A Stochastic Approximation for Fully Nonlinear Free Boundary Parabolic Problems

Erhan Bayraktar,[1] **Arash Fahim**[2]
[1] *Department of Mathematics, University of Michigan, Ann Arbor, Michigan 48109*

[2] *Department of Mathematics, Florida State University, Tallahassee, Florida 32306*

Received 24 July 2012; accepted 6 November 2013
Published online 3 December 2013 in Wiley Online Library (wileyonlinelibrary.com).
DOI 10.1002/num.21841

We present a stochastic numerical method for solving fully nonlinear free boundary problems of parabolic type and provide a rate of convergence under reasonable conditions on the nonlinearity. Numer Methods Partial Differential Eq 30: 902–929, 2014

Keywords: free boundary problems; fully nonlinear Partial Differential Equations; Monte Carlo method; rate of convergence; viscosity solutions

I. INTRODUCTION

When the option pricing problem is of several dimensions, for example, basket options, deterministic methods such as finite difference are almost intractable; because the complexity increases exponentially with the dimension and one almost inevitably needs to use Monte Carlo simulations. Moreover, many problems in finance, for example, pricing in incomplete markets and portfolio optimization, lead to fully nonlinear PDEs. Only very recently there has been some significant development in numerically solving these nonlinear PDEs using Monte Carlo methods, see, for examplpe, [1–6]. When the control problem also contains a stopper, for example, in determining the super hedging price of an American option, see [7], or solving controller-and-stopper games, see [8], the nonlinear PDEs have free boundaries.

For solving linear PDEs with free boundaries, that is, in the problem of American options, Longstaff–Shwartz [9], introduced a stochastic method in which American options are approximated by Bermudan options and least squares approximation is used for doing the backward induction. The major feature in [9] is the tractability of the implementation for the scheme proposed in terms of the CPU time in high dimensional problems. The most important feature of this model that facilitates the speed is that the number of paths simulated is fixed. Simulating

Correspondence to: Arash Fahim, Department of Mathematics, Florida State University, Tallahassee, FL 32306 (e-mail: fahim@math.fsu.edu)
Contract grant sponsor: National Science Foundation; contract grant numbers: DMS-0955463, DMS-1118673 (to E.B.)
Contract grant sponsor: Susan M. Smith Professorship (to E.B.)
Contract grant sponsor: National Science Foundation; contract grant number: DMS-1209519 (to A.F.)

the paths correspond to introducing a stochastic mesh for the space dimension and the Bermudan approximation to American options correspond to time discretization. Stochastic mesh makes sure that more important points in the state space are used in the computation of the value function, an important feature which increases the speed of convergence. So essentially, this algorithm can be thought of as an explicit finite difference scheme with stochastic mesh. One can in fact prove the convergence rate of the entire "stochastic" explicit finite difference scheme, see [10] for a survey of these results and some improvements to the original methodology of Longstaff–Shwartz.

For semilinear free boundary problems, a similar stochastic scheme is given through reflected backward stochastic differential equations in [4] and rate of convergence is derived to be $h^{1/4}$ assuming uniform ellipticity for the problem where h is the mesh size of the time discretizaton. Here, the number of paths, N, that one needs to simulate, increases with decreasing h and needs to be chosen in a certain way, see, for example, (3.16). This is similar to what we have in classical explicit finite difference schemes. To achieve stability, when we decrease the mesh size for time, we need to decrease the mesh size for the space variable to keep the ratio of time step over space step squared in a certain range. As we discussed above, the Monte Carlo simulation creates a stochastic mesh. The first result in this direction is due to [4]. Later [3] improved the result of [4] by removing the uniform ellipticity condition. Moreover, they improve the rate of convergence to $h^{1/2}$ by assuming more regularity on the obstacle function.

In this article, we generalize the Longstaff–Schwartz methodology for numerically solving a large class of fully nonlinear free boundary problems. We extend the idea in [5] to present a stochastic scheme for fully nonlinear Cauchy problems with obstacle. As described in Remark 3.11, our scheme is stochastic, that is, the outcome is a random variable which converges to the true solution pathwise. The convergence of our scheme follows from the methodology of [11], and the results of [5]. For the convenience of the reader, we sketch the convergence argument in Section III B. Under a concavity assumption on the nonlinearity and regularity of the coefficients, we obtain a rate of convergence using Krylov's method of shaking coefficients together with the switching system approximation as in [12], where a rate of approximation is obtained for classical finite difference schemes for elliptic problems with free boundaries. In [13], Cafarelli and Souganidis provide a rate of convergence without a concavity assumption on the nonlinearity but they consider elliptic problems with nonlinearity that depends only on the Hessian.

Appendix A is provided to establish the comparison, existence, and regularity results for a parabolic switching system with free boundary which is needed to provide the estimations in the rate of convergence proof. This appendix generalizes the result of [14] for parabolic obstacle problems to parabolic switching systems with obstacle. Also, it can be considered as the parabolic version of [12] where they study elliptic switching systems with obstacle. Appendix B contains a proof of the technical Lemma 3.9.

The rest of the article is organized as follows: In Section II, we present the stochastic numerical scheme. In Section III, we present the main results, the convergence rate, and its proof. Section IV is devoted to some numerical results illustrating our theoretical findings. The appendix is devoted to the analysis of nonlinear switching systems with obstacles, which is an essential ingredient in the proof of our main result and some technical proofs.

Notation. For scalars $a, b \in \mathbb{R}$, we write $a \wedge b := \min\{a, b\}$, and $a \vee b := \max\{a, b\}$. By $\mathbb{M}(n, d)$, we denote the collection of all $n \times d$ matrices with real entries. The collection of all symmetric matrices of size d is denoted $\mathbb{S}_d$, and its subset of nonnegative symmetric matrices is denoted by $\mathbb{S}_d^+$. For a matrix $A \in \mathbb{M}(n, d)$, we denote by A^{T} its transpose. For $A, B \in \mathbb{M}(n, d)$, we denote $A \cdot B := \mathrm{Tr}[A^{\mathrm{T}} B]$. In particular, for $d = 1$, A and B are vectors of $\mathbb{R}^n$ and $A \cdot B$

reduces to the Euclidean scalar product. For a suitably smooth function φ on $Q_T := (0,T] \times \mathbb{R}^d$, we define

$$|\varphi|_\infty := \sup_{(t,x)\in Q_T} |\varphi(t,x)| \text{and } |\varphi|_1 := |\varphi|_\infty + \sup_{\substack{Q_T\times Q_T\\ x\neq x', t\neq t'}} \frac{|\varphi(t,x)-\varphi(t',x')|}{|x-x'|+|t-t'|^{\frac{1}{2}}}.$$

Finally, by $\mathbb{E}_{t,x}$ we mean the conditional expectation given $X_t = x$ for a prespecified diffusion process X.

II. DISCRETIZATION

We consider the obstacle problem

$$\min\left\{-\mathcal{L}^X v - F\left(\cdot, v, Dv, D^2 v\right), v - g\right\} = 0, \text{on } [0,T) \times \mathbb{R}^d, \tag{2.1}$$

$$v = g, \text{on } \{T\} \times \mathbb{R}^d, \tag{2.2}$$

where

$$\mathcal{L}^X \varphi := \frac{\partial \varphi}{\partial t} + \mu \cdot D\varphi + \frac{1}{2} a \cdot D^2 \varphi,$$

and

$$F : (t,x,r,p,\gamma) \in \mathbb{R}_+ \times \mathcal{O} \times \mathbb{R} \times \mathbb{R}^d \times \mathbb{S}_d \mapsto F(x,r,p,\gamma) \in \mathbb{R},$$

is a nonlinear map, μ and σ are maps from $\mathbb{R}_+ \times \mathcal{O}$ to $\mathbb{R}^d$ and $\mathbb{M}(d,d)$, respectively, $a := \sigma\sigma^{\mathrm{T}}$, $g : [0,T) \times \mathbb{R}^d \to \mathbb{R}$. We consider an $\mathbb{R}^d$-valued Brownian motion W on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$, where the filtration $\mathbb{F} = \{\mathcal{F}_t\}_{t\in[0,T]}$ satisfies the usual conditions, and $\mathcal{F}_0$ is trivial. For a positive integer n, let $h := T/n, t_i = ih, i = 0, \ldots, n$, and consider the one step Euler discretization

$$\hat{X}_h^{t,x} := x + \mu(t,x)h + \sigma(t,x)(W_{t+h} - W_t), \tag{2.3}$$

of the diffusion X corresponding to the linear operator $\mathcal{L}^X$. Then, the Euler discretization of the process X is defined by:

$$\hat{X}_{t_{i+1}} := \hat{X}_h^{t_i, \hat{X}_{t_i}}.$$

We suggest the following approximation of the value function v

$$v^h(T,x) := g(T,x) \text{and } v^h(t_i,x) := \max\{\mathbf{T}_h[v^h](t_i,x), g(t_i,x)\} \text{for any } x \in \mathbb{R}^d, \tag{2.4}$$

where for a given test function $\psi : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}$, we denote

$$\mathbf{T}_h[\psi](t,x) := \mathbb{E}_{t,x}\left[\psi(t+h, \hat{X}_{t+h})\right] + hF\left(\cdot, \mathcal{D}_h\psi\right)(t,x), \tag{2.5}$$

$$\mathcal{D}_h\psi(t_i,x) = \mathbb{E}_{t,x}\left[\psi(t+h, \hat{X}_{t+h})H_h\right], \tag{2.6}$$

where $H_h = (H_0^h, H_1^h, H_2^h)^{\mathrm{T}}$ and

$$H_0^h = 1, H_1^h = \left(\sigma^{\mathrm{T}}\right)^{-1} \frac{W_h}{h}, \quad H_2^h = \left(\sigma^{\mathrm{T}}\right)^{-1} \frac{W_h W_h^{\mathrm{T}} - h\mathbf{I}_d}{h^2} \sigma^{-1},$$

provided σ is invertible. Notice that (2.6) comes from

$$\mathbb{E}_{t,x}[\psi(N)N] = \mathbb{E}_{t,x}[D\psi(N)] \text{and} \, \mathbb{E}_{t,x}[\psi(N)(N^2-1)] = \mathbb{E}_{t,x}[D^2\psi(N)], \tag{2.7}$$

where N is a standard Gaussian random variable. The details can be found in Lemma 2.1 of [5].

III. ASYMPTOTICS OF THE DISCRETE-TIME APPROXIMATION

In this section, we present the convergence and the rate of convergence result for the scheme introduced in (2.4), and the assumptions needed for these results.

A. The Main Results

The proof of the convergence follows the general methodology of Barles and Souganidis [11], and requires that the nonlinear PDE (2.1) satisfies the comparison principle in viscosity sense.

We recall that an upper-semicontinuous (resp. lower-semicontinuous) function $\underline{v}$ (resp. $\bar{v}$) on $[0, T] \times \mathbb{R}^d$, is called a viscosity subsolution (resp. supersolution) of (2.1) if for any $(t, x) \in [0, T] \times \mathbb{R}^d$ and any smooth function φ satisfying

$$0 = (\underline{v} - \varphi)(t,x) = \max_{[0,T]\times\mathbb{R}^d} (\underline{v} - \varphi) \left(\text{resp}.0 = (\bar{v} - \varphi)(t,x) = \min_{[0,T]\times\mathbb{R}^d} (\bar{v} - \psi) \right),$$

we have:

- if $t < T$

$$\min\left\{-\mathcal{L}^X\varphi - F(\cdot, \mathcal{D}\varphi), \varphi - g\right\}(t,x) \le (\text{resp}. \ge) \, 0,$$

- if $t = T, (\underline{v} - g)(T, x) \le 0$ (resp. $(\bar{v} - g)(T, x) \ge 0$).

Remark 3.1. Note that the above definition is not symmetric for subsolution and supersolutions. More precisely, for a subsolution, we need to have either

$$-\mathcal{L}^X\varphi - F(\cdot, \mathcal{D}\varphi) \le 0 \quad \text{or} \quad \varphi - g \le 0.$$

However, for a supersolutions, we need to have both

$$-\mathcal{L}^X\varphi - F(\cdot, \mathcal{D}\varphi) \ge 0 \quad \text{and} \quad \varphi - g \ge 0.$$

Definition 3.2. *We say that (2.1) has comparison for bounded functions if for any bounded upper semicontinuous subsolution $\underline{v}$ and any bounded lower semicontinuous supersolution $\bar{v}$ on $[0, T) \times \mathbb{R}^d$, satisfying $\underline{v}(T, \cdot) \le \bar{v}(T, \cdot)$, we have $\underline{v} \le \bar{v}$.*

We denote by F_r, F_p, and F_γ, the partial gradients of F with respect to r, p, and γ, respectively. We also denote by F_γ^- the pseudo-inverse of the nonnegative symmetric matrix F_γ.

Assumption *F.*

i. The nonlinearity F is Lipschitz-continuous with respect to (x, r, p, γ) *uniformly in t, and* $|F(\cdot, \cdot, 0, 0, 0)|_\infty < K$ *for some positive constant K;*
ii. is invertible and $|\mu|_1 + |\sigma|_1 < \infty$
iii. F is elliptic and dominated by the diffusion of the linear operator $\mathcal{L}^X$*, that is,*

$$a^{-1} \cdot \nabla_\gamma F \leq 1 \quad on\, \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d \times \mathbb{S}_d; \tag{3.1}$$

iv. $F_p \in$ *Image* (F_γ) *and* $|F_p^{\mathrm{T}} F_\gamma^- F_p|_\infty < K$*;*
V. $F_r - \frac{1}{4} F_p^T F_\gamma^- F_p \geq 0$.

Remark 3.3. Assumption **F**(v) is made for the sake of simplicity of the presentation. It implies the monotonicity of the above scheme. If this assumption is not made, one can carry out the analysis in [5, Remark 3.13, Theorem 3.12, and Lemma 3.19] and approximate the solution of the nonmonotone scheme with the solution of an appropriate monotone scheme.

