

## Gradient regularity for nonlinear parabolic equations

TUOMO KUUSI AND GIUSEPPE MINGIONE

*To Emmanuele DiBenedetto on his 65th birthday*

**Abstract.** We consider non-homogeneous degenerate and singular parabolic equations of the  $p$ -Laplacian type and prove pointwise bounds for the spatial gradient of solutions in terms of intrinsic parabolic potentials of the given datum. In particular, the main estimate found reproduces in a sharp way the behavior of the Barenblatt (fundamental) solution when applied to the basic model case of the evolutionary  $p$ -Laplacian equation with Dirac datum. Using these results as a starting point, we then give sufficient conditions to ensure that the gradient is continuous in terms of potentials; in turn these imply borderline cases of known parabolic results and the validity of well-known elliptic results whose extension to the parabolic case remained an open issue. As an intermediate result we prove the Hölder continuity of the gradient of solutions to possibly degenerate, homogeneous and quasilinear parabolic equations defined by general operators.

**Mathematics Subject Classification (2010):** 35K55 (primary); 31C45 (secondary).

### Contents

1	Introduction and results . . . . .	756
2	Basic notation and definitions . . . . .	768
3	The $C^{0,\alpha}$ -gradient theory . . . . .	772
4	Proof of the intrinsic potential estimate and consequences . . . . .	795
5	Continuity of the gradient via potentials . . . . .	809
	References . . . . .	820

The authors are supported by the ERC grant 207573 “Vectorial Problems” and by the Academy of Finland project “Potential estimates and applications for nonlinear parabolic partial differential equations”. Part of the paper was written while the two authors were attending the conference “21th Meeting on Calculus of Variations and Geometric Measure Theory”, at Levico Terme in February 2011, where some of the results were also presented.

Received March 21, 2011; accepted December 12, 2011.

**1. Introduction and results**

In this paper we deal with non-homogeneous, measure data and possibly degenerate/singular parabolic equations of the type

$$u_t - \operatorname{div} a(Du) = \mu, \tag{1.1}$$

considered in cylindrical domains  $\Omega_T = \Omega \times (-T, 0)$ , where  $\Omega \subset \mathbb{R}^n$  is a bounded domain and  $T > 0$ . We are here ultimately aiming at providing boundedness and continuity estimates of the spatial gradient of solutions  $Du$  in terms of suitable linear and nonlinear potentials of the right-hand side measure  $\mu$ . In turn, modulo standard approximation procedures, a priori estimates found allow  $\mu$  to be, in the most general case, a Borel measure with finite total mass:  $|\mu|(\Omega_T) < \infty$ . The  $C^1$ -vector field  $a: \mathbb{R}^n \rightarrow \mathbb{R}^n$  is assumed to satisfy the following assumptions

$$\begin{cases} |a(z)| + |\partial a(z)|(|z|^2 + s^2)^{1/2} \leq L(|z|^2 + s^2)^{(p-1)/2} \\ \nu(|z|^2 + s^2)^{(p-2)/2} |\xi|^2 \leq \langle \partial a(z)\xi, \xi \rangle \end{cases} \tag{1.2}$$

whenever  $z, \xi \in \mathbb{R}^n$ . The numbers  $s, \nu, L$  are assumed to satisfy  $0 < \nu \leq L$  and  $s \geq 0$ . In this paper we shall assume

$$2 - \frac{1}{n+1} < p. \tag{1.3}$$

The assumptions (1.2) are considered in order to generalize the main model example we have in mind, that is the evolutionary  $p$ -Laplacian equation

$$u_t - \operatorname{div} (|Du|^{p-2} Du) = \mu, \tag{1.4}$$

which clearly satisfies the assumptions considered in (1.2) with  $s \equiv 0$ . In this respect, let us observe that the number  $s$  serves to distinguish the degenerate case ( $s = 0$ ) from the non-degenerate one ( $s > 0$ ).

The lower bound (1.3) looks natural in view of the available existence theory that guarantees the existence of solutions in a Sobolev space; for this we refer to [4–6]. On the other hand, the sharpness of the lower bound in (1.3) with respect to such a property can be seen for instance by looking at the Barenblatt solution; see Section 1.3 below. For the same reason, as solutions to (1.1) are usually obtained via approximation with solutions to equations with more regular data and solutions, in the following we shall always assume to deal with energy solutions, *i.e.* we say that  $u$  is a solution to (1.1) if

$$u \in C^0(-T, 0; L^2(\Omega)) \cap L^p(-T, 0; W^{1,p}(\Omega)) \tag{1.5}$$

and  $u$  solves (1.1) in the distributional sense

$$- \int_{\Omega_T} u \varphi_t \, dx \, dt + \int_{\Omega_T} \langle a(Du), D\varphi \rangle \, dx \, dt = \int_{\Omega_T} \varphi \, d\mu \tag{1.6}$$

whenever  $\varphi \in C_c^\infty(\Omega_T)$ . Finally, without loss of generality we shall assume that

$$\mu \in L^1(\mathbb{R}^{n+1}).$$

In other words in this paper we shall confine ourselves to provide uniform a priori estimates for a priori regular (*i.e.* satisfying (1.5)–(1.6)) solutions, that in turn imply similar estimates for general solutions to (1.1) by following the by now classical approximation procedures proposed in [5, 6]. For more on integrability of solutions to measure data problems (that satisfy (1.6) but usually fail to match (1.5)) we again refer to [6, 21–23]. In particular, we refer to [30, Section 1.4] for a comprehensive discussion on the approximation methods in the present context.

The results in this paper fall into two categories. The first one includes (spatial) gradient pointwise bounds via potentials for solutions to singular with *i.e.*  $p \leq 2$  — parabolic equations which represent the counterpart of those obtained in [30] for the degenerate case  $p \geq 2$ . Several relevant differences — both in the type of results obtained and in the techniques required — occur when  $p < 2$ , as it will be explained in the next section. The second realm of results, instead, includes both degenerate and singular parabolic equations and deals with a characterization of gradient continuity of solutions in terms of linear and nonlinear potentials decay properties.

A point of independent interest here, that actually serves as a preliminary tool for proving the potential estimates and continuity results, is completely new proof of the Hölder continuity character of the spatial gradient of solutions in the case of homogeneous equations

$$w_t - \operatorname{div} a(Dw) = 0. \quad (1.7)$$

We indeed remark that, while in the case of the  $p$ -Laplacian equation (and system)

$$w_t - \operatorname{div} (|Dw|^{p-2} Dw) = 0, \quad (1.8)$$

this has been done in the fundamental work of DiBenedetto and Friedman [11, 12], the general case (1.7) has remained essentially untouched as since the work of Lieberman [34], who was dealing with boundary value problems (note that the proofs presented in [11] are based on linearization techniques possible only in the case of special structures as in (1.8)). The estimates obtained in [34] are anyway not sufficient to work as technical background for the results we are going to develop here, and a new approach to gradient estimates is needed. For the case  $p \geq 2$  the necessary estimates were obtained in [30], and here we present a yet novel approach, covering the subquadratic case  $p < 2$ , and based on the use of Harnack inequalities.

In particular, in order to prove the potential estimates, we need a peculiar, rigid form of the a priori Hölder continuity estimates; see Theorem 3.1 below. This theorem is indeed a central tool in our approach as it allows to prove both the gradient Hölder continuity for solutions to homogeneous equations and the regularity results and potentials estimates in the case of non-homogeneous ones.

In rest of the paper, when talking about the gradient of a solution  $u$ , we shall always mean the spatial gradient  $Du = (u_{x_i})_{1 \leq i \leq n}$ . Some of the results of this paper have been announced in [28].

**1.1. Pointwise potential estimates**

The results in this section refer to the singular case

$$2 - \frac{1}{n + 1} < p \leq 2 \tag{1.9}$$

and include pointwise bounds for  $Du$  in terms of suitable parabolic Riesz potentials of the right-hand side measure  $\mu$ . To clarify the situation let us recall the corresponding elliptic background established in [16] (see also [15, 29, 32] for the case  $p \geq 2$ ), and valid for solutions to stationary equations as

$$-\operatorname{div} a(Du) = \mu \quad \text{in } \Omega \tag{1.10}$$

with  $B(x_0, 2r) \subset \Omega$  denoting the usual Euclidean ball centered at  $x_0$  and with radius  $2r > 0$ , we have that

$$|Du(x_0)| \leq c \int_{B(x_0, r)} (|Du| + s) dx + c \left[ \mathbf{I}_1^{|\mu|}(x_0, 2r) \right]^{1/(p-1)} \tag{1.11}$$

holds when  $2 - 1/n < p \leq 2$  (this last condition being the usual one ensuring that elliptic measure data equations have Sobolev solutions). The last quantity on the right-hand side of (1.11) denotes the classical truncated Riesz potential, *i.e.*

$$\mathbf{I}_\beta^\mu(x_0, r) := \int_0^r \frac{|\mu|(B(x_0, \varrho))}{\varrho^{n-\beta}} \frac{d\varrho}{\varrho}, \quad \beta > 0. \tag{1.12}$$

We refer the reader to [20, 25, 45, 46] for zero-order nonlinear potentials estimates.

It is now readily seen that in the case of parabolic equations an estimate as (1.11) simply cannot hold for very elementary reasons: multiplying solutions by a constant does not yield a solution to a similar equation. For this reason all the a priori estimates available fail to show homogeneity and scaling properties typical of elliptic equations as (1.10) (see for instance [1, 11, 30]), exhibiting instead a *scaling deficit* which measures such a homogeneity failure. The analysis of this mechanism eventually leads to the basic, fundamental technique used to treat equations as (1.1): the one based on intrinsic geometry and scaling, firstly introduced by DiBenedetto. We will not go too much into the details of this method, but we rather refer to [11, 30] for a wider description. We just remark that the core point of this method prescribes to study the equations not on all parabolic cylinders, but rather on certain special ones, whose ratio between space and time lengths *depend on the size of the solutions itself on the same cylinder*, according to the regularity considered; from

this fact comes the use of the word “intrinsic”. Specifically, one considers so-called intrinsic cylinders defined by

$$Q = Q_r^\lambda(x_0, t_0) \equiv B(x_0, r) \times (t_0 - \lambda^{2-p}r^2, t_0). \tag{1.13}$$

The number  $\lambda > 0$  in turn obeys  $|Du| \approx \lambda > 0$  in the sense that

$$\frac{1}{|Q_r^\lambda|} \int_{Q_r^\lambda} |Du| dx dt = \int_{Q_r^\lambda} |Du| dx dt \approx \lambda.$$

Note that, when  $\lambda \equiv 1$  or when  $p = 2$ , the cylinder in (1.13) reduces to the standard parabolic cylinder given by

$$Q_r(x_0, t_0) \equiv Q_r^1(x_0, t_0) \equiv B(x_0, r) \times (t_0 - r^2, t_0).$$

Exactly in the same way one is alternatively led to consider cylinders of the type

$$Q_{r_\lambda}^\lambda(x_0, t_0) \equiv B(x_0, \lambda^{(p-2)/2}r) \times (t_0 - r^2, t_0), \quad r_\lambda = \lambda^{(p-2)/2}r. \tag{1.14}$$

The main effect of considering equations on such cylinders is that they locally look as isotropic ones, and estimates homogenize. As a matter of fact, all the a priori regularity estimates available for solutions can be expressed in terms of local geometries, see for instance [11, 13, 14].

In this paper we shall see that *Riesz potentials interact with intrinsic geometries* to make gradient potential estimates hold for singular parabolic equations. Indeed, the natural — in a sense that will be clear in a few lines — parabolic generalization of estimate (1.11) can be achieved using a family of *intrinsic Riesz potentials*. More precisely, let us recall that the parabolic Riesz potential of  $\mu$  is defined by

$$I_\beta^\mu(x_0, t_0; r) := \int_0^r \frac{|\mu|(Q_\varrho(x_0, t_0))}{\varrho^{N-\beta}} \frac{d\varrho}{\varrho}, \quad 0 < \beta \leq N := n + 2 \tag{1.15}$$

where in turn  $N$  is the usual parabolic dimension. Then the following holds:

**Theorem 1.1 (Intrinsic linear potential bound).** *Let  $u$  be a solution to (1.1) with  $2 - 1/(n + 1) < p \leq 2$ . Let  $\beta > 0$ . The following holds for a.e.  $(x_0, t_0) \in \Omega_T$ : There exists a constant  $c \geq 1$ , depending only on  $n, p, v, L$ , but not on  $(x_0, t_0)$ , the solution  $u$ , or the vector field  $a(\cdot)$ , such that if  $\lambda > 0$  is a generalized root of*

$$\lambda = c\beta + c \int_0^{2r_\lambda} \frac{|\mu|(Q_\varrho^\lambda(x_0, t_0))}{\varrho^{N-1}} \frac{d\varrho}{\varrho}, \quad r_\lambda = \lambda^{(p-2)/2}r, \tag{1.16}$$

and if

$$\int_{Q_{r_\lambda}^\lambda} (|Du| + s) dx dt \leq \beta,$$

where  $Q_{2r_\lambda}^\lambda \equiv Q_{2r_\lambda}^\lambda(x_0, t_0) \equiv B(x_0, 2\lambda^{(p-2)/2}r) \times (t_0 - 4r^2, t_0) \subset \Omega_T$  is an intrinsic cylinder with vertex at  $(x_0, t_0)$ , then

$$|Du(x_0, t_0)| \leq \lambda.$$

By saying that  $\lambda$  is a generalized root of equation (1.16), where  $\beta > 0$  and  $c \geq 1$  are given constants, we mean a positive solution to (1.16) (the smallest can be taken), with the word “generalized” referring to the possibility that no root exists, in which case we simply set  $\lambda = \infty$ . The finiteness of the integral on the right-hand side of (1.16) anyway rules this case out, see Remark 4.1 below. Although the formulation of Theorem 1.1 might appear at first sight involved, it is actually the natural parabolic version of the elliptic estimate (1.11). Indeed, when dealing with degenerate/singular parabolic problems, all estimates find their optimal form only using intrinsic formulations. As a matter of fact, Theorem 1.1 precisely allows to recast the behavior of the so-called Barenblatt (fundamental) solution. We refer to Section 1.3 for the precise computation. Moreover, we remark that when  $\mu \equiv 0$ , Theorem 1.1 gives back a classical result of DiBenedetto and Friedman asserting that

$$c \int_{Q_{r_\lambda}^\lambda} (|Du| + s) \, dx \, dt \leq \lambda \Rightarrow |Du(x_0, t_0)| \leq \lambda$$

for a constant  $c \equiv c(n, p, \nu, L) \geq 1$ ; see Theorem 4.2. As a matter of fact Theorem 1.1 can read as follows: there exists a constant  $c \equiv c(n, p, \nu, L)$  such that

$$c \int_{Q_{r_\lambda}^\lambda} (|Du| + s) \, dx \, dt + c \int_0^{2r_\lambda} \frac{|\mu|(Q_\varrho^\lambda(x_0, t_0))}{\varrho^{N-1}} \frac{d\varrho}{\varrho} \Rightarrow |Du(x_0, t_0)| \leq \lambda.$$

Theorem 1.1 provides estimates on intrinsic cylinders, and therefore its applicability does not look immediate as it prescribes to go through the intrinsic relation (1.16); nevertheless Theorem 1.1 always implies a priori estimates in standard parabolic cylinders:

**Corollary 1.2 (Parabolic Riesz potential bound).** *Let  $u$  be a solution to (1.1) with  $2 - 1/(n + 1) < p \leq 2$ . The following holds for a.e.  $(x_0, t_0) \in \Omega_T$ : There exists a constant  $c$ , depending only on  $n, p, \nu, L$ , but not on  $(x_0, t_0)$ , the solution  $u$ , or the vector field  $a(\cdot)$ , such that*

$$|Du(x_0, t_0)| \leq c \left( \int_{Q_r} (|Du| + s + 1) \, dx \, dt \right)^{2/[2-n(2-p)]} + c [\mathbf{I}_1^\mu(x_0, t_0; 2r)]^{2/[2-n(2-p)]} \tag{1.17}$$

holds whenever  $Q_{2r} \equiv Q_{2r}(x_0, t_0) \equiv B(x_0, 2r) \times (t_0 - 4r^2, t_0) \subset \Omega_T$  is a standard parabolic cylinder with vertex at  $(x_0, t_0)$ .

