

IMPLICIT AND FULLY DISCRETE APPROXIMATION OF THE SUPERCOOLED STEFAN PROBLEM IN THE PRESENCE OF BLOW-UPS*

CHRISTA CUCHIERO[†], CHRISTOPH REISINGER[‡], AND STEFAN RIGGER[§]

Abstract. We consider two approximation schemes of the one-dimensional supercooled Stefan problem and prove their convergence, even in the presence of finite time blow-ups. All proofs are based on a probabilistic reformulation recently considered in the literature. The first scheme is a version of the time-stepping scheme studied by Kaushansky et al. [*Ann. Appl. Probab.*, 33 (2023), pp. 274–298], but here the flux over the free boundary and its velocity are coupled implicitly. Moreover, we extend the analysis to more general driving processes than Brownian motion. The second scheme is a Donsker-type approximation, also interpretable as an implicit finite difference scheme, for which global convergence is shown under minor technical conditions. With stronger assumptions, which apply in cases without blow-ups, we obtain additionally a convergence rate arbitrarily close to $1/2$. Our numerical results suggest that this rate also holds for less regular solutions, in contrast to explicit schemes, and allow a sharper resolution of the discontinuous free boundary in the blow-up regime.

Key words. supercooled Stefan problem, approximation schemes, finite time blow-up, convergence analysis, Donsker approximation

MSC codes. 80A22, 35R35, 65N06, 65N75

DOI. 10.1137/22M1509722

1. Introduction. The classical PDE formulation of the one-dimensional one-phase Stefan problem is

$$(1) \quad \begin{aligned} \partial_t u &= \frac{1}{2} \partial_{xx} u, \quad x > \Lambda_t, \quad t \geq 0, \\ u(t, \Lambda_t) &= 0, \quad t \geq 0, \\ \dot{\Lambda}_t &= \frac{\alpha}{2} \partial_x u(t, \Lambda_t), \quad t \geq 0, \\ u(0, x) &= f(x), \quad x \geq 0, \quad \text{and} \quad \Lambda_0 = 0. \end{aligned}$$

In physics terminology, $-f$ is the initial temperature in a liquid, where we will consider $f \geq 0$ corresponding to the supercooled regime; $-u(t, \cdot)$ is the temperature in the liquid phase at time t ; and Λ_t is the location of the liquid-solid boundary at time t . The above relationship between the flux $\partial_x u$ at the free boundary Λ_t and the growth rate $\dot{\Lambda}_t$ of the solid phase is known as the *Stefan condition*. It expresses the energy conservation at the interface in integrated form,

*Received by the editors July 15, 2022; accepted for publication (in revised form) September 28, 2023; published electronically May 9, 2024.

<https://doi.org/10.1137/22M1509722>

Funding: This work was supported by Vienna Science and Technology Fund (WWTF) grant MA16-021 and Austrian Science Fund (FWF) grant Y 1235 of the START-program.

[†]Department of Statistics and Operations Research, Data Science, Vienna University, A-1090 Wien, Austria (christa.cuchiero@univie.ac.at).

[‡]Mathematical Institute, University of Oxford, OX2 6GG, Oxford, UK (christoph.reisinger@maths.ox.ac.uk).

[§]Department of Statistics and Operations Research, Vienna University, A-1090 Wien, Austria (stefan.rigger@univie.ac.at).

$$\Lambda_t = \alpha \left(1 - \int_{\Lambda_t}^{\infty} u(t, x) dx \right).$$

Variants of the Stefan problem, originally proposed in [38], have been studied in great detail in the applied mathematics literature. In particular, it has been established in a string of works starting in the 1970s (see, e.g., [37, 22, 16, 17, 12, 23, 28, 18, 39, 11, 24, 40]) that the supercooled Stefan problem may exhibit finite time *blow-ups*, whereby continuous solutions $t \mapsto \Lambda_t$ cease to exist.

As closed-form solutions are rarely available for this type of free boundary problem, one typically has to resort to numerical methods. We refer the reader to [4, 33] for a survey of numerical schemes. These works treat classical cases with different boundary conditions, where sign requirements (related to well-posedness without blow-ups) on the initial or boundary data—usually that the temperature of the material in the liquid phase is positive—are imposed. In the supercooled regime, where such conditions are not satisfied, regularization techniques to prevent the formation of blow-ups in finite time can be applied. For instance, in [27] the effect of kinetic undercooling as a regularizing mechanism is analyzed and it is shown how the (unregularized) supercooled Stefan problem can be recovered asymptotically. These asymptotics are also compared to numerical solutions of the unregularized problem. The authors there follow (for both the regularized and the unregularized cases) the so-called method of lines considered in [15]. More precisely, the continuous time PDE systems are discretized and solved at successive times as a sequence of free boundary problems for ordinary differential equations. They can in turn be tackled via the so-called method of invariant embedding described in [32], which gives rise to explicitly solvable Riccati equations. Note that in the numerical studies of the unregularized case the blow-up behavior cannot be fully reproduced due to truncation errors. Further alternative (less physical) regularization approaches are discussed in [19], albeit without a thorough numerical analysis. In [2], the ill-posed Stefan problem for melting a superheated solid, which is mathematically identical to the supercooled Stefan problem, is analyzed. There, similarly as in [27], the method of lines is applied and existing numerical results from the literature in the blow-up regime are reproduced. In [1], blow-up points to one-phase Stefan problems, however with Dirichlet boundary conditions, are treated and studied numerically. Let us finally mention the recent paper [31], which analyzes the Fisher–Stefan model, a generalization of the well-known Fisher–KPP model, in the context of biological invasion, where the speed of the moving boundary is related to the flux of population there. By rescaling this problem, it can be compared to the supercooled Stefan problem. In this context, the authors of [31] provide new links between these models and a numerical scheme for the Fisher–Stefan model.

Despite this plethora of articles on ill-posed Stefan problems, to our knowledge, with the exception of the particle simulation scheme of [26], no provably convergent algorithm is known for the blow-up scenario. We contribute to this literature by proposing a class of numerical schemes for which we can prove global convergence, more specifically that the discrete approximation of the free boundary converges to the true free boundary at all continuity points.

The following simple finite difference scheme is a canonical example of the schemes we will consider for (1), where u_i^k is an approximation to $u(kh, i\sqrt{h})$ for positive integers i and k and some fixed mesh widths $h > 0$ in time and \sqrt{h} in space, respectively:

$$\begin{aligned}
(2) \quad & \frac{u_i^k - u_i^{k-1}}{h} = \frac{1}{2} \frac{u_{i+1}^{k-1} - 2u_i^{k-1} + u_{i-1}^{k-1}}{(\sqrt{h})^2}, \quad k > 0, i > i_k, \\
& i_k = \left\lfloor \frac{\alpha}{\Delta x} \left(1 - \sum_{i=i_k+1}^{\infty} u_i^k \right) \right\rfloor, \quad k > 0, \\
& u_i^0 = \int_{i\sqrt{h}}^{(i+1)\sqrt{h}} f(x) dx, \quad i \geq 0, \quad \text{and} \quad u_i^k = 0, \quad k > 0, i \leq i_k.
\end{aligned}$$

We will discuss the implementation of a scheme of type (2) in section 4, with a slightly perturbed initial condition, as motivated later. Note that (2) is reminiscent of the forward Euler approximation of the heat equation but is nonetheless an implicit nonlinear equation through the dependence of the discrete free boundary $i_k \sqrt{h}$ on the solution $(u_i^k)_i$ at time kh .

Our analysis is based on the following probabilistic reformulation of the problem:

$$(3) \quad \begin{cases} X_t = X_{0-} + B_t - \Lambda_t, \\ \tau = \inf\{t \geq 0 : X_t \leq 0\}, \\ \Lambda_t = \alpha \mathbb{P}(\tau \leq t), \end{cases}$$

where X_{0-} is a real-valued nonnegative random variable, $\alpha > 0$, and B is a standard Brownian motion independent of X_{0-} . We will also study extensions where B is replaced by a more general continuous-time process.

The link between (3) and (1) is found by noticing that over sufficiently small times the forward density $p(t, \cdot)$, $t \in [0, T]$, of the absorbed process $X_t \mathbf{1}_{\{\tau > t\}}$, $t \in [0, T]$, on $(0, \infty)$ satisfies the initial boundary value problem

$$\begin{aligned}
(4) \quad & \partial_t p = \frac{1}{2} \partial_{xx} p + \dot{\Lambda}_t \partial_x p, \quad x \geq 0, \quad t \in [0, T], \\
& p(0, x) = f(x), \quad x \geq 0, \quad \text{and} \quad p(t, 0) = 0, \quad t \in [0, T], \\
& \Lambda_t = \alpha \left(1 - \int_0^\infty p(t, x) dx \right), \quad t \in [0, T],
\end{aligned}$$

as long as X_{0-} has a sufficiently regular density f supported on $[0, \infty)$. Applying the transformation $u(t, x) := p(t, x - \Lambda_t)$, $x \geq \Lambda_t$, then leads to (1) (see [10] for a rigorous proof).

The problem (3) and its variants have recently been used to model systemic risk of interconnected financial systems. There, Λ_t models the proportion of banks defaulted at time t and α the strength of interbank borrowing (see, e.g., [20, 34, 21, 7] for details and further references). Equations with similar mean-field effects through thresholding derive from integrate-and-fire models of neurons as considered in [9].

Besides the applicability of (3) to systemic risk modeling and neuroscience, which is justified by certain propagation of chaos results (see [9, 8]), the probabilistic formulation (3) also has several mathematical advantages. Most notably, it allows for a rigorous definition of global solutions to (1) even in the presence of blow-ups. For this, it is, however, necessary to first establish appropriate solution concepts for the probabilistic version of (3). Indeed, due to the singular interaction, (3) does not have unique solutions in general. Therefore, two physically and economically meaningful notions of solutions have been developed: so-called *physical solutions* and *minimal solutions*. A rigorous characterization of unique physical solutions was given under certain technical assumptions on the initial conditions in [10]. The concept of *minimal*

solutions, which are by definition unique, was treated in detail in [8]. In particular, it was shown there that the minimal solution is also a physical one. It is still an open question whether uniqueness of physical solutions holds in general, but it is clear that the minimal solution coincides with the minimal physical one.

Among the numerical schemes introduced recently for the approximation of (3) are those using time stepping (see [26]), iterations with heat kernels (see [30]), and particle systems (see [25]), where in the latter the numerical schemes are based on the corresponding propagation of chaos result.

Motivated by promising numerical experiments based on implicit instead of explicit schemes, we here first prove a global convergence result for an *implicit version of the Euler time-stepping scheme*, as considered in [26]. Furthermore, we extend this result to more general drivers than Brownian motion and allow for processes that satisfy the so-called crossing property (see (8)), which holds, e.g., for fractional Brownian motion with arbitrary Hurst parameter in $(0, 1)$ and some nondegenerate continuous semimartingales (see Example 2.1).

To further reduce the numerical complexity, we then consider a scheme that, in contrast to the time-stepping algorithm, does not require a (time-consuming) Monte Carlo simulation of the corresponding particle system. The main result in the second part of the paper is that convergence in the Brownian case still holds after a Donsker approximation of the driving Brownian motion, i.e., the increments over single time steps are approximated by random variables matching the first two moments of the Brownian increments. Under more restrictive conditions, in particular ruling out blow-ups, a rate of $1/2$ (modulo a log factor) is proven.

The computational significance of this result is that by use of suitable discrete random variables, the discrete measure is supported on a recombining binomial tree and can be computed deterministically by a recursion over time steps without simulation. The resulting discrete equations, of which (2) is a special case, are then interpretable as approximation schemes (finite difference schemes) for (1). We refer to these as *fully discrete* schemes, in contrast to the schemes based on time-stepping approximations alone, which require additional (particle) approximations.

We observe the following advantage in computational complexity. If the step size in time is h , the spacing of the tree nodes is $h^{1/2}$. On a given time horizon and finite spatial domain (by truncation of the real line), the total number of nodes, and hence the computational complexity, is therefore $O(h^{-3/2})$. Although we only prove an order of convergence in the case of no blow-ups, we find empirically that in all regimes studied (jumps, no jumps, different regularities) the error is of order $h^{1/2}$; see section 5 for details. Hence, to achieve an error of ϵ , the computational complexity is $O(\epsilon^{-3})$. In contrast to that, for the Monte Carlo time-stepping scheme of [25, 26], the simulation error using n particles is $O(n^{-1/2})$, at cost $O(n)$, and the time-stepping error in the most regular case is $h^{1/2}$. Setting n proportional to $[h^{-1}]$, the computational complexity is then $O(\epsilon^{-4})$ to achieve an error of ϵ . Moreover, for low regularity, the convergence of the (explicit) time-stepping scheme is found to be arbitrarily slow in [26], and hence the computational complexity for prescribed accuracy is arbitrarily high. For the Donsker scheme, we do not observe such a phenomenon but rather similar convergence rates in regular and irregular regimes. At the same time, the computational time is also considerably lower.

The remainder of the article is organized as follows. In section 1.1, we introduce relevant notation applied throughout the paper. Section 2 is dedicated to the convergence result of the implicit time-stepping scheme extended to more general driving processes than Brownian motion. In section 3, we then introduce the Donsker-type

approximation scheme, which has lower complexity and for which we can prove a rate of convergence in the regular case without blow-up. Section 4 details the implementation of the schemes, while section 5 concludes with numerical tests showing a better resolution of jumps by implicit schemes compared to explicit schemes in the blow-up regime, particularly on coarse meshes, and a uniform convergence rate of $1/2$ of the Donsker scheme across all scenarios, even where no such rate is proven.