Theorem 3.4 (Convergence)**.** *Suppose that Assumption* **F** *holds. Also, assume that the fully nonlinear PDE (2.1) has comparison for bounded functions. Then, for every bounded function g Lipschitz on x and* $\frac{1}{2}$*-Hölder on t, there exists a bounded function v such that* $v^h \to v$ *locally uniformly. Moreover, v is the unique bounded viscosity solution of problem (2.1)–(2.2).*

By imposing the following stronger assumption, we are able to derive a rate of convergence for the fully nonlinear PDE.

Assumption HJB.
The nonlinearity F satisfies Assumption ***F*** *(ii)–(v), and is of the Hamilton–Jacobi–Bellman type:*

$$\frac{1}{2} a \cdot \gamma + b \cdot p + F(t, x, r, p, \gamma) = \inf_{\alpha \in \mathcal{A}} \{\mathcal{L}^\alpha(t, x, r, p, \gamma)\},$$

$$\mathcal{L}^\alpha(t, x, r, p, \gamma) := \frac{1}{2} Tr[\sigma^\alpha \sigma^{\alpha \mathrm{T}}(t, x)\gamma] + b^\alpha(t, x)p + c^\alpha(t, x)r + f^\alpha(t, x),$$

where functions $\sigma^\alpha, b^\alpha, c^\alpha$ *and* f^α *satisfy:*

$$\sup_{\alpha \in \mathcal{A}} (|\sigma^\alpha|_1 + |b^\alpha|_1 + |c^\alpha|_1 + |f^\alpha|_1) < \infty.$$

Assumption HJB+.
The nonlinearity F satisfies HJB, and for any $\delta > 0$*, there exists a finite set* $\{\alpha_i\}_{i=1}^{M_\delta}$ *such that for any* $\alpha \in \mathcal{A}$

$$\inf_{1 \leq i \leq M_\delta} |\sigma^\alpha - \sigma^{\alpha_i}|_\infty + |b^\alpha - b^{\alpha_i}|_\infty + |c^\alpha - c^{\alpha_i}|_\infty + |f^\alpha - f^{\alpha_i}|_\infty \leq \delta.$$

Remark 3.5. Assumption HJB+ is satisfied if $\mathcal{A}$ is a compact separable topological space and $\sigma^\alpha(\cdot), b^\alpha(\cdot), c^\alpha(\cdot)$, and $f^\alpha(\cdot)$ are continuous maps from $\mathcal{A}$ to $C_b^{\frac{1}{2},1}$, the space of bounded maps which are Lipschitz in x and $\frac{1}{2}$-Hölder in t.

Theorem 3.6 (Rate of Convergence). *Assume that the boundary condition g is bounded Lipschitz on x and $\frac{1}{2}$-Hölder on t. Then, there is a constant $C > 0$ such that:*

i. *under Assumption **HJB**, we have $v - v^h \leq Ch^{1/4}$,*
ii. *under the stronger condition **HJB+**, we also have $-Ch^{1/10} \leq v - v^h$.*

It is worth mentioning that in the finite difference literature, the rate of convergence is usually stated in terms of the discretization in the space variable, that is, $|\Delta x|$, and the time step, that is, $|\Delta t|$ equals $|\Delta x|^2$. In our context, the stochastic numerical scheme (2.4) is only discretized in time with time step h. Therefore, the rates of convergence in Theorem 3.6 corresponds to the rates $|\Delta x|^{1/2}$ and $|\Delta x|^{1/5}$, respectively.

B. Proof of the Convergence Result

The proof Theorem 3.4, similar to the proof of Theorem 3.6 (i) of [5], is based on the result of [11] which requires the scheme to be consistent, monotone, and stable. To be consistence with the notation in [11], we define

$$S_h(t,x,r,\phi) := \min\{h^{-1}(r - \mathbf{T}_h[\phi](t,x)), r - g(t,x)\},$$

and then write the scheme (2.4) as $S_v(t,x,v^h(t,x),v^h) = 0$. Notice that by the discussions in [15] and in Section III of [16], the consistency and monotonicity for the scheme (2.4) for obstacle problem follows from the consistency and monotonicity of the scheme without obstacle provided by Lemmas 3.11 and 3.12 of [5]. More precisely, we have

i. **Consistency:** Let φ be a smooth function with bounded derivatives. Then, for all $(t,x) \in [0,T] \times \mathbb{R}^d$:

$$\lim_{\substack{(t',x') \to (t,x) \\ (h,c) \to (0,0) \\ t'+h \leq T}} S_h(t',x',c+\phi(t',x'),c+\phi)$$
$$= \min\{-\left(\mathcal{L}^X\varphi + F(\cdot,\varphi,D\varphi,D^2\varphi)\right), \varphi - g\}(t,x). \tag{3.2}$$

ii. **Monotonicity:** Let $\varphi, \psi : [0,T] \times \mathbb{R}^d \to \mathbb{R}$ be two bounded functions. Then:

$$\varphi \leq \psi \Rightarrow S_h(t,x,r,\varphi) \geq S_h(t,x,r,\psi). \tag{3.3}$$

Conversely, one can show the stability of the scheme (2.4) in the following Lemma. Throughout this section, Assumption of the Theorem 3.4 are enforced.

Lemma 3.7. *The family $(v^h)_h$ defined by (2.4) is bounded, uniformly in h*

Proof. Let $C_i = |v^h(t_i,\cdot)|_\infty$. By the argument in the proof of Lemma 3.14 in [5], $|T_h[v^h](t_i,\cdot)|_\infty \leq C_{i+1}(1+Ch) + Ch$ where $C > 0$ depends only on constant K in assumption **F**. Therefore,

$$C_i \leq \max\{|g|_\infty, C_{i+1}(1+Ch) + Ch\} \leq \max\{C_{i+1}, |g|_\infty\}(1+Ch) + Ch.$$

Using a backward induction, one could obtain that $C_i \leq |g|_\infty e^{CT} + \frac{e^{CT}}{C}$ for some constant C independent of h. ∎

The monotonicity, consistency, and stability of the scheme result to the following Lemma.

Lemma 3.8. *Let us define*

$$v_*(t,x) := \lim_{(\delta,h)\to(0,0)} \inf\{v^h(t,y) : |x-y| + |s-t| \leq \delta, s \in \{0,h,\cdots\} \cap [0,T]\},$$

$$v^*(t,x) := \lim_{(\delta,h)\to(0,0)} \sup\{v^h(t,y) : |x-y| + |s-t| \leq \delta, s \in \{0,h,\cdots\} \cap [0,T]\}.$$

Then, v_* and v^* are, respectively, a viscosity supersolution and a viscosity subsolution of (2.1)–(2.2).

Observe that thanks to Lemma 3.7, v_* and v^* are well-defined and bounded functions and we readily have $v_* \leq v^*$. Moreover, functions v_* and v^* are, respectively, lower semicontinuous and upper semicontinuous. Therefore, by the comparison principle for (2.1)–(2.2) and Lemma 3.8, it follows that $v^* = v_*$ and the function $v := v^* = v_*$ is a viscosity solution of (2.1)–(2.2) which completes the proof of Theorem 3.4. In addition, uniqueness in the class of bounded functions is a consequence of comparison principle for the problem.

Proof of Lemma 3.8. We show that v^* and v_* are subsolution and supersolution at any arbitrary point $(t_0, x_0) \in [0,T] \times \mathbb{R}^d$. We split the proof into the following steps.

Step 1 ($\mathbf{t}_0 < \mathbf{T}$). In this case, we only establish the supersolution property of v_*. The subsolution property of v^* follows from the same line of arguments. Let ϕ be a smooth function such that

$$0 = \min_{[0,T]\times\mathbb{R}^d} (v_* - \phi) = (v_* - \phi)(t_0, x_0).$$

As function v_* is bounded, by modifying ϕ outside a neighborhood of (t_0, x_0), one can assume that the (t_0, x_0) is a global strict minimum point. Notice that only the local property of the function ϕ matters in the definition of viscosity solution. Therefore, there exists a sequence $\{(t_n, x_n)\}$, such that $(t_n, x_n) \to (t_0, x_0), v^{h_n}(t_n, x_n) \to v^*(t_0, x_0), \xi_n := \min(v^{h_n} - \phi) = (v^{h_n} - \phi)(t_n, x_n) \to 0$, and (t_n, x_n) is a global minimum of $v^{h_n} - \phi$. (Obtaining this sequence is a standard technique in viscosity solution literature. For more details see [11] and the references therein.)

Therefore, $v^{h_n} \geq \phi + \xi_n$. By the monotonicity of the scheme, (3.3), we have

$$S_h(t_n, x_n, v^{h_n}(t_n, x_n), v^{h_n}) \leq S_h(t_n, x_n, \phi(t_n, x_n) + \xi_n, \phi + \xi_n).$$

Therefore, by the definition of v^h in (2.4),

$$0 \leq S_h(t_n, x_n, \phi(t_n, x_n) + \xi_n, \phi + \xi_n).$$

We divide both sides by h_n . Letting $n \to \infty$ and using (3.2) we obtain:

$$0 \leq \min\{-\left(\mathcal{L}^X \varphi + F(\cdot, \varphi, D\varphi, D^2\varphi)\right), \varphi - g\}(t_0, x_0).$$

Step 2 (Supersolution property when $t_0 = T$). Observe that since $g \le v^h$, we have

$$g(T, x_0) \le \liminf_{(h,t',x') \to (0,T,x_0)} v^h(t', x') = v_*(T, x_0),$$

which completes the proof of the supersolution argument at terminal time.
Step 3 (Subsolution property when $\mathbf{t}_0 = \mathbf{T}$). Observe that by the definition of vh and v^*, we readily have $v^* \ge g$. To complete the subsolution argument, we have to show that $v^*(T, \cdot) = g(T, \cdot)$. It is sufficient to show that

■

Lemma 3.9. *For all $x \in \mathbb{R}^d$ and $i = 0, \cdots, n$, we have*

$$|v^h(t_i, x) - g(T, x)| \le C\sqrt{T - t_i}.$$

The proof of Lemma 3.9 is similar to Lemma 4.2 [17] or Lemma 3.17 in [5]. For the convenience of the reader, we adjust the proof for free boundary problems. However, because the proof is technical and not related to the main result of the article, we prefer to present it in Appendix B.

C. Derivation of the Rate of Convergence

The proof of Theorem 3.6 is based on Barles and Jakobsen [18], which uses switching systems approximation and the Krylov method of shaking coefficients [19–22]. This has been adapted to classical finite difference schemes for elliptic obstacle (free boundary) problems in [12]. To use the method, we need to introduce a comparison principle for the scheme which we will undertake next.

Proposition 3.10. *Let Assumption **F** hold and set $\beta := |F_r|_\infty$. Consider two arbitrary bounded functions φ and ψ satisfying:*

$$\min\left\{h^{-1}\left(\varphi - \mathbf{T}_h[\varphi]\right), \varphi - g\right\} \le g_1 \quad \textit{and} \quad \min\left\{h^{-1}\left(\psi - \mathbf{T}_h[\psi]\right), \psi - g\right\} \ge g_2 \qquad (3.4)$$

for some bounded functions g_1 and g_2. Then, for every $i = 0, \ldots, n$:

$$(\varphi - \psi)(t_i, x) \le e^{\beta(T - t_i)}\left(|(\varphi - \psi)^+(T, \cdot)|_\infty + (1 + T - t_i)|(g_1 - g_2)^+|_\infty\right). \qquad (3.5)$$

Proof. This follows from Lemma 2.4 of [14] where it is provided for general monotone schemes for obstacle problems. ■

Proof of Theorem 3.6 (i). Under Assumption HJB, we can build a bounded subsolution v^ε of the nonlinear PDE, by the method of shaking coefficients, see [18], [12], [22], and the references therein. More precisely, consider the following equations

$$\min\left\{-\mathcal{L}^X v - \inf_{0<s<\varepsilon^2, |y|<\varepsilon} F\left(t - s, x + y, v, Dv, D^2 v\right), v - g\right\} = 0, \quad \text{on } [0, T) \times \mathbb{R}^d, \quad (3.6)$$

$$v = g, \text{on } \{T\} \times \mathbb{R}^d. \qquad (3.7)$$

By Theorem A.6, there exists a unique bounded solution v^ε to (3.6)–(3.7). Because $\inf_{0<s<\varepsilon^2,|y|<\varepsilon} F \le F, v^\varepsilon$ is a subsolution to (2.1)–(2.2) at all $(t,x) \in [0,T] \times \mathbb{R}^d$ and by Theorem A.4 (continuous dependence for Hamilton-Jacobi-Bellman equations), v^ε approximates v uniformly, that is, there exists a positive constant C such that $v - C\varepsilon \le v^\varepsilon \le v$.

