AN APPROXIMATED SOLUTION TO CONTINUOUS-TIME STOCHASTIC OPTIMAL CONTROL PROBLEMS THROUGH MARKOV DECISION CHAINS

JACEK B. KRAWCZYK AND ALISTAIR WINDSOR

Abstract. Strategies for constructing a Markov decision chain approximating a continuous-time finite-horizon optimal control problem are investigated. Some simple, analytically soluble, examples are treated and low computational complexity is reported. Extensions to the method and implementation are discussed. In particular, relevance of the approximated solution to a stochastic renewable resource valuation problem is examined.

Working Paper

Version 9b

JEL Classification: C8 (Computational Economics) C63 (Computational Techniques), D92 (Intertemporal Firm Choice and Growth), C87 (Economic Software), Q25 (Renewable Resources and Conservation).

AMS Categories: 93E20 (Optimal stochastic control), 93E25 (Computational methods in stochastic optimal control); 90C39 (Dynamic programming), 90C40 (Markov decision processes); 90A30 (Environmental economics: natural resource models, harvesting, pollution, etc.)

Authors' Keywords: Computational economics; natural resource valuation. Approximating Markov decision chains; simple noise discretisation.

Research supported by VUW GSBGM and IGC.

Correspondence should be addressed to: Jacek B. Krawczyk.

About the authors: Jacek B. Krawczyk. Commerce & Administration Faculty, Victoria University of Wellington, PO Box 600, Wellington, New Zealand; fax +64-4-4712200. Email: `Jacek.Krawczykvuw.ac.nz`; http://www.vuw.ac.nz/~jacek

Alistair Windsor. Honours Student, Mathematics, Victoria University of Wellington.
Email: `Alistair.Windsor@vuw.ac.nz`

An Approximated Solution to Continuous-Time Stochastic Optimal Control Problems Through Markov Decision Chains

INTRODUCTION

The objective of this paper is to examine an alternative method of arriving at numerical solutions to continuous-time finite-horizon stochastic optimal control problems. The method and the corresponding optimisation software we designed appear *more* user-friendly than the existing ones, and suitable for solution of stochastic resource valuation problems.

Dynamic programming is one of the most popular techniques for solving optimal control problems. In the case of continuous optimal control problems the dynamic programming technique "reduces" the optimal control problem to one of solving a partial differential equation, the Hamilton-Jacobi-Bellman (HJB) equation. In all but a few special cases this partial differential equation cannot be solved analytically. Thus at this point the most common approach is to discretise *à la* Kushner [4], [5] in time and space and seek a numerical solution. This procedure is quite popular in optimal control, and successful even in case of stochastic differential games [2]. However, the discretisation of the HJB, which is a second order "functional" partial differential equation, is a complex operation in its own right. The iterative search process for a fixed point required by that method makes it relatively difficult to apply.

Since the problem must be discretised eventually anyway[1] we investigate the consequence of an "early" discretisation (rather than "late" of the HJB as in the Kushner method). We discretise the original continuous optimal control problem and apply the dynamic programming technique to the Markov decision chain thus obtained. Here we propose a new method of approximating the continuous noise by a discretely valued noise. Solving a Markov decision chain, discrete in time and space, is a more elementary exercise than solving the Hamilton-Jacobi-Bellman partial differential equation. Via this procedure we can establish both a resource *value* (presumably easier than via solving the HJB equation) and a discrete-time optimal *policy* which appears as credible as its continuous-time counterpart. In other words, in this paper we solve a Markov decision chain approximating the original problem, rather than approximate the solution to the exact problem, as other approaches do.

The examples, which we solve, document low computational complexity of our method. Broadly speaking, the computation time doubles if the state-space, or time, discretisation step halves. The overall performance of the method could be improved by an "intelligent" search of the state space, see *e.g.* [7], [10]. However, we dedicate ourselves in this paper to examine how the process and noise approximations work with standard backward optimisation.

---

[1] We direct our attention to optimal control problems where the analytical solution is unavailable. However, for comparison purposes, the numerical examples treated later in the paper are problems for which analytical solutions exist.

Notice that our solution based on the discrete-time approximation of the continuous-time problem may come closer to "reality" than the continuous solution, in that the parameters of the discrete-time model are usually directly obtainable from statistical data. The continuous time model parameter "rates" are probably results of processing those data.

We will apply our method to estimate the discounted profit of resource utilisation which is an important problem in deregulated economies. Using our approach, we achieve a refinement of a particular model of forestry valuation presented in [6]. The refinement occurs mainly in the problem formulation where we make explicit a proper stochastic optimal control problem that is equivalent to the problem considered by those authors. We also show how the value realisations are distributed and plot the policy rules. These are not shown in [6].

The rest of the paper is organised as follows. The first part discusses how to formulate a Markov decision chain which approximates a given continuous-time stochastic optimal control problem. The method is illustrated by some simple examples and then some extensive numerical experiments are conducted on analytically soluble deterministic models in one and two dimensions. The application of the method to stochastic problems is illustrated by numerical experiments on a stochastic linear-quadratic problem in one dimension.

The second part is an application of the method to a model of resource valuation. The model is related to the model suggested in [6]. We analyse both a deterministic and stochastic system and compare the results with those derived in [6]. Matlab discretisation and optimisation routines appear as a separate document [11].

Part 1. Markov Decision Chain Approximations

1. THE MARKOVIAN APPROXIMATION

1.1. **Optimal Stochastic Control.** Consider the system with a state equation

$$d\mathbf{X}(t) = f\big(\mathbf{X}(t), \mathbf{u}(t), t\big)dt + b\big(\mathbf{X}(t), \mathbf{u}(t), t\big)d\mathbf{W}(t) \tag{1}$$

where

$$\mathbf{X} = \{\mathbf{X}(t) \in X \subset \mathbb{R}^n,\ t \geq 0,\ \mathbf{X}(0) = \mathbf{x}_0 - \text{given}\}$$

is the state process, $\mathbf{u}(t) \in U \subset \mathbb{R}^m$ is the control, $\mathbf{W}(t)$ is a Wiener process (whose "derivative" is white noise), $f\big(\mathbf{X}(t), \mathbf{u}(t), t\big)$ is a "slowly" varying continuous function (called the drift), and $b\big(\mathbf{X}(t), \mathbf{u}(t), t\big)d\mathbf{W}(t)$ is a "rapidly" varying continuous random component (called the diffusion). For the formal treatment of the optimally controlled diffusion process refer to [1].

We consider the control is *optimal* if it is applied via a Markovian feedback control rule μ

$$\mathbf{u}(t) = \mu(t, \mathbf{X}(t)) \tag{2}$$

and chosen so as to maximise a performance criterion[2]

$$\max_{\mathbf{u}} J(0, x_0; \mathbf{u}) \tag{3}$$

where

$$J(\tau, x; \mathbf{u}) = \mathbb{E}\left(\int_\tau^T g\big(\mathbf{X}(t), \mathbf{u}(t), t\big)\mathrm{d}t + h\big(\mathbf{X}(T)\big)\,\Big|\,\mathbf{X}(\tau) = x\right). \tag{4}$$

is the return-to-go function. For Markovian feedback controls the maximum value of (4)

$$H(\tau, x) = \max_{\mu(\cdot)} J(\tau, x; \mu(\cdot))$$

satisfies the Hamilton-Jacobi-Bellman equation

$$\max\,\{g(x, \mathbf{u}, t) + \mathcal{L}_u H(t, x)\} = 0 \tag{5}$$

with the boundary condition

$$H(T, \mathbf{X}) = h(\mathbf{X}_T) \tag{6}$$

where $\mathcal{L}_u$ is the operator

$$\mathcal{L}_u = \frac{\partial}{\partial\tau} + \sum_{i=1}^n f_i(\mathbf{x}, \mathbf{u}, \tau)\frac{\partial}{\partial x_i} + \frac{1}{2}\sum_{i,j=1}^n B_{i,j}(\mathbf{x}, \mathbf{u}, \tau)\frac{\partial^2}{\partial x_i \partial x_j} \tag{7}$$

and where f_i is the i-th component function of f and $B_{i,j}$ is the i,j-th entry of the covariance matrix $B = bb^T$.

The stochastic optimal control problem is formulated with a fixed finite horizon and free final state. This definition is motivated by our ultimate goal for this research of solving stochastic resource valuation problems. The class of problems with finite, fixed horizon and free terminal

---

[2] We are "maximisers" and associate the performance criterion with the return function.

state includes the majority of the typical resource valuation problems, including the leasehold problem examined in Part 2. The method should handle problems with a fixed finite horizon and fixed final state, however, the approach taken is not directly applicable to infinite horizon problems.

1.2. **An Equivalent Markov Decision Chain.** We discretise the state process equation (1) in time using the Euler-Maruyama approximation (see [3]). Then we restrict the state space to a finite dimensional discrete state grid and specify the transition probabilities for these discrete states. This formulation leads us to a Markov decision chain corresponding to the original continuous-time optimal control problem.

1.2.1. *Euler-Maruyama Approximation.* We define the approximation scheme for a one dimensional process. The extension of the scheme to $\mathbb{R}^n$ is obvious. Suppose $\mathbf{X}$, a stochastic process with $X \subset \mathbb{R}^1$, satisfies equation (1) written for $\mathbb{R}^1$. Given a strictly increasing sequence of real numbers, $\tau = \{\tau_\ell\}_{\ell=0}^{N}$, with $\tau_0 = 0$ and $\tau_N = T$ we partition the time interval $[0, T]$.

An Euler-Maruyama[3] approximation of process $\mathbf{X}$ is a stochastic process

$$\mathbf{Y} = \{\mathbf{Y}_\ell \in X,\ \ell \in \mathrm{N},\ 0 \le \ell \le N\}$$

satisfying the equation (called the iterative scheme)

$$\mathbf{Y}_{\ell+1} = \mathbf{Y}_\ell + f\big(\mathbf{Y}_\ell, \mathbf{u}_\ell, \tau_\ell\big)\big(\tau_{\ell+1} - \tau_\ell\big) + b\big(\mathbf{Y}_\ell, \mathbf{u}_\ell, \tau_\ell\big)\big(\mathbf{W}(\tau_{\ell+1}) - \mathbf{W}(\tau_\ell)\big) \tag{8}$$

The indices run $\ell = 0, 1, 2, \ldots, N-1$, the initial value is

$$\mathbf{Y}_0 = \mathbf{X}(0) = x_0, \tag{9}$$

the value of process $\mathbf{Y}$ at the discretisation time τ_ℓ we have written as

$$\mathbf{Y}_\ell = \mathbf{Y}(\tau_\ell)$$

For a time discretisation using a constant time step we write

$$\tau_\ell = \ell\,\delta \quad \text{where} \quad \delta = \tau_{\ell+1} - \tau_\ell = \frac{T}{N} \quad \text{for some } N \in \mathrm{N}. \tag{10}$$

1.2.2. *Markov decision chain.* We shall use this term to denote a Markov decision process in a finite dimensional space. As in the case of the definition of the Euler discretisation we will treat (8) as defining a one dimensional state process. However, the extension of the scheme to a multidimensional case is easy. Some hints regarding the extension are included in footnotes.

To determine a Markov decision process we need to define a discrete state space, transition probabilities and allowable controls[4] for each state , as well as a cost, or reward, function associated with each transition.

---

[3] Euler, for short.

[4] The constraints on the control for the Markov chain are the same constraints applied to the continuous system. The method requires no additional restriction on allowable controls.

Notation. The discretisation scheme, while intuitively simple, overlays several layers of discretisation; discretisation of time, discretisation of state, and discretisation of noise. Trying to capture all of these discretisations in a simple notation is difficult. In an attempt at clarity we have adopted the following conventions.

1. When discussing variables whose domain is in continuous-time we use standard functional notation, *i.e.* $x(t)$. When discussing variables whose domain is in discrete time we use subscripts, *i.e.* x_ℓ.
2. Points of the discrete state space $\overline{X}_\ell$ are given an overline, *i.e.* $\overline{x} \in \overline{X}_\ell$. Points of the continuous state space do not have an overline, *i.e.* $x \in X$.
3. The bold symbols refer to stochastic processes.

Discrete State Space. The method allows for the discrete state space at each stage to be any finite collection of distinct points[5]. The method does not require the discrete state space to consist of equally spaced points[6]. The discrete state space may be different for each stage. The discrete state space for stage ℓ is denoted by $\overline{X}_\ell \subset \mathbb{R}^1$. Let the upper and lower bounds of the state grid be

$$\overline{U}_\ell = \max \overline{X}_\ell \quad \text{and} \quad \overline{L}_\ell = \min \overline{X}_\ell.$$

respectively. A point $x \in X$ is defined to be within the grid $\overline{X}_\ell$ if $\overline{L}_\ell \leq x \leq \overline{U}_\ell$. The collection of the discrete state spaces for all the stages, $\{\overline{X}_\ell\}_{\ell=0}^N$, is denoted $\overline{X}$ and called the discrete state space.

Adjacency. Heuristically our scheme approximates a point of X at stage ℓ by the points of $\overline{X}_\ell$ which are "adjacent" to it. In order to make this heuristic definition rigorous we must define what we mean by "adjacent".

1. We define two states of $\overline{X}_\ell$ to be **adjacent** if no other state of $\overline{X}_\ell$ lies between them[7].
2. Given a point of the continuous state space, $x \in X$, a pair of states, $\overline{x}^\ominus \in \overline{X}_\ell$ and $\overline{x}^\oplus \in \overline{X}_\ell$, is defined to be **adjacent to** x if the states are adjacent and $\overline{x}^\ominus < x < \overline{x}^\oplus$[8].
3. Given $x \in X$ with $x \geq \overline{U}_\ell$ we define $\overline{U}_\ell$ to be **adjacent to** x.
4. Given $x \in X$ with $x \leq \overline{L}_\ell$ we define $\overline{L}_\ell$ to be **adjacent to** x.
5. Given $x \in X$ with $x \in \overline{X}_\ell$ we define x to be **adjacent to** itself.

The above definitions enable us to associate with each $x \in X$ a *non-empty* set of points adjacent to it.

---

[5] The extension to $\mathbb{R}^n$ is trivial: for each $i \in 1, \ldots n$ define a finite collection of distinct points $\overline{X}_\ell^i$ with the discrete state space then being $\overline{X}_\ell = \prod_{i=1}^n \overline{X}_\ell^i$.

[6] All our numerical experiments have state grid for which, for each i and ℓ, $\overline{X}_\ell^i$ consists of equally spaced points.

