APRIL 1988
Revised: August 1988

LIDS-P-1765

A CLASS OF ADAPTIVE CONTROL PROBLEMS SOLVED VIA STOCHASTIC CONTROL

by

Ofer Zeitouni [(1)]

ABSTRACT

Following a set up investigated by Rishel [7], we consider an adaptive control problem with unknown parameter x as a partially observed stochastic control problem. Exploiting the finite dimensionality of the estimator, we transform it to a fully observed stochastic optimal control problem to which we then find ε-optimal randomized feedback policies.

Keywords: Adaptive control; stochastic control; ε-optimal; Diffusion processes.

(1) Massachusetts Institute of Technology, Laboratory for Information and Decision Systems, Cambridge, MA 02139

This work was partially supported by the Air Force Office of Scientific Research under grant AFOSR-85-0227B and by the Weizmann Institute Postdoctoral Fellowship.

I. INTRODUCTION

Consider the following adaptive control problem introduced by Rishel [7].

Let $x \in \mathbf{R}^n$ be an unknown parameter. We consider a Bayesian set up, where x is distributed according to some prior density $p_0(x)$, which may or may not be compactly supported. Let w_t be an m dimensional Brownian motion, and define y_t by

$$y_t = y_0 + w_t \tag{1.1}$$

Let F_t be the sigma field generated by $\{y_s, 0 \le s \le t\}$ and G_t be the sigma field generated by $(\{y_s, 0 \le s \le t\} \vee x)$. Let U be a compact subset of $\mathbf{R}^q$. Let $A(x,y) \in \mathbf{R}^m$ and $B(x) \in \mathbf{R}^{m\times q}$ be matrices depending on the unknown parameter x and on y. Conditions on A(x,y) and B(x,y) will be imposed below.

Define an admissible control u(t) to be a U valued stochastic process satisfying:

(a) u_t is F_t adapted

(b) $E \exp\left(\frac{1}{2}\int_0^T \|A(x,y_t) + B(x)u_t\|^2 dt\right) < \infty \quad \forall x \in \text{supp } p_0(x)$

In particular, if A(x,y) and B(x) are bounded, (b) reduces to a trivial condition. The set of admissible controls will be denoted by U. Define:

$$\Lambda_T = \exp\left(\int_0^T \left(A(x,y_t) + B(x)u_t\right)^* dw_t - \frac{1}{2}\int_0^T \left\| A(x,y_t) + B(x)u_t \right\|^2 dt\right) \tag{1.2}$$

where * denotes the operation of taking transposes. By b), $E(\Lambda_T) = 1$, and we may define a new measure P^u such that $\frac{dp^y}{dp} = \Lambda_T$, under which

$$y_t = y_0 + \int_0^t (A(x,y_t) + B(x)u_t)ds + w_s^u \tag{1.3}$$

where w_s^u is a P^u Brownian motion.

The adaptive control goal is to minimize over the class of admissible control an objective cost of the form

$$J^u = E^u \int_0^T \ell(y_s, u_s, s)ds \tag{1.4}$$

We assume throughout that

$$|\ell(y,u,t)| \leq K(1 + |y|^r)$$

$$\left| \frac{\partial \ell(y,u,t)}{\partial y} \right| \leq K(1+|y|^r)$$

$$\left| \frac{\partial \ell(y,u,t)}{\partial u} \right| \leq K(1+|y|^r)$$

For some $r > 0$.

Note that (1.4) is a partially observed stochastic control problem, for under P^u x is not known to be controller.

