

Nonconvexities in a stochastic control problem with learning

Bruce Mizrach*

*Wharton School, University of Pennsylvania, Philadelphia, PA 19104, USA
Department of Economics, Boston College, Chestnut Hill, MA 02167, USA*

Received December 1989, final version received May 1990

This paper examines the benefits from active learning in a stochastic control problem. In a linear model with parametric uncertainty, there are gains to probing, but the probing component of the loss function often has nonconvexities. I show that they can arise for two reasons: (1) failure of the precision matrix of the parameters to increase monotonically with the control variable (the covariances between a state variable and the random parameters can reduce the information gained from probing) and (2) changes in the path of future state variables induced by modifying the certainty-equivalent control. If the parameter on the control variable is large, a small change in the control can cause a much larger change in future state which, for a given level of uncertainty, makes probing more costly.

1. Introduction

This paper examines the benefits from active learning in a stochastic control problem. In a dynamic programming framework with parametric uncertainty, there are often gains to probing, moving the control variable away from its certainty-equivalent value, so as to increase precision of parameter estimates. The controller willingly trades current-period losses for greater certainty about the future decisions he must make.

Kendrick (1978) has reported from computer simulations that costs do not decrease monotonically with increases in the control variable. The probing component of the loss function has nonconvexities. Since this greatly complicates the search for the optimal control, I try to pin down the source of these

*I would like to thank David Kendrick for bringing this problem to my attention and for his remarks on an earlier draft. I am also grateful to an anonymous referee for helpful comments. Any remaining deficiencies are entirely my own.

nonconvexities. I show below that the nonconvexities can arise for two reasons.

The first possible cause is the covariance between the unknown parameters and the state variable. An agent updates his certainty-equivalent beliefs recursively in the programming problem, despite receiving no actual realization on the state itself. Whether or not a particular realization of the state is informative depends upon its covariance with the parameters. If the correlation between the state and the random parameter falls, this can make increases in the control value less valuable. Increases in the control variable ultimately do reduce the parametric uncertainty to zero in the limit, but between the certainty-equivalent value and this limiting value there will often be at least one region in which increases in the control value raise parametric uncertainty before it begins to decline.

The second cause is the changes in the path of future state variables induced by a change in the certainty-equivalent control. If the parameter on the control variable is large, a small change in the control can cause a much larger change in the future state. Even if the increase in the control value reduces parametric uncertainty, the oscillations induced by the change in the first-period control can dominate this effect.

Section 2 states the adaptive control problem in a fairly general way. Following Bar-Shalom and Tse (1976) and Kendrick (1981), I divide the cost-to-go into three components – deterministic, cautionary, and probing. The probing term, which describes the gains from active learning, is studied in greater detail in section 3. The rather involved intuition about the nonconvexities is developed further in section 4. Section 5 has a numerical example and simulation of these results in the context of the McRae (1972) problem. Section 6 has a summary and conclusion.

2. A linear adaptive control problem

Consider a policy authority whose objective is to minimize an expected loss function L :

$$J_N = E \left[L_N(x_N) + \sum_{k=0}^{N-1} L_k(x_k, u_k) \right], \quad (1)$$

where u is the authority's control variable and x is a state variable, which evolves according to:

$$x_{k+1} = ax_k + bu_k + \varepsilon_k. \quad (2)$$

ε_k is an additive disturbance term, distributed normally with mean zero and

variance Q :

$$\varepsilon \sim N(0, Q). \quad (3)$$

a and b are assumed to be unknown parameters, with prior mean

$$\theta_{0|0} = \begin{bmatrix} a_{0|0} \\ b_{0|0} \end{bmatrix} \quad (4)$$

and covariance

$$\Sigma_{\theta\theta} = \begin{bmatrix} \text{var}(a)_{0|0} & \text{cov}(a, b)_{0|0} \\ \text{cov}(b, a)_{0|0} & \text{var}(b)_{0|0} \end{bmatrix}. \quad (5)$$

The recursive aspect of this problem can best be seen by stating this objective as a dynamic programming problem:

$$J_{N-k}^* \equiv \min_{u^k} E[L_k(u_k, x_k) + J_{N-k-1}^* | x^k, u^k], \quad (6)$$

with x and u realizations on the control and state variables through period k . We will refer to J_{N-k}^* as the cost-to-go. The first step of the optimal control is to solve (6) for the certainty-equivalent path: $(x_{k|0}, u_{k|0})_{k=0}^N$, which is the optimal control for the model with all random variables set to their expected values.

In Bar-Shalom and Tse and Kendrick, the cost-to-go is separated into three components – the deterministic or certainty-equivalent loss, the cautionary cost, and the probing cost. The certainty-equivalent loss evaluates (1) at the certainty-equivalent values:

$$J_{N-k}^{\text{CE}} = L_N(x_{N|0}) + \sum_{j=k+1}^{N-1} L_j(x_{j|0}, u_{j|0}). \quad (7)$$

The cautionary cost is given by

$$J_{N-k}^{\text{C}} = \text{tr}(K_{k+1} \Sigma_{k+1|k}) + \sum_{j=k+1}^{N-1} \text{tr}(K_{j+1} Q), \quad (8)$$

where K are Riccati matrices to be defined below, and $\Sigma_{k+1|k}$ is the one-period-ahead conditional covariance matrix for all the random elements in the system. This is developed further in section 3.1.

Of greatest interest in this paper will be the probing component:

$$J_{N-k}^P = \sum_{j=k+1}^{N-1} \text{tr}(\mathcal{R}_j \Sigma_{j|j}), \quad (9)$$

where \mathcal{R} is a complicated expression involving many system parameters and $\Sigma_{k|k}$ is the contemporaneous conditional covariance matrix.

Kendrick (1978) has noted that this term often is nonconvex. As this contributes to nonconvexities in the entire cost-to-go, it makes searching for the optimal control much more difficult. Understanding the source of these nonconvexities will be useful in figuring out their importance in the dual control problem.

3. The gains from probing

The nonconvexities in the probing cost term have not, as of yet, been understood analytically. Kendrick (1981) reports numerical results for the multi-period McRae problem that confirmed similar findings by Norman (1976). Their numerical analyses, however, have not yielded any insight into the source of the nonconvexities.

The intuition that volatility reduces uncertainty turns out to be correct only in special cases. With a single unknown parameter, $\Sigma^{\theta\theta}$ will decrease monotonically with the sum of squared residuals for the state. This is not always true with more than one parameter though. Section 3.1 shows that under highly plausible conditions, probing may be less informative over some range.

The Riccati matrix terms are developed in section 3.2. Even if probing reduces uncertainty, the path of future state variables will be affected by movements in the control variable. The Riccati matrices will tell us the cost, in future periods, of probing.

3.1. An examination of the precision matrix Σ

The system covariance matrix measures the uncertainty over the state and the parameters:

$$\Sigma_{k|k} = \begin{bmatrix} \Sigma_{k|k}^{xx} & : & \Sigma_{k|k}^{x\theta} \\ \cdots & & \cdots \\ \Sigma_{k|k}^{\theta x} & : & \Sigma_{k|k}^{\theta\theta} \end{bmatrix}, \quad (10)$$

where

$$\Sigma_{k|k}^{xx} = E[(x_k - x_{k|k})(x_k - x_{k|k}) | x^k], \quad (11)$$

$$\Sigma_{k|k}^{\theta x} = (\Sigma_{k|k}^{x\theta})^T = E[(x_k - x_{k|k})(\theta_k - \theta_{k|k}) | x^k], \quad (12)$$

and $\Sigma_{k|k}^{\theta\theta}$ is given by (5). Since x is not measured with error in our problem, $x_{k|k}$ will always equal x_k , and these two terms will always equal zero.

