

## DUAL CONTROL OF AN INTEGRATOR WITH UNKNOWN GAIN

K. J. ÅSTRÖM and A. HELMERSSON

Department of Automatic Control, Lund Institute of Technology, Lund, Sweden

**Abstract**—The dual control law for an integrator with constant but unknown gain is computed. Numerical problems associated with the solution of the Bellman equation are reviewed. Properties of the dual control law are discussed. A representation which makes it easy to compare dual control with certainty equivalence and cautious control is also introduced.

### 1. INTRODUCTION

The notion of dual control was introduced by Feldbaum[1] as a result of an attempt to formulate optimal control problems which would give adaptive control laws. Several attempts were made to compute dual control strategies for different problems after Feldbaum's original publication. See, e.g. Florentin[2], Åström[3], Bohlin[4], Åström and Wittenmark[5], Jacobs and Patchell[6], Hughes and Jacobs[7], Tse and Bar-Shalom[8,9], Bar-Shalom and Tse[10,11]. Although the dual control formulation is conceptually appealing it leads to excessive computations. The computational difficulties led some researchers to despair and abandon the field, Wonham[12]. The computational power available in the late sixties was limited. Since much larger computing power is available today it seems reasonable to reconsider computation of dual control laws. The main reason is to develop an understanding of how dual control works. Another reason is to understand the limitations of the heuristic approaches to adaptive control like model reference adaptive control and self-tuning regulators.

In this paper the dual control law is computed for a process which can be described as an integrator with constant but unknown gain. This is one of the simplest nontrivial adaptive control problems.

The problem is stated in Sec. 2 and the Bellman equation is derived. Suitably scaled variables which admit good physical interpretations are also introduced. It is a nontrivial problem to solve the Bellman equation numerically. Different aspects on the numerics are discussed in Sec. 3. The results obtained are summarized in Sec. 4. Simulations of the optimal control law are given in Sec. 5.

### 2. PROBLEM STATEMENT

Consider the first order discrete time system described by

$$y(t+1) = y(t) + b u(t) + \sigma e(t+1), \quad (2.1)$$

where  $y(t)$  is the output or the process state,  $u(t)$  is the control variable and  $\{e(t)\}$  is a sequence of independent normal  $(0, 1)$  random variables. Furthermore,  $\sigma$  is a known constant and  $b$  is a random variable with a normal  $[\hat{b}(0), P(0)]$  distribution. The model (2.1) represents an integrator whose gain  $b$  is an unknown constant. The control problem for such a plant is nontrivial because a constant feedback may not necessarily give a stable system.

Assume that the purpose of control is to minimize the loss function

$$J = E \left\{ \sum_{k=t+1}^{t+T} [y(k)]^2 | Y_t \right\}, \quad (2.2)$$

where  $Y_t$  denotes the sequence of observed outputs and inputs available at time  $t$ , i.e.  $Y_t = (x(t), x(t-1), \dots, u(t-1), u(t-2), \dots)$ , or more precisely the sigma algebra generated by

those variables. The symbol  $E(\cdot|Y_t)$  denotes the conditional expectation given  $Y_t$ . The admissible controls are such that  $u(t)$  is a function of  $Y_t$ .

### The hyperstate

To formulate the optimal control problem it is useful to introduce the hyperstate of the problem. See, e.g. Åström[13]. The hyperstate is the state  $y(t)$  of (2.1) and the conditional distribution of  $b$  given  $Y_t$ . In this case the conditional distribution is Gaussian with mean

$$\hat{b}(t) = E[b(t)|Y_t],$$

and covariance

$$P(t) = E\{[\hat{b}(t) - b(t)]^2|Y_t\}.$$

See Åström[13]. These moments can be computed recursively from

$$\hat{b}(t+1) = \hat{b}(t) + K(t)[y(t+1) - y(t) - \hat{b}(t)u(t)], \quad (2.3)$$

$$P(t+1) = [1 - K(t)u(t)]P(t), \quad (2.4)$$

where the estimator gain is

$$K(t) = u(t)P(t)\gamma^{-2}(t), \quad \gamma^2(t) = \sigma^2 + u^2P(t).$$

Introduce the normalized innovation

$$\epsilon(t) = [y(t+1) - y(t) - \hat{b}(t)u(t)]/\gamma(t),$$

and the normalized state variables

$$\begin{aligned} \eta(t) &= y(t)/\sigma, & \beta(t) &= \hat{b}(t)/\sqrt{P(t)}, \\ \xi(t) &= 1/\sqrt{P(t)}, & \nu(t) &= u(t)\sqrt{P(t)}/\sigma. \end{aligned}$$

