BULLETIN *of the*
MALAYSIAN
MATHEMATICAL
SCIENCES
SOCIETY

Bull. Malays. Math. Sci. Soc. (2) **27** (2004), 103–110

Aspects of Control for the Normal Markov Processes

NASROLLAH SAEBI
School of Mathematics, Kingston University, Penrhyn Road, London KT1 2EE - U.K.
e-mail: n.saebi@kingston.ac.uk

Abstract. The choice of optimal control policy for sequentially observed data studied in a Bayesian context is usually a dynamic programming problem that involves a backward iterative solution. In general, as in most sequential Bayes problems, optimal solutions are difficult to derive analytically in simple forms. The system of linear models examined here is, however, amongst the few cases with known explicit optimal solutions. This would allow analytical comparisons with the performance of sub-optimal control procedures. Certain sequence of myopic rules are introduced and applied to the control system. These rules, in general, will provide the user with good near-optimal control policies whenever optimal solutions are analytically difficult to determine. As the myopic rules do not involve backward iteration procedures, they are often convenient to apply, and in addition, the user has the option of improving the accuracy of any particular approximating solution by taking additional future costs into consideration. This approximation is, naturally, at its best when the complete future cost is considered and, for the Aoki (1967) linear control system, solutions are then proved to be optimal.

2000 Mathematics Subject Classification: 62C10

1. Introduction

Consider a process that is changing in time and can be observed at equidistant integer time points over a known time period $(0, N)$. Our principal objective is to control the stochastic parameter of the process at time point *t*, denoted by $\theta_t\,; 0 \le t \le N$. In general, θ_t is not observed and a value x_t (which is conditionally distributed about θ_t) is generated by the observation equation,

$$x_t = \theta_t + \eta_t\,, \tag{1}$$

and η_t's are independent and identically distributed with zero mean and known variance ρ^2. The decision space at time *t* is composed of all possible set of values of a control variable, u_t. It is assumed that the control u_t depends only on the current and past observations; that is

$$u_t = \phi(x^t); \quad x^t = (x_0, x_1, \cdots, x_t), \tag{2}$$

and the past control sequence $(u_0, u_1, u_2, \cdots, u_{t-1})$ is assumed to be known {closed-loop control}. The control model we consider here is a special case of the Kalman model [6] and is the control version of the linear Markov process with superimposed error; namely,

$$\theta_t = a(\theta_{t-1} - u_{t-1}) + \varepsilon_t; \quad (t = 1, 2, \cdots, N); \quad \varepsilon_t \sim N(0, \sigma^2) \tag{3}$$

and σ^2 is known constant. The Bayesian approach to the optimal control problems requires the assumption of a "priori distribution" for the parameter which can be updated by the Bayes rule (using controls and observation vector measurements) up to the current time point. The initial value, θ_0, is assumed to have a normal prior distribution with zero mean and known constant variance σ_0^2. The following model developed here can be viewed as "normal"; in the sense that, all the random variables involved are assumed to be normally distributed.

The problem here is to adjust the system output in such a way that it follows a specific target as closely as possible. For this we need to introduce an appropriate performance index or a defined loss structure *W*. We assume, naturally, that *W* depends on the stochastic parameter θ_t that we intend to adjust, and also on the control variable u_t. In addition we assume *W* to be additive over time. That is,

$$W = \sum_{t \le N} \{W_t(\theta_t, u_{t-1})\} \quad (W_t \ge 0). \tag{4}$$

Now that the structure of the system is described, we are in a position to outline a general procedure for deriving optimal solution to the problem.

2. Bayes' optimal solution to the control system

At time *t*, having observed x^t, the posterior density of θ_t is proportional to (written as, $\underset{\theta_t}{\propto}$), $f(x_t | \theta_t) \cdot f(\theta_t | x^{t-1})$. From the Markov property and Bayes' theorem, we have (for $t > 0$),

$$f(\theta_t | x^t) \underset{\theta_t}{\propto} f(x_t | \theta_t) \int f(\theta_t | \theta_{t-1}, u_{t-1})\, f(\theta_{t-1} | x^{t-1})\, d\theta_{t-1}, \tag{5}$$

which gives a recurrence relation between $f(\theta_t | x^t)$ and $f(\theta_{t-1} | x^{t-1})$, so that in general we can take $f(\theta_t | x^t)$ to be known for all t.

