# ANALYSIS & PDE

Volume 9 No. 3 2016



RUSSELL W. SCHWAB AND LUIS SILVESTRE


## REGULARITY FOR PARABOLIC INTEGRO-DIFFERENTIAL EQUATIONS WITH VERY IRREGULAR KERNELS


msp

ANALYSIS AND PDE
Vol. 9, No. 3, 2016
dx.doi.org/10.2140/apde.2016.9.727

# REGULARITY FOR PARABOLIC INTEGRO-DIFFERENTIAL EQUATIONS WITH VERY IRREGULAR KERNELS

RUSSELL W. SCHWAB AND LUIS SILVESTRE

We prove Hölder regularity for a general class of parabolic integro-differential equations, which (strictly) includes many previous results. We present a proof that avoids the use of a convex envelope as well as give a new covering argument that is better suited to the fractional order setting. Our main result involves a class of kernels that may contain a singular measure, may vanish at some points, and are not required to be symmetric. This new generality of integro-differential operators opens the door to further applications of the theory, including some regularization estimates for the Boltzmann equation.



## 1. Introduction

We study the Hölder regularity for solutions of integro-differential equations of the form

$$u_t + b(x,t)\cdot\nabla u - \int_{\mathbb{R}^d} \big(u(x+h,t)-u(x,t)\big)K(x,h,t)\,\mathrm{d}h = f(x,t). \tag{1-1}$$

The integral may be singular at the origin and must be interpreted in the appropriate sense. These equations now appear in many contexts. Most notably, they appear naturally in the study of stochastic processes with jumps, which traditionally has been the main motivation for their interest. In the same way that pure jump processes contain the class of diffusions (processes with continuous paths) as particular limiting cases, (1-1) contains the usual second-order parabolic equations as particular limiting cases. This is due to the fact

L. Silvestre was partially supported by NSF grants DMS-1254332 and DMS-1065979.
*MSC2010:* 47G20, 35R09.
*Keywords:* nonlocal equations, nonsymmetric kernels, covering lemma, crawling ink spots, regularity.

that the integral term becomes a second-order operator $a_{ij}(x,t)\,\partial_{ij}u$ as the order $\alpha$ (to be defined below) converges to 2. We note that the simplest choice of $K$ is $K(h) = C_{d,\alpha}|h|^{-d-\alpha}$, which results in the equation

$$u_t + (-\Delta)^{\alpha/2}u = 0,$$

and converges to the usual heat equation $u_t - \Delta u = 0$ as $\alpha \to 2$ (recall that $(-\Delta)^{\alpha/2}$ is the operator whose Fourier symbol is $|\xi|^\alpha$).

The Hölder estimates that we obtain in this article are an integro-differential version of the celebrated result by Krylov and Safonov [1980] for parabolic equations with measurable coefficients. There are, in fact, several versions of these Hölder estimates for integro-differential equations, which were obtained in the last 10 years, and we briefly review them in Section 1A. Besides the elliptic/parabolic distinction, the difference between each version of the estimates is in the level of generality in the possible choices of the kernels $K(x,h,t)$. In this article, we obtain the estimates for a very generic class of kernels $K$, including nearly all previous results of this type.

The most common assumption in the literature is that for all $x$ and $t$, the kernel $K$ is comparable pointwise in terms of $h$ to the kernel for the fractional Laplacian. More precisely,

$$(2-\alpha)\frac{\lambda}{|h|^{d+\alpha}} \le K(x,h,t) \le (2-\alpha)\frac{\Lambda}{|h|^{d+\alpha}}. \tag{1-2}$$

This is often accompanied by the symmetry assumption $K(x,h,t) = K(x,-h,t)$. It is important for the applications of these estimates that no regularity condition may be assumed for $K$ with respect to $x$ or $t$.

In this paper, we only assume a much weaker version of (1-2). The upper bound for $K$, in (1-2), is relaxed to hold only in average when we integrate all the values of $h$ on an annulus, and it appears as assumption (A2). Also, for our work, the lower bound in (1-2) only needs to hold in a subset of values of $h$ that has positive density, given as assumption (A3). We also make an assumption, (A4), which says that the odd part of $K$ is under control if $\alpha$ is close to 1. The exact conditions are listed in Section 2. We prove that solutions of (1-1) are uniformly Hölder continuous, which we state in an informal way here and revisit more precisely in Section 7.

**Theorem 1.1.** *Let $u$ solve* (1-1). *Assume that for every $x \in B_1$ and $t \in [-1,0]$, the kernel $K(x,\cdot,t)$ satisfies the assumptions* (A1), (A2), (A3) *and* (A4) *in Section 2. Assume also that $f$ is bounded, $b$ is bounded, and for $\alpha < 1$, we have $b \equiv 0$. Then for some $\gamma > 0$,*

$$[u]_{C^\gamma(Q_{1/2})} \le C(\|u\|_{L^\infty(\mathbb{R}^d\times[-1,0])} + \|f\|_{L^\infty(Q_1)}).$$

The constants $C$ and $\gamma$ depend on the constants $\mu$, $\lambda$ and $\Lambda$ in (A1)–(A4), on the dimension $d$, on a lower bound for $\alpha$ (in particular, $\alpha$ can be arbitrarily close to 2), and on $\|b\|_{L^\infty}$.

Our purpose in developing Theorem 1.1 is not merely for the sake of generalization. An estimate with the level of generality given here can be used to obtain a priori estimates for the homogeneous Boltzmann equation. This is a novel application. None of the previous Hölder estimates for integral equations are appropriate to be applied to the Boltzmann equation.

As a byproduct of our proof of Theorem 1.1, we simplify and clarify some of the details regarding parabolic covering arguments (see the crawling ink spots of Section 6) as well as present a proof that does not invoke a convex envelope. Rather, we circumvent the oft-used gradient mapping of the convex envelope by using a mapping that associates points via their correspondence through parameters in an inf-convolution, modeled on the arguments of [Imbert and Silvestre 2013a], originating in [Cabré 1997; Savin 2007].

In Section 8, we apply this result to derive the $C^{1,\alpha}$ regularity for the parabolic Isaacs equation. This is a rather standard application of Hölder estimates for equations with rough coefficients, as in Theorem 1.1.

**1A.** ***Comparison with previous results and some discussion of* (1-1).** The Hölder estimates for integro-differential equations that take the form of (1-1) are a fractional-order version of the classical theorem by Krylov and Safonov [1980]. This is a fundamental result in the study of regularity properties of parabolic equations in nondivergence form, and has consequences for many aspects of the subsequent PDE theory. The classical theorem of De Giorgi, Nash and Moser concerns second-order parabolic equations in *divergence form*, in contrast with the theorem of Krylov and Safonov. The basic results for integro-differential equations in divergence form were developed earlier, and a small survey of this subject can be found in [Kassmann and Schwab 2014].

The simplest case of $K$ would be $K(h) = (2-\alpha)|h|^{-d-\alpha}$, and this choice gives the operator $Lu(x) = -C_{d,\alpha}(-\Delta)^{\alpha/2}u(x)$, which is a multiple of the fractional Laplacian of order $\alpha$ (the operator whose Fourier symbol is $|\xi|^\alpha$). This operator (and its inverse, the Riesz potential of order $\alpha$) have a long history, and have been fundamental to potential theory for about a century; see, for example, Landkof's book [1966]. In fact, the appearance of nonlocal operators similar to the one in (1-1) is in some sense generic among all linear operators that satisfy the positive global maximum principle (that is, the operator is nonpositive whenever it is evaluated at a positive maximum of a $C^2$ function). This has been known since the work of Courrège [1965]. He proved that any linear operator with the positive maximum principle must be of the form

$$Lu(x)=-c(x)u(x)+b(x)\cdot\nabla u(x)+\mathrm{Tr}\big(A(x)D^2u(x)\big)+\int_{\mathbb{R}^d}\big(u(x+h)-u(x)-\mathbb{1}_{B_1}(h)\nabla u(x)\cdot h\big)\mu(x,\mathrm{d}h),$$

where $c \ge 0$ is a function, $A \ge 0$ is a matrix, $b$ is a vector, all of $A$, $b$, $c$ are bounded, and $\mu(x,\cdot)$ is a Lévy measure that satisfies

$$\sup_x \int_{\mathbb{R}^d} \min(|h|^2, 1)\mu(x,\mathrm{d}h) < +\infty.$$

Heuristically from the point of view of jump-diffusion stochastic processes, $b$ records the drift, $A$ records the local covariance (or $\sqrt{A}$ is the diffusion matrix), and $\mu$ records the jumps.

The first Hölder regularity result for an equation of the form (1-1) was obtained in [Bass and Levin 2002a]. In that paper, the authors consider the elliptic equation ($u$ constant in time), with symmetric kernels satisfying the pointwise bound (1-2) and without drift. Their proof uses probabilistic techniques involving a related Markov (pure jump) stochastic process. Other results using probabilistic techniques were [Bass and Kassmann 2005; Song and Vondraček 2004], where different assumptions on the kernels are considered. The first purely analytical proof was given in [Silvestre 2006]. This first generation of results consists only of elliptic problems. They are not *robust* in the sense that as order approaches 2, the constants in the estimates blow up (hence they do not recover the known second-order results). Furthermore, they all require a pointwise bound below for the kernels as in (1-2).

The first robust Hölder estimate for the elliptic problem was obtained in [Caffarelli and Silvestre 2009], which means that the estimate they proved has constants that do not blow up as the order $\alpha$ of the equation goes to 2. In that sense, it is the first true generalization of the theorem of Krylov and Safonov. It was the first of the series of papers [Caffarelli and Silvestre 2009; 2011a; 2011b] recreating the regularity theory for fully nonlinear elliptic equations in the nonlocal setting. As above, these results are only for the elliptic problem, and they require symmetric kernels that satisfy the pointwise assumption (1-2).

The first estimate for parabolic integro-differential equations, *in nondivergence form*, appeared, to the best of our knowledge, in [Silvestre 2011] (the divergence case had some earlier results such as [Bass and Levin 2002b; Chen and Kumagai 2003]). In this case, the kernels are symmetric and satisfy (1-2) with $\alpha = 1$. The focus of [Silvestre 2011] is on the interaction between the integro-differential part and the drift term. The proof can easily be extended to arbitrary values of $\alpha$, but the estimate is not robust (it blows up as $\alpha \to 2$), and the details of this proof are explained in the lecture notes by one of the authors [Silvestre 2012b]. It is even possible to extend this proof to kernels that satisfy the upper bound in average like in our assumption (A2) below (see [Silvestre 2014b]). However, the estimates are not robust, and the lower bound in (1-2) is required.

The first robust estimate for parabolic equations appeared in [Chang Lara and Dávila 2014], which is a parabolic version of the result in [Caffarelli and Silvestre 2009]. The kernels are required to be symmetric and to satisfy the two pointwise inequalities (1-2) as an assumption.

Elliptic integro-differential equations with nonsymmetric kernels are studied in the articles [Chang Lara 2012; Chang Lara and Dávila 2012]. There, the kernels are decomposed into the sum of their even (symmetric) and odd parts. The symmetric part is assumed to satisfy (1-2), and there are appropriate assumptions on the odd part so that the symmetric part of the equations controls the odd part. This effectively makes the contribution to the equation from the odd part of the kernel a lower-order term.

The only articles where the lower bound in the kernels (1-2) is not required to hold at all points are [Bjorland et al. 2012; Guillen and Schwab 2012; Kassmann and Mimica 2013a; Kassmann et al. 2014]. These papers concern elliptic equations and the upper bound in (1-2) is still assumed to hold. It is important to point out that under the conditions in [Bjorland et al. 2012; Kassmann et al. 2014], the Harnack inequality is not true. There is, in fact, a counterexample in [Bogdan and Sztonyk 2005] (also discussed in [Kassmann et al. 2014]). The assumption in these works that was made to replace the pointwise lower bound on the kernels is more restrictive than our assumption (A3) below.

The main result in this article (see Theorems 7.1 and 7.2) generalizes nearly all previous Hölder estimates (for both elliptic and parabolic equations) for integro-differential equations with rough kernels in nondivergence form. It strictly contains the Hölder regularity results in [Bass and Levin 2002a; Bjorland et al. 2012; Caffarelli and Silvestre 2009; Chang Lara 2012; Chang Lara and Dávila 2012; 2014; Guillen and Schwab 2012; Kassmann et al. 2014]. There is an interesting new result given in [Kassmann and Mimica 2013b] that allows for kernels with a *logarithmic* growth at the origin (among other cases), corresponding in our context to the limit $\alpha \to 0$, and it is not contained in the result of this paper.

Our approach draws upon ideas from several previous papers. Moreover, we haven been able to simplify the ideas substantially, especially how to handle parabolic equations, and we do not follow the method in

[Chang Lara and Dávila 2014]. Our method allows us to make more general assumptions on the class of possible kernels. We would like to point out that we do not make any assumption for *simplicity* in this paper. Extending these results to a more singular family of kernels would require new ideas.

There are two possible directions that we did not pursue in this paper. We did not try to analyze singularities of the kernels of order more general than a power of $|h|$, as in [Kassmann and Mimica 2013b]. Also, it might be possible to extend our regularity results for equations with Hölder continuous drifts and $\alpha < 1$, as in [Silvestre 2012a]; although, we do point out that this technique does not work right away with the methods in this paper. We also point out that the results in this paper and all of the others mentioned (except for [Kassmann and Schwab 2014]), require that the Lévy measure — referred to above as $\mu(x, \mathrm{d}h)$ — has a nontrivial absolutely continuous part, $K\,\mathrm{d}h$, with respect to Lebesgue measure (our work allows for a measure with a density plus some singular part). Verifying the validity of, and finding a proof for, results similar to Theorem 1.1 in the case when $\mu$ may not have a density with respect to Lebesgue measure remains a significant open question in the integro-differential theory.

The importance of not assuming any regularity in $x$ and $t$ for the ingredients of (1-1) — the case of so-called bounded measurable coefficients — is for much more than simply mathematical generality. For example, because equations such as (1-1) often lack a "divergence structure" — i.e., admitting a representation as a weak formulation for functions in an energy space such as $H^{\alpha/2}$ — they can usually only be realized as classical solutions or as viscosity solutions (weak solutions). (We note that uniqueness for equations related to (1-1) is still an open question for the theory of viscosity solutions of integro-differential equations, and recent progress has been made in [Mou and Święch 2015].) That means that one of the few tools available for compactness arguments involving families of solutions are those provided in the space of continuous functions via Theorem 1.1. This is relevant for both the possibility of proving the existence of classical solutions as well as for analyzing fully nonlinear equations in a way that doesn't depend on the regularity of the coefficients. Indeed, both situations can be viewed as morally equivalent to studying linear equations with bounded measurable coefficients. For studying regularity of translation invariant equations, this arises by effectively differentiating the equation, which results in coefficients that depend upon the solution. In the fully nonlinear case, many situations involve operators that are a min-max of linear operators, and so the bounded measurable linear coefficients arise from choosing the operators that achieve the min-max for the given function at each given point — a situation in which you cannot assume any regular dependence in the $x$-variable. Such min-max representations turn out to be somewhat generic for fully nonlinear elliptic equations, as was noted in the recent work [Guillen and Schwab 2014, Section 4].

**1B.** ***Application: the homogeneous non-cut-off Boltzmann equation.*** In this section, we briefly explain an important application of our main result, which is not possible to obtain with any of the previously known estimates for integro-differential equations. This result is explained in detail in [Silvestre 2014a].

The Boltzmann equation is a well-known integral equation that models the evolution of the density of particles in dilute gases. In the space homogeneous case, the equation is

$$f_t = Q(f, f) := \int_{\mathbb{R}^n} \int_{\partial B_1} \big(f(v', t) f(v'_\star, t) - f(v_\star, t) f(v, t)\big) B(|v - v_\star|, \theta)\,\mathrm{d}\sigma\,\mathrm{d}v_\star. \tag{1-3}$$

Here $v'$, $v'_\star$ and $\theta$ are defined by the relations

$$\begin{aligned} r &= |v_* - v| = |v'_* - v'|, & v' &= \tfrac{1}{2}(v + v_*) + \tfrac{1}{2} r\sigma, \\ \cos\theta &= \sigma \cdot \frac{v_* - v}{|v_* - v|}, & v'_* &= \tfrac{1}{2}(v + v_*) - \tfrac{1}{2} r\sigma. \end{aligned}$$

There are several modeling choices for the *cross-section* function $B$. From some physical considerations, it makes sense to consider $B(r,\theta) \approx r^\gamma |\theta|^{n-1+\alpha}$, with $\gamma > -n$ and $\alpha \in (0,2)$. Note that this cross section $B$ is never integrable with respect to the variable $\sigma \in \partial B_1$. In order to avoid this difficulty, sometimes a (non-physical) cross section is used that is integrable. This assumption is known as *Grad's cut-off assumption*.

Until the middle of the 1990s, most works on the Boltzmann equation used Grad's cut-off assumption. The non-cut-off case, despite its relevance for physical applications, was not studied so much due to its analytical complexity. An important result that caused a better understanding of the non-cut-off case came with the paper of Alexandre, Desvillettes, Villani, and Wennberg [Alexandre et al. 2000], in which they obtained a lower bound on the entropy dissipation in terms of the Sobolev norm $\|f\|^{\alpha/2}_{\mathrm{loc}}$. All regularity results for the non-cut-off case that came afterwards are based on a coercivity estimate that is a small variation of this entropy dissipation argument. So far, this was the only regularization mechanism that was known for the Boltzmann equation.

It turns our that we can split the right-hand side of the Boltzmann equation, (1-3), in two terms. The first one is an integro-differential operator, and the second is a lower-order term:

$$\begin{aligned} f_t &= Q_1(f,f) + Q_2(f,f) \\ &:= \int_{\mathbb{R}^n} \int_{\partial B_1} f(v'_\star,t)\big(f(v',t) - f(v,t)\big) B(|v - v_\star|,\theta)\, \mathrm{d}\sigma\, \mathrm{d}v_\star \\ &\qquad\qquad + f(v,t) \int_{\mathbb{R}^n} \int_{\partial B_1} \big(f(v'_\star,t) - f(v_\star,t)\big) B(|v - v_\star|,\theta)\, \mathrm{d}\sigma\, \mathrm{d}v_\star \\ &= \int_{\mathbb{R}^n} \big(f(v',t) - f(v,t)\big) K_f(v,v',t)\, \mathrm{d}v' + c f(v,t)[|v|^\gamma * f](v). \end{aligned}$$

The kernel $K_f$ depends on $f$ through a complicated change of variables given using the integral identity above. If one knew that $f$ was a smooth positive function vanishing at infinity, then indeed it could be proved that $K_f(v,v',t) \approx |v - v'|^{-n-\alpha}$, and the first term would correspond to an integro-differential operator of order $\alpha$ in the usual sense satisfying (1-2). Unfortunately, this is not practical for obtaining basic a priori estimates for (1-3). In fact, there is very little we can assume a priori from the solution $f$ to the Boltzmann equation, and it is not enough to conclude that $K_f$ satisfies (1-2). Instead, all we know a priori about $f$ is given by its *macroscopic quantities*: its mass (the integral of $f$), the energy (its second moment), and its entropy. The first two quantities are constant in time, whereas the third is monotone decreasing. It can be shown that $K_f$ satisfies the hypotheses (A1), (A2), (A3) and (A4), depending on these macroscopic quantities only. Therefore, the results in this article can be used to obtain a priori estimates for solutions of the homogeneous, non-cut-off, Boltzmann equation, which is explained in [Silvestre 2014a]. It is a new regularization effect for the Boltzmann equation that is not based on coercivity estimates, as in [Alexandre et al. 2000].

Interestingly enough, the macroscopic quantities do not give much more information about $K_f$ than what our assumptions (A1), (A2) and (A3) say. The kernels $K_f$ will be symmetric, so, in fact, (A4) is redundant. In terms of this generalization, almost the full power of our main result is needed. The only nonessential points are that the kernels can be assumed to be symmetric, and the robustness of the estimates does not necessarily play a role.

