On the rate of convergence for monotone numerical schemes for nonlocal Isaacs' equations

Imran H. Biswas[a], Indranil Chowdhury[a], Espen R. Jakobsen[b]

[a] *Centre for Applicable Mathematics, Tata Instiute of Fundamental Research, P.O. Box 6503, Bangalore 560065, India*
[b] *Norwegian University of Science and Technology, NO–7491, Trondheim, Norway*

Abstract

We study monotone numerical schemes for nonlocal Isaacs equations, the dynamic programming equations of stochastic differential games with jump-diffusion state processes. These equations are fully-nonlinear non-convex equations of order less than 2. In this paper they are also allowed to be degenerate and have non-smooth solutions. The main contribution is a series of new a priori error estimates: The first results for *nonlocal* Isaacs equations, the first general results for *degenerate* non-convex equations of order greater than 1, and the first results in the viscosity solution setting giving the *precise dependence* on the fractional order of the equation. We also observe a new phenomena, that the rates differ when the nonlocal diffusion coefficient depend on x and t, only on x, or on neither.

Keywords: Fractional and nonlocal equations, Isaacs equations, stochastic differential games, error estimate, viscosity solution, monotone scheme, rate of convergence.
2010 MSC: 45K05, 46S50, 49L20, 49L25, 91A23, 93E20

1. Introduction

In this paper we obtain error estimates for monotone approximation schemes for nonlocal Isaacs-Bellman equations originating from optimal stochastic control and differential game theory:

$$u_t + \inf_{\alpha\in\mathcal{A}} \sup_{\beta\in\mathcal{B}} \Big\{ -f^{\alpha,\beta}(t,x) + c^{\alpha,\beta}(t,x)u(t,x) - b^{\alpha,\beta}(t,x)\cdot\nabla u(t,x) - \mathcal{I}^{\alpha,\beta}[u](t,x) \Big\} = 0 \quad \text{in } Q_T, \tag{1.1}$$

$$u(0,x) = u_0(x) \quad \text{in } \mathbb{R}^N, \tag{1.2}$$

where $\mathcal{I}^{\alpha,\beta}$ is a nonlocal operator defined by

$$\mathcal{I}^{\alpha,\beta}[\phi](t,x) := \int_{\mathbb{R}^M\setminus\{0\}} \Big(\phi(t, x + \eta^{\alpha,\beta}(t,x;z)) - \phi(t,x) - \eta^{\alpha,\beta}(t,x;z)\cdot\nabla_x\phi(t,x) \Big)\, \nu(dz) \tag{1.3}$$

for smooth bounded functions ϕ. Here $Q_T := (0,T]\times\mathbb{R}^N$, $\mathcal{A}$ and $\mathcal{B}$ are metric spaces, $f^{\alpha,\beta}, c^{\alpha,\beta}, b^{\alpha,\beta}, \eta^{\alpha,\beta}$ are $\mathbb{R}, \mathbb{R}, \mathbb{R}^N$, and $\mathbb{R}^N$ valued functions respectively, while the Lévy measure ν is a nonnegative Radon measure satisfying the Lévy integrability assumption (**A**.4) in Section 2.

The diffusion part of this equation ($\mathcal{I}^{\alpha,\beta}$) is purely nonlocal, and under the assumptions of Section 2, $\mathcal{I}^{\alpha,\beta}$ is a non-positive fractional differential operator of order $\sigma\in[0,2)$. The fractional Laplacian $-(-\Delta)^{\frac{\sigma}{2}}$ is not covered, but all similar operators coming from tempered or truncated processes are. In particular almost all non-local operators appearing in finance are included [15]. In general this equation is a fully nonlinear, non-convex, nonlocal PDE (an integro-PDE) that may have any order $\sigma\in[0,2)$. In particular, it may have order greater than one. Moreover, since we also allow the equations to be degenerate, solutions are typically not smooth. Under Lipschitz type regularity assumptions on the coefficients and data, the problems are well-posed in the viscosity solution sense [16] having merely Hölder or Lipschitz continuous solutions. First and fractional derivatives need not exist. The precise assumptions and results can be found in Section 2. The literature on viscosity solutions and nonlocal PDEs is by now very large, but the results we will need here are mainly covered by [22, 1] and the references therein.

The study of Isaacs and Bellman equations is primarily motivated by their connection with the theory of stochastic differential games and stochastic control. Equations of the form (1.1) appear when the state dynamics is given by a controlled Lévy-Itô type SDE driven by a pure-jump Lévy process. By the dynamic programming principle (DPP), the value functions of such games satisfy nonlocal PDEs of the form (1.1). These equations are called the Isaacs or DPP equations for the differential games. We refer to [6, 18, 19] for more on differential games and dynamic programming equations. Note that if $\eta^{\alpha,\beta} \equiv 0$ or $\nu \equiv 0$, then there is no diffusion and (1.1) becomes the widely studied first order Isaacs equation corresponding to a deterministic game (see e.g. [18]). If the state process is driven by a Brownian motion, then the related DPP equation is a second order PDE (cf. [19]). This case will not be considered here.

The numerical approximations we consider here are monotone finite-difference quadrature methods in the spirit of e.g. [8]. We refer to (3.8) in Section 3 for the precise form of these approximations. The main contribution of this paper is a series of new and very accurate error estimates in this setting. If solutions are Lipschitz continuous, then these estimates may take the form

$$\|U-u\|_{L^\infty(Q_T)} \leq C_T \begin{cases} \Delta t^{\frac{1}{2}} + \Delta x^{\frac{1}{2}} & \text{if } \sigma \in [0,1), \\ \Delta t^{\frac{1}{2}} + \Delta x^{\frac{1}{2}} |\ln \Delta x| & \text{if } \sigma = 1, \\ \Delta t^{\frac{1}{2}} + \Delta x^{\frac{2-\sigma}{2\sigma}} & \text{if } \sigma \in (1,2), \end{cases} \tag{1.4}$$

where $\sigma \in [0,2)$ is the order of the nonlocal term, and $\Delta t > 0, \Delta x > 0$ are time and space grid parameters. In general solutions are only Hölder continuous in time, and then also the rates in time may depend on σ. Surprisingly, we also discover a new phenomenon. When $\sigma \in (1,2)$, the convergence rates differ depending on whether η depends on (x,t), only on x, or on neither! We find in Remark 3.4 that

$$\|U-u\|_{L^\infty(Q_T)} \leq C \begin{cases} (\Delta t)^{\frac{2-\sigma}{2\sigma}} + (\Delta x)^{\frac{2-\sigma}{2\sigma}} & \text{when } \eta \text{ depends on } x,t, \\ (\Delta t)^{\frac{1}{2\sigma}} + (\Delta x)^{\frac{2-\sigma}{2\sigma}} & \text{when } \eta \text{ only depends on } x, \\ (\Delta t)^{\frac{1}{2\sigma}} + (\Delta x)^{\frac{2-\sigma}{2}} & \text{when } \eta \text{ does not depend on } x,t. \end{cases} \tag{1.5}$$

Precise statements and results are given in Section 3.

The study of numerical approximation in the context of viscosity solutions began in the early eighties with pioneering papers of Lions, Crandall and others. In some of the early papers [13, 17, 29, 30], the authors obtained a priori error estimates for consistent monotone schemes for first order HJB equations. These results are derived through suitable modifications of the viscosity solution uniqueness proofs for the corresponding equations. These arguments can not be extended to 2nd order equations, and it took more than a decade before a solution was found by N. V. Krylov. In a series of articles [24, 25, 26], Krylov introduced the method of shaking the coefficients and was able to establish error estimates for a class of monotone schemes for convex second order HJB equations. These results were then extended and complemented by Barles & Jakobsen in [2, 3, 4]. In all of these papers, and the many others building upon them, convexity and a type of Jensen's inequality is crucial.

For non-convex equations like the Isaacs equation, there are no general results giving error estimates for numerical methods. However, in special cases there are some results: In one space dimension [20], for special types of non-convex equations [10, 21], and for uniformly elliptic/parabolic equations [12, 27, 31, 32]. In the two first cases the proofs rely on the special structure of the problems (one dimension and not too non-convex) and are not suitable for general equations/dimensions, while in the last case it relies on some type of elliptic regularity. This last direction of research was initiated by Cafferelli and Souganidis in [12] (but see also [27]), where they obtain an (unknown) algebraic rate of convergence for equations with rather general non-convex nonlinearities. In spite of all these results, it seems that the problem is very far from understood in the case of general, possibly degenerate, Isaacs equations.

The story of nonlocal Bellman-Isaacs equations is a more recent one and there is already a significant literature addressing the issues of numerical approximations and the related error analysis. Most of the development in this direction have taken place in the last ten years, see e.g. [7, 8, 23] for general error estimates for convex and nonlocal HJB equations. These results are extensions of the results for local 2nd order equations (Krylov-Barles-Jakobsen type theory) and convexity is again crucial. For non-convex nonlocal problems there are no results on error estimates as far as we know.

At this point, we note that convexity is not playing any role in the proof of the error estimates for first order equations. But, as we have already mentioned, these techniques do not work for 2nd order problems. However, for a different class of equations and weak solution concept (nonlinear convection-diffusion equations and entropy solutions), it was noticed in [14] that first order error estimation techniques surprisingly could work also for nonlocal/fractional problems of any order less than 2. At least for certain natural numerical approximations. Is it possible to do similar things also for the nonlocal Isaacs equations (1.1) and in a viscosity solution setting? The goal of this paper is to investigate if, and to what extent, we can extend first order error estimation techniques to nonlocal Isaacs equations (1.1) of any order less than two.

Because of the nonlocal term, the analysis necessarily becomes much more involved than in the first order case, and it leads (as usual) to 3 different cases: (i) The supercritical case where $\sigma \in [0, 1)$ and drift/convection dominates, (ii) the critical case $\sigma = 1$ where drift and diffusion is in balance, and (iii) the subcritical case where $\sigma \in (1, 2)$ and diffusion dominates. In this paper we give precise and rigorous error estimates in all cases, cf. e.g. (1.4) and (1.5). In case (i) we get the same (and hence the optimal) rate as for first order equations [13, 17, 30]. In case (ii) we get a rate with a logarithm, and in case (iii) we find a rate depending on σ. Under certain conditions these rates are consistent with the rates in [14]. Note that the rates go to 0 when $\sigma \to 2$. This behaviour is correct and is an artifact of the numerical method. Under our low regularity assumptions, these results are the best possible results for this method. In case (iii) (cf. (1.5)) we also observe that the rates differ according to whether η depend on x and t, only on x, or on neither of them. This is a new phenomenon that is not present for local equations. To summarize, the main novelties of this paper are:

1. The first error estimates for numerical schemes for *nonlocal* Isaacs equations.
2. The first error bounds for general *degenerate* non-convex equations of order greater than 1.
3. The first error bounds for a numerical scheme in the viscosity solution setting giving the *precise dependence* of the order σ of the nonlocal term.
4. The first error bounds where the rates depend on whether the jump term η depend on (x, t), only on x, or on neither.

As a part of our effort to get precise estimates correctly reflecting the fractional order σ of the nonlocal term, we also prove a new and refined time regularity result for viscosity solutions.

The rest of the paper is organised as follows. In Section 2, we list the assumptions and state the well-posedness result and a priori estimates for (1.1)–(1.2), including the new and more accurate time regularity result. In Section 3, we introduce the schemes, establish properties such as wellposedness, consistency, monotonicity and stability, and state our main results, the error estimates. The proof of the these estimates are given in Section 4. In Section 5, the last section of the paper, we explain how our techniques can be used to obtain error estimates for a larger class of monotone approximations of (1.1). But this extension comes at a price, the rates for more accurate schemes will be suboptimal.

2. Preliminaries

In this section we state our main assumptions, define the relevant concept of solutions – viscosity solutions, and state and partially prove a wellposedness result for (1.1)-(1.2). We start with some notation. By

C, K we mean various constants which may change from line to line. The Euclidean norm on any $\mathbb{R}^d$-type space is denoted by $|\cdot|$. For any subset $Q \subset \mathbb{R} \times \mathbb{R}^N$ and for any bounded, possibly vector valued, function on Q, we define the following norms,

$$\|w\|_0 := \sup_{(t,x)\in Q} |w(t,x)|,$$
$$\|w\|_1 := \|w\|_0 + \sup_{(t,x)\neq(s,y)} \frac{|w(t,x) - w(s,y)|}{|t-s| + |x-y|}.$$

Note that if w is independent of t, then $\|w\|_1$ is the Lipschitz (or $W^{1,\infty}$) norm of w. We use $C_b(Q)$ to denote the space of bounded continuous real valued functions on Q. We use the notation h to denote the vector $(\Delta t, \Delta x)$ involving the mesh parameters, and any dependence on Δt, Δx will be denoted by subscript h. The grid is denoted by $\mathcal{G}_h$ and is a subset of $\bar{Q}_T$ which need not be uniform or even discrete in general. We also set $\mathcal{G}_h^0 = \mathcal{G}_h \cap \{t = 0\}$ and $\mathcal{G}_h^+ = \mathcal{G}_h \cap \{t > 0\}$.

We now list the working assumptions of this paper. These are sufficient for the wellposedness and regularity results for (1.1)–(1.2).

(**A**.1) The sets $\mathcal{A}, \mathcal{B}$ are separable metric spaces, $c^{\alpha,\beta}(t,x) \geq 0$, and $c^{\alpha,\beta}(t,x)$, $f^{\alpha,\beta}(t,x)$, $b^{\alpha,\beta}(t,x)$ and $\eta^{\alpha,\beta}(t,x;z)$ are continuous in α, β, t, x and z.

(**A**.2) There exists a constant $K > 0$ such that for every α, β ,

$$\|u_0\|_1 + \|f^{\alpha,\beta}\|_1 + \|c^{\alpha,\beta}\|_1 + \|b^{\alpha,\beta}\|_1 \leq K.$$

(**A**.3) For $x, y \in \mathbb{R}^N$, $\alpha \in \mathcal{A}, \beta \in \mathcal{B}$ and $z \in \mathbb{R}^M$, there is a function $\rho(z) \geq 0$ such that

$$|\eta^{\alpha,\beta}(t,x;z) - \eta^{\alpha,\beta}(s,y;z)| \leq \rho(z)\,(|x-y| + |t-s|) \quad \text{and} \quad |\eta^{\alpha,\beta}(t,x;z)| \leq \rho(z)$$

and

$$|\rho(z)| \leq K|z| \quad \text{for} \quad |z| < 1 \qquad \text{and} \qquad 1 \leq \rho(z) \leq \rho(z)^2 \quad \text{for} \quad |z| > 1.$$

(**A**.4) The Lévy measure ν is a nonnegative Radon measure on $(\mathbb{R}^M, \mathcal{B}(\mathbb{R}^M))$ satisfying

$$\int_{|z|<1} |z|^2 \nu(\,dz) + \int_{|z|>1} \rho(z)^2 \nu(\,dz) < \infty.$$

(**A**.5) There is a $\sigma \in (0,2)$, a constant $C > 0$, and density $k(z)$ of $\nu(dz)$ for $|z| < 1$ satisfying

$$0 \leq k(z) \leq \frac{C}{|z|^{M+\sigma}} \qquad \text{for} \qquad |z| < 1.$$

Remark 2.1. (a) Typical examples are $\eta = \bar{\eta}(x)z$ and $\eta = \bar{\eta}(x)(e^z - 1)$, and for ν,

$$\nu(dz) = \frac{c_\sigma e^{-K|z|} dz}{|z|^{N+\sigma}} \quad \text{and} \quad \nu(dz) = 1_{|z|<1} \frac{c_\sigma dz}{|z|^{N+\sigma}}$$

for $\sigma \in (0,2)$, i.e. tempered or truncated σ-stable Lévy measures. Near $z = 0$ these Lévy measures behave as the Lévy measure associated to the fractional Laplacian $(-\Delta)^{\sigma/2}$, and their (pseudo-differential) orders is σ as it is for $(-\Delta)^{\sigma/2}$. We will see that we get different estimates when $\sigma < 1$, $\sigma = 1$, or $\sigma > 1$.

