Approximation Schemes for Mixed Optimal Stopping and Control Problems with Nonlinear Expectations and Jumps

Roxana Dumitrescu[1] · Christoph Reisinger[2] · Yufei Zhang[2]

Abstract

We propose a class of numerical schemes for mixed optimal stopping and control of processes with infinite activity jumps and where the objective is evaluated by a nonlinear expectation. Exploiting an approximation by switching systems, piecewise constant policy timestepping reduces the problem to nonlocal semi-linear equations with different control parameters, uncoupled over individual time steps, which we solve by fully implicit monotone approximations to the controlled diffusion and the nonlocal term, and specifically the Lax–Friedrichs scheme for the nonlinearity in the gradient. We establish a comparison principle for the switching system and demonstrate the convergence of the schemes, which subsequently gives a constructive proof for the existence of a solution to the switching system. Numerical experiments are presented for a recursive utility maximization problem to demonstrate the effectiveness of the new schemes.

Keywords Approximation schemes · Optimal stopping · Stochastic control · Nonlinear expectations · Piecewise constant policy timestepping · Jump processes

Mathematics Subject Classification 65M06 · 65M12 · 62L15 · 93E20 · 91G80

✉ Yufei Zhang
yufei.zhang@maths.ox.ac.uk

Roxana Dumitrescu
roxana.dumitrescu@kcl.ac.uk

Christoph Reisinger
christoph.reisinger@maths.ox.ac.uk

[1] Department of Mathematics, King's College London, London, UK

[2] Mathematical Institute, University of Oxford, Oxford, UK

1 Introduction

Classical Markovian mixed optimal stopping and control problems, where the target is to maximise the (linear) expectation of a payoff on a finite time horizon T, are defined as

$$u(t, x) = \sup_{\tau} \sup_{\alpha} \mathbb{E}^{t,x} \left[\int_t^\tau e^{-r(s-t)} f(\alpha_s, X_s^{\alpha,t,x}) \, ds + e^{-r(\tau-t)} \xi(\tau, X_\tau^{\alpha,t,x}) \right],$$

(1.1)

where $t \in [0, T]$ is the initial time of the control problem, α is an admissible control process and τ is a stopping time, and $X^{\alpha,t,x}$ is a controlled stochastic differential equation (SDE) of the form:

$$dX_s^{\alpha,t,x} = b(\alpha_s, X_s^{\alpha,t,x}) \, ds + \sigma(\alpha_s, X_s^{\alpha,t,x}) \, dW_s$$
$$+ \eta(\alpha_s, X_s^{\alpha,t,x}, e) \, \tilde{N}(ds, de), \ s \in [t, T]; \quad X_t^{\alpha,t,x} = x.$$

The positive constant r denotes the discount rate, and the functions ξ and f represent the terminal payoff and the instantaneous reward function, respectively. Under certain regularity assumptions on the coefficients, one can demonstrate that the value function u satisfies a nonlocal Hamilton–Jacobi–Bellman variational inequality (HJBVI) in the viscosity sense.

These results are extended in Reference [17] to a setting where the linear expectation $\mathbb{E}$ is replaced by a nonlinear expectation $\mathcal{E}^{\alpha,t,x}$ generated by a BSDE with jumps,

$$u(t, x) = \sup_{\tau} \sup_{\alpha} \mathcal{E}_{t,\tau}^{\alpha,t,x} [\xi(\tau, X_\tau^{\alpha,t,x})].$$

(1.2)

It has been demonstrated in Reference [17] that under suitable assumptions the value function u in (1.2) (after a change of time variable) can be characterized by the viscosity solution to a more complicated HJBVI (1.3), which involves an extra nonlinearity resulting from the nonlinear expectation:

$$\min \left\{ u(\mathbf{x}) - \zeta(\mathbf{x}), u_t + \inf_{\alpha \in \mathbf{A}} \left(-L^\alpha u - f(\alpha, \mathbf{x}, u, (\sigma^\alpha)^T Du, B^\alpha u) \right) \right\} = 0,$$
$$\mathbf{x} \in [0, T] \times \mathbb{R}^d; u(0, x) = g(x), \quad x \in \mathbb{R}^d,$$

(1.3)

with $\mathbf{x} = (t, x)$, nonlocal operators L^α and B^α, the driver f of the BSDE, and given functions ζ and g, which we will specify in Sect. 2. Particularly, in the case where the driver is additive in y and independent of z and k, i.e., $f(\alpha, \mathbf{x}, y, z, k) \equiv f(\alpha, x) - ry$, the generalized control problem (1.2) reduces to the classical linear expectation case (1.1), and (1.3) reduces to an HJB obstacle problem.[1]

Nonlinear expectations as in (1.2), and hence HJBVIs as in (1.3), arise naturally in financial mathematics, for instance as models for American options in a market with constrained portfolios [23], from recursive utility optimization problems [9], from

---

[1] Note a slight abuse of notation where $f(\cdot, \cdot)$ is the same as in (1.1).

dynamic risk measures [30], and from robust pricing and risk measures under probability model uncertainty [31]. In Sect. 5, we describe in more detail the situation of an ambiguity-averse investor who chooses the optimal wealth allocation and liquidation time of an asset whose price process has infinite activity jumps. This problem can be modelled naturally within the setup described above, and therefore leads directly to an HJBVI of the type (1.3) with all its key features present (i.e., singular non-local term, nonlinear driver, control optimisation, obstacle term). As it is usually difficult to obtain analytic solutions of HJBVIs, it is necessary to design efficient and robust numerical methods for solving these fully nonlinear PIDEs, of which we give a worked-out example in Sect. 5.

We also remark that, to the best of our knowledge, even for the case with linear expectations (i.e. f in the special form from above), there is no published numerical scheme covering the generality of (1.3). However, there is a vast literature on monotone approximations for local HJB equations (see, e.g., [2,11,13] and references therein) and a number of works covering specific extensions. For instance, monotone finite-difference quadrature schemes are proposed in References [5–7,10] for nonlocal HJB equations. We refer the reader also to Reference [14] for penalty approximations to nonlocal variational inequalities and to Reference [33] for an application of policy iteration together with penalization to solve HJB obstacle problems. Probabilistic methods for solving HJB equations (without jumps and optimal stopping) can be found, for example, in Reference [26].

All the aforementioned PDE methods solve (1.3) (with $f(\alpha, \mathbf{x}, y, z, k) \equiv f(\alpha, x) - ry$) by the standard "discretize, then optimize" approach, where one discretizes the operators in (1.3), and solves the resulting nonlinear discretized equations using policy iteration, or more generally semi-smooth Newton methods [21,28].

However, this standard approach cannot be easily extended to nonlinear f which is only assumed to be Lipschitz (and generally is not semi-smooth [28]), which prevents a direct application of Newton-like solvers (see Reference [34] for a special case of the discrete optimization problem with $f(\alpha, \mathbf{x}, y, z, k) \equiv f(\alpha, x, y)$ differentiable and concave in y, and $\mathbf{A}$ finite).

Moreover, at each step of policy iteration, one needs to identify the global optimal policy for each computational node. The nonlinear driver f and other PDE coefficients may have sufficiently complicate nonlinearities in the control variable such that the only way to construct a convergent algorithm is to discretize the admissible control set, and perform exhaustive search to determine the optimal policy at each node.

Another approach to solve (1.3) uses piecewise constant policy time stepping (PCPT) as in Reference [29]. It is based explicitly on a discrete approximation of the admissible set by a finite set, say with J elements, and then defines a piecewise decoupled system of PDEs corresponding to these J (constant) controls. The information from the different solutions is assembled at the end of each timestep by taking the pointwise maximum.

As a specific scheme for these semi-linear PDEs, we propose an implicit Euler time discretization, monotone (semi-Lagrangian) approximations for local diffusions, a monotone quadrature-based scheme for the nonlocal terms, and the Lax-Friedrichs scheme for the nonlinearity in the gradient. The different solutions may be defined on different discretization grids by possibly high order monotonicity preserving interpo-

lations. This approach not only avoids policy iteration, but also allows for an easier construction of convergent monotone schemes and an efficient parallel implementation of the individual semi-linear PDEs. Note that it is essential to obtain a monotone discretization, since it is well-known that non-monotone schemes may fail to converge or even converge to false "solutions" [13]. By Godunov's Theorem [20], in general, one can expect a monotone scheme to be at most first-order accurate.

The main contributions of our paper are:

- We formulate our algorithm by approximating the solution of (1.3) by the solution to a switching system with small switching cost. We shall establish a comparison principle for the switching system and demonstrate that as the switching cost tends to zero, the solution of the switching system converges to the viscosity solution of (1.3), which extends the results in Reference [6] to obstacle problems of switching systems and includes nonlinear drivers.
- We discretize the switching system piecewise in time by fully implicit monotone approximations. The convergence of the scheme is demonstrated, which subsequently gives a constructive proof for the existence of a viscosity solution to the switching system. Our results extend the one obtained in Reference [29] from the case of standard control problems. In contrast to there, PCPT leads to coupled semi-linear PDEs rather than linear PDEs due to the nonlinear expectations. The optimal stopping is treated as an additional control and included in the switching directly instead of the classical penalisation approach.
- By truncation of the singular jump measure and discretization of the control set, we obtain a stochastic control and optimal stopping problem whose value function is shown to converge to the value function of the initial problem and which satisfies a HJBVI equation. Our result extends earlier ones obtained only in the case of a linear expectation without control and optimal stopping (see e.g. [10]). This control-theoretical interpretation further enables us to establish the convergence order of these approximations through a probabilistic argument.
- For practical implementations, we propose a Picard-type iteration for the efficient numerical solution without the need to invert the dense matrices resulting from the nonlocal terms.
- Numerical examples for a recursive utility maximization problem are included to investigate the convergence order of the scheme with respect to different discretization parameters.

The remainder of this paper is organized as follows. In Sect. 2, we introduce the Markovian mixed optimal stopping and control problem with nonlinear expectations, and characterize its value function as the viscosity solution of a nonlocal HJBVI. We then derive numerical schemes in Sect. 3 by approximating the HJBVI with a switching system, PCPT, and ultimately fully discrete monotone schemes. Then we move on to the convergence analysis of our numerical schemes in Sect. 4. Numerical examples for a recursive utility maximization problem are presented in Sect. 5 to illustrate the effectiveness of our algorithms. In the Appendix, we include a rigorous proof of the comparison principle for the switching system and some complementary results that are used in this article.

2 Problem Formulation and Preliminaries

In this section, we formulate the mixed optimal stopping and control problem with nonlinear expectation and introduce the connection between such problems and HJB-VIs, which is crucial for the subsequent developments. We start with some useful notation that is needed frequently in the rest of this work.

We write by $T > 0$ the terminal time, and by $(\Omega, \mathcal{F}, P)$ a complete probability space, in which two mutually independent processes, a d-dimensional Brownian motion W and a Poisson random measure $N(dt, de)$ with compensator $v(de)dt$, are defined. We assume v is a σ-finite measure on $E := \mathbb{R}^n \setminus \{0\}$ equipped with its Borel field $\mathcal{B}(E)$ and satisfies

$$\int_E (1 \wedge |e|^2)\, v(de) < \infty. \tag{2.1}$$

We denote by $\mathbb{E}$ the usual expectation operator with respect to the measure P.

For any given $t \in [0, T]$, we define the t-translated Brownian motion $W^t := (W_s - W_t)_{s \geq t}$ and the t-translated Poisson random measure $N^t := N(]t, s], \cdot)_{s \geq t}$. We denote by $\tilde{N}^t(dt, de) = N^t(dt, de) - v(de)dt$ the compensated process of N^t, and by $\mathbb{F}^t = \{\mathcal{F}^t_s\}_{s \in [t, T]}$ be the filtration generated by W^t and N^t augmented by the P-null sets.

Furthermore, we introduce several spaces: L^2_v is the space of Borel functions $l : E \to \mathbb{R}$ with $\|l\|^2_v := \int_E |l(e)|^2\, v(de) < \infty$; $\mathbb{H}^2_t$ (resp. $\mathbb{H}^2_{t,v}$) is the space of $\mathbb{R}^d$-valued (resp. real-valued) $\mathbb{F}^t$-predictable processes (π_s) (resp. $(l_s(\cdot))$ with $\mathbb{E}\int_t^T |\pi_s|^2\, ds < \infty$ (resp. $\mathbb{E}\int_t^T \|l_s\|^2_{L^2_v}\, ds < \infty$); $\mathcal{S}^2$ is the space of real-valued $\mathbb{F}^t$-adapted càdlàg processes (ψ_s) with $\mathbb{E}[\sup_{t \leq s \leq T} \psi^2_s] < \infty$.

We now proceed to introduce the control problem of interest. For each $t \in [0, T]$, let $\mathcal{A}^t_t$ be a set of admissible controls, which are $\mathbb{F}^t$-predictable processes $(\alpha_s)_{s \in [t, T]}$ valued in a compact set $\mathbf{A}$, and $\mathcal{T}^t_t$ be the set of $\mathbb{F}^t$-stopping times which take values in $[t, T]$. For any given initial state $x \in \mathbb{R}^d$, and control $\alpha \in \mathcal{A}^t_t$, we consider the controlled jump-diffusion process $(X^{\alpha,t,x}_s)_{t \leq s \leq T}$ satisfying the following SDE: for each $s \in [t, T]$,

$$X^{\alpha,t,x}_s = x + \int_t^s b(\alpha_v, X^{\alpha,t,x}_v)\, dv + \int_t^s \sigma(\alpha_v, X^{\alpha,t,x}_v)\, dW^t_v$$

$$+ \int_t^s \int_E \eta(\alpha_v, X^{\alpha,t,x}_v, e)\, \tilde{N}^t(dv, de), \tag{2.2}$$

where b, $\eta \in \mathbb{R}^d$ and $\sigma \in \mathbb{R}^{d \times d}$ are given measurable functions. We remark that although our analyses are performed only for jump-diffusion processes with time-homogenous coefficients, similar results are valid for controlled dynamics with time-dependent coefficients.

The performance of the control problem, depending on α, is evaluated by a nonlinear expectation induced by a BSDE with a controlled driver $f(\alpha_s, s, X^{\alpha,t,x}_s, y, z, k)$. That is, for any given stopping time $\tau \in \mathcal{T}^t_t$ and any bounded Borel function ξ, we define the nonlinear expectation

$$\mathcal{E}_{t,\tau}^{\alpha,t,x}[\xi(\tau, X_{\tau}^{\alpha,t,x})] := Y_t^{\alpha,\tau,t,x},$$

where the process $(Y_s^{\alpha,\tau,t,x})_{s \leq \tau}$ is a solution in $\mathcal{S}_t^2$ of the following BSDE: for each $s \in [t, \tau]$,

$$\begin{cases} -dY_{s,\tau}^{\alpha,t,x} = f(\alpha_s, s, X_s^{\alpha,t,x}, Y_{s,\tau}^{\alpha,t,x}, Z_{s,\tau}^{\alpha,t,x}, K_{s,\tau}^{\alpha,t,x})ds - Z_{s,\tau}^{\alpha,t,x}dW_s^t \\ \quad - \int_E K_{s,\tau}^{\alpha,t,x} \tilde{N}^t(ds, de), \\ Y_{\tau,\tau}^{\alpha,t,x} = \xi(\tau, X_{\tau}^{\alpha,t,x}), \end{cases} \quad (2.3)$$

and $(Z_{s,\tau}^{\alpha,t,x})$, $(K_{s,\tau}^{\alpha,t,x})$ are two associated processes, if they exist, lying in $\mathbb{H}_t$ and $\mathbb{H}_t^{\nu}$, respectively.

Now we are ready to state the generalized mixed optimal stopping and control problem. For each initial time $t \in [0, T]$ and initial state $x \in \mathbb{R}^d$, we consider the following value function:

$$u(t, x) = \sup_{\tau \in \mathcal{T}_t^t} \sup_{\alpha \in \mathcal{A}_t^t} \mathcal{E}_{t,\tau}^{\alpha,t,x}[\xi(\tau, X_{\tau}^{\alpha,t,x})], \quad (2.4)$$

subject to the controlled SDE (2.2), where ξ is the terminal position given by

$$\xi(\tau, X_{\tau}^{\alpha,t,x}) = \zeta(\tau, X_{\tau}^{\alpha,t,x})1_{t \leq \tau < T} + g(X_T^{\alpha,t,x})1_{\tau=T},$$

for some reward functions ζ and g. Note that the value function of our control problem is constant up to a P-null set. Throughout this work, we shall perform the analysis under the following standard assumptions on the coefficients:

Assumption 2.1 The set of control values $\mathbf{A}$ is compact and the driver f is a measurable function of the form $f(\alpha, s, x, y, z, k) := \hat{f}(\alpha, s, x, y, z, \int_E k(e)\gamma(x, e)\nu(de))1_{s \geq t}$ for some functions $\hat{f}$ and γ. Moreover, there exists a constant $C > 0$ such that for any $\alpha, \alpha' \in \mathbf{A}, t \in [0, T], e \in E, x, x' \in \mathbb{R}^d, u, v \in \mathbb{R}, p, q \in \mathbb{R}^d, k, k' \in \mathbb{R}$, we have

(1) $|b(\alpha, x) - b(\alpha', x')| + |\sigma(\alpha, x) - \sigma(\alpha', x')| \leq C(|x - x'| + |\alpha - \alpha'|)$;
(2) $|\eta(\alpha, x, e) - \eta(\alpha', x', e)| \leq C(|x - x'| + |\alpha - \alpha'|)(1 \wedge |e|)$ and $|\eta(\alpha, x, e)| \leq C(1 \wedge |e|)$;
(3) $|\gamma(x, e) - \gamma(x', e)| \leq C|x - x'|(1 \wedge |e|^2)$; $|\gamma(x, e)| \leq C(1 \wedge |e|)$ and $\gamma(x, e) \geq 0$;
(4) $\hat{f} : \mathbf{A} \times [0, T] \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}$ is continuous in t and admits the properties:

   (a) (Boundedness.) $|\hat{f}(\alpha, t, x, 0, 0, 0)| \leq C$;
   (b) (Monotonicity.) $\hat{f}(\alpha, t, x, v, p, k) - \hat{f}(\alpha, t, x, u, p, k) \geq C(u - v)$ when $u \geq v$, and $k \to \hat{f}(\alpha, t, x, u, p, k)$ is non-decreasing in k;
   (c) (Lipschitz continuity.) $|\hat{f}(\alpha, t, x, u, p, k) - \hat{f}(\alpha', t, x', v, q, k')| \leq C(|\alpha - \alpha'| + |x - x'| + |u - v| + |p - q| + |k - k'|)$;

(5) the function $\zeta : [0, T] \times \mathbb{R}^d \to \mathbb{R}$ is continuous in t, and we have $|g(x) - g(x')| + |\zeta(t, x) - \zeta(t, x')| \leq C|x - x'|$, $|g(x)| + |\zeta(t, x)| \leq C$, and $g(x) \geq \zeta(0, x)$.

Remark 1 Assumption 2.1 is the same as that made in Reference [17]. Under assumptions (1), (2), equation (2.2) admits a unique solution. Assumptions (3), (4.a), (4.c), (5) guarantee the existence and uniqueness of the solution of (2.3).

Note that the Lipschitz continuity of $\hat{f}$ with respect to (x, y, z, u) is a key assumption in order to establish the "good" measurability properties of the value function $u^\alpha(t, x)$ with respect to α and x, where $u^\alpha(t, x) := \sup_{\tau \in \mathcal{T}_t^l} \mathcal{E}_{t,\tau}^{\alpha,t,x}[\xi(\tau, X_\tau^{\alpha,t,x})]$, which are needed in the proof of the dynamic programming principle in Reference [17] (see in particular Sect. 3.2, Theorem 3.7, which holds under Assumption 2.1).

The monotonicity of $\hat{f}$ with respect to the fourth variable y is a standard assumption for the comparison principle of the viscosity solutions to the HJB equations (see e.g. Assumption (C3) in Reference [22]). It can be made without loss of generality by performing an exponential in time scaling of the solution (see e.g. Remark 2.1 in Reference [22]).

The boundedness of $\hat{f}$, g and ζ ensures the boundedness of the value function u (see Lemma 6.2 in Reference [16]), which subsequently allows us to work with the infinity norm of the solution to the HJBVIs. It also plays an important role in deriving the error estimates of our approximations (see e.g. Theorem 4.3).

Finally, we remark that the Lipschitz continuity of γ is needed in Reference [16] for the proof of the comparison principle, which ensures the uniqueness of the solution to the corresponding HJBVIs with nonlinear drivers (see e.g. Lemma 4.3 in Reference [16] and (A.2 vi) in Reference [4]).

For notational convenience, in the sequel, we will write $\hat{f}$ as f, and denote by ψ^α a generic function ψ with control-dependence.