Let $\rho(t,x)$ be a C^∞ nonnegative function supported in $\{(t,x) : t \in [0,1], |x| \le 1\}$ with unit mass, and define

$$w^\varepsilon(t,x) := v^\varepsilon * \rho^\varepsilon \text{where } \rho^\varepsilon(t,x) := \frac{1}{\varepsilon^{d+2}} \rho\left(\frac{t}{\varepsilon^2}, \frac{x}{\varepsilon}\right). \tag{3.8}$$

It follows that $|w^\varepsilon - v| \le C\varepsilon$. From the concavity of the nonlinearity F, and Lemma A.3 in [23], $w^\varepsilon \in C^\infty, w^\varepsilon$ is a classical subsolution of (2.1) on $U := \{(t,x)|g(t-s,x+y) < v^\varepsilon(t-s,x+y); \text{for any} s \in [0,\varepsilon^2) \text{and} |y| < \varepsilon\}$.[1] Moreover, by Theorem A.7, v^ε is Lipschitz in x and 1/2-Hölder continuous in t. Thus, by Theorem 2.1 in [21],

$$|\partial_t^{\beta_0} D^\beta w^\varepsilon| \le C\varepsilon^{1-2\beta_0-|\beta|_1} \quad \text{for any } (\beta_0, \beta) \in \mathbb{N} \times \mathbb{N}^d \backslash \{0\}, \tag{3.9}$$

where $|\beta|_1 := \sum_{i=1}^d \beta_i$, and $C > 0$ is some constant. As a consequence of the consistency of $\mathbf{T}_h$, see Lemma 3.22 of [5], we know that

$$\mathcal{R}_h[w^\epsilon](t,x) := \frac{w^\varepsilon(t,x) - \mathbf{T}_h[w^\varepsilon](t,x)}{h} + \mathcal{L}^X w^\varepsilon(t,x) + F(\cdot, w^\varepsilon, Dw^\varepsilon, D^2 w^\varepsilon)(t,x) \le Ch\varepsilon^{-3}.$$

From this estimate together with the subsolution property of w^ε, we see that $w^\varepsilon \le \mathbf{T}_h[w^\varepsilon] + Ch^2\varepsilon^{-3}$ holds true on U. In addition, by the regularity properties of g, one can see that $w^\varepsilon \le g + C\varepsilon$ on $[0,T] \times \mathbb{R}^d \backslash U$. Therefore,

$$\min\left\{\frac{w^\varepsilon(t,x) - \mathbf{T}_h[w^\varepsilon](t,x)}{h}, w^\varepsilon - g\right\} \le C_1(\varepsilon + \varepsilon^{-3}h).$$

Then, it follows from Proposition 3.10 that

$$w^\varepsilon - v^h \le C|(w^\varepsilon - v^h)(T,\cdot)|_\infty + C_1(\varepsilon + h\varepsilon^{-3}) \le C(\varepsilon + h\varepsilon^{-3}). \tag{3.10}$$

Therefore, $v - v^h \le v - w^\varepsilon + w^\varepsilon - v^h \le C(\varepsilon + h\varepsilon^{-3})$. Minimizing the right-hand side estimate over $\varepsilon > 0$, we obtain $v - v^h \le Ch^{1/4}$. ■

Proof of Theorem 3.6 (ii). To prove the lower bound on the rate of convergence, we will use Assumption **HJB+** and build a switching system approximation to the solution of the nonlinear obstacle problem (2.1)–(2.2). This proof method has been used for Cauchy problems of [18] and [5]. For obstacle problems, this method is used in the elliptic case by [12] for the classical finite difference schemes. We apply this methodology for parabolic obstacle problems to prove the lower bound for the convergence rate of our stochastic finite difference scheme. We split the proof into the following steps:

1. Approximating the solution to (2.1)–(2.2) by a switching system, which relies on Theorem A.4, the continuous dependence result for switching systems with obstacle.

[1] This heuristically follows from $F\left(\cdot,\cdot,v^\varepsilon, Dv^\varepsilon, D^2v^\varepsilon\right) * \rho_\varepsilon(t,x) \le F\left(t,x,w^\varepsilon, Dw^\varepsilon, D^2w^\varepsilon\right)$.

2. Building an almost everywhere smooth supersolution to (2.1)–(2.2) using the mollification of the solution to the switching system.
3. Using Proposition 3.10, the comparison principle for the scheme, to bound the difference of the supersolution obtained in Step 2 and the approximate solution obtained from the scheme.

Step 1. Consider the following switching system:

$$\min\left\{\max\left\{-v_t^{\varepsilon,i}+\sup_{0<s<\varepsilon^2,|y|<\varepsilon}\mathcal{L}^{\alpha_i}v^{\varepsilon,i}(\cdot-s,\cdot+y),v^{\varepsilon,i}-\mathcal{M}^{(i)}v^{\varepsilon}\right\},v^{\varepsilon,i}-g\right\}(t,x)=0, \tag{3.11}$$

$$v^{\varepsilon,i}(T,\cdot)=g(T,\cdot), \tag{3.12}$$

where $v^\varepsilon=(v^{\varepsilon,i})_{i=1}^M,\mathcal{M}^{(i)}v^\varepsilon=\min_{j:j\neq i}\{v^{\varepsilon,j}+k\},k$ is a nonnegative constant, α_i's, for $i=1,\cdots,M$, are as in assumption **HJB+**, and $\mathcal{L}^{\alpha_i}\varphi:=\frac{1}{2}\mathrm{Tr}\,[a^{\alpha_i}(t,x)D^2\varphi]+b^{\alpha_i}(t,x)D\varphi+c^{\alpha_i}(t,x)\varphi+f^{\alpha_i}(t,x)$.

The above system of equations approximates (2.1)–(2.2). Intuitively speaking, Assumption **HJB+** introduces a set of approximating controls $\{\alpha_i\}_{i=1}^{M_\delta}$ in $\mathcal{A}$. In the corresponding optimization problem, the maximum cost of restricting controls to the set $\{\alpha_i\}_{i=1}^{M_\delta}$ is proportional to δ. In addition, the above switching system imposes a switching cost of k between controls in the finite set $\{\alpha_i\}_{i=1}^{M_\delta}$. If k goes to zero, then all functions $v^{\varepsilon,i}$ in the solution of problem (3.11)–(3.12) converges to the function v^ε, the solution of the problem without switching cost, that is, (3.6)–(3.7). Conversely, we have already seen that v^ε approximates function v, the solution of (2.1)–(2.2).

More rigorously, by Theorem A.6 the viscosity solution $\left(v^{\varepsilon,i}\right)_{i=1}^M$ to (3.11)–(3.12) exists and by Theorem A.7 is Lipschitz continuous on x and $\frac{1}{2}$-Hölder continuous on t. Moreover, by using Assumption **HJB+**, Theorem A.4 and Remark A.2, one can approximate the solution to (2.1)–(2.2) by the solution to (3.11)–(3.12), see Theorem 3.4 in [12] and the proof of Theorem 2.3 of [18] for more details. More precisely by setting $\delta=\varepsilon$, there exists a positive constant C such that

$$|v-v^{\varepsilon,i}|_\infty\leq C(\varepsilon+k^{\frac{1}{3}}).$$

Step 2. Let $v_\varepsilon^{(i)}:=v^{\varepsilon,i}*\rho^\varepsilon$, where $\{\rho^\varepsilon\}$ is as in (3.8). As in Lemma 4.2 of [12] and Lemma 3.4 of [18] for $\varepsilon\leq\left(12\sup_i|v_\varepsilon^{(i)}|_1\right)^{-1}k$, for $i_0\in\operatorname{argmin}_i v_\varepsilon^{(i)}(t,x)$, the function $v_\varepsilon^{(i_0)}$ is a supersolution to

$$-\mathcal{L}^X v_\varepsilon^{(i_0)}(t,x)-F\left(t,x,v_\varepsilon^{(i_0)},Dv_\varepsilon^{(i_0)},D^2v_\varepsilon^{(i_0)}\right)\geq 0. \tag{3.13}$$

Moreover, for any $(t,x)\in[0,T)\times\mathbb{R}^d$, we have $v_\varepsilon^{(i_0)}(t,x)<v_\varepsilon^{(i)}(t,x)+k$. Therefore, for all i we have

$$(w_\varepsilon-v)(t,x)=(v_\varepsilon^{(i_0)}-v)(t,x)\leq(v_\varepsilon^{(i_0)}-v_\varepsilon^{(i)})(t,x)+(v_\varepsilon^{(i)}-v)(t,x)\leq k+C(\varepsilon+k^{\frac{1}{3}}).$$

Choosing $k=C_1\varepsilon$ with $C_1=12\sup_i|v_\varepsilon^{(i)}|_1$, one can write

$$(w_\varepsilon-v)(t,x)\leq C\varepsilon^{\frac{1}{3}}. \tag{3.14}$$

Step 3. By the definition of w_ε, for any (t,x) and $i_0 \in \text{argmin}_i v_\varepsilon^{(i)}(t,x)$, we have $w_\varepsilon(t,x) = v_\varepsilon^{(i_0)}(t,x)$ and $w_\varepsilon \le v_\varepsilon^{(i_0)}$ elsewhere. Therefore, $\mathbf{T}_h[w_\varepsilon](t,x) \le \mathbf{T}_h[v_\varepsilon^{(i_0)}]$. Moreover, as (3.9) is satisfied by $v_\varepsilon^{(i_0)}$, by Lemma 3.22 of [5], one can conclude that

$$\mathcal{R}_h[v_\varepsilon^{(i_0)}](t,x) := \frac{v_\varepsilon^{(i_0)}(t,x) - \mathbf{T}_h[v_\varepsilon^{(i_0)}](t,x)}{h} + \mathcal{L}^X v_\varepsilon^{(i_0)}(t,x) + F(t,x,v_\varepsilon^{(i_0)}, Dv_\varepsilon^{(i_0)}, D^2 v_\varepsilon^{(i_0)}) \ge -Ch\varepsilon^{-3}.$$

Therefore, due to (3.13), $\frac{w_\varepsilon(t,x)-\mathbf{T}_h[w_\varepsilon](t,x)}{h} \ge -Ch\varepsilon^{-3}$ holds true. By Proposition 3.10, one can get

$$(v^h - w_\varepsilon)(t,x) \le Ch\varepsilon^{-3}. \tag{3.15}$$

Now, (3.14) and (3.15) yield

$$(v^h - v)(t,x) \le C(\varepsilon^{\frac{1}{3}} + \varepsilon^{-3}h).$$

By minimizing on $\varepsilon > 0$, the desired lower bound is obtained. ∎

Remark 3.11 (Stochastic scheme). Scheme (2.4) produces a deterministic approximate solution. However, in practice, we approximate the expectations in (2.5) based on a randomly generated set of sample paths of the process $\hat{X}$. As a result, the approximate solution is not deterministic anymore. By following the line of argument in Section IV of [5], one can show the almost sure convergence of this stochastic approximate solution and even provide the same rate of convergence in $\mathbb{L}^p(\Omega,\mathbb{P})$.

More precisely, assume that $\mathbb{E}$ is approximated by $\hat{\mathbb{E}}^N$ where N denotes the number of sample paths. Suppose that for some $p \ge 1$, there exist constants $C_b, \lambda, \nu > 0$ such that $||\hat{\mathbb{E}}^N[R] - \mathbb{E}[R]||_p \le C_b h^{-\lambda} N^{-\nu}$ for a suitable class of random variables R bounded by b. By replacing $\mathbb{E}$ with $\hat{\mathbb{E}}^N$ in the scheme (2.4), one obtains a stochastic approximate solution $\hat{v}_N^h$. Then, if we choose $N = N_h$ which is chosen to satisfy $\lim_{h\to 0} N_h^\nu h^{\lambda+2} = \infty$, then under assumptions of Theorem 3.4

$$\hat{v}_{N_h}^h(\cdot,\omega) \to v \text{ locally uniformly},$$

for almost every ω where v is the unique viscosity solution of (2.1)–(2.2). In addition, if

$$\lim_{h\to 0} N_h^\nu h^{\lambda+\frac{21}{10}} > 0, \tag{3.16}$$

we have that $||v - \hat{v}_{N_h}^h||_p \le Ch^{1/10}$, under the assumptions of Theorem 3.6.

IV. NUMERICAL RESULTS

A. Risk Neutral Pricing of Geometric American Put Option

We consider a geometric American put option on three risky assets each of which follows a Black–Scholes dynamics under risk neutral probability measure. The payoff of the option is given by $(K - \xi(T))_+$ where K and T are, respectively, the strike price and maturity and $\xi(t) := \prod_{i=1}^{3} S_i(t)$, where

$$dS(t) = \text{diag}\,(S(t))(rdt + \text{diag}\,(\Sigma) \cdot dW(t).$$

Here, $S(t) = (S_i(t))_{i=1}^3$ is the vector of asset prices, $W(t) = (W_i(t))_{i=1}^3$ is a three-dimensional (3D) Brownian motion, r is the risk free interest rate, and $\Sigma = (\sigma_i)_{i=1}^3$ where σ_i is the volatility of the ith asset.

The price of this option at time t and for asset price vector $s = (s_1, s_2, s_3)$ is given by

$$v(t,s) := \sup \mathbb{E}\left[e^{-r(\tau-t)}(K - \xi(\tau))_+ | S(t) = s\right] \tag{4.1}$$

where the supremum is over all stopping times $\tau \in [t, T]$ adapted to the filtration generated by the 3D Brownian motion and $\mathbb{E}$ is the risk neutral expectation. It is well-known that function v satisfies the following differential equation

$$0 = \min\left\{-\partial_t v - \frac{1}{2}\sum_{i=1}^{3} s_i^2\sigma_i^2\partial_{s_i s_i} v - r\sum_{i=1}^{3} s_i\partial_{s_i} v + rv, v - g\right\}$$

$$v(T,s) = g(s).$$

where $g(s) = (K - \prod_{i=1}^{3} s_i)_+$. We treat this linear equation as a fully nonlinear one by separating the linear second-order operator into two parts. More precisely, for some σ_0^2, we choose the linear and nonlinear parts to be $\mathcal{L}^X\phi := \frac{\sigma_0^2}{2}\sum_{i=1}^{3} s_i^2\sigma_i^2\partial_{s_i s_i}\phi + r\sum_{i=1}^{3} s_i\partial_{s_i}\phi$ and $F(\cdot,\cdot,r\phi, D\phi, D^2\phi) = \frac{1-\sigma_0^2}{2}\sum_{i=1}^{3} s_i^2\sigma_i^2\partial_{s_i s_i}\phi$, respectively. This leads to the choice of diffusion $X(t) := (X_i(t))_{i=1}^3$

$$dX(t) = \sigma_0 \text{diag}\,(X(t))\text{diag}\,(\sigma) \cdot dW(t)$$

for the approximation scheme (2.4). Conversely, the approximation of the second-order derivatives in (2.6) is given by

$$x_i^2\partial_{x_i x_i} v(t,x) \approx \frac{1-\sigma_0^2}{2\sigma_0^2}\mathbb{E}\left[v(t+h, x + \sigma_0 \text{diag}\,(x)\text{diag}\,(\sigma) \cdot W(h))\frac{W_i(h)^2 - h}{h^2}\right],$$

where $x = (x_i)_{i=1}^3$. For the numerical implementation, we choose the continuous-time interest rate $r = 0.03$, volatility of all assets $\sigma_i = 0.1$, $T = 1$, $K = 8$, and we evaluate the option at time $t = 0$ at $s = (2, 2, 2)$. The reference value for option price is obtained by applying the binomial tree algorithm to the one-dimensional (1D) optimal stopping problem

$$v(t,s) := \sup_{\tau} \mathbb{E}\left[e^{-r(\tau-t)}\xi(T) | \xi(T) = \prod_{i=1}^{3} s_i\right].$$

on the diffusion $\xi_t := \prod_{i=1}^{3} S_i(t)$ satisfying

$$d\xi(t) = \xi(t)(3rdt + \overline{\sigma}dB_t) \text{ with } \overline{\sigma} := \left(\sum_{i=1}^{3}\sigma_i^2\right)^{\frac{1}{2}}$$

TABLE I. The simulation results for geometric American put option regarded as a nonlinear problem.

