AN OPTIMISATION-BASED REPRESENTATION FOR REACTION-DIFFUSION EQUATIONS

DAVID P. DRIVER, MICHAEL R. TEHRANCHI
UNIVERSITY OF CAMBRIDGE

Abstract. This note is concerned with reaction-diffusion equations with a convex non-linearity. It is shown that a solution to such an equation can be represented as the value function of a particular stochastic optimal control problem. A consequence of this representation is that upper and lower bounds on the solution can be easily found. As an application, the speed of the right-most particle of a branching Lévy process is calculated.

1. Introduction

In this note, we study solutions $u : \mathbb{R}_+ \times E \to \mathbb{R}$ of certain reaction-diffusion equations of the form

$$\frac{\partial u}{\partial t} = \mathcal{L} u + f(t, x, u)$$

where the non-linearity $f(t, x, \cdot)$ is convex for all $(t, x)$, and where the operator $\mathcal{L}$ is the generator of a Markov process valued in some space $E$. The motivating example is the so-called Fisher–Kolmogorov–Petrovskii–Piskunov\(^1\) (FKPP) equation

$$\frac{\partial u}{\partial t} = \frac{1}{2} \frac{\partial^2 u}{\partial x^2} + u(u - 1)$$

in which the operator $\frac{1}{2} \frac{\partial^2}{\partial x^2}$ is the generator of a real-valued Brownian motion.

Reaction-diffusion equations with convex non-linearities are of interest to probabilists as they arise naturally in the study of branching processes. We briefly spell out the key result. Given a Markov process $X$ and a random non-negative integer, construct a branching process $\{X_i^t, i \in I_t, t \geq 0\}$ as follows. Initially, there is one particle following the process $X$. After an exponentially distributed time, this particle is replaced with $N$ particles located at the same position. This procedure then repeats, with each particle moving and branching independently\(^2\). Let $I_t$ be the set of particles alive at time $t$, and let $\mathcal{L}$ be the generator of

---

\(^1\)In the partial differential equations literature, it is not equation (2) but the related equation

$$\frac{\partial v}{\partial t} = \frac{1}{2} \frac{\partial^2 v}{\partial x^2} + v(1 - v)$$

that is usually called the FKPP equation. However, we can recover equation (2) by the substitution $v = 1 - u$.

\(^2\)More generally, the distribution of the number of offspring particles could depend on the current location of its parent, or the location of each offspring could be chosen randomly depending on the location of its parent. Further complications, such as immigration, are also possible. See, for instance, the book of Athreya & Ney [1]. For concreteness, we deal with the simple version of the branching process described here.
and
\[ G(s) = \mathbb{E}(s^N) \]
be the probability generating function of \( N \). Given a function \( u_0 : E \to [0, 1] \), consider the function
\[(3)\]
\[ u(t, x) = \mathbb{E}_x \left[ \prod_{i \in I_t} u_0(X^i_t) \right] \]
where \( \mathbb{E}_x \) denotes the conditional expectation given that the initial particle begins at \( X_0 = x \). Skorokhod [14] and later McKean [12] showed that \( u \) solves, in a certain sense to be made precise below, the FKPP-type equation
\[
\frac{\partial u}{\partial t} = \mathcal{L}u + G(u) - u \\
u(0, x) = u_0(x) \text{ for all } x \in E.
\]
It is possible to extract information directly from FKPP-type equations to apply to the study of branching process. For instance, let \( \{B^i_t, i \in I_t, t \geq 0\} \) be a dyadic (that is, \( N = 2 \) a.s.) branching Brownian motion starting at \( B^i_0 = 0 \), and set
\[ u(t, x) = \mathbb{P} \left( \max_{i \in I_t} B^i_t \leq x \right) \]
so that \( u(t, \cdot) \) is the distribution function of the right-most particle. Note if we set \( B^i_t = x - X^i_t \) then the process \( \{X^i_t, i \in I_t, t \geq 0\} \) is a dyadic branching Brownian motion starting at \( X^i_0 = x \) and
\[ u(t, x) = \mathbb{P}_x \left( \min_{i \in I_t} X^i_t \geq 0 \right) .\]
By the Skorokhod–McKean representation \((3)\), the function \( u \) satisfies the FKPP equation \((2)\) with the so-called Heaviside initial condition
\[ u_0(x) = 1_{\{x \geq 0\}}. \]
A well-known aspect of FKPP-type equations is their propensity to propagate travelling wave fronts. In fact, this property was the focus of the original papers of Fisher [5] and Kolmogorov–Petrovskii–Piskunov [10], which dealt with the spread of an advantageous gene through a population. Again, consider the solution \( u \) of equation \((2)\) with the Heaviside initial condition \( u_0(x) = 1_{\{x \geq 0\}} \). In this case, the equation propagates a travelling wave front in the following sense: defining the median \( m(t) \) for \( t > 0 \) by
\[ u(t, m(t)) = \frac{1}{2} \]
there exists an increasing function \( w \) such that
\[ u(t, x + m(t)) \to w(x) \]
as \( t \to \infty \). The asymptotic speed of the wave front is \( \sqrt{2} \) in sense that
\[(4)\]
\[ \frac{m(t)}{t} \to \sqrt{2} \text{ as } t \to \infty. \]
The McKean–Skorokhod representation combined with travelling wave speed calculation \((4)\) can be interpreted to say that the right-most particle \( \max_{i \in I_t} B^i_t \) of the branching Brownian
motion moves asymptotically linearly with speed $\sqrt{2}$. In particular, this is an example where studying a partial differential equation yields probabilistic information.

Conversely, it is possible to study the FKPP and related reaction-diffusion equations via their connection to branching processes. See, for example, the papers of Harris [7] and Henry-Labordere–Oudjane–Tan–Touzi–Warin [8], and the references therein, for applications of the Skorokhod–McKean representation to the analysis of semilinear partial differential equations.

We now outline the main contributions of this note.

1.1. **An optimisation-based representation.** Our main result is that a solution of equation (1) can be calculated as the value function of a certain minimisation problem. Indeed, for each $t \geq 0$ and each suitable function $u_0$, let $U_t(u_0)$ be the function defined by

$$U_t(u_0)(x) = \sup_{Z \in Z} \mathbb{E}_x[\Xi(t, X, Z; u_0)]$$

where $X$ is the Markov process with generator $\mathcal{L}$, the supremum is taken over a certain set $Z$ of adapted controls $Z = (Z_s)_{0 \leq s \leq t}$ and

$$\Xi(t, X, Z; u_0) = e^{\int_0^t Z_r dr} u_0(X_t) - \int_0^t e^{\int_0^r Z_r dr} \hat{f}(t - s, X_s, Z_s) ds$$

where, for each $(t, x)$ the function $\hat{f}(t, x, \cdot)$ is the Legendre transform of $f(t, x, \cdot)$, defined by

$$\hat{f}(t, x, z) = \sup_v [vz - f(t, x, v)].$$

Theorem 2.4 below says that, subject to some conditions, if $u$ solves equation (1) in a certain sense with initial condition $u(0, \cdot) = u_0$, then

$$u(t, x) = U_t(u_0)(x).$$

The proof, given in Section 2, amounts to verifying that equation (1) is essentially the Hamilton–Jacobi–Bellman equation for the stochastic optimal control problem.