In [7], Rishel has considered a version of the problem (1.3), (1.4) and proved a stochastic maximum principle. Explicit solutions for particular cases were derived in [1], [2]. Hijab [4] considered a modified version of (1.3) where under a linearity assumption and a more general model of x, he found an explicit solution to a problem where information cost is attached to J^u and the cost is a function of x and u. Here, we exploit, as in [1], [2], the finite dimensionality of the estimation problem, as follows: By taking conditional expectations in (1.3), one has [7]:

$$dy_t = E^u(A(x,y_t)|F_t)dt + E^u(B(x)|F_t)u_t dt + dv_t^u \tag{1.5}$$

where v_t^u is an F_t Brownian motion. In the sequel, ^ will denote conditional expectations w.r.t. F_t, i.e.

$$\hat{A}(x,y_t) \triangleq E^u(A(x,y_t)|F_t)$$

For simplicity and concreteness, we assume below that

$$A(x,y) = A_0 y + A_1(y)x \tag{1.6}$$

$$B(x) = B_0 + \tilde{B}_1(x), \quad \tilde{B}_1^{i,j}(x) = \sum_k b_{ij}^k \, x, \text{ i.e.}$$

$$B(x)u = B_0 u + B_1(u)x \text{ where } B_1^{ij}(u) \triangleq \sum_m b_{im}^j \, u^m \tag{1.7}$$

The linearity assumption of A(x,y) w.r.t. x can be dispensed of if one has a separation of variable of the form $A(x,y) = A_1(y)g(x) + A_0(y)$; similarly, one could include a y-dependence in B(x) in a separation of variable form. Since those extensions are easily handled, we do not consider them here.

Following now the argument of Liptser-Shiryayev [6, ch. 12], appropriately modified to our case due to the non-Gaussian assumptions on x_0, (c.f., e.g., [3], [9]) one has the following:

Lemma 2.1:

$$p(x|F_t) = \frac{\frac{p_0(x)}{N_x(0,I)} \exp(\frac{1}{2}(x-\gamma_t)^* \alpha_t^{-1}(x-\gamma_t))}{\int dx \frac{p_0(x)}{N_x(0,I)} \exp(\frac{1}{2}(x-\gamma_t)^* \alpha_t^{-1}(x-\gamma_t))} \tag{1.8}$$

where $N_x(0,I) \triangleq \dfrac{\exp(-\frac{1}{2}x^* x)}{(2\pi)^{n/2}}$, γ_t is an n-dimensional vector and α_t is an nxn dimensional matrix which satisfy:

$$d\gamma_t = \alpha_t [A_1(y_t) + B_1(u_t)]^* [dy_t - (A_0 y_t + B_0 u_t + (A_1(y_t) + B_1(u_t))\gamma_t dt]$$

$$\gamma_0 = 0 \tag{1.9a}$$

$$d\alpha_t = -\alpha_t [A_1(y_t) + B_1(u_t)]^* [A_1(y_t) + B_1(u_t)]\alpha_t; \ \alpha_0 = I \tag{1.9b}$$

Proof: Note that as in [1], [3], [9] the unnormalized density $dz\rho_x(z|F_t) = \text{Prob}(x \in (z,dz)|F_t) \bullet K_t$, where K_t is F_t adapted, is of the form:

$$dz\rho_x(z|F_t) = E(\Lambda_T 1_{x\varepsilon(z,z+dz)}|F_t)$$

$$= E(\Lambda_t 1_{x\varepsilon(z,z+dz)}|F_t) \ \alpha$$

$$dzp_0(z)\exp(-\frac{1}{2}\int_0^t z^*[A_1(y_s) + B_1(u_s)]^*[A_1(y_s) + B_1(u_s)] \ zds$$

$$+\int_0^t z^*[A_1(y_s) + B_1(u_s)]^*(dy_s - (A_0y_s + B_0u_s dt)$$

Dividing and multiplying by $N_x(0,I)$, one obtains (2.8). We remark that exactly as in [6], α_t is positive definite for $0 \le t \le T$. □

Note that y_t can be now rewritten as:

$$dy_t = (A_0y_t + B_0u_t)dt + (A_1(y_t) + B_1(u_t))F(\gamma_t, \alpha_t)dt + d\hat{\nu}_t; \ y_0 = 0 \quad (1.9c)$$

where

$$F(\gamma,\alpha) = \frac{\int_x \frac{xp_0(x)}{N_x(0,I)} N_x(\gamma,\alpha^{-1})dx}{\int_x \frac{p_0(x)}{N_x(0,1)} N_x(\gamma,\alpha^{-1})dx} \quad (1.10)$$

where $N_x(\gamma,\alpha)$ denotes a Gaussian distribution with mean γ, covariance matrix α^{-1}. Note that (1.9) together with (1.4) form a <u>completely observable</u> stochastic control problem. It is however somewhat a complicated one due to the degeneracity of the diffusion matrix, the fact that control enters the diffusion matrix and the non-Lipschitz coefficients of (1.9).

