Hölder regularity for generalized master equations with rough kernels

Luis Caffarelli and Luis Silvestre

September 4, 2012

Abstract

We prove Hölder estimates for integro-differential equations related to some continuous time random walks. These equations are nonlocal both in space and time and recover classical parabolic equations in limit cases. For some values of the parameters, the equations exhibit at the same time finite speed of propagation and C^α regularization.

1 Introduction

We study evolution problems that are related to continuous time random walks (CTRW), which are a discontinuous path for which both the jumps and the time elapsed in between them are random. These processes are governed by a generalized *master equation* which is nonlocal both in space and time.

We consider kernels $K(t,x,s,y)$ in $\mathbb{R}^n \times \mathbb{R}^n \times (0,\infty) \times (0,\infty)$. From this kernels, we define an integral operator which is nonlocal both in space and time.

$$Lu(t,x) = \int_{\mathbb{R}^n} \int_0^\infty (u(t,x) - u(t-s,x+y)) K(t,x,s,y) \, \mathrm{d}s \, \mathrm{d}y, \tag{1.1}$$

We will study equations that may or may not include a time derivation. The first model we are interested is the equation which is purely nonlocal.

$$Lu(t,x) = 0 \tag{1.2}$$

This was stated as equation (22) in [8], also in [9] and [5]. The function u in the equation above represents the distribution of particles following a CTRW that have are arriving at position x at time t.

Other physical models that study the evolution of a distribution of particles following a CTRW involve an equation of the form.

$$u_t + Lu(t,x) = 0 \tag{1.3}$$

Equations of this general form can be found in a variety of physical situations, for example see [2], [3], [4], [5], [10], [11], [14], [15] and [16].

A common simplifying assumption is that the jumps in space and the waiting times are decorrelated: $K(t,x,s,y) = \mu(x,y)\nu(t,s)$. However, studying correlated kernels provides a more flexible framework where more interesting physical phenomena can be observed (see for example the discussion by the end of [5]), and more subtle mathematical questions appear. The regularity estimates are in fact more interesting (harder mathematically) when the jumps in space and the waiting times are strongly correlated.

In order to obtain our regularity results, we need to make some assumptions on the kernels. We consider kernels K that are non degenerate in between two surfaces $c_1|y|^\beta \leq s \leq c_2|y|^\beta$, for some

$\beta > 0$. Moreover, our structural conditions, which follow below, can be interpreted as that the operator is of order σ in space, and σ/β in time. We assume

$$K(t,x,s,y) \geq \frac{\lambda}{|y|^{n+\sigma+\beta}} \qquad \text{when } c_1|y|^\beta \leq s \leq c_2|y|^\beta, \tag{1.4}$$

$$K(t,x,s,y) \leq \frac{\Lambda}{|y|^{n+\beta+\sigma} + s^{n/\beta+1+\sigma/\beta}}. \tag{1.5}$$

The hypothesis (1.5) assures that the integral expression in (1.1) is computable every time u is a smooth function (assuming $\beta > \sigma$). The hypothesis (1.4) is a non degeneracy condition that is necessary to obtain our regularity results.

No regularity is assumed with respect to any of the variables. For simplicity we assume that K is symmetric in y.

$$K(t,x,s,y) = K(t,x,s,-y) \tag{1.6}$$

Theorem 1.1. *Let $u : (-\infty,1) \times \mathbb{R}^n \to \mathbb{R}$ be a bounded solution to $Lu(t,x) = f(t,x)$ for all $(t,x) \in (0,1) \times B_1$, where L is an operator as above and f is a bounded function. Then the solution is in the class $C^\alpha((1/2,1) \times B_{1/2})$ for some $\alpha > 0$. Moreover an estimate holds:*

$$\|u\|_{C^\alpha((1/2,1)\times B_{1/2})} \leq C\left(\|u\|_{L^\infty((-\infty,1)\times\mathbb{R}^n)} + \|f\|_{L^\infty((0,1)\times B_1)}\right).$$

for some C depending on n, σ, β, λ and Λ.

Theorem 1.1 will be rephrased below in the article so that the equation $Lu = 0$ is understood in a weak (viscosity) sense.

We also provide a regularity result for the model with a time derivative.

Theorem 1.2. *Let $u : (-\infty,1) \times \mathbb{R}^n \to \mathbb{R}$ be a bounded solution to $u_t + Lu(t,x) = f(t,x)$ for all $(t,x) \in (0,1) \times B_1$, where L is an operator as above and f is a bounded function. Then the solution is in the class $C^\alpha((1/2,1) \times B_{1/2})$ for some $\alpha > 0$. Moreover an estimate holds:*

$$\|u\|_{C^\alpha((1/2,1)\times B_{1/2})} \leq C\left(\|u\|_{L^\infty((-\infty,1)\times\mathbb{R}^n)} + \|f\|_{L^\infty((0,1)\times B_1)}\right).$$

for some C depending on n, σ, β, λ and Λ.

The proofs of the Theorems 1.1 and 1.2 are given in the last section of the paper. In fact, we provide a restatement of these results as Theorems 7.1 and 7.2, in terms of viscosity solutions.

2 Analysis of assumptions and scaling

It is important to point out that the equation make sense only if $\sigma < \beta$. Otherwise the operator $Lu(t,x)$ has a non integrable function even if $u \in C^\infty_{x,t}$.

The case $\beta = 1$ has finite speed of propagation (meaning that the value of $u(t,x)$ depends only on the values of u in the cone $\{(s,y) : |y-x| < t-s\}$). It is interesting as an example of a parabolic equation, with regularization effects, and finite speed of propagation. It may be the first example of such equation.

On a first look at the main assumptions (1.4) and (1.5), it may seem strange that in the first assumption the lower bound is taken in between two surfaces only. The purpose of these two surfaces is to make the assumptions fairly general so as to accommodate singular kernels K which vanish for some values of s and y. Indeed, the more natural looking alternative

$$K(t,x,s,y) \geq \lambda\left(\frac{1}{|y|^{n+\beta+\sigma} + s^{n/\beta+1+\sigma/\beta}}\right)$$

is a strictly more restrictive assumption.

The kernels which are excluded from our assumptions are those which are made of a singular measure (instead of a locally L^1 density). For example, one can imagine an operator L of the form

$$Lu(t,x) = \int_{\mathbb{R}^n} (u(t,x) - u(t-|y|^\beta, x+y)) K(t,x,y) \, \mathrm{d}y.$$

This operator corresponds to a singular kernel K which is supported on $s = |y|^\beta$. In terms of a CTRW, it corresponds to a situation in which the time it takes for a particle to jump a distance y is always exactly $|y|^\beta$.

The assumptions (1.4) and (1.5) respect a natural scaling. Indeed, if we call $u_r(t,x) = u(r^\beta t, rx)$, then

$$L_r u_r(t,x) = r^\sigma Lu(r^\beta t, rx), \tag{2.1}$$

where L_r is an operator of the same form (1.1) with a kernel K_r satisfying the same assumptions (1.4) and (1.5).

For an operator that involves a time derivative $u_t + Lu$, the scaling is more complicated. In fact, there is no natural scaling that preserves the structure of the equation exactly. We can understand this since the operator L implicitly contains a lower order time derivative, so its scaling does not match the scaling of u_t. We consider the following scaling instead: $u_r(t,x) = u(r^\sigma t, rx)$. The function u_r satisfies the equation

$$\partial_t u_r(t,x) + L_r u_r(t,x) = r^\sigma \left(\partial_t u + Lu\right)(r^\sigma t, rx),$$

where L_r is the operator of the form (1.1) with the kernel $K_\rho(t,x,s,y) = \rho^{n+2\sigma} K(\rho^\sigma t, \rho x, \rho^\sigma s, \rho y)$. This kernel satisfies the bounds

$$K(t,x,s,y) \geq \frac{\rho^{\sigma-\beta}\lambda}{|y|^{n+\sigma+\beta}} \qquad \text{when } \rho^{\beta-\sigma} c_1 |y|^\beta \leq s \leq \rho^{\beta-\sigma} c_2 |y|^\beta, \tag{2.2}$$

$$K(t,x,s,y) \leq \frac{\Lambda \rho^{n+2\sigma}}{|\rho y|^{n+\beta+\sigma} + (\rho^\beta s)^{n/\beta+1+\sigma/\beta}}. \tag{2.3}$$

Note that since $\beta > \sigma$, the estimates two parabolas where K_r is bounded below in (2.2) become flat as $r \to 0$. Moreover, (2.3) says that K_r is concentrating close to $s = 0$ as $r \to 0$. That is, as $r \to 0$, the operator L_r tends to become local in time.