**1C. *Notation.***

- Our space variable $x$ belongs to $\mathbb{R}^d$.
- The annulus is $R_r := B_{2r} \setminus B_r$.
- The parabolic cylinder $Q_r$ is defined as

$$Q_r := B_r \times (-r^\alpha, 0], \text{ and with a different center, } Q_r(x,t) = Q_r + (x,t).$$

- The "$\alpha$-growth" class is

$$\mathrm{Growth}(\alpha) = \{v : \mathbb{R}^d \to \mathbb{R} \mid |v(x)| \le C(1+|x|)^{\alpha-\varepsilon} \text{ for some } C, \varepsilon > 0\}.$$

- Pointwise $C^{1,1}$ is defined as

$$C^{1,1}(x) := \{v : \mathbb{R}^d \to \mathbb{R} \mid \exists M(x) \text{ and } \varepsilon \text{ so that } |v(x+h) - v(x) - \nabla v(x)\cdot h| \le M(x)|h|^2 \text{ for } |h| < \varepsilon\},$$

  and over $\mathbb{R}^d$, we have

$$C^{1,1}(\mathbb{R}^d) := \{v : \mathbb{R}^d \to \mathbb{R} \mid \|v\|_{L^\infty(\mathbb{R}^d)} < \infty, \|\nabla v\|_{L^\infty(\mathbb{R}^d)} < \infty, \text{ and } v \in C^{1,1}(x) \forall x \text{ with } M(x) \text{ independent of } x\}.$$

- The difference operator for the different possibilities of $\alpha$ is

$$\delta_y u(x) := \begin{cases} u(x+y) - u(x) & \text{if } \alpha < 1, \\ u(x+y) - u(x) - \mathbb{1}_{B_1}(y)\nabla u(x) \cdot y & \text{if } \alpha = 1, \\ u(x+y) - u(x) - \nabla u(x) \cdot y & \text{if } \alpha > 1. \end{cases}$$

- The class of kernels and corresponding linear operators are

$$\mathcal{K} := \{K : \mathbb{R}^d \to \mathbb{R} \mid K \text{ satisfies assumptions (A1)–(A4)}\},$$

$$\mathcal{L} := \left\{ Lu(x) = \int_{\mathbb{R}^d} \delta_h u(x) K(h)\, \mathrm{d}h \;\middle|\; K \in \mathcal{K} \right\}.$$

We will try to stick to the following conventions for constants:

- Large constants will be upper case letters, e.g., $C$, and small constants will be lower case letters, e.g., $c$.
- If the value of a constant is not relevant for later arguments, then we will freely use the particular letter for the constant without regard to whether or not it was used previously or will be used subsequently.
- If the value of a constant is relevant to later arguments (e.g., in determining values of subsequent constants), then we will label the constant with a subscript, e.g., $C_0$, $C_1$, $C_2$, etc.

**Note 1.2.** The following observation is useful and applies for all values of $\alpha$: if $u(x) = \varphi(x)$ and $u \ge \varphi$ everywhere, then $\delta_h u(x) \ge \delta_h \varphi(x)$ for all $h$. This implicitly assumes that for $\alpha \ge 1$, both $u$ and $\varphi$ are differentiable at $x$.

## 2. Classes of kernels and extremal operators

The kernel $K(x,h,t)$ in (1-1) is not assumed to have any regularity with respect to $x$ or $t$. The best way to think about it is that for every value of $x$ and $t$, we have a kernel ($K_{x,t}(h) = K(x,\cdot,t)$) that belongs to a certain class. This class of kernels is what we describe below.

**2A.** ***Assumptions on K.*** For each value of $\lambda$, $\Lambda$, $\mu$ and $\alpha$, we consider the family of kernels $K : \mathbb{R}^d \to \mathbb{R}$ satisfying the following assumptions:

(A1) $K(h) \ge 0$ for all $h \in \mathbb{R}^d$.

(A2) For every $r > 0$,

$$\int_{B_{2r}\setminus B_r} K(h)\,\mathrm{d}h \le (2-\alpha)\Lambda r^{-\alpha}. \tag{2-1}$$

(A3) For every $r > 0$, there exists a set $A_r$ such that

- $A_r \subset B_{2r} \setminus B_r$,
- $A_r$ is symmetric in the sense that $A_r = -A_r$,
- $|A_r| \ge \mu |B_{2r} \setminus B_r|$,
- $K(h) \ge (2-\alpha)\lambda r^{-d-\alpha}$ in $A_r$.

Equivalently,

$$\bigl|\{y \in B_{2r} \setminus B_r \mid K(h) \ge (2-\alpha)\lambda r^{-d-\alpha} \text{ and } K(-h) \ge (2-\alpha)\lambda r^{-d-\alpha}\}\bigr| \ge \mu |B_{2r} \setminus B_r|. \tag{2-2}$$

(A4) For all $r > 0$,

$$\left|\int_{B_{2r}\setminus B_r} hK(h)\,\mathrm{d}h\right| \le \Lambda |1-\alpha| r^{1-\alpha}. \tag{2-3}$$

**2B.** ***Discussion of the assumptions.*** We stress that although our kernels can be zero for large sets of $h$, their corresponding integral operators are not rightfully described as "degenerate". One can draw an analogy with the second-order case in the context of diffusions. A diffusion process will satisfy uniform hitting-time estimates for measurable sets of positive measure whenever the diffusion matrix is comparable to the identity from below and above. In the context of our pure jump processes related to (1-1), these jump processes will still satisfy such uniform hitting-time estimates even though the kernels can be zero in many points (meaning that at the occurrence of any *one* jump, the process will have zero probability of jumping with certain values of $h$).

The first assumption, (A1), is unavoidable if one hopes to study examples of (1-1) that satisfy a comparison principle between sub- and supersolutions.

The second assumption, (A2), is mostly used to estimate an upper bound for the application of the operator, $L$, to a smooth test function. It is more general than assuming a pointwise upper bound, as was done in [Caffarelli and Silvestre 2009; Kassmann et al. 2014] and many others. It is also slightly more general than a corresponding bound obtained by integrating on spheres as

$$\int_{\partial B_r} K(h)\,\mathrm{d}S(h) \le (2-\alpha)\Lambda r^{-1-\alpha}.$$

It is, however, a stronger hypothesis than

$$\int_{B_r} |h|^2 K(h)\,\mathrm{d}h \le \Lambda r^{2-\alpha}.$$

It is worth pointing out that (A2) implies

$$\int_{\mathbb{R}^d \setminus B_r} K(h)\,\mathrm{d}h \le \frac{2^\alpha}{2^\alpha - 1}(2-\alpha)\Lambda r^{-\alpha}.$$

The first factor blows up as $\alpha \to 0$ but not as $\alpha \to 2$. In fact, the proofs of all our regularity results fail for $\alpha \le 0$ exactly because the tails of the integrals become infinite. The question of what happens as $\alpha \to 0$ is interesting for the nonlocal theory, and some results are obtained in [Kassmann and Mimica 2013b] (note, there they do not use the typical normalization constant as in potential theory, where $C_{d,\alpha} \approx \alpha$ as $\alpha \to 0$, so the limit operator is *not* a multiple of the identity). We also have

$$\int_{\mathbb{R}^d} (1 \wedge |h|^2) K(h)\,\mathrm{d}h \le C(\alpha)\Lambda \tag{2-4}$$

for a constant $C(\alpha)$ that stays bounded as $\alpha \to 2$, and (2-1) can be thought of as a scale invariant, of order $\alpha$, version of (2-4).

Note that the assumption (A2) does not preclude the kernel $K$ from containing a singular measure. For example, the measure given by

$$\int_A K(h)\,\mathrm{d}h = \int_{A\cap\{h_1=h_2=\cdots=h_{d-1}=0\}} (2-\alpha)\frac{\lambda}{|h_n|^{1+\alpha}}\,\mathrm{d}h_d$$

is a valid kernel $K$ that satisfies (A2) (but not (A3)). In this case, $K$ is a singular measure, but we abuse notation by writing it as if it was absolutely continuous with a density $K(h)$.

The example above corresponds to the operator

$$-\int_{\mathbb{R}^d} \delta_h u(x) K(h)\,\mathrm{d}h = (-\partial_{dd})^{\alpha/2} u.$$

As we mentioned before, this kernel satisfies the assumption (A2) but not (A3). However, the kernel of the operator

$$-\int_{\mathbb{R}^d} \delta_h u(x) K(h)\,\mathrm{d}h := (-\partial_{dd})^{\alpha/2} u(x) + (-\Delta)^{\alpha/2} u(x)$$

would satisfy both (A2) and (A3).

The third assumption, (A3), is stated in a form that does not require the kernel $K$ to be positive along some prescribed rays or cone-like sets, as was done in [Kassmann et al. 2014]. The relaxation to (A3) from previous works is important to allow for situations where the positivity set of $K$ may change from radius to radius. As mentioned above, it is equivalent to (2-2), which is the form we will actually invoke later on.

Finally, note that the assumption (A4) is automatic for symmetric kernels (i.e., when $K(h) = K(-h)$), since in that case the left-hand side is identically zero. This assumption is made in order to control the odd part of the kernels in a fashion that does not require us to split up $L$ into two pieces involving the even and odd parts of $K$. It is also worth pointing out that even for $\alpha < 1$, the kernel $K$ can have some asymmetry, but it must die out as $r \to \infty$.

There are two final facts that are important to point out. The first one is the observation that although each $K$ may not be such that

$$Lu(x) = \int_{\mathbb{R}^d} \delta_h u(x) K(h)\, \mathrm{d}h$$

results in an operator that is scale invariant, i.e., $Lu(r\,\cdot)(x) = r^\alpha Lu(rx)$, the *family of* $K$ that satisfy (A1)–(A4) is scale invariant. The second one is that some authors have worked with assumptions where the lower bound in (1-2) is only required for $|h| \le 1$. This does not effect our overall result because we can add and subtract the term

$$f(u;x) := (2-\alpha)\int_{\mathbb{R}^d} \delta_h u(x) \mathbb{1}_{\mathbb{R}^d\setminus B_1}(h)|h|^{-d-\alpha}\, \mathrm{d}h$$

from (1-1). Assuming $K$ satisfies the lower bound of (1-2) only for $|h| \le 1$, this would result in an operator governed by $\widetilde{K}(h) = K(h) + \mathbb{1}_{\mathbb{R}^d\setminus B_1}(h)|h|^{-d-\alpha}$, and now $\widetilde{K}$ does satisfy the lower bound of (1-2) for all $h$. Furthermore, the term $f(u;\cdot)$ is controlled by $\|u\|_{L^\infty}$ and possibly $C|\nabla u|$ (depending on $\alpha$) due to the fact that $\mathbb{1}_{\mathbb{R}^d\setminus B_1}(h)|h|^{-d-\alpha}$ is integrable, and hence these terms can be absorbed into the equation as a gradient term and bounded right-hand side. This pertains to, e.g., the results in [Chang Lara 2012].

**2C.** ***Extremal operators and useful observations.*** As mentioned above, $\mathcal{L}$ is the class of all integro-differential operators $Lu$ of the form

$$Lu(x) = \int_{\mathbb{R}^d} \delta_h u(x) K(h)\, \mathrm{d}h,$$

where $K$ is a kernel satisfying the assumptions (A1)–(A4) specified above. Sometimes we wish to refer to a kernel, $K$, instead of the operator, $L$, and so we also use $\mathcal{K}$ to denote the collection of all such kernels. Correspondingly, we define the extremal operators $M^+_{\mathcal{L}}$ and $M^-_{\mathcal{L}}$ as in [Caffarelli and Silvestre 2009]:

$$M^+_{\mathcal{L}} u(x) = \sup_{L\in\mathcal{L}} Lu(x),$$
$$M^-_{\mathcal{L}} u(x) = \inf_{L\in\mathcal{L}} Lu(x).$$

In order to avoid notational clutter, we omit the subscript $\mathcal{L}$ in the rest of the paper. We note that when (1-1) holds for some kernel $K$ satisfying the assumptions and with a bounded $b$ and $f$, this also implies that the pair of inequalities

$$u_t + C_0|\nabla u| - M^- u \ge -C_0,$$
$$u_t - C_0|\nabla u| - M^+ u \le C_0$$

is simultaneously satisfied. The advantage of this new formulation is that it can be understood in the viscosity sense, whereas the original equation (1-1) only makes sense for classical solutions. Unless otherwise noted, we use the terms solution, subsolution, and supersolution to be interpreted in the viscosity sense (made precise below, in Definition 3.2). There may be instances when we need equations to hold in a classical sense, and in those cases, we will explicitly mention that need.

**Remark 2.1.** We emphasize that although (1-1) allows for $K$ that are $x$-dependent, the class $\mathcal{L}$ — and hence the definition of $M^{\pm}$ — contains only those $K$ that are independent of $x$. The desired inequalities are obtained because $\mathcal{L}$ contains *all possible* such $K$, and hence, for each $x$ fixed, $K(x,\cdot) \in \mathcal{L}$.

It will be useful to know an important feature of $M^{\pm}$ regarding translations, rotations, and scaling. This is an important feature to keep in mind in the sense that for any *one* choice of a kernel to determine (1-1), $K$ may not have any symmetry or scaling properties on its own. However, it is controlled by an extremal operator that does enjoy these properties. This is particularly relevant for intuition on what to expect from solutions of these equations.

**Lemma 2.2.** *$M^+$ (and hence $M^-$) obey the following*:

(i) *If $z \in \mathbb{R}^d$ is fixed, and $Tu := u(\cdot + z)$, then $M^+Tu(x) = M^+u(x+z)$ (translation invariance).*

(ii) *If $R$ is a rotation or reflection on $\mathbb{R}^d$, then $M^+u(R\,\cdot)(x) = M^+u(Rx)$ (rotation invariance).*

(iii) *If $r > 0$, then $M^+u(r\,\cdot)(x) = r^{\alpha} M^+u(rx)$ (scaling).*

*Proof of Lemma 2.2.* Property (i) follows from a direct equality in $LTu(x) = Lu(x+z)$ whenever $K \in \mathcal{L}$ (importantly, note that $K \in \mathcal{L}$ requires $K(x,h) = K(h)$). Property (ii) follows because $\mathcal{L}$ is closed under composing $K$ with a rotation or reflection. Property (iii) follows from the observation that if $K \in \mathcal{L}$, then

$$\widetilde{K}(h) := r^{-d-\alpha} K\left(\frac{h}{r}\right) \in \mathcal{L}$$

as well, combined with the fact that for $L$, $\widetilde{L}$ corresponding to $K$, $\widetilde{K}$, we have $Lu(r\,\cdot)(x) = r^{\alpha}\widetilde{L}u(rx)$. It is worth remarking that when $\alpha = 1$, one must be careful with rescaling the integral due to the presence of $\mathbb{1}_{B_1}(h)$. However, in this case the rescaling still holds because (A4) implies that

$$\int_{B_1 \setminus B_r} hK(h)\,\mathrm{d}h = 0,$$

and this allows to keep the term $\mathbb{1}_{B_1}(h)$ fixed in $\widetilde{L}$ without effecting its value. □

In the rest of this section, we make some elementary estimates that give us some bounds on $Lu(x)$ in terms of bounds for $u$ and its derivatives. These estimates explain the need for the assumptions (2-1) and (2-3). We start with the following lemma.

**Lemma 2.3.** *Let $K$ be a kernel satisfying assumptions* (A2) *and* (A4). *Then the following inequalities hold*:

$$\int_{B_r} |h|^2 K(h)\,\mathrm{d}h \le C\Lambda r^{2-\alpha}, \tag{2-5}$$

$$\left|\int_{B_r} hK(h)\,\mathrm{d}h\right| \le C\Lambda r^{1-\alpha} \quad \text{if } \alpha < 1, \tag{2-6}$$

$$\left|\int_{\mathbb{R}^d \setminus B_r} hK(h)\,\mathrm{d}h\right| \le C\Lambda r^{1-\alpha} \quad \text{if } \alpha > 1, \tag{2-7}$$

$$\int_{\mathbb{R}^d \setminus B_r} K(h)\,\mathrm{d}h \le C\Lambda \frac{2-\alpha}{\alpha} r^{-\alpha}. \tag{2-8}$$

*In this lemma, the constant $C$ is independent of all the other constants.*

*Proof.* The four assertions are all proved in a similar fashion, and they follow from a straightforward decomposition of the integrals in dyadic rings $B_{2^{k+1}r} \setminus B_{2^k r}$ followed by applications of (2-1) and (2-3). We will only write down explicitly the proof of (2-7) as an example.

Assume $\alpha > 1$. We use (2-3) and decompose the integral in dyadic rings $B_{2^{k+1}r} \setminus B_{2^k r}$:

$$\begin{aligned}\left|\int_{\mathbb{R}^d \setminus B_r} hK(h)\,\mathrm{d}h\right| &\le \sum_{k=0}^{\infty}\left|\int_{B_{2^{k+1}r}\setminus B_{2^k r}} hK(h)\,\mathrm{d}h\right| \\ &\le \sum_{k=0}^{\infty} \Lambda|1-\alpha|(2^k r)^{1-\alpha} \\ &\le \Lambda r^{1-\alpha}\frac{|1-\alpha|}{1-2^{1-\alpha}}.\end{aligned}$$

Since the last factor on the right is bounded uniformly for $\alpha \in (1,2)$, we have finished the proof. □

**Lemma 2.4.** *Assume $\alpha \ge \alpha_0$. Let $K$ be any kernel that satisfies* (2-1) *and* (2-3). *Let $u$ be a function that is $C^2$ around the point $x$ and $p = \nabla u(x)$. Moreover, assume that $u$ satisfies the following bounds globally*:

$$D^2 u \le AI, \quad |u| \le B.$$

*Then*,

$$\int_{\mathbb{R}^d} \delta_h u(x) K(h)\,\mathrm{d}h \le C\left(\frac{B}{A}\right)^{-\alpha/2}\left(B + \left(\frac{B}{A}\right)^{1/2}|p|\right).$$

*Here $C$ is a constant that depends on $\Lambda$ and $\alpha_0$. Moreover*, *when $\alpha = 1$*, *we can drop the term depending on $p$ and get*

$$\int_{\mathbb{R}^d} \delta_h u(x) K(h)\,\mathrm{d}y \le C(AB)^{1/2}.$$

*Proof.* Since $\delta_h u(x)$ has a different form depending on $\alpha > 1$, $\alpha = 1$ and $\alpha < 1$, we must divide the proof into these three cases.

We start with the case $\alpha < 1$. In this case $\delta_h u(x) = u(x+h) - u(x)$. Let $r > 0$ be arbitrary. Then

$$\begin{aligned}\int_{\mathbb{R}^d} \delta_h u(x) K(h)\,\mathrm{d}y &= \int_{B_r} \delta_h u(x) K(h)\,\mathrm{d}h + \int_{\mathbb{R}^d\setminus B_r} \delta_h u(x) K(h)\,\mathrm{d}h \\ &\le \int_{B_r} (p\cdot h + A|h|^2) K(h)\mathrm{d}h + \int_{\mathbb{R}^d\setminus B_r} 2B\,K(h)\,\mathrm{d}h. \qquad (2\text{-}9)\end{aligned}$$

Using (2-6), (2-5) and (2-8), we get

$$\le C\left(|p|r^{1-\alpha} + Ar^{2-\alpha} + Br^{-\alpha}\right). \qquad (2\text{-}10)$$

We finish the proof in the case $\alpha < 1$ by picking $r = (B/A)^{1/2}$.

The case $\alpha > 1$ is similar. In this case $\delta_h u(x) = u(x+h) - u(x) - p\cdot h$ and we get

$$\begin{aligned}\int_{\mathbb{R}^d} \delta_h u(x) K(h)\,\mathrm{d}h &= \int_{B_r} \delta_h u(x) K(h)\,\mathrm{d}h + \int_{\mathbb{R}^d\setminus B_r} \delta_h u(x) K(h)\,\mathrm{d}h \\ &\le \int_{B_r} A|h|^2 K(h)\,\mathrm{d}h + \int_{\mathbb{R}^d\setminus B_r} (p\cdot h + 2B) K(h)\,\mathrm{d}h.\end{aligned}$$

This time using (2-5), (2-7) and (2-8), we again arrive at (2-10), and conclude by picking the same $r = (B/A)^{1/2}$.

We are left with the case $\alpha = 1$. In this case,

$$\delta_h u(x) = u(x+h) - u(x) - p \cdot h \, \mathbb{1}_{B_1}(h).$$

For arbitrary $r > 0$, we have

$$\int_{\mathbb{R}^d} \delta_y u(x) K(h) \, \mathrm{d}h = \int_{B_r} \big(u(x+h) - u(x) - p \cdot h\big) K(y) \, \mathrm{d}h + \int_{\mathbb{R}^d \setminus B_r} \big(u(x+h) - u(x)\big) K(h) \, \mathrm{d}h \pm \int_{B_1 \triangle B_r} h \cdot p \, K(h) \, \mathrm{d}h.$$

The last term on the right-hand side is equal to zero because of the assumption (2-3). Therefore, we can drop this term and use the other two to estimate the integral:

$$\begin{aligned} \int_{\mathbb{R}^d} \delta_h u(x) K(h) \, \mathrm{d}h &\le \int_{B_r} A|h|^2 K(h) \, \mathrm{d}h + \int_{\mathbb{R}^d \setminus B_r} 2B \, K(h) \, \mathrm{d}h. \\ &\le C(Ar + Br^{-1}), \end{aligned}$$

where the second inequality follows from (2-5) and (2-8). Picking $r = (B/A)^{1/2}$, we obtain

$$\int_{\mathbb{R}^d} \delta_h u(x) K(h) \, \mathrm{d}h \le C(AB)^{1/2}. \qquad \square$$

**Remark 2.5.** Lemma 2.4 requires an inequality to hold for $D^2 u$ in the full space $\mathbb{R}^d$. This does not require the function $u$ to be $C^2$ globally. What it means is that $u(x) - \frac{1}{2} A|x|^2$ is concave.