(b) Assumptions (**A**.3), (**A**.4), and (**A**.5) are quite general and encompass most models from finance [15], and under (**A**.3) and (**A**.4) there is a standard viscosity solution theory for (1.1). Note that assumption (**A**.5) only requires an upper bound on the density. This bound is needed to get an explicit convergence rate.

(c) All assumptions can be relaxed in such a way that our techniques and results would still apply: (**A**.3) and (**A**.5) can be replaced by more general integral conditions like $\int |\eta(t,x;z)-\eta(s,y;z)|^2\nu(dz) \le L(|x-y|+|t-s|)$, $\int_{|z|<r} |\eta(t,x;z)|^2\nu(dz) \le Kr^{2-\sigma}$, etc., and (**A**.4) can be relaxed when it comes to the integrability at infinity and absolute continuity. This is somehow straight forward, but we omit it since it would obscure the message and make the paper much longer and more technical.

We now give the definition of viscosity solution for (1.1)-(1.2). To this end, we define

$$
\begin{aligned}
\mathcal{I}^{\alpha,\beta}_{\kappa}[\phi](t,x) &= \int_{B(0,\kappa)} \Big(\phi(t,x+\eta^{\alpha,\beta}(t,x;z)) - \phi(t,x) - \eta^{\alpha,\beta}(t,x;z)\cdot\nabla_x\phi(t,x)\Big)\ \nu(dz)\ ,\\
\mathcal{I}^{\alpha,\beta,\kappa}[u;p](t,x) &= \int_{\mathbb{R}^M\setminus B(0,\kappa)} \Big(u(t,x+\eta^{\alpha,\beta}(t,x;z)) - u(t,x) - \eta^{\alpha,\beta}(t,x;z)\cdot p\Big)\ \nu(dz)\ ,
\end{aligned} \tag{2.1}
$$

for $(\alpha,\beta)\in\mathcal{A}\times\mathcal{B}$, $\kappa\in(0,1)$, $\phi\in C^2$, and bounded semicontinuous functions u. By (**A**.3)–(**A**.4), $\mathcal{I}^{\alpha,\beta,\kappa}[u;p]$ and $\mathcal{I}^{\alpha,\beta}_{\kappa}[\phi]$ are well-defined, in the first case since $\int_{|z|>\kappa}\nu(dz)<\infty$ and in the second case since

$$
|\mathcal{I}^{\alpha,\beta}_{\kappa}[\phi](x,t)| \le \frac{1}{2}\|D^2\phi(\cdot,t)\|_{L^\infty(B(x,\kappa))}\int_{|z|<\kappa} K^2\rho(z)^2\nu(dz) < \infty.
$$

Definition 2.1. (i) A function $u\in USC_b(Q_T)$ is a viscosity subsolution of (1.1) if for any $k\in(0,1)$, $\phi\in C^2(Q_T)$, and global maximum point $(t,x)\in Q_T$ of $u-\phi$,

$$
\begin{aligned}
\phi_t(t,x) + \inf_{\alpha}\sup_{\beta}\Big\{ &-f^{\alpha,\beta}(t,x) + c^{\alpha,\beta}(t,x)u(t,x) - b^{\alpha,\beta}(t,x).\nabla\phi(t,x)\\
&- \mathcal{I}^{\alpha,\beta}_{k}[\phi](t,x) - \mathcal{I}^{\alpha,\beta,k}[u,\nabla_x\phi(t,x)](t,x)\Big\} \le 0.
\end{aligned}
$$

(ii) A function $v\in LSC_b(Q_T)$ is a viscosity supersolution of (1.1) if for any $k\in(0,1)$, $\psi\in C^2(Q_T)$, global minimum point $(t,x)\in Q_T$ of $v-\psi$,

$$
\begin{aligned}
\psi_t(t,x) + \inf_{\alpha}\sup_{\beta}\Big\{ &-f^{\alpha,\beta}(t,x) + c^{\alpha,\beta}(t,x)v(t,x) - b^{\alpha,\beta}(t,x).\nabla\psi(t,x)\\
&- \mathcal{I}^{\alpha,\beta}_{k}[\psi](t,x) - \mathcal{I}^{\alpha,\beta,k}[v,\nabla_x\psi(t,x)](t,x)\Big\} \ge 0.
\end{aligned}
$$

(iii) A function $w\in C_b(Q_T)$ is a viscosity solution of (1.1) if it is both a sub and supersolution.

We then have the following wellposedness and Lipschitz/Hölder regularity results for (1.1).

Theorem 2.1. *Assume* (**A**.1)–(**A**.4) *hold.*

(a) If u and v are respectively viscosity sub and supersolutions of (1.1) *with $u(0,\cdot)\le v(0,\cdot)$, then $u\le v$.*

(b) There exists a unique bounded viscosity solution u of the initial value problem (1.1)–(1.2).

(c) There is a constant $K\ge 0$ such that the solution u from (b) satisfies for all $x,y\in\mathbb{R}^N$, $t,s\in[0,T]$,

$$
|u(x,t)-u(y,s)| \le K\Big(|x-y| + \bar\omega(t-s)\Big) \qquad \textit{where} \qquad \bar\omega(r) := \begin{cases} |r| & \textit{if } \sigma\in[0,1),\\ |r|(1+|\ln r|) & \textit{if } \sigma=1,\\ |r|^{\frac{1}{\sigma}} & \textit{if } \sigma\in(1,2). \end{cases} \tag{2.2}
$$

(d) Assume in addition

$$
K(u_0) := \sup_{\alpha,\beta}\Big\|\mathcal{I}^{\alpha,\beta}[u_0]\Big\|_{L^\infty([0,T]\times\mathbb{R}^N)} < \infty.
$$

Then there is $C\ge 0$ depending only on the data (**A**.1)–(**A**.5) *such that the solution u from (b) satisfies for all $x,y\in\mathbb{R}^N$, $t,s\in[0,T]$,*

$$
|u(x,t)-u(y,s)| \le C\Big(|x-y| + (1+K(u_0))|t-s|\Big).
$$

The wellposedness and x-regularity results are quite standard, but the time regularity results are new and more precise than earlier results. These time regularity results are somewhat parallel to the results in Lemma 5.4 in [14], but the equation, norm and solution concepts are different.

Remark 2.2. Under assumptions (A.3) and (A.4), either (i) $v \in W^{2,\infty}(\mathbb{R}^N)$, or (ii) $v \in W^{1,\infty}(\mathbb{R}^N)$ and (A.5) holds with $\sigma < 1$, are sufficient conditions for $K(v) < \infty$. See Lemma 2.2 below.

In the proof of Theorem 2.1 we will need the following lemma.

Lemma 2.2. *Assume (A.3)–(A.5). Then there is a constant $C > 0$ such that for all $\phi \in C_b^2(\mathbb{R}^N)$ and $\epsilon \in (0, 1)$,*

$$K(\phi) \leq \begin{cases} C\Big(\epsilon^{2-\sigma}\|D^2\phi\|_0 + (1 + \epsilon^{1-\sigma})\|D\phi\|_0\Big), & \text{if } \sigma \in (1, 2), \\ C\Big(\epsilon\|D^2\phi\|_0 + (1 + |\ln \epsilon|)\|D\phi\|_0\Big), & \text{if } \sigma = 1, \\ C\|D\phi\|_0, & \text{if } \sigma \in [0, 1). \end{cases}$$

Proof. When $\sigma < 1$, then $|\mathcal{I}^{\alpha,\beta}\phi(x)| \leq C\|D\phi\|_0 \int |\eta^{\alpha,\beta}(t, x, z)|\, \nu(dz)$. Since $\int \eta^{\alpha,\beta}\nu(dz) \leq \int \rho(z)\nu(dz) < \infty$ by (A.3) and (A.4), the bound on $K(\phi)$ follows by taking the supremum over x, α, β. For $\sigma \geq 1$, we split the integral in three parts and use Taylor's theorem:

$$\begin{aligned} \mathcal{I}^{\alpha,\beta}[\phi] &= \int \Big(\phi(x + \eta) - \phi(x) - \eta\nabla\phi(x)\Big)\, \nu(dz) \\ &= \int_{|z|<\epsilon} \int_0^1 (1 - t)\eta^T D^2\phi(x + t\eta)\eta\, dt\, \nu(dz) + \Big(\int_{\epsilon\leq|z|<1} + \int_{|z|\geq1}\Big) \int_0^1 \Big(\nabla\phi(x + t\eta) - \nabla\phi(x)\Big)\eta\, dt\, \nu(dz). \end{aligned}$$

By assumption (A.3) – (A.5), it follows that

$$\mathcal{I}^{\alpha,\beta}[\phi] \leq C\|D^2\phi\|_0 \int_{|z|<\epsilon} |z|^2 \frac{dz}{|z|^{N+\sigma}} + C\|D\phi\|_0\Big(\int_{\epsilon<|z|<1} |z|\frac{dz}{|z|^{N+\sigma}} + \int_{|z|\geq1} \rho(z)\nu(dz)\Big).$$

By (A.3) and (A.4), the last integral is finite, and the result then follows from computing the two first integrals in polar coordinates and taking the supremum over x, α, β. □

Proof of Theorem 2.1. We refer to Theorem 3.1 of the article [22] for a proof of part (a) and x-regularity part of (c) and (d). Part (b) then follows e.g. from Perron's method [9]. Time regularity in part (c) and (d) is new. We start by proving (d) and then use this result to prove (c).

(d) First we show Lipschitz in time at $t = 0$ by using the comparison principle and the fact that $w^\pm(t, x) = u_0(x) \pm Ct$ are super- and subsolutions of (1.1) if C is large enough. To see this, insert $w^\pm$ into the equation and use the regularity of u_0 to conclude. Here the assumption $K(u_0) < \infty$ is crucial and minimal. To get Lipschitz regularity for all times, we use a continuous dependence result and the t-Lipschitz regularity of the coefficients. See Theorem 5.1 and Theorem 5.3 of [22] for the details, and note that there is no growth in x of the estimates here since the coefficients and solutions are bounded.

(c) Let $0 < \epsilon < 1$ and regularize (by mollification) the initial data to get $u_0^\epsilon \in C_b^\infty(\mathbb{R}^N)$ satisfying $\|D^k u_0^\epsilon\|_0 \leq C\epsilon^{1-k}$ and $\|u_0 - u_0^\epsilon\|_0 \leq \epsilon$ (since u_0 is Lipschitz). Then let u^ϵ be the corresponding solution of (1.1)–(1.2). By (a) again $|u - u^\epsilon| \leq \|u_0^\epsilon - u_0\|_0 \leq C\epsilon$, and by the estimates on $D^k u_0^\epsilon$ and Lemma 2.2 with $\phi = u_0^\epsilon$,

$$K(u_0^\epsilon) \leq C \begin{cases} 1 & \text{if } \sigma \in [0, 1), \\ (1 + |\ln \epsilon|) & \text{if } \sigma = 1, \\ \epsilon^{1-\sigma} & \text{if } \sigma \in (1, 2). \end{cases} \tag{2.3}$$

By part (d) we have that $|u^\epsilon(t,x) - u_0^\epsilon(x)| \leq C(1 + K(u_0^\epsilon))t$, and by the triangle inequality

$$|u(t,x) - u_0(x)| \leq C(\epsilon + K(u_0^\epsilon)t + \epsilon).$$

When $\sigma < 1$, $\sigma = 1$, and $\sigma > 1$, we take $\epsilon = 0$, $\epsilon = t$, and $\epsilon = t^{\frac{1}{\sigma}}$ respectively. This proves the result for $s = 0$, $t \in [0,1]$ (and $x = y$). The result trivially holds for $s = 0, t > 1$, since then e.g. $|u(x,t) - u(x,0)| \leq 2\|u\|_0 t^{\frac{1}{\sigma}}$. The general result then follows from the t-Lipschitz regularity of the coefficients and the same continuous dependence result as in part (d). $\square$

3. The main results: Error estimates for a monotone scheme

In this section, we introduce a natural monotone difference-quadrature scheme for (1.1). The time discretizations include explicit, implicit and explicit-implicit schemes. For these schemes we prove wellposedness, L^∞-stability, and the main results, several estimates on their rates of convergence in L^∞.

For simplicity we consider a uniform grid in space and time. For $M > 0$, let $\Delta x > 0$ and $\Delta t := \frac{T}{M}$ be the discretization parameters/mesh size in the time and space and $h = (\Delta t, \Delta x)$. The corresponding mesh is

$$\mathcal{G}_h^N = \{(t_n, x_{\mathbf{m}}) : t_n = n\Delta t, x_{\mathbf{m}} = \mathbf{m}\,\Delta x;\ \mathbf{m} \in \mathbb{Z}^N, n = 0, 1, ..., M\}.$$

To obtain a full discretization of (1.1), we follow [8] and perform the following steps:

Step 1. Approximate singular diffusion by bounded diffusion. For $\delta \geq \Delta x$ we approximate $\mathcal{I}^{\alpha,\beta}[\phi]$ by replacing $\nu(dz)$ by the truncated non-singular measure $\nu_\delta(\,dz) := \mathbf{1}_{|z|>\delta}(z)\,\nu(dz)$ in (1.3):

$$\begin{aligned}\mathcal{I}^{\alpha,\beta,\delta}[\phi](t,x) &= \int_{|z|>\delta} \big(\phi(t, x + \eta^{\alpha,\beta}(t,x;z)) - \phi(t,x) - \eta^{\alpha,\beta}(t,x;z)\cdot\nabla_x\phi(t,x)\big)\,\nu(dz)\\ &= \mathcal{J}^{\alpha,\beta,\delta}[\phi](t,x) - b_\delta^{\alpha,\beta}(t,x)\cdot\nabla_x\phi(t,x),\end{aligned}$$

where

$$\mathcal{J}^{\alpha,\beta,\delta}[\phi](t,x) = \int_{|z|>\delta} \big(\phi(t, x + \eta^{\alpha,\beta}(t,x;z)) - \phi(t,x)\big)\,\nu(dz), \quad b_\delta^{\alpha,\beta}(t,x) = \int_{|z|>\delta} \eta^{\alpha,\beta}(t,x;z)\,\nu(dz).$$

This is a non-singular, nonnegative, consistent approximation of $\mathcal{I}^{\alpha,\beta}$, and a standard argument using Taylor's theorem gives the truncation error

$$\Big|\mathcal{I}^{\alpha,\beta}[\phi] - \mathcal{I}^{\alpha,\beta,\delta}[\phi]\Big| \leq \frac{1}{2}\|D^2\phi\|_0 \sup_{x,\alpha,\beta}\int_{|z|<\delta} |\eta^{\alpha,\beta}(t,x;z)|^2\nu(dz) \leq K\delta^{2-\sigma}\|D^2\phi\|_0 \qquad \text{for} \qquad \phi \in C_b^2(\mathbb{R}^N), \tag{3.1}$$

where the last inequality follows by (A.3)–(A.5). Let $\tilde{b}_\delta^{\alpha,\beta}(t,x) := b^{\alpha,\beta}(t,x) - b_\delta^{\alpha,\beta}(t,x)$. We approximate (1.1) by replacing $\mathcal{I}^{\alpha,\beta}$ by $\mathcal{I}^{\alpha,\beta,\delta} = \mathcal{J}^{\alpha,\beta,\delta} - b_\delta^{\alpha,\beta}\cdot\nabla$,

$$u_t^\delta + \inf_{\alpha\in\mathcal{A}}\sup_{\beta\in\mathcal{B}}\Big\{-f^{\alpha,\beta}(t,x) + c^{\alpha,\beta}(t,x)u^\delta(t,x) - \tilde{b}_\delta^{\alpha,\beta}(t,x).\nabla u^\delta(t,x) - \mathcal{J}^{\alpha,\beta,\delta}[u^\delta](t,x)\Big\} = 0 \quad \text{in } Q_T. \tag{3.2}$$