[7] In the $\mathbb{R}^n$ case two states are **adjacent** if their projections onto each of the n coordinate axes are adjacent in the sense just defined.

[8] In the $\mathbb{R}^n$ case two points are **adjacent to** x if their projections onto each of the n coordinate axes are adjacent to the projections of x in the sense just defined.

Transition Probabilities. Consider the stochastic process $\mathbf{Y} = \{\mathbf{Y}_\ell,\ \ell = 0, 1, 2, \dots N\}$ where $\mathbf{Y}_\ell$ is defined through equation (8) (or n of them if $\mathbf{Y}_\ell \in \mathbb{R}^n$). Notice that this process, though defined at discrete times, can take any real value (i.e. this is still a process on the continuous state space). For a given control sequence $\mathbf{u}_\ell$ and equidistant discretisation times, we can re-write the iterative scheme (8) in the abbreviated form

$$\mathbf{Y}_{\ell+1} = \mathbf{Y}_\ell + \delta f_\ell + b_\ell \Delta \mathbf{W}_\ell \tag{11}$$

where by f_ℓ and b_ℓ we understand, respectively,

$$f_\ell = f(\mathbf{Y}_\ell, \mathbf{u}_\ell, \tau_\ell) \qquad\qquad b_\ell = b(\mathbf{Y}_\ell, \mathbf{u}_\ell, \tau_\ell).$$

The increments

$$\Delta \mathbf{W}_\ell = \mathbf{W}(\tau_{\ell+1}) - \mathbf{W}(\tau_\ell), \text{ for } \ell = 0, 1, 2, \dots N$$

are ones of the Wiener process $\mathbf{W} = \{\mathbf{W}(t), t \geq 0\}$ and are known (*cf* [3]) to be independent Gaussian random variables with mean and variance:

$$\mathbb{E}(\Delta \mathbf{W}_\ell) = 0 \qquad\qquad \mathbb{E}((\Delta \mathbf{W}_\ell)^2) = \delta.$$

Suppose that at some time τ_ℓ, $\mathbf{Y}_\ell = \overline{\mathbf{Y}}_\ell \in \overline{X}_\ell$. Assume for the time being, that there is no noise in the process (11), so for a given control value $\mathbf{u}_\ell$, the process moves to $\mathbf{Y}_{\ell+1}$ which is defined by:

$$\mathbf{Y}_{\ell+1} = \overline{\mathbf{Y}}_\ell + \delta f_\ell. \tag{12}$$

If $\mathbf{Y}_{\ell+1}$ has only one state adjacent to it then the transition probability from $\mathbf{Y}_\ell$ to this single adjacent state given $\mathbf{u}_\ell$ is 1. If there is a pair of states of $\overline{X}_{\ell+1}$ adjacent to $\mathbf{Y}_{\ell+1}$ then we assign the transition probabilities using an inverse distance method. Let $\left(\overline{\mathbf{Y}}^{\ominus}_{\ell+1}, \overline{\mathbf{Y}}^{\oplus}_{\ell+1}\right)$ be the pair of states adjacent to $\mathbf{Y}_{\ell+1}$ with $\overline{\mathbf{Y}}^{\ominus}_{\ell+1} < \overline{\mathbf{Y}}^{\oplus}_{\ell+1}$. Define

$$h_\ell = \overline{\mathbf{Y}}^{\oplus}_{\ell+1} - \overline{\mathbf{Y}}^{\ominus}_{\ell+1}$$

and assign the following non-zero transition probabilities

$$p(\overline{\mathbf{Y}}_\ell, \overline{\mathbf{Y}}^{\oplus}_{\ell+1} | \mathbf{u}_\ell) = \frac{\mathbf{Y}_{\ell+1} - \overline{\mathbf{Y}}^{\ominus}_{\ell+1}}{h_\ell} \tag{13}$$

$$p(\overline{\mathbf{Y}}_\ell, \overline{\mathbf{Y}}^{\ominus}_{\ell+1} | \mathbf{u}_\ell) = \frac{\overline{\mathbf{Y}}^{\oplus}_{\ell+1} - \mathbf{Y}_{\ell+1}}{h_\ell}. \tag{14}$$

If the Gaussian noise is present in (11) we can no longer assume a deterministic value of $\mathbf{Y}_{\ell+1}$. For this situation, we employ the following approach. Partition the Gaussian process $\Delta \mathbf{W}_\ell$ realisations' domain $(-\infty, \infty)$ into three intervals[9]: $(-\infty, -\sqrt{\delta}], (-\sqrt{\delta}, \sqrt{\delta}], (\sqrt{\delta}, +\sqrt{\infty})$. Remember

---

[9] The partition into three intervals is arbitrary. It could be made finer if the current one did not deliver satisfactory results. However, notice that this partition depends on the time discretisation step δ in the following way: the smaller δ the more "accurately" the central value is defined. This promises that with δ decreasing the transition probabilities will reflect the feature of the Gaussian noise that with the standard deviation decreasing, the

that $\sqrt{\delta}$ is the standard deviation of $\Delta\mathbf{W}_\ell$. Compute the expected values of the noise for each interval in the following manner. Calculate the expression:

$$\omega = \frac{\frac{\sqrt{\delta}}{\sqrt{2\pi}}\displaystyle\int_1^\infty t e^{-\frac{t^2}{2}}\mathrm{d}t}{\frac{1}{\sqrt{2\pi}}\displaystyle\int_1^\infty e^{-\frac{t^2}{2}}\mathrm{d}t} = \frac{\sqrt{2\delta}}{\sqrt{\pi e}\,(1-\mathrm{erf}(\frac{1}{\sqrt{2}}))} \tag{15}$$

where

$$\mathrm{erf}(t) = \frac{2}{\sqrt{\pi}}\int_0^t e^{-t^2}\mathrm{d}t \tag{16}$$

is the the standard error function. We define the transition probabilities for an approximated situation in which process $\mathbf{Y}$ is perturbed by the discretely valued noise w_ℓ, $\ell = 0, 1, 2, \ldots N-1$ (rather than by the original Gaussian noise $\Delta\mathbf{W}_\ell$) of the following probability distribution[10]:

$$P(w_\ell = -\omega) = p_- = .1587 \tag{17}$$

$$P(w_\ell = 0) = p_0 = .6826 \tag{18}$$

$$P(w_\ell = +\omega) = p_+ = .1587. \tag{19}$$

This approximation of continuous perturbation by a three-value noise is of course arbitrary. As said in footnote "9", we can divide the perturbation realisations domain $(-\infty, \infty)$ into many intervals to get a more realistic discrete representation of $\Delta\mathbf{W}_\ell$. However, this might not be necessary if the time discretisation parameter $\delta = \tau_{\ell+1} - \tau_\ell$ is small. The intuitive argument is that tighter time discretisation results in (absolutely) lower expected values of the off-central noises, hence improves chances of the process to behave in line with the drift.

For this noise representation, the definition of the transition probabilities in the stochastic case is only slightly different from (13), (14). Let $\mathbf{Y}_{\ell+1}$ be determined through (11). Our noise discretisation method means that for $\delta \neq 0$ the process reaches, at $\ell + 1$:

$$\mathbf{Y}^-_{\ell+1} = \mathbf{Y}_{\ell+1} + b_\ell\,\omega_- \qquad \text{with probability } p_- \tag{20}$$

$$\mathbf{Y}^0_{\ell+1} = \mathbf{Y}_{\ell+1} \qquad \text{with probability } p_0 \tag{21}$$

$$\mathbf{Y}^+_{\ell+1} = \mathbf{Y}_{\ell+1} + b_\ell\,\omega_+ \qquad \text{with probability } p_+. \tag{22}$$

¿From this point we proceed much as we did in the deterministic case, for each one computing the states adjacent to it. Should there be only a single adjacent state then the transition probability to such a state is given by the corresponding (20), (21) or (22). If there are two adjacent states then we apply the inverse distance method as in (13), (14) but weight the two probabilities by the

---

central value becomes more probable. We will examine how well this noise representation captures the behaviour of the stochastic system in the numerical experiments carried out in Section 2. A variety of diffusion terms are considered to see what effect the "magnitude" of the noise has.

[10] The probabilities are of the standard normal distribution computed for the intervals: $(-\infty, -1], (-1, 1], (1, \infty)$.

appropriate (20), (21) or (22). Thus, for example, if $\mathbf{Y}^-_{\ell+1} \notin \overline{X}_{\ell+1}$ but there exist $\left(\overline{\mathbf{Y}}^{-\ominus}_{\ell+1}, \overline{\mathbf{Y}}^{-\oplus}_{\ell+1}\right)$ in $\overline{X}_{\ell+1}$ adjacent to $\mathbf{Y}^-_{\ell+1}$ with $\overline{\mathbf{Y}}^{-\ominus}_{\ell+1} < \overline{\mathbf{Y}}^{-\oplus}_{\ell+1}$ then the transition probabilities are

$$p(\overline{\mathbf{Y}}_\ell, \overline{\mathbf{Y}}^{-\oplus}_{\ell+1}|\mathbf{u}_\ell) \;=\; p_- \,\frac{\mathbf{Y}_{\ell+1} - \overline{\mathbf{Y}}^{-\ominus}_{\ell+1}}{h_\ell} \tag{23}$$

$$p(\overline{\mathbf{Y}}_\ell, \overline{\mathbf{Y}}^{-\ominus}_{\ell+1}|\mathbf{u}_\ell) \;=\; p_- \,\frac{\overline{\mathbf{Y}}^{-\oplus}_{\ell+1} - \mathbf{Y}_{\ell+1}}{h_\ell} \tag{24}$$

where

$$h_\ell = \overline{\mathbf{Y}}^{-\oplus}_{\ell+1} - \overline{\mathbf{Y}}^{-\ominus}_{\ell+1}.$$

If any of the states $\mathbf{Y}^{-\ominus}, \mathbf{Y}^{+\ominus}$, *etc.* overlaps another, the respective probabilities have to be summed up.

The above discretisation method has been motivated intuitively rather analytically. However, we observe that there is some *local consistency* between of $\mathbf{X}$ and $\overline{\mathbf{Y}}_\ell$ in the sense of [5] [11].

Transition Costs. So far we have defined a controlled Markov chain in the discrete state space $\overline{X}$ with the transition probabilities as described above. To define a Markov decision chain we need to assign a cost function to every transition.

Let the control strategy be Markovian (as in (2)) and action at state ℓ computed as

$$\mathbf{u}_\ell = \mu(\ell, \overline{\mathbf{Y}}_\ell), \qquad \overline{\mathbf{Y}}_\ell \in \overline{X}_\ell, \qquad \ell = 0, 1, 2, \ldots N-1. \tag{25}$$

Consider the performance criterion of the original optimisation problem (4). For the approximated problem, the decision maker incurs a cost that depends on the state at stage ℓ and on the action $\mathbf{u}_\ell$ (rather than on where the chain moves, as in the general case of non-controlled Markov chains). We define this cost as

$$\gamma(\overline{\mathbf{Y}}_\ell, \mathbf{u}_\ell, \ell) = \delta\, g(\overline{\mathbf{Y}}_\ell, \mathbf{u}_\ell, \ell),\;\; \ell = 0, 1, 2, \ldots N-1. \tag{26}$$

Assuming that $\mathbf{X}_T \in X^h$, we can determine the overall cost for the Markov decision chain $\mathbf{Y}$, starting from $\mathbf{Y}_0 = x_0$ and controlled by $\mathbf{u} = \{\mathbf{u}_0, \mathbf{u}_1, \ldots \mathbf{u}_{N-1}\}$ as

$$J(0, \overline{x}_0; \mathbf{u}) = \mathbb{E}\left(\sum_{\ell=0}^{N-1} \gamma(\overline{\mathbf{Y}}_\ell, \mathbf{u}_\ell, \ell) + h(\overline{\mathbf{Y}}_N) \,\middle|\, \overline{\mathbf{Y}}_0 = \overline{x}_0\right) \tag{27}$$

---

[11] See [5], page 1002. Conditions "1" and "2" (about continuity of the Markov chain expected value and variance) are easy to prove. However, because in our discretisation method the state step is independent of the time step, condition "3" (about continuity of the Markov chain increments) can only be satisfied if the grid $\overline{X}_l$ is allowed to become denser. Moreover, local consistency fails along the boundary of the discrete state space. This is not surprising since it would be impossible for a system constrained to a finite space to follow the behaviour of a system which is not similarly constrained at the points where the constraints become active.

We consider the problem:

$$\max_{\mathbf{u}} J(0, \overline{x}_0; \mathbf{u})$$

subject to

$$vecY_{\ell+1} = \mathbf{Y}_\ell + \delta f_\ell + b_\ell \Delta \mathbf{W}_\ell,$$

with the transition probabilities defined as above, the *Markov decision chain* approximating the original continuous-time optimisation problem of Section 1.1. To stress that the optimisation result depends on the discretisation parameters we will use the notation

$$J(\delta, \overline{X}) = J(0, \overline{x}_0; \hat{\mathbf{u}}) \tag{28}$$

where $\hat{\mathbf{u}}$ is a maximiser of the above optimisation.

2. NUMERICAL EXPERIMENTS

2.1. **The Algorithm and Computational Complexity.** With any numerical algorithm, and especially with one which claims to use the Dynamic Programming technique, the behaviour of the computation time as the various parameters change is critical. There are two crucial parameters for the solution method outlined above: the number of states and the number of time steps. Obviously increasing the number of states or the number of time steps will increase the computation time. It is our claim that the computation time increases linearly in both the number of states and the number of time steps.

Suppose we are computing a solution by backward induction for a state in stage ℓ and that the solution from stage $\ell + 1$ onwards has already been determined. The time required to compute the optimal decision for the state we are in is *largely* independent of both the number of time steps and the number of states. Its independence of the time step is essentially obvious. Its independence of the number of states is a consequence of our approximation scheme scanning only the adjacent states in the next stage. The number of adjacent states (in the interior of the state grid) is fixed by the dimension of the state space[12]. Doubling the number of states means that twice the time is taken for each stage (although the computation time for each state remains the same, there are twice as many states x_ℓ for each stage ℓ) and the computation time doubles. Doubling the number of time steps leaves the computation time for each stage fixed but doubles the number of stages and hence the computation time doubles. This exact linear relationship (reported in Figure 6 in Section 2.3.1) is spoiled by the vagaries of the computation time of the numerical minimisation, as well as the load dependence of the computer performance (see Figure 7).