N	Time	$\sigma_0^2 = 0.9$	$\sigma_0^2 = 1$
5	92	0.301173	0.326258
10	234	0.309001	0.334205
15	360	0.312642	0.337974
20	499	0.314978	0.340397
40	1050	0.320354	0.347909
50	1325	0.322115	0.346041

# of sample paths = 6 million, N = # of time steps, Time = time of the algorithm in seconds.

where B_t is a 1D Brownian motion. The binomial tree algorithm stabilizes to the value 0.338778 for more than 20,000 time steps. The numerical result as well as the run time of the algorithm[2] is provided in Table I for two different choices for σ_0. Here, we used projection method described in [10] with 8^5 locally linear basis functions with compact support. Notice that $\sigma_0^2 = 1$ corresponds to the Longstaff–Schwartz algorithm and for $\sigma_0^2 = 0.9$, the inequality (3.1) is satisfied. In Fig. 1, the red graph is the ratio of error for two consecutive time steps plotted against the ratio of the time steps, whereas the green graph is the theoretical rate of convergence; that is, $\frac{1}{4}$. Because of the analysis in [5, Section 3.4], one expects to have a higher the rate of convergence for the scheme on the linear equations. Therefore, the simulated red plot must lie below the theoretical green plot. Conversely, due to error of approximation of expectations in the scheme, the rate of convergence $\frac{1}{2}$ will never be obtained in practice.

B. Indifference Pricing of Geometric American Put Option

We consider a geometric put option on two nontradable risky assets with Black–Scholes dynamics given by $S(t) := (S_1(t), S_1(t))$

$$dS_i(t) = S(t)(\mu_i dt + \sigma_i dW_i(t)),$$

where $W(t) = (W_1(t), W_2(t))$ is a two-dimensional (2D) Brownian motion, and μ_i and σ_i are the drift and the volatility of the ith asset for $i = 1, 2$, respectively. We assume that there is a portfolio made of a tradable asset with Black–Scholes dynamics and money market with $r = 0$ interest rate which satisfies

$$dX_t^\theta = \theta_t(\mu_0 dt + \sigma_0 dB(t)),$$

where θ is the amount of money in risky asset, $B(t)$ is a one dimensional Brownian motion, and μ_0 and σ_0 are drift and volatility of the tradable asset, respectively. Here we assume that $dW_i(t) \cdot dB(t) = \rho_i dt$ for $i = 1, 2$. The indifference pricing with exponential utility leads to the controller-stopper problem below

$$v(t, x, s_1, s_2) := \sup_{\tau,\theta} \mathbb{E}\left[-\exp\left(-\gamma\left(X^\theta(\tau) + \left(K - \prod_{i=1}^{2} S_i(\tau)\right)_+\right)\right) | X(t) = x, S_i(t) = s_i,\ i = 1, 2\right]. \tag{4.2}$$

[2] Dual core i5 2.5 GHz, 4 GB of RAM

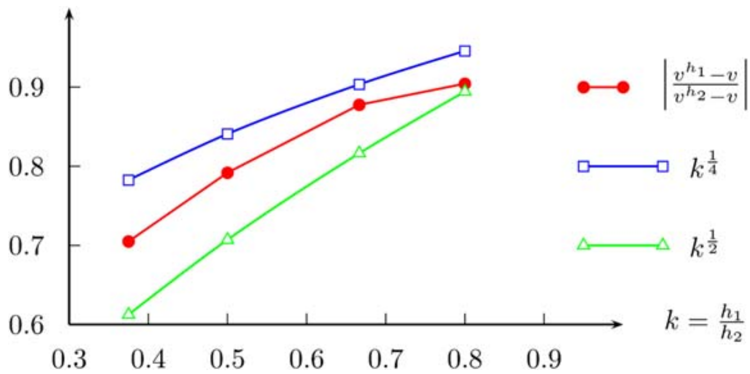

FIG. 1. Rate of convergence analysis: •: ratio of the error, that is, $|\frac{v^{h_1}-v}{v^{h_2}-v}|$, Δ: $(h_1/h_2)^{\frac{1}{2}}$, and □: $(h_1/h_2)^{\frac{1}{4}}$ (vertical axis) versus h_1/h_2 (horizontal axis). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

which satisfies the fully nonlinear obstacle problem below:

$$\min\left\{-\frac{\partial v}{\partial t}+\frac{\left(\mu v_x+\sum_{i=1}^{2}\sigma_0\rho_i\sigma_i s_i\partial_{xs_i}v\right)^2}{2\sigma_0^2\partial_{xx}v}-\mathcal{L}^S v, v-g\right\}=0$$
$$v(T,x,s_1,s_2)=-\exp(-\gamma(x+(K-\textstyle\prod_{i=1}^{2}s_i)_+))$$

where $\gamma>0$ is a constant and $g(t,x,s_1,s_2)=-\exp(-\frac{\mu_0^2}{2\sigma_0^2}(T-t)-\gamma(x+(K-\prod_{i=1}^{2}s_i)_+))$, K is the strike price and $\mathcal{L}^S=\sum_{i=1}^{2}s_i\mu_i\partial_{s_i}+\frac{1}{2}\sum_{i=1}^{2}s_i\sigma_i^2\partial_{s_is_i}$. To solve the above free boundary problem using scheme (2.4), we choose the linear and nonlinear parts as follow:

$$\overline{\mathcal{L}}\phi=\mathcal{L}^S\phi+\frac{1}{2}\varepsilon^2\partial_{xx}\phi$$
$$F(\cdot,\mathcal{D}\phi)=-\frac{\left(\mu\phi_x+\sum_{i=1}^{2}\sigma_0\rho_i\sigma_i s_i\partial_{xs_i}\phi\right)^2}{2\sigma_0^2\partial_{xx}\phi}-\frac{1}{2}\varepsilon^2\partial_{xx}\phi.$$

Thus, the appropriate diffusion to be used inside (2.4) is

$$d\bar{X}=\varepsilon d\bar{B}(t),$$
$$dS(t)=\text{diag}\,(S(t))(\mu dt+\text{diag}\,(\Sigma)\cdot dW(t),$$

where $\mu=(\mu_1,\mu_2)$, $\Sigma=(\sigma_1,\sigma_2)$, and $\bar{B}(t)$ is a 1D Brownian motion independent of $W(t)$.

To find a reference value for the solution, we follow the same idea as in Section IV A. Since $\xi_t:=\prod_{i=1}^{2}S_i(t)$ satisfies

$$d\xi(t)=\xi(t)(\overline{\mu}dt+\overline{\sigma}dB_t)\text{with}\,\overline{\mu}:=\sum_{i=1}^{2}\mu_i,\ \overline{\sigma}:=\left(\sum_{i=1}^{2}\sigma_i^2\right)^{\frac{1}{2}},$$

TABLE II. M = # of sample paths in million, N = # of time steps, Time = time of the algorithm on -dimensional problem in seconds, $\hat{v}$ and $\hat{u}$ value obtained by scheme (2.4) for the 3D problem (4.2) and the 2D problem with the same value function.

N	M	Time	$\hat{v}(0,1,1,1)$	$\hat{u}(0,1,1)$
5	2	34	−0.341675	−0.349489
10	1	36	−0.303425	−0.352110
	2	76	−0.356332	−0.351678
20	2	180	−0.356126	−0.351659
	3	273	−0.351659	−0.356126
	4	372	−0.348773	−0.350201
30	3	422	−0.353311	−0.353088
40	4	696	−0.322095	−0.361026

Notice that $v(0,1,1,1) = u(0,1,1)$.

we have $v(t,x,s_1,s_2) = u(t,x,\prod_{i=1}^{2} s_i)$ where the function u is the solution of the 2D controller-stopper problem

$$u(t,x,s) = \sup_{\tau,\theta}\mathbb{E}\left[-\exp(-\gamma(X^{\theta}(\tau) + (K-\xi(\tau))_+))|X(t)=x,\xi(t)=s\right]. \qquad (4.3)$$

Neither function v nor u have closed form solutions, but we expect that if the scheme converges numerically, it approximates the function u more accurately because of the reduction in the dimension. This is because the number of sample paths and time steps for a 2D problem for u can be chosen larger than the 3D problem for v. Therefore, to examine the convergence of scheme, we compare the approximation of these functions by the scheme (2.4). We set $K = 1, \gamma = 1, T = 1, \varepsilon = 0.05, X(0) = 1, \rho_i = 0.1, \mu_0 = \sigma_0 = \mu_i = \sigma_i = 0.1$, and $S_i(0) = 1$ for $i = 1, 2$.

The result of the simulation is summarized in Table II. In Fig. 2, we establish the convergence analysis for the nonlinear problem by using the approximations in Table II with the largest number of sample paths for each time step.

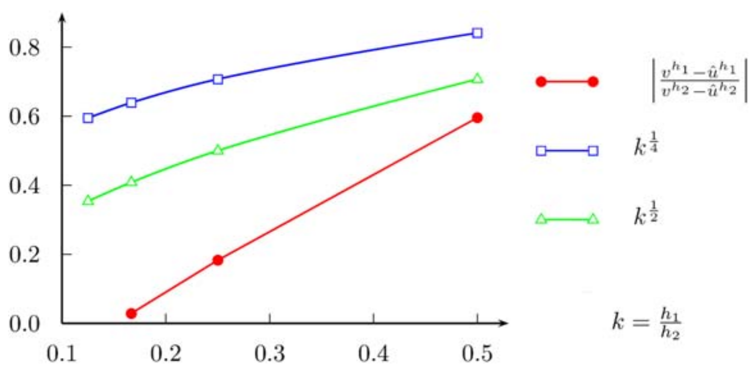

FIG. 2. Rate of convergence analysis: ●: ratio of the error, that is, $|\frac{v^{h_1}-\hat{u}^{h_1}}{v^{h_2}-\hat{u}^{h_2}}|$, Δ: $(h_1/h_2)^{\frac{1}{2}}$, and □: $(h_1/h_2)^{\frac{1}{4}}$ (vertical axis) versus $k = h_1/h_2$ (horizontal axis). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

APPENDIX A: A SWITCHING SYSTEM WITH AN OBSTACLE

In this section, we will provide some results needed in the Section III C. In particular, we present a continuous dependence result for the switching system with obstacle and as a corollary a comparison result, which provides the uniqueness of the solution. Then, the existence and regularity of the solutions to the switching systems are provided.

Consider the following system of PDEs for $v = (v^{(i)})_{i=1}^M$:

$$\min\left\{\max\left\{-v_t^{(i)} - F_i(\cdot, v^{(i)}, Dv^{(i)}, D^2v^{(i)}), v^{(i)} - \mathcal{M}^{(i)}v\right\}, v^{(i)} - g\right\} = 0, \text{for } i = 1, \cdots, M;$$
$$v^{(i)}(T, \cdot) = g(T, \cdot). \tag{A.1}$$

We also need to consider a variant of Eq. (A.1) as follows:

$$\min\left\{\max\left\{-v_t^{(i)} - \hat{F}_i(\cdot, v^{(i)}, Dv^{(i)}, D^2v^{(i)}), v^{(i)} - \mathcal{M}^{(i)}v\right\}, v^{(i)} - \hat{g}\right\} = 0, \text{for } i = 1, \cdots, M;$$
$$v^{(i)}(T, \cdot) = \hat{g}(T, \cdot). \tag{A.2}$$

Assumption HJB-S.

We assume that in (A.1) and (A.2)

$$F_i(\cdot, v^{(i)}, Dv^{(i)}, D^2v^{(i)}) = \inf_{\alpha\in\mathcal{A}^i} \mathcal{L}^{i,\alpha}v^{(i)} \text{ and } \hat{F}_i(\cdot, v^{(i)}, Dv^{(i)}, D^2v^{(i)}) = \inf_{\alpha\in\mathcal{A}^i} \hat{\mathcal{L}}^{i,\alpha}v^{(i)}, \tag{A.3}$$

$\mathcal{M}^{(i)}v = \min_{j:j\neq i}\{v_j + k\}, k$ *is a nonnegative constant, and*

$$\mathcal{L}^{i,\alpha}\varphi(x) := \frac{1}{2}Tr[a_i^\alpha(t,x)D^2\varphi] + b_i^\alpha(t,x)D\varphi + c_i^\alpha(t,x)\varphi + f_i^\alpha(t,x),$$
$$\hat{\mathcal{L}}^{i,\alpha}\varphi(x) := \frac{1}{2}Tr[\hat{a}_i^\alpha(t,x)D^2\varphi] + \hat{b}_i^\alpha(t,x)D\varphi + \hat{c}_i^\alpha(t,x)\varphi + \hat{f}_i^\alpha(t,x).$$

Moreover,

$$L := |g|_1 + |\hat{g}|_1 + \sup_{\alpha\in\bigcup_i\mathcal{A}^i}\left(|\sigma^\alpha|_1 + |b^\alpha|_1 + |c^\alpha|_1 + |f^\alpha|_1 + |\hat{\sigma}^\alpha|_1 + |\hat{b}^\alpha|_1 + |\hat{c}^\alpha|_1 + |\hat{f}^\alpha|_1\right) < \infty,$$

and for all $\alpha \in \bigcup_i \mathcal{A}^i$, *we have* $c_i^\alpha, \hat{c}_i^\alpha \leq -1$.

Remark A.1. $c_i^\alpha, \hat{c}_i^\alpha \leq -1$ in Assumption **HJB-S** is only to make the proofs simpler and is not a loss of generality. This can be seen by applying the change of variable $v^{(i)} \to e^{C(T-t)}v^{(i)}$ in Eqs. (A.1) and (A.2) for C large enough.

