE
CONTROL
SYSTEMS
SOCIETY

IEEE

CSS

Proceedings of the 32nd Conference
on Decision and Control
San Antonio, Texas • December 1993

FM13 - 2:10

OUTPUT FEEDBACK RISK-SENSITIVE CONTROL AND DIFFERENTIAL GAMES FOR CONTINUOUS-TIME NONLINEAR SYSTEMS*

M.R. James[1]   J.S. Baras[2]   R.J. Elliott[3]

[1] Department of Systems Engineering and Cooperative Research Centre for Robust and Adaptive Systems, Australian National University, Canberra, ACT 0200, Australia

[2] Martin Marietta Chair in Systems Engineering, Department of Electrical Engineering and Institute for Systems Research, University of Maryland, College Park, MD 20742, USA

[3] Department of Statistics and Applied Probability, University of Alberta, Edmunton, Alberta T6G 2G1, Canada

Abstract

In this paper we carry out a formal analysis of an output feedback risk-sensitive stochastic control problem. Using large deviation limits, this problem is related to a deterministic output feedback differential game. Both problems are solved using appropriate information states. The use of an information state for the game problem is new, and is the principal contribution of our work. Our results have important implications for the nonlinear robust stabilization problem.

1. Introduction

In a recent paper [9], the authors analyzed the output feedback risk-sensitive stochastic control problem for discrete-time nonlinear systems and related it to a deterministic output feedback dynamic game. In particular, it was shown how the use of an appropriate *information state* can be used to solve the game problem. While the use of information states in stochastic control theory is standard practice [13], we remark that the information state for the risk-sensitive problem is not the familiar (unnormalized) conditional density, and our use of an informations state in the context of differential games appears to be new. An important application of these results is to the output feedback robust control problem [12]. The purpose of this paper is to show at least formally that the same results can be obtained in the mathematically technical continuous-time setting.

The output feedback risk-sensitive stochastic optimal control problem for the system

$$dx_t^\epsilon = b(x_t^\epsilon, u_t)\, dt + \sqrt{\epsilon}\, d\tilde{w}_t$$
$$d\tilde{y}_t^\epsilon = h(x_t^\epsilon)\, dt + \sqrt{\epsilon}\, d\tilde{v}_t \tag{1.1}$$

is to minimize the functional

$$J^{\mu,\epsilon}(u) = \mathbf{E}\left[\exp\frac{\mu}{\epsilon}\left(\int_0^T L(x_s^\epsilon, u_s)\, ds + \Phi(x_T^\epsilon)\right)\right] \tag{1.2}$$

over the class $\tilde{\mathcal{U}}_{0,T}$ of control policies which are non-anticipating functionals of the observation process $\tilde{y}$. In (1.1), $\tilde{w}$ and $\tilde{v}$ are standard independent Wiener processes, and $b \in C_b^1(\mathbf{R}^n \times \mathbf{R}^m, \mathbf{R}^n)$, $h \in C_b^1(\mathbf{R}^n)$, $L \in C_b^1(\mathbf{R}^n \times \mathbf{R}^m)$, $\Phi \in C_b^1(\mathbf{R}^n)$, with $L \geq 0$, $\Phi \geq 0$.

In this paper we explain how this problem can be solved using an information state process, and we relate it to a deterministic output feedback dynamical game using large deviations type limits. The game is defined for the system

$$\dot{x}_t = b(x_t, u_t) + w_t,$$
$$y_t = h(x_t) + v_t, \tag{1.3}$$

where w and v are unknown square integrable disturbances on $[0,T]$, and the problem is to minimize the cost

$$J^\mu(u) = \sup_{(w,v)\in L_2([0,T],\mathbf{R}^{n+1})}\left\{\int_0^T L(x_t, u_t)\right.$$

$$\left. -\frac{1}{2\mu}\left(|w_t|^2 + |v_t|^2\right)\, dt + \Phi(x_T)\right\} \tag{1.4}$$

over the class $\mathcal{U}_{0,T}$ of control policies which are non-anticipating functionals of the observation y.

An important consequence of our analysis is the solution of the deterministic output feedback dynamic game in terms of an information state. We believe these results are new. The application of our results to the output feedback robust control problem will appear elsewhere.

The problems and solutions in this paper are discussed purely in heuristic terms. A detailed mathematical treatment appears to be quite difficult, as is the case for continuous-time partially observed stochastic control [5], [7]. This paper is also quite brief; a more complete paper will appear elsewhere. For rigorous results, we refer the reader to the paper [9], where the discrete-time case is considered.