Step 2. Discretize the local drift. We discretize $\tilde{b}_\delta^{\alpha,\beta}\cdot\nabla u$ by simple upwind finite differences:

$$\mathcal{D}_h^{\alpha,\beta,\delta}[u](t,x) := \sum_{i=1}^N \Big[\tilde{b}_{\delta,i}^{\alpha,\beta,+}(t,x)\frac{u(t,x+e_i\Delta x) - u(t,x)}{\Delta x} + \tilde{b}_{\delta,i}^{\alpha,\beta,-}(t,x)\frac{u(t,x-e_i\Delta x) - u(t,x)}{\Delta x}\Big]$$

$$= \sum_{\mathbf{j}\neq 0} d_{h,\mathbf{j}}^{\alpha,\beta,\delta}(t,x)\Big[u(t,x+x_{\mathbf{j}}) - u(t,x)\Big],$$

where $\{e_i\}_i \subset \mathbb{R}^N$ is the standard basis of $\mathbb{R}^N$, $b^{\pm} = \max(\pm b, 0)$, $d_{h,\pm e_i}^{\alpha,\beta,\delta}(t,x) = \frac{\tilde{b}_{\delta,i}^{\alpha,\beta,\pm}(t,x)}{\Delta x} \geq 0$ and $d_{h,\mathbf{j}}^{\alpha,\beta,\delta}(t,x) = 0$ otherwise. Hence the discretization is positive/monotone, and it is consistent since

$$\Big|\tilde{b}_{\delta}^{\alpha,\beta}(t,x)\cdot\nabla\phi(x) - \mathcal{D}_h^{\alpha,\beta,\delta}[\phi](t,x)\Big| \leq \frac{1}{2}\Delta x \sum_i \Big|\tilde{b}_{\delta,i}^{\alpha,\beta}(t,x)\Big| \|D^2\phi\|_0 \leq K\Delta x\, \Gamma(\sigma,\delta)\|D^2\phi\|_0 \quad \text{for} \quad \phi \in C_b^2(\mathbb{R}^N), \tag{3.3}$$

where

$$\Gamma(\sigma,\delta) = \begin{cases} \delta^{1-\sigma} & \text{when} \quad \sigma > 1, \\ -\log\delta & \text{when} \quad \sigma = 1, \\ 1 & \text{when} \quad \sigma < 1. \end{cases}$$

The last inequality follows by the definition of $\tilde{b}_{\delta}^{\alpha,\beta}$ since $\int_{|z|>\delta} |\eta^{\alpha,\beta}(t,x;z)|\,\nu(dz) \leq C\Gamma(\sigma,\delta)$ by (**A**.3)–(**A**.5).

Step 3. Discretize the nonlocal diffusion. We discretize $\mathcal{J}^{\alpha,\beta,\delta}$ by a quadrature formula obtained by replacing the integrand by a monotone interpolant (cf. [8]):

$$\mathcal{J}_h^{\alpha,\beta,\delta}[\varphi](t,x) := \int_{|z|>\delta} \mathfrak{i}_h[\varphi(t,x+\cdot) - \varphi(t,x)](\eta^{\alpha,\beta}(t,x;z))\nu(\,dz),$$

where $\mathfrak{i}_h$ is piecewise linear/multilinear interpolation on the spatial grid $\Delta x\mathbb{Z}^N$. That is,

$$\mathfrak{i}_h[\phi](x) = \sum_{\mathbf{j}\in\mathbb{Z}^N} \phi(x_{\mathbf{j}})\omega_{\mathbf{j}}(x;h) \quad \text{for} \quad x \in \mathbb{R}^N, \tag{3.4}$$

where the weights $\omega_{\mathbf{j}}$ are the standard "tent functions" satisfying $0 \leq \omega_{\mathbf{j}}(x;h) \leq 1$, $\omega_{\mathbf{j}}(x_{\mathbf{k}};h) = \delta_{\mathbf{j},\mathbf{k}}$, $\sum_{\mathbf{j}}\omega_{\mathbf{j}} = 1$, $\operatorname{supp}\omega_{\mathbf{j}} \subset B(x_{\mathbf{j}}, 2\Delta x)$, and $\|D\omega_{\mathbf{j}}\|_0 \leq C(\Delta x)^{-1}$. Note that the sum in (3.4) is always finite. We can rewrite the approximation in discrete monotone form:

$$\mathcal{J}_h^{\alpha,\beta,\delta}[\varphi](t,x) = \sum_{\mathbf{j}\in\mathbb{Z}^N} (\varphi(t,x+x_{\mathbf{j}}) - \varphi(t,x))\kappa_{h,\mathbf{j}}^{\alpha,\beta,\delta}(t,x); \quad \kappa_{h,\mathbf{j}}^{\alpha,\beta,\delta}(t,x;h) = \int_{|z|>\delta} \omega_{\mathbf{j}}(\eta^{\alpha,\beta}(t,x;z);h)\nu(dz),$$

where $\kappa_{h,\mathbf{j}}^{\alpha,\beta,\delta}$ is well-defined and nonnegative. This approximation is nonnegative, and since

$$\Big|\mathfrak{i}_h[\varphi](x) - \varphi(x)\Big| \leq K\|D^2\varphi\|_0(\Delta x)^2, \tag{3.5}$$

it is consistent with truncation error

$$|\mathcal{J}^{\alpha,\beta,\delta}[\phi] - \mathcal{J}_h^{\alpha,\beta,\delta}[\phi]| \leq K(\Delta x)^2\|D^2\phi\|_0 \int_{|z|>\delta} \nu(dz) \leq K_I\|D^2\phi\|_0 \frac{(\Delta x)^2}{\delta^{\sigma}} \qquad \text{for} \qquad \phi \in C_b^2(\mathbb{R}^N). \tag{3.6}$$

The last inequality follows from (**A**.5). We also note that since all $\omega_{\mathbf{j}}$'s have same diameter compact support and (**A**.3) and (**A**.5) hold with $\sigma \in (0,2)$, there is a constant K_N depending only on N such that

$$\sum_{\mathbf{j}\neq 0} \kappa_{h,\mathbf{j}}^{\alpha,\beta,\delta}(t,x) \leq \sum_{\mathbf{j}\neq 0} \|D\omega_{\mathbf{j}}\|_0 \int_{|z|>\delta} \Big|\eta^{\alpha,\beta}(t,x;z)\Big|\,\nu(dz) \leq \frac{K_N}{\Delta x}\Gamma(\sigma,\delta).$$

Step 4. The full discretization of (1.1). Combining the previous steps we obtain the following semidiscrete approximation of (1.1) (cf. (3.2)):

$$u_t + \inf_{\alpha\in\mathcal{A}} \sup_{\beta\in\mathcal{B}} \Big\{ -f^{\alpha,\beta}(t,x) + c^{\alpha,\beta}(t,x)u(t,x) - \mathcal{D}_h^{\alpha,\beta,\delta}[u](t,x) - \mathcal{J}_h^{\alpha,\beta,\delta}[u](t,x) \Big\} = 0. \tag{3.7}$$

To discretize in time we use a two-parameter monotone θ-like method that allows for explicit, implicit, and explicit-implicit versions (cf. [8]): For $\theta, \vartheta \in [0,1]$,

$$\begin{aligned} U_{\mathbf{j}}^n = U_{\mathbf{j}}^{n-1} - \Delta t \inf_{\alpha\in\mathcal{A}} \sup_{\beta\in\mathcal{B}} \Big\{ &-f_{\mathbf{j}}^{\alpha,\beta,n-1} + c_{\mathbf{j}}^{\alpha,\beta,n} U_{\mathbf{j}}^{n-1} - \theta \mathcal{D}_h^{\alpha,\beta,\delta}[U]_{\mathbf{j}}^n - (1-\theta)\mathcal{D}_h^{\alpha,\beta,\delta}[U]_{\mathbf{j}}^{n-1} \\ &-\vartheta \mathcal{J}_h^{\alpha,\beta,\delta}[U]_{\mathbf{j}}^n - (1-\vartheta)\mathcal{J}_h^{\alpha,\beta,\delta}[U]_{\mathbf{j}}^{n-1} \Big\} \\ &\text{for} \quad \mathbf{j} \in \mathbb{Z}^N, \quad 0 \le n \le M, \end{aligned} \tag{3.8}$$

$$U_{\mathbf{j}}^0 = u(0, x_{\mathbf{j}}) \quad \text{for} \quad \mathbf{j} \in \mathbb{Z}^N,$$

where $U_{\mathbf{j}}^n = U_h(t_n, x_{\mathbf{j}})$ is the solution of the scheme and $g_{\mathbf{j}}^n := g(t_n, x_{\mathbf{j}})$ for any function g and $(t_n, x_{\mathbf{j}}) \in \mathcal{G}_h^N$. With this convention,

$$\mathcal{D}_h^{\alpha,\beta,\delta}[\phi]_{\bar{\mathbf{j}}}^n = \sum_{\mathbf{j}\neq 0} d_{h,\mathbf{j},\bar{\mathbf{j}}}^{\alpha,\beta,\delta,n}\Big[\phi(t_n, x_{\bar{\mathbf{j}}} + x_{\mathbf{j}}) - \phi(t_n, x_{\bar{\mathbf{j}}})\Big] \quad \text{and} \quad \mathcal{J}_h^{\alpha,\beta,\delta}[\phi]_{\bar{\mathbf{j}}}^n = \sum_{\mathbf{j}\neq 0} \kappa_{h,\mathbf{j},\bar{\mathbf{j}}}^{\alpha,\beta,\delta,n}\Big[\phi(t_n, x_{\bar{\mathbf{j}}} + x_{\mathbf{j}}) - \phi(t_n, x_{\bar{\mathbf{j}}})\Big],$$

and we may rewrite our scheme (3.8) as

$$\inf_{\alpha\in\mathcal{A}} \sup_{\beta\in\mathcal{B}} \Big\{ a_{\bar{\mathbf{j}},\mathbf{0}}^{n,n}(\alpha,\beta) U_{\bar{\mathbf{j}}}^n - \sum_{\mathbf{j}\neq\mathbf{0}} a_{\bar{\mathbf{j}},\mathbf{j}}^{n,n}(\alpha,\beta) U_{\bar{\mathbf{j}}+\mathbf{j}}^n - \sum_{\mathbf{j}} a_{\bar{\mathbf{j}},\mathbf{j}}^{n,n-1}(\alpha,\beta) U_{\bar{\mathbf{j}}+\mathbf{j}}^{n-1} - \Delta t f_{\bar{\mathbf{j}}}^{\alpha,\beta,n} \Big\} = 0 \tag{3.9}$$

with

$$a_{\bar{\mathbf{j}},\mathbf{0}}^{n,m}(\alpha,\beta) = \begin{cases} 1 + \Delta t\theta \sum_{\mathbf{j}\neq 0} d_{h,\mathbf{j},\bar{\mathbf{j}}}^{\alpha,\beta,\delta,m} + \Delta t \vartheta \sum_{\mathbf{j}\neq 0} \kappa_{h,\mathbf{j},\bar{\mathbf{j}}}^{\alpha,\beta,\delta,m} & \text{if} \quad m = n \\ 1 - \Delta t[(1-\theta)\sum_{\mathbf{j}\neq 0} d_{h,\mathbf{j},\bar{\mathbf{j}}}^{\alpha,\beta,\delta,m} + (1-\vartheta)\sum_{\mathbf{j}\neq 0} \kappa_{h,\mathbf{j},\bar{\mathbf{j}}}^{\alpha,\beta,\delta,m} + c_{\bar{\mathbf{j}}}^{\alpha,\beta,n}] & \text{if} \quad m = n-1, \end{cases}$$

$$a_{\bar{\mathbf{j}},\mathbf{j}}^{n,m}(\alpha,\beta) = \begin{cases} \Delta t\theta d_{h,\mathbf{j},\bar{\mathbf{j}}}^{\alpha,\beta,\delta,m} + \Delta t\vartheta \kappa_{h,\mathbf{j},\bar{\mathbf{j}}}^{\alpha,\beta,\delta,m} & \text{if} \quad m = n \\ \Delta t[(1-\theta) d_{h,\mathbf{j},\bar{\mathbf{j}}}^{\alpha,\beta,\delta,m} + (1-\vartheta)\kappa_{h,\mathbf{j},\bar{\mathbf{j}}}^{\alpha,\beta,\delta,m}] & \text{if} \quad m = n-1. \end{cases}$$

Since $d, \kappa \ge 0$, we see that the scheme (3.9) has nonnegative coefficients and hence is monotone under the CFL condition:

$$\Delta t\Big[(1-\theta)\sum_{\mathbf{j}\neq 0} d_{h,\mathbf{j},\bar{\mathbf{j}}}^{\alpha,\beta,\delta,n-1} + (1-\vartheta)\sum_{\mathbf{j}\neq 0} \kappa_{h,\mathbf{j},\bar{\mathbf{j}}}^{\alpha,\beta,\delta,n-1} + c_{\bar{\mathbf{j}}}^{\alpha,\beta,n}\Big] \le 1. \tag{3.10}$$

By the discussion and definitions of d and κ above, for all $0 < \Delta x \le \delta \le 1$,

$$\sum_{\mathbf{j}\neq 0} d_{h,\mathbf{j},\bar{\mathbf{j}}}^{\alpha,\beta,\delta,n} \le \frac{K_D}{\Delta x}\Gamma(\sigma,\delta) \qquad \text{and} \qquad \sum_{\mathbf{j}\neq 0} \kappa_{h,\mathbf{j},\bar{\mathbf{j}}}^{\alpha,\beta,\delta,n} \le \frac{K_I}{\Delta x}\Gamma(\sigma,\delta)$$

for some constants K_D, K_I. Hence the CFL condition is satisfied when

$$\frac{\Delta t}{\Delta x}\Gamma(\sigma,\delta)\Big((1-\theta)K_D + (1-\vartheta)K_I\Big) + \Delta t \sup_{\alpha,\beta} |c^{\alpha,\beta}| \le 1. \tag{3.11}$$

Remark 3.1.

(a) The scheme is explicit when $\theta = 0 = \vartheta$, implicit when $\theta = 1 = \vartheta$, θ-method like when $\theta = \vartheta$, and explicit-implicit with explicit convection and implicit diffusion when $\theta = 0$ and $\vartheta = 1$.

(b) It is possible to use other monotone approximations in steps 1 – 4, and obtain schemes that can be analyzed using minor modifications of the arguments we present here.

(c) The CFL condition (3.11) gives a constraint on the relation between δ, Δx, Δt when the scheme is not completely implicit. In the "first order" case, when $\sigma \in (0,1)$ in (**A**.5), we get the usual CFL condition

$$\Delta t \leq K\Delta x.$$

In the critical case $\sigma = 1$, then $\Delta t \leq K\Delta x\,|\ln \Delta x|$. When the order of the equation is greater than 1, $\sigma \in (1,2)$ in (**A**.5), then

$$\Delta t \leq K\delta^{\sigma-1}\Delta x,$$

which when $\delta = (\Delta x)^{\frac{1}{\sigma}}$ (giving the optimal convergence rate, see below) gives

$$\Delta t \leq K\Delta x^{2-\frac{1}{\sigma}}.$$

We have the following existence, uniqueness and stability result for the scheme.

Theorem 3.1. *Assume (**A**.1)–(**A**.5), $0 < \Delta x \leq \delta \leq 1$, and the CFL condition (3.10).*

(a) (Monotone scheme) If U_h and V_h are bounded sub and supersolutions of (3.8) with $U_h(0,\cdot) \leq V_h(0,\cdot)$, then $U_h \leq V_h$.