A first application of the optimisation-based representation given by equation (8) is that equation (1) satisfies a comparison principle: if $u^1_0(x) \leq u^2_0(x)$ for all $x$, then $u^1(t, x) \leq u^2(t, x)$ for all $(t, x)$. To see why, simply note that

$$\Xi(t, X, Z; u^1_0) \leq \Xi(t, X, Z; u^2_0)$$

almost surely, for any control $Z$. In particular, the operator $U_t$ is increasing for each $t$.

A related application is that upper and lower bounds of a solution to equation (1) can be found. For instance, clearly we have

$$u(t, x) \geq \mathbb{E}_x[\Xi(t, X, Z; u_0)]$$

for any feasible control $Z \in Z$, and indeed

$$u(t, x) \geq \sup_{Z \in \mathcal{Z}_{\text{sub}}} \mathbb{E}_x[\Xi(t, X, Z; u_0)]$$

for any subset $\mathcal{Z}_{\text{sub}} \subseteq Z$ of feasible controls. On the other hand, we have

$$u(t, x) \leq \sup_{Z \in \mathcal{Z}_{\text{super}}} \mathbb{E}_x[\Xi(t, X, Z; u_0)]$$

for any set $\mathcal{Z}_{\text{super}} \supseteq Z$ of controls that contains the set of feasible controls.
1.2. The case of time and space independent non-linearity. In the case where the convex non-linearity $f(t, x, \cdot) = f(\cdot)$ does not depend on $(t, x)$ is particularly interesting. That is, we restrict attention to the equation

$$\frac{\partial u}{\partial t} = Lu + f(u).$$

In this case, rather explicit upper and lower bounds can be found.

Let $R_t(r_0)$ be the solution $r(t)$ be the solution of the ordinary differential equation

$$\frac{dr}{dt} = f(r)$$

$$r(0) = r_0.$$ 

By equation (8), applied to the constant Markov process $X = X_0$ with generator $L = 0$, we have

$$R_t(r_0) = \sup_{z \in Z_{\text{determ}}} \left( e^{f(\cdot) z, dr} r_0 - \int_0^t e^{f(\cdot) z, ds} \hat{f}(z_s) ds \right)$$

where the supremum is over a set $Z_{\text{determ}}$ of deterministic controls. Hence,

$$R_t(\mathbb{E}_x[u_0(X_t)]) = \sup_{z \in Z_{\text{determ}}} \mathbb{E}_x \left( e^{f(\cdot) z, dr} u_0(X_t) - \int_0^t e^{f(\cdot) z, ds} \hat{f}(z_s) ds \right)$$

$$\leq U_t(u_0)(x).$$

Similarly, we have

$$\Xi(t, X, Z; u_0) \leq R_t(u_0(X_t))$$

almost surely, for all (possibly anticipating) processes $Z$ with suitably integrable sample paths. Hence the above inequality holds for adapted controls $Z$, and so we have

$$U_t(u_0)(x) \leq \mathbb{E}_x[R_t(u_0(X_t))].$$

We can rewrite the calculations of the above paragraph as follows. The ‘diffusion’ term of equation (9) corresponds to a Markov (linear) semigroup $(P_t)_{t \geq 0}$ generated by $L$, while the ‘reaction’ term corresponds to the non-linear semigroup $(R_t)_{t \geq 0}$ generated by the convex (but independent of time and space) function $f$. Since each of the operators $R_t, P_t$ and $U_t$ are increasing, we have shown that

$$R_t \circ P_t \leq U_t \leq P_t \circ R_t$$

where $U$ is the non-linear ‘reaction-diffusion’ semigroup generated by the sum $L + f$.

Note that inequality (10) already says something interesting about branching processes. Indeed, note that in the case where the convex non-linearity is of the form $f(u) = G(u) - u$ where $G$ is the probability generating function of the offspring distribution of a branching process, then the Skorokhod–McKean representation says $R_t$ is the probability generating function of $|I_t|$, the number of particles alive at time $t$:

$$R_t(r_0) = \mathbb{E}[r_0^{\mid I_t \mid}].$$

Hence, inequality (10) says

$$\mathbb{E} \{ \mathbb{E}_x u_0(X_t) \mid \mid I_t \mid \} \leq \mathbb{E}_x \left[ \prod_{i \in I_t} u_0(X_t^i) \right] \leq \mathbb{E}_x [u_0(X_t) \mid I_t \mid]$$

where $I_t$ is the set of indices of the particles alive at time $t$. 

Note that $\Xi(t, X, Z; u_0) \leq R_t(u_0(X_t))$. 

Finally, we have

$$(9) \quad \frac{\partial u}{\partial t} = Lu + f(u).$$

In this case, rather explicit upper and lower bounds can be found.

Let $R_t(r_0)$ be the solution $r(t)$ be the solution of the ordinary differential equation

$$\frac{dr}{dt} = f(r)$$

$$r(0) = r_0.$$ 

By equation (8), applied to the constant Markov process $X = X_0$ with generator $L = 0$, we have

$$R_t(r_0) = \sup_{z \in Z_{\text{determ}}} \left( e^{f(\cdot) z, dr} r_0 - \int_0^t e^{f(\cdot) z, ds} \hat{f}(z_s) ds \right)$$

where the supremum is over a set $Z_{\text{determ}}$ of deterministic controls. Hence,

$$R_t(\mathbb{E}_x[u_0(X_t)]) = \sup_{z \in Z_{\text{determ}}} \mathbb{E}_x \left( e^{f(\cdot) z, dr} u_0(X_t) - \int_0^t e^{f(\cdot) z, ds} \hat{f}(z_s) ds \right)$$

$$\leq U_t(u_0)(x).$$

Similarly, we have

$$\Xi(t, X, Z; u_0) \leq R_t(u_0(X_t))$$

almost surely, for all (possibly anticipating) processes $Z$ with suitably integrable sample paths. Hence the above inequality holds for adapted controls $Z$, and so we have

$$U_t(u_0)(x) \leq \mathbb{E}_x[R_t(u_0(X_t))].$$

We can rewrite the calculations of the above paragraph as follows. The ‘diffusion’ term of equation (9) corresponds to a Markov (linear) semigroup $(P_t)_{t \geq 0}$ generated by $L$, while the ‘reaction’ term corresponds to the non-linear semigroup $(R_t)_{t \geq 0}$ generated by the convex (but independent of time and space) function $f$. Since each of the operators $R_t, P_t$ and $U_t$ are increasing, we have shown that

$$R_t \circ P_t \leq U_t \leq P_t \circ R_t$$

where $U$ is the non-linear ‘reaction-diffusion’ semigroup generated by the sum $L + f$.