To derive an optimal solution to the problem, we apply a backward iterative procedure and start at the last permissible control point in the process; namely, at $N-1$. Assuming x^{N-1} have been observed, and the previous controls have all been determined, we are then to find the last control u_{N-1}. As this control appears only in W_N, hence u_{N-1} is chosen to minimise the future expected loss $E\{W_N(\theta_N, u_{N-1}) \mid x^{N-1}\}$, which is

$$\iint_{\Theta} W_N\,(\theta_N, u_{N-1})\; f(\theta_N \mid \theta_{N-1}, u_{N-1})\; f\big(\theta_{N-1} \big| x^{N-1}\big)\, d\theta_N\;\; d\theta_{N-1}.$$

Since all the involved functions are known here, we can in principle minimise the above expected loss and determine the optimal control u^*_{N-1}. Denoting the minimised expected loss at $N-1$ by $V^*_{N-1}\,(x^{N-1})$, we then at the "two stages-to-go" choose u_{N-2} to minimise

$$E\{V^*_{N-1}(x^{N-1}) \; + \; W_{N-1}(\theta_{N-2}, u_{N-2}) \mid x^{N-2}\}.$$

This expectation, using equation (1), is with respect to $f(\eta_{N-1})\; f(\theta_{N-1} \mid x^{N-2})$, which can be written in terms of the known function $f(\theta_{N-2} \mid x^{N-2})$, so that there will be no difficulty in evaluating the optimal control u^*_{N-2} at stage $N-2$. This argument can be extended to the remaining stages of the process and, by application of the Bellman principle of recursive optimality [3], at the t-th stage we choose the optimal control u^*_t to minimise the future expected costs. We then apply u^*_t and sample x_{t+1}, and repeat the procedure to derive u^*_{t+1} and continue this procedure to the point when controls are determined at all stages of the process.

3. Optimal solution to the linear control system

We now proceed by considering a linear model with linear cost structure. This system has the advantage that the complete optimal solution can be derived explicitly, and subsequently it can be used as a measure of assessing the relative performance of sub-optimal rules introduced later in §4.

Theorem 1. *For the linear control system* (1) *to* (3) *with the additive linear cost structure* $W(4)$ *such that*

$$W_t \; = \theta_t \; + \omega^2 \, u^2_{t-1}\,, \tag{6}$$

where ω *is known constant, optimal control* u_t^* *at time* t *is deterministic and time variant only*

$$u_t^* = \frac{a(a^{N-t}-1)}{2(a-1)\omega^2}; \qquad (t = 1, 2, \cdots, N-1). \tag{7}$$

In addition, the minimised future expected loss measured at time t, {denoted by V_t^* *} is a linear function of the observations* x^t *and is determined by*

$$V_t^* = \alpha_t \mu_t + \beta_t, \tag{8}$$

where

$$\alpha_t = \frac{a(a^{N-t}-1)}{a-1}, \quad \beta_t = -\left\{\frac{a}{2(a-1)\omega}\right\}^2 \sum_{j=1}^{N-t} (a^j - 1)^2, \tag{9}$$

and

$$\mu_t = E\left(\theta_t \mid x^t\right) = \left\{\frac{x_t}{\rho^2} + \frac{a\left(\mu_{t-1} - u_{t-1}\right)}{\sigma^2 + a^2\sigma_{t-1}^2}\right\}\sigma_t^2. \tag{10}$$

Proof. Denoting the future expected loss at time t by V_t, we have

$$V_t = E\left\{\sum_{i=1}^{N-t}\left(\theta_{t+i} + \omega^2 u_{t+i-1}^2\right)\Big| x^t\right\}.$$

Using the generating equation for θ_t, we can write

$$V_t = E\left\{\theta_t\left(\sum_{j=1}^{N-t} a^j\right)\Big| x^t\right\} - \left(\sum_{j=1}^{N-t} a^j\right)u_t - \left(\sum_{j=1}^{N-t-1} a^j\right)u_{t+1} - \left(\sum_{j=1}^{N-t-2} a^j\right)u_{t+2}$$
$$- \cdots - a u_{N-1} + \omega^2 \sum_{j=0}^{N-t-1} u_{t+j}^2$$
$$\Rightarrow V_t = \frac{a\left(a^{N-t}-1\right)}{a-1}\mu_t - \sum_{j=0}^{N-t-1}\left\{\frac{a}{a-1}\left(a^{N-t-j}-1\right)u_{t+j} - \omega^2 u_{t+j}^2\right\}.$$

Minimising the above with respect to u_t gives the optimal control u_t^* at time *t*,

$$u_t^* = \frac{a\left(a^{N-t}-1\right)}{2(a-1)\omega^2}.$$

Substitution into V_t gives the minimised future expected cost at t, so that we have,

$$V_t^* = \frac{a\left(a^{N-t}-1\right)}{a-1}\mu_t - \left\{\frac{a}{2(a-1)\omega}\right\}^2 \sum_{j=0}^{N-t-1} \left(a^{N-t-j}-1\right)^2 ,$$

and by noting that the last summation is in fact equal to $\sum_{j=1}^{N-t} (a^j - 1)^2$, the proof of the theorem is thus established and the complete optimal solution is therefore available for future assessments with other procedures. We now introduce and define a sequence of near-optimal "myopic" rules, and proceed to evaluate their performance by examining their application to the linear system. This enables us to make comparative assessments in the light of the optimal solution derived in Theorem 1.