Eqs. (2.1), (2.3) and (2.4) can then be written as

$$\eta(t+1) = (\eta + \beta\nu + \sqrt{1 + \nu^2} \epsilon)(t), \quad (2.5)$$

$$x(t+1) = (\beta\sqrt{1 + \nu^2} + \nu\epsilon)(t), \quad (2.6)$$

$$\xi(t+1) = (\xi\sqrt{1 + \nu^2})(t). \quad (2.7)$$

### The Bellman equation

Assume that the minimum of the loss function is assumed for an admissible control law and introduce the function

$$W_T(y, \hat{b}, P, t) = \min E \left\{ \sum_{k=t+1}^{t+T} y(k)^2 | Y_t \right\}.$$

A standard Dynamic Programming argument gives the following Bellman equation:

$$W_T(y, \hat{b}, P, t) = \min_{u(t)} E \{ y(t)^2 + W_{T-1}(y(t+1), \hat{b}(t+1), P(t+1), t+1) | Y_t \}.$$

The conditional distributions of  $y(t+1)$  and  $\hat{b}(t+1)$  are normal, and the conditional distri-

bution of  $P(t + 1)$  is a point distribution. Hence

$$W_T(y, \hat{b}, P, t) = \min_u \left\{ (y + \hat{b}u)^2 + \sigma^2 + u^2P + \int_{-\infty}^{\infty} \varphi(\epsilon) W_{T-1} \left( y + bu\sqrt{\sigma^2 + u^2P}, b + \frac{uP}{\sqrt{\sigma^2 + u^2P}}, \frac{\sigma^2P}{\sigma^2 + u^2P}, t + 1 \right) d\epsilon \right\},$$

where

$$\varphi(\epsilon) = (1/\sqrt{2\pi}) e^{-\epsilon^2/2}.$$

Using the normalized variables and the corresponding loss function  $V$  the equation can be written as

$$V_T(\eta, \beta, \xi, t) = \min_v \left\{ (\eta + \beta v)^2 + 1 + v^2 + \int_{-\infty}^{\infty} \varphi(\epsilon) V_{T-1}(\eta + \beta v + \sqrt{1 + v^2} \epsilon, \beta\sqrt{1 + v^2} + v\epsilon, \xi\sqrt{1 + v^2}, t + 1) d\epsilon \right\}. \quad (2.8)$$

For  $T = 1$  the loss function becomes

$$W_1(y, \hat{b}, P, t) = \min_u \{(y + \hat{b}u)^2 + \sigma^2 + u^2P\}.$$

The corresponding equation in normalized variables is

$$v_1(\eta, \beta, \xi, t) = \min_v \{(\eta + \beta v)^2 + 1 + v^2\} = 1 + \eta^2/(1 + \beta^2).$$

The optimization is easy to do because the right-hand side is quadratic in  $v$ . Notice that  $V_1$  does not depend on  $\xi$  and  $t$ .

It follows by induction that  $V_T$  also does not depend explicitly on  $\xi$  and  $t$ .

Introducing

$$U_T(\eta, \beta, v) = (\eta + \beta v)^2 + 1 + v^2 + \int_{-\infty}^{\infty} \varphi(\epsilon) V_{T-1}(\eta + \beta v + \sqrt{1 + v^2} \epsilon, \beta\sqrt{1 + v^2} + v\epsilon) d\epsilon, \quad (2.9)$$

Eq. (2.8) can thus be written as

$$V_T(\eta, \beta) = \min_v U_T(\eta, \beta, v). \quad (2.10)$$

The solution to the stochastic optimal control problem is thus reduced to solving the recursive optimization problem given by the *Bellman equation* (2.10). The solution of this equation defines an optimal control as

$$v = v_0(\eta, \beta), \quad (2.11)$$

which gives the normalized  $v$  control as a function of the normalized hyperstate  $\eta$  and  $\beta$ .