In some simple cases (and specifically, in the case where $B(x) = B$). Benes and Rishel [1] have been able to compute <u>explicitly</u> optimal controls via the Hamilton-Jacobi-Bellman equation and the maximum principle. In the general case, however, the Bellman equation does not seem solvable and we are led to consider ε-optimal approximations.

<u>Remark</u>. Note that if $p_0(x) = N_x(0,I)$, $F(\gamma,\alpha) = \gamma$.

2. ε-OPTIMAL RANDOMIZED MARKOV STRATEGIES.

In this section, we construct ε-optimal randomized Markov strategies for the problem posed in section 1, i.e. for (1.9) and (1.4). Those strategies are defined in terms of a classical solution of an associated Bellman equation. For simplicity, we make the following structural restrictions. Those restrictions are not crucial and could be avoided at the expense of more cumbersome expressions and proofs. Additional restrictions of more technical nature (boundedness etc.) will be imposed later (c.f. lemma 2.1).

Assumptions.

$$A_0 = B_0 = 0 \qquad (2.1a)$$

$$p_0(x) \sim N_x(0,I) \qquad (2.1b)$$

We will seek to apply the method of [5, ch. 5]. To do that, it will however be convenient to rewrite (1.9) in a different way: Let $\beta_t = \alpha_t^{-1}\gamma_t$

$$d\alpha_t = -\alpha_t[A_1(y_t) + B_1(u_t)]^*[A_1(y_t) + B_1(u_t)]\alpha_t; \quad \alpha_0 = I \qquad (2.2a)$$

$$d\beta_t = +[A_1(y_t) + B_1(u_t)]^*[A_1(y_t) + B_1(u_t)]\alpha_t\beta_t dt + [A_1(y_t) + B_1(u_t)]^* d\hat{v}_t$$
$$\beta_0 = 0 \qquad (2.2b)$$

$$dy_t = +[A_1(y_t) + B_1(u_t)]\alpha_t\beta_t dt + d\hat{v}_t \qquad (2.2c)$$

In the sequel, b will denote the drift vector in (2.2) and σ will denote the diffusion matrix there. Note that (2.2) does not satisfy the conditions of [5, ch. 5] and therefore the methods described there have to be modified to be applicable.

Note that (2.2) is locally Lipshitz; however, we will need global Lipschitz conditions. Towards this end, let

$$\Omega^R \triangleq \{\beta \varepsilon \mathbf{R}^n,\ y \varepsilon \mathbf{R}^m |\ |\beta| \le \mathbf{R},\ |y| \le R\}$$

$$\tau^R \triangleq \{\inf t > 0 |\ (\beta_t, y_t)\ \varepsilon\ \partial\Omega^R\}$$

For g denoting either $b(\alpha, \beta, y)$ or $\sigma(\alpha, \beta, y)$, let

$$g^R(\alpha,\beta,y) = g(\alpha, \frac{\beta}{|\beta|}(|\beta| \wedge R), \frac{y}{|y|}(|y| \wedge R)$$

Let now α^R, β^R, y^R denote the solution to (2.2) when b^R, σ^R is substituted instead of b, σ. Finally, let J^u_R denote J^u in (2.4) with y^R substituted instead of y.

We claim

Lemma 2.1. Assume $|A_1(y)| < K$, $|B_1(u)| < K$ for $u \varepsilon U$. Then $|J^u_R - J^u| \underset{R\to\infty}{\to} 0$ uniformly on all admissible strategies.