(b) There exists a unique bounded solution U_h of the initial value problem (3.8)–(1.2).

(c) (L^∞-stability) The solution U_h from (b) satisfies $|U_h(t_n,x)| \leq \|u_0\|_0 + t_n \sup_{\alpha,\beta}\|f^{\alpha,\beta}\|_0$.

(d) There is a constant $K \geq 0$ such that the solution U_h from (b) satisfies for all x, t_n,

$$|U_h(t_n,x) - u_0(x)| \leq K\bar{\omega}(t_n), \qquad \text{where } \bar{\omega} \text{ is defined in (2.2).}$$

*(e) Assume in addition that $K(u_0) < \infty$ (cf. Theorem 2.1). Then there is $C \geq 0$ only depending on the data (**A**.1)–(**A**.5) such that the solution U_h from (b) satisfies for all x, t_n,*

$$|U_h(t_n,x) - u_0(x)| \leq C(1 + K(u_0))t_n.$$

Proof of Theorem 3.1. The proofs of (a)–(c) are standard. Part (a) is a direct consequence of the scheme having positive coefficients, and part (c) follow from (a) since $\|u\|_0 \pm t_n \sup_{\alpha,\beta}\|f^{\alpha,\beta}\|$ are super- and subsolutions. Part (b), existence and uniqueness, can be proved using time-iteration and Banach fixed point theorem. The proof is essentially the same as the proof of Theorem 3.1 in [8]. Part (d) and (e) are new and non-standard. We will prove these results in the same way as for the solution of the continuous problem (1.1)–(1.2), cf. Theorem 2.1 (c) and (d). First we prove (e), and then use this result to prove (d).

(e) Note that $V^\pm(x,t_n) = u_0(x) \pm Ct_n$ are super- and subsolutions of the scheme (3.8)–(1.2) if h is sufficiently small and

$$C \geq 1 + K(u_0) + \sup_{\alpha,\beta}\Big(\|u_0\|_1\|b^{\alpha,\beta}\|_0 + \|u_0\|_0\|c^{\alpha,\beta}\|_0 + \|f^{\alpha,\beta}\|_0\Big).$$

The result then follows since $V^- \leq U_h \leq V^+$ by comparison (part (a)).

(d) We regularize (by mollification) the initial data to get u_0^ϵ and let U_h^ϵ be the corresponding solution of (3.8)–(1.2). By (a) again $|U_h - U_h^\epsilon| \leq \|u_0^\epsilon - u_0\|_0 \leq C\epsilon$, and the estimate (2.3) for $K(u_0^\epsilon)$ still holds. Hence by part (e) we have that $|U_h^\epsilon(t_n,x) - u_0^\epsilon(x)| \leq CK(u_0^\epsilon)t_n$, and then by the triangle inequality

$$|U_h(t_n,x) - u_0(x)| \leq C(\epsilon + K(u_0^\epsilon)t_n + \epsilon).$$

In view of (2.3), we conclude by taking $\epsilon = 0$, $\epsilon = t_n$, $\epsilon = t_n^{\frac{1}{\sigma}}$ when $\sigma < 1$, $\sigma = 1$, $\sigma > 1$ respectively. $\square$

Convergence of U_h to the unique viscosity solution of (1.1)-(1.2) follows from (an easy nonlocal extension of) the Barles-Perthame-Souganidis half-relaxed limits method [5] in view of monotonicity, stability, consistency of the scheme and strong comparison of the limit equation.

We now give precise estimates on the rate of convergence of our method for our low-regularity solutions. These are the main contributions of the paper. They are the first such result for non-convex degenerate equations of order greater than one, the first result for nonlocal non-convex equations, and these estimates are more accurate than previous results for the non-local operators $\mathcal{I}$: First, as expected, the rates depend on the maximal fractional order of the operator $\mathcal{I}$, or equivalently, on σ in assumptions (**A**.5). But we also see a surprising phenomenon that does not seem to have been observed before: We have 3 different results depending on whether η depends on (x,t), only on x, and on none of them. We devote one theorem to each case:

Theorem 3.2 (General case)**.** *Assume (**A**.1)–(**A**.5), $0 < \Delta x \le \delta \le 1$, the CFL condition (3.10) holds, u solves the equation (1.1)-(1.2), and U_h^δ solves the scheme (3.8)-(1.2). Then there exists a constant $C > 0$ (only depending on the constants in (**A**.1)–(**A**.5)) such that for all $(t,x) \in \mathcal{G}_h^N$,*

$$\left|U_h^\delta(t,x) - u(t,x)\right| \le C(1+T)\begin{cases}(T\wedge 1)^{\frac{1}{2}}\left(\Delta t^{\frac{1}{2}} + \Delta x^{\frac{1}{2}} + \delta^{1-\frac{\sigma}{2}}\right) & \text{if} \quad \sigma \in [0,1),\\[2mm] (T\wedge 1)^{\frac{1}{2}}\left(\Delta t^{\frac{1}{2}}|\log \Delta t| + \Delta t^{\frac{1}{2}}|\log \delta| + \Delta x^{\frac{1}{2}}|\log\delta| + \delta^{\frac{1}{2}}\right) & \text{if} \quad \sigma = 1,\\[2mm] (T\wedge 1)^{\frac{1}{2\sigma}}\Delta t^{\frac{1}{2\sigma}} + (T\wedge 1)^{\frac{1}{2}}\left(\Delta t^{\frac{1}{2}}\delta^{1-\sigma} + \Delta x^{\frac{1}{2}}\delta^{1-\sigma} + \delta^{1-\frac{\sigma}{2}}\right) & \text{if} \quad \sigma\in(1,2).\end{cases}$$

Remark 3.2. (a) These results imply the convergence of the scheme, and *optimal* error estimates for local first order Hamilton-Jacobi equations (cf. [17, 30]) follows as a special case since $\mathcal{I}^{\alpha,\beta} \equiv 0$ is allowed. This also means that the rate in the case $\sigma \in (0,1)$ is optimal because of the first order drift term in our equation.

(b) The results for $\sigma \in [1,2)$ are also optimal. The principal error term is $\delta^{1-\frac{\sigma}{2}}$ since $\Delta x \le \delta$. This term comes from the truncation of the singularity and is optimal in view of the low regularity of our problem. See (3.1) for the rate for smooth solutions and Lemma 4.1 below for the rate under our assumptions.

Theorem 3.3 (No t-dependence)**.** *Let the assumptions of Theorem 3.2 hold and $\eta^{\alpha,\beta}$ be independent of t.*

(a) Then there is a constant C such that for all $(t,x) \in \mathcal{G}_h^N$,

$$|U_h^\delta - u| \le C(1+T)\begin{cases}(T\wedge 1)^{\frac{1}{2}}\left(\Delta t^{\frac{1}{2}} + \Delta x^{\frac{1}{2}} + \delta^{1-\frac{\sigma}{2}}\right) & \text{if} \quad \sigma \in [0,1),\\[2mm] (T\wedge 1)^{\frac{1}{2}}\left(\Delta t^{\frac{1}{2}}|\log \Delta t| + \Delta x^{\frac{1}{2}}|\log\delta| + \delta^{\frac{1}{2}}\right) & \text{if} \quad \sigma = 1,\\[2mm] (T\wedge 1)^{\frac{1}{2\sigma}}\Delta t^{\frac{1}{2\sigma}} + (T\wedge 1)^{\frac{1}{2}}\left(\Delta x^{\frac{1}{2}}\delta^{1-\sigma} + \delta^{1-\frac{\sigma}{2}}\right) & \text{if} \quad \sigma\in(1,2).\end{cases}$$

(b) If $K(u_0) < \infty$ (cf. Theorem 2.1), then there is a constant C such that for all $(x,t) \in \mathcal{G}_h^N$,

$$|U_h^\delta - u| \le C(1+T)(T\wedge 1)^{\frac{1}{2}}\left(\Delta t^{\frac{1}{2}} + \Delta x^{\frac{1}{2}}\Gamma(\sigma,\delta) + \delta^{1-\frac{\sigma}{2}}\right) \qquad \text{if} \qquad \sigma \in [0,2).$$

*All the constants C only depend on the constants in (**A**.1)–(**A**.5) and (3.11), and for (b), also on $K(u_0)$.*

The Theorem also holds for η depending on t if $\Delta t/\Delta x \le K$.

Remark 3.3. (a) Since η depends on time, the convergence in Δt and δ is coupled in Theorem 3.2 for $\sigma \in [1, 2)$! When η does not depend on t, there is no coupling and a better rate by Theorem 3.3 (a).

(b) When $\sigma \geq 1$, there is a reduction of rate in Δt because the solution of (1.1) no longer is Lipschitz in t. However, assuming more regularity of the initial data will make the solution t-Lipschitz again, and then we get back the full rate $\frac{1}{2}$ in Theorem 3.3 (b).

Theorem 3.4 (No x, t dependence)**.** *The assumptions of Theorem 3.2 hold and $\eta^{\alpha,\beta}$ is independent of x, t.*

(a) There is a constant C such that for all $(t, x) \in \mathcal{G}_h^N$,

$$\left|U_h^\delta(t,x) - u(t,x)\right| \leq C(1+T)\begin{cases} (T \wedge 1)^{\frac{1}{2}}\Big(\Delta t^{\frac{1}{2}} + \Delta x^{\frac{1}{2}} + \delta^{1-\frac{\sigma}{2}}\Big) & \text{if} \quad \sigma = [0,1), \\ (T \wedge 1)^{\frac{1}{2}}\Big(\Delta t^{\frac{1}{2}}|\log \Delta t| + \Delta x^{\frac{1}{2}}|\log \delta|^{\frac{1}{2}} + \delta^{\frac{1}{2}}\Big) & \text{if} \quad \sigma = 1, \\ (T \wedge 1)^{\frac{1}{2\sigma}}\Delta t^{\frac{1}{2\sigma}} + (T \wedge 1)^{\frac{1}{2}}\Big(\Gamma(\sigma,\delta)^{\frac{1}{2}}\Delta x^{\frac{1}{2}} + \delta^{1-\frac{\sigma}{2}}\Big) & \text{if} \quad \sigma \in (1,2). \end{cases}$$

(b) If also $K(u_0) < \infty$ (cf. Theorem 2.1), then there is a constant C such that for all $(t, x) \in \mathcal{G}_h^N$,

$$\left|U_h^\delta(t,x) - u(t,x)\right| \leq C(1+T)(T \wedge 1)^{\frac{1}{2}}\Big(\Delta t^{\frac{1}{2}} + \Gamma(\sigma,\delta)^{\frac{1}{2}}\Delta x^{\frac{1}{2}} + \delta^{1-\frac{\sigma}{2}}\Big)$$

*All the constants C only depend on the constants in (**A**.1)–(**A**.5) and (3.11), and for (b), also on $K(u_0)$.*

The proofs these results are given in Section 4.

Remark 3.4. (a) Our estimates hold for solutions that are merely Lipschitz in x and Lipschitz or Hölder in t. In general this is the best regularity for our problem under our assumptions. For more regular solutions, better estimates should hold. However, the maximal rate or accuracy of our scheme is $O(\Delta x^{1\wedge(2-\sigma)})$. The dominant error term comes from truncation of the measure (cf. (3.1), (3.3), (3.6) and recall that $\Delta x \leq \delta$).

(b) The choices of δ that optimize the error are $\delta = \Delta x$ for $\sigma \in (0, 1)$ and when $\sigma \in (1, 2)$ that are $\delta = \max(\Delta t^{\frac{1}{\sigma}}, \Delta x^{\frac{1}{\sigma}})$ in Theorem 3.2, $\delta = \Delta x^{\frac{1}{\sigma}}$ in Theorem 3.3 and $\delta = \Delta x$ in Theorem 3.4. Assume now $K(u_0) = \infty$. When $\sigma \leq 1$, all cases then lead to the estimate

$$|u - U_h^\delta| \leq C \begin{cases} (\Delta t)^{\frac{1}{2}} + (\Delta x)^{\frac{1}{2}} & \text{when} \quad \sigma \in [0,1), \\ (\Delta t)^{\frac{1}{2}}|\log(\Delta t)| + (\Delta x)^{\frac{1}{2}}|\log(\Delta x)| & \text{when} \quad \sigma = 1. \end{cases}$$

However the estimates for $\sigma \in (1, 2)$ are different in each case:

$$|u - U_h^\delta| \leq C \begin{cases} (\Delta t)^{\frac{2-\sigma}{2\sigma}} + (\Delta x)^{\frac{2-\sigma}{2\sigma}} & \text{in Theorem 3.2}, \\ (\Delta t)^{\frac{1}{2\sigma}} + (\Delta x)^{\frac{2-\sigma}{2\sigma}} & \text{in Theorem 3.3 (a)}, \\ (\Delta t)^{\frac{1}{2\sigma}} + (\Delta x)^{\frac{2-\sigma}{2}} & \text{in Theorem 3.4 (a)}. \end{cases}$$

Note the improvement in rate in each line! When $K(u_0) < \infty$, the solution is Lipschitz in t, and the time rate improves to $O(\Delta t^{\frac{1}{2}})$ in Theorems 3.3 and 3.4. In particular, the rate of Theorem 3.4 (b) becomes

$$O\Big((\Delta t)^{\frac{1}{2}} + (\Delta x)^{\frac{2-\sigma}{2}}\Big).$$

This latter spatial rate is consistent with the rates (for the implicit scheme) of Theorem 6.3 in [14] where other types of (x, t)-independent nonlocal nonlinear equations are considered.

4. Proof of the main results - Theorems 3.2, 3.3 and 3.4

4.1. Reduction to finite Lévy measures

Since the two problems (1.1) and (3.2) have the same data and coefficients except for the Lévy measures ν and ν_δ, we can use the continuous dependence results of [22] to bound the distance between u and u^δ.

Lemma 4.1. *Assume (A.1)–(A.5). If u and u^δ solve (1.1) and (3.2), then*

$$|u(t,x) - u^\delta(t,x)| \leq CT^{\frac{1}{2}}\delta^{1-\frac{\sigma}{2}} \qquad \text{for all} \qquad (t,x) \in Q_T.$$

Proof. In a similar way as Theorem 4.1 in [22] follows from Corollary 3.2 in [22], we can use Corollary 3.2 in [22] and the fact that all coefficients are bounded to show that

$$|u(t,x) - u^\delta(t,x)| \leq CT^{\frac{1}{2}}\sqrt{\int_{\mathbb{R}^M\setminus\{0\}} |\eta^{\alpha,\beta}(t,x;z)|^2\,|\nu-\nu_\delta|(dz)}.$$

Note that as opposed to Theorem 4.1 in [22], there is no growth in x in our estimate. The result then follows by (A.3)–(A.5) and $\int_{|z|<\delta}|z|^2\nu(dz) = C\delta^{2-\sigma}$. □

In view of the result, it is enough to prove Theorem 3.2 when the Lévy measure ν is replaced by the bounded measure ν_δ. Therefore, in the rest of the proof we only work with $u = u^\delta$, the solution of (3.2)–(1.2).