Projecting ahead though, they will not equal zero because of parametric uncertainty and the disturbance term on the state equation:

$$\Sigma_{k+1|k}^{xx} = E[(x_{k+1} - x_{k+1|k})(x_{k+1} - x_{k+1|k}) | x^k], \quad (13)$$

$$\Sigma_{k+1|k}^{\theta x} = (\Sigma_{k+1|k}^{x\theta})^T = E[(x_k - x_{k+1|k})(\theta_{k+1} - \theta_{k+1|k}) | x^k]. \quad (14)$$

To obtain approximations to these terms, Kendrick takes a second-order expansion around $x_{k+1|k} = [x_k \ u_k] \theta_{k+1|k} = [x_k \ u_k] [a_{k+1|k} \ b_{k+1|k}]^T$. This yields

$$\Sigma_{k+1|k}^{xx} = x_k^2 \text{var}(a)_{k|k} + 2x_k u_k \text{cov}(a, b)_{k|k} + u_k^2 \text{var}(b)_{k|k} + Q, \quad (15)$$

which is just the conditional variance of the normally distributed multivariate random variable x_{k+1} , given x_k . Similarly, one finds

$$\Sigma_{k+1|k}^{\theta x} = \begin{bmatrix} x_k \text{var}(a)_{k|k} + u_k \text{cov}(a, b)_{k|k} \\ x_k \text{cov}(a, b)_{k|k} + u_k \text{var}(b)_{k|k} \end{bmatrix}. \quad (16)$$

Prior to receiving new data, the parameter covariance matrix is unchanged: $\Sigma_{k+1|k}^{\theta\theta} = \Sigma_{k|k}^{\theta\theta}$.

Updating the parameter covariance matrix is a Kalman filter problem, with future variables set to their certainty-equivalent values. For our problem:

$$\Sigma_{k+1|k+1}^{\theta\theta} = \Sigma_{k+1|k}^{\theta\theta} - \Sigma_{k+1|k}^{\theta x} (\Sigma_{k+1|k}^{xx})^{-1} \Sigma_{k+1|k}^{x\theta}. \quad (17)$$

From (3.8) the standard intuition is straightforward. Assume only b is an unknown parameter, making $\Sigma^{\theta\theta}$ a scalar. Eq. (17) reduces to

$$\begin{aligned} \text{var}(b)_{k+1|k+1} &= \text{var}(b)_{k+1|k} \\ &\quad - u_k \text{var}(b)_{k|k} (u_k^2 \text{var}(b)_{k|k} + Q)^{-1} u_k \text{var}(b)_{k|k}, \end{aligned} \quad (18)$$

which we will show below is decreasing monotonically with the control variable u . This will, all other things equal (which is not the case), reduce the probing cost. I now turn to look at those other things.

3.2. The Riccati matrix terms \mathcal{R}

I begin this section by defining the underlying elements of the term \mathcal{R} . To save on notation, all random variables will be assumed to equal their period k certainty-equivalent values, e.g., $b = b_{j|k}$, $x_j = x_{j|k}$, $j = k, \dots, N$.

$$\begin{aligned} \mathcal{R}_k = & \left[\left[K_{k+1}^{\theta x} + f_{\theta k}^{xT} K_{k+1}^{xx} \right] b + p_k^x b_{\theta} \right] \mu_k \\ & \times \left[b \left[K_{k+1}^{xx} f_{\theta k}^x + K_{k+1}^{\theta x} \right] + \left[p_k^x b_{\theta} \right]^T \right], \end{aligned} \quad (19)$$

where p_k^x and K_{k+1} are Riccati matrices, $f_{\theta k}^x = [x_k \ u_k]$, $\mu_k = [A_k + bK_{k+1}^{xx}b]^{-1}$, and $b_{\theta} = [0 \ 1]^T$. A_k is a penalty matrix for deviations from a desired control value $\bar{\mu}$. The Riccati matrices may be found recursively, working back from terminal values, $K_N^{xx} = W_N$, where W_N is a penalty matrix for deviations of x_k from its target value \bar{x} . In turn, one can then calculate:

$$K_k^{xx} = aK_{k+1}^{xx}a - [aK_{k+1}^{xx}b] \mu_k [bK_{k+1}^{xx}a] + W_k. \quad (20)$$

If $A_k = 0$ and $W_k = 1, \forall k$, then this term is simply equal to one.

For p , I first pin down the terminal value of a component term, $\bar{p}_N^x = -W_N \bar{x}_N$, where \bar{x}_k is the desired target value for x in period k . The \bar{p} 's may then be found recursively according to

$$\bar{p}_k = -[aK_{k+1}^{xx}b] \mu_k [b\bar{p}_{k+1}] + a\bar{p}_{k+1} - W_k \bar{x}_k. \quad (21)$$

The p 's relate to the \bar{p} 's by

$$p_k = K_k^{xx} x_k + \bar{p}_k. \quad (22)$$

These give, along with the terminal condition $K_N^{\theta x} = 0$, the Riccati matrices:

$$\begin{aligned} K_k^{\theta x} = & \left[f_{\theta k}^{xT} K_{k+1}^{xx} + K_{k+1}^{\theta x} \right] a + p_{k+1}^x a_{\theta} \\ & - \left[\left[f_{\theta k}^{xT} K_{k+1}^{xx} + K_{k+1}^{\theta x} \right] b + p_{k+1}^x b_{\theta} \right] \mu_k [bK_{k+1}^{xx}a], \end{aligned} \quad (23)$$

with $a_{\theta} = [1 \ 0]^T$. The last set of Riccati matrices, $K_k^{\theta\theta}$, also have a terminal

value of zero. They are then found by recursions:

$$\begin{aligned}
 K_k^{\theta\theta} &= f_{\theta k}^x T [K_{k+1}^{xx} f_{\theta k}^x + K_{k+1}^{x\theta}] + [K_{k+1}^{\theta x} f_{\theta k}^x + K_{k+1}^{\theta\theta}] \\
 &\quad - [(f_{\theta k}^x T K_{k+1}^{xx} + K_{k+1}^{\theta x})b + p_{k+1}^x b_{\theta}] \mu_k \\
 &\quad \times [b(K_{k+1}^{xx} f_{\theta k}^x + K_{k+1}^{x\theta}) + p_{k+1}^x b_{\theta}]. \tag{24}
 \end{aligned}$$

These expressions are not particularly intuitive, except in some simple cases which I will look at below. The probing cost term, in the absence of measurement error, will only depend on K_k^{xx} and $K_k^{\theta x}$.

4. Comparative statics of a change in the control value

This section examines changes in the system in response to an increase in the control variable u , away from its certainty-equivalent value. The probing cost will then depend upon how the uncertainty in the system changes, $\partial \Sigma_{k|k} / \partial u_0$, and how the Riccati terms change. Random variables are equal to their period 0 values, unless otherwise noted. I will regard as the leading example that $\text{cov}(a, b)$ is negative and larger in absolute value than $\text{var}(b)$.

