An iterative procedure for constructing subsolutions of discrete-time optimal control problems

Markus Fischer*†

version of November, 2011

Abstract

An iterative procedure for constructing subsolutions of deterministic or stochastic optimal control problems in discrete time with continuous state space is introduced. The procedure generates a non-decreasing sequence of subsolutions, giving true lower bounds on the minimal costs. Convergence of the values at any fixed initial state is shown.

2000 AMS subject classifications: 60J05, 90-08, 90C39, 90C40, 93E20.

Key words and phrases: deterministic / stochastic optimal control, Markov decision process, dynamic programming, discrete-time subsolution.

1 Introduction

In this article we introduce a scheme for constructing subsolutions of deterministic or stochastic optimal control problems in discrete time with general state space.

The optimal control problems considered here belong to the class of semi-continuous models studied in Bertsekas and Shreve (1996, Sect. 8.3). We restrict

*Department of Pure and Applied Mathematics, University of Padua, via Trieste 63, 35121 Padova, Italy. Research supported by the German Research Foundation (DFG research fellowship), the National Science Foundation (DMS-0706003), and the Air Force Office of Scientific Research (FA9550-07-1-0544, FA9550-09-1-0378).

†Discussions with Paul Dupuis and Hui Wang were the starting point for the present work; the author thanks them for their support. The author is grateful to Wendell H. Fleming and Konstantinos Spiliopoulos for their comments and helpful suggestions.

attention to systems controlled over a finite time horizon. Performance is measured in terms of expected costs, which are to be minimized. The procedure generates a non-decreasing sequence of subsolutions. In each step, state trajectories starting from a fixed initial state are computed which serve to update the current subsolution. The values of the subsolutions at the initial state will be shown to converge from below to the minimal costs, or value function, of the control problem.

Using subsolutions as approximations to the value function automatically yields one-sided *a posteriori* error bounds, namely, lower bounds on the value function. For minimization problems, upper bounds can always be obtained by choosing any control policy and calculating the associated costs. The subsolutions produced by our scheme are also used for control synthesis, and the costs associated with the resulting policies turn out, under mild conditions, to converge to the minimal costs. In the case of stochastic problems, calculating expected costs for any fixed control policy will typically involve Monte Carlo simulation: State trajectories are computed according to the dynamics and the given control policy based on samples of the noise random variables. The resulting cost estimates are upper bounds on the value function to within a Monte Carlo error margin only. In the case of deterministic dynamics, there is no Monte Carlo error. Subsolutions, on the other hand, nowhere exceed the value function and thus always give perfectly reliable lower bounds.

For the class of control problems considered here, the Principle of Dynamic Programming (PDP) holds. Since time is discrete, the control problems can in theory be solved by backward recursion according to the PDP. This yields an implementable and feasible method only if the state space is finite and the number of states not too big. The state space would have to be discretized if it were, say, a region in $\mathbb{R}^d$, as is the case for a large class of controlled continuous-time stochastic processes and their discrete-time analogues, and the number of discrete states would grow exponentially in the dimension d. This is related to the so-called curse of dimensionality, which affects broad classes of stochastic optimal control problems (cf. Chow and Tsitsiklis, 1989).

The scheme proposed here avoids discretization of the state space. By producing two-sided *a posteriori* error bounds, it provides approximation guarantees for each individual control problem. While the method is not simply an application of dynamic programming, the PDP is an essential ingredient in the proof of convergence. We present the scheme and give a proof of convergence in a general context, our main assumption on the control problems being that

of preservation of Lipschitz continuity by the associated Bellman operators. As discussed in Section 6, the method can be adapted to other classes of control problems, which are characterized in terms of certain regularity properties preserved by their Bellman operators (Lipschitz continuity here, convexity in the case of convex control problems).

Subsolutions will be represented as pointwise maxima of certain base functions, in this paper, conical surfaces, a choice related to the assumption of preservation of Lipschitz continuity. Representation of functions as pointwise maxima or minima of certain elementary functions plays an important role in the application of max-plus (or min-plus) methods to optimal control; see, for instance, McEneaney (2011), where a numerical procedure for a class of discrete-time stochastic optimal control problems is developed.

In Dupuis and Wang (2007) efficient dynamic importance sampling schemes are constructed based on subsolutions. The optimization problem there is a deterministic differential game in continuous time, and subsolutions are constructed for the outer maximization problem. The corresponding concept in the present context would be that of supersolutions.

A classical approach to optimal control is via the linear programming formulation (e.g. Hernández-Lerma and Lasserre, 1996, Ch. 6). The approach replaces the original control problem by a pair of infinite-dimensional constrained linear problems. This representation can serve as a starting point for building approximation procedures; see, for instance, Helmes and Stockbridge (2008) and the references therein. A numerical procedure yielding two-sided *a posteriori* error bounds for a special class of controlled continuous-time deterministic systems is developed in Lasserre et al. (2008).

A reformulation for problems of optimal control of possibly degenerate Itô diffusions in terms of linear minimization problems with convex constraints is given in Fleming and Vermes (1989). Based on convex duality and the equivalence between the original problem and its reformulation, it is shown that the value function of the original problem can be represented as the pointwise supremum over all smooth subsolutions of the associated Hamilton-Jacobi-Bellman equation. In Hernández-Hernández et al. (1996), based on linear programming, the authors show that for a broad class of continuous-time, finite horizon deterministic models, the value function is the limit of a sequence of smooth approximate subsolutions.

A different approach to discrete-time stochastic control problems, which allows to compute two-sided bounds, has recently been proposed by Rogers (2007),

also see Belomestny et al. (2010). It is assumed that the effect of the control on the dynamics can be represented as a change of measure w.r.t. some reference probability measure which can be thought of as the law of an uncontrolled Markov chain. Under this hypothesis, the problem of minimizing expected costs (maximizing expected gain) for a fixed initial state over all non-anticipating strategies can be rewritten as a pathwise optimization problem over deterministic strategies. The constraint of non-anticipativity of strategies is accounted for by a Lagrangian martingale term.

A Lagrange multiplier formulation is also the basis for the approach taken in Kuhn (2009), where certain continuous-time stochastic control problems are approximated using stage aggregation, producing upper as well as lower bounds on the minimal costs. The resulting optimization problems are meant to be solved by scenario-tree methods.

The rest of the paper is organized as follows. In Section 2 we describe the class of discrete-time optimal control problems and discuss basic properties. The procedure for constructing subsolutions is introduced in Section 3. Convergence is first studied for the special and simpler case of deterministic systems in Section 4, and then in Section 5 in greater generality for stochastic systems. Section 6 contains remarks on implementation, modifications and possible extensions of the procedure. In Appendix A we compare our procedure to the classical method known as policy iteration or approximation in policy space (cf. Puterman, 1994, p. 264). Appendix B contains some calculations regarding discrete-time approximations to controlled Itô diffusions.

2 The class of control problems

The dynamics of the optimal control problems are described as controlled time inhomogeneous Markov chains with *state space* $\mathcal{X}$, *action space* Γ, and *disturbance space* $\mathcal{Y}$. Here, $\mathcal{X}$ and $\mathcal{Y}$ are Borel subsets of complete and separable metric spaces $(\bar{\mathcal{X}}, \rho_X)$ and $(\bar{\mathcal{Y}}, \rho_Y)$, respectively, and (Γ, ρ_Γ) is a separable metric space assumed to be compact. The evolution of the system is determined by a Borel measurable function $\Psi\colon \mathbb{N}_0 \times \mathcal{X} \times \Gamma \times \mathcal{Y} \to \mathcal{X}$, the *system function*.

Let μ be a probability measure on the Borel σ-algebra of $\mathcal{Y}$; μ will be called *noise distribution*. Let us call *disturbance* or *noise variable* any $\mathcal{Y}$-valued random variable that has distribution μ. Define the set $\mathcal{U}$ of *Markov feedback control policies* as the set of all functions $u : \mathbb{N}_0 \times \mathcal{X} \to \Gamma$ which are Borel measurable.

In order to determine the pathwise evolution of the system state, choose a

complete probability space $(\Omega, \mathcal{F}, \mathbf{P})$ carrying an independent sequence $(\xi_j)_{j\in\mathbb{N}}$ of noise variables. Given any control policy $u \in \mathcal{U}$, initial state $x_0 \in \mathcal{X}$, and initial time $j_0 \in \mathbb{N}_0$, the corresponding *state sequence* is recursively defined, for each $\omega \in \Omega$, by

$$X_{j_0}(\omega) \doteq x_0, \quad X_{j+1}(\omega) \doteq \Psi\big(j, X_j(\omega), u(j, X_j(\omega)), \xi_{j+1}(\omega)\big), \quad j \geq j_0. \tag{1}$$

Performance is measured in terms of expected costs over a finite deterministic *time horizon*, denoted by $N \in \mathbb{N}$. Let $f : \mathbb{N}_0 \times \mathcal{X} \times \Gamma \to \mathbb{R}$, $F : \mathcal{X} \to \mathbb{R}$ be lower semicontinuous functions bounded from below, the *running costs* and the *terminal costs*, respectively. Define the *cost functional* $J\colon \{0, \ldots, N\} \times \mathcal{X} \times \mathcal{U} \to \mathbb{R} \cup \{\infty\}$ by setting

$$J(j_0, x_0, u) \doteq \mathbf{E}\left[\sum_{j=j_0}^{N-1} f\big(j, X_j, u(j, X_j)\big) + F(X_N)\right], \tag{2}$$

where $(X_j)_{j\geq j_0}$ is the state sequence generated according to Eq. (1) with control policy u and $X_{j_0} = x_0$. Notice that J depends on the noise variables only through their distribution and does not depend on the particular choice of the underlying probability space.