**Corollary 2.6.** *Let $M^+_{\mathcal{L}}$ and $M^-_{\mathcal{L}}$ be the extremal operators defined above. Let $p = \nabla u(x)$ and assume that $u$ satisfies the global bounds*

$$-A_- I \le D^2 u \le A_+ I, \qquad |u| \le B.$$

*Then*

$$\begin{aligned} M^+_{\mathcal{L}} u(x) &\le C\Big(\frac{B}{A_+}\Big)^{-\alpha/2}\Big(B + \Big(\frac{B}{A_+}\Big)^{1/2} |p|\Big), \\ M^-_{\mathcal{L}} u(x) &\ge -C\Big(\frac{B}{A_-}\Big)^{-\alpha/2}\Big(B + \Big(\frac{B}{A_-}\Big)^{1/2} |p|\Big). \end{aligned}$$

*Moreover, if $\alpha = 1$, the estimate can be reduced to*

$$\begin{aligned} M^+_{\mathcal{L}} u(x) &\le C(BA_+)^{1/2}, \\ M^-_{\mathcal{L}} u(x) &\ge -C(BA_-)^{1/2}. \end{aligned}$$

*Proof.* The estimate for $M^+_{\mathcal{L}}$ follows from taking the supremum in $K$ in Lemma 2.4. The estimate for $M^-_{\mathcal{L}}$ follows then since

$$M^-_{\mathcal{L}} u(x) = -M^+_{\mathcal{L}}[-u](x). \qquad \square$$

## 3. Viscosity solutions

We use a standard definition of viscosity solutions for integral equations that is the parabolic version of the one in [Caffarelli and Silvestre 2009] and equivalent under most conditions to the parabolic version of [Barles and Imbert 2008].

**Definition 3.1** (cf. [Caffarelli and Silvestre 2011b, Definition 21 and (1.2)]). We say $I$ is a nonlocal operator that is elliptic with respect to the class of operators in this article if $Iu(x)$ is well-defined for any function $u \in \text{Growth}(\alpha)$ such that $u \in C^2(x)$ and moreover,

$$M^-(u_1-u_2)(x) - C|\nabla(u_1-u_2)(x)| \le Iu_1(x) - Iu_2(x) \le M^+(u_1-u_2)(x) + C|\nabla(u_1-u_2)(x)|.$$

The constant $C$ must be equal to zero if $\alpha \le 1$.

We say that $I$ is translation invariant if $I[u(\cdot - x_0)] = Iu(\cdot - x_0)$.

Note that the operators $M^+$ and $M^-$ in particular are nonlocal operators, uniformly elliptic with respect to this class. These are the only operators that are needed for the main result in this article (Theorem 1.1). The main result has implications to nonlinear equations in terms of operators, as in Definition 3.1, which are given in Section 8.

**Definition 3.2** (cf. [Caffarelli and Silvestre 2009, Definition 2.2; Caffarelli and Silvestre 2011b, Definition 25]). Let $I$ be a nonlocal operator as in Definition 3.1. Assume that $u \in \text{Growth}(\alpha)$. We say $u : \mathbb{R}^d \times [T_1, T_2]$ satisfies the following inequality in the viscosity sense, and also refer to it as a viscosity supersolution of

$$u_t - Iu \ge 0 \quad \text{in } \Omega \subset \mathbb{R}^d \times \mathbb{R}$$

if every time there exist a $C^{1,1}$ function $\varphi : D \subset \Omega \to \mathbb{R}$ so that $\varphi(x_0, t_0) = u(x_0, t_0)$ and also $u \ge \varphi$ in $D \cap \{t \le t_0\}$, then the auxiliary function

$$v(x) = \begin{cases} \varphi(x, t_0) & \text{if } (x, t_0) \in D, \\ u(x, t_0) & \text{if } (x, t_0) \notin D \end{cases}$$

satisfies

$$v_t(x_0, t_0) - Iv(x_0, t_0) \ge 0.$$

One of the most characteristic properties of viscosity solutions is that they obey the comparison principle. In the context of this article, we state it as follows.

**Proposition 3.3.** *Let $I$ be a translation invariant nonlocal operator that is uniformly elliptic in the sense of Definition 3.1. Let $u, v \in \mathbb{R}^n \times [0, T]$ be two continuous functions such that*

- *for all $x \in \mathbb{R}^n$, we have $u(x, 0) \ge v(x, 0)$,*
- *for all $x \in \mathbb{R}^n \setminus B_1$ and $t \in [0, T]$, we have $u(x, t) \ge v(x, t)$,*
- *$u_t - Iu \ge 0$ and $v_t - Iv \le 0$ in $B_1 \times [0, T]$.*

*The $u(x, t) \ge v(x, t)$ for all $x \in B_1$ and $t \in [0, T]$.*

The proof of Proposition 3.3 is by now standard. We refer the reader to [Chang Lara and Dávila 2014, Corollary 3.1; Silvestre 2011, Lemmas 3.2, 3.3; Caffarelli and Silvestre 2009, Theorem 5.2; Barles and Imbert 2008] for the main ideas. For the purposes of this article, we do not use the full power of Proposition 3.3. We only use the comparison principle to compare a supersolution $u$ with a special barrier function constructed in Section 5. This barrier function is explicit and is smooth, except on a sphere where it has an *angle* singularity. The comparison principle follows easily from Definition 3.2 when $v$ is this special barrier function or any smooth subsolution of the equation.

In [Caffarelli and Silvestre 2009], and many subsequent works, it was frequently used that wherever a viscosity solution $u$ can be touched with a $C^2$ test function from one side, the equation can be evaluated classically with the original $u$ at that particular point (a notable departure from the second-order theory!). This fact plays a role in some measure estimates used to prove the regularity results in those works. With our current setting, it is not possible to evaluate the equation pointwise in $u$ because of the gradient terms; however, many possible useful variations on that theme can be shown — similar to [Kassmann et al. 2014, Appendix 7.2]. In this case, the following lemma is what we will use to obtain pointwise evaluation of the regularized supersolution.

**Lemma 3.4.** *Assume $u$ satisfies the following inequality in the viscosity sense*:

$$u_t + C_0|\nabla u| - M^- u \geq -C \quad \text{in } \Omega.$$

*Assume also that there is a test function $\varphi : \mathbb{R}^d \times [t_1, t_2] \to \mathbb{R}$ so that $\varphi(x_0, t_0) = u(x_0, t_0)$ and $\varphi(x, t) \leq u(x, t)$ for all $t \in (t_0 - \varepsilon, t_0]$.*

*Then, the following inequality holds*:

$$\varphi_t(x_0,t_0)+C_0|\nabla\varphi(x_0,t_0)|-M^-\varphi(x_0,t_0)-\inf\left\{\int_{\mathbb{R}^d}\big(u(x+y,t_0)-\varphi(x+y,t_0)\big)K(y)\,\mathrm{d}y \,:\, K\in\mathcal{K}\right\}\geq -C.$$

*Proof.* We can use $\varphi$ as the test function for Definition 3.2 in any small domain $D = B_r(x_0) \times (t_0 - \varepsilon, t_0]$. Constructing the auxiliary function $v$, we observe that

$$v_t(x_0, t_0) = \varphi_t(x_0, t_0), \quad \nabla v(x_0, t_0) = \nabla\varphi(x_0, t_0),$$

$$\begin{aligned} M^- v(x_0, t_0) &= \inf\left\{\int_{\mathbb{R}^d} \delta_y\varphi(x)K(y)\,\mathrm{d}y + \int_{\mathbb{R}^d\setminus B_r} \big(u(x+y) - \varphi(x+y)\big)K(y)\,\mathrm{d}y \;:\; K \in \mathcal{K}\right\} \\ &\geq M^-\varphi(x_0, t_0) + \inf\left\{\int_{\mathbb{R}^d\setminus B_r} \big(u(x+y) - \varphi(x+y)\big)K(y)\,\mathrm{d}y \;:\; K \in \mathcal{K}\right\}. \end{aligned}$$

Observe that the last term is monotone increasing as $r \to 0$.

From Definition 3.2, for any $r > 0$, we have that $v_t(x_0, t_0) + C_0|\nabla v(x_0, t_0)| - M^- v(x_0, t_0) \geq -C_1$. The result of the lemma follows by taking $r \to 0$. □

## 4. Relating a pointwise value with an estimate in measure: the growth lemma

In order to obtain the Hölder continuity of $u$, we need to show the following point-to-measure lemma, which seems to originate in the work of Landis [1971] (in some circles, it is known as the *growth lemma*).

It is a cornerstone of regularity theory, it leads to the weak Harnack inequality, and it is one of the few places where the equation plays a fundamental role.

**Lemma 4.1.** *There exist positive constants $A_0$ and $\delta_0$ depending on $\lambda$, $\Lambda$, $d$, $\alpha_0$ and $C_0$ so that if $\alpha > \alpha_0$ and if $u : \mathbb{R}^d \times (-1, 0] \to \mathbb{R}$ is a function such that*

(1) *$u \geq 0$ in the whole space $\mathbb{R}^d \times (-1, 0]$,*

(2) *$u$ is a supersolution in $Q_1$, i.e.,*

$$u_t + C_0|\nabla u| - M^- u(x) \geq 0 \quad \text{in } Q_1, \tag{4-1}$$

(3) $\min_{Q_{1/4}} u \leq 1$,

*then*

$$|\{u \leq A_0\} \cap Q_1| \geq \delta_0.$$

The following function, $q$, plays an important role in the proof of Lemma 4.1. It is actually an inf-convolution of $u$ with a quadratic, and it is defined as

$$q(x, t) = \min_{y \in \overline{B}_1} u(y, t) + 64|x - y|^2. \tag{4-2}$$

Note that $q$ is a nonnegative function. We will prove a collection of properties of the function $q$, which will lead us to the proof of Lemma 4.1.

The next barrier is used to find a bound for the rate at which $q$ can decrease with respect to $t$.

**Lemma 4.2.** *For a universal constant $C_1$, the function*

$$\varphi(x, t) = \max(0, f(t) - 64|x|^2)$$

*is a subsolution to*

$$\varphi_t + C_0|\nabla \varphi| - M^- \varphi \leq 0 \quad \text{in } \mathbb{R}^n \times (-\infty, 0].$$

*The inequality holds classically at all points where $\varphi > 0$.*

*Here $f(t)$ is the (unique) positive solution to the (backward) ODE*

$$\begin{cases} f(0) = 0, \\ f'(t) = -C_1\big(f(t)^{1/2} + f(t)^{1-\alpha/2}\big), \end{cases} \tag{4-3}$$

*where $C_1$ is a constant depending on $\Lambda$ and $\alpha_0$ (such that $\alpha \geq \alpha_0$).*

*Proof.* Note that for every fixed value of $t \in (-\infty, 0]$, it holds that

$$\|\varphi\|_{L^\infty} = f(t), \quad \|\nabla \varphi\|_{L^\infty} \leq C\sqrt{f(t)}, \quad \text{and} \quad 0 \geq D^2\varphi \geq -128I.$$

Applying Corollary 2.6,

$$M^- \varphi \geq -C f(t)^{1-\alpha/2}.$$

Then, at all points where $\varphi > 0$, we have

$$\varphi_t + C_0|\nabla \varphi| - M^- \varphi \leq f'(t) + C_0 C f(t)^{1/2} + C f(t)^{1-\alpha/2}.$$

The lemma then follows by choosing $C_1$ so that $f'$ dominates the right-hand side. ☐

It is worth commenting that the ODE for $f$ in Lemma 4.2 has a unique solution that is strictly positive for $t < 0$. This function $f$ is differentiable and locally Lipschitz. The universal constant $C_2$ of the following result is the Lipschitz constant of $f$ in the interval $[-T, 0]$, where $f(T) = -4$.

**Corollary 4.3.** *Assume $x \in B_{1/8}$ and $q(x,t) < 3$. Then there are positive universal constants $\tau$ and $C_2$ such that for $s \in (t-\tau, t)$, we have $q(x,s) - q(x,t) < C_2(t-s)$.*

*Proof.* We let $x$, $t$, and $s$ be fixed as stated. Let $y$ be the point where the minimum for $q(x,t)$ is achieved in (4-2). Using the definition of $q$, we note that for all values of $z \in B_1$, we have $u(z,s) \geq q(x,s) - 64|x-z|^2$.

The point of the proof is to use the fact that $u$ and $\varphi$ are respectively super- and subsolutions of (4-1) on the time interval $(s, 0]$. In order to invoke a comparison result between them, we will make various choices involving $\tau$ and $f$ to enforce $\varphi$ to be below $u$ at the initial time, $s$, and on the boundary, which is $\mathbb{R}^d \setminus B_1$.

We define the function

$$\bar{\varphi}(\bar{x}, \bar{t}) := \varphi(\bar{x} - x, \bar{t} - s + t_0),$$

where $t_0$ is a fixed time, yet to be chosen. We fix the constant $\tau$ so that

$$\tau < f^{-1}(3) - f^{-1}(4),$$

and we fix the time $t_0 < 0$ so that

$$f(t_0) = \min(q(x,s), 4).$$

Checking the boundary condition for $\bar{x} \notin B_1$ and $\bar{t} > s$, we see that $|x - \bar{x}| \geq \frac{7}{8}$ (as $x \in B_{1/8}$), and hence since $f(t_0) \leq 4 \leq 49$, we have (note $f$ is decreasing)

$$\bar{\varphi}(\bar{x}, \bar{t}) = \varphi(\bar{x} - x, \bar{t} - s + t_0) = \max\bigl(0, f(\bar{t} - s + t_0) - 64|x - \bar{x}|^2\bigr) \leq \max(0, f(t_0) - 49) \leq 0.$$

Checking the initial condition at $\bar{t} = s$, we have (by the definition of $t_0$)

$$\bar{\varphi}(\bar{x}, s) = \varphi(\bar{x} - x, t_0) = \max(0, f(t_0) - 64|x - \bar{x}|^2) \leq \max(0, q(x,s) - 64|x - \bar{x}|^2) \leq u(\bar{x}, s),$$

from the definition of $q$.

Comparison therefore tells us that $u \geq \bar{\varphi}$ on $B_1 \times (s, 0)$, and, in particular, for $\bar{x} = y$ and $\bar{t} = t$,

$$u(y,t) \geq \varphi(x - y, t - s + t_0) \geq f(t - s + t_0) - 64|x - y|^2.$$

Hence

$$q(x,t) = u(y,t) + 64|x - y|^2 \geq f(t - s + t_0),$$

and we will use

$$q(x,t) \geq f(t - s + t_0) \geq f(t_0) - |f'(t_0)|(t - s).$$

In the case that $f(t_0) = q(x,s)$, we can conclude the corollary with $C_2 := \max\{f'(t) : t \in (-f^{-1}(4), 0)\}$. However, $\tau$ was chosen specifically so that it is impossible for $f(t_0) < q(x,s)$. Indeed we see that if it occurred that $f(t_0) = 4$ then because $f$ is decreasing and $t - s \leq \tau$, it holds that

$$3 > q(x,t) \geq f(t - s + t_0) \geq f(t_0) + f(\tau + t_0) - f(t_0) \geq 4 + f(f^{-1}(3)) - 4 = 3,$$

which is a contradiction. Thus $f(t_0) = q(x,s)$ is the only possibility, and we conclude. ∎

Corollary 4.3 should be interpreted as $q_t \geq -C_2$ everywhere. The next lemma gives us a bound above for $q_t$ in a set of positive measure.

**Lemma 4.4.** *Under the assumptions of Lemma 4.1,* (*but assuming here* $u(0,0)=1$) *the function* $q$ *from* (4-2) *satisfies* $|\{q_t \leq A_1\} \cap Q_1| \geq \delta_1 > 0$, *where* $A_1$ *and* $\delta_1$ *are universal constants.*

*Proof.* Since $u(0,0)=1$, for any $x \in B_{1/4}$, we have $q(x,0) \leq 1 + 64|x|^2 < 5$. Moreover, the minimum is achieved at some $y \in B_{1/2}$ since $1 + 64|y-x|^2 > 5$ if $|y| > \frac{1}{2}$. By similar reasoning, we also have that for every $x \in B_{1/8}$, it holds that $q(x,0) < 2$. Corollary 4.3 implies that for $t \in (-\tau, 0]$,

$$q(x,t) \leq q(x,0) + C_2|t| < 2 + C_2|t|.$$

Thus if we restrict $t \in (-\tau', 0]$, where $\tau' = 1/C_2$, then we have that $q(x,t) < 3$ and a second application of Corollary 4.3 shows that $q(x,t) + C_2 t$ is monotone increasing. Thus $q_t(x,t)$ exists pointwise for a.e. $t \in (-\tau', 0]$ and $q_t$ exists as a signed measure. Furthermore,

$$q_t(x,t) \geq -C_2 \quad \text{for a.e. } t \in (-\tau', 0].$$

Integrating the measure $q_t(x,t)$ and ignoring its singular part shows (note, $q \geq 0$ always)

$$\begin{aligned} C = 2|B_{1/8}| &\geq \int_{B_{1/8}} q(x,0) - q(x,-\tau')\,\mathrm{d}x \\ &\geq \int_{-\tau'}^{0} \int_{B_{1/8}} q_t(x,s)\,\mathrm{d}x\,\mathrm{d}s \\ &\geq A_1\big|\big((-\tau',0] \times B_{1/8}\big) \cap \{q_t > A_1\}\big| - C_2\big|\big((-\tau',0] \times B_{1/8}\big) \setminus \{q_t > A_1\}\big| \\ &= -C_2\tau'|B_{1/8}| + (A_1 + C_2)\big|\big((-\tau',0] \times B_{1/8}\big) \cap \{q_t > A_1\}\big|. \end{aligned}$$

Therefore, rearranging shows that

$$\big|\big((-\tau',0] \times B_{1/8}\big) \cap \{q_t > A_1\}\big| \leq \frac{C + C_2\tau'}{A_1 + C_2}.$$

We can make the right-hand side arbitrarily small by choosing $A_1$ large. In particular, we choose $A_1$ sufficiently large (depending only on universal constants) so that we have

$$\big|\big((-\tau',0] \times B_{1/8}\big) \cap \{q_t \leq A_1\}\big| \geq \tfrac{1}{2}\tau'|B_{1/8}| =: \delta_1. \qquad \square$$

After Corollary 4.3 and Lemma 4.4, we obtain a set of positive measure where $|q_t|$ is bounded. At this point, we can use ideas from the stationary case to proceed with the rest of the proof.

The next lemma replaces Lemma 8.1 in [Caffarelli and Silvestre 2009]. We, in fact, prove a slightly modified version of the lemma, which enforces a quadratic growth of $\delta_h u$ simultaneously on *two* rings. In the proofs of Theorem 8.7 and Lemma 10.1 in [Caffarelli and Silvestre 2009], there is a cube decomposition plus a covering argument. It could be replaced by a double covering argument. In this paper, we will have a simpler covering argument using Vitali's lemma only once. This is possible thanks to the stronger measure estimate in the next lemma (in two simultaneous rings).

**Lemma 4.5.** *Let $\mu$ be the constant in* (2-2) *and $c_0 < 1$ be an arbitrary constant. Let $y$ be the point in $B_{1/2}$ where the minimum of* (4-2) *is achieved and $u$ satisfies* (4-1). *Assume that $x \in B_{1/4}$, $q(x,t) < 3$ and $q_t(x,t) < A_1$. Then, for $A_2$ sufficiently large* (*depending on $C_1$, $\mu_1$, $\lambda$, $\Lambda$, $c_0$ and $\alpha_0$ but not on $\alpha$*), *we have that there exists some $r \le r_0$ so that both*

$$\left|\left\{h \in B_{2r} \setminus B_r : \delta_h u(y,t) \le A_2 r^2 \text{ and } \delta_{-h} u(y,t) \le A_2 r^2\right\}\right| \ge \tfrac{1}{2}\mu |B_{2r} \setminus B_r| \tag{4-4}$$

*and*

$$\left|\left\{h \in B_{2c_0r} \setminus B_{c_0r} : \delta_h u(y,t) \le A_2 (c_0r)^2 \text{ and } \delta_{-h} u(y,t) \le A_2 (c_0r)^2\right\}\right| \ge \tfrac{1}{2}\mu |B_{2c_0r} \setminus B_{c_0r}| \tag{4-5}$$

*hold simultaneously for $r$ and $c_0r$. Here $r_0 = 4^{-1/(2-\alpha)}$, and we note that $r_0 \to 0$ as $\alpha \to 2$.*

In Lemma 4.5, we abuse notation by writing

$$\delta_h u(y,t) = u(y+h,t) - u(y,t) - 128(x-y)\cdot h,$$

even though $\nabla u(y,t)$ may not exist. Note that if $u$ happens to be differentiable at $(y,t)$, then $\nabla u(y,t) = 128(x-y)$ because of (4-2). The value of $c_0$ will be selected as a universal constant in Lemma 4.7.

*Proof.* From the construction of $x$ and $y$, we have that $u(y,t) = q(x,t) - 64|x-y|^2$. Moreover, $u(z,s) \ge q(x,s) - 64|x-z|^2$ for any $z \in \mathbb{R}^n$ and $s \le t$. Since we are assuming that $q_t(x,t) < A_1$ (in particular, that $q_t$ exists at that point), there is an $\varepsilon > 0$ so that $q(x,s) > q(x,t) - A_1(t-s)$ for $s \in (t-\varepsilon, t]$. Consequently, $u(z,s) \ge q(x,t) - 64|x-z|^2 - A_1(t-s)$ for $s \in (t-\varepsilon, t]$.

