ANZIAM J. 57(2016), 339–351
doi:10.1017/S1446181115000279

SOLUTIONS AND DIAGNOSTICS OF SWITCHING PROBLEMS WITH LINEAR STATE DYNAMICS

J. HINZ[✉1] and N. YAP[2]

(Received 1 October, 2014; accepted 8 March, 2015; first published online 28 January 2016)

Abstract

Optimal control problems of stochastic switching type appear frequently when making decisions under uncertainty and are notoriously challenging from a computational viewpoint. Although numerous approaches have been suggested in the literature to tackle them, typical real-world applications are inherently high dimensional and usually drive common algorithms to their computational limits. Furthermore, even when numerical approximations of the optimal strategy are obtained, practitioners must apply time-consuming and unreliable Monte Carlo simulations to assess their quality. In this paper, we show how one can overcome both difficulties for a specific class of discrete-time stochastic control problems. A simple and efficient algorithm which yields approximate numerical solutions is presented and methods to perform diagnostics are provided.

2010 *Mathematics subject classification*: primary 49M29; secondary 49M25, 91G80, 90C39.

Keywords and phrases: Markov decisions, approximate dynamic programming, stochastic control, duality.

1. Introduction

Stochastic switching problems with linear state dynamics are common in applications. Surprisingly, most of the classical numerical solution methodologies do not take advantage of their special structure. In what follows, we explain how to adapt the philosophy of function- and regression-based methods to obtain an approximation of the value function in a rather generic way. The main thrust of the present paper is to demonstrate that by exploiting the assumption of linear state dynamics, it becomes possible to achieve remarkably efficient numerical schemes to obtain approximate solutions and assess their distance to optimality.

[1]School of Mathematical and Physical Sciences, University of Technology Sydney, PO Box 123, Broadway, NSW 2007, Australia; e-mail: juri.hinz@uts.edu.au.
[2]Finance Discipline Group, UTS Business School, University of Technology Sydney, PO Box 123, Broadway, NSW 2007, Australia; e-mail: yapsgna@gmail.com.

When making decisions under uncertainty in discrete time, mathematical problems are usually formulated within the framework of *Markov decision theory* (see [2, 4, 6, 12]). In this work, we consider a specific Markov decision model whose state evolution consists of one discrete and one continuous component. To be more specific, we assume that the state space $E = P \times \mathbb{R}^d$ is the product of a finite space P and the Euclidean space $\mathbb{R}^d$. We suppose that the first component $p \in P$ is deterministically driven by a finite set A of actions in terms of a function

$$\alpha : P \times A \to A, \quad (p, a) \mapsto \alpha(p, a),$$

where $\alpha(p, a) \in A$ is the new value of the discrete component of the state, if its previous discrete component value was p and the action $a \in A$ was taken by the controller. Furthermore, we assume that the continuous state component evolves as an uncontrolled Markov process $(Z_t)_{t=0}^T$ on $\mathbb{R}^d$, whose evolution is driven by random linear transformations

$$Z_{t+1} = W_{t+1} Z_t$$

with pre-specified independent and integrable disturbance matrices $(W_t)_{t=1}^T$. In this setting, the transition operators are given by

$$\mathcal{K}_t^a v(p, z) = \mathbb{E}(v(\alpha(p, a), W_{t+1} z)) \quad \text{for } t = 0, \dots, T-1,\ a \in A,$$

for all $(p, z) \in E$ acting on each function $v : E \to \mathbb{R}$, where the above expectation exists. At each time t, we are given the *t-step reward function* $r_t : E \times A \mapsto \mathbb{R}$, where $r_t(x, a)$ represents the reward for applying an action $a \in A$ when the state of the system is $x \in E$ at time t. At the end of the time horizon, at time T, it is assumed that no action can be taken. Here, if the system is in a state x, a *scrap value* $r_T(x)$, which is described by a pre-specified *scrap function* $r_T : E \to \mathbb{R}$, is collected.

The calculation of the optimal policy is addressed in the following setting. We introduce for $t = 0, \dots, T-1$ the so-called *Bellman operator*

$$\mathcal{T}_t v(x) = \sup_{a \in A} (r_t(x, a) + \mathcal{K}_t^a v(x)), \quad x \in E,$$

which acts on each measurable function $v : E \to \mathbb{R}$, where the integrals $\mathcal{K}_t^a v$ for all $a \in A$ exist. Further, consider the *Bellman recursion*

$$v_T = r_T, \quad v_t = \mathcal{T}_t v_{t+1} \quad \text{for } t = T-1, \dots, 0.$$

Under appropriate assumptions, there exists a recursive solution $(v_t^*)_{t=0}^T$ to the Bellman recursion, which gives the so-called *value functions* and determines an optimal policy π^* via

$$\pi_t^*(x) = \operatorname*{argmax}_{a \in A} (r_t(x, a) + \mathcal{K}_t^a v_{t+1}^*(x)), \quad x \in E,$$

for all $t = 0, \dots, T-1$. In this work, we concentrate on Markov decision problems which satisfy specific additional assumptions under which the solutions to the Bellman recursion exist. This enables us to focus on finding efficient numerical solutions.

https://doi.org/10.1017/S1446181115000279 Published online by Cambridge University Press

If we assume that all reward functions

$$r_t(p,\cdot,a) \quad \text{for } t = 0,\dots,T-1,\ p \in P,\ a \in A,$$

and scrap functions $r_T(p,\cdot)$, $p \in P$, are convex and globally Lipschitz continuous in the second component, then we obtain a specific situation. Markov decision problems satisfying these assumptions are referred to as *convex switching systems* in [8]. The remainder of the paper is organized as follows. In Section 2, we describe a modified dynamic programming algorithm that approximates value functions by exploiting the linearity of the evolution of the state variables. In order to assess the distance to optimality of these approximations, diagnostics methods are provided in Section 3 that can be used to perform solution assessment. In Section 4, we present numerical results in the context of a storage management problem. The paper ends with some concluding remarks in Section 5.