4.1. Effects on the system covariance

Given the Kalman filter update (17), the posterior precision will depend upon the update term $-\Sigma_{k+1|k}^{\theta x} (\Sigma_{k+1|k}^{xx})^{-1} \Sigma_{k+1|k}^{x\theta}$. Since Σ^{xx} is a scalar-valued expression, I will rewrite the update term as

$$- \begin{bmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \end{bmatrix} (\Sigma_{1|0}^{xx})^{-1}, \tag{25}$$

where the d_{ij} are the elements of the cross-product $\Sigma_{1|0}^{\theta x} \Sigma_{1|0}^{x\theta}$. Differentiating with respect to the control value, I get

$$\begin{aligned}
 \partial \Sigma_{1|1}^{\theta\theta} / \partial u_0 &= - \begin{bmatrix} \partial d_{11} / \partial u_0 & \partial d_{12} / \partial u_0 \\ \partial d_{21} / \partial u_0 & \partial d_{22} / \partial u_0 \end{bmatrix} (\Sigma_{1|0}^{xx})^{-1} \\
 &\quad + \begin{bmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \end{bmatrix} (\Sigma_{1|0}^{xx})^{-2} \partial (\Sigma_{1|0}^{xx}) / \partial u_0. \tag{26}
 \end{aligned}$$

Let's consider first the d_{ij} :

$$d_{11} = [x_0^2 \text{var}^2(a) + 2x_0u_0 \text{cov}(a, b)\text{var}(a) + u_0^2 \text{cov}^2(a, b)], \quad (27)$$

$$\partial d_{11}/\partial u_0 = [2x_0 \text{cov}(a, b)\text{var}(a) + 2u_0 \text{cov}^2(a, b)]. \quad (28)$$

A large variance of a combined with a negative covariance of a, b can, for example, make this negative. In this instance, the precision of a will fall with increases in the control value, contrary to the standard intuition.

To understand better why, consider the term $d_{11}/\Sigma_{1|0}^{xx}$. d_{11} is the square of the conditional covariance between a and x , and recall Σ^{xx} is the conditional variance of x . I can rewrite this as

$$d_{11}/\Sigma_{1|0}^{xx} = \rho_{1|0} \text{cov}(a, x)_{1|0}, \quad (29)$$

where $\text{cov}(a, x)_{1|0} = x_0 \text{var}(a) + u_0 \text{cov}(a, b)$, which is just the upper row of $\Sigma_{1|0}^{ax}$, and $\rho_{1|0} = \text{cov}(a, x)_{1|0}/\Sigma_{1|0}^{xx}$. One can interpret the term $\rho_{1|0}$ as the pseudo-correlation between the random variables $x_{1|0}$ and $a_{1|0}$. The derivative with respect to u_0 contains all the intuition,

$$\partial[\rho_{1|0} \text{cov}(a, x)_{1|0}]/\partial u_0 = \rho_{1|0} \text{cov}(a, b) + \text{cov}(a, x)_{1|0} \partial \rho_{1|0}/\partial u_0. \quad (30)$$

The first term indicates, given current information, what happens to the conditional covariance of a and x . For our leading example, $\rho_{1|0} > 0$ and $\text{cov}(a, b) < 0$, this produces additional uncertainty and lowers precision. The second term tells us whether the pseudo-correlation improves with increases in the control value, i.e., does it provide more information. Expanding it and simplifying a bit:

$$\begin{aligned} & \text{cov}(a, x)_{1|0} \partial \rho_{1|0}/\partial u_0 \\ &= \rho_{1|0} \left[\text{cov}(a, b) - \frac{\text{cov}(a, x)_{1|0}(2u_0 \text{var}(b) + 2x_0 \text{cov}(a, b))}{\Sigma_{1|0}^{xx}} \right]. \end{aligned} \quad (31)$$

When the expression in parentheses is negative, all the terms in (30) are negative. This indicates that realizations of x are less informative to the controller. The posterior variance of a will increase with increases in u_0 until

it drives $\text{cov}(a, x)_{1|0}$ negative. The intuition is identical for the other posterior variance. The argument is the same, but with the signs reversed for the posterior covariances.

For the off-diagonal terms:

$$d_{12} = d_{21} = \left[x_0^2 \text{cov}(a, b) \text{var}(a) + u_0^2 \text{cov}(a, b) \text{var}(b) + x_0 u_0 \text{cov}^2(a, b) + x_0 u_0 \text{var}(a) \text{var}(b) \right], \quad (32)$$

$$\partial d_{12} / \partial u_0 = \left[2u_0 \text{cov}(a, b) \text{var}(b) + x_0 \text{cov}^2(a, b) + x_0 \text{var}(a) \text{var}(b) \right]. \quad (33)$$

This term can again be positive, meaning that the covariance will be larger in absolute value, if the variance of a is very large.

And finally:

$$d_{22} = \left[x_0^2 \text{cov}^2(a, b) + 2x_0 u_0 \text{cov}(a, b) \text{var}(b) + u_0^2 \text{var}^2(b) \right], \quad (34)$$

$$\partial d_{22} / \partial u_0 = \left[2x_0 \text{cov}(a, b) \text{var}(b) + 2u_0 \text{var}^2(b) \right]. \quad (35)$$

The variance of a does not enter this term, but if the $\text{cov}(a, b)$ is large, in absolute value, relative to $\text{var}(b)$, again this term can decrease with u_0 .

The other effect of a change in the control value at period zero is to change the one-period-ahead conditional variance of x ,

$$\partial \Sigma_{1|0}^{xx} / \partial u_0 = 2u_0 \text{var}(b) + 2x_0 \text{cov}(a, b). \quad (36)$$

The sign of this term depends upon the parameters. The uncertainty over x though can be reduced for some u_0 if the $\text{cov}(a, b)$ is negative. For large enough u_0 , the sign will always be positive if $\text{var}(b) > 0$.

In the case of a single unknown parameter b , (26) reduces to

$$- \left[\frac{2u_0 \text{var}^2(b)Q}{(u_0^2 \text{var}(b) + Q)^2} \right], \quad (37)$$

which is clearly negative, confirming the single-parameter intuition.

As for the second derivative, $\partial^2 \Sigma^{\theta\theta} / \partial u_0^2$, the diagonal entries will be negative for large enough u . Eventually then, the variances of a and b will

decrease with the control value at a decreasing rate. The off-diagonal terms are positive. The covariance will decrease to zero, also at a decreasing rate.

4.2. Effect on the Riccati terms

Some of the Riccati terms will be independent of the control value. K_k^{xx} depends only upon the parameters of the state and the penalty matrix W . μ depends only upon K_{k+1}^{xx} and b . \tilde{p}_k^x depends only upon the state parameters and K_k^{xx} . As these don't vary with u , they are not impacted in any way.

All of the other terms will be affected, principally through their effect on the change in the certainty equivalent values of x and u . If, for example, b is negative, an increase in u_0 will lower $x_{1|0}$, and this in turn requires an adjustment of u_1 . These effects are all registered in the Riccati matrix term $K_k^{\theta x}$. To save on notation, as in section 3.2, random parameters are equal to their period 0 certainty equivalents, $b = b_{j|0}$.

Next period's expected value of x is the first to change. It moves by $b\Delta u_0$. I then re-evaluate the first-period certainty-equivalent control:

$$u_{1|0} = -\mu_1 [bK_2^{xx}ax_{1|0} + \tilde{p}_2^x], \quad (38)$$

with first derivative

$$\partial u_{1|0}/\partial u_0 = -\mu_1 ab^2 K_2^{xx}. \quad (39)$$

This then enables one to find the second-period expected value of x :

$$\partial x_{2|0}/\partial u_0 = ab - \mu_1 ab^3 K_2^{xx}, \quad (40)$$

and so on:

$$\partial u_{k|0}/\partial u_0 = -\mu_k ab K_{k+1}^{xx} \partial x_{k|0}/\partial u_0, \quad (41)$$

$$\partial x_{k|0}/\partial u_0 = a \partial x_{k-1}/\partial u_0 + b \partial u_{k-1}/\partial u_0. \quad (42)$$

For large enough b , these changes can come to dominate the probing cost.