1.1. Notation. Throughout the paper, $D([-1, \infty))$ denotes the space of càdlàg functions on $[-1, \infty)$ endowed with the Skorokhod M_1 -topology. For properties of the M_1 -topology, we refer the reader to [41], [8, Appendix A]. For $x \in D([-1, \infty))$, we denote by $\text{Disc}(x)$ the set of discontinuity points of x . We also define the path functional $\lambda_t(x) := \mathbb{1}_{(0, \infty)}(\inf_{0 \leq s \leq t} x_s)$, such that $\lambda_t(x) = 0$ if $x_s > 0$ on $[0, t]$ and $\lambda_t(x) = 1$ otherwise. The space of continuous functions on $[0, \infty)$ is denoted by $C([0, \infty))$ and endowed with the topology of compact convergence, i.e., $w_n \rightarrow w$ in $C([0, \infty))$ if and only if $w_n|_K \rightarrow w|_K$ uniformly for every compact $K \subseteq [0, \infty)$. We define the following space of cumulative distribution functions on \mathbb{R} , supported on $[0, \infty]$, where \mathbb{R} denotes the two-point compactification of \mathbb{R} :

$$(5) \quad M := \{\ell: \mathbb{R} \rightarrow [0, 1] \mid \ell \text{ càdlàg and increasing, } \ell_t = 0 \text{ for } t < 0 \text{ and } \ell_\infty = 1\}.$$

The reason why we use as domain of definition \mathbb{R} instead of $[0, \infty]$ is to avoid that 0 necessarily becomes a point of continuity. We endow M with the topology induced by the Lévy metric, i.e., $\ell^n \rightarrow \ell$ in M if and only if $\ell_t^n \rightarrow \ell_t$ for every $t \in [0, \infty] \setminus \text{Disc}(\ell)$. This topology turns M into a compact Polish space. Furthermore, we define $\bar{E} := C([0, \infty)) \times M$ and endow it with the product topology. For $\alpha > 0$, we define $\iota_\alpha: \bar{E} \rightarrow D([-1, \infty))$ for $w \in C([0, \infty))$ and $\ell \in M$ as

$$(6) \quad \iota_\alpha(w, \ell)_t := \begin{cases} w_0, & t \in [-1, 0) \\ w_t - \alpha \ell_t, & t \in [0, \infty). \end{cases}$$

Throughout the paper, we denote by x a generic element of $D([-1, \infty))$, by w a generic element of $C([0, \infty))$, and by ℓ a generic element of M . If S is a Polish space, we denote the space of probability measures on S by $\mathcal{P}(S)$ and endow it with the topology of weak convergence; i.e., we say that $\mu_n \rightarrow \mu$ in $\mathcal{P}(S)$ if and only if $\int_S F(x) d\mu_n(x) \rightarrow \int_S F(x) d\mu(x)$ for all $F \in C_b(S; \mathbb{R})$, where $C_b(S; \mathbb{R})$ denotes the space of continuous bounded functions from S to \mathbb{R} . If $\mu \in \mathcal{P}(S)$ and $F: S \rightarrow \mathbb{R}$, we denote the integral of F with respect to μ also with brackets; i.e., we write $\int_S F(x) d\mu(x) = \langle \mu, F \rangle$. Furthermore, if ν is the pushforward of the measure μ with respect to the map T , we denote this by $T(\mu) = \nu$. In particular, with a slight abuse of notation, if ξ denotes a measure on \bar{E} , $\iota_\alpha(\xi)$ denotes its pushforward on $D([-1, \infty))$.

2. Convergence of the time-stepping scheme. In this section, we prove convergence of an implicit version of the time-stepping scheme considered in [26] for more general driving processes than Brownian motion. Specifically, consider the following McKean–Vlasov problem:

$$(7) \quad \begin{cases} X_t = X_{0-} + Z_t - \Lambda_t, \\ \tau = \inf\{t \geq 0 : X_t \leq 0\}, \\ \Lambda_t = \alpha \mathbb{P}(\tau \leq t), \end{cases}$$

where $X_{0-} \in \mathbb{R}$, $\alpha > 0$, and Z is a continuous stochastic process with $Z_0 = 0$ independent of X_{0-} satisfying the following *crossing property*:

$$(8) \quad \mathbb{P}\left(\tau < \infty, \inf_{0 \leq s \leq h} (Z_{\tau+s} - Z_\tau) = 0\right) = 0, \quad h > 0,$$

for every stopping time τ with respect to the natural filtration of $X_{0-} + Z$. Note that the crossing property we use here is slightly more general than the one defined in [9, 8, 35] since we require (8) to hold for any stopping time, not just for the hitting time of zero. Note also that (8) holds for every stopping time if and only if it holds for every bounded stopping time.

Example 2.1.

- (a) Let $Z := M + Y$, where M is a continuous local martingale and Y is 1/2-Hölder continuous on compacts. Suppose additionally that for every $K \in \mathbb{N}$ there is a strictly positive random variable ϵ_K such that $\langle M \rangle_t - \langle M \rangle_s \geq \epsilon_K(t - s)$ for $0 \leq s \leq t \leq K$, where $\langle M \rangle$ is the quadratic variation of M . Then the proof of Theorem 3.5 in [3] shows that Z satisfies the crossing property (8). In particular, we may choose Z to be Brownian motion.
- (b) The process $Z := B^H$, where B^H is fractional Brownian motion with Hurst parameter $H \in (0, 1)$, satisfies the crossing property (8) by [36, Theorem 1.1].

We start by establishing existence (and uniqueness) of minimal solutions which are defined as follows. We call a solution $(\underline{X}, \underline{\tau}, \underline{\Lambda})$ to the McKean–Vlasov problem (7) *minimal* if for every solution (X, τ, Λ) to (7) we have

$$(9) \quad \underline{\Lambda}_t \leq \Lambda_t, \quad t \geq 0.$$

We introduce the fixed-point operator Γ , defined for $\ell \in M$,

$$(10) \quad \begin{cases} X_t[\ell] = X_{0-} + Z_t - \alpha \ell_t, \\ \tau[\ell] = \inf\{t \geq 0 : X_t[\ell] \leq 0\}, \\ \Gamma[\ell]_t = \mathbb{P}(\tau[\ell] \leq t). \end{cases}$$

As the next proposition shows, Γ is continuous on M .

PROPOSITION 2.2. *The operator $\Gamma : M \rightarrow M$ is continuous.*

Proof. Suppose that $\ell^n \rightarrow \ell$ in M . Define

$$\xi^n := \text{law}((X_{0-} + Z, \ell^n)), \quad \xi := \text{law}((X_{0-} + Z, \ell)).$$

Then $\xi^n \rightarrow \xi$ in $\mathcal{P}(\bar{E})$. Set $\mu_n := \iota_\alpha(\xi^n)$ and $\mu := \iota_\alpha(\xi)$, and note that $\langle \mu_n, \lambda_t \rangle = \Gamma[\ell^n]_t$ and $\langle \mu, \lambda_t \rangle = \Gamma[\ell]_t$, where λ was defined in section 1.1.

By Theorem 4.2 in [8] and the continuous mapping theorem, $\mu_n \rightarrow \mu$ in $\mathcal{P}(D([-1, \infty)))$. For $x \in D([-1, \infty))$, set $\tau_0(x) := \inf\{t \geq 0 : x_t \leq 0\}$ and define $\tau_\ell(w) := \inf\{t \geq 0 : w_t - \alpha \ell_t \leq 0\}$. Fixing $h > 0$, we compute

$$(11) \quad \begin{aligned} & \mu(\{x \in D([-1, \infty)) : \tau_0(x) < \infty, \inf_{0 \leq s \leq h} (x_{\tau_0+s} - x_{\tau_0}) = 0\}) \\ &= \xi(\{(w, \bar{\ell}) \in \bar{E} : \tau_0(w - \alpha \bar{\ell}) < \infty, \inf_{0 \leq s \leq h} [(w_{\tau_0+s} - w_{\tau_0}) - \alpha(\bar{\ell}_{\tau_0+s} - \bar{\ell}_{\tau_0})] = 0\}) \\ &\leq \xi\left(\{(w, \bar{\ell}) \in \bar{E} : \tau_\ell < \infty, \inf_{0 \leq s \leq h} (w_{\tau_\ell+s} - w_{\tau_\ell}) = 0\}\right) \\ &= \mathbb{P}\left(\tau_\ell < \infty, \inf_{0 \leq s \leq h} (Z_{\tau_\ell+s} - Z_{\tau_\ell}) = 0\right) = 0, \end{aligned}$$

where the inequality is due to the fact that $\bar{\ell} \in M$ is increasing. Noting that the analogue of Lemma 5.5 in [8] holds with the assumption that (11) vanishes for every $h > 0$, it follows that

$$\lim_{n \rightarrow \infty} \Gamma[\ell^n] = \lim_{n \rightarrow \infty} \langle \mu_n, \lambda \rangle = \langle \mu, \lambda \rangle = \Gamma[\ell]. \quad \square$$

THEOREM 2.3. *There is a (unique) minimal solution to (7), and it is given by*

$$(12) \quad \underline{\Lambda} = \alpha \lim_{k \rightarrow \infty} \Gamma^{(k)}[0],$$

where the convergence is understood as $\Gamma^{(k)}[0] \rightarrow \alpha^{-1} \underline{\Lambda}$ in M .

Proof. This is analogous to the proof of Proposition 2.3 in [8], making use of Proposition 2.2 above. \square

For $\Delta > 0$, define a time-discretized version Z^Δ by $Z_t^\Delta := Z_{\lfloor \frac{t}{\Delta} \rfloor \Delta}$ for $t \geq 0$. In analogy to the situation in continuous time, we define a fixed-point operator and use it to construct the minimal solution in this time-discretized version of the McKean–Vlasov problem and prove continuity of the operator in a suitable sense.

LEMMA 2.4. *For $\ell \in M$, set*

$$(13) \quad \begin{cases} X_t^\Delta[\ell] = X_{0-} + Z_t^\Delta - \alpha \ell_{\lfloor \frac{t}{\Delta} \rfloor \Delta}, \\ \tau^\Delta[\ell] = \inf\{t \geq 0 : X_t^\Delta[\ell] \leq 0\}, \\ \Gamma_\Delta[\ell]_t = \mathbb{P}(\tau^\Delta[\ell] \leq t). \end{cases}$$

Suppose that $\ell_{\lfloor \frac{\cdot}{\Delta} \rfloor \Delta}^n \rightarrow \ell_{\lfloor \frac{\cdot}{\Delta} \rfloor \Delta}$ in M and that $\ell_0^n = \ell_0$. Assume in addition that either $\text{law}(X_{0-})$ is atomless or that $\text{law}(Z_t)$ is atomless for every $t > 0$. Then $\lim_{n \rightarrow \infty} \Gamma_\Delta[\ell^n] = \Gamma_\Delta[\ell]$ in M .

Proof. Note that the condition $\ell_{\lfloor \frac{\cdot}{\Delta} \rfloor \Delta}^n \rightarrow \ell_{\lfloor \frac{\cdot}{\Delta} \rfloor \Delta}$ in M implies $\ell_{k\Delta}^n \rightarrow \ell_{k\Delta}$ for $k \in \mathbb{N}$. The assumption $\ell_0^n = \ell_0$ implies $\mathbb{P}(\tau^\Delta[\ell^n] = 0, \tau^\Delta[\ell] > 0) = 0$. For any $t \geq 0$,

$$\begin{aligned} \Gamma_\Delta[\ell^n]_t - \Gamma_\Delta[\ell]_t &\leq \mathbb{P}(\tau^\Delta[\ell^n] \leq t, \tau^\Delta[\ell] > t) \\ &= \sum_{k=1}^{\lfloor \frac{t}{\Delta} \rfloor} \mathbb{P}(\tau^\Delta[\ell^n] = k\Delta, \tau^\Delta[\ell] > t) \\ &\leq \sum_{k=1}^{\lfloor \frac{t}{\Delta} \rfloor} \mathbb{P}(X_{k\Delta}^\Delta[\ell^n] \leq 0, X_{k\Delta}^\Delta[\ell] > 0) \\ &= \sum_{k=1}^{\lfloor \frac{t}{\Delta} \rfloor} \mathbb{P}(X_{0-} + Z_{k\Delta} - \alpha \ell_{k\Delta}^n \leq 0, X_{0-} + Z_{k\Delta} - \alpha \ell_{k\Delta} > 0) \\ &= \sum_{k=1}^{\lfloor \frac{t}{\Delta} \rfloor} \mathbb{P}(X_{0-} + Z_{k\Delta} \in (\alpha \ell_{k\Delta}, \alpha \ell_{k\Delta}^n]). \end{aligned}$$

Now if $\text{law}(Z_t)$ is atomless for every $t > 0$, using the independence of Z and X_{0-} we may rewrite this as

$$\sum_{k=1}^{\lfloor \frac{t}{\Delta} \rfloor} \mathbb{E}[\mathbb{P}(x + Z_{k\Delta} \in (\alpha \ell_{k\Delta}, \alpha \ell_{k\Delta}^n)) | x = X_{0-}],$$

and the dominated convergence theorem yields that $\limsup_{n \rightarrow \infty} \Gamma_\Delta[\ell^n]_t \leq \Gamma_\Delta[\ell]_t$. The same argument works if we assume that $\text{law}(X_{0-})$ is atomless, as we see by noticing that $\mathbb{P}(X_{0-} + Z_{k\Delta} \in (\alpha \ell_{k\Delta}, \alpha \ell_{k\Delta}^n]) = \mathbb{E}[\mathbb{P}(X_{0-} + z \in (\alpha \ell_{k\Delta}, \alpha \ell_{k\Delta}^n]) | z = Z_{k\Delta}]$. Interchanging the roles of ℓ^n and ℓ in the estimates then yields the claim. \square

DEFINITION 2.5. We say that Λ^Δ solves the discretized McKean–Vlasov problem if $\alpha \Gamma_\Delta[\alpha^{-1} \Lambda^\Delta] = \Lambda^\Delta$. If $\underline{\Lambda}^\Delta$ is such that $\underline{\Lambda}^\Delta \leq \Lambda^\Delta$ for every Λ^Δ that solves the discretized McKean–Vlasov problem, we call $\underline{\Lambda}^\Delta$ minimal. We shall refer to the scheme as an implicit time-stepping scheme.

Remark 2.6. With Definition 2.5, Λ^Δ is a solution if and only if

$$(14) \quad \Lambda_{k\Delta}^\Delta = \alpha \mathbb{P} \left(\min_{0 \leq i \leq k} \{X_{0-} + Z_{i\Delta} - \Lambda_{i\Delta}^\Delta\} \leq 0 \right)$$

for every $k \in \mathbb{N}$. Since $\Lambda_{k\Delta}^\Delta$ appears on both sides of (14), this is an implicit equation for Λ^Δ (the solution of which is not unique in general); therefore, we call our time-stepping scheme implicit. If we define an alternative notion of solution through (14) by taking the minimum on the right-hand side over $\{0, \dots, k-1\}$ instead of $\{0, \dots, k\}$,

$$(15) \quad \tilde{\Lambda}_{k\Delta}^\Delta = \alpha \mathbb{P} \left(\min_{0 \leq i \leq k-1} \{X_{0-} + Z_{i\Delta} - \Lambda_{i\Delta}^\Delta\} \leq 0 \right)$$

for every $k \in \mathbb{N}$, we obtain the time-stepping scheme of [26], which we refer to as explicit in the following.