2. Solution techniques

The first step in obtaining a numerical solution to the backward induction

$$v_T = r_T, \quad v_t = \mathcal{T}_t v_{t+1} \quad \text{for } t = T-1,\dots,0$$

is an appropriate discretization of the Bellman operator

$$\mathcal{T}_t v(p,z) = \max_{a\in A}(r_t(p,z,a) + \mathbb{E}(v(\alpha(p,a), W_{t+1}z))).$$

For this reason, we consider a modified Bellman operator

$$\mathcal{T}_t^n v(p,z) = \max_{a\in A}\left(r_t(p,z,a) + \sum_{k=1}^{n} \nu_{t+1}^n(k) v(\alpha(p,a), W_{t+1}(k)z)\right)$$

with the expectation replaced by its numerical counterpart, which is defined in terms of an appropriate distribution sampling $(W_{t+1}(k))_{k=1}^n$ of each disturbance W_{t+1} with corresponding probability weighting $(\nu_{t+1}^n(k))_{k=1}^n$. In the resulting modified backward induction

$$v_T^{(n)} = r_T, \quad v_t^{(n)} = \mathcal{T}_t^n v_{t+1}^{(n)} \quad \text{for } t = T-1,\dots,0 \tag{2.1}$$

the functions $(v_t^{(n)})_{t=0}^T$ need to be described by algorithmically tractable objects. Note that if all reward and scrap functions are convex in the continuous variable, then the modified value functions (2.1) are also convex. Then we approximate these value functions in terms of piecewise linear and convex functions in the following manner. First, we introduce the so-called sub-gradient envelope $\mathcal{S}_G f$ of a convex function $f : \mathbb{R}^d \to \mathbb{R}$ on a grid $G \subset \mathbb{R}^d$ as $\mathcal{S}_G f = \bigvee_{g\in G}(\nabla_g f)$, which is a maximum of the sub-gradients $\nabla_g f$ of f on all grid points $g \in G$. Using the sub-gradient envelope operator, we define the double-modified Bellman operator as

$$\mathcal{T}_t^{m,n} v(p,\cdot) = \mathcal{S}_{G^m} \max_{a\in A}\left(r_t(p,\cdot,a) + \sum_{k=1}^{n} \nu_{t+1}^n(k) v(\alpha(p,a), W_{t+1}(k)\,\cdot)\right),$$

https://doi.org/10.1017/S1446181115000279 Published online by Cambridge University Press

where the operator $\mathcal{S}_{G^m}$ stands for the sub-gradient envelope on the grid $G^m = \{g^1, \dots, g^m\}$. The corresponding backward induction

$$v_T^{(m,n)}(p,\cdot) = \mathcal{S}_{G^m} r_T(p,\cdot), \quad p \in P, \tag{2.2}$$

$$v_t^{(m,n)}(p,\cdot) = \mathcal{T}_t^{m,n} v_{t+1}^{(m,n)}(p,\cdot) \quad \text{for } t = T-1, \dots, 0 \tag{2.3}$$

yields the so-called double-modified value functions $(v_t^{n,m})_{t=0}^T$, which enjoy excellent asymptotic and algorithmic properties. Namely, under appropriate additional assumptions, the double-modified value functions converge uniformly to the true value functions almost surely on compact sets (see [8]). These assumptions include the convexity and global Lipschitz continuity of the rewards and scraps, the integrability of all disturbances and some restrictions on the distribution sampling and grid density.

However, the algorithmic properties of the scheme (2.2) and (2.3) are most essential for the purpose of this work. Since the double-modified value functions $(v_t^{(m,n)})_{t=0}^T$ are piecewise linear and convex, they can be expressed in a compact and appealing form, using matrix representations. Note that any piecewise convex function f can be described by a matrix where each linear functional will be represented by one of the matrix's rows. To denote this relation, let us agree on the following notation: given a function f and a matrix F, we write $f \sim F$ whenever $f(z) = \max(Fz)$ holds for all $z \in \mathbb{R}^d$. Let us emphasize that the sub-gradient envelope operation $\mathcal{S}_G$ on a grid G is reflected in terms of a matrix representative by a specific row-rearrangement operator

$$f \sim F \quad \Leftrightarrow \quad \mathcal{S}f \sim \Upsilon_G[F],$$

where the row-rearrangement operator Υ_G associated with the grid $G = \{g^1, \dots, g^m\} \subset \mathbb{R}^d$ acts on the matrix F with d columns as follows:

$$(\Upsilon_G F)_{i,\cdot} = L_{\operatorname{argmax}(Fg^i),\cdot} \quad \text{for all } i = 1, \dots, m. \tag{2.4}$$

Remark 2.1. The implementation of the row-rearrangement operator Υ_G defined in (2.4) is easily obtained when the grid is represented by a matrix G, whose rows contain grid elements as row vectors. Having represented a convex function in terms of a matrix F, the row-rearrangement operator is illustrated by the following command in the language *R*:

```
F[apply(F%*%t(G), FUN=which.max, MARGIN=2)].
```