4. Near-optimal myopic procedures

General outline of the rules

At time point *t*, having observed $x^t \left\{= (x_0, x_1, \cdots, x_t)\right\}$ the Bayes' optimal control u_t^* is determined by applying a backward iterative procedure that starts at $N-1$, the last permissible control point of the process. Because of complications involved in this approach, in most cases u_t^* cannot be derived analytically and is determined by computation. To find a method that provides us with near-optimal analytical solution to the problem, we would need to investigate alternative approach to the backward iterative procedure. By applying the following myopic rules at time *t*, we can derive a sequence of approximating solutions by minimising certain future expected costs that are measured at time *t* and are conditioned on the available information, x^t.

As these myopic procedures do not involve backward iteration, they are often convenient to apply. In addition, the user has the option of increasing the accuracy of any particular approximating solution by taking further future costs into account than those currently considered, and, by applying a myopic rule of some higher "order" than that being used at the time. (The "order" of the myopic rule refers to the number of stages ahead of the present stage *t* that we wish to consider.) This means that, by applying the myopic rule of the *r-th* order (for *r* a positive integer; $r = 1, 2, \cdots, N-t$), we can determine the current policy at time *t* {denoted by $u_t^{(r)}$} and the future provisional controls {denoted by $u_{t;t+j}^{(r)}$; $j = 1, 2, \cdots, r-1$} estimated at time *t*. These controls are determined by minimising the future expected costs at *t*, and for the myopic rule of the *r-th* order future cost is composed of costs at $(t+1, t+2, \cdots, t+r)$.

Myopic rule of the "*r-th* order"– definition

Having observed x^t at time *t*, current control $u_t^{(r)}$ and the future provisional controls $u_{t;\,t+j}^{(r)}$, $\{ j = 1, 2, \cdots, r-1\}$, are determined respectively by minimising the future expected cost

$$E\left\{ \sum_{j=1}^{r} W_{t+j} \,\middle|\, x^t \right\},$$

with respect to u_{t+i} for $i = 0, 1, 2, \cdots, r-1$, $0 \le t \le N-1$, and integer *r*, $1 \le r \le N-t$.

As the myopic rules provide us with a sequence of approximating solutions at time *t*, we can compare and "rank" them according to their merits of closeness to the current Bayes' optimal solution u_t^*. We now proceed to examine the performance of these rules in the context of the linear control system used in Theorem 1, when optimal control was also found to be analytically obtainable.

5. Application of myopic rules to the linear control system

Theorem 2. *At time t, by applying the myopic rule of the r-th order* $\{ r = 1, 2, \cdots, N-t\}$ *to the linear control system* (1) *to* (3) *with the additive linear cost structure* (4) *and* (6), *we obtain a sequence of sub-optimal controls* $\{ u_t^{(1)}, u_t^{(2)}, \cdots, u_t^{(N-t)}\}$. *When these are compared with the optimal control* u_t^* *derived in Theorem* 1, *it follows that,*

$$u_{N-r}^{(r)} = u_{N-r}^{*} \qquad \left(r = 1, 2, \cdots, N-t \right). \tag{11}$$

In addition, these controls can be ranked with respect to the optimal control u_t^* *and, we have*

$$u_t^{(1)} < u_t^{(2)} < u_t^{(3)} < \cdots < u_t^{(N-t)} = u_t^{*}; \tag{12}$$

where

$$u_t^{(r)} = \frac{a\left(a^r - 1\right)}{2(a-1)\,\omega^2} \qquad \left(0 \le t \le N-1\right). \tag{13}$$

Proof. With reference to the definition of the myopic rule of the *r-th* order we can write

$$E\left\{ \sum_{j=1}^{r} W_{t+j} \,\middle|\, x^t \right\} = \frac{a(a^r - 1)}{a-1}\mu_t - \frac{a}{a-1}\left\{ \sum_{j=0}^{r-1}\left(a^{r-j} - 1\right)u_{t+j} \right\} + \omega^2\left(\sum_{j=0}^{r-1} u_{t+j}^2 \right).$$