The rest of this section is devoted to the equivalence between the mixed control problem and a generalized nonlocal HJBVI. Specifically, we now consider a Hamilton-Jacobi-Bellman variational inequality of the following form:

$$
\begin{aligned}
0 &= F(\mathbf{x}, u, Du, D^2u, \{K^\alpha u\}_{\alpha \in \mathbf{A}}, \{B^\alpha u\}_{\alpha \in \mathbf{A}}) \\
&= \begin{cases} \min\left\{u(\mathbf{x}) - \zeta(\mathbf{x}), u_t + \inf_{\alpha \in \mathbf{A}}\left(-L^\alpha u - f^\alpha[u]\right)\right\}, & \mathbf{x} \in \mathcal{Q}_T, \\ u(\mathbf{x}) - g(x), & \mathbf{x} \in \{0\} \times \mathbb{R}^d, \end{cases}
\end{aligned}
\tag{2.5}
$$

where $f^\alpha[u] = f(\alpha, \mathbf{x}, u, (\sigma^\alpha)^T Du, B^\alpha u)$, $\mathcal{Q}_T = (0, T] \times \mathbb{R}^d$, $\mathbf{x} = (t, x)$ contains both the time t and the spatial coordinate $x \in \mathbb{R}^d$, and the nonlocal operators $L^\alpha := A^\alpha + K^\alpha$ and B^α satisfy, for $\phi \in C^{1,2}(\bar{\mathcal{Q}}_T)$:

$$
A^\alpha\phi(\mathbf{x}) = \frac{1}{2}\text{tr}(\sigma^\alpha(x)(\sigma^\alpha(x))^T D^2\phi(\mathbf{x})) + b^\alpha(x) \cdot D\phi(\mathbf{x}),
\tag{2.6}
$$

$$
K^\alpha\phi(\mathbf{x}) = \int_E \left(\phi(t, x + \eta^\alpha(x, e)) - \phi(\mathbf{x}) - \eta^\alpha(x, e) \cdot D\phi(\mathbf{x})\right)\nu(de),
\tag{2.7}
$$

$$
B^\alpha\phi(\mathbf{x}) = \int_E \left(\phi(t, x + \eta^\alpha(x, e)) - \phi(\mathbf{x})\right)\gamma(x, e)\,\nu(de),
\tag{2.8}
$$

where $E = \mathbb{R}^n \setminus \{0\}$ is defined at the beginning of Sect. 2 and the nonlocal operators $K^\alpha u$ and $B^\alpha u$ are well-defined under Assumption 2.1.

We emphasize that since the matrix $\sigma^\alpha (\sigma^\alpha)^T$ is only assumed to be nonnegative definite, both the diffusion coefficient $\sigma^\alpha (\sigma^\alpha)^T$ and the jump intensity η of (2.5) are allowed to vanish at some points. Consequently, there is no Laplacian smoothing from the second-order differential operator nor fractional Laplacian smoothing from the nonlocal operator to this degenerate equation (2.5). Therefore, in general, this HJBVI will not admit classical solutions, and we shall interpret the equation in the following viscosity sense based on semi-continuous envelopes of the equation [1,29].

Definition 2.2 (*Viscosity solution of HJBVI*) An upper (resp. lower) semicontinuous function u is said to be a viscosity subsolution (resp. supersolution) of (2.5) if and only if for any point $\mathbf{x}_0$ and for any $\phi \in C^{1,2}(\bar{\mathcal{Q}}_T)$ such that $\phi(\mathbf{x}_0) = u(\mathbf{x}_0)$ and $u - \phi$ attains its global maximum (resp. minimum) at $\mathbf{x}_0$, one has

$$F_*(\mathbf{x}_0, u(\mathbf{x}_0), D\phi(\mathbf{x}_0), D^2\phi(\mathbf{x}_0), \{K^\alpha \phi(\mathbf{x}_0)\}_{\alpha \in \mathbf{A}}, \{B^\alpha \phi(\mathbf{x}_0)\}_{\alpha \in \mathbf{A}}) \leq 0,$$

$$\left(resp. \quad F^*(\mathbf{x}_0, u(\mathbf{x}_0), D\phi(\mathbf{x}_0), D^2\phi(\mathbf{x}_0), \{K^\alpha \phi(\mathbf{x}_0)\}_{\alpha \in \mathbf{A}}, \{B^\alpha \phi(\mathbf{x}_0)\}_{\alpha \in \mathbf{A}}) \geq 0 \right).$$

A continuous function is a viscosity solution of the HJBVI (2.5) if it is both a a viscosity sub- and supersolution.

Under Assumption 2.1, the HJBVI (2.5) is well-posed in the class of bounded continuous functions (see References [16,17]). The unique viscosity solution of (2.5) (after a change of time variable) can be further characterized as the optimal value function (2.4) of the mixed control problem. In other words, to obtain the optimal value function for all initial times t and initial states x, it is equivalent to design effective numerical schemes to solve (2.5).

Moreover, a strong comparison principle holds for the HJBVI (2.5), the proof of which is similar to that in Reference [16] (without controls) and hence omitted. In particular, if U is a bounded viscosity subsolution and V is a bounded viscosity supersolution to (2.5) with $U(0, \cdot) \leq V(0, \cdot)$, we have $U(\mathbf{x}) \leq V(\mathbf{x})$ for all $\mathbf{x} \in \bar{\mathcal{Q}}_T$.

3 Construction of Numerical Schemes

In this section, we will design numerical schemes for solving HJBVI (2.5). We carry out the following string of approximations to construct our numerical algorithm:

- truncation of the singular jump measure (Eq. (3.3) and Sect. 4.1);
- approximation of the control set with a finite set (Eq. (3.5) and Theorem 4.3);
- approximation of the discretized control problem with a switching system (Eq. (3.6) and Theorem 4.6);
- discretization in time and space (Eqs. (3.8, 3.9) and Theorem 4.7).

We start the derivations of our schemes by approximating the singular measure ν with a truncated non-singular measure and a modified diffusion coefficient as suggested in Reference [10]. This can be done by introducing an approximative jump-diffusion dynamics and an approximative backward SDE (see Sect. 4.1). More precisely, for any given $\varepsilon > 0$, let us define the truncated measure $\nu_\varepsilon(de) =$

$1_{|e|>\varepsilon}\nu(de)$ and the modified diffusion coefficient $\tilde{\sigma}^\alpha(x)$ such that $\tilde{\sigma}_{ij}^\alpha(x) = \sigma_{ij}^\alpha(x)$ for $i \neq j$ and

$$\tilde{\sigma}_{ii}^\alpha(x) = \text{sgn}(\sigma_{ii}^\alpha(x))\left((\sigma_{ii}^\alpha(x))^2\right.$$
$$\left.+ \int_{|e|<\varepsilon} |\eta_i^\alpha(x, e)|^2 \nu(de)\right)^{1/2}, \quad i = 1, \ldots, d, \ x \in \mathbb{R}^d. \quad (3.1)$$

We further introduce the modified local operator A_ε^α as:

$$A_\varepsilon^\alpha \phi(\mathbf{x}) := \frac{1}{2}\text{tr}(\tilde{\sigma}^\alpha(x)(\tilde{\sigma}^\alpha(x))^T D^2\phi(\mathbf{x})), \quad \phi \in C^{1,2}([0, T] \times \mathbb{R}^d), \quad (3.2)$$

and truncated nonlocal operators K_ε^α and B_ε^α by replacing ν with ν_ε in (2.7) and (2.8), respectively.

With these operators in hand, we consider the following modified HJBVI:

$$0 = F^\varepsilon(\mathbf{x}, u, Du, D^2u, \{K_\varepsilon^\alpha u\}_{\alpha \in \mathbf{A}}, \{B_\varepsilon^\alpha u\}_{\alpha \in \mathbf{A}})$$
$$= \begin{cases} \min\left\{u - \zeta, u_t + \min_{\alpha \in \mathbf{A}}\left(-L_\varepsilon^\alpha u - f_\varepsilon^\alpha[u]\right)\right\}, & \mathbf{x} \in \mathcal{Q}_T, \\ u(\mathbf{x}) - g(x), & \mathbf{x} \in \{0\} \times \mathbb{R}^d, \end{cases} \quad (3.3)$$

where we have $f_\varepsilon^\alpha[u] = f^\alpha(\mathbf{x}, u, (\tilde{\sigma}^\alpha)^T Du, B_\varepsilon^\alpha u)$ and $L_\varepsilon^\alpha \phi := A_\varepsilon^\alpha \phi + K_\varepsilon^\alpha \phi$ for any $\phi \in C^{1,2}([0, T] \times \mathbb{R}^d)$. These modified coefficients clearly satisfy Assumption 2.1, and hence (3.3) is well-posed in the viscosity sense.

We then approximate the admissible control set in (3.3) by a finite set. More precisely, for a finite subset $\mathbf{A}_\delta$ of the compact set $\mathbf{A}$ such that

$$\max_{\alpha \in \mathbf{A}} \min_{\tilde{\alpha} \in \mathbf{A}_\delta} |\alpha - \tilde{\alpha}| < \delta, \quad (3.4)$$

we introduce the finite control HJBVI by

$$0 = F^{\varepsilon,\delta}(\mathbf{x}, u, Du, D^2u, \{K_\varepsilon^\alpha u\}_{\alpha \in \mathbf{A}_\delta}, \{B_\varepsilon^\alpha u\}_{\alpha \in \mathbf{A}_\delta})$$
$$= \begin{cases} \min\left\{u - \zeta, u_t + \min_{\alpha \in \mathbf{A}_\delta}\left(-L_\varepsilon^\alpha u - f_\varepsilon^\alpha[u]\right)\right\}, & \mathbf{x} \in \mathcal{Q}_T, \\ u(\mathbf{x}) - g(x), & \mathbf{x} \in \{0\} \times \mathbb{R}^d, \end{cases} \quad (3.5)$$

Since (3.5) is a special case of (2.5) with a finite admissible set, it is clear that (3.5) admits a unique bounded viscosity solution.

Remark 2 In Sect. 4.1, we provide an alternative interpretation of the above two approximations by identifying the viscosity solution of (3.5) as the value function of a mixed control problem in terms of modified SDE and BSDE. This characterization further enables us to establish the convergence of these approximations through a probabilistic argument.

Next, we approximate the finite control equation (3.5) by a switching system (
[2,6]). Suppose the finite control set is given by $\mathbf{A}_\delta = \{\alpha_1, \alpha_2, \ldots, \alpha_J\}$. We denote
by $U_j^{\varepsilon,\delta,c}$, $j = 1, \ldots, J$ the solution of the following system of HJB equations:

$$
\begin{aligned}
0 &= F_j^{\varepsilon,\delta,c}(\mathbf{x}, U_j, DU_j, D^2 U_j, \{K_\varepsilon^\alpha u\}_{\alpha \in \mathbf{A}_\delta}, \{B_\varepsilon^\alpha u\}_{\alpha \in \mathbf{A}_\delta}, \{U_k\}_{k \neq j}) \\
&= \begin{cases} \min \left[U_j - \zeta, \ \min \left(U_{j,t} - L_\varepsilon^{\alpha_j} U_j - f_\varepsilon^{\alpha_j}[U_j]; \ U_j - \mathcal{M}_j U \right) \right], & \mathbf{x} \in \mathcal{Q}_T, \\ U_j(\mathbf{x}) - g(x), & \mathbf{x} \in \{0\} \times \mathbb{R}^d, \end{cases}
\end{aligned}
\tag{3.6}
$$

where we define $\mathcal{M}_j U := \max_{k \neq j} \left(U_k - c \right)$ for any $c > 0$. The positive switching
cost is needed for the well-posedness of the switching system (3.6).

We now proceed to introduce a discrete approximation to the switching system
based on the idea of piecewise constant policy timestepping. Define a set of nodes
$\{x_{j,i}\}$ and timesteps t^n, with discretization parameters h and Δt, i.e.,

$$
\max_{1 \leq j \leq J, x \in \mathbb{R}^d} \min_i |x - x_{j,i}| = h, \quad \max_n (t^{n+1} - t^n) = \Delta t.
\tag{3.7}
$$

By parameterizing the grid $\Omega_{j,h} = \{\mathbf{x}_{j,i}\}_i$ with the control index j, we are allowing
the usage of different discretization grids for different controls. We denote by $U_{j,i}^n$
the discrete approximation to U_j at the point $\mathbf{x}_{j,i}^n = (t^n, x_{j,i})$, and extend it to the
computational domain by interpolation.

Let $L_{\varepsilon,h}^{\alpha_j}$ and $f_{\varepsilon,h}^{\alpha_j}$ be the discrete form of the operators $L_\varepsilon^{\alpha_j}$ and $f_\varepsilon^{\alpha_j}$, respectively.
We discretize (3.6) on the grid $\Omega_{j,h}$ with the uniform time partition $t^{n+1} - t^n = \Delta t$
by performing piecewise constant policy timestepping and applying the constraints at
the beginning of a new timestep,

$$
U_{j,i}^{n+\frac{1}{2}} = \max \left[\zeta_i^{n+1}, \ U_{j,i}^n, \ \max_{k \neq j} (\tilde{U}_{k,i(j)}^n - c)) \right],
\tag{3.8}
$$

$$
U_{j,i}^{n+1} - \Delta t \left(L_{\varepsilon,h}^{\alpha_j} U_{j,i}^{n+1} + f_{\varepsilon,h}^{\alpha_j}[U_{j,i}^{n+1}] \right) = U_{j,i}^{n+\frac{1}{2}}, \qquad j = 1, \ldots, J,
\tag{3.9}
$$

where $\tilde{U}_{k,i(j)}^n$ is the value of the interpolant of $\{U_{k,l}^n\}_{l \in \Omega_{k,l}}$ at the i-th point of the grid
$\Omega_{j,h}$.

Now by rearranging the terms of (3.9), we obtain the following numerical scheme:
for $\mathbf{x}_{j,i}^n \in \mathcal{Q}_T$ and $j = 1, \ldots, J$,

$$
\begin{aligned}
0 &= G_j(\mathbf{x}_{j,i}^{n+1}, h, U_{j,i}^{n+1}, \{U_{j,a}^{b+1}\}_{(a,b) \neq (i,n)}, \{\tilde{U}_k^n\}_{k \neq j}) \\
&= \min \left[U_{j,i}^{n+1} - \zeta_i^{n+1} - \Delta t \left(L_{\varepsilon,h}^{\alpha_j} U_{j,i}^{n+1} + f_{\varepsilon,h}^{\alpha_j}[U_{j,i}^{n+1}] \right), \ \frac{U_{j,i}^{n+1} - U_{j,i}^n}{\Delta t} - L_{\varepsilon,h}^{\alpha_j} U_{j,i}^{n+1} \right. \\
&\quad \left. - f_{\varepsilon,h}^{\alpha_j}[U_{j,i}^{n+1}], U_{j,i}^{n+1} - \max_{k \neq j} (\tilde{U}_{k,i(j)}^n - c) - \Delta t \left(L_{\varepsilon,h}^{\alpha_j} U_{j,i}^{n+1} + f_{\varepsilon,h}^{\alpha_j}[U_{j,i}^{n+1}] \right) \right]. \tag{3.10}
\end{aligned}
$$

As seen from (3.10), performing switching at the beginning of a new timestep introduces two additional terms to both the switching part and the obstacle part of the equation, which will not appear in a straightforward discretization of the switching system (3.6). However, we will demonstrate in Sect. 4.4 that these terms vanish as $\Delta t, h \to 0$, and consequently our scheme (3.10) forms a consistent approximation to (3.6).

For notational simplicity, we label our approximations only by h and assume in the sequel that Δt is a given function of h with $\Delta t \to 0$ as $h \to 0$.

Remark 3 To determine the optimal stopping strategy and the optimal controls, one can simply compare values of the obstacle and all components of the switching system at each grid point. As we will see in Sect. 4.2, each component of the switching system converges to the solution of the HJBVI (2.5) as the discretization parameters tend to 0; therefore, although we have no guarantee for the convergence of the control approximation, this numerically found control is close to optimal for the mixed control problem (2.4).

We now describe in detail how we perform spatial discretizations for L_ε^α and B_ε^α to construct a monotone discrete operator $L_{\varepsilon,h}^\alpha$ and $f_{\varepsilon,h}^\alpha$, for a fixed control parameter $\alpha \in \{\alpha_1, \ldots, \alpha_J\}$. To simplify the presentation, we consider a piecewise linear or multilinear interpolation $\mathcal{I}_h$ on a uniform spatial grid $h\mathbb{Z}^d$. That is,

$$\mathcal{I}_h[\phi](x) = \sum_{m \in \mathbb{Z}^d} \phi(x_m)\omega_m(x; h), \quad x \in \mathbb{R}^d, \tag{3.11}$$

for the standard "tent functions" ω_m satisfying $0 \le \omega_m(x; h) \le 1$, $\omega_m(x_i; h) = \delta_{mi}$, $\sum_m \omega_m = 1$, $\text{supp}\,\omega_m \subset B(x_m, 2h)$, and $|D\omega_m| \le C/h$.

We start with the nonlocal terms. The definition of K_ε^α gives

$$K_\varepsilon^\alpha \phi(\mathbf{x}) = \int_{|e| \ge \varepsilon} \left(\phi(t, \mathbf{x} + \eta^\alpha(x, e)) - \phi(x) - \eta^\alpha(x, e) \cdot D\phi(\mathbf{x}) \right) \nu(de)$$

$$= \int_{|e| \ge \varepsilon} \left(\phi(t, x + \eta^\alpha(x, e)) - \phi(x) \right) \nu(de)$$

$$+ \int_{|e| \ge \varepsilon} -\eta^\alpha(x, e)\,\nu(de) \cdot D\phi(\mathbf{x})$$

$$:= K_\varepsilon^{\alpha,1}\phi(\mathbf{x}) + b_\varepsilon^\alpha(x) \cdot D\phi(\mathbf{x}).$$

Then, by replacing the integrands by their monotone interpolants (c.f. [7]), we derive the following approximations for $K_\varepsilon^{\alpha,1}$ and B_ε^α (where we have dropped the mesh index j in x for simplicity):

$$K_{\varepsilon,h}^{\alpha,1}\phi(t^n, x_i) := \int_{|e| \ge \varepsilon} \mathcal{I}_h[\phi(t^n, x_i + \cdot) - \phi(t^n, x_i)](\eta^\alpha(x_i, e))\,\nu(de)$$

$$= \sum_{m \in \mathbb{Z}^d} \kappa_{h,m,i}^{\alpha,n}[\phi(t^n, x_i + x_m) - \phi(t^n, x_i)] \tag{3.12}$$

$$
\begin{aligned}
B_{\varepsilon,h}^{\alpha}\phi(t^n, x_i) &:= \int_{|e|\geq\varepsilon} \mathcal{I}_h[\phi(t^n, x_i + \cdot) - \phi(t^n, x_i)](\eta^{\alpha}(x_i, e))\gamma(x_i, e)\,\nu(de) \\
&= \sum_{m\in\mathbb{Z}^d} \beta_{h,m,i}^{\alpha,n}[\phi(t^n, x_i + x_m) - \phi(t^n, x_i)],
\end{aligned}
\tag{3.13}
$$

with the coefficients

$$
\begin{aligned}
\kappa_{h,m,i}^{\alpha,n} &:= \int_{|e|\geq\varepsilon} \omega_m(\eta^{\alpha}(x_i, e); h)\,\nu(de), \quad \beta_{h,m,i}^{\alpha,n} \\
&:= \int_{|e|\geq\varepsilon} \omega_m(\eta^{\alpha}(x_i, e); h)\gamma(x_i, e)\,\nu(de),
\end{aligned}
\tag{3.14}
$$

which are well-defined and nonnegative, and consequently result in monotone approximations. These coefficients can be efficiently evaluated by using quadrature rules with positive weights, such as Gauss methods of appropriate order [7].

We then turn to discretizing the local terms. We introduce the modified drift by $\tilde{b}^{\alpha}(x) = b^{\alpha}(x) + b_{\varepsilon}^{\alpha}(x)$, and write the modified diffusion $\tilde{\sigma}^{\alpha} = (\tilde{\sigma}_1^{\alpha}, \ldots, \tilde{\sigma}_d^{\alpha})$, where $\tilde{\sigma}_l^{\alpha}$, $l = 1, \ldots, d$ is the l-th column of $\tilde{\sigma}^{\alpha}$ defined in (3.1).

With these modified coefficients, we are ready to construct the following approximations of the local operators: for any $k > 0$,

$$
\begin{aligned}
\frac{1}{2}\mathrm{tr}(\tilde{\sigma}^{\alpha}(x)(\tilde{\sigma}^{\alpha}(x))^T D^2\phi(x)) &\approx \frac{1}{2}\sum_{l=1}^{d} \frac{\mathcal{I}_h[\phi](x + k\tilde{\sigma}_l^{\alpha}) - 2\mathcal{I}_h[\phi](x) + \mathcal{I}_h[\phi](x - k\tilde{\sigma}_l^{\alpha})}{k^2} \\
\tilde{b}^{\alpha}(x) \cdot D\phi &\approx \frac{1}{2}\frac{\mathcal{I}_h[\phi](x + k^2\tilde{b}^{\alpha}) - 2\mathcal{I}_h[\phi](x) + \mathcal{I}_h[\phi](x + k^2\tilde{b}^{\alpha})}{k^2},
\end{aligned}
$$

which, by using (3.11) and the fact that $\sum_m \omega_m = 1$, can be further written in the discrete monotone form

$$
A_{\varepsilon,h,k}^{\alpha}\phi(t^n, x_i) = \sum_{m\in\mathbb{Z}^d} d_{h,k,m,i}^{\alpha,n}[\phi(t^n, x_m) - \phi(t^n, x_i)],
\tag{3.15}
$$

with non-negative coefficients

$$
\begin{aligned}
d_{h,k,m,i}^{\alpha,n} &= \frac{1}{2}\sum_{l=1}^{d} \frac{\omega_m(x_i + k\tilde{\sigma}_l^{\alpha}(x_i); h) + \omega_m(x_i - k\tilde{\sigma}_l^{\alpha}(x_i); h)}{k^2} \\
&+ \frac{\omega_m(x_i + k^2\tilde{b}^{\alpha}(x_i); h)}{k^2} \geq 0.
\end{aligned}
\tag{3.16}
$$

The approximation to the local operator A_{ε}^{α} falls into the class of semi-Lagrangian schemes (see e.g. [13]), and provides a consistent monotone approximation for possibly degenerate, non-diagonally dominant diffusion coefficients.