4.2. The doubling of variables argument

Recall that U_h is defined on $\mathcal{G}_h^N$ as $U_h(t_n, x_{\mathbf{j}}) = U_{\mathbf{j}}^n$, and $u = u^\delta$ solves (3.2)–(1.2). We want to bound $|U_h(t_n,x_{\mathbf{j}}) - u(t_n,x_{\mathbf{j}})|$ in $\mathcal{G}_h^N$ and start by deriving a nonnegative upper bound on

$$\mu = \sup_{\mathbf{j}\in\mathbb{Z}^N,\, n\leq M} (U_{\mathbf{j}}^n - u(t_n,x_{\mathbf{j}})).$$

Assume that $\mu > 0$ (if not $\mu \leq 0$ and we are done). Since u and U_h are bounded uniformly in h,

$$R := \max\{\|u\|_{L^\infty}, \|U_h\|_{L^\infty}\} < \infty.$$

We will use the method of doubling of variables (e.g [17]) and to do that we introduce $\Psi : \mathcal{G}_h^N \times Q_T \to \mathbb{R}$,

$$\Psi(t,x,s,y) = U_h(t,x) - u(s,y) - \phi(x,y) - \xi(t,s) - \frac{\mu}{4T}(t+s),$$

where $\phi : \mathbb{R}^N\times\mathbb{R}^N \to \mathbb{R}$ and $\xi : [0,T]\times[0,T] \to \mathbb{R}$ are defined by

$$\phi(x,y) = \frac{\gamma}{2}|x-y|^2 + \frac{\varepsilon}{2}(|x|^2+|y|^2) \qquad \text{and} \qquad \xi(t,s) = \frac{\eta}{2}|t-s|^2,$$

for $\gamma,\eta,\varepsilon > 0$. From the boundedness of U_h and u, it follows that Ψ has a maximum at $(t_0,x_0,s_0,y_0) \in \mathcal{G}_h^N \times Q_T$ such that

$$\Psi(t_0,x_0,s_0,y_0) \geq \Psi(t,x,s,y) \tag{4.1}$$

for any $(t,x,s,y) \in \mathcal{G}_h^N \times Q_T$. Since $0 = \Psi(0,0,0,0) \leq \Psi(t_0,x_0,s_0,y_0)$, it follows that

$$\frac{\gamma}{2}|x_0-y_0|^2 + \frac{\varepsilon}{2}(|x_0|^2+|y_0|^2) + \frac{\eta}{2}|t_0-s_0|^2 \leq U_h(t_0,x_0) - u(s_0,y_0),$$

and hence $U_h(t_0, x_0) - u(s_0, y_0) \geq 0$ and

$$\varepsilon(|x_0|^2 + |y_0|^2) \leq 4R. \tag{4.2}$$

Moreover, since the map $y \to u(s_0, y) + \phi(x_0, y)$ has a minimum at y_0 and u is Lipschitz, $\phi(x_0, y_0) - \phi(x_0, y) \leq u(s_0, y) - u(s_0, y_0) \leq L|y - y_0|$, and hence $|D_y\phi(x_0, y_0)| \leq L$. Then by the definition of ϕ, and since $\varepsilon|y_0| \leq \sqrt{4R\varepsilon}$ by the last bound in (4.2), we have

$$|x_0 - y_0| \leq \frac{1}{\gamma}(L + \sqrt{4R\varepsilon}). \tag{4.3}$$

By the inequality $\Psi(t_0, x_0, t_0, y_0) \leq \Psi(t_0, x_0, s_0, y_0)$ and the regularity of u in Theorem 2.1, we find that

$$\frac{\eta}{2}|t_0 - s_0|^2 \leq u(t_0, y_0) - u(s_0, y_0) \leq K\omega(t_0 - s_0) \tag{4.4}$$

where $\omega(r) = |r|$ if $K(u_0) < \infty$ and $\omega = \bar{\omega}$ from Theorem 2.1 (c) if not. For $\sigma \neq 1$, we get that

$$|t_0 - s_0| \leq \frac{2K}{\eta^q} \tag{4.5}$$

with $q = 1$ when $K(u_0) < \infty$ and u is Lipschitz in t and $q = \frac{\sigma}{2\sigma - 1}$ when $\sigma \in (1, 2)$ and u is Hölder $\frac{1}{\sigma}$ in t.

If either t_0 or s_0 is 0, then we get a bound on μ using only the regularity of the u and U_h at $t = 0$. If $s_0 = 0$ and $t_0 > 0$, then for any point $(t, x) \in \mathcal{G}_h^N$,

$$\begin{aligned} U_h(t, x) - u(t, x) - \varepsilon|x|^2 - \frac{\mu}{2T}t &= \Psi(t, x, t, x) \leq \Psi(t_0, x_0, 0, y_0) \\ &= U_h(t_0, x_0) - u_0(y_0) - \phi(x_0, y_0) - \xi(t_0, 0) - \frac{\mu}{4T}t_0 \leq U_h(t_0, x_0) - u_0(y_0). \end{aligned}$$

If either (**A**.5) holds with $\sigma \in (0, 1)$ or $K(u_0) < \infty$, then u and U_h are Lipschitz in t at $t = 0$. By Theorem 3.1 (e) and the regularity of u_0, $U_h(t_0, x_0) - u_0(x_0) + u_0(x_0) - u_0(y_0) \leq C(t_0 + |x_0 - y_0|)$. Hence by estimates (4.3) and (4.5) with $q = 1$ and since $t_0 = |t_0 - s_0|$, we find that $U_h(t, x) - u(t, x) - \varepsilon|x|^2 - \frac{\mu}{2T}t \leq C(\frac{1}{\gamma} + \frac{1}{\eta})$. If we first send ε to 0 and then take the supremum over $\mathcal{G}_h^N$, by the definition of μ we get that

$$\frac{\mu}{2} \leq \sup_{\mathbf{j} \in \mathbb{Z}^N, n \leq M} (U_{\mathbf{j}}^n - u(t_n, x_{\mathbf{j}})) - \frac{\mu}{2} \leq C\Big(\frac{1}{\gamma} + \frac{1}{\eta}\Big).$$

When $\sigma \in (1, 2)$, u and U_h are only Hölder $\frac{1}{\sigma}$ in t at $t = 0$ (cf. Theorem 2.1 (c) and Theorem 3.1 (d)). In this case e.g. $U_h(t_0, x_0) - u_0(y_0) \leq C(t_0^{\frac{1}{\sigma}} + |x_0 - y_0|)$, and hence by (4.3) and (4.5) with $q = \frac{\sigma}{2\sigma - 1}$, we find that

$$\frac{\mu}{2} \leq C\Big(\frac{1}{\gamma} + \frac{1}{\eta^{\frac{1}{2\sigma - 1}}}\Big).$$

A similar argument using time regularity of u, shows that these bounds also hold when $t_0 = 0$ and $s_0 \geq 0$.

Only the case $t_0 > 0$ and $s_0 > 0$ remains. Here we have to use the equations and the argument is long so we divide it into several steps.

Step 1: It is easily seen from (4.1) that (s_0, y_0) is a global minimum point on Q_T of

$$(s, y) \to u(s, y) - \Big(- \phi(x_0, y) - \xi(t_0, s) - \frac{\mu}{4T}(t_0 + s)\Big).$$

By the supersolution inequalities for u (cf. (3.2)) with $\kappa = \delta$,

$$-D_s\xi(t_0, s_0) - \frac{\mu}{4T} + \inf_{\alpha} \sup_{\beta} \Big\{ - f^{\alpha,\beta}(s_0, y_0) + c^{\alpha,\beta}(s_0, y_0)u(s_0, y_0)$$

$$- \tilde{b}_{\delta}^{\alpha,\beta}(s_0, y_0)(-D_y\phi(x_0, y_0)) - \mathcal{J}^{\alpha,\beta,\delta}[u](s_0, y_0)\Big\} \geq 0. \tag{4.6}$$

We now get an analogous relation for the scheme at the grid-point (t_0, x_0). By (4.1) again $\Psi(t, x, s_0, y_0) \leq \Psi(t_0, x_0, s_0, y_0)$, and hence the function

$$W(t, x) := U_h(t_0, x_0) + \phi(x, y_0) - \phi(x_0, y_0) + \xi(t, s_0) - \xi(t_0, s_0) + \frac{\mu}{4T}(t - t_0)$$

satisfies

$$U_h \leq W \quad \text{in} \quad \mathcal{G}_h^N \qquad \text{and} \qquad U_h(t_0, x_0) = W(t_0, x_0).$$

By the definition and monotonicity of the scheme (under the CFL condition (3.10)) we then get at the maximum point $(t_0, x_0) = (p\Delta t, \Delta x\, \mathbf{k})$ that

$$\begin{aligned}
U_{\mathbf{k}}^{p} = U_{\mathbf{k}}^{p-1} - \Delta t \inf_{\alpha} \sup_{\beta} \Big\{ & -f_{\mathbf{k}}^{\alpha,\beta,p} + c_{\mathbf{k}}^{\alpha,\beta,p} U_{\mathbf{k}}^{p-1} - \theta \mathcal{D}_h^{\alpha,\beta,\delta}[U]_{\mathbf{k}}^{p} - (1-\theta)\mathcal{D}_h^{\alpha,\beta,\delta}[U]_{\mathbf{k}}^{p-1} \\
& - \vartheta \mathcal{J}_h^{\alpha,\beta,\delta}[U]_{\mathbf{k}}^{p} - (1-\vartheta)\mathcal{J}_h^{\alpha,\beta,\delta}[U]_{\mathbf{k}}^{p-1} \Big\} \\
\leq W_{\mathbf{k}}^{p-1} - \Delta t \inf_{\alpha} \sup_{\beta} \Big\{ & -f_{\mathbf{k}}^{\alpha,\beta,p} + c_{\mathbf{k}}^{\alpha,\beta,p} W_{\mathbf{k}}^{p-1} - \theta \mathcal{D}_h^{\alpha,\beta,\delta}[W]_{\mathbf{k}}^{p} - (1-\theta)\mathcal{D}_h^{\alpha,\beta,\delta}[W]_{\mathbf{k}}^{p-1} \\
& - \vartheta \mathcal{J}_h^{\alpha,\beta,\delta}[U]_{\mathbf{k}}^{p} - (1-\vartheta)\tilde{\mathcal{J}}_h^{\alpha,\beta,\delta}[U, W]_{\mathbf{k}}^{p-1} \Big\}
\end{aligned}$$

where

$$\tilde{\mathcal{J}}_h^{\alpha,\beta,\delta}[U, W]_{\mathbf{k}}^{p-1} = \sum_{\mathbf{j} \in \mathbb{Z}^N} \left[U_{\mathbf{k}+\mathbf{j}}^{p-1} - W_{\mathbf{k}}^{p-1} \right] \int_{|z| \geq \delta} \omega_{\mathbf{j}} \left(\eta^{\alpha,\beta}(t_0 - \Delta t, x_0; z); h \right) \, \nu(dz)$$

and this non-standard term is admissible by the monotonicity of the full scheme in the $U_{\mathbf{k}}^{p-1}$-argument. We will see later that we really need the term in this form. By definition of W, $\mathcal{D}_h^{\alpha,\beta}[W] = \mathcal{D}_h^{\alpha,\beta}[\phi(., y_0)]$ etc., and we divide by Δt and rewrite the above inequality as

$$\begin{aligned}
\frac{\mu}{4T} \leq \frac{\xi(t_0 - \Delta t, s_0) - \xi(t_0, s_0)}{\Delta t} - \inf_{\alpha} \sup_{\beta} \Big\{ & - f_{\mathbf{k}}^{\alpha,\beta,p} + c_{\mathbf{k}}^{\alpha,\beta,p} \Big[U_{\mathbf{k}}^{p} + \xi(t_0 - \Delta t, s_0) - \xi(t_0, s_0) - \frac{\mu}{4T}\Delta t \Big] \\
& - \theta \mathcal{D}_h^{\alpha,\beta,\delta}[\phi(., y_0)](t_0, x_0) - (1-\theta)\mathcal{D}_h^{\alpha,\beta,\delta}[\phi(., y_0)](t_0 - \Delta t, x_0) \\
& - (1-\vartheta)\tilde{\mathcal{J}}_h^{\alpha,\beta,\delta}[U_h, W](t_0 - \Delta t, x_0) - \vartheta \mathcal{J}_h^{\alpha,\beta,\delta}[U_h](t_0, x_0) \Big\}.
\end{aligned} \tag{4.7}$$

Subtracting inequalities (4.6) and (4.7) and using the fact that $\inf\sup f - \inf\sup g \leq \sup\sup(f - g)$,

$$\begin{aligned}
\frac{\mu}{2T} \leq & \frac{\xi(t_0 - \Delta t, s_0) - \xi(t_0, s_0)}{\Delta t} - D_s\xi(t_0, s_0) \\
& + \sup_{\alpha} \sup_{\beta} \Big\{ f_{\mathbf{k}}^{\alpha,\beta,p} - f^{\alpha,\beta}(s_0, y_0) - c_{\mathbf{k}}^{\alpha,\beta,p} \Big[U_{\mathbf{k}}^{p} + \xi(t_0 - \Delta t, s_0) - \frac{\mu}{4T}\Delta t - \xi(t_0, s_0) \Big] + c^{\alpha,\beta}(s_0, y_0) u(s_0, y_0) \\
& \quad + \theta \mathcal{D}_h^{\alpha,\beta,\delta}[\phi(., y_0)](t_0, x_0) + (1-\theta)\mathcal{D}_h^{\alpha,\beta,\delta}[\phi(., y_0)](t_0 - \Delta t, x_0) - \tilde{b}_{\delta}^{\alpha,\beta}(s_0, y_0)(-D_y\phi(x_0, y_0)) \\
& \quad + (1-\vartheta)\tilde{\mathcal{J}}_h^{\alpha,\beta,\delta}[U_h, W](t_0 - \Delta t, x_0) + \vartheta \mathcal{J}_h^{\alpha,\beta,\delta}[U_h](t_0, x_0) - \mathcal{J}^{\alpha,\beta,\delta}[u](s_0, y_0) \Big\} \\
= & I_1 + \sup_{\alpha} \sup_{\beta} \Big\{ I_2 + I_3 + I_4 \Big\}.
\end{aligned} \tag{4.8}$$

Step 2: We now estimate the terms I_1, I_2, I_3 in (4.8). First note that $\xi(t_0-\Delta t, s_0)-\xi(t_0, s_0) = -\partial_t\xi(t_0, s_0)\Delta t + \frac{\eta}{2}\Delta t^2$, and hence since $\partial_t\xi = -\partial_s\xi$

$$I_1 = \frac{\xi(t_0-\Delta t, s_0)-\xi(t_0, s_0)}{\Delta t} - \partial_s\xi(t_0, s_0) = \frac{\eta}{2}\Delta t.$$

We estimate I_2 using $c \geq 0$, $U^p_{\mathbf{k}} - u(s_0, y_0) \geq 0$, regularity of the coefficients c and f, the estimate on I_1, and the bounds on $|x_0 - y_0|$ and $|t_0 - s_0|$,

$$\begin{aligned}
I_2 &= -c^{\alpha,\beta,p}_{\mathbf{k}}\Big[U^p_{\mathbf{k}} + \xi(t_0-\Delta t, s_0) - \frac{\mu}{4T}\Delta t - \xi(t_0, s_0)\Big] + c^{\alpha,\beta}(s_0, y_0)u(s_0, y_0) + f^{\alpha,\beta}(t_0, x_0) - f^{\alpha,\beta}(s_0, y_0)\\
&\leq 0 + |u(s_0, y_0)||c^{\alpha,\beta,p}_{\mathbf{k}} - c^{\alpha,\beta}(s_0, y_0)| + K\Big(|\xi(t_0-\Delta t, s_0) - \xi(t_0, s_0)| + \frac{\mu}{4T}\Delta t\Big) + \Big|f^{\alpha,\beta}(t_0, x_0) - f^{\alpha,\beta}(s_0, y_0)\Big|\\
&\leq C\Big(|x_0 - y_0| + |t_0 - s_0| + \Delta t + \eta\Delta t^2\Big). \qquad (4.9)
\end{aligned}$$