Proof. Note first that by the boundedness of $|A_1(y)|$, $|B_1(u)|$ and of x_t, one has by standard arguments that $E[\xi_t^{2p}] < \infty$, where ξ_t stands for either β_t, y_t, β^R_t, y^R_t. Next, we note that

$$|J^u - J^u_R| \le E^u\left(\int_{\tau_R \wedge T}^{T} |\ell(y_s,u_s,s) - \ell(y^R_s, u_s, s)|ds \right)$$

$$\le K_1 P(\tau_R < T) \bullet \sup_{y_0,\, \beta_0 \varepsilon \Omega^R} \int_0^T E(|y_s|^r + |y^R_s|^r)ds$$

Note that by standard estimates,

$$\sup_s \sup_{y_0,\beta_0 \varepsilon \Omega^R} E(|y_s|^r) \le K_2 \quad \sup_s \sup_{y_0,\beta_0 \varepsilon \Omega^R} E(|\beta_s|^r) \le K_3 R^r \tag{2.4}$$

with a similar bound on $E(|y^R_s|^r)$. On the other hand,

$$P(\tau_R < T) \le P(K_4 \sup_{t\varepsilon[0,T]} v_t > R) \tag{2.5}$$

$$\le K_5 \frac{E(\sup v_t)^{2r}}{R^{2r}} \le \frac{K_6}{R^{2r}}$$

where the constants K_1 - K_6 do not depend on u. Therefore,

$$|J^u - J^u_R| \le \frac{K_7}{R^r} \underset{R\to\infty}{\to} 0 \tag{2.6}$$

and the lemma is proved. □

In view of the lemma above, it is enough to build ε-optimal strategies for the system indexed by R, for R large enough. We will attempt to do that by perturbing (2.2a) and (2.3b) with an auxilliary Brownian motion. For reasons to become clear below, we will not perturb (2.2c). That is the main point where we depart from the classical treatment [5]. Note however that when perturbing (2.2a) such that α_t is no longer positive definite, $b^R(\alpha,\beta,y)$ is no longer Lipschitz continuous and moreover, (3.2a) may have a finite explosion time. To remedy that, we modify $b^R(\alpha, \beta, y)$ on the set of non-positive α'-s; This will not affect the solution of (2.2) since in (2.2), $\alpha > 0$ a.s.

Let $\tilde{b}^R(\alpha,\beta,y) \triangleq b^R(P\alpha,\beta,y)$ where $P\alpha$ denotes the projection of the symmetric matrix α on the convex set $\{|\alpha| < R \wedge \alpha \geq 0\}$. Note that the system (2.2) with $(\tilde{b}^R, \sigma^R)$ is identical to the system with (b^R, σ^R), and that $(\tilde{b}^R,\sigma^R)$ are globally Lipshitz. Consider the following perturbed system:

$$d\alpha^\varepsilon_t = \tilde{b}^R_\alpha(u, \alpha^\varepsilon, y^\varepsilon)dt + \varepsilon I dw^1_t \tag{2.7a}$$

$$d\beta^\varepsilon_t = \tilde{b}^R_\beta(u, \alpha^\varepsilon, \beta^\varepsilon, y^\varepsilon)dt + \sigma^R_\beta(u,y^\varepsilon)d\hat{\nu}_t + \varepsilon I dw^2_t \tag{2.7b}$$

$$dy^\varepsilon_t = \tilde{b}^R_y(u,\alpha^\varepsilon,\beta^\varepsilon, y^\varepsilon)dt + d\hat{\nu}_t \tag{2.7c}$$

where w^t_1, w^2_t are independent Brownian motions of appropriate dimensions. Note that (2.7) is uniformly nondegenerate, and that for $\varepsilon=0$ (2.7) reduces to (2.2) with (b^R, σ^R) instead of (b,σ). Let $v^R_\varepsilon(s,x)$ denote the value function of the control problem (2.7) together with (2.4), i.e.

$$v^R_\varepsilon(s,x) = \inf_{\substack{u\varepsilon U \\ (\alpha_0,\beta_0,y_0)=x}} J^u_R(\alpha^\varepsilon, \beta^\varepsilon, y^\varepsilon) \tag{2.8}$$

and let L^u_R be the Backward Kolmogorov operator associated with (2.7). By [5, Thm. 4.7.7], we have that $v^R_\varepsilon(s,x) \in W^{1,2}(C_{T,R^1})$ for each $R^1 > 0$. Moreover, by