2. Risk-Sensitive Stochastic Control

The stochastic system (1.1) is assumed defined on a probability space $(\Omega, \mathcal{F}, \mathbf{P}^u)$, and we write $\mathcal{G}_t = \sigma(x_s, \tilde{y}_s; 0 \le s \le t)$ and $\mathcal{Y}_t = \sigma(\tilde{y}_s; 0 \le s \le t)$. There is an equivalent probability measure $\mathbf{P}^\dagger$ under which $\tilde{y}$ is a standard Wiener process independent of the state process [4]. For $u \in \tilde{\mathcal{U}}_{0,T}$ we have

$$\frac{d\mathbf{P}^u}{d\mathbf{P}^\dagger}\big|_{\mathcal{G}_t} =$$

$$\exp\left(-\frac{1}{\epsilon}\left[\frac{1}{2}\int_0^t |h(x_s^\epsilon)|^2 \, ds - \int_0^t h(x_s^\epsilon) \, d\tilde{y}_s^\epsilon\right]\right).$$

Write Z_t^ϵ for the right hand side of this expression.

Following an idea of [2] for linear systems, we seek to express $J^{\mu,\epsilon}(u)$ in terms of a new "state" process and then solve the resulting optimal stochastic control problem with complete "state" information. To this end, we define an information state $\sigma_t^{\mu,\epsilon}(x)$ in $L_1(\mathbf{R}^n)$ by

$$\langle \sigma_t^{\mu,\epsilon}, \eta \rangle =$$

$$\mathbf{E}^\dagger\left[\eta(x_t^\epsilon)Z_t^\epsilon \exp\left(\frac{\mu}{\epsilon}\int_0^t L(x_s^\epsilon, u_s)\, ds\right)\,\big|\, \mathcal{Y}_t\right],$$
$$(2.5)$$

for all test functions η, where

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

The evolution of the information state process is governed by the SPDE

$$d\sigma_t^{\mu,\epsilon} = \left(A^{u_t *} + \frac{\mu}{\epsilon}L^{u_t}\right)\sigma_t^{\mu,\epsilon}\, dt + \frac{1}{\epsilon}h\sigma_t^{\mu,\epsilon}\, d\tilde{y}_t^\epsilon,$$

$$\sigma_0^{\mu,\epsilon} = \rho \quad \text{in } L_1(\mathbf{R}^n),$$
$$(2.6)$$

where $A^u\phi = \frac{\epsilon}{2}\Delta\phi + b^u \cdot D\phi$, $L^u = L(\cdot, u)$, and $b^u = b(\cdot, u)$.

Using the information state process, one obtains the representation

$$J^{\mu,\epsilon}(u) = \mathbf{E}^\dagger\left[\langle \sigma_T^{\mu,\epsilon}, e^{\frac{\mu}{\epsilon}\Phi}\rangle\right]. \quad (2.7)$$

Remark 2.1 We note that there is an adjoint process $\nu_t^{\mu,\epsilon}$ such that $\langle \sigma_t^{\mu,\epsilon}, \nu_t^{\mu,\epsilon} \rangle$ is independent of t, and formally satisfies a backward SPDE.

Thanks to the representation (2.7), the problem now is to minimize the RHS of (2.7) with dynamics (2.6) which are completely observable. We use dynamic programming to solve this problem. The value function is defined by

$$S^{\mu,\epsilon}(\sigma, t) = \inf_{u\in\tilde{\mathcal{U}}_{t,T}} \mathbf{E}_{\sigma,t}^\dagger\left[\langle \sigma_T^{\mu,\epsilon}, e^{\frac{\mu}{\epsilon}\Phi}\rangle\right]. \quad (2.8)$$

The Hamilton-Jacobi-Bellman (HJB) equation for $S^{\mu,\epsilon}$ is (c.f. [L]):

$$\frac{\partial S^{\mu,\epsilon}}{\partial t} + \frac{1}{2\epsilon}D^2 S^{\mu,\epsilon}(h\sigma, h\sigma)$$

$$+ \inf_{u\in U}\left\{DS^{\mu,\epsilon} \cdot \left(A^{u *}\sigma + \frac{\mu}{\epsilon}L^u\sigma\right)\right\} = 0$$

$$\text{in } L_1(\mathbf{R}^n)\times[0,T],$$

$$S^{\mu,\epsilon}(\sigma, T) = \langle \sigma, e^{\frac{\mu}{\epsilon}\Phi}\rangle \quad \text{in } L_1(\mathbf{R}^n).$$
$$(2.9)$$

A verification theorem for this problem would say that if there exists a sufficiently smooth solution $S^{\mu,\epsilon}$ to (2.9) and if $u^* \in \tilde{\mathcal{U}}_{0,T}$ is a policy such that $u_t^* = \bar{u}_t^*(\sigma_t^{\mu,\epsilon})$, where $\bar{u}_t^*(\sigma)$ achieves the minimum in (2.9), then u^* is optimal.