PROPOSITION 2.7. There is a (unique) minimal solution of the discretized McKean–Vlasov problem, and it is given by

$$(16) \quad \underline{\Lambda}^\Delta = \alpha \lim_{k \rightarrow \infty} \Gamma_\Delta^{(k)}[0].$$

Proof. Γ_Δ is monotone in the sense that if $\ell^1 \leq \ell^2$, then $\Gamma_\Delta[\ell^1] \leq \Gamma_\Delta[\ell^2]$. Therefore,

$$0 \leq \Gamma_\Delta[0] \leq \Gamma_\Delta^{(2)}[0] \leq \dots$$

This allows us to define $\hat{\ell}_t^\Delta := \lim_{k \rightarrow \infty} \Gamma_\Delta^{(k)}[0]_t$. Then $\tilde{\ell}_t^\Delta := \hat{\ell}_{t+}^\Delta$ lies in M and by construction $\lim_{k \rightarrow \infty} \Gamma_\Delta^{(k)}[0] = \tilde{\ell}^\Delta$ in M . Note that $\Gamma_\Delta^{(k)}[0]$ is a step function with jumps at times $\Delta\mathbb{N}$, and so the same holds for $\tilde{\ell}^\Delta$. This implies $\Gamma_\Delta^{(k)}[0]_{\lfloor \frac{t}{\Delta} \rfloor \Delta} = \Gamma_\Delta^{(k)}[0]_t$ and $\tilde{\ell}_{\lfloor \frac{t}{\Delta} \rfloor \Delta}^\Delta = \tilde{\ell}_t^\Delta$ for $t \geq 0$. As $\Gamma_\Delta^{(k)}[0]_s = \Gamma_\Delta[0]_s$ for $0 \leq s < \Delta$, it follows that $\tilde{\ell}_0^\Delta = \Gamma_\Delta^{(k)}[0]_0$ for every $k \in \mathbb{N}$, and therefore we may apply Lemma 2.4 to find

$$\Gamma_\Delta[\tilde{\ell}^\Delta] = \Gamma_\Delta \left[\lim_{k \rightarrow \infty} \Gamma_\Delta^{(k)}[0] \right] = \lim_{k \rightarrow \infty} \Gamma_\Delta[\Gamma_\Delta^{(k)}[0]] = \tilde{\ell}^\Delta,$$

and so $\tilde{\Lambda} := \alpha \tilde{\ell}^\Delta$ is a solution of the discretized McKean–Vlasov problem. If Λ^Δ is another solution, then since $0 \leq \alpha^{-1} \Lambda^\Delta$ and Γ_Δ is monotone we have $\alpha \Gamma_\Delta[0] \leq \alpha \Gamma_\Delta[\alpha^{-1} \Lambda^\Delta] = \Lambda^\Delta$. Proceeding inductively we obtain $\alpha \Gamma_\Delta^{(k)}[0] \leq \Lambda^\Delta$ and therefore $\tilde{\Lambda}_t^\Delta \leq \Lambda_t^\Delta$ for every continuity point of $\tilde{\Lambda}^\Delta$. By right continuity, it follows that $\tilde{\Lambda}^\Delta$ is minimal. \square

Remark 2.8. In contrast to the explicit scheme in [26], the time-stepping scheme we propose here requires us to solve the implicit condition that $\underline{\Lambda}^\Delta$ is the minimal

solution of (14) through the iteration (16). The analogous results we prove here hold as well for the explicit version of the scheme. It is already shown in [26] for a driving Brownian motion that the explicit scheme converges in a suitable sense to discontinuous solutions. However, in the explicit version, the loss over a single time step is determined purely by the minimum of Z , and in the case of Brownian motion this induces a diffusive behavior that smooths out jumps. In contrast, the implicit scheme allows for contagious feedback within a single time step and therefore has the potential to lead to a sharper approximation of the discontinuities for the same step size. We observe this empirically in section 5, in particular Figure 1(a).

We are now ready to state the main result of this section.

THEOREM 2.9. *Choose a sequence $\Delta_n > 0$ such that the resulting discretizations are nested, i.e., $\Delta_n \mathbb{N} \subseteq \Delta_{n+1} \mathbb{N}$ for all $n \in \mathbb{N}$, and such that $\lim_{n \rightarrow \infty} \Delta_n = 0$. Assume in addition that either $\text{law}(X_{0-})$ is atomless or that $\text{law}(Z_t)$ is atomless for every $t > 0$. Then $\lim_{n \rightarrow \infty} \underline{\Lambda}^{\Delta_n} = \underline{\Lambda}$ in M .*

We collect some properties of the space M in the next lemma.

LEMMA 2.10. *Let $\ell^n, \ell \in M$. The following statements are equivalent:*

- (a) $\ell^n \rightarrow \ell$ in M .
- (b) Let t be a continuity point of ℓ . Then, for every $\epsilon > 0$, there is a $\delta > 0$ with

$$(17) \quad \limsup_{n \rightarrow \infty} \sup_{s \in [t-\delta, t+\delta]} |\ell_s^n - \ell_s| \leq \epsilon.$$

- (c) There is a co-countable set $D \subseteq [0, \infty)$ such that $\ell_t^n \rightarrow \ell_t$ for $t \in D$.

Proof. [(a) \implies (b)]: Fix $\epsilon > 0$. As ℓ is continuous at t and ℓ has at most countably many discontinuity points, there is $\delta > 0$ such that both $t + \delta$ and $t - \delta$ are continuity points and $|\ell_s - \ell_r| < \frac{\epsilon}{2}$ for $s, r \in [t - \delta, t + \delta]$. Choose $n_0 \in \mathbb{N}$ large enough such that $|\ell_{t+\delta}^n - \ell_{t+\delta}| < \frac{\epsilon}{2}$ and $|\ell_{t-\delta}^n - \ell_{t-\delta}| < \frac{\epsilon}{2}$ for all $n \geq n_0$. Then for $s \in [t - \delta, t + \delta]$ we have

$$\begin{aligned} \ell_s^n - \ell_s &\leq \ell_{t+\delta}^n - \ell_s < \ell_{t+\delta} - \ell_s + \frac{\epsilon}{2} < \epsilon, \\ \ell_s^n - \ell_s &\geq \ell_{t-\delta}^n - \ell_s > \ell_{t-\delta} - \ell_s - \frac{\epsilon}{2} > -\epsilon, \end{aligned}$$

proving that $\sup_{s \in [t-\delta, t+\delta]} |\ell_s^n - \ell_s| \leq \epsilon$ for $n \geq n_0$. The reverse implication is clear.

[(c) \implies (a)]: By the compactness of M , after passing to a subsequence if necessary, we may assume that $\ell^n \rightarrow \bar{\ell}$ for some $\bar{\ell} \in M$. Since D is co-countable, the set $D \setminus \text{Disc}(\bar{\ell})$ is co-countable as well and therefore dense in $[0, \infty)$. For $t \in D \setminus \text{Disc}(\bar{\ell})$, we obtain $\ell_t = \lim_{n \rightarrow \infty} \ell_t^n = \bar{\ell}_t$, and by right continuity it follows that $\bar{\ell} = \ell$. Hence, $\ell^n \rightarrow \ell$ in M . The reverse implication is clear. \square

PROPOSITION 2.11. *Let $\Delta_n \rightarrow 0$, and suppose that $\ell^n \rightarrow \ell$ in M . Then*

$$(18) \quad \lim_{n \rightarrow \infty} \Gamma_{\Delta_n}[\ell^n] = \Gamma[\ell].$$

Proof.

Step 1. We first show that $(X_{0-} + Z^{\Delta_n}, \ell^n_{\lfloor \frac{\cdot}{\Delta_n} \rfloor \Delta_n}) \rightarrow (X_{0-} + Z, \ell)$ in \bar{E} . Fix $T > 0$. For $\omega \in \Omega$ and $\epsilon > 0$, by uniform continuity there almost surely is $\delta > 0$ such that $|Z_t(\omega) - Z_s(\omega)| < \epsilon$ whenever $|s - t| < \delta$ and $s, t \in [0, T]$. For $\Delta_n < \delta$, we then have

$$|Z_t(\omega) - Z_t^{\Delta_n}(\omega)| < \epsilon, \quad t \in [0, T],$$

and hence $\lim_{n \rightarrow \infty} \|Z - Z^{\Delta_n}\|_{C([0, T])} = 0$ almost surely. Let t be a continuity point of ℓ . Write $g^n := \ell^n_{\lfloor \frac{t}{\Delta_n} \rfloor \Delta_n}$. Fix $\epsilon > 0$, and choose $\delta > 0$ as in Lemma 2.10(b). We then have

$$\limsup_{n \rightarrow \infty} |g_t^n - \ell_t| \leq \limsup_{n \rightarrow \infty} \sup_{s \in [t-\delta, t+\delta]} |\ell_s^n - \ell_s| + \limsup_{n \rightarrow \infty} |\ell_{\lfloor \frac{t}{\Delta_n} \rfloor \Delta_n} - \ell_t| \leq \epsilon.$$

As $\epsilon > 0$ was arbitrary, we see that $g^n \rightarrow \ell$ in M .

Step 2. Setting $\xi_n := \text{law}((X_{0-} + Z^{\Delta_n}, g^n))$ and $\xi := \text{law}((X_{0-} + Z, \ell))$, by Step 1 we have $\xi_n \rightarrow \xi$ in $\mathcal{P}(\bar{E})$. By the same reasoning as in the proof of Proposition 2.2 we obtain that $\langle \iota_\alpha(\xi^n), \lambda \rangle \rightarrow \langle \iota_\alpha(\xi), \lambda \rangle$ in M with λ being defined in section 1.1. Since $\Gamma_{\Delta_n}[\ell^n]_t = \langle \iota_\alpha(\xi^n), \lambda_t \rangle$ and $\Gamma[\ell]_t = \langle \iota_\alpha(\xi), \lambda_t \rangle$, this yields the claim. \square

We are now prepared to prove Theorem 2.9.

Proof of Theorem 2.9. A straightforward induction using Proposition 2.11 shows that $\lim_{n \rightarrow \infty} \Gamma_{\Delta_n}^{(k)}[0] = \Gamma^{(k)}[0]$. Since $\Delta_n \mathbb{N} \subseteq \Delta_{n+1} \mathbb{N}$, it holds that $\Gamma_{\Delta_n}[\ell] \leq \Gamma_{\Delta_{n+1}}[\ell]$ for any $\ell \in M$. Set $J := \text{Disc}(\underline{\Lambda}) \cup \{\text{Disc}(\Gamma^{(k)}[0]) \mid k \in \mathbb{N}\} \cup \bigcup_{n \geq 0} \Delta_n \mathbb{N}$. Then J is countable, and for $t \notin J$ we have by Proposition 2.7

$$\underline{\Lambda}_t = \alpha \lim_{k \rightarrow \infty} \Gamma^{(k)}[0]_t = \alpha \lim_{k \rightarrow \infty} \lim_{n \rightarrow \infty} \Gamma_{\Delta_n}^{(k)}[0]_t = \alpha \lim_{n \rightarrow \infty} \lim_{k \rightarrow \infty} \Gamma_{\Delta_n}^{(k)}[0]_t = \lim_{n \rightarrow \infty} \underline{\Lambda}^{\Delta_n}_t,$$

where we may exchange the order of the limits because $\Gamma_{\Delta_n}^{(k)}[0]_t$ is increasing both in n and in k . Lemma 2.10(c) yields the claim. \square

3. A Donsker-type approximation. In this section, we consider the numerical approximation of a special case of the McKean–Vlasov problem (7) with $Z = B$, with B a Brownian motion independent of X_{0-} . The idea is to replace the Brownian motion by its Donsker approximation and discretize the initial condition in space and the loss in space and time. For $h > 0$, we introduce the discretization operator \mathcal{D}_h acting on càdlàg paths as¹

$$(19) \quad \mathcal{D}_h[x]_t = \left\lfloor \frac{x_{\lfloor \frac{t}{h} \rfloor h}}{\sqrt{h}} \right\rfloor \sqrt{h}, \quad t \in [-1, \infty), \quad x \in D([-1, \infty)).$$

In words, \mathcal{D}_h acts by first discretizing x in time on a mesh of width h and then discretizes the image of x in space on a mesh of width \sqrt{h} . Note that \mathcal{D}_h is a projection and maps M to M . The discretised McKean–Vlasov equation for $h > 0$ and $X_{0-}^h \geq 0$ then reads

$$(20) \quad \begin{cases} X_t^h = X_{0-}^h + B_t^h - \Lambda_t^h, \\ \tau^h = \inf\{t \geq 0 : X_t^h \leq 0\}, \\ \Lambda_t^h = \alpha \mathbb{P}(\tau^h \leq t), \end{cases}$$

where

$$(21) \quad B_t^h = \sqrt{h} \sum_{k=0}^{\lfloor t/h \rfloor} Y_k,$$

and $(Y_i)_{i \in \mathbb{N}}$ is an i.i.d. sequence of random variables satisfying $\mathbb{E}[Y_1] = 0$ and $\mathbb{E}[Y_1^2] = 1$. We repeat the same reasoning as previously, obtaining the minimal solution of (20) through a fixed-point iteration. We call a solution $\underline{\Lambda}^h$ of (20) minimal if $\underline{\Lambda}^h \leq \Lambda^h$ for every Λ^h that solves (20). The minimal solution will then give rise to the implicit Donsker approximation scheme presented in section 4.

¹To distinguish the two schemes, we use here h rather than Δ for the time step.