Remark A.2. For the sake of simplicity in Assumption **HJB-S**, we only included the nonlinearities of infimum type. However, all the results of this appendix still hold if we assume that

$$F_i(\cdot, v^{(i)}, Dv^{(i)}, D^2v^{(i)}) = \inf_{\alpha\in\mathcal{A}^i}\sup_{\beta\in\mathcal{B}^i}\mathcal{L}^{i,\alpha,\beta} \text{ and } \hat{F}_i(\cdot, v^{(i)}, Dv^{(i)}, D^2v^{(i)}) = \inf_{\alpha\in\mathcal{A}^i}\sup_{\beta\in\mathcal{B}^i}\hat{\mathcal{L}}^{i,\alpha,\beta},$$

$$|g|_1 + |\hat{g}|_1 + \sup_{\substack{\alpha \in \cup_i \mathcal{A}^i \\ \beta \in \cup_i \mathcal{B}^i}}$$

$$\left(|\sigma_i^{\alpha,\beta}|_1 + |b_i^{\alpha,\beta}|_1 + |c_i^{\alpha,\beta}|_1 + |f_i^{\alpha,\beta}|_1 + |\hat{\sigma}_i^{\alpha,\beta}|_1 + |\hat{b}_i^{\alpha,\beta}|_1 + |\hat{c}_i^{\alpha,\beta}|_1 + |\hat{f}_i^{\alpha,\beta}|_1\right) < \infty, \quad \text{(A.4)}$$

and for all $\alpha \in \cup_i \mathcal{A}^i$ and $\beta \in \cup_i \mathcal{B}^i$, we have $c_i^{\alpha,\beta}, \hat{c}_i^{\alpha} \leq -1$. This remark is also valid if we change the order of inf and sup in (4).

Lemma A.3. *Let $u = (u^{(i)})_i$ and $v = (v^{(i)})_i$ be, respectively, the upper semicontinuous subsolution and the lower semicontinuous supersolution of (A.1) and (A.2), and assume that $\varphi(t,x,y)$ is a smooth function bounded from below. Define*

$$\psi^{(i)}(t,x,y) \equiv u^{(i)}(t,x) - v^{(i)}(t,y) - \varphi(t,x,y),$$

$$\mathcal{J}_1 := \left\{ j \,|\, \exists (t',x',y') : \sup_{i,t,x,y} \psi^{(i)}(t,x,y) = \psi^{(j)}(t',x',y') \right\},$$

$$\mathcal{J}_2(t,x) := \left\{ j | u^{(j)}(t,x) \leq g(t,x) \right\}.$$

Suppose that there exists an (i_0', t_0, x_0, y_0) such that $\sup_{i,t,x,y} \psi^{(i)}(t,x,y) = \psi^{(i_0')}(t_0,x_0,y_0)$ and $\mathcal{J}_1 \cap \mathcal{J}_2(t_0,x_0) = \emptyset$. Then, there exists an i_0 such that $\psi^{(i_0)}(t_0,x_0,y_0) = \psi^{(i_0')}(t_0,x_0,y_0)$ and

$$v^{(i_0)}(t_0,y_0) < \mathcal{M}^{(i_0)}v(t_0,y_0). \quad \text{(A.5)}$$

Moreover, if in a neighborhood of (t_0,x_0,y_0) there are some continuous functions $h_0(t,x,y) > 0, h(t,x)$ and $\hat{h}(t,y)$ such that

$$D^2\varphi(t,x,y) \leq h_0(t,x,y) \begin{pmatrix} I & -I \\ -I & I \end{pmatrix} + \begin{pmatrix} h(t,x) & 0 \\ 0 & \hat{h}(t,y) \end{pmatrix},$$

then, there are $a, b \in \mathbb{R}$ and $X, Y \in \mathbb{S}_+^d$ such that

$$a - b = \varphi_t(t_0,x_0,y_0), \quad \text{(A.6)}$$

$$\begin{pmatrix} X & 0 \\ 0 & -Y \end{pmatrix} \leq 2h_0(t_0,x_0,y_0) \begin{pmatrix} I & -I \\ -I & I \end{pmatrix} + \begin{pmatrix} h(t_0,x_0) & 0 \\ 0 & \hat{h}(t_0,y_0) \end{pmatrix}, \quad \text{(A.7)}$$

$$-a - \inf_{\alpha\in\mathcal{A}_{i_0}} \left\{ \frac{1}{2} Tr[a_{i_0}^{\alpha}(t_0,x_0)X] + b_{i_0}^{\alpha}(t_0,x_0) D_x\varphi(t_0,x_0,y_0) \right.$$
$$\left. + c_{i_0}^{\alpha}(t_0,x_0)u^{(i_0)}(t_0,x_0) + f_{i_0}^{\alpha}(t_0,x_0) \right\} \leq 0, \quad \text{(A.8)}$$

$$-b - \inf_{\alpha\in\mathcal{A}_{i_0}} \left\{ \frac{1}{2} Tr[\hat{a}_{i_0}^{\alpha}(t_0,y_0)Y] + \hat{b}_{i_0}^{\alpha}(t_0,y_0)(-D_y\varphi(t_0,x_0,y_0)) \right.$$
$$\left. + \hat{c}_{i_0}^{\alpha}(t_0,y_0)v^{(i_0)}(t_0,y_0) + \hat{f}_{i_0}^{\alpha}(t_0,y_0) \right\} \geq 0. \quad \text{(A.9)}$$

Proof. The first part of the proof is similar to those of Lemma A.2 of [24], Lemma A.1 of [12]. The second part follows as a result of Theorem 2.2 of [25]. ∎

The following theorem on continuous dependence is used in Section Proof of Theorem 3.6 (ii) (ii) and in the regularity result, Theorem A.7 below. Intuitively speaking, continuous dependence result asserts that a slight change in the coefficients of (A.1) changes the solution only slightly.

Theorem A.4 (Continuous dependence). *Let* **HJB-S** *hold. Suppose that* $u = (u^{(i)})_i$ *and* $v = (v^{(i)})_i$ *are a bounded upper semicontinuous subsolution of (A.1) and a bounded lower semicontinuous supersolution of (A.2), respectively. Then, for any* $i = 1, \cdots, N$,

$$u^{(i)} - v^{(i)} \leq \mathcal{B} := C\max_j\{|(g-\hat{g})(\cdot,\cdot)|_\infty$$
$$+T\sup_\alpha\left\{|f^{j,\alpha} - \hat{f}^{j,\alpha}|_\infty + (|u|_\infty \vee |v|_\infty)|c^{j,\alpha} - \hat{c}^{j,\alpha}|_\infty\right\}$$
$$+\sqrt{T}\sup_\alpha\left\{|\sigma^{j,\alpha} - \hat{\sigma}^{j,\alpha}|_\infty + |b^{j,\alpha} - \hat{b}^{j,\alpha}|_\infty\right\}\Bigg\}.$$

Proof. Let $\varphi(t,x,y) = e^{\lambda(T-t)}\frac{\theta}{2}|x-y|^2 + e^{\lambda(T-t)}\frac{\varepsilon}{2}(|x|^2+|y|^2)$ and define

$$\mathcal{D} := \sup_{t,i,x,y}\left\{u^{(i)}(t,x) - v^{(i)}(t,y) - \varphi(t,x,y) - \frac{\overline{\varepsilon}}{t}\right\},$$

where $\varepsilon, \overline{\varepsilon} > 0$ are arbitrary constants and constants $\lambda, \theta > 0$ will be determined later in the proof. We will show that $\mathcal{D}$ is bounded by a constant $B(\varepsilon,\overline{\varepsilon},\theta)$ which is bounded by $\mathcal{B}$ mentioned in the theorem as $(\varepsilon,\overline{\varepsilon}) \to (0,0)$ and θ is set appropriately. Then, it would follow that

$$u^{(i)}(t,x) - v^{(i)}(t,x) \leq \mathcal{D} + \frac{\overline{\varepsilon}}{t} + e^{\lambda(T-t)}\frac{2\varepsilon|x|^2}{2} \leq B(\varepsilon,\overline{\varepsilon},\theta) + \frac{\overline{\varepsilon}}{t} + e^{\lambda(T-t)}\frac{2\varepsilon|x|^2}{2}.$$

Sending $\varepsilon, \overline{\varepsilon} \to 0$, one would then obtain

$$u^{(i)}(t,x) - v^{(i)}(t,x) \leq \mathcal{B}, \text{for } t > 0.$$

Note that the above inequality is also valid for $t = 0$ by considering $[-\delta, T]$ as the time interval and by changing T to $T + \delta$.

Define

$$\psi^{(i)}(t,x,y) = u^{(i)}(t,x) - v^{(i)}(t,y) - \varphi(t,x,y) - \frac{\sigma(T-t)}{2T} - \frac{\overline{\varepsilon}}{t},$$

where $\sigma = \mathcal{D} - \sigma_T$ with $\sigma_T = \sup_{i,x,y}\left\{u^{(i)}(T,x) - v^{(i)}(T,y) - \varphi(T,x,y) - \frac{\overline{\varepsilon}}{T}\right\}^+$. Let

$$\overline{\mathcal{D}} := \sup_{i,t,x,y}\psi^{(i)}(t,x,y). \tag{A.10}$$

Since $u^{(i)}$ and $v^{(i)}$ are bounded, we have $\overline{\mathcal{D}} < \infty$. Conversely, by semicontinuity of $u^{(i)}$ and $v^{(i)}$, one can conclude that the supremum in the definition of $\overline{\mathcal{D}}$ is attained at some point (i_0, t_0, x_0, y_0). In other words, $\mathcal{J}_1 \neq \emptyset$ (see Lemma A.3 for the definition of $\mathcal{J}_1$).

If $\sigma \le 0$, then $\mathcal{D} \le \sigma_T$. Since

$$\sigma_T \le |g-\hat{g}|_\infty + \sup_{x,y}\{|g|_1|x-y| - \varphi(T,x,y)\} - \frac{\overline{\varepsilon}}{T}$$
$$\le |g-\hat{g}|_\infty + \sup_{x,y}\{|g|_1|x-y| - \frac{\theta}{2}|x-y|^2\} \le |g-\hat{g}|_\infty + \frac{|g|_1^2}{2\theta},$$

one can conclude that $\mathcal{D} \le |g-\hat{g}|_\infty + \frac{|g|_1^2}{2\theta}$. Therefore, we may assume that $\sigma > 0$. From the definition of $\overline{\mathcal{D}}$, we have $t_0 > 0$. Conversely, $\sigma > 0$ implies $t_0 < T$. Because if $t_0 = T$, then $\sigma_T \ge \overline{\mathcal{D}}$ which implies

$$\sigma_T \ge \mathcal{D} - \frac{\sigma}{2} \ge \sigma_T + \frac{\sigma}{2} > \sigma_T$$

which is a contradiction. So, we have $0 < t_0 < T$. We continue the proof by considering two different cases.

Case 1. $\mathcal{J}_1 \cap \mathcal{J}_2(t_0,x_0) \ne \emptyset$. The supremum in (A.10) is attained at some point (i_0,t_0,x_0,y_0) with $u^{(i)}(t_0,x_0) \le g(t_0,x_0)$ and $v^{(i)}(t_0,y_0) \ge \hat{g}(t_0,y_0)$. Therefore,

$$\overline{\mathcal{D}} \le g(t_0,x_0) - \hat{g}(t_0,y_0) - \varphi(t_0,x_0,y_0) - \frac{\sigma\,(T-t_0)}{2T} - \frac{\overline{\varepsilon}}{t_0}$$
$$\le |g-\hat{g}|_\infty + |g|_1|x_0-y_0| - \frac{\theta}{2}|x_0-y_0|^2 \le |g-\hat{g}|_\infty + \frac{|g|_1^2}{2\theta}.$$

Conversely, since $\mathcal{D} \le \overline{\mathcal{D}} + \frac{\sigma}{2} \le \overline{\mathcal{D}} + \frac{1}{2}(\mathcal{D}-\sigma_T) \le \overline{\mathcal{D}} + \frac{1}{2}\mathcal{D}$, we have $\mathcal{D} \le 2|g-\hat{g}|_\infty + \frac{|g|_1^2}{\theta}$.
Case 2. $\mathcal{J}_1 \cap \mathcal{J}_2(t_0,x_0) = \emptyset$. In this case, (A.5) is satisfied and by Lemma A.3 and the same line of argument as Theorem A.1 in [14] the result is provided. For the convenience of the reader, we present a sketch of the proof.

By subtracting (A.9) from (A.8), we have

$$\lambda\phi(t_0,x_0,y_0) + \frac{\mathcal{D}-\sigma_T}{2T} + \frac{\overline{\varepsilon}}{t_0^2} - \inf_{\alpha\in\mathcal{A}_{i_0}} \{\frac{1}{2}(\mathrm{Tr}\,[a_{i_0}^\alpha(t_0,x_0)X] - \mathrm{Tr}\,[\hat{a}_{i_0}^\alpha(t_0,y_0)Y])$$
$$-b_{i_0}^\alpha(t_0,x_0)D_x\varphi(t_0,x_0,y_0) - \hat{b}_{i_0}^\alpha(t_0,y_0)D_y\varphi(t_0,x_0,y_0)$$
$$-c_{i_0}^\alpha(t_0,x_0)(u^{(i_0)}(t_0,x_0) - v^{(i_0)}(t_0,y_0))$$
$$-(c_{i_0}^\alpha(t_0,x_0) - \hat{c}_{i_0}^\alpha(t_0,y_0))v^{(i_0)}(t_0,y_0)$$
$$-f_{i_0}^\alpha(t_0,x_0) - \hat{f}_{i_0}^\alpha(t_0,y_0)\} \le 0$$

Now, using $c_{i_0}^\alpha \le -1$ and $u^{(i_0)}(t_0,x_0) - v^{(i_0)}(t_0,y_0) \ge \overline{\mathcal{D}} \ge \frac{1}{2}\mathcal{D} > 0$ together with (A.6)-(A.7), one can obtain the following bound for $\mathcal{D}$

$$\mathcal{D} \le CT(\theta\sup_\alpha\left\{|\sigma^{j,\alpha} - \hat{\sigma}^{j,\alpha}|_\infty^2 + |b^{j,\alpha} - \hat{b}^{j,\alpha}|_\infty^2\right\}$$
$$+\sup_\alpha\left\{|f^{j,\alpha} - \hat{f}^{j,\alpha}|_\infty + (|u|_\infty \vee |v|_\infty)|c^{j,\alpha} - \hat{c}^{j,\alpha}|_\infty\right\} + \sigma_T) + C_1|x_0-y_0|^2 - \lambda\phi(t_0,x_0,y_0),$$

where C_1 is the constant depending only on L in Assumption **HJB-S**. After we choose $\lambda \geq C_1$ in the above and maximize the right-hand side with respect to θ, the proof is complete. ∎

The following result is a straightforward consequence of Theorem A.4, and will be used to establish the existence and the regularity of the solution to (A.1).