For piecewise convex functions, the result of maximization, summation and composition with linear mapping, followed by the sub-gradient envelope, can be obtained using their matrix representatives. More precisely, if $f_1 \sim F_1$ and $f_2 \sim F_2$, then it follows that

$$\mathcal{S}_G(f_1 + f_2) \sim \Upsilon_G(F_1) + \Upsilon_G(F_2),$$
$$\mathcal{S}_G(f_1 \vee f_2) \sim \Upsilon_G(F_1 \sqcup F_2),$$
$$\mathcal{S}_G(f_1(W_{t+1}(k)\cdot) \sim \Upsilon_G(F_1 W_{t+1}(k)),$$

where the operator $\sqcup$ stands for binding matrices by rows, which yields a matrix whose rows contain all rows from each participating matrix.

https://doi.org/10.1017/S1446181115000279 Published online by Cambridge University Press

Remark 2.2. The row-maximization operator $\sqcup$ is implemented by simple binding-by-row of the matrices $F1$, $F2$ representing the corresponding functions, which can be implemented in *R* by

```
rbind(F1, F2).
```

Under the assumptions of global Lipschitz continuity and convexity for scraps and reward functions, the backward induction (2.2) and (2.3) can be expressed in terms of the matrix representatives $V_t^{m,n}(p)$ of the value functions $v_t^{(m,n)}(p,\cdot)$ for $p \in P$, $t = 0, \ldots, T$. Since the double-modified backward induction involves maximization, summations and compositions with linear mappings applied to piecewise linear convex functions, it can be rewritten in terms of matrix operations. Now, let us present the resulting algorithm.

Algorithm 1: Double-modified backward induction

1 **Pre-calculations:** Given a grid $G^m = \{g^1, \ldots, g^m\}$, implement the row-rearrangement operator $\Upsilon = \Upsilon_{G^m}$ and the row-maximization operator $\bigsqcup_{a\in A}$. Determine a distribution sampling $(W_t(k))_{k=1}^n$ of each disturbance W_t with the corresponding weights $(\nu_t(k))_{k=1}^n$ for $t = 1, \ldots, T$. Given reward functions $(r_t)_{t=0}^{T-1}$ and scrap value r_T, determine the normal form of the matrix representatives of their sub-gradient envelopes

$$\mathcal{S}_{G^m} r_t(p,\cdot,a) \sim R_t(p,a), \quad \mathcal{S}_{G^m} r_T(p,\cdot) \sim R_T(p)$$

for $t = 0, \ldots, T-1$, $p \in P$ and $a \in A$. Introduce matrix representatives of each value function

$$v_t^{(m,n)}(p,\cdot) \sim V_t(p) \quad \text{for } t = 0, \ldots, T, p \in P$$

which are obtained via the following matrix form of the backward induction:

2 **Initialization:** Start with the matrices

$$V_T(p) = R_T(p) \quad \text{for all } p \in P.$$

3 **Recursion:** For $t = T-1, \ldots, 0$, calculate for $p \in P$

$$V_t(p) = \bigsqcup_{a\in A}\Big(R_t(p,a) + \sum_{k=1}^{n} \nu_{t+1}(k)\Upsilon[V_{t+1}(\alpha(p,a)) \cdot W_{t+1}(k)]\Big). \tag{2.5}$$

Remark 2.3. The choice of the weights $(\nu_{t+1}(k))_{k=1}^n$ can be derived from appropriate discretization. Here, a sub-martingale based sampling has been suggested by Hinz [8]. However, for Monte Carlo based procedures, the matrices $(W_{t+1}(k))_{k=1}^n$ can be obtained from independent identically distributed realizations of W_{t+1} with equal weights $(\nu_{t+1}(k) = 1/n)_{k=1}^n$.

https://doi.org/10.1017/S1446181115000279 Published online by Cambridge University Press

Remark 2.4. Numerical solutions for stochastic control problems have been considered in numerous publications (see [11]). For switching and stopping problems, function-based methods of the least-squares approach [10] are popular. In one of our previous works [9], these least-squares methods are compared to our convex switching approach.

3. Solution diagnostics

Convergence properties of regression based methods for numerical solutions of Markov decision problems have been extensively studied by Belomestny et al. [3].

Bounds estimation for optimal stopping problems has been discussed in the literature [1, 7, 13] and extended by Rogers [14] to a certain class of discrete-time stochastic control problems. The technical note [15] discusses optimal stopping problems in the framework of partial observation, and uses particle filtering techniques to assess the quality of a numerical solution in terms of duality bounds. It also addresses the connection to the variance reduction technique.

In what follows, we present an adaptation of the work of Rogers [14] to obtain upper and lower bound estimation of an approximate solution. We shall use ω_k to represent a realization of the random experiment. Given a policy $\pi = (\pi_t)_{t=0}^{T-1}$, all actions and positions can be determined recursively via

$$a_t^\pi := \pi_t(p_t^\pi, Z_t), \quad p_{t+1}^\pi := \alpha(p_t^\pi, a_t^\pi) \quad \text{for } t = 0, \dots, T-1.$$

For such a policy, having started at $p_0^\pi = p_0$ and $Z_0 = z_0$, the *policy value*

$$v_0^\pi(p_0, z_0) = \mathbb{E}\Big(\sum_{s=0}^{T-1} r_s(p_s^\pi, Z_s, a_s^\pi) + r_T(p_T^\pi, Z_T)\Big)$$

is the expected value of a *test run*

$$\mathcal{V}_0^\pi(p_0, z_0) = \sum_{s=0}^{T-1} r_s(p_s^\pi, Z_s, a_s^\pi) + r_T(p_T^\pi, Z_T).$$

