Pathwise stochastic optimal control

L.C.G. Rogers*
University of Cambridge

August 15, 2006

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, in particular: 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 [4], 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 [11], Wets [13] 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 [11],[13], [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 [11], [13], [2], the subsequent results do not.

The approach of this paper develops the recent result of Rogers [12], proved independently by Haugh & Kogan [6], 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 [12] 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 [7] proposes a 'multiplicative' version of the result of [12], [6].

[1] See Davis & Karatzas [5] 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 forms of the main idea. 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. The different forms of the main result that we derive suggest different techniques for approaching the problem of Monte Carlo approximation of the solution. There are also links to the 'occupation measure' approach to optimal control of a Markov process (which Kurtz & Stockbridge [8] trace back to Manne [10]); this we discuss in an appendix.

2 The problem and its solution.

We shall consider the optimal control of a discrete-time Markov process with a finite time horizon T. The Markov process X takes values in some measurable space $(\mathcal{X}, \mathcal{G})$, and the control process $\mathbf{a} \equiv (a_0, a_1, \ldots, a_{T-1})$ belongs to the class $\mathcal{A}$ of adapted processes with values in some measurable space $(A, \mathcal{B})$ of permitted controls. The objective is

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

to be maximised over $\mathbf{a} \in \mathcal{A}$. For simplicity, we shall make the assumption that the functions f_j and F are *bounded* measurable, to avoid having to worry over finiteness of objectives and other such inessential issues; this restriction is made solely for ease of exposition. We shall suppose that there is some reference measure m over $(\mathcal{X}, \mathcal{G})$ such that for each $a \in A$ the transition under control a has density $p(x, x'; a)$ with respect to m, and that there is some reference Markovian transition density $p^*(x, x')$. We write

$$\varphi(x, x'; a) = \frac{p(x, x'; a)}{p^*(x, x')}$$

for the controlled transition density with respect to the reference Markovian transition p^*. We write $V_j(x)$ for the value function of the problem starting from state x at time j:

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

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 $(0 \leq k \leq t < T)$

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

we may recast the optimisation problem in the form

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

where the expectation is now taken with respect to the fixed reference probability P^*. We shall have need of the notations (for (bounded) measurable $g, h_j : \mathcal{X} \mapsto \mathbb{R}$) :

$$\begin{aligned} Pg(x,a) &= E^*[\, g(X_1)\varphi(x, X_1; a) \mid X_0 = x], \quad (x \in \mathcal{X},\ a \in A), & (6)\\ (\mathcal{L}h)_j(x) &= \sup_a \big[\, f_j(x,a) + Ph_{j+1}(x,a)\,\big], \quad (x \in \mathcal{X}, j = 0, \ldots, T-1). & (7) \end{aligned}$$

The first notation is just the expectation of $g(X_1)$ if at time 0 we are in state x and use action a; the second defines the one-step Bellman operator.

The first result is the following.

Theorem 1

$$\begin{aligned} V_0(X_0) &= \min_{(h_j)\in\mathcal{H}} E^* \Big[\sup_{\mathbf{a}} \Big\{ \sum_{j=0}^{T-1} \Lambda_j(\mathbf{a}) \{ f_j(X_j, a_j) + E_j^*(\eta_{j+1}) - \eta_{j+1} \} + \Lambda_T(\mathbf{a}) F(X_T) \Big\} \Big] & (8)\\ &= \min_{(h_j)\in\mathcal{H}} \Big[h_0(X_0) + \sum_{j=0}^{T-1} E^* \sup_{\mathbf{a}} \Lambda_j(\mathbf{a})\ \{\, (\mathcal{L}h)_j(X_j) - h_j(X_j)\,\}^+ \Big] & (9) \end{aligned}$$

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

$$\eta_{j+1} \equiv h_{j+1}(X_{j+1})\varphi(X_j, X_{j+1}; a_j), \tag{10}$$

and the set $\mathcal{H}$ is the set of sequences $(h_j)_{j=0}^T$ of (bounded) measurable functions from $\mathcal{X}$ to $\mathcal{X}$, satisfying the terminal condition

$$h_T = F.$$

Remarks. (i) To get from the form (5) to (8), we add a martingale-difference sequence $E_j^*(\eta_{j+1}) - \eta_{j+1}$ to the objective, then do a *pathwise* optimisation over the controls, take expectations, and finally minimise 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 $\mathbf{a}$. 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. Notice that because of the form (10) of η_{j+1}, the conditional expectation appearing in (8) can as well be expressed as

$$E_j^*(\eta_{j+1}) = Ph_{j+1}(X_j, a_j). \tag{11}$$

(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, the result allows us to obtain *upper* bounds on the value function.

