Deterministic stochastic optimal control

L.C.G. Rogers*
University of Cambridge

October 17, 2005

Abstract

This paper approaches optimal control problems for discrete-time controlled Markov processes by representing the value of the problem in a dual Lagrangian form; the value is expressed as an infimum over a family of Lagrangian martingales of an expectation of a pathwise supremum of the objective adjusted by the Lagrangian martingale term. This representation opens up the possibility of numerical methods based on Monte Carlo simulation which may be advantageous in high-dimensional problems, or in problems with complicated constraints.

Keywords: Deterministic, stochastic, dynamic programming, pathwise optimisation, dual.

AMS Subject Classifications: 90C40, 90C46, 90C39, 93E20, 93E25.

1 Introduction

The title of this paper refers to this: we intend to show that the solution of a stochastic optimal control problem can be characterised in terms of a *pathwise* optimisation. In simple terms, this means that we can randomly generate a sample path, and then solve a *deterministic* optimisation for that sample path on its own. Repeating, we can get an approximation to the solution of the problem.

This approach is in contrast to the more familiar method of trying to find the value function of the problem, and the associated optimal control; this more familiar approach requires consideration of all possible future evolutions of the process at each time that a

*Statistical Laboratory, University of Cambridge, Wilberforce Road, Cambridge CB3 0WB, UK; The author thanks participants at the Isaac Newton Institute programme Developments in Quantitative Finance 2005, and at the Cambridge Finance Seminar, for helpful discussions and comments. We thank particularly Mark Broadie, Mark Davis, Michael Dempster, Paul Glasserman, David Hodge, Stan Pliska, Jose Scheinkman, Nizar Touzi, Richard Weber, and Peter Whittle.

control choice is to be made. This method is well developed, and generally effective, but there are certainly problems (such as the optimal control of a diffusion in high dimensions) where the approach is impractical.

The approach we follow is foreshadowed by various papers in the control literature, where the relationship between deterministic and stochastic optimal control is explored. There is for example the paper of Davis & Burstein [5], where the theme of optimal control of a diffusion process is considered. The tools applied, notably the use of the stochastic flow of a 'null' solution to the optimal control problem, are strongly specific to that particular context, but the form of the solution, involving a pathwise optimsation of the original objective modified by a Lagrangian term, invites extension. Other interesting papers around this theme are by Rockafellar & Wets [10], Wets [12] and by Back & Pliska [2], who present the maximisation of some concave path functional over a family of adapted processes in terms of the maximisation of the same functional modified by a linear (Lagrangian) functional over the larger family of *measurable* processes. The linear functional is of course the gradient of the objective at the optimum, in some suitable sense.

Both of these contributions leave the representation of the Lagrangian form of the solution in quite abstract terms. By contrast, the approach to be followed in this paper derives simple and quite explicit representations which may be the basis for effective numerical techniques. This approach does not require any convexity assumptions on the objective, unlike [10],[12], [2], and the proofs are simple and completely elementary. Although our first result has the appearance of the 'Lagrangian form' of the problem studied by [10], [12], [2], the subsequent results do not.

The approach of this paper is inspired by the recent result of Rogers [11], proved independently by Haugh & Kogan [7], on Monte Carlo pricing of American options[1]. This result says the following. Given an adapted process[2] $(Z_t)_{0\le t\le T}$, the value Y_0^* at time 0 of the optimal stopping problem satisfies

$$\begin{aligned} Y_0^* &\equiv \sup_{\tau\in\mathcal{T}} EZ_\tau \\ &= \inf_{M\in\mathcal{M}_0} E\big[\sup_{0\le t\le T}(Z_t - M_t)\big], \end{aligned} \tag{1}$$

where $\mathcal{T}$ is the family of stopping times, and $\mathcal{M}_0$ is the space of uniformly-integrable martingales started at 0. The importance of this result is that it gives a way to find the value of an American option via Monte Carlo simulation; given the sample path of $Z-M$, we simply stop at the best place, without considering what might be happening on any other path, and in particular without considering what the value function might be at any time. The numerical methods presented in [11] are crude, but good enough to get upper and lower bounds in a number of interesting examples which were different by about 0.5%–2%. Andersen & Broadie [1] present a more systematic way to search out 'good' martingales, and achieve bounds that are generally better. Jamshidian [8] proposes a 'multiplicative' version of the result of [11], [7].

[1] See Davis & Karatzas [6] for a weaker partial result.

[2] ... satisfying mild integrability conditions ...

Now the optimal stopping problem is a particularly simple class of optimal control problems; *could any variant of the result (1) be used for more general stochastic control problems?* Passing to complete generality introduces a couple of major complications; the first is that the space of possible controls is no longer a two-point set, but can be very large; and the second is that the choice of controls now affects the law of the process, and there is no canonical choice. However, the main message of this paper is that we *can* extend the dual methodology that worked so well for optimal stopping problems; we present a number of different ways of doing this, and explore some numerical examples. We present results only in a discrete-time setting; there are doubtless continuous-time analogues, but we prefer to present the main ideas in the technically simplest form. Our main focus is on the development of Monte Carlo methodologies that use the main ideas of the paper to solve optimal control problems. Existing techniques for solving Hamilton-Jacobi-Bellman equations by PDE methods are reasonably satisfactory provided the problem is not too involved, but it does not take much imagination to come up with examples that are so complicated that only a simulation methodology could possibly work.