The *value function* $V\colon \{0, \ldots, N\} \times \mathcal{X} \to \mathbb{R} \cup \{\infty\}$ of the problem is defined by

$$V(j_0, x_0) \doteq \inf_{u\in\mathcal{U}} J(j_0, x_0, u). \tag{3}$$

For any $j \in \mathbb{N}_0$, define an operator $\mathcal{L}_j$, the *one-step Bellman operator* at time j, acting on functions $\varphi\colon \mathcal{X} \to \mathbb{R} \cup \{\infty\}$ according to

$$\mathcal{L}_j(\varphi)(x) \doteq \inf_{\gamma\in\Gamma} \left\{ f(j, x, \gamma) + \int_{\mathcal{Y}} \varphi\big(\Psi(j, x, \gamma, y)\big)\mu(dy) \right\}, \qquad x \in \mathcal{X}. \tag{4}$$

The application of $\mathcal{L}_j$ to φ produces a function $\mathcal{L}_j(\varphi) : \mathcal{X} \to \mathbb{R} \cup \{\infty\}$ that is lower semicontinuous and bounded from below whenever $\varphi\colon \mathcal{X} \to \mathbb{R} \cup \{\infty\}$ is lower semicontinuous and bounded from below.

Let us make the following assumptions:

(A1) The system function Ψ is Borel measurable and $\Psi(j, x, ., .)$ is continuous on $\Gamma \times \mathcal{Y}$ for each $(j, x) \in \mathbb{N}_0 \times \mathcal{X}$.

(A2) The space of control actions Γ is compact.

(A3) The cost coefficients f, F are nonnegative and lower semicontinuous.

(A4) Costs are finite: $J(0,x,u) < \infty$ for all $x \in \mathcal{X}$, $u \in \mathcal{U}$.

(A5) Lipschitz continuity is preserved: There are constants $c_0, c_1 > 0$ such that, whenever $\varphi\colon \mathcal{X} \to \mathbb{R}$ is globally Lipschitz continuous with Lipschitz constant L_φ, then, for any $j \in \{0,\dots,N-1\}$, $\mathcal{L}_j(\varphi)$ is globally Lipschitz continuous with Lipschitz constant not greater than $c_0 + L_\varphi(1+c_1)$.

Assumptions (A1)–(A4) guarantee that optimal Markov feedback policies exist and that the Principle of Dynamic Programming holds.

Lemma 1 (PDP)**.** *Grant Assumptions (A1)–(A4). Then for all $x \in \mathcal{X}$ and all $j \in \{0,\dots,N-1\}$,*

$$V(j,x) = \mathcal{L}_j\big(V(j+1,.)\big)(x).$$

Moreover, an optimal Markov feedback control policy exists, that is, there is $u \in \mathcal{U}$ such that $V(j,x) = J(j,x,u)$ for all $(j,x) \in \{0,\dots,N-1\} \times \mathcal{X}$.

Proof. The assertion follows from Proposition 8.6 in Bertsekas and Shreve (1996, p. 209). Notice that, in our set-up, there is an explicit time variable, namely, the index $j \in \{0,\dots,N\}$. □

The class of control policies can be enlarged in different ways without changing the value function. In Section 5 we will need to consider the following weak formulation of our control problems; cf. Definition 2.4.2 in Yong and Zhou (1999, p. 64) in the context of continuous-time models.

Let $\hat{\mathcal{U}}$ denote the set of all quadruples consisting of a complete probability space $(\Omega, \mathcal{F}, \mathbf{P})$, a filtration $(\mathcal{F}_j)_{j\in\mathbb{N}_0}$ in $\mathcal{F}$, a sequence $(\xi_j)_{j\in\mathbb{N}}$ of noise variables, and a sequence $(\gamma_j)_{j\in\mathbb{N}_0}$ of Γ-valued random variables on $(\Omega,\mathcal{F})$ such that ξ_j is $\mathcal{F}_j$-measurable, ξ_{j+l} is independent of $\mathcal{F}_j$, all $j,l \in \mathbb{N}$, and γ_j is $\mathcal{F}_j$-measurable, all $j \in \mathbb{N}_0$. The sequence (γ_j) will be referred to as a *random control policy.* The dependence of (γ_j) on the stochastic basis (including the noise variables) may be suppressed, and we will write $(\gamma_j) \in \hat{\mathcal{U}}$ instead of $((\Omega,\mathcal{F},\mathbf{P}),(\mathcal{F}_j),(\xi_j),(\gamma_j)) \in \hat{\mathcal{U}}$.

Given a random control policy $(\gamma_j)_{j\in\mathbb{N}_0} \in \hat{\mathcal{U}}$, $x_0 \in \mathcal{X}$, $j_0 \in \mathbb{N}_0$, the corresponding *state sequence* is recursively defined, for each $\omega \in \Omega$, by

$$X_{j_0}(\omega) \doteq x_0, \qquad X_{j+1}(\omega) \doteq \Psi\big(j, X_j(\omega), \gamma_j(\omega), \xi_{j+1}(\omega)\big), \qquad j \geq j_0. \tag{5}$$

The costs associated with such a state sequence and random control policy are given by

$$\hat{J}\big(j_0, x_0, (\gamma_j)\big) \doteq \mathbf{E}\left[\sum_{j=j_0}^{N-1} f(j, X_j, \gamma_j) + F(X_N)\right], \tag{6}$$

where expectation is taken w.r.t. the probability measure of the stochastic basis coming with (γ_j). The *value function* in the weak formulation is defined by

$$\hat{V}(j_0, x_0) \doteq \inf_{(\gamma_j)\in\hat{\mathcal{U}}} \hat{J}\big(j_0, x_0, (\gamma_j)\big), \quad j_0 \in \{0, \dots, N\},\ x_0 \in \mathcal{X}. \tag{7}$$

Under Assumptions (A1) – (A4), the value function V defined in (3) over the class of Markov feedback control policies with fixed stochastic basis coincides with $\hat{V}$, the value function defined in (7) over the class of random control policies with varying stochastic basis. Though results of this type are standard, we give a proof for the sake of completeness.

Lemma 2. *Grant Assumptions (A1) – (A4). Then $V = \hat{V}$.*

Proof. Any Markov feedback strategy $u \in \mathcal{U}$ induces a random control policy $(\gamma_j) \in \hat{\mathcal{U}}$ on the given probability space which gives rise to the same state sequence and the same associated costs. It follows that $\hat{V} \le V$.

We show that $\hat{V} \ge V$ by backward induction. By construction, $V(N, .) = F(.) = \hat{V}(N, .)$. Suppose that $\hat{V}(i{+}1, .) \ge V(i{+}1, .)$ for some $i \in \{0, \dots, N-1\}$. It is enough to show that this implies $\hat{J}(i, ., (\gamma_j)) \ge V(i, .)$ for all $(\gamma_j) \in \hat{\mathcal{U}}$. Let $((\Omega, \mathcal{F}, \mathbf{P}), (\mathcal{F}_j), (\xi_j), (\gamma_j)) \in \hat{\mathcal{U}}$, $x_0 \in \mathcal{X}$, and let X be determined by (5) with $j_0 = i$. Then

$$\begin{aligned}
\hat{J}\big(i, x_0, (\gamma_j)\big) &= \mathbf{E}\left[f(i, x_0, \gamma_i) + \sum_{j=i+1}^{N-1} f(j, X_j, \gamma_j) + F(X_N)\right] \\
&= \mathbf{E}\big[f(i, x_0, \gamma_i)\big] \\
&\quad + \int_{\mathcal{X}} \mathbf{E}\left[\sum_{j=i+1}^{N-1} f(j, X_j, \gamma_j) + F(X_N) \;\Big|\; X_{i+1} = x\right] \mathbf{P}_{X_{i+1}}(dx) \\
&= \mathbf{E}\big[f(i, x_0, \gamma_i)\big] + \int_{\mathcal{X}} \hat{J}\big(i{+}1, x, (\gamma_j^{i,x})\big)\, \mathbf{P}_{X_{i+1}}(dx) \\
&\ge \mathbf{E}\big[f(i, x_0, \gamma_i)\big] + \int_{\mathcal{X}} V(i{+}1, x)\, \mathbf{P}_{X_{i+1}}(dx),
\end{aligned}$$

where $(\gamma_j^{i,x})$ indicates the random control policy induced by (γ_j) under the probability measure $\mathbf{P}(.|X_{i+1} = x)$. More precisely, $(\gamma_j^{i,x})$ stands for the random control policy $((\Omega, \mathcal{F}, \mathbf{P}(.|X_{i+1} = x), (\mathcal{F}_j), (\xi_j), (\gamma_j)) \in \hat{\mathcal{U}}$. Notice that the expression $\sum_{j=i+1}^{N-1} f(j, X_j, \gamma_j) + F(X_N)$ depends on a noise variable ξ_j only if $j > i+1$, and in this case the law of ξ_j under $\mathbf{P}(.|X_{i+1} = x)$ is μ by independence. The inequality in the last line of the display above holds because $\hat{J}(i+1, x, (\tilde{\gamma}_j)) \geq V(i+1, x)$ for all $x \in \mathcal{X}$ and all $(\tilde{\gamma}_j) \in \hat{\mathcal{U}}$ by induction hypothesis. It follows that

$$\begin{aligned}
\hat{J}\big(i, x_0, (\gamma_j)\big) &\geq \mathbf{E}\big[f(i, x_0, \gamma_i)\big] + \int_{\mathcal{X}} V(i+1, x)\, \mathbf{P}_{X_{i+1}}(dx) \\
&= \mathbf{E}\left[f(i, x_0, \gamma_i) + \int_{\mathcal{Y}} V\big(i+1, \Psi(i, x_0, \gamma_i, y)\big)\mu(dy)\right] \\
&\geq \inf_{\gamma \in \Gamma}\left\{f(i, x_0, \gamma) + \int_{\mathcal{Y}} V\big(i+1, \Psi(i, x_0, \gamma, y)\big)\mu(dy)\right\}.
\end{aligned}$$

The right-hand side of the last line above equals $\mathcal{L}_i(V(i+1, .))(x_0)$, which by Lemma 1 is equal to $V(i, x_0)$. □

An important class of discrete-time control problems results from continuous-time problems by discretization of time. Consider the optimal control of an Itô diffusion over a finite horizon $T > 0$. The dynamics of the continuous-time problem are of the form

$$dX(t) = b\big(t, X(t), u(t)\big)dt + \sigma\big(t, X(t), u(t)\big)dW(t), \quad t > 0, \tag{8}$$

where $W(.)$ is a d_1-dimensional standard Wiener process, the drift coefficient b is a function $[0, \infty) \times \mathbb{R}^d \times \Gamma \to \mathbb{R}^d$, and the diffusion coefficient σ is a function $[0, \infty) \times \mathbb{R}^d \times \Gamma \to \mathbb{R}^{d \times d_1}$. The process $u(.)$ in Eq. (8) is any Γ-valued process which is progressively measurable w.r.t. the filtration induced by the Wiener process. The corresponding costs are given by

$$J(t_0, x_0, u(.)) \doteq \mathbf{E}\left[\int_{t_0}^{T} \tilde{f}\big(t, X(t), u(t)\big)dt + \tilde{F}(X(T))\right], \tag{9}$$

where the cost coefficients $\tilde{f}$ and $\tilde{F}$ are functions $[0, \infty) \times \mathbb{R}^d \times \Gamma \to \mathbb{R}$ and $\mathbb{R}^d \to \mathbb{R}$, respectively, and $X(.)$ solves Eq.(8) with initial condition $(t_0, x_0) \in [0, T] \times \mathbb{R}^d$ under policy $u(.)$.