Since $v_0^\pi(p_0, z_0) = \mathbb{E}(\mathcal{V}_0^\pi(p_0, z_0))$, the value function

$$v_0^*(p_0, z_0) = \sup_\pi v_0^\pi(p_0, z_0) = v_0^{\pi^*}(p_0, z_0)$$

can be estimated from below via a Monte Carlo method in the following way. By the strong law of large numbers,

$$\lim_{K\to\infty} \frac{1}{K} \sum_{k=1}^{K} \mathcal{V}_0^\pi(p_0, z_0)(\omega_k) = v_0^\pi(p_0, z_0) \le v_0^*(p_0, z_0),$$

given $K \in \mathbb{N}$ independent realizations $\{\omega_1, \dots, \omega_K\}$. In principle, each Monte Carlo trial can be computed backwards recursively as

$$\mathcal{V}_T^\pi(p, z) = r_T(p, z), \tag{3.1}$$

$$\mathcal{V}_t^\pi(p, z) = r_t(p, z, \pi_t(p, z)) + \mathcal{V}_{t+1}^\pi(\alpha(p, \pi_t(p, z)), W_{t+1}z) \quad \text{for } t = T-1, \dots, 0. \tag{3.2}$$

Such a procedure is frequently referred to as *back-testing*. Although the back-testing may be useful for practical purposes, convincing a practitioner by satisfactory results,

https://doi.org/10.1017/S1446181115000279 Published online by Cambridge University Press

it does not clarify how far these outcomes are from the theoretically best possible results.

In the remainder of the section, we suggest a sound solution to this question by means of a diagnostic method. Given a starting point (p_0, z_0), we show how the gap

$$[v_0^{\pi}(p_0, z_0), v_0^{\pi^*}(p_0, z_0)] \tag{3.3}$$

between a given strategy π and the optimal strategy π^* can be assessed. The methodology is based on a finite sample $\{\omega_1, \dots, \omega_K\}$ of trajectory realizations and utilizes a built-in variance reduction technique to derive tight (close) confidence bounds for upper and lower estimates of the interval (3.3).

Let us focus on the upper bound first. Consider a sequence $\varphi = (\varphi_t)_{t=1}^T$ of random mappings

$$\varphi_t : P \times \mathbb{R}^d \times A \times \Omega \to \mathbb{R}, \quad (p, z, a, \omega) \mapsto \varphi_t(p, z, a)(\omega), \tag{3.4}$$

which, for $t = 1, \dots, T$, satisfy

$$\mathbb{E}(\varphi_t(p, z, a)) = 0, \quad p \in P, z \in \mathbb{R}^d, a \in A, \tag{3.5}$$

such that the σ-algebras

$$\sigma(\varphi_t(p, z, a), W_t; a \in A, p = P, z \in \mathbb{R}^d) \quad \text{for } t = 1, \dots, T \text{ are independent.} \tag{3.6}$$

Given such $\varphi = (\varphi_t)_{t=1}^T$, introduce random functions $(\bar{v}_t^{\varphi})_{t=0}^T$, with $\bar{v}_t^{\varphi} : P \times \mathbb{R}^d \times \Omega \to \mathbb{R}$ for $t = 0, \dots, T$, which are recursively defined for $t = T, \dots, 1$ via

$$\bar{v}_T^{\varphi}(p, z) = r_T(p, z), \tag{3.7}$$

$$\bar{v}_t^{\varphi}(p, z) = \max_{a \in A}(r_t(p, z, a) + \varphi_{t+1}(p, z, a) + \bar{v}_{t+1}^{\varphi}(\alpha(p, a), W_{t+1}z)). \tag{3.8}$$

Using $(\bar{v}_t^{\varphi})_{t=0}^T$, the following result holds (see [9]).

THEOREM 3.1. (i) *Given $\varphi = (\varphi_t)_{t=1}^T$ as in (3.4) satisfying (3.5) and (3.6), introduce $(\bar{v}_t^{\varphi})_{t=0}^T$ by (3.8). For each policy $\pi = (\pi_t)_{t=0}^{T-1}$, its value $(v_t^{\pi})_{t=0}^T$ is dominated from above as*

$$v_t^{\pi}(p, z) \le \mathbb{E}(\bar{v}_t^{\varphi}(p, z)) \quad \textit{for all } t = 0, \dots, T, p \in P, z \in \mathbb{R}^d. \tag{3.9}$$

(ii) *Given the value $(v_t^{\pi^*})_{t=0}^T$ of the optimal policy $\pi^* = (\pi_t^*)_{t=0}^{T-1}$, define $(\varphi_t^*)_{t=1}^T$ by*

$$\varphi_{t+1}^*(p, z, a) = \mathbb{E}(v_{t+1}^{\pi^*}(\alpha(p, a), W_{t+1}z)) - v_{t+1}^{\pi^*}(\alpha(p, a), W_{t+1}z) \tag{3.10}$$

for all $p \in P$, $z \in \mathbb{R}^d$, $a \in A$ and $t = 0, \dots, T-1$. Then the mappings $(\varphi_t^*)_{t=1}^T$ satisfy (3.4)–(3.6) such that inequality (3.9) holds with

$$v_t^{\pi^*}(p, z) = \bar{v}_t^{\varphi^*}(p, z) \quad \textit{for all } t = 0, \dots, T, p \in P, z \in \mathbb{R}^d.$$