Finally, when  $\mu$  is time-independent or admits a favorable decomposition — see (1.18) below — the elliptic Riesz potentials come back exactly as in (1.11):

**Corollary 1.3 (Parabolic/Elliptic Riesz potential bound).** *Let  $u$  be a solution to (1.1) with  $2 - 1/(n + 1) < p \leq 2$  and assume that the following decomposition holds:*

$$\mu = \mu_0 \otimes f, \tag{1.18}$$

where  $\mu_0$  is a finite mass Borel measure on  $\mathbb{R}^n$  and  $f \in L^\infty(-T, 0)$ . The following holds for a.e.  $(x_0, t_0) \in \Omega_T$ : There exists a constant  $c$ , depending only on  $n, p, v, L$ , but not on  $(x_0, t_0)$ , the solution  $u$ , or the vector field  $a(\cdot)$ , such that

$$|Du(x_0, t_0)| \leq c \left( \int_{Q_r} (|Du| + s + 1) dx dt \right)^{2/[2-n(2-p)]} + c \|f\|_{L^\infty}^{1/(p-1)} [\mathbf{I}_1^{\mu_0}(x_0, 2r)]^{1/(p-1)} \tag{1.19}$$

whenever  $Q_{2r}(x_0, t_0) \equiv B(x_0, 2r) \times (t_0 - 4r^2, t_0) \subset \Omega_T$  is a standard parabolic cylinder having  $(x_0, t_0)$  as vertex. The (elliptic) Riesz potential  $\mathbf{I}_1^{\mu_0}$  is defined in (1.12).

**Remark 1.4 (Structure of the exponents).** It is worthwhile to analyze the exponents appearing in (1.17), and in particular to make a comparison with the one appearing in (1.19), as they precisely reflect the structure properties of the equation, and in particular of the Barenblatt (fundamental) solutions. The number  $2/[2 - n(2 - p)]$  is the same one appearing in the typical gradient estimates for homogeneous equations ( $\mu = 0$ ) and reflects the gradient nature of the estimate in question. Indeed, when  $\mu \equiv 0$  estimate (1.17) reduces to the classical one obtained in [11, Chapter 8, Theorem 5.2']. The exponent  $2/[(n + 1)p - 2n]$  instead blows up as  $p \rightarrow 2n/(n + 1)$  and reflects the non-homogeneity of the equation studied, as well as the structure of the Barenblatt solution; see Section 1.3 below. Such exponent indeed intervenes in those estimates related to the Barenblatt solution, as for instance the Harnack inequalities in [13, 14, 26]. For the very same reason the exponent  $2/[(n + 1)p - 2n]$  relates to the fact that the right-hand side measure  $\mu$  in general depends on time, and it disappears when  $\mu$  is time-independent. This is completely natural as in this case it is possible to consider stationary solutions, for which (1.19) reduces to the elliptic estimate in (1.11). Yet, it is interesting to compare estimate (1.17) with the main result in [1], where a completely similar dependence on the exponents appear.

The techniques used to prove Theorem 1.1, although relying on the general scheme already introduced in [30] for the case  $p \geq 2$ , differ from those of [30] in several relevant aspects, which are essentially linked to the fact that we are here dealing with the singular case. Singular equations are indeed more delicate to treat when proving gradient bounds: as a matter of fact, as estimates as (1.16) and (1.17) essentially aim at establishing an upper bound for the gradient, they become more difficult to prove, because the equations as for instance (1.4) become more degenerate as  $|Du|$  increases when  $p < 2$ . In other words, the structure of the equation itself poses additional obstructions to the proof of gradient bounds; for the same reason the word *singular* appears somehow misleading when used in the present context. As mentioned earlier in this paper, in order to prove Theorem 1.1 we shall develop a novel approach to the Hölder continuity of the gradient of homogeneous equations as (1.7) for general vector field  $a: \mathbb{R}^n \rightarrow \mathbb{R}^n$  satisfying (1.2). Here, avoiding any linearization as used in [11, 12] we treat general singular equations

as (1.7). The approach in this paper prescribes to use as a central tool a homogeneous excess decay bound found in Theorems 3.1-3.5 below. In turn, these imply the Hölder continuity of the gradient  $Dw$  together with suitable a priori estimates in intrinsic cylinders, as eventually shown in Theorem 3.3. It is worth noticing here that our proof does not make of logarithmic type inequalities but rather relies on the classical weak Harnack inequality for linear parabolic equations.

**1.2. Continuity of the gradient via potentials**

In this section we shall consider parabolic equations under the more general bound in (1.3), therefore treating both degenerate and singular parabolic equations. For the sake of exposition we shall first give the continuity criteria in the singular case (1.9), and eventually will give the statements for the degenerate one  $p \geq 2$ , where, according to [30], nonlinear Wolff potentials come into the play. The proof of the results in the case  $p \geq 2$  can be indeed achieved combining the new arguments presented here with those developed in [30] in a way which appears then clear to the careful reader of both the papers.

To understand the role that potentials play in establishing the continuity of the gradient, let us observe that Corollary 1.2 establishes that  $Du$  is locally bounded provided so  $(x_0, t_0) \mapsto \mathbf{I}_1^\mu(x_0, t_0; \cdot)$  is. In particular, in this case we have that  $\mathbf{I}_1^\mu(x_0, t_0; r) \rightarrow 0$ , equiboundedly with respect to  $(x_0, t_0)$ . By strengthening this condition in uniform converge finally leads to the continuity of  $Du$ .

**Theorem 1.5 (Gradient continuity via linear potentials).** *Let  $u$  be a solution to (1.1) in  $\Omega_T$  and assume (1.2) with  $2 - 1/(n + 1) < p \leq 2$ . If  $\mathbf{I}_1^\mu(x, t; r) \rightarrow 0$  uniformly in  $(x, t) \in \Omega_T$  as  $r \rightarrow 0$ , that is*

$$\lim_{r \rightarrow 0} \sup_{(x,t) \in \Omega_T} \int_0^r \frac{|\mu|(Q_\varrho(x, t))}{\varrho^{N-1}} \frac{d\varrho}{\varrho} = 0 \tag{1.20}$$

*holds, then  $Du$  is continuous in  $\Omega_T$ .*

Weakening the convergence in (1.20) still leads to VMO gradient regularity.

**Theorem 1.6 (Gradient VMO-regularity).** *Let  $u$  be a solution to (1.1) in  $\Omega_T$  and assume (1.2) with  $2 - 1/(n + 1) < p \leq 2$ . If the function*

$$(x, t) \mapsto \mathbf{I}_1^\mu(x, t; r) = \int_0^r \frac{|\mu|(Q_\varrho(x, t))}{\varrho^{N-1}} \frac{d\varrho}{\varrho}$$

*is locally bounded in  $\Omega_T$  for some  $r > 0$ , and if furthermore*

$$\lim_{r \rightarrow 0} \frac{|\mu|(Q_r(x, t))}{r^{N-1}} = 0 \quad \text{uniformly in } (x, t) \in \Omega_T, \tag{1.21}$$

*then  $Du$  is locally VMO-regular in  $\Omega_T$ .*



We recall that the local VMO-regularity of  $Du$  means that for every subcylinder  $\tilde{Q} \Subset \Omega_T$ , we have that

$$\lim_{r \rightarrow 0} \sup_{Q_\varrho \subset \tilde{Q}, \varrho \leq r} \int_{Q_\varrho} |Du - (Du)_{Q_\varrho}| dx dt = 0. \quad (1.22)$$

The case  $p \geq 2$  instead involves the use of Wolff potentials, and connects to the results obtained in [30].

**Theorem 1.7 (Gradient continuity via nonlinear potentials).** *Let  $u$  be a solution to (1.1) in  $\Omega_T$  and assume (1.2) with  $p \geq 2$ . If*

$$\lim_{r \rightarrow 0} \sup_{(x,t) \in \Omega_T} \int_0^r \left( \frac{|\mu|(Q_\varrho(x,t))}{\varrho^{N-1}} \right)^{p/[2(p-1)]} \frac{d\varrho}{\varrho} = 0 \quad (1.23)$$

*holds, then  $Du$  is continuous in  $\Omega_T$ . Finally, if the function*

$$(x, t) \mapsto \int_0^r \left( \frac{|\mu|(Q_\varrho(x,t))}{\varrho^{N-1}} \right)^{p/[2(p-1)]} \frac{d\varrho}{\varrho}$$

*is locally bounded in  $\Omega_T$  for some  $r > 0$ , and if furthermore (1.21) holds uniformly with respect to  $(x, t)$  in  $\Omega_T$ , then  $Du$  is locally VMO-regular in  $\Omega_T$ .*

We remark that the previous statements extend to the parabolic case the gradient continuity criteria found in [17], the proof being here considerably more difficult and involved — actually new ideas are needed. For statements relating VMO-regularity and conditions as in (1.21) see also [39,40].

The previous three theorems admit a number of corollaries, allowing to set a few questions that, being classical in the elliptic case, remained still open in the parabolic one, at least in the case  $p \neq 2$ . A first corollary concerns a Lorentz Spaces criterion ensuring the continuity of the gradient. The one in the next lines is actually a borderline version of a result of DiBenedetto [11] asserting that for solutions to the model equation (1.4), we have

$$\mu \in L^{N+\delta} \implies Du \text{ is Hölder continuous}$$

whenever  $\delta > 0$ , where  $N = n + 2$  is the usual parabolic dimension (see [11, Chapter IX]). A borderline version of DiBenedetto's result, valid now for general equations as in (1.1) and given in terms of natural Lorentz spaces, is the following:

**Corollary 1.8 (Lorentz spaces criterion).** *Let  $u$  be a solution to (1.1) in  $\Omega_T$ , under the assumptions (1.2) with  $p > 2 - 1/(n + 1)$ ; assume also that*

$$\mu \in L(N, q), \quad \text{where } q := \min\{1, p/[2(p - 1)]\}.$$

*Then  $Du$  is continuous in  $\Omega_T$ .*

Needless to say,  $L^{N+\delta} \subset L(N, q)$  for every  $q > 0$ . We refer to Section 2 below for the relevant definitions on Lorentz spaces in this setting. See also [27] for a related gradient boundedness result.

The last result we propose extends to the parabolic case by now classical elliptic results of Lieberman [35] asserting that, when considering elliptic equations as (1.10), the density condition

$$|\mu|(B_\varrho) \leq c\varrho^{n-1+\delta}$$

for some  $\delta > 0$  implies the local Hölder continuity of the gradient (see also the work of Kilpeläinen [19] for similar statements concerning  $u$  rather than  $Du$ ). Here we find that a similar condition still implies Hölder continuity, obviously replacing  $n$  with the parabolic dimension  $N = n + 2$ .

**Theorem 1.9 (Measure density criterion).** *Let  $u$  be a solution to (1.1) in  $\Omega_T$ , under the assumptions (1.2) with  $p > 2 - 1/(n + 1)$ ; assume also that*

$$|\mu|(Q_\varrho) \leq c_D \varrho^{N-1+\delta} \tag{1.24}$$

*holds for some  $c_D \geq 1$  and  $\delta > 0$ , whenever  $Q_\varrho$  is a standard parabolic cylinder having width equal to  $\varrho$ . Then there exists an exponent  $\beta \in (0, 1)$ , depending only on  $n, p, v, L$  and  $\delta$ , such that  $Du \in C_{loc}^{0,\beta}(\Omega_T, \mathbb{R}^n)$ .*

The borderline case of the previous result also follows:

**Corollary 1.10 (Borderline measure density criterion).** *Let  $u$  be a solution to (1.1) in  $\Omega_T$ , under the assumptions (1.2) with  $p > 2 - 1/(n + 1)$ . Assume that, for a non-negative function  $h : [0, \infty) \rightarrow [0, \infty)$  such that*

$$\int_0^\infty [h(\varrho)]^q \frac{d\varrho}{\varrho} < \infty, \quad \text{where } q = \min\{1, p/[2(p - 1)]\},$$

*the inequality*

$$|\mu|(Q_\varrho) \leq c\varrho^{N-1}h(\varrho) \tag{1.25}$$

*holds for some  $c \geq 1$ , whenever  $Q_\varrho$  is a standard parabolic cylinder having width equal to  $\varrho$ . Then  $Du$  is continuous in  $\Omega_T$ .*

The proof of the gradient continuity criteria of Theorems 1.5-1.9 involves some delicate new ideas. In particular, we do not use the stopping time arguments on chains on shrinking cylinders, where we measure the oscillations of  $Du$  (as for the proof of Theorem 1.1 – a method introduced earlier in [30]). On the contrary, we build a new iteration scheme using an a priori large number of stopping times, by instead recovering the regularity estimates using excess decay properties that hold exactly on chains of cylinders bounded by the exit times considered. We shall call these as *maximal iteration chains* reflecting the degeneracy/singularity of the equation and they are defined in due course of the proof of Theorem 1.6. The first result

we prove is indeed Theorem 1.6; after proving the VMO-regularity of the gradient we use this result to prove its continuity under the more restrictive assumption of the convergence of the potentials. A particular feature of the proof is that it is made in order to keep track of the constant dependence in a very precise way. Indeed, intrinsic in the proof of Theorem 1.6 is a control of the modulus of continuity of the gradient in terms of certain radii that, when considering (1.24), leads to the proof of Theorem 1.9. This last results indeed relies in turn on the method of Theorem 1.6 together with a *very careful analysis of the constants dependence in the a priori estimates for solutions to homogeneous equations*, when considered with respect to a certain exit time argument. Indeed, for Theorem 1.9 it is necessary to develop a suitable version of Theorem 3.1 below, namely, Theorem 3.5, where a more careful constants dependence is connected to a certain double sided inequality for the intrinsic geometry.

**Remark 1.11.** The techniques of this paper are suitable to treat more general equations of the type

$$u_t - \operatorname{div} a(x, t, Du) = \mu$$

under suitable regularity assumptions on the partial map  $x \mapsto a(x, \cdot)$ . More precisely, we assume that for every point  $(x, t)$  the vector field  $z \mapsto a(x, \cdot)$  satisfies assumptions (1.2), while the map  $x \mapsto a(\cdot, t, \cdot)$  is just assumed to be measurable. Instead, with respect to to the  $x$  variable we assume a Dini continuity dependence, that is, we assume that

$$|a(x, t, z) - a(y, t, z)| \leq \omega(|x - y|)$$

holds whenever  $x, y \in \Omega, t \in (-T, 0)$  and  $z \in \mathbb{R}^n$ , where

$$\int_0^\infty \omega(\varrho) \frac{d\varrho}{\varrho} < \infty. \quad (1.26)$$

Note that we do not need to assume any regularity on the partial map  $t \rightarrow a(\cdot, t, \cdot)$  other than measurability. For the sake of brevity we shall not report the necessary modifications; this can be obtained by using the techniques introduced in this paper with those of forthcoming [31]. Indeed, in this last paper we shall further exploit the ideas used here to prove a few regularity results for solutions to  $p$ -Laplacian type evolutionary systems with coefficients. Moreover, we shall also show how to use the techniques of this paper to prove the Lipschitz continuity of solutions to systems which are asymptotically close to the  $p$ -Laplacian one. See also the paper [16], where condition (1.26) has been used in the elliptic case.