A straightforward discretization of time with constant mesh size $h > 0$ leads to the following discrete-time control problem: Choose the state space $\mathcal{X}$ to

be $\mathbb{R}^d$, the disturbance space $\mathcal{Y}$ to be $\mathbb{R}^{d_1}$, each with the metric induced by Euclidean distance, and take the original space of control actions Γ. Define the system function $\Psi\colon \mathbb{N}_0 \times \mathbb{R}^d \times \Gamma \times \mathbb{R}^{d_1} \to \mathbb{R}^d$ by

$$\Psi(j, x, \gamma, y) \doteq x + h \cdot b(jh, x, \gamma) + \sqrt{h} \cdot \sigma(jh, x, \gamma) y. \tag{10}$$

As noise distribution μ we may choose the d_1-variate standard normal distribution. An alternative approximation is obtained when the normal distributions are replaced with the d_1-fold product of Bernoulli distributions concentrated on $\{-1, 1\}$. More generally, we may take $\mu \doteq \otimes^{d_1} \nu$, where ν is any probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ with mean zero and variance one. Last, define the time horizon N and the cost coefficients f, F of the discrete-time problem by setting $N \doteq \lfloor \frac{T}{h} \rfloor$ and, for $j \in \mathbb{N}_0$, $x \in \mathbb{R}^d$, $\gamma \in \Gamma$, $f(j, x, \gamma) \doteq h \cdot \tilde{f}(jh, x, \gamma)$, $F(x) \doteq \tilde{F}(x)$.

Suppose that the coefficients b, σ, $\tilde{f}$ of the continuous-time problem are jointly measurable, continuous in the state and control variable, and Lipschitz continuous in the state variable, uniformly in time and control, with Lipschitz constants L_b, L_σ, $L_{\tilde{f}}$, respectively. Suppose, in addition, that $\tilde{f}$, $\tilde{F}$ are non-negative and $\tilde{F}$ has at most polynomial growth. Then the discrete-time control problem just described satisfies Assumptions (A1) – (A5). In particular, Assumption (A5) is fulfilled if one chooses

$$c_0 \doteq L_{\tilde{f}} h, \qquad\qquad c_1 \doteq (2L_b + L_\sigma^2)h + L_b^2 h^2, \tag{11}$$

see Appendix B. Notice that the constants c_0, c_1 are both of order one in h as h tends to zero. This result can be seen as a simplified version of Theorem 4.1 in Krylov (1980, p. 165), where a bound on the norm of the (generalized) gradient of the value function of the original continuous-time model is given. If the diffusion coefficient σ does not depend on the state, then we simply take $c_1 \doteq L_b h$.

3 A scheme for constructing subsolutions

Consider a control problem of the form described in Section 2. Recall that the function F gives the terminal costs.

Definition 1. A function $w\colon \{0, \ldots, N\} \times \mathcal{X} \to \mathbb{R}$ which is lower semicontinuous and bounded from below is called a *subsolution* of the control problem iff the following two properties hold:

(i) $w(N, x) \leq F(x)$ for all $x \in \mathcal{X}$,

(ii) $w(j,x) \le \mathcal{L}_j\big(w(j+1,.)\big)(x)$ for all $j \in \{0,\dots,N-1\}$, $x \in \mathcal{X}$.

As a consequence of Lemma 1 and the monotonicity of the one-step Bellman operators $\mathcal{L}_j$, we have $w \le V$ for any subsolution w, that is, a subsolution nowhere exceeds the value function of the problem. Also observe that if w, $\tilde{w}$ are subsolutions, then so is $\max\{w,\tilde{w}\}$, the pointwise maximum of w, $\tilde{w}$.

By Assumption (A3), all costs are nonnegative. We can easily construct a subsolution w by setting $w(j,.) = 0$ for all $j \in \{0,\dots,N-1\}$ and choosing $w(N,.)$ such that $w(N,.)$ is lower semicontinuous and $0 \le w(N,x) \le F(x)$ for all $x \in \mathcal{X}$. In particular, $w \equiv 0$ as well as $\tilde{w}(j,.) = 0$, $j \in \{0,\dots,N-1\}$, $\tilde{w}(N,.) = F(.)$ are both subsolutions of the control problem.

Starting from a nonnegative Lipschitz continuous subsolution, the scheme iteratively produces a non-decreasing sequence of Lipschitz continuous subsolutions. Each iteration step of the scheme consists of two parts. In the first part, the simulation part, state trajectories are generated by forward simulation (forward in time) with fixed initial state according to the dynamics under a control policy which is selected based on the current subsolution. In the second part, the update part, a new subsolution is constructed from the previous one by backward recursion along the state trajectories computed in the simulation part. For the sake of simplicity, we first introduce the scheme using exactly one state trajectory at each iteration step.

Choose $x_0 \in \mathcal{X}$, the fixed initial state. Assume that the function F quantifying the terminal costs is Lipschitz continuous with Lipschitz constant L_F. Let $w^{(0)}$ be a subsolution such that $w^{(0)}(j,.)$ is Lipschitz continuous with constant $L_j^{(0)}$, each $j \in \{0,\dots,N\}$, and $L_N^{(0)} \le L_F$ and $L_j^{(0)} \le c_0 + L_{j+1}^{(0)}(1+c_1)$ if $j \in \{0,\dots,N-1\}$; for instance, take $w^{(0)} \equiv 0$.

The building elements for the subsolutions will be functions corresponding to the surface of an upward pointing (or downward opening) cone in $\mathcal{X} \times \mathbb{R}$. More precisely, given $\tilde{x} \in \mathcal{X}$, $v \in \mathbb{R}$, $L > 0$, we define the *cone function* with center $\tilde{x}$, height v, and slope L by setting

$$\mathrm{Cone}(\tilde{x}, v, L)(x) \doteq v - L \cdot \rho_X(\tilde{x}, x), \quad x \in \mathcal{X},$$

where ρ_X is the metric on $\mathcal{X}$. Recall that, according to Assumption (A5), the one-step Bellman operators preserve Lipschitz continuity. If $\varphi : \mathcal{X} \to \mathbb{R}$ is Lipschitz continuous with constant not exceeding L, then, fixing $\tilde{x} \in \mathcal{X}$, we have $\varphi(x) \ge \mathrm{Cone}(\tilde{x}, \varphi(\tilde{x}), L)(x)$ for all $x \in \mathcal{X}$, and $\varphi(\tilde{x}) = \mathrm{Cone}(\tilde{x}, \varphi(\tilde{x}), L)(\tilde{x})$. If $\varphi_1, \dots, \varphi_n$ are Lipschitz continuous with Lipschitz constants $L_1, \dots, L_n$, then

the mapping $\mathcal{X} \ni x \mapsto \max\{\varphi_1(x), \ldots, \varphi_n(x)\}$ is Lipschitz continuous with constant not greater than $\max\{L_1, \ldots, L_n\}$. A Lipschitz continuous function can therefore be arbitrarily well approximated from below on any compact set by the pointwise maximum of a finite number of cone functions. This together with Assumption (A5) is the reason we use maxima of cone functions in constructing subsolutions. Building elements different from cone functions can be used if the Bellman operators preserve other types of regularity instead of (or in addition to) Lipschitz continuity; cf. the discussion in Section 6.

