Numerical Comparison of Controls and Verification of Optimality for Stochastic Control Problems [1]

Kurt Helmes[2] and Richard H. Stockbridge[3]

August 12, 1997

Abstract. We provide two approaches to the numerical analysis of stochastic control problems. These analyses rely on linear programming formulations of the control problem and allow numerical comparison between controls and numerical verification of optimality. The formulations characterize the processes through the moments of the induced occupation measures. We deal directly with the processes rather than with some approximations to the processes. Excellent software is readily available since the computations involve finite dimensional linear programs.

Key words. stochastic control, linear programming, numerical comparison, numerical verification, moments, bounded follower.

Abbreviated Title. Numerical Analysis of Stochastic Control Problems.

AMS(MOS) subject classifications. 49M35, 90C05, 93E20, 93E25.

1. Introduction. The purpose of this paper is to provide a numerical method for comparing the efficiency of various controls and to present a verification approach to the solution of stochastic control problems. We provide a numerical approach for which excellent software is readily available.

We use a linear programming formulation for the stochastic control problem which involves optimizing over stationary distributions on the (state,control) product space. The comparison occurs by specifying the various controls and running the LP to determine the corresponding values. The verification approach specifies the candidate for the optimal control (rather than the value function as in dynamic programming) and after modifying the LP, seeks evidence that the candidate is, in fact, optimal.

In each case, we characterize the set of feasible distributions for the LP through linear conditions which their moments must satisfy. For numerical computation, the problem is reduced to finite dimensions by considering only a finite number of moments. An important aspect of this approach is that we are analyzing the stochastic processes of the control problem rather than some approximating processes.

Formulating the stochastic control problem as a linear program over a space of stationary distributions was initiated by Manne [11] for the long-term average control of finite-state Markov chains and has been extensively studied in a discrete setting (see for example [4, 5, 16, 13]). Recently Hernandez-Lerma *et. al.* [9] has utilized this approach for Markov decision processes. The LP formulation for the long-term average control problem has been extended to continuous time and continuous state and control spaces in [14, 15]. Recently, LP formulations have been established ([10, 3]) in this general setting for infinite-horizon discounted, finite-horizon and first exit control problems. An alternate numerical approach to the solution of stochastic control problems using the LP approach has been studied in [12].

This paper is organized as follows. In the rest of this section we give the general formulation of the control problem. We first state the problem in stochastic process terms and

[1] This research is partially supported by NSF under grant DMS 9404990.
[2] Institute of Operations Research, Humboldt University of Berlin, Berlin, Germany
[3] Department of Statistics, University of Kentucky, Lexington, KY 40506–0027

then give the linear programming formulation. We refer the reader to [10] for the equivalence of the formulations. Beginning with section 2, we specialize to a particular example in order to illustrate the approach. Section 2 defines the example as a control problem for stochastic processes and displays the linear program. Section 3 discusses how to use the linear program to compare controls and writes the LP in terms of the moments of the distribution. By considering a renewal argument, it also identifies various characteristics of the controlled process and provides numerical results. Section 4 discusses the verification method. In section 5, we provide the analytical solution of the problem and compare the numerical results to this. Section 6 briefly examines the problem with a different criterion and in section 7 we discuss some issues which have arisen from the numerical experiments.

The linear programs have been run using the LP procedure in SAS, CPLEX, GAMS and NAG. Each has performed well and the results are consistent.

1.1. Stochastic process formulation. For a metric space S, we let $M(S)$ denote the space of Borel measurable functions, $C(S)$ denote the space of continuous functions, $\widehat{C}(S)$ denote the space of continuous functions which vanish at infinity, and $\mathcal{P}(S)$ the collection of probability measures on S.

Dynamics. The processes are specified as *relaxed* solutions for a controlled martingale problem which we now define.

Let the state space E and control space U be locally compact, complete, separable, metric spaces and let $E^\Delta = E \cup \{\Delta\}$ be the one-point compactification of E. Let $A : \mathcal{D}(A) \subset \widehat{C}(E) \to C(E \times U)$ and $\nu \in \mathcal{P}(E)$. Then an $E \times \mathcal{P}(U)$-valued process (X, Λ) is a relaxed solution of the controlled martingale problem for (A, ν) if there exists a filtration $\{\mathcal{F}_t\}$ such that (X, Λ) is $\{\mathcal{F}_t\}$-progressive, X has initial distribution ν and for every $f \in \mathcal{D}(A)$,

$$f(X(t)) - \int_0^t \int_U Af(X(s), u)\, \Lambda_s(du)\, ds \tag{1.1}$$

is an $\{\mathcal{F}_t\}$-martingale.

We assume that the generator A satisfies the following conditions:

(i) $\mathcal{D}(A)$ is dense in $\widehat{C}(E)$,
(ii) for each $f \in \mathcal{D}(A)$ and $u \in U$, $Af(\cdot, u) \in \widehat{C}(E)$,
(iii) for each $f \in \mathcal{D}(A)$ and compact $K \subset U$, $\limsup_{x \to \Delta} \sup_{u \in K} |Af(x, u)| = 0$,
(iv) for each $u \in U$, $A_u f = Af(\cdot, u)$ satisfies the positive maximum principle,
(v) $\mathcal{D}(A)$ is an algebra,
(vi) there exists $\psi \in C(U)$, $\psi > 0$, such that for each $f \in \mathcal{D}(A)$ there exist constants a_f and b_f satisfying

$$|Af(x, u)| \le a_f + b_f \psi(u). \tag{1.2}$$

The first four conditions are sufficient to guarantee existence of solutions (at least in E^Δ) to the controlled martingale problem for each $\nu \in \mathcal{P}(E)$, condition (v) is required for the LP reformulation and condition (vi) is for compactness when the control space is not compact.