2 The problem and its solution.

Consider the problem of controlling a Markov process X taking values in some set $\mathcal{X}$ over choice of control processes a in the class $\mathcal{A}$ of adapted processes with values in some set $\mathcal{U}$ of permitted controls. The problem has finite time horizon T, and objective

$$E\Big[\sum_{j=0}^{T-1} f_j(X_j, a_j) + F(X_T)\Big]. \tag{2}$$

to be maximised over adapted $a \in \mathcal{A}$, where the controlled transitions have density $\varphi(x, y; a)$ with respect to some reference Markovian transition P^*. We write $V_j(x)$ for the value function of the problem starting from state x at time j:

$$V_j(x) = \sup_{a\in\mathcal{A}} E\Big[\sum_{r=j}^{T-1} f_r(X_r, a_r) + F(X_T) \;\Big|\; X_j = x\Big]. \tag{3}$$

We may view the effect of control as being an alteration of the law of the underlying process X. If we do this, introducing the notation

$$\Lambda_t(a) \equiv \prod_{j=0}^{t-1} \varphi(X_r, X_{r+1}; a_r), \tag{4}$$

we may recast the optimisation problem in the form

$$V_0(X_0) = \sup_{a\in\mathcal{A}} E^*\Big[\sum_{j=0}^{T-1} \Lambda_j(a) f_j(X_j, a_j) + \Lambda_T(a) F(X_T)\Big], \tag{5}$$

where the expectation is now taken with respect to the fixed reference probability P^*. The first result is the following.

Theorem 1

$$V_0(X_0) = \min_{(h_j)} E^* \left[\sup_a \left\{ \sum_{j=0}^{T-1} \Lambda_j(a)\{f_j(X_j, a_j) - \eta_{j+1} + E_j^*(\eta_{j+1})\} + \Lambda_T(a)F(X_T) \right\} \right], \quad (6)$$

where the random variables η_j are defined in terms of the functions (h_j) via

$$\eta_{j+1} \equiv h_{j+1}(X_{j+1})\varphi(X_j, X_{j+1}; a_j). \quad (7)$$

Remarks. (i) To get from the form (5) to (6), we have subtracted a martingale-difference sequence $\eta_{j+1} - E_j^*(\eta_{j+1})$ from the objective, then done a *pathwise* optimisation over the controls, taken expectations, and finally minimised over choice of the martingale difference sequence. This is formally similar to what we did at (1); as there, the martingale-difference sequence can be interpreted as a Lagrangian process to account for the adapted constraint on the controls a. Once we have included this term in the objective, we optimise pathwise, allowing ourselves to see the entire path and pick controls in an anticipative way.

(ii) As we shall see, the minimum is attained, when we take $h_j = V_j$. This fact is of little practical value, since we cannot assume that we know V - it is after all the solution we seek! Nevertheless, this result will allow us to obtain upper bounds on the value function.

(iii) The choice of reference measure can be critical in practice. We cannot expect a simulation method to work well if most of the paths simulated are quite unlike the paths of the optimally-controlled process.

Proof. The problem is to find

$$V_0(X_0) = \sup_{a \in \mathcal{A}} v_0(X_0; a)$$

where of course we define

$$v_0(X_0; a) \equiv \sup_{a \in \mathcal{A}} E^* \left[\sum_{j=0}^{T-1} \Lambda_j(a) f_j(X_j, a_j) + \Lambda_T(a)F(X_T) \right].$$

Now fixing $a \in \mathcal{A}$, for any martingale M,

$$\begin{aligned} v_0(X_0; a) &= E^* \left[\sum_{j=0}^{T-1} \Lambda_j(a) f_j(X_j, a_j) + \Lambda_T(a)F(X_T) \right] \\ &= E^* \left[\sum_{j=0}^{T-1} \Lambda_j(a)\{f_j(X_j, a_j) + \Delta M_{j+1}\} + \Lambda_T(a)F(X_T) \right] \\ &= E^* \left[\sum_{j=0}^{T-1} \Lambda_j(a)\{f_j(X_j, a_j) - \eta_{j+1} + E_j^*(\eta_{j+1})\} + \Lambda_T(a)F(X_T) \right] \end{aligned}$$

where in the final expression we have specialized somewhat by writing

$$\eta_{j+1} \equiv h_{j+1}(X_{j+1})\varphi(X_j, X_{j+1}; a_j).$$

Hence

$$\begin{aligned} V_0(X_0) &= \sup_{a\in\mathcal{A}} v_0(X_0; a) \\ &= \sup_{a\in\mathcal{A}} E^*\Big[\sum_{j=0}^{T-1} \Lambda_j(a)\{f_j(X_j, a_j) - \eta_{j+1} + E_j^*(\eta_{j+1})\} + \Lambda_T(a)F(X_T) \Big] \\ &\le E^*\Big[\sup_a \Big\{ \sum_{j=0}^{T-1} \Lambda_j(a)\{f_j(X_j, a_j) - \eta_{j+1} + E_j^*(\eta_{j+1})\} + \Lambda_T(a)F(X_T) \Big\} \Big]. \end{aligned}$$

Taking the infimum over the functions h_j, we get

$$V_0(X_0) \le \inf_{(h_j)} E^*\Big[\sup_a \Big\{ \sum_{j=0}^{T-1} \Lambda_j(a)\{f_j(X_j, a_j) - \eta_{j+1} + E_j^*(\eta_{j+1})\} + \Lambda_T(a)F(X_T) \Big\} \Big]. \quad (8)$$

In fact, there is equality in (8). To see this, we use the Bellman equation

$$\begin{aligned} V_j(x) &= \sup_a E\big[\, f_j(x,a) + V_{j+1}(X_{j+1}) \,\big|\, X_j = x, a_j = a\big] \\ &= \sup_a \big\{\, f_j(x,a) + E^*\big[V_{j+1}(X_{j+1})\varphi(x, X_{j+1}; a)\big] \,\big\} \\ &\ge f_j(x,a) + E^*\big[V_{j+1}(X_{j+1})\varphi(x, X_{j+1}; a)\big] \end{aligned} \quad (9)$$

for any a. Now we take $h_j = V_j$; the right-hand side of (8) is at most

$$\begin{aligned} RHS &\le E^*\Big[\sup_a \Big\{ \Lambda_T(a)F(X_T) + \sum_{j=0}^{T-1} \Lambda_j(a)\{V_j(X_j) - V_{j+1}(X_{j+1})\varphi(X_j, X_{j+1}; a_j)\} \Big\} \Big] \\ &= V_0(X_0) \end{aligned}$$

since $V_T = F$; the sum telescopes. Combining with (8) gives the result. ∎

The study [11] of Monte Carlo valuation of American options showed that the optimal policy was in some sense a 'minimum-variance' policy, and there is an analogue in this setting too. Writing

$$Y(X; h) \equiv \sup_a \Big\{ \sum_{j=0}^{T-1} \Lambda_j(a)\{f_j(X_j, a_j) - \eta_{j+1} + E_j^*(\eta_{j+1})\} + \Lambda_T(a)F(X_T) \Big\}$$

(where the η_j are as at (7)), Theorem 1 says that $V(X_0) = \inf_{(h_j)} E^*[\, Y(X; h)\,]$. Moreover, the infimum is attained by taking $h_j = V_j$, and in that case the proof of Theorem 1 shows that the random variable $Y(X; V)$ is almost surely constant. We therefore have the following alternative characterisation of the optimal solution.