[5, Lemma 5.1.1], for each $\delta > 0$ and $R_1 > 0$ there exists an infinitely differentiable feedback strategy $u^{\varepsilon,\delta}_{R_1}(x^{\varepsilon},t)$ with uniformly bounded spacial first derivatives such that $\sigma^R(u^{\varepsilon,\delta}_{R_1}(x^{\varepsilon},t), x)$ and $\tilde{b}^R(u^{\varepsilon,\delta}_{R_1}(x^{\varepsilon},t),x)$ are uniformly Lipshitz continuous and

such that

$$\sup_{t\varepsilon(0,T)} |F[v^R_\varepsilon] - [L_R^{u^{\varepsilon,\delta}_{R_1}} v^R_\varepsilon + \ell^{u^{\varepsilon,\delta}_{R_1}}]| \le \delta \tag{2.9}$$

where

$$F[u^R_\varepsilon] \triangleq \sup_{u\varepsilon U} [L^u_R u^R_\varepsilon + \ell^u]$$

Finally, note that again by [5, Corollary 3.1,13], $v^R_\varepsilon(s,x) \to v^R(s,x)$ uniformly in $C_{T,R}$.

We can state now our main result::

Theorem. Assume the conditions in lemma 2.1. Let

$u^{\varepsilon,\delta}_{R_1}(x,t) = u^{\varepsilon,\delta}_{R_1}(\alpha+\varepsilon w_1, \beta+\varepsilon w_2, y)$.

Then

$$\overline{\lim_{\varepsilon\to 0}}\ \overline{\lim_{R\uparrow\infty}}\ \overline{\lim_{\delta\downarrow 0}}\ J_R^{\hat{u}^{\varepsilon,\delta}_{R_1}} = \inf_{u\in U} J^u_R = \lim_{\varepsilon\to 0} v^R_\varepsilon(0,x_0)$$

i.e., one can construct feedback randomized control for the system (2.2) which will be as close as desired to the optimal control.

Proof. The proof is an adaptation to the proof of [5, Thm. 5.2.5)]. Note that under $u^{\varepsilon,\delta}_{R_1}$ the system (2.2a) (with $\tilde{b}^R$, σ^R instead of (b,σ)) is transformed into the system:

$$d\alpha^\varepsilon = \tilde{b}^R_\alpha(u^{\varepsilon,\delta}_{R_1}(\alpha^\varepsilon, \beta^\varepsilon, y^\varepsilon), \ \alpha^\varepsilon - \varepsilon w_1, \beta^\varepsilon - \varepsilon w_2, y^\varepsilon)\, dt \ + \varepsilon dw_1 \tag{2.10a}$$

$$d\beta^\varepsilon = \tilde{b}^R_\beta(u^{\varepsilon,\delta}_{R_1}(\alpha^\varepsilon, \beta^\varepsilon, y^\varepsilon), \alpha^\varepsilon - \varepsilon w_1, \beta^\varepsilon - \varepsilon w_2, y^\varepsilon)dt + \varepsilon dw_2 \tag{2.10b}$$

$$+ \sigma^R_\beta(y^\varepsilon, u^{\varepsilon,\delta}_{R_1}(\alpha^\varepsilon, \beta^\varepsilon, y^\varepsilon))d\lozenge_t$$

$$dy^\varepsilon = \tilde{b}^R_y(u^{\varepsilon,\delta}_{R_1}(\alpha^{\varepsilon,\delta}(\alpha^\varepsilon, \beta^\varepsilon, y^\varepsilon), \alpha^\varepsilon - \varepsilon w_1, \beta^\varepsilon - \varepsilon w_2, y^\varepsilon)\, dt + d\hat{v}_t \qquad (2.10c)$$