Given a sequence $\varphi = (\varphi_t)_{t=1}^T$ satisfying (3.5) and (3.6), we introduce the random functions $(\underline{v}_t^{\pi,\varphi})_{t=0}^T$ with $\underline{v}_t^{\pi,\varphi} : P \times \mathbb{R}^d \times \Omega \to \mathbb{R}$ for $t = 0, \dots, T$, which are recursively defined for $t = T, \dots, 1$ via

$$\underline{v}_T^{\pi,\varphi}(p, z) = r_T(p, z), \tag{3.11}$$

$$\underline{v}_t^{\pi,\varphi}(p, z) = r_t(p, z, \pi_t(p, z)) + \varphi_{t+1}(p, z, \pi_t(p, z)) + \underline{v}_{t+1}^{\pi,\varphi}(\alpha(p, \pi_t(p, z)), W_{t+1}z). \tag{3.12}$$

Using $(\underline{v}_t^{\pi,\varphi})_{t=0}^T$, the following theorem holds (see [9]).

https://doi.org/10.1017/S1446181115000279 Published online by Cambridge University Press

THEOREM 3.2. (i) *Given* $\varphi = (\varphi_t)_{t=1}^T$ *as in* (3.4) *satisfying* (3.5) *and* (3.6) *and a policy* $\pi = (\pi_t)_{t=0}^{T-1}$, *introduce* $(\underline{v}_t^{\pi,\varphi})_{t=0}^T$. *We have*

$$v_t^\pi(p,z) = \mathbb{E}(\underline{v}_t^{\pi,\varphi}(p,z)) \quad \text{for all } t = 0,\dots,T,\ p \in P, z \in \mathbb{R}^d. \tag{3.13}$$

(ii) *Given the value* $(v_t^{\pi^*})_{t=0}^T$ *of the optimal policy* $\pi^* = (\pi_t^*)_{t=0}^{T-1}$, *define* $(\varphi_t^*)_{t=1}^T$ *by*

$$\varphi_{t+1}^*(p,z,a) = \mathbb{E}(v_{t+1}^{\pi^*}(\alpha(p,a), W_{t+1}z)) - v_{t+1}^{\pi^*}(\alpha(p,a), W_{t+1}z)$$

for all $p \in P$, $z \in \mathbb{R}^d$, $a \in A$ *and* $t = 0,\dots,T-1$. *Then the mappings* $(\varphi_t^*)_{t=1}^T$ *satisfy* (3.4) *and* (3.6) *such that equation* (3.13) *holds with*

$$v_t^{\pi^*}(p,z) = \underline{v}_t^{\pi^*,\varphi^*}(p,z) \quad \text{for all } t = 0,\dots,T,\ p \in P, z \in \mathbb{R}^d.$$

Let us elaborate on a practical application of this technique. Suppose that we attempt to assess the distance to optimality of an approximate policy $\tilde{\pi}$, obtained by a numerical procedure described previously. According to Theorem 3.1(i), any arbitrary $(\varphi_t)_{t=1}^T$ satisfying (3.5) and (3.6) yields an upper bound

$$v_0^{\tilde{\pi}}(p,z) \le v_0^{\pi^*}(p,z) \le \mathbb{E}(\bar{v}_0^\varphi(p,z)) \quad p \in P, z \in \mathbb{R}^d.$$

Note that the expectation $\mathbb{E}(\bar{v}_0^\varphi(p,z))$ can be approached via a Monte Carlo average. Thus, we obtain the following estimation procedure.

Algorithm 2: Upper bound estimation

1 Given a switching system, find a $(\varphi_t)_{t=1}^T$ which satisfies (3.4), (3.5) and (3.6).

2 Choose a number $K \in \mathbb{N}$ of Monte Carlo trials and obtain for $k = 1,\dots,K$ independent realizations $(W_t(\omega_k))_{t=1}^T$ of disturbances.

3 Starting at $z_0^k := z_0 \in \mathbb{R}^d$, define for $k = 1,\dots,K$ the trajectories $(z_t^k)_{t=0}^T$ recursively by

$$z_{t+1}^k = W_{t+1}(\omega_k) z_t^k \quad \text{for } t = 0,\dots,T-1$$

and determine realizations

$$\varphi_{t+1}(p, z_t^k, a)(\omega_k) \quad \text{for } t = 0,\dots,T-1 \text{ and } k = 1,\dots,K.$$

4 For each $k = 1,\dots,K$, initialize the recursion at $t = T$ as

$$\bar{v}_T^\varphi(p, z_T^k) = r_T(p, z_T^k) \quad \text{for all } p \in P$$

and continue for $t = T-1,\dots,0$ by

$$\bar{v}_t^\varphi(p, z_t^k) = \max_{a \in A}(r_t(p, z_t^k, a) + \varphi_{t+1}(p, z_t^k, a)(\omega_k) + \bar{v}_{t+1}^\varphi(\alpha(p,a), z_{t+1}^k)).$$

Store the value $\bar{v}_0^\varphi(p, z_0^k)$ for $k = 1,\dots,K$.