Remark 2.2 Note that $\bar{u}_t^*(\sigma)$ is an *information state feedback* policy.

3. Small Noise Limit

The deterministic differential game is related to the risk-sensitive stochastic control problem via large deviation type small noise limits. For the full state feedback case, this connection was established in [10], [6]. For linear systems, the relationship is more direct [8], [16]. In the nonlinear output feedback case, we need to evaluate two limits; one for the information state, and one for the value function.

As in [11], one can show that

$$\lim_{\epsilon\to 0}\frac{\epsilon}{\mu}\log\sigma_t^{\mu,\epsilon}(x) = p_t^\mu(x), \quad (3.10)$$

provided $\tilde{y}_t^\epsilon \to \int_0^t y_s\, ds$, and p_t^μ satisfies the first order nonlinear PDE

$$\frac{\partial p^\mu}{\partial t} = \sup_{w\in\mathbf{R}^n}\left[-Dp^\mu(b^{u_t} + w) - \frac{1}{2\mu}|w|^2\right] + L^{u_t}$$

$$- \frac{1}{\mu}\left[\frac{1}{2}|h|^2 - hy_t\right],$$

$$p_0^\mu = \alpha.$$
$$(3.11)$$

The corresponding limit result for the value function is

$$\lim_{\varepsilon \to 0} \frac{\varepsilon}{\mu} \log S^{\mu,\varepsilon}(e^{\frac{\mu}{\varepsilon}p}, t) = W^\mu(p,t), \qquad (3.12)$$

where W^μ satisfies the Hamilton-Jacobi-Isaacs (HJI) equation

$$\frac{\partial W^\mu}{\partial t} + \inf_{u \in U} \sup_{y \in \mathbf{R}} \Big\{ DW^\mu \cdot$$

$$\Big(\sup_{w \in \mathbf{R}^n} \Big[-Dp^\mu(b^u + w) - \frac{1}{2\mu}|w|^2 \Big] + L^u$$

$$- \frac{1}{\mu}\Big[\frac{1}{2}|h|^2 - hy\Big]\Big) - \frac{1}{2\mu}|y|^2 \Big\} = 0,$$

$$W^\mu(p,T) = (p, \Phi).$$

$$(3.13)$$

Here, we have used the pairing

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

$$= \lim_{\varepsilon \to 0} \frac{\varepsilon}{\mu} \log\langle e^{\frac{\mu}{\varepsilon}p}, e^{\frac{\mu}{\varepsilon}q} \rangle.$$

Equations (3.11) and (3.13) are obtained by taking logarithmic transformations, writing down the appropriate PDE, and formally sending $\varepsilon \to 0$. In addition, we have expressed (3.13) in a way which uses the linearity of the operator DW^μ. The information state limit can be made precise using the robust version of (2.6) and standard viscosity solution methods [11]. The value function limit requires an analogous framework. These issues will be dealt with in more detail elsewhere.

Remark 3.3 There is an "adjoint" quantity q_t^μ such that the quantity (p_t^μ, q_t^μ) is independent of t, and satisfies a backward first order nonlinear PDE.

4. Output Feedback Differential Game

In this section we interpret the HJI equation (3.13) as a dynamic programming equation for the deterministic output feedback differential game defined in §1, and p_t^μ as an information state with dynamics (3.11). In this way the game problem can be solved.

Indeed, the cost function $J^\mu(u)$ can be represented in terms of the information state p^μ (with $\alpha = 0$):

$$J^\mu(u) = \sup_{y \in L_2([0,T],\mathbf{R})} \Big\{ (p_T^\mu, \Phi) - \frac{1}{2\mu} \int_0^T |y_s|^2 \, ds \Big\}.$$

$$(4.14)$$

This can be seen with the aid of the following repre-

sentation of the information state:

$$p_t^\mu(x) = \sup_{\xi \in C([0,T],\mathbf{R}^n)} \Big\{ \alpha(\xi_0) + \int_0^t L(\xi_s, u_s)$$

$$- \frac{1}{\mu}\Big(\frac{1}{2}|\dot\xi - b(\xi_s, u_s)|^2 + \frac{1}{2}|h(\xi_s)|^2 - h(\xi_s)y_s\Big) \, ds$$

$$: \xi_t = x \Big\}$$

$$(4.15)$$

The problem now is to minimize the RHS of (4.14) with dynamics (3.11) which are completely observable. Using dynamic programming, the value function is given by

$$W^\mu(p,t) = \inf_{u \in \mathcal{U}_{t,T}} \sup_{y \in L_2([t,T],\mathbf{R})} \Big\{ (p_T^\mu, \Phi)$$

$$- \frac{1}{2\mu} \int_t^T |y_s|^2 \, ds \ : \ p_t^\mu = p \Big\},$$

$$(4.16)$$

and formally satisfies the HJI equation (3.13).