(iii) The choice of reference measure must be expected to 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.

(iv) The form (8) is well suited to Monte Carlo, since it involves an expectation of a pathwise supremum. The second form (9) can be evaluated with no backward recursion. It can be reworked in the situation where

$$\psi(x) \equiv \int \sup_{a\in A}\, p(x, x'; a)\, m(dx) < \infty \tag{12}$$

for all x. This allows us to define a new transition density

$$\bar{p}(x, x') = \frac{\sup_{a\in A} p(x, x'; a)}{\psi(x)}$$

with corresponding path probability $\bar{P}$. Writing $\Psi_j \equiv \prod_{i=0}^{j-1} \psi(X_i)$, the final form (9) becomes simply

$$\min_{(h_j)} \Big[\, h_0(X_0) + \sum_{j=0}^{T-1} \bar{E}\Psi_j \{\, (\mathcal{L}h)_j(X_j) - h_j(X_j)\,\}^+ \,\Big]. \tag{13}$$

This is of interest because it expresses the solution in terms of a *fixed* measure, which we could call the *maximum-likelihood measure*, together with a reweighting factor which is independent of any choice of controls.

PROOF.The problem is to find

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

where of course we define

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

Now fixing $\mathbf{a} \in \mathcal{A}$, for any P^*-martingale M,

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

since for $\mathbf{a} \in \mathcal{A}$ the process $\Lambda(\mathbf{a})$ is adapted. We shall specialise the martingale slightly by expressing the martingale-differences as

$$\Delta M_{j+1} = E_j^*(\eta_{j+1}) - \eta_{j+1}, \quad \eta_{j+1} \equiv h_{j+1}(X_{j+1})\varphi(X_j, X_{j+1}; a_j). \tag{14}$$

Notice that

$$\Lambda_j(\mathbf{a})\eta_{j+1} = \Lambda_{j+1}(\mathbf{a})h_{j+1}(X_{j+1}); \tag{15}$$

this fact is used in the following reworking. The first inequality comes by relaxing the constraint that $\mathbf{a} \in \mathcal{A}$:

$$\begin{aligned}
V_0(X_0) &= \sup_{\mathbf{a}\in\mathcal{A}} v_0(X_0;\mathbf{a}) \\
&= \sup_{\mathbf{a}\in\mathcal{A}} E^*\Big[\sum_{j=0}^{T-1} \Lambda_j(\mathbf{a})\{f_j(X_j,a_j) + E_j^*(\eta_{j+1}) - \eta_{j+1}\} + \Lambda_T(\mathbf{a})F(X_T) \Big] \\
&= \sup_{\mathbf{a}\in\mathcal{A}} E^*\Big[\sum_{j=0}^{T-1} \Lambda_j(\mathbf{a})\{f_j(X_j,a_j) + Ph_{j+1}(X_j,a_j) - \eta_{j+1}\} + \Lambda_T(\mathbf{a})F(X_T) \Big] \\
&\le E^*\Big[\sup_{\mathbf{a}}\Big\{ \sum_{j=0}^{T-1} \Lambda_j(\mathbf{a})\{f_j(X_j,a_j) + Ph_{j+1}(X_j,a_j) - \eta_{j+1}\} + \Lambda_T(\mathbf{a})F(X_T)\Big\} \Big] \\
&= E^*\Big[\sup_{\mathbf{a}}\Big\{ \sum_{j=0}^{T-1} \big\{\Lambda_j(\mathbf{a})\{f_j(X_j,a_j) + Ph_{j+1}(X_j,a_j)\} - \Lambda_{j+1}(\mathbf{a})h_{j+1}(X_{j+1})\big\} \\
&\qquad\qquad + \Lambda_T(\mathbf{a})F(X_T)\Big\} \Big] \\
&= E^*\Big[\sup_{\mathbf{a}}\Big\{ h_0(X_0) + \sum_{j=0}^{T-1} \Lambda_j(\mathbf{a})\{f_j(X_j,a_j) + Ph_{j+1}(X_j,a_j) - h_j(X_j)\} \Big\} \Big] \\
&\le E^*\Big[h_0(X_0) + \sum_{j=0}^{T-1} \sup_{\mathbf{a}} \Lambda_j(\mathbf{a})\{f_j(X_j,a_j) + Ph_{j+1}(X_j,a_j) - h_j(X_j)\} \Big] \\
&= h_0(X_0) + \sum_{j=0}^{T-1} E^*\big[\sup_{\mathbf{a}} \Lambda_j(\mathbf{a})\{ (\mathcal{L}h)_j(X_j) - h_j(X_j)\} \big] \\
&\le h_0(X_0) + \sum_{j=0}^{T-1} E^*\big[\sup_{\mathbf{a}} \Lambda_j(\mathbf{a})\{ (\mathcal{L}h)_j(X_j) - h_j(X_j)\}^+ \big].
\end{aligned}$$