Corollary 1 *Assuming that V_0 is non-negative*[3]*, the problem*

$$\inf_{(h_j)} E^*[\,Y(X;h)^2\,]$$

is solved by taking $h_j = V_j$.

As in the case of Jamshidian's version of the optimal stopping result, we have a multiplicative form of Theorem 1.

Theorem 2

$$V_0(X_0) \le \inf_{\eta>0} E^*\left[\ \sup_a\left\{\sum_{j=0}^{T-1}\Lambda_j(a)\,f_j(X_j,a_j)\,\frac{\eta_{j+1}}{E_j^*[\eta_{j+1}]} + \Lambda_T(a)F(X_T)\ \right\}\ \right], \tag{10}$$

where the random variables η_j are positive. Provided

$$g_j^*(X_j, X_{j+1}, a_j) \equiv V_j(X_j) - V_{j+1}(X_{j+1})\varphi(X_j, X_{j+1}; a) > 0, \tag{11}$$

the result (10) can be strengthened to the statement

$$V_0(X_0) = \min_{\eta>0} E^*\left[\ \sup_a\left\{\sum_{j=0}^{T-1}\Lambda_j(a)\,f_j(X_j,a_j)\,\frac{\eta_{j+1}}{E_j^*[\eta_{j+1}]} + \Lambda_T(a)F(X_T)\ \right\}\ \right], \tag{12}$$

with the minimising choice of η_{j+1} being $\eta_{j+1} = g_j^*(X_j, X_{j+1}, a_j)$.

REMARK. Condition (11) could be weakened to non-negativity; we simply need to change f_j to $f_j - j$, and apply the Theorem to this modified problem (whose value is $T(T-1)/2$ less than the value of the original problem).

PROOF. The proof follows similar lines to the proof of Theorem 1. Fixing $a \in \mathcal{A}$, and letting η be any strictly positive adapted process,

$$\begin{aligned} v_0(X_0;a) &= E^*\left[\ \sum_{j=0}^{T-1}\Lambda_j(a)f_j(X_j,a_j) + \Lambda_T(a)F(X_T)\ \right] \\ &= E^*\left[\ \sum_{j=0}^{T-1}\Lambda_j(a)\,f_j(X_j,a_j)\,\frac{\eta_{j+1}}{E_j^*[\eta_{j+1}]} + \Lambda_T(a)F(X_T)\ \right] \end{aligned}$$

Just as before,

$$\begin{aligned} V_0(X_0) &= \sup_{a\in\mathcal{A}} v_0(X_0;a) \\ &= \sup_{a\in\mathcal{A}} E^*\left[\ \sum_{j=0}^{T-1}\Lambda_j(a)\,f_j(X_j,a_j)\,\frac{\eta_{j+1}}{E_j^*[\eta_{j+1}]} + \Lambda_T(a)F(X_T)\ \right] \\ &\le E^*\left[\ \sup_a\left\{\sum_{j=0}^{T-1}\Lambda_j(a)\,f_j(X_j,a_j)\,\frac{\eta_{j+1}}{E_j^*[\eta_{j+1}]} + \Lambda_T(a)F(X_T)\ \right\}\ \right]. \end{aligned}$$

[3]Non-negativity is needed only because we use the reasoning $E^*Y(X;h)^2 = \mathrm{var}(Y(X;h)) + E^*(Y(X;h))^2 \ge E^*(Y(X;h))^2 \ge (\min E^*Y(X;h))^2$, and the final step is not true unless we have $E^*Y(X;h) \ge 0$ for all h.

Taking the infimum over all choices of η leads to the first statement (10).

For the second statement (12), we again use the inequality (9) of the Bellman equation; positivity of g_j^* allows us to conclude that

$$\frac{f_j(X_j, a_j)}{E_j^*[\eta_{j+1}]}\,\eta_{j+1} \le \eta_{j+1}$$

and once again the sum telescopes to $V_0(X_0)$. ∎

In both of these results, the effect of the controls is to modify the measure; if we simulate paths according to the measure P^*, then the controls applied do not affect the path of X, they simply affect the value assigned to the path. It may sometimes be more helpful to be able to allow the controls to act on the path directly, for which we need to formulate the problem slightly differently.

We shall suppose that if some control sequence $(a_j)_{j=0}^{T-1}$ is chosen, and the initial value X_0 for the process is given, then the trajectory X is determined by the relations

$$X_{j+1} = \xi(j, X_j, a_j, \varepsilon_{j+1}), \qquad (j = 0, \ldots, T-1) \tag{13}$$

where the ε_j are independent random variables with common distribution, which we could take to be uniform on $[0, 1]$ if we wish. The function ξ expresses the Markovian evolution; for example, we could have a controlled AR(1) process

$$X_{j+1} = \beta X_j + a_j + \varepsilon_{j+1} \tag{14}$$

with IID $N(0, \sigma^2)$ noises ε_j, in which case the function ξ is just linear in its arguments; or it could be that the process X was a finite-state controlled Markov chain, in which case the function ξ has a straightforward though slightly involved form. From a theoretical point of view it may be a little unusual to specify a Markov process in this way, rather than through the transition kernel, but from the point of view of simulating the paths of the process, this is *exactly* the way we think of the controlled Markov process!