Let

$$\varphi(z,s) := \max\bigl(q(x,t) - 64|x-z|^2 - A_1(t-s), -256\bigr).$$

The choice of the number $-256$ is made so that the maximum is always achieved by the paraboloid every time $z \in B_1$. From the analysis above, we have that $u \ge \varphi$ in $\mathbb{R}^n \times (t-\varepsilon, t]$ and $u(y,t) = \varphi(y,t)$. Note that since $q(x,t) < 3$, we have $|\nabla\varphi(y,t)| \le 16\sqrt{3}$. Also, from Lemma 2.4, since $D^2\varphi \ge -128I$, we have $M^-\varphi(y,t) \ge -C$ for some universal constant $C$. We apply Lemma 3.4 and we get

$$\begin{aligned}
0 &\le \varphi_t(y,t) + C_0|\nabla\varphi(y,t)| - M^-\varphi(y,t) - \inf\left\{\int_{\mathbb{R}^d} \bigl(u(y+h,t) - \varphi(y+h,t)\bigr)K(h)\,\mathrm{d}h : K \in \mathcal{K}\right\} \\
&\le A_1 + C_0|\nabla\varphi(y,t)| - M^-\varphi(y,t) - \inf\left\{\int_{\mathbb{R}^d} \bigl(u(y+h,t) - \varphi(y+h,t)\bigr)K(h)\,\mathrm{d}h : K \in \mathcal{K}\right\} \\
&\le C - \inf\left\{\int_{\mathbb{R}^d} \bigl(u(y+h,t) - \varphi(y+h,t)\bigr)K(h)\,\mathrm{d}h : K \in \mathcal{K}\right\}.
\end{aligned}$$

Note that $u(y+h,t) - \varphi(y+h,t) \ge 0$ for all values of $h \in \mathbb{R}^n$. We abuse notation by saying

$$\delta_h u(y,t) = u(y+h,t) - u(y,t) - h\cdot\nabla\varphi(y,t).$$

Note that

$$u(y+h,t) - \varphi(y+h,t) = \delta_h u(y,t) - \delta_h\varphi(y,t),$$

and $\delta_h\varphi(y,t) = -64|h|^2$ whenever $y+h \in B_1$.

Using that the integrand is positive, we can reduce its domain of integration to an arbitrary subset of $\mathbb{R}^n$:

$$C \ge \inf\left\{ \int_{B_{r_0}} \big(u(y+h,t) - \varphi(y+h,t)\big) K(h)\,\mathrm{d}h : K \in \mathcal{K} \right\}$$
$$= \inf\left\{ \int_{B_{r_0}} \big(\delta_h u(y,t) + 64|h|^2\big) K(h)\,\mathrm{d}h : K \in \mathcal{K} \right\}.$$

Let us define

$$w(h) := \delta_h u(x,t) + 64|h|^2 \ge 0$$

for $h \in B_{r_0}$. We have that there exists an admissible kernel $K$ such that

$$C \ge \int_{B_{r_0}} w(h) K(h)\,\mathrm{d}h. \tag{4-6}$$

Let $r \le r_0 = 4^{-1/(2-\alpha)}$. From (2-2), we know that

$$\big|\{h \in B_{2r} \setminus B_r : K(h) > (2-\alpha)\lambda r^{-d-\alpha} \text{ and } K(-h) > (2-\alpha)\lambda r^{-d-\alpha}\}\big| > \mu |B_{2r} \setminus B_r|. \tag{4-7}$$

To obtain a contradiction, let us assume that the result of the lemma is false. That is, for all $r \le r_0$, either

$$\big|\{h \in B_{2r} \setminus B_r : w(h) > (A+64)r^2 \text{ or } w(-h) > (A+64)r^2\}\big| > \big(1 - \tfrac{1}{2}\mu\big)|B_{2r} \setminus B_r| \tag{4-8}$$

or

$$\big|\{h \in B_{2c_0 r} \setminus B_{c_0 r} : w(h) > (A+64)(c_0 r)^2 \text{ or } w(-h) > (A+64)(c_0 r)^2\}\big| > \big(1-\tfrac{1}{2}\mu\big)|B_{2c_0 r} \setminus B_{c_0 r}|. \tag{4-9}$$

Therefore, the intersection of the set in (4-7) — with $r$ appropriately chosen in each case — with either of that in (4-8) or (4-9) must have measure at least $\frac{1}{2}\mu|B_{2r} \setminus B_r|$ or $\frac{1}{2}\mu|B_{2c_0 r} \setminus B_{c_0 r}|$, depending on which of the two possibilities occurred. Let us set $\tilde{r}$ to be either $r$ or $c_0 r$, depending upon whether we will invoke (4-8) or (4-9). Let us call $G_{\tilde{r}}$ this intersection between the sets (4-7) and either (4-8) or (4-9). Note that $G_{\tilde{r}} \subset B_{2\tilde{r}} \setminus B_{\tilde{r}}$ and $G_{\tilde{r}}$ is symmetric (i.e., $G_{\tilde{r}} = -G_{\tilde{r}}$). Moreover, for all $h \in G_{\tilde{r}}$, either $w(h) > (A+64)\tilde{r}^2$ and $K(h) > (2-\alpha)\lambda\tilde{r}^{-d-\alpha}$ or $w(-h) > (A+64)\tilde{r}^2$ and $K(-h) > (2-\alpha)\lambda\tilde{r}^{-d-\alpha}$. Therefore

$$\begin{aligned}
\int_{B_{2\tilde{r}} \setminus B_{\tilde{r}}} w(h)K(h)\,\mathrm{d}h &\ge \int_{G_{\tilde{r}}} w(h)K(h)\,\mathrm{d}h \\
&= \frac{1}{2}\int_{G_{\tilde{r}}} w(h)K(h) + w(-h)K(-h)\,\mathrm{d}h \\
&\ge \frac{1}{2}\int_{G_{\tilde{r}}} A\lambda(2-\alpha)\tilde{r}^{-d+2-\alpha}\,\mathrm{d}h \\
&\ge A\lambda(2-\alpha)\tilde{r}^{2-\alpha}\mu\omega_d,
\end{aligned}$$

where $\omega_d$ is a constant depending on dimension only.

We invoke the contradiction assumption for each of the radii $r_j = 2^{-j-1} r_0$, with $j = 0, 1, 2, \ldots$. For each $r_j$, we get the estimates corresponding to $\tilde{r}_j$, which is either $r_j$ or $c_0 r_j$, depending on the case of the contradiction assumption. Partitioning $B_{r_0}$, we get

$$\begin{aligned}\int_{B_{r_0}} w(h)K(h)\,\mathrm{d}h &= \sum_{j=0}^{\infty}\int_{B_{2r_j}\setminus B_{r_j}} w(h)K(h)\,\mathrm{d}h\\ &\geq \frac{1}{2}\sum_{j=0}^{\infty}\int_{B_{2\tilde r_j}\setminus B_{\tilde r_j}} w(h)K(h)\,\mathrm{d}h\\ &\geq (A+64)\lambda(2-\alpha)\mu\omega_d\sum_{j=0}^{\infty}(\tilde r_j)^{2-\alpha}\\ &\geq (A+64)\lambda(2-\alpha)\mu\omega_d\sum_{j=0}^{\infty}(c_0 2^{-j-1}r_0)^{2-\alpha}\\ &= C(d)c_0^{2-\alpha}(A+64)\mu\lambda\frac{2-\alpha}{1-2^{\alpha-2}}.\end{aligned}$$

We get a contradiction with (4-6) if $A$ is large enough. Note that the last factor is bounded away from zero, independently of $\alpha$ as long as $\alpha \in (0,2)$. Thus the value of $A = A_2$ is independent of $\alpha$, and it is chosen to obtain this contradiction. □

The following geometric statement about functions will play a role in the proof of Lemma 4.1.

**Lemma 4.6.** *Let $u : \mathbb{R}^d \to \mathbb{R}$ be a continuous bounded function such that $\nabla u(0)$ exists. Let $q(x) = \min_{y\in\bar B_1} u(y) + 64|x-y|^2$. Assume the following conditions hold true*:

- *There is at least one point $x_0 \in \mathbb{R}^d$ for which $q(x_0) = u(0)+64|x_0|^2 = \min_{y\in\bar B_1}\{u(y)+64|x_0-y|^2\}$.*
- *If we consider the (symmetric) set*

$$G := \{h \in B_2\setminus B_1 : \delta_h u(0) \leq A \text{ and } \delta_{-h}u(0) \leq A\},$$

*then $|G| \geq \frac{1}{2}\mu|B_2\setminus B_1|$. (Here, as in Lemma 4.5, $\delta_h u(y,t) = u(y+h,t)-u(y,t)-128(x-y)\cdot h$.)*

*Then there are constants $c_0$ and $C_4$ depending on $A$ and $\mu$ and $d$ so that if for some pair of points $x_1$, $y_1$ we have*

$$q(x_1) = u(y_1) + 64|x_1-y_1|^2,$$

*then $|y_1| < c_0$ implies $|x_1-x_0| < C_4$.*

*Proof.* Assume $|y_1| < c_0$. Let $p_0$ and $p_1$ be the quadratic polynomials

$$p_0(z) = q(x_0) - 64|x_0-z|^2,$$
$$p_1(z) = q(x_1) - 64|x_1-z|^2.$$

From the definition of $q$, we have that $p_0(z) \leq u(z)$ and $p_1(z) \leq u(z)$ for all $z \in \mathbb{R}^d$. Moreover, $p_0(y_0) = u(y_0)$ and $p_1(y_1) = u(y_1)$.

Observe that $p_1 - p_0$ is the affine function

$$p_1(z) - p_0(z) = q(x_1) - q(x_0) + 64(|x_0|^2 - |x_1|^2) + 128(x_1-x_0)\cdot z.$$

Since $p_1(y_1) = u(y_1) \geq p_0(y_1)$, we have

$$p_1(y_1+z) - p_0(y_1+z) \geq 128(x_1-x_0)\cdot z.$$

Using that $u(y_1+z) \geq p_1(y_1+z) \geq p_0(y_1+z) + 128(x_1-x_0)\cdot z$, we get that

$$\begin{aligned}\delta_{(y_1+z)}u(0) &\geq \delta_{(y_1+z)}p_0(0) + 128(x_1-x_0)\cdot z \\ &\geq -64 + 128(x_1-x_0)\cdot z \quad \text{for } z \in B_1.\end{aligned}$$

Let us consider the following set, which is the intersection of a cone (whose vertex is at $y_1$, recall that $|y_1| < c_0$) and the ring $B_2 \setminus B_1$:

$$H = \bigl\{h \in B_2 \setminus B_1 \,:\, h = y_1 + z \text{ with } z\cdot(x_1-x_0) > c_0|z||x_1-x_0|\bigr\}.$$

Observe that as $c_0 \to 0$, the set $H$ approximates the intersection of the ring $B_2 \setminus B_1$ with the half-space $\{z \,:\, z\cdot(x_1-x_0) > 0\}$. More precisely

$$\bigl|B_2 \setminus B_1 \setminus H \setminus -H\bigr| \leq Cc_0$$

for some constant $C$ depending on dimension only.

Let us choose $c_0$ so that $Cc_0 < \frac{1}{2}\mu|B_2 \setminus B_1|$. Then $H \cap G$ must have a positive measure (also $G \cap -H$, recall that $G$ is symmetric), and so there exists some $h \in H \cap G$. Then

$$\begin{aligned}A \geq \delta_h u(0) &\geq -64 + 128(x_1-x_0)\cdot z \\ &> -64 + 128c_0|x_1-x_0||z| \\ &\geq -64 + 64c_0|x_1-x_0|.\end{aligned}$$

Therefore $|x_1-x_0| < \bigl(\frac{1}{64}A+1\bigr)/c_0 =: C_4$. □

In the proof of Lemma 4.1, we will use the map $m : y \mapsto x$, which assigns the point $x$ where the minimum is achieved in the definition of $q$. This maps plays the same role as the gradient map of the convex envelope of $u$ in $B_r$ does in an ABP-based proof of the growth lemma. This would be the purpose of [Caffarelli and Silvestre 2009, Lemma 8.4] or [Bjorland et al. 2012, Lemma 3.6]. In those cases, we would need to adjust $u$ by a supporting hyperplane and argue using a convex envelope. In our approach, we work without invoking a convex envelope.

Note that after Corollary 4.3 and Lemma 4.4, where we obtain that $|q_t|$ is bounded in a set of positive measure, the rest of the proof of Lemma 4.1 should be interpreted as a nonlocal version of the method in [Savin 2007]. It is more flexible, and arguably more natural, than an ABP-based proof.

We are now in a position to prove Lemma 4.1.

*Proof of Lemma 4.1.* We assume $u(0,0) = 1$. The result follows for the assumption $\min_{Q_{1/2}} u = 1$ by a simple translation argument.

Let $G$ be the set of points $(x,t) \in B_{1/8} \times (-\tau, 0]$ so that $q_t \leq A_1$. From Lemma 4.4, we have a universal lower bound on its measure: $|G| > \delta_1$. For each point $(x,t) \in G$, there is at least one point $y \in B_1$ that realizes the minimum value for $q(x,t)$ in (4-2). For each fixed value of $t$, we define the map $m : y \mapsto x$.

This is a well-defined function if $u \in C^1$. In general, the function nature of $m$ is not necessary, and we should think of $m$ as a set mapping that sends values of $y$ into a set of possible values of $x$ (like the subdifferential of a convex function).

We note that if $y \in m^{-1}(G)$, we have $q_t(x,t) \le A_1$ for some $x \in G$, and we can apply Lemma 4.5, which was presented above. This gives a ball around $y$ and a collection of points where $u$ does not grow too much. For example, we can control the set

$$E_y := \{z \in B_{c_0 r}(y) : u(z,t) < A_2 + 43\}. \tag{4-10}$$

This is possible by starting with the ring from Lemma 4.5 and then noting that $r \le 1$, $u(y) < 3$ (since $q(x) \le 3$, see first line of the proof of Lemma 4.5), $\delta_{\pm h} u(y) \le A_2 r^2$, $|h| \le \frac{1}{2}$, $|x-y| \le \frac{5}{8}$, and $128|x-y||h| \le 40$. Thus from Lemma 4.5, we see that

$$|E_y| = \left|\{z \in B_{c_0 r}(y) : u(z,t) < A_2 + 43\}\right| > \delta |B_{c_0 r}|. \tag{4-11}$$

Here $\delta$ is a constant that depends on dimension and the $\mu$ from Lemma 4.5. We note that we use the $r^2$-growth of $\delta_h u$ from Lemma 4.5 in a very rough fashion at this step. The importance of the $r^2$ comes later, in relationship to an upper bound on $|m(B_r)|$. We also note that we have used the ball $B_{c_0 r}$ instead of $B_r$. At this stage, both balls have the same estimate regarding the growth of $u$ on a universal proportion of the set. However, only $B_{c_0 r}$ also has the necessary estimate for the size of $m(B_{c_0 r})$. This choice will be further illuminated below.

We need to estimate a set where $u$ is not too large, and given the choice of $E_y$ above, we see that a good candidate is

$$NL := \bigcup_{y \in m^{-1}(G)} E_y.$$

Thanks to (4-10) and (4-11), the measure of $NL$ can be equivalently estimated via the size of

$$NLB := \bigcup_{y \in m^{-1}(G)} B_{c_0 r(y)}(y),$$

where $B_{c_0 r(y)}(y)$ is the good ball given in Lemma 4.5. Therefore, the only question is whether or not the set, $NLB$, has a measure that is comparable to $B_1$.

If $\{B_j\}$ is a Vitali subcovering of the collection $\{B_{r(y)}(y)\}_{y \in m^{-1}(G)}$, then we have

$$\bigcup_j 5B_j \supset m^{-1}(G),$$

and hence

$$m\Bigl(\bigcup_j 5B_j\Bigr) \supset m(m^{-1}(G)).$$

Also by subadditivity, we have that

$$\Bigl|m\Bigl(\bigcup_j B_j\Bigr)\Bigr| \le \sum_j |m(B_j)|.$$

In order to conclude, it would suffice to know that

$$|m(B_j)| \le C_3 |B_j|, \tag{4-12}$$

which allows us to compare $|NLB|$ back to $|G|$.

The inequality (4-12) follows from the following lemma.

**Lemma 4.7.** *Under the same conditions as in Lemma 4.5, $|m(B_{c_0 r}(y))| \le C_3 r^d$. Here $r$ is the same value as in Lemma 4.5, $c_0$ is fixed from Lemma 4.6 and depends only on other universal constants, and $C_3$ depends on $c_0$, $C_4$ (of Lemma 4.6) and the constant $A_1$ of Lemma 4.5.*

In order to prove Lemma 4.7, we only use the equation through Lemma 4.5. Indeed, after fixing a time $t$ and rescaling, it reduces to Lemma 4.6.

We simply sketch the main idea to show how Lemma 4.7 follows from Lemma 4.6.

*Sketch of the proof of Lemma 4.7.* Assume that $u$ and $q$ are as given in the statements of Lemmas 4.5 and 4.7. After a translation, we can assume that $y = 0$. We would then define the rescaled functions

$$\hat{u}(z) = r^{-2} u(rz) \quad \text{and} \quad \hat{q}(z) = r^{-2} q(rz) \quad \text{for } z \in B_2.$$

We note the definition of $\hat{q}$ will be through a minimum over $B_{1/r}$, but, in fact, restricting the minimum to $B_1$ changes nothing since $y = 0$ is such a point that gives the minimum for $\hat{x} = x/r$. Then Lemma 4.6 is applicable with the functions $\hat{u}$ and $\hat{q}$, with the point $x_0 = \hat{x} = x/r$, and the set $\hat{G} = r^{-1} G$, with $G$ being the set arising from the outcome of Lemma 4.5. □

Lemma 4.7 gives (4-12) via the result of Lemma 4.5 and the choice of $c_0 r(y)$.

We will use the fact that $m$ maps onto $G$ as well as the fact that by construction of the subcover $\{B_j\}$, $m^{-1}(G)$ is contained in its union. Thus we see that

$$G = m(m^{-1}(G)) \subset m\Big(\bigcup_j B_j\Big) = \bigcup_j m(B_j),$$

and hence by the choice of $c_0 r(y)$ and the definition of $E_y$, with Lemmas 4.5 and 4.7, it holds that

$$|G| \le \Big|\bigcup_j m(B_j)\Big| \le \sum_j |m(B_j)| \le \sum_j C_3 |B_j| \le \sum_j \frac{C_3}{\delta} |E_{y_j}|.$$

Since the $B_j$ were chosen to be disjoint, so are the corresponding $E_{y_j}$, and thus we can conclude

$$|NL| \ge \Big|\bigcup_j E_{y_j}\Big| = \sum_j |E_{y_j}| \ge \frac{\delta}{C_3} |G| \ge \frac{\delta \delta_1}{C_3}.$$

This finishes the proof of Lemma 4.1. □

## 5. A special barrier function

This section is concerned with the construction of a barrier function that is essential for all of the results regarding regularity of parabolic (and elliptic) equations in nondivergence form. In principle, one would expect our construction to be similar to the one presented in [Chang Lara and Dávila 2014, Lemma 4.2], but this is not actually the case. We deviate in some significant respects due to the additional generality allowed by assumptions (A2) and (A3). In this regard, our construction is more accurately described as a parabolic version of the barrier from [Kassmann et al. 2014, Section 5], where similar lower bounds on only small sets were allowed. Significant detail is required to carry over the ideas from [loc. cit.] to the parabolic setting. These additional difficulties involved in the construction of the barrier are, in fact, also related to the conditions under which the Harnack inequality fails for equations such as (1-1).

Because of the relative strength of the terms $|\nabla p|$ and $M^- p$ under rescaling, it is necessary to break the construction of the special barrier function into two cases: one with $\alpha \geq 1$ and the other with $\alpha < 1$. For the second case, we must remove the gradient term from the equation.

## 5A. *The main lemmas and the barrier.*

**Lemma 5.1.** *Let $\alpha \in [1,2)$ and suppose $r \in (0,1)$ is given. There exists $\varepsilon_0 > 0$, $q_0 > 0$ and a function $p : \mathbb{R}^d \times (0,\infty) \to \mathbb{R}$ such that for all $\alpha \geq 1$,*

$$p_t + C_0|\nabla p| - M^- p \leq 0 \quad \text{in } \big(B_1 \times (0,\infty)\big) \setminus \big(B_r \times (0, r^\alpha]\big), \tag{5-1}$$

$$p \leq 1 \quad \text{in } B_r \times (0, r^\alpha], \tag{5-2}$$

$$p \leq 0 \quad \text{in } \big(\mathbb{R}^d \setminus B_1\big) \times (0,\infty) \text{ and } \big(\mathbb{R}^d \setminus B_r\big) \times \{0\}, \tag{5-3}$$

$$p \geq \varepsilon_0 r^{q_0} e^{-C_5(T - r^\alpha)} \quad \text{in } B_{3/4} \times [r^\alpha, T]. \tag{5-4}$$

*The constants $\varepsilon_0$ and $q_0$ depend only on $\lambda$, $\Lambda$, $\mu$, $C_0$, $\alpha_0$ and dimension.*

**Lemma 5.2.** *Let $\alpha \in [\alpha_0, 2)$ and suppose $r \in (0,1)$ is given. Then the same statement of Lemma 5.1 remains true except* (5-1) *is replaced by*

$$p_t - M^- p \leq 0 \quad \text{in } \big(B_1 \times (0,\infty)\big) \setminus \big(B_r \times (0, r^\alpha]\big). \tag{5-5}$$

**Remark 5.3.** Note that the same constants $\varepsilon_0$ and $q_0$ can be chosen to work for both Lemmas 5.1 and 5.2.