These new certainty-equivalent values also provide the updated terms p_k^x . p_1^x moves by $K_1^{xx}b\Delta u_0$, p_2^x moves by $K_2^{xx}(a\Delta x_{1|0} + b\Delta u_{1|0})$, etc. With these in hand, I can then evaluate $K_{N-1}^{\theta x}$ and begin that set of Riccati recursions. f_{θ}^x will have the new certainty-equivalent values for x and u , and the p^x terms will reflect the new x values as well.

Consider first the penultimate Riccati term, with $K_N^{xx} = 1$, $K_N^{\theta x} = 0$, and $p_N^x = x_{N|0}$:

$$K_{N-1}^{\theta x} = \begin{bmatrix} ax_{N-1|0} \\ au_{N-1|0} \end{bmatrix} - \begin{bmatrix} bx_{N-1|0} \\ bu_{N-1|0} + x_{N|0} \end{bmatrix} \times \mu_{N-1} [bK_N^{xx}a] + \begin{bmatrix} x_{N|0} \\ 0 \end{bmatrix}. \quad (43)$$

Using (41) and (42), we can evaluate the partials with respect to x and u :

$$\begin{aligned} \partial K_{N-1}^{\theta x} / \partial u_0 &= \begin{bmatrix} (a - \mu_{N-1} ab^2 K_N^{xx}) \partial x_{N-1|0} / \partial u_0 + \partial x_{N|0} / \partial u_0 \\ (a - \mu_{N-1} ab^2 K_N^{xx}) \partial u_{N-1|0} / \partial u_0 - \mu_{N-1} ab K_N^{xx} \partial x_{N|0} / \partial u_0 \end{bmatrix}. \end{aligned} \quad (44)$$

The signs of these terms obviously depend upon the parameters, but since they enter the probing cost as a cross-product, their magnitude is probably more important. Both a and b enter exponentially, meaning that a very small change in u_0 can lead to a great deal of fluctuation in the future certainty-equivalent path.

Working backwards:

$$\begin{aligned} \partial K_{N-2}^{\theta x} / \partial u_0 &= \begin{bmatrix} (a - \mu_{N-2} ab^2 K_{N-1}^{2xx}) \partial x_{N-2|0} / \partial u_0 + K_{N-1}^{xx} \partial x_{N-1|0} / \partial u_0 \\ (a - \mu_{N-2} ab^2 K_{N-1}^{2xx}) \partial u_{N-2|0} / \partial u_0 - \mu_{N-2} ab K_{N-1}^{2xx} \partial x_{N-1|0} / \partial u_0 \end{bmatrix} \\ &\quad + (a - ab^2 \mu_{N-2} K_{N-1}^{xx}) \partial K_{N-1}^{\theta x} / \partial u_0. \end{aligned} \quad (45)$$

There exists a possibility that the recursive Riccati term could dominate, working its way back to all the earlier terms.

The overall change in \mathcal{R}_k defies easy explanation in all but some simplifying cases. In the example below, I look at a two-period problem, which drops out all the terms involving $K^{\theta x}$.

5. Numerical examples and simulation

This section takes up the MacRae (1972) problem. As this has been widely studied in the literature, it will be an excellent vehicle to show how nonconvexities arise in the cost-to-go. Nothing pathological is required. Going from one unknown to two unknowns or choosing a large b will prove sufficient.

Consider first the penultimate Riccati term, with $K_N^{xx} = 1$, $K_N^{\theta x} = 0$, and $p_N^x = x_{N|0}$:

$$K_{N-1}^{\theta x} = \begin{bmatrix} ax_{N-1|0} \\ au_{N-1|0} \end{bmatrix} - \begin{bmatrix} bx_{N-1|0} \\ bu_{N-1|0} + x_{N|0} \end{bmatrix} \times \mu_{N-1} [bK_N^{xx}a] + \begin{bmatrix} x_{N|0} \\ 0 \end{bmatrix}. \quad (43)$$

Using (41) and (42), we can evaluate the partials with respect to x and u :

$$\begin{aligned} \partial K_{N-1}^{\theta x} / \partial u_0 &= \begin{bmatrix} (a - \mu_{N-1} ab^2 K_N^{xx}) \partial x_{N-1|0} / \partial u_0 + \partial x_{N|0} / \partial u_0 \\ (a - \mu_{N-1} ab^2 K_N^{xx}) \partial u_{N-1|0} / \partial u_0 - \mu_{N-1} ab K_N^{xx} \partial x_{N|0} / \partial u_0 \end{bmatrix}. \end{aligned} \quad (44)$$

The signs of these terms obviously depend upon the parameters, but since they enter the probing cost as a cross-product, their magnitude is probably more important. Both a and b enter exponentially, meaning that a very small change in u_0 can lead to a great deal of fluctuation in the future certainty-equivalent path.

Working backwards:

$$\begin{aligned} \partial K_{N-2}^{\theta x} / \partial u_0 &= \begin{bmatrix} (a - \mu_{N-2} ab^2 K_{N-1}^{2xx}) \partial x_{N-2|0} / \partial u_0 + K_{N-1}^{xx} \partial x_{N-1|0} / \partial u_0 \\ (a - \mu_{N-2} ab^2 K_{N-1}^{2xx}) \partial u_{N-2|0} / \partial u_0 - \mu_{N-2} ab K_{N-1}^{2xx} \partial x_{N-1|0} / \partial u_0 \end{bmatrix} \\ &\quad + (a - ab^2 \mu_{N-2} K_{N-1}^{xx}) \partial K_{N-1}^{\theta x} / \partial u_0. \end{aligned} \quad (45)$$

There exists a possibility that the recursive Riccati term could dominate, working its way back to all the earlier terms.

The overall change in \mathcal{R}_k defies easy explanation in all but some simplifying cases. In the example below, I look at a two-period problem, which drops out all the terms involving $K^{\theta x}$.

5. Numerical examples and simulation

This section takes up the MacRae (1972) problem. As this has been widely studied in the literature, it will be an excellent vehicle to show how nonconvexities arise in the cost-to-go. Nothing pathological is required. Going from one unknown to two unknowns or choosing a large b will prove sufficient.

be repeated one more time: $u_1 = -(0.8)[(-0.5)(1.0)(0.7)(3.272) + (-0.5)(1.0)(3.5) + (-0.5)(0)] = 2.316$. $x_{2|0} = (0.7)(3.272) + (-0.5)(2.316) + 3.5 = 4.632$. $\Sigma_{1|0}^{xx} = (2)^2(5)^2 + (2)(2)(3.257)(-2) + (3.257)^2(1)^2 + 0.2 = 4.752$. Note the negative covariance term. Increases in u_0 will, over some range, actually lower this term,

$$\Sigma_{1|0}^{\theta x} = \begin{bmatrix} 5 & -2 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 3.257 \end{bmatrix} = \begin{bmatrix} 3.486 \\ -0.743 \end{bmatrix}.$$

This is going to enter the $\Sigma^{\theta\theta}$ calculation as a cross-product.

Now, I can update the parameter covariance matrix:

$$\begin{aligned} \Sigma_{1|1}^{\theta\theta} &= \begin{bmatrix} 5 & -2 \\ -2 & 1 \end{bmatrix} - \begin{bmatrix} 3.486 \\ -0.743 \end{bmatrix} (4.752)^{-1} [3.486 \quad -0.743] \\ &= \begin{bmatrix} 2.442 & -1.445 \\ -1.445 & 0.884 \end{bmatrix}. \end{aligned} \quad (50)$$

The choice of u has indeed reduced parametric uncertainty, but not nearly as dramatically in percentage terms as Kendrick did with a single unknown.