5 Determine the sample mean $(1/K)\sum_{k=1}^K \bar{v}_0^\varphi(p, z_0^k)$ to estimate $v_0^{\pi^*}(p, z_0)$ from above, possibly using confidence bounds.

https://doi.org/10.1017/S1446181115000279 Published online by Cambridge University Press

To obtain a tight (close) upper bound, $(\varphi_t)_{t=1}^T$ must be chosen accordingly. Thereby, the assertion (ii) of Theorem 3.1 suggests an appropriate choice. Namely, in the hypothetical ideal case that the value functions $(v_t^{\pi*})_{t=0}^T$ are known, $(\varphi_t^*)_{t=1}^T$ is obtained via (3.10), which gives an exact and nonrandom upper bound. In practice, this situation is not feasible, since an optimal strategy π^* is not known. Instead, we suggest using an approximate value function $(\tilde{v}_t)_{t=0}^T$, returned by one of the algorithms described in this work. That is, following (3.10), a reasonable candidate for $t = 0, \dots, T-1$ can be given as

$$\varphi_{t+1}(p, z, a) = \mathbb{E}(\tilde{v}_{t+1}(\alpha(p, a), W_{t+1}z)) - \tilde{v}_{t+1}(\alpha(p, a), W_{t+1}z). \tag{3.14}$$

However, note that this choice involves an exact calculation of the expectation $\mathbb{E}(\tilde{v}_{t+1}(\alpha(p, a), W_{t+1}z))$, which is not possible in practice. For this reason, we suggest a slight modification. We introduce φ_{t+1} similar to (3.14), with the expectation replaced by an arithmetic mean over a number I of independent copies $(W_{t+1}^{(i)})_{i=1}^I$ of W_{t+1}. That is, given independent random variables W_{t+1} and $W_{t+1}^{(i)}$ for $i = 1, \dots, I$ and $t = 0, \dots, T-1$ such that the distribution of $W_{t+1}^{(i)}$ equals that of W_{t+1}, we define

$$\varphi_{t+1}(p, z, a) = \frac{1}{I} \sum_{i=1}^{I} \tilde{v}_{t+1}(\alpha(p, a), W_{t+1}^{(i)}z) - \tilde{v}_{t+1}(\alpha(p, a), W_{t+1}z) \tag{3.15}$$

for all $t = 0, \dots, T-1$, $a \in A$, $p \in P$ and $z \in \mathbb{R}^d$. With this definition, $(\varphi_t)_{t=1}^T$ satisfies (3.5) and (3.6) and the above algorithm of upper bound estimation can be used in practice.

Having suggested the estimation of the upper bound in (3.3), let us turn now to the estimation of the lower bound of this interval. Given a strategy $\pi = (\pi_t)_{t=0}^{T-1}$, the value $v_0^\pi(p_0, z_0)$ can in principle be approached from test runs of the strategy in a series of independent back-testing experiments. However, it turns out that a slight adaptation of the upper bound technique provides far better results, due to a built-in variance reduction. Similarly to (ii) of the previous theorem, which indicates that the variance of Monte Carlo trials reduces if the approximate solution is close to the optimal, we establish a recursive procedure with a built-in variance reduction.

The idea is simple; given a nearly optimal policy $\pi = (\pi_t)_{t=0}^{T-1}$, we alter the recursion (3.7) and (3.8) by replacing the maximization with choices of actions that are consistent with the policy $\pi = (\pi_t)_{t=0}^{T-1}$.

The practical implementation of the lower bound estimation is based on the same realization of $(\varphi_t)_{t=1}^T$ as in (3.15), using independent copies of disturbances. Let us summarize this procedure as follows.

https://doi.org/10.1017/S1446181115000279 Published online by Cambridge University Press

Algorithm 3: Lower bound estimation

1 Given approximate value functions $(\tilde{v}_t)_{t=0}^T$ and a corresponding strategy $\tilde{\pi} = (\tilde{\pi}_t)_{t=0}^{T-1}$, choose $\varphi = (\varphi_t)_{t=0}^{T-1}$ as in (3.15).

2 For $K \in \mathbb{N}$ Monte Carlo trials, obtain for $k = 1, \dots, K$ independent realizations $(W_t(\omega_k))_{t=1}^T$ of disturbances.

3 Starting at $z_0^k := z_0 \in \mathbb{R}^d$, define for $k = 1, \dots, K$ trajectories $(z_t^k)_{t=0}^T$ recursively by

$$z_{t+1}^k = W_{t+1}(\omega_k) z_t^k \quad \text{for } t = 0, \dots, T-1$$

and determine realizations

$$\varphi_{t+1}(p, z_t^k, a)(\omega_k) \quad \text{for } t = 0, \dots, T-1, \text{ and } k = 1, \dots, K.$$

4 For each $k = 1, \dots, K$, initialize the recursion at $t = T$ as

$$\underline{v}_T^{\tilde{\pi},\varphi}(p, z_T^k) = r_T(p, z_T^k) \quad \text{for all } p \in P$$

and continue for $t = T-1, \dots, 0$ and for all $p \in P$ by

$$\begin{aligned}\underline{v}_t^{\tilde{\pi},\varphi}(p, z_t^k) &= r_t(p, z_t^k, \tilde{\pi}_t(p, z_t^k)) \\ &\quad + \varphi_{t+1}(p, z_t^k, \tilde{\pi}_t(p, z_t^k))(\omega_k) + \underline{v}_{t+1}^{\tilde{\pi},\varphi}(\alpha(p, \tilde{\pi}_t(p, z_t^k)), z_{t+1}^k).\end{aligned}$$

Store the value $\underline{v}_0^{\tilde{\pi},\varphi}(p, z_0^k)$ for $k = 1, \dots, K$, $p \in P$.