Cost. For the purposes of this paper, we illustrate the approach using long-term average and discounted criteria. The approach of this paper can also be applied to a finite horizon criterion and a first passage or first exit criterion associated with solutions of the controlled martingale problem.

Let $c \in M(E \times U)$ be lower semicontinuous and bounded below. The long-term average cost associated with a solution (X, Λ) of the controlled martingale problem for A is

$$\limsup_{t \to \infty} E\left[t^{-1} \int_0^t \int_U c(X(s), u) \Lambda_s(du)\, ds \right]. \tag{1.3}$$

1.2. LP reformulation. The LP reformulation depends upon the following additional assumptions:

(vii) there exist constants a and b such that

$$\psi(u) \leq a + b\, c(x,u); \tag{1.4}$$

(viii) $\{(x,u) : c(x,u) \leq a\}$ is compact for each $a > 0$.

REMARK 1.1. *Observe that conditions (vi) and (vii) are satisfied whenever the space of controls U is compact.*

It is shown in [10] that the stochastic control problem is equivalent to the linear program

$$\begin{cases} \text{Minimize } \langle c, \mu \rangle := \displaystyle\int_{E\times U} c(x,u)\,\mu(dx \times du) \\ \\ \text{Subject to } \ \langle Af, \mu \rangle := \displaystyle\int_{E\times U} Af(x,u)\mu(dx \times du) = 0 \quad \forall\, f \in \mathcal{D}(A) \ , \\ \qquad\qquad\quad \mu \in \mathcal{P}(E \times U). \end{cases}$$

(Denote the set of distributions which satisfy the constraints by the set $\mathcal{A}$.) Moreover, an optimal control can be characterized in terms of an optimizing μ^* as the conditional distribution $\eta^*(x,du)$ on the control space U given the state x; that is, for $\mu^* \in \mathcal{A}$ satisfying $\langle c, \mu^* \rangle = \inf_{\mu\in\mathcal{A}} \langle c, \mu \rangle$, $\eta^*(x,du)$ gives an optimal Markov (relaxed) control where η^* satisfies

$$\mu^*(dx \times du) = \eta^*(x,du)\mu_0^*(dx)$$

and μ_0^* denotes the state marginal distribution of μ^*.

2. The bounded follower problem. We illustrate the verification method by considering a modification of the bounded follower problem studied by Beneš, Shepp and Witsenhausen [1]. The bounded follower model consists of a drifted Brownian motion in which the control action is to choose the drift. The intuition is that the standard Brownian motion is an object which it is desired to follow as closely as possible. The follower acts by choosing its velocity subject to strict limits, hence the name.

We modify this bounded follower model by placing a limit on the distance between the two objects (the drifted Brownian motion process). We consider a process which, when it is in the unit interval $[0,1)$, is a reflected diffusion starting at the origin:

$$dx_t = u_t dt + dw_t + d\xi_t$$

in which w is a standard Brownian motion process and ξ is a monotone process which, when $x_t = 0$, increases enough so as to keep the process in the interval. The control specifies the drift of the diffusion and is subject to the hard constraints

$$|u_t| \leq 1.$$

When the process hits $\{1\}$ it sticks for an exponential length of time (with parameter λ) and then jumps back to a position specified by the initial distribution of the diffusion. There is no control of the process while it is at the boundary.

The long-term average objective is to minimize

$$\limsup_{t\to\infty} t^{-1} E\left[\int_0^t x_s^2\, ds\right]. \tag{2.1}$$

Since x is reflected at the origin, the test functions in the domain of the generator must satisfy the boundary condition $f'(0) = 0$. Letting $\mathcal{D}(A) = \{f \in C^2([0,1]) : f'(0) = 0\}$, the generator is

$$Af(x,u) = (uf'(x) + 1/2f''(x))\, I_{[0,1)}(x) + \lambda(f(0) - f(x))\, I_{\{1\}}(x). \tag{2.2}$$

The linear program for the modified bounded follower problem is

$$\begin{cases} \text{Minimize } \langle x^2, \mu\rangle = \displaystyle\int_{[0,1]\times[-1,1]} x^2\, \mu(dx \times du) \\ \\ \text{Subject to } \ \langle Af, \mu\rangle = \displaystyle\int_{[0,1]\times[-1,1]} Af(x,u)\, \mu(dx \times du) = 0, \qquad \forall f \in \mathcal{D}(A) \\ \qquad\qquad\quad \mu \in \mathcal{P}([0,1] \times [-1,1]). \end{cases}$$

REMARK 2.1. *The use of a* reflected *model is for numerical convenience since the optimal unreflected process will behave symmetrically about the origin. The number of variables for the unreflected problem would be double that of the reflected one.*