3 Second order parabolic equations as asymptotic limits

Second order parabolic equations can be formally obtained as limits of master equations in different ways. In this section we demonstrate how an operator of the form (1.1) converges to a parabolic operator of the form $c(t,x)\partial_t u - a_{ij}(t,x)\partial_{x_i x_j} u$ in some assymptotic regimes.

3.1 Limit with $\beta = 2$ and $\sigma \to 2$

Proposition 3.1. *Let $a(s,y)$ be a fixed kernel so that*

- $\lambda \leq a(s,y) \leq \Lambda$ *(i.e. a is bounded above and below).*
- $a(s,y) = a(s,\lambda y)$ *for all $\lambda \in \mathbb{R}$, $\lambda \neq 0$ (i.e. a is homogeneous of degree zero). Note we also assume this for $\lambda < 0$, so a is even in y.*

Let us consider the following family of kernels K_σ:

$$K_\sigma(s,y) = \begin{cases} (2-\sigma)\frac{a(s,y)}{|y|^{n+2+\sigma}} & \text{if } c_1|y|^2 \leq s \leq c_2|y|^2, \\ 0 & \text{otherwise.} \end{cases}$$

Let L_σ be the corresponding operator as in (1.1)*. Then, there exist a constant $c > 0$ and a positive definite matrix a_{ij}, such that for any C^2 function u,*

$$\lim_{\sigma \to 2} L_\sigma u(t,x) = c u_t(t,x) - a_{ij} \partial_{ij} u(t,x).$$

Proof. We perform the direct computation.

$$\lim_{\sigma \to 2} L_\sigma u(t,x) = \lim_{\sigma \to 2} \left(\int_{\mathbb{R}^n} \int_0^\infty (u(t,x) - u(t-s, x+y))(2-\sigma) \frac{a(s,y)}{|y|^{n+2+\sigma}} \, \mathrm{d}s \, \mathrm{d}y \right),$$

For any $r > 0$, we split the domain of integration,

$$= \lim_{\sigma \to 2} \left(\int_{B_r} \int_0^\infty (u(t,x) - u(t-s, x+y))(2-\sigma) \frac{a(s,y)}{|y|^{n+2+\sigma}} \, \mathrm{d}s \, \mathrm{d}y + (2-\sigma) \int_{\mathbb{R}^n \setminus B_r} \dots \, \mathrm{d}x \right)$$

For any $r > 0$, the second integral is bounded independently of σ, then we drop that term in the limit.

$$= \lim_{\sigma \to 2} \left(\int_{B_r} \int_0^\infty (u(t,x) - u(t-s, x+y))(2-\sigma) \frac{a(s,y)}{|y|^{n+2+\sigma}} \, \mathrm{d}s \, \mathrm{d}y \right)$$

We now use a second order Taylor expansion for u at (t,x). Note that in the region of integration $s \approx |y|^2$.

$$= \lim_{\sigma \to 2} \left(\int_{B_r} \int_0^\infty (s\, u_t(t,x) - y \cdot \nabla_x u(t,x) - y^t D_x^2 u(t,x) y + o(|y|^2))(2-\sigma) \frac{a(s,y)}{|y|^{n+2+\sigma}} \, \mathrm{d}s \, \mathrm{d}y \right)$$

Integrating in s,

$$= \lim_{\sigma \to 2} \left(\int_{B_r} (c(y)|y|^2 u_t(t,x) - y \cdot \nabla_x u(t,x) - y^t D_x^2 u(t,x) y + o(|y|^2))(2-\sigma) \frac{A(y)}{|y|^{n+\sigma}} \, \mathrm{d}s \, \mathrm{d}y \right)$$

where $A(y) = \frac{1}{|y|^2} \int_{c_1|y|^2}^{c_2|y|^2} a(s,y) \, \mathrm{d}s$, which is bounded above and below. And $c(y) = \frac{1}{A(y)|y|^4} \int_{c_1|y|^2}^{c_2|y|^2} s a(s,y) \, \mathrm{d}s$, which is also bounded above and below. By the assumptions on a, both A and c are radially symmetric in y. We now observe that the term which involved $\nabla_x u$ is odd, and thus integrates to zero.

$$= \lim_{\sigma \to 2} \left(\int_{B_r} (c(y)|y|^2 u_t(t,x) - y^t D_x^2 u(t,x) y + o(|y|^2))(2-\sigma) \frac{A(y)}{|y|^{n+\sigma}} \, \mathrm{d}s \, \mathrm{d}y \right)$$

We now use polar coordinates. Recall that A and c are homogeneous of degree zero.

$$= \lim_{\sigma \to 2} \left(\int_0^r \int_{\partial B_1} (c(\theta) u_t(t,x) - \theta^t D_x^2 u(t,x) \theta + o(1))(2-\sigma) \frac{A(\theta)}{\rho^{n-2+\sigma}} \rho^{n-1} \, \mathrm{d}\theta \, \mathrm{d}\rho \right)$$
$$= \lim_{\sigma \to 2} \left(\int_0^r (2-\sigma) \rho^{1-\sigma} \, \mathrm{d}\rho \right) \left(\int_{\partial B_1} (c(\theta) u_t(t,x) - \theta^t D_x^2 u(t,x) \theta + o(1)) A(\theta) \, \mathrm{d}\theta \right)$$

In the last expression the term $o(1)$ represents a quantity that goes to zero as $r \to 0$. Recall that $r > 0$ is arbitrary.

$$= \lim_{\sigma \to 2} r^{2-\sigma} \left(\int_{\partial B_1} (c(\theta) u_t(t,x) - \theta^t D_x^2 u(t,x) \theta + o(1)) A(\theta) \, \mathrm{d}\theta \right)$$
$$= \int_{\partial B_1} (c(\theta) u_t(t,x) - \theta^t D_x^2 u(t,x) \theta) A(\theta) \, \mathrm{d}\theta + o(1),$$

making $r \to 0$,

$$= cu_t(t,x) - a_{ij}\partial_{x_i x_j} u(t,x).$$

For some positive coefficients c and a_{ij}. This is just because the last expression is a linear function in $u_t(t,x)$ and $D^2u(t,x)$, so it must correspond to some coefficients. The positiveness of c and a_{ij} is a consequence of the monotonicity of the last expression given that $A(\theta) > 0$ and $c(\theta) > 0$. □

3.2 Limit with finite speed of propagation and $c_1 \to 0$

If we relax the assumptions on the kernels, one can consider an asymptotic regime with a fixed β which converges to a second order parabolic equation. Let us consider the following example.

$$K_{\varepsilon\sigma}(s,y) = \frac{(2-\sigma)}{|y|^{n+2+\sigma} + s^{n/2+1+\sigma/2}} \chi_{\{|y|>\varepsilon s\}}.$$

A computation as in the previous subsection shows that the corresponding operator $L_{\varepsilon\sigma}$ will converge to the heat operator $\partial_t - \triangle$ as $\varepsilon \to 0$ and $\sigma \to 0$.

The fact that for any $\varepsilon > 0$ the kernel is supported in $\{|y| > \varepsilon s\}$ effectively defines a cone of dependence, and naturally the equation has a finite speed of propagation (equal to ε^{-1}).

The method presented in this paper does not provide a Hölder continuity estimate for these operators $L_{\varepsilon\sigma}$, even for fixed $\varepsilon > 0$ and $\sigma > 0$. The problem is that the family of kernels defined above is not scale invariant. The cone of dependence would degenerate in small scales when using the parabolic scaling. The parabolic scaling is the only one which is compatible with the first factor.

4 Maximal operators and Viscosity solutions

We define Pucci-like extremal operators and their corresponding viscosity solutions.

The maximal and minimal operators M^+u and M^-u are by definition the maximal and minimal values that an operator $Lu(x)$ of the form (1.1) can achieve under the restrictions (1.4) and (1.5). More explicitly

$$M^+u(t,x) = \max \left\{ \int_{\mathbb{R}^n} \int_0^\infty (u(t,x) - u(t-s,x+y))K(s,y) \, \mathrm{d}s \, \mathrm{d}y : \text{for all K satsifying (1.4) and (1.5)} \right\},$$
$$M^-u(t,x) = \min \left\{ \int_{\mathbb{R}^n} \int_0^\infty (u(t,x) - u(t-s,x+y))K(s,y) \, \mathrm{d}s \, \mathrm{d}y : \text{for all K satsifying (1.4) and (1.5)} \right\}.$$

Note that the maximum and minimum above are typically achieved at different kernels K depending on the point (t,x), thus, for any smooth function u, M^+u and M^-u coincide with some Lu for a kernel $K(t,x,s,y)$.