Taking the infimum over the functions $(h_j) \in \mathcal{H}$, we get

$$\begin{aligned} V_0(X_0) &\leq \inf_{(h_j)\in\mathcal{H}} E^*\Big[\sup_{\mathbf{a}}\Big\{ \sum_{j=0}^{T-1} \Lambda_j(\mathbf{a})\{f_j(X_j,a_j) - \eta_{j+1} + E_j^*(\eta_{j+1})\} + \Lambda_T(\mathbf{a})F(X_T)\Big\}\Big]. \\ &\leq \inf_{(h_j)\in\mathcal{H}}\Big[h_0(X_0) + \sum_{j=0}^{T-1} E^*\big[\sup_{\mathbf{a}} \Lambda_j(\mathbf{a})\{\,(\mathcal{L}h)_j(X_j) - h_j(X_j)\,\}\,\big]\Big] \\ &\leq \inf_{(h_j)\in\mathcal{H}}\Big[h_0(X_0) + \sum_{j=0}^{T-1} E^*\big[\sup_{\mathbf{a}} \Lambda_j(\mathbf{a})\{\,(\mathcal{L}h)_j(X_j) - h_j(X_j)\,\}^+\,\big]\Big]. \end{aligned} \tag{16}$$

In fact, there is equality throughout. To see this, we use the Bellman equation for the value function

$$V_j = (\mathcal{L}V)_j,$$

so if we take $h_j = V_j$, the sum in (16) vanishes and leaves only $h_0(X_0) = V_0(X_0)$. ∎

REMARK. The proof also shows that

$$V_0(x_0) = \min_{(h_j)\in\mathcal{H}}\Big[h_0(X_0) + \sum_{j=0}^{T-1} E^* \sup_{\mathbf{a}} \Lambda_j(\mathbf{a})\,\{\,(\mathcal{L}h)_j(X_j) - h_j(X_j)\,\}\Big], \tag{17}$$

a fact that we will refer back to later.

Theorem 1 gives 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 (8). However, it is important that this optimisation, over the sequence $\mathbf{a}$, 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 (8). We can rewrite the expression inside the expectation on the right-hand side,

$$\sum_{j=0}^{T-1} \Lambda_j(\mathbf{a})\{f_j(X_j,a_j) + Ph_{j+1}(X_j,a_j) - \eta_{j+1}\} + \Lambda_T(\mathbf{a})F(X_T) \tag{18}$$

$$= \sum_{j=0}^{m-1} \Lambda_j(\mathbf{a})\{f_j(X_j,a_j) + Ph_{j+1}(X_j,a_j) - \eta_{j+1}\} + \Lambda_m(\mathbf{a})Z_m, \tag{19}$$

where

$$Z_m \equiv \sum_{j=m}^{T-1} \Lambda_{m,j}(\mathbf{a})\{f_j(X_j,a_j) + Ph_{j+1}(X_j,a_j) - \eta_{j+1}\} + \Lambda_{m,T}(\mathbf{a})F(X_T)$$

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

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

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

$$V_k^{(n+1)}(x) = E^*\Bigg[\sup_{\mathbf{a}} \Bigg\{ \sum_{j=k}^{T-1} \Lambda_{k,j}(\mathbf{a}) \{ f_j(X_j, a_j) - V_{j+1}^{(n)}(X_{j+1}) \varphi(X_j, X_{j+1}; a_j) + P V_{j+1}^{(n)}(X_j, a_j) \} + \Lambda_{k,T}(\mathbf{a}) F(X_T) \Bigg\} \Bigg| X_k = x \Bigg], \quad (20)$$

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

$$\Delta_k^{(n)} \leq (1 + B) \sum_{r=k+1}^{T} \Delta_r^{(n-1)}. \quad (21)$$

REMARKS. 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 (21) 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 (20) 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 (20) gives us

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

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.