Corollary A.5. *Assume that **HJB-S** holds. Suppose that $u = (u^{(i)})_i$ and $v = (v^{(i)})_i$ are, respectively, a bounded upper semicontinuous subsolution and a bounded lower semicontinuous supersolution of (A.1). Then, for any $i = 1, \cdots, N, u^{(i)} \leq v^{(i)}$*

Theorem A.6 (Existence). *Assume that **HJB-S** holds. Then, there exists a unique continuous viscosity solution in the class of bounded functions to (1)*

Proof. We follow Perron's method (see e.g., Section IV of [25]). Observe that by Assumption **HJB-S**, $\underline{u} = -K$ and $\bar{v} = K$ are, respectively, subsolution and supersolution of (A.1) for a suitable choice of positive constant K. Define $v^{(i)}(t,x) := \sup\{u^{(i)}(t,x)\,;\, u \text{ is a subsolution to (A.1)}\}$ and

$$v^{(i)*}(t,x) := \lim_{\delta\to 0} \sup\{v^{(i)}(s,y) : |x-y| + |s-t| \leq \delta, s \in [0,T]\},$$

and

$$v^{(i)}_*(t,x) := \lim_{\delta\to 0} \inf\{v^{(i)}(s,y) : |x-y| + |s-t| \leq \delta, s \in [0,T]\}.$$

It is straight forward that $-K \leq v^{(i)}_* \leq v^{(i)*} \leq K$. We want to show that $(v^{(i)*})_{i=1}^M$ and $(v^{(i)}_*)_{i=1}^M$ are, respectively, a subsolution and a supersolution to (A.1) which by comparison, Corollary A.5, yields the desired result.

Step 1: Subsolution property of $v^{(i)*}$. We start by showing that $(U, \cdots, U)$ with

$$U(t,x) := a_\varepsilon(T-t) + g(T,z) + |g|_1\big(T - t + |x-z|^2 + \varepsilon\big)^{\frac{1}{2}}$$

is a supersolution to (A.1) for a suitable positive constant a_ε. Observe that since

$$U(t,x) - g(t,x) \geq g(T,z) - g(t,x) + |g|_1\big(T - t + |x-z|^2 + \varepsilon\big)^{\frac{1}{2}} \geq 0,$$

we have that $U(t,x) \geq g(t,x)$, and in particular, $U(T,x) \geq g(T,x)$. Conversely, by simple calculations, one can show that, for an appropriate choice of a_ε, we have $-U_t - \inf_{\alpha\in\mathcal{A}^i} \mathcal{L}^{i,\alpha} U \geq 0$.

Therefore, by comparison, Corollary A.5, for any subsolution $u, u \leq U$ which implies $v^{(i)*} \leq U$; specially $v^{(i)*}(T,x) \leq U(T,x)$. Sending $\varepsilon \to 0$ and setting $x = z$, $v^{(i)*}(T,x) \leq g(T,x)$.

Now, for fixed i, we suppose $t < T$ and φ is a test function such that

$$0 = \max_{[0,T]\times\mathbb{R}^d}\{v^{(i)*} - \varphi\} = (v^{(i)*} - \varphi)(t,x).$$

It follows from the definition of $v^{(i)*}$ that there exists a sequence $\{(u_n, t_n, x_n)\}_n$ with $t_n < T$ such that u_n is a subsolution to (A.1), $(t_n, x_n) \to (t, x), u_n^{(i)}(t_n, x_n) \to v^{(i)*}(t, x)$, and (t_n, x_n) is the global strict maximum of $u_n^{(i)} - \varphi$. Let $\delta_n := \max_{[0,T]\times\mathbb{R}^d}\{u_n^{(i)} - \varphi\}$. By the subsolution property of u_n , we have

$$\min\left\{\max\left\{-\varphi_t - \inf_{\alpha\in\mathcal{A}^i}\mathcal{L}^{i,\alpha}(\varphi+\delta_n), u_n^{(i)} - \mathcal{M}^{(i)}u_n\right\}, \varphi+\delta_n - g\right\}(t_n, x_n) \le 0.$$

Because $\mathcal{M}^{(i)}u_n \le \mathcal{M}^{(i)}v^*$, by sending $n \to \infty$,

$$\min\left\{\max\left\{-\varphi_t - \inf_{\alpha\in\mathcal{A}^i}\mathcal{L}^{i,\alpha}\varphi, v^{(i)*} - \mathcal{M}^{(i)}v^*\right\}, v^{(i)*} - g\right\}(t, x) \le 0.$$

Step 2: Supersolution property of $v_*^{(i)}$. Since $(g, \cdots, g)$ is a subsolution to (A.1), $v_*^{(i)}(t, x) \ge g(t, x)$. In particular, $v_*^{(i)}(T, x) \ge g(T, x)$. Therefore, we only need to show that

$$\max\left\{-(v_*^{(i)})_t - \inf_{\alpha\in\mathcal{A}^i}\mathcal{L}^{i,\alpha}v_*^{(i)}, v_*^{(i)} - \mathcal{M}^{(i)}v_*\right\} \ge 0, \tag{A.11}$$

on $[0, T) \times \mathbb{R}^d$ in the viscosity sense. We will prove (A.11) by a contradiction argument. Assume that there are a test function φ and (i, t, x) with $t < T$ such that (t, x) is the global strict minimum of $v_*^{(i)} - \varphi$ and $(v_*^{(i)} - \varphi)(t, x) = 0$ but $\max\left\{-\varphi_t - \inf_{\alpha\in\mathcal{A}^i}\mathcal{L}^{i,\alpha}\varphi, \varphi - \mathcal{M}^{(i)}v_*\right\}(t, x) < 0$. Then, by continuity of φ and the equation and lower semicontinuity of v_*, one can find $\varepsilon > 0$ and $\delta > 0$ small enough, such that for $|x - y| + |s - t| < \delta$ we have that $\varphi + \varepsilon < v_*^{(i)}$ and that

$$\max\left\{-(\varphi+\varepsilon)_t - \inf_{\alpha\in\mathcal{A}^i}\mathcal{L}^{i,\alpha}(\varphi+\varepsilon), (\varphi+\varepsilon) - \mathcal{M}^{(i)}v_*\right\}(s, y) < 0. \tag{A.12}$$

Define

$$w^{(j)}(s, y) := \begin{cases}\max\{\varphi+\varepsilon, v^{(j)*}\}(s, y), & j = i \text{ and } |x-y|+|s-t| < \delta;\\ v^{(j)*}(s, y), & \text{otherwise}.\end{cases}$$

Since v^* is a subsolution to (A.1) and by (A.11), one can show that w is a subsolution to (A.1). By the definition of $v_*^{(i)}$, we must have $v_*^{(i)} \ge w^{(i)}$, which contradicts the fact that $w^{(i)}(t, x) = \varphi(t, x) + \varepsilon < v_*^{(i)}(t, x)$ for $|x - y| + |s - t| < \delta$. ∎

Theorem A.7 (Regularity). *Assume that* ***HJB-S*** *holds. Let $(u^{(i)})_{i=1}^M$ be the solution to (A.1). Then, $(u^{(i)})_{i=1}^M$ is Lipschitz continuous with respect to x and $\frac{1}{2}$-Hölder continuous with respect to t on $\mathbb{R}^d \times [0, T]$*

Proof. Lipschitz continuity with respect to x: For fixed $y \in \mathbb{R}^d$, $v^{(i)}(x) = u^{(i)}(t, x+y)$ is the solution of a switching system obtained from (A.1) by replacing $\mathcal{L}^{i,\alpha}$ with

$$\mathcal{L}^{i,\alpha,y}\varphi(x) := \frac{1}{2}\mathrm{Tr}\,[a_i^\alpha(t, x+y)D^2\varphi] + b_i^\alpha(t, x+y)D\varphi + c_i^\alpha(t, x+y)\varphi + f_i^\alpha(t, x+y),$$

with the terminal condition given by $v^{(i)}(T,x) = g(T, x+y)$. By Theorem A.4, there is a positive constant C such that

$$\sup_{t,x} |u^{(i)}(t,x) - u^{(i)}(t,x+y)| = \sup_{t,x} |u^{(i)}(t,x) - v^{(i)}(t,x)| \le C|y|.$$

$\frac{1}{2}$-Hölder continuity with respect to t: For $t < s$, define $\bar{u} = (\bar{u}^{(i)})_{i=1}^M$ to be the solution to

$$\max\left\{-\bar{u}_t^{(i)} - F_i(\cdot, \bar{u}^{(i)}, D\bar{u}^{(i)}, D^2\bar{u}^{(i)}), \bar{u}^{(i)} - \mathcal{M}^{(i)}\bar{u}\right\} = 0, \text{for } i = 1, \cdots, M;$$
$$\bar{u}^{(i)}(s,\cdot) = u^{(i)}(s,\cdot).$$

Since $\bar{u}$ is a subsolution of (1) on $[0,s] \times \mathbb{R}^d$ with terminal condition $u^{(i)}(s,\cdot)$, by comparison result, Corollary A.5, we have $\bar{u}^{(i)} \le u^{(i)}$. Therefore, $u^{(i)}(t,x) - u^{(i)}(s,x) \ge \bar{u}^{(i)}(t,x) - \bar{u}^{(i)}(s,x)$. By Theorem A.1 of [18], $\bar{u}^{(i)}$ is $\frac{1}{2}$-Hölder continuous in t which provides

$$u^{(i)}(t,x) - u^{(i)}(s,x) \ge -C\sqrt{s-t}.$$

Now, for fixed $y \in \mathbb{R}^d$, define

$$\psi^{(i)}(t,x) := \frac{\lambda L}{2} e^{A(s-t)} \left(|x-y|^2 + B(s-t)\right) + \frac{L}{\lambda} + B(s-t) + g(s,y),$$

where A, B, and λ are positive constants which will be given later and L is the same as in Assumption **HJB-S**. We will show that for an appropriate choice of A and B, $(\psi^{(i)})_{i=1}^M$ is a supersolution of (1) with terminal condition $g(s,x)$. Then, comparison, Corollary A.5, would then imply that $u^{(i)} \le \psi^{(i)}$. Therefore,

$$u^{(i)}(t,y) - u^{(i)}(s,y) \le \psi^{(i)}(t,y) - g(s,y) \le \frac{\lambda L}{2} e^{A(s-t)} B(s-t) + \frac{L}{\lambda} + B(s-t).$$

By setting $\lambda = \frac{1}{\sqrt{s-t}}$, we have $u^{(i)}(t,y) - u^{(i)}(s,y) \le C\sqrt{s-t}$, where C is a positive constant. Therefore, it remains to show that for A and B large enough, we have

$$\min\left\{\max\left\{-\psi_t^{(i)} - \inf_{\alpha\in\mathcal{A}^i} \mathcal{L}^{i,\alpha}\psi^{(i)}, \psi^{(i)} - \mathcal{M}^{(i)}\psi^{(i)}\right\}, \psi^{(i)} - g\right\} \ge 0,$$

on $[0,s] \times \mathbb{R}^d$. Since $\psi^{(i)} - \mathcal{M}^{(i)}\psi^{(i)} < 0$, one needs to show that

$$-\psi_t^{(i)} - \inf_{\alpha\in\mathcal{A}^i} \mathcal{L}^{i,\alpha}\psi^{(i)} \ge 0 \text{and } \psi^{(i)} - g \ge 0.$$

Observe that if $B \ge 1$, by the regularity assumption on g, we have

$$\psi^{(i)}(t,x) - g(t,x) \ge \frac{L}{2}\left(\lambda|x-y|^2 + \lambda(s-t) + \frac{2}{\lambda}\right) + g(s,y) - g(t,x) \ge 0.$$

Conversely,

$$-\psi_t^{(i)} - \inf_{\alpha\in\mathcal{A}^i} \mathcal{L}^{i,\alpha}\psi^{(i)} = \sup_{\alpha\in\mathcal{A}^i}\{\frac{L\lambda}{2} e^{A(s-t)}(A|x-y|^2 + AB(s-t) + B$$
$$-\frac{1}{2}\mathrm{Tr}\,[a^{\alpha,i}] - b^{\alpha,i}\cdot(x-y)) + B - c^{\alpha,i}\psi^{(i)} - f^{\alpha,i}\}$$
$$\ge \frac{L\lambda}{2} e^{A(s-t)}\left(A|x-y|^2 - L|x-y| + LB - \frac{L}{2}\right) + B - CL.$$

By choosing A and B large enough, the right hand side in the above inequality is positive which completes the argument. ∎

APPENDIX B: PROOF OF LEMMA 3.9

If $v^h(t_i,x) = g(t_i,x)$ holds true, then as function g is $\frac{1}{2}$-Hölder continuous on t, the proof is done. So, we assume that $v^h(t_i,x) > g(t_i,x)$. We introduce the discrete stopping time $\hat{\tau} := \min\{t_j | j \geq i, v^h(t_j,\hat{X}_{t_j}) = g(t_j,\hat{X}_{t_j})\}$. Observe that $t_i < \hat{\tau}$.