Given a sequence (h_j) of functions of the Markovian state variable, we define

$$Ph_{j+1}(x, a) = E\, h_{j+1}(\xi(j, x, a, \varepsilon_{j+1})).$$

Then we have the following result.

Theorem 3

$$V_0(X_0) = \min_{(h_j)} E\left[\, \sup_a \left\{ \sum_{j=0}^{T-1} (f_j(X_j, a_j) - h_{j+1}(X_{j+1}) + Ph_{j+1}(X_j, a_j)) + F(X_T) \right\} \right], \tag{15}$$

where the X_j and a_j are related through (13). The minimum is attained by taking $h_j = V_j$.

REMARKS. The Monte Carlo approach to evaluating the right-hand side of (15) would generate a sequence of ε values, then find the optimal controls. In effect, what this means is that we

have to solve a deterministic optimisation problem along each path, where the choice of control will now affect where the path goes to, and doing this is arguably no easier than solving the Bellman equation for the original stochastic control problem. However, in situations where this deterministic control problem can be dealt with more simply, there may be value in this result.

PROOF. This follows closely the lines of the proof of Theorem 1; we leave this to the reader to check. ∎

Theorems 1 and 2 give us a way to approach a stochastic optimal control problem by Monte Carlo methods, by simulating paths repeatedly, and computing the expressions inside the expectations (6), (10). However, it is important that this optimisation, over the sequence $(a_j)_{j=0}^{T-1}$, can be done efficiently, otherwise the method will be too slow. Fortunately, it turns out that the optimisation required may be performed *recursively*, so we have a sequence of optimisation problems over the choice of only one a_j at a time.

To explain this in more detail, let us focus on the form (6). We can rewrite the expression inside the expectation on the right-hand side,

$$\sum_{j=0}^{T-1} \Lambda_j(a)\{f_j(X_j,a_j) - \eta_{j+1} + E_j^*(\eta_{j+1})\} + \Lambda_T(a)F(X_T) \tag{16}$$

$$= \sum_{j=0}^{m-1} \Lambda_j(a)\{f_j(X_j,a_j) - \eta_{j+1} + E_j^*(\eta_{j+1})\} + \Lambda_m(a)Z_m, \tag{17}$$

where

$$Z_m \equiv \sum_{j=m}^{T-1} \frac{\Lambda_j(a)}{\Lambda_m(a)}\{f_j(X_j,a_j) - \eta_{j+1} + E_j^*(\eta_{j+1})\} + \frac{\Lambda_T(a)}{\Lambda_m(a)}F(X_T)$$

contains all dependence on $a_m, \ldots, a_{T-1}$. Recursively,

$$\begin{aligned} Z_m &= f_m(X_m,a_m) - \eta_{m+1} + E_m^*(\eta_{m+1}) + \frac{\Lambda_{m+1}(a)}{\Lambda_m(a)}Z_{m+1} \\ &= f_m(X_m,a_m) + E_m^*(\eta_{m+1}) + \varphi(X_m,X_{m+1};a_m)\big[\, Z_{m+1} - h_{m+1}(X_{m+1})\,\big]. \end{aligned}$$

Assuming we have already got the maximising values of $a_{m+1}, \ldots, a_{T-1}$, this is a maximisation over a_m only!

3 Towards an algorithm.

It is clear from the statement of Theorem 1 that the choice of the Lagrangian functions (h_j) is critical. The following little result offers a possible approach to finding good choices.

Proposition 1 *Suppose that*

$$B \equiv \sup_{a,x,x'} \varphi(x,x';a) < \infty$$

and suppose given a sequence $(V_j^{(0)})_{j=0}^T$ *of functions from* $\mathcal{X}$ *to* $\mathcal{X}$*, with* $V_T^{(0)} = F$*. Define recursively the functions* $(V_k^{(n)})_{k=0}^T$ *for* $n = 1, 2, \ldots$ *by*

$$V_k^{(n+1)}(x) = E^*\Bigg[\sup_a \Bigg\{ \sum_{j=k}^{T-1} \Lambda_{k,j}(a)\{f_j(X_j, a_j) - V_{j+1}^{(n)}(X_{j+1})\varphi(X_j, X_{j+1}; a_j) + PV_{j+1}^{(n)}(X_j, a_j)\} + \Lambda_{k,T}(a)F(X_T) \Bigg\} \Bigg| X_k = x \Bigg], \tag{18}$$

for $x \in \mathcal{X}$*,* $k = 0, \ldots, T$*, where*

$$\Lambda_{k,j}(a) \equiv \prod_{r=k}^{j-1} \varphi(X_r, X_{r+1}; a_r),$$

and

$$P\psi(x, a) \equiv E^*[\varphi(x, X_1; a)\psi(X_1) \mid X_0 = x\,]. \tag{19}$$

Defining

$$\Delta_k^{(n)} \equiv \sup_x |V_k^{(n)}(x) - V_k^{(n-1)}(x)|,$$

$k = 0, \ldots, T$*,* $n \geq 1$*, we have*

$$\Delta_k^{(n)} \leq (1 + B) \sum_{r=k+1}^{T} \Delta_r^{(n-1)}. \tag{20}$$

REMARKS. (i) The result may be vacuous if the $\Delta^{(n)}$ are infinite; a sufficient condition for finiteness would be the boundedness of the f_j and F, but this is not of course necessary.

(ii) The impact of Proposition 1 lies in the fact that $V_T^{(n)} = F$ for all n, so $\Delta_T^{(n)} = 0$ for all n. Hence from (20) we conclude that (provided that the $\Delta_k^{(n-1)}$ are finite)

$$\Delta_k^{(n)} = 0 \quad \forall n \geq T - k.$$

Thus by applying the recursive construction of Proposition 1 we compute the true value function step by step back from the end. Now in one sense all we have done is to re-express the familiar backward recursion of the Bellman equation in a more complicated form, but there is nevertheless something gained; if we are not able to compute the recursive recipe (18) exactly (as would be the case where we were using Monte Carlo in a high-dimensional problem, for example), we can still use the *approximate* output of the n^{th} stage to begin on the $(n+1)^{\text{th}}$.