∎

Discussion. For the purposes of this discussion, we assume for ease of exposition that $f_j = f$ for all j, and that there exists a sequence of functions ψ_k such that the integral $P\psi_k(x, a)$ is known in closed form. The reason for this is to permit approximation of the value function as linear combinations of the ψ_k; this is like what Longstaff & Schwartz [9] do.

WHEN MIGHT WE USE THIS APPROACH? When the steps of the dynamic programming algorithm are numerically intensive, as for example in a situation where $\mathcal{X}$ is very high dimensional and the required integrations are difficult to do, or when the pointwise optimisation over $a \in A$ is hard, then the simulation-based approach of Theorem 1 may be of value. One advantage is that this approach only seeks the solution starting from a particular x_0, whereas the dynamic programming approach is calculating the solution from *all* starting points.

The first thing to do will be to simulate some paths of the process.

WHAT LAW SHOULD WE USE FOR THE INITIAL SIMULATION? Probably not the reference Markovian law P^*, as the paths of X under P^* can't be expected to look very much like the

paths of the optimally-controlled process, and so we will get little relevant information about the objective if we just simulate from P^*. This problem becomes more acute the larger T, so it may be worth simulating initially only out to some $T_1 < T$, and gradually increasing T_1 as the algorithm proceeds. Since the intermediate rewards f_j could all be zero (or very small), we should not forget to include a term $F(X_{T_1})$ in the objective, as we will ultimately be steering towards this. The *maximum likelihood measure* $\bar{P}$ is also not a very promising candidate, as the law does not depend in any way on f_j, F, but it suggests something we might try instead. Since the objective is

$$\begin{aligned} v_0(X_0;\mathbf{a}) &\equiv E^*\Big[\sum_{j=0}^{T-1}\Lambda_j(\mathbf{a})f(X_j,a_j)+\Lambda_T(\mathbf{a})F(X_T)\Big] \\ &= E^*\Big[\Lambda_T(\mathbf{a})\,\{\sum_{j=0}^{T-1}f(X_j,a_j)+F(X_T)\}\Big] \\ &\simeq \varepsilon^{-1}E^*\Big[\Lambda_T(\mathbf{a})\exp\Big\{\varepsilon\sum_{j=0}^{T-1}f(X_j,a_j)+\varepsilon F(X_T)\Big\}-1\Big] \\ &= \varepsilon^{-1}E^*\Big[\prod_{j=0}^{T-1}\varphi(X_j,X_{j+1};a_j)e^{\varepsilon f(X_j,a_j)}\,.e^{\varepsilon F(X_T)}-1\Big], \end{aligned}$$

this suggests we might modify the definition of $\bar{P}$ by defining

$$\begin{aligned} \psi(x) &\equiv \int \sup_{a\in A}\, p(x,x';a)e^{\varepsilon f(x,a)}\,m(dx), \\ \bar{p}(x,x') &= \frac{\sup_{a\in A}p(x,x';a)e^{\varepsilon f(x,a)}}{\psi(x)}. \end{aligned}$$

The effect of this is to lead the process in directions where the running reward is higher. The choice of ε will need to be tuned a bit.

How do we move from one simulation to the next? We suppose that our current estimate $V_t^{(n)}$ of the value is expressed as a linear combination of the ψ_k:

$$V_t^{(n)} = \sum_k c_{t,k}^{(n)}\psi_k$$

which allows us to write down expressions for $PV_t^n(x,a)$. Once we have simulated sample paths $(X_0^{(i)}, X_1^{(i)}, \ldots, X_T^{(i)})$ for $i = 1, \ldots, N$, we perform the pathwise optimisation in (8), and then have some estimate of the value at the points $X_t^{(i)}$ at time t. We now regress these values onto the functions ψ_k to get a next approximation to the value. The next simulation should be according to what we now think is an approximation to the optimal path law, and one way to do this would be as follows. Suppose that at time t on the simulated path we have reached x; first choose $x' \in \{X_t^{(i)}, i = 1, \ldots, N\}$ at random, points 'nearer' to x being chosen with higher probability, and jump to that point, $x' = X_t^{(q)}$, say. Then make the move to y at

time $t+1$ according to the density $p(x', \cdot; a_t^{(q)})$, where $a_t^{(q)}$ was the control optimally chosen at $X_t^{(q)}$.