### 1.3. Comparison with the Barenblatt solution

Here we are going to demonstrate the sharpness of the intrinsic potential estimate of Theorem 1.1 by showing that Theorem 1.1 provides for solutions to general equations as

$$u_t - \operatorname{div} a(Du) = \delta \quad (1.27)$$

in  $\mathbb{R}^n \times \mathbb{R}$ , where  $\delta$  is the unit Dirac mass charging the origin, the same asymptotic estimates that hold for the fundamental solution of

$$u_t - \operatorname{div}(|Du|^{p-2}Du) = \delta. \tag{1.28}$$

More precisely, for the choice of  $p$  and  $\theta$  as follows:

$$\frac{2n}{n+1} < 2 - \frac{1}{n+1} < p < 2 \quad \text{and} \quad \theta = n(p-2) + p, \tag{1.29}$$

and for a suitable choice of the positive constant  $c_b \equiv c_b(n, p)$ , the function

$$\mathcal{B}_p(x, t) := \begin{cases} t^{-n/\theta} \left[ \frac{2-p}{p} \theta^{-\frac{1}{p-1}} \left( c_b + \left( \frac{|x|}{t^{1/\theta}} \right)^{\frac{p}{p-1}} \right) \right]^{\frac{p-1}{p-2}} & (x, t) \in \mathbb{R}^n \times (0, \infty) \\ 0 & \text{otherwise} \end{cases}$$

is a very weak solution – in the sense of [6] – to (1.28) in  $\mathbb{R}^n \times \mathbb{R}$ . See for instance [47, page 192]. Specifically,  $\mathcal{B}_p$  still verifies (1.6), but fails to match the integrability conditions in (1.5), as it usually happens when dealing with solutions to problems involving measure data. The function  $\mathcal{B}_p$  in (1.30) is indeed called the Barenblatt solution to (1.28) (see [2] for the original motivations). A direct calculation gives

$$|D\mathcal{B}_p(x_0, t_0)| \leq c t_0^{-(n+1)/\theta}. \tag{1.30}$$

As for the bounds in (1.29) notice that indeed  $D\mathcal{B}_p(x, t) \in L^1$ . In fact, observe that  $\mathcal{B}_p(x, t) = t^{-n/\theta} v(|x|t^{-1/\theta})$  with  $v$  being a function in  $C^{1,\alpha}(\mathbb{R}^+)$  for some  $\alpha > 0$ . Then, co-area formula and change of variables give

$$\int_0^T \int_{\mathbb{R}^n} |D\mathcal{B}_p(x, t)| dx dt = \omega_{n-1} \int_0^\infty |v'(r)| r^{n-1} dr \int_0^T t^{-1/\theta} dt < \infty$$

for all  $T > 0$  if and only if

$$-\frac{1}{\theta} > -1 \iff n(p-2) + p > 1 \iff p > \frac{2n+1}{n+1} = 2 - \frac{1}{n+1},$$

which is exactly the bound in (1.9). Note that here we have also used the estimate

$$\int_0^\infty |v'(r)| r^{n-1} dr \leq c \left( 1 + \int_1^\infty r^{\frac{1}{p-1} + \frac{p}{p-1} \left( \frac{p-1}{p-2} - 1 \right) + n-1} dr \right) < \infty$$

for some  $c \equiv c(n, p)$ , and in turn this is implied by

$$\frac{1}{p-1} + \frac{p}{p-1} \left( \frac{p-1}{p-2} - 1 \right) < -n \iff \frac{2}{p-2} < -n \iff p > 2 - \frac{2}{n}.$$

It follows that

$$A_r(\lambda) := \frac{1}{|Q_r|} \int_{Q_{r\lambda}^\lambda} |D\mathcal{B}_p| \, dx \, dt \rightarrow 0 \tag{1.31}$$

as  $r \rightarrow \infty$ , uniformly with respect to  $\lambda > 0$  and  $(x_0, t_0) \in \mathbb{R}^n \times \mathbb{R}$ . In fact, this is true for solutions  $u$  to more general Cauchy problems whenever the initial trace of  $u$  is compactly supported, *i.e.* that the source term is concentrated on  $t = 0$  and has a compact support, see for example [11, Chapter 11] and [33].

The point we are interested in here is that it is possible to show the pointwise bound appearing in (1.30) as a direct consequence of Theorem 1.1 and convergence property (1.31), thereby proving its sharpness. For this, let us start observing that, with  $(x_0, t_0) \in \mathbb{R}^n \times (0, \infty)$ , we have  $t_0 - \lambda^{2-p} \varrho^2 < 0$  iff  $\varrho > \sqrt{\lambda^{p-2} t_0}$ . Therefore, whenever  $\lambda > 0$ , we obtain

$$\int_0^\infty \left( \frac{\delta(Q_\varrho^\lambda(x_0, t_0))}{\varrho^{N-1}} \right) \frac{d\varrho}{\varrho} \leq \int_{\sqrt{\lambda^{p-2} t_0}}^\infty \left( \frac{1}{\varrho^{N-1}} \right) \frac{d\varrho}{\varrho} = c \left( \lambda^{p-2} t_0 \right)^{-(n+1)/2}. \tag{1.32}$$

Now, let us define, for  $\lambda > 0$ , the function

$$h_r(\lambda) := \lambda - c\lambda^{n(2-p)/2} A_r(\lambda) - c \int_0^{2r\lambda} \left( \frac{\delta(Q_\varrho^\lambda(x_0, t_0))}{\varrho^{N-1}} \right) \frac{d\varrho}{\varrho},$$

where  $A_r(\cdot)$  has been defined in (1.31). Observe that under the assumption (1.3) we have  $n(2-p)/2 < 1$  and thus the function  $h_r(\cdot)$  is negative in a neighborhood of 0 (if  $A_r(\lambda) \equiv 0$  for some  $\lambda > 0$ , there is nothing to prove). On the other hand, for the same reason, and noticing also that again by (1.3) we have  $(2-p)(n+1)/2 < 1$  by (1.32), we conclude with

$$\lim_{\lambda \rightarrow \infty} h_r(\lambda) \geq \lim_{\lambda \rightarrow \infty} \left[ \lambda - \lambda^{n(2-p)/2} A_r(1) - c\lambda^{(2-p)(n+1)/2} t_0^{-(n+1)/2} \right] = \infty.$$

The function  $h_r(\cdot)$  is, on the other hand, easily seen to be continuous and therefore there exists a solution  $\lambda_r > 0$  of the equation  $h_r(\lambda_r) = 0$ . We can hence apply Theorem 1.1 to  $\mathcal{B}_p$  in order to obtain

$$|D\mathcal{B}_p(x_0, t_0)| + \lambda_r \leq 2c\lambda_r^{n(2-p)/2} A_r(\lambda_r) + 2c \int_0^\infty \left( \frac{\delta(Q_\varrho^{\lambda_r}(x_0, t_0))}{\varrho^{N-1}} \right) \frac{d\varrho}{\varrho}.$$

By using Young’s inequality with conjugate exponents

$$\left( \frac{2}{n(2-p)}, \frac{2}{2-n(2-p)} \right)$$

we estimate

$$\lambda_r^{n(2-p)/2} A_r(\lambda_r) \leq \frac{\lambda_r}{4} + c[A_r(\lambda_r)]^{2/[2-n(2-p)]}.$$

On the other hand, noting  $\theta/2 + (n + 1)(2 - p)/2 = 1$ , again Young’s inequality implies

$$2c \int_0^\infty \left( \frac{\delta(Q_\varrho^{\lambda_r}(x_0, t_0))}{\varrho^{N-1}} \right) \frac{d\varrho}{\varrho} \leq \frac{\lambda_r}{4} + ct_0^{-(n+1)/\theta}.$$

Merging the last three estimates and reabsorbing terms yields

$$|D\mathcal{B}_p(x_0, t_0)| \leq c[A_r(\lambda_r)]^{2/[2-n(2-p)]} + ct_0^{-(n+1)/\theta},$$

from which (1.30) finally follows letting  $r \rightarrow \infty$ , thanks to (1.31).

**1.4. Plan of the paper**

The paper is organized as follows; after collecting the basic notation and terminology in Section 2, we present in Section 3 the results about Hölder gradient continuity for solutions to homogeneous equations. The central result is at this stage Theorem 3.1, which will be a main tool both in the proof of the potential estimates and in that of the gradient continuity results. In turn, Theorem 3.1 is used here to give a proof of certain a priori Hölder continuity estimates for solutions of homogeneous equations (see Theorem 3.3), that will be also an important tool in the subsequent proof. The proof of the potential estimates in Theorems 1.1-1.3 are then in Section 4. Finally, Section 5 is devoted to the proof of Theorems 1.6-1.9 and their consequences.

ACKNOWLEDGEMENTS. The authors thank Paolo Baroni for his remarks on a preliminary version of the paper.

**2. Basic notation and definitions**

**2.1. Notation**

In what follows we denote by  $c$  a general positive constant, possibly varying from line to line; special occurrences will be denoted by  $c_1, c_2$  etc and relevant dependencies on parameters will be emphasized using parentheses. *All such constants will be greater than or equal to 1.* We also denote by

$$B(x_0, r) := \{x \in \mathbb{R}^n : |x - x_0| < r\}$$

the open ball of  $\mathbb{R}^n$  with center  $x_0$  and radius  $r > 0$ ; when not important, or clear from the context, we shall omit denoting the center as follows:  $B_r \equiv B(x_0, r)$ . Unless otherwise stated, different balls in the same context will have the same center. We shall also denote  $B \equiv B_1 = B(0, 1)$ . In a similar fashion we shall denote by

$$Q_r(x_0, t_0) := B(x_0, r) \times (t_0 - r^2, t_0)$$

the standard parabolic cylinder with vertex  $(x_0, t_0)$  and width  $r > 0$ . When the vertex will not be important in the context or it will be clear that all the cylinders occurring in a proof will share the same vertex, we shall omit to indicate it, simply denoting  $Q_r$ . With  $\lambda > 0$  being a free parameter, we shall often consider cylinders of the type

$$Q_r^\lambda(x_0, t_0) := B(x_0, r) \times (t_0 - \lambda^{2-p}r^2, t_0). \tag{2.1}$$

These will be called “intrinsic cylinders” as they will be usually employed in a context when the parameter  $\lambda$  is linked to the behavior of the solution to some equation on the same cylinder  $Q_r^\lambda$ . Again, when specifying the vertex will not be essential we shall simply denote  $Q_r^\lambda \equiv Q_r^\lambda(x_0, t_0)$ . Observe that the intrinsic cylinders reduce to the standard parabolic ones when either  $p = 2$  or  $\lambda = 1$ . In the rest of the paper  $\lambda$  will always denote a constant larger than zero and will be considered in connection to intrinsic cylinders as (2.1). We shall often denote

$$\delta Q_r^\lambda(x_0, t_0) \equiv Q_{\delta r}^\lambda(x_0, t_0) = B(x_0, \delta r) \times (t_0 - \lambda^{2-p}\delta^2r^2, t_0)$$

the intrinsic cylinder with width magnified of a factor  $\delta > 0$ . Finally, with  $Q = A \times (t_1, t_2)$  being a cylindrical domain, we denote by

$$\partial_{\text{par}}Q := \bar{A} \times \{t_1\} \cup \partial A \times (t_1, t_2)$$

the usual parabolic boundary of  $Q$ , and this is nothing else but the standard topological boundary without the upper cap  $A \times \{t_2\}$ .

The parabolic metric is defined as

$$|(x_1, t_1) - (x_2, t_2)|_{\text{par}} := \max \left\{ |x_1 - x_2|, \sqrt{|t_1 - t_2|} \right\}$$

and the corresponding parabolic distance between sets  $E_1$  and  $E_2$  in  $\mathbb{R}^{n+1}$  as

$$\text{dist}_{\text{par}}(E_1, E_2) = \inf \left\{ |(x_1, t_1) - (x_2, t_2)|_{\text{par}} : (x_1, t_1) \in E_1, (x_2, t_2) \in E_2 \right\}.$$

With  $A \subset \mathbb{R}^{n+1}$  being a measurable subset with positive measure, and with  $g : A \rightarrow \mathbb{R}^n$  being a measurable map, we shall denote by

$$\int_A g(x) dx dt := \frac{1}{|A|} \int_A g(x) dx dt$$

its integral average; of course  $|A|$  denotes the Lebesgue measure of  $A$ . A similar notation is adopted if the integral is only in space or time. In the following, with  $g \equiv (g_i)_{1 \leq i \leq n} \in \mathbb{R}^n$  being this time a vector, we define

$$\|g\| := \max_i |g_i|,$$

which is a norm equivalent to the usual one defined by  $|g|^2 := \sum |g_i|^2$  via

$$\|g\| \leq |g| \leq \sqrt{n}\|g\|.$$

Moreover, everywhere in the following, when considering the sup operator we shall actually mean  $\text{esssup}$ , and similarly for inf and osc operators.

Given a real-valued function  $h$  and a real number  $k$ , we shall denote

$$(h - k)_+ := \max\{h - k, 0\} \quad \text{and} \quad (h - k)_- := \max\{k - h, 0\}.$$

As in this paper we are considering only a priori estimates – see the discussion in the Introduction – we shall restrict ourselves to examine the case when  $\mu$  is an integrable function. In particular, with  $g: A \rightarrow \mathbb{R}^n$  being a vector-valued map we denote

$$\text{osc}_A g := \sup_{(x,t), (x_0,t_0) \in A} |g(x, t) - g(x_0, t_0)|.$$

In the treatment of parabolic equations, a standard difficulty in using test functions arguments involving the solution is that we start with solutions that, enjoying the regularity in (1.5), do not have in general time derivatives in any reasonable sense. There are several, by now standard, ways to overcome this point, for instance using a regularization procedure via so-called Steklov averages. See for instance [11, Chapter 2] for their definition and their standard use. In this paper, in order to concentrate the attention only on significant issues, following a by now standard custom (see for instance [12]), we shall argue on a formal level, that is assuming when using test functions argument, that the solution has for instance square integrable time derivatives. Such arguments can easily be made rigorous using in fact Steklov averages as for instance in [11] or using convolutions with mollifiers.

**2.2. Lorentz spaces and nonlinear potentials**

We start by recalling a few basic definitions concerning Lorentz spaces; in this section  $\mu: \mathbb{R}^{n+1} \rightarrow \mathbb{R}$  will in general denote a measurable map, while  $Q \subset \mathbb{R}^{n+1}$  will denote an open subset. We assume  $|\{(x, t) \in Q : |\mu(x, t)| > t\}| < \infty$  for  $t \geq 0$ . The decreasing rearrangement  $\mu^*: [0, \infty] \rightarrow [0, \infty]$  of  $\mu$  is defined as the (unique) non-increasing, right-continuous function which is equi-distributed with  $|\mu(\cdot)|$ , that is

$$\mu^*(s) := \sup \{h \geq 0 : |\{(x, t) \in \mathbb{R}^{n+1} : |\mu(x, t)| > h\}| > s\}.$$

Now, the usual definition of the Lorentz space  $L(\gamma, q)(Q) \equiv L(\gamma, q)$ , for  $\gamma \in (0, \infty)$  and  $q \in (0, \infty)$  prescribes that

$$[\mu]_{\gamma, q} := \left( \frac{q}{\gamma} \int_0^\infty \left( \mu^*(\varrho) \varrho^{1/\gamma} \right)^q \frac{d\varrho}{\varrho} \right)^{1/q} < \infty. \tag{2.2}$$

The local version of Lorentz spaces is defined in the usual way by saying that  $\mu \in L(\gamma, q)$  locally in  $Q$  iff  $\chi_A \mu \in L(\gamma, q)$  for every open subset  $A \Subset Q$ .