We now estimate I_3. By the consistency estimate (3.3), the definition of $\tilde{b}^{\alpha,\beta}_\delta$, the time regularity and bounds on b and η and integrability assumptions (A.2)–(A.5) of ν, the definition and gradient bound of ϕ,

$$\begin{aligned}
&\theta\mathcal{D}^{\alpha,\beta,\delta}_h[\phi(.,y_0)](t_0, x_0) + (1-\theta)\mathcal{D}^{\alpha,\beta,\delta}_h[\phi(.,y_0)](t_0-\Delta t, x_0)\\
&\leq \Big(\theta\tilde{b}^{\alpha,\beta}_\delta(t_0, x_0) + (1-\theta)\tilde{b}^{\alpha,\beta}_\delta(t_0-\Delta t, x_0)\Big)\cdot D_x\phi(x_0, y_0) + C\|\tilde{b}^{\alpha,\beta}_\delta\|_0\|D^2\phi\|_0\Delta x\\
&\leq \tilde{b}^{\alpha,\beta}_\delta(t_0, x_0)\cdot D_x\phi(x_0, y_0) + C\Big(1 + \int_{|z|>\delta}\rho(z)\nu(dz)\Big)\Big((1-\theta)L\Delta t + (\gamma+\varepsilon)\Delta x\Big).
\end{aligned}$$

Hence since $D_x\phi = -D_y\phi + \varepsilon(x+y)$ and b is Lipschitz continuous, by the maximum point estimates, the definition of $\tilde{b}^{\alpha,\beta}_\delta$, and the Lipschitz bound on ϕ,

$$\begin{aligned}
I_3 &= \theta\mathcal{D}^{\alpha,\beta,\delta}_h[\phi(.,y_0)](t_0, x_0) + (1-\theta)\mathcal{D}^{\alpha,\beta,\delta}_h[\phi(.,y_0)](t_0-\Delta t, x_0) - \tilde{b}^{\alpha,\beta}_\delta(s_0, y_0)\cdot(-D_y\phi(x_0, y_0))\\
&\leq \Big(\tilde{b}^{\alpha,\beta}_\delta(t_0, x_0) - \tilde{b}^{\alpha,\beta}_\delta(s_0, y_0)\Big)\cdot D_x\phi(x_0, y_0) + \varepsilon|x_0 + y_0||\tilde{b}^{\alpha,\beta}_\delta(s_0, y_0)|\\
&\quad + C\Big(1 + \int_{|z|>\delta}\rho(z)\nu(dz)\Big)\Big((1-\theta)\Delta t + (\gamma+\varepsilon)\Delta x\Big)\\
&\leq C\Big(1 + \int_{|z|>\delta}\rho(z)\nu(dz)\Big)\Big((|x_0 - y_0| + |t_0 - s_0|)L + (1-\theta)\Delta t + (\gamma+\varepsilon)\Delta x\Big) + o_\varepsilon(1). \qquad (4.10)
\end{aligned}$$

In the case that η does not depend on t, then a recomputation of the above estimate using the fact that $\tilde{b}^{\alpha,\beta}_\delta(x,t) := b^{\alpha,\beta}(x,t) - \int_{|z|>\delta}\eta^{\alpha,\beta}(x;z)\,\nu(dz)$, leads to

$$I_3 \leq C\Big(|t_0 - s_0| + (1-\theta)\Delta t + \Big(1 + \int_{|z|>\delta}\rho(z)\nu(dz)\Big)\Big(|x_0 - y_0| + (\gamma+\varepsilon)\Delta x\Big)\Big) + o_\varepsilon(1). \qquad (4.11)$$

When η does not depend on both x and t then

$$I_3 \leq C\Big(|t_0 - s_0| + |x_0 - y_0| + (1-\theta)\Delta t + \Big(1 + \int_{|z|>\delta}\rho(z)\nu(dz)\Big)(\gamma+\varepsilon)\Delta x\Big) + o_\varepsilon(1). \qquad (4.12)$$

Step 3: It remains to estimate I_4. We rewrite this term as

$$\begin{aligned}
I_4 &= \vartheta\Big[\mathcal{J}^{\alpha,\beta,\delta}_h[U_h](t_0, x_0) - \mathcal{J}^{\alpha,\beta,\delta}[u](s_0, y_0)\Big] + (1-\vartheta)\Big[\tilde{\mathcal{J}}^{\alpha,\beta,\delta}_h[U_h, W](t_0-\Delta t, x_0) - \mathcal{J}^{\alpha,\beta,\delta}[u](s_0, y_0)\Big]\\
&\equiv \vartheta J_1 + (1-\vartheta)J_2.
\end{aligned}$$

By the definition of W and since $\sum \omega_{\mathbf{j}}(x;h) = 1$, we find that

$$J_2 = \int_{|z|>\delta} \sum_{\mathbf{j}\in\mathbb{Z}^N} \Big\{ u(s_0, y_0) - u(s_0, y_0 + \eta^{\alpha,\beta}(s_0, y_0; z)) \\ -\Big(U_{\mathbf{k}}^p - \xi(t_0, s_0) + \xi(t_0 - \Delta t, s_0) - \frac{\mu}{4T}\Delta t\Big) + U_{\mathbf{k}+\mathbf{j}}^{p-1} \Big\}\, \omega_{\mathbf{j}}(\eta^{\alpha,\beta}(t_0 - \Delta t, x_0; z); h)\, \nu(dz).$$

In the following argument, it is essential that we have $U_{\mathbf{k}}^p$ in the integral defining J_2 and not $U_{\mathbf{k}}^{p-1}$, and this explains why we introduced the strange quantity $\tilde{\mathcal{J}}_h^{\alpha,\beta,\delta}[U_h, W]$ in the first place. Recall that (t_0, x_0, s_0, y_0) is a global maximum point of Ψ, so $\Psi(t_0, x_0, s_0, y_0) \geq \Psi(t_0 - \Delta t, x_0 + x_{\mathbf{j}}, s_0, y_0 + \eta^{\alpha,\beta}(s_0, y_0; z))$, and hence

$$u(s_0, y_0) - u(s_0, y_0 + \eta^{\alpha,\beta}(s_0, y_0; z)) - \Big(U_{\mathbf{k}}^p - \xi(t_0, s_0) + \xi(t_0 - \Delta t, s_0) - \frac{\mu}{4T}\Delta t\Big) + U_{\mathbf{k}+\mathbf{j}}^{p-1} \\ \leq \phi(x_0 + x_{\mathbf{j}}, y_0 + \eta^{\alpha,\beta}(s_0, y_0; z)) - \phi(x_0, y_0).$$

By the nonnegativity of $\omega_{\mathbf{j}}$, the definition of the interpolation $\mathfrak{i}_h$, the error bound (3.5), and assumptions (**A**.3) and (**A**.4), we may use these inequalities to estimate J_2:

$$\begin{aligned}
J_2 &\leq \int_{|z|>\delta} \sum_{\mathbf{j}\in\mathbb{Z}^N} \Big\{ \phi(x_0 + x_{\mathbf{j}}, y_0 + \eta^{\alpha,\beta}(s_0, y_0; z)) - \phi(x_0, y_0) \Big\}\, \omega_{\mathbf{j}}(\eta^{\alpha,\beta}(t_0 - \Delta t, x_0; z); h)\, \nu(dz) \\
&= \int_{|z|>\delta} \Big\{ \mathfrak{i}_h[\phi(x_0 + \cdot, y_0 + \eta^{\alpha,\beta}(s_0, y_0; z))](\eta^{\alpha,\beta}(t_0 - \Delta t, x_0; z)) - \phi(x_0, y_0) \Big\}\; \nu(dz) \\
&\leq \int_{|z|>\delta} \Big\{ \phi(x_0 + \eta^{\alpha,\beta}(t_0 - \Delta t, x_0; z), y_0 + \eta^{\alpha,\beta}(s_0, y_0; z)) - \phi(x_0, y_0) + K(\gamma + \varepsilon)\, (\Delta x)^2 \Big\}\; \nu(dz) \\
&= \int_{|z|>\delta} \Big\{ \gamma(x_0 - y_0) \cdot (\eta^{\alpha,\beta}(s_0, y_0; z) - \eta^{\alpha,\beta}(t_0 - \Delta t, x_0; z)) + \frac{\gamma}{2}|\eta^{\alpha,\beta}(s_0, y_0; z) - \eta^{\alpha,\beta}(t_0 - \Delta t, x_0; z)|^2 \\
&\qquad + \varepsilon\Big(x_0 \cdot \eta^{\alpha,\beta}(t_0 - \Delta t, x_0; z) + y_0 \cdot \eta^{\alpha,\beta}(s_0, y_0; z)\Big) + \frac{\varepsilon}{2}(|\eta^{\alpha,\beta}(t_0 - \Delta t, x_0; z)|^2 + |\eta^{\alpha,\beta}(s_0, y_0; z)|^2) \\
&\qquad + K(\gamma + \varepsilon)\, (\Delta x)^2 \Big\}\, \nu(dz) \\
&\leq C\gamma \Big\{ |x_0 - y_0|\Big(|x_0 - y_0| + |t_0 - s_0| + \Delta t\Big) \int_{|z|>\delta} \rho(z)\, \nu(dz) + \Big(|x_0 - y_0|^2 + |t_0 - s_0|^2 + \Delta t^2\Big) \int_{|z|>\delta} \rho(z)^2\, \nu(dz) \Big\} \\
&\qquad + C\varepsilon(|x_0| + |y_0|) \int_{|z|>\delta} \rho(z)\, \nu(dz) + C\varepsilon \int_{|z|>\delta} \rho(z)^2\, \nu(dz) + C(\gamma + \varepsilon)(\Delta x)^2 \int_{|z|>\delta} \nu(dz).
\end{aligned} \tag{4.13}$$

In the case that η does not depend on t, an easy recomputation of the above estimate shows that

$$J_2 \leq C\Big(\gamma|x_0 - y_0|^2 + \varepsilon(1 + |x_0| + |y_0|)\Big) \int_{|z|>\delta} (\rho(z) + \rho(z)^2)\, \nu(dz) + C(\gamma + \varepsilon)(\Delta x)^2 \int_{|z|>\delta} \nu(dz), \tag{4.14}$$

and when η does not depend on both x and t then

$$J_2 \leq C\varepsilon(1 + |x_0| + |y_0|) \int_{|z|>\delta} (\rho(z) + \rho(z)^2)\, \nu(dz) + C(\gamma + \varepsilon)(\Delta x)^2 \int_{|z|>\delta} \nu(dz). \tag{4.15}$$

Similar but simpler arguments, using the fact that $\Psi(t_0, x_0, s_0, y_0) \geq \Psi(t_0, x_0 + x_{\mathbf{j}}, s_0, y_0 + \eta^{\alpha,\beta}(s_0, y_0; z))$, shows that J_1, and hence also I_4, satisfy the same upper bounds as J_2.

Step 4: By (**A**.3)-(**A**.5) and the definition of $\Gamma(\sigma, \delta)$,

$$\int_{|z|>\delta} \rho(z)^2\, \nu(dz) \leq K^2 \int_{0<|z|<1} |z|^2\, \nu(dz) + \int_{|z|>1} \rho(z)^2\, \nu(dz) \leq C,$$

$$\int_{|z|>\delta} \rho(z)\, \nu(dz) \leq K^2 \int_{\delta<|z|<1} |z|\, \nu(dz) + \int_{|z|>1} \rho(z)^2\, \nu(dz) \leq C(1+\Gamma(\sigma,\delta)),$$
$$\int_{|z|>\delta} \nu(dz) \leq C \int_{\delta<|z|<1} \frac{dz}{|z|^{M+\sigma}} + C \leq C(1+\delta^{-\sigma}).$$

Now we get a bound on μ from (4.8) by using these estimates along with the estimates of steps 1–3 (which are independent of α and β). If we also take into account the fact that $0 < \Delta x < \delta \leq 1$, $\Gamma(\sigma,\delta) \geq 1$, and that we may take $\eta, \gamma \geq 1$ and $\Delta t \leq 1$, we find after combining (4.9), (4.10) &(4.13) and dropping all non-dominant terms that

$$\begin{aligned}
\frac{\mu}{2T} &\leq I_1 + \sup_\alpha \sup_\beta \Big\{ I_2 + I_3 + I_4 \Big\} \\
&\leq C\Big\{\eta\Delta t + \gamma\Delta t^2 + \gamma|t_0 - s_0|^2 + \gamma\frac{\Delta x^2}{\delta^\sigma}\Big\} \\
&\quad + C\Gamma(\sigma,\delta)\Big\{|x_0 - y_0| + |t_0 - s_0| + \Delta t + \gamma\Delta x + \gamma|x_0 - y_0|\Big(|x_0 - y_0| + |t_0 - s_0| + \Delta t\Big)\Big\} \\
&\quad + C\varepsilon\Big\{1 + \Gamma(\sigma,\delta)\Big(|x_0| + |y_0| + \Delta x\Big) + \frac{\Delta x^2}{\delta^\sigma}\Big\}.
\end{aligned} \tag{4.16}$$

Note that by (4.2), all ε-terms go to 0 as $\varepsilon \to 0$ and γ, η, δ are fixed, and $\gamma\frac{\Delta x^2}{\delta^\sigma} \leq \gamma\Delta x\, \delta^{1-\sigma} \leq \Gamma(\sigma,\delta)(\gamma\Delta x)$ since $\Delta x \leq \delta$. Hence in view of estimates (4.2)–(4.5),

$$\frac{\mu}{2T} \leq C\Big(\eta\Delta t + \gamma\Delta t^2 + \frac{\gamma}{\eta^{2q}} + \Gamma(\sigma,\delta)\Big(\frac{1}{\gamma} + \frac{1}{\eta^q} + \Delta t + \gamma\Delta x\Big)\Big) + o_\varepsilon(1).$$

In the case that η does not depend on t, we combine (4.9), (4.11) and (4.14) and find

$$\begin{aligned}
\frac{\mu}{2T} &\leq C\Big\{\eta\Delta t + |t_0 - s_0| + \gamma\frac{\Delta x^2}{\delta^\sigma} + \Gamma(\sigma,\delta)\Big(|x_0 - y_0| + \gamma\Delta x + \gamma|x_0 - y_0|^2\Big)\Big\} + o_\varepsilon(1) \\
&\leq C\Big\{\eta\Delta t + \frac{1}{\eta^q} + \Gamma(\sigma,\delta)\Big(\frac{1}{\gamma} + \gamma\Delta x\Big)\Big\} + o_\varepsilon(1),
\end{aligned}$$

and when η does not depend on both x and t then (4.9), (4.12) and (4.15) are combined to have

$$\begin{aligned}
\frac{\mu}{2T} &\leq C\Big\{\eta\Delta t + |t_0 - s_0| + |x_0 - y_0| + \gamma\frac{\Delta x^2}{\delta^\sigma} + \Gamma(\sigma,\delta)\gamma\Delta x\Big\} + o_\varepsilon(1) \\
&\leq C\Big\{\eta\Delta t + \frac{1}{\eta^q} + \frac{1}{\gamma} + \Gamma(\sigma,\delta)\gamma\Delta x\Big\} + o_\varepsilon(1).
\end{aligned}$$

Conclusion: Sending $\varepsilon \to 0$ and combining the above estimates for μ in the cases whether t_0 and/or s_0 are positive or zero, we find that

$$\mu \leq C\Big(\frac{1}{\gamma} + \frac{1}{\eta^{\tilde{q}}}\Big) + CT\Big(\eta\Delta t + \gamma\Delta t^2 + \frac{\gamma}{\eta^{2q}} + \Gamma(\sigma,\delta)\Big(\frac{1}{\gamma} + \frac{1}{\eta^q} + \Delta t + \gamma\Delta x\Big)\Big), \tag{4.17}$$

when η does not depend on t then

$$\mu \leq C\Big(\frac{1}{\gamma} + \frac{1}{\eta^{\tilde{q}}}\Big) + CT\Big(\eta\Delta t + \frac{1}{\eta^q} + \Gamma(\sigma,\delta)\Big(\frac{1}{\gamma} + \gamma\Delta x\Big)\Big), \tag{4.18}$$

and finally when η does not depend on both x and t then

$$\mu \leq C\Big(\frac{1}{\gamma} + \frac{1}{\eta^{\tilde{q}}}\Big) + CT\Big(\eta\Delta t + \frac{1}{\eta^q} + \frac{1}{\gamma} + \Gamma(\sigma,\delta)\gamma\Delta x\Big). \tag{4.19}$$

Here $q = 1 = \tilde{q}$ if $K(u_0) < \infty$, otherwise $q = \frac{\sigma}{2\sigma - 1}$ and $\tilde{q} = \frac{1}{2\sigma - 1}$ (when $\sigma \neq 1$).