Step 1. Let $\overline{\theta}$ be such that $F(t_j,\hat{X}^x_{t_j},\mathcal{D}_h v^h(t_j,X^x_{t_j})) - F(t_j,\hat{X}^x_{t_j},0,0,0) = \nabla F(t_j,\hat{X}^x_{t_j},\overline{\theta}) \cdot \mathcal{D}_h v^h(t_j,X^x_{t_j}))$ For all $j = i,\cdots,n-1$, on the event $\{t_j < \hat{\tau}\}$ one can write

$$v^h(t_j,x) = \mathbb{E}_{t_j,x}[v^h(t_{j+1},\hat{X}^x_{t_{j+1}})P_{j+1}] + hF_j,$$

where $F_j = F(t_j,\hat{X}^x_{t_j},0,0,0)$, $\Delta W_{j+1} = W_{t_{j+1}} - W_{t_j}$, $P_{j+1} = 1-\alpha_j + \sqrt{h}\beta_j \cdot \Delta W_{j+1} + h^{-1}\alpha_j \cdot \Delta W_{j+1}\Delta W^T_{j+1}$, and $\alpha_j := F_\gamma \cdot a^{-1}(t_j,\hat{X}^x_{t_j},\overline{\theta}) < 1$ and $\beta_j := F_p \cdot \sigma^{-1}(t_j,\hat{X}^x_{t_j},\overline{\theta})$ are $\mathcal{F}_{t_{j+1}}$-measurable. We can rewrite the above equality in the following form.

$$v^h(t_j,x)\mathbb{1}_{\{t_j<\hat{\tau}\}} = \mathbb{E}_{t_j,x}[v^h(t_{j+1},\hat{X}^x_{t_{j+1}})P_{j+1}\mathbb{1}_{\{t_{j+1}<\hat{\tau}\}} + g(\hat{\tau},\hat{X}^x_{\hat{\tau}})P_{j+1}\mathbb{1}_{\{t_{j+1}=\hat{\tau}\}}] + h\mathbb{1}_{\{t_j<\hat{\tau}\}}F_j \tag{B.1}$$

Notice that the first term in the right hand side of (1) is zero if $\{t_j \geq \hat{\tau}\}$. We define $Q_j := \prod_{k=i+1}^{j} P_k$ with $Q_i := 1$ and $V_j := v^h(t_j,\hat{X}^x_{t_j})Q_j\mathbb{1}_{\{t_j<\hat{\tau}\}}$. Observe that Q_j is a discrete martingale with respect to $\{W_{t_j}\}^n_{j=i}$. Multiplying (B.1) by Q_j, one can write

$$V_j = \mathbb{E}_{t_j,x}[V_{j+1} + g(\hat{\tau},\hat{X}^x_{\hat{\tau}})Q_{j+1}\mathbb{1}_{\{t_{j+1}=\hat{\tau}\}}] + h\mathbb{1}_{\{t_j<\hat{\tau}\}}Q_jF_j.$$

By summing the above equality over $j = i,\cdots,n-1$ and taking expectation $\mathbb{E}_{t_i,x}$, we have

$$\begin{aligned} v^h(t_i,x) = V_i &= \mathbb{E}_{t_i,x}\left[V_n + \sum_{j=i}^{n-1} g(\hat{\tau},\hat{X}^x_{\hat{\tau}})Q_{j+1}\mathbb{1}_{\{t_{j+1}=\hat{\tau}\}} + h\sum_{j=i}^{n-1}\mathbb{1}_{\{t_j<\hat{\tau}\}}Q_jF_j\right] \\ &= \mathbb{E}_{t_i,x}\left[\sum_{j=i}^{n-1} g(\hat{\tau},\hat{X}^x_{\hat{\tau}})Q_{j+1}\mathbb{1}_{\{t_{j+1}=\hat{\tau}\}} + h\sum_{j=i}^{n-1}\mathbb{1}_{\{t_j<\hat{\tau}\}}Q_jF_j\right]. \end{aligned}$$

Observe that here we used $V_n = 0$ by the definition of $\hat{\tau}$. Thus, we can write

$$v^h(t_i,x) - g(t_i,x) = \mathbb{E}_{t_i,x}\left[(g(\hat{\tau},\hat{X}^x_{\hat{\tau}}) - g(t_i,x))\sum_{j=i}^{n-1} Q_{j+1}\mathbb{1}_{\{t_{j+1}=\hat{\tau}\}} + h\sum_{j=i}^{n-1}\mathbb{1}_{\{t_j<\hat{\tau}\}}Q_jF_j\right], \tag{B.2}$$

where in the above we used optional stopping theorem for $\mathbb{E}_{t_i,x}[\sum_{j=i}^{n-1} Q_{j+1}\mathbb{1}_{\{t_{j+1}=\hat{\tau}\}}] = 1$. Our goal is to show that the right-hand side of (B.2) is bounded by $C\sqrt{T-t_i}$. First observe that by Assumption **F**(i), F_j is bounded. Then, because Q_j is a positive martingale, we bounded the second term in (2) by $C(T-t_i)$:

$$\left|\mathbb{E}_{t_i,x}\left[h\sum_{j=i}^{n-1}\mathbb{1}_{\{t_j<\hat{\tau}\}}Q_jF_j\right]\right|\leq Ch\sum_{j=i}^{n-1}\mathbb{E}_{t_i,x}[Q_j]\leq C(T-t_i).$$

We continue by bounding the other term in (B.2) in the next step.

Step 2. To bound $\mathbb{E}_{t_i,x}\left[(g(\hat{\tau},\hat{X}^x_{\hat{\tau}})-g(t_i,x))\sum_{j=i}^{n-1}Q_{j+1}\mathbb{1}_{\{t_{j+1}=\hat{\tau}\}}\right]$, we want to apply Itô formula on $g(\hat{\tau},\hat{X}^x_{\hat{\tau}})-g(t_i,x)$. But, because g is not a C^2 function, we first approximate g by a smooth function uniformly, that is, $|g-g_\varepsilon|_\infty\leq C\varepsilon$. This can be done by $g_\varepsilon:=g*\rho_\varepsilon$ where $\{\rho_\varepsilon\}_\varepsilon$ is a family of mollifiers. Because g is Lipschitz on x and $\frac{1}{2}$-Hölder on t, we have

$$|\partial_t g_\varepsilon|_\infty\leq\varepsilon^{-1},\ |Dg_\varepsilon|_\infty\leq C,\text{ and }|D^2g_\varepsilon|_\infty\leq\varepsilon^{-1}. \tag{B.3}$$

Therefore, we write

$$\begin{aligned}g(\hat{\tau},\hat{X}^x_{\hat{\tau}})-g(t_i,x)&=(g_\varepsilon(\hat{\tau},\hat{X}^x_{\hat{\tau}})-g_\varepsilon(t_i,x))\\&\quad+(g(\hat{\tau},\hat{X}^x_{\hat{\tau}})-g_\varepsilon(\hat{\tau},\hat{X}^x_{\hat{\tau}}))+(g_\varepsilon(t_i,x)-g(t_i,x)).\end{aligned}$$

Observe that since $|g_\varepsilon-g|_\infty\leq C\varepsilon$, one has

$$\begin{aligned}\left|\mathbb{E}_{t_i,x}\left[(g_\varepsilon(\hat{\tau},\hat{X}^x_{\hat{\tau}})-g(\hat{\tau},\hat{X}^x_{\hat{\tau}}))\sum_{j=i}^{n-1}Q_{j+1}\mathbb{1}_{\{t_{j+1}=\hat{\tau}\}}\right]\right|&\leq C\varepsilon,\\\left|\mathbb{E}_{t_i,x}\left[(g_\varepsilon(t_i,x)-g(t_i,x))\sum_{j=i}^{n-1}Q_{j+1}\mathbb{1}_{\{t_{j+1}=\hat{\tau}\}}\right]\right|&\leq C\varepsilon.\end{aligned} \tag{B.4}$$

In the following steps, we find a bound on $\mathbb{E}_{t_i,x}\left[(g_\varepsilon(\hat{\tau},\hat{X}^x_{\hat{\tau}})-g_\varepsilon(t_i,x))\sum_{j=i}^{n-1}Q_{j+1}\mathbb{1}_{\{t_{j+1}=\hat{\tau}\}}\right]$ in terms of $T-t_i$ and ϵ.

Step 3. We apply Itô formula on $g_\varepsilon(\hat{\tau},\hat{X}^x_{\hat{\tau}})-g_\varepsilon(t_i,x)$:

$$g_\varepsilon(\hat{\tau},\hat{X}^x_{\hat{\tau}})-g_\varepsilon(t_i,x)=\int_{t_i}^{\hat{\tau}}\mathcal{L}^{\hat{X}}g_\varepsilon(s,\hat{X}^x_s)ds+\int_{t_i}^{\hat{\tau}}Dg_\varepsilon(s,\hat{X}^x_s)\cdot dW_s,$$

where $\mathcal{L}^{\hat{X}}$ is the infinitesimal generator for the processe $\hat{X}$. Thus,

$$\begin{aligned}&\mathbb{E}_{t_i,x}\left[(g_\varepsilon(\hat{\tau},\hat{X}^x_{\hat{\tau}})-g_\varepsilon(t_i,x))\sum_{j=i}^{n-1}Q_{j+1}\mathbb{1}_{\{t_{j+1}=\hat{\tau}\}}\right]\\&=\sum_{j=i}^{n-1}\mathbb{E}_{t_i,x}\left[\left(\int_{t_i}^{\hat{\tau}}\mathcal{L}^{\hat{X}}g_\varepsilon(s,\hat{X}^x_s)ds+\int_{t_i}^{\hat{\tau}}Dg_\varepsilon(s,\hat{X}^x_s)\cdot dW_s\right)Q_{j+1}\mathbb{1}_{\{t_{j+1}=\hat{\tau}\}}\right].\end{aligned} \tag{B.5}$$

We proceed by calculating the term in the above summation for $j = n-1$.

$$
\begin{aligned}
&\mathbb{E}_{t_i,x}\left[\left(\int_{t_i}^{\hat{\tau}} \mathcal{L}^{\hat{X}} g_\varepsilon(s,\hat{X}_s^x)ds + \int_{t_i}^{\hat{\tau}} Dg_\varepsilon(s,\hat{X}_s^x)\cdot dW_s\right) Q_n \mathbb{1}_{\{t_n=\hat{\tau}\}}\right] \\
&= \mathbb{E}_{t_i,x}\left[\left(\int_{t_i}^{t_n} \mathcal{L}^{\hat{X}} g_\varepsilon(s,\hat{X}_s^x)ds + \int_{t_i}^{t_n} Dg_\varepsilon(s,\hat{X}_s^x)\cdot dW_s\right) Q_n \mathbb{1}_{\{t_n=\hat{\tau}\}}\right] \\
&= \mathbb{E}_{t_i,x}\left[\left(\int_{t_i}^{t_{n-1}} \mathcal{L}^{\hat{X}} g_\varepsilon(s,\hat{X}_s^x)ds + \int_{t_i}^{t_{n-1}} Dg_\varepsilon(s,\hat{X}_s^x)\cdot dW_s\right) Q_n \mathbb{1}_{\{t_n=\hat{\tau}\}}\right] \\
&+\mathbb{E}_{t_i,x}\left[\left(\int_{t_{n-1}}^{t_n} \mathcal{L}^{\hat{X}} g_\varepsilon(s,\hat{X}_s^x)ds + \int_{t_{n-1}}^{t_n} Dg_\varepsilon(s,\hat{X}_s^x)\cdot dW_s\right) Q_n \mathbb{1}_{\{t_n=\hat{\tau}\}}\right] \qquad \text{(B.6)}
\end{aligned}
$$

We first bound the second term in the right-hand side in the next step.

Step 4. Since $Q_{t_{n-1}}$ and $\mathbb{1}_{\{t_n=\hat{\tau}\}}$ are $\mathcal{F}_{t_{n-1}}$ measurable, the second term in the right-hand side can be written as

$$
\begin{aligned}
&\mathbb{E}_{t_i,x}\left[\left(\int_{t_{n-1}}^{t_n} \mathcal{L}^{\hat{X}} g_\varepsilon(s,\hat{X}_s^x)ds + \int_{t_{n-1}}^{t_n} Dg_\varepsilon(s,\hat{X}_s^x)\cdot dW_s\right) Q_n \mathbb{1}_{\{t_n=\hat{\tau}\}}\right] \\
&= \mathbb{E}_{t_i,x}\left[Q_{n-1}\mathbb{1}_{\{t_n=\hat{\tau}\}}\mathbb{E}_{t_{n-1}}\left[\left(\int_{t_{n-1}}^{t_n} \mathcal{L}^{\hat{X}} g_\varepsilon(s,\hat{X}_s^x)ds + \int_{t_{n-1}}^{t_n} Dg_\varepsilon(s,\hat{X}_s^x)\cdot dW_s\right) P_n\right]\right],
\end{aligned}
$$

where $\mathbb{E}_{t_j}[\cdot] = \mathbb{E}[\cdot|\mathcal{F}_{t_j}]$. Notice that we can write $P_n = 1 + h^{\frac{1}{2}}\beta_{n-1}\cdot\int_{t_{n-1}}^{t_n} dW_s + h^{-1}\alpha_{n-1}\cdot\int_{t_{n-1}}^{t_n}(W_s - W_{t_{n-1}})dW_s^{\mathrm{T}}$. Thus, one can calculate $\mathbb{E}_{t_{n-1}}\left[\left(\int_{t_{n-1}}^{t_n}\mathcal{L}^{\hat{X}} g_\varepsilon(s,\hat{X}_s^x)ds + \int_{t_{n-1}}^{t_n} Dg_\varepsilon(s,\hat{X}_s^x)\cdot dW_s\right)P_n\right]$ using It$\hat{o}$ isometry and the fact that the expected value of stochastic integrals is zero:

$$
\begin{aligned}
&\mathbb{E}_{t_{n-1}}\left[\left(\int_{t_{n-1}}^{t_n} \mathcal{L}^{\hat{X}} g_\varepsilon(s,\hat{X}_s^x)ds + \int_{t_{n-1}}^{t_n} Dg_\varepsilon(s,\hat{X}_s^x)\cdot dW_s\right) P_n\right] \\
&= \mathbb{E}_{t_{n-1}}\left[\int_{t_{n-1}}^{t_n} \mathcal{L}^{\hat{X}} g_\varepsilon(s,\hat{X}_s^x)ds + h^{\frac{1}{2}}\beta_{n-1}\cdot\int_{t_{n-1}}^{t_n} Dg_\varepsilon(s,\hat{X}_s^x)ds\right. \\
&\qquad\left. + h^{-1}\int_{t_{n-1}}^{t_n}\alpha_{n-1}Dg_\varepsilon(s,\hat{X}_s^x)\cdot W_s ds\right]
\end{aligned}
$$