Lorentz spaces refine the standard Lebesgue spaces and it follows from the definition that  $L(\gamma, \gamma) \equiv L^\gamma$ . For more on Lorentz spaces we refer for instance



to [44]. A classical fact due to Hunt [18] states that when considering the maximal operator of  $\mu^*$  it is possible to obtain a quantity, in fact equivalent to that in (2.2) when  $\gamma > 1$ , that defines a norm in  $L(\gamma, q)$ . More precisely, defining for  $\xi > 0$  the maximal operator

$$\mu^{**}(\xi) := \frac{1}{\xi} \int_0^\xi \mu^*(y) dy, \tag{2.3}$$

for  $q < \infty$  the quantity

$$\|\mu\|_{\gamma,q} := \left( \frac{q}{\gamma} \int_0^\infty \left( \mu^{**}(\varrho) \varrho^{1/\gamma} \right)^q \frac{d\varrho}{\varrho} \right)^{1/q}$$

is such that

$$[\mu]_{\gamma,q} \leq \|\mu\|_{\gamma,q} \leq c(\gamma, q)[\mu]_{\gamma,q} \quad \text{for } \gamma > 1; \tag{2.4}$$

see for instance [44, Theorem 3.21]. The following inequality, which is a straightforward corollary of the definition of rearrangement of a function, holds whenever  $A \subseteq Q$  is a measurable set:

$$\int_A |\mu(x, t)| dx dt \leq \int_0^{|A|} \mu^*(y) dy. \tag{2.5}$$

We next define the following family of Wolff-type nonlinear potentials:

$$\mathbf{W}_q^\mu(x, t; r) := \int_0^r \left( \frac{|\mu|(Q_\varrho(x, t))}{\varrho^{N-1}} \right)^q \frac{d\varrho}{\varrho}, \quad q > 0. \tag{2.6}$$

Obviously

$$\mathbf{W}_1^\mu \equiv \mathbf{I}_1^\mu \quad \text{and} \quad \mathbf{W}_{p/[2(p-1)]}^\mu(x, t; r) = \int_0^r \left( \frac{|\mu|(Q_\varrho(x, t))}{\varrho^{N-1}} \right)^{p/[2(p-1)]} \frac{d\varrho}{\varrho}. \tag{2.7}$$

The following lemma is rather straightforward consequence of the definitions:

**Lemma 2.1.** *Let  $\mu \in L^1(\mathbb{R}^{n+1}, \mathbb{R})$  and  $q > 0$ ; for every  $r > 0$  it holds that*

$$\sup_{(x,t) \in Q} \mathbf{W}_q^\mu(x, t; r) \leq c_1 \int_0^{\omega_n r^N} \left( \mu^{**}(\varrho) \varrho^{1/N} \right)^q \frac{d\varrho}{\varrho}, \tag{2.8}$$

where the constant  $c_1$  depends only on  $n, q$  and  $\omega_n$  denotes the measure of the unit ball in  $\mathbb{R}^n$ . In particular

$$\mu \in L(N, q) \implies \lim_{r \rightarrow 0} \sup_{(x,t) \in Q} \mathbf{W}_q^\mu(x, t; r) = 0. \tag{2.9}$$

*Proof.* We give the proof for completeness. Observe that (2.5) and the definition in (2.3) give, whenever  $(x_0, t_0) \in Q$

$$\begin{aligned} \left( \frac{|\mu|(Q_\varrho(x_0, t_0))}{\varrho^{N-1}} \right)^q &= \left( \omega_n \varrho \int_{Q_\varrho(x_0, t_0)} |\mu| \, dx \, dt \right)^q \leq \left( \frac{\varrho}{\varrho^N} \int_0^{\omega_n \varrho^N} \mu^*(\xi) \, d\xi \right)^q \\ &\leq \left[ \omega_n \varrho \mu^{**}(\omega_n \varrho^N) \right]^q. \end{aligned}$$

Therefore, integrating the previous inequality in  $(0, r)$ , and changing variables, leads to (2.8). □

### 3. The $C^{0,\alpha}$ -gradient theory

In this section we concentrate on homogeneous equations of the type

$$w_t - \operatorname{div} a(Dw) = 0 \tag{3.1}$$

in a given cylinder  $Q = B \times (t_1, t_2)$ , where  $B \subset \mathbb{R}^n$  is a given ball. We remark that in this section the exponent  $p$  will always be, unless otherwise stated, such that  $p \in (1, 2]$ . In particular, in this section the lower bound (1.3) plays no role. In the rest of the section  $w$  will denote an energy solution to (3.1) in a given cylinder  $Q$ .

In the following, with  $q \geq 1$  and  $g \in L^q(Q')$  being a function defined on the measurable set  $Q' \subset \mathbb{R}^{n+1}$ , we define the excess functional

$$E_q(g, Q') := \left( \int_{Q'} |g - (g)_{Q'}|^q \, dx \, dt \right)^{1/q} \tag{3.2}$$

which is the  $L^q$ -mean deviation of  $g$  from its mean value on  $Q'$ . An elementary property of excess functionals we shall often use in the following is

$$E_q(g, Q') \leq 2 \left( \int_{Q'} |g - \gamma|^q \, dx \, dt \right)^{1/q} \quad \forall \gamma \in \mathbb{R}^n. \tag{3.3}$$

A central tool in our approach to both the Hölder continuity of the spatial gradient of solutions to homogeneous equations and to gradient potential estimates in the case of non-homogeneous equations, is provided by the following:

**Theorem 3.1.** *Suppose that  $w$  is a weak solution to (3.1) in  $Q_r^\lambda$ , and consider numbers  $A, B, q \geq 1$  and  $\gamma \in (0, 1)$ . Then there exist constants  $\delta_\gamma \in (0, 1/2)$  and  $\xi \in (0, 1/4)$ , both depending only on  $n, p, \nu, L, A, B, \gamma$ , but otherwise independent of  $s, q$ , the solution  $w$ , and the vector field  $a(\cdot)$ , such that if*

$$\frac{\lambda}{B} \leq \max \left\{ \frac{s}{\xi B}, \sup_{Q_{\delta_\gamma r}^\lambda} \|Dw\| \right\}, \quad s + \sup_{Q_r^\lambda} \|Dw\| \leq A\lambda \tag{3.4}$$

hold, then

$$E_q(Dw, \delta_\gamma Q_r^\lambda) \leq \gamma E_q(Dw, Q_r^\lambda) \tag{3.5}$$

holds as well. Moreover, there exist two constants  $\alpha \in (0, 1)$  and  $c(A, B) \geq 1$ , both depending only on  $n, p, v, L, A, B$ , such that

$$\delta_\gamma = \frac{\gamma^{1/\alpha}}{c(A, B)}, \tag{3.6}$$

and  $c(A, B)$  is a nondecreasing function of  $A, B$ . Finally, (3.5) still holds when replacing  $\delta_\gamma$  by a smaller number  $\delta$ .

**Remark 3.2.** The proof of Theorem 3.1 will be given first in the case

$$s > 0, \tag{3.7}$$

with estimates that will be on the other hand uniform with respect to  $s$ . A suitable approximation argument then allows to treat the case  $s = 0$  starting from the one  $s > 0$ , and for this we refer to Section 3.3 below.

Theorem 3.1 is a key ingredient in the proof of the intrinsic potential estimates of Theorems 1.1-1.3 but it also incorporates enough information to allow for a proof of the Hölder continuity of  $Dw$ , as shown in the next Theorem 3.3. In other words, Theorem 3.1 allows for a unified approach to the gradient regularity for both homogeneous and non-homogeneous  $p$ -Laplacian equations. In turn, the a priori estimate in Theorem 3.3 below is another key tool in the proof of the potential estimates. We notice that a slightly different version of Theorem 3.1 has already been proved in [30] for the case  $p \geq 2$ ; here we show the necessary modifications of the case  $p < 2$  — which are at several stages non-trivial — and then show how to apply Theorem 3.1 to prove the following:

**Theorem 3.3.** *Let  $w$  be a weak solution to (3.1) in a given cylinder  $Q$  under assumptions (1.2). Then*

- If

$$\frac{2n}{n+2} < p \leq 2$$

holds, then  $Dw$  has is locally Hölder continuous in  $Q$

- If  $1 < p \leq 2$  and  $Q_r^\lambda \subset Q$  is an intrinsic cylinder such that

$$s + \sup_{Q_r^\lambda} \|Dw\| \leq A\lambda \tag{3.8}$$

holds for a certain constant  $A \geq 1$ , then

$$|Dw(x, t) - Dw(x_1, t_1)| \leq 4\sqrt{n}A\lambda \left(\frac{\varrho}{r}\right)^\alpha \tag{3.9}$$

holds as well whenever  $(x, t), (x_1, t_1) \in Q_\varrho^\lambda$  for a Hölder exponent  $\alpha \equiv \alpha(n, p, v, L, A) \in (0, 1)$ , which, in particular, is independent of  $s$ , the solution  $w$  considered and the vector field  $a(\cdot)$ . Here  $Q_\varrho^\lambda \subset Q_r^\lambda$  are intrinsic cylinders sharing the same vertex.

*Proof.*

*Step 0: Reduction to the second assertion.* Let us observe that it is sufficient to prove the second assertion; in fact by Theorem 4.2 below, we know that  $Dw$  is locally bounded in  $Q$  and the Hölder continuity follows from the second assertion via a standard covering argument (compare for instance with [11]).

*Step 1: Framework of Theorem 3.1.* Our approach to prove the second part of the Theorem is to apply Theorem 3.1 with particular choices of  $B$  and  $\gamma$ , and with  $q = \infty$ , i.e. letting  $q \rightarrow \infty$  in the corresponding inequalities of Theorem 3.1 as all the constants are independent of  $q$ . Choose

$$B = 8\sqrt{n}, \quad \gamma = \frac{1}{32n^{3/2}A}, \tag{3.10}$$

and let  $\delta_\gamma \equiv \delta_\gamma(n, p, v, L, A) \in (0, 1/2)$  be the constant in Theorem 3.1 corresponding  $A, B$ , and  $\gamma$ . We start defining

$$\lambda_k := 2^{-k}\lambda, \quad r_k := \delta_\gamma^k r, \quad Q_k := Q_{r_k}^{\lambda_k},$$

$k = 0, 1, \dots$ , so that  $Q_{k+1} \subset \delta_\gamma Q_k$  holds whenever  $k \geq 0$ . Define the number

$$m := \inf \left\{ k \geq 0 : \max \left\{ \frac{8s}{B}, \sup_{Q_{k+1}} \|Dw\| \right\} > \frac{2\lambda_k}{B} \right\}.$$

In case  $m < \infty$  we *redefine* the intrinsic geometry parameter

$$\lambda_k = \lambda_m, \quad k = m + 1, m + 2, \dots,$$

and accordingly, we redefine  $Q_k := Q_{r_k}^{\lambda_m}$  for  $k = m + 1, m + 2, \dots$

*Step 2: The equation becomes uniformly parabolic immediately.* This means that  $m = 0$ . We then have that at least one of the two following holds true:

$$\lambda_0/4 < s \leq s + \sup_{Q_0} \|Dw\| \leq A\lambda_0, \tag{3.11}$$

and

$$\lambda_0/B \leq 2\lambda_0/B < \sup_{Q_1} \|Dw\| \leq s + \sup_{Q_0} \|Dw\| \leq A\lambda_0. \tag{3.12}$$

Let us prove that, in any case, we have

$$\operatorname{osc}_{Q_{k+1}} Dw \leq 2\sqrt{n}\gamma \operatorname{osc}_{Q_k} Dw \quad \forall k \geq 0. \tag{3.13}$$

Indeed, Theorem 3.1 implies (applied with  $q = \infty$  on the fourth line below)

$$\begin{aligned} \operatorname{osc}_{Q_{k+1}} Dw &\leq \sqrt{n} \max_{1 \leq j \leq n} \operatorname{osc}_{Q_{k+1}} w_{x_j} \\ &\leq 2\sqrt{n} \max_{1 \leq j \leq n} \|w_{x_j} - (w_{x_j})_{Q_{k+1}}\|_{L^\infty(Q_{k+1})} \\ &\leq 2\sqrt{n} \|Dw - (Dw)_{Q_{k+1}}\|_{L^\infty(Q_{k+1})} \\ &\leq 2\sqrt{n}\gamma \|Dw - (Dw)_{Q_k}\|_{L^\infty(Q_k)} \\ &\leq 2\sqrt{n}\gamma \operatorname{osc}_{Q_k} Dw \end{aligned} \tag{3.14}$$

for all  $k = 0, 1, \dots$  in the case (3.11) holds (recall that  $\lambda_k = \lambda_0$ ), and for  $k = 0$  in the case (3.12) holds. Let us further analyze the occurrence of (3.12) in order to prove that (3.13) holds when (3.12) is in force. We have by (3.12) and (3.14) (this last one used only for  $k = 0$ ) that

$$\operatorname{osc}_{Q_1} Dw \leq 2\sqrt{n}\gamma \operatorname{osc}_{Q_0} Dw \leq 4n\gamma \sup_{Q_0} \|Dw\| \leq 4n\gamma A\lambda_0.$$

Therefore

$$\sup_{Q_1} w_{x_i} - \inf_{Q_1} w_{x_i} = \operatorname{osc}_{Q_1} w_{x_i} \leq \operatorname{osc}_{Q_1} Dw \leq 4n\gamma A\lambda_0 \tag{3.15}$$

holds for every  $i \in \{1, \dots, n\}$ . Next, we have by (3.12) that either

$$\sup_{Q_1} w_{x_i} = \sup_{Q_1} \|Dw\| \geq 2\lambda_0/B \quad \text{or} \quad \inf_{Q_1} w_{x_i} = -\sup_{Q_1} \|Dw\| \leq -2\lambda_0/B$$

holds for some  $i \in \{1, \dots, n\}$ . If the first inequality from the previous line is true, *i.e.*

$$\sup_{Q_1} w_{x_i} = \sup_{Q_1} \|Dw\| \geq 2\lambda_0/B$$

holds for some  $i \in \{1, \dots, n\}$ , then by (3.15)

$$2\lambda_0/B - \inf_{Q_1} w_{x_i} \leq \sup_{Q_1} w_{x_i} - \inf_{Q_1} w_{x_i} \leq 4n\gamma A\lambda_0,$$

and by (3.10) it follows that

$$\inf_{Q_1} w_{x_i} \geq \frac{\lambda_0}{B} (2 - 4nAB\gamma) = \frac{\lambda_0}{B}. \tag{3.16}$$

Similarly, if

$$\inf_{Q_1} w_{x_i} = -\sup_{Q_1} \|Dw\| \leq -2\lambda_0/B$$

for some  $i \in \{1, \dots, n\}$ , then we have

$$\sup_{Q_1} w_{x_i} \leq -\lambda_0/B. \tag{3.17}$$

Inequalities (3.16) and (3.17) readily imply that

$$\|Dw\| \geq \lambda_0/B \quad \text{a.e. in } Q_1.$$

By this and (3.12) we have that

$$\lambda_k/B \leq \sup_{Q_{k+1}} \|Dw\|, \quad s + \sup_{Q_k} \|Dw\| \leq A\lambda_k \tag{3.18}$$

for all  $k = 0, 1, \dots$  (again recall that in this case it is  $\lambda_k = \lambda_0$ ) and Theorem 3.1 applied with  $q = \infty$  then implies (3.13) for all  $k = 0, 1, \dots$

*Step 3: The equation becomes uniformly parabolic in  $Q_m$ .* This means that  $m \geq 1$ . We have that at least one of the following two inequalities holds:

$$s > \lambda_m/4 > \xi \lambda_m \quad \text{and} \quad \sup_{Q_{m+1}} \|Dw\| > 2\lambda_m/B.$$

Moreover, since  $m \geq 1$ , we have by the very definition of  $m$  that

$$\sup_{Q_k} \|Dw\| \leq \frac{1}{4\sqrt{n}} \lambda_{k-1} = \frac{1}{2\sqrt{n}} \lambda_k, \quad k = 1, \dots, m, \tag{3.19}$$

and, in particular, it follows that

$$\operatorname{osc}_{Q_k} Dw \leq 2\sqrt{n} \sup_{Q_k} \|Dw\| \leq \lambda_k, \quad k = 1, \dots, m. \tag{3.20}$$

We also have

$$s + \sup_{Q_m} \|Dw\| \leq \lambda_m \leq A\lambda_m, \tag{3.21}$$

because  $s \leq \lambda_{m-1}/4 = \lambda_m/2$ . Now, if on the one hand  $s > \lambda_m/4$ , then we get (3.14) by Theorem 3.1 for all  $k = m, m + 1, \dots$ , in view of (3.21). If, on the other hand,

$$2\lambda_m/B \leq \sup_{Q_{m+1}} \|Dw\|,$$

then we obtain, analogously to Step 2, that

$$\|Dw\| \geq \lambda_m/B \quad \text{a.e. in } Q_{m+1}.$$

This, together with (3.21) allows us to apply Theorem 3.1 with  $q = \infty$  as in Step 2; we obtain (3.13) for all  $k = m, m + 1, \dots$ .

*Step 4: The equation never becomes nondegenerate parabolic.* This means that  $m = \infty$  and therefore, as in the previous Step, we have that (3.20) holds for every  $k \in \mathbb{N}$ .