Before presenting our fully discrete scheme, we shall point out that by considering a truncated problem, one can without loss of generality assume that σ^α is bounded, which consequently implies the Hamiltonian

$$\bar{f}^\alpha(\mathbf{x}, u, p, k) := f^\alpha(\mathbf{x}, u, (\tilde{\sigma}^\alpha(x))^T p, k)$$

is Lipschitz continuous with respect to p. Indeed, suppose σ^α is unbounded, then for any given $\mu > 0$, we define the cut-off function

$$\xi_\mu : \mathbb{R}^d \to [0, 1], \quad \xi_\mu \equiv 1 \text{ for } |x| < \frac{1}{\mu}, \text{ and } \xi_\mu \in C_0^\infty(\mathbb{R}^d),$$

and consider a truncated HJBVI (3.5) by replacing σ^α with the bounded diffusion coefficient $\sigma_\mu^\alpha(x) := \xi_\mu(x)\sigma^\alpha(x)$. Using the fact that $1 - \xi_\mu \to 0$ uniformly on compact sets as $\mu \to 0$, one can easily prove this additional approximation is consistent with (3.5), and hence its viscosity solution converges to the solution to (3.5) uniformly on bounded sets.

The Lipschitz continuous Hamiltonian enables an approximation by the implicit Lax-Friedrichs numerical flux [11]. For each $l = 1, \ldots, d$, we denote by $\Delta_+^{(l)} U_{j,i}^n$ (resp. $\Delta_-^{(l)} U_{j,i}^n$) the one-step forward (resp. backward) difference operator along the l-th coordinate, and by $\Delta U_{j,i}^n = (\Delta_+^{(1)} U_{j,i}^n + \Delta_-^{(1)} U_{j,i}^n, \ldots, \Delta_+^{(d)} U_{j,i}^n + \Delta_-^{(d)} U_{j,i}^n)^T$ the central difference operator at the grid point $\mathbf{x}_{j,i}^n$. Then for any $\theta > 0$, the Lax-Friedrichs numerical flux is given for any $(\mathbf{x}_{j,i}^n, u, k) \in \Omega_{j,h} \times \mathbb{R} \times \mathbb{R}$ by

$$\tilde{f}^\alpha(\mathbf{x}_{j,i}^n, u, \Delta U_{j,i}^n, k) := \bar{f}^\alpha\left(\mathbf{x}_{j,i}^n, u, \frac{\Delta U_{j,i}^n}{2h}, k\right) + \sum_{l=1}^d \frac{\theta}{\lambda}\left(\frac{\Delta_+^{(l)} U_{j,i}^n - \Delta_-^{(l)} U_{j,i}^n}{h}\right),$$

$$(3.17)$$

where we define $\lambda = \Delta t/h$.

A fully implicit time discretisation is finally given by

$$U_{j,i}^n - \Delta t\left(A_{\varepsilon,h,k}^\alpha U_{j,i}^n + K_{\varepsilon,h}^{\alpha,1} U_{j,i}^n \right. $$
$$\left. + \tilde{f}^\alpha(\mathbf{x}_{j,i}^n, U_{j,i}^n, \Delta U_{j,i}^n, B_{\varepsilon,h}^\alpha U_{j,i}^n)\right) - U_{j,i}^{n-\frac{1}{2}} = 0. \tag{3.18}$$

Substituting (3.8) into (3.18), one can reformulate this implicit scheme in its equivalent form (3.10).

We end this section with a remark about the implementation of the implicit scheme (3.18). To avoid solving linear systems with the dense matrices resulting from the discretization of the nonlocal operators, we write the solution to (3.18) as the fixed point of a sparse contraction mapping T, such that sufficient accuracy is achieved in practice by a few fixed point iterations.

Given bounded functions $U_j^{n-\frac{1}{2}}$ and $U_j^{n,(k)}$, we define the following mapping T on $\ell^\infty(\mathbb{Z}^d)$, i.e., the Banach space of bounded functions on $h\mathbb{Z}^d$ employed with the sup-norm $|\cdot|_0$:

$$(1 - \Delta t A_{\varepsilon,h,k}^\alpha)(TU_{j,i}^{n,(k)}) = \Delta t\big(K_{\varepsilon,h}^{\alpha,1}U_{j,i}^{n,(k)}$$
$$+ \tilde{f}^\alpha(\mathbf{x}_{j,i}^n, U_{j,i}^{n,(k)}, \Delta U_{j,i}^{n,(k)}, B_{\varepsilon,h}^\alpha U_{j,i}^{n,(k)})\big) + U_{j,i}^{n-\frac{1}{2}}. \qquad (3.19)$$

It is clear that a fixed point U_j^n with $TU_j^n = U_j^n$ is a solution to (3.18). Moreover, for any given functions $U_j^{n,(k)}$ and $V_j^{n,(k)}$ in $\ell^\infty(\mathbb{Z}^d)$, we obtain from the Lipschitz continuity of f and the ℓ^∞ stability of the numerical flux $\tilde{f}$ (see Lemma 4.11) that

$$|TU_j^{n,(k)} - TV_j^{n,(k)}|_0 \leq \bigg(\Delta t\Big[\sum_{m \neq 0}\kappa_{h,m,i}^{\alpha,n}$$
$$+ C\big(1 + \sum_{m \neq 0}\beta_{h,m,i}^{\alpha,n}\big)\Big] + 4d\theta\bigg)|U_j^{n,(k)} - V_j^{n,(k)}|_0.$$

Since we need $h = o(\varepsilon)$ in general to achieve consistency of our scheme (see Lemma 4.9), it suffices to require $\frac{\Delta t}{\varepsilon^2} < 1$ and $4d\theta < 1$ to ensure T is a contraction mapping on $\ell^\infty(\mathbb{Z}^d)$. This establishes the well-posedness of (3.18) and enables us to solve the nonlinear equation (3.18) through Picard iterations by setting

$$U_j^{n,(0)} = U_j^{n-\frac{1}{2}}, \qquad U_j^{n,(k+1)} = TU_j^{n,(k)}, \quad k \geq 0.$$

We emphasize that the criterion $\frac{\Delta t}{\varepsilon^2} < 1$ is a sufficient condition in the worst case, but is often far from computationally optimal since we have used no information about the exact behavior of the singular measure ν around zero (see Remark 6 for details). For typical Lévy measures from finance [10], we only need $\Delta t = O(\varepsilon)$ (such as in the variance gamma case in our tests) or even Δt independent of ε (for instance, for a Gaussian density) to guarantee T is a contraction mapping.

Moreover, since in practice we can evaluate the discrete nonlocal operators $K_{\varepsilon,h}^{\alpha,1}U_j^n$ and $B_{\varepsilon,h}^\alpha U_j^n$ at all grid points in $O(N\log N)$ operations using a FFT (see e.g. [14]), and in each iteration a sparse linear system is solved (in one dimension it is tridiagonal), the total complexity of the implicit scheme is still close to linear.

It is also worth pointing out that even if we chose explicit approximations for the nonlocal operators, due to the nonlinearity of f, one still has to perform iterations to solve the resulting nonlinear equations. Because we only assume Lipschitz continuity but no higher regularity of f, we adopt the Picard iteration to solve for U^n.

4 Convergence Analysis

In this section, we establish the convergence of the numerical approximations in Sect. 3 to the viscosity solution of the HJBVI (2.5). To simplify the notation, we will occasionally drop the terms $\{K^\alpha u\}_{\alpha \in \mathbf{A}}$ and $\{B^\alpha u\}_{\alpha \in \mathbf{A}}$ in (3.3), (3.5) and (3.6), and simply denote them by $F(\mathbf{x}, u, Du, D^2 u) = 0$ in the sequel.

4.1 Approximation by Non-singular Measures and Finite Control Sets

In this section, we shall study the approximations of HJBVI (2.5) with a non-singular measure and finite control set.

In fact, it is not difficult to see that (3.3) and (3.5) are consistent approximations of (2.5) in the viscosity sense (see e.g. [22,29]), such that the comparison principle of (2.5) enables us to conclude that the solutions of (3.3) and (3.5) converge to that of (2.5) on compact sets as $\varepsilon, \delta \to 0$.

The remainder of this section thus focuses on obtaining an error estimate for these approximations in terms of the jump truncation size ε and control mesh size δ. Although it should be possible to extend the analytic arguments in Reference [22] to the present nonlinear setting, we follow a different path by first identifying the viscosity solution of (3.5) as the value function of a mixed control problem in terms of modified SDE and BSDE. This control-theoretical interpretation further enables us to develop a shorter proof of the convergence order of these approximations through a probabilistic argument.

We start with the truncation of singular measures. A possible way to work with a nonsingular jump measure is to introduce a Backward SDE with a modified driver and an approximative jump-diffusion dynamics where the small jumps part has been substituted by a rescaled diffusion coefficient of the Brownian motion W.

More precisely, we adopt the same probability space as introduced in Sect. 2, which supports the Brownian motion process W and the independent Poisson measure $N(dt, de)$. For a given jump truncation size $\varepsilon > 0$, we define a modified diffusion coefficient $\tilde{\sigma}^\alpha$ as in (3.1), and also introduce a modified driver $f^\varepsilon(\alpha, t, x, y, z, k) := \hat{f}(\alpha, t, x, y, z, \int_{|e| \geq \varepsilon} k(e)\gamma(x, e)\nu(de))$, where $\hat{f}$ is given in Assumption 2.1.

Next we shall modify the coefficients by including the control discretization. For any given control mesh size $\delta > 0$, we denote by ϕ^δ the piecewise constant approximation (on the control variable α) of a generic function ϕ based on its value on $\mathbf{A}_\delta$, where $\phi = b, \tilde{\sigma}, \eta, f^\varepsilon$. Note that the Lipschitz continuity of the coefficients on the control parameter α (see Assumption 2.1) and the condition (3.4) imply that there exists a constant $C \geq 0$, such that it holds for any given $(\alpha, t, x, u, p, k) \in \mathbf{A} \times [0, T] \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d \times \mathbb{R}$ and discretization parameters $\varepsilon, \delta > 0$ that

$$|b(\alpha, x) - b^\delta(\alpha, x)| + |\sigma(\alpha, x) - \tilde{\sigma}^\delta(\alpha, x)| \leq C\left(\delta + \left|\int_{|e| < \varepsilon} (1 \wedge |e|^2)\, \nu(de)\right|^{\frac{1}{2}}\right),$$
(4.1)

$$|\eta(\alpha, x, e) - \eta^\delta(\alpha, x, e)| \leq C\delta(1 \wedge |e|),$$
(4.2)

$$|f(\alpha, t, x, u, p, k) - f^{\varepsilon, \delta}(\alpha, t, x, u, p, k)| \le C\left(\delta + \int_{|e|<\varepsilon} k(e)\gamma(x, e)\,\nu(de)\right). \tag{4.3}$$

For any given initial state $x \in \mathbb{R}^d$, control $\alpha \in \mathcal{A}_t^t$ and $\tau \in \mathcal{T}_t^t$, we consider the modified controlled jump-diffusion process $(X_s^{\varepsilon, \delta, \alpha, t, x})_{t \le s \le T}$ satisfying the following SDE: for each $s \in [t, T]$,

$$
\begin{aligned}
X_s = x &+ \int_t^s b^{\delta}(\alpha_v, X_v)\,dv + \int_t^s \tilde{\sigma}^{\delta}(\alpha_v, X_v)\,dW_v^t \\
&+ \int_t^s \int_{|e|>\varepsilon} \eta^{\delta}(\alpha_v, X_v, e)\,\tilde{N}^t(dv, de),
\end{aligned} \tag{4.4}
$$

and the BSDE with the modified controlled driver $f^{\varepsilon, \delta}(\alpha_s, s, X_s^{\varepsilon, \delta, \alpha, t, x}, y, z, k)$:

$$
\begin{cases}
-dY_{s, \tau}^{\varepsilon, \delta, \alpha, t, x} = f^{\varepsilon, \delta}(\alpha_s, s, X_s^{\varepsilon, \delta, \alpha, t, x}, Y_{s, \tau}^{\varepsilon, \delta, \alpha, t, x}, Z_{s, \tau}^{\varepsilon, \delta, \alpha, t, x}, K_{s, \tau}^{\varepsilon, \delta, \alpha, t, x})ds \\
\qquad\qquad - Z_{s, \tau}^{\varepsilon, \delta, \alpha, t, x}dW_s^t - \int_E K_{s, \tau}^{\varepsilon, \delta, \alpha, t, x}(e)\,\tilde{N}^t(ds, de), \qquad s \in [t, \tau); \\
Y_{\tau, \tau}^{\varepsilon, \alpha, t, x} = \xi(\tau, X_\tau^{\varepsilon, \alpha, t, x}).
\end{cases} \tag{4.5}
$$

The coefficients of the above SDE and BSDE satisfy Assumption 2.1, and therefore the equations are well-posed.

Now for any each time $t \in [0, T]$ and state $x \in \mathbb{R}^d$, we consider the following value function:

$$u^{\varepsilon, \delta}(t, x) = \sup_{\tau \in \mathcal{T}_t^t} \sup_{\alpha \in \mathcal{A}_t^t} \mathcal{E}_{t, \tau}^{f^{\varepsilon, \delta, \alpha}}[\xi(\tau, X_\tau^{\varepsilon, \delta, \alpha, t, x})], \tag{4.6}$$

subject to the controlled SDE (4.4), where the nonlinear expectation is induced by (4.5).

The following theorem shows that $u^{\varepsilon, \delta}$ is the unique viscosity solution of the HJBVI equation (3.5) introduced in Sect. 3.

Proposition 4.1 *The function $(t, x) \mapsto u^{\varepsilon, \delta}(T - t, x)$, with $u^{\varepsilon, \delta}$ defined by (4.6), is the unique viscosity solution of the HJBVI (3.5).*

Proof Due to the compactness of the set **A**, one can follow the same arguments in Reference [16,17] and characterize $u^{\varepsilon, \delta}$ as the unique viscosity solution of an HJBVI with coefficients b^{δ}, $\tilde{\sigma}^{\delta}$, η^{δ}, $f^{\varepsilon, \delta}$. Then it suffices to observe that such HJBVI is equivalent to (3.5) since its coefficients are piecewise linear in the control parameter. □

Remark 4 Contrary to the case without control and optimal stopping studied in Reference [15], it is not clear that one can use a different approximation of the forward backward system by introducing an independent Brownian motion scaled with the

standard deviation of small jumps. Indeed, the equations would be well-posed in an enlarged filtration $\mathbb{G}$, but the control process is $\mathbb{F}$-predictable, with $\mathbb{F} \subset \mathbb{G}$, which leads to difficulties in the derivation of the dynamic programming principle.

We now exploit the above control-theoretical characterization of the viscosity solution $u^{\varepsilon,\delta}$ and obtain a convergence order in terms of the jump truncation size ε and control mesh size δ. Let us first show the following uniform convergence result of the forward component $X^{\varepsilon,\delta,\alpha,t,x}$ towards $X^{\alpha,t,x}$ when ε, δ tends to 0.

Lemma 4.2 *For each $\varepsilon, \delta \in (0, 1)$, $p \geq 2$, $t \in [0, T]$, $x \in \mathbb{R}^d$ and $\alpha \in \mathcal{A}_t^t$, it holds that*

$$
\mathbb{E}\left[\sup_{t \leq s \leq T} |X_s^{\varepsilon,\delta,\alpha,t,x} - X_s^{\alpha,t,x}|^p\right]
$$
$$
\leq C\left\{\delta^p + \left(\int_{|e| \leq \varepsilon} |e|^2 \nu(de)\right)^{p/2} + \left(\int_{|e| \leq \varepsilon} |e|^p \nu(de)\right)\right\}, \qquad (4.7)
$$

where C is a constant independent of t, x, α, ε and δ.

Proof Fix $\varepsilon, \delta \in (0, 1)$, $\alpha \in \mathcal{A}_t^t$ and $v \in [t, T]$. We have:

$$
\mathbb{E}\left[\sup_{t \leq u \leq v} |X_u^{\varepsilon,\delta,\alpha,t,x} - X_u^{\alpha,t,x}|^p\right]
$$
$$
\leq C\mathbb{E}\left[\sup_{t \leq u \leq v}\left|\int_t^u (b^\delta(\alpha_s, X_s^{\varepsilon,\delta,\alpha,t,x}) - b(\alpha_s, X_s^{\alpha,t,x}))ds\right|^p\right]
$$
$$
+ C\mathbb{E}\left[\sup_{t \leq u \leq v}\left|\int_t^u (\tilde{\sigma}^\delta(\alpha_s, X_s^{\varepsilon,\delta,\alpha,t,x}) - \sigma(\alpha_s, X_s^{\alpha,t,x}))dW_s\right|^p\right]
$$
$$
+ C\mathbb{E}\left[\sup_{t \leq u \leq v}\left|\int_t^u \int_{|e|>\varepsilon} (\eta^\delta(\alpha_s, X_s^{\varepsilon,\delta,\alpha,t,x}, e) - \eta(\alpha_s, X_s^{\alpha,t,x}, e))\tilde{N}(ds, de)\right|^p\right]
$$
$$
+ C\mathbb{E}\left[\sup_{t \leq u \leq v}\left|\int_t^u \int_{|e| \leq \varepsilon} (\eta(\alpha_s, X_s^{\alpha,t,x}, e))\tilde{N}(ds, de)\right|^p\right],
$$

where C is a constant independent of α. The Burkholder–Davis–Gundy inequality, together with (4.1), (4.2), and the Lipschitz assumptions on the coefficients b, σ, η (see Assumption 2.1) lead to:

$$
\mathbb{E}\left[\sup_{t \leq u \leq v} |X_u^{\varepsilon,\delta,\alpha,t,x} - X_u^{\alpha,t,x}|^p\right]
$$
$$
\leq C\left\{\delta^p + \left(\int_{|e| \leq \varepsilon} (1 \wedge |e|^2)\nu(de)\right)^{\frac{p}{2}}\right.
$$
$$
\left. + \mathbb{E}\left[\int_t^v \left(\sup_{t \leq u \leq s} |X_u^{\varepsilon,\delta,\alpha,t,x} - X_u^{\alpha,t,x}|^p\right)ds\right]\right\}
$$

$$
+ C\mathbb{E}\left[\left(\int_t^v \int_{|e|>\varepsilon} |\eta^\delta(\alpha_s, X_s^{\varepsilon,\delta,\alpha,t,x}, e) - \eta(\alpha_s, X_s^{\alpha,t,x}, e)|^2 \nu(de)ds\right)^{\frac{p}{2}}\right]
$$

$$
+ C\mathbb{E}\left[\int_t^v \int_{|e|>\varepsilon} |\eta^\delta(\alpha_s, X_s^{\varepsilon,\delta,\alpha,t,x}, e) - \eta(\alpha_s, X_s^{\alpha,t,x}, e)|^p \nu(de)ds\right]
$$

$$
+ C\mathbb{E}\left[\left(\int_t^v \int_{|e|\leq\varepsilon} |\eta(\alpha_s, X_s^{\alpha,t,x}, e)|^2 \nu(de)ds\right)^{\frac{p}{2}}\right.
$$

$$
+ \left.\left(\int_t^v \int_{|e|\leq\varepsilon} |\eta(\alpha_s, X_s^{\alpha,t,x}, e)|^p \nu(de)ds\right)\right]
$$

$$
\leq C\left\{\mathbb{E}\left[\int_t^v \left(\sup_{t\leq u\leq s} |X_u^{\varepsilon,\delta,\alpha,t,x} - X_u^{\alpha,t,x}|^p\right)ds\right]\right.
$$

$$
+ \delta^p + \left.\left(\int_{|e|\leq\varepsilon} |e|^2 \nu(de)\right)^{\frac{p}{2}} + \left(\int_{|e|\leq\varepsilon} |e|^p \nu(de)\right)\right\},
$$

where the last inequality follows by the integrability assumption on the measure ν. Then we obtain the desired result (4.7) from the Gronwall's inequality.   □

Based on the above estimate, we now show the convergence order of the viscosity function $u^{\varepsilon,\delta}$ of (3.5) towards u in terms of ε and δ.