4.3. *Proof of Theorem 3.2 when $\sigma \in [0,1)$*

In this case $\sigma \in (0,1)$, $K(u_0) < \infty$, $\Gamma(\sigma,\delta) = 1$, and $q = 1$ in (4.17) since u is Lipschitz in t by Theorem 2.1 (d). From estimate (4.17) and our assumptions (note that $\Delta x \le \delta \le 1$), we see that the optimal parameter values are $\eta = \gamma$. This leads to the following bound

$$\mu \le C\frac{1}{\gamma} + CT\Big(\frac{1}{\gamma} + \gamma\big(\Delta t + \Delta x\big)\Big).$$

We conclude the proof of Theorem 3.2 (a) by taking $\gamma = (T \wedge 1)^{-\frac{1}{2}}(\Delta t + \Delta x)^{-\frac{1}{2}}$ and then adding the estimate from Lemma 4.1.

4.4. *Proof of Theorem 3.2 when $\sigma \in (1,2)$*

In this case $\sigma \in (1,2)$, $\Gamma(\sigma,\delta) > 1$, and $q = \frac{\sigma}{2\sigma-1}$ and $\tilde{q} = \frac{1}{2\sigma-1}$ in (4.17) since u and U_h are only Hölder $\frac{1}{\sigma}$ in t at $t = 0$ by Theorem 2.1 (c) and Theorem 3.1 (d). The optimal values for η and γ in (4.17) can be chosen by balancing the principal terms. This leads to

$$\gamma = \min\Big\{((T \wedge 1)\Delta x)^{-\frac{1}{2}}, ((T \wedge 1)\Delta t^2)^{-\frac{1}{2}}, \eta^q\Big\} \qquad \text{and} \qquad \eta = ((T \wedge 1)\Delta t)^{-\frac{1}{1+\tilde{q}}}.$$

Then $\frac{1}{1+\tilde{q}} = \frac{2\sigma-1}{2\sigma}$, $(T \wedge 1)\eta\Delta t = \frac{1}{\eta^{\tilde{q}}} = ((T \wedge 1)\Delta t)^{\frac{1}{2\sigma}}$, $\frac{\gamma}{\eta^{2q}} \le \frac{1}{\eta^q} = ((T \wedge 1)\Delta t)^{\frac{1}{2}}$, and by our assumptions (including $\delta, \Delta x, \Delta t \le 1$), (4.17) implies that

$$\begin{aligned}
\mu &\le C\Big(\frac{1}{\gamma} + \frac{1}{\eta^{\tilde{q}}}\Big) + CT\Big(\eta\Delta t + \Gamma(\sigma,\delta)\Big(\frac{1}{\gamma} + \frac{1}{\eta^q} + \gamma\Delta x\Big)\Big)\\
&\le C(1+T)\Big(((T \wedge 1)\Delta t)^{\frac{1}{2\sigma}} + \Gamma(\sigma,\delta)\Big(((T \wedge 1)\Delta x)^{\frac{1}{2}} + ((T \wedge 1)\Delta t)^{\frac{1}{2}}\Big)\Big).
\end{aligned}$$

We conclude the proof of Theorem 3.2 (b) by adding the estimate from Lemma 4.1.

4.5. *Proof of Theorem 3.2 when $\sigma = 1$*

The proof is a combination of the proof of the case $\sigma \in (1,2)$ and the regularization argument of the proof of Theorem 3.1*(e)*. Let $u_0^{\tilde{\varepsilon}}$ be the mollified initial data and $u^{\tilde{\varepsilon}}$ and $U_h^{\tilde{\varepsilon}}$ be the corresponding solutions of (3.2) and (3.8) both with initial condition $u_0^{\tilde{\varepsilon}}$. Then we double the variables by redefining Ψ to be

$$\Psi(t,x,s,y) = U_h^{\tilde{\varepsilon}}(t,x) - u^{\tilde{\varepsilon}}(s,y) - \phi(x,y) - \xi(t,s) - \frac{\mu}{4T}(t+s)$$

where $\mu = \sup_{\mathcal{G}_h^N}(U_h^{\tilde{\varepsilon}} - u^{\tilde{\varepsilon}})$ and ϕ and ξ are the same as before. As before, there exists a maximum point (x_0, y_0, t_0, s_0) of Ψ satisfying (4.1)–(4.4). By Theorem 2.1, $|u^{\tilde{\varepsilon}}(t,y) - u^{\tilde{\varepsilon}}(s,y)| \le K(u_0^{\tilde{\varepsilon}})|t-s|$ for $K(u_0^{\tilde{\varepsilon}}) = C(1 + |\log\tilde{\varepsilon}|)$, and hence by (4.4)

$$|t_0 - s_0| \le \frac{K(u_0^{\tilde{\varepsilon}})}{\eta}. \tag{4.20}$$

At this point the proof continues as for the case $\sigma \in (1,2)$ but with (4.20) replacing (4.5). If either $t_0 = 0$ or $s_0 = 0$ we use as before regularity to estimate μ. E.g. if $s_0 = 0$, then since $\Psi(t,x,t,x) \le \Psi(t_0,x_0,0,y_0)$,

$$U_h^{\tilde{\varepsilon}}(t,x) - u^{\tilde{\varepsilon}}(t,x) - \varepsilon|x|^2 - \frac{\mu}{2T}t \le U_h^{\tilde{\varepsilon}}(t_0,x_0) - u_0^{\tilde{\varepsilon}}(y_0) \le C\Big(K(u_0^{\tilde{\varepsilon}})t_0 + |x_0 - y_0|\Big) \le C\Big(\frac{K(u_0^{\tilde{\varepsilon}})^2}{\eta} + \frac{1}{\gamma}\Big),$$

where we used (4.3) and (4.20) for the last inequality. We send $\varepsilon \to 0$ and take the supremum over $\mathcal{G}_h^N$ to find that

$$\mu \leq C\Big(\frac{K(u_0^{\tilde{\epsilon}})^2}{\eta} + \frac{1}{\gamma}\Big). \tag{4.21}$$

The same bound holds also when $t_0 = 0$. When both $t_0 > 0$ and $s_0 > 0$, the proof for $\sigma \in (1,2)$ is valid also for $\sigma = 1$ up until and including the bound (4.16). We add the estimates on μ, (4.16) and (4.21), use estimates (4.3) and (4.20), and send $\varepsilon \to 0$ (compare with (4.17)), to get

$$\mu \leq C\Big(\frac{1}{\gamma} + \frac{K(u_0^{\tilde{\epsilon}})^2}{\eta}\Big) + CT\Big(\eta\Delta t + \gamma\Delta t^2 + \frac{\gamma K(u_0^{\tilde{\epsilon}})^2}{\eta^2} + |\log\delta|\Big(\frac{1}{\gamma} + \frac{K(u_0^{\tilde{\epsilon}})}{\eta} + \Delta t + \gamma\Delta x\Big)\Big). \tag{4.22}$$

Taking optimal values of γ and η in (4.22) by balancing the principal terms, then leads to

$$\gamma = \min\Big\{((T\wedge 1)\Delta x)^{-\frac{1}{2}}, ((T\wedge 1)\Delta t^2)^{-\frac{1}{2}}, \frac{\eta}{K(u_0^{\tilde{\epsilon}})}\Big\} \qquad \text{and} \qquad \eta = \frac{K(u_0^{\tilde{\epsilon}})}{((T\wedge 1)\Delta t)^{1/2}},$$

and hence

$$(U_h^{\tilde{\epsilon}} - u^{\tilde{\epsilon}}) \leq \mu \leq C(1+T)(T\wedge 1)^{\frac{1}{2}}\Big(\Delta x^{\frac{1}{2}} + |\log\tilde{\epsilon}|\Delta t^{\frac{1}{2}} + |\log\delta|(\Delta x^{\frac{1}{2}} + \Delta t^{\frac{1}{2}})\Big).$$

A bound for $(u^{\tilde{\epsilon}} - U_h^{\tilde{\epsilon}})$ can be found by interchanging the roles of $u^{\tilde{\epsilon}}$ and $U_h^{\tilde{\epsilon}}$. By comparison, Theorems 2.1 *(a)* and 3.1, *(a)*, $|U_h - U_h^{\tilde{\epsilon}}|, |u^{\tilde{\epsilon}} - u^\delta| \leq |u_0^{\tilde{\epsilon}} - u_0| \leq C\tilde{\epsilon}$, and then

$$\begin{aligned}|U_h(t,x) - u^\delta(t,x)| &\leq |U_h(t,x) - U_h^{\tilde{\epsilon}}(t,x)| + |U_h^{\tilde{\epsilon}}(t,x) - u^{\tilde{\epsilon}}(t,x)| + |u^{\tilde{\epsilon}}(t,x) - u^\delta(t,x)| \\ &\leq 2\tilde{\epsilon} + C(1+T)(T\wedge 1)^{\frac{1}{2}}\Big(\Delta x^{\frac{1}{2}} + |\log\tilde{\epsilon}|\Delta t^{\frac{1}{2}} + |\log\delta|(\Delta x^{\frac{1}{2}} + \Delta t^{\frac{1}{2}})\Big).\end{aligned}$$

The proof of Theorem 3.2 *(b)* for $\sigma = 1$ is complete by taking $\tilde{\epsilon} = \Delta t$ and adding the estimate of Lemma 4.1.

4.6. Proof of Theorem 3.3 (a).

We only do the case $\sigma \in (1,2)$. The case $\sigma = 1$ follows in a similar way, cf. proof of Theorem 3.2 for $\sigma = 1$, and the case $\sigma \in [0,1)$ follows directly from Theorem 3.2. Now $\Gamma(\sigma,\delta) > 1$, and $q = \frac{\sigma}{2\sigma-1}$ and $\tilde{q} = \frac{1}{2\sigma-1}$ in (4.17) since u and U_h are only Hölder $\frac{1}{\sigma}$ in t at $t = 0$ by Theorem 2.1 (c) and Theorem 3.1 (d). Note that when $\Delta t \leq \Delta x$, $\gamma \leq \eta^q$, and $\gamma \geq 1$ – then $\frac{\gamma}{\eta^{2q}} \leq \frac{1}{\eta^q}$, $\frac{1}{\eta^q} \leq \frac{1}{\gamma}$ and $\Delta t \leq \gamma\Delta x$. By our assumptions, both (4.17) with $\Delta t \leq \Delta x$ (and then $(1 \leq)\, \gamma \leq \eta^q$, see below!) and (4.18) implies that

$$\mu \leq C\Big(\frac{1}{\gamma} + \frac{1}{\eta^{\tilde{q}}}\Big) + CT\Big(\eta\Delta t + \frac{1}{\eta^q} + \Gamma(\sigma,\delta)\Big(\frac{1}{\gamma} + \gamma\Delta x\Big)\Big).$$

We conclude the proof of Theorem 3.3 (a) by taking taking $\eta = ((T\wedge 1)\Delta t)^{-\frac{1}{1+\tilde{q}}}$, $\gamma = ((T\wedge 1)\Delta x)^{-\frac{1}{2}}$, and then adding the estimate from Lemma 4.1.

4.7. Proof of Theorem 3.3 (b).

In this case $\sigma \in (1,2)$, $\Gamma(\sigma,\delta) > 1$, and $q = 1$ in (4.17) and (4.18) since u and U_h are Lipschitz in t at $t = 0$ by Theorem 2.1 (d) and Theorem 3.1 (e). By our assumptions (note that $\Delta x \leq \delta \leq 1$), both (4.17) with $\Delta t \leq \Delta x$ (and then $\gamma \leq \eta$, see below!) and (4.18) implies that

$$\mu \leq C\Big(\frac{1}{\gamma} + \frac{1}{\eta}\Big) + CT\Big(\eta\Delta t + \frac{1}{\eta} + \Gamma(\sigma,\delta)\Big(\frac{1}{\gamma} + \gamma\Delta x\Big)\Big).$$

We conclude the proof of Theorem 3.3 (b) by taking $\eta = ((T\wedge 1)\Delta t)^{-\frac{1}{2}}$, $\gamma = ((T\wedge 1)\Delta x)^{-\frac{1}{2}}$, and then adding the estimate from Lemma 4.1.

4.8. Proof of Theorem 3.4 (a).

Again we only do the case $\sigma \in (1, 2)$. The case $\sigma = 1$ follows in a similar way, cf. proof of Theorem 3.2 for $\sigma = 1$, and the case $\sigma \in [0, 1)$ follows directly from Theorem 3.2. Again $\Gamma(\sigma, \delta) > 1$, and $q = \frac{\sigma}{2\sigma-1}$ and $\tilde{q} = \frac{1}{2\sigma-1}$ in (4.19). We conclude the proof by taking taking $\eta = ((T \wedge 1)\Delta t)^{-\frac{1}{1+\tilde{q}}}$, $\gamma = \Big((T \wedge 1)\Gamma(\sigma, \delta)\Delta x\Big)^{-\frac{1}{2}}$, and then adding the estimate from Lemma 4.1.

5. On suboptimal rates for general monotone schemes

A close inspection of our proofs shows that our methods can handle a large class of monotone approximations of (1.1)-(1.2) that allow for truncation errors involving derivatives of at most order two. In most numerical approximations it is possible to use *suboptimal* truncation errors that satisfy this condition. The resulting error estimates will not be optimal in general, but at this point there are no alternative methods to get error estimates for general Isaacs equations.