Let $(\xi_j^{(n)})_{j,n\in\mathbb{N}}$ be an independent collection of noise variables defined on a complete probability space $(\Omega, \mathcal{F}, \mathbf{P})$. Starting from $w^{(0)}$ the scheme produces, for each scenario $\omega \in \Omega$, a sequence $(w_\omega^{(n)})_{n\in\mathbb{N}}$ of functions $\{0, \ldots, N\} \times \mathcal{X} \to \mathbb{R}$. At stage $n \in \mathbb{N}_0$, the function $w_\omega^{(n)}$ is nonnegative and, for each $j \in \{0, \ldots, N\}$, $w_\omega^{(n)}(j, .)$ is Lipschitz continuous with Lipschitz constant bounded by $L_j^{(n)}$. The following procedure, which is a particular instance of the general scheme and will be referred to as the *full cone construction* with one trajectory at each iteration, constructs $w_\omega^{(n+1)}$ from $w_\omega^{(n)}$, $n \in \mathbb{N}_0$, as follows:

a) Generate one trajectory of the state sequence starting in x_0 at time zero using a control policy induced by $w_\omega^{(n)}$: Set $X_0^n(\omega) \doteq x_0$ and for $j \in \{0, \ldots, N-1\}$,

$$X_{j+1}^n(\omega) \doteq \Psi\left(j, X_j^n(\omega), \gamma_j^n(\omega), \xi_{j+1}^{(n+1)}(\omega)\right),$$

$$\gamma_j^n(\omega) \in \operatorname{argmin}_{\gamma\in\Gamma}\Big\{ f(j, X_j^n(\omega), \gamma) + \int_{\mathcal{Y}} w_\omega^{(n)}\big(j+1, \Psi(j, X_j^n(\omega), \gamma, y)\big)\mu(dy)\Big\}.$$

b) Construct a function $w_\omega^{(n+1)} : \{0, \ldots, N\} \times \mathcal{X} \to \mathbb{R}$ by backward recursion in the following way:

– Set for $x \in \mathcal{X}$

$$w_\omega^{(n+1)}(N, x) \doteq \max\left\{ w_\omega^{(n)}(N, x), \operatorname{Cone}\left(X_N^n(\omega), F(X_N^n(\omega)), L_F\right)(x)\right\}.$$

– For j running from $N-1$ down to 0 do:

- compute $v_j^n(\omega) \doteq \mathcal{L}_j(w_\omega^{(n+1)}(j+1, .))(X_j^n(\omega))$,
- let $L \doteq c_0 + L_{j+1}^{(n+1)}(1 + c_1)$ and set for $x \in \mathcal{X}$

$$w_\omega^{(n+1)}(j, x) \doteq \max\left\{ w_\omega^{(n)}(j, x), \operatorname{Cone}\left(X_j^n(\omega), v_j^n(\omega), L\right)(x)\right\}.$$

The constants c_0, c_1 appearing above are chosen according to Assumption (A5). Some comments concerning the procedure are in order.

Remark 1. The functions $w_\omega^{(n)} : \{0,\dots,N\} \times \mathcal{X} \to [0,\infty)$, $n \in \mathbb{N}$, depend on $\omega \in \Omega$; we might occasionally suppress this dependence and simply write $w^{(n)}$ in place of $w_\omega^{(n)}$. The functions $w_\omega^{(n)}(j,.) : \mathcal{X} \to [0,\infty)$ are Lipschitz continuous with Lipschitz constants uniformly bounded in $n \in \mathbb{N}$, $\omega \in \Omega$, since for all $j \in \{0,\dots,N\}$,

$$L_j^{(n)} \le L_F \left(1 + \frac{c_0}{c_1}\right)(1+c_1)^{N-j} \tag{12}$$

by construction, the assumption that F is Lipschitz continuous with constant L_F and the choice of $w^{(0)}$.

Remark 2. The only possible source of underspecification in the full cone construction is the argmin operation, that is, the choice of the minimizing control actions in the simulation part. We will assume that the procedure uses some fixed measurable deterministic rule to select, given $x \in \mathcal{X}$, $j \in \mathbb{N}_0$, and a nonnegative continuous function φ, exactly one control action $\gamma_\varphi^*(j,x)$ such that

$$\gamma_\varphi^*(j,x) \in \operatorname{argmin}_{\gamma\in\Gamma} \left\{ f(j,x,\gamma) + \int_{\mathcal{Y}} \varphi\big(\Psi(j,x,\gamma,y)\big)\mu(dy) \right\}.$$

Thus, we assume that we have chosen a Borel measurable mapping

$$\mathbf{C}_+(\mathcal{X}) \times \mathbb{N}_0 \times \mathcal{X} \ni (\varphi,j,x) \mapsto \gamma_\varphi^*(j,x) \in \Gamma,$$

where $\mathbf{C}_+(\mathcal{X})$ is the space of all nonnegative continuous functions endowed with the maximum norm topology. Assumptions (A1)–(A4) entail that such measurable selectors exist; see, for instance, Section 3.3 in Hernández-Lerma and Lasserre (1996, pp. 27-31). As a consequence, the full cone construction is a measurable construction in the sense that the mapping $\Omega \times \{0,\dots,N\} \times \mathcal{X} \to [0,\infty)$ defined by $(\omega,j,x) \mapsto w_\omega^{(n)}(j,x)$ is measurable w.r.t. $\mathcal{F}$ and the Borel σ-algebras involved. Moreover, $w^{(n)}(j,.)$ is measurable w.r.t. $\mathcal{F}_0^n \otimes \mathcal{B}(\mathcal{X})$, where $\mathcal{F}_0^n$ is the σ-algebra generated by the noise variables $\xi_i^{(k)}$, $k \in \{1,\dots,n\}$, $i \in \mathbb{N}$.

Remark 3. The control actions $\gamma_j^n(\omega)$, $\omega \in \Omega$, $j \in \{0,\dots,N\}$, chosen at iteration step $n \in \mathbb{N}_0$ correspond to a random control policy in the sense of Section 2. To see this, for $n \in \mathbb{N}_0$ and $j \in \mathbb{N}_0$, let $\mathcal{F}_j^n$ denote the σ-algebra generated by the noise variables $\xi_i^{(k)}$, $k \in \{1,\dots,n\}$, $i \in \mathbb{N}$, and $\xi_l^{(n+1)}$, $l \in \{1,\dots,j\}$; in particular $\mathcal{F}_0^0 = \{\emptyset,\Omega\}$. Notice that $\xi_l^{(n+1)}$ is independent of $\mathcal{F}_j^n$ if $l > j$.

For $j > N$ let γ_j^n be an arbitrary $\mathcal{F}_j^n$-measurable Γ-valued random variable on $(\Omega, \mathcal{F}, \mathbf{P})$. Then $((\Omega, \mathcal{F}, \mathbf{P}), (\mathcal{F}_j^n)_{j\in\mathbb{N}_0}, (\gamma_j^n)_{j\in\mathbb{N}_0}, (\xi_j^{(n+1)})_{j\in\mathbb{N}})$ is in $\hat{\mathcal{U}}$. Moreover, the state sequence X_j^n, $j \in \{0, \ldots, N\}$, computed in the simulation part of stage n satisfies Eq. (5) under (γ_j^n) with $X_0^n = x_0$.

In the update part of the full cone construction we define the function $w_\omega^{(n+1)}(j, .)$ to be equal to the pointwise maximum of $w_\omega^{(n)}(j, .)$ and the cone function $\mathrm{Cone}(X_j^n(\omega), v_j^n(\omega), L)(.)$. This condition in conjunction with Assumption (A5) guarantees that the subsolution property of the functions $w^{(n)}$ is preserved by the update part, as the proof of Proposition 1 shows. Actually, in order to produce a non-decreasing sequence of subsolutions we need only inequalities in the update part. In addition, instead of using just one trajectory at each iteration step, we may use $M \in \mathbb{N}$ state trajectories. A potential advantage of using several state trajectories simultaneously (instead of just one) lies in a possibly faster convergence of the scheme in the sense that the values computed in the update part through application of the one-step Bellman operators ($v_j^{i,n}$ below) might be higher and thus closer to the minimal costs than in a one-trajectory version that uses the same total number of state trajectories.

Choose $M \in \mathbb{N}$, the number of simultaneous state trajectories, and let $\xi_j^{(i,n)}$, $j, n \in \mathbb{N}$, $i \in \{1, \ldots, M\}$, be an independent collection of noise variables defined on some complete probability space $(\Omega, \mathcal{F}, \mathbf{P})$. Fix an initial state $x_0 \in \mathcal{X}$ and choose a nonnegative subsolution $w^{(0)}$ as above. The *general scheme* with M trajectories at each iteration then produces $w_\omega^{(n+1)}$ from $w_\omega^{(n)}$, $n \in \mathbb{N}_0$, as follows:

a) Generate M trajectories of the state sequence starting in x_0 at time zero using a control policy induced by $w_\omega^{(n)}$: Set $X_0^{i,n}(\omega) \doteq x_0$ and for $j \in \{0, \ldots, N-1\}$,

$$X_{j+1}^{i,n}(\omega) \doteq \Psi\left(j, X_j^{i,n}(\omega), \gamma_j^{i,n}(\omega), \xi_{j+1}^{(i,n+1)}(\omega)\right),$$

$$\gamma_j^{i,n}(\omega) \in \mathrm{argmin}_{\gamma\in\Gamma}\Big\{ f(j, X_j^{i,n}(\omega), \gamma) + \int_{\mathcal{Y}} w_\omega^{(n)}\big(j+1, \Psi(j, X_j^{i,n}(\omega), \gamma, y)\big)\mu(dy)\Big\},$$

where $i \in \{1, \ldots, M\}$.

b) Construct a function $w_\omega^{(n+1)} : \{0, \ldots, N\} \times \mathcal{X} \to \mathbb{R}$ by backward recursion in the following way:

– Choose $w_\omega^{(n+1)}(N, .)$ Lipschitz continuous with constant $L_N^{(n+1)}$ such that $w_\omega^{(n)}(N, x) \le w_\omega^{(n+1)}(N, x) \le F(x)$ for all $x \in \mathcal{X}$.

– For j running from $N-1$ down to 0 do:
  - compute $v_j^{i,n}(\omega) \doteq \mathcal{L}_j(w_\omega^{(n+1)}(j+1,.))(X_j^{i,n}(\omega))$, $i \in \{1,\ldots,M\}$,
  - let $L \doteq c_0 + L_{j+1}^{(n+1)}(1+c_1)$ and choose $w_\omega^{(n+1)}(j,.)$ Lipschitz continuous with Lipschitz constant $L_j^{(n+1)}$ such that, for all $x \in \mathcal{X}$, $w_\omega^{(n+1)}(j,x) \geq w_\omega^{(n)}(j,x)$ and

$$w_\omega^{(n+1)}(j,x) \leq \max\left\{w_\omega^{(n)}(j,x), \max_{i\in\{1,\ldots,M\}} \mathrm{Cone}(X_j^{i,n}(\omega), v_j^{i,n}(\omega), L)(x)\right\}.$$

The functions generated according to the general scheme, which comprises the full cone construction as a special case, are subsolutions.