---

[12] In numerical examples provided below, for those which are in $\mathbb{R}^1$ and whose noise discretisation is as explained in footnote "9", there will in general be *two*, or *six*, non-zero entries in each row of the transition probability matrix, for a deterministic or stochastic case, respectively.

Recent papers [8] and [9] report breaking the curse of dimensionality for a certain subclass of Markov decision chains through use of randomisation. Notice that the results, which we have obtained, indicate that also "our" randomisation leads to a similar conclusion.

2.2. **The Role of Simulation.** The expected value obtained by the approximation process is not directly comparable to the expected value of the continuous-time stochastic optimal control problem being modelled. To compare, we compute a "simulated" value. More precisely, we use the control rule, derived by solving the Markovian approximation (and interpolate it, if necessary), for controlling a continuous-state process. The shorter the simulation time step, the closer the simulated process to its continuous time counterpart.

Many interpolation schemes could be used to infer a continuous-time continuous-state control rule. For simplicity we have used linear interpolation to determine the control at states which lie between points on the state grid and have used no interpolation in time. If we are at time t, and $t_{\ell-1} \leq t < t_{\ell}$, then the control rule computed for stage ℓ of the Markov decision chain is used by the simulation routine. All the other facets of the simulation routine we used are similarly simple. In particular, the integral for the performance criterion is computed via the left-hand endpoint rectangles method and the differential equation is solved via Euler's method. The number of simulation time steps can be different from the number of time steps used to compute the Markovian approximation and is typically much higher.

With a deterministic problem simulating the system gives, for a given number of simulations steps and a given control rule, an unambiguous value. The expected value of such a stochastic problem is much harder to compute from the control rule. Before we can simulate we must choose a noise realisation. To get an approximation to the expected value of the system we perform a large number of simulation runs, each with a noise realisation chosen at random from the appropriate normal distributions, and then average the values thus obtained. This process only gives us an approximation to the expected value. We shall use "Markov value" to denote an expected value obtained directly from the Markov decision chain and use "simulated value" to denote a value computed via simulation, as explained above.

The lack of any interpolation in time means that if the control rule, derived from the Markovian approximation, is time inhomogeneous strange "jumps" can occur in the control time path at $t = t_1, t_2, \ldots$ The fewer the number of time steps used in computing the Markovian approximation the more pronounced this effect can be, see Figure 8.

In the numerical experiments below the timepaths and the values shown in the convergence graphs result from simulating the continuous-time system as above. For deterministic problems see Figures 1, 4, 5, 8, 10, and 11; for a stochastic problem see Figures 14 and 15. The continuous-state approximated control rules can be directly compared with the analytical solution control rules, see Figures 2, 9, and 12. Similarly we can show the optimal value function of the Markov decision chain together with the optimal value function of the associated continuous-time model, see Figures 3 and 13. As said before, unlike the control rules, however, the value functions are not directly comparable; in Figure 13 the Markovian approximations value function lies below the optimal value function of the continuous-time model.

2.3. **The Deterministic Case.** We shall consider the case of a deterministic system in more detail. This case clearly motivates the process and forms the base of the procedure for stochastic problems. In this section, we provide the numerical solution to two optimal control problems and compare them to their analytical counterparts.

2.3.1. *A Test Problem in* $\mathbb{R}^1$. Consider the system with equation of motion (30), performance criterion (29), with finite time horizon $T = 1$, state space $X = \mathbb{R}^1$, and control space $U = \mathbb{R}^1$.

$$\min_{u(\cdot)} \ J \ = \frac{1}{2}\int_0^1 \Big(u(t)^2 + x(t)^2\Big)\mathrm{d}t + \frac{1}{2}x^2(1) \tag{29}$$

subject to

$$\dot{x} = u \tag{30}$$

$$x(0) = \frac{1}{2}\,. \tag{31}$$

This is a linear-quadratic problem for which we can obtain an analytical optimal solution. This was chosen for the ease with which the analytical solution could be obtained. The numerical procedure does not explicitly exploit any of the special structure of linear-quadratic problems[13].

The corresponding Markov decision chain is:

$$\min_{u_\ell} \ J \ = \frac{1}{2}\sum_{\ell=0}^{N-1} \delta\,(u_\ell^2 + \overline{x}_\ell^2) + \frac{1}{2}\overline{x}_N^2 \tag{32}$$

subject to

$$x_{\ell+1} = \overline{x}_\ell + \delta u_\ell \tag{33}$$

$$\overline{x}_0 = \frac{1}{2} \tag{34}$$

with $\ell = 0, 1, \ldots, N = 1$.

Finally we specify the transition probabilities and the boundary conditions. As introduced in Section 1.2, the probabilities are based on an inverse distance method applied to those points in $\overline{X}$ which are “adjacent” to a given target point in X. For example, consider an equidistant time discretisation

$$\tau_{\ell+1} - \tau_\ell = \delta = .1,$$

---

[13] *I.e.* the method makes no use of the Ricatti equation

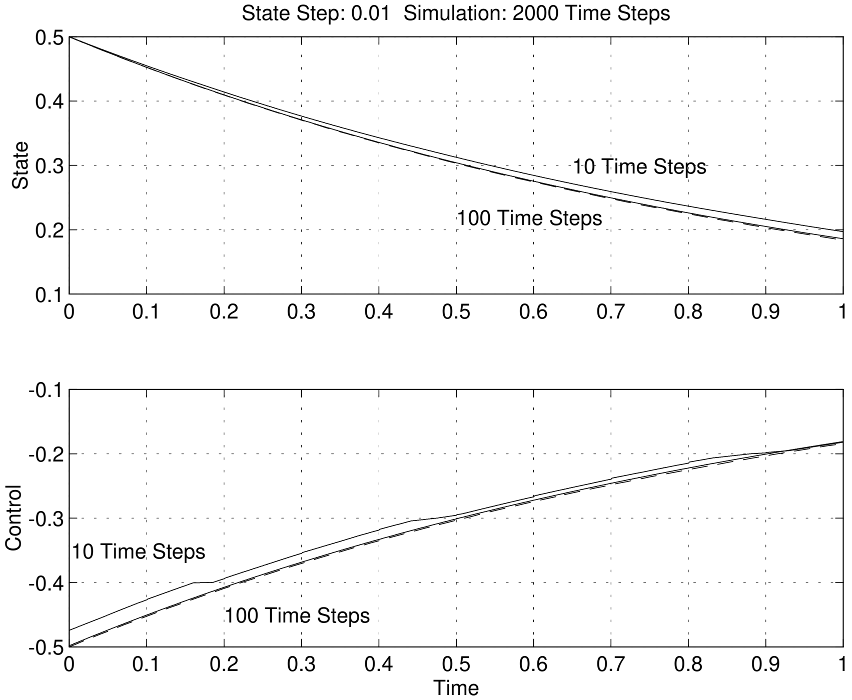

FIGURE 1. Optimal and approximated trajectories.

and an equidistant, constant, space discretisation of $[0, .5]$ with $h = .05$. Suppose $\overline{x}_\ell = 0.4$. For control $u_\ell = -0.3$, the system moves to

$$x_{\ell+1} = .4 + .1 \cdot (-0.3) = .37 \notin \overline{X}_{\ell+1}.$$

The adjacent states are

$$\overline{x}^{\oplus}_{\ell+1} = .4 \quad \text{and} \quad \overline{x}^{\ominus}_{\ell+1} = .35.$$

We define the following transition probabilities:

$$p(.40, .40| -0.3) = \frac{.37 - .35}{.05} = .4$$
$$p(.40, .35| -0.3) = \frac{.40 - .37}{.05} = .6.$$

The boundary conditions are defined using the concept of $\overline{U}_\ell$ and $\overline{L}_\ell$ adjacent states, see Section 1.2. For example, consider the problem in the example above, but suppose $\overline{x}_\ell = 0.1$. For control $u_\ell = -2$, the system moves to

$$x_{\ell+1} = .1 + .1 \cdot (-2) = -.1 \notin \overline{X}_{\ell+1}.$$

The only state adjacent to $x_{\ell+1}$ is 0 and the transition probability is $p(0.1, 0| -2) = 1$.

The optimal solution to min (29), subject to (30) and (31) provides us with the optimal

$$\hat{J} = 0.125.$$

For comparison purposes the optimal control and state time paths as well as the optimal control rule and value function are displayed as dashed lines in figures where the approximated solutions are presented.

The control and state time paths, resulting from min (32), subject to (33) are shown in Figure 1 for two discretisation combinations. The solution computed with 100 time steps is almost indistinguishable from the optimal solution. The "waves" in the 10 time step control timepath result from "waves" in the control rule for the 10 time step approximation, see Figure 2.

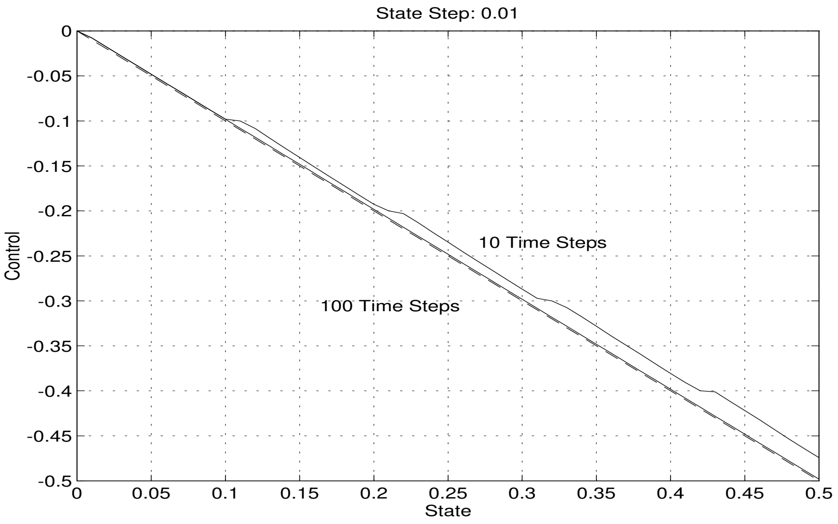

FIGURE 2. Optimal policy rules (t=0).

The optimal control rules resulting from two Markovian approximations are displayed in Figure 2, and the optimal value functions of these Markovian approximations are displayed in Figure 3. The test problem has a time-homogeneous linear optimal control rule, $u(t) = -x(t)$ and a quadratic optimal value function $J(0, x) = x^2$. The control rules for the Markovian approximation are not exactly time homogeneous. Note that the optimal control rules displayed are the ones associated with $t = 0$. The accuracy, with which the optimal policy for the 100 time step Markovian approximation matches the optimal policy for the continuous-time test problem, results in the simulated timepaths and, though not shown, the simulated value accurately mirroring those of the test problem.

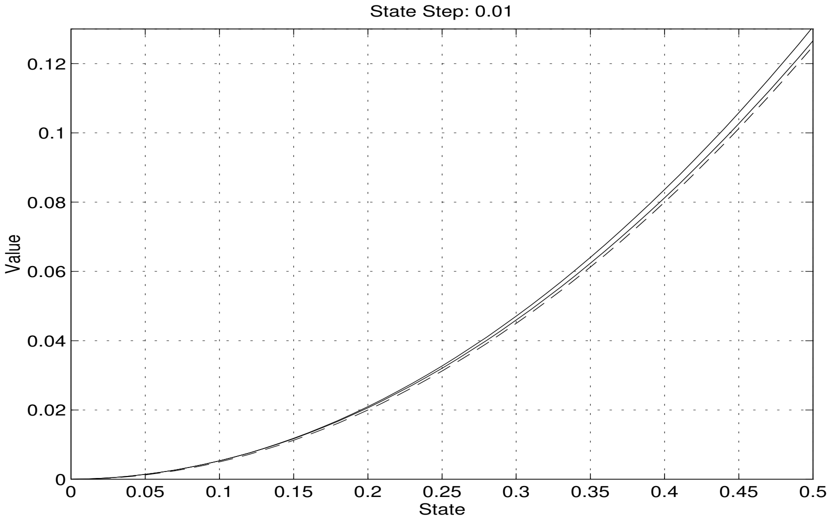

FIGURE 3. Optimal value function ($t = 0$).

The most important result, from the application viewpoint, is the convergence of $J(\delta, \overline{X})$ exhibited in Figures 4 and 5. Figure 4 shows the convergence of the value to the actual optimal value as the number of *time* steps is increased. Figure 5 shows the convergence of the value to the actual optimal value as the number of *state* steps is increased. Also of interest in these graphs is the "fast" nature of this convergence.

Notice that the results displayed in Figures 1, 4 and 5 are obtained through the use of the control rules of Figure 2 for a given simulation time step (which is different from δ). This means that the values of the sum (32) approximate the integral (29) by the rectangles method, with an error resulting from that method. The discrepancy between the best approximation of the optimal value and the one of the continuous-time case, are attributed to both integration error and the intrinsic error of the method.

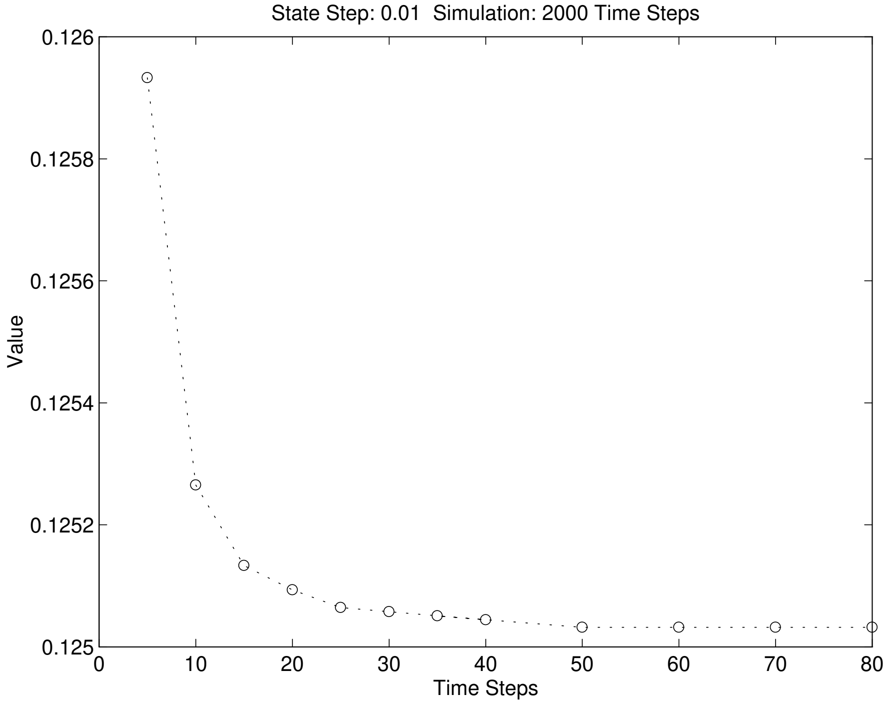

Figure 4. Simulated value as a function of time discretisation steps.