Note that inequality (10) already says something interesting about branching processes. Indeed, note that in the case where the convex non-linearity is of the form $f(u) = G(u) - u$ where $G$ is the probability generating function of the offspring distribution of a branching process, then the Skorokhod–McKean representation says $R_t$ is the probability generating function of $|I_t|$, the number of particles alive at time $t$:

$$R_t(r_0) = \mathbb{E}[r_0^{\mid I_t \mid}].$$

Hence, inequality (10) says

$$\mathbb{E} \{ \mathbb{E}_x u_0(X_t) \mid \mid I_t \mid \} \leq \mathbb{E}_x \left[ \prod_{i \in I_t} u_0(X_t^i) \right] \leq \mathbb{E}_x [u_0(X_t) \mid I_t \mid]$$

where $I_t$ is the set of indices of the particles alive at time $t$. 

Note that $\Xi(t, X, Z; u_0) \leq R_t(u_0(X_t))$. 

Finally, we have

$$(9) \quad \frac{\partial u}{\partial t} = Lu + f(u).$$
where $X$ is an independent copy of the Markov process.

However, for our application, the power of inequality (10) is that it can be iterated, leading to the following non-asymptotic form

\[(R_{t/n} \circ P_{t/n})^{(n)} \leq U_t \leq (P_{t/n} \circ R_{t/n})^{(n)}\]

of the Trotter product formula.

For another implication of the representation (8), consider the case when $E = \mathbb{R}$. Suppose that if $x \geq y$ then the conditional law of $X_t$ given $X_0 = x$ stochastically dominates the conditional law of $X_t$ given $X_0 = y$. Examples of processes $X$ satisfying this stochastic dominance condition is a Lévy process or the solution to a stochastic differential equation satisfying a comparison principle. Then an application of representation (8), is that if $u_0$ is decreasing then $U_t(u_0)$ is decreasing for all $t \geq 0$. This fact will also be useful for the discussion of the propagation of wave fronts.

1.3. An application to branching Lévy processes. Consider a branching Lévy process \( \{L_i, i \in I_t, t \geq 0\} \) starting at $L_0 = 0$ where $N$ is the random number of particles produced at each branching event.

Our Theorem 4.1 below says that, conditional on the event that the population does not go extinct, the right-most particle $\max_{i \in I_t} L_i$ moves asymptotically linearly with speed $q$ defined by

\[q = \inf_{\theta > 0} \Lambda(\theta) + \mathbb{E}(N) - 1\]

where $\Lambda$ is the cumulant generating function of the underlying Lévy process, defined by

\[\mathbb{E}[e^{\theta L_t}] = e^{t\Lambda(\theta)} \]

Versions of Theorem 4.1 have appeared before, notably in Biggins [2, Corollary 2]. In contrast to the probabilistic approach taken there, the novelty of our approach is to study the appropriate FKPP-type equation. The key insight is that inequality (11) above can be used to find good upper bounds on the solution of the equation.

1.4. Roadmap. The remainder of the paper is structured as follows. Section 2 contains the statement and proof of Theorem 2.4 which gives precise conditions under which the representation of equation (8) holds true. Section 3 reviews existence and comparison results for the reaction-diffusion equations of interest. Section 4 contains the statement and proof of Theorem 4.1 which finds the speed of the right-most particle of a branching Lévy process. The proof is based on Proposition 4.4 which shows that a certain FKPP-type equation propagates a wave front at the required speed.

2. An optimisation-based representation

The main theorem of this paper says that the solution of the reaction-diffusion equation (1) can be represented as the value function of a certain optimal control problem. Before giving a precise statement of this result, we give a heuristic account of the structure of the proof.
Suppose $u$ is a solution of equation (1) with initial condition $u(0, \cdot) = u_0$. Fix a time horizon $t \geq 0$ and for $0 \leq s \leq t$ let

$$M_s = u(t - s, X_s) + \int_0^s \left( \frac{\partial u}{\partial t} - \mathcal{L}u \right) (t - r, X_r) dr$$

$$= u(t - s, X_s) + \int_0^s f(t - r, X_r, u_r) dr$$

where $u_r = u(t - r, X_r)$. Notice that the process $(M_s)_{0 \leq s \leq t}$ is a local martingale in the filtration generated by $X$. Now pick a suitable adapted process $Z$, and consider the process

$$\xi_s = \exp\int_0^s Z_r dr u(t - s, X_s) - \int_0^s \exp\int_0^r Z_q dq \hat{f}(t - s, X_r, Z_s) ds + dM_s$$

where $\hat{f}$ is the Legendre transform of $f$ defined by equation (7). By construction, the terminal condition is $\xi_t = \Xi(t, X, Z; u_0)$ as defined by equation (6). Also, the dynamics are given by

$$d\xi_s = \exp\int_0^s Z_r dr \left( [u_s Z_s - f(t - s, X_s, u_s) - \hat{f}(t - s, X_s, Z_s)] ds + dM_s \right).$$

The key observation is Young’s inequality

$$\hat{f}(\tau, x, z) + f(\tau, x, u) \geq uz$$

with equality if

$$z = \frac{\partial f}{\partial v}(\tau, x, u),$$

assuming $f$ is differentiable. In particular, the drift of $\xi$ is non-negative in general and vanishes if $Z = Z^*$ where

$$Z^*_s = \frac{\partial f}{\partial v}(t - s, X_s, u_s).$$

Assuming enough integrability so that $\xi$ is a supermartingale, we have

$$u(t, x) = \xi_0 \geq \mathbb{E}_x(\xi_t)$$

$$= \mathbb{E}_x \left( \exp\int_0^t Z_r dr u_0(X_t) - \int_0^t \exp\int_0^r Z_q dq \hat{f}(t - s, X_s, Z_s) ds \right)$$

with equality if $Z = Z^*$.

It remains to exhibit sufficient conditions under which the above argument can be made rigorous. Rather than strive for maximum generality, we present a reasonable set of sufficient conditions. Other choices are possible, but the conditions listed below are sufficient for our application.

The argument appears to require that the stochastic integral $\int \exp\int Z dM$ to be a martingale, or at least a supermartingale for any feasible control and a martingale for the optimal control. However, so far we have made few assumptions on the underlying filtered probability space, so the stochastic integral may not even be defined. Rather than invoking the technicalities of stochastic integration theory, it is easier to avoid stochastic integrals altogether. This is achieved by noting that, in fact, the key step is is the inequality $\mathbb{E}(\xi_t) \leq M_0$ for any feasible
control with equality for the optimal control. By applying the integration by parts formula we have the candidate identity

\[ \xi_t = \int_0^t e^{\int_s^t Z_r dr} [u_s Z_s - f(t - s, X_s, u_s) - \hat{f}(t - s, X_s, Z_s)] ds \]

\[ + M_t + \int_0^t (M_t - M_s) Z_s e^{\int_s^t Z_r dr} ds. \]

This identity can be verified by path-wise Lebesgue integration and Fubini’s theorem, assuming sufficient integrability. It is enough to assume that, almost surely, \( s \rightarrow f(t - s, X_s, u_s) \) is Lebesgue integrable and \( Z \) is bounded.