Proposition 1. *Grant Assumptions (A1) – (A5). Let $w^{(0)}$ be a nonnegative Lipschitz continuous subsolution, let $x_0 \in \mathcal{X}$, $M \in \mathbb{N}$, $\omega \in \Omega$, and let $(w_\omega^{(n)})_{n\in\mathbb{N}}$ be any sequence of functions constructed according to the general scheme starting from $w^{(0)}$ with parameters x_0, M, ω. Then, for all $n \in \mathbb{N}_0$, $0 \leq w_\omega^{(n)} \leq w_\omega^{(n+1)}$, and $w_\omega^{(n)}$ is a subsolution.*

Proof. The inequalities $0 \leq w_\omega^{(n)} \leq w_\omega^{(n+1)}$ hold by construction. Since $w^{(0)}$ is a (nonnegative) subsolution by hypothesis, it is enough to show that if $w_\omega^{(n)}$ is a subsolution for some $n \in \mathbb{N}_0$, then $w_\omega^{(n+1)}$ is a subsolution. By construction, $w_\omega^{(n)}(N,.) \leq w_\omega^{(n+1)}(N,.) \leq F(.)$. Let $j \in \{0,\ldots,N-1\}$. We have to show that for all $x \in \mathcal{X}$, $\mathcal{L}_j(w_\omega^{(n+1)}(j+1,.))(x) - w_\omega^{(n+1)}(j,x) \geq 0$. Set

$$v(x) \doteq \mathcal{L}_j(w_\omega^{(n+1)}(j+1,.))(x), \quad x \in \mathcal{X}.$$

Since $w_\omega^{(n+1)}(j+1,.) \geq w_\omega^{(n)}(j+1,.)$, monotonicity of the one-step Bellman operator implies that $v(x) \geq \mathcal{L}_j(w_\omega^{(n)}(j+1,.))(x)$ for all $x \in \mathcal{X}$. From the subsolution property of $w_\omega^{(n)}$ it follows that $v(.) \geq w_\omega^{(n)}(j,.)$. By construction, $w_\omega^{(n+1)}(j+1,.)$ is Lipschitz continuous with Lipschitz constant $L_{j+1}^{(n+1)}$. This together with Assumption (A5) implies that $v(.)$ is Lipschitz with Lipschitz constant not greater than $L \doteq c_0 + L_{j+1}^{(n+1)}(1+c_1)$. Consequently, given any $\tilde{x} \in \mathcal{X}$, we have $v(.) \geq \mathrm{Cone}(\tilde{x}, v(\tilde{x}), L)(.)$. Since $v_j^{i,n}(\omega) = v(X_j^{i,n}(\omega))$, $i \in \{1,\ldots,M\}$, by construction and $v(.) \geq w_\omega^{(n)}(j,.)$, it follows that for all $x \in \mathcal{X}$,

$$v(x) \geq \max\left\{w_\omega^{(n)}(j,x), \max_{i\in\{1,\ldots,M\}} \mathrm{Cone}(X_j^{i,n},(\omega), v_j^{i,n}(\omega), L)(x)\right\} \geq w_\omega^{(n+1)}(j,x),$$

and hence

$$\mathcal{L}_j(w_\omega^{(n+1)}(j+1,.))(x) - w_\omega^{(n+1)}(j,x) \geq \mathcal{L}_j(w_\omega^{(n+1)}(j+1,.))(x) - v(x) = 0.$$

□

4 Convergence for deterministic systems

In this section, we consider the case where the controlled system is deterministic. The system function Ψ thus does not depend on the noise variables. The class of control policies, in this case, may be restricted to the set of deterministic open-loop controls, that is, to the set of Γ-valued sequences. With slight abuse of notation, the recursion in (1) for the state dynamics starting at time zero can be rewritten as

$$x_{j+1} \doteq \Psi(j, x_j, \gamma_j), \quad j \in \mathbb{N}_0, \tag{13}$$

where $x_0 \in \mathcal{X}$ and $(\gamma_j)_{j\in\mathbb{N}_0} \subset \Gamma$. The mathematical expectations in the definition of the cost functional and the one-step Bellman operator become redundant. The costs for applying control policy $(\gamma_j) \subset \Gamma$ to the system starting at time zero in initial state x_0 are simply

$$J\big(0, x_0, (\gamma_j)\big) = \sum_{j=0}^{N-1} f(j, x_j, \gamma_j) + F(x_N),$$

where (x_j) is computed according to recursion (13). Let V denote the corresponding value function.

Let $(w^{(n)})_{n\in\mathbb{N}}$ be the sequence of subsolutions generated according to the full cone construction with one trajectory at each iteration and initial state x_0, starting from a nonnegative Lipschitz continuous subsolution $w^{(0)}$.

Denote by $\mathcal{X}_N(x_0)$ the set of all points in $\mathcal{X}$ which a state sequence starting in x_0 at time zero reaches in at most N steps under some control policy $(\gamma_j) \subset \Gamma$. If $\mathcal{X}_N(x_0)$ is contained in a compact subset of $\mathcal{X}$ and if the terminal costs F are Lipschitz continuous, then the subsolutions produced by the full cone construction converge to the value of the control problem at the initial state. The hypothesis that $\mathcal{X}_N(x_0)$ be (contained in) a compact set is automatically satisfied if the system function Ψ is continuous.

Theorem 1. *Grant Assumptions (A1) – (A5). Assume in addition that F is Lipschitz continuous and that the closure of $\mathcal{X}_N(x_0)$ in $\mathcal{X}$ is compact. Then $w^{(n)}(0, x_0)$ converges to $V(0, x_0)$ from below as n tends to infinity.*

Proof. Lemma 1, Assumption (A5), and the Lipschitz continuity of F imply that $V(j, .)$ is Lipschitz continuous for each $j \in \{0, \ldots, N\}$. By Proposition 1, $(w^{(n)})_{n\in\mathbb{N}_0}$ is indeed a sequence of subsolutions; in particular, $w^{(n)}(j, .) \leq V(j, .)$

for all $n \in \mathbb{N}_0$. By construction, $w^{(n)}(j,.) \leq w^{(n+1)}(j,.)$. As a consequence of the theorem of monotone convergence of sequences,

$$w(j,x) \doteq \lim_{n\to\infty} w^{(n)}(j,x), \quad j \in \{0,\dots,N\},\ x \in \mathcal{X},$$

defines a real-valued function by pointwise limits. By construction and (12), the functions $w^{(n)}(j,.)$ are Lipschitz continuous with Lipschitz constant uniformly bounded over $n \in \mathbb{N}_0$ and $j \in \{0,\dots,N\}$. The family $(w^{(n)})_{n\in\mathbb{N}}$ thus is equicontinuous, and it is uniformly bounded on compact sets (by zero from below, by V from above). By the Arzelà-Ascoli theorem, $w^{(n)}(j,.)$ converges to $w(j,.)$ uniformly on compact subsets of $\mathcal{X}$, and $w(j,.)$ is Lipschitz continuous. Since $\mathrm{cl}(\mathcal{X}_N(x_0))$ is compact by hypothesis, $w^{(n)}(j,.)$ converges to $w(j,.)$ uniformly on $\mathcal{X}_N(x_0)$.

Let $x^n_{j+1} \in \mathcal{X}_N(x_0)$, $\gamma^n_j \in \Gamma$, $j \in \{0,\dots,N-1\}$, be the states and control actions generated by the algorithm at iteration step n. Thus,

$$x^n_{j+1} = \Psi(j,x^n_j,\gamma^n_j), \quad \gamma^n_j \in \mathrm{argmin}_{\gamma\in\Gamma}\left\{f(j,x^n_j,\gamma) + w^{(n)}\big(j{+}1,\Psi(j,x^n_j,\gamma)\big)\right\}.$$

Let $\varepsilon > 0$. Choose $n \in \mathbb{N}$ such that $w(j,x) - w^{(n)}(j,x) \leq \varepsilon$ for all $j \in \{0,\dots,N\}$, $x \in \mathcal{X}_N(x_0)$. By construction,

$$w^{(n+1)}(j,x^n_j) \geq \mathrm{Cone}(x^n_j,v^n_j,L)(x^n_j) = v^n_j = \mathcal{L}_j\big(w^{(n+1)}(j{+}1,.)\big)(x^n_j).$$

Actually, $w^{(n+1)}(j,x^n_j) = \mathcal{L}_j\big(w^{(n+1)}(j{+}1,.)\big)(x^n_j)$ since $\mathcal{L}_j\big(w^{(n+1)}(j{+}1,.)\big)(x^n_j) \geq w^{(n+1)}(j,x^n_j)$ by the subsolution property of $w^{(n+1)}$. Using again monotonicity, we find that for $j \in \{0,\dots,N-1\}$,

$$\begin{aligned}
w(j,x^n_j) &\geq w^{(n+1)}(j,x^n_j)\\
&= \mathcal{L}_j\big(w^{(n+1)}(j{+}1,.)\big)(x^n_j)\\
&\geq \mathcal{L}_j\big(w^{(n)}(j{+}1,.)\big)(x^n_j)\\
&= \min_{\gamma\in\Gamma}\left\{f(j,x^n_j,\gamma) + w^{(n)}\big(j{+}1,\Psi(j,x^n_j,\gamma)\big)\right\}\\
&= f(j,x^n_j,\gamma^n_j) + w^{(n)}(j{+}1,x^n_{j+1})\\
&\geq f(j,x^n_j,\gamma^n_j) + w(j{+}1,x^n_{j+1}) - \varepsilon.
\end{aligned}$$

Consequently, observing that $w(N,x_N) \geq w^{(n+1)}(N,x_N) = F(x_N)$, we have

$$\begin{aligned}
w(0,x_0) &\geq \sum_{j=0}^{N-1} f(j,x^n_j,\gamma^n_j) + F(x_N) - N\cdot\varepsilon\\
&= J\big(0,x_0,(\gamma^n_j)\big) - N\cdot\varepsilon\\
&\geq V(0,x_0) - N\cdot\varepsilon.
\end{aligned}$$

Since $\varepsilon > 0$ was arbitrary, it follows that $w(0, x_0) \geq V(0, x_0)$. On the other hand, $w(0, x) \leq V(0, x)$ for all $x \in \mathcal{X}$. Therefore $w(0, x_0) = V(0, x_0)$. $\square$

The proof of Theorem 1 also shows that $(w^{(n)})$ converges to the value function V along the optimal trajectories starting in x_0. Moreover, the costs associated with the control policy (γ_j^n), the policy induced by $w^{(n)}$ according to the fixed selection mechanism, converge to the minimal costs, that is,

$$J\big(0, x_0, (\gamma_j^n)\big) \overset{n\to\infty}{\longrightarrow} V(0, x_0).$$

The convergence of the costs is in general not monotonic.