Now, turn to the Riccati matrices. The two-period problem obviously simplifies things a great deal.

$$K_2^{\theta x} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

and

$$\begin{aligned} K_1^{\theta x} &= \begin{bmatrix} 3.272 \\ 2.316 \end{bmatrix} (0.7) + \begin{bmatrix} 0 \\ 4.632 \end{bmatrix} - \left(\begin{bmatrix} 3.272 \\ 2.316 \end{bmatrix} (-0.5) + \begin{bmatrix} 0 \\ 4.632 \end{bmatrix} \right) \\ &\quad \times [(0.8)(-0.5)(1.0)(0.7)] \\ &= \begin{bmatrix} 6.464 \\ 2.594 \end{bmatrix}, \end{aligned} \quad (51)$$

where $\tilde{p}_2^x = (1)(4.632) + 0.0 = 4.632$. What really is of interest is the third term:

$$\begin{aligned} \mathcal{R}_1 &= (\cdot)(\cdot)^T \mu_1 = \begin{bmatrix} -1.636 \\ 3.474 \end{bmatrix} [-1.636 \quad 3.474] (0.8) \\ &= \begin{bmatrix} 2.141 & -4.546 \\ -4.546 & 9.655 \end{bmatrix}. \end{aligned}$$

This term will not figure very prominently until the next example.

Table 1
Example 1 – Precision decreases with control value.

State equation	Variance of state disturbance
$x_{k+1} = 0.70 * x_k + -0.50 * u_k + 3.50$	$Q = 0.200$
Target values for state and control	Penalty parameters
$\bar{x} = 0.000, \bar{u} = 0.000$	$\Lambda = 1.0, W = 1.0$
State and control variables	
$x_{k 0} \quad u_{k 0}$	
$k = 0 \quad 2.000 \quad 3.757$	
$k = 1 \quad 3.022 \quad 2.246$	
$k = 2 \quad 4.492 \quad 0.000$	
Conditional covariances	
Σ_{jk}^{xx}	
$j = 1, k = 0 \quad 4.259$	
$j = 2, k = 1 \quad 12.354$	$\Sigma_{1 0}^{\theta x} = \begin{bmatrix} 2.486 \\ -0.243 \end{bmatrix}$
Parameter covariance matrix $\Sigma_k^{\theta\theta}$	
$k = 0$	$k = 1$
$\begin{bmatrix} 5.00 & -2.00 \\ -2.00 & 1.00 \end{bmatrix}$	$\begin{bmatrix} 3.55 & -1.86 \\ -1.86 & 0.986 \end{bmatrix}$
Riccati matrix K_k^{xx}	
$k = 1 \quad 1.392$	
$k = 2 \quad 1.000$	
Riccati matrix $K_k^{\theta x}$	
$k = 1$	$k = 2$
$\begin{bmatrix} 6.184 \\ 2.156 \end{bmatrix}$	$\begin{bmatrix} 0.000 \\ 0.000 \end{bmatrix}$
Riccati matrix \mathcal{R}	Probing cost
$\begin{bmatrix} 1.826 & -4.072 \\ -4.072 & 9.080 \end{bmatrix}$	15.2832050

Now compute the probing cost:

$$\begin{aligned}
 Jp_1 &= 0.5 \operatorname{tr} \left(\begin{bmatrix} 2.141 & -4.546 \\ -4.546 & 9.655 \end{bmatrix} \begin{bmatrix} 2.443 & -1.455 \\ -1.455 & 0.883 \end{bmatrix} \right) \\
 &= 0.5 \operatorname{tr} \left(\begin{bmatrix} 11.844 & \dots \\ \dots & 15.139 \end{bmatrix} \right) = 13.495. \quad (52)
 \end{aligned}$$

Clearly, having two parameters and a larger *a priori* covariance matrix makes the probing cost a more important component of the total cost-to-go.

Table 1 reports values for all of the variables above after increasing the control value u_0 by 0.5. While $(\Sigma_{1|0}^{xx})^{-1}$ has risen, the precision of the posterior covariance matrix has actually fallen. All other things equal, this will produce nonconvexities in the cost-to-go.

Figs. 1 to 3 plot the elements of the posterior covariance matrix as u_0 is increased from its certainty equivalent of 3.257 by steps of 0.05, up to 6.257. From (28), $\partial d_{11}/\partial u_0 = 0$ at $u_0 = 5$, $\partial d_{12}/\partial u_0 = 0$ at $u_0 = 4.5$, and $\partial d_{22}/\partial u_0 = 0$ at $u_0 = 5$, for these parameters. Σ^{xx} reaches its minima at $u_0 = 4$. The d_{ij} are the dominant factors though as can be seen in the figures. The minima and the maxima for the parameter covariance elements all occur at the zeroes of the d_{ij} .

The \mathcal{R}_k are considerably easier to analyze in the two-period problem. Eq. (19) reduces to

$$\begin{aligned} \mathcal{R}_1 &= [f_{\theta_1}^{xxT} K_2^{xx} b + p_2^x b_{\theta}] \mu_1 \times [b K_2^{xx} f_{\theta_1}^x + (p_2^x b_{\theta})^T] \\ &= \begin{bmatrix} K_2^{xx} b x_{1|0} \\ K_2^{xx} b u_{1|0} + x_{2|0} \end{bmatrix} \mu_1 \times [b K_2^{xx} x_{1|0} b K_2^{xx} u_{1|0} + x_{2|0}] \\ &= \begin{bmatrix} -1.636 \\ 3.474 \end{bmatrix} (0.8) \times [(-0.5)(1)[3.272 \quad 2.316] + [0 \quad 4.632]] \\ &= \begin{bmatrix} 2.141 & -4.546 \\ -4.546 & 9.655 \end{bmatrix}. \end{aligned} \quad (53)$$

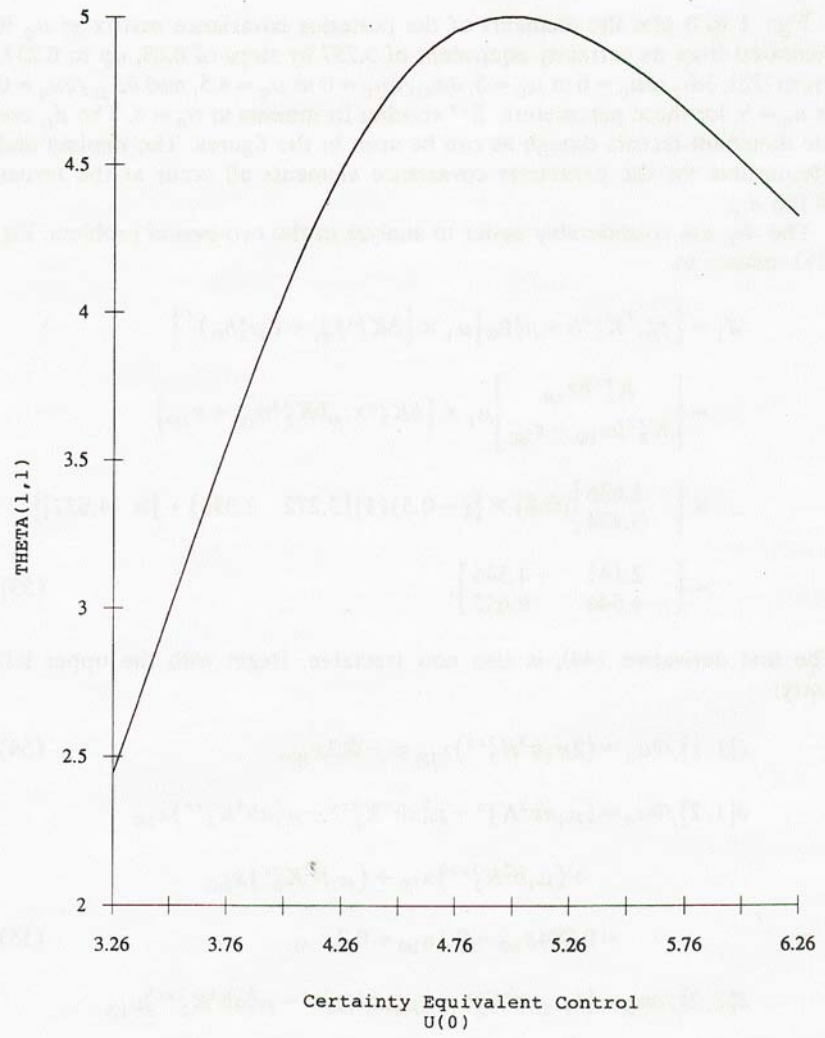
The first derivative, (44), is also now tractable. Begin with the upper left entry:

$$\partial[1, 1]/\partial u_0 = (2\mu_1 b^3 K_2^{2xx}) x_{1|0} = -0.2 x_{1|0}, \quad (54)$$

$$\begin{aligned} \partial[1, 2]/\partial u_0 &= (\mu_1 a b^2 K_2^{xx} - \mu_1^2 a b^4 K_2^{2xx} - \mu_1^2 a b^4 K_2^{3xx}) x_{1|0} \\ &\quad + (\mu_1 b^3 K_2^{2xx}) u_{1|0} + (\mu_1 b^2 K_2^{xx}) x_{2|0} \\ &= 0.084 x_{1|0} - 0.1 u_{1|0} + 0.2 x_{2|0}, \end{aligned} \quad (55)$$

$$\begin{aligned} \partial[2, 2]/\partial u_0 &= (2\mu_1 a b^2 K_2^{xx} - 2\mu_1^2 a b^4 K_2^{3xx} - \mu_1^2 a b^4 K_2^{2xx}) u_{1|0} \\ &\quad + (2ab\mu_1 - 2\mu_1^2 a b^3 K_2^{xx}) x_{2|0} \\ &= 0.168 u_{1|0} - 0.448 x_{2|0}. \end{aligned} \quad (56)$$

Given the parameters in this example, these terms decline, in absolute value, monotonically. As in Kendrick's problem, they are contributing to a reduction in the probing cost. But as fig. 4 testifies, the overall probing cost rises for sixteen steps until the posterior variance of b begins to fall.

Fig. 1. Example 1 - Posterior var(a).

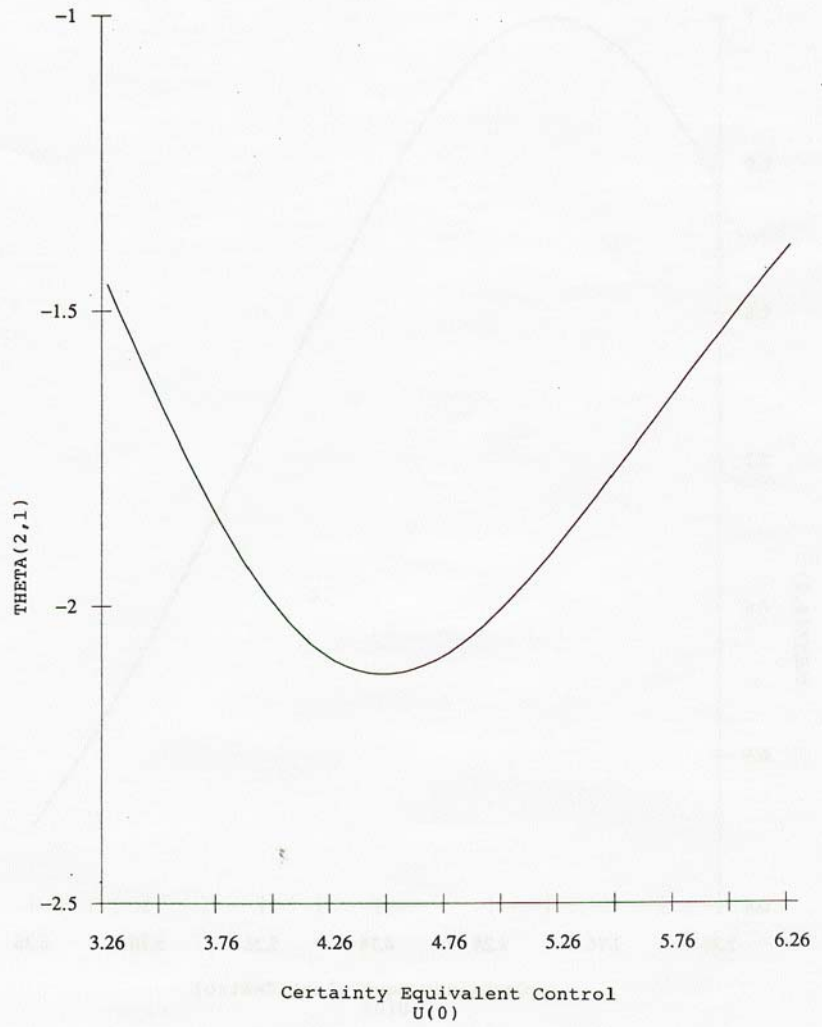
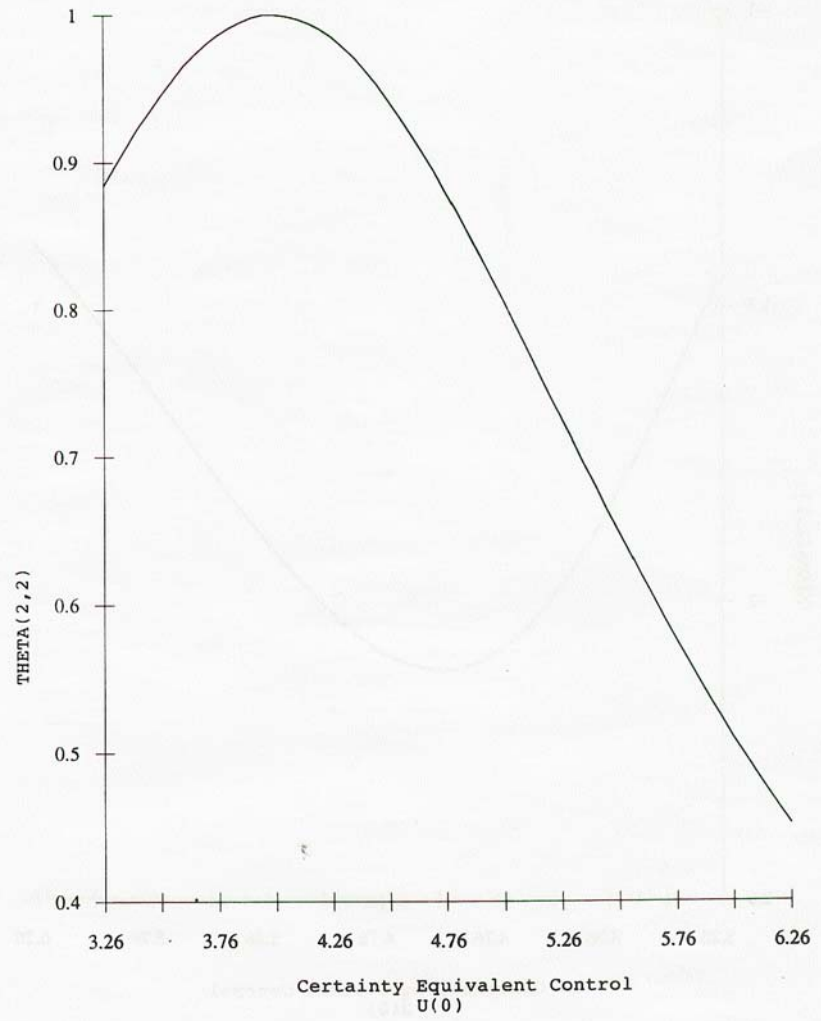


Fig. 2. Example 1 - Posterior $\text{cov}(a, b)$.