PROOF. Clearly,

$$\begin{aligned} -V_{j+1}^{(n)}(X_{j+1})\varphi(X_j, X_{j+1}; a_j) &\leq -V_{j+1}^{(n-1)}(X_{j+1})\varphi(X_j, X_{j+1}; a_j) + \Delta_{j+1}^{(n)}\varphi(X_j, X_{j+1}; a_j) \\ &\leq -V_{j+1}^{(n-1)}(X_{j+1})\varphi(X_j, X_{j+1}; a_j) + B\Delta_{j+1}^{(n)} \end{aligned}$$

and

$$PV_{j+1}^{(n)}(X_j, a_j) \leq PV_{j+1}^{(n-1)}(X_j, a_j) + \Delta_{j+1}^{(n)}$$

so using this in (18) gives us

$$\begin{aligned}
V_k^{(n+1)}(x) &\equiv E^*\Bigg[\sup_a \Bigg\{ \sum_{j=k}^{T-1} \Lambda_{k,j}(a)\{f_j(X_j, a_j) - V_{j+1}^{(n)}(X_{j+1})\varphi(X_j, X_{j+1}; a_j) \\
&\qquad + PV_{j+1}^{(n)}(X_j, a_j)\} + \Lambda_{k,T}(a)F(X_T) \Bigg\} \;\Bigg|\; X_k = x \Bigg] \\
&\leq E^*\Bigg[\sup_a \Bigg\{ \sum_{j=k}^{T-1} \Lambda_{k,j}(a)\{f_j(X_j, a_j) - V_{j+1}^{(n-1)}(X_{j+1})\varphi(X_j, X_{j+1}; a_j) \\
&\qquad + PV_{j+1}^{(n-1)}(X_j, a_j)\} + \Lambda_{k,T}(a)F(X_T) \Bigg\} \;\Bigg|\; X_k = x \Bigg] + (1+B) \sum_{r=k+1}^{T} \Delta_r^{(n)} \\
&= V_k^{(n)}(x) + (1+B) \sum_{r=k+1}^{T} \Delta_r^{(n)}.
\end{aligned}$$

Thus

$$V_k^{(n+1)}(x) - V_k^{(n)}(x) \leq (1+B) \sum_{r=k+1}^{T} \Delta_r^{(n)},$$

and a similar bound on the other side establishes the result.

∎

4 Infinite horizon.

So far we have been considering only finite-horizon problems, but it is at least as important to develop methods for infinite-horizon discounted problems, as these will generate time-independent strategies that are easier to interpret and implement. Throughout this section, we will assume that f is uniformly bounded, and that we aim to find the value function $V : \mathcal{X} \to \mathcal{X}$ solving

$$V(x) = \sup_a E^*\Big[f(x, a) + \beta\varphi(x, X_1; a)V(X_1) \;\Big|\; X_0 = x \Big]. \tag{21}$$

Under the assumptions that $0 < \beta < 1$ and that f is uniformly bounded, it is well known that the Bellman operator $\mathcal{L}_B : L^\infty(\mathcal{X}) \to L^\infty(\mathcal{X})$ defined by

$$\mathcal{L}_B g(x) \equiv \sup_{a \in \mathcal{A}} E^*\Big[f(x, a) + \beta\varphi(x, X_1; a)g(X_1) \;\Big|\; X_0 = x \Big] \tag{22}$$

is a monotone contraction with unique fixed point the value function V solving (21).

To see where the dual method leads in this infinite-horizon setting, we need to introduce for each $h \in L^\infty(\mathcal{X})$ the operator $\mathcal{L}_h : L^\infty(\mathcal{X}) \to L^\infty(\mathcal{X})$ defined by

$$\mathcal{L}_h g(x) \equiv E^*\left[\sup_a \big\{ f(x,a) - h(X_1)\varphi(x,X_1;a) + Ph(x,a) + \beta\varphi(x,X_1;a)g(X_1)\big\} \;\Big|\; X_0 = x \right]. \tag{23}$$

Just as for $\mathcal{L}_B$, the operator $\mathcal{L}_h$ is a monotone contraction with a unique fixed point, which we denote by g_h^*. The analogue of Theorem 1 for the infinite-horizon setting is the following.

Theorem 4 *Assuming that f is uniformly bounded, the value function V is characterised as*

$$V = \inf_h \; g_h^* = \min_h \; g_h^*, \tag{24}$$

where the infimum is attained by taking $h = \beta V$.

Proof. Evidently, the supremum in the definition of $\mathcal{L}_h g$ will be reduced if we insist that a must be a function only of X_0 and not of X_1; therefore

$$\begin{aligned}
\mathcal{L}_h g(x) &\geq \sup_a E^*\left[f(x,a) - h(X_1)\varphi(x,X_1;a) + Ph(x,a) + \beta\varphi(x,X_1;a)g(X_1) \;\Big|\; X_0 = x \right] \\
&= \sup_a E^*\left[f(x,a) + \beta\varphi(x,X_1;a)g(X_1) \;\Big|\; X_0 = x \right] \\
&\equiv \mathcal{L}_B g(x).
\end{aligned}$$

Since $\mathcal{L}_B V = V$, we deduce immediately that whatever h we shall have $\mathcal{L}_h V \geq V$, and by induction we conclude that for all n,

$$\mathcal{L}_h^n V \geq V.$$

By the Contraction Mapping Principle, $\mathcal{L}_h^n V \to g_h^*$ as $n \to \infty$, and so we have for any h

$$g_h^* \geq V,$$

hence $V \leq \inf_h g_h^*$.