5 Convergence for stochastic systems

Here we return to the general setup as introduced in Section 2. We will prove convergence of the full cone construction with one state trajectory at each step of the iterative procedure (parameter $M = 1$). The proof in the case that several state trajectories are used to update the current subsolution (parameter $M > 1$) is only notationally different.

Convergence of the sequence of subsolutions to the value of the control problem at the initial state will be proved under two additional assumptions.

(A6) Given $x_0 \in \mathcal{X}$ and any sequence $(\gamma_j^n)_{n\in\mathbb{N}} \subset \hat{\mathcal{U}}$ of random control policies, the family $(X_j^n)_{n\in\mathbb{N}, j\in\{0,\dots,N\}}$ of $\mathcal{X}$-valued random variables is tight, where X^n is the solution to Eq. (5) under (γ_j^n) with initial state x_0 at time zero.

(A7) The cost coefficients f, F are bounded from above by some constant $K > 0$.

Assumption (A6) about the tightness of the state sequences is to be understood in the usual sense that the family of laws of the $\mathcal{X}$-valued random variables X_j^N, $n \in \mathbb{N}$, $j \in \{0, \dots, N\}$, is tight (e.g. Billingsley, 1999, p.59); here $\mathcal{X}$ is endowed with its Borel σ-algebra. The assumption is clearly satisfied if the state space $\mathcal{X}$ itself is compact. The assumption is needed in situations where the system function is not continuous in the state variable or the noise distribution μ has non-compact support. Assumption (A7) could be replaced by boundedness on compact sets and an additional growth condition on the controlled state sequences.

Let $x_0 \in \mathcal{X}$. Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a complete probability space carrying an independent collection $(\xi_j^{(n)})_{j,n\in\mathbb{N}}$ of noise variables. Let $w^{(0)}$ be a nonnegative

Lipschitz continuous subsolution (for instance, $w^{(0)} \equiv 0$). For each $\omega \in \Omega$, let $(w_\omega^{(n)})_{n\in\mathbb{N}}$ be the sequence of subsolutions generated according to the full cone construction with initial state x_0, parameter $M = 1$, and noise samples $\xi_j^{(n)}(\omega)$, $j \in \{1, \dots, N\}$, $n \in \mathbb{N}$, starting from $w^{(0)}$.

Theorem 2. *Grant Assumptions (A1) – (A7) and assume in addition that F is Lipschitz continuous. Then, for $\mathbf{P}$-almost every $\omega \in \Omega$, $w_\omega^{(n)}(0, x_0)$ converges to $V(0, x_0)$ from below as n tends to infinity.*

Proof. Lemma 1, Assumption (A5), and the Lipschitz continuity of F imply that $V(j, .)$ is Lipschitz continuous for each $j \in \{0, \dots, N\}$.

Let $\omega \in \Omega$. By Proposition 1, $(w_\omega^{(n)})_{n\in\mathbb{N}_0}$ is a sequence of subsolutions. In particular, $w_\omega^{(n)}(j, .) \le V(j, .)$ and, by construction, $w_\omega^{(n)}(j, .) \le w_\omega^{(n+1)}(j, .)$ for all $n \in \mathbb{N}_0$. As a consequence of the theorem of monotone convergence of sequences,

$$w_\omega(j, x) \doteq \lim_{n\to\infty} w_\omega^{(n)}(j, x), \quad j \in \{0, \dots, N\},\ x \in \mathcal{X},$$

defines a real-valued function by pointwise limits, and $w_\omega \le V$. By construction and (12), the functions $w_\omega^{(n)}(j, .)$ are Lipschitz continuous with Lipschitz constant uniformly bounded over $n \in \mathbb{N}_0$, $j \in \{0, \dots, N\}$, not depending on ω. It follows that $w_\omega^{(n)}(j, .)$ converges to $w_\omega(j, .)$ as n tends to infinity uniformly on compact subsets of $\mathcal{X}$, and that $w_\omega(j, .)$ is Lipschitz continuous.

Let $X_{j+1}^n(\omega) \in \mathcal{X}$, $\gamma_j^n(\omega) \in \Gamma$, $j \in \{0, \dots, N-1\}$, denote the states and control actions computed at iteration $n \in \mathbb{N}$; thus, with $X_0^n(\omega) \doteq x_0$,

$$\begin{aligned} X_{j+1}^n(\omega) &= \Psi\big(j, X_j^n(\omega), \gamma_j^n(\omega), \xi_{j+1}^{(n+1)}(\omega)\big), \\ \gamma_j^n(\omega) \in \operatorname{argmin}_{\gamma\in\Gamma}\Big\{ & f(j, X_j^n(\omega), \gamma) \\ & + \int_{\mathcal{Y}} w_\omega^{(n)}\big(j+1, \Psi(j, X_j^n(\omega), \gamma, y)\big)\mu(dy)\Big\}. \end{aligned}$$

Let $\varepsilon > 0$. We show that there is $n(\varepsilon) \in \mathbb{N}$ such that the event

$$A_{n,\varepsilon} \doteq \Big\{\omega \in \Omega : w_\omega\big(j, X_j^n\big)(\omega) - w_\omega^{(n)}\big(j, X_j^n\big)(\omega) \le \varepsilon \text{ for all } j \in \{0, \dots, N\}\Big\}$$

has probability $\mathbf{P}(A_{n(\varepsilon),\varepsilon}) \ge 1 - \varepsilon$. Thanks to Assumption (6) we can find a compact set $G_\varepsilon \subset \mathcal{X}$ such that

$$\inf_{n\in\mathbb{N}} \mathbf{P}\left\{\omega \in \Omega : X_j^n(\omega) \in G_\varepsilon \text{ for all } j \in \{0, \dots, N\}\right\} \ge 1 - \frac{\varepsilon}{2}.$$

Since $w_\omega^{(n)}(j,.)$ converges to $w_\omega(j,.)$ uniformly on compact subsets of $\mathcal{X}$, for each $\omega \in \Omega$ there is $n_0(\varepsilon,\omega) \in \mathbb{N}$ such that $w_\omega(j,x) - w_\omega^{(n)}(j,x) \leq \varepsilon$ for all $n \geq n_0(\varepsilon,\omega)$, $j \in \{0,\dots,N\}$, $x \in G_\varepsilon$. Now we can choose $n(\varepsilon) \in \mathbb{N}$ such that

$$\mathbf{P}\left\{\omega \in \Omega : n_0(\varepsilon,\omega) \leq n(\varepsilon)\right\} \geq 1 - \frac{\varepsilon}{2}.$$

Then $\mathbf{P}(A_{n(\varepsilon),\varepsilon}) \geq 1-\varepsilon$, because

$$A_{n(\varepsilon),\varepsilon} \supseteq \left\{\omega \in \Omega : n_0(\varepsilon,\omega) \leq n(\varepsilon),\ X_j^{n(\varepsilon)}(\omega) \in G_\varepsilon \text{ for all } j \in \{0,\dots,N\}\right\}.$$

Set $n \doteq n(\varepsilon)$, and let $\omega \in A_{n,\varepsilon}$. By construction, monotonicity of the Bellman operators, and the subsolution property of $w_\omega^{(n)}$,

$$w_\omega^{(n+1)}(j, X_j^n(\omega)) = \mathrm{Cone}(X_j^n(\omega), v_j^n(\omega), L)(x_j^n) = v_j^n(\omega),$$

where $v_j^n(\omega) = \mathcal{L}_j\big(w_\omega^{(n+1)}(j+1,.)\big)(X_j^n(\omega))$. Using again monotonicity of the Bellman operators we find that for $j \in \{0,\dots,N-1\}$,

$$\begin{aligned}
w_\omega(j, X_j^n(\omega)) &\geq w_\omega^{(n+1)}(j, X_j^n(\omega)) \\
&= \mathcal{L}_j\big(w_\omega^{(n+1)}(j+1,.)\big)(X_j^n(\omega)) \\
&\geq \mathcal{L}_j\big(w_\omega^{(n)}(j+1,.)\big)(X_j^n(\omega)) \\
&= \min_{\gamma\in\Gamma}\left\{f(j, X_j^n(\omega), \gamma) + \int_{\mathcal{Y}} w_\omega^{(n)}\big(j+1, \Psi(j, X_j^n(\omega), \gamma, y)\big)\mu(dy)\right\} \\
&\geq f(j, X_j^n(\omega), \gamma_j^n(\omega)) + \int_{\mathcal{Y}} w_\omega^{(n)}\big(j+1, \Psi(j, X_j^n(\omega), \gamma_j^n(\omega), y)\big)\mu(dy) \\
&\geq f(j, X_j^n(\omega), \gamma_j^n(\omega)) + \int_{\mathcal{Y}} w_\omega^{(n)}\big(j+1, \Psi(j, X_j^n(\omega), \gamma_j^n(\omega), y)\big)\mu(dy) \\
&\quad - w_\omega^{(n)}\big(j+1, X_{j+1}^n(\omega)\big) + w_\omega(j+1, X_{j+1}^n(\omega)) - \varepsilon.
\end{aligned}$$