**Remark 5.4.** The existence of the barrier is closely related to uniform estimates on hitting times of a Markov process, which are crucial to the proofs of the weak Harnack inequality and Hölder regularity in the probabilistic framework. These hitting-time estimates appear in the original work of Krylov and Safonov [1980; 1979], and they have become a standard technique in the probability literature (see the presentation in, e.g., the lecture notes [Bass 2004]). In other contexts, there exists an explicit barrier and this lemma looks deceivingly simple. For nonlocal equations whose kernels are allowed to vanish, this step is, in fact, highly nontrivial. Lemmas 5.1 and 5.2 have a probabilistic interpretation as the lower bound for the probability of the process to hit a ball between time 0 and $r^\alpha$.

The strategy for this construction is to start with a yet-to-be-determined function, $\Phi$, supported in $B_1$, and rescale $\Phi$ on the time interval $t \in (0, r^\alpha)$ as

$$p(x,t) = t^{-q_0} \Phi\left(\frac{rx}{t^{1/\alpha}}\right), \tag{5-6}$$

and then to use

$$p(x,t) = e^{-C_5(t-r^\alpha)} p(x, r^\alpha) = e^{-C_5(t-r^\alpha)} r^{-\alpha q_0} \Phi(x) \tag{5-7}$$

for $t \in (r^\alpha, \infty)$. The choice of $rx/t^{1/\alpha}$ is to make sure that $p$ will be positive for all $|x| < 1$ when $t \geq r^\alpha$. The constants $q_0$ and $C_5$ are there to force the subsolution property in the regions where $M^- p$ cannot be made to be as large as we like. We now make some initial computations to illuminate our subsequent choices (note the use of Lemma 2.2):

$$p_t = -q_0 t^{-q_0-1} \Phi\left(\frac{rx}{t^{1/\alpha}}\right) - \tfrac{1}{\alpha} t^{-q_0-1/\alpha-1} \nabla\Phi\left(\frac{rx}{t^{1/\alpha}}\right) \cdot rx, \tag{5-8}$$

$$\nabla p = r t^{-q_0-1/\alpha} \nabla\Phi\left(\frac{rx}{t^{1/\alpha}}\right), \tag{5-9}$$

$$M^- p = t^{-q_0-1} r^\alpha M^- \Phi\left(\frac{rx}{t^{1/\alpha}}\right). \tag{5-10}$$

We want to satisfy (5-1), which then can be transformed to the new goal (at least for $t \in (0, r^\alpha)$):

$$t^{-q_0-1}\left(-q_0\Phi\left(\frac{rx}{t^{1/\alpha}}\right) - \tfrac{1}{\alpha} t^{-1/\alpha} \nabla\Phi\left(\frac{rx}{t^{1/\alpha}}\right) \cdot rx + r t^{1-1/\alpha} C_0 \left|\nabla\Phi\left(\frac{rx}{t^{1/\alpha}}\right)\right| - r^\alpha M^-\Phi\left(\frac{rx}{t^{1/\alpha}}\right)\right) \leq 0. \tag{5-11}$$

Switching out variables

$$z = \frac{rx}{t^{1/\alpha}},$$

for an appropriate set of $z$, we want

$$t^{-q_0-1}\big(-q_0\Phi(z) - \tfrac{1}{\alpha}\nabla\Phi(z)\cdot z + r t^{1-1/\alpha} C_0 |\nabla\Phi(z)| - r^\alpha M^-\Phi(z)\big) \leq 0. \tag{5-12}$$

We can now turn to the requirement for $p$ to satisfy (5-1) when $t \geq r^\alpha$. The computations are similar to the case of $t \in [0, r^\alpha]$. Using (5-6),

$$\begin{aligned} p_t &= -C_5 e^{-C_5(t-r^\alpha)} r^{-\alpha q_0} \Phi(x), \\ \nabla p &= r^{-\alpha q_0} e^{-C_5(t-r^\alpha)} \nabla\Phi(x), \\ M^- p &= r^{-\alpha q_0} e^{-C_5(t-r^\alpha)} M^-\Phi(x). \end{aligned}$$

Then the goal (5-12) becomes

$$e^{-C_5(t-r^\alpha)} r^{-\alpha q_0}\big(-C_5\Phi(x) + C_0|\nabla\Phi(x)| - M^-\Phi(x)\big) \leq 0. \tag{5-13}$$

The function $\Phi$ and subsequently $p$ will be built in a many-staged process. One of the key components is a special bump function, which acts as a barrier in the stationary setting. This construction proceeds similarly to that of [Kassmann et al. 2014], and we would like to point out that there, just as here, there are significant challenges for this construction due to the generality of the lower bound assumption in (2-2) (cf. the bump function in [Caffarelli and Silvestre 2009], where the lower bound on $K$ holds globally).

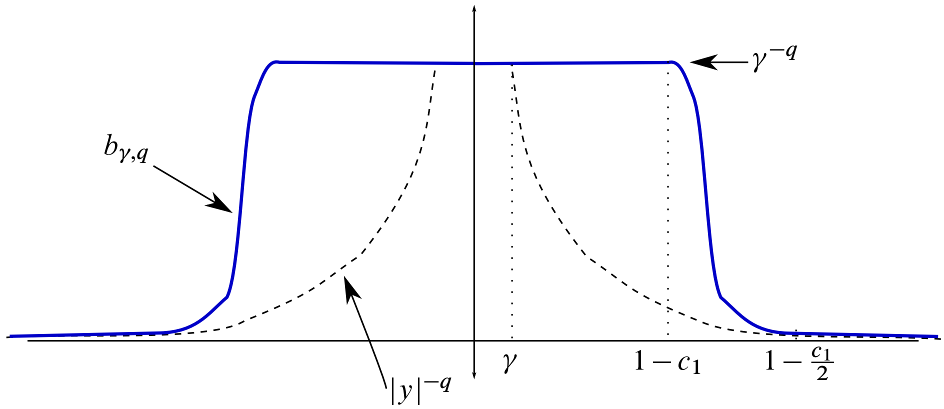


**Figure 1.** The function $b_{\gamma,q}$.

We start with a two-parameter family of auxiliary functions

$$b_{\gamma,q}(y) = \hat{b}(|y|)$$

and

$$\hat{b}(r) = \begin{cases} r^{-q} & \text{if } r \geq 1 - \frac{c_1}{2}, \\ m_{\gamma,q}(r) & \text{if } 1 - c_1 \leq r \leq 1 - \frac{c_1}{2}, \\ \gamma^{-q} & \text{if } r \leq 1 - c_1, \end{cases} \tag{5-14}$$

with $m_{\gamma,q}$ smooth and monotonically decreasing (so there will be a restriction between $\gamma$ and $c_1$ both being small enough), and without loss of generality $m_{\gamma,q}$ will be such that

$$b_{\gamma,q}(y) \geq \min\{\gamma^{-q}, |y|^{-q}\} \quad \text{for all } \ y \in \mathbb{R}^d.$$

See Figure 1 for the graph of $b_{\gamma,q}$.

The key part of the construction is that there are choices of $\gamma$ and $q$ that make $b$ a subsolution in a given small strip (and a subsequent truncation allows the equation to hold in a large set). We state this result for the choices of $\gamma$ and $q$, and then we will prove it in Section 5B.

**Lemma 5.5.** *Let $C > 0$ be given. Then there exist a small constant $c_1$ and choices of $\gamma_1$ and $q_1$ (depending on $C$ plus all other universal objects) such that*

$$M^- b_{\gamma_1,q_1}(x) \geq C q_1 |x|^{-q_1-\alpha} \quad \textit{for all } \ 1 - \frac{c_1}{2} \leq |x| \leq 1, \tag{5-15}$$

*for all $\alpha \in (\alpha_0, 2)$. The constant $c_1$ depends on the lower bound of $K$ in* (2-2).

**Remark 5.6.** Lemma 5.5 provides a subsolution to a stationary problem. It is a generalized version of [Caffarelli and Silvestre 2009, Corollary 9.2; Bjorland et al. 2012, Lemma 3.10; Kassmann et al. 2014, Lemmas 5.2 and 5.3] to the more general class of kernels in this article.

Now that we know the details of an equation for $b = b_{\gamma,q}$, we will continue the calculations, which will be useful to construct $p$. For the following, we assume that $1 - \frac{c_1}{2} \leq |z| \leq 1$. We also note that

$\gamma_1$, $q_1$, and $C$ will be determined subsequently:

$$b(z) = b_{\gamma_1,q_1}(z) = |z|^{-q_1} \quad \text{if } 1 - \tfrac{c_1}{2} < |z|, \tag{5-16}$$

$$\nabla b(z) = -q_1 z |z|^{-q_1-2}, \tag{5-17}$$

$$-\tfrac{1}{\alpha}\nabla b(z)\cdot z = \tfrac{1}{\alpha} q_1 |z|^{-q_1}, \tag{5-18}$$

$$C_0|\nabla b(z)| = C_0 q_1 |z|^{-q_1-1}, \tag{5-19}$$

$$-M^- b(z) \le -C q_1 |z|^{-q_1-\alpha}. \tag{5-20}$$

Now that we have sorted out the details regarding $b_{\gamma_1,q_1}$, we can proceed with the proof of Lemma 5.1. Some complications arise from the need to satisfy the boundary conditions in (5-3).

We will give the proof of Lemma 5.1 and then afterwards indicate the few steps that are modified to prove Lemma 5.2.

*Proof of Lemma 5.1.* We proceed with defining $p$ in terms of $\Phi$ as described in (5-6) and (5-7). Note that this construction gives a function $p$ that is unbounded around the origin $(0, 0)$. To fix that, at the end of the proof, we have an extra truncation step.

In order to satisfy the boundary conditions (5-3), $\Phi$ will be the following truncated version of $b_{\gamma,q}$:

$$\Phi(z) = \max\{b_{\gamma,q}(z) - b_{\gamma,q}(e_1), 0\}.$$

This function $\Phi$ is zero outside of $B_1$ and strictly positive inside $B_1$. The properties of the function $b$ will be used to make the value of $M^-\Phi$ large in $B_1 \setminus B_{1-c_2/2}$.

Recall the variable $z$,

$$z = \frac{rx}{t^{1/\alpha}}. \tag{5-21}$$

We need to verify (5-12) and (5-13) in order to account for the regions $t \in [0, r^\alpha]$ and $t \in (r^\alpha, \infty)$. We will need to select parameters and constants to work for both ranges of $t$. But we note that all of the parameters are such that they can be chosen to satisfy both conditions simultaneously.

**Part 1:** $t \in [0, r^\alpha]$.

Note the following relations for $z \in B_1$:

$$\nabla\Phi(z) = \nabla b(z),$$

$$M^-\Phi(z) \ge M^- b(z).$$

We need to find parameters so that (5-12) holds. The computation will be different in the three regions $|z| \le 1 - \frac{c_1}{2}$, $1 - \frac{c_1}{2} < |z| < 1$, and $|z| \ge 1$.

Replacing (5-17), (5-18), (5-19) and (5-20) in the left-hand side of (5-12), we get

$$\begin{aligned} -q_0\Phi(z) - \tfrac{1}{\alpha}\nabla\Phi(z)\cdot z + rt^{1-1/\alpha}C_0|\nabla\Phi(z)| - r^\alpha M^-\Phi(z) \\ \le -q_0\Phi(z) - \tfrac{1}{\alpha}\nabla b(z)\cdot z + rC_0|\nabla b(z)| - r^\alpha M^- b(z). \end{aligned} \tag{5-22}$$

For the last inequality, we used that $t^{1-1/\alpha} \le 1$. This is because $t \le r^\alpha \le 1$ and $\alpha \ge 1$. When $\alpha < 1$, the negative power of $t$ cannot be controlled and that is why we assume $C_0 = 0$ in those cases.

When $1-\frac{c_1}{2}<|z|<1$, we can ignore $-q_0 b(z)$, and instead focus on

$$-\tfrac{1}{\alpha}\nabla b(z)\cdot z + rC_0|\nabla b(z)| - r^\alpha M^- b(z) \le 0. \tag{5-23}$$

In light of (5-18), (5-19), (5-20), it will suffice to choose $b$ so that

$$\tfrac{1}{\alpha} q_1 |z|^{-q_1} + rC_0 q_1 |z|^{-q_1-1} - Cq_1 r^\alpha |z|^{-q_1-\alpha} \le 0,$$

or more succinctly

$$q_1|z|^{-q_1}\big(\tfrac{1}{\alpha} + rC_0|z|^{-1} - Cr^\alpha |z|^{-\alpha}\big) \le 0. \tag{5-24}$$

After $C$ is chosen to obtain (5-24) (recall $|z|\le 1$), then $b=b_{\gamma_1,q_1}$ can be fixed by Lemma 5.5. The resulting $b$ will be smooth and bounded.

Switching now to the set $|z|\le 1-\frac{c_1}{2}$, inequality (5-22) then follows from

$$\big(-q_0\Phi(z) - \tfrac{1}{\alpha}\nabla b(z)\cdot z + rC_0|\nabla b(z)| - r^\alpha M^- b(z)\big) \le 0. \tag{5-25}$$

The function $\Phi$ is strictly positive in $B_1$ and, in particular, it is bounded below by a positive constant in $B_{1-c_1/2}$. Since $C$, $\gamma_1$, $q_1$ have all been fixed and all of the terms are bounded, we can then choose $q_0$ large enough so that (5-25) will also hold.

We are only left with the case $|z|\ge 1$. Note that because of the angle singularity of the function $\Phi$ on $|z|=1$, we cannot touch the function $\Phi$ from above with any smooth function at those points. Therefore, the points $|z|=1$ play no role in $\Phi$ satisfying (5-12) in the viscosity sense. If $|z|>1$, then $\Phi(z)=|\nabla\Phi(z)|=0$ and $M^-\Phi(z)\ge 0$ because $z$ will be at a global minimum of $\Phi$, and so (5-12) trivially holds.

**Part 2:** $t\in(r^\alpha,\infty)$.

We now need to make sure (5-13) holds. The procedure is similar to the first part.

In the region $1-\frac{c_1}{2}<|x|<1$, using (5-19) and (5-20), we get

$$-C_5\Phi(x) + C_0|\nabla\Phi(x)| - M^-\Phi(x) = -C_5\Phi(x) + C_0 q_1|z|^{-q_1-1} - Cq_1|z|^{-q_1-\alpha}.$$

We ignore the term $-C_5\Phi(x)\le 0$ and use

$$-C_5\Phi(x) + C_0|\nabla\Phi(x)| - M^-\Phi(x) \le q_1|x|^{-q_1}(C_0|x|^{-1} - C|x|^{-\alpha}) \le 0.$$

The last inequality holds provided that we choose $C$ large enough (which can be done by choosing appropriate values of $\gamma$ and $q$ from Lemma 5.5).

In the region $|x|<1-\frac{c_1}{2}$, we use that $b$ (note that $\gamma$ and $q$ are fixed in the previous step) is a given smooth function and $\Phi(x)\ge \big|1-\frac{c_1}{2}\big|^{-q}-1>0$. Therefore, picking a large enough $C_5$, we can make (5-13) hold.

If $|x|\ge 1$, then the equation holds just as in the first part of this proof, owing to the fact that $z$ will be at a global minimum of $\Phi$. Note that the constant $C$, which we use for picking $\gamma_1$ and $q_1$ in Lemma 5.5, needs to be large enough to satisfy the requirements of both Part 1 ($t\in[0,r^\alpha]$) and Part 2 ($t>r^\alpha$) of this proof.

**Part 3:** The truncation step.

Now there is one last step of truncation. At this stage, the function $t^{-q_0}\Phi(rx/t^{1/\alpha})$ has a singularity at $x=0$ and $t\to 0$, which of course violates requirement (5-2).

We define the function
$$\tilde{p}(x,t) := t^{-q_0}\Phi\left(\frac{rx}{t^{1/\alpha}}\right),$$
and $p$ will be defined as a truncation of $\tilde{p}$ to be compatible with (5-2). Importantly, in this truncation we need to not destroy the equation satisfied by our choice of $\tilde{p}$ outside of $B_r \times [0, r^\alpha]$. That means that we should only truncate at a small enough $t$ so that the support of $\tilde{p}(\cdot, t)$ is contained in $B_r$. This way, for such $x$ outside of $B_r$, the desired equation is trivially satisfied because the equation will be evaluated where $\tilde{p}_t = 0$ and $\tilde{p}(x,t) = 0$, which is the global minimum for $\tilde{p}$, giving $\nabla\tilde{p} = 0$ and $M^-\tilde{p} \geq 0$. Given the scaling $z = rx/t^{1/\alpha}$ and that the support of $\Phi$ is in $B_1$, we see that a convenient choice for truncation will be when the graph of $t = (r|x|)^\alpha$ intersects the line $|x| = r$; hence at $t = r^{2\alpha}$.

Accordingly, we define (note for each $t$, we know that $\tilde{p}$ has its max at $x = 0$)
$$\begin{aligned} p(x,t) &= \frac{\min\{\tilde{p}(x,t), \tilde{p}(0, r^{2\alpha})\}}{\tilde{p}(0, r^{2\alpha})} \\ &= (r^{-2\alpha q_0}\Phi(0))^{-1}\min\{\tilde{p}(x,t), r^{-2\alpha q_0}\Phi(0)\}. \end{aligned}$$
This now gives a complete description of $p$ for $t$ in both $(0, r^\alpha)$ and $[r^\alpha, \infty)$ via (5-6) and (5-7) respectively.

The inequality (5-4) follows by a direct inspection using the expression (5-7) for $\tilde{p}$. We get that for $t > r^\alpha$ and $|x| \leq \frac{3}{4}$,
$$p(x,t) = \left(r^{-2\alpha q_0}\Phi(0)\right)^{-1} e^{-C_5(t-r^\alpha)} r^{-\alpha q_0}\Phi(x) \geq r^{\alpha q_0} e^{-C_5(t-r^\alpha)} \min_{B_{3/4}} \Phi.$$

We note that the truncation expression has shown that the choice of $q$ for the lower bound requirement in (5-4) will be $q = \alpha q_0$. The choice of radius $\frac{3}{4}$ in (5-4) is irrelevant, since a similar lower bound would hold if $\frac{3}{4}$ is replaced by any other number smaller than 1.

This completes the proof of Lemma 5.1. ☐

We now mention where the proof of Lemma 5.2 deviates from the previous one.

*Proof of Lemma 5.2.* One needs to go back and remove the term $C_0|\nabla p|$ from all of the calculations. Note this was the only term affected by the factor $t^{1-\alpha/2}$, which would be unbounded if $\alpha < 1$. ☐

**5B. *The proof of Lemma 5.5.*** Lemma 5.5 will be attained in two stages, Lemmas 5.10 and 5.11. First we develop some auxiliary results related to $b$. We begin by making a useful observation about the behavior of $\delta_h b$.

**Lemma 5.7.** *Assume* $\alpha \in [1,2)$. *If* $b = b_{\gamma,q}$ *is as in* (5-14), *then for some universal* $r_0$ *and* $C(q)$, *where* $|h| \leq r_0$ *and* $1 - \frac{c_1}{2} < |x| < 1$, *we have*
$$\delta_h b(x) \geq -q\frac{|h|^2}{|x|^{q+2}} + q(q+2)\frac{(h_1)^2}{|x|^{q+2}} - C(q)|h|^3$$
(*this is only relevant*, *and only invoked*, *for* $\alpha > 1$; *otherwise we would use a different expansion for* $\alpha < 1$).

*Proof.* This follows from Taylor's theorem. Note that $h$ is restricted to be in a small set, $B_{r_0}$, and so actually $b(x) = |x|^{-q}$ and $b(x+h) \geq |x+h|^{-q}$. ☐

The next lemma says that our assumptions allow that for all $r \le r_1$, the set $A_r$ intersects annuli centered at $-e_1$ in a uniformly nontrivial fashion. This feature is essential to be able to utilize the lower bounds on $K$ in (2-2).

**Lemma 5.8.** *There exist constants $c_1$, $c_2$ and $r_1$ (all small), so that*

(i) *for any $x$ so that $1-c_1 < |x| < 1$,*

$$\left|A_{r_1} \cap B_{1-c_1}(-x)\right| \ge \tfrac{1}{4}\mu\left|B_{2r_1} \setminus B_{r_1}\right|,$$

(ii) *for all $r$,*

$$\left|A_r \cap \{h : (h_1)^2 \ge c_2|h|^2\}\right| \ge \tfrac{1}{2}\mu\left|B_{2r} \setminus B_r\right|.$$

*Proof.* We first note that by the symmetry of $A_r$,

$$\left|A_r \cap (B_{2r} \setminus B_r) \cap \{h : h\cdot x \le 0\}\right| \ge \tfrac{1}{2}\mu\left|B_{2r} \setminus B_r\right|. \tag{5-26}$$

Now we will establish (i). We first choose $r_1$ small enough so that

$$\left|\left((B_{2r_1} \setminus B_{r_1}) \cap \{h : h\cdot x \le 0\}\right) \setminus B_{|x|}(-x)\right| \le \tfrac{1}{8}\mu\left|(B_{2r_1} \setminus B_{r_1}) \cap \{h : h\cdot x \le 0\}\right|.$$

Note that this choice of $r_1$ can be done uniformly for all $1-c_1 < |x| < 1$.