*Learning loss*

If the parameter  $b$  is known the control problem has a simple solution given by the minimum variance control law

$$u = -x/b.$$

See Åström[14]. This gives the normalized loss function

$$V_T = T.$$

The difference

$$L_T(\eta, \beta) = V_T(\eta, \beta) - T$$

can thus be interpreted as the part of the loss function which is due to the uncertainty in the parameter  $b$ . This is called the *learning loss*.

### 3. NUMERICS

The Bellman equation does not have an analytical solution. It is therefore necessary to solve (2.10) numerically. One iteration of (2.10) can be decomposed into three steps.

- (1) Representation of the functions  $V_{T-1}(\eta, \beta)$  by their values in discrete points. Values for other arguments are obtained by interpolation.
- (2) Evaluation of the integral in (2.9) using a quadrature formula.
- (3) Minimization of the function  $U(\eta, \beta, \nu)$ .

The steps are closely interrelated. The selection of methods in each step requires insight into the properties of the function  $V(\eta, \beta)$ .

#### *Properties of the loss function*

For  $T = 1$  there is an analytical solution to the optimization problem. Although there is no known analytical solution for larger  $T$  there are some symmetry properties which are useful. The loss function  $V$  is symmetric in both  $\eta$  and  $\beta$ . The optimal control function  $\nu$  is asymmetric in  $\eta$  and  $\beta$ . This can be derived recursively from Eq. (2.10). It can also be understood intuitively because if the output  $\eta$  of the process changes sign then the control variable  $\nu$  must also change sign.

For  $T = 1$  the loss function is equal zero both when  $\eta = 0$  and  $\beta = 0$ . For  $T \geq 1$  we can expect that  $\nu$  is not equal zero for either  $\eta = 0$  or  $\beta = 0$  because of probing. We may thus expect discontinuities in  $\nu$  when  $\eta$  or  $\beta$  change sign, due to asymmetry. The function  $\nu(\eta, \beta)$  may also have other discontinuities. To understand how these occur notice that it follows from (2.10) that the value of the function  $\nu$  for the arguments  $\eta$  and  $\beta$  are obtained by searching  $U(\eta, \beta, \nu)$  for the value  $\nu$  which gives the smallest value. If the function  $U(\eta, \beta, \nu)$  for fixed  $\eta$  and  $\beta$  has several local minima to another when  $\eta$  or  $\beta$  is changed a little. This occurs, e.g., for  $U_3(0, \beta, \nu)$  as is shown in Fig. 1. The function  $V(\eta, \beta)$  may have a discontinuity in its first partial derivatives where  $\nu(\eta, \beta)$  is discontinuous.

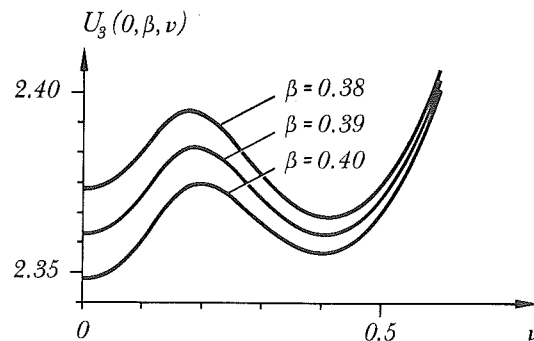


Fig. 1. Illustrates why the control law is discontinuous.

*Some difficulties*

In ordinary quadrature algorithms, like Simpson's rule, a continuous integral is replaced by a weighted sum of the values of the integrand at discrete points in the integration interval.

To compute the integral in Eq. (2.10), the loss function  $V_{T-1}$  is first evaluated in several old coordinates  $(\eta, \beta)$  dependent on  $\nu$  and  $\epsilon$ . For each  $\nu$ , a weighted sum is computed on a set of data points of  $V_{T-1}(\eta, \beta)$ . When  $\nu$  is varied the set of data points, for which the sum is computed, is also varied. Such a computation may introduce artificial local minima in the computed function.

It is thus important to consider the interplay between interpolation, quadrature and optimization. It is also very important to remember that the optimal control may be discontinuous. Having understood some of the difficulties involved it is now possible to tailor the algorithms to the problem.

*Minimization*

The minimization algorithm used is straightforward. Three equally spaced node points are fitted to a parabola and the minimum of this parabola is tested as a potential global minimum. Since the function  $U$  may have several local minima it is necessary to combine the local optimization with a global search.