Note that if \( M \) is a martingale and \( Z \) is bounded, then the inequality \( \mathbb{E}(\xi_t) \leq M_0 \) follows quickly from Young’s inequality, by Fubini’s theorem and the tower property of conditional expectation. Furthermore, if \( M \) is a martingale then

\[ u(t, x) = \mathbb{E}_x(M_t) \]

\[ = \mathbb{E}_x \left[ u_0(X_t) + \int_0^t f(t - r, X_r, u(t - r, X_r)) dr \right]. \]

This observation leads to the following definition:

**Definition 2.1.** Given the non-linearity \( f : \mathbb{R}_+ \times E \times \mathbb{R} \rightarrow \mathbb{R} \), the operator \( \mathcal{L} \) which is the generator of the time-homogeneous Markov process \( X \) and the initial condition \( u_0 : E \rightarrow \mathbb{R} \), a mild solution to equation (13) is a measurable function \( u : \mathbb{R}_+ \times E \rightarrow \mathbb{R} \) such that for all \( (t, x) \)

\[ u(t, x) = \mathbb{E}_x \left[ u_0(X_t) + \int_0^t f(t - s, X_s, u(t - s, X_s)) ds \right]. \]

**Remark 2.2.** Note that the notion of mild solution implicitly assumes certain measurability and integrability of the data \( f, X \) and \( u_0 \). To be explicit, we assume the process \( X \) is measurable, that

\[ \mathbb{E}_x \left[ |u_0(X_t)| + \int_0^t |f(t - s, X_s, u(t - s, X_s))| ds \right] < \infty \]

for all \( (t, x) \).

One advantage of the mild formulation of equation (11), rather than the classical one, is that it not necessary to give meaning to the expression \( \mathcal{L}u \). Indeed, in most cases of interest, the operator \( \mathcal{L} \) is not everywhere-defined. Some results on the existence and uniqueness of mild solutions are reviewed in Section 3.

Another other benefit of the mild formulation is that the process \( M \) defined above is a true martingale:

**Proposition 2.3.** Suppose \( u \) is a mild solution of equation (11). Fix \( t \geq 0 \) and for \( 0 \leq s \leq t \)

\[ M_s = u(t - s, X_s) + \int_0^s h(t - r, X_r) dr \]

where for all \( \tau \geq 0, x \in E \) we let

\[ h(\tau, x) = f(\tau, x, u(\tau, x)). \]
Then \((M_s)_{0 \leq s \leq t}\) is a martingale in the filtration generated by \(X\).

**Proof.** Fix \(0 \leq s \leq t\). We have
\[
u(t - s, X_s) = \mathbb{E} \left[ u_0(\tilde{X}_{t-s}) + \int_0^{t-s} h(t-s-r, \tilde{X}_r) dr \mid \tilde{X}_0 = y \right]
\]
\[
= \mathbb{E} \left[ u_0(X_t) + \int_s^t h(t-r, X_r) dr \mid X_s \right]
\]
\[
= \mathbb{E} \left[ u_0(X_t) + \int_s^t h(t-r, X_r) dr \mid \mathcal{F}_s \right]
\]
where \(\tilde{X}\) is a Markov process independent of \(X\), which of course has the same law as a shifted copy of \(X\), and where \(\mathcal{F}_s = \sigma(X_r : 0 \leq r \leq s)\) defines the filtration generated by \(X\). Hence
\[
\mathbb{E}(M_t | \mathcal{F}_s) = \mathbb{E} \left[ u_0(X_t) + \int_0^t h(t-r, X_r) dr \mid \mathcal{F}_s \right]
\]
\[
= u(t - s, X_s) + \int_0^s h(t-r, X_r) dr
\]
\[
= M_s
\]
as desired. □

Finally, to ensure that the maximum is attained by a feasible control, we note that by equation (12) it is sufficient to assume that \(f\) is differentiable and uniformly Lipschitz.

To summarise, we have proven the following:

**Theorem 2.4.** Suppose that the measurable function \(f : \mathbb{R}_+ \times E \times \mathbb{R} \to \mathbb{R}\) is such that \(f(t, x, \cdot)\) is convex and differentiable. Suppose that the function \(u : \mathbb{R}_+ \times E \to K\) is a mild solution of equation (1), for some subset \(K \subseteq \mathbb{R}\). Suppose there exists a constant \(C > 0\) such that
\[
\left| \frac{\partial f}{\partial v}(t, x, u) \right| < C
\]
for all \((t, x, u) \in \mathbb{R}_+ \times E \times K\). Let \(\Xi(t, X, Z; u_0)\) be defined by equation (6) where \(\hat{f}\) is the Legendre transform of \(f\), defined by equation (7). Finally, let \(U_t(u_0)\) be defined by equation (5) where \(Z\) is the set of bounded measurable processes adapted to the filtration generated by \(X\). Then
\[
u(t, x) = U_t(u_0)(x)
\]
holds for all \((t, x)\). Furthermore, the supremum is achieved for the process \(Z^*\) given by equation (12).

### 3. Existence, comparison and an extension

In this section we address the question of whether equation (11) has a mild solution, and under what conditions the solution is valued in a certain subset \(K \subseteq \mathbb{R}\). Fortuitously, the uniform Lipschitz assumption of Theorem 2.4 also guarantees the existence of the mild solution. The following is a standard existence and comparison result. A proof using Picard iteration can be found in the paper of Cabré & Roquejoffre[4, Section 2.3].
Proposition 3.1. Suppose that the measurable function \( f : \mathbb{R}_+ \times E \times \mathbb{R} \to \mathbb{R} \) is such that, uniformly in \((t, x)\), the function \( f(t, x, \cdot) \) is Lipschitz and \( f(t, x, 0) \) is bounded. For every bounded measurable \( u_0 : E \to \mathbb{R} \) there exists a unique mild solution \( u : \mathbb{R}_+ \times E \to \mathbb{R} \) to equation (1) with initial condition \( u(0, \cdot) = u_0 \). Furthermore, \( u \) is bounded on \([0, T] \times E\) for any time horizon \( T > 0 \).

Let \( u_i^1 \) for \( i = 1, 2 \) be two bounded functions, and let \( u^1 \) be the mild solution to equation (1) with initial condition \( u_0^1 \). If \( u_0^1(x) \leq u_0^2(x) \) for all \( x \in E \), then \( u^1(t, x) \leq u^2(t, x) \) for all \((t, x) \in \mathbb{R}_+ \times E\).