Since $w_\omega(N, X_N^n(\omega)) \geq w_\omega^{(n+1)}(N, X_N^n(\omega)) = F(X_N^n(\omega))$, it follows that

$$\begin{aligned}
&w_\omega(0, x_0) \\
&\geq \sum_{j=0}^{N-1} f(j, X_j^n(\omega), \gamma_j^n(\omega)) + F(X_N^n(\omega)) - N\cdot\varepsilon \\
&\quad + \sum_{j=0}^{N-1}\left(\int_{\mathcal{Y}} w_\omega^{(n)}\big(j+1, \Psi(j, X_j^n(\omega), \gamma_j^n(\omega), y)\big)\mu(dy) - w_\omega^{(n)}\big(j+1, X_{j+1}^n(\omega)\big)\right).
\end{aligned}$$

By Assumption (A7), f, F are bounded from above by K. Since $\mathbf{P}(A_{n,\varepsilon}) \geq 1-\varepsilon$, it follows that

$$\begin{aligned}
&\mathbf{E}\,[w(0,x_0)] \\
&\geq \mathbf{E}\left[\sum_{j=0}^{N-1} f(j,X_j^n,\gamma_j^n) + F(X_N^n)\right] - K(N+1)\cdot\varepsilon \\
&\quad + \sum_{j=0}^{N-1} \mathbf{E}\left[\int_{\mathcal{Y}} w^{(n)}\big(j+1,\Psi(j,X_j^n,\gamma_j^n,y)\big)\mu(dy) - w^{(n)}\big(j+1,X_{j+1}^n\big)\right].
\end{aligned}$$

By construction, $X_{j+1}^n(\omega) = \Psi(j,X_j^n(\omega),\gamma_j^n(\omega),\xi_{j+1}^{(n+1)}(\omega))$, $\omega\in\Omega$, where $\xi_{j+1}^{(n+1)}$ has distribution μ and is independent of X_j^n, γ_j^n, and $\omega\mapsto w_\omega^{(n)}(j,.)$. Conditioning and an application of Fubini's theorem show that for all $j\in\{0,\ldots,N-1\}$,

$$\mathbf{E}\left[\int_{\mathcal{Y}} w^{(n)}\big(j+1,\Psi(j,X_j^n,\gamma_j^n,y)\big)\mu(dy) - w^{(n)}\big(j+1,X_{j+1}^n\big)\right] = 0.$$

Therefore, taking into account Remark 3,

$$\begin{aligned}
\mathbf{E}\,[w(0,x_0)] &\geq \mathbf{E}\left[\sum_{j=0}^{N-1} f(j,X_j^n,\gamma_j^n) + F(X_N^n)\right] - K(N+1)\cdot\varepsilon \\
&= \hat{J}\big(0,x_0,(\gamma_j^n)\big) - K(N+1)\cdot\varepsilon \\
&\geq \hat{V}(0,x_0) - K(N+1)\cdot\varepsilon.
\end{aligned}$$

By Lemma 2, $\hat{V}(0,x_0) = V(0,x_0)$. Since $\varepsilon>0$ was arbitrary, we have $\mathbf{E}[w(0,x_0)]\geq V(0,x_0)$. On the other hand, $w_\omega(0,x)\leq V(0,x)$ for all $x\in\mathcal{X}$ and all $\omega\in\Omega$. Therefore, for $\mathbf{P}$-almost every $\omega\in\Omega$, $w_\omega(0,x_0)=V(0,x_0)$. □

In analogy with Section 4, the proof of Theorem 2 shows that the costs associated with the random control policy (γ_j^n), which is the policy induced by $w^{(n)}$, converge to the minimal costs:

$$\hat{J}\big(0,x_0,(\gamma_j^n)\big) \overset{n\to\infty}{\longrightarrow} \hat{V}(0,x_0) = V(0,x_0).$$

The convergence of the costs is in general not monotonic.

6 Remarks on implementation and extensions

The procedure for constructing subsolutions has been introduced in a general setting, which nonetheless can be extended further. In particular, state-dependent

constraints on the control actions can be included. The one-step Bellman operators given in (4) would have to be redefined accordingly, namely, as

$$\mathcal{L}_j(\varphi)(x) \doteq \inf_{\gamma\in\Gamma(j,x)} \left\{ f(j,x,\gamma) + \int_{\mathcal{Y}} \varphi\big(\Psi(j,x,\gamma,y)\big)\mu(dy) \right\}, \qquad x \in \mathcal{X},$$

where $\Gamma(j,x) \subset \Gamma$ is the set of admissible control actions at step j in state x. If the sets $\Gamma(j,x)$ depend on the state x in such a way that Assumption (A5) holds, then the same construction as before can be used to produce convergent sequences of subsolutions. It is also possible to deal with a non-compact control space Γ, provided that the PDP is still valid. In this case, nearly optimal control actions instead of minimizing control actions would have to be selected; the tolerance in non-optimality would have to tend to zero as the number of iterations goes to infinity in order to ensure convergence.

Convergence of the scheme has been shown in the sense of monotone convergence of the subsolutions to the value function at the fixed initial state. For deterministic systems we have, in addition, convergence to the value function along optimal trajectories. Since the subsolution property is a global property, the subsolutions produced by the scheme, though dependent on the initial state, are always global lower approximations to the value function. A subsolution $w^{(n)} = w^{(n,x_0)}$ produced after n iterations of the scheme with initial state x_0 can be used to compute state trajectories and costs starting from a different initial state $\tilde{x}_0$ (control synthesis). Moreover, $w^{(n,x_0)}$ may be taken as the initial subsolution for running the scheme with initial state $\tilde{x}_0$.

The full cone construction is not always implementable as it stands. Consider the case $\mathcal{X} = \mathbb{R}^d$, and let us assume that floating point numbers are acceptable substitutes for real numbers. The main reason why the procedure in this situation might not be directly implementable is related to the computation of $\mathcal{L}_j(\varphi)(x)$, where φ is a nonnegative Lipschitz function and $x \in \mathbb{R}^d$ is a given state. Indeed, the one-step Bellman operator involves a global minimization over the action space Γ and the evaluation of an integral over the noise space $\mathcal{Y}$. If Γ, $\mathcal{Y}$ are finite sets of moderate cardinality, then computing $\mathcal{L}_j(\varphi)(x)$ poses no difficulty and our procedure can be implemented as it stands. The same is true for more general noise spaces $\mathcal{Y}$ and noise distributions μ as long as integration w.r.t. μ can be performed efficiently. The subsolutions produced by the full cone construction are pointwise maxima of downward opening symmetric cones. Functions of this type can be represented efficiently. Indeed, it is enough to store, for each cone, its center, its height, and its slope, which amounts to

one element of $\mathcal{X}$ plus two real numbers. The maximum is computed only when evaluating the function at any given point.

Assumption (A5) about preservation of Lipschitz continuity by the one-step Bellman operators holds for a broad class of discrete-time optimal control problems, and it could be relaxed even further to preservation of just local Lipschitz continuity. The price to pay for this generality is a procedure which in general is too conservative. In the construction of the subsolutions no regularity of the value function other than Lipschitz continuity is exploited; in using downward opening cones of maximal slope one presumes worst-case regularity at every point.

Modifications in the update step of the scheme are possible and can lead to more efficient variants for more specific models. A special but important case are convex control problems, that is, dynamic minimization problems whose one-step Bellman operators preserve convexity; see, for instance, Hernández-Lerma et al. (1995), where approximations to the value function are constructed which are monotone from above. In the case of convex control problems, subsolutions can be represented as maxima of hyperplanes instead of conic surfaces. A detailed description of this variant of the scheme, including numerical experiments, is in preparation. Another particular class of problems to which the scheme can be adapted consists of discrete-time control problems derived from continuous-time non-degenerate stochastic control problems, which admit classical (i.e. $\mathbf{C}^{1,2}$ continuously differentiable) solutions.

The proofs of convergence given in Sections 4 and 5 use essentially three properties of the sequence of functions $(w^{(n)})$ generated by the full cone construction: the subsolution property, monotonicity (i.e., $w^{(n)} \leq w^{(n+1)}$ for all n) plus uniform continuity on compacts (here guaranteed by the Lipschitz property), and the equality $w^{(n+1)}(j, X_j^n) = \mathcal{L}_j(w^{(n+1)}(j{+}1, .))(X_j^n)$ along the state trajectories produced by the simulation part. Any modification to the update part of the scheme that preserves these three properties also preserves convergence of the scheme.

Appendix

A Relation to policy iteration

Policy iteration or approximation in policy space is ordinarily used to numerically solve stationary Markov control problems (cf. Puterman, 1994, Sect. 6.4).

In Fleming and Rishel (1975, pp. 168-169), a non-stationary version is employed to prove existence of classical solutions to the Hamilton-Jacobi-Bellman equation associated with a class of non-degenerate controlled Itô diffusions.

Applied in the present context, policy iteration would recursively compute strategies $u^{(m)} \in \mathcal{U}$ and functions $W^{(m+1)} : \{0, \dots, N\} \times \mathcal{X} \to \mathbb{R}$, $m \in \mathbb{N}_0$, as follows. To start, at step zero, choose any feedback strategy $u^0 \in \mathcal{U}$. At step $m \in \mathbb{N}$ do the following:

a) Given $u^{(m-1)} \in \mathcal{U}$, compute the function $W^{(m)}$ by backward recursion in the following way:

- Set $W^{(m)}(N, x) \doteq F(x)$, all $x \in \mathcal{X}$.
- For j running from $N-1$ down to 0, set for $x \in \mathcal{X}$

$$W^{(m)}(j,x) \doteq f\big(j, x, u^{(m-1)}(j,x)\big) + \int_{\mathcal{Y}} W^{(m)}\big(j{+}1, \Psi(j, x, u^{(m-1)}(j,x), y)\big)\mu(dy).$$

b) Given $W^{(m)}$, choose $u^{(m)} \in \mathcal{U}$ such that for all $j \in \{0, \dots, N\}$, $x \in \mathcal{X}$,

$$u^m(j,x) \in \operatorname{argmin}_{\gamma \in \Gamma} \left\{ f\big(j, x, \gamma\big) + \int_{\mathcal{Y}} W^{(m)}\big(j{+}1, \Psi(j, x, \gamma, y)\big)\mu(dy) \right\}.$$

The sequence $(W^{(m)})_{m \in \mathbb{N}}$ produced by policy iteration is non-increasing, that is, $W^{(m)} \geq W^{(m+1)}$, and it converges to the value function V from above, uniformly over the state space. In contrast, the sequence of subsolutions produced by our procedure is non-decreasing, and it converges to the value function from below at any fixed initial state.