3. Comparison of controls. Observe that since the control enters linearly in A, by decomposing each μ as $\mu(dx \times du) = \eta(x, du)\mu_0(dx)$ where μ_0 is the state marginal and η is the regular conditional distribution of u given x, the condition $\langle Af, \mu\rangle = 0$ is equivalent to

$$\langle Af(x, u(x)), \mu_0\rangle = 0$$

where $u(x) = \int u\eta(x, du)$ gives the mean value of the control used in state x. Since the objective function can also be expressed in terms of μ_0, we can consider the linear program

$$\text{LP0}\begin{cases} \text{Minimize } \langle x^2, \mu_0\rangle \\ \\ \text{Subject to } \ \langle Af(x,u(x)), \mu_0\rangle = 0 \qquad \forall f \in \mathcal{D}(A) \\ \qquad\qquad\quad \mu_0 \in \mathcal{P}([0,1]). \end{cases}$$

Once $u(x)$ has been specified, the linear program determines the induced invariant distribution μ_0 with the smallest cost. We can then compare controls by solving the LP for the different $u(x)$.

We illustrate this by considering three controls:

$$u_1(x) = -1, \tag{3.3}$$

$$u_2(x) = \begin{cases} -1, & x < a \\ 1, & x > a, \end{cases} \tag{3.4}$$

$$u_3(x) = 1, \tag{3.5}$$

where the switch location a is at some value in $[0, 1]$.

The first control u_1 is the optimal control for the original bounded follower problem. The second control is selected because for some values of the parameter λ it will be an optimal control (see section 5) and the third control is selected merely for illustrative purposes.

3.1. Numerical solution of LP0. This linear program is still not in a form which can be solved numerically. Rather than consider the distributions μ_0 directly, we characterize them in terms of their moments since these uniquely determine the distributions on the compact interval $[0,1]$ ([2, Theorem 30.1]).

We restrict the test functions to the class $\mathcal{D}_1 = \{f(x) = x^n : n \geq 2\}$ and for $n = 0, 1, 2, \ldots$ let

$$m_n = \int_{[0,1]} x^n \, \nu(dx),$$

and

$$a_n = \int_{[0,a]} x^n \, \nu(dx) \qquad \text{and} \qquad b_n = \int_{(a,1]} x^n \, \nu(dx).$$

For $f \in \mathcal{D}_1$, the constraints of LP0 using u_1, u_2 and u_3, respectively, become

$$0 \;=\; \int_{[0,1)} \left[n u_i(x) x^{n-1} + \frac{n(n-1)}{2} x^{n-2} \right] \mu_0(dx) - \lambda \mu_0(\{1\})$$

$$(3.6) \qquad = \begin{cases} -n m_{n-1} + \frac{n(n-1)}{2} m_{n-2} + \left(n - \frac{n(n-1)}{2} - \lambda\right) \mu_0(\{1\}) \\[2ex] -n a_{n-1} + \frac{n(n-1)}{2} a_{n-2} + n b_{n-1} + \frac{n(n-1)}{2} b_{n-2} - \left(n + \frac{n(n-1)}{2} + \lambda\right) \mu_0(\{1\}) \\[2ex] n m_{n-1} + \frac{n(n-1)}{2} m_{n-2} - \left(n + \frac{n(n-1)}{2} + \lambda\right) \mu_0(\{1\}). \end{cases}$$

With this notation, the objectives become one of minimizing m_2 when $u = u_1$ or u_3 and minimizing $m_2 = a_2 + b_2$ when $u = u_2$. Note that the linear program determines the minimum cost over all switching controls of the form $u = u_2$ so picks the partial moments $\{a_n, b_n\}$ corresponding to the best switching location.

At this point we must be quite careful. We derived the constraints (3.6) using the moments of the distribution μ_0. However, these constraints on their own do not necessarily require that the collections $\{m_n\}$ and $\{a_n + b_n\}$ be the moments of a distribution. We must incorporate additional constraints, the Hausdorff moment conditions, to guarantee that the collections correspond to the moments of some distribution. The Hausdorff conditions come from the observation that for any distribution μ_0 concentrated on $[0,1]$,

$$\int x^k (1-x)^n \, \mu_0(dx) \geq 0, \qquad k = 0, 1, 2, \ldots, n = 1, 2, \ldots$$

since the integrand is non-negative. These conditions become, for $k = 0, 1, 2, \ldots$ and $n = 1, 2, 3, \ldots$,

$$(3.7) \qquad \sum_{j=0}^{n} (-1)^j \binom{n}{j} m_{j+k} \geq 0.$$

The conditions (3.7) are clearly necessary for the moments. They are also sufficient to guarantee that the sequence corresponds to the moments of some distribution ([6, p. 222]). *The crucial property of these moment conditions is that they are linear* so they fit directly into the linear programming constraints.

Finally, in order to numerically solve this linear program, we restrict analysis to a finite number of moments.

3.2. Additional analysis. Under any control u, the stochastic process has natural cycles consisting of two parts, the reflected diffusion on the interval $[0,1)$ and the time when it is at the boundary $\{1\}$. A renewal argument for the process indicates that the long-term average cost equals the expected cost over a cycle divided by the expected cycle length. In addition, the optimal measure $\mu_0{}^*$ which solves LP0 under the control u gives the long-term average occupation of the process in the various states. We combine these observations as follows.