For the application to branching processes, we need to consider the case \( f(u) = G(u) - u \), where \( G \) is the probability generating function of the offspring distribution. The function \( f \) is convex, but unfortunately, it is not necessarily globally Lipschitz. Indeed, in the case of dyadic branching we have \( f(u) = u^2 - u \). Hence, it is not clear a priori whether Theorem 2.4 is applicable. The idea is to restrict attention to the compact set \( K = [0, 1] \). Fortunately, we can use the fact that \( f(0) \geq 0 \) and \( f(1) = 0 \) together with the comparison result in Proposition 3.1 to show that there are solutions valued in \( K \).

Proposition 3.2. Suppose that \( f \) is such that \( f(0) \geq 0 \) and \( f(1) = 0 \), with \( f \) Lipschitz on \([0, 1]\). Given \( u_0 : E \to [0, 1] \), there exists a mild solution \( u \) to equation (1) with initial condition \( u_0 \) such that \( u(t, x) \in [0, 1] \) for all \((t, x)\).

Proof. Let \( \bar{f} \) be a globally Lipschitz function such that \( \bar{f} = f \) on \([0, 1]\). Given any bounded \( u_0 \), there exists a unique mild solution \( u \) to the modified equation

\[
\frac{\partial u}{\partial t} = \mathcal{L}u + \bar{f}(u)
\]

with initial condition \( u(0, \cdot) = u_0 \). Note \( u^1(t, x) = 1 \) is a solution to this equation. Hence by the comparison principle of Proposition 3.1 if \( u_0(x) \leq 1 = u^1(x) \) for all \( x \) then \( u(t, x) \leq 1 \) for all \((t, x)\). Similarly, the function \( u^0(t, x) = r(t) \) is a solution to the equation, where \( r(t) \) is the solution to the ordinary differential equation

\[
\frac{dr}{dt} = \bar{f}(r)
\]

\( r(0) = 0 \).

By familiar phase-plane analysis, we have \( r(t) \geq 0 \) for \( t \geq 0 \). Hence if \( u_0(x) \geq 0 \) for all \( x \) then by the comparison principle \( u(t, x) \geq r(t) \geq 0 \) for all \((t, x)\). Since \( f \) and \( \bar{f} \) agree on \([0, 1]\), we have shown that \( u \) is a mild solution of equation (1). \( \square \)

Remark 3.3. The upshot of Proposition 3.2 is that Theorem 2.4 is applicable to the FKPP-type equations arising in the study of branching processes if \( \mathbb{E}(N) < \infty \). Indeed, for any initial condition \( u_0 \) taking values in \([0, 1]\), there exists a unique solution \( u \) to the FKPP equation valued in \([0, 1]\). Since

\[
|f'(v)| = |G'(v) - 1| \leq \mathbb{E}(N) + 1
\]

for all \( v \in [0, 1] \), the hypotheses of Theorem 2.4 are verified.
Remark 3.4. As an example, consider the classical FKPP equation (2). Then for any bounded, non-negative $u_0$ we have
\[ u(t, x) = \max_z \mathbb{E}_x \left[ e^{\int_0^t Z_s \frac{\partial}{\partial x} u_0(X_t) - \frac{1}{4} \int_0^t e^{\int_0^r Z_r dr} (Z_s + 1)^2 ds} \right] \]
where $X$ is a Brownian motion. By letting $Y_s = e^{\frac{1}{2} \int_0^s (Z_r + 1)} dr$ we have the alternative representation
\[ u(t, x) = \max_Y \mathbb{E}_x \left[ e^{\int_0^t Y_t^2 u_0(X_t) - \int_0^t e^{\int_0^s Y_s^2 ds} \right] \]
where the maximum is over absolutely continuous adapted, positive and bounded processes $Y$ with $Y_0 = 1$, where $\dot{Y}$ is its weak derivative.

4. The right-most particle of a branching Lévy process

Consider a branching Lévy process $\{L_i^t, i \in I_t, t \geq 0\}$ starting at $L^0 = 0$ where $N$ is the random number of particles produced at each branching event. See, for instance, the paper Kyprianou [11] for a general discussion of this process. Recall from the introduction that we are considering the special case where, at a branching event, the offspring particles are born at the current location of its parent.

Let $E = \{|I_t| \to 0\}$ be the event that the branching process eventually becomes extinct, and $E^c$ the complementary event of non-extinction. We will assume here that the branching process is supercritical, meaning that $\mathbb{P}(E) < 1$. Recall that the necessary and sufficient condition of supercriticality is $\mathbb{E}(N) > 1$. Furthermore, recall that the probability of extinction $\mathbb{P}(E)$ is the smallest non-negative root of the equation $G(s) = s$ where
\[ G(s) = \mathbb{E}(s^N) \]
is the probability generating function of $N$. See the book of Athreya & Ney [1, Theorem III.4.1].

We will also assume that $\mathbb{E}(N) < \infty$. This is a sufficient condition that the branching process does not explode in finite time, so that $\mathbb{P}(|I_t| < \infty) = 1$ for all $t \geq 0$. Again, see the book of Athreya & Ney [1, Theorem III.2.1]

Let $\Lambda$ be the cumulant generating function of the underlying Lévy process, defined by
\[ \mathbb{E}[e^{\theta L^t}] = e^{\Lambda(\theta)}. \]
Suppose that $\Lambda$ is finite in a neighbourhood of $\theta = 0$. Recall that by the Lévy–Khintchine formula we have
\[ \Lambda(\theta) = b\theta + \frac{1}{2} \sigma^2 \theta^2 + \int_{\mathbb{R}\setminus\{0\}} [e^{\theta y} - 1 - \theta y \mathbb{I}_{|y|\leq1}] \nu(dy) \]
for some constants $b, \sigma$ and measure $\nu$, where we are supposing that $\int (e^{\theta y} \wedge y^2) \nu(dy) < \infty$ for all $\theta$ in some neighbourhood of $\theta = 0$.

Theorem 4.1. Set
\[ q = \inf_{\theta > 0} \frac{\Lambda(\theta) + \mathbb{E}(N) - 1}{\theta}. \]
Then for any $\varepsilon > 0$ we have
\[ \mathbb{P}(\frac{1}{|I_t|} \sup_{i \in I_t} L^i_t - q > \varepsilon \mid E^c) \to 0 \]
as \( t \to \infty \), where \( \sup \emptyset = -\infty \) as usual.

**Remark 4.2.** Note that if \( L \) is degenerate, in the sense that \( L_t = bt \) for some constant \( b \), then there is no branching and hence the right-most particle moves with speed \( b \). This agrees with Theorem 4.1 since in this case \( \Lambda(\theta) = b\theta \) and hence \( q = b + \inf_{\theta > 0} \frac{1}{\theta}(\mathbb{E}(N) - 1) = b \). Therefore, for the remainder of this section, we will assume that \( L \) is non-degenerate, so that \( \text{Var}(L_t) > 0 \) for \( t > 0 \).