*Interpolation*

The functions  $V$  and  $\nu$  were represented by their values in discrete points. The variables  $\eta$  and  $\beta$  were transformed by the mapping  $x \rightarrow x/(1+x)$ . A uniform grid in the transformed variables was used. The grid sizes were varied from  $16 \times 16$  to  $128 \times 128$ . The function values were interpolated using an interpolation polynomial having sixteen coefficients. Because of the symmetry properties of  $V$  and  $\nu$  the functions are tabulated for one octant only.

*Quadrature*

The range of integration is reduced to the interval  $[-5, 5]$ , without any significant loss of accuracy. The quadrature algorithm is modified to cope with discontinuities in the first derivative of  $V$  for  $\eta = 0$  and  $\beta = 0$ . The discontinuity points are, e.g.,  $\epsilon = -(\eta + \beta\nu)/\sqrt{1 + \nu^2}$  and  $\epsilon = -\beta\sqrt{1 + \nu^2}/\nu$ .

The integral is evaluated by an ordinary Simpson's algorithm, except for those intervals which containing discontinuities. These intervals are divided into two subintervals, one on either side of the discontinuity. The subintervals are then evaluated using Simpson's algorithm and accumulated to the total integral.

## 4. RESULTS

The algorithms were programmed in Pascal on a Vax-11/780 computer. Simple commands were introduced to give flexibility to change parameters, sequences of iterations, input/output handling, etc.

*The dual control law*

The dual control law (2.11) which is a function of two variables may be represented by surface and level plots. It is useful to introduce some scaling. The control variable  $u$  or  $\nu$  is represented as

$$\mu = u\hat{b}/y = -\nu\beta/\eta. \quad (4.1)$$

To obtain finite plots the state variables  $\eta$  and  $\beta$  are also represented using the scales  $\eta/(1+\eta)$  and  $\beta^2/(1+\beta^2)$ . The advantage of this representation is that the *certainty equivalence* control is the plane