Let τ_1 denote the time at which the process first hits $\{1\}$ and τ_2 the time the process stays at the boundary. Then for each continuous f,

$$\int_{[0,1]} f(x)\mu_0{}^*(dx) = \frac{E\left[\int_0^{\tau_1+\tau_2} f(x_s)ds\right]}{E\left[\tau_1+\tau_2\right]} = \frac{E\left[\int_0^{\tau_1} f(x_s)\,ds\right] + f(1)/\lambda}{E\left[\tau_1\right] + 1/\lambda} \tag{3.8}$$

since τ_2 is exponential with parameter λ. In particular, this implies that

$$\mu_0{}^*(\{1\}) = \frac{1/\lambda}{E\left[\tau_1\right] + 1/\lambda} \tag{3.9}$$

and with $f(x) = x^n$,

$$m_n = K^{-1}\left(E\left[\int_0^{\tau_1} x_s^n ds\right] + 1/\lambda\right) \tag{3.10}$$

where $K = E[\tau_1] + 1/\lambda$. Rearranging (3.9) and (3.10) we obtain $K = 1/(\lambda\mu_0{}^*\{1\})$,

$$E[\tau_1] = \frac{1-\mu_0{}^*\{1\}}{\lambda\mu_0{}^*\{1\}} \tag{3.11}$$

and

$$E\left[\int_0^{\tau_1} x_s^n ds\right] = \frac{m_n - \mu_0{}^*\{1\}}{\lambda\mu_0{}^*\{1\}}\ . \tag{3.12}$$

We can therefore derive the moments of the process up to the first hitting time of $\{1\}$ from the LP results. We will illustrate the values of (3.12) for $n=2$ and refer to it as the L^2-norm of the process.

3.3. Numerical results. Table 1 displays, for the three controls and for various values of λ, the long-term average second moment m_2 of the processes, the mass on the boundary $\mu_0\{1\}$, and the corresponding mean hitting time of the process at $\{1\}$, $E[\tau_1]$, and the L^2-norm, $E\left[\int_0^{\tau_1} x_s^2 ds\right]$.

Notice that the mean exit times and the L^2-norms of the processes under u_1 and u_3 are consistent for all values of the parameter λ. Moreover, as λ increases, the process spends less time on the boundary resulting in smaller values of $\mu_0\{1\}$ and m_2.

The mean exit time and L^2-norm for the process under u_2 are not constant. This is due to the fact that the LP optimizes over the location a of the switch. Thus as λ changes so does the optimal switch location.

Now to compare controls. For small values of λ, we see that the long-term average costs of controls u_1 and u_2, the mean exit times, and the L^2-norms are the same. This indicates that the optimal switch location $a \in [0,1]$ occurs at $a = 1$. However, for larger values of λ,

λ	u_i	m_2	$\mu_0\{1\}$	$E[\tau_1]$	$E\left[\int_0^{\tau_1} x_s^2 ds\right]$
0.10	u_1	0.8416833066	0.8200399358	2.1945280494	0.2639306914
	u_2	0.8416833066	0.8200399359	2.1945280493	0.2639306914
	u_3	0.9573699227	0.9462825988	0.5676676416	0.1171671541
1.00	u_1	0.3956549048	0.3130352854	2.1945280494	0.2639306913
	u_2	0.3956549048	0.3130352855	2.1945280487	0.2639306912
	u_3	0.7126301037	0.6378903113	0.5676676416	0.1171671541
2.00	u_1	0.2835118719	0.1855612525	2.1945280494	0.2639306913
	u_2	0.2835117903	0.1855610974	2.1945303029	0.2639311102
	u_3	0.5780517551	0.4683105309	0.5676676412	0.1171671539
5.00	u_1	0.1937461921	0.0835237658	2.1945280495	0.2639306914
	u_2	0.1937461921	0.0835237658	2.1945280494	0.2639306914
	u_3	0.4131568624	0.2605294129	0.5676676416	0.1171671541
6.00	u_1	0.1823641884	0.0705857359	2.1945280495	0.2639306914
	u_2	0.1822436923	0.0724345038	2.1342625582	0.2526631710
	u_3	0.3865185347	0.2269629306	0.5676676416	0.1171671541
10.00	u_1	0.1586080812	0.0435819470	2.1945280493	0.2639306913
	u_2	0.1573813291	0.0478394774	1.9903238347	0.2289779434
	u_3	0.3252623619	0.1497751182	0.5676676416	0.1171671541
20.00	u_1	0.1398649000	0.0222763979	2.1945280494	0.2639306913
	u_2	0.1360774009	0.0265725598	1.8316403160	0.2060487237
	u_3	0.2706425638	0.0809496833	0.5676676416	0.1171671541
50.00	u_1	0.1282127320	0.0090312696	2.1945280494	0.2639306913
	u_2	0.1209275305	0.0122298521	1.6153427513	0.1777579604
	u_3	0.2334094042	0.0340328420	0.5676676416	0.1171671541

Table 1: Values of the processes with controls u_1,u_2 and u_3 using 24 moments.

the long-term average second moment $m_2 = a_2 + b_2$ under control u_2 is smaller than that under control u_1 and the mean exit time and the L^2-norm are also smaller. This indicates that the constant control is no longer a better control but switching actually improves the costs.