The verification theorem then says that if there exists a sufficiently smooth solution W^μ to the HJI equation (3.13) and if $u^* \in \mathcal{U}_{0,T}$ is a policy such that $u_t^* = \bar u_t^*(p_t^\mu)$, where $\bar u_t^*(p)$ achieves the minimum in (3.13), then u^* is optimal.

Remark 4.4 As was the case for the output feedback risk-sensitive stochastic control problem, $\bar u_t^*(p)$ is an *information state feedback* policy for the deterministic output feedback differential game. Thus the information state concept provides the appropriate framework for solving such differential games.

Finally, we provide a condition under which a certainty equivalence principle holds (c.f. [16], [17], [9]).

Consider the solution $f_t^\mu(x)$ of the HJI equation

$$\frac{\partial f^\mu}{\partial t} + \inf_u \sup_w \Big\{ Df^\mu \cdot (b^u + w) - \frac{1}{2\mu}|w|^2 + L^u \Big\}$$

$$= 0 \ \text{in} \ \mathbf{R}^n \times [0,T],$$

$$f_T^\mu = \Phi \ \text{in} \ \mathbf{R}^n.$$

$$(4.17)$$

f^μ is the value function associated with a state feedback differential game. Note that in (4.17), the Isaacs condition holds. Let $\bar u_t^*(x)$ denote a control value which achieves the minimum in (4.17). This defines an optimal policy for the corresponding game.

The *minimum stress estimate* $\bar x_t$ of x_t is defined by

$$\bar x_t \in \underset{x \in \mathbf{R}^n}{\operatorname{argmax}} \{ p_t^\mu(x) + f_t^\mu(x) \}, \qquad (4.18)$$

the RHS being a set valued map.

Assume that the following condition holds:

$$\inf_u \sup_y \{DW \cdot$$

$$\left(\sup_w \left[-Dp^\mu(b^u + w) - \tfrac{1}{2\mu}|w|^2\right] + L^u\right.$$

$$\left.-\tfrac{1}{\mu}\left[\tfrac{1}{2}|h|^2 - hy\right]\right) - \tfrac{1}{2\mu}|y|^2\right\} =$$

$$DW \cdot \left(\inf_u \sup_w \left[-Dp^\mu(b^u + w) - \tfrac{1}{2\mu}|w|^2 + L^u\right]\right).$$

$$(4.19)$$

Then the value function has the explicit form

$$W^\mu(p,t) = (p, f_t^\mu) \qquad (4.20)$$

and the certainty equivalence policy $u^* \in \mathcal{U}_{0,T}$ defined by

$$u_t^* = \tilde{u}_t^*(\bar{x}_t) \qquad (4.21)$$

is an optimal policy for the deterministic output feedback differential game. This follows from representation of the HJI (3.13) using the condition (4.19) and the verification theorem.

Remark 4.5 The equality (4.19) is perhaps an *Isaacs* condition for the deterministic output feedback differential game. In general, if (4.19) fails, we expect that $W^\mu(p,t) \geq (p, f_t^\mu)$, and consequently the certainty equivalence policy will be suboptimal.

5. References

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

[2] 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.

[3] 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.

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

[5] W.H. Fleming and E. Pardoux, *Optimal Control for Partially Observed Diffusions*, SIAM J. Control Optim., **20** (2) (1982) 261—285.

[6] 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.

[7] O. Hijab, *Partially Observed Control for Markov Processes*, Proc. Conf. on Applied Stochastic Analysis, Rutgers University, Springer Verlag, 1991 (I. Karatzas and D. Ocone, eds.).

[8] 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.

[9] M.R. James, J.S. Baras and R.J. Elliott, *Risk–Sensitive Control and Dynamic Games for Partially Observed Discrete–Time Nonlinear Systems*, to appear, IEEE Trans. Automatic Control.

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

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

[12] M.R. James and J.S. Baras, *Robust Output Feedback Control for Nonlinear Discrete–Time Systems*, submitted for publication, 1993.

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

[14] P.L. Lions, *Viscosity Solutions of Fully Nonlinear Second Order Equations and Optimal Stochastic Control in Infinite Dimensions, Part II: Optimal Control of Zakai's Equation*, CEREMADE Report 8825, 1989.

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

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