To conclude, we observe that taking $h = \beta V$ gives for any x, a

$$f(x,a) + Ph(x,a) \leq \sup_{a'} \{ f(x,a') + Ph(x,a') \} = V(x).$$

Hence,

$$\begin{aligned}
\mathcal{L}_h V(x) &\equiv E^*\left[\sup_a \big\{ f(x,a) - h(X_1)\varphi(x,X_1;a) + Ph(x,a) + \beta\varphi(x,X_1;a)V(X_1)\big\} \;\Big|\; X_0 = x \right] \\
&\leq V(x) + E^*\left[\sup_a \big\{ -h(X_1)\varphi(x,X_1;a) + \beta\varphi(x,X_1;a)V(X_1)\big\} \;\Big|\; X_0 = x \right] \\
&= V(x).
\end{aligned}$$

By induction, $\mathcal{L}_h^n V \leq V$, and so taking the limit as $n \to \infty$ leads to the conclusion that $g_h^* \leq V$. ∎

As in the finite-horizon case, we can ask about possible recursive methods for generating a better approximation to the solution from an existing one. The following result, proved only under rather restrictive conditions, shows that something can be done.

Proposition 2 *Suppose that f is uniformly bounded, and that*

$$B \equiv \sup_{x,x',a} \varphi(x,x';a) < \infty,$$

and that β is so small that

$$\frac{\beta(1+B)}{1-\beta B} < 1.$$

Then the sequence $(g_n)_{n=0}^{\infty}$ generated by taking an arbitrary $g_0 \in L^\infty(\mathcal{X})$ and letting g_{n+1} be the unique fixed point of $\mathcal{L}_{\beta g_n}$ converges to the value function.

Proof. The relation linking g_{n+1} and g_n can be expressed as

$$\begin{aligned} g_{n+1}(x) &= E^*\Bigg[\sup_a\big\{f(x,a) - \beta g_n(X_1)(X_1)\varphi(x,X_1;a) + \beta P g_n(x,a) \\ &\qquad +\beta\varphi(x,X_1;a)g_{n+1}(X_1)\big\} \;\Bigg|\; X_0 = x\Bigg]. \end{aligned}$$

If we set $\Delta_n \equiv \sup_x |g_n(x) - g_{n-1}(x)|$, then this leads to

$$\begin{aligned} g_{n+1}(x) &\leq E^*\Bigg[\sup_a\big\{f(x,a) - \beta g_{n-1}(X_1)(X_1)\varphi(x,X_1;a) + \beta P g_{n-1}(x,a) \\ &\qquad +\beta(1+B)\Delta_n + \beta\varphi(x,X_1;a)g_{n+1}(X_1)\big\} \;\Bigg|\; X_0 = x\Bigg], \end{aligned}$$

so if we set $\tilde{g}_{n+1} \equiv g_{n+1} + A$, we have

$$\begin{aligned} \tilde{g}_{n+1}(x) + A &\leq E^*\Bigg[\sup_a\big\{f(x,a) - \beta g_{n-1}(X_1)(X_1)\varphi(x,X_1;a) + \beta P g_{n-1}(x,a) \\ &\qquad +\beta(1+B)\Delta_n + \beta\varphi(x,X_1;a)(\tilde{g}_{n+1}(X_1) + A)\big\} \;\Bigg|\; X_0 = x\Bigg] \\ &\leq E^*\Bigg[\sup_a\big\{f(x,a) - \beta g_{n-1}(X_1)(X_1)\varphi(x,X_1;a) + \beta P g_{n-1}(x,a) \\ &\qquad +\beta(1+B)\Delta_n + \beta BA + \beta\varphi(x,X_1;a)\tilde{g}_{n+1}(X_1)\big\} \;\Bigg|\; X_0 = x\Bigg] \end{aligned}$$

Taking

$$A \equiv \frac{\beta(1+B)\Delta_n}{1-\beta B}$$

gives us

$$\tilde{g}_{n+1}(x) \;\leq\; E^*\Big[\sup_a\big\{f(x,a) - \beta g_{n-1}(X_1)(X_1)\varphi(x,X_1;a) + \beta P g_{n-1}(x,a) \\ +\beta\varphi(x,X_1;a)\tilde{g}_{n+1}(X_1)\big\} \;\Big|\; X_0 = x\Big],$$

from which we conclude that $\tilde{g}_{n+1} \equiv g_{n+1} - A \leq g_n$. A similar argument for the lower bound gives

$$\Delta_{n+1} \leq \frac{\beta(1+B)}{1-\beta B}\,\Delta_n,$$

and the result follows.

∎

Remarks. Proposition 2 shows how we may recursively construct approximations to the solution using this methodology, provided the discount factor β is small enough. The assumptions of Proposition 2 will be unlikely to be satisfied in most applications, but at least the methodology can be tried; the conditions are sufficient but not necessary!

5 Numerical issues.

We have been discussing characterisations of the optimal solution in terms of pathwise optima of an objective modified by some 'Lagrangian' term. Propositions 1 and 2 offer some template for the possible numerical solution of the problem, which can be summarised as

(i) propose some approximation (h_j) to the value;

(ii) evaluate $E^*[\sup_a \ldots]$;

(iii) improve the approximation (h_j).

It is unrealistic to suppose that in a general problem we would be able to make a good guess for V at the first stage, and so we are forced to envisage some *doubly* recursive scheme, where we perform a recursion over the approximations to V, each of which is obtained by a backward recursion in the time variable. At first sight, this appears a lot more computationally intensive than the classical dynamic programming approach, which appears to need only a backward recursion over the time variable; *when might the approach presented here be better?*

One possible answer to that question is, '*When the space $\mathcal{X}$ is very large.*' Imagine that $\mathcal{X} = \mathbb{R}^N$ for some moderately large N (30, say), and consider the familiar Bellman equations:

$$\begin{aligned} V_{n-1}(x) &= \sup_a E^*\Big[f(x,a) + \varphi(x,X_1;a)V_n(X_1) \;\Big|\; X_{n-1} = x\Big], \quad (1 \leq n \leq T) \qquad (25)\\ V_T(x) &= F(x). \end{aligned}$$