In each case, the control u_3 has higher long-term average second moments than either of the other controls.

4. The verification approach. It is possible to make adjustments to LP0 and derive a method of numerical verification for the optimal control problem. We investigate this approach in this section.

Consider as a candidate for the optimal control u_1 given in (3.3) and fix this as the control in (2.2). Thus, for each $f \in \mathcal{D}_1$, let $Bf(x) = Af(x, -1)$ and consider a second linear program

$$\text{LP1} \begin{cases} \text{Minimize } \langle x^2, \nu \rangle \\ \\ \text{Subject to } \ \langle Bf(x), \nu \rangle \leq 0, \qquad \forall f \in \mathcal{D}_1 \\ \qquad\qquad\quad \nu \in \mathcal{P}([0,1]). \end{cases}$$

Let $\mathcal{A}_0$ denote the set of feasible distributions $\mu_0 \in \mathcal{P}([0,1])$ for LP0 for any $u(x)$ and similarly $\mathcal{A}_1$ denote the set of feasible distributions $\nu \in \mathcal{P}([0,1])$ for LP1.

PROPOSITION 4.1. $\inf_{\nu \in \mathcal{A}_1} \langle x^2, \nu \rangle \leq \inf_{\mu \in \mathcal{A}_0} \langle x^2, \mu \rangle$

Proof. This follows immediately from the observation that for each $f \in \mathcal{D}_1$, $f' \geq 0$ so $Bf(x) \leq Af(x, u(x))$ and thus for each feasible μ_0 of LP0

$$0 = \langle Af(x, u(x)), \mu_0 \rangle \geq \langle Bf(x), \mu_0 \rangle.$$

Thus $\mathcal{A}_0 \subset \mathcal{A}_1$. $\square$

REMARK 4.2. *Observe that the distribution μ_0 induced by the candidate control (3.3) satisfies the constraints of LP1 with equality.*

The idea of the verification approach is to numerically solve LP1 and check to see that the constraints are tight. This corresponds to (3.3) being the optimal control for LP1 and therefore the optimal control for LP0.

The same modifications as for LP0 (involving writing the constraints in terms of the moments and adding the Hausdorff moment conditions) are used on LP1 to enable it to be solved numerically.

Table 2 gives the values of the nth-moments and, for the stationary constraints on the test functions $f(x) = x^n$, the slack variables from solving LP1 in the cases when the parameter $\lambda = 0.1$ and $\lambda = 10$. Recall, the test functions used for the stationarity constraints require $n \geq 2$ so tightness is only relevant for these constraints.

	$\lambda = 0.1$		$\lambda = 10$	
n	nth-moment	Slack Variables	nth-moment	Slack Variables
0	1.0000000000	.	1.0000000000	.
1	0.8690179712	.	0.3036801507	.
2	0.8416833066	0.0000000000	0.1582580401	0.0000000000
3	0.8320039936	0.0000000000	0.1072760943	0.0000000000
4	0.8275672526	0.0000000000	0.0836722890	0.0030757346
5	0.8251908953	0.0000000000	0.0710297175	0.0000000000
6	0.8237779579	0.0000000000	0.0635121245	0.0000000000
7	0.8228725136	0.0000000000	0.0586925615	0.0000000000
8	0.8222586917	0.0000000000	0.0554273255	0.0000000000
9	0.8218239376	0.0000000000	0.0531161734	0.0000991442
10	0.8215050357	0.0000000000	0.0514207454	0.0001305796
11	0.8212643186	0.0000000000	0.0501404281	0.0000768566
12	0.8210782324	0.0000000000	0.0491504544	0.0000204654
13	0.8209314351	0.0000000000	0.0483695901	0.0000000000
14	0.8208134971	0.0000000000	0.0477429766	0.0000000000
15	0.8207163950	0.0000000000	0.0472325633	0.0000010046
16	0.8206321056	0.0000000000	0.0468113486	0.0000000000

Table 2: Numerical results for LP1 using 17 moments.

Observe that when $\lambda = 0.1$ the slack variables are zero which indicates that the stationarity constraints are tight. This numerically verifies that u_1 is the optimal control for LP1 and hence for LP0, when $\lambda = 0.1$. When the parameter $\lambda = 10$, it is no longer the case that the constraints are tight and thus there is evidence that u_1 is no longer the optimal control.

5. The analytical solution. We now present the optimal solution of the modified bounded follower problem and compare the numerical results to them. The details of the solution are given in the appendix.

The optimal control for the original bounded follower problem under the long-term average criterion is to always choose a maximum drift towards the origin, $u^* = u_1$. The optimal control for the *modified* model depends on the exponential holding time λ. For small λ, the optimal control is again to choose maximum drift toward the origin, $u^* = u_1$; but for larger λ, the optimal control has a switch, $u^* = u_2$. For $x < a$, where $a < 1$, the drift should be toward the origin at maximum level but for $x > a$, it is optimal to push away from the origin with maximum force.