The comparison of Figures 4 and 5 tells us that approximately the same solution accuracy can be obtained using different discretisation combinations. Here is where we observe the linear growth of the computational burden as the discretisation steps become smaller. Figure 6 shows how many minimisations of the value function are required for a few discretisation combinations.

Those comparisons also raise the question of how to most efficiently compute a solution of a given accuracy. For example, the same accuracy of solution $J(\delta, \overline{X}) = .12505$ can be obtained by 1600 minimisations resulting from 100 time steps on a 16 node state grid (state step 0.03333) or by 2040 minimisations resulting from 40 time steps on a 51 node state grid (state step 0.01). The tradeoff between the various discretization parameters is problem specific.

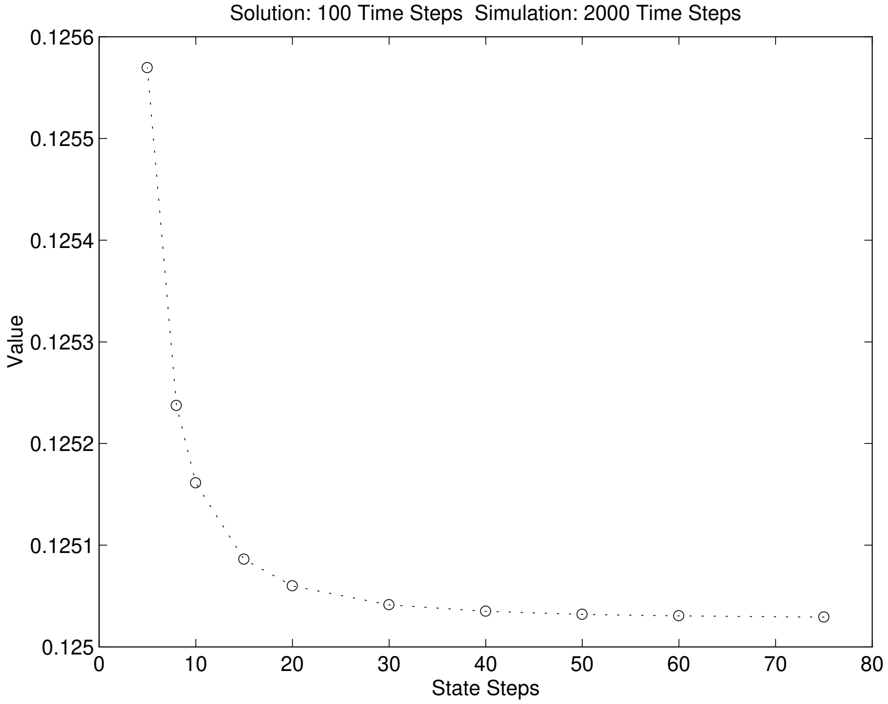

Figure 5. Simulated value as a function of state discretisation.

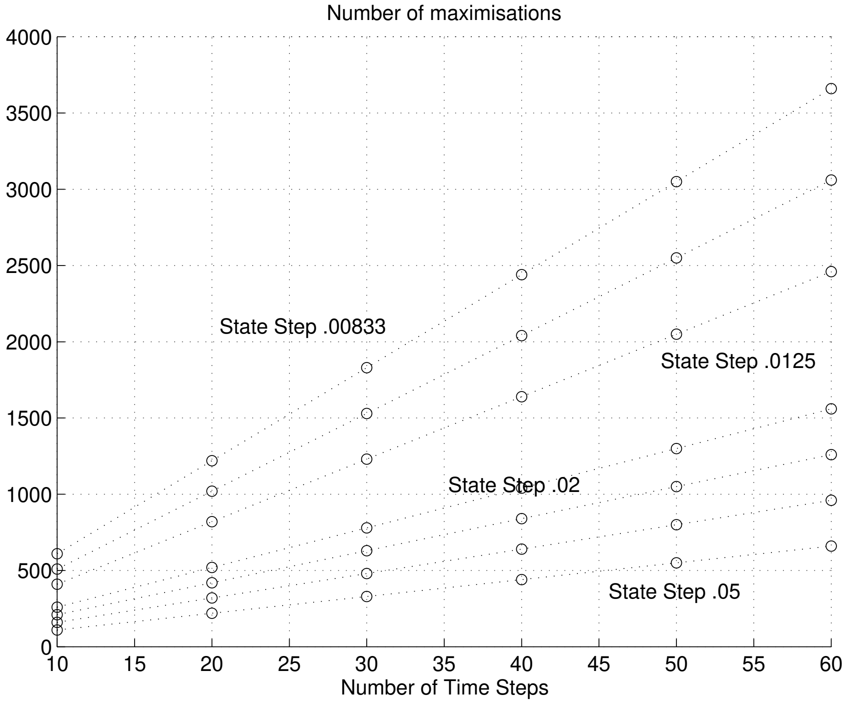

Figure 6. Computation burden in Time Steps.

We also have measured the computation times of generating a solution to a problem, see Figure 7. Despite using CPU seconds as our measure of computation time, the measures were still affected by the load of the machine (which used virtual memory). Moreover, each discretisation combination modifies the "geometry" of the value function. Consequently, we observe some deviations from linearity in this figure. However, this figure gives also some evidence for us to conclude that the computation time increases not faster than linearly in the number of time, or state-space, steps.

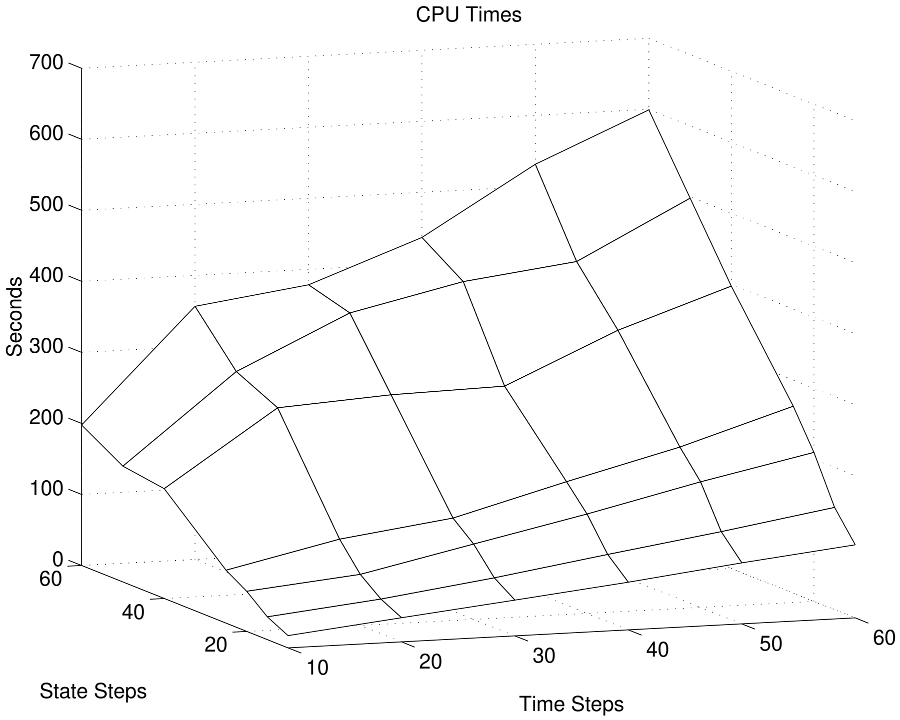

FIGURE 7. Computation times.

2.3.2. *A Test Problem in $I\!R^2$.* We want to test our method using a two dimensional example. The continuous-time optimisation problem is now composed of two equations of motion (36) and performance criterion (35). The time horizon is finite $T = 1$, state space $X = \mathbb{R}^2$, and control space $U = \mathbb{R}^1$ The model is as follows:

$$\min_{u(\cdot)} \; J \; = \frac{1}{2}\int_0^1 \left(\frac{1}{2}u(t)^2 + y(t)^2\right) \mathrm{d}t + \frac{1}{2}x^2(1) \tag{35}$$

subject to

$$\begin{cases} \dot{x} = y \\ \dot{y} = u \end{cases} \tag{36}$$

$$\begin{cases} x(0) = 1 \\ y(0) = 0 \end{cases} . \tag{37}$$

This is a linear-quadratic problem for which we can obtain an analytical optimal solution. The optimal solution to min (35), subject to (36) and (37) provides us with the optimal

$$\hat{J} = .3644802.$$

The optimal policy rule for $t = 0$ is

$$\hat{u}_0 = -0.917x_0 - 1.833y_0$$

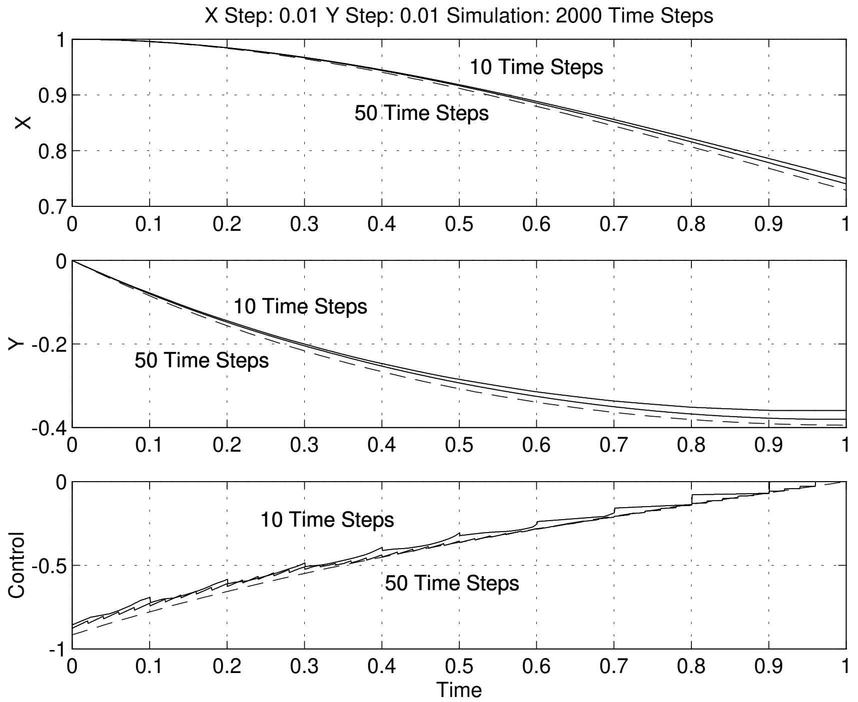

FIGURE 8. 2D Optimal and approximated trajectories.

which was computed via the Ricatti equation. The optimal timepaths were computed using the Maximum Principle of Pontryagin.

The corresponding Markov decision chain is:

$$\min_{u_\ell} J = \frac{1}{2} \sum_{\ell=0}^{N-1} \delta \left(\frac{1}{2} u_\ell^2 + \overline{y}_\ell^2 \right) + \frac{1}{2} \overline{x}_N^2 \tag{38}$$

subject to

$$\begin{cases} x_{\ell+1} = \overline{x}_\ell + \delta \overline{y}_\ell \\ y_{\ell+1} = \overline{y}_\ell + \delta u_\ell \end{cases} \tag{39}$$

with $\ell = 0, 1, \ldots, N = 1$.

The transition probabilities and the boundary conditions are defined in a similar way to those in the one dimensional case.

As before, for comparison purposes, the optimal control and state time paths as well as the optimal control rule and value function are displayed as dashed lines in figures where the approximated solutions are also presented.

The control and state time paths, resulting from min (38), subject to (39) are shown in Figure 8 for two discretisation combinations. The solution computed with 50 time steps appears to be a relatively good approximation to the optimal solution. The jumps in the control timepaths are caused by the time inhomogeneity of the Markov chain control rules.

The optimal control rules resulting from a Markovian approximation is displayed in Figure 9. Note that the optimal control rules displayed are the ones associated with $t = 0$. The control rule graph reveals that the Markovian approximation's control rule is a relatively good approximation in a neighborhood of $(x_0, y_0) = (1, 0)$. Notice that these control rule graphs are for $t = 0$ and

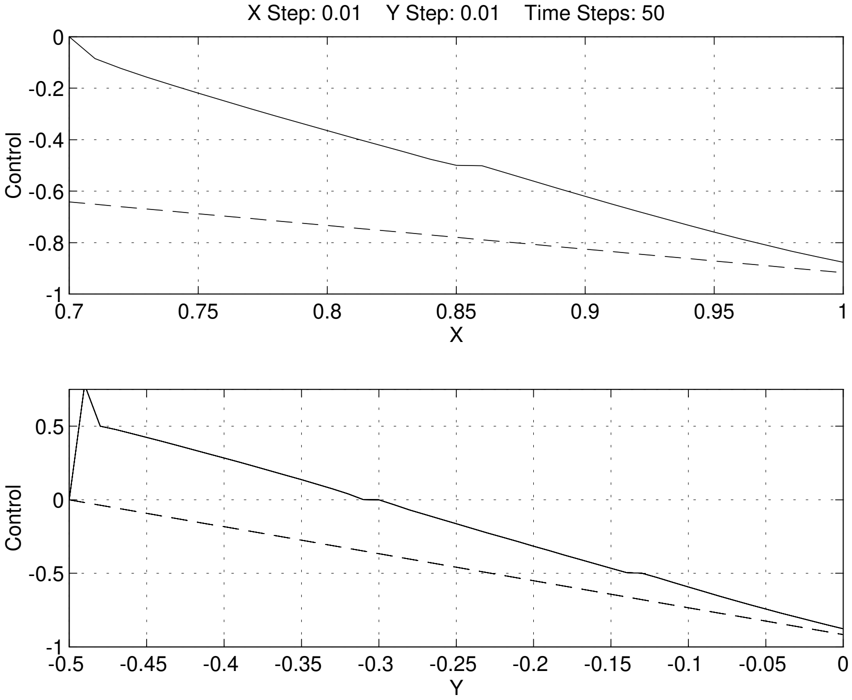

FIGURE 9. 2D optimal policy rules.

our problem has initial condition $(x_0, y_0) = (1, 0)$. The state grid for this problem was selected to produce convergence for the initial condition $(x_0, y_0) = (1, 0)$ and there is no reason why we should get, or need, convergence away from this initial condition.