**Remark 4.3.** It is possible to express the speed \( q \) in several ways. For instance, following an idea in the paper of Hiriart-Urruty & Martínez-Legaz [9], define a new function \( \Lambda^\circ \) by the formula

\[
\Lambda^\circ(\theta) = \begin{cases} 
+\infty & \text{if } \theta \geq 0 \\
-\theta \Lambda(-1/\theta) & \text{if } \theta < 0.
\end{cases}
\]

Note that the function \( \Lambda^\circ \) is convex, and indeed, it is related to the perspective function of the cumulant generating function \( \Lambda \). Define its Legendre transform in the usual fashion

\[
\hat{\Lambda}^\circ(v) = \sup_{\theta} [v\theta - \Lambda^\circ(\theta)].
\]

We can then see that \( q \) can be rewritten as

\[
q = -\hat{\Lambda}^\circ(\mathbb{E}(N) - 1).
\]

Alternatively, we will see in the following proof that \( q \) can be rewritten as

\[
q = \sup\{ r : \hat{\Lambda}(r) < \mathbb{E}(N) - 1 \},
\]

where \( \hat{\Lambda} \) is the Legendre transform of \( \Lambda \). This formulation for the speed of the right-most particle appears in the paper of Biggins [2] or, more recently, in the paper of Groisman & M. Jonckheere [6].

To prove Theorem 4.1, we let

\[
u(t, x) = \mathbb{P}\left( \sup_{i \in I_t} L_i^t \leq x \right)
\]

be the distribution function of the right-most particle. By the trick outlined in the introduction, we can let \( X_i^t = x - L_i^t \) so that the process \( \{X_i^t, i \in I_t, t \geq 0\} \) is a branching Lévy motion starting at \( X_0 = x \) and

\[
u(t, x) = \mathbb{P}_x\left( \inf_{i \in I_t} X_i^t \geq 0 \right).
\]

By the Skorokhod–McKean representation [3], the function \( \nu \) satisfies, in a certain sense, the FKPP-type equation

\[
\frac{\partial \nu}{\partial t} = \mathcal{L} \nu + G(\nu) - \nu
\]

with the Heaviside initial condition

\[
u_0(x) = \mathbb{1}_{\{x \geq 0\}},
\]

where and \( \mathcal{L} \) is the generator of \( X \). Note that by the Lévy–Khintchine formula the generator \( \mathcal{L} \) of \( X \) is given by

\[
\mathcal{L} \phi = -b \frac{\partial \phi}{\partial x} + \frac{1}{2} \sigma^2 \frac{\partial^2 \phi}{\partial x^2} + \int_{\mathbb{R}\setminus\{0\}} \left[ \phi(\cdot - y) - \phi + \frac{\partial \phi}{\partial x} \mathbb{1}_{\{|y| \leq 1\}} \right] \nu(dy)
\]
for compactly supported smooth functions $\phi$.

So far, we have not specified in what sense the FKPP-type equation holds in this setting. Fortunately for us, Skorokhod [14, Equation (4)] showed that the precise sense in which the FKPP-type equation holds is the mild sense. In fact, letting $f(u) = G(u) - u$, the equation

$$u(t, x) = \mathbb{E}_x \left[ u_0(X_t) + \int_0^t f(u(t - s, X_s))ds \right]$$

is called the S-equation by Sawyer [13]. Furthermore, since we have assumed that $\mathbb{E}(N) < \infty$, the non-linearity $f$ is Lipschitz on $[0, 1]$. In particular, Theorem 2.4 and its corollaries are applicable.

Theorem 4.1 will follow from the following result.

**Proposition 4.4.** Let $u$ be the unique mild solution of the equation

$$\frac{\partial u}{\partial t} = Lu + f(u)$$

with Heaviside initial condition $u_0(x) = \mathbb{1}_{\{x \geq 0\}}$ where $L$ is the generator of a Lévy process $X$ with Laplace exponent $\Lambda$ defined by the equation

$$\mathbb{E}_x[e^{-\theta X}] = e^{-\theta x + t\Lambda(\theta)},$$

and assume that $\Lambda$ is finite in a neighbourhood of the origin. Suppose $f$ is convex and differentiable on $[0, 1]$, with $f(0) \geq 0$ and $f(1) = 0$. Suppose $f'(1) = \gamma > 0$ and let $\alpha \in [0, 1)$ be the smaller root of $f$. Set

$$q = \inf_{\theta > 0} \frac{\Lambda(\theta) + \gamma}{\theta}$$

Then we have

$$u(t, rt) \to \begin{cases} \alpha & \text{if } r < q \\ 1 & \text{if } r > q \end{cases}$$

**Proof of Theorem 4.1.** Let $M_t = \sup_{s \in I_t} L_s$, where $M_t = -\infty$ when $I_t$ is empty.

First note $\mathbb{P}\left( \left\{ M_t \leq rt \right\} \cap E \right) \to \mathbb{P}(E)$ since

$$\mathbb{P}(E) \geq \mathbb{P}\left( \left\{ M_t \leq rt \right\} \cap E \right) \geq \mathbb{P}\left( \left\{ M_t \leq rt \right\} \cap \{ I_t = \emptyset \} \right) = \mathbb{P}(I_t = \emptyset) \to \mathbb{P}(E).$$

On the other hand, applying Proposition 4.4, we have

$$\mathbb{P}(M_t \leq rt) \to \begin{cases} \mathbb{P}(E) & \text{if } r < q \\ 1 & \text{if } r > q \end{cases}$$

The conclusion follows since

$$\mathbb{P}\left( M_t \leq rt \mid E^c \right) = \frac{1}{\mathbb{P}(E^c)} \left[ \mathbb{P}(M_t \leq rt) - \mathbb{P}\left( \left\{ M_t \leq rt \right\} \cap E \right) \right] \to \begin{cases} 0 & \text{if } r < q \\ 1 & \text{if } r > q \end{cases}.$$
It remains to prove Proposition 4.4. We will use inequality (11) to establish upper and lower bounds. However, it will be more convenient to work with $v = 1 - u$. Let us introduce some new notation. Let

$$g(v) = -f(1 - v).$$

Note that $g$ is concave, that $g(0) = 0$ and $g(1) \leq 0$. Also, $g'(0) = \gamma$ and the larger root of $g$ is $\beta = 1 - \alpha \in (0, 1]$. Let

$$Q_t(q) = 1 - R_t(1 - q)$$

and note that if $q(t) = Q_t(q_0)$ then $q$ satisfies the ordinary differential equation

$$\frac{dq}{dt} = g(q)$$

$$q(0) = q_0.$$

In this notation, inequality (11) becomes

$$(P_{t/n} \circ Q_{t/n})^{(n)}v_0 \leq v(t, \cdot) \leq (Q_{t/n} \circ P_{t/n})^{(n)}v_0.\quad (14)$$

Our goal is to show

$$v(t, rt) \rightarrow \begin{cases} \beta & \text{if } r < q \\ 0 & \text{if } r > q. \end{cases}$$

Of the two bounds, the upper bound is easier to obtain. Using the $n = 1$ case we have

$$v(t, x) \leq Q_t(\mathbb{P}_x(X_t < 0)).$$

By the concavity of $g$ we have

$$g(v) \leq \gamma v$$

and hence by Grönwall’s inequality

$$Q_t(q_0) \leq q_0 e^{\gamma t}.$$

Now by standard large deviation calculations we have

$$\mathbb{P}_x(X_t < 0) \leq e^{-x\theta + t\Lambda(\theta)}$$

for any $\theta > 0$. Putting this together, we have shown

$$v(t, rt) \leq e^{t(\Lambda(\theta) + \gamma - r\theta)}.$$

If $r > \frac{1}{\theta}(\Lambda(\theta) + \gamma)$ then the right-hand side vanishes as $t \rightarrow \infty$, as claimed.