We illustrate this approach by proving suboptimal error estimates for an improved version of our previous scheme. The idea is to compensate for the truncation of the nonlocal operator $\mathcal{I}^{\alpha,\beta}$ by a vanishing local diffusion. To do so, first note that $\mathcal{I}^{\alpha,\beta}[\phi] = \mathcal{I}^{\alpha,\beta,\delta}[\phi] + \mathcal{I}^{\alpha,\beta}_{\delta}[\phi]$ where $\mathcal{I}^{\alpha,\beta,\delta}[\phi]$ is defined in (2.1) and

$$\mathcal{I}^{\alpha,\beta}_{\delta}[\phi](t, x) = \int_{|z|\leq\delta} \left(\phi(t, x + \eta^{\alpha,\beta}(t, x; z)) - \phi(t, x) - \eta^{\alpha,\beta}(t, x; z) \cdot \nabla_x \phi(t, x)\right) \nu(dz).$$

By Taylor expansion we see that we can approximate this term by the local term (cf. e.g. [23])

$$\mathrm{tr}\Big[a^{\alpha,\beta}_{\delta}(t, x)D^2\phi(t, x)\Big] \qquad \text{with} \qquad a^{\alpha,\beta}_{\delta}(t, x) = \frac{1}{2}\int_{|z|\leq\delta} \eta^{\alpha,\beta}(t, x; z)\eta^{\alpha,\beta}(t, x; z)^T \nu(dz)$$

and the error is $C\|D^3\phi\|_\infty \int_{|z|\leq\delta} |\eta^{\alpha,\beta}(t, x; z)|^3\nu(dz) \leq C\|D^3\phi\|_\infty\delta^{3-\sigma}$ in view of (**A**.3) and (**A**.5). Next we take a monotone finite difference approximation $\mathcal{L}^{\alpha,\beta}_{\delta,h}[\phi]$ of this local term with error $K\|a^{\alpha,\beta}_{\delta}\|_0\|D^4\phi\|_0(\Delta x)^2 \leq K\delta^{2-\sigma}(\Delta x)^2\|D^4\phi\|_0$. Note that to ensure this rate, we have to assume e.g. that $a^{\alpha,\beta}_{\delta}$ is diagonally dominant. Under this assumption, the (wide stencil) schemes of Kushner [28], Bonnans-Zidani [11] or Krylov [26] would give this error. Combining these results, we conclude that $\mathcal{L}^{\alpha,\beta}_{\delta,h}$ is an approximation of $\mathcal{I}^{\alpha,\beta}_{\delta}$ with error

$$\Big|\mathcal{I}^{\alpha,\beta}_{\delta}[\phi] - \mathcal{L}^{\alpha,\beta}_{\delta,h}[\phi]\Big| \leq C\Big(\|D^3\phi\|_0\delta^{3-\sigma} + \|D^4\phi\|_0\Delta x^2\delta^{2-\sigma}\Big).$$

Now we discretize equation (1.1) as in Section 3 except that we do not throw away the $\mathcal{I}^{\alpha,\beta}_{\delta}$-term but rather approximate it by $\mathcal{L}^{\alpha,\beta}_{\delta,h}$. The resulting semidiscrete approximation is then (compare with (3.7))

$$u_t + \inf_{\alpha\in\mathcal{A}}\sup_{\beta\in\mathcal{B}}\Big\{-f^{\alpha,\beta}(t, x) + c^{\alpha,\beta}(t, x)u(t, x) - \mathcal{D}^{\alpha,\beta,\delta}_h[u](t, x) - \mathcal{L}^{\alpha,\beta}_{\delta,h}[u](t, x) - \mathcal{J}^{\alpha,\beta,\delta}_h[u](t, x)\Big\} = 0. \tag{5.1}$$

In view of the discussion above and in Section 3, the truncation error for this scheme is

$$\begin{aligned} E &:= \Big\|b^{\alpha,\beta} \cdot \nabla\phi + \mathcal{I}^{\alpha,\beta}[\phi] - (\mathcal{D}^{\alpha,\beta,\delta}_h + \mathcal{L}^{\alpha,\beta}_{\delta,h} + \mathcal{J}^{\alpha,\beta,\delta}_h)[\phi]\Big\|_0 \\ &\leq C\bigg(\Delta x\,\Gamma(\sigma, \delta)\|D^2\phi\|_0 + \|D^3\phi\|_0\delta^{3-\sigma} + \|D^4\phi\|_0\Delta x^2\delta^{2-\sigma} + \|D^2\phi\|_0\frac{\Delta x^2}{\delta^{\sigma}}\bigg). \end{aligned}$$

For $\sigma \in [0, 1)$ or $\sigma = 1$, the optimal choice of δ is $\delta = \Delta x$ and then $E = O(\Delta x)$ or $E = O(\Delta x|\ln\Delta x|)$ as in the previous section. But when $\sigma \in (1, 2)$, then the two first terms in the bound on E dominate and the optimal choice is $\delta = \Delta x^{\frac{1}{2}}$. The corresponding error $E = O(\Delta x^{\frac{3-\sigma}{2}})$ is better than the (optimal) truncation error $O(\Delta x^{2-\sigma})$ from Section 3 (see Remark 3.4), especially when $\sigma \approx 2$. To find a useful suboptimal bound, note that $|\mathcal{L}_{\delta,h}^{\alpha,\beta}[\phi]| \leq C\|a_\delta^{\alpha,\beta}\|_0\|D^2\phi\|_0 \leq C\delta^{2-\delta}\|D^2\phi\|_0$ and $|\mathcal{L}_\delta^{\alpha,\beta}[\phi]| \leq C\delta^{2-\sigma}\|D^2\phi\|_0$, and then

$$
\begin{aligned}
\tilde{E} &:= \left|b_\delta^{\alpha,\beta}\cdot\nabla\phi - \mathcal{D}_h^{\alpha,\beta,\delta}[\phi]\right| + |\mathcal{J}^{\alpha,\beta,\delta}[\phi] - \mathcal{J}_h^{\alpha,\beta,\delta})[\phi]\big| + \left|\mathcal{L}_\delta^{\alpha,\beta}[\phi]\right| + \left|\mathcal{L}_{\delta,h}^{\alpha,\beta}[\phi]\right| \\
&\leq C\|D^2\phi\|_0\Big(\Delta x\,\Gamma(\sigma,\delta) + \frac{\Delta x^2}{\delta^\sigma} + \delta^{2-\sigma}\Big).
\end{aligned}
$$

This is same estimate that was optimal for the scheme in Section 3.

A fully discrete scheme is then obtained by discretizing (5.1) in time as in (3.8). For simplicity, we only consider an implicit scheme here:

$$
U_{\mathbf{j}}^n = U_{\mathbf{j}}^{n-1} - \Delta t \inf_{\alpha\in\mathcal{A}}\sup_{\beta\in\mathcal{B}}\Big\{-f_{\mathbf{j}}^{\alpha,\beta,n} + c_{\mathbf{j}}^{\alpha,\beta,n}U_{\mathbf{j}}^n - \mathcal{D}_h^{\alpha,\beta,\delta}[U]_{\mathbf{j}}^n - \mathcal{J}_h^{\alpha,\beta,\delta}[U]_{\mathbf{j}}^n - \mathcal{L}_{\delta,h}^{\alpha,\beta}[U]_{\mathbf{j}}^n\Big\}. \tag{5.2}
$$

We have the following result.

Theorem 5.1. *Assume $\mathcal{L}_{\delta,h}^{\alpha,\beta}$ is as explained above and $\eta^{\alpha,\beta}$ does not depend on x, t. Then Theorem 3.4 remains true when the scheme (3.8) is replaced by the scheme (5.2).*

Outline of proof. We follow the proof of Section 4.2 without doing the truncation step in Section 4.1 first. The idea is to estimate separately the terms $\mathcal{L}_\delta^{\alpha,\beta}[\phi]$ and $\mathcal{L}_{\delta,h}^{\alpha,\beta}[\phi]$. By the discussion above and the definition of the test function ϕ, both terms are bounded by the vanishing viscosity like bound $C(\gamma+\varepsilon)\delta^{2-\sigma}$, and in the proof this term would appear as new term I_5 on the right hand side of (4.8). Continuing the proof, the bound on μ in (4.19) will have this additional term i.e.

$$
\mu \leq C\Big(\frac{1}{\gamma} + \frac{1}{\eta^{\tilde{q}}}\Big) + CT\Big(\eta\Delta t + \frac{1}{\eta^q} + \Gamma(\sigma,\delta)\gamma\Delta x + \gamma\delta^{2-\sigma}\Big).
$$

To conclude the (same) error estimates, we now have to modify the choice of γ and take

$$
\gamma = \min\Big(((T\wedge 1)\Gamma(\sigma,\delta)\Delta x)^{-\frac{1}{2}}, (T\wedge 1)^{-\frac{1}{2}}\delta^{-(1-\frac{\sigma}{2})}\Big)
$$

which leads to the bound $\mu \leq \Delta t\text{-term} + C(1+T)(T\wedge 1)^{\frac{1}{2}}(\Gamma(\sigma,\delta)^{\frac{1}{2}}\Delta x^{\frac{1}{2}} + \delta^{1-\frac{\sigma}{2}})$. □

Remark 5.1. (a) If η does not depend on (x, t), then our approach gives error bounds for arbitrary monotone schemes that admit possibly suboptimal error expansion involving no higher order derivatives than order 2.

(b) If η depends on x, then the results will not be so good. Redoing the proof outlined above, we have to replace (4.19) by (4.17) or (4.18) which contain an additional $O(\Gamma(\sigma,\delta)\frac{1}{\gamma})$ term. To get the final error bound, we now have to take a γ that minimize

$$
\Gamma(\sigma,\delta)\Big(\frac{1}{\gamma} + \gamma\Delta x\Big) + \gamma\delta^{2-\sigma}.
$$

This leads to $\gamma = \min\Big(\Delta x^{-1/2}, \frac{\Gamma(\sigma,\delta)^{1/2}}{\delta^{\frac{2-\sigma}{2}}}\Big) = \min(\Delta x^{-1/2}, \delta^{-1/2}) = \delta^{-1/2}$ since $\Delta x \leq \delta < 1$, and then

$$
\mu \leq \cdots + C\Big(\delta^{1-\sigma}(\delta^{1/2} + \Delta x^{1/2}) + \delta^{-1/2}\delta^{2-\sigma}\Big) = \cdots + C\Big(\delta^{1-\sigma}\Delta x^{1/2} + \delta^{\frac{3}{2}-\sigma}\Big).
$$

This error bound is worse than before, and only valid for $\sigma \leq \frac{3}{2}$.

(c) A possible way to obtain general (suboptimal) results when η depends on (x, t), is via continuous dependence results like in [23]. But now such results are also needed for the scheme. Obtaining such results can be challenging in general and will not be considered here.

Acknowledgments

E. R. Jakobsen is supported by the Toppforsk (research excellence) project Waves and Nonlinear Phenomena (WaNP), grant no. 250070 from the Research Council of Norway. I. H. Biswas acknowledges the support received from INSA via *INSA Young Scientist Project*. In addition, Indranil Chowdhury is supported by the research fellowship of *Department of Atomic Energy, India*.

References

[1] G. Barles and C. Imbert. Second-Order Elliptic Integro-Differential Equations: Viscosity Solutions' Theory Revisited. *Ann. Inst. H. Poincar Anal. Non Linaire* 25 (2008), no. 3, 567–585.

[2] G. Barles and E. R. Jakobsen. On the convergence rate of approximation schemes for Hamilton-Jacobi-Bellman equations. *M2AN Math. Model. Numer. Anal.* 36 (2002), no. 1, 33–54.

[3] G. Barles and E. R. Jakobsen. Error bounds for monotone approximation schemes for Hamilton-Jacobi-Bellman equations. *SIAM J. Numer. Anal.* 43 (2005), no. 2, 540–558 (electronic)

[4] G. Barles and E. R. Jakobsen, Error bounds for monotone approximation schemes for parabolic Hamilton-Jacobi-Bellman equations, *Math. Comp.*, 76 (2007), 1861–1893.

[5] G. Barles and P. E. Souganidis, Convergence of approximation schemes for fully nonlinear second order equations, *Asymptotic Anal.*, 4 (1991), 271–283.

[6] I. H. Biswas. On zero-sum stochastic differential games with jump-diffusion driven state: a viscosity solution framework. *SIAM J. Control Optim.* 50 (2012), no. 4, 1823–1858.

[7] I.H. Biswas, E.R. Jakobsen and K.H. Karlsen. Error estimates for a class of finite difference-quadrature schemes for fully nonlinear degenerate parabolic integro-PDEs. *J. Hyperbolic Differ. Equ.* 5 (2008), no. 1, 187–219.

[8] I.H. Biswas, E.R. Jakobsen and K.H. Karlsen. Difference-quadrature schemes for nonlinear degenerate parabolic integro-PDE. *SIAM J. Numer. Anal.* 48 (2010), no. 3, 1110–1135.

[9] I. H. Biswas, E. R. Jakobsen, and K. H. Karlsen. Viscosity solutions for a system of integro-PDEs and connections to optimal switching and control of jump-diffusion processes. *Appl. Math. Optim.* 62 (2010), no. 1, 47–80.

[10] F. Bonnans, S. Maroso and H. Zidani. Error estimates for stochastic differential game: the adverse stopping case Frdric Bonnans, *IMA Journal of Numerical Analysis.* 28(2006), 188–212.

[11] J. F. Bonnans and H. Zidani. Consistency of generalized finite difference schemes for the stochastic HJB equation. *SIAM J. Numer. Anal.* 41 (2003), no. 3, 1008–1021.

[12] L. A. Caffarelli and P. E. Souganidis. A rate of convergence for monotone finite difference approximations to fully nonlinear, uniformly elliptic PDEs. *Comm. Pure Appl. Math.* 61 (2008), no. 1, 1–17.

[13] I. Capuzzo-Dolcetta and M. Falcone. Discrete dynamic programming and viscosity solutions of the Bellman equation. *Ann. Inst. H. Poincare Anal. Non Lineaire* 6 (1989), suppl., 161–183.

[14] S. Cifani and E. R. Jakobsen. On numerical methods and error estimates for degenerate fractional convection-diffusion equations. *Numer. Math.* 127(2014), no. 3, 447–483.

[15] R. Cont and P. Tankov. Financial Modelling with Jump Processes. *Chapman and Hall/CRC*, 2003.

[16] M. G. Crandall, H. Ishii, and P. L. Lions. User's guide to viscosity solutions of second order partial differential equations. *Bull. Amer. Math. Soc. (N.S.)* 27 (1992), no. 1, 1–67.

[17] M. G. Crandall and P.-L. Lions, Two approximations of solutions of Hamilton-Jacobi equations, *Math. Comp.*, 43 (1984), 1–19.

[18] L. C. Evans and P. E. Souganidis. Differential games and representation formulas solutions for Hamilton-Jacobi-Isaacs equations, *Indiana Univ. Math. J.*, 33(1984), 773–797.

[19] W. H. Fleming and P.E. Souganidis, On the existence of value functions of two player, zero-sum stochastic differential Games, *Indiana University Mathematics Journal*, 38 (1989), 293–314.

[20] E. R. Jakobsen. On error bounds for approximation schemes for non-convex degenerate elliptic equations. BIT 44 (2004), no. 2, 269–285.

[21] E. R. Jakobsen. On error bounds for monotone approximation schemes for multi-dimensional Isaacs equations. *Asymptotic Analysis* 49 (2006), no. 3, 249–273.

[22] E. R. Jakobsen and K. H. Karlsen. Continuous dependence estimates for viscosity solutions of integro-PDEs. *J. Differential Equations*, 212 (2005), no. 2, 278–318.

[23] E. R. Jakobsen, K. H. Karlsen, and C. La Chioma. Error estimates for approximate solutions to Bellman equations associated with controlled jump-diffusions. *Numer. Math.* 110 (2008), no. 2, 221–255.

[24] N. V. Krylov. On the rate of convergence of finite-difference approximations for Bellman's equations, *Algebra i Analiz*, 9 (1997), 245–256.

[25] N. V. Krylov, On the rate of convergence of finite-difference approximations for Bellman's equations with variable coefficients, *Probab. Theory Related Fields*, 117 (2000), 1–16.

[26] N. V. Krylov, The rate of convergence of finite-difference approximations for Bellman equations with Lipschitz coefficients, *Appl. Math. Optim.*, 52 (2005), 365–399.

[27] N. V. Krylov. On the rate of convergence of finite-difference approximations for elliptic Isaacs equations in smooth domains. *Comm. Partial Differential Equations* 40 (2015), no. 8, 1393–1407.

[28] H. J. Kushner and P. G. Dupuis. *Numerical methods for stochastic control problems in continuous time*. Springer-Verlag, New York, 1992.

[29] P.-L. Lions and B. Mercier, Approximation numérique équations de Hamilton-Jacobi-Bellman, *RAIRO Anal. Numér.*, 14 (1980), 369–393.

[30] P. E. Souganidis. Approximation schemes for viscosity solutions of Hamilton-Jacobi equations. *J. Differential Equations* 59 (1985), no. 1, 1–43.

[31] O. Turanova. Error estimates for approximations of nonhomogeneous nonlinear uniformly elliptic equations. *Calc. Var. Partial Differential Equations* 54 (2015), no. 3, 2939–2983.

[32] O. Turanova. Error estimates for approximations of nonlinear uniformly parabolic equations. *NoDEA Nonlinear Differential Equations Appl.* 22 (2015), no. 3, 345–389.