Theorem 4.3 *For any $p > 2$, there exists a constant C_p depending on p, such that for all $\varepsilon, \delta \in (0, 1)$, we have*

$$
\left|u^{\varepsilon,\delta}(t, x) - u(t, x)\right| \leq C_p \left\{\delta + \left(\int_{|e|\leq\varepsilon} |e|^2 \nu(de)\right)^{\frac{1}{2}}\right.
$$

$$
+ \left.\left(\int_{|e|\leq\varepsilon} |e|^p \nu(de)\right)^{\frac{1}{p}}\right\}, \quad (t, x) \in [0, T] \in \mathbb{R}^d.
$$

Proof Fix $\varepsilon, \delta \in (0, 1)$, $t \in [0, T]$ and $x \in \mathbb{R}^d$. The definitions of $u^{\varepsilon,\delta}$ and u imply that

$$
|u^{\varepsilon,\delta}(t, x) - u(t, x)|^2 = \Big| \sup_{\alpha\in\mathcal{A}_t^t} \sup_{\tau\in\mathcal{T}_t^t} \mathcal{E}_{t,\tau}^{f^{\varepsilon,\delta,\alpha}}\left[\zeta(\tau, X_\tau^{\varepsilon,\delta,\alpha,t,x})\right]
$$

$$
- \sup_{\alpha\in\mathcal{A}_t^t} \sup_{\tau\in\mathcal{T}_t^t} \mathcal{E}_{t,\tau}^{f^\alpha}\left[\zeta(\tau, X_\tau^{\alpha,t,x})\right] \Big|^2
$$

$$
\leq \sup_{\alpha\in\mathcal{A}_t^t} \sup_{\tau\in\mathcal{T}_t^t} \left|\mathcal{E}_{t,\tau}^{f^{\varepsilon,\delta,\alpha}}\left[\zeta(\tau, X_\tau^{\varepsilon,\delta,\alpha,t,x})\right] - \mathcal{E}_{t,\tau}^{f^\alpha}\left[\zeta(\tau, X_\tau^{\alpha,t,x})\right]\right|^2. \quad (4.8)
$$

Recall that, since $\alpha \in \mathcal{A}_t^t$ and $\tau \in \mathcal{T}_t^t$, $\left|\mathcal{E}_{t,\tau}^{f^{\varepsilon,\delta,\alpha}}\left[\zeta(\tau, X_\tau^{\varepsilon,\delta,\alpha,t,x})\right] - \mathcal{E}_{t,\tau}^{f^\alpha}\left[\zeta(\tau, X_\tau^{\alpha,t,x})\right]\right|$ is deterministic. By the a priori estimates on the spread between the first component of the solutions of two BSDEs with jumps (see Proposition A.4. in Reference [32]),

we derive from (4.3) that there exist $\beta > 0$ and $\eta > 0$ independent on $\tau \in \mathcal{T}_t^t$ and $\alpha \in \mathcal{A}_t^t$, such that

$$
\begin{aligned}
\Big| \mathcal{E}_{t,\tau}^{f^{\varepsilon,\delta,\alpha}} & \left[\zeta(\tau, X_\tau^{\varepsilon,\delta,\alpha,t,x}) \right] - \mathcal{E}_{t,\tau}^{f^\alpha} \left[\zeta(\tau, X_\tau^{\alpha,t,x}) \right] \Big|^2 \\
& \leq \mathbb{E}\left[e^{\beta(\tau-t)} \left(\zeta(\tau, X_\tau^{\varepsilon,\delta,\alpha,t,x}) - \zeta(\tau, X_\tau^{\alpha,t,x}) \right)^2 \right] \\
& \quad + \eta \mathbb{E}\left[\int_t^\tau e^{\beta(s-t)} \left(f(s, \alpha_s, X_s^{\alpha,t,x}, Y_{s,\tau}^{\alpha,t,x}, Z_{s,\tau}^{\alpha,t,x}, K_{s,\tau}^{\alpha,t,x}) \right. \right. \\
& \quad \left. \left. - f^{\varepsilon,\delta}(s, \alpha_s, X_s^{\varepsilon,\delta,\alpha,t,x}, Y_{s,\tau}^{\alpha,t,x}, Z_{s,\tau}^{\alpha,t,x}, K_{s,\tau}^{\alpha,t,x}) \right)^2 ds \right] \\
& \leq C\left\{ \mathbb{E}\left[\sup_{t \leq u \leq T} |X_u^{\varepsilon,\delta,\alpha,t,x} - X_u^{\alpha,t,x}|^2 \right] + \delta^2 \right. \\
& \quad + \mathbb{E}\left[\int_t^\tau \left(\int_{|e| \leq \varepsilon} K_{s,\tau}^{\alpha,t,x}(e) \gamma(X_s^{\alpha,t,x}, e) \nu(de) \right)^2 ds \right] \\
& \quad \left. + \mathbb{E}\left[\int_t^\tau \left(\int_{|e| > \varepsilon} K_{s,\tau}^{\alpha,t,x}(e) (\gamma(X_s^{\alpha,t,x}, e) - \gamma(X_s^{\varepsilon,\delta,\alpha,t,x}, e)) \nu(de) \right)^2 ds \right] \right\},
\end{aligned}
$$

where C is a constant independent on $t, x, \varepsilon, \delta, \alpha, \tau$, only depending on β, η, T and the Lipschitz constant of f.

Now we estimate the last two terms in the above inequality. For any given $p \geq 2$, the uniform boundedness of ζ, g and f with respect to t, x, α and τ (see Assumption 2.1), together with the a priori estimates for L^p solutions of BSDEs (see Proposition 2 in Reference [27]) gives us an uniform control on the $\mathbb{H}_{t,\nu}^p$ norm of $K_{\cdot,\tau}^{\alpha,t,x}$ (which only depends on p and the bounds of ζ, g, f and T). Using this result and the boundedness of the map γ (see Assumption 2.1), we derive from Hölder's inequality that there exists a constant C independent on τ and α such that

$$
\begin{aligned}
\mathbb{E}\left[\int_t^\tau \left(\int_{|e| \leq \varepsilon} K_{s,\tau}^{\alpha,t,x}(e) \gamma(X_s^{\alpha,t,x}, e) \nu(de) \right)^2 ds \right] & \\
\leq \mathbb{E}\left[\int_t^\tau \left(\int_{|e| \leq \varepsilon} (K_{s,\tau}^{\alpha,t,x})^2(e) \nu(de) \right) \left(\int_{|e| \leq \varepsilon} \gamma^2(X_s^{\alpha,t,x}, e) \nu(de) \right) ds \right] & \\
\leq C \left(\int_{|e| \leq \varepsilon} |e|^2 \nu(de) \right). &
\end{aligned}
$$

Furthermore, for any given $p > 1$, by using the Lipschitz continuity of the map γ, we can obtain

$$
\begin{aligned}
\mathbb{E}\left[\int_t^\tau \left(\int_{|e| > \varepsilon} K_{s,\tau}^{\alpha,t,x}(e) (\gamma(X_s^{\alpha,t,x}, e) - \gamma(X_s^{\varepsilon,\delta,\alpha,t,x}, e)) \nu(de) \right)^2 ds \right] & \\
\leq C \mathbb{E}\left[\int_t^\tau \left(\int_{|e| > \varepsilon} |K_{s,\tau}^{\alpha,t,x}(e)| |X_s^{\alpha,t,x} - X_s^{\varepsilon,\delta,\alpha,t,x}| (1 \wedge |e|^2) \nu(de) \right)^2 ds \right] &
\end{aligned}
$$

$$\leq C\mathbb{E}\left[\sup_{t\leq u\leq T}|X_u^{\alpha,t,x}-X_u^{\varepsilon,\delta,\alpha,t,x}|^2\int_t^\tau\left(\int_E|K_{s,\tau}^{\alpha,t,x}(e)|(1\wedge|e|^2)\nu(de)\right)^2ds\right]$$

$$\leq C\mathbb{E}\left[\sup_{t\leq u\leq T}|X_u^{\alpha,t,x}-X_u^{\varepsilon,\delta,\alpha,t,x}|^2\left(\int_t^\tau\int_E|K_{s,\tau}^{\alpha,t,x}(e)|^2\nu(de)ds\right)\left(\int_E(1\wedge|e|^4)\nu(de)\right)\right]$$

$$\leq C_p\left(\mathbb{E}\left[\sup_{t\leq u\leq T}|X_u^{\alpha,t,x}-X_u^{\varepsilon,\delta,\alpha,t,x}|^{2p}\right]\right)^{1/p}, \tag{4.9}$$

where we have used Hölder's inequality and the boundedness of the $\mathbb{H}_{t,\nu}^{2p/(p-1)}$ norm of $K_{\cdot,\tau}^{\alpha,t,x}$ for the last line. Consequently, by summarizing all the above estimates and using Lemma 4.2, we can obtain that

$$\left|\mathcal{E}_{t,\tau}^{f^{\varepsilon,\delta,\alpha}}\left[\zeta(\tau,X_\tau^{\varepsilon,\delta,\alpha,t,x})\right]-\mathcal{E}_{t,\tau}^{f^\alpha}\left[\zeta(\tau,X_\tau^{\alpha,t,x})\right]\right|^2$$

$$\leq C\left\{\mathbb{E}\left[\sup_{t\leq u\leq T}|X_u^{\varepsilon,\delta,\alpha,t,x}-X_u^{\alpha,t,x}|^2\right]+\delta^2\right.$$

$$\left.+\int_{|e|\leq\varepsilon}|e|^2\nu(de)+\left(\mathbb{E}\left[\sup_{t\leq u\leq T}|X_u^{\varepsilon,\delta,\alpha,t,x}-X_u^{\alpha,t,x}|^{2p}\right]\right)^{\frac{1}{p}}\right\}$$

$$\leq C\left\{\delta^2+\int_{|e|\leq\varepsilon}|e|^2\nu(de)+\left[\delta^{2p}+\left(\int_{|e|\leq\varepsilon}|e|^2\nu(de)\right)^p\right.\right.$$

$$\left.\left.+\left(\int_{|e|\leq\varepsilon}|e|^{2p}\nu(de)\right)\right]^{\frac{1}{p}}\right\}$$

$$\leq C\left\{\delta^2+\left(\int_{|e|\leq\varepsilon}|e|^2\nu(de)\right)+\left(\int_{|e|\leq\varepsilon}|e|^{2p}\nu(de)\right)^{\frac{1}{p}}\right\},$$

from which we can conclude the desired estimate by taking the supremum over α and τ.   □

Theorem 4.3 extends the continuous dependence result for classical nonlocal HJB-VIs in Reference [22, Theorem 4.4] to the HJBVIs with nonlinear drivers and *state-dependent measures* (the operator B^α defined by (2.8) involves the measure $\gamma(x,e)\nu(de)$, which depends on the spatial variable x).

Due to the presence of the state-dependent measure, in particular the term (4.9), our error estimate has an additional term $\left(\int_{|e|\leq\varepsilon}|e|^p\nu(de)\right)^{1/p}$, which in general cannot be compared to the term $\left(\int_{|e|\leq\varepsilon}|e|^2\nu(de)\right)^{1/2}$ without further information. For example, if ν is finite around zero, then for small enough ε, the fact that $p>2$ and Jensen's inequality lead to the following estimate:

$$\left(\int_{|e|\leq\varepsilon}|e|^2\nu(de)\right)^{\frac{p}{2}}=\nu(B_\varepsilon)^{\frac{p}{2}}\left(\int_{|e|\leq\varepsilon}|e|^2\frac{\nu(de)}{\nu(B_\varepsilon)}\right)^{\frac{p}{2}}$$

$$\leq\nu(B_\varepsilon)^{\frac{p}{2}-1}\left(\int_{|e|\leq\varepsilon}|e|^p\nu(de)\right)\leq\int_{|e|\leq\varepsilon}|e|^p\nu(de),$$

where we denote $\nu(B_\varepsilon) = \nu(\{e \in E \mid 0 < |e| \leq \varepsilon\})$. On the other hand, if we assume that the singular measure ν behaves similar to the Lévy measures of α-stable processes around zero, in the sense that ν admits a density $\rho(e)$ such that it holds for some constants $C > 0$ and $\kappa \in [0, 2)$ that

$$0 \leq \rho(e) \leq C|e|^{-n-\kappa}, \quad |e| < 1, \, e \in E = \mathbb{R}^n \setminus \{0\},$$

then a direct computation shows $\left(\int_{|e| \leq \varepsilon} |e|^p \nu(de) \right)^{1/p} = \mathcal{O}(\varepsilon^{(p-\kappa)/p})$, $p \geq 2$, which implies the jump truncation error is dominated by the term $\left(\int_{|e| \leq \varepsilon} |e|^2 \nu(de) \right)^{1/2}$, and consequently we recover the same convergence rate as that for the classical setting.

4.2 Approximation by Switching Systems

In this section, we study the approximation of (3.5) by switching systems. We adopt the following standard definition of a viscosity solution to switching systems of the form (3.6) (see [1,6,29] and references therein).

Definition 4.4 (*Viscosity solution of switching system*) A $\mathbb{R}^J$-valued upper (resp. lower) semicontinuous function U is said to be a viscosity subsolution (resp. supersolution) of (3.6) if and only if for any point $\mathbf{x}_0$ and for any $\phi \in C^{1,2}(\bar{\mathcal{Q}}_T)$ such that $U_j - \phi$ attains its global maximum (resp. minimum) at $\mathbf{x}_0$, one has

$$\begin{aligned} &F_{j*}^{\varepsilon,\delta,c}(\mathbf{x}_0, U_j(\mathbf{x}_0), D\phi(\mathbf{x}_0), D^2\phi(\mathbf{x}_0), \{K_\varepsilon^\alpha \phi(\mathbf{x}_0)\}_{\alpha \in \mathbf{A}_\delta}, \\ &\quad \{B_\varepsilon^\alpha \phi(\mathbf{x}_0)\}_{\alpha \in \mathbf{A}_\delta}, \{U_k(\mathbf{x}_0)\}_{k \neq j}) \leq 0 \\ &\left(resp. \quad F_j^{\varepsilon,\delta,c*}(\mathbf{x}_0, U_j(\mathbf{x}_0), D\phi(\mathbf{x}_0), D^2\phi(\mathbf{x}_0), \{K_\varepsilon^\alpha \phi(\mathbf{x}_0)\}_{\alpha \in \mathbf{A}_\delta}, \right. \\ &\quad \left. \{B_\varepsilon^\alpha \phi(\mathbf{x}_0)\}_{\alpha \in \mathbf{A}_\delta}, \{U_k(\mathbf{x}_0)\}_{k \neq j}) \geq 0 \right). \end{aligned}$$

A continuous function is a viscosity solution of the HJBVI (3.6) if it is both a a viscosity sub- and supersolution.

Note that in the definition of the viscosity solution of F_j, the test function only replaces U_j in the integrals and derivatives, while leaving the terms $\{U_k\}_{k \neq j}$ unchanged.

Now we present the comparison principle for bounded semicontinuous viscosity solutions of (3.6), which not only implies the uniqueness of bounded viscosity solutions of (3.6), but is also essential for our convergence analysis. The proof will be given in Appendix 1.

Theorem 4.5 *Let* $U = (U_1, U_2, ..., U_J)$ *and* $V = (V_1, V_2, ..., V_J)$ *be bounded viscosity sub- and supersolutions, respectively, of* (3.6) *with* $U(0, \cdot) \leq V(0, \cdot)$. *Then it holds under Assumption* 2.1 *that* $U_j(\mathbf{x}) \leq V_j(\mathbf{x})$ *for all* $j = 1, \ldots, J$.

The following theorem demonstrates the convergence of the switching system to the finite control HJBVI (3.5) as the switching cost goes to 0. Convergence with order 1/3 is proved in Reference [6] by a different technique, for nonlocal Bellman equations without obstacles and nonlinear source terms.

We momentarily assume the switching system (3.6) to admit a viscosity solution bounded independently of the (small enough) switching cost c. We give a constructive proof of existence through our numerical schemes in Sect. 4.3.

Theorem 4.6 *Under Assumption 2.1, let $U^{\varepsilon,\delta,c} = (U_1^{\varepsilon,\delta,c}, \ldots, U_J^{\varepsilon,\delta,c})$ and $u^{\varepsilon,\delta}$ be the viscosity solution of (3.6) and (3.5), respectively. Then for fixed $\varepsilon, \delta > 0$, we have for each $j = 1, \ldots, J$ that $U_j^{\varepsilon,\delta,c} \to u^{\varepsilon,\delta}$ uniformly on compact sets as $c \to 0$.*

Proof Since ε and δ are fixed for our analysis, we shall omit the dependence on ε and δ, and simply denote by U^c the solution of (3.6). Consider a sequence of switching costs $c_m \to 0$ as $m \to \infty$, and the corresponding viscosity solution $U^{c_m} = (U_1^{c_m}, \ldots, U_J^{c_m})$. We shall first prove by contradiction that

$$U_j^{c_m}(\mathbf{x}) \geq \mathcal{M}_j U^{c_m}, \quad \mathbf{x} \in \mathcal{Q}_T, \ j = 1, \ldots, J. \tag{4.10}$$

Suppose the statement is false, then there would exist $k \neq j$ and $\mathbf{x}_0 \in \mathcal{Q}_T$ such that $U_j^{c_m}(\mathbf{x}_0) < U_k^{c_m}(\mathbf{x}_0) - c_m$. We then obtain from the continuity of $U_j^{c_m}$ and $U_k^{c_m}$ that there exists a nonempty open ball B around $\mathbf{x}_0$ such that

$$U_j^{c_m}(\mathbf{x}) < U_k^{c_m}(\mathbf{x}) - c_m, \quad \mathbf{x} \in B.$$

On the other hand, due to the fact that semi-jets are nonempty on a dense set (see e.g. [25, Lemma 8 on pp. 23]), there exists a C^2 function ϕ such that $U_j^{c_m} - \phi$ attains its minimum at some point in B, say $\mathbf{x}_1$. Hence we deduce from the fact that $U_j^{c_m}$ is a supersolution that

$$U_j^{c_m}(\mathbf{x}_1) \geq \mathcal{M}_j U^{c_m}(\mathbf{x}_1) \geq U_k^{c_m}(\mathbf{x}_1) - c_m,$$

which leads to a contradiction.

We now introduce the following functions through a relaxed limit: for $j = 1, \ldots, J$,

$$\overline{U}_j(x) = \lim_{r \to \infty} \sup_{m > r} \sup_{|\mathbf{y}-\mathbf{x}| < 1/r} U_j^{c_m}(\mathbf{y}), \quad \underline{U}_j(x) = \lim_{r \to \infty} \inf_{m > r} \inf_{|\mathbf{y}-\mathbf{x}| < 1/r} U_j^{c_m}(\mathbf{y}). \tag{4.11}$$

It is not hard to check $\overline{U}_1 = \ldots = \overline{U}_J \equiv \overline{U}$ and $\underline{U}_1 = \ldots = \underline{U}_J \equiv \underline{U}$. In fact, for any given $j, k \in \{1, \ldots, J\}$, $j \neq k$, $\mathbf{x} \in \mathcal{Q}_T$, and $m, r \in \mathbb{N}$, we obtain from (4.10) that $U_j^{c_m}(\mathbf{y}) \geq U_k^{c_m}(\mathbf{y}) - c_m$ for $\mathbf{y} \in \mathcal{Q}_T$, and hence

$$\sup_{m > r} \sup_{|\mathbf{y}-\mathbf{x}| < 1/r} U_j^{c_m}(\mathbf{y}) \geq \sup_{m > r} \sup_{|\mathbf{y}-\mathbf{x}| < 1/r} U_k^{c_m}(\mathbf{y}) - \sup_{m > r} c_m.$$

Letting $r \to \infty$ leads to the fact that $\overline{U}_j \geq \overline{U}_k$ for all $j \neq k$. The statement for $\{\underline{U}_j\}$ can be shown similarly.

Since it is clear that $\overline{U}$ and $\underline{U}$ is bounded upper and lower semicontinuous, respectively, we now aim to show $\overline{U}$ and $\underline{U}$ is respectively a sub- and supersolution of (3.5). Then the strong comparison principle gives us $\overline{U} \leq \underline{U}$, which implies $U = \overline{U} = \underline{U}$ is

the unique viscosity solution of (3.5). Uniform convergence on compact sets follows from a variation of Dini's theorem (See Remark 6.4 in Reference [12]).