The intuition is that when the holding time on the boundary is short, it is better to move the process to the boundary quickly and then jump to the origin. The location at which the switch occurs is given as the solution of

$$a^2 + a + \left(r - \frac{1}{2}\right)(e^{2a} - 1) = 0 \tag{5.1}$$

where

$$r = \frac{\frac{1}{12}[-8a^3 - 6a^2 - 6a + 6a^2 e^{(2a-2)} - 6ae^{(2a-2)} + 3e^{(2a-2)} + 3e^{2a} - 2] + \frac{1}{\lambda}}{\frac{1}{2}[-4a + e^{2a} + e^{(2a-2)}] + \frac{1}{\lambda}} \tag{5.2}$$

Note that this root is less than 1 when $\lambda > 5.55471$. In this event the value of r in (5.2) is the long-term average cost (2.1) for the optimally controlled process. Note that r has the form of (3.8).

5.1. Comparison of numerical and analytical results. When $\lambda > 5.55471$, the long-term average second moment is given by r in (5.2) with a being the solution of (5.1). Furthermore, the expected hitting time and L^2-norm are given by the expressions in (5.2) and the mass $\mu_0\{1\}$ can be derived from (5.2). When $\lambda \le 5.55471$, u_1 is the optimal control and the various quantities are obtained by setting $a = 1$ in (5.2) since there is no switch. In the latter case, we immediately see that

$$E[\tau_1] = \frac{e^2 - 3}{2} = 2.1945280494$$

and

$$E\left[\int_0^{\tau_1} x_s^2 ds\right] = \frac{3e^2 - 19}{12} = 0.26393069123,$$

though the values of m_2 and $\mu_0\{1\}$ vary with λ. Table 3 gives all these values for the values of λ in Table 1.

When $\lambda \le 5.55471$, the numerical results agree exactly with the theoretical values. We should emphasize that our numerical analysis works directly with the original processes rather than with approximating processes. It should therefore not be surprising that we have such excellent accuracy.

When $\lambda > 5.55471$, the values of LP0 are slightly off the theoretical values for the optimal process. In particular, the long-term average second moments m_2 obtained numerically are slightly lower than for the optimal processes and the masses $\mu_0\{1\}$ are slightly higher. These errors therefore result in differences in the mean hitting times and the L^2-norms.

λ	m_2	$\mu_0\{1\}$	$E[\tau_1]$	$E\left[\int_0^{\tau_1} x_s^2 ds\right]$
0.10	0.8416833066	0.8200399358	2.1945280494	0.2639306913
1.00	0.3956549048	0.3130352854	2.1945280494	0.2639306913
2.00	0.2835118719	0.1855612525	2.1945280494	0.2639306913
5.00	0.1937461920	0.0835237657	2.1945280494	0.2639306913
6.00	0.1823615576	0.0707056796	2.1905225652	0.2631940334
10.00	0.1580372084	0.0461427770	2.0671864240	0.2424960926
20.00	0.1372709223	0.0259198279	1.8790251491	0.2147990615
50.00	0.1230912029	0.0114300966	1.7297664806	0.1953808610

Table 3: Exact values for the optimal process.

6. Discounted Criterion. In this section, we briefly discuss the numerics for the modified bounded follower process under a discounted criterion. In stochastic processes terms, the objective is to minimize

$$E\left[\int_0^\infty e^{-\alpha s} x_s^2 ds\right].$$

The equivalent linear programming formulation (see [10]) is

$$\begin{cases} \text{Minimize } \alpha^{-1}\langle x^2, \mu\rangle \\ \\ \text{Subject to } \langle A^\alpha f, \mu\rangle, \qquad \forall f \in \mathcal{D}(A) \\ \qquad\qquad\quad \mu \in \mathcal{P}([0,1]\times[-1,1]), \end{cases}$$

where $A^\alpha f(x,u) = Af(x,u) + \alpha\left(\int f d\nu_0 - f(x)\right)$.

Exactly the same analysis can be applied to this LP as for the long-term average LP, the only difference being that the stationarity conditions (3.6) on the distribution μ_0 is

$$0 = \int_{[0,1)} \left[\frac{n(n-1)}{2} x^{n-2} + nu(x)x^{n-1} - \alpha x^n\right] \mu_0(dx) - (\lambda+\alpha)\mu_0\{1\}.$$

We refer to the discounted version of LP1 as DLP1.

λ	u_1	u_2	u_3
0.001	0.3543373355	0.3543373355	0.6794825745
0.01	0.3527479369	0.3527479369	0.6777467904
0.10	0.3379165671	0.3379165671	0.6610423927
1.00	0.2512760753	0.2512760753	0.5408490415
2.00	0.2091110243	0.2091108045	0.4633168750
5.00	0.1619134371	0.1619134371	0.3542808256
6.00	0.1547071186	0.1546817076	0.3349794073
10.00	0.1385458559	0.1378041037	0.2885825450
20.00	0.1245984269	0.1216802000	0.2446678102
50.00	0.1153304093	0.1106759634	0.2132431911

Table 4: m_2-values of the processes with controls u_1,u_2 and u_3; $\alpha = 1$ using 24 moments.

Controls can be compared by specifying $u(x)$ and running the LP to determine the cost as in section 3. Table 4 compares controls u_1, u_2 and u_3 for various values of λ.

Similarly, the verification method can be applied as in section 4. Table 5 seeks to determine whether u_1 is optimal for two values of λ. The same comments hold as before for the long-term average criterion.