The right-hand side of (25) requires an integration over $\mathcal{X} = \mathbb{R}^N$ of the value function V_n, which presents two problems: firstly, how do we characterise V_n numerically, and secondly, how do we integrate it? Both of these are issues for the alternative approach of this paper

as well. In either approach we will try to characterise V_n numerically, by storing its value at a (necessarily sparse) finite set of points, or by approximating it in terms of a parametric family of functions (for example, as a linear combination of some suitable finite set of 'basis' functions). Now in implementing the dynamic-programming approach, the first step in the algorithm *must* be to evaluate (25) when $n = T$ at some finite set of points in $\mathcal{X} = \mathbb{R}^N$ - *how are those points to be placed?* And as we consider this question, it becomes apparent that the simple dynamic programming approach is not as simple as it appears; we clearly would like to place evaluation points *in regions where the optimally-controlled process is most likely to go,* but when we are starting out on the dynamic programming algorithm, *we do not know where these are.* It is not good enough to pick randomly from the law at time T of the process under some default control - this law may be completely different from the law of the optimally-controlled process. It is clear then that we must make some first guess at the law of the optimally-controlled process, before solving the Bellman equation and then improving our guess - in other words, we are forced to consider doubly-recursive schemes even if trying the conventional dynamic-programming approach. An obvious candidate is policy improvement; at each stage, we have a policy, and we could use the law of the process under the current policy as the reference measure to determine where to place points, as in the stochastic mesh method of Broadie & Glasserman [3]. This approach is not obviously hopeless, though as Broadie & Glasserman emphasise, the generation of the stochastic mesh needs careful handling; and many of the components of this approach are common to the alternative methodology presented in this paper.

Even assuming we have made good choices of the evaluation points, the numerical integration is still an issue. If we are storing V_n at a finite set of points, the integration will involve the calculation of weights, which will be different for each x and a, and this may be quite a time-consuming business. An alternative is to approximate V_n by some member of a parametric family of functions, and holds out more hope; if we can find 'nice' functions $\psi(\cdot, \theta)$ for which the expectation

$$E^*[\varphi(x, X_1; a)\psi(X_1, \theta)|X_0 = x] \equiv P\psi(x, a, \theta)$$

can be given explicitly, then the integration in (25) becomes a function evaluation. While it may be asking a lot for there to exist such a family of basis functions, if we are working in a large space $\mathcal{X}$ without some such strong regularity, then we will not have a chance of finding the solution; this assumption is implicit in the approach of Carriere [4][4], for example.

Let us now compare with the numerical approach that would come from the Lagrangian theory of this paper; for concreteness, we discuss the finite-horizon situation of Theorem 1. The lesson of Andersen & Broadie [1] in the context of optimal stopping is that in order to find good dual solutions it is important to consider good primal solutions at the same time. With this in mind, Proposition 1 suggests a recursive approach; in outline, the algorithm would run as follows:

STEP 0: Set $k = 0$;

STEP 1: Set reference measure $P^{(k)}$ ($= P^*$ for $k = 0$);

[4]This approach was rediscovered and popularised by Longstaff & Schwartz [9].

STEP 2: Propose $h_n^{(k)}$, approximations to $(V_n^{(k)})$;

STEP 3: Simulate N paths, and optimise pathwise - at each time n, we obtain an approximation $\hat{V}_n^{(k+1)}(X_n^{(j)})$ to $V_n^{(k+1)}$ at each of the points $X_n^{(1)}, \ldots, X_n^{(N)}$ visited by the simulated paths;

STEP 4: Regress approximate values onto basis - find some linear combination of basis functions $\psi(\cdot, \theta)$ that matches the values $\hat{V}_n^{(k+1)}(X_n^{(j)})$ at the points $X_n^{(j)}$;

STEP 5: Make up $P^{(k+1)}$. In more detail, the transitions from position x at time n will be determined by selecting a point $X_n^{(j)}$ from $\{X_n^{(1)}, \ldots, X_n^{(N)}\}$ at random, points nearer to x being chosen with higher probability, and then jumping from the chosen point according to the transition function for the action a which was optimal for the pathwise optimisation of that jth path at that time;

STEP 6: $k = k + 1$; goto Step 3.

While many of the more computationally-intensive elements of this plan are common to the numerical approach to the Bellman equation[5], the selection of the points at which the value function is to be well approximated is done naturally, and using our current best idea about where the optimally-controlled process should be.

6 A numerical example.

We shall illustrate the methodology presented by considering a controlled Markov process on the unit circle $\mathbb{T} \equiv [0, 2\pi]$, whose dynamics are given by

$$X_{t+1} = X_t + \varepsilon_{t+1} + a_t \pmod{2\pi} \tag{26}$$

and where the ε_t have density proportional to $\cos(x)$. The control a lies in $\mathbb{T}$, and the objective to be maximised is

$$\sum_{t=0}^{T} \beta^t \big[\cos(X_t) + \cos(a_t)\big]. \tag{27}$$

For the results reported, we took $T = 15$, $\beta = 0.9$, and discretised the circle into 40 equally-spaced points. Approximating the dynamics by the corresponding discretisation gave a controlled Markov process with 40 states and 40 possible actions. We then applied the algorithm outline of Section 5. The initial paths were generated using $a \equiv 0$, and our first guess at the value function was the value of using this null policy. Simulating 800 paths at each pass through the algorithm, we performed the pathwise optimisations and stored the actions used at each step along each path, as well as the values obtained. This information was then used to determine the path law for the next simulation, and the next approximation to the value. We display the results in Figure 1; the true values are displayed as continuous curves (one plotted for each value of time-to-go), and the values computed by the algorithm are superimposed

[5] in particular, the step-by-step optimisation over a, and the optimisation over θ in approximating the value functions

as crosses. As is apparent, the agreement is effectively perfect. The calculation took 21s in Scilab on a 3.4GHz Pentium 4 processor. Of course, the calculation by standard dynamic programming took only a tiny fraction of this time, but that is not the point; the point is that we have shown that at least in a very simple example the Monte Carlo implementation of the algorithm suggested by the main result of this paper actually does work correctly.