5 Calculate the sample mean $(1/K) \sum_{k=1}^K \underline{v}_0^{\tilde{\pi},\varphi}(p, z_0^k)$ to estimate $v_0^{\pi^*}(p, z_0)$ for $p \in P$ from below, possibly using confidence bounds.

4. Valuation of storage

In this section, we demonstrate the applicability of our method in solving a storage management problem. This is an example of practical importance for which the setup of a stochastic switching system is natural (see [5]). At each time, a decision has to be made on the level of the commodity stored in a facility, where each change incurs an associated transaction cost. Let us now describe the problem in detail and formulate it as a convex switching system.

Let P be the set of possible levels of the commodity in the storage facility. Given storage costs and random price fluctuations, the controller has to decide when to purchase the commodity to fill the storage and when to withdraw from storage and sell it at the market price. Here, A is the set of actions which can be taken in order to change the level in the storage facility. The action a yields a transition from the previous storage level p to the new level $\alpha(p, a)$.

For the purpose of illustration, we consider a simple case in which the storage level can be full, half-full or empty. At each time, the controller must make a decision based on the price of the commodity whether to fill or draw down on the storage facility, and this sequence of decisions can be viewed as an optimal switching problem. For the

https://doi.org/10.1017/S1446181115000279 Published online by Cambridge University Press

sake of definiteness, we set $P = \{1, 2, 3\}$, where 1 stands for 'empty', 2 for 'half-full' and 3 for 'full', and $A = \{1, 2, 3\}$, where 1 stands for 'empty the storage', 2 for 'half-fill the storage' and 3 for 'fill the storage', with changes in the first component given by the function α whose values are determined by the following matrix:

$$\begin{bmatrix} \alpha(1,1) & \alpha(1,2) & \alpha(1,3) \\ \alpha(2,1) & \alpha(2,2) & \alpha(2,3) \\ \alpha(3,1) & \alpha(3,2) & \alpha(3,3) \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \\ 1 & 2 & 3 \end{bmatrix}.$$

In the simplest form of this example, $(Z_t)_{t=0}^T$ describes the Markovian evolution of the market (spot) price of the underlying commodity. More generally, the state Z_t at time t could be multivariate, in which case one of the components of Z_t is usually the market (spot) price of the commodity at time t. The other components may be latent variables representing the current market conditions or stochastic factors which are needed to ensure the Markov property of the dynamics.

For the sake of concreteness, we assume that the commodity price follows a univariate autoregressive model of order $d = 2$ with coefficients 0.3 and 0.65, driven by a unit variance noise. We form such a scalar process as the second component $(Z_t^{(2)})_{t\in\mathbb{N}}$ of the linear state space process $(Z_t)_{t\in\mathbb{N}}$ defined by the recursion

$$\underbrace{\begin{bmatrix} Z_{t+1}^{(1)} \\ Z_{t+1}^{(2)} \\ 1 \end{bmatrix}}_{Z_{t+1}} = \underbrace{\begin{bmatrix} 0 & 1 & 0 \\ 0.65 & 0.3 & N_{t+1} \\ 0 & 0 & 1 \end{bmatrix}}_{W_{t+1}} \underbrace{\begin{bmatrix} Z_t^{(1)} \\ Z_t^{(2)} \\ 1 \end{bmatrix}}_{Z_t},$$

where $(N_t)_{t\in\mathbb{N}}$ is a sequence of independent identically distributed random variables.

In this three-level storage example, the reward is given by the affine linear functions

$$r_t(p, (z^{(1)}, z^{(2)}, z^{(3)}), a) = (p - \alpha(p, a)) \cdot z^{(2)} - c|p - \alpha(p, a)|,$$

where $(p - \alpha(p, a))z^{(2)}$ represents the proceeds from the sale or purchase of the commodity and $c > 0$ denotes the transaction cost. Assuming that the storage must be returned to the owner at the pre-specified level $p = 2$, the terminal cash flow is then given by the scrap value

$$r_T(p, (z^{(1)}, z^{(2)}, z^{(3)})) = (p - \alpha(p, 2)) \cdot z^{(2)} - c|p - \alpha(p, 2)|.$$

Remark 4.1. Note that the convexity of the rewards is required in the continuous variable $z = (z^{(1)}, z^{(2)}, z^{(3)}) \in \mathbb{R}^3$ and satisfied because of affine-linearity in $z \in \mathbb{R}^3$.

The time horizon $t = 0, \ldots, T$ of the system is given by $T = 20$ and transaction costs are set at $c = 0.50$. The dynamics of the spot price is given by $(Z)_{t=0}^T$ and the innovations of the autoregressive process are distributed uniformly on the interval $[-1, 1]$. The grid is created by simulating samples of $(Z)_{t=0}^T$ and storing their values at each of the time steps. We shall initialize the first two elements of our state variable, $Z_0^{(1)}$ and $Z_0^{(2)}$, with the same initial price, $z_0 = Z_0^{(1)} = Z_0^{(2)}$.

https://doi.org/10.1017/S1446181115000279 Published online by Cambridge University Press

TABLE 1. 99% confidence intervals for the value of storage facility for different starting positions and commodity prices.