The most important, however, from the application viewpoint of the approximation method is again the converging behaviour of $J(\delta, \overline{X} \times \overline{X})$ shown in Figures 10 and 11 which display the procedure converging as the number of time steps, and states, are increased. Notice that Figure 11 has been obtained using the same step size in both x and y directions. Also note the "speed" of convergence in the state step, from what we can see the convergence appears to be linear. Convergence in time step length is also linear. Figure 10 is plotted in the number of time steps, $1/\delta$, and looks like a hyperbola. If we plotted the value against the time step, δ, then a linear relationship would be evident.

Here too, the discrepancy between the best approximation of the optimal value and the one of the continuous-time case, are attributed to both integration error and the intrinsic error of the method. The jumps in the control timepath are caused by the time inhomogeneity of the Markov chain control rules.

The computation times of getting a 2D-problem solution, and their dependence on discretisation, were not greatly different from those reported in Section 2.3.1, and will not be discussed.

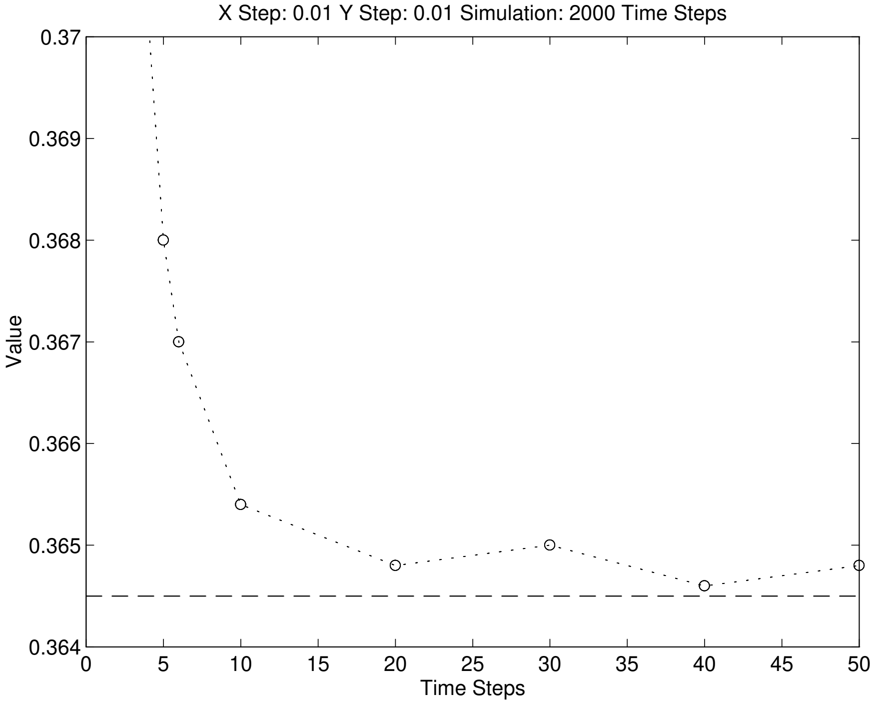

FIGURE 10. 2D simulated value as a function of time discretisation steps.

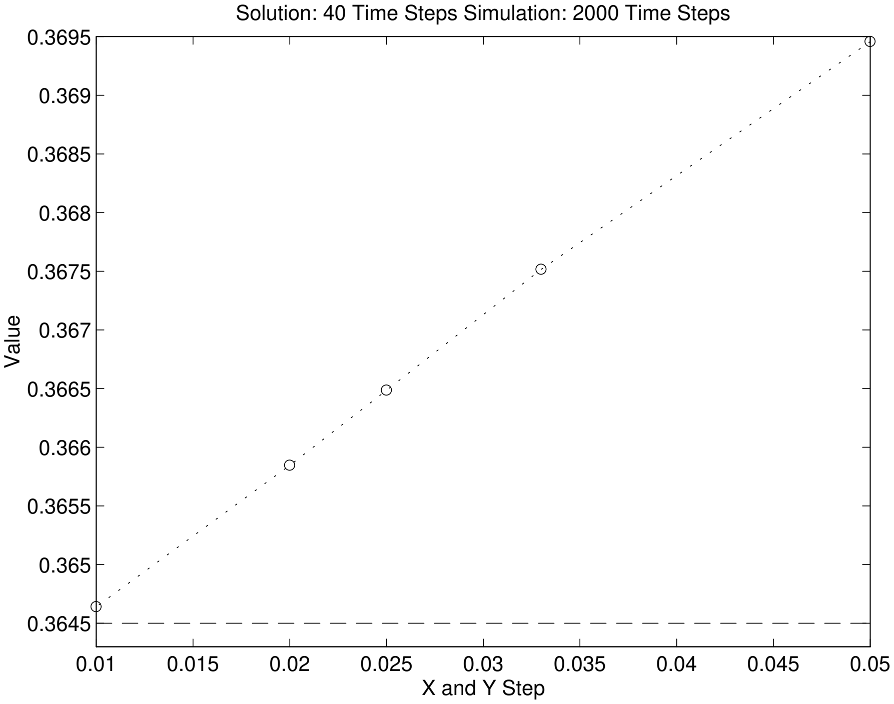

FIGURE 11. 2D simulated value as a function of state discretisation.

2.4. **The Stochastic Case.** We are going to examine only a one dimensional stochastic test case since the resource valuation problem in Part 2 is a two dimensional stochastic problem. We have already concluded from 2.3 that the Markovian approximation behaves well in the absence of noise. The purpose of this section is to test how well our scheme for representing the noise in the system captures the behaviour of the continuous-time stochastic system.

We extend the test problem used in the deterministic case, Section 2.3.1. Consider a system with the equation of motion

$$d\mathbf{X} = \mathbf{u}\,dt + \sigma\,d\mathbf{W}.$$

The discretised system has equation of motion

$$\mathbf{Y}_{\ell+1} = \mathbf{Y}_\ell + \delta \mathbf{u}_\ell + \sigma \Delta \mathbf{W}_\ell$$

where δ is the time step, $\Delta\mathbf{W}_\ell$ is a discrete random variable, and σ is a parameter determining the magnitude of the noise. For example consider the system with $\delta = 0.1$ and $\sigma = 0.30$. Then, as defined on page 8, $\Delta\mathbf{W}_\ell$ takes values -0.48229, 0, and 0.48229 with probabilities 0.1587, 0.6826, and 0.1587. Let $\mathbf{Y}_\ell = 0.4$, $\mathbf{u}_\ell = -0.3$ and the state step be 0.05 as in the earlier example. Taking into account the three different values of $\Delta\mathbf{W}_\ell$ we get three different values of $\mathbf{Y}_{\ell+1}$

$$\begin{aligned}
\mathbf{Y}^-_{\ell+1} &= 0.4 + 0.1\cdot(-0.3) + 0.2\cdot(-0.48229) \\
&= 0.273542 && \text{with probability } 0.1587 \\
\mathbf{Y}^0_{\ell+1} &= 0.4 + 0.1\cdot(-0.3) + 0.2\cdot 0 \\
&= 0.37 && \text{with probability } 0.6826 \\
\mathbf{Y}^+_{\ell+1} &= 0.4 + 0.1\cdot(-0.3) + 0.2\cdot 0.48229 \\
&= 0.466458 && \text{with probability } 0.1587
\end{aligned}$$

which each have two adjacent states. Thus there are six points to which the system can evolve.

$$\begin{aligned}
p(.40, .50|-0.3) &= 0.1587\cdot\frac{0.466458-0.45}{0.05} = 0.0522377 \\
p(.40, .45|-0.3) &= 0.1587\cdot\frac{0.50-0.466458}{0.05} = 0.106462 \\
p(.40, .40|-0.3) &= 0.6826\cdot\frac{0.37-0.35}{0.05} = 0.27304 \\
p(.40, .35|-0.3) &= 0.6826\cdot\frac{0.40-0.37}{0.05} = 0.40956 \\
p(.40, .30|-0.3) &= 0.1587\cdot\frac{0.273542-0.25}{0.05} = 0.0747223 \\
p(.40, .25|-0.3) &= 0.1587\cdot\frac{0.30-0.273542}{0.05} = 0.0839777
\end{aligned}$$

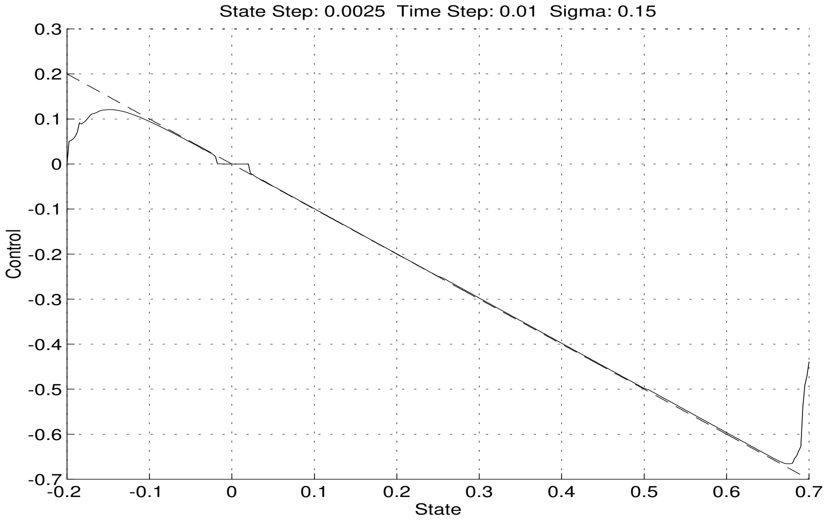

Figure 12. Optimal policy rules.

Now we apply our approximation scheme to a stochastic optimal control problem, compare Section 2.3.1:

$$\min_{\mathbf{u}(\cdot)} J = \mathbb{E}\left(\frac{1}{2}\int_0^1 (\mathbf{u}(t)^2 + \mathbf{X}(t)^2)\mathrm{d}t + \frac{1}{2}\mathbf{X}^2(1)\,\Big|\,\mathbf{X}(0) = \frac{1}{2}\right) \tag{40}$$

subject to

$$\mathrm{d}\mathbf{X} = \mathbf{u}\,\mathrm{d}t + \sigma\,\mathrm{d}\mathbf{W}. \tag{41}$$

The optimal control rules resulting from the Markovian approximation are displayed in Figure 12, and the optimal value function in Figure 13.

The Markovian control rule is almost indistinguishable from the analytical solution except at the edge of the state grid. It is not surprising that the optimal control rule differs near the edges of the state grid. There is a penalty function included in the solution algorithm to prevent the solution from crossing outside the state grid and this will have an effect near the edge of the state grid.

Here, as in the one dimensional deterministic case, the value function of the Markovian approximation is extremely close to the value function of the continuous-time problem. However Figure 13 seems to show that the Markovian approximation yields a “better” value than the optimal solution. There is no paradox here, the values are obtained from distinct, though related, performance criteria, see Section 2.2.

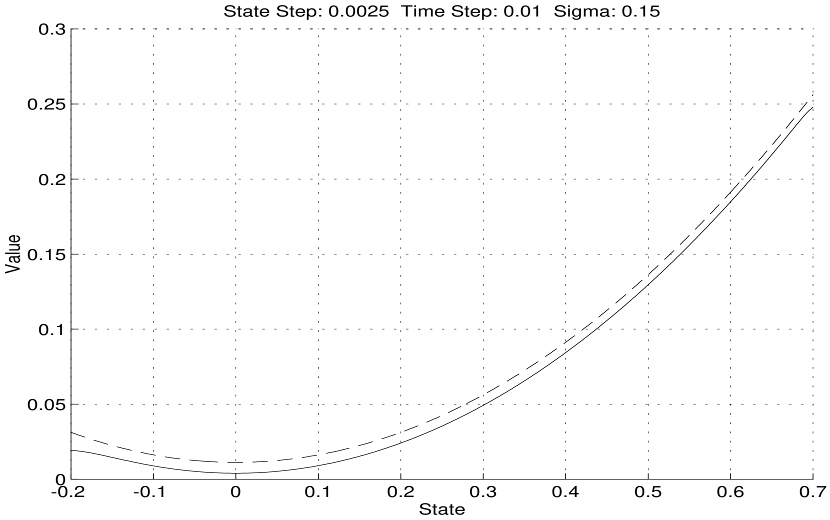

FIGURE 13. Optimal value function ($t = 0$).

The convergence in the time step of the simulated value to the optimal continuous-time value is shown in Figure 14 for three different values of σ. The dashed lines are again the values obtained from the analytical solutions.

The convergence is clear and fast. For each point 4000 simulation runs were performed, each with 100 time steps[14], using a noise realisation chosen at random and then the average of the values yielded by these simulation runs was taken. The vagaries of this method of computing the expected value explains the strange point for $\sigma = 0.15$ where the approximated solution has apparently yielded a better value than the analytical optimal solution.

The convergence in state of the simulated value to the optimal continuous-time value is shown in Figure 15 for three different values of σ. Notice that even with 100 state steps the $\sigma = .30$ simulated value has, strictly speaking, failed to converge. Nevertheless even without the proper convergence, the obtained simulation values are close to the continuous time optimal one. The convergence shown by the other cases leads us to conclude that the larger the diffusion term in relation to the drift term the more solution timesteps are required for convergence (if indeed we get convergence at all for this combination of state step, simulation time step, and number of simulations).

---

[14] These are arbitrary numbers. We chose 4000 runs of 100 steps so that the simulations were completed in a reasonable time and the variance of the simulated values was relatively small.