$$\mu = 1, \quad (4.2)$$

and the cautious control is the plane

$$\mu = \beta^2/1 + \beta^2. \quad (4.3)$$

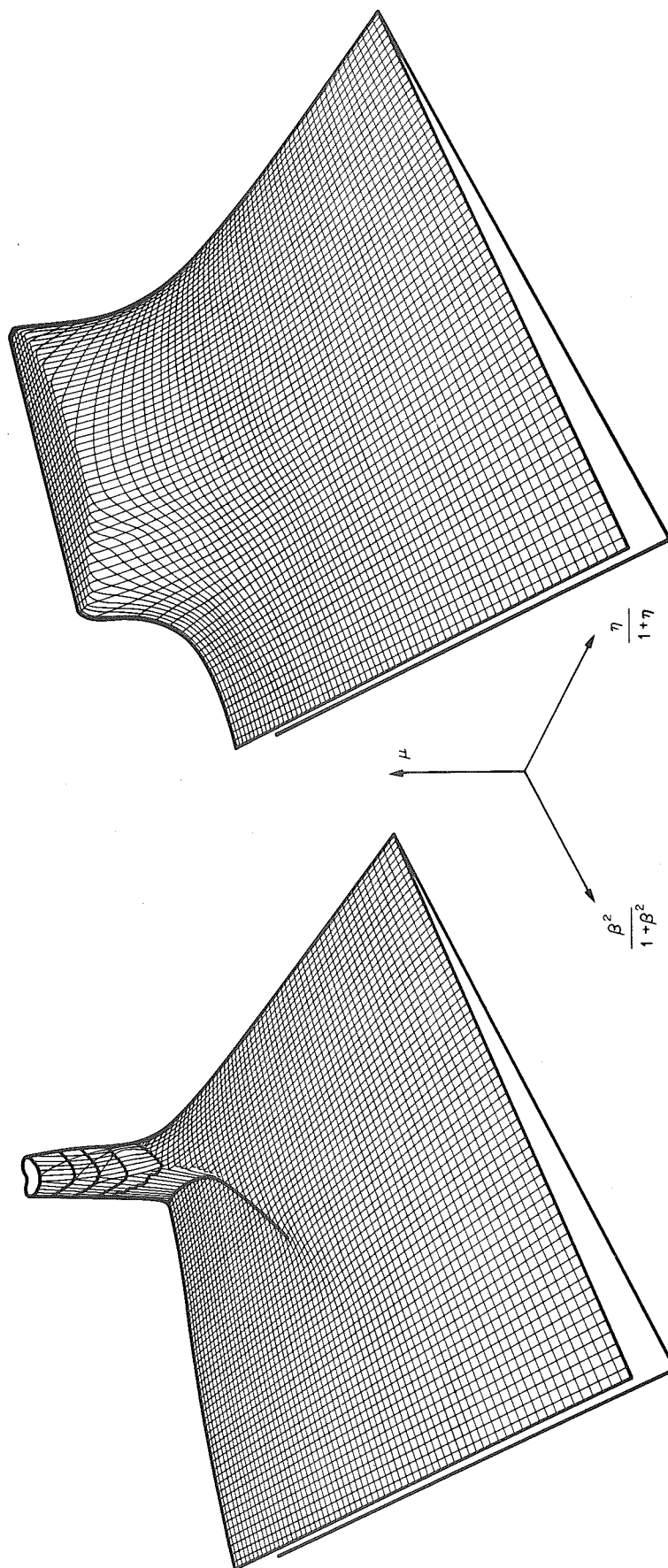


Fig. 2. Dual control laws for (A)  $T = 3$  and (B)  $T = 31$ .

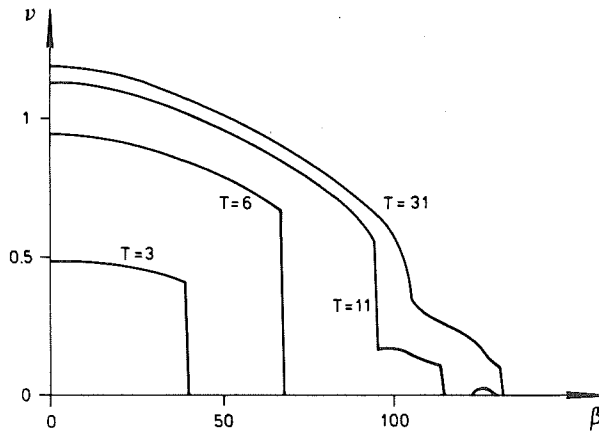


Fig. 3. The dual control law  $v(0, \beta)$  for  $T = 3, 6, 11$  and  $31$ .

Using the chosen representation it is easy to compare dual control with certainty equivalence and cautious control. The dual control laws for the time horizons  $T = 3$ , and  $T = 31$  are shown in Fig. 2. Notice that the dual control function agrees with the certainty equivalence control when the parameter  $\beta$ , which represents normalized parameter uncertainty is very large. Dual control also agrees with cautious control when the parameter  $\eta$ , which represents normalized control error, is very large.

For small values of  $\eta$  and  $\beta$  the dual control is much larger than the cautious control. Notice that there is a zone where dual control is larger than the certainty equivalence control. This zone grows with increasing time horizon  $T$ .

The dual control is larger than the cautious control except for  $2 \leq T \leq 6$  when there exists a small region where the dual control is smaller than the cautious control.

It appears from Fig. 2(A) that the dual control law is discontinuous. This is shown more clearly in Fig. 3 where  $v(0, \beta)$  is plotted versus  $\beta$ . Notice that both the certainty equivalence and the cautious controls are zero for  $\eta = 0$ . For  $T = 3$  there is a discontinuity for  $\beta = 0.39$ . The reason for this is the switch from one local minimum to another which was demonstrated in Fig. 1. The discontinuity is more pronounced for  $T = 6$  but it appears that the discontinuities are less pronounced for increasing  $T$ . Computations up to  $T = 100$  indicate little difference between the control policies for  $T = 100$  and  $T = 31$ . The difference  $|v_{100} - v_{30}|$  is less than 0.012 for all  $\eta$  and  $\beta$ .

The question whether a limiting control policy exists  $T \rightarrow \infty$  is open. A simple approximation, Helmersson[15] indicates that no limit learning loss function exists, i.e. the learning loss tends to increase infinitely.