If a smooth function u satisfies the equation (1.1) for some kernel $K(t,x,s,y)$ satisfying (1.4) and (1.5), then it also satisfies $M^+u \geq 0$ and $M^-u \leq 0$ in the same domain. This is because at every point (t,x), the kernel $K(t,x,\cdot,\cdot)$ is one candidate in the maximum and minimum defining M^+ and M^-.

Conversely, if for some smooth function u, we have $M^+u \geq 0$ and $M^-u \leq 0$ in some domain, then we can find two kernels K^1 and K^2, satisfying the assumptions (1.4) and (1.5), such that the corresponding operators L^1 and L^2 satisfy $L^1u \geq 0$ and $L^2u \leq 0$. It is not hard to see that there will be an intermediate kernel K (for example $K = ((-M^-u)K_1 + (M^+u)K_2)/(M^+u + M^-u))$ for which the corresponding operator satisfies $Lu = 0$.

Therefore, assuming that $Lu = 0$ for some operator L as in (1.1) with (1.4) and (1.5) is the same as assuming that $M^+u \geq 0$ and $M^-u \leq 0$. The technical advantage of the latter formulation is that the two inequalities can be defined in the viscosity sense, whereas it is hard to define the meaning of $Lu = 0$ for a given kernel K if u is not a smooth function.

In order for us to be able to compute the values classically in the definition of $M^+u(t,x)$ and $M^-u(t,x)$, we need some regularity of u at least from one side. More precisely, assume that there exists $A \in \mathbb{R}^n$ and $C > 0$ such that in a neighborhood of (t,x),

$$u(s,y) \leq u(t,x) + A \cdot (y-x) + C(|x-y|^2 + t - s).$$

Then, for any **symmetric** kernel K satisfying (1.4) and (1.5), we have

$$\begin{aligned} Lu(t,x) &= \int_{\mathbb{R}^n} \int_0^\infty (u(t,x) - u(t-s, x+y)) K(s,y) \, \mathrm{d}s \, \mathrm{d}y, \\ &= \frac{1}{2} \int_{\mathbb{R}^n} \int_0^\infty (2u(t,x) - u(t-s,x+y) - u(t-s,x-y)) K(s,y) \, \mathrm{d}s \, \mathrm{d}y, \\ &= \frac{1}{2} \int_{\mathbb{R}^n} \int_0^\infty (2u(t,x) - \dots)^+ K(s,y) \, \mathrm{d}s \, \mathrm{d}y - \frac{1}{2} \int_{\mathbb{R}^n} \int_0^\infty (\dots)^- K(s,y) \, \mathrm{d}s \, \mathrm{d}y \end{aligned}$$

In the first term, $(2u(t,x) - u(t-s,x+y) - u(t-s,x-y))^+ \leq C(|x-y|^2 + t - s)$ by the assumption, and then it is integrable. We do not have any bound for the second integral, so the value of $Lu(t,x)$ could be $-\infty$. There is no problem with this, it is just that the value of $Lu(t,x)$ is in $[-\infty, +\infty]$. The only case that we would be unable to compute the value of $Lu(t,x)$ is when both the positive and negative parts of the integrals are infinity and we end up with the undetermined difference $+\infty - \infty$. The quadratic control on one side prevents this to happen. The definition of viscosity solution that we give below evaluates the equation only at those points when it is possible to do it.

Definition 4.1. *Let $u : (-\infty, T) \to \mathbb{R}^n$ be a bounded function which is upper (resp. lower) semicontinuous in an* open *domain $D \subset (-\infty, T) \to \mathbb{R}^n$. We say $M^-u \leq B$ (resp. $M^+u \geq -B$) in D if the following happens. For every point (t,x) in D such that there exists $A \in \mathbb{R}^n$, $r > 0$ (small) and $C > 0$ (large) such that*

$$u(s,y) \leq u(t,x) + A \cdot (y-x) + C(|x-y|^2 + t - s) \qquad \text{for all } (s,y) \in Q_r(t,x). \qquad \text{(resp. } \geq\text{)}$$

Then $M^-u(t,x) \leq B$ (resp. $M^+u(t,x) \geq -B$).

This definition of viscosity solution looks somewhat unusual because the equation is evaluated in the original function u and not on smooth test functions φ which are tangent to u from one side. The definition is in fact equivalent to the usual one. This is a characteristic of nonlocal equations. See the discussion in [1] to understand this equivalence. Note that the points (t,x) for which we evaluate the operators M^+ and M^- are exactly those for which a smooth tangent function φ can be found touching the graph of u from either above or below respectively.

For equations which depend on time derivatives, we need to treat that term with the usual idea of viscosity solutions of evaluating the derivatives in the test functions.

Definition 4.2. *Let $u : (-\infty, T) \to \mathbb{R}^n$ be a bounded function which is upper (resp. lower) semicontinuous in an* open *domain $D \subset (-\infty, T) \to \mathbb{R}^n$. We say $u_t + M^-u \leq B$ (resp.$u_t + M^+u \geq -B$) in D if the following happens. For every point (t,x) in D such that there exists $V \in \mathbb{R}^n$, $z \in \mathbb{R}$, $r > 0$ (small) and $C > 0$ (large) such that*

$$u(s,y) \leq u(t,x) + A \cdot (y-x) + z(t-s) + C(|x-y|^2 + |t-s|^2) \qquad \text{for all } (s,y) \in Q_r(t,x). \qquad \text{(resp. } \geq\text{)}$$

Then $z + M^-u(t,x) \leq B$ (resp. $z + M^+u(t,x) \geq -B$).

Note that if u is a smooth function, then $\partial_t u + M^-u \leq B$ in the viscosity sense if and only if there exists a kernel K satisfying (1.4) and (1.5) such that

$$u_t(t,x) + Lu(t,x) \leq B$$

holds point-wise. Conversely, $\partial_t u + M^+ u \geq B$ if and only if there exist K such that

$$u_t(t,x) + Lu(t,x) \geq B.$$

Both inequalities hold at the same time if there exists a function f with $\|f\|_{L^\infty} \leq B$ such that

$$u_t(t,x) + Lu(t,x) = f(t,x).$$

If $u_t + M^- u \leq B$, we refer to u as a *subsolution*. If $\partial_t u + M^+ u \geq B$, we refer to u as a *supersolution*. A function for which both inequalities hold is called a solution.

In order to handle the scaling of the equation $u_t + Lu$, we also introduce the scaled version of M^+ and M^-.

$$M^+_\rho u(t,x) = \max \left\{ \int_{\mathbb{R}^n} \int_0^\infty (u(t,x) - u(t-s, x+y))K(s,y) \, ds \, dy : \text{for all K satsifying (2.2) and (2.3)} \right\},$$
$$M^-_\rho u(t,x) = \min \left\{ \int_{\mathbb{R}^n} \int_0^\infty (u(t,x) - u(t-s, x+y))K(s,y) \, ds \, dy : \text{for all K satsifying (2.2) and (2.3)} \right\}.$$

The definition of viscosity solutions applies to M^+_ρ and M^-_ρ is analogous.

5 Growth lemma - without time derivatives

We start by defining a cylinder Q_r with a scale and proportion which is compatible with our assumptions (1.4) and (1.5) and is convenient for the upcoming proofs.

Definition 5.1. *We write Q_r to denote the cylinder $(-\frac{c_1+c_2}{2} r^\beta, 0) \times B_r$. Also $Q_r(t,x) := (t,x) + Q_r$.*

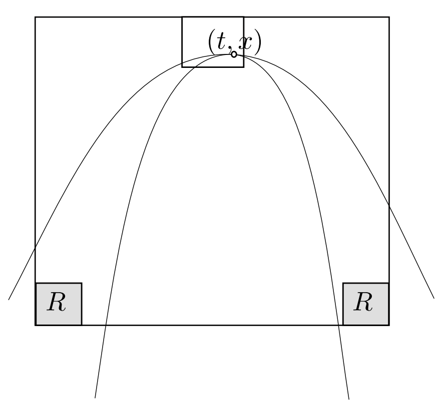
The following Lemma is a simple geometric observation. It is a technical result which will be used in the growth lemma.