4 Variants of the main result.

4.1 Least-squares characterisation.

The study [12] 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

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

(where the η_j are as at (10)), 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*

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

is solved by taking $h_j = V_j$.

4.2 Multiplicative form of the main result.

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

Theorem 2

$$V_0(X_0) \le \inf_{\eta>0} E^* \left[\sup_{\mathbf{a}} \left\{ \sum_{j=0}^{T-1} \Lambda_j(\mathbf{a})\, f_j(X_j, a_j)\, \frac{\eta_{j+1}}{E_j^*[\eta_{j+1}]} + \Lambda_T(\mathbf{a}) F(X_T) \right\} \right], \tag{22}$$

where the random variables η_j are positive. Provided

$$g_j^*(X_j, X_{j+1}, a_j) \equiv V_j(X_j) - V_{j+1}(X_{j+1}) \varphi(X_j, X_{j+1}; a) > 0, \tag{23}$$

the result (22) can be strengthened to the statement

$$V_0(X_0) = \min_{\eta>0} E^* \left[\sup_{\mathbf{a}} \left\{ \sum_{j=0}^{T-1} \Lambda_j(\mathbf{a})\, f_j(X_j, a_j)\, \frac{\eta_{j+1}}{E_j^*[\eta_{j+1}]} + \Lambda_T(\mathbf{a}) F(X_T) \right\} \right], \tag{24}$$

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

[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.

REMARK. Condition (23) 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 $\mathbf{a} \in \mathcal{A}$, and letting η be any strictly positive adapted process,

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

Just as before,

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

Taking the infimum over all choices of η leads to the first statement (22).

For the second statement (24), we again use 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 $V_0(X_0)$. ∎

4.3 'Strong' form of the main result.

In Theorems 1 and 2, 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

$$X_{j+1} = \xi(j, X_j, a_j, \varepsilon_{j+1}), \qquad (j = 0,\ldots,T-1) \tag{25}$$

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;

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! The difference is exactly the difference between a strong solution of a stochastic differential equation, constructed over a given driving process, and a weak solution, constructed in law on some probability space (as in Theorems 1 and 2).

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

$$V_0(X_0) = \min_{(h_j)\in\mathcal{H}} E\Big[\sup_{\mathbf{a}}\Big\{\sum_{j=0}^{T-1}(f_j(X_j,a_j) - h_{j+1}(X_{j+1}) + Ph_{j+1}(X_j,a_j)) + F(X_T)\Big\}\Big], \quad (26)$$

where the X_j and a_j are related through (25). The minimum is attained by taking $h_j = V_j$.

Remarks. The Monte Carlo approach to evaluating the right-hand side of (26) 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. ∎

4.4 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

$$V(x) = \sup_a E^*\Big[f(x,a) + \beta\varphi(x,X_1;a)V(X_1) \;\Big|\; X_0 = x \Big]. \quad (27)$$

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

$$\mathcal{L}g(x) \equiv \sup_{a\in\mathcal{A}} E^*\Big[f(x,a) + \beta\varphi(x,X_1;a)g(X_1) \;\Big|\; X_0 = x \Big] \quad (28)$$

is a monotone contraction with unique fixed point the value function V solving (27).

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

$$\mathcal{L}_h g(x) \equiv E^*\left[\sup_a \big\{ f(x,a) + Ph(x,a) - h(X_1)\varphi(x,X_1;a) + \beta\varphi(x,X_1;a)g(X_1)\big\} \;\middle|\; X_0 = x \right]. \tag{29}$$

Just as for $\mathcal{L}$, 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*

$$V = \inf_h \; g_h^* = \min_h \; g_h^*, \tag{30}$$

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

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

Since $\mathcal{L}V = V$, 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 for any h we have $g_h^* \geq V$, 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,

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

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

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

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

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

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 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 there remain important issues in performing the simulations and related calculations in an efficient manner. The whole study of numerical implementation is barely begun.

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] A. G. Bhatt and V. S. Borkar. Occupation measures for controlled Markov processes: characterization and optimality. *Annals of Probability*, 24:1531–1562, 1996.

[4] 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.

[5] 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.

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

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

[8] T. G. Kurtz and R. H. Stockbridge. Existence of Markov controls and characterization of optimal Markov controls. *SIAM Journal on Control and Optimization*, 36:609–653, 1998.

[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] A. S. Manne. Linear programming and sequential decisions. *Management Science*, 6:259–267, 1960.