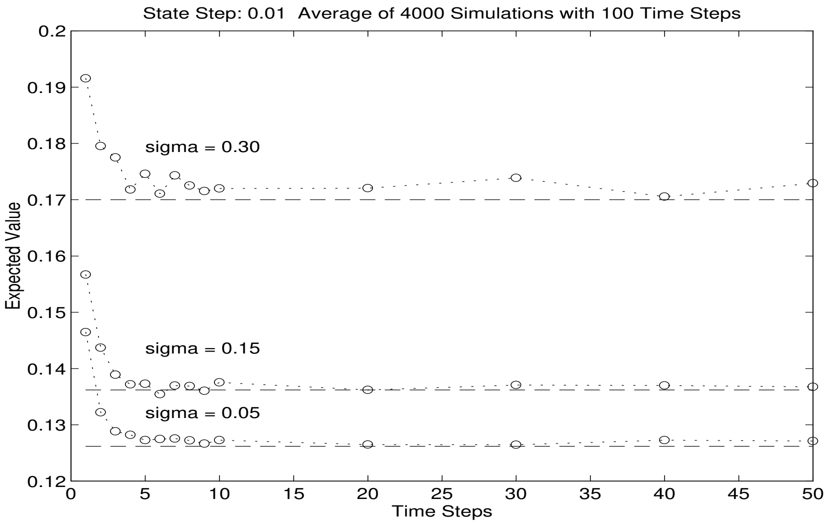

Figure 14. Simulated value as a function of time discretisation.

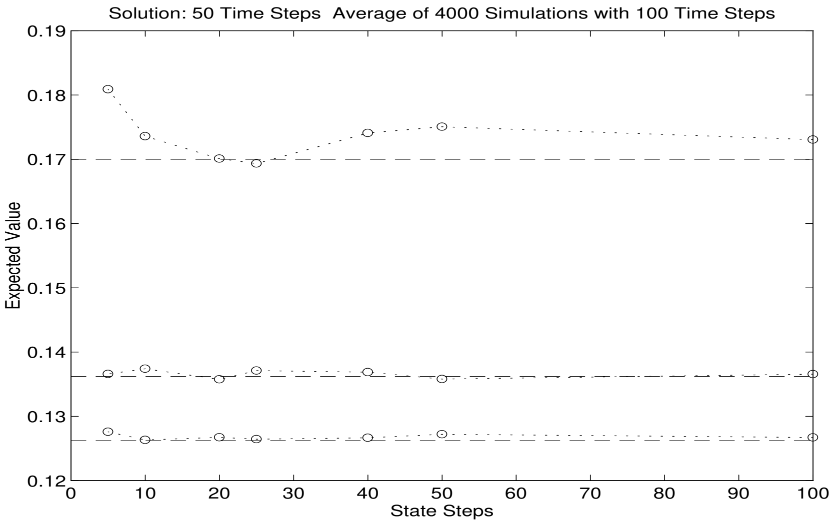

Figure 15. Simulated value as a function of state discretisation.

In our examples the approximation scheme appears to capture the behaviour of stochastic systems well. We expect that the cases where the drift component does not dominate the diffusion component can still be addressed by this method by a combination of increasing the number of solution time steps, and the number of intervals into which we partition the noise. Increasing the number of intervals we partition the noise into makes a trivial difference to the implementation of the method but could result in a large increase in computation time. Such a "better" noise discretisation should make the method capture the stochastic behaviour of the continuous-time stochastic problem better, but at the expense of longer computation time. A balance must be sought between the "accuracy" of the noise representation and the speed of the method.

A question may arise as to what impact the stochastic parameter σ has on the computation times. However, comparing solutions obtained for different σ values is not easy. Different σ values imply different grids on which the solutions are obtained. Also, the larger σ, the higher the probability of process "hitting" the state space boundary. This might actually *reduce* the computation time as the number of adjacent states will be lower. However, the accuracy of solution will suffer. To improve it, the state space should be extended resulting in a different optimisation problem. Allowing for different state-space dimensions, in our experiments (in $\mathbb{R}^1$), accuracy comparable solutions (to within 5% of the continuous-time optimal value, say) required approximately the same computation time, for $\sigma \in [.05, .3]$. This time was about 5-6 times that of the deterministic problem. These runs were done on a state grid twice the size (101 states vs 51 states) of that which was used for the deterministic runs. The three interval noise discretisation results in each state of a stochastic system taking about 3 times more of the computation time than in the deterministic case. This theoretical argument is in line with what was actually observed.

Part 2. Stochastic Renewable Resource Valuation Problem

3. THE MODEL

3.1. **The computational economics problem.** The valuation of resources in situations where uncertainty and managerial reactions to changing conditions are important factors can be handled better by contingent claims methods than by traditional methods. The latter budgeting techniques would require an estimation of future cash-flows to be made, with the uncertainty inherent in the project and managers' strategic reactions being considered only superficially. At best they are "factored" into a risk adjusted discount rate.

The problem chosen in [6] to illustrate the contingent claims approach is that of continuous-time optimal forest management. The problem consists of finding the value of a 10 year forest leasehold. The value of the lease is the value of a limited duration ("American") option on the timber in the forest. In [6], this problem is modelled as a stochastic optimisation problem where the solution is obtained through discretisation of the Hamilton-Jacobi-Bellman equation. We will use our approximation method developed in Part 1 to solve that lease problem.

3.2. **The mathematical model.** We proceed with the model as presented in [6]. The value of a forestry leasehold is to be determined when both price and inventory are stochastic. Inventory

$$\mathbf{I} = \{\mathbf{I}(t) \in X_{\mathbf{I}} \subset \mathbb{R}^1,\ t \geq 0,\ \mathbf{I}(0) = I_0 \ - \text{ given}\},$$

and prices

$$\mathbf{P} = \{\mathbf{P}(t) \in X_{\mathbf{P}} \subset \mathbb{R}^1,\ t \geq 0,\ \mathbf{P}(0) = P_0 \ - \text{ given}\}$$

are stochastic processes modelled according to the stochastic differential equations

$$d\mathbf{P}(t) = (r-\kappa)\mathbf{P}(t)dt + \sigma_{\mathbf{P}}\mathbf{P}(t)d\mathbf{W}_{\mathbf{P}} \tag{42}$$

$$d\mathbf{I}(t) = (\gamma I - \mathbf{u}(\mathbf{P},\mathbf{I},t))dt + \sigma_{\mathbf{I}}\mathbf{I}(t)d\mathbf{W}_{\mathbf{I}} \tag{43}$$

where $\mathbf{W}_{\mathbf{P}}$ and $\mathbf{W}_{\mathbf{I}}$ are Wiener processes in $\mathbb{R}$ with correlation coefficient ρ; r is the risk free interest rate, κ is the convenience yield, $\sigma_{\mathbf{P}}$ and $\sigma_{\mathbf{I}}$ are instantaneous diffusion coefficients, γ is inventory growth rate and $\mathbf{u}(\mathbf{P},\mathbf{I},t)$ is the cutting rate policy depending on the observed prices and the state of inventory. Equations (42) and (43) model the price and inventory processes. In particular, (42) tells us that the price variations become larger, the higher the price level is. Regarding inventory, (43) is a proportional growth equation with the stochastic variations being proportional to the forest stock.

We seek to find the cutting policy, $\mathbf{u}(\mathbf{P},\mathbf{I},t)$ that will maximise the functional

$$J(t,\mathbf{P}_0,\mathbf{I}_0;\mathbf{u}) = \mathbb{E}\int_0^T e^{-(r+\lambda)t}\Pi(\mathbf{u},t)dt \tag{44}$$

$$\Pi(\mathbf{u},t) = (1-T_C)[(1-T_R)\mathbf{P}(t)\mathbf{u}(\mathbf{P},\mathbf{I},t) - A(\mathbf{u},t)] \tag{45}$$

where T_R is the royalty rate on timber sales, T_C is the corporate tax rate, λ is the property tax rate and $A(\mathbf{u}, t)$ is the cost of logging. Next we formulate the optimal return function, $\mathcal{H}(t, \mathbf{P}, \mathbf{I})$, in the standard way

$$\mathcal{H}(\tau, \mathbf{P}, \mathbf{I}) = \max_{\mathbf{u}} \mathbb{E} \int_{\tau}^{T} e^{-(r+\lambda)t} \Pi(\mathbf{u}, t) dt. \tag{46}$$

The optimal return function, $\mathcal{H}(\tau, \mathbf{P}, \mathbf{I})$, satisfies the HJB equation for stochastic optimal control problems (compare (5) - (7))

$$\begin{aligned} -\frac{\partial \mathcal{H}}{\partial \tau} &= \max_{\mathbf{u}} \left(e^{-(r+\lambda)\tau} \Pi(\mathbf{u}, \tau) + (r - \kappa) \frac{\partial \mathcal{H}}{\partial \mathbf{P}} + (\gamma \mathbf{I}(\tau) - \mathbf{u}(\mathbf{P}, \mathbf{I}, \tau)) \frac{\partial \mathcal{H}}{\partial \mathbf{I}} \right. \\ &+ \left. \sigma_{\mathbf{P}}^2 \mathbf{P}^2(\tau) \frac{1}{2} \frac{\partial^2 \mathcal{H}}{\partial \mathbf{P}^2} + \sigma_{\mathbf{I}}^2 \mathbf{I}^2(\tau) \frac{1}{2} \frac{\partial^2 \mathcal{H}}{\partial \mathbf{I}^2} + \rho \sigma_{\mathbf{P}} \sigma_{\mathbf{I}} \mathbf{P}(\tau) \mathbf{I}(\tau) \frac{\partial^2 \mathcal{H}}{\partial \mathbf{P} \partial \mathbf{I}} \right). \end{aligned} \tag{47}$$

We can give $\mathcal{H}(\tau, \mathbf{P}(\tau), \mathbf{I}(\tau))$ an economic interpretation of the present value of the firm at time τ in state $(\mathbf{P}(\tau), \mathbf{I}(\tau))$.

3.3. **The optimisation model equivalence.** We will now show that our optimisation problem: max (44) subject to (42) and (43) is equivalent to the model of §3 of [6], and is explicit in showing what performance criterion for this problem is, only implicitly dealt with in [6].

The benefits of this approach are two-fold:

1. having an explicit performance criterion it is easier for us to calculate the value of sub-optimal policies as well as optimal policies;
2. having an explicit formulation of the continuous-time optimisation problem we can analyse its discrete-time approximations.

Given the interpretation of $\mathcal{H}$ as the present value of the firm it is natural to consider making the substitution

$$\mathcal{H}(t, \mathbf{P}, \mathbf{I}) = e^{-(r+\lambda)t} V(t, \mathbf{P}, \mathbf{I})$$

to get the partial differential equation governing the evolution of the nominal (*i.e.* current) value of the firm, $V(t, \mathbf{P}, \mathbf{I})$. Following the substitution the Bellman equation becomes

$$\begin{aligned} &-(e^{-(r+\lambda)t} \frac{\partial V}{\partial t} - (r + \lambda) e^{-(r+\lambda)t} V) \\ &= \max_{\mathbf{u}} \left(e^{-(r+\lambda)t} \Pi(\mathbf{u}, t) + (r - \kappa) e^{-(r+\lambda)t} \frac{\partial V}{\partial \mathbf{P}} + (\gamma I - \mathbf{u}(\mathbf{P}, \mathbf{I}, t)) e^{-(r+\lambda)t} \frac{\partial V}{\partial \mathbf{I}} \right. \\ &\left. + \frac{1}{2} \sigma_{\mathbf{P}}^2 \mathbf{P}^2 e^{-(r+\lambda)t} \frac{\partial^2 V}{\partial \mathbf{P}^2} + \frac{1}{2} \sigma_{\mathbf{I}}^2 \mathbf{I}^2 e^{-(r+\lambda)t} \frac{\partial^2 V}{\partial \mathbf{I}^2} \right) \end{aligned} \tag{48}$$

which upon simplification becomes

$$\begin{aligned} -\frac{\partial V}{\partial t} &= \max_{\mathbf{u}} \left(\Pi(\mathbf{u}, t) - \lambda V - rV + (r - \kappa) \frac{\partial V}{\partial \mathbf{P}} + (\gamma I - \mathbf{u}(\mathbf{P}, \mathbf{I}, t)) \frac{\partial V}{\partial \mathbf{I}} \right. \\ &+ \left. \frac{1}{2} \sigma_{\mathbf{P}}^2 \mathbf{P}^2 \frac{\partial^2 V}{\partial \mathbf{P}^2} + \frac{1}{2} \sigma_{\mathbf{I}}^2 \mathbf{I}^2 \frac{\partial^2 V}{\partial \mathbf{I}^2} \right) \end{aligned} \tag{49}$$

parameter	value	unit	name	comment
r	.0122		RiskFreeRate	$\ln(1+.05)/4$
γ	0.421428	% / quarter	InventoryDrift	$\ln(1+0.017)/4$; μ in [6]
$\sigma_{\mathbf{P}}$	2.473498696373	% / quarter	PriceProportionalSD	$\ln(1+0.104)/4$
$\sigma_{\mathbf{I}}$	2.223155229860	% / quarter	InvenProportionalSD	$\ln(1+0.093)/4$
$r-\kappa$	4.6	% / quarter	$r-\kappa$ is PriceDrift	$\ln(1+.004)/4$
λ	2	%	property tax rate	
T_c	33	%	corporate tax rate	
T_r	5	%	royalty rate	
a_0	25000	\$ / quarter	cost function parameter	
a_1	28	\$ / 000 m^3	cost function parameter	
a_2	2	\$ / (000 $\mathrm{m}^3)^2$	cost function parameter	
$\mathbf{I}(0)$	150	000 m^3	initial inventory	
$\mathbf{P}(0)$	50	\$	initial price	
T	40	quaters	horizon	

TABLE I. Forest lease problem parameters.

This is identical to equation (21) in [6]. Thus solution of our stochastic optimal control problem should be equivalent to the solution of the forestry valuation model presented in §3 of [6].

Following [6], we define the operation cost of logging quantity $\mathbf{u}$ as

$$A(\mathbf{u},t) = \begin{cases} a_0 + a_1\mathbf{u} + \frac{1}{2}\mathbf{u}^2 & \mathbf{u} > 0 \\ 0 & \mathbf{u} = 0 \end{cases} \tag{50}$$

and assign the following values to the problem parameters, see Table I.

4. THE DISCRETE-TIME APPROXIMATION

In Part 1, we showed how to find a solution to a continuous-time stochastic optimal control problem via a Markov chain approximation. Here, we will apply our method to the stochastic resource valuation problem max (45) in policy $\mathbf{u}(t,\mathbf{P},\mathbf{I})$, subject to (42), (43).

4.1. **Boundary Conditions and Control Constraints.** Boundary conditions for the discrete state space are established in Section 1.2.2 of Part 1. It is clear that something must be done to prevent harvesting of non-existent forest. This was accomplished via constraints on the control

$$0 \le \mathbf{u}(t) \le \mathbf{I}(t). \tag{51}$$

Such *point* constraints of the control are easily allowed for in the numerical procedure through the use of a constrained optimisation procedure.