Lemma 5.2. *There exists $r_0 > 0$ (depending on β, c_1 and c_2) such that for all (t,x) in Q_{2r_0}, the ring*

$$R = (-\frac{c_1+c_2}{2}, -\frac{c_1+c_2}{2} + r_0] \times (B_1 \setminus B_{1-r_0}) \tag{5.1}$$

is contained in the set

$$\left\{(s,y) : c_1|x-y|^\beta \leq (t-s) \leq c_2|x-y|^\beta\right\}. \tag{5.2}$$



Proof. We observe that the points (s,y) for which $s = -\frac{c_1+c_2}{2}|y|^\beta$ are in the interior of the open set

$$\{(s,y) : c_1|y|^\beta < s < c_2|y|^\beta\}.$$

Therefore, for (t,x) sufficiently close to $(0,0)$, those points will also be in the interior of the open set (5.2). In equivalent words, there is a neighborhood of $(0,0)$ (say Q_{r_1} for some $r_1 > 0$) so that for any $(t,x) \in Q_{r_1}$, the set (5.2) contains a neighborhood of $\{s = -\frac{c_1+c_2}{2}|y|^\beta\}$. In particular it contains

$$(-\frac{c_1+c_2}{2}, -\frac{c_1+c_2}{2} + r_2] \times (B_1 \setminus B_{1-r_2}),$$

for some $r_2 > 0$.

We conclude the proof of the lemma by choosing $r_0 = \min(r_1/2, r_2)$. □

Lemma 5.3. *There exist $\alpha > 0$ and $\varepsilon > 0$ sufficiently small such that the following holds. Let $u : \mathbb{R}^n \times (-\infty, 1) \to \mathbb{R}$ be a subsolution to $M^- u(t,x) \le \varepsilon$ in Q_1. Let r_0 and R be the ones from Lemma 5.2. Assume that*

$$u \le 2r_0^{-k\alpha} - 1 \qquad \text{in } Q_{r_0^{-k}} \text{ for } k = 0,1,2,\dots, \tag{5.3}$$

$$|\{u \le 0\} \cap R| \ge \mu. \tag{5.4}$$

Then $u \le (1-\theta)$ in Q_{r_0} for some $\theta > 0$ sufficiently small depending on n, σ, β, λ and Λ.

Proof. Let $b : (-\infty, 0] \times \mathbb{R}^n \to \mathbb{R}$ be a smooth function such that

- $b \ge 0$ everywhere.
- $b(t,x) = 0$ if $(t,x) \notin Q_{2r_0}$.
- $b(t,x) = 1$ if $(t,x) \in Q_{r_0}$.

We will show that if α and θ are sufficiently small, then $u \le 1 - \theta b$ in Q_1, which clearly implies the result in the Lemma.

The proof is by contradiction. The appropriate values of α and h will be chosen later. Let (t_0, x_0) be the point where $u + \theta b$ achieves its maximum in $\overline{Q_1}$. By the assumption that we are trying to contradict, we have that $u(t_0,x_0) + \theta b(t_0,x_0) > 1$. Thus, $(t_0,x_0) \in B_{2r_0}$, since otherwise $b = 0$ and $u \le 1$.

To obtain the contradiction, we match our estimates against the negativity of the integral. Indeed, on one hand the difference $u(t_0,x_0) - u(\cdot,\cdot)$ is smaller than $\theta(b(\cdot,\cdot) - b(t_0,x_0))$ in Q_1, and outside Q_1 the estimate (5.3) gives a control of the tails of the integral. On the other hand, $u \le 0$ in a substantial part of the domain of integration, which would make the integral too negative as we will see below.

Let $u(t_0,x_0) + \theta b(t_0,x_0) = W > 1$. By the choice of (t_0,x_0), we have that $u \le W - \theta b$ in Q_1. Since b is a smooth function, (t_0,x_0) is a point where the equation can be evaluated at the function u (recall Definition 4.1). Therefore, we have that $M^- u(t_0,x_0) \le 0$. In other words, there exists some kernel K satisfying (1.4) and (1.5) such that

$$\int_{\mathbb{R}^n} \int_0^\infty (u(t_0,x_0) - u(t_0-s, x_0+y)) K(s,y) \, \mathrm{d}s \, \mathrm{d}y \le \varepsilon.$$

We split the domain of integration first.

$$\int_{(t_0-s,x_0+y)\in Q_1} (u(t_0,x_0) - u(t_0-s,x_0+y))K(s,y) \, \mathrm{d}s \, \mathrm{d}y \\ + \sum_{k=0}^\infty \int_{(t_0-s,x_0+y)\in(Q_{k+1}\setminus Q_k)} (u(t_0,x_0) - u(t_0-s,x_0+y))K(s,y) \, \mathrm{d}s \, \mathrm{d}y \le \varepsilon.$$

We start by estimating the second term, which corresponds to the values so that $(t_0-s, x_0+y) \notin Q_1$. We call this the *tail of the integral.* We note that by the assumption (5.3), $u(t_0-s,x_0+y) \le 2r_0^{-\alpha k} - 1$

if $(t_0 - s, x_0 + y) \in Q_{r_0^{-k}}$, and $u(t_0, x_0) \geq 1 - \theta$. Therefore, if $(t_0 - s, x_0 + y) \notin Q_1$, we have

$$
\begin{aligned}
u(t_0, x_0) - u(t_0 - s, x_0 + y) &\geq 1 - \theta + 1 - 2r_0^{-\alpha k} \qquad \text{for } k \text{ s.t. } (t_0 - s, x_0 + y) \in Q_{r_0^{-k}} \\
&\geq -\theta - \left(1 - (C(|y| + s^{1/\beta}))^\alpha\right).
\end{aligned}
$$

Where $C = r_0^{-1}$. Note that this lower bound can be arbitrarily close to zero if we take θ and α small. We estimate the second term as

$$
\begin{aligned}
&\sum_{k=0}^{\infty} \int_{(t_0-s,x_0+y)\in(Q_{k+1}\setminus Q_k)} (u(t_0, x_0) - u(t_0 - s, x_0 + y))K_0(s, y) \,\mathrm{d}s \,\mathrm{d}y \\
&\qquad \geq \int_{Q_{1/4}^c} (-\theta + 2 - 2(C(|x| + s^{\frac{1}{\beta}}))^\alpha)\Lambda \left(\frac{1}{|y|^{n+\sigma+\beta}} + \frac{1}{s^{n/\beta+1+\sigma/\beta}}\right) \mathrm{d}s \,\mathrm{d}y \\
&\qquad \geq -C\theta - \delta(\alpha).
\end{aligned}
$$

Where $\delta(\alpha)$ can be made arbitrarily small by picking $\alpha \ll 1$.

For the fist term, we stress that by the choice of (t_0, x_0), $u(t_0, x_0) - u(t_0 - s, x_0 + y) \geq \theta(b(t_0 - s, x_0 + y) - b(t_0, x_0))$ in Q_1. Therefore, we can estimate the first term by

$$
\begin{aligned}
&\int_{(t_0-s,x_0+y)\in Q_1} (u(t_0, x_0) - u(t_0 - s, x_0 + y))K_0(s, y) \,\mathrm{d}s \,\mathrm{d}y \\
&\qquad\qquad \geq \int_{(t_0-s,x_0+y)\in Q_1} \theta(b(t_0 - s, x_0 + y) - b(t_0, x_0))K_0(s, y) \,\mathrm{d}s \,\mathrm{d}y \\
&\qquad\qquad \geq -C\theta
\end{aligned}
$$

In the last inequality we use that b is a smooth function and then $\int(b(t_0-s, x_0+y)-b(t_0, x_0))K_0(s, y) \,\mathrm{d}s \,\mathrm{d}y$ is a bounded function of (t_0, x_0). This bound depends on b, Λ and n. Recall that we assumed that K is symmetric in y.