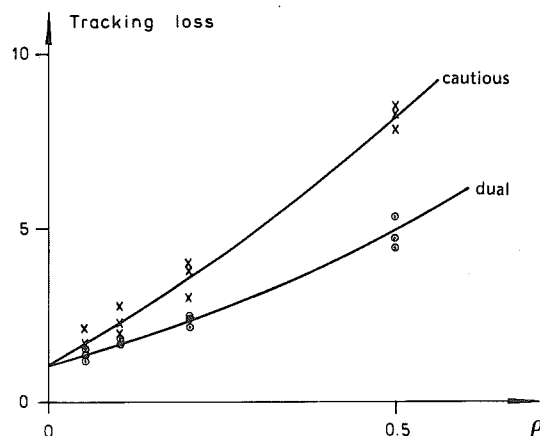


Fig. 4. Comparison between dual and cautious control for a process with time varying gain. The tracking loss is plotted versus the gain noise parameter  $\rho$ .

## 5. SIMULATIONS

A few simulations were performed to get further insight into the properties of the dual control law. This includes comparisons with self-tuners and cautious control laws. Differences between the control laws may be expected only during the transient phase when there are significant parameter uncertainties. The steady state performance of the control laws will be more or less the same.

*Transient behaviour*

To perform the simulations it is necessary to have some representation of the dual control law. A table and an interpolation formula could be used. It is, however, more convenient to have an analytic approximation. The following approximation of the dual control law obtained for  $T = 31$  was used.

$$v(\eta, \beta) = \frac{0.56 + \beta}{2.2 + 0.08\beta + \beta^2} \eta + \frac{1.9}{1.7 + \beta^4}, \quad \eta > 0, \quad \beta > 0. \quad (5.1)$$

Notice that  $v$  is asymmetric in  $\eta$  and  $\beta$ . The first term can be interpreted as a modified certainty equivalence and cautious term. The second term represents active probing. It excites the process even when  $\eta = 0$ . The loss function of this suboptimal control was only 0.2 units larger than the optimal loss function for  $T = 31$ . This corresponds to an increase in the loss function of less than 1%.

Simulations were made using the interactive simulation language SIMNON. See Elmqvist[16]. The simulations showed that the dual regulator is very robust. The loss function did not vary much for different simulations with different parameters and initial values. The learning period for dual control is in most cases less than five steps. There are no problems even if the algorithm is initialized with the wrong sign of the gain. The initial value of the variance  $P(0)$  should, however, not be too small, because the learning period will then be extended and the loss increases.

The dual regulator was compared with certainty equivalence control and cautious control.

The certainty equivalence control is sometimes good. During start-up periods it gives, however, too large control signals when the gain estimate is close to zero. It may therefore give very large losses during start-up. The asymptotic behaviour is, however, very close to optimal. The cautious control is too cautious which results in long learning periods when gain estimate is close to zero because of turn-off. The dual control does not have any tendencies of turn-off. The simulations indicate clearly that it may be useful to make some modifications of the self-tuning regulator in order to improve its transient behaviour.

*Tracking*

Even if the dual controller discussed in this paper is not intended for time varying gain, it can be expected to be a good approximation to the optimal controller if the parameters do not change too rapidly. Suppose that the gain can be modelled as

$$b(t + 1) = \phi b(t) + \rho v(t), \quad (5.2)$$

where  $\{v(t)\}$  is white Gaussian noise with zero mean and unit variance.

The estimator must then be modified. The Bellman equation is more difficult to solve, because it has three independent variables instead of two. The execution time for the new problem is significantly longer. It has been investigated how well the regulator (5.1) behaves in combination with the correct state estimator for (5.2).