	$\lambda = 0.1$		$\lambda = 10$	
n	nth-moment	Slack Variables	nth-moment	Slack Variables
0	1.0000000000	.	1.0000000000	.
1	0.7745925209	.	0.3009331794	.
2	0.7278447422	0.0000000000	0.1557275800	0.0000000000
3	0.7113561196	0.0000000000	0.1054915045	0.0000000000
4	0.7038177099	0.0000000000	0.0820719206	0.0058889074
5	0.6997872307	0.0000000000	0.0695397005	0.0000000000
6	0.6973938025	0.0000000000	0.0621207409	0.0000000000
7	0.6958614515	0.0000000000	0.0573672430	0.0004105538
8	0.6948233558	0.0000000000	0.0541427177	0.0001744158
9	0.6940884920	0.0000000000	0.0518599229	0.0000000000
10	0.6935496790	0.0000000000	0.0501865293	0.0000000000
11	0.6931431037	0.0000000000	0.0489236039	0.0000137213
12	0.6928288964	0.0000000000	0.0479473831	0.0000000000
13	0.6925811506	0.0000000000	0.0471776838	0.0000000000
14	0.6923826153	0.0000000000	0.0465603257	0.0000202830
15	0.6922227127	0.0000000000	0.0460575423	0.0000303232
16	0.6920975391	0.0000000000	0.0456425575	0.0000170282

Table 5: Numerical results for DLP1 using 17 moments with $\alpha = 0.1$.

7. Stability issues. We now discuss some of the issues that have arisen from our numerical experiments. These issues involve the dependence of the method on the number of moments used.

We have observed that when too few moments are used the values of the moments become identical after a certain point, but when more moments are added this degeneracy is corrected. This is illustrated in Table 6.

There is also a limit to the number of moments that the program can handle. If too many moments are requested, the problem seems to become infeasible. We conjecture that the reason for this is that the Hausdorff moment conditions involve binomial coefficients and that numerical problems arise when these coefficients are too large. Perhaps a different collection of test functions such as the Laguerre polynomials would provide more robustness.

There is also some sensitivity of the results on the number of moments, even when there is no degeneracy of the values. This is illustrated in the Table 7 which gives the results of LP1 when $\lambda = 2$.

8. Appendix. In this section, we provide full details of the optimal solution using a dynamic programming approach. (We refer the reader to the standard texts [7] and [8].) Thus we seek a C^2 function f and a constant r (the long-term average value) such that

$$\min_{|u|\le 1}(Af(x,u) + x^2) = r. \tag{8.1}$$

Clearly, the value of u which will minimize this expression will be -1 when $f'(x) > 0$ and will be $+1$ when $f'(x) < 0$.

Moment	26 moments	36 moments
0	1.0000000000	1.0000000000
1	0.1821561325	0.3036801507
2	0.0799256505	0.1582580401
3	0.0710360436	0.1072760943
4	0.0706319706	0.0836722890
5	0.0706319706	0.0710297175
6	0.0706319705	0.0635121245
7	0.0706319705	0.0586925615
8	0.0706319705	0.0554273255
9	0.0706319705	0.0531161734
10	0.0706319705	0.0514207454
11	0.0706319705	0.0501404281
12	0.0706319705	0.0491504544
13	0.0706319705	0.0483695901
14	0.0706319705	0.0477429766
15	0.0706319705	0.0472325633
16	0.0706319705	0.0468113486
17	0.0706319705	0.0464597373
18	0.0706319705	0.0461632168
19	0.0706319705	0.0459108614
20	0.0706319705	0.0456943281
21	0.0706319705	0.0455071545
22	0.0706319705	0.0453442668
23	0.0706319705	0.0452016390
24	0.0706319705	0.0450760496
25	0.0706319705	0.0449648944
26		0.0448660446
27		0.0447777463
28		0.0446985506
29		0.0446272520
30		0.0445628350
31		0.0445044393
32		0.0444513390
33		0.0444029162
34		0.0443586924
35		0.0443186675

Table 6: Values of the moments for LP1 using 26 and 36 moments; $\lambda = 10$.

Due to the form of the generator A, the HJB equation (8.1) consists of a differential equation on the interval $[0, 1)$ and requires

$$\lambda(f(0) - f(1)) + 1 = r. \tag{8.2}$$

We seek the function f by solving the differential equation $1/2 f''(x) + u f'(x) + x^2 = r$ with the switching control (3.4) where the choice of switching point a needs to be determined. When $a \geq 1$, the optimal control is actually (3.3) and the value r is obtained by setting $a = 1$ in the formula.

No. of moments	m_2	$\mu_0\{1\}$
3	0.2307692307	0.2307692307
5	0.2307692307	0.2307692307
7	0.2407407407	0.2222222222
10	0.2434988179	0.2198581560
13	0.2456140350	0.2180451127
15	0.2469879518	0.2168674698
20	0.2486187845	0.2154696132
25	0.2495637002	0.2146596855
28	0.2494246182	0.2147788986
30	0.2494246182	0.2147788986
31	0.2822731209	0.1866230391
32	0.2830136312	0.1859883160
33	0.2812853179	0.1874697274
34	0.2833149458	0.1857300464
35	0.2834852170	0.1855840996

Table 7: Dependence of results of LP1 on the number of moments computed; $\lambda = 2$.