We start by showing $\overline{U}$ is a subsolution of (3.5). Let $\phi \in C^{1,2}$ and $\overline{U} - \phi$ have a strict global maximum at $\hat{\mathbf{x}}_0 \in \bar{\mathcal{Q}}_T$, then there will be a sequence $c_m \to 0$ such that for each $j \in \{1, \ldots, J\}$, we have $\hat{\mathbf{x}}_m^j \to \hat{\mathbf{x}}_0$, $U_j^{c_m}(\hat{\mathbf{x}}_m^j) \to \overline{U}(\hat{\mathbf{x}}_0)$, and $U_j^{c_m} - \phi$ attains a global maximum at $\hat{\mathbf{x}}_m^j$. Since $U_j^{c_m}$ is a subsolution of (3.6) with c_m, if we have $\hat{\mathbf{x}}_0 \in \{0\} \times \mathbb{R}^d$, $U_j^{c_m}(\hat{\mathbf{x}}_m^j) \leq g(\hat{\mathbf{x}}_m^j)$ for infinitely many m and a fixed j, then it is clear that $\overline{U}(\hat{\mathbf{x}}_0) \leq g(\hat{\mathbf{x}}_0)$. Therefore, without loss of generality, we assume for all m and j that

$$\min \left[U_j^{c_m}(\hat{\mathbf{x}}_m^j) - \zeta(\hat{\mathbf{x}}_m^j), \ \min \left(\phi_t(\hat{\mathbf{x}}_m^j) - L_\varepsilon^{\alpha_j} \phi(\hat{\mathbf{x}}_m^j) \right. \right.$$
$$- f^{\alpha_j}(\hat{\mathbf{x}}_m^j, U_j^{c_m}(\hat{\mathbf{x}}_m^j), \tilde{\sigma}^{\alpha_j} \cdot D\phi(\hat{\mathbf{x}}_m^j), B_\varepsilon^{\alpha_j} \phi(\hat{\mathbf{x}}_m^j));$$
$$\left. \left. U_j^{c_m}(\hat{\mathbf{x}}_m^j) - \mathcal{M}_j U^{c_m}(\hat{\mathbf{x}}_m^j) \right) \right] \leq 0. \tag{4.12}$$

We have two cases. If there exists $j \in \{1, \ldots, J\}$ and a subsequence of c_m such that $U_j^{c_m}(\hat{\mathbf{x}}_m^j) - \zeta(\hat{\mathbf{x}}_m^j) \leq 0$, then by passing to the limit $m \to \infty$, we have $\overline{U}(\hat{\mathbf{x}}_0) - \zeta(\hat{\mathbf{x}}_0) \leq 0$. Otherwise, by passing to subsequence, without loss of generality we can assume $U_j^{c_m}(\hat{\mathbf{x}}_m^j) - \zeta(\hat{\mathbf{x}}_m^j) > 0$ holds for all j and m. Then for each $m \in \mathbb{N}$, we can choose $j_m \in \{1, \ldots, J\}$ and $\hat{\mathbf{x}}_m^{j_m}$ such that

$$(U_{j_m}^{c_m} - \phi)(\hat{\mathbf{x}}_m^{j_m}) = \max_{j=1,\ldots,J} (U_j^{c_m} - \phi)(\hat{\mathbf{x}}_m^j) = \max_{j=1,\ldots,J} \max_{\mathbf{x}} (U_j^{c_m} - \phi)(\mathbf{x}),$$

and deduce from (4.12) that

$$\min \left(\phi_t(\hat{\mathbf{x}}_m^{j_m}) - L_\varepsilon^{\alpha_{j_m}} \phi(\hat{\mathbf{x}}_m) - f^{\alpha_{j_m}}(\hat{\mathbf{x}}_m^{j_m}, U_{j_m}^{c_m}(\hat{\mathbf{x}}_m^{j_m}), \tilde{\sigma}^{\alpha_{j_m}} \cdot D\phi(\hat{\mathbf{x}}_m^{j_m}), B_\varepsilon^{\alpha_{j_m}} \phi(\hat{\mathbf{x}}_m^{j_m})); \right.$$
$$\left. U_{j_m}^{c_m}(\hat{\mathbf{x}}_m^{j_m}) - \mathcal{M}_{j_m} U^{c_m}(\hat{\mathbf{x}}_m^{j_m}) \right) \leq 0. \tag{4.13}$$

Our choice of j_m implies $(U_{j_m}^{c_m} - \phi)(\hat{\mathbf{x}}_m^{j_m}) \geq (U_k^{c_m} - \phi)(\hat{\mathbf{x}}_m^{j_m})$ for all $k \neq j_m$, and thus $U_{j_m}^{c_m}(\hat{\mathbf{x}}_m^{j_m}) > \mathcal{M}_{j_m} U^{c_m}(\hat{\mathbf{x}}_m^{j_m})$. Consequently we obtain from (4.13) that

$$\phi_t(\hat{\mathbf{x}}_m^{j_m}) - L_\varepsilon^{\alpha_{j_m}} \phi(\hat{\mathbf{x}}_m) - f^{\alpha_{j_m}}(\hat{\mathbf{x}}_m^{j_m}, U_{j_m}^{c_m}(\hat{\mathbf{x}}_m^{j_m}), \tilde{\sigma}^{\alpha_{j_m}} \cdot D\phi(\hat{\mathbf{x}}_m^{j_m}), B_\varepsilon^{\alpha_{j_m}} \phi(\hat{\mathbf{x}}_m^{j_m})) \leq 0.$$

Since we only have finite many choices of j_m, by passing to a further subsequence if necessary, we can assume that $j_m \to j_0$, then letting $m \to \infty$ and using the continuity of the equation, we have

$$\phi_t(\hat{\mathbf{x}}_0) - L_\varepsilon^{\alpha_{j_0}} \phi(\hat{\mathbf{x}}_0) - f^{\alpha_{j_0}}(\hat{\mathbf{x}}_0, \overline{U}(\hat{\mathbf{x}}_0), \tilde{\sigma}^{\alpha_{j_0}} \cdot D\phi(\hat{\mathbf{x}}_0), B_\varepsilon^{\alpha_{j_0}} \phi(\hat{\mathbf{x}}_0)) \leq 0.$$

Since $\alpha_{j_0} \in \mathbf{A}_\delta$ is an admissible control, we obtain

$$\min_{\alpha \in \mathbf{A}_\delta} \left\{ \phi_t(\hat{\mathbf{x}}_0) - L^\alpha_\varepsilon \phi(\hat{\mathbf{x}}_0) - f^{\alpha_j}(\hat{\mathbf{x}}_0, \overline{U}(\hat{\mathbf{x}}_0), \tilde{\sigma}^\alpha \cdot D\phi(\hat{\mathbf{x}}_0), B^\alpha_\varepsilon \phi(\hat{\mathbf{x}}_0)) \right\} \leq 0,$$

and conclude that $\overline{U}$ is a subsolution of (3.6).

We now proceed to show $\underline{U}$ is a supersolution. If $\phi \in C^{1,2}$ and $\overline{U} - \phi$ has a strict global mimimum at $\hat{\mathbf{x}}_0 \in \overline{\mathcal{Q}}_T$, then for any given $j \in \{1, \ldots, J\}$, there will be sequences $c_m \to 0$, $\hat{\mathbf{x}}_m \to \hat{\mathbf{x}}_0$, $U^{c_m}_j(\hat{\mathbf{x}}_m) \to \underline{U}(\hat{\mathbf{x}}_0)$, and $U^{c_m}_j - \phi$ attains a global mimimum at $\hat{\mathbf{x}}_m$. Using the fact that $U^{c_m}_j$ is a supersolution to (3.6), we have (by ignoring the term $U^{c_m}_j(\hat{\mathbf{x}}^j_m) - \mathcal{M}_j U^{c_m}(\hat{\mathbf{x}}^j_m)$):

$$\min \left[U^{c_m}_j(\hat{\mathbf{x}}_m) - \zeta(\hat{\mathbf{x}}_m), \ \phi_t(\hat{\mathbf{x}}_m) - L^{\alpha_j}_\varepsilon \phi(\hat{\mathbf{x}}_m) \right.$$
$$\left. - f^{\alpha_j}(\hat{\mathbf{x}}_m, U^{c_m}_j(\hat{\mathbf{x}}_m), \tilde{\sigma}^{\alpha_j} \cdot D\phi(\hat{\mathbf{x}}_m), B^{\alpha_j}_\varepsilon \phi(\hat{\mathbf{x}}_m)) \right] \geq 0,$$

then passing $m \to \infty$ enables us to conclude for any $j \in \{1, \ldots, J\}$,

$$\min \left[\underline{U}(\hat{\mathbf{x}}_0) - \zeta(\hat{\mathbf{x}}_0), \ \phi_t(\hat{\mathbf{x}}_0) - L^{\alpha_j}_\varepsilon \phi(\hat{\mathbf{x}}_0) \right.$$
$$\left. - f^{\alpha_j}(\hat{\mathbf{x}}_0, \underline{U}(\hat{\mathbf{x}}_0), \tilde{\sigma}^{\alpha_j} \cdot D\phi(\hat{\mathbf{x}}_0), B^{\alpha_j}_\varepsilon \phi(\hat{\mathbf{x}}_0)) \right] \geq 0,$$

which completes our proof. □

4.3 General Discrete Approximation to the Switching System

In this section, we establish the convergence of the piecewise constant policy approximation of (3.10) to the solution of the switching system (3.6). We will first summarize all the required conditions to guarantee the convergence, and perform the analysis under these assumptions. Then we will demonstrate in Sect. 4.4 that these conditions are in fact satisfied by the numerical scheme (3.18) proposed in Sect. 3 .

We assume the scheme (3.10) satisfies the following conditions introduced in Reference [29]:

Condition 1 *(1) (Positive interpolation.) Let $\tilde{U}^n_{k,i(j)}$ be the interpolant of the k-th grid onto the i-th point $\mathbf{x}^n_{j,i}$ of the j-th grid, and $N^k(j, i, n)$ be the neighbours[2] to the point $\mathbf{x}^n_{j,i}$ on the k-th grid $\Omega_{k,h}$. Then there exist weights $\{\omega^n_{k,i(j),a}\}_{a \in N^k(j,i,n)}$ satisfying $\omega^n_{k,i(j),a} \geq 0$ and $\sum_{a \in N^k(j,i,n)} \omega^n_{k,i(j),a} = 1$, such that we can write*

$$\tilde{U}^n_{k,i(j)} = \sum_{a \in N^k(j,i,n)} \omega^n_{k,i(j),a} U^n_{k,a}. \tag{4.14}$$

---

[2] "Neighbours" can be any a set of indices a such that $\mathbf{x}^n_{k,a} \to \mathbf{x}^n_{j,i}$.

(2) (Weak monotonicity.) The scheme (3.10) *is monotone with respect to* $U^n_{j,i}$ *and* $\tilde{U}^n_{k,i(j)}$, *i.e., if*

$$V^n_{j,i} \geq U^n_{j,i}, \quad \forall(i,j,n); \quad \tilde{V}^n_{k,i(j)} \geq \tilde{U}^n_{k,i(j)}, \quad \forall(i,k,n),$$

then we have

$$G_j(\mathbf{x}^n_{j,i}, h, U^{n+1}_{j,i}, \{V^{b+1}_{j,a}\}_{(a,b)\neq(i,n)}, \{\tilde{V}^n_k\}_{k\neq j})$$
$$\leq G_j(\mathbf{x}^n_{j,i}, h, U^{n+1}_{j,i}, \{U^{b+1}_{j,a}\}_{(a,b)\neq(i,n)}, \{\tilde{U}^n_k\}_{k\neq j}). \tag{4.15}$$

(3) (ℓ^∞ stability.) The solution $U^{n+1}_{j,i}$ *of the scheme* (3.10) *exists and is bounded uniformly in h and c.*

(4) (Consistency.) Let ε, δ, c *be fixed. For any test functions* $\phi_j \in C^{1,2}(\bar{\mathcal{Q}}_T)$ *and continuous* φ_k, *there exist function* $\omega_1(h)$ *and* $\omega_2(\xi)$, *possibly depending on* ε, *such that* $\omega_1(h) \to 0$ *as* $h \to 0$, $\omega_2(\xi) \to 0$ *as* $\xi \to 0$, *and*

$$|G_j(\mathbf{x}^{n+1}_{j,i}, h, \phi^{n+1}_{j,i} + \xi, \{\phi^{b+1}_{j,a}\}_{(a,b)\neq(i,n)} + \xi, \{\tilde{\varphi}^n_k\}_{k\neq j})$$
$$- F^{\varepsilon,\delta,c}_j(\mathbf{x}^{n+1}_{j,i}, \phi_j(\mathbf{x}^{n+1}_{j,i}), D\phi_j(\mathbf{x}^{n+1}_{j,i}), D^2\phi_j(\mathbf{x}^{n+1}_{j,i}), \tag{4.16}$$
$$\{\tilde{\varphi}_k(\mathbf{x}^n_{j,i})\}_{k\neq j})| \leq \omega_1(h) + \omega_2(\xi).$$

Remark 5 As pointed out in Reference [29], Condition 1 (1)–(2) are weaker than the standard condition that the scheme is monotone in $U^n_{k,\alpha}$ (see e.g. [3]). By only requiring that the interpolation has positive coefficients and that the numerical scheme is monotone in the interpolant $\tilde{U}^n_{k,\alpha}$, we are allowing the usage of high order nonlinear interpolations among different grids (e.g., the monotonicity preserving interpolations in Reference [19]).

Also note the contrast to the linear interpolant (3.11) used in (3.12) and (3.13) for the construction of a monotone approximation to the integral operators.

We now present the convergence of the discrete approximation to the switching system.

Theorem 4.7 *Under Assumptions* 2.1, *the solution to any scheme of the form* (3.10) *satisfying Condition* 1 *converges to the viscosity solution of* (3.6) *uniformly on bounded domains.*

The proof is essentially the same as that in Reference [29] and is omitted. We remark that in the proof, we construct the solution of the switching system directly from the numerical solutions. Since the solution of the scheme (3.10) is uniformly bounded, Theorems 4.5 and 4.7 immediately give the existence and uniqueness of a bounded viscosity solution to the switching system (3.6).

Corollary 4.8 *Under Assumption* 2.1 *and the existence of a scheme satisfying Condition* 1, *the switching system* (3.6) *admits a unique viscosity solution bounded uniformly in c.*

4.4 A Specific Implicit Scheme for the Switching System

In this section, we analyze the implicit scheme (3.18) and demonstrate that it satisfies Condition 1, which subsequently implies its convergence to the switching system.

The following estimates are essential for our consistency and stability analysis.

Lemma 4.9 *Under Assumption 2.1, there exists C independent of h, k, ε, δ such that for any test functions $\phi_j \in C^{1,2}(\bar{Q}_T)$ and $\varepsilon < 1$ that*

$$|A_{\varepsilon,h,k}^\alpha \phi_{j,i}^{n+1} + K_{\varepsilon,h}^{\alpha,1} \phi_{j,i}^{n+1} - A_\varepsilon^\alpha \phi_j(\mathbf{x}_{j,i}^{n+1})$$

$$- K_\varepsilon^\alpha \phi_j(\mathbf{x}_{j,i}^{n+1})| \leq C \left(\frac{h^2}{k^2} + \frac{h^2}{\varepsilon^2} + \omega(\mathbf{x}_{j,i}^{n+1}, k) \right),$$

$$|B_{\varepsilon,h}^\alpha \phi_{j,i}^{n+1} - B_\varepsilon^\alpha \phi(\mathbf{x}_{j,i}^{n+1})| \leq C \frac{h^2}{\varepsilon}.$$

for some $\omega(\mathbf{x}_{j,i}^{n+1}, k)$ such that $\omega(\cdot, k) \to 0$ as $k \to 0$ uniformly on compact neighbourhoods of $\mathbf{x}_{j,i}^{n+1}$.

Proof We first derive the estimate for $B_{\varepsilon,h}^\alpha \phi_{j,i}^{n+1}$. It follows from $|\eta^\alpha| \leq C$ and the definitions of $B_{\varepsilon,h}^\alpha \phi$ and $B_\varepsilon^\alpha \phi$ that

$$|B_{\varepsilon,h}^\alpha \phi_{j,i}^{n+1} - B_\varepsilon^\alpha \phi(\mathbf{x}_{j,i}^{n+1})|$$

$$\leq \int_{|e| \geq \varepsilon} |\mathcal{I}_h[\phi(t^{n+1}, x_{j,i} + \cdot)](\eta^\alpha(x_{j,i}, e))$$

$$- \phi(t^{n+1}, x_{j,i} + \eta^\alpha(x_{j,i}, e))| \gamma(x_{j,i}, e) \, \nu(de)$$

$$\leq Ch^2 |D^2\phi|_{B(\mathbf{x}_{j,i}^{n+1}, C)} \int_{|e| \geq \varepsilon} (1 \wedge |e|) \, \nu(de) \leq C \frac{h^2}{\varepsilon},$$

where we have used the fact that $|\mathcal{I}_h[\phi] - \phi|_{B(\mathbf{x}_{j,i}^{n+1}, C)} \leq C|D^2\phi|_{B(\mathbf{x}_{j,i}^{n+1}, C)} h^2$. Similar arguments give us that $|K_{\varepsilon,h}^{\alpha,1} \phi_{j,i}^{n+1} - K_\varepsilon^{\alpha,1} \phi(\mathbf{x}_{j,i}^{n+1})| \leq Ch^2 |D^2\phi|_{B(\mathbf{x}_{j,i}^{n+1}, C)} \int_{|e| \geq \varepsilon} \nu(de) \leq C \frac{h^2}{\varepsilon^2}$.

We then infer from Taylor's theorem with an integral remainder that the truncation errors of the local terms can be bounded by

$$|A_{\varepsilon,h,k}^\alpha \phi_{j,i}^{n+1} - A_\varepsilon^\alpha \phi(\mathbf{x}_{j,i}^{n+1}) - b_\varepsilon^\alpha(x_{j,i}) \cdot D\phi(\mathbf{x}_{j,i}^{n+1})|$$

$$\leq C|D^2\phi|_{B(\mathbf{x}_{j,i}^{n+1}, C)} \frac{h^2}{k^2} + \omega(\mathbf{x}_{j,i}^{n+1}, k)$$

for some function $\omega(\mathbf{x}_{j,i}^{n+1}, k)$ such that $\omega(\cdot, k) \to 0$ as $k \to 0$ uniformly on compact neighbourhoods of $\mathbf{x}_{j,i}^{n+1}$, which enables us to deduce that

$$|A_{\varepsilon,h,k}^{\alpha}\phi_{j,i}^{n+1} + K_{\varepsilon,h}^{\alpha,1}\phi_{j,i}^{n+1} - A_{\varepsilon}^{\alpha}\phi_j(\mathbf{x}_{j,i}^{n+1}) - K_{\varepsilon}^{\alpha}\phi_j(\mathbf{x}_{j,i}^{n+1})|$$

$$\leq |A_{\varepsilon,h,k}^{\alpha}\phi_{j,i}^{n+1} - A_{\varepsilon}^{\alpha}\phi(\mathbf{x}_{j,i}^{n+1}) - b_{\varepsilon}^{\alpha}(x_{j,i}) \cdot D\phi(\mathbf{x}_{j,i}^{n+1})| + |K_{\varepsilon,h}^{\alpha,1}\phi_{j,i}^{n+1} - K_{\varepsilon}^{\alpha,1}\phi(\mathbf{x}_{j,i}^{n+1})|$$

$$\leq C\left(\frac{h^2}{k^2} + \frac{h^2}{\varepsilon^2} + \omega(\mathbf{x}_{j,i}^{n+1}, k)\right).$$

$\square$

Lemma 4.10 *Under Assumption* 2.1 *there exists C independent of $h, k, \varepsilon, \delta$ such that for all $\varepsilon < 1$*

$$\sum_{m\neq 0}\kappa_{h,m,i}^{\alpha,n} \leq \frac{C}{h\varepsilon} \wedge \frac{1}{\varepsilon^2}, \quad \sum_{m\neq 0}\beta_{h,m,i}^{\alpha,n} \leq \frac{C}{h} \wedge \frac{1}{\varepsilon}, \quad \sum_{m\in\mathbb{Z}^d}d_{h,k,m,i}^{\alpha,n} \leq \frac{C}{k^2},$$

where $\kappa_{h,m,i}^{\alpha,n}$, $\beta_{h,m,i}^{\alpha,n}$, and $d_{h,k,m,i}^{\alpha,n}$ are defined in (3.14) *and* (3.16), *respectively.*

Proof We shall only prove the estimate for $\kappa_{h,m,i}^{\alpha,n}$, since the estimate for $\beta_{h,m,i}^{\alpha,n}$ follows from a similar argument, and the estimate for $d_{h,k,m,i}^{\alpha,n}$ follows directly from the fact that $\sum_m \omega_m = 1$.

The definition of $\kappa_{h,m,i}^{\alpha,n}$ and the integrability property (2.1) of ν imply that

$$\sum_{m\neq 0}\kappa_{h,m,i}^{\alpha,n} = \sum_{m\neq 0}\int_{|e|>\varepsilon}\omega_m(\eta^{\alpha}(x_i,e);h)\,\nu(de)$$

$$= \sum_{m\neq 0}\int_{|e|>\varepsilon}\omega_m(\eta^{\alpha}(x_i,e);h)\mathbf{1}_{\{\eta^{\alpha}(x_i,e)\in\operatorname{supp}\omega_m\}}\,\nu(de)$$

$$= \sum_{m\neq 0}\int_{|e|>\varepsilon}\left(\omega_m(\eta^{\alpha}(x_i,e);h) - \omega_m(0;h)\right)\mathbf{1}_{\{\eta^{\alpha}(x_i,e)\in\operatorname{supp}\omega_m\}}\,\nu(de)$$

$$\leq \int_{|e|>\varepsilon}\sum_{m\neq 0}|D\omega_m|_0|\eta^{\alpha}(x_i,e)|\mathbf{1}_{\{\eta^{\alpha}(x_i,e)\in\operatorname{supp}\omega_m\}}\,\nu(de)$$

$$\leq \frac{C}{h}\int_{|e|>\varepsilon}(1\wedge|e|)\,\nu(de)$$

$$\leq \frac{C}{h}\int_{|e|>\varepsilon}\frac{1\wedge|e|}{\varepsilon}(1\wedge|e|)\,\nu(de) = \frac{C}{h\varepsilon}\int_{|e|>\varepsilon}(1\wedge|e|^2)\,\nu(de) \leq \frac{C}{h\varepsilon}.$$

Alternatively, it follows directly from the identity $\sum_{m\in\mathbb{Z}^d}\omega_m(\cdot;h)\equiv 1$ that

$$\sum_{m\neq 0}\kappa_{h,m,i}^{\alpha,n} = \sum_{m\neq 0}\int_{|e|>\varepsilon}\omega_m(\eta^{\alpha}(x_i,e);h)\,\nu(de)$$

$$\leq \int_{|e|>\varepsilon}\nu(de) \leq \frac{1}{\varepsilon^2}\int_{|e|>\varepsilon}(1\wedge|e|^2)\,\nu(de),$$

which leads us to the desired estimates. $\square$

Remark 6 Since we have not used any information on the exact behavior of the non-singular measure ν around zero, the estimates for the nonlocal terms in Lemmas 4.9 and 4.10 are not optimal for many specific cases. If one can estimate upper bounds of the density of the Lévy measure, or equivalently estimate the (pseudo-differential) orders of the nonlocal operators K^α and B^α, more precise results for the truncation error of the singular measure can be deduced (Reference [5]).

The next lemma presents some important properties of the Lax–Friedrichs numerical flux for Lipschitz continuous Hamiltonian, which are crucial for our subsequent analysis. We refer readers to Reference [11] for a proof of these statements. Then the following hold:

Lemma 4.11 Let $\tilde{f}$ as in (3.17) and $(\mathbf{x}^n_{j,i}, u, k) \in \Omega_{j,h} \times \mathbb{R} \times \mathbb{R}$, and suppose Assumption 2.1 and the condition $\theta > C\lambda$ hold, where C is the Lipschitz constant of the Hamiltonian $\bar{f}$.