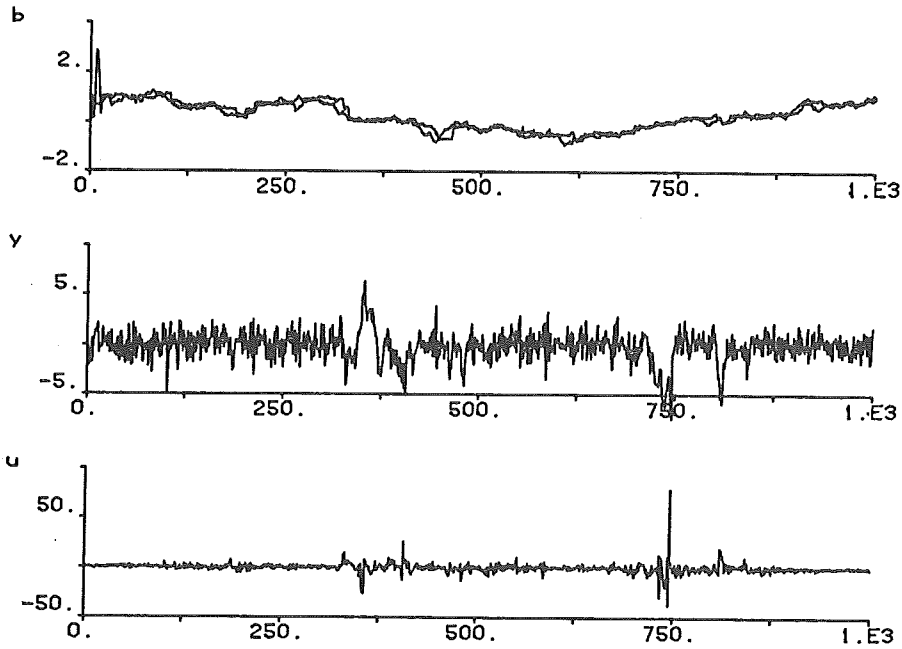
$$\frac{1}{N} \sum_{t=1}^N y(t)^2$$

was evaluated for different noise realizations and different values of  $\rho$ . The parameter  $\phi$  was chosen so that the gain  $b$  would get unit variance:

$$\phi = \sqrt{1 - \rho^2}.$$



Cautious



Dual

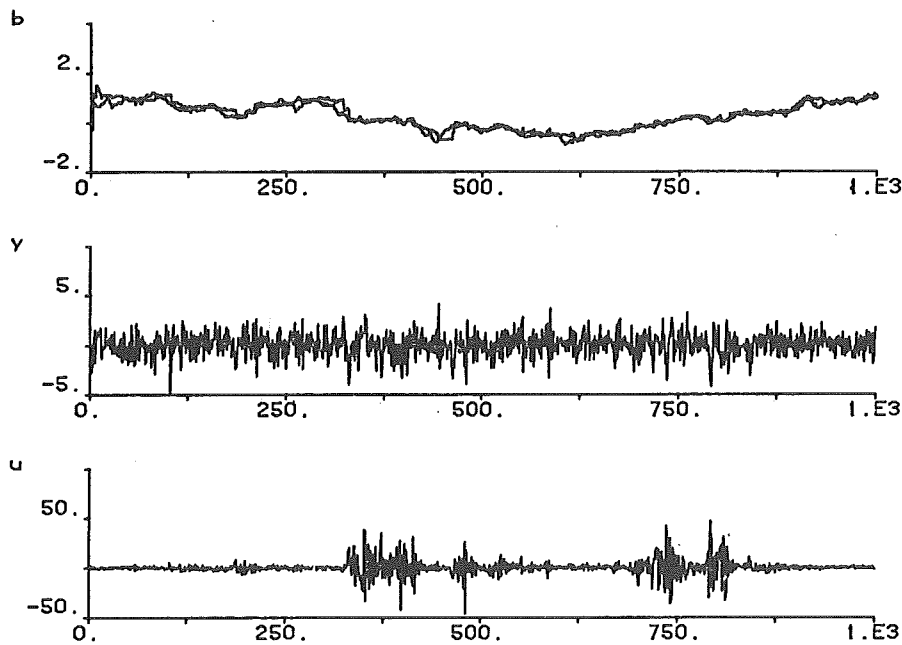


Fig. 5. Simulation showing the difference between dual and cautious control. Gain noise parameter  $\rho = 0.05$ .

The tracking loss is  $\sigma^2$  as  $N \rightarrow \infty$  in the constant gain case, i.e., when  $\rho = 0$  and  $\phi = 1$ . In the simulations the estimator equations were modified to cope with nonconstant gain as modelled above. Three simulations were made for each  $\rho$  to compare the steady state performance of the regulators. In Fig. 4 the tracking loss versus  $\rho$  is plotted for cautious control and dual control. Certainty equivalence control deteriorated and gave much larger tracking losses than the other regulators. This was due to too large control signals when the gain estimate was close to zero.