Since (2.10a) is uniformly nondegenerate and $v^R_\varepsilon(s,x) \in W^{1,2}(C_{T,R})$, one has from the weak Ito formula ([5, thm. 2.10.1]), that, for each $(\varepsilon,\delta,R_1) = \eta$

$$v^R_\varepsilon(0,x_0) = E\left(\int_0^{\tau_{R_1}\wedge T} \ell^{u^{\varepsilon,\delta}_{R_1}}\, dt\right) + v^R_\varepsilon(T\wedge\tau^{R_1}, x_{\tau^{R_1}\wedge T}) \qquad (2.11)$$

$$-E\int_0^{\tau_{R_1}\wedge T} [L^{u^{\varepsilon,\delta}_{R_1}} v_\varepsilon + \ell^{u^{\varepsilon,\delta}_{R_1}}]dt$$

$$+E\int_0^{\tau_{R_1}\wedge T} \Big[\tilde{b}^R(u^{\varepsilon,\delta}_{R_1}(s,x^\varepsilon), x^\varepsilon)$$

$$-\tilde{b}^R(u^{\varepsilon,\delta}_{R_1}(s,x^\varepsilon), x^\varepsilon - \varepsilon w)\Big]\nabla_x v_\varepsilon(s,x^\varepsilon)ds$$

By the same arguments as in [5, thm. 5.2.5], the first terms converge to $J_R^{u^{\varepsilon,\delta}_{R_1}}$, the second term converges to zero, whereas since $|\nabla_x v_\varepsilon(s,x^\varepsilon)| < K\,(1+|x^\varepsilon|^r)$, one has also the required convergence for the last term. The theorem is proved. □

Remarks: 1) Note that the main difference from [5] is that we have used ε-perturbation only in some of the components of the diffusion. Perturbing all the components would have violated the structural condition under which one may trade controls by randomized controls.

2) The theorem proved allows one to actually build $\tilde{\varepsilon}$-optimal control. Indeed, for a given $\tilde{\varepsilon}$, pick up $\delta, R, R_1, \varepsilon$ such that

$$J_R^{\hat{u}^{\varepsilon,\delta}_{R_1}} < \inf_{u\in U} J^u + \tilde{\varepsilon}$$

Such a choice exists due to the theorem. Then $\hat{u}^{\varepsilon,\delta}_{R_1}$ is the required randomized feedback control. Note that $u^{\varepsilon,\delta}_{R_1}$, which is the feedback function needed to build $\hat{u}^{\varepsilon,\delta}_{R_1}$, is obtained via a classical solution of the Bellman equation associated with the system (2.7)-(2.8).

Acknowledgements. I would like to thank Prof. Vivek Borkar from the Tata Institute for his very generous help, without which the results in this paper would not have been derived. This work was done while both of us were visiting the Department of Mathematics, Scuola Normale Superiore, Pisa, Italy.

REFERENCES

[1] V.E. Benes, and R.W. Rishel, Optimality of full bang to reduce predicted miss for some partially observed stochastic control problems, In: W.H. Fleming and P.L. Lions, Eds., Proc. Workshop Stoch. Diff. Systems, IMA Vol. 10, pp. 1-15, 1988.

[2] V.E. Benes, and R.W. Rishel, A Stochastic Adaptive Control Problem and Related Non Adaptive Problems, Preprint, 1988.

[3] U. Haussaman and E. Pardoux, Stochastics, (23) 1988.

[4] O. Hijab, Stabilization of Control Systems, Springer, Berlin-New York 1987.

[5] N.V. Krylov, Controlled Diffusion Processes, Springer, Berlin-New York, 1980.

[6] R.S. Liptser and A.N. Shiryayev, Statistics of Random Processes, Springer, Berlin-New York, 1977.

[7] R.W. Rishel, An exact formula for a linear quadratic adaptive stochastic optimal control law, SIAM J. Cont. Opt., 24 (1986), pp. 667-674.

[8] E. Wong, SIAM J. Control 9(1971), pp. 621-633.

[9] O. Zeitouni and B.Z. Bobrovsky, On the joint filtering-smoothing of diffusion processes, Systems and Control letters 7(1986) pp. 317-321.