LEMMA 3.1. Define the operator Γ_h via

$$\begin{aligned} X_t^h[\ell] &= X_{0-}^h + B_t^h - \alpha \ell_t, \\ \tau^h[\ell] &= \inf\{t \geq 0 : X_t^h[\ell] \leq 0\}, \\ \Gamma_h[\ell]_t &:= \mathcal{D}_h[\mathbb{P}(\tau^h[\ell] \leq \cdot)]_t. \end{aligned}$$

Then, it holds that

$$(22) \quad \underline{\Lambda}^h = \alpha \lim_{k \rightarrow \infty} \Gamma_h^{(k)}[0].$$

Proof. By the same arguments given in Lemma 2.4, the iteration $\Gamma_h^{(k)}[0]$ is increasing in k and dominated by any solution Λ^h to (20). Let $t = 0$, set $j := \lfloor 1/\sqrt{h} \rfloor$ then since $\Gamma_h^{(k)}[0]_0$ is increasing in k and takes values in the set $\{0, \sqrt{h}, \dots, j\sqrt{h}\}$ we must have $\Gamma_h^{(l+1)}[0]_0 = \Gamma_h^{(l)}[0]_0$ for some $l \in \{0, 1, \dots, j\}$ and hence $\Gamma_h^{(j+1)}[0]_0 = \Gamma_h^{(j)}[0]_0$. Analogously, since $\Gamma_h^{(k)}[0]_h$ depends on k only through $\Gamma_h^{(k-1)}[0]_0$ and $\Gamma_h^{(k-1)}[0]_h$, we obtain $\Gamma_h^{(2j+1)}[0]_h = \Gamma_h^{(2j)}[0]_h$. Proceeding inductively along $t \in h\mathbb{N}$, the claim follows. \square

Remark 3.2. If we were to pose problem (20) on a finite time horizon $[0, T]$, the result analogous to Lemma 3.1 would hold by the same reasoning. In particular, since the solution operator for the finite time horizon is the restriction to $[0, T]$ of Γ_h , the representation (22) shows that the minimal solution for the problem on $[0, T]$ is simply the restriction of the global minimal solution.

We collect some properties of the operator \mathcal{D}_h for future use in the next Lemma.

LEMMA 3.3. The following statements hold for the operator \mathcal{D}_h defined in (19):

(a) Let $t > 0$, and x, y be càdlàg paths, then we have the estimate

$$|\mathcal{D}_h[x]_t - \mathcal{D}_h[y]_t| \leq \sup_{s \in [t-h, t]} |x_s - y_s| + \sqrt{h}.$$

(b) Suppose that $\lim_{h \rightarrow 0} \ell^h = \ell$ in M . Then, $\lim_{h \rightarrow 0} \mathcal{D}_h[\ell^h] = \ell$ in M .

Proof. [(a)]: This follows readily from the fact that, for any $z, z' \in \mathbb{R}$

$$\left\lfloor \frac{z}{\sqrt{h}} \right\rfloor - \left\lfloor \frac{z'}{\sqrt{h}} \right\rfloor \leq \frac{z}{\sqrt{h}} - \left(\frac{z'}{\sqrt{h}} - 1 \right) \leq \frac{|z - z'|}{\sqrt{h}} + 1.$$

[(b)]: Let t be a continuity point of ℓ . First, note that $\mathcal{D}_h[\ell]_t \rightarrow \ell_t$, since

$$|\ell_t - \mathcal{D}_h[\ell]_t| \leq |\ell_t - \ell_{\lfloor \frac{t}{h} \rfloor h}| + \sqrt{h}.$$

We then estimate

$$|\ell_t - \mathcal{D}_h[\ell^h]_t| \leq |\ell_t - \mathcal{D}_h[\ell]_t| + |\mathcal{D}_h[\ell]_t - \mathcal{D}_h[\ell^h]_t|$$

and note that the second term on the right-hand side vanishes as $h \rightarrow 0$ by (a) and Lemma 2.10 (b). \square

From here on, we will denote by $\underline{\Lambda}^h$ the minimal solution in the Donsker approximation with initial condition

$$(23) \quad X_{0-}^h = \left(\left\lfloor \frac{X_{0-}}{\sqrt{h}} \right\rfloor + \lfloor \log(h)^2 \rfloor \right) \sqrt{h}.$$

The reasons for this choice are twofold: first, to make the theoretical results applicable to the numerical scheme of section 4, we replace X_{0-} by a space-discretized version of itself, given by $\left\lfloor \frac{X_{0-}}{\sqrt{h}} \right\rfloor \sqrt{h}$; second, for technical reasons, we require a perturbation term of order $\mathcal{O}(\sqrt{h} \log(h)^2)$ in the initial condition.

The goal of this section is to show that, with the above choice of initial condition (23), $\underline{A}^h \rightarrow \underline{A}$ in M as $h \rightarrow 0$. We require some additional assumptions.

Assumptions 3.4. We assume the following:

- (i) X_{0-} has a bounded density, which we denote by f .
- (ii) The moment generating function of Y_1 exists in a neighborhood of 0, i.e., $\mathbb{E}[\exp(uY_1)] < \infty$ for some $\delta > 0$ and $|u| < \delta$.

3.1. Convergence results. On an abstract level, the proof of convergence has two main parts: (i) proving that limit points of solutions to the Donsker problem are solutions to the McKean–Vlasov problem, and (ii) identifying limit points of minimal solutions of the Donsker problem as the minimal solution of the McKean–Vlasov problem. Typically, part (ii) is the more challenging one. Fortunately, an approach similar to that in [8] works.

The central idea of this approach is to find a sequence of optimization problems that allows us to relate the minimal solution of the (perturbed) Donsker problem to the minimal solution of the McKean–Vlasov equation, showing that the latter dominates the former asymptotically as $h \rightarrow 0$.

The following lemma constitutes the main technical ingredient of this proof. As a rough intuition, setting $\ell = \underline{A}$, the lemma allows us to quantify how far away the minimal solution of the McKean–Vlasov problem is from being a solution to the Donsker problem when h is small.

LEMMA 3.5. *Suppose that Assumptions 3.4 are satisfied. Then there is a constant $C > 0$ depending on T and $\|f\|_\infty$ such that for all sufficiently small $h > 0$ it holds that*

$$(24) \quad \sup_{\ell \in M} \sup_{t \in [0, T]} |\Gamma_h[\ell]_t - \mathcal{D}_h[\Gamma[\ell]]_t| \leq C\sqrt{h} \log(T/h),$$

where Γ_h is defined as in Lemma 3.1 with $X_{0-}^h = \left\lfloor \frac{X_{0-}}{\sqrt{h}} \right\rfloor \sqrt{h}$.

The proof of the above lemma requires some preparation. The convergence rate in (24) is essentially the convergence rate of the Donsker approximation with respect to the Prokhorov metric, which we transfer to the estimate above through a Lipschitz mapping theorem.

DEFINITION 3.6. *Let (S, d_S) be a Polish space. For $\mu, \nu \in \mathcal{P}(S)$, define the Prokhorov distance between μ and ν as*

$$(25) \quad \rho_P^S(\mu, \nu) = \inf\{\epsilon > 0 : \mu(B) \leq \nu(B^\epsilon) + \epsilon \text{ for all Borel sets } B\},$$

where $B^\epsilon = \{x \in S : d_S(x, B) < \epsilon\}$. If X, Y are random variables on S , we will write $\rho_P^S(X, Y)$ in place of $\rho_P(\text{law}(X), \text{law}(Y))$.

It is well known that the Prokhorov metric metrizes weak convergence (in the probabilistic sense) and $(\mathcal{P}(S), \rho_P^S)$ is a Polish space. By [41, Theorem 3.4.2], Lipschitz maps preserve the Prokhorov metric in the sense that $\rho_P^S(g(X), g(Y)) \leq (1 \vee K)\rho_P^S(X, Y)$ for any K -Lipschitz map g . We will need a slight generalization of this result, which allows g to depend on an additional random variable that is independent of the Lipschitz component.

THEOREM 3.7. Let $(S, d_S), (S', d_{S'}), (R, d_R)$ be Polish spaces, and suppose that $g: S' \times S \rightarrow R$ is K -Lipschitz in the second variable, i.e.,

$$(26) \quad d_R(g(x, y), g(x, z)) \leq K d_S(y, z), \quad x \in S', y, z \in S.$$

Suppose furthermore that the random variables X, Y, Z satisfy $X \perp Y, X \perp Z$. Then

$$(27) \quad \rho_P^R(g(X, Y), g(X, Z)) \leq (1 \vee K) \rho_P^S(Y, Z).$$

Proof. We may assume w.l.o.g. that $K \geq 1$. By the Strassen–Dudley representation theorem [13, Corollary 11.6.4], there is a probability space $(\tilde{\Omega}, \tilde{\mathbb{P}})$ with random variables \tilde{Y}, \tilde{Z} satisfying $\text{law}(\tilde{Y}) = \text{law}(Y), \text{law}(\tilde{Z}) = \text{law}(Z)$ such that $\rho_P^S(Y, Z) = \rho_K(\tilde{Y}, \tilde{Z})$, where ρ_K is the Ky Fan metric, i.e.,

$$\rho_K(\tilde{Y}, \tilde{Z}) = \inf\{\epsilon > 0: \tilde{\mathbb{P}}(d_S(\tilde{Y}, \tilde{Z}) > \epsilon) \leq \epsilon\}.$$

Now set $\bar{\Omega} = \Omega \times \tilde{\Omega}$ and $\bar{\mathbb{P}} = \mathbb{P} \otimes \tilde{\mathbb{P}}$; then $\text{law}((X, \tilde{Y})) = \text{law}((X, Y))$ and $\text{law}((X, \tilde{Z})) = \text{law}((X, Z))$ by the construction of $(\bar{\Omega}, \bar{\mathbb{P}})$ and the independence assumption. In particular, we have $\rho_P^R(g(X, Y), g(X, Z)) = \rho_P^R(g(X, \tilde{Y}), g(X, \tilde{Z}))$. Noting the general fact that $\rho_P \leq \rho_K$, we consider $\epsilon > 0$ such that $\tilde{\mathbb{P}}(d_S(\tilde{Y}, \tilde{Z}) > \epsilon/K) \leq \epsilon/K$. Then

$$\bar{\mathbb{P}}(d_R(g(X, \tilde{Y}), g(X, \tilde{Z})) > \epsilon) \leq \bar{\mathbb{P}}(d_S(\tilde{Y}, \tilde{Z}) > \epsilon/K) \leq \epsilon/K \leq \epsilon.$$

It follows that

$$\rho_K(g(X, \tilde{Y}), g(X, \tilde{Z})) \leq \inf\{\epsilon > 0: \tilde{\mathbb{P}}(d_S(\tilde{Y}, \tilde{Z}) > \epsilon/K) \leq \epsilon/K\} = K \rho_K(\tilde{Y}, \tilde{Z}).$$

We have obtained that

$$\rho_P^R(g(X, Y), g(X, Z)) \leq \rho_K(g(X, \tilde{Y}), g(X, \tilde{Z})) \leq K \rho_K(\tilde{Y}, \tilde{Z}) = K \rho_P^S(Y, Z). \quad \square$$

Example 3.8. We illustrate with an example that Theorem 3.7 does not hold in general without the independence assumption. We consider $S = S' = R = \mathbb{R}$, and we set $\text{law}((X, Y)) = \frac{1}{2}\delta_{(0,0)} + \frac{1}{2}\delta_{(1,1)}$ and $\text{law}((X, Z)) = \frac{1}{2}\delta_{(0,1)} + \frac{1}{2}\delta_{(1,0)}$. Then $\text{law}(X) = \text{law}(Y) = \text{law}(Z) = \frac{1}{2}\delta_0 + \frac{1}{2}\delta_1$, so in particular $\rho_P^S(Y, Z) = 0$. However, taking $g(x, y) := |x - y|$, we see that $\text{law}(g(X, Y)) = \delta_0$ and $\text{law}(g(X, Z)) = \delta_1$, so $\rho_P^R(g(X, Y), g(X, Z)) = 1$.

COROLLARY 3.9. Suppose $g: S' \times S \rightarrow \mathbb{R}$ is K -Lipschitz in the second variable. Suppose that $X \perp Y, X \perp Z$ and that the cdf of $g(X, Y)$ is Lipschitz with constant L . Then, for all $x \in \mathbb{R}$, we have

$$(28) \quad |\mathbb{P}(g(X, Y) \leq x) - \mathbb{P}(g(X, Z) \leq x)| \leq (1 + L)(1 \vee K) \rho_P^R(Y, Z).$$

Proof. Denote the cdf of $g(X, Y)$ by F and that of $g(X, Z)$ by G . It holds that

$$\begin{aligned} \sup_{x \in \mathbb{R}} |G(x) - F(x)| &\leq (1 + L) \inf\{\epsilon > 0: F(x - \epsilon) - \epsilon \leq G(x) \leq F(x + \epsilon) + \epsilon \forall x \in \mathbb{R}\} \\ &\leq (1 + L) \rho_P^R(g(X, Y), g(X, Z)). \end{aligned}$$

The claim now follows from Theorem 3.7. \square

We state two technical lemmata for future reference.

LEMMA 3.10. *Let x^1, x^2 be càdlàg paths. Then*

$$\left| \inf_{0 \leq s \leq t} x_s^1 - \inf_{0 \leq s \leq t} x_s^2 \right| \leq \sup_{0 \leq s \leq t} |x_s^1 - x_s^2|.$$

Proof. Let $\epsilon > 0$, and choose $r \in [0, t]$ such that $x_r^2 \leq \inf_{0 \leq s \leq t} x_s^2 + \epsilon$. Then

$$\inf_{0 \leq s \leq t} x_s^1 - \inf_{0 \leq s \leq t} x_s^2 \leq x_r^1 - x_r^2 + \epsilon \leq \sup_{0 \leq s \leq t} |x_s^1 - x_s^2| + \epsilon.$$

Reversing the roles and noting that $\epsilon > 0$ was arbitrary proves the claim. \square

LEMMA 3.11. *Let X_{0-} admit a bounded density, denoted by f , and let Y^1, Y^2 be càdlàg and such that X_{0-} is independent of (Y^1, Y^2) . Then it holds that*

$$\left| \mathbb{P} \left(\inf_{0 \leq s \leq t} X_{0-} + Y_s^1 \leq 0 \right) - \mathbb{P} \left(\inf_{0 \leq s \leq t} X_{0-} + Y_s^2 \leq 0 \right) \right| \leq \|f\|_\infty \mathbb{E} \left[\sup_{0 \leq s \leq t} |Y_s^1 - Y_s^2| \right].$$

Proof. Set $I := [-\inf_{s \leq t} Y_s^1, -\inf_{s \leq t} Y_s^2] \cup [-\inf_{s \leq t} Y_s^2, -\inf_{s \leq t} Y_s^1]$, and note that by Lemma 3.10, the length of I is bounded by $\sup_{0 \leq s \leq t} |Y_s^1 - Y_s^2|$. We argue as in the proof of Lemma 2.1 in [29], using the assumption on X_{0-} and the general inequality $\mathbb{P}(A) - \mathbb{P}(B) \leq \mathbb{P}(A \setminus B)$:

$$\begin{aligned} & \left| \mathbb{P} \left(\inf_{0 \leq s \leq t} X_{0-} + Y_s^1 \leq 0 \right) - \mathbb{P} \left(\inf_{0 \leq s \leq t} X_{0-} + Y_s^2 \leq 0 \right) \right| \\ & \leq \mathbb{P}(X_{0-} \in I) \\ & = \mathbb{E} \left[\int_0^\infty \mathbf{1}_I(x) f(x) dx \right] \\ & \leq \|f\|_\infty \mathbb{E} \left[\sup_{0 \leq s \leq t} |Y_s^1 - Y_s^2| \right]. \end{aligned} \quad \square$$

Proof of Lemma 3.5.