However, we can improve the pointwise bound on $u(t_0-s, x_0+y)$ at those points where $(t_0-s, x_0+y) \in R$ and $u(t_0-s, x_0+y) \leq 0$. In the estimate above we used that $u(t_0, x_0)-u(t_0-s, x_0+y) \geq \theta(b(t_0-s, x_0+y)-b(t_0, x_0))$, whereas in this set we can use the better estimate $u(t_0, x_0)-u(t_0-s, x_0+y) \geq 1-\theta$. By assumption, the measure of this set is larger than μ and K is boudned below there by (1.4). We add the difference between these two bounds to the estimate above:

$$
\begin{aligned}
&\int_{(t_0-s,x_0+y)\in Q_1} (u(t_0, x_0) - u(t_0 - s, x_0 + y))K_0(s, y) \,\mathrm{d}s \,\mathrm{d}y \\
&\qquad \geq -C\theta + \int_{(t_0-s,x_0+y)\in\{u\leq 0\}\cap R} (1 - \theta - \theta(b(t_0 - s, x_0 + y) - b(t_0, x_0)))K_0(s, y) \,\mathrm{d}s \,\mathrm{d}y \\
&\qquad \geq -C\theta + c_0\mu
\end{aligned}
$$

Adding up the estimates, we obtain $-C\theta + c_0\mu - \delta(\alpha) \leq \varepsilon$. This is a contradiction if we choose ε, θ and α small enough, since the positive term in the middle is independent of both constants. □

Corollary 5.4. *There exists an $\alpha, \varepsilon \in (0, 1)$ (small enough depending on n, σ, β, λ and Λ) so that the following result holds. Let $u : \mathbb{R}^n \times (-\infty, 1) \to \mathbb{R}$ be a subsolution to $M^-u(t, x) \leq eps$ in Q_1. Assume that*

$$
\begin{aligned}
u &\leq 2r_0^{-k\alpha} - 1 \qquad \text{in } Q_{r_0^{-k}} \text{ for } k = 0, 1, 2, \ldots, \\
|\{u \leq 0\} \cap R| &\geq \mu.
\end{aligned}
$$

Then $u \leq 2r_0^\alpha - 1$ in Q_{r_0}.

Proof. The value of α which makes the statement of the corollary true is the minimum between the value of α of Lemma 5.3 and $\log_2(1-\theta)$. □

Lemma 5.5. *There exist $\alpha > 0$ and $\varepsilon > 0$ sufficiently small (depending on n, σ, β, λ and Λ) such that the following holds. Let $u : \mathbb{R}^n \times (-\infty, 1) \to \mathbb{R}$ be a viscosity solution to both $M^- u(t,x) \le \varepsilon$ and $M^+ u \ge -\varepsilon$ in Q_1. Assume that*

$$\operatorname*{osc}_{Q_{r_0^{-k}}} u \le C_0 r_0^{-k\alpha} \quad \text{for } k = 0, 1, 2, \dots$$

Then $\operatorname{osc}_{B_{r_0}} u \le C_0 r_0^{\alpha}$.

Proof. We use the same α as in Corollary 5.4.

Let $m = \min_{Q_1} u$ and $M = \max_{Q_1} u$. Either

$$|\{u \le \frac{m+M}{2}\} \cap R| \ge \frac{1}{2} R,$$

or

$$|\{u \ge \frac{m+M}{2}\} \cap R| \ge \frac{1}{2} R.$$

Let us assume the former (otherwise we do the same with $-u$ instead of u).

Let v be the normalized function

$$v(t,x) = \frac{2}{C_0}\left(u(t,x) - m\right) - 1.$$

It is easy to check that v satisfies the hypothesis of Corollary 5.4. Then, $v \le 2r_0^{-\alpha} - 1$ in Q_{r_0}. In terms of u, this means that $u \le C_0 r_0^{\alpha} + m$ in Q_{r_0}. Thus, since $u \ge m$ in $Q_{r_0} \subset Q_1$, we conclude the proof of the Lemma. □

Note that in the proof of Lemma 5.5 we are applying Corollary 5.4 to a normalized version of either u or $-u$. Even though we only need u to be a subsolution in Corollary 5.4, we use that u is a solution in Lemma 5.5, since we cannot say a priori which of the two alternatives will apply.

6 Growth lemma - with time derivatives

When we study the regularity of solutions to equations that involve time derivatives as in Theorem 1.2, we need to consider a different scaling, as explained in section 2. The proof of Lemma 5.3 in the previous section can be easily adapted at unit scale to equations that involve a time derivative. However, that proof is not invariant by the scaling, since it depends on the assumptions (1.4) and (1.5) and we cannot replace them by (2.2) and (2.3) without affecting the result.

We must introduce a different scaling and a different version of Lemma 5.3 that uses (2.2) and (2.3) for $\rho \ll 1$ instead of (1.4) and (1.5).

We start by defining the cylinder $\tilde{Q}_r$ with a this new scaling

Definition 6.1. *We write $\tilde{Q}_r$ to denote the cylinder $(-r^\sigma, 0) \times B_r$. Also $\tilde{Q}_r(t,x) := (t,x) + \tilde{Q}_r$.*

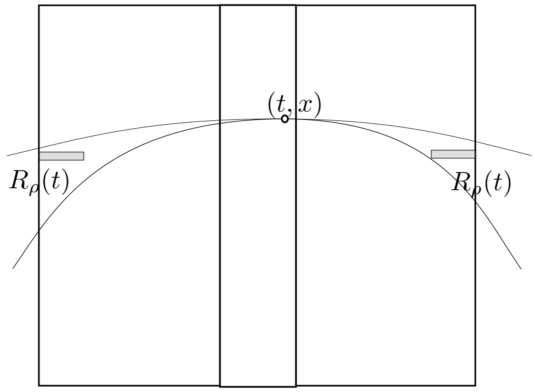
Lemma 6.2. *Let r_0 be small as in Lemma 5.2. There exist $d_1, d_2 > 0$ depending on c_1, c_2 and r_0 such that for any $\rho > 0$, $t \in \mathbb{R}$, and $x \in B_{r_0}$, the ring*

$$R_\rho(t) = (-d_2 \rho^{\beta-\sigma} + t, -d_1 \rho^{\beta-\sigma} + t) \times (B_1 \setminus B_{1-r_0}),$$

is contained in the set

$$\{(s,y) : \rho^{\beta-\sigma} c_1 |x-y|^\beta \le (t-s) \le \rho^{\beta-\sigma} c_2 |x-y|^\beta\}.$$

Proof. After a time translation of t and a scaling in time multiplying time $\rho^{\sigma-\beta}$, the Lemma reduces to Lemma 5.2 taking $t=0$ only, with $d_1 = (c_1+c_2)/2 - r_0$ and $d_2 = (c_1+c_2)/2$. □



It is important to realize that the hypothesis (2.3) gives a bound for $L\varphi$ for any smooth function φ that does not depend on ρ, which is justified in the following lemma.

Lemma 6.3. *Let L be a linear operator whose kernel satisfies the estimate* (2.3)*. Let φ be a bounded function, C^2 around x, such that $\varphi(t,\cdot) - \varphi(s,\cdot) \le C_0(t-s)$ for all $s<t$. Then*

$$L\varphi(t,x) \le C\Lambda C_0 \rho^\beta + \int_{\mathbb{R}^n} (\varphi(t,x) - \varphi(t,y))\tilde{K}(t,x,y)\, \mathrm{d}y.$$

where

$$\tilde{K}(t,x,y) = \int_0^\infty K(t,x,s,y)\, \mathrm{d}s.$$

Moreover, (2.2) *and* (2.3) *imply that*

$$\frac{c\lambda}{|y|^{n+\sigma}} \le \tilde{K}(t,x,y) \le \frac{C\lambda}{|y|^{n+\sigma}}.$$

for some constants c and C depending on c_1, c_2 and dimension, but not on ρ.

Remark 6.4. Note that the condition $\varphi(t,\cdot) - \varphi(s,\cdot) \le C_0(t-s)$ would be implied by $\varphi_t \le C_0$.