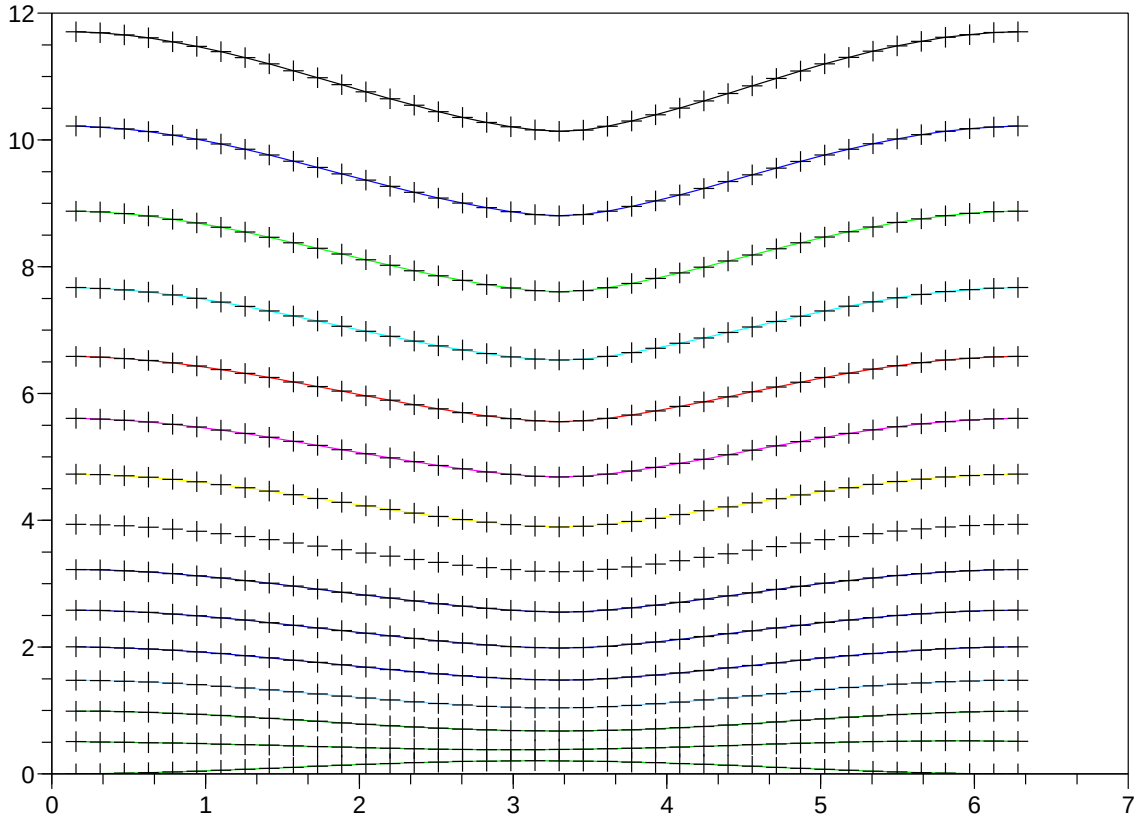

Figure 1: True value function, with MC results superimposed

7 Conclusions.

This paper has presented a novel strategy for solving stochastic optimal control problems, using duality ideas. This approach is completely general, but is particularly well suited to problems

where the statespace is so large that it is hard to determine where the value function should be approximated closely. The methodology involves modifying the objective by adding in appropriate martingale differences, and then carrying out a *pathwise* optimisation, an approach that is well suited to Monte Carlo evaluation. We have shown that under suitable regularity conditions a recursive method for improving the martingale difference sequence converges to the true solution.

Choosing the martingale difference sequence well is of course key to the success of the method, but we have shown how the characterisation of the solution leads naturally to a Monte Carlo algorithm for computing the value, and have demonstrated on a very simple example that this algorithm works correctly.

There remains much work to be done in exploring the usefulness of this approach in more challenging examples.

References

[1] L. Andersen and M. Broadie. A primal-dual simulation algorithm for pricing multi-dimensional American options. *Management Science*, 50:1222–1234, 2004.

[2] K. Back and S. R. Pliska. The shadow price of information in continuous time decision problems. *Stochastics*, 22:151–186, 1987.

[3] M. Broadie and P. Glasserman. A stochastic mesh method for pricing high-dimensional American options. *Journal of Computational Finance*, 7:35–72, 2004.

[4] J. Carriere. Valuation of early-exercise price of options using simulations and nonparametric regression. *Insurance: Mathematics and Economics*, 19:19–30, 1996.

[5] M. H. A. Davis and G. Burstein. A deterministic approach to stochastic optimal control with application to anticipative optimal control. *Stochastics and Stochastics Reports*, 40:203–256, 1992.

[6] M. H. A. Davis and I. Karatzas. A deterministic approach to optimal stopping, with applications. In F. P. Kelly, editor, *Probability, Statistics and Optimisation: a Tribute to Peter Whittle*, pages 455–466. Wiley, New York and Chichester, 1994.

[7] M. Haugh and L. Kogan. Pricing American options: A duality approach. *Operations Research*, 52:258–270, 2004.

[8] F. Jamshidian. Numeraire-invariant option pricing and American, Bermudan and trigger stream rollover. Technical report, University of Twente, 2004.

[9] F. A. Longstaff and E. A. Schwartz. Valuing American options by simulation: a simple least-squares approach. *Review of Financial Studies*, 14:113–147, 2001.

[10] R. T. Rockafellar and R. J. B. Wets. Nonanticipativity and L^1 martingales in stochastic optimization problems. *Mathematical Programming Study*, 6:170–187, 1976.

[11] L. C. G. Rogers. Monte Carlo valuation of American options. *Mathematical Finance*, 12:271–286, 2002.

[12] R. J. B. Wets. On the relation between stochastic and deterministic optimization. In *Numerical Methods and Computer Systems Modelling*, volume 107 of *Lecture Notes in Economics and Mathematical Systems*, pages 350–361, Berlin, 1975. Springer.