The solution of $Af(x,-1) + x^2 = r$ on the interval $[0, a]$ has the form

$$f(x) = -rx - r/2 + x^3/3 + x^2/2 + x/2 + K_1 e^{2x} + K_2 \tag{8.3}$$

and the solution of $Af(x,1) + x^2 = r$ on the interval $[a, 1)$ is

$$f(x) = rx - r/2 - x^3/3 + x^2/2 - x/2 + K_3 e^{-2x} + K_4 \tag{8.4}$$

The constants K_1, K_2, K_3, K_4 and r are chosen so that f is C^2, $f'(0) = 0$ and (8.2) is satisfied. Furthermore, the switch location a is chosen so that $f'(x) \geq 0$ for $0 \leq x \leq a$ and $f'(x) \leq 0$ for $a \leq x \leq 1$. Note that this implies $f'(a) = 0$ and as a result $f''(a-) = r - a^2 = f''(a+)$.

The condition $f'(0) = 0$ determines $K_1 = r/2 - 1/4$.

Using (8.4) with $f'(a+) = 0$, we identify $K_3 = (r/2 - 1/4 - a^2/2 + a/2)e^{2a}$.

Since the function f is only defined up to an additive constant, we set $K_2 = 0$ which leaves the continuity of f at a to give

$$K_4 = -2ra - r/2 + 2a^3/3 + a^2/2 + 1/4 + (r/2 - 1/4)e^{2a}.$$

The condition (8.2) can be used to solve for r in terms of the switch point a, yielding (5.2).

Finally, the condition $f'(a) = 0$ with f given by (8.3) means that, as claimed, a is a root of the equation (5.1).

Observe that for $x < a$, $f'(x) = x^2 + x + (r - 1/2)(e^{2x} - 1)$. The expression (5.2) for r can be used to see that $r < 1/2$ when $\lambda > 1$, so f' increases monotonely from 0 then eventually decreases monotonely. As a result, there is a unique point $a > 0$ where $f'(a) = 0$. When a is set equal to 1, (5.1) is satisfied by $\lambda = 5.55471$. For $\lambda > 5.55471$, the value of a satisfying (5.1) is in the interval (0,1), so f and r defined by (8.3), (8.4) and (5.2) satisfy (8.1) and the control which switches at a is optimal. For $\lambda < 5.55471$, the root a is greater than 1. Hence f defined by (8.3) on $[0, 1]$ with r given by (5.2) using $a = 1$ satisfies (8.1) and (3.3) is the optimal control.

Acknowledgements. We wish to acknowledge Marta Mendiondo, Werner Kunz, and Stefan Roehl for their help with various implementations of the linear programs in this work.

REFERENCES

[1] V.E. Beneš, L.A. Shepp and H.S. Witsenhausen, Some solvable stochastic control problems, *Stochastics*, **4** (1980), 39–83.
[2] P. Billingsley, *Probability and Measure,* 2nd ed., Wiley, New York (1985).
[3] A. G. Bhatt and V. S. Borkar, Occupation measures for controlled Markov processes: Characterization and optimality. Preprint (1993).
[4] C. Derman, On sequential decisions and Markov Chains, *Management Sci.*, **9**, 16–24, 1962.
[5] E.V. Denardo, On linear programming in a Markov decision problem, *Management Sci.*, **16** (1970), pp. 281–288.
[6] W. Feller, *An Introduction to Probability Theory and its Applications*, Vol. 2, Wiley, New York (1965).
[7] W.H. Fleming and R.W. Rishel, *Deterministic and Stochastic Optimal Control*, Springer–Verlag, New York, Heidelberg, Berlin, 1975.
[8] W.H. Fleming and H.M. Soner, *Controlled Markov Processes and Viscosity Solutions*, Springer–Verlag, New York, Heidelberg, Berlin, 1993.
[9] O. Hernandez-Lerma, J.C. Hennet and J.B. Lasserre, Average Cost Markov Decision Processes: Optimality Conditions, *J. Math. Anal. Appl.*, **158** (1991), pp.396–406.
[10] T.G. Kurtz and R.H. Stockbridge, *Existence of Markov Controls and Characterization of Optimal Markov Controls*, to appear in SIAM J. Control. Optim. .
[11] A.S. Manne, Linear programming and sequential decisions, *Management Sci.*, **6** (1960), 259–267.
[12] M.S. Mendiondo and R.H. Stockbridge, *Approximation of Infinite-dimensional Linear Programming Problems which Arise in Stochastic Control*, submitted for publication.
[13] B.G. Pittel, A linear programming problem connected with optimal stationary control in a dynamic decision problem, *Theory Probab. Appl.*, **16**, 724–728, 1971.
[14] R.H. Stockbridge, Time-average control of martingale problems: Existence of a stationary solution, *Ann. Probab.*, **18** (1990), pp. 190–205.
[15] R.H. Stockbridge, Time-average control of martingale problems: A linear programming formulation, *Ann. Probab.*, **18** (1990), pp. 206–217.
[16] P. Wolfe and G.B. Dantzig, Linear programming in a Markov chain, *Oper. Res.*, **10**, 702–710, 1962.