Proof. Without loss of generality, we will prove the Lemma for $(t,x) = (0,0)$. Recall that L has the form (1.1).

$$\begin{aligned}
L\varphi(0,0) &= \iint (\varphi(0,0) - \varphi(-s,y))K(0,0,s,y)\, \mathrm{d}s\, \mathrm{d}y, \\
&\le \iint (\varphi(0,0) - \varphi(0,y) - C_0 s)\, K(0,0,s,y)\, \mathrm{d}s\, \mathrm{d}y, \\
&= \iint C_0 s K(s,y)\, \mathrm{d}s\, \mathrm{d}y + \iint (\varphi(0,0) - \varphi(0,y))\, K(0,0,s,y)\, \mathrm{d}s\, \mathrm{d}y, \\
&= \iint C_0 s K(0,0,,s,y)\, \mathrm{d}s\, \mathrm{d}y + \int (\varphi(0,0) - \varphi(0,y))\, \tilde{K}(0,0,y)\, \mathrm{d}y,
\end{aligned}$$

In order to estimate the first term, we use (2.2).

$$\int_{B_1}\int_{\{s>0\}} C_0 s K(s,y)\ \mathrm{d}s\ \mathrm{d}y$$

$$\leq \int_{B_1} C_0 s \int_{\{s>0\}} \frac{\Lambda \rho^{n+2\sigma}}{|\rho y|^{n+\beta+\sigma} + (\rho^\beta s)^{n/\beta+1+\sigma/\beta}}\ \mathrm{d}s\ \mathrm{d}y,$$

using the formula $\int_0^\infty \frac{s\ \mathrm{d}s}{a+(bs)^\gamma} = Ca^{\frac{2-\gamma}{\gamma}} b^{-1}$, we get

$$= \Lambda C_0 \rho^\beta \int_{B_1} C|y|^{-n-\sigma+\beta}\ \mathrm{d}y = C\Lambda C_0 \rho^\beta$$

The estimates for $\tilde{K}$ follows by the elementary computation of $\int K\ \mathrm{d}s$ using (2.2) and (2.3). Indeed,

$$\int_{\{s>0\}} K(s,y)\ \mathrm{d}s \leq \int_{\{s>0\}} \frac{\Lambda \rho^{n+2\sigma}}{|\rho y|^{n+\beta+\sigma} + (\rho^\beta s)^{n/\beta+1+\sigma/\beta}}\ \mathrm{d}s,$$

using the formula $\int_0^\infty \frac{\mathrm{d}s}{a+(bs)^\gamma} = Ca^{\frac{1-\gamma}{\gamma}} b^{-1}$, the powers of ρ cancel out and we get

$$\leq \Lambda C|y|^{-n-\sigma}$$

A similar computation gives the other inequality. □

Corollary 6.5. *Let L be a linear operator whose kernel satisfies the estimate* (2.3). *Let φ be a bounded function, C^2 in a neighborhood of x, such that $\varphi(t,\cdot) - \varphi(s,\cdot) \leq C_0(t-s)$ for all $s < t$. Then*

$$L\varphi(t,x) \leq C(C_0\rho^\beta + \|\varphi\|_{C^2_x})\Lambda.$$

for a constant C independent of ρ.

In particular, $L\varphi$ is uniformly bounded for $\rho \in (0,1]$.

Proof. The Corollary follows from Lemma 6.3 since the second term is an integro-differential operator of order σ (as in [1] or [12]) which is bounded by the C^2_x norm of φ □

Corollary 6.6. *If φ is C^3_x and $\nabla\varphi$ is Lipschitz in time, then $M^+_\rho\varphi$ and $M^-_\rho\varphi$ are Lipschitz functions in x, independently of ρ as long as $\rho \in (0,1)$.*

Proof. Applying Corollary 6.5 to every directional derivative in x of φ, we see that $L\varphi$ is uniformly Lipschitz for all K satisfying (2.3). Thus, both $M^+_\rho\varphi$ and $M^-_\rho\varphi$ are Lipschitz since they are a supremum and an infimum of Lipschitz functions. □

We now proceed with the crucial growth lemma which is the heart of the C^α regularity proof. Recall that when we scale $\tilde{Q}_r$ to $\tilde{Q}_1$, the scaling of the equation makes the integral operator approximate an operator that is local in time. It is natural that this proof of the growth lemma below will borrow ideas from regularity results for nonlocal equations in space but local in time. The idea of this proof was inspired by [13].

Lemma 6.7. *There exist $\alpha > 0$ and $r_0, \rho, \varepsilon > 0$ sufficiently small such that the following holds. Let $u : \mathbb{R}^n \times (-\infty, 1) \to \mathbb{R}$ be a subsolution to $u_t + M^-_r u(t,x) \leq \varepsilon$ in $\tilde{Q}_1$. Let $R = (-3/4, -1/4) \times (B_1 \setminus B_{1-r_0})$. Assume that*

$$u \leq 2r_0^{-k\alpha} - 1 \qquad \text{in } \tilde{Q}_{r_0^{-k}} \text{ for } k = 0,1,2,\ldots, \tag{6.1}$$

$$|\{u \leq 0\} \cap R| \geq \mu. \tag{6.2}$$

Then $u \leq (1-\theta)$ in $\tilde{Q}_{r_0}$ for some $\theta > 0$ sufficiently small depending on n, σ, β, λ and Λ, but not on ρ.

Proof. We take r_0 to be the minimum between the value from Lemma 6.2 and $8^{-1/\sigma}$.

Let $\tilde{b} : \mathbb{R}^n \to \mathbb{R}$ be a smooth function such that

- $\tilde{b} \geq 0$ everywhere.
- $\tilde{b}(x) = 0$ if $x \notin B_{2r_0}$.
- $\tilde{b}(x) = 1$ if $x \in B_{r_0}$.

We will show that the function u stays below the function

$$b(x,t) = \begin{cases} 1 + \xi + \delta(t+1) - m(t)\tilde{b}(x) & \text{in}[-3/4, 0] \times B_1, \\ 1 + \xi & \text{in } [-1, -3/4] \times B_1, \\ 2r_0^{-k\alpha} - 1 + \xi & \text{in } \tilde{Q}_{r_0^{-k}} \setminus \tilde{Q}_{r_0^{-k+1}} \text{ for } k = 1, 2, \dots \end{cases} \tag{6.3}$$

where $\xi > 0$ is an arbitrarily small constant, $\delta > 0$ will be chosen below and m is the solution to the ODE:

$$\begin{aligned} &m(-3/4) = 0, \\ &\quad m'(t) = c_0 \rho^{\sigma-\beta} |\{x \in R_\rho(t) \cap \tilde{Q}_1 : u(x,t) \leq 0\}| - C_1 m(t) \text{ for } t > -3/4. \end{aligned}$$

for constants c_0 and C_1 to be chosen later.

The ODE for m can be solved explicitly. Indeed, for any $t \geq -3/4$,

$$m(t) = c_0 \rho^{\sigma-\beta} \int_{-3/4}^{t} |\{x \in R_\rho(s) \cap \tilde{Q}_1 : u(x,s) \leq 0\}| e^{-C_1(t-s)} \,\mathrm{d}s.$$

Assuming that r_0 and ρ are small enough, we can relate the values of $m(t)$ for $t > -r_0$ to μ. Indeed, for $t > -r_0$, and assuming that $d_2 \rho^\beta < 1/4 - r_0$,

$$\begin{aligned} m(t) &\geq c_0 e^{-C_1} \rho^{\sigma-\beta} \int_{-3/4}^{t} |\{x \in R_\rho(t) \cap \tilde{Q}_1 : u(x,t) \leq 0\}| \,\mathrm{d}s, \\ &= c_0 e^{-C_1} \rho^{\sigma-\beta} \int_{-3/4}^{t} \int_{-d_2\rho^{\beta-\sigma}+t}^{-d_1\rho^{\beta-\sigma}+t} \int_{B_1 \setminus B_{1-r_0}} \chi_{\tilde{Q}_1 \cap \{u \leq 0\}}(\zeta, y) \,\mathrm{d}y \,\mathrm{d}\zeta \,\mathrm{d}s, \\ &\geq c_0 e^{-C_1} \rho^{\sigma-\beta} \int_{B_1 \setminus B_{1-r_0}} \int_{-3/4}^{-1/4} \int_{\zeta+d_1\rho^{\beta-\sigma}}^{\zeta+d_2\rho^{\beta-\sigma}} \chi_{\tilde{Q}_1 \cap \{u \leq 0\}}(\zeta, y) \,\mathrm{d}s \,\mathrm{d}\zeta \,\mathrm{d}y, \\ &= c_0 e^{-C_1} (d_2 - d_1) \int_{B_1 \setminus B_{1-r_0}} \int_{-3/4}^{-1/4} \chi_{\tilde{Q}_1 \cap \{u \leq 0\}}(\zeta, y) \,\mathrm{d}\zeta \,\mathrm{d}y, \\ &\geq c_0 e^{-C_1} (d_2 - d_1) \mu. \end{aligned}$$

Therefore, by showing that $u \leq b$, we prove the lemma as long as δ is small enough.

The inequality $u(t,x) < b(t,x)$ follows directly from our assumptions anywhere outside $(-3/4, 0] \times B_1$. We show that it holds for $t > -3/4$ and $x \in B_1$ by proving that it can never be invalidated for the first time. Indeed, assume there was a point (t_0, x_0) where equality holds. This point must be in the support of $\tilde{b}$ (strict inequality holds in the rest since $u \leq 1$ in $\tilde{Q}_1$), thus $x_0 \in B_{2r_0}$.