Step 1. We transfer the convergence rate on $[0, 1]$ to $[0, T]$ for $T > 0$. Set

$$(29) \quad S := \{x \in D[0, 1] : \text{Disc}(x) \subseteq h\mathbb{Q}\}, \quad R := \left\{x \in D[0, T] : \text{Disc}(x) \subseteq \frac{h}{T}\mathbb{Q}\right\}.$$

In words, S is the set of càdlàg paths on $[0, 1]$ that can only jump at times that are rational multiples of h . Note that if we endow the sets S and R with the uniform metric, they are separable and hence Polish spaces. By Theorem 1.16 in [6],² we have $\rho_P^S(B^{h'}, W) \leq C\sqrt{h'} \log(1/h')$ for sufficiently small $h' > 0$. Define $f : S \rightarrow R$ by $f(x) = \sqrt{T}x_{./T}$; then f is \sqrt{T} -Lipschitz. Theorem 3.7 and the self-similarity of the Donsker approximation and Brownian motion give

$$\begin{aligned} \rho_P^R(B^h, W) &= \rho_P^R(\sqrt{T}B_{./T}^{h/T}, \sqrt{T}W_{./T}) = \rho_P^R(f(B^{h/T}), f(W)) \\ &\leq C \frac{(1 \vee \sqrt{T})}{\sqrt{T}} \sqrt{h} \log(T/h) \end{aligned}$$

for $h > 0$ sufficiently small, which concludes Step 1.

Step 2. Fix $\ell \in M$, and define the map $g_t : \mathbb{R} \times R \rightarrow \mathbb{R}$ by $g_t(x_0, x) := \inf_{0 \leq s \leq t} x_0 + x_s - \alpha \ell_s$. Note that $g(x_0, \cdot)$ is 1-Lipschitz for every $x_0 \in \mathbb{R}$ by Lemma 3.10. Denote

²Note that in [6], the result is stated to hold for all $h > 0$, but the proof reveals that it only holds for sufficiently small $h > 0$.

the cdf of $g(X_{0-}, B)$ by F . We check that F is Lipschitz continuous: for $x, y \in \mathbb{R}$, setting $Y^1 := B - \alpha\ell - x$ and $Y^2 := B - \alpha\ell - y$ in Lemma 3.11, we obtain

$$(30) \quad |F(x) - F(y)| \leq \|f\|_\infty |x - y|,$$

where f denotes the bounded density of X_{0-} . By Corollary 3.9, we therefore obtain for $t \in [0, T]$ and $h > 0$ small enough

$$|\mathbb{P}(g_t(X_{0-}, B^h) \leq 0) - \mathbb{P}(g_t(X_{0-}, B) \leq 0)| \leq C(1 + \|f\|_\infty) \frac{(1 \vee \sqrt{T})}{\sqrt{T}} \sqrt{h} \log(T/h).$$

Since with $X_{0-}^h := \left\lfloor \frac{X_{0-}}{\sqrt{h}} \right\rfloor \sqrt{h}$ it holds that $|X_{0-} - X_{0-}^h| \leq \sqrt{h}$ almost surely, setting $Y^1 := -\sqrt{h} + B^h - \alpha\ell$ and $Y^2 := B^h - \alpha\ell$ in Lemma 3.11 we find

$$\begin{aligned} & \mathbb{P}(g_t(X_{0-}^h, B^h) \leq 0) - \mathbb{P}(g_t(X_{0-}, B^h) \leq 0) \\ & \leq \mathbb{P}(g_t(X_{0-} - \sqrt{h}, B^h) \leq 0) - \mathbb{P}(g_t(X_{0-}, B^h) \leq 0) \\ & \leq \|f\|_\infty \sqrt{h}, \end{aligned}$$

and arguing analogously for the downwards estimate we find

$$(31) \quad |\mathbb{P}(g_t(X_{0-}^h, B^h) \leq 0) - \mathbb{P}(g_t(X_{0-}, B^h) \leq 0)| \leq \|f\|_\infty \sqrt{h}.$$

Taking the supremum over $t \in [0, T]$ and absorbing some terms into the constant C , we obtain

$$\sup_{t \in [0, T]} |\mathbb{P}(g_t(X_{0-}^h, B^h) \leq 0) - \mathbb{P}(g_t(X_{0-}, B) \leq 0)| \leq C\sqrt{h} \log(T/h).$$

Noting that $\Gamma_h[\ell]_t = \mathcal{D}_h[\mathbb{P}(g_t(X_{0-}^h, B^h) \leq 0)]_t$ and $\Gamma[\ell]_t = \mathbb{P}(g_t(X_{0-}, B) \leq 0)$, we obtain the claim by point (a) of Lemma 3.3 with $x = [\mathbb{P}(g_t(X_{0-}^h, B^h) \leq 0)]$ and $y = \Gamma[\ell]$. \square

LEMMA 3.12. Let $\underline{\Lambda}^h$ be the minimal solution to (20) with the initial condition given by (23), and suppose that Assumptions 3.4 are satisfied. Then, for any $t > 0$, it holds that

$$(32) \quad \limsup_{h \rightarrow 0} \underline{\Lambda}_t^h \leq \underline{\Lambda}_t.$$

Proof. Define the sequence of optimal values

$$(33) \quad V_t^h = \inf_{\ell \in M} c_h(\ell), \quad c_h(\ell) = \alpha\ell_t + \frac{\alpha}{\sqrt{h} \lfloor \log(h)^2 \rfloor} \sup_{s \in [0, t]} |\Gamma_h[\ell]_s - \mathcal{D}_h[\ell]_s|,$$

where Γ_h is defined as in Lemma 3.1.

Step 1. We show that V_t^h is asymptotically dominated by $\underline{\Lambda}$ as $h \rightarrow 0$. By Lemma 3.5, for small enough $h > 0$ we have

$$(34) \quad V_t^h \leq c_h(\alpha^{-1} \underline{\Lambda}) = \underline{\Lambda}_t + \frac{\alpha}{\sqrt{h} \lfloor \log(h)^2 \rfloor} \sup_{s \in [0, t]} |\Gamma_h[\underline{\Lambda}]_s - \mathcal{D}_h[\underline{\Lambda}]_s|$$

$$(35) \quad \leq \underline{\Lambda}_t + \alpha C \frac{\log(t/h)}{\lfloor \log(h)^2 \rfloor}$$

and therefore $\limsup_{h \rightarrow 0} V_t^h \leq \underline{\Lambda}_t$.

Step 2. We show that V_t^h asymptotically dominates $\underline{\Lambda}_t^h$. Choose a sequence $(L^h)_{h \geq 0}$ with $L^h \in M$ such that $c_h(L^h) - V_t^h \rightarrow 0$ as $h \rightarrow 0$. Suppose that $\sup_{s \in [0, t]} |\Gamma_h[L^h]_s - \mathcal{D}_h[L^h]_s| > h^{1/2} \lfloor \log(h)^2 \rfloor$ for some h , then we have $c_h(L^h) > c_h(\alpha^{-1} \tilde{\Lambda}^h)$, where $\tilde{\Lambda}^h$ is a solution to (20) with $X_{0-}^h := \left\lfloor \frac{X_{0-}}{\sqrt{h}} \right\rfloor \sqrt{h}$. We may therefore assume without loss of generality that $\sup_{s \in [0, t]} |\Gamma_h[L^h]_s - \mathcal{D}_h[L^h]_s| \leq h^{1/2} \lfloor \log(h)^2 \rfloor$ for all $h > 0$. For the sake of brevity, we write $d_h := h^{1/2} \lfloor \log(h)^2 \rfloor$ in the following. Using the fact that $\mathcal{D}_h[\ell] \leq \ell$ for $\ell \in M$ and the monotonicity of Γ_h , we find

$$(36) \quad \mathcal{D}_h[L^h]_s \geq \Gamma_h[L^h]_s - d_h \geq \Gamma_h[\mathcal{D}_h[L^h]]_s - d_h, \quad s \in [0, t].$$

For $\ell \in M$, define the operator

$$(37) \quad \tilde{\Gamma}_h[\ell]_t := \mathcal{D}_h \left[\mathbb{P} \left(\inf_{0 \leq s \leq \cdot} \left[\left(\left\lfloor \frac{X_{0-}}{\sqrt{h}} \right\rfloor + \lfloor \log(h)^2 \rfloor \right) \sqrt{h} + B_s^h - \alpha \ell_s \right] \leq 0 \right) \right]_t.$$

Inequality (36), the monotonicity of Γ_h and \mathcal{D}_h , and the fact that $\mathcal{D}_h[d_h] = d_h$ imply

$$(38) \quad \mathcal{D}_h[L^h]_s \geq \Gamma_h[-d_h]_s - d_h = \tilde{\Gamma}_h[0]_s - d_h, \quad s \in [0, t].$$

Substituting this again into inequality (36) yields

$$\mathcal{D}_h[L^h]_s \geq \Gamma_h[\tilde{\Gamma}_h[0] - d_h]_s - d_h = \tilde{\Gamma}_h^{(2)}[0]_s - d_h, \quad s \in [0, t].$$

Proceeding inductively, we obtain $\mathcal{D}_h[L^h]_s \geq \tilde{\Gamma}_h^{(k)}[0]_s - d_h$ for every $s \in [0, t]$, $\mathcal{D}_h[L^h]_s \geq \tilde{\Gamma}_h^{(k)}[0]_s - d_h$ for every $s \in [0, t]$ and $k \in \mathbb{N}$, and taking the limit as $k \rightarrow \infty$, by Lemma 3.1 we obtain $\alpha L_s^h \geq \alpha \mathcal{D}_h[L^h]_s \geq \underline{\Lambda}_s^h - \alpha d_h$ for $s \in [0, t]$. Since $d_h \rightarrow 0$ as $h \rightarrow 0$, combined with the definition of c_h and Step 1, this yields

$$\limsup_{h \rightarrow 0} \underline{\Lambda}_t^h \leq \limsup_{h \rightarrow 0} \alpha L_t^h \leq \limsup_{h \rightarrow 0} c_h(L^h) = \limsup_{h \rightarrow 0} V_t^h \leq \underline{\Lambda}_t. \quad \square$$

THEOREM 3.13. *Suppose that Assumptions 3.4 are satisfied. Then the sequence of (perturbed) minimal solutions to the Donsker problem converges to the minimal solution to the McKean–Vlasov problem; i.e., it holds that*

$$(39) \quad \lim_{h \rightarrow 0} \underline{\Lambda}_t^h = \underline{\Lambda}_t$$

for every $t \geq 0$ that is a continuity point of $\underline{\Lambda}$.

Proof. Choose a sequence $(h_n)_{n \geq 1}$ such that $h_n > 0$ and $\lim_{n \rightarrow \infty} h_n = 0$. By the compactness of M , after passing to a subsequence if necessary, we may assume that $\underline{\Lambda}^{h_n} \rightarrow \Lambda$ for some Λ with $\alpha^{-1} \Lambda \in M$. If x is a càdlàg path, we introduce the notation $\Gamma^\alpha[x] := \alpha \Gamma[\alpha^{-1}x]$, such that Λ solves (7) iff $\Gamma^\alpha[\Lambda] = \Lambda$. We use the notation Γ_h^α in an analogous way. We show that Λ solves the McKean–Vlasov problem. By Proposition 2.2 and Lemma 3.3, for $t \notin \text{Disc}(\Lambda) \cup \text{Disc}(\Gamma[\Lambda])$ we have

$$\begin{aligned} |\Gamma^\alpha[\Lambda]_t - \Lambda_t| &= \lim_{n \rightarrow \infty} \left| \mathcal{D}_{h_n} \left[\Gamma^\alpha[\underline{\Lambda}^{h_n}] \right]_t - \underline{\Lambda}_t^{h_n} \right| \\ &\leq \limsup_{n \rightarrow \infty} \left[\left| \mathcal{D}_{h_n} \left[\Gamma^\alpha[\underline{\Lambda}^{h_n}] \right]_t - \Gamma_{h_n}^\alpha[\underline{\Lambda}^{h_n}]_t \right| + \left| \Gamma_{h_n}^\alpha[\underline{\Lambda}^{h_n}]_t - \Gamma^\alpha[\underline{\Lambda}^{h_n}]_t \right| \right], \end{aligned}$$

where $\tilde{\Gamma}_{h_n}^\alpha[x]$ is defined as $\alpha \tilde{\Gamma}_{h_n}[\alpha^{-1}x]$ with $\tilde{\Gamma}_h$ defined as in (37) and we used that $\underline{\Lambda}^{h_n} = \tilde{\Gamma}^\alpha[\underline{\Lambda}^{h_n}]$ since $\underline{\Lambda}^{h_n}$ is a solution to the perturbed Donsker problem. Due to

Lemma 3.5, the first term above vanishes as $n \rightarrow \infty$. For the second term, setting $Y^1 := \sqrt{h_n}(1 + \lfloor \log(h_n)^2 \rfloor) + B^h - \underline{\Lambda}^{h_n}$ and $Y^2 := -\sqrt{h_n} + B^h - \underline{\Lambda}^{h_n}$, we have for any $r \geq 0$

$$\begin{aligned} \alpha \mathcal{D}_{h_n} \left[\mathbb{P} \left(X_{0-} + \inf_{s \leq \cdot} Y_s^1 \leq 0 \right) \right]_r &\leq \Gamma^\alpha[\underline{\Lambda}^{h_n}]_r \leq \Gamma_{h_n}^\alpha[\underline{\Lambda}^{h_n}]_r \\ &\leq \alpha \mathcal{D}_{h_n} \left[\mathbb{P} \left(X_{0-} + \inf_{s \leq \cdot} Y_s^2 \leq 0 \right) \right]_r, \end{aligned}$$

so by Lemma 3.11 and Lemma 3.3.a, it follows that

$$|\Gamma_{h_n}^\alpha[\underline{\Lambda}^{h_n}]_t - \Gamma^\alpha[\underline{\Lambda}^{h_n}]_t| \leq \alpha \|f\|_\infty \sqrt{h_n}(2 + \lfloor \log(h_n)^2 \rfloor) + 2\sqrt{h_n}.$$

Therefore, $\Gamma^\alpha[\Lambda]_t = \Lambda_t$. By the right continuity of Λ and $\Gamma^\alpha[\Lambda]$, we find $\Lambda = \Gamma^\alpha[\Lambda]$. By definition, this implies $\Lambda \geq \underline{\Lambda}$. On the other hand, for $t \notin \text{Disc}(\Lambda)$, Lemma 3.12 yields

$$\Lambda_t = \limsup_{n \rightarrow \infty} \underline{\Lambda}_t^{h_n} \leq \underline{\Lambda}_t.$$

By the right continuity, we find $\Lambda \leq \underline{\Lambda}$ and hence $\Lambda = \underline{\Lambda}$. \square

In the case $\alpha \|f\|_\infty < 1$, we will also prove a rate of convergence. By [29, Theorem 2.2 and the comment below], this assumption corresponds to the weak feedback regime, where uniqueness and, in particular, the continuity of solutions hold true. To obtain the convergence rate, we first prove that the solution is globally Hölder continuous in this case.