*Step 5: Conclusion.* In all cases we have established

$$\operatorname{osc}_{Q_k} Dw \leq \lambda_k, \quad 1 \leq k \leq m,$$

when  $m \geq 1$ , and

$$\operatorname{osc}_{Q_{k+1}} Dw \leq 2\sqrt{n}\gamma \operatorname{osc}_{Q_k} Dw \quad \text{for } k \geq m, \quad \operatorname{osc}_{Q_0} Dw \leq 2\sqrt{n}A\lambda,$$

when  $m \geq 0$ . Thus, because  $2\sqrt{n}\gamma \leq 1/2$ , above inequalities yield

$$\operatorname{osc}_{Q_k} Dw \leq \frac{\sqrt{n}A}{2^{k-1}} \lambda$$

for all  $k = 0, 1, \dots$ . But now, since

$$Q^\lambda_{(2^{(p-2)/2}\delta_\gamma)^k r} \subset Q^k$$

we have - by taking  $\varrho \leq r$  as in the statement of the Theorem - by choosing  $k \geq 0$  such that  $(2^{(p-2)/2}\delta_\gamma)^{k+1} < \varrho/r \leq (2^{(p-2)/2}\delta_\gamma)^k$ ,

$$\operatorname{osc}_{Q^\lambda_\varrho} Dw \leq \operatorname{osc}_{Q^k} Dw \leq 4\sqrt{n}A\lambda \left(\frac{\varrho}{r}\right)^\alpha$$

with

$$\alpha = -\frac{\log 2}{\log (2^{(p-2)/2}\delta_\gamma)}.$$

Clearly  $\alpha \equiv \alpha(n, p, \nu, L, A)$ , concluding the proof. □

**Remark 3.4.** Theorem 3.3 still holds when  $p > 2$ , and for this we refer to [30], where the needed version of Theorem 3.1 is featured.

In the following we shall also need a different version of Theorem 3.1, that is:

**Theorem 3.5.** *Suppose that  $w$  is a weak solution to (3.1) in  $Q^\lambda_r$ , and consider numbers  $A, B, q \geq 1$  and  $\gamma \in (0, 1)$ . Then there exists a constant  $\delta_\gamma \in (0, 1/2)$  depending only on  $n, p, \nu, L, A, B, \gamma$ , but otherwise independent of  $s, q$ , of the solution  $w$ , and of the vector field  $a(\cdot)$ , such that if*

$$\frac{\lambda}{B} \leq \sup_{Q^\lambda_{\delta_\gamma r}} \|Dw\| \quad \text{and} \quad s + \sup_{Q^\lambda_r} \|Dw\| \leq A\lambda \tag{3.22}$$

hold, then

$$E_q(Dw, \delta_\gamma Q^\lambda_r) \leq \gamma E_q(Dw, Q^\lambda_r) \tag{3.23}$$

holds as well. Moreover, there exist two constants  $\alpha \in (0, 1)$  and  $c(A) \geq 1$ , both depending only on  $n, p, \nu, L, A$ , but not on  $B$  and  $q$ , such that

$$\delta_\gamma = \frac{1}{c(A)} \left(\frac{\gamma}{B}\right)^{1/\alpha}, \tag{3.24}$$

and  $c(A)$  is a nondecreasing function of  $A$ . Finally, (3.23) still holds when replacing  $\delta_\gamma$  by a smaller number  $\delta$ .

This version, which differs from Theorem 3.1 for a different, more peculiar, dependence of the constant (3.24), will be necessary in order to prove Theorem 1.9 with the right dependence upon the parameters.

**Remark 3.6.** The assertions of Theorems 3.1-3.5 still hold for solutions to more general equations with coefficients of the type

$$w_t - \operatorname{div} a(t, Dw) = 0$$

where the vector field  $a(\cdot)$  is supposed to be only measurable with respect to the time variable, and such that  $z \mapsto a(t, z)$  satisfies assumptions (1.2) for every choice of  $t$ . This fact can be easily derived by observing that the proof of the above results solely relies on the analysis of the differentiated (with respect to the space variable) equation; see (3.35) below. The possibility of such a differentiation procedure is clearly not affected by an additional, even measurable, dependence on time.

**3.1. Basic Gradient Hölder continuity estimates**

In this section  $w$  denotes a solution to (3.1) in a cylinder of the type  $Q_r^\lambda \equiv Q$ ; throughout we shall assume that (3.7) holds; see the comments in Remark 3.2. We denote

$$\|v\|_{V^2(Q)}^2 := \sup_{t_1 < t < t_2} \int_B |v(x, t)|^2 dx + \int_Q |Dv(x, t)|^2 dx dt \tag{3.25}$$

whenever we are considering a cylinder of the type  $Q = B \times (t_1, t_2)$ . The space  $V^2(Q)$  is that defined by all those  $L^2(t_1, t_2; W^{1,2}(B))$  functions  $v$  such that the previous quantity is finite. Moreover we set  $V_0^2(Q) = V^2(Q) \cap L^2(t_1, t_2; W_0^{1,2}(B))$ . The following Poincaré-type inequality is then classical (see [11, Chapter 1, Corollary 3.1])

$$\|v\|_{L^2(Q_1)}^2 \leq c(n) |\{ |v| > 0 \} \cap Q_1|^{2/(n+2)} \|v\|_{V^2(Q_1)}^2, \quad Q_1 = B_1 \times (-1, 0), \tag{3.26}$$

and holds for all functions  $v \in V_0^2(Q_1)$ . We start with a sort of a height bound, already derived for the case  $p \geq 2$  in [30]; dealing with the singular case  $p < 2$  necessitates a different proof, and poses additional difficulties. In particular, we are forced to derive a Caccioppoli inequality which is different from the usual one involving an additional integration-by-parts argument.

**Proposition 3.7 (Height bound).** *Assume that*

$$s + \sup_{Q_r^\lambda} \|Dw\| \leq A\lambda \tag{3.27}$$

*holds for some constant  $A \geq 1$ . There exists a number  $\sigma \equiv \sigma(n, p, v, L, A) \in (0, 1/2)$  such that if either*

$$|Q_r^\lambda \cap \{w_{x_i} < \lambda/2\}| \leq \sigma |Q_r^\lambda| \quad \text{or} \quad |Q_r^\lambda \cap \{w_{x_i} > -\lambda/2\}| \leq \sigma |Q_r^\lambda| \tag{3.28}$$

*holds for some  $i \in \{1, \dots, n\}$ , then  $|w_{x_i}| \geq \lambda/4$  holds almost everywhere in  $Q_{r/2}^\lambda$ .*



*Proof.* Without loss of generality, we may assume that

$$|Q_r^\lambda \cap \{w_{x_i} < \lambda/2\}| \leq \sigma |Q_r^\lambda|, \tag{3.29}$$

then proving that  $w_{x_i} \geq \lambda/4$  holds almost everywhere in  $Q_{r/2}^\lambda$ . Indeed, assuming that the second inequality in (3.28) holds, by defining  $\tilde{w} := -w$  we have

$$|Q_r^\lambda \cap \{w_{x_i} > -\lambda/2\}| \leq \sigma |Q_r^\lambda| \iff |Q_r^\lambda \cap \{\tilde{w}_{x_i} < \lambda/2\}| \leq \sigma |Q_r^\lambda|. \tag{3.30}$$

Observing that  $\tilde{w}$  solves the equation  $\tilde{w}_t - \operatorname{div} \tilde{a}(D\tilde{w}) = 0$ , where  $\tilde{a}(z) := -a(-z)$  and the vector field  $\tilde{a}(\cdot)$  still satisfies assumptions (1.2), we reduce to the case (3.29). Again without loss of generality, we shall assume that the vertex of all the cylinders considered in the following coincides with the origin. The rest of the proof now goes in three steps.

*Step 1: Rescaling.* We now perform the standard intrinsic scaling by defining

$$v(x, t) := \frac{w(rx, \lambda^{2-p}r^2t)}{r}, \quad (x, t) \in Q_1 := B_1 \times (-1, 0), \tag{3.31}$$

so that the newly defined function  $v$  solves

$$\lambda^{p-2}v_t - \operatorname{div} a(Dv) = 0 \quad \text{in } Q_1. \tag{3.32}$$

From now on all the estimates will be recast in terms of the function  $v$ . Notice that with the new definition we still have

$$s + \|Dv\|_{L^\infty(Q_1)} \leq A\lambda \tag{3.33}$$

and assumption (3.29) translates into

$$|\{(x, t) \in Q_1 : v_{x_i}(x, t) < \lambda/2\}| \leq \sigma |Q_1|.$$

Our next aim is to show that

$$v_{x_i} \geq \lambda/4 \quad \text{a.e. in } Q_{1/2}. \tag{3.34}$$

The statement of the Proposition will then follow by scaling back to  $w$ .

*Step 2: A Caccioppoli-type estimate.* In the following we shall proceed formally (assuming for instance that solutions are differentiable with respect to time), and all the details can be justified in a standard way using Steklov averages [11]. The main point here is that since we are dealing with singular equations the treatment must be different from the one in [30]. We start by differentiating equation (3.32) in the  $x_i$ -direction; recall that we are assuming that  $p < 2$ ,  $s > 0$  and  $Dv$  is bounded. It follows, indeed

$$Dv \in L^2_{\text{loc}}(-1, 0; W^{1,2}_{\text{loc}}(B_1, \mathbb{R}^n)) \cap C^0(-1, 0; L^2_{\text{loc}}(B_1, \mathbb{R}^n)).$$

The details can be found for instance in [11, Chapter 8, Section 3]. We obtain the following equation:

$$\lambda^{p-2}(v_{x_i})_t - \operatorname{div} \partial_{x_i} (a(Dv)) = 0. \tag{3.35}$$

Multiplying the previous equation by  $(v_{x_i} - k)_-\phi^2$ , where  $\phi \in C_c^\infty(Q_1)$  and  $k \in (0, A\lambda)$ , and integrating by parts yields (actually using the weak formulation and in fact Steklov averages)

$$\begin{aligned} 0 &= \lambda^{p-2} \int_{Q_1} (v_{x_i})_t (v_{x_i} - k)_-\phi^2 \, dx \, dt \\ &\quad + \int_{Q_1} \langle \partial_{x_i} a(Dv), D\phi^2 \rangle (v_{x_i} - k)_- \, dx \, dt \\ &\quad + \int_{Q_1} \langle \partial_{x_i} a(Dv), D((v_{x_i} - k)_-) \rangle \phi^2 \, dx \, dt =: I_1 + I_2 + I_3. \end{aligned}$$

We integrate the first term by parts with respect to  $t$  and the second onewith respect to  $x_i$ . We obtain

$$I_1 = -\lambda^{p-2} \int_{Q_1} \partial_t \int_{v_{x_i}}^k (y - k)_- \, dy \, \phi^2 \, dx \, dt = \lambda^{p-2} \int_{Q_1} (v_{x_i} - k)_-^2 \phi_t \phi \, dx \, dt$$

and

$$\begin{aligned} I_2 &= - \int_{Q_1} \langle a(Dv), \partial_{x_i} D\phi^2 \rangle (v_{x_i} - k)_- \, dx \, dt \\ &\quad - 2 \int_{Q_1 \cap \{v_{x_i} < k\}} \langle a(Dv), D\phi \rangle \phi v_{x_i x_i} \, dx \, dt =: I_{2,1} + I_{2,2}. \end{aligned}$$

We estimate the terms  $I_{2,1}$ ,  $I_{2,2}$ , and  $I_3$  using the structure of  $a(\cdot)$  described in (1.2) as follows: the growth condition on  $a(\cdot)$  implies

$$|I_{2,1}| \leq 2L \int_{Q_1} (|Dv|^2 + s^2)^{(p-1)/2} (v_{x_i} - k)_- (|\phi| |D^2\phi| + |D\phi|^2) \, dx \, dt \tag{3.36}$$

and

$$|I_{2,2}| \leq 2L \int_{Q_1} (|Dv|^2 + s^2)^{(p-1)/2} |D(v_{x_i} - k)_-| |D\phi| |\phi| \, dx \, dt, \tag{3.37}$$

and the coercivity of  $\partial a(\cdot)$  gives

$$I_3 \leq -v \int_{Q_1} (|Dv|^2 + s^2)^{(p-2)/2} |D(v_{x_i} - k)_-|^2 \phi^2 \, dx \, dt. \tag{3.38}$$

Furthermore, from Young’s inequality we deduce

$$\begin{aligned} & 2L \left( |Dv|^2 + s^2 \right)^{(p-1)/2} |D(v_{x_i} - k)_-| |D\phi| |\phi| \\ & \leq \frac{\nu}{2} \left( |Dv|^2 + s^2 \right)^{(p-2)/2} |D(v_{x_i} - k)_-|^2 \phi^2 + \frac{2L^2}{\nu} \left( |Dv|^2 + s^2 \right)^{p/2} |D\phi|^2. \end{aligned}$$

Therefore,

$$\begin{aligned} I_3 + I_{2,2} & \leq -\frac{\nu}{2} \int_{Q_1} \left( |Dv|^2 + s^2 \right)^{(p-2)/2} |D(v_{x_i} - k)_-|^2 \phi^2 \, dx \, dt \\ & \quad + \frac{2L^2}{\nu} \int_{Q_1 \cap \{v_{x_i} < k\}} \left( |Dv|^2 + s^2 \right)^{p/2} |D\phi|^2 \, dx \, dt. \end{aligned}$$

Rearranging terms yields

$$\begin{aligned} & \frac{\nu}{2} \int_{Q_1} \left( |Dv|^2 + s^2 \right)^{(p-2)/2} |D(v_{x_i} - k)_-|^2 \phi^2 \, dx \, dt \\ & \quad + \lambda^{p-2} \int_{Q_1} (v_{x_i} - k)_-^2 (-\phi_t) \phi \, dx \, dt \\ & \leq \frac{2L^2}{\nu} \int_{Q_1 \cap \{v_{x_i} < k\}} \left( |Dv|^2 + s^2 \right)^{p/2} |D\phi|^2 \, dx \, dt \\ & \quad + 2L \int_{Q_1} \left( |Dv|^2 + s^2 \right)^{(p-1)/2} (v_{x_i} - k)_- \left( |\phi| |D^2\phi| + |D\phi|^2 \right) \, dx \, dt. \end{aligned} \tag{3.39}$$

Replace then  $\phi$  with  $\phi\theta$ , where  $\phi \in C^\infty(Q_1)$  vanishes on the parabolic boundary of  $Q_1$  and  $\theta \in C^\infty(\mathbb{R})$ , being nonincreasing and vanishing for  $t \geq 0$ , approximates the characteristic function of  $(-\infty, \tau)$ ,  $\tau \in (-1, 0)$ . In view of (3.33) and of the fact that we have chosen  $k \in (0, A\lambda)$ , we have bounds

$$\left( |Dv|^2 + s^2 \right)^{(p-2)/2} \geq (2A\lambda)^{p-2}, \quad \max\left\{ \left( |Dv|^2 + s^2 \right)^{1/2}, (v_{x_i} - k)_- \right\} \leq 2A\lambda,$$

and thus (3.39) yields, with  $\tau$  realizing (actually approaching) the supremum below,

$$\begin{aligned} & \lambda^{p-2} \| (v_{x_i} - k)_- \phi \|^2_{V^2(Q_1)} \\ & \leq 2\lambda^{p-2} \int_{Q_1} \left( |D(v_{x_i} - k)_-|^2 \phi^2 + (v_{x_i} - k)_-^2 |D\phi|^2 \right) \, dx \, dt \\ & \quad + \lambda^{p-2} \sup_{t \in (-1, 0)} \int_{B_1} (v_{x_i} - k)_-^2 \phi^2 \, dx \\ & \leq c\lambda^p \sup_{Q_1} \left( |\phi| |D^2\phi| + |D\phi|^2 + |\phi| |\phi_t| \right) |Q_1 \cap \{v_{x_i} < k\} \cap \{\phi > 0\}| \end{aligned} \tag{3.40}$$

for a constant  $c \equiv c(p, \nu, L, A)$ . This is the Caccioppoli-type inequality we were looking for.