(1) *(Consistency.) For any test functions $\phi \in C^{1,2}([0, T] \times \mathbb{R}^d)$, we have*

$$|\tilde{f}^\alpha(\mathbf{x}^n_{j,i}, u, \Delta\phi^n_{j,i}, k) - \bar{f}^\alpha(\mathbf{x}^n_{j,i}, u, D\phi(\mathbf{x}^n_{j,i}), k)| \leq Ch^2/\Delta t.$$

(2) *(Monotonicity.) If $V^n_{j,i} \geq U^n_{j,i}$, for all i, j, n, then we have*

$$\Delta t\, \tilde{f}^\alpha(\mathbf{x}^n_{j,i}, u, \Delta V^n_{j,i}, k) + 2d\theta V^n_{j,i} \geq \Delta t\, \tilde{f}^\alpha(\mathbf{x}^n_{j,i}, u, \Delta U^n_{j,i}, k) + 2d\theta U^n_{j,i}.$$

(3) *(Stability.) For any bounded functions U and V, we have*

$$|(\Delta t\, \tilde{f}^\alpha(\mathbf{x}^n_{j,i}, u, \Delta V^n_{j,i}, k) + 2d\theta V^n_{j,i})$$
$$- (\Delta t\, \tilde{f}^\alpha(\mathbf{x}^n_{j,i}, u, \Delta U^n_{j,i}, k) + 2d\theta U^n_{j,i})| \leq 2d\theta|U - V|_0.$$

Proposition 4.12 *Suppose Assumption 2.1, the positive interpolation property in Condition 1 and the condition $\theta > C\lambda$ hold. Then we have the following:*

(1) *There exists a unique bounded solution U^n of the scheme (3.18).*
(2) *The scheme is ℓ^∞ stable and weakly monotone. It is consistent with the switching system (3.6) provided $h^2/\Delta t \to 0$ and $h/k \to 0$ as $h, k, \Delta t \to 0$ (ε is fixed here).*

Proof We start to establish the existence and uniqueness of a bounded solution of (3.18) in (1) by an induction argument. It is clear the statement holds for $t^0 = 0$ since $U^0 = g$ is bounded. Now we assume that $\{U^{n-1}_j\}^J_{j=1}$ are bounded functions on $h\mathbb{Z}^d$ and consider the time point t^n. The positive interpolation property implies the interpolation step among different grids does not increase the ℓ^∞ norm of the solution, and hence $U^{n-\frac{1}{2}}_j$ is bounded for each $j = 1, \ldots, J$.

For each $\rho > 0$ and $j = 1, \ldots, J$, we define the operator $\mathcal{P} : U^n_j \to U^n_j$ by

$$\mathcal{P}U^n_{j,i} = U^n_{j,i} - \rho \cdot (\text{left-hand side of (3.18)}), \quad i \in \mathbb{Z}^d,$$

with a given function $U_j^{n-\frac{1}{2}}$. By virtue of the fact that fixed points to the equation $\mathcal{P}U_j^n = U_j^n$ are precisely the solutions to (3.18), it suffices to establish that for small enough ρ, the operator $\mathcal{P}$ is a contraction on $\ell^\infty(\mathbb{Z}^d)$, i.e., the Banach space of bounded functions on $h\mathbb{Z}^d$ employed with the sup-norm, which along with the contraction mapping theorem leads to the desired results. (Similar contraction operators have been introduced in References [7,13] to demonstrate the well-posedness of their numerical schemes).

For any bounded functions U_j^n and V_j^n, the definitions of $\mathcal{P}$, $A_{\varepsilon,h,k}^\alpha$ and $K_{\varepsilon,h}^{\alpha,1}$ give that

$$\mathcal{P}U_{j,i}^n - \mathcal{P}V_{j,i}^n$$

$$\leq (1-\rho)(U_{j,i}^n - V_{j,i}^n) + \rho\Delta t\Bigg[\sum_{m\in\mathbb{Z}^d} d_{h,k,m,i}^{\alpha,n}[(U_{j,m}^n - V_{j,m}^n) - (U_{j,i}^n - V_{j,i}^n)]$$

$$+ \sum_{m\neq 0} \kappa_{h,m,i}^{\alpha,n}[(U_{j,i+m}^n - V_{j,i+m}^n) - (U_{j,i}^n - V_{j,i}^n)]$$

$$+ \tilde{f}^\alpha(\mathbf{x}_{j,i}^n, U_{j,i}^n, \Delta U_{j,i}^n, B_{\varepsilon,h}^\alpha U_{j,i}^n) - \tilde{f}^\alpha(\mathbf{x}_{j,i}^n, V_{j,i}^n, \Delta V_{j,i}^n, B_{\varepsilon,h}^\alpha V_{j,i}^n)\Bigg]$$

$$\leq (1-\rho - \rho\Delta t \sum_{m\in\mathbb{Z}^d} d_{h,k,m,i}^{\alpha,n} - \rho\Delta t \sum_{m\neq 0} \kappa_{h,m,i}^{\alpha,n})(U_{j,i}^n - V_{j,i}^n)$$

$$+ \rho\Delta t(\sum_{m\in\mathbb{Z}^d} d_{h,k,m,i}^{\alpha,n} + \sum_{m\neq 0} \kappa_{h,m,i}^{\alpha,n})|U_j^n - V_j^n|_0$$

$$+ \rho\Delta t\big(\tilde{f}^\alpha(\mathbf{x}_{j,i}^n, U_{j,i}^n, \Delta U_{j,i}^n, B_{\varepsilon,h}^\alpha U_{j,i}^n) - \tilde{f}^\alpha(\mathbf{x}_{j,i}^n, V_{j,i}^n, \Delta U_{j,i}^n, B_{\varepsilon,h}^\alpha U_{j,i}^n)\big)$$

$$\tag{4.17}$$

$$+ \rho\Delta t\big(\tilde{f}^\alpha(\mathbf{x}_{j,i}^n, V_{j,i}^n, \Delta U_{j,i}^n, B_{\varepsilon,h}^\alpha U_{j,i}^n) - \tilde{f}^\alpha(\mathbf{x}_{j,i}^n, V_{j,i}^n, \Delta U_{j,i}^n, B_{\varepsilon,h}^\alpha V_{j,i}^n)\big)$$

$$\tag{4.18}$$

$$+ \rho\Delta t\big(\tilde{f}^\alpha(\mathbf{x}_{j,i}^n, V_{j,i}^n, \Delta U_{j,i}^n, B_{\varepsilon,h}^\alpha V_{j,i}^n) - \tilde{f}^\alpha(\mathbf{x}_{j,i}^n, V_{j,i}^n, \Delta V_{j,i}^n, B_{\varepsilon,h}^\alpha V_{j,i}^n)\big).$$

$$\tag{4.19}$$

It remains to estimate (4.17), (4.18) and (4.19). Lemma 4.11 (3) enables us to bound (4.19) by $-\rho 2d\theta(U_{j,i}^n - V_{j,i}^n) + \rho 2d\theta|U_j^n - V_j^n|_0$. We then derive upper bounds for (4.17) and (4.18) depending on whether $U_{j,i}^n - V_{j,i}^n$ or $B_{\varepsilon,h}^\alpha U_{j,i}^n - B_{\varepsilon,h}^\alpha V_{j,i}^n$ is positive. If $U_{j,i}^n - V_{j,i}^n > 0$, the monotonicity of f in y implies that (4.17) is bounded above by $-\rho\Delta t C(U_{j,i}^n - V_{j,i}^n)$, while if $U_{j,i}^n - V_{j,i}^n < 0$, the Lipschitz continuity of f in y enables us to bound (4.17) by $\rho\Delta t C|U_{j,i}^n - V_{j,i}^n| = -\rho\Delta t C(U_{j,i}^n - V_{j,i}^n)$.

We then discuss the sign of $B_{\varepsilon,h}^\alpha U_{j,i}^n - B_{\varepsilon,h}^\alpha V_{j,i}^n$. Suppose $B_{\varepsilon,h}^\alpha U_{j,i}^n - B_{\varepsilon,h}^\alpha V_{j,i}^n < 0$, then we obtain from the monotonicity of f in k that (4.18)≤ 0. Consequently we obtain that

$$\mathcal{P}U_{j,i}^n - \mathcal{P}V_{j,i}^n \leq (1-\rho - \rho\Delta t \sum_{m\in\mathbb{Z}^d} d_{h,k,m,i}^{\alpha,n}$$

$$- \rho \Delta t \sum_{m \neq 0} \kappa_{h,m,i}^{\alpha,n} - \rho \Delta t C - \rho 2 d\theta)(U_{j,i}^n - V_{j,i}^n)$$

$$+ \rho(\Delta t \sum_{m \in \mathbb{Z}^d} d_{h,k,m,i}^{\alpha,n} + \Delta t \sum_{m \neq 0} \kappa_{h,m,i}^{\alpha,n} + 2 d\theta)|U_j^n - V_j^n|_0$$

$$\leq (1 - \rho - \rho \Delta t C)|U_j^n - V_j^n|_0, \qquad (4.20)$$

provided that $1 - \rho(1 + 2d\theta) - \rho \Delta t \left(\sum_{m \in \mathbb{Z}^d} d_{h,k,m,i}^{\alpha,n} + \sum_{m \neq 0} \kappa_{h,m,i}^{\alpha,n} + C \right) > 0$, which is satisfied for small enough ρ.

On the other hand, if $B_{\varepsilon,h}^\alpha U_{j,i}^n - B_{\varepsilon,h}^\alpha V_{j,i}^n > 0$, the Lipschitz continuity of f in k enables us to bound (4.18) by $C(B_{\varepsilon,h}^\alpha U_{j,i}^n - B_{\varepsilon,h}^\alpha V_{j,i}^n)$, which along with (3.13) implies again (4.20) provided that the the following condition is satisfied:

$$1 - \rho(1 + 2d\theta) - \rho \Delta t \left(\sum_{m \in \mathbb{Z}^d} d_{h,k,m,i}^{\alpha,n} + \sum_{m \neq 0} (\kappa_{h,m,i}^{\alpha,n} + \beta_{h,m,i}^{\alpha,n}) + C \right) > 0, \quad (4.21)$$

which holds for small enough ρ. This completes the proof that $\mathcal{P}$ is a contraction operator.

We now proceed to establish the ℓ^∞ stability of the scheme. Let $\{U_j^{n-1}\}_{j=1}^J$ be the solutions to (3.18). By expressing the discrete operators $A_{\varepsilon,h,k}^\alpha$ and $K_{\varepsilon,h}^{\alpha,1}$ in the monotone form (3.15) and (3.12), and substituting them into (3.18), we obtain that

$$[1 + 2d\theta + \Delta t \left(\sum_{m \in \mathbb{Z}^d} d_{h,k,m,i}^{\alpha,n} + \sum_{m \neq 0} \kappa_{h,m,i}^{\alpha,n} \right)]U_{j,i}^n$$

$$- \Delta t \left(\sum_{m \in \mathbb{Z}^d} d_{h,k,m,i}^{\alpha,n} U_{j,m}^n + \sum_{m \neq 0} \kappa_{h,m,i}^{\alpha,n} U_{j,i+m}^n \right)$$

$$= U_{j,i}^{n-\frac{1}{2}} + \Delta t \tilde{f}^\alpha(\mathbf{x}_{j,i}^n, U_{j,i}^n, \Delta U_{j,i}^n, B_{\varepsilon,h}^\alpha U_{j,i}^n) + 2d\theta U_{j,i}^n,$$

from which we can deduce

$$[1 + 2d\theta + \Delta t \left(\sum_{m \in \mathbb{Z}^d} d_{h,k,m,i}^{\alpha,n} + \sum_{m \neq 0} \kappa_{h,m,i}^{\alpha,n} \right)]U_{j,i}^n$$

$$- \Delta t \left(\sum_{m \in \mathbb{Z}^d} d_{h,k,m,i}^{\alpha,n} + \sum_{m \neq 0} \kappa_{h,m,i}^{\alpha,n} \right)|U_{j,i}^n|_0$$

$$\leq \Delta t \left[f^\alpha(\mathbf{x}_{j,i}^n, U_{j,i}^n, \Delta U_{j,i}^n, B_{\varepsilon,h}^\alpha U_{j,i}^n) - f^\alpha(\mathbf{x}_{j,i}^n, 0, \Delta U_{j,i}^n, B_{\varepsilon,h}^\alpha U_{j,i}^n) \right] \quad (4.22)$$

$$+ \Delta t \left[f^\alpha(\mathbf{x}_{j,i}^n, 0, \Delta U_{j,i}^n, B_{\varepsilon,h}^\alpha U_{j,i}^n) - f^\alpha(\mathbf{x}_{j,i}^n, 0, \Delta U_{j,i}^n, 0) \right]$$

$$+ |U_j^{n-\frac{1}{2}}|_0 + (\Delta t [\tilde{f}^\alpha(\mathbf{x}_{j,i}^n, 0, \Delta U_{j,i}^n, 0) - \tilde{f}^\alpha(\mathbf{x}_{j,i}^n, 0, 0, 0)]$$

$$+ 2d\theta U_{j,i}^n) + \Delta t \tilde{f}^\alpha(\mathbf{x}_{j,i}^n, 0, 0, 0). \qquad (4.23)$$

Using similar arguments as those for the upper bound of (4.17), we deduce that (4.22) is bounded above by $-\Delta t C U_{j,i}^n$ independent of the sign of $U_{j,i}^n$.

Suppose now $B^\alpha_{\varepsilon,h} U^n_{j,i} < 0$, then we obtain from the monotonicity of f in k that (4.23) is nonpositive. Then the ℓ^∞ stability of the numerical flux and the boundedness of $f^\alpha(\mathbf{x}, 0, 0, 0)$ yield that

$$(1 + \Delta t C)|U^n_j|_0 \le |U^{n-\frac{1}{2}}_j|_0 + \Delta t C_1. \tag{4.24}$$

Here C is the constant from Assumption 2.1 and $C_1 > 0$ is a large enough constant that we will choose later. On the other hand, if $B^\alpha_{\varepsilon,h} U^n_{j,i} > 0$, the Lipschitz continuity of f in k enables us to bound (4.23) by $C B^\alpha_{\varepsilon,h} U^n_{j,i}$, which along with (3.12) implies again (4.24).

With the estimate (4.24) in hand, we are ready to derive a uniform bound for the solutions $\{U^n_j\}$, which is independent of h and c. The proof follows from an inductive argument. Let us introduce the notation $|U^n|_0 = \max_{1 \le j \le J} |U^n_j|_0$ for each n and define the term $a_0 = \max(|g|_0, |\zeta|_0)$, then it is clear that $a_0 \ge \max(|U^0|_0, |\zeta|_0)$. Suppose we have a_{n-1} such that $a_{n-1} \ge \max(|U^{n-1}|_0, |\zeta|_0)$. Then the definition of $U^{n-\frac{1}{2}}_{j,i}$ implies that $|U^{n-\frac{1}{2}}_j|_0 \le \max(|\zeta|_0, |U^{n-1}|_0) \le a_{n-1}$. Define the term

$$a_n := \frac{1}{1 + \Delta t C} a_{n-1} + \Delta t C_1,$$

with the same constants as those in (4.24), then we have $|U^n|_0 \le a_n$. To proceed by induction, we further require $a_n \ge |\zeta|_0$. Since $a_{n-1} \ge |\zeta|_0$ and C is fixed, it suffices to require $C_1 \ge C|\zeta|_0$. In this way, we can construct a sequence $\{a_n\}$, such that $|U^n|_0 \le a_n$, but a_n is uniformly bounded independent of c, h and Δt, and hence this completes the proof of ℓ^∞ stability.

We now study the weak monotonicity of the scheme. Let $V^n_{j,i} \ge U^n_{j,i}$ and $\tilde{V}^n_{k,i(j)} \ge \tilde{U}^n_{k,i(j)}$ for all i, j, k, n, then we have $V^{n+\frac{1}{2}}_{j,i} \ge U^{n+\frac{1}{2}}_{j,i}$. Moreover the monotonicity of f in k and the weak monotonicity of $\tilde{f}$ imply that

$$\sum_{m \in \mathbb{Z}^d} d^{\alpha,n}_{h,k,m,i} U^{n+1}_{j,m} + \sum_{m \ne 0} \kappa^{\alpha,n}_{h,m,i} U^{n+1}_{j,i+m}$$
$$+ \tilde{f}^\alpha(\mathbf{x}^{n+1}_{j,i}, U^{n+1}_{j,i}, \Delta U^{n+1}_{j,i}, \sum_{m \ne 0} \beta^{\alpha,n}_{h,m,i}[U^{n+1}_{j,i+m} - U^{n+1}_{j,i}])$$

is nondecreasing with $\{U^{b+1}_{j,a}\}_{(a,b) \ne (i,n)}$, which gives the weak monotonicity of the scheme (3.18).

Finally we study the consistency of the scheme. By using the Lipschitz continuity of $x \to \min(x, a)$, it is clear that it suffices to bound

$$(I_1) := \Delta t \big(A^{\alpha}_{\varepsilon,h,k} \phi^{n+1}_{j,i} + K^{\alpha,1}_{\varepsilon,h} \phi^{n+1}_{j,i} + \tilde{f}^{\alpha}(\mathbf{x}^{n+1}_{j,i}, \phi^{n+1}_{j,i} + \xi, \Delta \phi^{n+1}_{j,i}, B^{\alpha}_{\varepsilon,h} \phi^{n+1}_{j,i}) \big)$$

$$(I_2) := \bigg| \frac{\phi^{n+1}_{j,i} - \phi^{n}_{j,i}}{\Delta t} - \big(A^{\alpha}_{\varepsilon,h,k} \phi^{n+1}_{j,i} + K^{\alpha,1}_{\varepsilon,h} \phi^{n+1}_{j,i}$$

$$+ \tilde{f}^{\alpha}(\mathbf{x}^{n+1}_{j,i}, \phi^{n+1}_{j,i} + \xi, \Delta \phi^{n+1}_{j,i}, B^{\alpha}_{\varepsilon,h} \phi^{n+1}_{j,i}) \big)$$

$$- \phi_{j,t}(\mathbf{x}^{n+1}_{j,i}) - A^{\alpha}_{\varepsilon} \phi_j(\mathbf{x}^{n+1}_{j,i}) - K^{\alpha}_{\varepsilon} \phi_j(\mathbf{x}^{n+1}_{j,i})$$

$$- f^{\alpha}(\mathbf{x}^{n+1}_{j,i}, \phi(\mathbf{x}^{n+1}_{j,i}), D\phi(\mathbf{x}^{n+1}_{j,i}), B^{\alpha}_{\varepsilon} \phi(\mathbf{x}^{n+1}_{j,i})) \bigg|,$$

which can be estimated by using Lemmas 4.9, 4.11, and the Lipschitz continuity of f. □

Remark 7 The contraction operator $\mathcal{P}$ is introduced to demonstrate our scheme admits a unique solution for any given discretization parameters Δt, h, k and ε. However, due to its low convergence rate, it is not advisable to implement this contraction mapping directly to solve the nonlinear equation (3.18). In fact, Lemma 4.10 and the stability condition (4.20) restrict the contraction constant of $\mathcal{P}$ to admit a lower bound depending on the spatial discretization of the diffusion operator. This undesirable dependence of Δt on k can be avoided by considering the mapping T defined by (3.19), which is implicit in the local terms. It has been shown that for small enough h, the contraction constant of T is proportional to θ, which can be chosen to achieve a rapid convergence.

5 Numerical Experiments

In this section, we present several numerical experiments to analyse the effectiveness of the numerical scheme proposed in Sect. 3. We shall investigate the convergence of numerical solutions with respect to the switching cost, timestep, and mesh size, and show that a relatively coarse discretization of the admissible control set already leads to an accurate approximation.

We consider a portfolio optimization problem over a time interval $[0, T]$, in a framework of recursive utility. An investor can control his wealth process $X^{t,x,\alpha}$ through a selection of the control process $\alpha \in \mathcal{A}^{t}_{t}$, say his or her portfolio strategy, and can also choose the duration of the investment via a stopping time τ. If the agent chooses a strategy pair (α, τ), then the associated terminal reward is given by

$$\xi^{t,x,\alpha}_{\tau} = \zeta(\tau, X^{t,x,\alpha}_{\tau}) 1_{t \leq \tau < T} + g(X^{t,x,\alpha}_{\tau}) 1_{\tau=T}$$

for some utilities ζ and g, and where $\tau \in \mathcal{T}^{t}_{t}$, the set of $\mathbb{F}^{t}$-stopping times valued in $[t, T]$.

The performance of this investment is evaluated under a particular nonlinear expectation, called the recursive utility process (see e.g. [9]), which is associated with a BSDE (with Lipschitz continuous drivers). It generalizes the standard additive utilities by including a dependence on the future utility (corresponding to the future

wealth). Roughly speaking, the recursive utility depends on the future utility through the dependance of the driver f on y, and can also depend on the "variability" or "volatility" of future utility through the dependance of f on z and k.

Let x be the wealth at the initial time t, (α, τ) be the chosen strategy, $\mathcal{E}^{\alpha,t,x}[\cdot]$ be a recursive utility function associated with the BSDE with driver f^α. The aim of the investor is to maximize the utility of the investment:

$$u(t, x) := \sup_{\tau \in \mathcal{T}_t^t} \sup_{\alpha \in \mathcal{A}_t^t} \mathcal{E}_{t,\tau}^{t,\alpha}[\xi_\tau^{t,x,\alpha}],$$

over all admissible choices of (α, τ). Under Assumption 2.1, it can be shown that the value function u of this mixed optimization problem coincides with the unique bounded viscosity solution of the (backward) HJBVI (2.5).