Let $u(t_0, x_0) = b(t_0, x_0)$. By the choice of (t_0, x_0), we have that $u \leq b$ in Q_1 whereas $u(t_0, x_0) = b(t_0, x_0)$. To apply the definition of viscosity subsolution, we must check that $b_t(t_0, x_0) + M^- u(t_0, x_0) \leq \varepsilon$.

From the definition of M_ρ^-, there exists a kernel K satisfying (2.2) and (2.3) such that

$$b_t(t_0, x_0) + \int_{\mathbb{R}^n} \int_0^\infty (u(t_0, x_0) - u(t_0 - s, x_0 + y)) K(s, y) \,\mathrm{d}s \,\mathrm{d}y \leq \varepsilon.$$

We have the simple inequality

$$b_t(t_0, x_0) = -m'(t_0)\tilde{b}(x_0) + \delta.$$

Replacing the value of $b_t(t_0, x_0)$ above we obtain

$$\int_{\mathbb{R}^n} \int_0^\infty (u(t_0, x_0) - u(t_0 - s, x_0 + y))K(s, y) \,\mathrm{d}s \,\mathrm{d}y \leq m'(t_0)\tilde{b}(x_0) + \varepsilon - \delta. \tag{6.4}$$

In order to get a contradiction, we estimate the integral on the left hand side from below as in the proof of Lemma 5.3.

We have that by the choice of (t_0, x_0), $u(t_0, x_0) - u(t_0 - s, x_0 + y) \geq b(t_0, x_0) - b(t_0 - s, x_0 + y)$ in Q_1. Therefore, we can estimate the $Lu(t_0, x_0)$ by

$$\begin{aligned} Lu(t_0, x_0) &= \int (u(t_0, x_0) - u(t_0 - s, x_0 + y))K_0(s, y) \,\mathrm{d}s \,\mathrm{d}y \\ &\geq \int (b(t_0, x_0) - b(t_0 - s, x_0 + y))K_0(s, y) \,\mathrm{d}s \,\mathrm{d}y \\ &= Lb(t_0, x_0) \end{aligned}$$

We can improve the pointwise bound on $u(t_0 - s, x_0 + y)$ at those points where $(t_0 - s, x_0 + y) \in R_\rho(t)$ and $u(t_0 - s, x_0 + y) \leq 0$. In the previous estimate above we used that $u(t_0, x_0) - u(t_0 - s, x_0 + y) \geq (b(t_0, x_0) - b(t_0 - s, x_0 + y))$, whereas in this set we can use the better estimate $u(t_0, x_0) - u(t_0 - s, x_0 + y) \geq b(t_0, x_0)$. From (2.2) K is boudned below there. We add the difference $(b(t_0 - s, x_0 + y))$ between these two bounds to the estimate above:

$$Lu(t_0, x_0) - Lb(t_0, x_0) \geq \int_{(t_0 - s, x_0 + y) \in \{u \leq 0\} \cap R_\rho(t) \cap \tilde{Q}_1} b(t_0 - s, x_0 + y)K_0(s, y) \,\mathrm{d}s \,\mathrm{d}y,$$

using that $b \geq 1/2$ everywhere,

$$\geq c_0 \rho^{\sigma - \beta} |\{x \in R_\rho(t) \cap \tilde{Q}_1 : u(x, t) \leq 0\}|$$

for a constant c_0 depending on λ, but not on ρ.

Plugging this estimate into the inequality (6.4), we obtain

$$\begin{aligned} Lb(t_0, x_0) + c_0 \rho^{\sigma - \beta} |\{x \in R_\rho(t) \cap \tilde{Q}_1 : u(x, t) \leq 0\}| &\leq m'(t_0)\tilde{b}(x_0) + \varepsilon - \delta, \\ &\leq \Big(c_0 \rho^{\sigma - \beta} |\{x \in R_\rho(t) \cap \tilde{Q}_1 : u(x, t) \leq 0\}| - C_1 m(t)\Big) b(x_0) + \varepsilon - \delta. \end{aligned}$$

Since $b \leq 1$, we cancel out the obvious terms to obtain

$$Lb(t_0, x_0) \leq -C_1 m(t_0)\tilde{b}(x_0) + \varepsilon - \delta. \tag{6.5}$$

In $\tilde{Q}_1$, the function b is smooth in x and Lipchitz in t. Since $(t_0, x_0) \in (-3/4, 0] \times B_{2r_0}$, for $\alpha \leq \sigma$ we can apply Lemma 6.3 and obtain

$$Lb(t_0, x_0) \geq \tilde{-} C c_0 \rho^\beta + Lb(t_0, x_0).$$

The operator $\tilde{L}$ is the integro-differential operator with kernel $\tilde{K}$ described in Lemma 6.3. It depends on the values of $b(t_0, \cdot)$ only. Moreover, the estimates on $\tilde{K}$ given on Lemma 6.3 are independent of ρ. We estimate by a direct computation

$$\tilde{L}b(t_0, x_0) \geq -m(t_0)\tilde{L}\tilde{b}(x_0) - \eta(\alpha).$$

Here $\eta(\alpha)$ is a bound on the contribution of the points $y \notin B_1$ in the integral representation of $\tilde{L}b(t_0, x_0)$. The value of $\eta(\alpha)$ is arbitrarily small as $\alpha \to 0$ (this is the same computation, but only in space, as in Lemma 5.3, it is also done explicitly in [13]). We conclude that

$$Lb(t_0, x_0) \geq -\eta(\alpha) - Cm(t_0)\tilde{L}\tilde{b}(x_0) - Cc_0\rho^\beta.$$

Therefore, recalling (6.5),

$$-\eta(\alpha) - Cm(t_0)\tilde{L}\tilde{b}(x_0) - Cc_0\rho^\beta \leq -C_1 m(t_0)\tilde{b}(x_0) + \varepsilon - \delta.$$

We choose $\delta = \varepsilon + \eta(\alpha) + Cc_0\rho^\beta$. Note that the choice of c_0, which depends only on dimension and λ, and C_1, which will be chosen below, are independent of ρ, α and ε. We now have

$$-m(t_0)\tilde{L}\tilde{b}(x_0) \leq -C_1 m(t_0)\tilde{b}(x_0).$$

Let M_x^- be the *monster Pucci operator* which is defined in [12] and [1]. Since $-M_x^+\tilde{b}(x_0) \leq \tilde{L}\tilde{b}(x_0)$, we have

$$m(t_0)M_x^+\tilde{b}(x_0) \leq -C_1 m(t_0)\tilde{b}(x_0).$$

This is clearly possible if we know a lower bound for $\tilde{b}(x_0)$ and ρ is small. However, we must also consider that x_0 may be a point where $\tilde{b}$ is very small. The final trick is to note that since $M_x^+\tilde{b} > 0$ wherever $\tilde{b} = 0$ and $M_x^+\tilde{b}$ is a Lipchitz function, then $M_x^+\tilde{b} > 0$ as long as $\tilde{b}$ is smaller than some constant b_0. Therefore, a those values of x_0 we get a contradiction immediately regardless of the choice of C_1. If $\tilde{b}(x_0) > b_0$, then we can choose C_1 large to get a contradiction.

We proved that $u \leq b$ everywhere. Recall that b is given by (6.3). We have $\tilde{b} \equiv 1$ in B_{r_0} and that $m(t) \geq c\mu$ for $t \geq -r_0$, where c depends on c_0, C_1, d_1 and d_2 only. So, if δ (which depends on α and ε only) is small enough, then we finish the proof. □

Arguing as in section 5, we derive the following two results from Lemma 6.7.

Corollary 6.8. *There exists an $\alpha, \varepsilon, \rho \in (0,1)$ (small enough depending on n, σ, β, λ and Λ) so that the following result holds. Let $u : \mathbb{R}^n \times (-\infty, 1) \to \mathbb{R}$ be a subsolution to $u_t + M_\rho^- u(t,x) \leq eps$ in Q_1. Assume that*

$$u \leq 2r_0^{-k\alpha} - 1 \qquad \text{in } \tilde{Q}_{r_0^{-k}} \text{ for } k = 0, 1, 2, \ldots,$$
$$|\{u \leq 0\} \cap R| \geq \mu,$$

where R is as in Lemma 6.7. Then $u \leq 2r_0^\alpha - 1$ in $\tilde{Q}_{r_0}$.