PROPOSITION 3.14. *Suppose that X_{0-} has a bounded density f such that $\alpha \|f\|_\infty < 1$. Then, the unique solution Λ to (7) is $\frac{1}{2}$ -Hölder continuous with constant $\sqrt{\frac{2}{\pi} \frac{\alpha \|f\|_\infty}{1 - \alpha \|f\|_\infty}}$.*

Proof. Applying Lemma 3.11 with $Y^1 = B - \alpha \ell^1$ and $Y^2 = B - \alpha \ell^2$ shows

$$(40) \quad \sup_{t \in [0, T]} |\Gamma[\ell^1]_t - \Gamma[\ell^2]_t| \leq \alpha \|f\|_\infty \sup_{t \in [0, T]} |\ell_t^1 - \ell_t^2|.$$

Fix $t, h \geq 0$. Define the extensions

$$\hat{\Lambda}_s := \begin{cases} \Lambda_s, & s \in [0, t] \\ \Lambda_t, & s \in [t, t+h] \end{cases} \quad \hat{B}_s := \begin{cases} B_s, & s \in [0, t] \\ B_t, & s \in [t, t+h]. \end{cases}$$

Noting that $\sup_{s \in [0, t+h]} |\Lambda_s - \hat{\Lambda}_s| = \Lambda_{t+h} - \Lambda_t$, and using (40) it follows that

$$\begin{aligned} \Lambda_{t+h} - \Lambda_t &= \Gamma^\alpha[\Lambda]_{t+h} - \Gamma^\alpha[\Lambda]_t = \Gamma^\alpha[\Lambda]_{t+h} - \Gamma^\alpha[\Lambda]_{t+h} - \Gamma^\alpha[\hat{\Lambda}]_{t+h} + \Gamma^\alpha[\hat{\Lambda}]_{t+h} - \Gamma^\alpha[\Lambda]_t \\ &\leq \alpha \|f\|_\infty (\Lambda_{t+h} - \Lambda_t) + \Gamma^\alpha[\hat{\Lambda}]_{t+h} - \Gamma^\alpha[\Lambda]_t. \end{aligned}$$

We may write the second term as

$$\begin{aligned} \Gamma^\alpha[\hat{\Lambda}]_{t+h} - \Gamma^\alpha[\Lambda]_t &= \alpha \mathbb{P} \left(\inf_{0 \leq s \leq t+h} X_{0-} + B_s - \hat{\Lambda}_s \leq 0 \right) \\ &\quad - \alpha \mathbb{P} \left(\inf_{0 \leq s \leq t+h} X_{0-} + \hat{B}_s - \hat{\Lambda}_s \leq 0 \right). \end{aligned}$$

Arguing as in the proof of Lemma 3.11, we find

$$\Gamma^\alpha[\hat{\Lambda}]_{t+h} - \Gamma^\alpha[\Lambda]_t \leq \alpha \|f\|_\infty \mathbb{E} \left[\sup_{t \leq s \leq t+h} B_s - B_t \right] \leq \sqrt{\frac{2}{\pi}} \alpha \|f\|_\infty \sqrt{h}.$$

Rearranging, we obtain

$$\Lambda_{t+h} - \Lambda_t \leq \sqrt{\frac{2}{\pi}} \frac{\alpha \|f\|_\infty}{1 - \alpha \|f\|_\infty} \sqrt{h}. \quad \square$$

THEOREM 3.15. *Suppose that Assumptions 3.4 are satisfied and that additionally $\alpha \|f\|_\infty < 1$. Then there is a constant C depending on T, α , and $\|f\|_\infty$ such that*

$$\sup_{t \in [0, T]} |\underline{\Lambda}_t - \underline{\Lambda}_t^h| \leq C \log(h)^2 \sqrt{h}$$

for $h > 0$ sufficiently small.

Proof. Recalling the notation of the proof of Theorem 3.13, we have $\underline{\Lambda}^h = \tilde{\Gamma}_h^\alpha[\underline{\Lambda}^h]$, since $\underline{\Lambda}^h$ is a solution to the perturbed Donsker problem and $\underline{\Lambda}^h$ is a solution to the perturbed

$$\begin{aligned} \sup_{t \in [0, T]} |\underline{\Lambda}_t - \underline{\Lambda}_t^h| &= \sup_{t \in [0, T]} |\Gamma^\alpha[\underline{\Lambda}]_t - \tilde{\Gamma}_h^\alpha[\underline{\Lambda}^h]_t| \\ &\leq \sup_{t \in [0, T]} |\Gamma^\alpha[\underline{\Lambda}]_t - \mathcal{D}_h[\Gamma^\alpha[\underline{\Lambda}]]_t| + \sup_{t \in [0, T]} |\mathcal{D}_h[\Gamma^\alpha[\underline{\Lambda}]] - \mathcal{D}_h[\Gamma^\alpha[\underline{\Lambda}^h]]_t| \\ &\quad + \sup_{t \in [0, T]} |\mathcal{D}_h[\Gamma^\alpha[\underline{\Lambda}^h]]_t - \Gamma^\alpha[\underline{\Lambda}^h]_t| + \sup_{t \in [0, T]} |\Gamma^\alpha[\underline{\Lambda}^h]_t - \tilde{\Gamma}_h^\alpha[\underline{\Lambda}^h]_t| \end{aligned}$$

The third and fourth terms can be estimated as in the proof of Theorem 3.13, yielding

$$\begin{aligned} &\sup_{t \in [0, T]} |\mathcal{D}_h[\Gamma^\alpha[\underline{\Lambda}^h]]_t - \Gamma^\alpha[\underline{\Lambda}^h]_t| + \sup_{t \in [0, T]} |\Gamma^\alpha[\underline{\Lambda}^h]_t - \tilde{\Gamma}_h^\alpha[\underline{\Lambda}^h]_t| \\ &\leq \alpha C \sqrt{h} \log(T/h) + \alpha \|f\|_\infty \sqrt{h} (2 + \lfloor \log(h)^2 \rfloor). \end{aligned}$$

We apply Proposition 3.14 to estimate the first term:

$$\begin{aligned} |\Gamma^\alpha[\underline{\Lambda}]_t - \mathcal{D}_h[\Gamma^\alpha[\underline{\Lambda}]]_t| &= |\underline{\Lambda}_t - \mathcal{D}_h[\underline{\Lambda}]_t| \leq |\Lambda_t - \Lambda_{\lfloor \frac{t}{h} \rfloor h}| + \sqrt{h} \\ &\leq \sqrt{\frac{2}{\pi}} \frac{\alpha \|f\|_\infty}{1 - \alpha \|f\|_\infty} \sqrt{h} + \sqrt{h}. \end{aligned}$$

For the second term, we employ (40) together with Lemma 3.3(a) to find

$$\sup_{t \in [0, T]} |\mathcal{D}_h[\Gamma^\alpha[\underline{\Lambda}]]_t - \mathcal{D}_h[\Gamma^\alpha[\underline{\Lambda}^h]]_t| \leq \alpha \|f\|_\infty \sup_{t \in [0, T]} |\underline{\Lambda}_t - \underline{\Lambda}_t^h| + 2\sqrt{h}.$$

Rearranging terms and absorbing some into the constant C , the claim follows. \square

Remark 3.16. Note that the choice $\sqrt{h} \lfloor \log(h)^2 \rfloor$ for the perturbation of the initial condition was somewhat arbitrary since the proofs of the previous theorems work for any function $\delta(h)$ that converges to zero slowly enough such that

$$\lim_{h \rightarrow 0} \frac{\sqrt{h} \log(T/h)}{\delta(h)} = 0$$

for every $T > 0$. In particular, we could consider $\sqrt{h} \lfloor \epsilon \log(h)^2 \rfloor$ for any $\epsilon > 0$, which suggests that the perturbation can be ignored in practice, and we will indeed ignore it in the numerical tests in the next section. It seems natural to conjecture that the result should also hold with $\epsilon = 0$, and this would indeed follow from Conjecture 6.10 in [8]; however, we do not have a proof at this point.

4. Implementation. In this section, we first describe a particle approximation of the implicit and explicit time-stepping schemes from section 2 and then our implementation of the Donsker scheme from section 3.

4.1. Particle approximation for the explicit and implicit time stepping schemes. Here, we briefly discuss the implementation of the time-stepping scheme from [26] and the new scheme specified by (13) and Definition 2.5.

The precursor work [26] considers the time-stepping scheme (15) and, for a fixed number $\mathbb{N} \ni n \geq 1$, a particle approximation

$$\begin{aligned}\tilde{X}_{k\Delta}^{\Delta,n,(m)} &= X_{0-} + Z_{k\Delta}^{n,\Delta,(m)} - \tilde{\Lambda}_{k\Delta}^{\Delta,n}, & k \geq 0, 1 \leq m \leq n, \\ \tilde{\Lambda}_{k\Delta}^{\Delta,n} &= \frac{\alpha}{n} \sum_{m=1}^n \mathbb{1}_{\{\min_{0 \leq i < k} X_{i\Delta}^{\Delta,n,(m)} \leq 0\}}, & k > 0, \quad \tilde{\Lambda}_0^{\Delta,n} = 0.\end{aligned}$$

As the estimator for $\tilde{\Lambda}_{k\Delta}^{\Delta,n}$ can be written explicitly in terms of the simulated paths of X up to time $(k-1)\Delta$, we refer to it here for brevity as the explicit particle scheme (see also Remark 2.6).

For the implicit scheme, we define a corresponding approximation to (14) by $\Lambda_{k\Delta}^{\Delta,n}$, the smallest solution to

$$\begin{aligned}X_{k\Delta}^{\Delta,n,(m)} &= X_{0-} + Z_{k\Delta}^{n,\Delta,(m)} - \Lambda_{k\Delta}^{\Delta,n}, & k \geq 0, 1 \leq m \leq n, \\ \Lambda_{k\Delta}^{\Delta,n} &= \frac{\alpha}{n} \sum_{m=1}^n \mathbb{1}_{\{(\min_{0 \leq i < k} X_{i\Delta}^{\Delta,n,(m)}) \wedge (X_{0-} + Z_{k\Delta}^{n,\Delta,(m)} - \Lambda_{k\Delta}^{\Delta,n}) \leq 0\}}, & k \geq 0,\end{aligned}$$

where we interpret the minimum over an empty set as ∞ . Here, Λ^Δ is implicitly defined and can be found iteratively:

$$\begin{aligned}\Lambda_{k\Delta}^{\Delta,n,(0)} &= \frac{\alpha}{n} \sum_{m=1}^n \mathbb{1}_{\{\min_{0 \leq i < k} X_{i\Delta}^{\Delta,n,(m)} \leq 0\}}, & k \geq 0, \\ \Lambda_{k\Delta}^{\Delta,n,(j)} &= \frac{\alpha}{n} \sum_{m=1}^n \mathbb{1}_{\{(\min_{0 \leq i < k} X_{i\Delta}^{\Delta,n,(m)}) \wedge (X_{0-} + Z_{k\Delta}^{n,\Delta,(m)} - \Lambda_{k\Delta}^{\Delta,n,(j-1)}) \leq 0\}}, & k \geq 0, j > 0.\end{aligned}$$

It is clear that $\Lambda_{k\Delta}^{\Delta,n,(j)}$ is increasing in j and that it terminates in finitely many iterations in a fixed point, which has to be the minimal solution.

4.2. A tree-type scheme for the Donsker approximation. The Donsker approximation (21) used in (20) does not specify the increments Y_k but only imposes conditions on its first and second moments. In particular, this construction accommodates $Y_k \sim N(0, 1)$ i.i.d. and therefore the time-stepping scheme (14) as a special case. If, however, Y_k are chosen as i.i.d. discrete random variables, the support of $X_{k\Delta}^\Delta$ is also discrete and its distribution can be expressed in terms of the probability mass function of Y_1 . Hence, no estimation by particle simulation is necessary in this case. Most generally, the cardinality of the support and therefore the computational complexity grow exponentially in the number of time steps but with the following choice only linearly.