For the lower bound, we will introduce some more notation. Let

$$F_t(y) = \mathbb{P}_0(X_t \leq y)$$

be the conditional distribution function of the random variable $X_t$ given $X_0 = 0$. Note that by spatial homogeneity of the Lévy process, we have

$$\mathbb{P}_x(X_t \leq y) = F_t(y - x).$$

Let $F_t^{-1}$ be the quantile function, defined as

$$F_t^{-1}(p) = \inf\{x : F_t(x) \geq p\},$$

so that $F_t(x) \geq p \iff x \geq F_t^{-1}(p)$.

The key estimates are the following:
Lemma 4.5. For all $0 < b < \beta$, $n \geq 1$, $t > 0$ and $x \in \mathbb{R}$ we have

$$v(t, x) \geq b F_\delta \left( -x - (n - 1) F_\delta^{-1} \left( \frac{Q^{-1}_\delta(b)}{b} \right) \right)$$

where $\delta = t/n$.

Remark 4.6. It is interesting to note that Lemma 4.5 actually holds with no assumption on law of the Lévy process. In particular, it holds for processes, such as stable processes, for which the cumulant generating function $\Lambda(\theta)$ is infinite for all $\theta \neq 0$.

Proof. We fix $\delta$ and use induction on $n$. We first consider the $n = 1$ case.

Since the points 0 and $\beta \leq 1$ are fixed points of $g$, we have $Q^{-1}_\delta(0) = 0$ and $Q^{-1}_\delta(1) \geq \beta$. In particular, we have

$$v(\delta, x) \geq P_\delta \circ Q_\delta 1_{(-\infty, 0]}(x) \geq \beta P_x(X_\delta \leq 0) = \beta F_\delta(-x)$$

To do the inductive step, we will make use of the following observation: for any $0 < b < \beta$ and $k \in \mathbb{R}$ we have

$$Q_\delta[b F_\delta(k)] \geq b 1_{\{F_\delta(k) \geq Q^{-1}_\delta(b)/b\}}$$

since $Q_\delta$ is increasing on $[0, \beta]$. Now suppose the claim is true for $n = k$, we have

$$v((N + 1)\delta, x) \geq P_\delta \circ Q_\delta \left[ b F_\delta \left( -x - (k - 1) F_\delta^{-1} \left( \frac{Q^{-1}_\delta(b)}{b} \right) \right) \right] (x)$$

$$\geq b \mathbb{P}_x \left[ F_\delta \left( -X_\delta - (k - 1) F_\delta^{-1} \left( \frac{Q^{-1}_\delta(b)}{b} \right) \right) \geq \frac{Q^{-1}_\delta(b)}{b} \right]$$

$$= b F_\delta \left( -x - k F_\delta^{-1} \left( \frac{Q^{-1}_\delta(b)}{b} \right) \right).$$

Lemma 4.7. For all $0 < c < \gamma = g'(0)$ and all $0 < b < \beta$, where $\beta$ is the larger root of $g$, there exists $\delta^* > 0$ such that $Q^{-1}_\delta(b) \leq be^{-c\delta}$ for all $\delta \geq \delta^*$.

Proof. Fix a $q^* \in (0, \beta)$, for instance $q^* = \beta/2$ and let

$$H(q) = \int_{q^*}^q \frac{ds}{g(s)}.$$

Note that the differential equation defining $Q$ can be solved as

$$Q_\delta(q_0) = H^{-1}(H(q_0) + \delta)$$

for $0 < q_0 < \beta$, and hence

$$Q^{-1}_\delta(q_0) = H^{-1}(H(q_0) - \delta).$$

In this notation, we must prove that

$$H(b) - \delta \leq H(be^{-c\delta})$$
or equivalently
\[
\frac{1}{\delta} \int_0^{\delta} \frac{bce^{-cx}dx}{g(be^{-cx})} \leq 1
\]
for \(\delta\) large enough. To do this, note that the limit of the left-hand side as \(\delta \to \infty\) is \(c/\gamma < 1\) by l'Hôpital's rule.

**Lemma 4.8.** For all \(r < q\) there exists a \(c < \gamma\) and a \(\delta^* > 0\) such that \(F_\delta^{-1}(e^{-c\delta}) \leq -r\delta\) for all \(\delta \geq \delta^*\).

**Proof.** Note that \(q > -\mathbb{E}_0(X_1)\) with strict inequality since since \(\Lambda(\theta) > -\theta\mathbb{E}_0(X_1)\) by Jensen’s inequality. Hence we need only consider \(r\) such that
\[
-\mathbb{E}_0(X_1) < r < q.
\]
In particular, we may invoke Cramér large deviation principle to conclude that,
\[
\log F_\delta(-r\delta) = -\hat{\Lambda}(r)\delta(1 + o(1))
\]
as \(\delta \to \infty\), where the large deviation rate function \(\hat{\Lambda}\) is the Legendre transform of \(\Lambda\), defined by
\[
\hat{\Lambda}(v) = \sup_{\theta} [v\theta - \Lambda(\theta)].
\]
Hence, it is enough to show that
\[
\hat{\Lambda}(r) < \gamma.
\]
Now, since \(r > \Lambda'(0) = -\mathbb{E}_0(X_1)\), there exists an \(\varepsilon > 0\) such that \(r > \Lambda'(\varepsilon)\), since \(\Lambda'\) is continuous and increasing in a neighbourhood of \(\theta = 0\). By the convexity of \(\Lambda\) we have the inequality
\[
r\theta - \Lambda(\theta) \leq r\varepsilon - \Lambda(\varepsilon)
\]
for \(\theta < \varepsilon\) and hence
\[
\hat{\Lambda}(r) = \sup_{\theta \geq \varepsilon} [r\theta - \Lambda(\theta)]
\]
\[
\leq -\varepsilon(q - r) + \sup_{\theta \geq \varepsilon} [q\theta - \Lambda(\theta)].
\]
The conclusion follows since \(q\theta - \Lambda(\theta) \leq \gamma\) for all \(\theta > 0\) by the definition of \(q\). \(\square\)

**Proof of Proposition 4.4.** Fix \(0 < b < \beta\) and \(r < q\). Pick \(\bar{r}\) such that \(r < \bar{r} < q\). By Lemma 4.8 there exists a \(c\) and \(\delta^*_1\) such that \(F_\delta^{-1}(e^{-c\delta}) \leq -\bar{r}\delta\) for all \(\delta \geq \delta^*_1\). By Lemma 4.7 there exists \(\delta^*_2\) such that \(Q_\delta^{-1}(b) \leq be^{-c\delta}\) for all \(\delta \geq \delta^*_2\).