*Step 3: Iteration.* Define first truncation levels

$$k_m := \frac{\lambda}{4} \left( 1 + \frac{1}{2^m} \right), \quad \text{so that} \quad \frac{\lambda}{4} \leq k_m \leq \frac{\lambda}{2},$$

for all integers  $m = 0, 1, \dots$ . We then consider the following non-negative cut-off functions:  $\phi_m \in C^\infty(Q_1)$ , where

$$Q^{(m)} := Q_{\varrho_m} \quad \varrho_m := \frac{1}{2} + \frac{1}{2^{m+1}}, \quad m \geq 0,$$

defined in such a way to vanish outside  $Q^{(m)}$  and smoothly on the parabolic boundary of  $Q^{(m)}$  and, moreover, satisfy the following conditions:

$$\begin{aligned} 0 \leq \phi_m \leq 1, \quad \phi_m \equiv 1 \quad \text{on } Q^{(m+1)} \\ [5pt] |D^2 \phi_m| + |D \phi_m|^2 + |(\phi_m)_t| \leq c(n) 4^m. \end{aligned} \tag{3.41}$$

Now, let us introduce, again for  $m \geq 0$ , the sets

$$A_m := Q^{(m)} \cap \{v_{x_i} < k_m\}.$$

Taking  $\phi \equiv \phi_m$  and  $k \equiv k_m$ , by (3.41) estimate (3.40) gives

$$\lambda^{p-2} \|(v_{x_i} - k_m) - \phi_m\|_{V^2(Q_1)}^2 \leq c 4^m \lambda^p |A_m|.$$

Using (3.26) and then the previous inequality, we have, as  $\phi_m \equiv 1$  on  $Q^{(m+1)}$ , that

$$\begin{aligned} \lambda^{p-2} (k_m - k_{m+1})^2 |A_{m+1}| &\leq \lambda^{p-2} \|(v_{x_i} - k_m) - \phi_m\|_{L^2(A_{m+1})}^2 \\ &\leq \lambda^{p-2} \|(v_{x_i} - k_m) - \phi_m\|_{L^2(Q^{(m)})}^2 \\ &\leq c \lambda^{p-2} \|(v_{x_i} - k_m) - \phi_m\|_{V^2(Q^{(m)})}^2 |A_m|^{2/(n+2)} \\ &\leq c 4^m \lambda^p |A_m|^{1+2/(n+2)} \end{aligned}$$

with a constant  $c = c(n, p, \nu, L, A)$ . In turn, observing that

$$k_m - k_{m+1} = 2^{-m-3} \lambda,$$

we arrive at

$$|A_{m+1}| \leq c 8^m |A_m|^{1+2/(n+2)},$$

for a constant  $c$  still depending only on  $n, p, \nu, L, A$ , and for every  $m \geq 1$ . At this stage by using a standard iteration lemma (see for instance [11, Chapter 1, Lemma 4.1]) we have that there exists a number  $\sigma \equiv \sigma(n, p, \nu, L, A) \in (0, 1)$  such that if

$$|Q_1 \cap \{v_{x_i} < \lambda/2\}| = |A_0| \leq \sigma |Q_1|,$$

then  $|A_m| \rightarrow 0$  and this implies (3.34) by the inclusion

$$Q_{1/2} \cap \{v_{x_i} \leq \lambda/4\} \subset \cap_m A_m.$$

The proof is complete. □

The previous result allows us to use the same machinery employed in [30, Section 3] for the case  $p \geq 2$ . The outcome is the following statement, whose proof is completely similar to the one for the case  $p \geq 2$ .

**Proposition 3.8.** *Assume that (3.27) is in force. Then there are numbers  $\sigma \in (0, 1/2)$ ,  $\beta \in (0, 1)$ , and  $c_d \geq 1$ , all depending only on  $n, p, v, L, A$ , such that (3.28) implies*

$$\left( \int_{Q_{\delta r}^\lambda} |Dw - (Dw)_{Q_{\delta r}^\lambda}|^q dx dt \right)^{1/q} \leq c_d \delta^\beta \left( \int_{Q_r^\lambda} |Dw - (Dw)_{Q_r^\lambda}|^q dx dt \right)^{1/q} \quad (3.42)$$

whenever  $\delta \in (0, 1)$  and  $q \geq 1$ .

Similarly, we have to the following result, that, proved in [30] in the case  $p \geq 2$ , extends to the case  $1 < p < 2$  with a completely similar proof.

**Proposition 3.9.** *Assume that*

$$\sup_{Q_r^\lambda} \|Dw\| \leq A\lambda \quad \text{and} \quad \xi\lambda \leq s \leq \xi_1 A\lambda, \quad \text{where } 0 < \xi \leq \xi_1. \quad (3.43)$$

Then

$$\left( \int_{Q_{\delta r}^\lambda} |Dw - (Dw)_{Q_{\delta r}^\lambda}|^q dx dt \right)^{1/q} \leq \tilde{c}_d \delta^{\beta_1} \left( \int_{Q_r^\lambda} |Dw - (Dw)_{Q_r^\lambda}|^q dx dt \right)^{1/q} \quad (3.44)$$

holds whenever  $\delta \in (0, 1)$  for constants  $\beta_1 \equiv \beta_1(n, p, v, L, A, \xi, \xi_1) \in (0, 1)$  and  $\tilde{c}_d \equiv \tilde{c}_d(n, p, v, L, A, \xi, \xi_1) \geq 1$ .

The next proposition analyzes the case ruled out by the previous Proposition 3.8. Usually such type of result is achieved via arguments based on logarithmic type inequalities (see for instance [11, 30]). Here we propose an alternative approach, using the weak Harnack inequality for supersolutions to linear parabolic equations, which turns out to be faster and probably more elegant. To this aim, we recall a classical result.

**Theorem 3.10 (Weak Harnack principle).** *Let  $z \in L^2(-1, 0; W^{1,2}(B_1))$  be a non-negative weak supersolution in  $Q_1 \equiv B_1 \times (-1, 0)$  to the linear parabolic equation*

$$z_t - \operatorname{div}(\tilde{B}(x, t)Dz) = 0,$$

where the matrix  $\tilde{B}$  has measurable coefficients and satisfies the conditions

$$v_1|\xi|^2 \leq \langle \tilde{B}(x, t)\xi, \xi \rangle \leq L_1|\xi|^2$$

whenever  $\xi \in \mathbb{R}^n$ , and where  $0 < v_1 \leq L_1$ . If

$$Q^* = B(0, 1 - \delta_1) \times (-\delta_2 + \delta_3 + \delta_4), -(\delta_2 + \delta_3))$$

and

$$Q^{**} = B(0, 1 - \delta_1) \times (-\delta_2, 0),$$

where  $\delta_i$  for  $i \in \{1, 2, 3, 4\}$  are positive parameters such that  $\delta_1 + \dots + \delta_4 < 1$ , then there exists a constant  $c$ , depending only on  $n, v_1, L_1$  and  $\{\delta_i\}$ , such that

$$\int_{Q^*} z \, dx \, dt \leq c \inf_{Q^{**}} z.$$

The previous theorem roughly tells that, in order to realize a suitable Harnack principle, we need a waiting time  $\delta_3 > 0$ . We refer to [36] for a proof of the previous theorem; see also [13, 26, 42] for more on Harnack inequalities in the nonlinear parabolic setting.

**Proposition 3.11.** *Assume that (3.27) holds and let  $\sigma \in (0, 1)$  be described in Proposition 3.7. Suppose that (3.28) is not satisfied for any  $i \in \{1, \dots, n\}$ . Then it is possible to find  $\eta \in (1/2, 1)$ , depending only on  $n, p, v, L, A$ , such that*

$$\|Dw\| \leq \eta A\lambda \quad \text{a.e. in } Q_{\sigma r}^\lambda. \tag{3.45}$$

*Proof.* Assume that for any  $i \in \{1, \dots, n\}$  the assumption (3.28) does not hold; therefore for every  $i \in \{1, \dots, n\}$  it happens that

$$|\{(x, t) \in Q_r^\lambda : w_{x_i}(x, t) \geq \lambda/2\}| < (1 - \sigma)|Q_r^\lambda| \tag{3.46}$$

and

$$|\{(x, t) \in Q_r^\lambda : w_{x_i}(x, t) \leq -\lambda/2\}| < (1 - \sigma)|Q_r^\lambda|. \tag{3.47}$$

Here  $\sigma \equiv \sigma(n, p, v, L, A)$  is the number determined in Proposition 3.7. We consider the occurrence of (3.46) for single  $i \in \{1, \dots, n\}$ ; we later describe how to argue for (3.47). Defining the scaled function

$$v(x, t) := \frac{w(rx, \lambda^{2-p}r^2t)}{r}, \quad (x, t) \in Q_1, \tag{3.48}$$

solving (3.32), inequality (3.46) takes the form

$$|\{(x, t) \in Q_1 : v_{x_i}(x, t) < \lambda/2\}| \geq \sigma|Q_1|. \tag{3.49}$$

In particular, by setting

$$Q^* = B(0, 1 - \sigma/(80n^4)) \times (-1 + \sigma/4, -\sigma/4) \subset Q_1 \tag{3.50}$$

we still have

$$|Q^* \cap \{v_{x_i} < \lambda/2\}| \geq \frac{\sigma}{4} |Q^*|. \tag{3.51}$$

Furthermore, it is quite standard to show that  $\tilde{v} := (v_{x_i} - \lambda/2)_+$  is a weak subsolution to the linear parabolic equation with measurable coefficients

$$z_t - \lambda^{2-p} \operatorname{div}(\partial a(Dv)Dz) = 0. \tag{3.52}$$

For the sake of the non-Finnish reader we briefly recall the argument in Remark 3.12 below. Next, we extend the coefficients of the equation of  $\tilde{v}$  outside of  $\{\tilde{v} > 0\}$  by setting

$$\tilde{B}(x, t) = \begin{cases} \lambda^{2-p} \partial a(Dv(x, t)) & \text{in } Q_1 \cap \{\tilde{v} > 0\} \\ v Id & \text{on } Q_1 \cap \{\tilde{v} = 0\}, \end{cases}$$

$Id$  being the identity matrix. By the upper bound (3.27) and the fact that  $|Dv| \geq \lambda/2$  in the support of  $\tilde{v}$ , we obtain

$$\frac{1}{c} \left( \frac{A\lambda}{\lambda} \right)^{p-2} |\xi|^2 \leq \left\langle \lambda^{2-p} \partial a(Dv)\xi, \xi \right\rangle \leq c \left( \frac{\lambda/2}{\lambda} \right)^{p-2} |\xi|^2$$

almost everywhere in  $Q_1 \cap \{\tilde{v} > 0\}$ , for a constant  $c \equiv c(n, p, v, L) \geq 1$ . Therefore the eigenvalues of measurable coefficients  $\tilde{B}$  belong to  $[A^{p-2}/c, c]$  and  $\tilde{v}$  is a weak subsolution to the linear parabolic equation

$$z_t - \operatorname{div}(\tilde{B}(x, t)Dz) = 0. \tag{3.53}$$

Moreover, it follows by (3.27) (look also at (3.33)) that  $z \equiv A\lambda - \lambda/2 - \tilde{v}$  is a non-negative weak supersolution to (3.53) in  $Q_1$ . Therefore we apply Theorem 3.10 with the choice  $Q^{**} \equiv B_{1/2} \times (-\sigma/8, 0)$  and  $Q^*$  as defined in (3.50); the outcome is

$$\frac{1}{c} \int_{Q^*} (A\lambda - \lambda/2 - \tilde{v}) \, dx \, dt \leq A\lambda - \lambda/2 - \sup_{B_{1/2} \times (-\sigma/8, 0)} \tilde{v}, \tag{3.54}$$

for a constant  $c \equiv c(n, p, v, L, A, \sigma) \equiv c(n, p, v, L, A)$ . By (3.51) we have the lower bound

$$\begin{aligned} \int_{Q^*} (A\lambda - \lambda/2 - \tilde{v}) \, dx \, dt &\geq \frac{1}{|Q^*|} \int_{Q^* \cap \{\tilde{v}=0\}} (A\lambda - \lambda/2) \, dx \, dt \\ &\geq \frac{\sigma}{4} (A\lambda - \lambda/2) \geq \frac{\sigma}{8} A\lambda, \end{aligned}$$

and combining the previous inequality with (3.54) yields

$$\sup_{B_{\sigma/2} \times (-\sigma/4)^2, 0)} v_{x_i} \leq \max \left\{ \frac{1}{2}, \left( 1 - \frac{\sigma}{8c} \right) \right\} A\lambda =: \eta A\lambda.$$

Finally, interchanging  $v_{x_i}$  with  $-v_{x_i}$  in the above reasoning - keep in mind the change of variable made at the beginning of the proof of Proposition 3.7 - and using (3.47), we actually obtain the same bound instead for the supremum of  $-v_{x_i}$ . Since the above holds for arbitrary  $i \in \{1, \dots, n\}$ , inequality (3.45) follows.  $\square$

**Remark 3.12 (Proof that  $\tilde{v}$  is a weak subsolution of (3.52)).** We shall argue modulo a standard time regularization via Steklov averages. We shall test the weak formulation with  $\eta_\varepsilon := \varepsilon^{-1} \min(\varepsilon, \tilde{v})\phi$ ,  $\varepsilon > 0$ , where  $\phi \in C_c^\infty(Q_1)$  is non-negative. Firstly,

$$(v_{x_i})_t \varepsilon^{-1} \min(\varepsilon, \tilde{v}) = \partial_t \int_{\lambda/2}^{v_{x_i}} \varepsilon^{-1} \min(\varepsilon, (y - \lambda/2)_+) dy$$

holds for all  $\varepsilon > 0$  and therefore integration by parts yields

$$\begin{aligned} \int_{Q_1} (v_{x_i})_t \eta_\varepsilon dx dt &= - \int_{Q_1} \int_{\lambda/2}^{v_{x_i}} \varepsilon^{-1} \min(\varepsilon, (y - \lambda/2)_+) dy \phi_t dx dt \\ &\rightarrow - \int_{Q_1} \tilde{v} \phi_t dx dt \end{aligned}$$

as  $\varepsilon \rightarrow 0$ . Secondly, since

$$D\eta_\varepsilon = \varepsilon^{-1} \min(\varepsilon, \tilde{v}) D\phi + \varepsilon^{-1} \phi \chi_{\{0 < \tilde{v} < \varepsilon\}} Dv_{x_i},$$

we have that

$$\begin{aligned} \langle \partial a(Dv) Dv_{x_i}, D\eta_\varepsilon \rangle &\geq \varepsilon^{-1} \min(\varepsilon, \tilde{v}) \langle \partial a(Dv) Dv_{x_i}, D\phi \rangle \\ &\rightarrow \langle \partial a(Dv) D\tilde{v}, D\phi \rangle \end{aligned}$$

almost everywhere as  $\varepsilon \rightarrow 0$ . Collecting calculations, we have showed

$$\begin{aligned} 0 &= \lim_{\varepsilon \downarrow 0} \int_{Q_1} \left( \lambda^{p-2} (v_{x_i})_t \eta_\varepsilon + \langle \partial a(Dv) Dv_{x_i}, D\eta_\varepsilon \rangle \right) dx dt \\ &\geq \int_{Q_1} \left( -\lambda^{p-2} \tilde{v} \phi_t + \langle \partial a(Dv) D\tilde{v}, D\phi \rangle \right) dx dt. \end{aligned}$$

Since  $\phi$  is an arbitrary non-negative test function, it follows that  $\tilde{v}$  is a weak subsolution.



### 3.2. Proof of Theorem 3.1

We shall give here the proof in the case  $s > 0$ , while the case  $s = 0$  follows via approximation, as explained in Section 3.3 below. Note that anyway we shall never use the fact that  $s > 0$  but when appealing to Propositions 3.8-3.11. Indeed, assuming that the statements of these last two results remain true for  $s = 0$ , then the rest of the proof below remains unchanged. The proof goes now in five steps.

*Step 1: Stopping time for the singular iteration and the choice of  $\xi$ .* With  $\eta \equiv \eta(n, p, \nu, L, A) \in (0, 1)$  being defined in Proposition 3.11, we set

$$\eta_1 := \frac{1 + \eta}{2}, \quad \text{so that } \eta < \eta_1 < 1 \quad \text{and} \quad \eta_1 - \eta = \frac{1 - \eta}{2}. \quad (3.55)$$

Obviously  $\eta_1 \equiv \eta_1(n, p, \nu, L, A) \in (0, 1)$ . We define  $m \in \mathbb{N}$  as the smallest integer such that

$$\eta_1^m A \lambda < \frac{\lambda}{2B}. \quad (3.56)$$

Observe that this determines  $m \geq 1$  as a function of the parameters  $n, p, \nu, L, A, B$ . Define  $\xi$  accordingly as

$$\xi := \min\{1/8, (\eta_1 - \eta)\eta_1^{2m} A\}, \quad (3.57)$$

which then also depends only on parameters  $n, p, \nu, L, A, B$ . The choices in (3.56) and (3.57) fix the number  $\xi$  in the statement of Theorem 3.1.