We specify Y_k to be Rademacher random variables, i.e., $\mathbb{P}(Y_k = -1) = \mathbb{P}(Y_k = 1) = 1/2$. We introduce a time mesh $t_k = kh$, $k \geq 0$, an integer, and

$$u_i^k = \mathbb{P}(X_{t_k}^h = i\sqrt{h} - \underline{\Lambda}_{t_k}^h, \tau^h > t_k)$$

for $i \in \mathbb{Z}$, where

$$\underline{\Lambda}_{t_k}^h = \alpha \left(1 - \sum_{i=i_k+1}^{\infty} u_i^k \right), \quad i_k = \left\lfloor \frac{\underline{\Lambda}_{t_k}^h}{\sqrt{h}} \right\rfloor,$$

similar³ to the notation defined in Section 3. We have that

$$(41) \quad u_i^k = \begin{cases} \frac{1}{2}u_{i-1}^{k-1} + \frac{1}{2}u_{i+1}^{k-1}, & i > i_k + 1, \\ \frac{1}{2}u_{i+1}^{k-1}, & i = i_k + 1, \\ 0, & i < i_k + 1, \end{cases}$$

for $k > 0$. This is an implicit scheme, as u^k and $\underline{\Lambda}^h$ are implicitly coupled. The recurrence relation (41) resembles a binomial tree, shifted by the interaction term, and can be rearranged into the finite difference scheme (2) for (1). It can also be interpreted as a special type of semi-Lagrangian scheme (see [5, 14]) for the forward equation (4).

Note that (41) is a nonlinear equation through the dependence of i_k on u^k via $\underline{\Lambda}_{t_k}^h$. We propose in the following an iteration similar to that in subsection 4.1 for the implicit particle scheme.

4.3. Iterative solution of the discretized equations. Lemma 3.1 suggests a fixed-point iteration to solve simultaneously for $\underline{\Lambda}_{t_k}^h$ and the vector u^k for each k . We assume that we know the cdf of X_{0-} exactly, and therefore we can calculate

$$\mathbb{P}(X_{0-}^h = i\sqrt{h}) = \mathbb{P}(X_{0-} \in [i\sqrt{h}, (i+1)\sqrt{h}))$$

for all $i \in \mathbb{Z}$. To calculate u_i^0 for $i \in \mathbb{Z}$, we need to determine $\underline{\Lambda}_0^h$ first, which we obtain as in Lemma 3.1 through the iteration initialized with $\lambda^0 = 0$ and

$$\lambda^{n+1} = \alpha \sum_{j=0}^{\iota^n} \mathbb{P}(X_{0-}^h = j\sqrt{h}), \quad \iota^n = \left\lfloor \frac{\lambda^n}{\sqrt{h}} \right\rfloor.$$

The iteration terminates when $\iota^{n+1} = \iota^n$, which happens after at most $\lfloor \alpha/\sqrt{h} \rfloor$ iterations, since $\lambda^n \leq \lambda^{n+1}$ and hence $\iota^n \leq \iota^{n+1}$ as well. This yields $\underline{\Lambda}_0^h$. We then calculate u_i^0 via

$$u_i^0 = 0 \text{ for } i \leq \left\lfloor \frac{\underline{\Lambda}_0^h}{\sqrt{h}} \right\rfloor \quad \text{and} \quad u_i^0 = \mathbb{P}(X_{0-}^h = i\sqrt{h}) \text{ for } i > \left\lfloor \frac{\underline{\Lambda}_0^h}{\sqrt{h}} \right\rfloor.$$

The vectors $u^k = (u_i^k)_i$ are then calculated recursively through (41), using a local in time version of the iteration in Lemma 3.1 that iterates only over the scalar loss at each time point. Set $\lambda^0 = \underline{\Lambda}_{t_{k-1}}^h$ and $\iota^0 = \lfloor \lambda^0/\sqrt{h} \rfloor$, and for $n \geq 0$,

$$(42) \quad u_i^{k,n+1} = \begin{cases} \frac{1}{2}u_{i-1}^{k-1} + \frac{1}{2}u_{i+1}^{k-1}, & i > \iota^n + 1, \\ \frac{1}{2}u_{i+1}^{k-1}, & i = \iota^n + 1, \\ 0, & i < \iota^n + 1, \end{cases}$$

$$\lambda^{n+1} = \alpha \left(1 - \sum_{i=\iota^n+1}^{\infty} u_i^{k,n+1} \right), \quad \iota^{n+1} = \left\lfloor \frac{\lambda^{n+1}}{\sqrt{h}} \right\rfloor.$$

³Note that we have α inside $\lfloor \cdot \rfloor$, in contrast to (19) and (20), such that the mesh size here is directly \sqrt{h} rather than $\alpha\sqrt{h}$. The theoretical properties are unchanged.

Expressing this in terms of Γ_h , if we define

$$(43) \quad \hat{\Lambda}^{(0)} := \begin{cases} \underline{\Lambda}_t^h, & t \in [0, t_{k-1}], \\ \underline{\Lambda}_{t_{k-1}}^h, & t \in [t_{k-1}, t_k], \end{cases}$$

and set $\hat{\Lambda}^{(n+1)} = \Gamma_h[\hat{\Lambda}^{(n)}]$, then λ^{n+1} above is equal to $\Gamma_h[\hat{\Lambda}^{(n)}]_{t_k}$. We convince ourselves that this computes the minimal solution: As in the case of the initial condition, the iteration is increasing in n for the losses, $\lambda^{n+1} \geq \lambda^n$ and $\iota^{n+1} \geq \iota^n$, and the iteration terminates in at most $\lfloor \alpha/\sqrt{h} \rfloor$ iterations (because ι^n can only take values in $\{i_{k-1}, \dots, \lfloor \alpha/\sqrt{h} \rfloor\}$). If n_0 is the smallest n such that $\lambda^{n+1} = \lambda^n$, then $\hat{\Lambda}^{(n_0)}$ solves (20) on $[0, t_k]$, and therefore $\Lambda_t^{(n_0)} \geq \underline{\Lambda}_t^h$ for $t \in [0, t_k]$ by Remark 3.2. On the other hand, since $\underline{\Lambda}^h$ is increasing, it holds that $\hat{\Lambda}^{(n_0)} \leq \underline{\Lambda}^h$, and by the monotonicity of Γ_h and a straightforward induction it follows that $\hat{\Lambda}^{(n)} \leq \underline{\Lambda}^h$ for $n \in \mathbb{N}$ and hence $\hat{\Lambda}^{(n_0)} = \underline{\Lambda}_t^h$ for $t \in [0, t_k]$.

5. Numerical tests. In this section, we analyze the computational performance and, especially, the numerical accuracy of the schemes.

As a first example, we consider a $\Gamma(k, \theta)$ distribution for X_{0-} with $k = 2$ and $\theta = 1/3$. Note that the initial density f is globally Lipschitz in this case. Furthermore, we fix the time interval $[0, T] = [0, 0.02]$.

5.1. Explicit and implicit time-stepping schemes. We first consider the implicit and explicit Euler-type time-stepping schemes from section 2 (i.e., without Donsker approximation), whose precise difference is explained in Remark 2.6. We analyze the example above for $\alpha = 1.5$, where a jump is observed. In the first two experiments, we fix a seed to generate $n = 100\,000$ sample paths for the particle method detailed in section 4.1, and vary the number N of time points, setting $\Delta = 1/N$. The relatively small number of particles is chosen to keep the computational time similar to the Donsker scheme below.

Figure 1(a) shows how the explicit scheme smooths out the jump and takes a significant time for the losses to “catch up” in a way already seen in [26] (see, in particular, Figure 3 there). More precisely, the implicit scheme with $N = 100$ behaves similarly to the explicit scheme with $N = 400$.

Concerning the convergence of $L_T^\Delta = \underline{\Lambda}_T^\Delta/\alpha$ (for $\Delta = 1/N$), both the implicit and the explicit time-stepping schemes show more irregular behavior than the implicit Donsker approximation as illustrated in Figure 1(b). This is likely a consequence of the Monte Carlo error which is quite high due to the relatively small sample size. Figure 2 quantifies this further by showing the error estimator $4|t_*^{2\Delta} - t_*^\Delta|$ for the jump times,⁴ where t_*^Δ denotes the jump time for mesh size Δ , identified by $t_*^\Delta = \Delta \arg\max_{0 < k \leq N} \{\underline{\Lambda}_{t_k}^\Delta - \underline{\Lambda}_{t_{k-1}}^\Delta\}$. Also shown is the error estimator $4|J^{2\Delta} - J^\Delta|$ for the jump size, where $J^\Delta = L_{t_*}^\Delta - L_{t_*-\Delta}^\Delta$. Here, we choose between $N = 2$ and approximately 2000 time steps, and 2000N samples, to reduce the Monte Carlo error together with the time-stepping error.

The jump times appear to converge with order $1/2$ for both the explicit and the implicit schemes, where the implicit scheme is slightly more accurate by a constant factor.⁵ This is consistent with the earlier observation in Figure 1(a). The simple estimate J^Δ of the jump size does not converge to the true jump size of around 0.78

⁴The factor 4 accounts for extrapolation of the time-stepping error and additional Monte Carlo error.

⁵The missing data points in the implicit case are explained by 0 change of the jump time for coarse time steps and the resulting undefined values in the log-log plot.

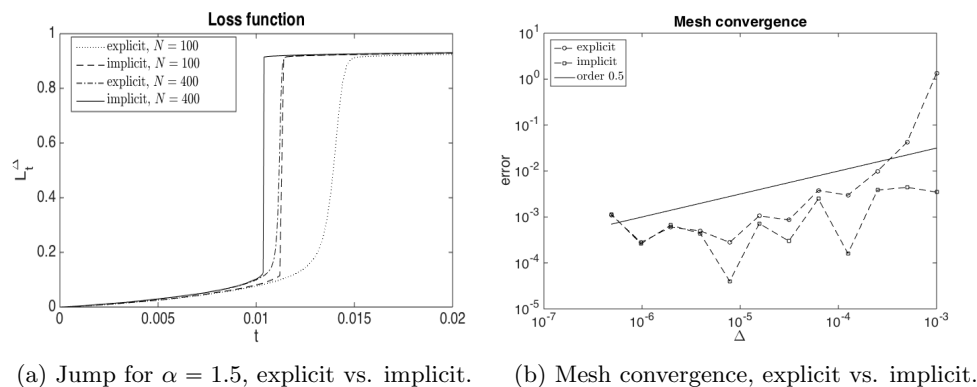


FIG. 1. Convergence of $L^\Delta = \underline{L}^\Delta / \alpha$ in the time-stepping schemes for $\alpha = 1.5$.

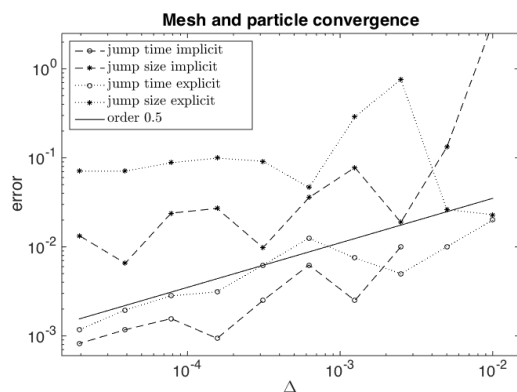


FIG. 2. Convergence of jump time and jump size: explicit and implicit time stepping.

for the explicit scheme but fluctuates around 0.27. We conjecture that this is due to the fact that the estimator only considers changes over a single time step, which are limited by the transmission mechanism discussed in Remark 2.8. Although we have theoretical convergence guarantees of the time-stepping scheme at continuity points, this does not imply convergence of the largest increment over a *single time step* to the physical jump size. Empirically (i.e., without theoretical guarantees), the convergence of the jump size in the implicit scheme is consistent with order 1/2, but the high variance due to the particle estimator prevents a reliable conclusion.

5.2. Convergence properties of the Donsker scheme. We now revisit the same example given at the beginning of the section with the Donsker scheme from section 4.2 with its iterative solution from section 4.3. Here, $h = 1/N$, the number of time points, and for simplicity, we set $u_i^0 = f(i\sqrt{h})$.

Figure 3(a) shows the numerical free boundary $(t_k, \underline{L}_{t_k}^h / \alpha)$, labeled “implicit” and scaled by $1/\alpha$ to show how much mass was absorbed at the boundary. The solution is compared to an approximation computed with a simplified scheme where the iterative procedure from section 4.3 is stopped after the first iteration. The latter can be considered as an explicit treatment, where, analogous to the explicit particle scheme, the interaction term computed at each time point is used to compute the

density at the following point, and is hence labeled “explicit,” in contrast to scheme (41), where p^k and $\underline{\Delta}_{t_k}^h$ are implicitly coupled.

In Figure 3(a), we observe the same phenomenon as in Figure 1(a) for the particle scheme when comparing the resolution of the jump between explicit and implicit schemes.

Here, the implicit scheme (41) reproduces a sharp jump; i.e., for sufficiently small h , the increment $\underline{\Delta}_{t_{k+1}}^h - \underline{\Delta}_{t_k}^h$ is small for all but one k , while this largest increase does not go to 0 as h goes to 0 but converges to the true jump size.⁶

Let us now turn to the convergence of $L^h = \underline{\Delta}^h/\alpha$ as the step size h goes to 0 in the Donsker scheme. Figure 3(b) shows the error estimator $2(L_T^h - L_T^{2h})$ for decreasing h , for $\alpha = 0.5$ and $\alpha = 1.5$, and otherwise the same parameters as earlier.⁷ In both cases, the order of convergence appears to be 0.5, irrespective of the jump that occurs for $\alpha = 1.5$.

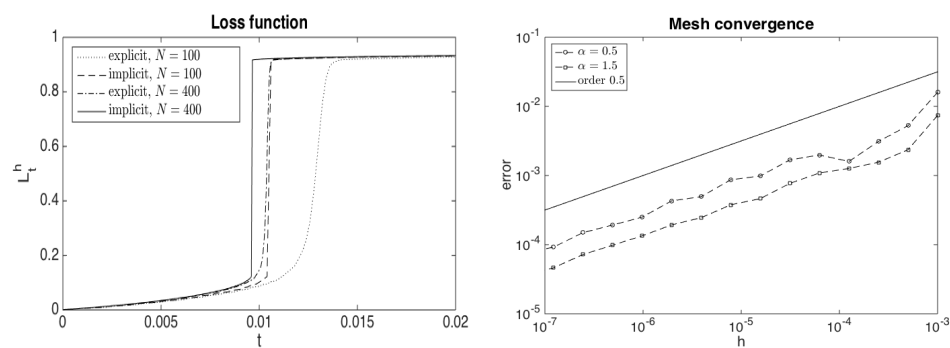
One key advantage of the Donsker scheme is the avoidance of Monte Carlo sampling, which explains its outperformance over the time-stepping algorithm in terms of computational complexity. We shall therefore focus in the following examples solely on the implicit Donsker approximation, for which we investigate two further cases of intermediate regularity, inspired by [26].