Let us define the failed set where $A_r$ cannot reach $B_{1-c_1}(-x)$ as

$$F := \left((B_{2r_1} \setminus B_{r_1}) \cap \{h : h\cdot x \le 0\}\right) \setminus B_{1-c_1}(-x).$$

With $r_1$ fixed, we can choose $c_1$ small enough so that

$$|F| \le \tfrac{1}{4}\mu\left|(B_{2r_1} \setminus B_{r_1}) \cap \{h : h\cdot x \le 0\}\right|. \tag{5-27}$$

This is possible because

$$\begin{aligned} |F| &\le \left|\left((B_{2r_1} \setminus B_{r_1}) \cap \{h : h\cdot x \le 0\}\right) \setminus B_{|x|}(-x)\right| + \left|B_{|x|} \setminus B_{1-c_1}\right| \\ &\le \tfrac{1}{8}\mu\left|(B_{2r_1} \setminus B_{r_1}) \cap \{h : h\cdot x \le 0\}\right| + C(1-(1-c_1)^d). \end{aligned}$$

Finally, combining (5-26) with (5-27) we obtain (i).

To establish (ii), we note that

$$\begin{aligned} \left|A_r \cap \{h : (h_1)^2 \ge c_2|h|^2\}\right| &\ge |A_r| - \left|\{h \in B_{2r} \setminus B_r : h_1^2 < c_2|h|^2\}\right| \\ &\ge (\mu - Cc_2)|B_{2r} \setminus B_r| \end{aligned}$$

for a universal constant $C$. Thus, we simply take $c_2$ small enough so that $(\mu - Cc_2) \ge \frac{1}{2}\mu$. □

**Note 5.9.** If $\gamma_1 < \gamma_2$ and $q$ is fixed, then for all $y$,

$$b_{\gamma_1,q}(y) \ge b_{\gamma_2,q}(y),$$

and the two functions are equal when $|y| \ge 1 - \frac{c_1}{2}$; hence

$$M^- b_{\gamma_1,q}(x) \ge M^- b_{\gamma_2,q}(x)$$

for all $|x| \ge 1 - \frac{c_1}{2}$.

Next we make the first choice of parameter for $b$. It is the selection of the exponent, $q$, and it only uses the information about the family $\mathcal{K}$ for $\alpha$ very close to 2.

**Lemma 5.10.** *Let $\gamma \leq \gamma_0 = \frac{1}{4}$ be fixed. Let $C > 0$ be given. Then, there exist a $q_1 \geq 1$ and an $\alpha_1$, depending only on $C$, $\gamma_0$, $C_0$, $\mu$, $d$, $\lambda$, $\Lambda$, such that*

$$M^- b_{\gamma,q_1}(x) \geq C q_1 |x|^{-q_1-\alpha} \quad \text{for all } 1 - \tfrac{c_1}{2} < |x| < 1,$$

*for all orders, $\alpha \in (\alpha_1, 2)$ and for all $\gamma \leq \gamma_0$.*

Then once the $q$ has been chosen, we can finish the definition of $b$ by fixing the truncation height, $\gamma^{-q}$, to be large enough (so $\gamma$ small enough). This allows us to fix one function that satisfies the special subsolution property for all $\alpha \in [\alpha_0, 2)$.

**Lemma 5.11.** *Let $C > 0$ and $q_1$ be as in Lemma 5.10. Then there exists a $\gamma_1 \leq \gamma_0 = \frac{1}{4}$ such that*

$$M^- b_{\gamma_1,q_1}(x) \geq C q_1 |x|^{-q_1-\alpha} \quad \text{for all } 1 - \tfrac{c_1}{2} < |x| < 1,$$

*for all orders, $\alpha \in (\alpha_0, \alpha_1]$.*

First we give the proof of Lemma 5.10.

*Proof of Lemma 5.10* . Let $x$ be any point such that $1 - \frac{c_1}{2} < |x| < 1$. We begin with a few simplifying observations. First of all, there is no loss of generality in assuming $\alpha > 1$ for this lemma — indeed the end of the proof culminates with a choice of $\alpha_1$ that is sufficiently close to 2 (hence $\delta_h b(x)$ uses only one case for $\alpha > 1$). Second, to simplify notation, we drop the $\gamma, q$ dependence and denote $b_{\gamma,q}$ by $b$.

To obtain the bound we want, we only need the contribution of $\delta_h b(x)$ to $M^- b(x)$ in a small ball, $h \in B_{r_2}$, for some $r_2$ fixed with, say, $r_2 = \min\{r_0, \frac{c_1}{2}\}$, where $r_0$ originates in Lemma 5.7 and $c_1$ comes from Lemma 5.8. This is because the large curvature of the graph of $b$ in the $h_1$-direction can be used to dominate the integral at the expense of all the other terms.

We also note that for $h \in \mathbb{R}^d \setminus B_{r_2}$, we have

$$\delta_h b(x) \geq \inf_{h \in \mathbb{R}^d \setminus B_{r_2}} \big(b(x+h) - b(x) - q|x|^{-q-2} x \cdot h\big) \geq -C_q \Big(1 + \frac{x}{|x|} \cdot h\Big).$$

Here $C_q = \max\big(q\big(1 - \frac{c_1}{2}\big)^{-q-1}, \big(1 - \frac{c_1}{2}\big)^{-q}\big)$.

Therefore, by Lemma 2.3, we see that

$$\int_{\mathbb{R}^d \setminus B_{r_2}} \delta_h b(x) K(h) \, \mathrm{d}h \geq -(2-\alpha) C_q \Lambda \bigg(\frac{r_2^{-\alpha}}{\alpha} + r_2^{1-\alpha}\bigg). \tag{5-28}$$

Furthermore, combining Lemmas 5.8 and 5.7, we see that on each ring, $B_{2^{-k}r} \setminus B_{2^{-k-1}r}$, we can enhance the positive contribution to $M^- f(x)$ by manipulating the term

$$\frac{q(q+2)}{|x|^{q+2}} \int_{B_{2^{-k}r} \setminus B_{2^{-k-1}r}} (h_1)^2 K(h) \, \mathrm{d}h.$$

By Lemma 5.8 and assumption (A3), we see that

$$\begin{aligned}\int_{B_{2^{-k}r}\setminus B_{2^{-k-1}r}} (h_1)^2 K(h)\,\mathrm{d}h &\ge \int_{A_{2^{-k-1}r}} (h_1)^2 K(h)\,\mathrm{d}h \\ &\ge \int_{A_{2^{-k-1}r}\cap\{h:(h_1)^2\ge c_2|h|^2\}} c_2|h|^2 K(h)\,\mathrm{d}h \\ &\ge c_2(2^{-k-1}r)^2\lambda(2-\alpha)(2^{-k-1}r)^{-d-\alpha}\left|A_{2^{-k-1}r}\cap\{h : (h_1)^2\ge c_2|h|^2\}\right| \\ &\ge c_2(2^{-k-1}r)^2\lambda(2-\alpha)(2^{-k-1}r)^{-d-\alpha}\tfrac{1}{2}\mu\left|B_{2^{-k}r}\setminus B_{2^{-k-1}r}\right| \\ &= c_2\lambda(2-\alpha)\mu 2c(d)r^{2-\alpha}2^{-k(\alpha-2)},\end{aligned}$$

where $c(d)$ is a purely dimensional constant that we use temporarily during this proof. Hence adding up the contribution along all of the rings, we see

$$\int_{B_{r_2}} (h_1)^2 K(h)\,\mathrm{d}h = \sum_{k=0}^{\infty}\int_{B_{2^{-k}r_2}\setminus B_{2^{-k-1}r_2}} (h_1)^2 K(h)\,\mathrm{d}h \ge \big(\lambda\mu c_2 c(d)\big) r_2^{2-\alpha}, \tag{5-29}$$

where we have collected various dimensional constants into $c(d)$ in such a way that is uniform for $\alpha\in(0,2)$. Note that

$$\sum_{k=0}^{\infty}(2-\alpha)2^{k(\alpha-2)} = \frac{2-\alpha}{1-2^{\alpha-2}} \le 2$$

for all $\alpha\in(1,2)$.

We also estimate the following integral using assumption (A2):

$$\int_{B_{r_2}} |h|^3 K(h)\,\mathrm{d}h = \sum_k \int_{B_{2^{-k}r_2}\setminus B_{2^{-k-1}r_1}} |h|^3 K(h)\,\mathrm{d}h \le \frac{(2-\alpha)2^{\alpha}}{1-2^{\alpha-3}} r_2^{3-\alpha}\Lambda. \tag{5-30}$$

Now we need to put all of the pieces together. We will use Lemma 5.7 to balance the terms of different orders in both $|h|$ and $q$. We will be invoking Lemma 2.3 as well as the bounds from (5-28)–(5-30):

$$\begin{aligned}&\int_{\mathbb{R}^d}\delta_h b(x)K(h)\,\mathrm{d}h \\ &= \int_{B_{r_2}}\delta_h b(x)K(h)\,\mathrm{d}h + \int_{\mathbb{R}^d\setminus B_{r_2}}\delta_h b(x)K(h)\,\mathrm{d}h \\ &\ge \frac{q(q+2)}{|x|^{q+2}}\int_{B_{r_2}}(h_1)^2K(h)\,\mathrm{d}h - \frac{q}{|x|^{q+2}}\int_{B_{r_2}}|h|^2K(h)\,\mathrm{d}h - C(q)\int_{B_{r_2}}|h|^3K(h)\,\mathrm{d}h + \int_{\mathbb{R}^d\setminus B_{r_2}}\delta_h f(x)K(h)\,\mathrm{d}h \\ &\ge \frac{q}{|x|^{q+2}}\big((q+2)(\lambda\mu c_2 c(d)) - C_d\Lambda\big)r_2^{2-\alpha} - (2-\alpha)\left(C_q\Lambda\left(\frac{r_2^{-\alpha}}{\alpha}+r_2^{1-\alpha}\right) - C(q)\frac{2^{\alpha}}{1-2^{\alpha-3}}r_2^{3-\alpha}\Lambda\right).\end{aligned} \tag{5-31}$$

At this point, we note that the first term is the one that does not have the factor $(2-\alpha)$ in front. We will first choose $q$ large to control the sign of this term. Hence we can choose $q=q_1$ large enough, depending only on the given constant $C$ and the universal parameters, so that (recall $C$, with no subscript, was the parameter given in the statement of this lemma and $|x| < 1$)

$$\frac{q}{|x|^{q+2}}\big((q+2)(\lambda\mu c_2 c(d)) - C_d\Lambda\big)r_2^{2-\alpha} \geq 3Cq|x|^{-q-\alpha}r_2^{2-\alpha}.$$

Once $q_1$ has been fixed, we can now choose $\alpha_1$ close enough to 2 so that the rest of the expression in (5-31) is small:

$$(2-\alpha)\left(C_q\Lambda\left(\frac{r_2^{-\alpha}}{\alpha}+r_2^{1-\alpha}\right) - C(q)\frac{2^\alpha}{1-2^{\alpha-3}}r_2^{3-\alpha}\Lambda\right) \leq Cq_1|x|^{-q_1-\alpha}r_2^{2-\alpha}.$$

(Recall that $r_2 = \min\{r_0, \frac{c_1}{2}\}$.) Thus we have achieved

$$\int_{\mathbb{R}^d}\delta_h b(x)K(h)\,\mathrm{d}h \geq 2Cq_1|x|^{-q_1-\alpha}r_2^{2-\alpha}.$$

The chosen value of $\alpha$ is sufficiently close to 2. We may choose $\alpha$ even closer to 2 so that $r_2^{2-\alpha} > \frac{1}{2}$ and

$$\int_{\mathbb{R}^d}\delta_h b(x)K(h)\,\mathrm{d}h \geq Cq_1|x|^{-q_1-\alpha}.$$

Taking an infimum over $K$ yields the result. □

**Remark 5.12.** The underlying reason why the previous proof works is because if we fix the values of $\Lambda$, $\lambda$ and $\mu$, the following limit holds:

$$\lim_{\alpha\to 2} M^- b(x) = \mathcal{M}^-_{\tilde{\lambda},\tilde{\Lambda}}(D^2 b(x)),$$

where $\mathcal{M}^-$ is the classical minimal Pucci operator of order 2 and $\tilde{\lambda}$, $\tilde{\Lambda}$ are ellipticity constants that depend on $\lambda$, $\Lambda$, $\mu$ and dimension. The proof of this fact goes along the same lines as the proof of Lemma 5.10.

**Remark 5.13.** We note that the statement and proof of Lemma 5.10 here, combined with step 1 of the proof of Lemma 5.1, corrects an error in the construction of the similar barrier used in [Kassmann et al. 2014, Section 5], where the truncation step should have been done first, not at the end of the construction.

Now we can conclude this section with the proof of Lemma 5.11.

*Proof of Lemma 5.11.* Let $x$ be any point such that $1-\frac{c_1}{2} < |x| < 1$. First of all, we note that $q_1$ has been fixed already, so we will drop it from the notation. Since we will be manipulating the choice of $\gamma$ to obtain the desired bound on $M^- b_{\gamma,q_1}(x)$, it will be convenient to have bounds that transparently do not depend on $\gamma$. Therefore, as above, we keep $\gamma_0 = \frac{1}{4}$ fixed and we will use an auxiliary function to make some of the estimates. Let $\varphi$ be any function in $C^2(\mathbb{R}^d)$ such that

$$0 \leq \varphi \leq b_{\gamma_0,q_1} \quad \text{in } \mathbb{R}^d,$$

and

$$\varphi(x) = |x|^{-q_1} \quad \forall\, |x| \geq 1-\tfrac{c_1}{2}.$$

We note that these definitions imply $\|\varphi\|_{C^2}$ can be chosen to be independent of $\gamma$ (depending on universal parameters plus $\gamma_0$, $q_1$).

We now estimate the contributions from the positive and negative parts of $(\delta_h b(x))^{\pm}$ separately. The first estimate below is simply a use of the fact that by construction, $\varphi$ touches $b$ from below at $x$, and the second one uses (5-28):

$$\begin{aligned}\int_{\mathbb{R}^d} (\delta_h f(x))^- K(h)\,\mathrm{d}h &\le \int_{B_{r_1}} C(d)\big(\|\varphi\|_{C^{1,1}(B_{1/2}(x))}\big)|h|^2 K(h)\,\mathrm{d}h + \int_{\mathbb{R}^d\setminus B_{r_1}} (\delta_h f(x))^- K(h)\,\mathrm{d}h \\ &\le C_d C(d)\|\varphi\|_{C^{1,1}(B_{1/2}(x))}\Lambda r_1^{2-\alpha} + C_d\frac{\Lambda}{\alpha} r_1^{-\alpha} + q_0 C_d \Lambda r_1^{1-\alpha}. \qquad (5\text{-}32)\end{aligned}$$

Now we move to $(\delta_h b(x))^+$. Here we will use Lemma 5.8(i), the important feature being that there is *at least one* good ring where $(\delta_h b(x))^+$ will see the influence of the value of $b$ on the set $B_{1-c_2}$. We alert the reader to a strange term in line (5-33) below, which arises simply as a worst case scenario of the three definitions of $\delta_h$, and, for example, if $\alpha < 1$, the term would not even be necessary. It does not harm the computation, and so we leave it there for any of the possible three cases of $\delta_h$ via $\alpha$. Finally we note the important feature that we may only integrate on the set $h \in B_{1-c_1}(-x)$, which allows us to avoid the singularity of $K$ at $h = 0$. Also note if $h \in B_{1-c_1}(-x)$, then $|h| \le 2$:

$$\begin{aligned}&\int_{\mathbb{R}^d} (\delta_h f(x))^+ K(h)\,\mathrm{d}h \\ &\ge \int_{A_{r_1}\cap B_{1-c_1}(-x)} (\delta_h f(x))^+ K(h)\,\mathrm{d}h \\ &\ge \int_{A_{r_1}\cap B_{1-c_1}(-x)} (\gamma^{-q_1} - |x|^{-q_1}) K(h)\,\mathrm{d}h - q_1|x|^{-q_1-1}\int_{B_{1-c_1}(-x)} |h| K(h)\,\mathrm{d}h \qquad (5\text{-}33) \\ &\ge \big(\gamma^{-q_1} - \big(1-\tfrac{c_1}{2}\big)^{-q_1}\big)(2-\alpha)\lambda \int_{A_{r_1}\cap B_{1-c_1}(-e_1)} |h|^{-d-\alpha}\,\mathrm{d}h - q_1\big(1-\tfrac{c_1}{2}\big)^{-q_1-1}\int_{B_{1-c_1}(-e_1)} 2K(h)\,\mathrm{d}h \qquad (5\text{-}34) \\ &\ge \big(\gamma^{-q_1} - \big(1-\tfrac{c_1}{2}\big)^{-q_1}\big)(2-\alpha)\lambda r_1^{-d-\alpha}\tfrac{1}{4}\mu\big|B_{2r_1}\setminus B_{r_1}\big| - q_1\big(1-\tfrac{c_1}{2}\big)^{-q_1-1}(2-\alpha)C(d,\alpha_0). \qquad (5\text{-}35)\end{aligned}$$

We note the use of (2-8) in the transition between the last two lines.

Recall that the values of $c_1$ and $q_1$ were fixed in Lemmas 5.8 and 5.10. In order to conclude the proof, we see that we can choose $\gamma = \gamma_1$ large enough so that when we add together the contributions from (5-32) and (5-35), the final estimate becomes greater than $C > q_1$ for all $\alpha \in (\alpha_0, \alpha_1)$. We note that it is crucial to have $\alpha \le \alpha_1 < 2$ in this case in order to keep $\alpha$ uniformly away from 2, which would cause problems. □

## 6. An estimate in $L^\varepsilon$ — the weak Harnack inequality

The purpose of this section is to combine the point-to-measure estimate with the special barrier to prove the $L^\varepsilon$ estimate, also called the *weak Harnack inequality*.

**Theorem 6.1** (the $L^\varepsilon$ estimate). *Assume $\alpha \ge \alpha_0 > 0$. Let $u$ be a function such that*

$$u \ge 0 \quad \text{in } \mathbb{R}^d \times [-1,0],$$
$$u_t + C_0|\nabla u| - M^- u \le C \quad \text{in } Q_1,$$

*and for the case $\alpha < 1$, further assume $C_0 = 0$. Then there are constants $C_6$ and $\varepsilon$ such that*

$$\left( \int_{B_{1/4}\times[-1,-2^{-\alpha}]} u^\varepsilon \mathrm{d}x\, \mathrm{d}t \right)^{1/\varepsilon} \le C_6 (\inf_{Q_{1/4}} u + C).$$

*The constants $C_6$ and $\varepsilon$ depend on $\alpha_0$, $\lambda$, $\Lambda$, $C_0$, $d$ and $\mu$.*

Note that the $L^\varepsilon$ norm of $u$ is computed in the cylinder $B_{1/4} \times [-1, -2^{-\alpha}]$. This cylinder lies earlier in time than the cylinder $Q_{1/2}$, where the infimum is taken in the right-hand side of the inequality. This is natural due to the causality effect of parabolic equations. What should be noted in this case is that, due to the scaling of the equation, the size of these cylinders varies. Indeed, if $\alpha \in (1, 2)$, then the time interval $[-1, -2^{-\alpha}]$ is longer than $\frac{1}{2}$ and certainly longer than $[-4^{-\alpha}, 0]$, which is the time span of $Q_{1/4}$. However, for small values of $\alpha$, the length of $[-1, -2^{-\alpha}]$ becomes arbitrarily small and the time span of $Q_{1/4}$ is almost 1. We still have uniform choices of the constants $C$ and $\varepsilon$ because of the assumption $\alpha \ge \alpha_0 > 0$.

The basic building block of this proof is Lemma 4.1, which needs to be combined with Lemmas 5.1 and 5.2 as well as a covering argument. Since the work of Krylov and Safonov [1980], it is known that these ingredients lead to Theorem 6.1. However, there are several ways to organize the proof and there are some subtleties that we want to point out. Thus, we describe the full proof explicitly. We start with some preparatory lemmas.

The following lemma plays the role of Corollary 4.26 in [Imbert and Silvestre 2013b], which the reader can compare with [Chang Lara and Dávila 2014, Corollary 5.2]. Recall the notation $Q_r(x,t) = B_r(x) \times [t - r^\alpha, t]$. We now define a time shift of the cylinder $Q$, which we call $\bar{Q}^m$. For any positive number $m$, we write $\bar{Q}^m$ to denote

$$\bar{Q}^m = B_r(x) \times (t, t + m r^\alpha).$$

The cylinder $\bar{Q}^m$ starts exactly where $Q$ ends (see Figure 3). Moreover, its time span is enlarged by a factor $m$. Because of the order of causality, the information we have about the solution $u$ in $Q$ propagates to $\bar{Q}^m$. This is reflected in the following lemma.