*Step 2: The first nonsingular case.* In this case we assume that  $\xi \lambda \leq s$ . The equation becomes immediately nonsingular and Proposition 3.9 then implies that

$$\left( \int_{Q_{\delta r}^\lambda} |Dw - (Dw)_{Q_{\delta r}^\lambda}|^q dx dt \right)^{1/q} \leq \tilde{c}_d \delta^{\beta_1} \left( \int_{Q_r^\lambda} |Dw - (Dw)_{Q_r^\lambda}|^q dx dt \right)^{1/q}$$

for all  $\delta \in (0, 1)$  and constants  $\tilde{c}_d$  and  $\beta_1$  depending only on  $n, p, \nu, L, A, B$ , since  $\xi$  above depends only on such parameters. Thus we will eventually obtain (3.5) with  $\delta \equiv \delta_\gamma$  whenever

$$\delta_\gamma \equiv \delta \leq \left( \frac{\gamma}{\tilde{c}_d} \right)^{1/\beta_1} \equiv \frac{\gamma^{1/\alpha}}{c(A, B)}. \quad (3.58)$$

For this reason, we shall from now on argue under the assumption that

$$s \leq (\eta_1 - \eta)\eta_1^{2m} A \lambda = \xi \lambda \quad (3.59)$$

holds for  $m$  defined as in (3.56).

*Step 3: The Singular Iteration.* Given a cylinder  $Q_r^\lambda$ , where the intrinsic inequality (3.27) holds, by Propositions 3.8 and 3.11 we then have two possibilities:

- **The Nonsingular Alternative.** This means that we can apply Proposition 3.8 and therefore we have that

$$\left( \int_{Q_{\delta r}^\lambda} |Dw - (Dw)_{Q_{\delta r}^\lambda}|^q dx dt \right)^{1/q} \leq c_d \delta^\beta \left( \int_{Q_r^\lambda} |Dw - (Dw)_{Q_r^\lambda}|^q dx dt \right)^{1/q} \quad (3.60)$$

for every  $\delta \in (0, 1)$ , where the constants  $\beta \equiv \beta(n, p, \nu, L, A) \in (0, 1)$  and  $c_d \equiv c_d(n, p, \nu, L, A) \geq 1$  are those defined in Proposition 3.8.

- **The Singular Alternative.** In this case we can instead apply Proposition 3.11 that in turn yields

$$\sup_{Q_{\sigma_1 r}^\lambda} \|Dw\| \leq \eta A \lambda,$$

where  $\eta \equiv \eta(n, p, \nu, L, A) \in (0, 1)$  and  $\sigma_1 \equiv \sigma_1(n, p, \nu, L, A) \in (0, 1)$ .

The previous alternatives can be now combined in order to build a basic iteration scheme, that we call *the Singular Iteration*. Let  $\eta_1$  be as in (3.55). We define the sequences

$$\begin{cases} \lambda_{i+1} := \eta_1 \lambda_i \\ \lambda_0 := \lambda, \end{cases} \quad \begin{cases} R_{i+1} := \sigma_1 R_i \\ R_0 := r. \end{cases}$$

With such a choice the following inclusions trivially hold:

$$Q_{R_{i+1}}^{\lambda_{i+1}} \subset Q_{\sigma_1 R_i}^{\lambda_i} \subset Q_{R_i}^{\lambda_i} \subset Q_r^\lambda \quad \forall i \in \mathbb{N}. \quad (3.61)$$

From now on we shall also set

$$Q_i := Q_{R_i}^{\lambda_i}.$$

Here, as in the following, all the cylinders share the same vertex. We now proceed building the iteration scheme by (finite) induction: to this aim, let us assume that the Singular Alternative holds in the cylinders  $Q_{R_{i-1}}^{\lambda_{i-1}}$  for  $i \in \{1, \dots, j\}$  for some integer  $j$  such that  $1 \leq j \leq 2m$ . Therefore we have that

$$s + \sup_{Q_{R_{j-1}}^{\lambda_{j-1}}} \|Dw\| \leq A \lambda_{j-1} \quad \text{and} \quad \sup_{Q_{\sigma_1 R_{j-1}}^{\lambda_{j-1}}} \|Dw\| \leq \eta A \lambda_{j-1}$$

hold. It follows that for all such  $j$  we have

$$\begin{aligned} s + \sup_{Q_{R_j}^{\lambda_j}} \|Dw\| &\leq (\eta_1 - \eta) \eta_1^{2m} A \lambda + \sup_{Q_{\sigma_1 R_{j-1}}^{\lambda_{j-1}}} \|Dw\| \\ &\leq (\eta_1 - \eta) A \lambda_{j-1} + \eta A \lambda_{j-1} = A \lambda_j. \end{aligned} \quad (3.62)$$

In particular, the needed upper bound (3.27) is verified with  $Q_r^\lambda \equiv Q_{R_j}^{\lambda_j}$ , and we can proceed inductively to verify whether the Singular Alternative or the Nonsingular Alternative occurs in  $Q_{R_j}^{\lambda_j}$ , provided  $j \leq 2m$ .

*Step 4: The second nonsingular case.* Now, consider all such numbers  $\delta_\gamma$  such that

$$\delta_\gamma \leq \underline{\delta} := \left( \eta_1^{(2-p)/2} \sigma_1 \right)^{m+1}, \tag{3.63}$$

so that

$$Q_{\delta_\gamma r}^\lambda \subset Q_{\underline{\delta} r}^\lambda \subset Q_{R_{m+1}}^{\lambda_{m+1}} \tag{3.64}$$

holds. Assume, according to the first inequality in (3.4) and to (3.59) (and towards the final determination of  $\delta_\gamma$  in the statement), that

$$s \leq \xi \lambda, \quad \frac{\lambda}{B} \leq \sup_{Q_{\underline{\delta} r}^\lambda} \|Dw\| \tag{3.65}$$

hold. This in turn implies

$$A\lambda_m \equiv \eta_1^m A\lambda < \frac{\lambda}{2B} \leq \sup_{Q_{\underline{\delta} r}^\lambda} \|Dw\| \leq \sup_{Q_{R_{m+1}}^{\lambda_{m+1}}} \|Dw\| \tag{3.66}$$

by (3.56) and (3.65). Let us now define

$$\tilde{m} := \min \left\{ k \in \mathbb{N} : \text{The Singular Alternative does not occur on } Q_{R_k}^{\lambda_k} \right\}. \tag{3.67}$$

Observe that by definition this means that the Singular Iteration can be performed  $\tilde{m}$  times, but that it cannot be applied on the cylinder  $Q_{R_{\tilde{m}}}^{\lambda_{\tilde{m}}}$ . We have

$$\tilde{m} \leq m. \tag{3.68}$$

Indeed, by assuming by contradiction that  $m < \tilde{m}$ , then the Singular Alternative would hold at  $m$  and therefore

$$\sup_{Q_{R_{m+1}}^{\lambda_{m+1}}} \|Dw\| \leq \eta A\lambda_m < A\lambda_m,$$

contradicting (3.66). Thus (3.68) holds provided (3.65) is assumed, and from now on, we shall choose, when providing  $\delta_\gamma$ , a number suitably smaller than  $\underline{\delta}$ , where  $\underline{\delta}$  has been introduced in (3.63). *Indeed, notice that if the above reasoning still holds upon decreasing the value of  $\underline{\delta}$ .* Moreover, as in the case of (3.62), we have

$$s + \sup_{Q_{R_j}^{\lambda_j}} \|Dw\| \leq A\lambda_j, \quad \text{for } j \in \{0, 1, \dots, \tilde{m}\}. \tag{3.69}$$

Next we shall apply the result of the Nonsingular Alternative. First, since the Singular Alternative could not be applied in  $Q_{\tilde{m}}$  and (3.69) in particular holds for  $j = \tilde{m}$ , then the Nonsingular Alternative must occur in  $Q_{\tilde{m}}$  and we have

$$\begin{aligned} & \left( \int_{\delta Q_{\tilde{m}}} |Dw - (Dw)_{\delta Q_{\tilde{m}}}|^q dx dt \right)^{1/q} \\ & \leq c_d \delta^\beta \left( \int_{Q_{\tilde{m}}} |Dw - (Dw)_{Q_{\tilde{m}}}|^q dx dt \right)^{1/q} \end{aligned} \tag{3.70}$$

for all  $\delta \in (0, 1]$ . Second, by (3.68), we have

$$\begin{aligned} & \left( \int_{Q_{\tilde{m}}} |Dw - (Dw)_{Q_{\tilde{m}}}|^q dx dt \right)^{1/q} \\ & \leq 2 \left( \frac{|Q_0|}{|Q_{\tilde{m}}|} \right)^{1/q} \left( \int_{Q_0} |Dw - (Dw)_{Q_0}|^q dx dt \right)^{1/q} \\ & \leq 2 \left( \frac{|Q_0|}{|Q_m|} \right)^{1/q} \left( \int_{Q_0} |Dw - (Dw)_{Q_0}|^q dx dt \right)^{1/q} \\ & \leq 2\sigma_1^{-m(n+2)} \eta_1^{-m(2-p)} \left( \int_{Q_r^\lambda} |Dw - (Dw)_{Q_r^\lambda}|^q dx dt \right)^{1/q}. \end{aligned} \tag{3.71}$$

Third, since  $\delta \underline{\delta} Q_r^\lambda \subset \delta Q_m \subset \delta Q_{\tilde{m}}$  - recall the definition of  $\underline{\delta}$  in (3.63) and again (3.68) - we have

$$\begin{aligned} & \left( \int_{\delta \underline{\delta} Q_r^\lambda} |Dw - (Dw)_{\delta \underline{\delta} Q_r^\lambda}|^q dx dt \right)^{1/q} \\ & \leq 2 \left( \frac{|\delta Q_{\tilde{m}}|}{|\delta \underline{\delta} Q_r^\lambda|} \right)^{1/q} \left( \int_{\delta Q_{\tilde{m}}} |Dw - (Dw)_{\delta Q_{\tilde{m}}}|^q dx dt \right)^{1/q} \\ & \leq 2 \underline{\delta}^{-(n+2)} \left( \int_{\delta Q_{\tilde{m}}} |Dw - (Dw)_{\delta Q_{\tilde{m}}}|^q dx dt \right)^{1/q} \end{aligned} \tag{3.72}$$

for all  $\delta \in (0, 1]$ . Combining (3.70)-(3.72) we arrive at

$$\begin{aligned} & \left( \int_{\delta \underline{\delta} Q_r^\lambda} |Dw - (Dw)_{\delta \underline{\delta} Q_r^\lambda}|^q dx dt \right)^{1/q} \\ & \leq 4c_d \left[ \underline{\delta}^{-(n+2)} \sigma_1^{-m(n+2)} \eta_1^{-m(2-p)} \right] \delta^\beta \left( \int_{Q_r^\lambda} |Dw - (Dw)_{Q_r^\lambda}|^q dx dt \right)^{1/q} \end{aligned} \tag{3.73}$$

for all  $\delta \in (0, 1]$ . Taking thus any

$$\delta \leq \left( \frac{\underline{\delta}^{(n+2)} \sigma_1^{m(n+2)} \eta_1^{m(2-p)} \gamma}{4c_d} \right)^{1/\beta}, \quad \delta_\gamma := \delta \underline{\delta}, \tag{3.74}$$

we conclude with (3.5) in this case.

*Step 5: Determining the number  $\delta_\gamma$ .* By looking at conditions (3.58) and (3.74), we are led to define

$$\delta_\gamma := \left( \eta_1^{(2-p)/2} \sigma_1 \right)^{m+1} \min \left\{ \left( \frac{\delta^{(n+2)} \sigma_1^{m(n+2)} \eta_1^{m(2-p)} \gamma}{4c_d} \right)^{1/\beta}, \left( \frac{\gamma}{\tilde{c}_d} \right)^{1/\beta_1} \right\} \tag{3.75}$$

and notice that  $\delta_\gamma$  depends indeed only on  $n, p, v, L, A, B, \gamma$ , as  $\sigma_1, \eta_1$  and  $m$  only depend on such parameters. The number  $\delta_\gamma$  just defined in (3.75) is the one we are looking for and it does not depend on the solution  $w$  (neither on the vector field  $a(\cdot)$ ) since it works both in Step 2 and Step 3. Moreover, it has the form required in (3.6). The proof is now complete when  $s > 0$  (although all the previous consideration formally holds for  $s = 0$  provided one can use the two Alternatives in Step 3). The case  $s = 0$  now follows from the case  $s > 0$  via the approximation argument described in the next Section 3.3.

*Proof of Theorem 3.5.* The proof closely follows of the one found for Theorem 3.1, therefore, we shall indicate the necessary modifications step by step.

*Step 1.* Here the crucial remark is that the number  $\eta_1$  does not depend on  $B$ . We notice that  $m$  as defined in (3.56) satisfies

$$m \approx \frac{\log(4AB)}{-\log \eta_1} =: \tilde{c}_1(A) \log(AB) = \log(AB)^{c_1(A)}. \tag{3.76}$$

for suitable constant  $c_1(A)$ , which is non-decreasing in  $A$ , and also depends on  $n, p, v, L$ .

*Step 2.* Since we are now assuming (3.22) rather than (3.4), the content of Step 2 becomes irrelevant. A different argumentation will occur later, when considering the stopping time of the Singular Iteration. In particular we shall not assume (3.59).

*Step 3.* In fact, as now (3.59) is not any longer in force, there is an additional reason for the Singular Iteration to stop. In other words, even before checking the occurrence of the Singular Alternative at step, let's say,  $j$ , we have to verify that the starting condition

$$s + \sup_{Q_{R_j}^{\lambda_j}} ||Dw|| \leq A\lambda_j$$

holds. Observe that were (3.59) in force, this would no longer be necessary, as shown in (3.62). Therefore, considering the number  $\tilde{m}$  defined in (3.67), the iteration may stop at  $\tilde{m}$  either because the Nonsingular Alternative occurs or simply because the initial condition

$$s + \sup_{Q_{R_{\tilde{m}}}^{\lambda_{\tilde{m}}}} ||Dw|| \leq A\lambda_{\tilde{m}} \tag{3.77}$$

does not hold. Treating the two different cases will be done in the next step.

*Step 4.* Exactly as for Theorem 3.1 we have (3.68) and we proceed as done there. In the case the Singular Alternative stops at  $\tilde{m}$  as the Nonsingular Alternative holds (together with (3.77)) we proceed as in Theorem 3.1. In particular, we take  $\delta_\gamma$  as indicated in (3.74). In the other case the Singular Iteration stops at step  $\tilde{m}$  as (3.77) fails to hold, that is

$$s + \sup_{Q_{R_{\tilde{m}}}^{\lambda_{\tilde{m}}}} \|Dw\| > A\lambda_{\tilde{m}} = \eta_1^{\tilde{m}} A\lambda. \tag{3.78}$$

This implies that  $\tilde{m} \geq 1$  as on the other hand the second inequality in (3.22) is assumed. Note also that since the Singular Alternative holds at level  $\tilde{m} - 1$ , we have

$$s + \sup_{Q_{R_{\tilde{m}-1}}^{\lambda_{\tilde{m}-1}}} \|Dw\| \leq A\lambda_{\tilde{m}-1} = \frac{A}{\eta_1} \lambda_{\tilde{m}} \tag{3.79}$$

and

$$\sup_{Q_{R_{\tilde{m}}}^{\lambda_{\tilde{m}}}} \|Dw\| \leq \eta A\lambda_{\tilde{m}-1} = \eta \eta_1^{\tilde{m}-1} A\lambda. \tag{3.80}$$

Comparing (3.78) and (3.80) yields

$$s > \eta_1^{\tilde{m}} A\lambda - \sup_{Q_{R_{\tilde{m}}}^{\lambda_{\tilde{m}}}} \|Dw\| \geq \eta_1^{\tilde{m}-1} (\eta_1 - \eta) A\lambda = (\eta_1 - \eta) A\lambda_{\tilde{m}-1}. \tag{3.81}$$

In conclusion, by (3.79) and (3.81), we find

$$\sup_{Q_{R_{\tilde{m}-1}}^{\lambda_{\tilde{m}-1}}} \|Dw\| \leq A\lambda_{\tilde{m}-1} \quad \text{and} \quad (\eta_1 - \eta) A\lambda_{\tilde{m}} \leq s \leq \frac{A}{\eta_1} \lambda_{\tilde{m}}$$

and we can therefore apply Proposition 3.9 with the choice

$$Q_r^\lambda \equiv Q_{R_{\tilde{m}}}^{\lambda_{\tilde{m}}}, \quad \xi \equiv (\eta_1 - \eta) A, \quad \xi_1 \equiv 1/\eta_1.$$

The outcome is that

$$\left( \int_{\delta Q_{\tilde{m}}} |Dw - (Dw)_{\delta Q_{\tilde{m}}}|^q dx dt \right)^{1/q} \leq \tilde{c}_d \delta^{\beta_1} \left( \int_{Q_{\tilde{m}}} |Dw - (Dw)_{Q_{\tilde{m}}}|^q dx dt \right)^{1/q}$$

holds whenever  $\delta \in (0, 1)$ , where  $\tilde{c}_d \geq 1$  and  $\beta_1 \in (0, 1)$  are constants depending only on  $n, N, p, v, L$  and  $A$ , but *not* on  $B$  and  $q$ . The last estimate is the analog of (3.70) and from this point on we may argue exactly as after there, to obtain (3.73) with  $c_d$  replaced by  $\tilde{c}_d$ , and  $\beta$  replaced by  $\beta_1$ . Therefore, by choosing

$$\delta \leq \left( \frac{\delta^{(n+2)} \sigma_1^{m(n+2)} \eta_1^{m(2-p)} \gamma}{4\tilde{c}_d} \right)^{1/\beta}, \quad \delta_\gamma := \delta \underline{\delta}, \tag{3.82}$$

we conclude with (3.5).