For the numerical tests, we consider a financial market with a risk-free asset with an interest rate r and a risky asset whose price follows

$$dS_t = S_{t-}\left[b \, dt + \sigma \, dW_t + \int_E \eta(e) \, \tilde{N}(dt, de)\right],$$

where W is a Brownian motion and $\tilde{N}(dt, de) = N(dt, de) - \nu(de)dt$ is a compensated jump measure. If we denote by α_t the percentage of the portfolio held in the risky asset at time t, then the dynamics of the portfolio is given by

$$dX_t = \alpha_t X_{t-}\left[b \, dt + \sigma \, dW_t + \int_E \eta(e) \, \tilde{N}(dt, de)\right], \quad X_0 = x_0.$$

The performance will be evaluated by the recursive utility function induced by the BSDE with the following driver:

$$f(t, x, y, z) = \psi(t, x) - \beta y - \kappa|z|.$$

for some instantaneous reward function ψ. Recall that any concave utility function admits a dual representation via a set of probability measures absolutely continuous with respect to the original probability measure P (see e.g. [24]). This result allows us to interpret $\kappa \geq 0$ as an ambiguity-aversion coefficient relative to the Brownian motion as suggested in [9, Sect. 3.3].

The value function of this control problem satisfies the following HJBVI:

$$\begin{cases} \min\left\{u(t, x) - \zeta(t, x), u_t + \inf_{\alpha \in [0,1]}\left(-L^\alpha u - \psi + \beta u + \alpha \kappa \sigma |x u_x|\right)\right\} = 0, \\ u(0, x) - g(x) = 0 \end{cases}$$

$$(5.1)$$

for $(t, x) \in [0, T] \times \mathbb{R}$, where the nonlocal operator $L^\alpha = A^\alpha + K^\alpha$ satisfies for $\phi \in C^2([0, T] \times \mathbb{R})$

Table 1 Model parameters for the recursive utility maximization problem

β	κ	b	σ	μ	T	x_0
0.2	1	0.1	0.15	6	1	1

$$A^\alpha \phi(t, x) = \frac{1}{2}\alpha^2\sigma^2 x^2 \phi_{xx}(t, x) + (\alpha b + (1 - \alpha)r)x\phi_x(t, x),$$

$$K^\alpha \phi(t, x) = \int_{\mathbb{R}\setminus\{0\}} \big(\phi(t, x + \alpha x\eta(e)) - \phi(t, x) - \alpha x\eta(e)\phi_x(t, x)\big)\,\nu(de). \quad (5.2)$$

We then specify the choice of data for our numerical experiments. We use the exponential utility function $\zeta(t, x) = g(x) = (1 - e^{-x})^+$, which determines both the intermediate and terminal payoff, and acts as the initial condition and the obstacle to the HJBVI. Moreover, we consider the tempered stable Lévy measure $\nu(de) = \frac{e^{-\mu|e|}}{|e|}de$ on $\mathbb{R}$ with intensity $\eta(e) = 1 \wedge |e|$ for the jump component (which is a special case of the variance Gamma model in Reference [10]). For simplicity, we choose a zero interest rate, i.e., $r = 0$.

We further choose the function $\psi(t, x) = 0.8\exp(-(T - t))\exp(-x/2)$ as the instantaneous reward. As we will see later, this choice of ψ implies that the optimal control α varies in the state space and evolves in time, and there can be non-trivial stopping. The resulting HJBVI will be localized to the domain $(0, 2)$ with $u(t, x) = g(x)$ for $(t, x) \in (0, T) \times \mathbb{R} \setminus (0, 2)$. The numerical values for the parameters used in the experiments are given in Table 1.

Now we are ready to discuss the selection of the discretization parameters in detail. The density of the tempered stable measure ν enables us to improve the estimates in Lemma 4.10 to $\sum_{m \neq 0} \kappa_{h,m,i}^{\alpha,n} \leq \log(\varepsilon)$, and hence choosing $\varepsilon = h$ and $\Delta t = O(h)$ leads us to a consistent approximation to the switching system (3.6). Moreover, choosing $\theta = \frac{1}{40}$ and $\Delta t = \frac{h}{15}$ ensures the numerical flux is stable and the contraction constant of T in (3.19) is less than $\frac{1}{10}$.

The coefficients of the nonlocal terms are evaluated by the midpoint quadrature formula, which is clearly monotone and consistent. We observe that for the control problem with the parameters as in Table 1, the optimal strategy α^* will always be obtained at one of the endpoints of $[0, 1]$. In fact, using Taylor's theorem, we are able to approximate the nonlocal term $K^\alpha u$ by

$$K^\alpha u(t, x) \approx \frac{1}{2}\alpha^2 x^2 \int_{\mathbb{R}\setminus\{0\}} (1 \wedge |e|)^2\,\nu(de)u_{xx}(t, x) = \frac{1}{2}\alpha^2 x^2 C u_{xx}(t, x),$$

at any given (t, x) for which the value function lies above the obstacle and is sufficiently smooth. Then we infer from the HJBVI (5.1) that the optimal control α^* is the maximizer of a quadratic function on $[0, 1]$, which is attained in the interior only if

$$u_{xx}(t, x) < 0, \quad -\frac{bu_x - \sigma|u_x|}{(\sigma^2 + C)xu_{xx}} \in (0, 1).$$

Table 2 Numerical solutions for the recursive utility maximization problem with different mesh sizes and switching costs

h		1/25	1/50	1/100	1/200	1/400	1/800	1/1600
$c = \frac{1}{40}$	(a)	0.77538	0.77770	0.77869	0.77912	0.77933	0.77942	**0.77947**
	(b)		2321.35	984.37	437.19	203.03	97.30	47.53
	(c)			2.3582	2.2516	2.1533	2.0867	2.0471
$c = \frac{1}{160}$	(a)	0.79003	0.79194	0.79290	0.79338	0.79361	0.79373	**0.79379**
	(b)		1916.14	954.91	476.01	237.63	118.73	59.34
	(c)			2.0066	2.0061	2.0031	2.0015	2.0008
$c = \frac{1}{640}$	(a)	0.79471	0.79663	0.79759	0.79806	0.79830	0.79842	**0.79848**
	(b)		1917.40	955.54	476.33	237.79	118.80	59.38
	(c)			2.0066	2.0061	2.0031	2.0015	2.0008
$c = \frac{1}{2560}$	(a)	0.79588	0.79780	0.79876	0.79923	0.79947	0.79959	**0.79965**
	(b)		1917.71	955.70	476.40	237.83	118.82	59.39
	(c)			2.0066	2.0061	2.0031	2.0015	2.0008
$c = \frac{1}{10240}$	(a)	0.79618	0.79810	0.79905	0.79953	0.79977	0.79988	**0.79994**
	(b)		1917.79	955.74	476.42	237.84	118.83	59.39
	(c)			2.0066	2.0061	2.0031	2.0015	2.0007

Shown are: (a) the numerical solutions U_h at (T, x_0); (b) the increments $U_h - U_{2h}$ (in 10^{-6}) ; (c) the rate of increments $(U_h - U_{2h})/(U_{2h} - U_{4h})$

However, since we have $b < \sigma$, the above conditions can never hold for any $x > 0$. Consequently, we deduce that the admissible set is already finite, and replacing [0, 1] by $\mathbf{A}_\delta = \{0, 1\}$ in (5.1) will not introduce any discretiztion error. This has been confirmed with our numerical experiments. For the sake of simplicity, we discretise each component of the switching system on a single uniform mesh, thus Condition 1 (1) is trivially satisfied.

Table 2 contains the numerical solutions to the last component of the switching system at the grid point (T, x_0) with different mesh size h and switching cost c. We examine the convergence of the numerical solutions, denoted as U_h, in h for fixed c, as well as their convergence with respect to the cost c. For any fixed positive switching cost c, we infer from the lines (a) that the numerical solutions converge monotonically to the exact solution. Moreover, the lines (c) indicate the approximation error admits an asymptotic magnitude $O(h) + O(\Delta t)$, which seems not to be affected by the size of the cost c. By considering the boldface values in Table 2 as an accurate approximation to the exact solution of the switching system with a given cost c, we can further conclude that the switching system is consistent to the HJBVI (5.1) with order 1. This follows from the approximate factor of four between the differences 0.00469, 0.00117, and 0.00029 between the last three pairs of values, proportional to the reduction in c. Therefore, by taking $c = O(h)$ and $\Delta t = O(h)$, we can obtain a first-order scheme for the HJBVI.

Table 3 Numerical results for the recursive utility maximization problem with different control refinements

J		2	11	21	41	81
Average error ($\times 10^{-7}$)	(a)	1.1392	0.2398	0.0570	0.0133	0
Runtime in seconds	(b)	200.014	1106.266	2142.714	4204.336	8536.232
	(c)	155.977	228.594	331.294	501.895	809.178
	(d)	53.728	124.390	192.634	230.019	262.488
	(e)	102.249	104.204	138.660	271.876	546.690
Speed-up rate (b/e)	(f)	1.9561	10.61	15.45	15.46	15.61

Shown are: (a) the difference to the solution with the finest refinement; (b) total runtime time without parallelization; (c) total runtime time with parallelization on 16 processors; (d) total communication time among processors; (e) net parallel computation time without communication; (f) speed-up rate of parallelization

We then proceed to analyze the effect of the control discretization. We pick the same parameters as those in Table 1, except that $b = 0.25$, which is chosen such that it is now possible that the optimal control is attained in the interior of $(0, 1)$ (as seen from a similar argument as earlier). Computations are performed using MATLAB R2016b on a 3.30GHz Intel Xeon ES-2667 16-Core processor with 256 GB RAM to enable parallelization. Table 3 illustrates the numerical results for different control meshes ($J = 1/\delta + 1$) with a fixed mesh size $h = 0.005$ and switching cost $c = 1/2560$, and also compares the runtime with or without parallelization.

We can clearly observe from line (a) second order convergence of the numerical solutions, and a relatively coarse control mesh has already yielded an accurate approximation with a negligible control discretization error.

Next, we discuss lines (b)–(f) which analyse the algorithm's parallel efficiency. Hereby, the implicit finite difference scheme for individual components of the switching system (i.e., (3.9), for different j) is solved independently on different processors, while the maximisation step (3.8) requires communication between processors.

The total execution time with and without parallelization are presented in line (b) and (c), respectively, which indicate a significant reduction of computational times. Moreover, by subtracting the communication time among clusters, as shown in line (d), from the total runtime, we can obtain the actual time spent on executing the numerical scheme (line (e)). The speed-up rate of the parallelization is shown in line (f), which grows with the number of controls, and remains stable at the number of cores. Therefore, together with parallelization, piecewise constant timestepping enables us to achieve a high accuracy in the control discretization without significantly increasing the computational time, which is an advantage over policy iterations, which do not parallelise naturally.

We finally examine the impact of the computational domain by performing computations on $(0, 3)$ with $h = 1/400$, $\Delta t = h/20$, $c = 1/640$ and the parameters as in Table 1. Compared to the results in Table 2, this larger domain leads to a relative difference of $7.53 \cdot 10^{-7}$, which is negligible compared to the time and spatial discretization errors.
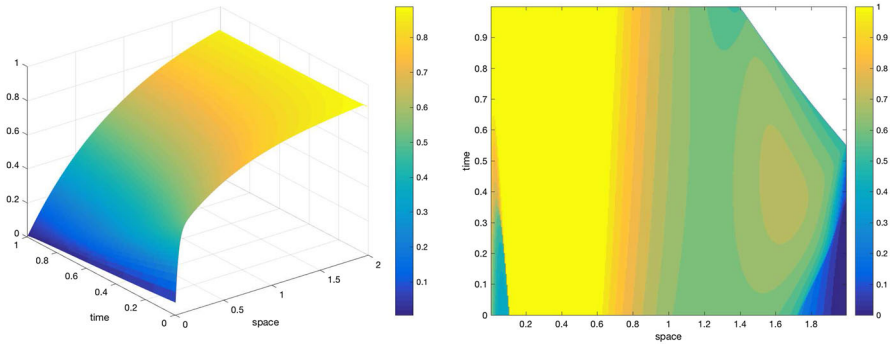
Fig. 1 Numerical value functions (left) and corresponding control strategies with $J = 21$ (right), where the early stopping region is white

The numerical value function and the corresponded feedback control strategy with $J = 21$ are presented in Figure 1, in which the white area represents the region where the obstacle is active, and otherwise the colour indicates the value of the optimal control, as shown in the panel on the right. The approximation to the optimal control pair (τ, α) was found from the numerical solution as follows (see also (3.8) and Remark 3), noting that in our tests $x_{j,i} = x_{k,i}$ for all j, k, and therefore no interpolation is needed:

$$i_n^* \in \operatorname{argmax}_k U_{k,i}^n,$$

$$\theta_i^n = \begin{cases} 0 & \max_k U_{k,i}^n > \xi(t_n, x_{1,i}), \\ 1 & \max_k U_{k,i}^n \le \xi(t_n, x_{1,i}), \end{cases}$$

where $\alpha_i^n = \alpha_{i_n^*}$ is an approximation to the optimal policy and $\{(t_n, x_{1,i}) : \theta_i^n = 1\}$ is an approximation to the stopping region.

6 Conclusions

This paper provides a PDE approximation scheme for the value function of a mixed stochastic control/optimal stopping problem with nonlinear expectations and infinite activity jumps, which is the unique viscosity solution of a nonlocal HJB variational inequality. The approach that we have adopted is based on piecewise constant policy time stepping (PCPT), which reduces the problem to a system of semi-linear PDEs, and a monotone approximation scheme. We prove the convergence of the numerical scheme and illustrate the theoretical results with some numerical examples in the case of a recursive utility maximisation problem.

To the best of our knowledge, this is the first paper which proposes a numerical approximation for a control problem in such a generality. Natural next steps would be to derive theoretical results on the convergence rate and to extend this approach to the case of Hamilton–Jacobi–Bellman–Isaac equations obtained in Reference [8].

Appendix: Comparison Principle for Switching Systems

In this section, we establish the comparison principle for switching system (3.6), cf. Theorem 4.5. We consider a slightly more general switching system with no truncation of the singular measure in K_ε and B_ε, which includes as a special case the switching system (3.6). We first use a classical no-loop argument to reduce the problem into scalar cases, and then analyze the scalar HJBVI by extending the results for continuous solutions in Reference [16] to semicontinuous viscosity solutions. For simplicity, we denote by σ the modified diffusion coefficient $\tilde{\sigma}^\alpha$ defined as (3.1).

Proof of Theorem 4.5 Set

$$M = \sup_{1 \le j \le J, \mathbf{x}, \mathbf{y} \in \mathcal{Q}_T} (U_j(\mathbf{x}) - V_j(\mathbf{y})).$$

It suffices to show that $M \le 0$. For any given $\varepsilon, \rho > 0$, we introduce the functions

$$\psi_j^{\varepsilon,\rho}(t, s, x, y) = U_j(t, x) - V_j(s, y) - \frac{|x - y|^2}{\varepsilon^2}$$

$$- \frac{|t - s|^2}{\varepsilon^2} - \rho^2(|x|^2 + |y|^2), \quad j = 1, \ldots, J, \tag{7.1}$$

for each $t, s \in [0, T]$ and $x, y \in \mathbb{R}^d$, and define the quantity

$$M^{\varepsilon,\rho} := \sup_{j, t, s, x, y} \psi_j^{\varepsilon,\rho}(t, s, x, y).$$

The upper semicontinuity and boundedness of $U_j - V_j$, along the penalization terms, imply that the supremum is obtained at some point $(j^{\varepsilon,\rho}, t^{\varepsilon,\rho}, s^{\varepsilon,\rho}, x^{\varepsilon,\rho}, y^{\varepsilon,\rho})$. Then as in Reference [16], one can find a constant C such that $|t^{\varepsilon,\rho} - s^{\varepsilon,\rho}| + |x^{\varepsilon,\rho} - y^{\varepsilon,\rho}| \le C\varepsilon$, $|x^{\varepsilon,\rho}| \le \frac{C}{\rho}$, and $|y^{\varepsilon,\rho}| \le \frac{C}{\rho}$. Passing to a subsequence if necessary, we may assume that for each ρ, the sequences $\{t^{\varepsilon,\rho}\}_\varepsilon$ and $\{s^{\varepsilon,\rho}\}_\varepsilon$ converge to a common limit t^ρ, while the sequences $\{x^{\varepsilon,\rho}\}_\varepsilon$ and $\{y^{\varepsilon,\rho}\}_\varepsilon$ converge to a common limit x^ρ as ε tends to 0. Moreover, $j^{\varepsilon,\rho}$ lies in a finite set, we may assume $j^{\varepsilon,\rho} = j^\rho$ for all ε. The following lemma gives the convergence of these sequences, whose proof will be deferred after the proof of Theorem 4.5.

Lemma 7.1 *Extracting a further subsequence if necessary, we have*

$$\lim_{\varepsilon \to 0} \frac{|x^{\varepsilon,\rho} - y^{\varepsilon,\rho}|^2}{\varepsilon^2} = \lim_{\varepsilon \to 0} \frac{|t^{\varepsilon,\rho} - s^{\varepsilon,\rho}|^2}{\varepsilon^2} = 0, \quad \lim_{\varepsilon \to 0} U_{j^\rho}(t^{\varepsilon,\rho}, x^{\varepsilon,\rho}) = U_{j^\rho}(t^\rho, x^\rho)$$

$$\lim_{\varepsilon \to 0} V_{j^\rho}(s^{\varepsilon,\rho}, y^{\varepsilon,\rho}) = V_{j^\rho}(t^\rho, x^\rho), \quad \lim_{\rho \to 0} \rho^2 |x^\rho|^2 = 0, \quad \lim_{\rho \to 0} \lim_{\varepsilon \to 0} M^{\varepsilon,\rho} = M.$$

We now divide our analysis into three cases to establish $M \leq 0$.

If there exists a subsequence of $\{t^\rho\}$ such that $t^\rho = 0$ for all ρ, we then deduce $M \leq 0$ along this subsequence by adapting the arguments in Reference [16] to semicontinuous solutions.

On the other hand, if t^ρ is different from 0 for all ρ, then for any fixed ρ and small enough ε, using Lemma 7.2, which can be proved similarly as Lemma 4.1 in Reference [6], we know there exists $j_0^{\varepsilon,\rho} \in \{1, \ldots, J\}$, which for simplicity is still denoted as $j^{\varepsilon,\rho}$, such that $U_{j^{\varepsilon,\rho}}(t^{\varepsilon,\rho}, x^{\varepsilon,\rho}) > \mathcal{M}_{j^{\varepsilon,\rho}} U(t^{\varepsilon,\rho}, x^{\varepsilon,\rho})$. In other words, at the point $(t^{\varepsilon,\rho}, s^{\varepsilon,\rho}, x^{\varepsilon,\rho}, y^{\varepsilon,\rho})$, by considering the $j^{\varepsilon,\rho}$ component of the switching system, we can without loss of generality ignore the term $U_{j^{\varepsilon,\rho}} - \mathcal{M}_{j^{\varepsilon,\rho}} U$ in the definition of subsolutions and get back to the scalar HJBVI.

In this case, if we further assume for each ρ, there exists a subsequence of $\{x^{\varepsilon,\rho}\}_\varepsilon$ such that $(U_{j^\rho} - \zeta)(t^{\varepsilon,\rho}, x^{\varepsilon,\rho}) \leq 0$, then the same arguments in Reference [16] enables us to derive that $M \leq 0$.