Fig. 3. Example 1 - Posterior var(b).

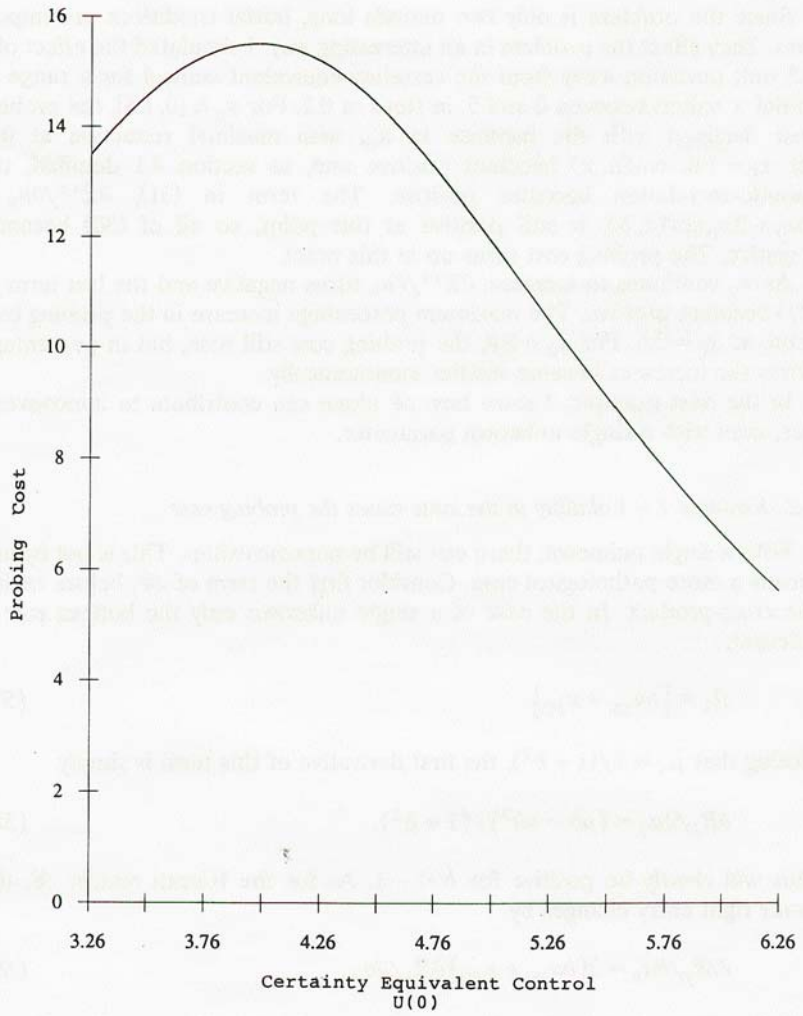


Fig. 4. Example 1 - Probing cost.

Since the problem is only two periods long, initial conditions are important. They affect the problem in an interesting way. I simulated the effect of a 0.5 unit deviation away from the certainty-equivalent control for a range of initial x values between 0 and 5, in steps of 0.1. For $x_0 \in [0, 1.5]$, the probing cost declined with the increase in u_0 , with maximal reduction at 0.8. At $x_0 = 1.6$, $\text{cov}(a, x)$ becomes positive and, as section 4.1 detailed, the pseudo-correlation becomes positive. The term in (31), $\partial \Sigma^{xx} / \partial u_0 = 2u_0 + 2x_0 \text{cov}(a, b)$, is still positive at this point, so all of (30) becomes negative. The probing cost turns up at this point.

As x_0 continues to increase, $\partial \Sigma^{xx} / \partial u_0$ turns negative and the last term in (31) becomes positive. The maximum percentage increase in the probing cost came at $x_0 = 2.6$. For $x_0 > 2.6$, the probing cost still rose, but in percentage terms the increases became smaller monotonically.

In the next example, I show how \mathcal{R} alone can contribute to nonconvexities, even with a single unknown parameter.

5.2. Example 2 – Volatility in the state raises the probing cost

With a single unknown, there can still be nonconvexities. This is not by any means a more pathological case. Consider first the term of \mathcal{R} , before taking the cross-product. In the case of a single unknown only the bottom row is relevant:

$$R_1 \equiv [bu_{10} + x_{20}]. \quad (57)$$

Noting that $\mu_1 = 1/(1 + b^2)$, the first derivative of this term is simply

$$\partial R_1 / \partial u_0 = (ab - ab^3) / (1 + b^2). \quad (58)$$

This will clearly be positive for $b < -1$. As for the Riccati matrix \mathcal{R} , the lower right entry changes by

$$\partial \mathcal{R}_1 / \partial u_0 = 2(bu_{10} + x_{20}) \partial R_1 / \partial u_0. \quad (59)$$

If b is sufficiently negative, the term in parentheses will go positive. To see this, express the term as

$$bu_{10} + x_{20} = 2b \frac{-ab^2x_0 - ab^2u_0 - abc - bc}{a + b^2} + a^2x_0 + abu_0 + 2c. \quad (60)$$

Once this turns positive, the probing cost will start to increase for given $\Sigma^{\theta\theta}$.

Table 2
Example 2 – Volatility in the state increases the probing cost.

State equation $x_{k+1} = 0.70x_k + -3.50u_k + 3.50$	Variance of state disturbance $Q = 0.200$
Target values for state and control $\bar{x} = 0.000, \bar{u} = 0.000$	Penalty parameters $\Lambda = 1.0, W = 1.0$
State and control variables	
$x_{k 0}$ $u_{k 0}$	
$k = 0$ 2.000 1.345	
$k = 1$ 0.192 0.960	
$k = 2$ 0.274 0.000	
Conditional covariances Σ_{jk}^{xx}	
$j = 1, k = 0$ 4.723	$\Sigma_{1 0}^{\theta x} = \begin{bmatrix} 0.000 \\ 3.363 \end{bmatrix}$
$j = 2, k = 1$ 0.298	
Parameter covariance matrix $\Sigma_k^{\theta\theta}$	
$k = 0$	$k = 1$
$\begin{bmatrix} 0.00 & 0.00 \\ 0.00 & 2.50 \end{bmatrix}$	$\begin{bmatrix} 0.00 & 0.000 \\ 0.00 & 0.106 \end{bmatrix}$
Riccati matrix K_k^{xx}	
$k = 1$ 1.037	
$k = 2$ 1.000	
Riccati matrix $K_k^{\theta x}$	
$k = 1$	$k = 2$
$\begin{bmatrix} 0.284 \\ 0.101 \end{bmatrix}$	$\begin{bmatrix} 0.000 \\ 0.000 \end{bmatrix}$
Riccati matrix \mathcal{R}	
$\begin{bmatrix} 0.034 & 0.157 \\ 0.157 & 0.719 \end{bmatrix}$	Probing cost 0.0380448

Let's choose parameters similar to the first example. Since a is no longer an unknown, $cov(a, b) = var(a) = 0$. The value of b is decreased to -3.5 and its variance is raised to 2.5 . Everything else is unchanged. Table 2 reports all the summary statistics.

At u_0 's certainty-equivalent value of 1.345 , (60) is negative. The expression only turns positive at $u_0 = 2.845$. Surprisingly, as fig. 5 shows, $var(b)_{1|1}$ drops quite dramatically. At the certainty-equivalent value, it equals 0.1059 . At the point where R_1 starts to increase, it has fallen to 0.0245 . The change in R_1 is large enough to overcome the drop in the posterior variance of b , since its proportionate reduction is quite small at that time.