4.2. **The deterministic case.** We first found a solution to a deterministic problem defined as max (45) in policy $\mathbf{u}(t, \mathbf{P}, \mathbf{I})$, subject to (42), (43) and (51) with all stochastic variables set at mean values and zeroed variances.

The two different time paths of the control and state variables are shown in Figure 16 for three time steps[15]. The main feature of the solution [6]: "cut now rather than later" is visible for each of them. This is not surprising as the price drift is "small" compared to the discount factor, for this basic run.

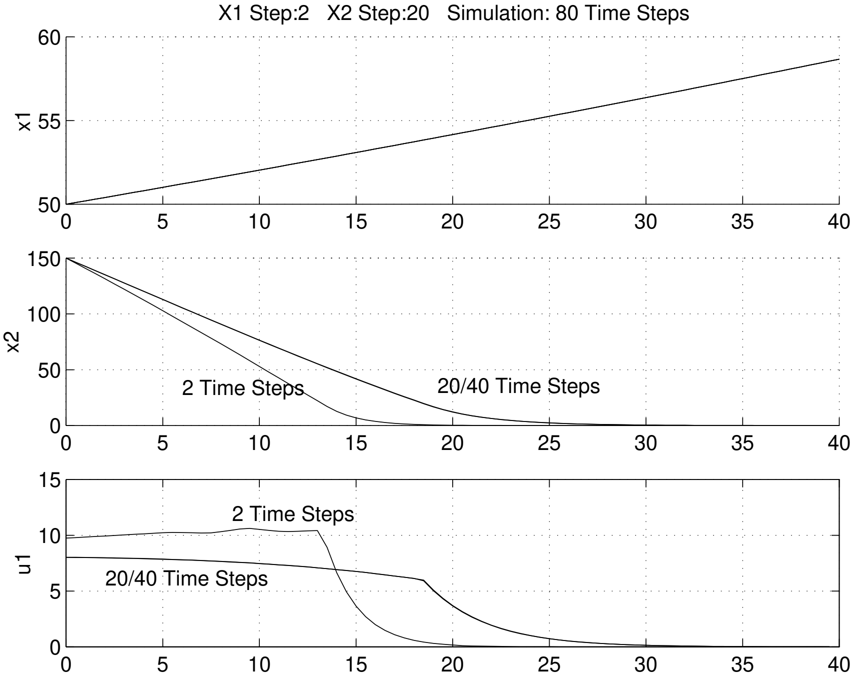

FIGURE 16. Optimal deterministic paths.

The corresponding forest's values are different for either of the control paths and are \$460,272 for 2 `TimeSteps` and \$744,308 for 20 `TimeSteps` (\$744,405 for 40 `TimeSteps`), respectively. ¿From here, and from Figure 17, we conclcude that the state space and time discretisation: `PriceStep=2`, `InventoryStep=20` (*i.e.* 100 grid nodes), `TimeStep=2` are giving us (for 80 simulation time steps) a reasonable approximation of the deterministic value of the optimally managed forest. An apparent worsening of the optimal value for finer state grids (see right pannel of Figure 17) is attributed to the decrease of the integration error.

Consider an effect the action of augmenting the price drift (ten times) has got on the state and control time paths. Keeping all the other parameters fixed we obtain a path which indicates that

---

[15] However, there is practically no difference between the 20 `TimeStep` and 40 `TimeStep` runs.

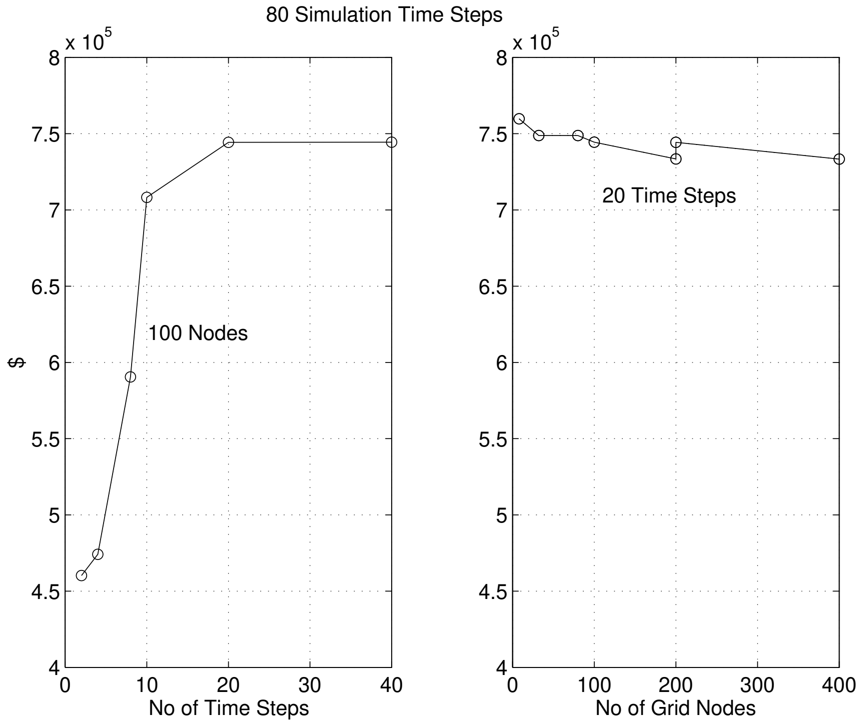

Figure 17. Optimal value as a function of discretisation grids.

it is optimal to let the forest grow and cut when the price is high, see Figure 18.[16] The optimal policy graphs are presented in Figures 22 and 23, together with the stochastic relationships.

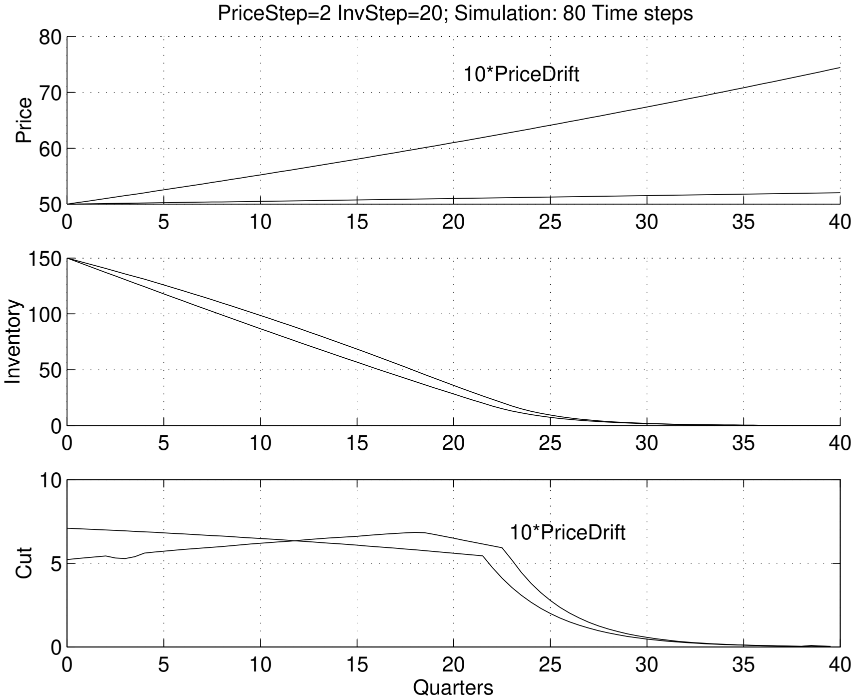

Figure 18. Determistic optimal paths for different price drifts.

---

[16] We did more experiments (not reported here) like that; each time the model results were economically explainable. This is gives some degree of certainty that the model captures the basic forest economics well.

4.3. **The stochastic case.** The stochastic resource valuation problem consists of finding a net present value of an optimally managed process of the resource utilisation in an uncertain environment. This means that we have to solve a SOC problem and estimate the Bellman function value at the intial value of the resource (and the remaining state variables).

We have solved the stochastic resource valuation problem max (45) in policy $\mathbf{u}(t, \mathbf{P}, \mathbf{I})$, subject to (42), (43) and (51). Figures 19 and 20 provide some evidence for us to assess the forest value as \$520,920 (\$545,600 in [6]): the former shows the convergence of the value averaged over 300 noise realisations; the latter shows the distribution of the value realisations.

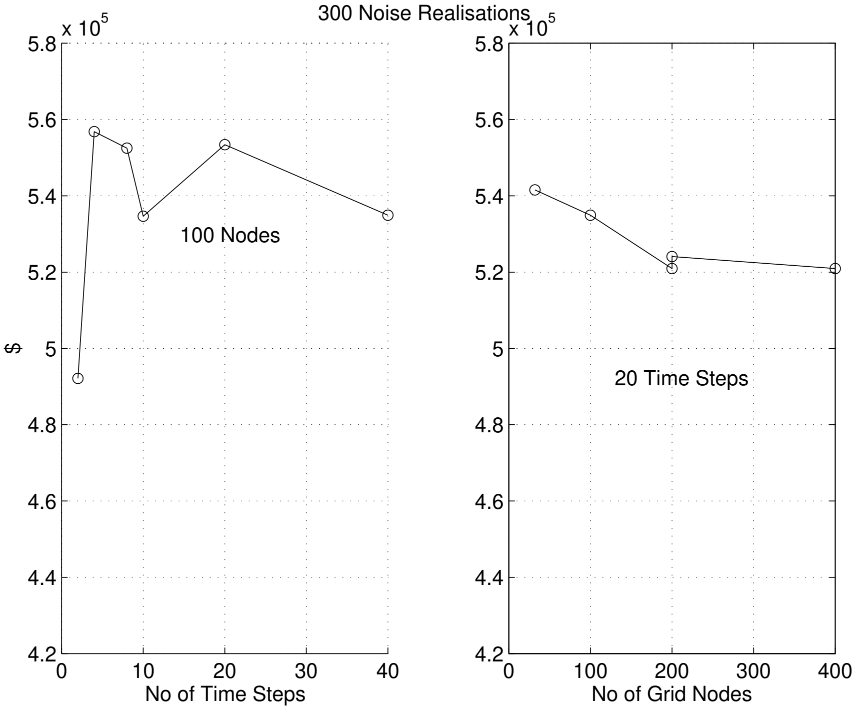

FIGURE 19. Stochastic optimal value as a function of discretisation grids.

A "by-product" of the optimal valuation are the sample paths of the optimal cutting rate, prices and inventory, see 21. These were obtained using the policy rules shown in Figures 22 and 23. The graphs show how the policy evolves as the discretision grids become denser. Notice that the second "smooth" lines were used for the paths in Figure 21. The policy graphs are important for the forest managers, to "control" the forest *on-line*.

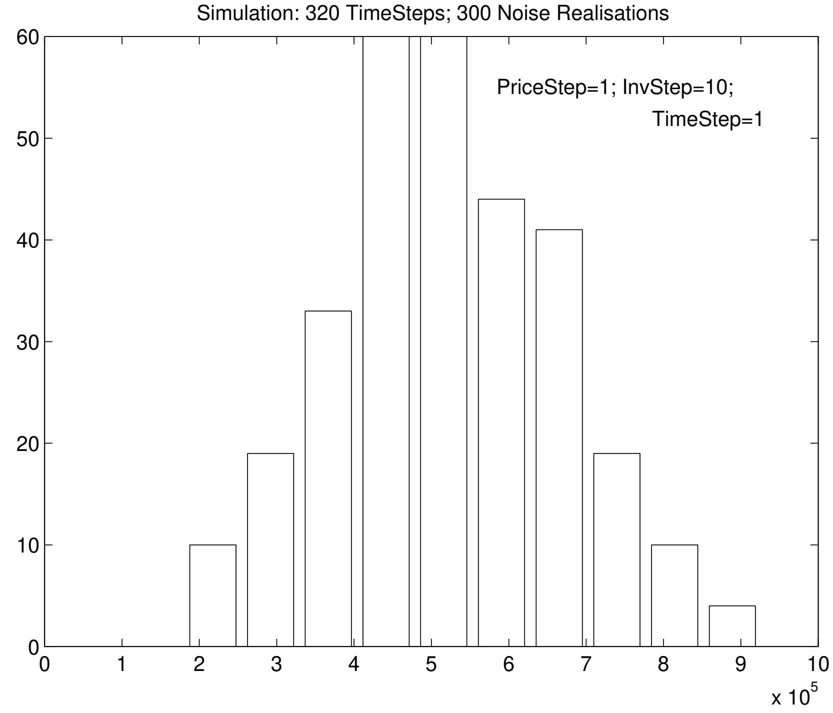

Figure 20. Distribution of value realisations.

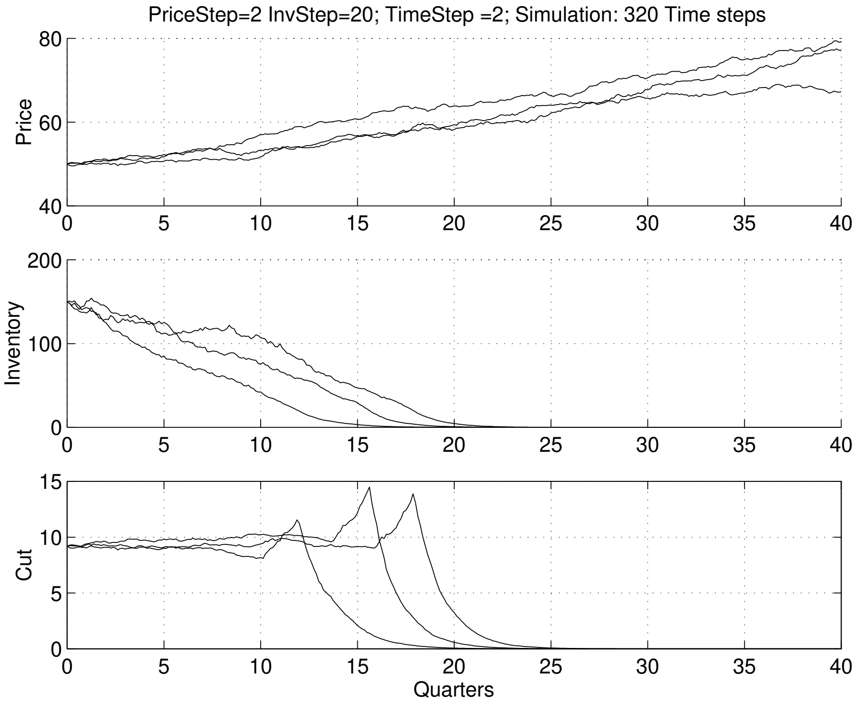

Figure 21. Optimal stochastic sample paths.