**Lemma 6.2** (stacked point estimate). *Let $m$ be a positive integer. There exist $\delta_2 > 0$ and $N > 0$ depending only on $\lambda$, $\Lambda$, $d$, $\alpha_0$ and $m$ such that if for some cylinder $Q = Q_\rho(x_0, t_0) \subset Q_1$, we have*

$$u \ge 0 \text{ in } \mathbb{R}^d \times [-1, 0], \tag{6-1}$$

$$u_t + C_0 |\nabla u| - M^- u \ge 0 \quad \text{in } Q_1, \tag{6-2}$$

$$\left| \{u \ge N\} \cap Q_\rho(x_0, t_0) \right| \ge (1 - \delta_2) |Q_\rho|, \tag{6-3}$$

$$B_{2\rho}(x_0) \times [t_0 - \rho^\alpha, t_0 + m\rho^\alpha] \subset Q_1, \tag{6-4}$$

*then $u \ge 1$ in $\bar{Q}^m = B_\rho(x_0) \times [t_0, t_0 + m\rho^\alpha]$.*

*Proof.* Let $\tilde{u}$ be the scaled function

$$\tilde{u}(x,t) = \frac{A_0}{N} u(\rho x + x_0, \rho^\alpha t + t_0),$$

where $A_0$ is the constant from Lemma 4.1.

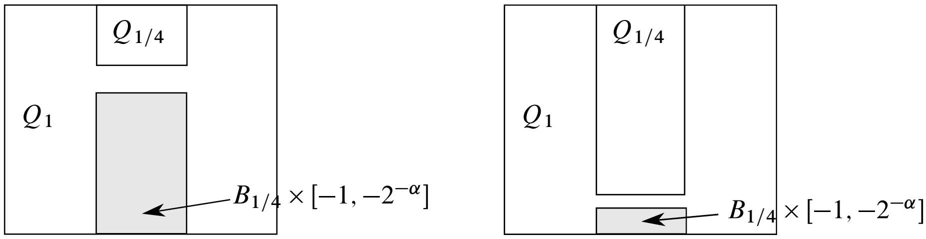


**Figure 2.** The cylinders $Q_{1/4}$ and $B_{1/4} \times [-1, -2^{-\alpha}]$ with large $\alpha$ (left) and small $\alpha$ (right).

Both $u$ and $\tilde{u}$ satisfy (6-2). From our assumption (6-3), we have that

$$\left|\{\tilde{u} > A_0\} \cap Q_1\right| \geq (1-\delta_2)|Q_1|.$$

Applying the contrapositive of Lemma 4.1, we obtain that $\tilde{u} \geq 1$ in $Q_{1/4}$. Thus,

$$u \geq \frac{N}{A_0} \quad \text{in } Q_{\rho/4}(x_0, t_0).$$

Recall that $u$ is a supersolution in $Q_1$ and $u \geq 0$ everywhere. We apply Lemmas 5.1 or 5.2 with $r = \frac{1}{2}$ to obtain the subsolution, $p$, and we can compare the functions $\tilde{u}$ and $p$. Writing this in terms of $u$ gives

$$u(x,t) \geq \frac{N}{M} p\left(\frac{(x-x_0)}{\rho}, \frac{(t-t_0+(\rho/4)^\alpha)}{\rho^\alpha}\right).$$

The conclusion follows from taking $N$ large enough, combined with the lower bound for $p$ given in Lemma 5.1. □

The point of the previous lemma is that it can be combined with the *crawling ink spots* theorem. This is a covering argument that can be used as an alternative to the Calderón–Zygmund decomposition, and it is close to the original argument by Krylov and Safonov in [1980]. It has the cosmetic advantage that it does not use cubes but only balls. Moreover, the Calderón–Zygmund decomposition uses that we can tile the space with cubes, which is only true for $\alpha = 1$. In [Chang Lara and Dávila 2014], this difficulty is overcome by a special tiling with variable scaling, which is explained by the beginning of Section 4.2. It is a cumbersome construction to define rigorously. The use of the crawling ink spots theorem completely avoids this difficulty.

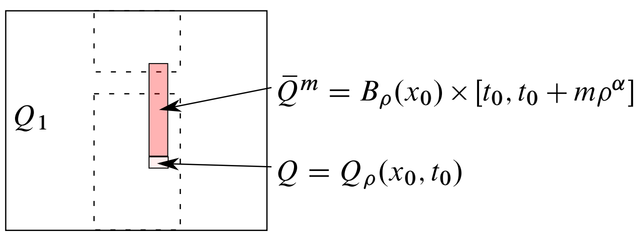


**Figure 3.** The cylinders involved in Lemma 6.2.

**Theorem 6.3** (crawling ink spots). *Let $E \subset F \subset B_{1/2} \times \mathbb{R}$. We make the following two assumptions*:

- *For every point $(x,t) \in F$, there exists a cylinder $Q \subset B_1 \times \mathbb{R}$ so that $(x,t) \in Q$ and $|E \cap Q| \le (1-\mu)|Q|$.*
- *For every cylinder $Q \subset B_1 \times \mathbb{R}$ such that $|E \cap Q| > (1-\mu)|Q|$, we have $\bar{Q}^m \subset F$.*

*Then*

$$|E| \le \frac{m+1}{m}(1-c\mu)|F|.$$

*Here $c$ is an absolute constant depending on dimension only.*

The proof of Theorem 6.3 will be presented in the Appendix. The crawling ink spots theorem is used with a value of $m$ sufficiently large so that $\frac{m+1}{m}(1-c\delta) < 1$. In order to prove the $L^\varepsilon$ estimate, we would want to apply Theorem 6.3 with

$$E = \{u \ge N^{k+1}\} \cap B_{1/2} \cap (-1, -2^{-\alpha}) \quad \text{and} \quad F = \{u \ge N^k\} \cap B_{1/2} \cap (-1, -2^{-\alpha}).$$

The problem is that the assumption of Theorem 6.3 is not implied by Lemma 6.2 because there is no way to ensure that $t + mr^\alpha \le -2^{-\alpha}$. This is a difficulty that is nonexistent in the elliptic setting. Because of the time shift in all the point estimates, the conclusion of the crawling ink spots theorem may be spilling outside of the time interval $[-1, -2^{-\alpha}]$. There is no trivial workaround for this.

The purpose of the following lemma is to show that the cylinders $Q_\rho(x_0, \rho_0)$ that satisfy the condition of the crawling ink spots theorem are necessarily small, and consequently the amount of measure that *leaks* outside the cylinder $B_{1/4} \times [-1, -2^{-\alpha}]$ will decay exponentially.

**Lemma 6.4.** *Assume that*

$$\inf_{Q_{1/4}} u \le 1,$$

$$u \ge 0 \quad \text{in } \mathbb{R}^d \times [-1, 0],$$

$$u_t + C_0|\nabla u| - M^- u \ge 0 \quad \text{in } Q_1,$$

*and that there is a cylinder $Q_\rho(x_0, t_0)$ such that*

$$Q_\rho(x_0, t_0) \subset B_{1/4} \times [-1, -2^{-\alpha}],$$

$$\left|\{u \ge N\} \cap Q_\rho(x_0, t_0)\right| \ge (1-\delta_2)|Q_\rho|.$$

*Then $\rho < C\, N^{-\gamma}$ for some universal $\gamma > 0$ and $C > 0$.*

*Proof.* Applying Lemma 4.1 rescaled to $Q_\rho(x_0, t_0)$, we obtain that $u \ge N/M$ in $Q_{\rho/4}(x_0, t_0)$. Just as in the proof of Lemma 6.2, we get

$$u(x,t) \ge \frac{N}{M} p\left(\tfrac{4}{3}(x - x_0), \left(\tfrac{4}{3}\right)^\alpha \left(t - t_0 + \left(\tfrac{1}{4}\rho\right)^\alpha\right)\right),$$

where $p$ is the function from Lemmas 5.1 or 5.2 with $r = \frac{1}{3}\rho$. The reason for the factor $\frac{4}{3}$ is that since $x_0 \in B_{1/4}$, we know that $B_{3/4}(x_0) \subset B_1$.

We have that $x_0 \in B_{1/4}$, $t_0 \in [-1, -2^{-\alpha}]$ and $\rho \le \min\left(\frac{1}{4}, (1-2^{-\alpha})^{1/\alpha}\right)$. Since $\inf_{Q_{1/4}} u \le 1$, we have

$$\frac{M}{N} \ge \inf\left\{p(x,t) \,:\, x \in B_{2/3} \wedge t \in \left[(3^{-\alpha}(2^\alpha - 1) + \left(\tfrac{1}{3}\rho\right)^\alpha, \left(\tfrac{4}{3}\right)^\alpha + \left(\tfrac{1}{3}\rho\right)^\alpha\right]\right\} \ge c\rho^q,$$

which holds by (5-4) in Lemmas 5.1 and 5.2. Therefore $\rho < CN^{-\gamma}$, where $\gamma = \frac{1}{q}$ and $q$ is the exponent from Lemma 5.1 or 5.2. □

*Proof of Theorem 6.1.* We start by noting that we can assume $C = 0$. Otherwise we consider $\tilde{u}(x,t) = u(x,t) - Ct$ instead. For every positive integer $k$, let

$$A_k := \{u > N^k\} \cap (B_{1/4} \times (-1, -2^{-\alpha})),$$

where $N$ is the constant from Lemma 6.2. We apply Theorem 6.3 with

$$E = \{u \geq N^{k+1}\} \cap \big(B_{1/4} \times (-1, -2^{-\alpha})\big) \quad \text{and} \quad F = \{u \geq N^k\} \cap \big(B_{1/4} \times (-1, -2^{-\alpha} + CmN^{-\gamma\alpha k})\big),$$

where $C$ and $\gamma$ are the constants from Lemma 6.4.

Let us verify that both assumptions of Theorem 6.3 are satisfied. The first assumption in Theorem 6.3 is implied by Lemma 6.4 (at least when $N$ and/or $k$ are large). Indeed, any point

$$(x,t) \in B_{1/4} \times (-1, -2^{-\alpha} + mN^{-\gamma\alpha k})$$

is contained in some cylinder $Q_r(x_0, t_0)$ with large enough $\rho$ so that $\rho > CN^{-k\gamma}$. Because of Lemma 6.2, whenever there is a cylinder $Q$ such that $|A_{k+1} \cap Q| \geq (1-\delta)|Q|$, we know that $\bar{Q}^m \subset \{u > N^k\}$. Moreover, because of Lemma 6.4, the length in time of $\bar{Q}^m$ is less than $mCN^{-\gamma k}$. Therefore $\bar{Q}^m \subset F$. Thus, the second assumption of Theorem 6.3 holds as well.

Note that we allow the result of the crawling ink spots theorem to spill to the time interval

$$[-2^{-\alpha}, -2^{-\alpha} + CmN^{-\gamma\alpha k}].$$

Therefore,

$$|A_{k+1}| \leq \frac{m+1}{m}(1-c\delta)\big(|A_k| + CmN^{-\gamma\alpha k}\big).$$

We first pick $m$ sufficiently large so that

$$\frac{m+1}{m}(1-c\delta) := 1 - \mu < 1.$$

Thus, we have

$$|A_{k+1}| \leq (1-\mu)\big(|A_k| + CmN^{-\gamma\alpha k}\big).$$

This already implies an exponential decay on $|A_k|$, which proves the theorem. □

## 7. Hölder continuity of solutions

We first state a Hölder continuity for parabolic integral equations without drift. In this case, $\alpha \in (0,2)$ can be arbitrarily small, although the estimates depend on its lower bound $\alpha_0$.

**Theorem 7.1** (Hölder estimates without drift). *Assume $\alpha \geq \alpha_0 > 0$. Let $u$ be a bounded function in $\mathbb{R}^d \times [-1,0]$ such that*

$$u_t - M^+ u \leq C \quad \text{in } Q_1,$$
$$u_t - M^- u \geq -C \quad \text{in } Q_1.$$

*Then there are constants $C_7$ and $\gamma$, depending on $n$, $\lambda$, $\Lambda$ and $\alpha_0$, such that*

$$\|u\|_{C^\gamma(Q_{1/2})} \le C_7\big(\|u\|_{L^\infty(\mathbb{R}^d\times[-1,0])} + C\big).$$

We can also include a drift term in the equation when $\alpha \ge 1$. This is stated in the next result.

**Theorem 7.2** (Hölder estimates with drift). *Assume $\alpha \ge 1$. Let $u$ be a bounded function in $\mathbb{R}^d \times [-1,0]$ such that*

$$\begin{aligned} u_t - C_0|\nabla u| - M^+u &\le C \quad \text{in } Q_1,\\ u_t + C_0|\nabla u| - M^-u &\ge -C \quad \text{in } Q_1. \end{aligned}$$

*Then there are constants $C_7$ and $\gamma$, depending on $n$, $\lambda$, $\Lambda$, $C_0$, such that*

$$\|u\|_{C^\gamma(Q_{1/2})} \le C_7\big(\|u\|_{L^\infty(\mathbb{R}^d\times[-1,0])} + C\big).$$

The proofs of these two theorems are essentially the same. The only difference is that when $\alpha \ge 1$, we can include a nonzero drift term in Theorem 6.1. Because of this, we write the proof only once, for Theorem 7.2, which applies to both theorems.

*Proof of Theorem 7.2.* We start by observing that we can reduce to the case $C \le \varepsilon_0$ and $\|u\|_{L^\infty} \le \frac{1}{2}$ by considering the function

$$\frac{1}{C/\varepsilon_0 + 2\|u\|_{L^\infty}} u(x,t).$$

We choose $\varepsilon_0$ sufficiently small, which will be specified below.

Our objective is to prove that for some $\gamma > 0$, which will also be specified below,

$$\operatorname*{osc}_{Q_r} u \le 2r^\gamma \tag{7-1}$$

for all $r \in (0,1)$. This proves the desired modulus of continuity at the point $(0,0)$. Since there is nothing special about the origin, we obtain the result of the theorem at every point in $Q_{1/2}$ using a standard scaling and translation argument. Note that since $\|u\|_{L^\infty} \le \frac{1}{2}$, we know a priori that (7-1) holds for all $r < 2^{-1/\gamma}$. We can make this threshold arbitrarily small by choosing a small value of $\gamma$.

In order to prove that (7-1) holds for all values of $r \in (0,1)$, we use induction. We assume that it holds for all $r \ge 8^{-k}$ and we show that it then holds for all $r \ge 8^{-(k+1)}$. Because of the observation in the previous paragraph, we can guarantee this inequality for the first few values of $k$ by choosing a small value of $\gamma$. Thus, we are left to prove the inductive step.

Let

$$\tilde u(x,t) = \frac{1}{2}\,\frac{1}{8^{\gamma(k-1)}} u\left(\frac{8^{-(k-1)}}{2}x, \frac{8^{-\alpha(k-1)}}{2^\alpha}t\right).$$

This function $\tilde u$ is a scaled version of $u$ so that the values of $\tilde u$ in $Q_2$ correspond to the values of $u$ in $Q_{8^{-k+1}}$. Moreover, since (7-1) holds for $r \ge 8^{-k}$, we have that

$$\operatorname*{osc}_{Q_{2r}} \tilde u \le \min(r^\gamma, 1) \quad \text{for all } r \ge \tfrac{1}{8}. \tag{7-2}$$

Since $\operatorname{osc}_{Q_2} \tilde u \le 1$, for all $(x,t) \in Q_2$, we have that $\tilde u(x,t) \ge \max_{Q_2} \tilde u - \frac{1}{2}$ or $\tilde u(x,t) \le \min_{Q_2} \tilde u + \frac{1}{2}$. There may be points where both inequalities hold. The important thing is that at least one of the two inequalities holds at every point $(x,t) \in Q_2$. Therefore, one of the two inequalities will hold in at least half of the points (in measure) of the cylinder $B_{1/4}\times[-1,-2^{-\alpha}]$. Without loss of generality, let us assume it is the first of these inequalities that holds for most points (a similar argument works otherwise). That is, we have

$$\left|\left\{\tilde u \ge \max_{Q_2}\tilde u - \tfrac12\right\}\cap\left(B_{1/4}\times[-1,-2^{-\alpha}]\right)\right| \ge \tfrac12|B_{1/4}|\times(1-2^{-\alpha}).$$

Let $v$ be the truncated function

$$v(x,t) := \left(\tilde u(x,t)-\max_{Q_2}\tilde u+1\right)^+.$$

Note that $v\ge 0$ everywhere and $v=\tilde u(x,t)-\max_{Q_2}\tilde u+1$ in $Q_2$. If $x\notin B_2$ and $t\in[-1,0]$, it can happen that $v(x,t)>\tilde u(x,t)-\max_{Q_2}\tilde u+1$. We can estimate their difference using (7-2):

$$v(x,t)-\left(\tilde u(x,t)-\max_{Q_2}\tilde u+1\right)\le \operatorname*{osc}_{B_{|x|}\times[-1,0]}\tilde u-1\le\left(\frac{|x|}{2}\right)^{\gamma}-1 \quad\text{for any } x\notin B_2,\ t\in[-1,0]. \tag{7-3}$$

Note that for any fixed $R$, the right-hand side converges to zero uniformly for $2\le|x|\le R$ as $\gamma\to 0$.

Inside $Q_1$, the function $v$ satisfies the equation

$$\begin{aligned} v_t + C_0|\nabla v| - M^- v &\ge \tilde u_t + C_0|\nabla\tilde u| - M^-\tilde u + M^-(\tilde u - v)\\ &\ge -\varepsilon_0 + M^-(\tilde u - v)\\ &= -\varepsilon_0 + M^-\left((\tilde u-\max\tilde u+1)-v\right)\\ &\ge -\varepsilon_0 - c(\gamma). \end{aligned}$$

Here $c(\gamma)=-\min_{Q_1} M^-\left((\tilde u-\max\tilde u+1)-v\right)=\max_{Q_1} M^+\left(v-(\tilde u-\max\tilde u+1)\right)$. We can estimate $c(\gamma)$ using (7-3) and assumption (A2), because

$$\begin{aligned} L\left(v-(\tilde u-\max\tilde u+1)\right)(x) &= \int_{\mathbb{R}^d}\delta_h\left(v-(\tilde u-\max\tilde u+1)\right)(x)K(h)\,\mathrm{d}h\\ &= \int_{|h|\ge 2}\left(v-(\tilde u-\max\tilde u+1)\right)(h)K(h)\,\mathrm{d}h\\ &\le C\int_{2\le|h|\le R}(|h|^{\gamma}-1)K(h)\,\mathrm{d}h + \int_{|h|\ge R}2\|\tilde u\|_{L^\infty}K(h)\,\mathrm{d}h, \end{aligned} \tag{7-4}$$

where we note the use of the fact that $v-(\tilde u-\max\tilde u+1)\equiv 0$ and also $\nabla\left(v-(\tilde u-\max\tilde u+1)\right)\equiv 0$ in $Q_2$. Thus given any $\rho$, we can make $c(\gamma)<\rho$ by first choosing $R$ large enough so that the tails of $K$ are negligible outside of $B_R$ — hence controlling the second term of (7-4) — and then choosing $\gamma$ small enough so that second term of (7-4) is small enough. Since none of these choices depend upon the kernel, $K$, they hold for $M^+$, and hence $c(\gamma)$, as well.

Applying Theorem 6.1,

$$\begin{aligned} \min_{Q_{1/4}} v + \varepsilon_0 + c(\gamma) &\ge \frac{1}{C_6}\left(\int_{B_{1/4}\times[-1,-2^{-\alpha}]} v^{\varepsilon}\mathrm{d}x\,\mathrm{d}t\right)^{1/\varepsilon}\\ &\ge \frac{1}{C_6}\left(\frac12|B_{1/4}|(1-2^{-\alpha})\right)^{1/\varepsilon}\frac12. \end{aligned}$$

Let us choose $\varepsilon_0>0$ and $\gamma>0$ sufficiently small so that

$$\delta := \frac{1}{C_6}\left(\frac12|B_{1/4}|(1-2^{-\alpha})\right)^{1/\varepsilon}\frac12 - \varepsilon_0 - c(\gamma) > 0.$$

Therefore, we obtained $\min_{Q_{1/4}} v \geq \delta$, which implies that $\operatorname{osc}_{Q_{1/4}} \tilde{u} \leq 1-\delta$. In terms of the original variables, this means that

$$\operatorname*{osc}_{Q_{8^{-k}}} u \leq 2 \times 8^{-\gamma(k-1)}(1-\delta).$$

Consequently, for any $r \in (8^{-k-1}, 8^{-k})$,

$$\operatorname*{osc}_{Q_r} u \leq 2 \times 8^{-\gamma(k-1)}(1-\delta).$$

Choosing $\gamma$ sufficiently small so that

$$8^{-2\gamma} \geq (1-\delta)$$

implies that (7-1) holds for all $r > 2^{-k-1}$. This finishes the inductive step, and hence the proof.

Note that there is no circular dependence between the constants $\gamma$ and $\varepsilon_0$. All conditions required in the proof are satisfied for any smaller value. We choose $\varepsilon_0$ and $\gamma$ sufficiently small so that all these conditions are met. □

## 8. $C^{1,\gamma}$ regularity for nonlinear equations

It is by now standard that a Hölder regularity result as in Theorem 1.1 for kernels $K$ that have rough dependence in $x$ and $t$ implies a $C^{1,\alpha}$ estimate for solutions to nonlinear equations. The following is a more precise statement.