Let \(m = 1 + \mathbb{E}_0(X_1)\). By the weak law of large numbers
\[
F_\delta(m\delta) = \mathbb{P}_0(X_\delta/\delta \leq m) \to 1.
\]
So given \(\varepsilon > 0\), there exists \(\delta^*_3\) such that \(F_\delta(m\delta) \geq 1 - \varepsilon\) for \(\delta \geq \delta^*_3\).

Let \(n \geq \frac{m + \bar{r}}{r - \bar{r}}\) and \(t \geq n \max_i\{\delta^*_i\}\).
Let $\delta = t/n$, so $\delta \geq \max_i \{\delta^*_i\}$ and hence
\[
v(t, rt) = v(n\delta, rn\delta) \\
\geq bF_\delta(-rn\delta - (n-1)F^{-1}_\delta(Q^{-1}_\delta(b)/b)) \\
\geq bF_\delta(-rn\delta - (n-1)F^{-1}(e^{-c\delta})) \\
\geq bF_\delta(-rn\delta + (n-1)\bar{r}\delta) \\
= bF_\delta([n(\bar{r} - r) - \bar{r}]\delta) \\
\geq bF_\delta(m\delta) \\
\geq b(1 - \varepsilon).
\]

Since $b < \beta$ and $\varepsilon > 0$ are arbitrary, the conclusion follows. \hfill \Box

Remark 4.9. Consider the case of dyadic branching Brownian motion, where $G(s) = s^2$ and $\Lambda(\theta) = \frac{1}{2}\theta^2$. Letting $m(t)$ be the median, defined by $u(t, m(t)) = \frac{1}{2}$, we have
\[
-m(t) \geq -\sqrt{t} \Phi^{-1}\left(\frac{1}{e^t + 1}\right) \\
\geq \frac{1}{\sqrt{n}} \Phi^{-1}\left(\frac{1}{2b}\right) - \frac{n-1}{n} \sqrt{t} \Phi^{-1}\left(\frac{1}{e^{t/n}(1 - b) + b}\right)
\]
for all $1/2 < b < 1, n \geq 1$, where
\[
\Phi(z) = \int_{-\infty}^{z} \frac{e^{-s^2/2}}{\sqrt{2\pi}} ds
\]
is the standard normal distribution function. Indeed, the upper bound follows from the upper bound $1/2 \leq v(t, m) = Q_t[P(B_t \leq -m)]$ and the calculation $Q_t(q_0) = \frac{q_0}{q_0 + e^{-t(1-q_0)}}$ in the case when $g(v) = v(1-v)$. The lower bound is implied by Lemma 4.5.

Using $\Phi^{-1}(\varepsilon) = -\sqrt{2\log(1/\varepsilon)}(1 + o(1))$ as $\varepsilon \downarrow 0$ yields
\[
m(t) = \sqrt{2t} + o(t).
\]

On the other hand, a famous result of Bramson \cite{3} says
\[
m(t) = \sqrt{2t} - \frac{3}{2\sqrt{2}} \log t + O(1).
\]

Since $\Phi^{-1}(\varepsilon) = -\sqrt{2\log(1/\varepsilon)} + O\left(\frac{\log \log(1/\varepsilon)}{\log(1/\varepsilon)}\right)$ the upper bound in equation (15) actually recovers the correct order of magnitude of the second term of the expansion. It would be interesting to see if, by optimising over the free parameters $b$ and $n$, it is possible to recover the $\log t$ term in the lower bound.

References

[1] K.B. Athreya and P.E. Ney. Branching Processes. Die Grundlehren der mathematischen Wissenschaften in Einzeldarstellungen, Band 196. Springer-Verlag. (1972)
[2] J.D. Biggins. The growth and spread of the general branching random walk. Annals of Applied Probability 5(4): 1008–1024. (1995)
[3] M. Bramson. Maximal displacement of branching Brownian motion. Communications on Pure and Applied Mathematics 31(5): 531–581. (1978)
[4] X. Cabrè and J.M. Roquejoffre. The influence of fractional diffusion in Fisher-KPP equations. Communications in Mathematical Physics 320(3): 679–722. (2013)
[5] R.A. Fisher. The wave of advance of advantageous genes. Annals of Eugenics 7(4): 355–369. (1937)
[6] P. Groisman and M. Jonckheere. Front propagation and quasi-stationary distributions for one-dimensional Lévy processes. arXiv:1609.09338

[7] S.C. Harris. Travelling-waves for the FKPP equation via probabilistic arguments. In Proceedings of the Royal Society of Edinburgh: Section A Mathematics 129(3): 503–517. (1998)

[8] P. Henry-Labordere, N. Oudjane, X. Tan, N. Touzi, X. Warin. Branching diffusion representation of semilinear PDEs and Monte Carlo approximation. arXiv:1603.01727 (2016)

[9] J.-B. Hiriart-Urruty and J.-E. Martínez-Legaz. New formulas for the Legendre–Fenchel transform. Journal of Mathematical Analysis and Applications, 288: 544–555 (2003)

[10] A.N. Kolmogorov, I.G. Petrovskii and N.S. Piskunov. A study of the diffusion equation with increase in the amount of substance, and its application to a biological problem. Moscow University Bulletin of Mathematics 1(6): 1–26. (1937) In Selected Works of A.N. Kolmogorov. Volume 1: Mathematics and Mechanics. V.M. Tikhomirov editor, V.M. Volosov translator. Mathematics and Its Applications. Kluwer Academic Press. (1991)

[11] A.E. Kyprianou. A note on branching Lévy processes. Stochastic Processes and their Applications 82: 1–14. (1999)

[12] H.P. McKean. Application of Brownian motion to the equation of Kolmogorov-Petrovskii-Piskunov. Communications on Pure and Applied Mathematics 28(3): 323–331. (1975)

[13] S. Sawyer. A remark on the S-equation for branching processes. Proceedings of the Japan Academy 46(5): 427–429. (1970)

[14] A.V. Skorokhod. Branching diffusion processes. Theory of Probability & Its Applications 9(3): 445–449. (1964)

Statistical Laboratory, Centre for Mathematical Sciences, Wilberforce Road, Cambridge CB3 0WB, UK

E-mail address: m.tehranchi@statslab.cam.ac.uk

17