Because of (B.4), the first two term in the above are bounded by $C(h+\frac{h}{\varepsilon})$. The third term can be calculated by using (2.6)

$$
\mathbb{E}_{t_{n-1}}\left[\int_{t_{n-1}}^{t_n}\alpha_{n-1}Dg_\varepsilon(s,\hat{X}_s^x)\cdot W_s ds\right] = \int_{t_{n-1}}^{t_n}\alpha_{n-1}\mathbb{E}_{t_{n-1}}[Dg_\varepsilon(s,\hat{X}_s^x)\cdot W_s]ds
$$

By (2.7), we have $\mathbb{E}_{t_{n-1}}[Dg_\varepsilon(s,\hat{X}_s^x)W_s] = s\mathbb{E}_{t_{n-1}}[D^2 g_\varepsilon(s,\hat{X}_s^x)]$ which is bounded by $C\frac{s}{\varepsilon}$. Thus,

$$\left|\mathbb{E}_{t_{n-1}}\left[\int_{t_{n-1}}^{t_n} \alpha_{n-1} Dg_\varepsilon(s,\hat{X}_s^x)\cdot W_s ds\right]\right| = \frac{C}{\varepsilon}\int_{t_{n-1}}^{t_n} s ds = \frac{Ch^2}{2\varepsilon}.$$

Therefore,

$$\mathbb{E}_{t_{n-1}}\left[\left(\int_{t_{n-1}}^{t_n} \mathcal{L}^{\hat{X}} g_\varepsilon(s,\hat{X}_s^x)ds + \int_{t_{n-1}}^{t_n} Dg_\varepsilon(s,\hat{X}_s^x)\cdot dW_s\right) P_n\right] \le Ch(1+\varepsilon^{-1}).$$

Step 5. Because $\{Q_j\}_{j=i}^n$ is a martingale and $\mathbb{1}_{\{t_n=\hat{\tau}\}}$ is $\mathcal{F}_{t_{n-1}}$-measurable, one can write the first term in the right-hand side of (B.6) as

$$\begin{aligned}
&\mathbb{E}_{t_i,x}\left[\left(\int_{t_i}^{t_{n-1}} \mathcal{L}^{\hat{X}} g_\varepsilon(s,\hat{X}_s^x)ds + \int_{t_i}^{t_{n-1}} Dg_\varepsilon(s,\hat{X}_s^x)\cdot dW_s\right) Q_n \mathbb{1}_{\{t_n=\hat{\tau}\}}\right]\\
&= \mathbb{E}_{t_i,x}\left[\left(\int_{t_i}^{t_{n-1}} \mathcal{L}^{\hat{X}} g_\varepsilon(s,\hat{X}_s^x)ds + \int_{t_i}^{t_{n-1}} Dg_\varepsilon(s,\hat{X}_s^x)\cdot dW_s\right) \mathbb{1}_{\{t_n=\hat{\tau}\}}\mathbb{E}_{t_{n-1}}[Q_n]\right]\\
&= \mathbb{E}_{t_i,x}\left[\left(\int_{t_i}^{t_{n-1}} \mathcal{L}^{\hat{X}} g_\varepsilon(s,\hat{X}_s^x)ds + \int_{t_i}^{t_{n-1}} Dg_\varepsilon(s,\hat{X}_s^x)\cdot dW_s\right) Q_{n-1} \mathbb{1}_{\{t_n=\hat{\tau}\}}\right],
\end{aligned}$$

Thus, from (B.5) we have

$$\begin{aligned}
&\left|\mathbb{E}_{t_i,x}\left[(g_\varepsilon(\hat{\tau},\hat{X}_{\hat{\tau}}^x) - g_\varepsilon(t_i,x))\sum_{j=i}^{n-1} Q_{j+1}\mathbb{1}_{\{t_{j+1}=\hat{\tau}\}}\right]\right| \le Ch(1+\frac{1}{\varepsilon})\\
&+\left|\sum_{j=i}^{n-2}\mathbb{E}_{t_i,x}\left[\left(\int_{t_i}^{\hat{\tau}} \mathcal{L}^{\hat{X}} g_\varepsilon(s,\hat{X}_s^x)ds + \int_{t_i}^{\hat{\tau}} Dg_\varepsilon(s,\hat{X}_s^x)\cdot dW_s\right) Q_{j+1}\mathbb{1}_{\{t_{j+1}=\hat{\tau}\}}\right]\right.\\
&\left.+\mathbb{E}_{t_i,x}\left[\left(\int_{t_i}^{t_{n-1}} \mathcal{L}^{\hat{X}} g_\varepsilon(s,\hat{X}_s^x)ds + \int_{t_i}^{t_{n-1}} Dg_\varepsilon(s,\hat{X}_s^x)\cdot dW_s\right) Q_{n-1}\mathbb{1}_{\{t_n=\hat{\tau}\}}\right]\right|\\
&= Ch(1+\frac{1}{\varepsilon}) + \left|\sum_{j=i}^{n-3}\mathbb{E}_{t_i,x}\left[\left(\int_{t_i}^{\hat{\tau}} \mathcal{L}^{\hat{X}} g_\varepsilon(s,\hat{X}_s^x)ds + \int_{t_i}^{\hat{\tau}} Dg_\varepsilon(s,\hat{X}_s^x)\cdot dW_s\right) Q_{j+1}\mathbb{1}_{\{t_{j+1}=\hat{\tau}\}}\right]\right.\\
&\left.+\mathbb{E}_{t_i,x}\left[\left(\int_{t_i}^{t_{n-1}} \mathcal{L}^{\hat{X}} g_\varepsilon(s,\hat{X}_s^x)ds + \int_{t_i}^{t_{n-1}} Dg_\varepsilon(s,\hat{X}_s^x)\cdot dW_s\right) Q_{n-1}\mathbb{1}_{\{t_{n-1}\le\hat{\tau}\}}\right]\right|.
\end{aligned}$$

By repeating the argument in Step 3 and Step 4 inductively over $k = n-1, \cdots, i+1$, one can write the above as

$$\left|\mathbb{E}_{t_i,x}\left[(g_\varepsilon(\hat{\tau},\hat{X}^x_{\hat{\tau}})-g_\varepsilon(t_i,x))\sum_{j=i}^{n-1}Q_{j+1}\mathbb{1}_{\{t_{j+1}=\hat{\tau}\}}\right]\right|\le C(n-k)h(1+\frac{1}{\varepsilon})$$

$$+\left|\sum_{j=i}^{k-2}\mathbb{E}_{t_i,x}\left[\left(\int_{t_i}^{\hat{\tau}}\mathcal{L}^{\hat{X}}g_\varepsilon(s,\hat{X}^x_s)ds+\int_{t_i}^{\hat{\tau}}Dg_\varepsilon(s,\hat{X}^x_s)\cdot dW_s\right)Q_{j+1}\mathbb{1}_{\{t_{j+1}=\hat{\tau}\}}\right]\right.$$

$$\left.+\mathbb{E}_{t_i,x}\left[\left(\int_{t_i}^{t_k}\mathcal{L}^{\hat{X}}g_\varepsilon(s,\hat{X}^x_s)ds+\int_{t_i}^{t_k}Dg_\varepsilon(s,\hat{X}^x_s)\cdot dW_s\right)Q_k\mathbb{1}_{\{t_k\le\hat{\tau}\}}\right]\right|.$$

Specially for $k=i+1$ (the term containing $\sum_{j=i}^{k-2}$ disappears), we have

$$\left|\mathbb{E}_{t_i,x}\left[(g_\varepsilon(\hat{\tau},\hat{X}^x_{\hat{\tau}})-g_\varepsilon(t_i,x))\sum_{j=i}^{n-1}Q_{j+1}\mathbb{1}_{\{t_{j+1}=\hat{\tau}\}}\right]\right|\le C(n-i-1)h(1+\frac{1}{\varepsilon})$$

$$+\left|\mathbb{E}_{t_i,x}\left[\left(\int_{t_i}^{t_{i+1}}\mathcal{L}^{\hat{X}}g_\varepsilon(s,\hat{X}^x_s)ds+\int_{t_i}^{t_{i+1}}Dg_\varepsilon(s,\hat{X}^x_s)\cdot dW_s\right)Q_{i+1}\mathbb{1}_{\{t_{i+1}\le\hat{\tau}\}}\right]\right|$$

$$\le C(n-i)h(1+\frac{1}{\varepsilon})=C(T-t_i)(1+\varepsilon^{-1}).$$

Step 6. By using (B.4) and the bound found in Step 5 in (B.2), one has

$$|v^h(t_i,x)-g(t_i,x)|\le C(\varepsilon+\frac{T-t_i}{\varepsilon}+T-t_i).$$

By choosing $\varepsilon=\sqrt{T-t_i}$, we conclude that

$$|v^h(t_i,x)-g(t_i,x)|\le C\sqrt{T-t_i}.$$

Then, the result follows from x-Lipschitz continuity and t-$\frac{1}{2}$-Hölder continuity of g.

The authors are grateful to Xavier Warin and anonymous referees for their helpful comments and suggestions.

References

1. B. Bouchard and N. Touzi, Discrete-time approximation and Monte-Carlo simulation of backward stochastic differential equations, Stoch Process Appl 111 (2004), 175–206.
2. J. Zhang, A numerical scheme for BSDEs, Ann Appl Probab 14 (2004), 459–488.
3. B. Bouchard and J.-F. Chassagneux, Discrete-time approximation for continuously and discretely reflected BSDEs, Stoch Process Appl 118 (2008), 2269–2293.
4. J. Ma and J. Zhang, Representations and regularities for solutions to BSDEs with reflections, Stoch Process Appl 115 (2005), 539–569.
5. A. Fahim, N. Touzi, and X. Warin, A probabilistic numerical method for fully non-linear parabolic pdes, Ann Appl Probab 21 (2011), 1322–1364.
6. W. Guo, J. Zhang, and J. Zhuo, A monotone scheme for high dimensional fully nonlinear pdes, arXiv preprint arXiv:1212.0466.

7. I. Karatzas and S. G. Kou, Hedging American contingent claims with constrained portfolios, Finance Stoch 2(1998), 215–258.
8. E. Bayraktar and Y.-J. Huang, On the multidimensional controller-and-stopper games, SIAM J Control Optim 51 (2013), 1263–1297.
9. F. A. Longstaff and E. S. Schwartz, Valuing American options by simulation: a simple least-squares approach, Rev Financ Stud 14 (2001), 113–147.
10. B. Bouchard and X. Warin, Monte-Carlo valorisation of American options: facts and new algorithms to improve existing methods, R. Carmona, P. Del Moral, P. Hu, and N. Oudjane, editors, Numerical methods in finance, Springer Proceedings in Mathematics, Berlin Heidelberg, 2012, pp. 215–255.
11. G. Barles and P. E. Souganidis, Convergence of approximation schemes for fully nonlinear second order equations, Asymptotic Anal 4 (1991), 271–283.
12. J. F. Bonnans, S. Maroso, and H. Zidani, Error estimates for stochastic differential games: the adverse stopping case, IMA J Numer Anal 26 (2006), 188–212.
13. L. A. Caffarelli and P. E. Souganidis, A rate of convergence for monotone finite difference approximations to fully nonlinear, uniformly elliptic pdes, Commun Pure Appl Math 61 (2008), 1–17.
14. E. R. Jakobsen, On the rate of convergence of approximation schemes for Bellman equations associated with optimal stopping time problems, Math Models Methods Appl Sci 13 (2003), 613–644.
15. A. M. Oberman and T. Zariphopoulou, Pricing early exercise contracts in incomplete markets, Comput Manage Sci 1 (2003), 75–107.
16. A. M. Oberman, Convergent difference schemes for degenerate elliptic and parabolic equations: Hamilton-Jacobi equations and free boundary problems, SIAM J Numer Anal 44 (2006), 879–895 (electronic).
17. A. Fahim, Convergence of a Monte Carlo method for fully non-linear elliptic and parabolic PDEs in some general domains. Preprint available at http://www.math.fsu.edu/~fahim/Bounded_Domain_v0.pdf.
18. G. Barles and E. R. Jakobsen, Error bounds for monotone approximation schemes for parabolic Hamilton-Jacobi-Bellman equations. Math Comput 76 (2007), 1861–1893 (electronic).
19. N. V. Krylov, On the rate of convergence of finite-difference approximations for Bellman's equations, Algebra i Analiz 9 (1997), 245–256.
20. N. V. Krylov, Approximating value functions for controlled degenerate diffusion processes by using piece-wise constant policies, Electron J Probab 4 (1999), 1–19 (electronic).
21. N. V. Krylov, On the rate of convergence of finite-difference approximations for Bellman's equations with variable coefficients, Probab Theory Relat Fields 117 (2000), 1–16.
22. N. V. Krylov, The rate of convergence of finite-difference approximations for Bellman equations with Lipschitz coefficients, Appl Math Optim 52 (2005), 365–399.
23. G. Barles and E. R. Jakobsen, On the convergence rate of approximation schemes for Hamilton-Jacobi-Bellman equations. M2AN Math Model Numer Anal 36 (2002), 33–54.
24. G. Barles and E. R. Jakobsen, Error bounds for monotone approximation schemes for Hamilton-Jacobi-Bellman equations. SIAM J Numer Anal 43 (2005), 540–558 (electronic).
25. M. G. Crandall, H. Ishii, and P.-L. Lions, User's guide to viscosity solutions of second order partial differential equations, Bull Am Math Soc (N.S.) 27 (1992), 1–67.