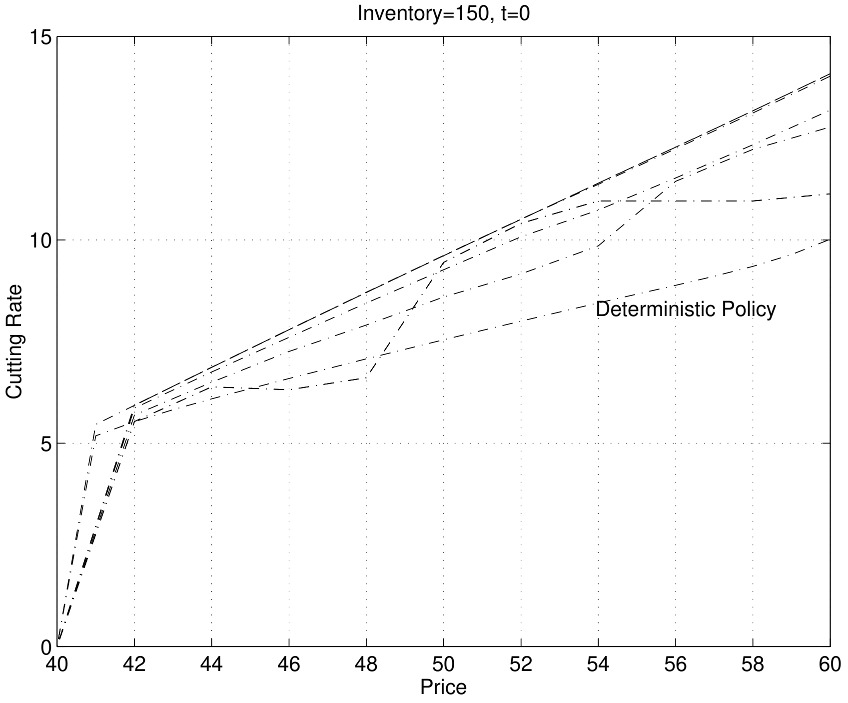

Figure 22. Optimal policy versus price.

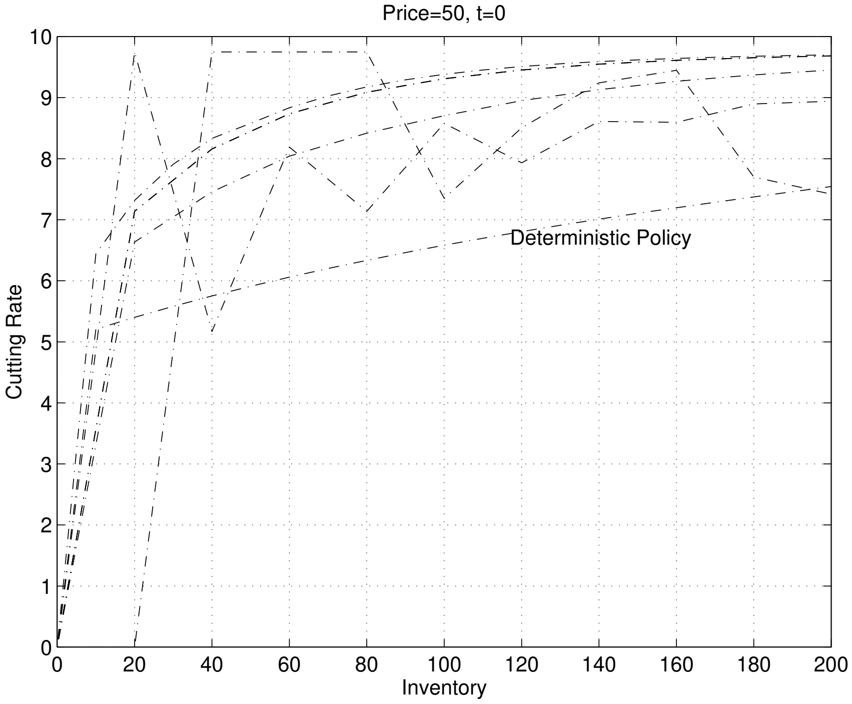

Figure 23. Optimal policy versus inventory.

4.4. **Forest lease economics.** We computed policy rules for different planning horizons (20, 40, 80, 100) and conluded that they were almost indistinguishable (not shown). The corresponding value realisations were all of the same order. We can therefore conjecture that the solution that we obtained for the planning horizon of 40 quartes is near stationary. Figure 24 shows sample paths for the different horizon lengths.

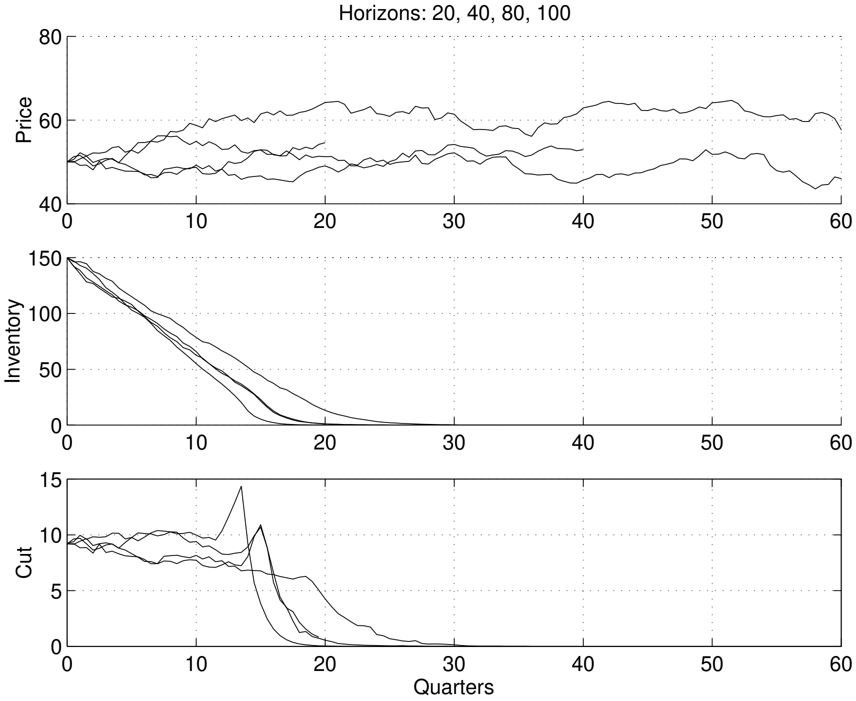

FIGURE 24. Sample paths for different horizons.

The solution procedure of the stochastic resource valuation problem can be repeated for different "scenarios" of prices, uncertainties, parameters etc. This will provide a decision maker with the resource valuation for each such scenario.

For example, we can see under what conditions logging for the whole period T could be economical (and feasible). We expect this to happen for a low discount rate and a high growth rate. Indeed, for $\lambda = .002$, $r = .005$ and $\gamma = .042$ the cutting rate and states' realisations[17] are as in Figure 25. They are results of the new optimal policy, see Figure 26, which is more conservative than the "optimal" in Figure 22, regarding the advice on how much to cut at what price. It is evident from figures that for the modified parameter set, it is optimal to cut the forest less intensively at the beginning (than for the basic parameter set, compare Figure 21) allowing it to grow and sell at higher prices.

---

[17] Actually, we have also modified a few of the remaining parameters to make the sample paths disctinctly different from the basic runs. We increased the price drift to 0.025, $\sigma_{\mathbf{P}} = 0.0495$, $\sigma_{\mathbf{I}} = .01$ and $a_2 = 3$.

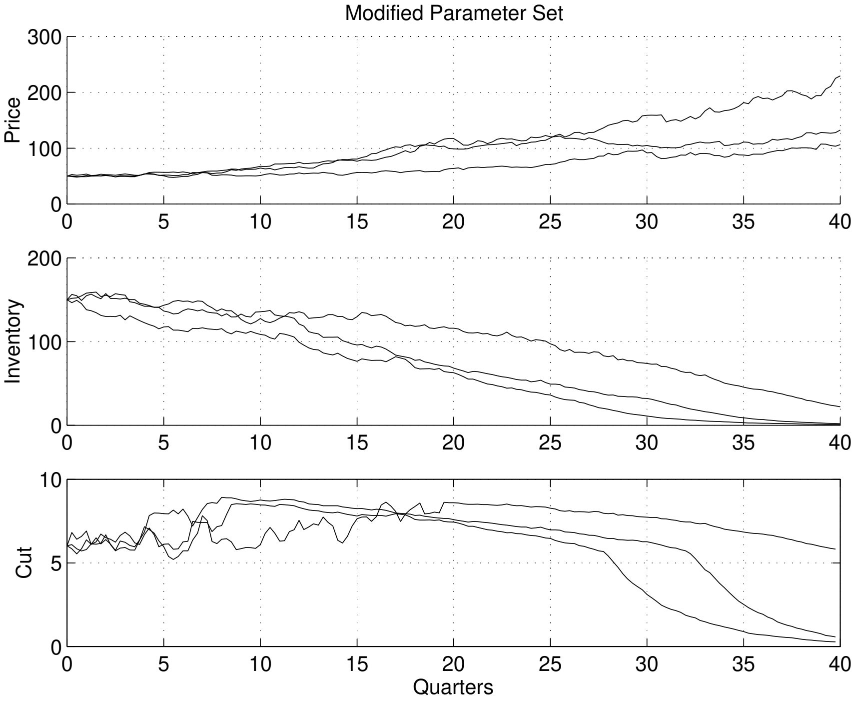

FIGURE 25. Sample paths for modified parameters.

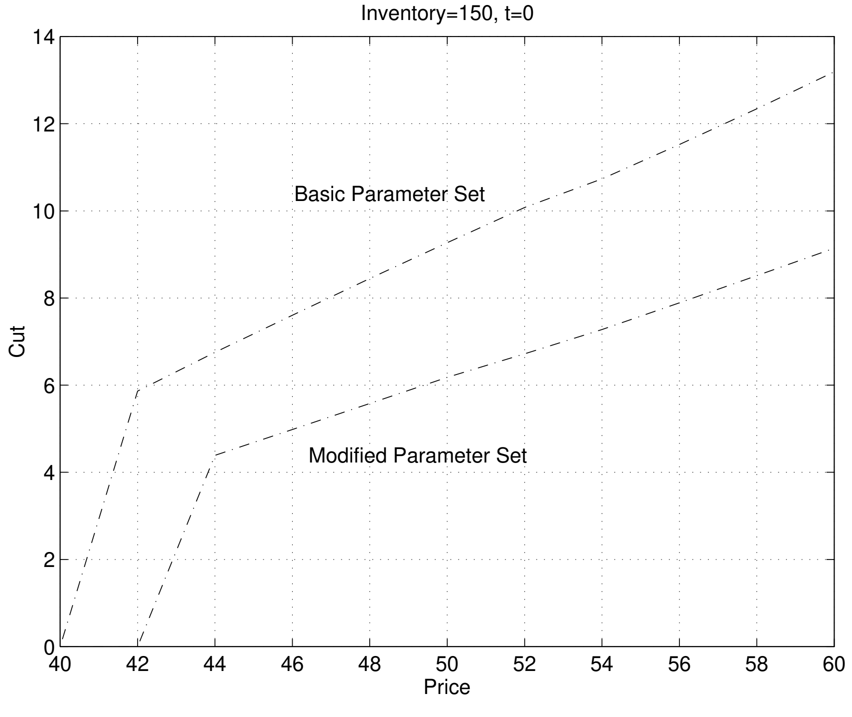

FIGURE 26. Optimal policy for the modified parameter set.

5. CONCLUSION

We have proposed a new discretisation method usful for Markovian approximation of continous-time stochastic optimal-control problems. We developed an algorithm optimising a Markov decision chain thus obtained. We have successfully applied the algorithm to a range of problems, including the stochastic resource valuation problem.

REFERENCES

[1] FLEMING, W. H., R. W. RISHEL, 1975, **Deterministic and Stochastic Optimal Control**, Springer-Verlag, New York *etc.*

[2] HAURIE, A., J. B. KRAWCZYK, M. ROCHE, 1994, "Monitoring Cooperative Equilibria in a Stochastic Differential Game", *Journal of Optimization Theory & Applications*, Vol. 81, No 1 (April), pp. 73-95.

[3] KLOEDEN, P. E., E. PLATEN, 1992, **Numerical Solution of Stochastic Differential Equations**, Springer-Verlag, Berlin *etc.*

[4] H.J. KUSHNER, 1977, **Probability Methods for Approximation in Stochastic Control and for Elliptic Equations**, Academic Press, New York.

[5] H.J. KUSHNER, 1990, "Numerical Methods for Stochastic Control Problems in Continuous Time", *SIAM J. Control & Optimization*, Vol. 28, No. 5, pp. 999-1048.

[6] MORCK, R., E. SCHWARZ & D. STANGELAND, 1989, "The Valuation of Forestry Resources under Stochastic Prices and Inventories", *Journal of Financial and Quantitative Analysis*, Vol. 24, No. 4, pp. 473-487.

[7] PEREIRA, M.V.F. & L.M.V.G PINTO , 1991, "Multi-stage Stochastic Optimization Applied to Energy Planning", *Mathematical Programming*, Series B, 52, pp. 359-375.

[8] RUST, J, 1997, "Using Randomization to Break the Curse of Dimensionality", *Econometrica*, Vol. 65, No. 3 (May), pp. 487-516.

[9] RUST, J, 1997, A Comparison of Policy Iteration Methods for Solving Continuous-State, Infinite-Horizon Markovian Decision Problems Using Random, Quasi-random, and Deterministic Discretizations, Economic Working Papers Archive, comp/9704001.

[10] TOLWINSKI, B. & R. UNDERWOOD, 1991, An Algorithm to Estimate the Optimal Evolution of an Open Pit Mine, Working Paper MCS-91-05, Colorado School of Mines, Golden, Colorado.

[11] WINDSOR, A. & J. B. KRAWCZYK, 1997, SOC Sol-I
a Matlab Package for Approximating the Solution to a Continuous-Time Stochastic Optimal Control Problem , Working Paper, Version 1a, School of Economics and Finance, Victoria University of Wellington.