B Lipschitz bound for diffusion models

Consider a continuous-time controlled diffusion as given at the end of Section 2 with dynamics according to Eq. (8) and cost functional (9). Let $h > 0$, and define a corresponding discrete-time problem with uniform time step h. The system function, in particular, is defined according to (10). Let the noise distribution μ be a product measure $\otimes^{d_1}\nu$ on $\mathcal{B}(\mathbb{R}^{d_1})$, where ν is a probability measure on $\mathcal{B}(\mathbb{R})$ with mean zero and variance one. Let ξ be an $\mathbb{R}^{d_1}$-valued random variable with distribution μ.

If the drift and diffusion coefficients b, σ and the running costs $\tilde{f}$ of the original problem are Lipschitz continuous in the state variable (uniformly in time and control) with Lipschitz constants L_b, L_σ, and $L_{\tilde{f}}$, respectively, then Assumption (A5) is satisfied for the one-step Bellman operators $\mathcal{L}_j$ of the discrete-time problem. More precisely, if φ is a Lipschitz continuous function $\mathbb{R}^d \to \mathbb{R}$ with Lipschitz constant L, then for all $x, \hat{x} \in \mathbb{R}^d$, $j \in \{1, \dots, N\}$,

$$
\begin{aligned}
&|\mathcal{L}_j(\varphi)(x) - \mathcal{L}_j(\varphi)(\hat{x})| \\
&\le \sup_{\gamma \in \Gamma} \Big\{ h |\tilde{f}(j, x, \gamma) - \tilde{f}(j, \hat{x}, \gamma)| \Big\} \\
&\quad + \sup_{\gamma \in \Gamma} \left\{ \int_{\mathcal{Y}} \big| \varphi\big(\Psi(j, x, \gamma, y)\big) - \varphi\big(\Psi(j, \hat{x}, \gamma, y)\big) \big| \, \mu(dy) \right\} \\
&\le L_{\tilde{f}} h |x - \hat{x}| + L \sup_{\gamma \in \Gamma} \left\{ \int_{\mathcal{Y}} |\Psi(j, x, \gamma, y) - \Psi(j, \hat{x}, \gamma, y)| \, \mu(dy) \right\} \\
&\le L_{\tilde{f}} h |x - \hat{x}| + L \sup_{\gamma \in \Gamma} \sqrt{\mathbf{E}\left[|\Psi(j, x, \gamma, \xi) - \Psi(j, \hat{x}, \gamma, \xi)|^2 \right]}.
\end{aligned}
$$

Now, for all $\gamma \in \Gamma$, recalling that ξ has law μ with mean zero and covariance matrix the identity,

$$
\begin{aligned}
&\mathbf{E}\left[|\Psi(j, x, \gamma, \xi) - \Psi(j, \hat{x}, \gamma, \xi)|^2 \right] \\
&= \mathbf{E}\left[\Big| x - \hat{x} + h\big(b(jh, x, \gamma) - b(jh, \hat{x}, \gamma)\big) + \sqrt{h}\big(\sigma(jh, x, \gamma) - \sigma(jh, \hat{x}, \gamma)\big)\xi \Big|^2 \right] \\
&= \big| x - \hat{x} + h\big(b(jh, x, \gamma) - b(jh, \hat{x}, \gamma)\big) \big|^2 + \mathbf{E}\left[\Big| \sqrt{h}\big(\sigma(jh, x, \gamma) - \sigma(jh, \hat{x}, \gamma)\big)\xi \Big|^2 \right] \\
&= |x - \hat{x}|^2 + 2h \left\langle x - \hat{x}, b(jh, x, \gamma) - b(jh, \hat{x}, \gamma) \right\rangle + h^2 \big| b(jh, x, \gamma) - b(jh, \hat{x}, \gamma) \big|^2 \\
&\quad + h \big| \sigma(jh, x, \gamma) - \sigma(jh, \hat{x}, \gamma) \big|^2 \\
&\le \big(1 + 2L_b h + L_b^2 h^2 + L_\sigma^2 h\big) |x - \hat{x}|^2.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
&|\mathcal{L}_j(\varphi)(x) - \mathcal{L}_j(\varphi)(\hat{x})| \\
&\le L_{\tilde{f}} h |x - \hat{x}| + L \sqrt{\big(1 + 2L_b h + L_b^2 h^2 + L_\sigma^2 h\big) |x - \hat{x}|^2} \\
&\le \Big(L_{\tilde{f}} h + L\big(1 + 2L_b h + L_b^2 h^2 + L_\sigma^2 h\big) \Big) |x - \hat{x}|,
\end{aligned} \tag{14}
$$

showing that Assumption (A5) is satisfied if c_0, c_1 are chosen according to (11).

Let T be the finite time horizon of the continuous-time problem. Suppose the terminal cost function $F = \tilde{F}$ is Lipschitz continuous with constant L_F. For $N \in \mathbb{N}$, let V_N denote the value function of the corresponding discrete-time problem with step size $h = T/N$. By Lemma 1, $V_N(j,.) = \mathcal{L}_j \circ \ldots \circ \mathcal{L}_{N-1}(F)$ for all $j \in \{0, \ldots, N-1\}$. Denote by $L_j^N \in (0, \infty]$ the Lipschitz constant of $V_N(j,.)$. By (14),

$$L_j^N \leq \hat{c}_0 \frac{T}{N} + L_{j+1}^N \left(1 + \hat{c}_1 \frac{T}{N}\right),$$

where $\hat{c}_0 \doteq L_{\tilde{f}}$, $\hat{c}_1 \doteq 2L_b + L_b^2 T + L_\sigma^2$ are independent of $N \in \mathbb{N}$. It follows that the Lipschitz constant of $V_N(j,.)$ is bounded from above by $L_F(1+\frac{\hat{c}_0}{\hat{c}_1})(1+\hat{c}_1\frac{T}{N})^{N-j}$. If N tends to infinity and $j \cdot T/N$ tends to $t \in [0,T]$, then the above bound on the Lipschitz constant (or on the norm of the weak sense first derivative) converges to $L_F(1+\frac{\hat{c}_0}{\hat{c}_1})e^{\hat{c}_1(T-t)}$.

References

D. Belomestny, A. Kolodko, and J. Schoenmakers. Regression methods for stochastic control problems and their convergence analysis. *SIAM J. Control Optim.*, 48(5):3562–3588, 2010.

D. P. Bertsekas and S. E. Shreve. *Stochastic Optimal Control: The Discrete-Time Case*. Athena Scientific, Belmont, Massachusetts, reprint of the 1978 edition, 1996.

P. Billingsley. *Convergence of Probability Measures*. Wiley series in Probability and Statistics. John Wiley & Sons, New York, 2nd edition, 1999.

C.-S. Chow and J. N. Tsitsiklis. The complexity of dynamic programming. *J. Complexity*, 5:466–488, 1989.

P. Dupuis and H. Wang. Subsolutions of an Isaacs equation and efficient schemes for importance sampling. *Math. Oper. Res.*, 32(3):723–757, 2007.

W. H. Fleming and R. W. Rishel. *Deterministic and Stochastic Optimal Control*, volume 1 of *Applications of Mathematics*. Springer, New York, 1975.

W. H. Fleming and D. Vermes. Convex duality approach to the optimal control of diffusions. *SIAM J. Control Optim.*, 27(5):1136–1155, 1989.

K. H. Helmes and R. H. Stockbridge. Determining the optimal control of singular stochastic processes using linear programming. In *Markov Processes and Related Topics: A Festschrift for Thomas G. Kurtz*, volume 4 of *IMS Collections*, pages 137–153. Institute of Mathematical Statistics, Beachwood, OH, 2008.

D. Hernández-Hernández, O. Hernández-Lerma, and M. Taksar. The linear programming approach to deterministic optimal control problems. *Appl. Math.*, 24(1):17–33, 1996.

O. Hernández-Lerma and J. B. Lasserre. *Discrete-Time Markov Control Processes: Basic Optimality Criteria*, volume 30 of *Applications of Mathematics*. Springer, New York, 1996.

O. Hernández-Lerma, C. Piovesan, and W. J. Runggaldier. Numerical aspects of monotone approximations in convex stochastic control problems. *Ann. Oper. Res.*, 56(1):135–156, 1995.

N. V. Krylov. *Controlled Diffusion Processes*, volume 14 of *Applications of Mathematics*. Springer, New York, 1980.

D. Kuhn. An information-based approximation scheme for stochastic optimization problems in continuous time. *Math. Oper. Res.*, 34(2):428–444, 2009.

J. B. Lasserre, D. Henrion, C. Prieur, and E. Trélat. Nonlinear optimal control via occupation measures and LMI-relaxations. *SIAM J. Control Optim.*, 47 (4):1643–1666, 2008.

W. H. McEneaney. Distributed Dynamic Programming for discrete-time stochastic control, and idempotent algorithms. *Automatica*, 47(3):443–451, 2011.

M. L. Puterman. *Markov Decision Processes*. Wiley series in Probability and Statistics. John Wiley & Sons, Hoboken, New Jersey, 1994.

L. C. G. Rogers. Pathwise stochastic optimal control. *SIAM J. Control Optim.*, 46(3):1116–1132, 2007.

J. Yong and X. Y. Zhou. *Stochastic Controls. Hamiltonian Systems and HJB Equations*, volume 43 of *Applications of Mathematics*. Springer, New York, 1999.