First, we consider the jump regime with $\alpha = 1.5$ and vary k in the $\Gamma(k, 1/3)$ initial distribution of X_{0-} to $k = 3/2$ and $5/4$, such that the density is only Hölder $1/2$ and $1/4$, respectively. As seen from Figure 4(a), the empirical convergence order is still 0.5 in all cases, even though for small N the lower regularity of the initial density is noticeable.

Second, we consider the initial density

$$(44) \quad f(x) = \begin{cases} \frac{1}{\alpha} - cx^a, & 0 \leq x \leq A, \\ 0, & x > A, \end{cases}$$

for $\alpha > 0$ and $a > 0$, which we vary in the tests, and where $A > 0$ is determined by $\int_0^\infty f(x) dx = 1$ for given $c > 0$, the latter being sufficiently small. Moreover, we let $T = 10^{-4}$ be small enough to precede a possible discontinuity. Here, the convergence



(a) Jump for $\alpha = 1.5$, explicit vs. implicit. (b) Implicit Donsker, $\alpha = 0.5$ and $\alpha = 1.5$.

FIG. 3. Convergence of $L^h = \underline{\Delta}^h/\alpha$ in the Donsker scheme.

⁶We thank Andreas Søjmark for a discussion on the implicit treatment of jumps.

⁷Assuming the error to be $L_T - L_T^h \approx ch^{1/2}$, we find more precisely that $L_T^h - L_T^{2h} \approx (\sqrt{2} - 1)(L_T - L_T^h)$.

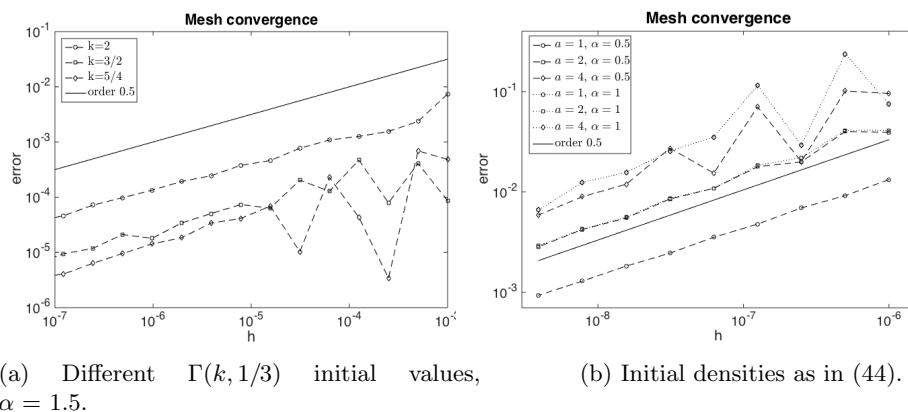


FIG. 4. Mesh convergence, implicit Donsker.

TABLE 1

Average and maximum number of iterations over all N time points for $\alpha = 0.5$ (no jump) and $\alpha = 1.5$ (jump).

	Time points N	100	200	400	800	1600	3200
$\alpha = 0.5$,	av. iter.	1.0200	1.0150	1.0125	1.0088	1.0063	1.0047
	max. iter.	2	2	2	2	2	2
$\alpha = 1.5$,	av. iter.	1.2800	1.1800	1.1175	1.0775	1.0513	1.0341
	max. iter.	16	18	20	20	22	23

order of the explicit time-stepping scheme in [26] is $1/(2(a+1))$ (see Theorem 1.5 and Table 1 there). Remarkably, the asymptotic order of our scheme appears to be 0.5 irrespective of a ; see Figure 4(b).

5.3. Iterative solution. Finally, we examine iteration (42) for the Donsker scheme for the example introduced at the start of section 5. For two different values of α and an increasing number of time points N such that $h = 1/N$, Table 1 gives the number of iterations before termination, first averaged over all time steps, and then the maximum number of iterations for any time step.

In the regular case without a jump ($\alpha = 0.5$), there are never more than 2 iterations needed, while the average number is close to 1. This is explained by the fact that $i_k \neq i_{k-1}$ only if $\underline{\Lambda}_{t_k}^h - \underline{\Lambda}_{t_{k-1}}^h > \alpha\sqrt{h}$, so in the regular regime, where $\underline{\Lambda}_t$ is differentiable, a change of i_k will only happen every $O(1/\sqrt{h})$ time steps, and usually by only 1. Put differently, because of the monotonicity of $\underline{\Lambda}_{t_k}^h$ in k and that of the iteration, the total number of iterations summed up over all time steps is bounded by α/\sqrt{h} . So the average number of iterations is $1 + O(\sqrt{h})$.

In the presence of a jump ($\alpha = 1.5$), a larger number of iterations is needed at the time of the jump, but this number only grows mildly under mesh refinement, and on average the number of iterations is still close to 1.

In either case, therefore, the computational cost of the iteration amounts to less than 10% of the overall cost for reasonably fine time meshes.

REFERENCES

- [1] T. AIKI AND H. IMAI, *Blow-up points to one phase Stefan problems with Dirichlet boundary conditions*, in Modelling and Optimization of Distributed Parameter Systems Applications to Engineering, Chapman & Hall, New York, 1996, pp. 83–89.

- [2] J. M. BACK, S. McCUE, AND T. MORONEY, *Numerical study of two ill-posed one phase Stefan problems*, ANZIAM J., 52 (2010), pp. C430–C446.
- [3] C. BENDER, *Simple arbitrage*, Ann. Appl. Probab., 22 (2012), pp. 2067–2085, <https://doi.org/10.1214/11-AAP830>.
- [4] J. CALDWELL AND Y. KWAN, *Numerical methods for one-dimensional Stefan problems*, Comm. Numer. Methods Engrg., 20 (2004), pp. 535–545.
- [5] F. CAMILLI AND M. FALCONE, *An approximation scheme for the optimal control of diffusion processes*, ESAIM Math. Model. Numer. Anal., 29 (1995), pp. 97–122.
- [6] M. CSÖRGÖ AND L. HORVÁTH, *Weighted Approximations in Probability and Statistics*, Wiley Ser. Probab. Math. Statist. Probab. Math. Statist., John Wiley & Sons, Chichester, UK, 1993.
- [7] C. CUCHIERO, C. REISINGER, AND S. RIGGER, *Optimal bailout strategies resulting from the drift controlled supercooled Stefan problem*, Ann. Oper. Res., (2023), <https://doi.org/10.1007/s10479-023-05293-7>.
- [8] C. CUCHIERO, S. RIGGER, AND S. SVALUTO-FERRO, *Propagation of minimality in the supercooled Stefan problem*, Ann. Appl. Probab., 33 (2023), pp. 1588–1618.
- [9] F. DELARUE, J. INGLIS, S. RUBENTHALER, AND E. TANRÉ, *Particle systems with a singular mean-field self-excitation. Application to neuronal networks*, Stochastic Process. Appl., 125 (2015), pp. 2451–2492.
- [10] F. DELARUE, S. NADTOCHIY, AND M. SHKOLNIKOV, *Global solutions to the supercooled Stefan problem with blow-ups: Regularity and uniqueness*, Probab. Math. Phys., 3 (2022), pp. 171–213.
- [11] J. DEWYNNE, S. HOWISON, J. OCKENDON, AND W. XIE, *Asymptotic behavior of solutions to the Stefan problem with a kinetic condition at the free boundary*, J. Austral. Math. Soc. Ser. B, 31 (1989), pp. 81–96, <https://doi.org/10.1017/S0334270000006494>.
- [12] E. DiBENEDETTO AND A. FRIEDMAN, *The ill-posed Hele–Shaw model and the Stefan problem for supercooled water*, Trans. Amer. Math. Soc., 282 (1984), pp. 183–204, <https://doi.org/10.2307/1999584>.
- [13] R. M. DUDLEY, *Real Analysis and Probability*, CRC Press, Boca Raton, FL, 2018.
- [14] M. FALCONE AND R. FERRETTI, *Semi-Lagrangian Approximation Schemes for Linear and Hamilton–Jacobi Equations*, SIAM, Philadelphia, 2013.
- [15] A. FASANO, G. H. MEYER, AND M. PRIMICERIO, *On a problem in the polymer industry: Theoretical and numerical investigation of swelling*, SIAM J. Math. Anal., 17 (1986), pp. 945–960, <https://doi.org/10.1137/0517067>.
- [16] A. FASANO AND M. PRIMICERIO, *New results on some classical parabolic free-boundary problems*, Quart. Appl. Math., 81 (1980), pp. 439–460, <https://doi.org/10.1090/qam/614552>.
- [17] A. FASANO AND M. PRIMICERIO, *A critical case for the solvability of Stefan-like problems*, Math. Methods Appl. Sci., 5 (1983), pp. 84–96, <https://doi.org/10.1002/mma.1670050107>.
- [18] A. FASANO, M. PRIMICERIO, S. HOWISON, AND J. OCKENDON, *On the singularities of one-dimensional Stefan problems with supercooling*, in Mathematical Models for Phase Change Problems (Óbidos, 1988), Internat. Ser. Numer. Math. 88, Birkhäuser, Basel, 1989, pp. 215–226.
- [19] A. FASANO, M. PRIMICERIO, S. HOWISON, AND J. OCKENDON, *Some remarks on the regularization of supercooled one-phase Stefan problems in one dimension*, Quart. Appl. Math., 48 (1990), pp. 153–168.
- [20] B. HAMBLY, S. LEDGER, AND A. SØJMARK, *A McKean–Vlasov equation with positive feedback and blow-ups*, Ann. Appl. Probab., 29 (2019), pp. 2338–2373, <https://doi.org/10.1214/18-AAP1455>.
- [21] B. HAMBLY AND A. SØJMARK, *An SPDE model for systemic risk with endogenous contagion*, Finance Stoch., 23 (2019), pp. 535–594.
- [22] M. HERRERO AND J. VELÁZQUEZ, *Singularity formation in the one-dimensional supercooled Stefan problem*, European J. Appl. Math., 7 (1996), pp. 119–150.
- [23] S. HOWISON, J. OCKENDON, AND A. LACEY, *Singularity development in moving-boundary problems*, Quart. J. Mech. Appl. Math., 38 (1985), pp. 343–360, <https://doi.org/10.1093/qjmam/38.3.343>.
- [24] S. HOWISON AND W. XIE, *Kinetic undercooling regularization of supercooled Stefan problems*, in Mathematical Models for Phase Change Problems (Óbidos, 1988), Internat. Ser. Numer. Math. 88, Birkhäuser, Basel, 1989, pp. 227–237.
- [25] V. KAUSHANSKY AND C. REISINGER, *Simulation of a simple particle system interacting through hitting times*, Discrete Contin. Dyn. Syst. Ser. B, 24 (2019), pp. 5481–5502.
- [26] V. KAUSHANSKY, C. REISINGER, M. SHKOLNIKOV, AND Z. Q. SONG, *Convergence of a time-stepping scheme to the free boundary in the supercooled Stefan problem*, Ann. Appl. Probab., 33 (2023), pp. 274–298.

- [27] J. R. KING AND J. D. EVANS, *Regularization by kinetic undercooling of blow-up in the ill-posed Stefan problem*, SIAM J. Appl. Math., 65 (2005), pp. 1677–1707, <https://doi.org/10.1137/04060528X>.
- [28] A. LACEY AND J. OCKENDON, *Ill-posed free boundary problems*, Control Cybernet., 14 (1985), pp. 275–296.
- [29] S. LEDGER AND A. SØJMARK, *Uniqueness for contagious McKean–Vlasov systems in the weak feedback regime*, Bull. Lond. Math. Soc. (2), 52 (2020), pp. 448–463, <https://doi.org/10.1112/blms.12337>.
- [30] A. LIPTON, V. KAUSHANSKY, AND C. REISINGER, *Semi-analytical solution of a McKean–Vlasov equation with feedback through hitting a boundary*, European J. Appl. Math., 32 (2021), pp. 1035–1068.
- [31] S. MCCUE, M. EL-HACHEM, AND M. SIMPSON, *Traveling waves, blow-up, and extinction in the Fisher–Stefan model*, Stud. Appl. Math., 148 (2022), pp. 964–986.
- [32] G. H. MEYER, *One-dimensional parabolic free boundary problems*, SIAM Rev., 19 (1977), pp. 17–34, <https://doi.org/10.1137/1019003>.
- [33] S. MITCHELL AND M. VYNNYCKY, *Finite-difference methods with increased accuracy and correct initialization for one-dimensional Stefan problems*, Appl. Math. Comput., 215 (2009), pp. 1609–1621.
- [34] S. NADTOCHIY AND M. SHKOLNIKOV, *Particle systems with singular interaction through hitting times: Application in systemic risk modeling*, Ann. Appl. Probab., 29 (2019), pp. 89–129, <https://doi.org/10.1214/18-AAP1403>.
- [35] S. NADTOCHIY AND M. SHKOLNIKOV, *Mean field systems on networks, with singular interaction through hitting times*, Ann. Probab., 48 (2020), pp. 1520–1556, <https://doi.org/10.1214/19-AOP1403>.
- [36] R. PEYRE, *Fractional Brownian motion satisfies two-way crossing*, Bernoulli, 23 (2017), pp. 3571–3597, <https://doi.org/10.3150/16-BEJ858>.
- [37] B. SHERMAN, *A general one-phase Stefan problem*, Quart. Appl. Math., 28 (1970), pp. 377–382.
- [38] J. STEFAN, *Über einige Probleme der Theorie der Wärmeleitung*, Sitzungber. Wien. Akad. Mat. Natur., 98 (1889), pp. 473–484.
- [39] A. VISINTIN, *Stefan problem with a kinetic condition at the free boundary*, Ann. Mat. Pura Appl. (4), 146 (1987), pp. 97–122, <https://doi.org/10.1007/BF01762361>.
- [40] X. WEIQING, *The Stefan problem with a kinetic condition at the free boundary*, SIAM J. Math. Anal., 21 (1990), pp. 362–373, <https://doi.org/10.1137/0521020>.
- [41] W. WHITT, *Stochastic-Process Limits: An Introduction to Stochastic-Process Limits and Their Application to Queues*, Springer, New York, 2002.