**Theorem 8.1.** *Assume $\alpha_0 > 1$, $\alpha \in [\alpha_0, 2]$ and $I$ is a translation-invariant nonlocal operator that is uniformly elliptic with respect to the class of kernels that satisfy* (A1), (A2), (A3) *and* (A4). *Let $u : \mathbb{R}^n \times [-T, 0] \to \mathbb{R}$ be a bounded viscosity solution of the equation*

$$u_t - Iu = f \quad \text{in } B_1 \times [-T, 0].$$

*Then $u(\cdot, t) \in C^{1+\gamma}(B_{1/2})$ for all $t \in [-T/2, 0]$ and $u(x, \cdot) \in C^{(1+\gamma)/\alpha}([-T/2, 0])$ for all $x \in B_{1/2}$. Moreover, the following regularity estimate holds*:

$$\sup_{t\in[-T/2,0]} \|u(\cdot,t)\|_{C^{1+\gamma}(B_{1/2})} + \sup_{x\in B_{1/2}} \|u(x,\cdot)\|_{C^{(1+\gamma)/\alpha}([-T/2,0])} \leq C\big(\|u\|_{L^\infty(\mathbb{R}^n\times[-T,0])} + \|f\|_{L^\infty(B_1\times[-T,0])} + I\big).$$

*The constants $C$ and $\gamma$ depend only on $\lambda$, $\Lambda$, $\mu$, $n$ and $\alpha_0$. Here $\gamma > 0$ is the minimum between $\alpha_0 - 1$ and the constant $\gamma$ from Theorem 1.1* (*or Theorem 7.2*).

The proof of Theorem 8.1 is given in [Serra 2015] for the smaller class of symmetric kernels satisfying (1-2). His proof uses the main result in [Chang Lara and Dávila 2014], and the proof of Theorem 8.1 follows simply by replacing it with Theorem 7.2 in this paper. There is only one comment that needs to be made. In [Serra 2015], the following quantity is used a few times to control the tail of an integral operator

$$\|u\|_{L^1(\mathbb{R}^n,\omega_0)} := \int_{\mathbb{R}^n} u(x)(1+|x|)^{-n-\alpha_0}\, dx.$$

Because of our assumption (2-1), this quantity is not sufficient and needs to be replaced by

$$\max\big\{x \in \mathbb{R}^n \,:\, (1+|x|)^{\varepsilon-\alpha_0} u(x)\big\}$$

for some arbitrary small $\varepsilon > 0$. After this small modification, the proof in [Serra 2015] straightforwardly applies to prove Theorem 8.1 using Theorem 7.2.

The main example of a nonlinear integral operator $I$ is the Isaacs operator from stochastic games:

$$I u(x) = \inf_i \sup_j \int_{\mathbb{R}^n} \delta_h u(x,t) K^{ij}(h)\, \mathrm{d}h.$$

Here, the kernels $K^{ij}$ must satisfy the hypotheses (A1), (A2), (A3) and (A4) uniformly in $i$ and $j$.

The result can also be extended to kernels $K^{ij}(x,h,t)$ that are not translation-invariant provided that they are continuous with respect to $x$ and $t$. See [Serra 2015] for a discussion on this extension.

## Appendix: The crawling ink spots theorem

We prove a version of the *crawling ink spots theorem* for fractional parabolic equations, which is a covering argument that first appeared in the original work of Krylov and Safonov [1979]. There it is indicated that the result was previously known by Landis, and it was Landis himself who came up with its suggestive name.

Let $d_\alpha$ be the parabolic distance of order $\alpha$. By definition, it is

$$d_\alpha\big((x_0,t_0),(x_1,t_1)\big) = \max\big((2|t_1 - t_2|)^{1/\alpha}, |x_1 - x_2|\big).$$

The parabolic cylinders $Q_r(x,t)$ are balls of radius $r$ centered at $\big(x, t - \frac{1}{2} r^\alpha\big)$ with respect to the distance $d_\alpha$. The importance of this characterization is that it allows us to use the Vitali covering lemma, since this result is valid in arbitrary metric spaces.

**Lemma A.1.** *Let $\mu > 0$ and $E \subset F \subset B_1 \times \mathbb{R}$ be two open sets that satisfy the following two assumptions*:

- *For every point $(x,t) \in F$, there exists a cylinder $Q \subset B_1 \times \mathbb{R}$ so that $(x,t) \in Q$ and $|E \cap Q| \le (1-\mu)|Q|$.*
- *For every cylinder $Q \subset B_1 \times \mathbb{R}$ such that $|E \cap Q| > (1-\mu)|Q|$, we have $Q \subset F$.*

*Then $|E| \le (1 - c\mu)|F|$, where $c$ is a constant depending on dimension only.*

*Proof.* For every point $(x,t) \in F$, let $Q^0$ be the cylinder such that $(x,t) \in Q^0$ and $|E \cap Q^0| < (1-\mu)|Q^0|$.

Recall that $F$ is an open set. Let us choose a maximal cylinder $Q^{(x,t)}$ such that $(x,t) \in Q^{(x,t)}$, $Q^{(x,t)} \subset Q^0$ and $Q^{(x,t)} \subset F$. Two things may happen; either $Q^{(x,t)} = Q^0$, in which case $|Q^{(x,t)} \cap E| < (1-\mu)|Q^{(x,t)}|$, or for any larger cylinder $Q^{(x,t)} \subset Q \subset Q^0$, we would have $Q \not\subset F$. In the latter case, we would have $|E \cap Q| \le (1-\mu)|Q|$ for *any* cylinder $Q$ so that $Q^{(x,t)} \subset Q \subset Q^0$. In particular, the inequality holds for a decreasing sequence converging to $Q^{(x,t)}$ and therefore $|E \cap Q^{(x,t)}| \le (1-\mu)|Q^{(x,t)}|$.

In any case, we have constructed a cover $Q^{(x,t)}$ of the set $F$ so that for all $(x,t) \in F$,

- $(x,t) \in Q^{(x,t)}$,
- $Q^{(x,t)} \subset F$,
- $|Q^{(x,t)} \cap E| \le (1-\mu)|Q^{(x,t)}|$.

Using the Vitali covering lemma, we can select a countable subcollection of cylinders $Q_j$ such that $F \subset \bigcup_{j=1}^\infty 5Q_j$. Here each $Q_j$ is one of the cylinders $Q^{(x,t)}$. We write $5Q_j$ to denote the cylinder expanded as a ball with respect to the metric $d_\alpha$ with the same center and five times the radius.

Since $Q_j \subset F$ and $|E \cap Q_j| \leq (1-\mu)|Q_j|$, we have $|Q_j \cap (F \setminus E)| \geq \mu|Q_j|$. Therefore,

$$\begin{aligned}|F \setminus E| &\geq \sum_{j=1}^{\infty} |Q_j \cap (F \setminus E)| \\ &\geq \sum_{j=1}^{\infty} \mu|Q_j| \\ &= 5^{-d-\alpha}\mu \sum_{j=1}^{\infty} |5Q_j| \geq 5^{-d-\alpha}\mu|F|.\end{aligned}$$

The lemma follows with $c = 5^{-d-\alpha}$. □

Lemma A.1 is not applicable directly to parabolic equations. What we need is a covering lemma so that if $|E \cap Q| \geq (1-\mu)|Q|$, then a time-shift of the cylinder $Q$ is included in $F$ instead of $Q$ itself. This time-shift is given by the cylinders $\bar{Q}^m$, which we defined in Section 6.

We now give the proof of the *crawling ink spots* theorem.

*Proof of Theorem 6.3.* Let $\mathcal{Q}$ be the collection of cylinders $Q \subset B_1 \times \mathbb{R}$ such that $|E \cap Q| > (1-\mu)|Q|$. Let $G = \bigcup_{Q\in\mathcal{Q}} Q$. By construction, $E$ and $G$ satisfy the assumptions of Lemma A.1; thus $|E| \leq (1-c\mu)|G|$. In order to prove this theorem, we are left to show that $|G| \leq (m+1)/m|F|$. For that, we will see that

$$\Bigl|\bigcup_{Q\in\mathcal{Q}} \bar{Q}^m\Bigr| \geq \frac{m}{m+1}\Bigl|\bigcup_{Q\in\mathcal{Q}} Q \cup \bar{Q}^m\Bigr| \geq \frac{m}{m+1}|G|.$$

The second inequality above is trivial by the inclusion of the sets. The first inequality is not obvious since the cylinders may overlap. We justify this first inequality below.

From Fubini's theorem, the measure of any set $A \in B_1 \times \mathbb{R}$ is given by

$$|A| = \int_{B_1} \mathcal{L}_1(A \cap (\{x\} \times \mathbb{R}))\,\mathrm{d}x,$$

where $\mathcal{L}_1$ stands for the one-dimensional Lebesgue measure.

We finish the proof applying Fubini's theorem and noticing that for all $x \in B_1$,

$$\mathcal{L}_1\Bigl(\bigcup_{Q\in\mathcal{Q}} \bar{Q}^m \cap (\{x\} \times \mathbb{R})\Bigr) \geq \frac{m}{m+1}\mathcal{L}_1\Bigl(\bigcup_{Q\in\mathcal{Q}} (Q \cup \bar{Q}^m) \cap (\{x\} \times \mathbb{R})\Bigr).$$

This inequality follows from Lemma A.2, which is described below. □

The following lemma is copied directly from [Imbert and Silvestre 2013b, Lemma 2.4.25]. An elementary proof is given there, which is independent of the rest of the text.

**Lemma A.2.** *Consider two (possibly infinite) sequences of real numbers $(a_k)_{k=1}^N$ and $(h_k)_{k=1}^N$ for $N \in \mathbb{N} \cup \{\infty\}$ with $h_k > 0$ for $k = 1, \dots, N$. Then*

$$\Bigl|\bigcup_{k=1}^{N} (a_k, a_k + (m+1)h_k)\Bigr| \leq \frac{m}{m+1}\Bigl|\bigcup_{k=1}^{N} (a_k + h_k, a_k + (m+1)h_k)\Bigr|.$$

## References


[Alexandre et al. 2000] R. Alexandre, L. Desvillettes, C. Villani, and B. Wennberg, "Entropy dissipation and long-range interactions", *Arch. Ration. Mech. Anal.* **152**:4 (2000), 327–355. MR 1765272 Zbl 0968.76076

[Barles and Imbert 2008] G. Barles and C. Imbert, "Second-order elliptic integro-differential equations: viscosity solutions' theory revisited", *Ann. Inst. H. Poincaré Anal. Non Linéaire* **25**:3 (2008), 567–585. MR 2422079 Zbl 1155.45004

[Bass 2004] R. F. Bass, "PDE from a probability point of view", Lecture notes, University of Connecticut, Storrs, CT, 2004, available at http://homepages.uconn.edu/∼rib02005/pdeprob.pdf.

[Bass and Kassmann 2005] R. F. Bass and M. Kassmann, "Hölder continuity of harmonic functions with respect to operators of variable order", *Comm. Partial Differential Equations* **30**:7-9 (2005), 1249–1259. MR 2180302 Zbl 1087.45004

[Bass and Levin 2002a] R. F. Bass and D. A. Levin, "Harnack inequalities for jump processes", *Potential Anal.* **17**:4 (2002), 375–388. MR 1918242 Zbl 0997.60089

[Bass and Levin 2002b] R. F. Bass and D. A. Levin, "Transition probabilities for symmetric jump processes", *Trans. Amer. Math. Soc.* **354**:7 (2002), 2933–2953. MR 1895210 Zbl 0993.60070

[Bjorland et al. 2012] C. Bjorland, L. Caffarelli, and A. Figalli, "Non-local gradient dependent operators", *Adv. Math.* **230**:4-6 (2012), 1859–1894. MR 2927356 Zbl 1252.35099

[Bogdan and Sztonyk 2005] K. Bogdan and P. Sztonyk, "Harnack's inequality for stable Lévy processes", *Potential Anal.* **22**:2 (2005), 133–150. MR 2137058 Zbl 1081.60055

[Cabré 1997] X. Cabré, "Nondivergent elliptic equations on manifolds with nonnegative curvature", *Comm. Pure Appl. Math.* **50**:7 (1997), 623–665. MR 1447056 Zbl 0878.58054

[Caffarelli and Silvestre 2009] L. Caffarelli and L. Silvestre, "Regularity theory for fully nonlinear integro-differential equations", *Comm. Pure Appl. Math.* **62**:5 (2009), 597–638. MR 2494809 Zbl 1170.45006

[Caffarelli and Silvestre 2011a] L. Caffarelli and L. Silvestre, "The Evans–Krylov theorem for nonlocal fully nonlinear equations", *Ann. of Math.* (2) **174**:2 (2011), 1163–1187. MR 2831115 Zbl 1232.49043

[Caffarelli and Silvestre 2011b] L. Caffarelli and L. Silvestre, "Regularity results for nonlocal equations by approximation", *Arch. Ration. Mech. Anal.* **200**:1 (2011), 59–88. MR 2781586 Zbl 1231.35284

[Chang Lara 2012] H. A. Chang Lara, "Regularity for fully non linear equations with non local drift", preprint, 2012. arXiv 1210.4242

[Chang Lara and Dávila 2012] H. A. Chang Lara and G. Dávila, "Regularity for solutions of nonlocal, nonsymmetric equations", *Ann. Inst. H. Poincaré Anal. Non Linéaire* **29**:6 (2012), 833–859. MR 2995098 Zbl 1317.35278

[Chang Lara and Dávila 2014] H. A. Chang Lara and G. Dávila, "Regularity for solutions of non local parabolic equations", *Calc. Var. Partial Differential Equations* **49**:1-2 (2014), 139–172. MR 3148110 Zbl 1292.35068

[Chen and Kumagai 2003] Z.-Q. Chen and T. Kumagai, "Heat kernel estimates for stable-like processes on $d$-sets", *Stochastic Process. Appl.* **108**:1 (2003), 27–62. MR 2008600 Zbl 1075.60556

[Courrège 1965] P. Courrège, "Sur la forme intégro-différentielle des opérateurs de $\mathbb{C}_k^\infty$ dans $\mathbb{C}$ satisfaisant au principe du maximum", *Séminaire Brelot-Choquet-Deny. Théorie du Potentiel* **10**:1 (1965), Exposé 2. Zbl 0155.17402

[Guillen and Schwab 2012] N. Guillen and R. W. Schwab, "Aleksandrov–Bakelman–Pucci type estimates for integro-differential equations", *Arch. Ration. Mech. Anal.* **206**:1 (2012), 111–157. MR 2968592 Zbl 1256.35190

[Guillen and Schwab 2014] N. Guillen and R. W. Schwab, "Neumann homogenization via integro-differential operators", preprint, 2014. To appear in *Discrete Contin. Dyn. Syst.* arXiv 1403.1980

[Imbert and Silvestre 2013a] C. Imbert and L. Silvestre, "Estimates on elliptic equations that hold only where the gradient is large", preprint, 2013. To appear in *J. Euro. Math. Soc.* (*JEMS*). arXiv 1306.2429

[Imbert and Silvestre 2013b] C. Imbert and L. Silvestre, "An introduction to fully nonlinear parabolic equations", pp. 7–88 in *An introduction to the Kähler–Ricci flow*, edited by S. Boucksom et al., Lecture Notes in Mathematics **2086**, Springer, Cham, 2013. MR 3185332 Zbl 1282.35004

[Kassmann and Mimica 2013a] M. Kassmann and A. Mimica, "Analysis of jump processes with nondegenerate jumping kernels", *Stochastic Process. Appl.* **123**:2 (2013), 629–650. MR 3003366 Zbl 1259.60100

[Kassmann and Mimica 2013b] M. Kassmann and A. Mimica, "Intrinsic scaling properties for nonlocal operators", preprint, 2013. To appear in *J. Eur. Math. Soc.* (*JEMS*). arXiv 1310.5371

[Kassmann and Schwab 2014] M. Kassmann and R. W. Schwab, "Regularity results for nonlocal parabolic equations", *Riv. Math. Univ. Parma* (*N.S.*) **5**:1 (2014), 183–212. MR 3289601 Zbl 06450555

[Kassmann et al. 2014] M. Kassmann, M. Rang, and R. W. Schwab, "Integro-differential equations with nonlinear directional dependence", *Indiana Univ. Math. J.* **63**:5 (2014), 1467–1498. MR 3283558 Zbl 1311.35047

[Krylov and Safonov 1979] N. V. Krylov and M. V. Safonov, "An estimate for the probability of a diffusion process hitting a set of positive measure", *Dokl. Akad. Nauk SSSR* **245**:1 (1979), 18–20. In Russian; translated in *Sov. Math. Dokl.* **20** (1979), 253–256. MR 525227 Zbl 0459.60067

[Krylov and Safonov 1980] N. V. Krylov and M. V. Safonov, "Некоторое свойство решений параболических уравнений с измеримыми коэффициентами", *Izv. Akad. Nauk SSSR Ser. Mat.* **44**:1 (1980), 161–175. Translated as "A certain property of solutions of parabolic equations with measurable coefficients" in *Math. USSR Izv.* **16**:1 (1981), 151–164. MR 563790 Zbl 0439.35023

[Landis 1971] E. M. Landis, *Уравнения второго порялка эллиптического и параболического типов*, Nauka, Moscow, 1971. Translated as *Second order equations of elliptic and parabolic type*, Translations of Mathematical Monographs **171**, American Mathematical Society, Providence, RI, 1998. MR 1487894 Zbl 0226.35001

[Landkof 1966] N. S. Landkof, *Основы современной теории потенциала*, Nauka Fizmatgiz, Moscow, 1966. Translated as *Foundations of modern potential theory*, Grundlehren der Mathematischen Wissenschaften **180**, Springer, New York, 1972. MR 214795 Zbl 0148.10301

[Mou and Święch 2015] C. Mou and A. Święch, "Uniqueness of viscosity solutions for a class of integro-differential equations", *Nonlinear Differential Equations Appl.* **22**:6 (2015), 1851–1882. MR 3415025 Zbl 06506047

[Savin 2007] O. Savin, "Small perturbation solutions for elliptic equations", *Comm. Partial Differential Equations* **32**:4-6 (2007), 557–578. MR 2334822 Zbl 1221.35154

[Serra 2015] J. Serra, "Regularity for fully nonlinear nonlocal parabolic equations with rough kernels", *Calc. Var. Partial Differential Equations* **54**:1 (2015), 615–629. MR 3385173 Zbl 1327.35170

[Silvestre 2006] L. Silvestre, "Hölder estimates for solutions of integro-differential equations like the fractional Laplace", *Indiana Univ. Math. J.* **55**:3 (2006), 1155–1174. MR 2244602 Zbl 1101.45004

[Silvestre 2011] L. Silvestre, "On the differentiability of the solution to the Hamilton–Jacobi equation with critical fractional diffusion", *Adv. Math.* **226**:2 (2011), 2020–2039. MR 2737806 Zbl 1216.35165

[Silvestre 2012a] L. Silvestre, "Hölder estimates for advection fractional-diffusion equations", *Ann. Sc. Norm. Super. Pisa Cl. Sci.* (5) **11**:4 (2012), 843–855. MR 3060702 Zbl 1263.35056

[Silvestre 2012b] L. Silvestre, "Lecture notes on nonlocal equations", preprint, University of Texas, Austin, TX, 2012, available at http://www.ma.utexas.edu/mediawiki/index.php/Lecture_notes_on_nonlocal_equations.

[Silvestre 2014a] L. Silvestre, "A new regularization mechanism for the Boltzmann equation without cut-off", preprint, 2014. arXiv 1412.4706

[Silvestre 2014b] L. Silvestre, "Regularity estimates for parabolic integro-differential equations and applications", pp. 873–894 in *Proceedings of the International Congress of Mathematicians* (Seoul, 2014), vol. III, edited by S. Y. Jang et al., Kyung Moon Sa, Seoul, 2014.

[Song and Vondraček 2004] R. Song and Z. Vondraček, "Harnack inequality for some classes of Markov processes", *Math. Z.* **246**:1-2 (2004), 177–202. MR 2031452 Zbl 1052.60064

Received 3 Oct 2015. Accepted 16 Dec 2015.

RUSSELL W. SCHWAB: rschwab@math.msu.edu
*Department of Mathematics, Michigan State University, East Lansing, MI 48824, United States*

LUIS SILVESTRE: luis@math.uchicago.edu
*Department of Mathematics, The University of Chicago, Chicago, IL 60637, United States*

# Analysis & PDE
msp.org/apde




See inside back cover or msp.org/apde for submission instructions.


The subscription price for 2016 is US $/year for the electronic version, and $/year (+$, if shipping outside the US) for print and electronic. Subscriptions, requests for back issues from the last three years and changes of subscribers address should be sent to MSP.

Analysis & PDE (ISSN 1948-206X electronic, 2157-5045 printed) at Mathematical Sciences Publishers, 798 Evans Hall #3840, c/o University of California, Berkeley, CA 94720-3840, is published continuously online. Periodical rate postage paid at Berkeley, CA 94704, and additional mailing offices.

APDE peer review and production are managed by EditFlow® from MSP.

PUBLISHED BY

mathematical sciences publishers

nonprofit scientific publishing

http://msp.org/

# ANALYSIS & PDE

Volume 9 No. 3 2016



2157-5045(2016)9:3;1-9