z_0	p_0	Lower bound confidence intervals	Upper bound confidence intervals
0	1	(0.3090, 0.3213)	(0.3111, 0.3238)
	2	(0.3260, 0.3384)	(0.3279, 0.3407)
	3	(0.3097, 0.3219)	(0.3118, 0.3244)
5	1	(−2.906, −2.8898)	(−2.9003, −2.8825)
	2	(1.594, 1.6102)	(1.5997, 1.6175)
	3	(6.094, 6.1102)	(6.0997, 6.1175)
10	1	(−5.7056, −5.6903)	(−5.7037, −5.6875)
	2	(3.7944, 3.8097)	(3.7963, 3.8125)
	3	(13.2944, 13.3097)	(13.2963, 13.3125)
15	1	(−8.442, −8.4319)	(−8.4413, −8.431)
	2	(6.058, 6.0681)	(6.0587, 6.069)
	3	(20.558, 20.5681)	(20.5587, 20.569)
20	1	(−11.1678, −11.1582)	(−11.1675, −11.1579)
	2	(8.3322, 8.3418)	(8.3325, 8.3421)
	3	(27.8322, 27.8418)	(27.8325, 27.8421)

Results were computed using $m = 1024$ grid points, $n = 1024$ disturbances and $K = 1024$ paths for diagnostics.

For bound computations, we use confidence intervals based on K simulated trajectories. More precisely, we quote the intervals as

$$\left[\underline{\mu} - \Phi^{-1}\left(1 - \frac{x}{2}\right)\frac{\underline{\sigma}}{\sqrt{K}}, \overline{\mu} + \Phi^{-1}\left(1 - \frac{x}{2}\right)\frac{\overline{\sigma}}{\sqrt{K}}\right],$$

where $1 - x$ denotes the confidence level and $(\underline{\mu}, \underline{\sigma})$ and $(\overline{\mu}, \overline{\sigma})$ denote the sample mean and sample standard deviation of $(\underline{v}_0^{\tilde{\pi},\varphi}(p, z_0^k))_{k=1}^K$ and $(\overline{v}_0^{\varphi}(p, z_0^k))_{k=1}^K$, respectively.

Results for different starting levels of z_0 and p_0 are given in Table 1. We included the result for $z_0 = 0$ in order to verify if the method works. In this case, the values for $p_0 = 1$ should be the same as those for $p_0 = 3$ due to the symmetry of the problem. In Table 1, we indeed observe an agreement of the results up to three decimal places. For other values of z_0, we obtain fairly tight confidence intervals, even though a relatively low number of sample paths have been used to perform diagnostics.

5. Conclusion

In this paper, we have demonstrated a new method of solving stochastic switching problems when the continuous component of state variables is high dimensional with linear state dynamics. By using duality and variance reduction techniques, we have

https://doi.org/10.1017/S1446181115000279 Published online by Cambridge University Press

provided methods to assess the distance to optimality of the approximated solutions. Finally, we have shown that these methods can give appropriate numerical results for storage management problems.

Acknowledgement

This research was supported under the Australian Research Council's Discovery Projects funding scheme (project no. DP130103315).

References

[1] L. Andersen and M. Broadie, "A primal–dual simulation algorithm for pricing multidimensional American options", *Manag. Sci.* **50** (2004) 1222–1234; doi:10.1287/mnsc.1040.0258.

[2] N. Bäuerle and U. Rieder, *Markov decision processes with applications to finance* (Springer, Heidelberg, 2011); doi:10.1007/978-3-642-18324-9 2.

[3] N. Belomestny, A. Kolodko and J. Schoenmakers, "Regression methods for stochastic control problems and their convergence analysis", *SIAM J. Control Optim.* **48** (2010) 3562–3588; doi:10.1137/090752651.

[4] D. P. Bertsekas, *Dynamic programming and optimal control* (Athena Scientific, Belmont, MA, 2005).

[5] R. Carmona and M. Ludkovski, "Valuation of energy storage: an optimal switching approach", *Quant. Finance* **10** (2010) 359–374; doi:10.1080/14697680902946514.

[6] E. A. Feinberg and A. Schwartz, "Handbook of Markov decision processes", *Internat. Ser. Oper. Res. Management Sci.* (2002); doi:10.1007/978-1-4615-0805-2.

[7] M. Haugh and L. Kogan, "Pricing American options: a duality approach", *Oper. Res.* **52** (2004) 258–270; doi:10.1287/opre.1030.0070.

[8] J. Hinz, "Optimal stochastic switching under convexity assumptions", *SIAM J. Control Optim.* **52** (2014) 164–188; doi:10.1137/13091333X.

[9] J. Hinz and N. Yap, "Algorithms for optimal control of stochastic switching systems", *Theory Probab. Appl.* (to appear).

[10] F. Longstaff and E. Schwartz, "Valuing American options by simulation: a simple least-squares approach", *Rev. Financ. Stud.* **14** (2001) 113–147; http://links.jstor.org/sici?sici=0893-94542820012129143A13C1133AVAOBSA3E2.0.CO3B2-W.

[11] W. B. Powell, *Approximate dynamic programming: solving the curses of dimensionality* (Wiley, Hoboken, NJ, 2007).

[12] M. L. Puterman, *Markov decision processes: discrete stochastic dynamic programming* (Wiley, New York, 1994).

[13] L. C. G. Rogers, "Monte Carlo valuation of American options", *Math. Finance* **12** (2002) 271–286; doi:10.1111/1467-9965.02010.

[14] L. C. G. Rogers, "Pathwise stochastic optimal control", *SIAM J. Control Optim.* **46** (2007) 1116–1132; doi:10.1137/050642885.

[15] F. Ye and E. Zhou, "Optimal stopping of partially observable Markov processes: a filtering-based duality approach", *IEEE Trans. Automat. Control* **58** (2013) 2698–2704; doi:10.1109/TAC.2013.2257970.