Lemma 6.9. *There exist $\rho > 0$, $\alpha > 0$ and $\varepsilon > 0$ sufficiently small (depending on n, σ, β, λ and Λ) such that the following holds. Let $u : \mathbb{R}^n \times (-\infty, 1) \to \mathbb{R}$ be a viscosity solution to both $u_t + M_\rho^- u(t,x) \leq \varepsilon$ and $u_t + M_\rho^+ u \geq -\varepsilon$ in Q_1. Assume that*

$$\operatorname*{osc}_{Q_{r_0^{-k}}} u \leq C_0 r_0^{-k\alpha} \quad \text{for } k = 0, 1, 2, \ldots$$

Then $\operatorname{osc}_{B_{r_0}} u \leq C_0 r_0^\alpha$.

7 Hölder regularity

The following theorems are a more precise restatement of Theorem 1.1 and Theorem 1.2 in terms of viscosity solutions. These restatements makes sense even when u is not smooth enough for the operator Lu to be computable classically.

Theorem 7.1. *Let $u : (-\infty, 0) \times \mathbb{R}^n \to \mathbb{R}$ be a bounded function, continuous in Q_1, such that $M^+u \geq -B$ and $M^-u \leq B$ in the viscosity sense in Q_1 for some constant $B \geq 0$.*

Then $u \in C^\alpha((-1/2, 0] \times B_{1/2})$ for some $\alpha > 0$ with an estimate

$$\|u\|_{C^\alpha((-1/2,0]\times B_{1/2})} \leq C\left(\|u\|_{L^\infty((-\infty,0)\times\mathbb{R}^n)} + B\right)$$

for some C and $\alpha > 0$ depending on n, σ, β, λ and Λ.

Proof. The proof follows by a more or less standard iterative application of Lemma 5.5. We prove that a Hölder modulus of continuity applies at the origin $(0, 0)$. The same argument centered at other points gives the estimate of the Theorem.

We point out that we can assume without loss of generality that $\|u\|_{L^\infty} = 1$ and $B \leq \varepsilon$. Otherwise, we replace u by the normalized function

$$\frac{1}{\|u\|_{L^\infty} + B/\varepsilon}.$$

We prove by induction in k that

$$\operatorname*{osc}_{Q_{r_0^k}} u \leq 2r_0^{\alpha k}. \tag{7.1}$$

This clearly implies the Hölder modulus of continuity at $(0, 0)$.

For $k \leq 0$, we have that $\operatorname{osc}_{Q_{r_0^k}} u \leq 2$ since $\|u\|_{L^\infty} = 1$, and this is enough to obtain (7.1). This is the base case of the induction.

Now, assume we know (7.1) holds up to some value of k. We will show that it holds for $k + 1$ as well. For that we apply Lemma 5.5 to the rescaled function

$$v(t, x) = r_0^{-k\alpha} u(r_0^{-\beta k} t, r_0^{-k} x),$$

with $C_0 = 2$. The application of Lemma 5.5 provides (7.1) for $k + 1$. □

Theorem 7.2. *Let $u : (-\infty, 0] \times \mathbb{R}^n \to \mathbb{R}$ be a bounded function, continuous in $\tilde{Q}_1$, such that $u_t + M^+u \geq -B$ and $u_t + M^-u \leq B$ in the viscosity sense in $\tilde{Q}_1$ for some constant $B \geq 0$.*

Then $u \in C^\alpha((-1/2, 0] \times B_{1/2})$ for some $\alpha > 0$ with an estimate

$$\|u\|_{C^\alpha((-1/2,0]\times B_{1/2})} \leq C\left(\|u\|_{L^\infty((-\infty,0)\times\mathbb{R}^n)} + B\right)$$

for some C and $\alpha > 0$ depending on n, σ, β, λ and Λ.

Proof. We first scale the equation by considering $u(\rho^\sigma t, \rho x)$ instead of u for some ρ small enough. Then, the proof of Theorem 7.2 is identical to the proof of Theorem 7.1 but replacing every Q_r by $\tilde{Q}_r$ and the application of Lemma 5.5 by Lemma 6.9. Note that every rescaling to $\tilde{Q}_r$ in the iteration makes the scale ρ even smaller. □

8 Future directions

The constants C and α in Theorem 1.1 depend on β and σ. We did not prove that in the asymptotic regimes described in section 3 there is a uniform choice of constants C and α which passes to the limit. In fact, our method does not provide such uniform estimates in terms of σ. For parabolic equations in nondivergence form, as those obtained in the asymptotic limits, these Hölder estimates are known to hold true (this was proved by Krylov and Safonov in [6]). The pursue of uniform estimates with constants C and α which do not deteriorate as we pass to the limit to obtain classical parabolic equations will be the subject of future research. We anticipate that in order to have an estimate that passes to the limit, the proof must contain some form of the parabolic ABP estimate, in the spirit of the one from [7]. One of the main differences with the proof in [7] would be that the convex envelope that they consider is automatically Lipschitz in time, whereas for type of equations in this paper it would not be the case. This would force a different scaling in the building blocks of the L^ε estimate.

It would be very interesting to understand the case in which the kernel $K(s,y)\mathrm{d}s\mathrm{d}y$ is replaced by a singular measure $\mathrm{d}\mu(s,y)$. The model case is when the measure is supported entirely on the parabola $s = |y|^2$ and the operator L has the form

$$Lu(t,x) = \int_{\mathbb{R}^n} (u(t,x) - u(t-|y|^2, x+y)) \frac{a(t,x,y)}{|y|^{n+\sigma}} \, \mathrm{d}y,$$

where $a(t,x,y)$ is bounded from above and below.

In future work, we would also like to address the nonlinear equations arising from stochastic control problems and stochastic games related to CTRW.

Acknowledgements

Luis Caffarelli and Luis Silvestre were partially supported by an NSF FRG project (DMS-1065926 and DMS-1065979)

Luis Caffarelli was also partially supported by NSF grant DMS-1160802. Luis Silvestre was also partially supported by NSF grant DMS-1001629 and the Sloan fellowship.

References

[1] L. Caffarelli and L. Silvestre. Regularity theory for fully nonlinear integro-differential equations. *Communications on Pure and Applied Mathematics*, 62(5):597–638, 2009.

[2] FusionWiki. Continuous time random walk, 2009.

[3] VM Kenkre and RS Knox. Generalized-master-equation theory of excitation transfer. *Physical Review B*, 9(12):5279, 1974.

[4] VM Kenkre, EW Montroll, and MF Shlesinger. Generalized master equations for continuous-time random walks. *Journal of Statistical Physics*, 9(1):45–50, 1973.

[5] J. Klafter, A. Blumen, and MF Shlesinger. Stochastic pathway to anomalous diffusion. *Physical Review A*, 35(7):3081, 1987.

[6] N. V. Krylov and M. V. Safonov. A property of the solutions of parabolic equations with measurable coefficients. *Izv. Akad. Nauk SSSR Ser. Mat.*, 44(1):161–175, 239, 1980.

[7] H.A.C. Lara and G. Dávila. Regularity for solutions of non local parabolic equations. *Arxiv preprint arXiv:1109.3247*, 2011.

[8] R. Metzler and J. Klafter. The random walk's guide to anomalous diffusion: a fractional dynamics approach. *Physics Reports*, 339(1):1–77, 2000.

[9] R. Metzler and J. Klafter. The restaurant at the end of the random walk: recent developments in the description of anomalous transport by fractional dynamics. *Journal of Physics A: Mathematical and General*, 37:R161, 2004.

[10] H. Scher and E.W. Montroll. Anomalous transit-time dispersion in amorphous solids. *Physical Review B*, 12(6):2455, 1975.

[11] M.F. Shlesinger, J. Klafter, and YM Wong. Random walks with infinite spatial and temporal moments. *Journal of Statistical Physics*, 27(3):499–512, 1982.

[12] L. Silvestre. Holder estimates for solutions of integro-differential equations like the fractional laplace. *Indiana University mathematics journal*, 55(3):1155–1174, 2006.

[13] L. Silvestre. On the differentiability of the solution to the hamilton–jacobi equation with critical fractional diffusion. *Advances in mathematics*, 226(2):2020–2039, 2011.

[14] B.P. van Milligen, I. Calvo, and R. Sánchez. Continuous time random walks in finite domains and general boundary conditions: some formal considerations. *Journal of Physics A: Mathematical and Theoretical*, 41:215004, 2008.

[15] B.P. van Milligen, BA Carreras, and R. Sánchez. Uphill transport and the probabilistic transport model. *Physics of plasmas*, 11:3787, 2004.

[16] B.P. Van Milligen, R. Sanchez, and BA Carreras. Probabilistic finite-size transport models for fusion: Anomalous transport and scaling laws. *Physics of Plasmas*, 11:2272, 2004.