Figure 5 shows the difference between dual and cautious control for the case  $\rho = 0.05$ . The same noise sequences were used for both the controllers. The cautious controller regulates well, except in intervals where the gain is close to zero. It can clearly be seen that the cautious controller deteriorates in these intervals. The dual controller, on the other hand, copes well with these problems using active probing.

## 6. CONCLUSIONS

Calculation of dual control laws gives interesting insight into the nature of the optimal solutions to adaptive control problems. In particular it is shown that the optimal control may be discontinuous. This may be one possibility to define the probing action. The comparisons with cautious and certainty equivalence control explain why the certainty equivalence control behaves so well asymptotically. It also explains the poor behaviour of cautious control under certain conditions (turn off). The results also indicate that the heuristic algorithms may be improved. There have been suggestions for improved strategies based on approximative solutions to problems with a short time horizon, e.g.  $T = 2$ . See Sternby[17]. Such approximations are of limited value for the example discussed in this paper because there is a considerable difference between the strategies obtained for  $T = 2$  and  $T = 30$ . Based on the results of this paper it seems important to use a much longer time horizon to obtain the full benefits of learning.

## REFERENCES

1. A. A. Feldbaum, Dual control therapy I-IV. *Auto. Remote Control* **21**, 874-880 (1960); **21**, 1033-1039 (1960); **22**, 1-12 (1960); **22**, 109-121 (1960).
2. J. J. Florentin, Optimal probing, adaptive control of a simple Bayesian system. *Int. J. Control* **13**, 165-177 (1962).
3. K. J. Åström, Optimal control of Markov processes with incomplete state information. *J. Math. Anal. Appl.* **10**, 174-205 (1965).
4. T. Bohlín, Optimal dual control of a simple process with unknown gain. Technical paper TP 18.196. IBM Nordic Laboratory, Lidingö, Sweden (1969).
5. K. J. Åström and B. Wittenmark, Problems of identification and control. *J. Math. Anal. Appl.* **34**, 90-113 (1971).
6. O. L. R. Jacobs and J. W. Patchell, Caution and probing in stochastic control. *Int. J. Control* **16**, 189-199 (1972).
7. D. J. Hughes and O. L. R. Jacobs, Turn-off, escape and probing in nonlinear stochastic control, *Proc. IFAC Symp. on Adaptive Control*, Budapest, Hungary (1974).
8. E. Tse and Y. Bar-Shalom, An actively adaptive control for discrete-time systems with random parameters. *IEEE Trans. Auto. Control* **AC-18**, 109-117 (1973).
9. E. Tse and Y. Bar-Shalom, Generalized certainty equivalence and dual effect in stochastic control. *IEEE Trans. Auto. Control*, **AC 20**, 817-819 (1975).
10. Y. Bar-Shalom and E. Tse, Dual effect, certainty equivalence, and separation in stochastic control. *IEEE Trans. Auto. Control* **AC-19**, 494-500 (1974).
11. Y. Bar-Shalom and E. Tse, Caution, probing, and the values of information in the control of uncertain systems. *Ann. Econ. Soc. Meas.* **5**, 323-337 (1976).
12. M. Wonham, Optimal stochastic control. In *Stochastic Problems in Control* (Edited by B. Friedland). ASME, New York (1968).
13. K. J. Åström, Stochastic control problems. In *Mathematical Control Theory*. (Edited by W. A. Coppel). Lecture Notes in Mathematics, Springer-Verlag, Berlin (1978).
14. K. J. Åström, *Introduction to Stochastic Control Theory*. Academic, New York (1970).
15. A. Helmersson, (Dual control—An optimal dual regulator for an integrator with constant but unknown gain. In Swedish). CODEN:LUTFD2/(TRFT-5245)/1-86/(1981). Department of Automatic Control, Lund Institute of Technology, Lund, Sweden (1981).
16. H. Elmqvist, SIMNON—An interactive simulation program for nonlinear systems. User's Manual. CODEN:LUTFD2/(TFRT-3091)/. Department of Automatic Control, Lund Institute of Technology, Lund, Sweden (1975).
17. J. Sternby, Topics in dual control. CODEN:LUTFD2/(TFRT-1012)/1-135(1977). Department of Automatic Control, Lund Institute of Technology, Lund, Sweden (1977).