The nonconvexity this gives rise to in the probing cost can be seen in fig. 6. If for a given level of uncertainty, the path of state variables is more volatile as you increase the control value, the probing cost will have nonconvexities.

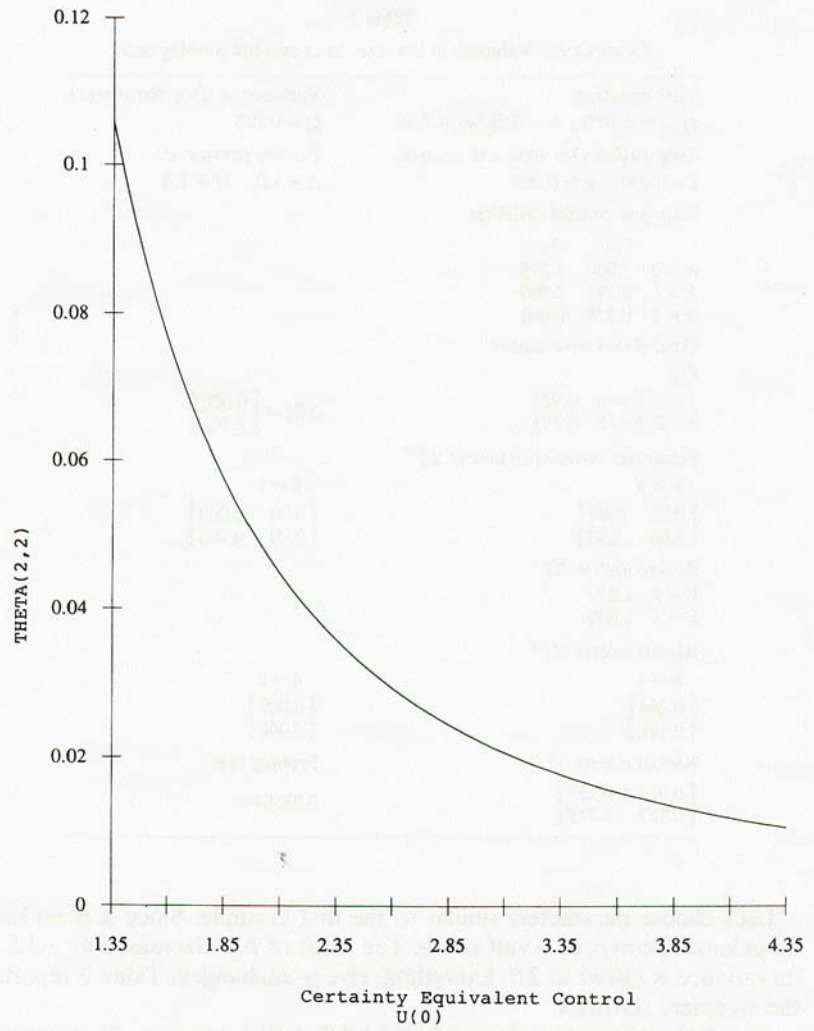


Fig. 5. Example 2 - Posterior var(b).

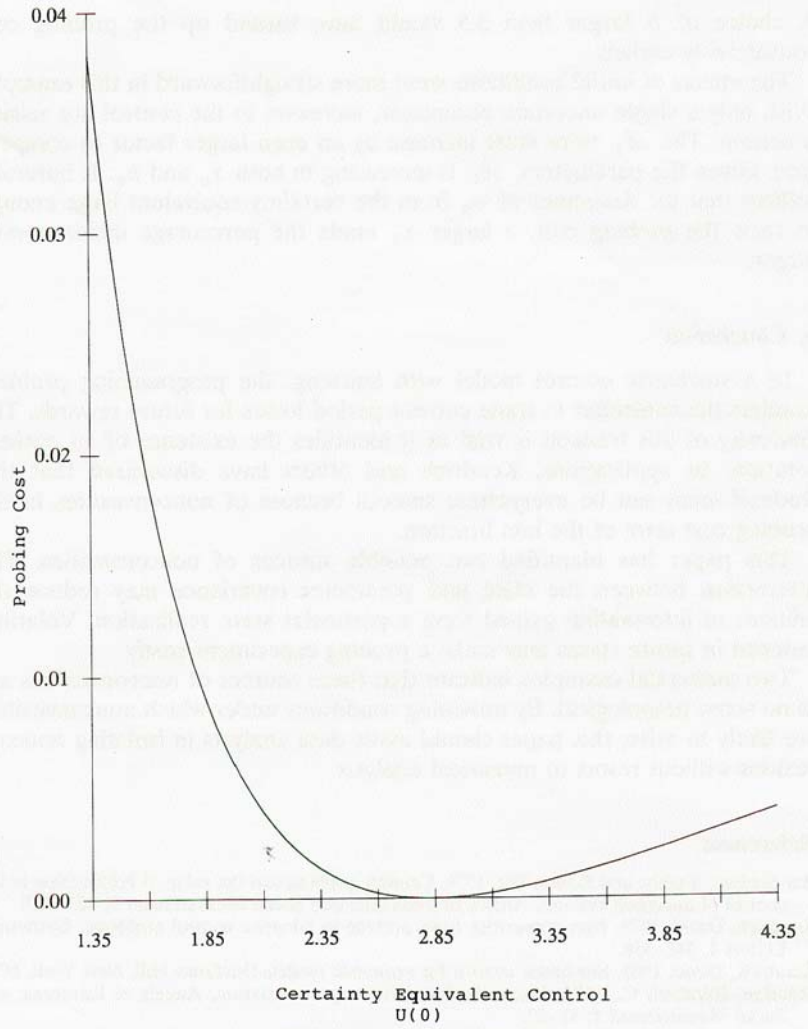


Fig. 6. Example 2 - Probing cost.

A choice of b larger than 3.5 would have turned up the probing cost considerably earlier.

The effects of initial conditions were more straightforward in this example. With only a single uncertain parameter, increases in the control are raising precision. The \mathcal{R}_1 term must increase by an even larger factor to compensate. Given the parameters, \mathcal{R}_1 is increasing in both x_0 and u_0 . It naturally follows that for deviations of u_0 from the certainty-equivalent large enough to raise the probing cost, a larger x_0 made the percentage increase even larger.

6. Conclusion

In a stochastic control model with learning, the programming problem requires the controller to trade current-period losses for future rewards. The convexity of this tradeoff is vital as it identifies the existence of an optimal solution. In applications, Kendrick and others have discovered that this tradeoff may not be everywhere smooth because of nonconvexities in the probing cost term of the loss function.

This paper has identified two possible sources of nonconvexities. The interaction between the state and parameter covariance may reduce the amount of information gained from a particular state realization. Volatility induced in future states may make a probing experiment costly.

Two numerical examples indicate that these sources of nonconvexities are in no sense pathological. By providing conditions under which nonconvexities are likely to arise, this paper should assist data analysts in isolating nonconvexities without resort to numerical analysis.

References

- Bar-Shalom, Yaakov and Edison Tse, 1976, Caution, probing and the value of information in the control of uncertain systems, *Annals of Economic and Social Measurement* 5, 323-337.
- Kendrick, David, 1978, Non-convexities from probing in adaptive control problems, *Economics Letters* 1, 347-351.
- Kendrick, David, 1981, *Stochastic control for economic models* (McGraw Hill, New York, NY).
- MacRae, Elizabeth C., 1972, Linear decision with experimentation, *Annals of Economic and Social Measurement* 1, 43-47.
- Norman, Alfred, 1976, First order dual control, *Annals of Economic and Social Measurement* 6, 275-300.