Control at time t is hence determined by

$$u_t^{(r)} = \frac{a(a^r - 1)}{2(a-1)\,\omega^2} < u_t^* .$$

so that, we can intuitively expect myopic rules of the higher orders to provide us with closer approximations to the optimal control solution. That is

$$u_t^{(r-1)} < u_t^{(r)} \qquad \left(r \neq N - t \right).$$

Furthermore, the sequence of future provisional controls estimated at time t is derived as

$$u_{t;\,t+i}^{(r)} = \frac{a(a^{r-i} - 1)}{2(a-1)\,\omega^2} \qquad \left(i = 1, 2, \cdots, r-1 \right). \tag{14}$$

and when at future stages more information becomes available, these policies will be then revised accordingly.

Clearly for the myopic rule of the *r-th* order, the minimised future expected loss at t, namely

$$W_t^{(r)} = E\left\{ \sum_{j=1}^{r} W_{t+j} \,\middle|\, x^t,\ u_t^{(r)},\ u_{t;\,t+1}^{(r)}, \cdots, u_{t;\,t+r-1}^{(r)} \right\} \quad \left(r \neq N - t \right) \tag{15}$$

exceeds the optimum Bayes loss V_t^* [8] as required.

As an implication of Theorem 2, it is interesting to note that when applying the myopic rule of the "full order" {i.e. when $r = N - t$ } to the linear control system, current and future control policies are optimal and we have,

$$u_{t;t+i}^{(N-t)} = \frac{a\left(a^{N-t-i} - 1\right)}{2(a-1)\,\omega^2} \qquad \left(i = 0, 1, 2, \cdots, N - t - 1 \right). \tag{16}$$

The reason for the optimality of these solutions is twofold. Firstly, since solutions are found by using the full order myopic rule, we naturally expect to obtain the closest possible approximation to the optimal policies. Secondly, because of the deterministic structure of the optimal control policies in the present linear control system, both the Bayes optimal rule and the myopic rule of the full order are seen to be identical.

In a non-deterministic system, such as that with a non-linear cost structure, however, the two rules can be demonstrated to be completely different.

6. Conclusion

In general, the Bayes' optimal control policy for the normal Markov process involves a backward iterative procedure and in most cases optimal solutions are difficult to derive analytically. The Aoki linear control system studied here is, however, amongst the few cases with known explicit optimal solution. This allows assessment of the performance of certain sequence of introduced myopic procedures in relation to the Bayes optimal control solution to the system. These procedures are designed in such a way that by taking more future costs into account, we are able to obtain better approximations to optimal solution. The approximation is at its best when the complete future cost is considered. As these myopic rules do not involve a dynamic programming approach, they often have the advantage of being easier to apply in practice. In addition to the determination of current policy, by applying these rules we can also estimate the future provisional control policies conditioned to the current information. These provisional policies can be updated during the process as more information becomes available to the control system. Myopic rules, in certain cases provide the user with optimal solutions to the problem. Such a case is demonstrated in a deterministic system where solutions are independent of the current available information and are time-variant only.

References

1. M. Aoki, *Optimisation of Stochastic Systems (Topics in Discrete-Time Systems)*, Academic Press, New York, 1967.
2. J.A. Bather, Control charts and the minimisation of costs, *J. R. Stat. Soc. Ser. B Stat. Methodol.* **25** (1963), 49–80.
3. R. Bellman, *Dynamic Programming*, University Press, Princeton, 1957.
4. G.E.P. Box and G.C. Tiao, *Bayesian Inference in Statistical Analysis*, Addison-Wesley, London, 1973.
5. M.H. DeGroot, *Optimal Statistical Decisions*, McGraw-Hill, New York, 1970.
6. R.E. Kalman, A new approach to linear filtering and prediction problems, *A.S.M.E. Trans. J. Basic Eng.* **83, D,** (1960), 35–45.
7. J.P. La Salle, The time-optimal control problem, *Ann. of Math. Stud.* **45** (1960), 1–25.
8. D.V. Lindley, *Introduction to Probability and Statistics from a Bayesian View Point, Part 2, Inference*, University Press, Cambridge, 1980.
9. P.A. Whittle, A view of stochastic control theory, *J. Roy. Statist. Soc. Ser. A* **132** (1969), 320–334.
10. D.M.G. Wishart, A survey of control theory, *J. Roy. Statist. Soc. Ser. A* **132** (1969), 293–319.

Keywords: Bayes' optimal control policies; myopic control rules; sequentially observed data; control of stochastic parameter; linear Markov processes; additive cost structure; Bayesian analysis.