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

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

[13] 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.

A Links to the occupation measure approach

The approach of Theorem 1 is in some sense a dual approach, but how is it related to another dual approach, the occupation measure approach, as explained and studied in [10], [3], [8]? In this approach, the original optimization problem is re-expressed as

$$\sup_{(\mu_t),(\kappa_t)} \sum_{t=0}^{T} \int_{\mathcal{X}} \mu_t(dx) \int_A \kappa_t(x,da) f_t(x,a), \tag{A1}$$

subject to the constraints

$$\begin{aligned} \mu_0(dx) &= \delta_{x_0}(dx), & \text{(A2)} \\ \mu_{t+1}(dx) &= \int_{\mathcal{X}} \mu_t(dx') \int_A \kappa_t(x',da) p(x',x;a) m(dx), \quad (t=0,\ldots,T-1) & \text{(A3)} \end{aligned}$$

where we write $f_T(x,a) \equiv F(x)$, and each of the measures μ_t is a probability measure, and κ_t is a Markov kernel from $\mathcal{X}$ into A for each t.

The interpretation of this is that μ_t is the law of the controlled process at time t under controls given by the Markov kernels κ_t; frequently, the Markov kernels will be degenerate, in the sense that $\kappa_t(x,da) = \delta_{\alpha(t,x)}(da)$ for all t,x, but this formulation allows randomised decision rules also.

Introducing Lagrangian multiplier functions $v_t : \mathcal{X} \to \mathbb{R}$ for each $t = 0,\ldots,T$ changes the optimisation problem into the Lagrangian form

$$\begin{aligned} &\sup_{\mu_t,\kappa_t \geq 0} \Big[v_0(x_0) + \sum_{t=0}^{T-1} \int_{\mathcal{X}} \mu_t(dx) \left\{ -v_t(x) + \int_A \kappa_t(x,da) f_t(x,a) + \int_A \kappa_t(x,da) P v_{t+1}(x,a) \right\} \\ &\qquad\qquad + \int_{\mathcal{X}} \mu_T(dx) \int_A \kappa_T(x,da) \{F(x) - v_T(x)\} \Big] \\ = &\sup_{\mu_t,\kappa_t \geq 0} \Big[v_0(x_0) + \sum_{t=0}^{T-1} \int_{\mathcal{X}} \mu_t(dx) \int_A \kappa_t(x,da) \left\{ -v_t(x) + f_t(x,a) + P v_{t+1}(x,a) \right\} \\ &\qquad\qquad + \int_{\mathcal{X}} \mu_T(dx) \{F(x) - v_T(x)\} \Big]. \end{aligned}$$

We deduce the dual-feasibility conditions

$$\begin{aligned} v_t(x) &\geq f_t(x,a) + P v_{t+1}(x,a) \quad (x \in \mathcal{X}, a \in A, t = 0,\ldots,T-1) & \text{(A4)} \\ v_T(x) &\geq F(x) & \text{(A5)} \end{aligned}$$

and the dual problem is now to minimise $v_0(x_0)$ subject to (A4),(A5). These conditions are obviously equivalent to

$$\begin{aligned} v_t(x) &\geq \sup_{a \in A} \left\{ f_t(x,a) + P v_{t+1}(x,a) \right\} \equiv (\mathcal{L}v)_t(x) \quad (x \in \mathcal{X}, t = 0,\ldots,T-1) & \text{(A6)} \\ v_T(x) &\geq F(x) & \text{(A7)} \end{aligned}$$

which is solved by taking $v_T = F$, and $v_t = (\mathcal{L}v)_t$ for $0 \leq t < T$ - the Bellman equations. The value of the dual problem is also evidently equal to the value of the primal problem. However, it will often be the case that the operations involved in the Bellman equations (taking expectations, and pointwise maximisation) will be hard to do numerically, so casting the dual problem in Lagrangian form gives us

$$L(\{g_t\}) \equiv \inf_{(v_t)} \left\{ v_0(x_0) + \sum_{t=0}^{T-1} \int g_t(x)\{(\mathcal{L}v)_t(x) - v_t(x)\}\, m(dx) \right\}$$

for non-negative multiplier functions (g_t). The dual form of this programming problem is

$$\sup_{g_t \geq 0} L(\{g_t\}) \leq V_0(x_0),$$

and (9) is the same expression, for a particular choice of the multipliers (g_t), attaining the value.