Now we come to the final case, where for each $\rho, \varepsilon > 0$, we have $(U_{j^\rho} - \zeta)(t^{\varepsilon,\rho}, x^{\varepsilon,\rho}) > 0$. Applying the nonlocal Jensen-Ishii's lemma as in Reference [16] enables us to obtain $a \in \mathbb{R}$, $\bar{p}, \bar{q} \in \mathbb{R}^d$, and $X, Y \in \mathbb{R}^{d \times d}$ such that it holds for any $\delta > 0$ that

$$\begin{aligned} H_{j^\rho}(t^{\varepsilon,\rho}, x^{\varepsilon,\rho}, U_{j^\rho}(t^{\varepsilon,\rho}, x^{\varepsilon,\rho}), a, \bar{p}, X, l_K, l_B) \\ - H_{j^\rho}(s^{\varepsilon,\rho}, y^{\varepsilon,\rho}, V_{j^\rho}(s^{\varepsilon,\rho}, y^{\varepsilon,\rho}), a, \bar{q}, Y, l'_K, l'_B) \leq 0, \end{aligned} \tag{7.2}$$

where l_K, l'_K, l_B, l'_B are defined as in Reference [16] for each $\delta > 0$ and $H_{j^\rho}(t, x, u, a, p, X, l_1, l_2)$ is given by:

$$\begin{aligned} H_{j^\rho} := a - \mathrm{tr}(\sigma(\alpha_{j^\rho}, x)(\sigma(\alpha_{j^\rho}, x))^T X) \\ - b(\alpha_{j^\rho}, x)^T p - l_1 - f(\alpha_{j^\rho}, t, x, u, \sigma(\alpha_{j^\rho}, x)^T p, l_2). \end{aligned}$$

We now extend the arguments in Reference [16] to semicontinuous subsolution U (resp. supersolution V) and argue by contradiction by assuming $M > 0$. Then for small enough $\rho, \varepsilon, \delta > 0$, we obtain from the monotonicity of f in u that there exists a constant $C > 0$ satisfying

$$\begin{aligned} 0 < \frac{C}{2} M \leq C M^{\varepsilon,\rho} &\leq C(U_{j^\rho}(t^{\varepsilon,\rho}, x^{\varepsilon,\rho}) - V_{j^\rho}(s^{\varepsilon,\rho}, y^{\varepsilon,\rho})) \\ &\leq H_{j^\rho}(s^{\varepsilon,\rho}, y^{\varepsilon,\rho}, U_{j^\rho}(t^{\varepsilon,\rho}, x^{\varepsilon,\rho}), a, \bar{q}, Y, l'_K, l'_B) \\ &\quad - H_{j^\rho}(s^{\varepsilon,\rho}, y^{\varepsilon,\rho}, V_{j^\rho}(s^{\varepsilon,\rho}, y^{\varepsilon,\rho}), a, \bar{q}, Y, l'_K, l'_B) \\ &= H_{j^\rho}(s^{\varepsilon,\rho}, y^{\varepsilon,\rho}, U_{j^\rho}(t^{\varepsilon,\rho}, x^{\varepsilon,\rho}), a, \bar{q}, Y, l'_K, l'_B) \\ &\quad - H_{j^\rho}(t^{\varepsilon,\rho}, x^{\varepsilon,\rho}, U_{j^\rho}(t^{\varepsilon,\rho}, x^{\varepsilon,\rho}), a, \bar{p}, X, l'_K, l'_B) \qquad (7.3) \\ &\quad + H_{j^\rho}(t^{\varepsilon,\rho}, x^{\varepsilon,\rho}, U_{j^\rho}(t^{\varepsilon,\rho}, x^{\varepsilon,\rho}), a, \bar{p}, X, l'_K, l'_B) \\ &\quad - H_{j^\rho}(t^{\varepsilon,\rho}, x^{\varepsilon,\rho}, U_{j^\rho}(t^{\varepsilon,\rho}, x^{\varepsilon,\rho}), a, \bar{p}, X, l_K, l_B) \qquad (7.4) \end{aligned}$$

$$+ H_{j^\rho}(t^{\varepsilon,\rho}, x^{\varepsilon,\rho}, U_{j^\rho}(t^{\varepsilon,\rho}, x^{\varepsilon,\rho}), a, \bar{p}, X, l_K, l_B)$$
$$- H_{j^\rho}(s^{\varepsilon,\rho}, y^{\varepsilon,\rho}, V_{j^\rho}(s^{\varepsilon,\rho}, y^{\varepsilon,\rho}), a, \bar{q}, Y, l'_K, l'_B), \tag{7.5}$$

from which, by expanding (7.3), using the fact that f is Lipschitz continuous and monotone in k for (7.4), and applying (7.2) to (7.5), we can derive that

$$
\begin{aligned}
0 \le CM/2 \le\ & l_K - l'_K + C(l_B - l'_B) + [b(\alpha_{j^\rho}, x^{\varepsilon,\rho})^T \bar{p} - b(\alpha_{j^\rho}, y^{\varepsilon,\rho})^T \bar{q}] \\
& + \frac{1}{2}\big[\mathrm{tr}(\sigma(\alpha_{j^\rho}, x^{\varepsilon,\rho})(\sigma(\alpha_{j^\rho}, x^{\varepsilon,\rho}))^T X \\
& \qquad - \sigma(\alpha_{j^\rho}, y^{\varepsilon,\rho})(\sigma(\alpha_{j^\rho}, y^{\varepsilon,\rho}))^T Y)\big] \\
& + f(\alpha_{j^\rho}, t^{\varepsilon,\rho}, x^{\varepsilon,\rho}, U_{j^\rho}(t^{\varepsilon,\rho}, x^{\varepsilon,\rho}), \sigma(\alpha_{j^\rho}, x^{\varepsilon,\rho})^T \bar{p}, l'_B) \\
& \quad - f(\alpha_{j^\rho}, s^{\varepsilon,\rho}, x^{\varepsilon,\rho}, U_{j^\rho}(t^{\varepsilon,\rho}, x^{\varepsilon,\rho}), \sigma(\alpha_{j^\rho}, x^{\varepsilon,\rho})^T \bar{p}, l'_B) \\
& + f(\alpha_{j^\rho}, s^{\varepsilon,\rho}, x^{\varepsilon,\rho}, U_{j^\rho}(t^{\varepsilon,\rho}, x^{\varepsilon,\rho}), \sigma(\alpha_{j^\rho}, x^{\varepsilon,\rho})^T \bar{p}, l'_B) \\
& \quad - f(\alpha_{j^\rho}, s^{\varepsilon,\rho}, y^{\varepsilon,\rho}, U_{j^\rho}(t^{\varepsilon,\rho}, x^{\varepsilon,\rho}), \sigma(\alpha_{j^\rho}, x^{\varepsilon,\rho})^T \bar{p}, l'_B) \\
& + f(\alpha_{j^\rho}, s^{\varepsilon,\rho}, y^{\varepsilon,\rho}, U_{j^\rho}(t^{\varepsilon,\rho}, x^{\varepsilon,\rho}), \sigma(\alpha_{j^\rho}, x^{\varepsilon,\rho})^T \bar{p}, l'_B) \\
& \quad - f(\alpha_{j^\rho}, s^{\varepsilon,\rho}, y^{\varepsilon,\rho}, U_{j^\rho}(t^{\varepsilon,\rho}, x^{\varepsilon,\rho}), \sigma(\alpha_{j^\rho}, y^{\varepsilon,\rho})^T \bar{q}, l'_B) \\
\le\ & l_K - l'_K + C(l_B - l'_B) + \frac{1}{2}\big[\mathrm{tr}(\sigma(\alpha_{j^\rho}, x^{\varepsilon,\rho})(\sigma(\alpha_{j^\rho}, x^{\varepsilon,\rho}))^T X \\
& \qquad - \sigma(\alpha_{j^\rho}, y^{\varepsilon,\rho})(\sigma(\alpha_{j^\rho}, y^{\varepsilon,\rho}))^T Y)\big] \\
& + f(\alpha_{j^\rho}, t^{\varepsilon,\rho}, x^{\varepsilon,\rho}, U_{j^\rho}(t^{\varepsilon,\rho}, x^{\varepsilon,\rho}), \sigma(\alpha_{j^\rho}, x^{\varepsilon,\rho})^T \bar{p}, l'_B) \\
& \quad - f(\alpha_{j^\rho}, s^{\varepsilon,\rho}, x^{\varepsilon,\rho}, U_{j^\rho}(t^{\varepsilon,\rho}, x^{\varepsilon,\rho}), \sigma(\alpha_{j^\rho}, x^{\varepsilon,\rho})^T \bar{p}, l'_B) \\
& + |x^{\varepsilon,\rho} - y^{\varepsilon,\rho}| + [b(\alpha_{j^\rho}, x^{\varepsilon,\rho})^T \bar{p} - b(\alpha_{j^\rho}, y^{\varepsilon,\rho})^T \bar{q}] \\
& + [\sigma(\alpha_{j^\rho}, x^{\varepsilon,\rho})^T \bar{p} - \sigma(\alpha_{j^\rho}, y^{\varepsilon,\rho})^T \bar{q}].
\end{aligned}
$$

Then noticing the estimates derived in Reference [16] for each term on the right-hand side of the above expression are uniform in the control α_{j^ρ}, and successively passing δ, ε and ρ to 0, we deduce that $0 < M \le 0$, which leads to a contradiction. Thus we conclude $M \le 0$ and complete the proof. □

Proof of Lemma 7.1 For each $\rho > 0$ and $j = 1, \ldots, J$, we introduce the functions $\hat{U}_j^\rho(t, x) = U_j(t, x) - \eta^2 |x|^2$ and $\hat{V}_j^\rho(t, x) = V_j(t, x) - \eta^2 |x|^2$. Then we define

$$M^\rho = \sup_{j,t,x}(\hat{U}_j^\rho - \hat{V}_j^\rho),$$

which is attained at some point $(\hat{j}^\rho, \hat{t}^\rho, \hat{x}^\rho)$. Recall that for any ρ, we can assume without loss of generality that $\{(t^{\varepsilon,\rho}, s^{\varepsilon,\rho}, x^{\varepsilon,\rho}, y^{\varepsilon,\rho})\}_\varepsilon$ converges to $(t^\rho, t^\rho, x^\rho, x^\rho)$ as $\varepsilon \to 0$ and $j^{\varepsilon,\rho} = j^\rho$ for all ε. Then the definition of $M^{\varepsilon,\rho}$ gives us that

$$M^\rho = (\hat{U}_{\hat{j}^\rho}^\rho - \hat{V}_{\hat{j}^\rho}^\rho)(\hat{t}^\rho, \hat{x}^\rho) \le M^{\varepsilon,\rho} = U_{j^\rho}(t^{\varepsilon,\rho}, x^{\varepsilon,\rho}) - V_{j^\rho}(s^{\varepsilon,\rho}, y^{\varepsilon,\rho})$$

$$- \frac{|x^{\varepsilon,\rho} - y^{\varepsilon,\rho}|^2}{\varepsilon^2} - \frac{|t^{\varepsilon,\rho} - s^{\varepsilon,\rho}|^2}{\varepsilon^2} - \rho^2(|x^{\varepsilon,\rho}|^2 + |y^{\varepsilon,\rho}|^2). \tag{7.6}$$

Define $\bar{l}_\rho = \limsup_{\varepsilon \to 0} \frac{|x^{\varepsilon,\rho} - y^{\varepsilon,\rho}|^2}{\varepsilon^2}$ and $\underline{l}_\rho = \liminf_{\varepsilon \to 0} \frac{|x^{\varepsilon,\rho} - y^{\varepsilon,\rho}|^2}{\varepsilon^2}$, then we obtain from (7.6) that

$$
\begin{aligned}
0 \le \underline{l}_\rho \le \bar{l}_\rho &\le \limsup_{\varepsilon \to 0} \big(U_{j^\rho}(t^{\varepsilon,\rho}, x^{\varepsilon,\rho}) - V_{j^\rho}(s^{\varepsilon,\rho}, y^{\varepsilon,\rho}) \\
&\quad - \rho^2(|x^{\varepsilon,\rho}|^2 + |y^{\varepsilon,\rho}|^2)\big) - (\hat{U}^\rho_{\hat{j}\rho} - \hat{V}^\rho_{\hat{j}\rho})(\hat{t}^\rho, \hat{x}^\rho) \\
&\le (\hat{U}^\rho_{j^\rho} - \hat{V}^\rho_{j^\rho})(t^\rho, x^\rho) - (\hat{U}^\rho_{\hat{j}\rho} - \hat{V}^\rho_{\hat{j}\rho})(\hat{t}^\rho, \hat{x}^\rho) \le 0,
\end{aligned}
$$

where we have used the semicontinuity of U_{j^ρ} and V_{j^ρ}. Similarly, we can derive $\lim_{\varepsilon \to 0} \frac{|t^{\varepsilon,\rho} - s^{\varepsilon,\rho}|^2}{\varepsilon^2} = 0$, which along with (7.6) implies that $\lim_{\varepsilon \to 0} M^{\varepsilon,\rho} = M^\rho$. The fact that $\lim_{\rho \to 0} M^\rho = M$ can be shown as in Reference [16].

Let us now prove $\lim_{\rho \to 0} \rho^2 |x^\rho|^2 = 0$. It holds for each $\rho > 0$ that

$$M^\rho = \lim_{\varepsilon \to 0} M^{\varepsilon,\rho} = \limsup_{\varepsilon \to 0} M^{\varepsilon,\rho} \le U_{j^\rho}(t^\rho, x^\rho) - V_{j^\rho}(t^\rho, x^\rho) - 2\rho^2 |x^\rho|^2 \le M^\rho, \tag{7.7}$$

and hence all inequalities in the above expression are in fact equalities. Thus we have

$$
\begin{aligned}
M^{\rho/2} = \sup_{j,t,x} \left[U_j(t,x) - V_j(t,x) - 2(\tfrac{\rho}{2})^2 |x|^2 \right] &\ge U_{j^\rho}(t^\rho, x^\rho) - V_{j^\rho}(t^\rho, x^\rho) \\
&\quad - 2\rho^2 |x^\rho|^2 + \tfrac{3}{2}\rho^2 |x^\rho|^2 \\
&= M^\rho + \tfrac{3}{2}\rho^2 |x^\rho|^2,
\end{aligned}
$$

which implies $0 \le \limsup_{\rho \to 0} \rho^2 |x^\rho|^2 \le \limsup_{\rho \to 0} \tfrac{2}{3}(M^{\rho/2} - M^\rho) = 0$.

Finally, we obtain from (7.7) and $\lim_{\varepsilon \to 0} \frac{|t^{\varepsilon,\rho} - s^{\varepsilon,\rho}|^2}{\varepsilon^2} = \lim_{\varepsilon \to 0} \frac{|x^{\varepsilon,\rho} - y^{\varepsilon,\rho}|^2}{\varepsilon^2} = 0$ that we have $\lim_{\varepsilon \to 0} U_{j^\rho}(t^{\varepsilon,\rho}, x^{\varepsilon,\rho}) - V_{j^\rho}(s^{\varepsilon,\rho}, y^{\varepsilon,\rho}) = U_{j^\rho}(t^\rho, x^\rho) - V_{j^\rho}(t^\rho, x^\rho)$, which together the semicontinuity of U_{j^ρ} and V_{j^ρ} implies $\limsup_{\varepsilon \to 0} U_{j^\rho}(t^{\varepsilon,\rho}, x^{\varepsilon,\rho}) = U_{j^\rho}(t^\rho, x^\rho)$ and $\liminf_{\varepsilon \to 0} V_{j^\rho}(s^{\varepsilon,\rho}, y^{\varepsilon,\rho}) = V_{j^\rho}(t^\rho, x^\rho)$. By extracting further subsequences if necessary, we complete our proof. $\qquad \square$

Lemma 7.2 *Let U (resp. V) be a bounded subsolution (resp. supersolution) of (3.6). For any given $\varepsilon, \rho > 0$, we consider the function $\psi_j^{\varepsilon,\rho}(t, s, x, y)$ as defined in (7.1) and $M^{\varepsilon,\rho} = \sup_{j,t,s,x,y} \psi_i^{\varepsilon,\rho}(t, s, x, y)$. If there exists an index $j^{\varepsilon,\rho}$ and a point $(t^{\varepsilon,\rho}, s^{\varepsilon,\rho}, x^{\varepsilon,\rho}, y^{\varepsilon,\rho}) \in (0, T]^2 \times \mathbb{R}^{2d}$ such that $\psi_{j^{\varepsilon,\rho}}(s^{\varepsilon,\rho}, t^{\varepsilon,\rho}, x^{\varepsilon,\rho}, y^{\varepsilon,\rho}) = M^{\varepsilon,\rho}$, then there exists an index $j_0^{\varepsilon,\rho} \in \{1, \dots, J\}$ such that*

$$\psi_{j_0^{\varepsilon,\rho}}(s^{\varepsilon,\rho}, t^{\varepsilon,\rho}, x^{\varepsilon,\rho}, y^{\varepsilon,\rho}) = M^{\varepsilon,\rho} \text{ and } U_{j_0^{\varepsilon,\rho}}(t^{\varepsilon,\rho}, x^{\varepsilon,\rho}) > \mathcal{M}_{j_0^{\varepsilon,\rho}}(t^{\varepsilon,\rho}, x^{\varepsilon,\rho}).$$

References

1. Barles, G.: Convergence of numerical schemes for degenerate parabolic equations arising in finance. In: Rogers, L.C.G., Talay, D. (eds.) Numerical Methods in Finance, pp. 1–21. Cambridge University Press, Cambridge (1997)
2. Barles, G., Jakobsen, E.R.: Error bounds for monotone approximation schemes for parabolic Hamilton–Jacobi–Bellman equations. Math. Comp. **76**, 1861–1893 (2007)
3. Barles, G., Souganidis, P.E.: Convergence of approximation schemes for fully nonlinear second order equations. Asymptot. Anal. **4**, 271–283 (1991)
4. Barles, G., Buckdahn, R., Pardoux, E.: Backward stochastic differential equations and integral-partial differential equations. Stoch. Stoch. Rep. **60**, 57–83 (1997)
5. Biswas, I., Chowdhury, I., Jakobsen, E.R.: On the Rate of Convergence for Monotone Numerical Schemes for Nonlocal Isaacs' Equations, arXiv preprint arxiv:1709.07743 [math.AP]
6. Biswas, I., Jakobsen, E.R., Karlsen, K.H.: Viscosity solutions for a system of integro-PDEs and connections to optimal switching and control of jump-diffusion processes. Appl. Math. Optim. **62**, 47–80 (2010)
7. Biswas, I.H., Jakobsen, E.R., Karlsen, K.H.: Difference-quadrature schemes for nonlinear degenerate parabolic integro-PDE. SIAM J. Numer. Anal. **48**, 1110–1135 (2010)
8. Buckdahn, R., Li, J.: Stochastic differential games and viscosity solutions of Hamilton–Jacobi–Bellman-Isaacs equations. SIAM J. Contr. Opt. **47**, 444–475 (2008)
9. Chen, Z., Epstein, L.: Ambiguity, risk, and asset returns in continuous time. Econometrica **70**, 1403–1443 (2002)
10. Cont, R., Voltchkova, E.: A finite difference scheme for option pricing in jump diffusion and exponential Lévy models. SIAM J. Numer. Anal. **43**, 1596–1626 (2005)
11. Crandall, M.G., Lions, P.-L.: Two approximations of solutions of Hamilton–Jacobi equations. Math. Comput. **43**, 1–19 (1984)
12. Crandall, M.G., Ishii, H., Lions, P.-L.: User's guide to viscosity solutions of second order partial differential equations. Bull. Am. Math. Soc. **27**, 1–67 (1992)
13. Debrabant, K., Jakobsen, E.R.: Semi-Lagrangian schemes for linear and fully nonlinear diffusion equations. Math. Comp. **82**, 1433–1462 (2012)
14. d'Halluin, Y., Forsyth, P.A., Labahn, G.: A penalty method for American options with jump diffusion processes. Numer. Math. **97**, 321–352 (2004)
15. Di Nunno, G., Khedher, A., Vanmaele, M.: Robustness of quadratic hedging strategies in finance via backward stochastic differential equations with jumps. Appl. Math. Optim. **72**, 353–389 (2015)
16. Dumitrescu, R., Quenez, M.-C., Sulèm, A.: Optimal stopping for dynamic risk measures with jumps and obstacle problems. J. Optim. Theory Appl. **167**, 219–242 (2015)
17. Dumitrescu, R., Quenez, M.-C., Sulèm, A.: A weak dynamic programming principle for combined optimal stopping/ stochastic control with $\mathcal{E}^f$-expectations. SIAM J. Control Optim. **54**, 2090–2115 (2016)
18. Dumitrescu, R., Quenez, M.-C., Sulèm, A.: Mixed generalized Dynkin game and stochastic control in a Markovian framework. Stochastics **89**, 400–429 (2016)
19. Fritsch, F.N., Carlson, R.E.: Monotone piecewise cubic interpolation. SIAM J. Numer. Anal. **17**, 238–246 (1980)
20. Godunov, S.K.: A difference scheme for the numerical computation of a discontinuous solution of the hydrodynamic equations. Math. Sb. **47**, 271–306 (1959)
21. Ito, K., Kunisch, K.: Semi-smooth Newton methods for variational inequalities of the first kind, M2AN Math. Model. Numer. Anal. **37**, 41–62 (2003)
22. Jakobsen, E.R., Karlsen, K.H.: Continuous dependence estimates for viscosity solutions of integro-PDEs. J. Differ. Equ. **212**, 278–318 (2005)
23. El Karoui, N., Pardoux, E., Quenez, M.-C.: Reflected backward SDEs and American options. In: Robers, L., Talay, D. (eds.) Numerical Methods in Finance, pp. 215–231. Cambridge University Press, Cambridge (1997)
24. Karoui, N. El., Peng, S., Quenez, M.-C.: Backward stochastic differential equations in finance, Pre-publication No. 260 du Lab. de Probabilité (1994)
25. Katzourakis, N.: An Introduction to Viscosity Solutions for Fully Nonlinear PDE with Applications to Calculus of Variations in L^∞. Springer, Cham (2015)

26. Kharroubi, I., Langrene, N., Pham, H.: Discrete time approximation of fully nonlinear HJB equations via BSDEs with nonpositive jumps. Ann. Appl. Probab. **25**, 2301–2338 (2015)
27. Kruse, T., Popier, A.: BSDEs with monotone generator driven by Brownian and Poisson noises in a general filtration. Stochastics **88**(4), 491–539 (2016)
28. Qi, L., Sun, J.: A nonsmooth version of Newton's method. Math. Progr. **58**, 353–367 (1993)
29. Reisinger, C., Forsyth, P.A.: Piecewise constant policy approximations to Hamilton–Jacobi–Bellman equations. Appl. Numer. Math. **103**, 27–47 (2016)
30. Gianin, E.R.: Risk measures via g-expectations. Insur. Math. Econom. **39**, 19–34 (2006)
31. Royer, M.: Backward stochastic differential equations with jumps and related non-linear expectations. Stoch. Process. Appl. **116**, 1358–1376 (2006)
32. Quenez, M.-C., Sulèm, A.: BSDEs with jumps, optimization and applications to dynamic risk measures. Stoch. Process. Appl. **123**, 3328–3357 (2013)
33. Witte, J.H., Reisinger, C.: A penalty method for the numerical solution of Hamilton–Jacobi–Bellman (HJB) equations in finance. SIAM J. Numer. Anal. **49**, 213–231 (2011)
34. Xu, H., Xie, S.: A semismooth Newton method for a kind of HJB equation. Comput. Math. Appl. **73**, 2581–2586 (2017)