Step 5. In view of (3.74) and (3.82) we finally determine  $\delta$  as follows

$$\delta_\gamma := \left( \eta_1^{(2-p)/2} \sigma_1 \right)^{m+1} \left( \frac{\delta^{(n+2)} \sigma_1^{m(n+2)} \eta_1^{m(2-p)} \gamma}{4(c_d + \tilde{c}_d)} \right)^{1/\min\{\beta, \beta_1\}}. \tag{3.83}$$

Notice here that the only parameter containing a dependence on  $B$  is  $m$ . Using the dependence described in (3.76) and in (3.83) we obtain (3.24), for a suitable choice of the constant  $c(A)$ .  $\square$

### 3.3. The approximation scheme

Here we briefly discuss the approximation method needed to reduce the proof of Theorems 3.1-3.5 to the nonsingular case  $s > 0$ ; when switching to the case  $s = 0$  the constants will increase, in a universal way. We shall of course confine ourselves to the case of Theorem 3.1. We start mollifying the vector field  $a(\cdot)$  as follows. Let  $\varepsilon > 0$  (actually denoting a sequence converging to zero) and let  $\theta_\varepsilon \in C_c^\infty(B_\varepsilon(0))$  be a standard mollifier with  $B_\varepsilon(0) \subset \mathbb{R}^n$ , such that  $\int_{\mathbb{R}^n} \theta_\varepsilon(z) dz = 1$ . Define the regularized vector fields  $a_\varepsilon : \mathbb{R}^n \rightarrow \mathbb{R}^n$  as

$$a_\varepsilon(z) := \int_{\mathbb{R}^n} \theta_\varepsilon(z-\xi) a(\xi) d\xi.$$

We then define  $w_\varepsilon$  as the unique solution to the following Cauchy-Dirichlet problem:

$$\begin{cases} (w_\varepsilon)_t - \operatorname{div} a_\varepsilon(Dw_\varepsilon) = 0 & \text{in } Q_r^\lambda \\ w_\varepsilon = w & \text{on } \partial_{\text{par}} Q_r^\lambda. \end{cases}$$

Exactly as for instance in [1] or [27], it follows that  $a_\varepsilon(\cdot)$  satisfies (1.2) with new constants  $\nu, L$ , and with  $s$  (here assumed to be equal to zero) replaced now by  $s_\varepsilon = \varepsilon$ ; without loss of generality we shall consider  $\varepsilon$  small enough to have  $s_\varepsilon \leq \lambda$ . Again as in [1,27] it follows that up to not relabeled subsequences (*i.e.* we still keep the notation  $\varepsilon$ )

$$\begin{cases} Dw_\varepsilon \rightarrow Dw \quad \text{strongly in } L^p \text{ and a.e.} \\ \int_{Q_r^\lambda} (|Dw_\varepsilon| + s_\varepsilon)^p dx dt \leq c_a^p \int_{Q_r^\lambda} (|Dw| + s_\varepsilon)^p dx dt, \end{cases} \tag{3.84}$$

where  $c_a$  depends only on  $n, p, \nu, L$ . Before going on let us recall a basic result (see [11, Chapter 8, Theorem 5.1']) asserting

$$\int_{Q_r^\lambda} (|Dw_\varepsilon| + s_\varepsilon)^p dx dt \leq c^p \lambda^p \implies \sup_{Q_{r/2}^\lambda} \|Dw_\varepsilon\| \leq \tilde{c}_a c \lambda \tag{3.85}$$

for a new constant  $\tilde{c}_a \equiv \tilde{c}_a(n, p, \nu, L)$ . Now we assume that the Theorem 3.1 holds for the case  $s > 0$ , and fix  $A, B, \gamma$  in the “ $s = 0$ ”-version of Theorem 3.1 we want to prove. Take the choice

$$\gamma \rightarrow 2^{-(n+2)}\gamma =: \tilde{\gamma}, \quad A \rightarrow 2\tilde{c}_a c_a A =: \tilde{A}, \quad B \rightarrow 2B \tag{3.86}$$

and determine the number  $\delta_{\tilde{\gamma}}(\tilde{A})$  in Theorem 3.1 for the case  $s > 0$  (remark that  $\delta_{\tilde{\gamma}}(\tilde{A})$  also depends on  $n, p, \nu, L$  via the new constants in (1.2) for  $a_\varepsilon(\cdot)$ ; it of course also depends on  $B$ ). We claim that now Theorem 3.1 for the case  $s = 0$  holds with the choice

$$\delta_\gamma(A) := \delta_{\tilde{\gamma}}(\tilde{A})/2, \tag{3.87}$$

and indeed the assumptions in question are now

$$\frac{\lambda}{B} \leq \sup_{Q_{\delta_{\tilde{\gamma}}r/2}^\lambda} \|Dw\| \leq \sup_{Q_r^\lambda} \|Dw\| \leq A\lambda. \tag{3.88}$$

Let us observe that for yet another not relabeled subsequence we may assume that

$$\frac{\lambda}{2B} \leq \sup_{Q_{\delta_{\tilde{\gamma}}r/2}^\lambda} \|Dw_\varepsilon\|.$$

Indeed, were this not the case, by using the convergence in (3.84) we would immediately contradict the first inequality in (3.88). On the other hand, thanks to (3.84)-(3.85), it follows that

$$s_\varepsilon + \sup_{Q_{r/2}^\lambda} \|Dw_\varepsilon\| \leq \tilde{A}\lambda. \tag{3.89}$$

We can therefore apply Theorem 3.1 in the case  $s \equiv s_\varepsilon > 0$ , thereby obtaining

$$E_q(Dw_\varepsilon, (\delta_{\tilde{\gamma}}/2)Q_r^\lambda) \leq 2^{-(n+2)}\gamma E_q(Dw_\varepsilon, Q_{r/2}^\lambda).$$

Letting  $\varepsilon \rightarrow 0$ , (3.84) and (3.89) yield

$$E_q(Dw, (\delta_{\tilde{\gamma}}/2)Q_r^\lambda) \leq 2^{-(n+2)}\gamma E_q(Dw, Q_{r/2}^\lambda) \leq \gamma E_q(Dw, Q_r^\lambda)$$

and this proves Theorem 3.1 in the case  $s = 0$ , with the choice in (3.87). Notice that the choices in (3.86) and (3.87) do keep the structure of the number  $\delta_\gamma$  with respect to  $\gamma$  and  $B$  described in (3.6) and (3.24).



#### 4. Proof of the intrinsic potential estimate and consequences

In this section we give the proof of Theorem 1.2, which in turn implies Corollaries 1.2 and 1.3. First, in Section 4.1 we propose a few comparison estimates necessary to implement the iteration procedure that will lead, in Section 4.2 below, to the proof of the intrinsic potential estimate (1.16).

Let us first clarify that, with  $\beta > 0$  being fixed, generalized solutions as in (1.16) always exist provided the Riesz potential is finite:

$$\mathbf{I}_1^\mu(x_0, t_0; 2r) := \int_0^{2r} \left[ \frac{|\mu|(Q_\varrho(x_0, t_0))}{\varrho^{N-1}} \right] \frac{d\varrho}{\varrho} < \infty. \quad (4.1)$$

**Remark 4.1 (Finiteness of potentials implies generalized roots exist).** Let us assume that (4.1) holds with  $Q_\varrho(x_0, t_0)$  denoting a standard parabolic cylinder. Changing variable (*i.e.*  $\varrho \rightarrow \lambda^{(2-p)/2}\varrho$ ) in (1.16), we obtain

$$\lambda = c\beta + c\lambda^{(n+1)(2-p)/2} \int_0^{2r} \left[ \frac{|\mu|(Q_{\lambda^{(p-2)/2}\varrho}^\lambda(x_0, t_0))}{\varrho^{N-1}} \right] \frac{d\varrho}{\varrho}$$

and accordingly define the function

$$h(\lambda) := \lambda - c\beta - c\lambda^{(n+1)(2-p)/2} \int_0^{2r} \left[ \frac{|\mu|(Q_{\lambda^{(p-2)/2}\varrho}^\lambda(x_0, t_0))}{\varrho^{N-1}} \right] \frac{d\varrho}{\varrho}$$

defined for  $\lambda > 0$ . Observe that  $h(\cdot)$  is continuous function in  $(0, \infty)$  and, moreover,  $h(\lambda) < 0$  for  $0 < \lambda < c\beta$ . On the other hand, using that

$$|\mu|(Q_{\lambda^{(p-2)/2}\varrho}^\lambda) \leq |\mu|(Q_\varrho), \quad \text{for } \lambda \geq 1,$$

we have

$$\lim_{\lambda \rightarrow \infty} h(\lambda) \geq \lim_{\lambda \rightarrow \infty} \left[ \lambda - c\beta - c\lambda^{(n+1)(2-p)/2} \mathbf{I}_1^\mu(x_0, t_0; 2r) \right] = \infty,$$

since  $(n+1)(2-p)/2 < 1$  for  $p > 2 - 1/(n+1)$ . Therefore there must exist  $\lambda \in [c\beta, \infty)$  solving  $h(\lambda) = 0$ , that is, a solution to (1.16).

#### 4.1. Comparison results

In the rest of the section, until the proof of Theorem 1.1, we consider in the fixed parabolic cylinder  $Q_\varrho^\lambda(x_0, t_0) \subseteq \Omega_T$ ,  $\varrho, \lambda > 0$ , the unique solution

$$\begin{aligned} w \in C^0([t_0 - \lambda^{2-p}\varrho^2, t_0]; L^2(B(x_0, \varrho))) \\ \cap L^p(t_0 - \lambda^{2-p}\varrho^2, t_0; W^{1,p}(B(x_0, \varrho))) \end{aligned} \quad (4.2)$$

to the following Cauchy-Dirichlet problem:

$$\begin{cases} w_t - \operatorname{div} a(Dw) = 0 & \text{in } Q_\rho^\lambda \\ w = u & \text{on } \partial_{\text{par}} Q_\rho^\lambda. \end{cases} \tag{4.3}$$

The following result is taken from [11, Chapter 8, Theorem 5.2], and in the form suitable for general equations can be also retrieved from [27].

**Theorem 4.2.** *With  $w$  being defined in (4.3), there exists a constant  $c_3 \geq 1$ , depending only on  $n, p, v, L$ , but otherwise independent of  $s$ , of the solution  $w$  considered and of the vector field  $a(\cdot)$ , such that*

$$\sup_{\frac{1}{2}Q_\rho^\lambda} \|Dw\| \leq \frac{c_3}{2}(\lambda + s) + \frac{c_3}{2} \lambda^{\frac{n(p-2)}{n(p-2)+2}} \left( \int_{Q_\rho^\lambda} (|Dw| + s) \, dx \, dt \right)^{\frac{2}{n(p-2)+2}}.$$

Then we establish a comparison estimate between  $u$  and  $w$  in the next result:

**Lemma 4.3.** *Let  $u$  be as in Theorem 1.2 and  $w$  as in (4.3); then there exists a constant  $c$  depending only on  $n, p, v, L$  such that the following inequality holds:*

$$\begin{aligned} \int_{Q_\rho^\lambda} |Du - Dw| \, dx \, dt &\leq c \left[ \frac{|\mu|(Q_\rho^\lambda)}{|Q_\rho^\lambda|^{\frac{n+1}{n+2}}} \right]^{\frac{n+2}{(p-1)n+p}} \\ &+ c \left[ \frac{|\mu|(Q_\rho^\lambda)}{|Q_\rho^\lambda|^{\frac{n+1}{n+2}}} \right] \left( \int_{Q_\rho^\lambda} (|Du| + s) \, dx \, dt \right)^{(2-p)\frac{n+1}{n+2}}. \end{aligned} \tag{4.4}$$

*Proof.*

*Step 1: Preliminary estimates.* We go back to the proof of Lemma 4.1, Step 1, in [30] and obtain the following preliminary estimates, which in fact hold also for the whole  $p > 1$ :

$$\sup_{\tau} \int_{Q_\tau} |u - w| \, dx \leq |\mu|(Q) \tag{4.5}$$

and

$$\int_Q \frac{\langle a(Du) - a(Dw), Du - Dw \rangle}{(\alpha + |u - w|)^\xi} \, dx \, dt \leq c \frac{\alpha^{1-\xi}}{\xi - 1} |\mu|(Q) \tag{4.6}$$

for  $\alpha > 0$  and  $\xi > 1$ , where  $c \equiv c(n, p, v) \geq 1$ .

*Step 2: Comparison estimates.* Fix now

$$\xi = \frac{n + 1}{n}(p - 1), \tag{4.7}$$

so that

$$\xi > 1 \quad \Longleftrightarrow \quad p > 2 - \frac{1}{n+1}.$$

Define

$$\alpha = \left( \int_Q |u-w|^{\frac{n+1}{n}} dx dt \right)^{\frac{n}{n+1}}$$

and assume that  $\alpha > 0$ , for if it is not, then  $u = w$  and (4.4) follows. The parabolic Sobolev inequality (see for example [11, Chapter 1, Proposition 3.1]) gives

$$\alpha \leq c(n) \left[ \int_Q |Du - Dw| dx dt \left( \sup_{\tau} \int_{Q_{\tau}} |u-w| dx \right)^{\frac{1}{n}} \right]^{\frac{n}{n+1}},$$

and thus by (4.5) that

$$\alpha \leq c|\mu|(Q)^{\frac{1}{n+1}} \left( \int_Q |Du - Dw| dx dt \right)^{\frac{n}{n+1}}. \quad (4.8)$$

Next (see for instance [39]) we recall some of the basic properties of the map

$$V(z) = V_s(z) := (s^2 + |z|^2)^{\frac{p-2}{4}} z, \quad z \in \mathbb{R}^n,$$

that are the inequalities

$$|V(z_1) - V(z_2)|^2 \leq \tilde{c} \langle a(z_1) - a(z_2), z_1 - z_2 \rangle \quad (4.9)$$

and

$$\frac{1}{\tilde{c}} \left( s^2 + |z_1|^2 + |z_2|^2 \right)^{\frac{p-2}{2}} \leq \frac{|V(z_1) - V(z_2)|^2}{|z_1 - z_2|^2} \leq \tilde{c} \left( s^2 + |z_1|^2 + |z_2|^2 \right)^{\frac{p-2}{2}} \quad (4.10)$$

holding for some universal constant  $\tilde{c} = \tilde{c}(n, \nu, L) \geq 1$ , whenever  $z_1, z_2 \in \mathbb{R}^n$ . Applying (4.6) together with (4.9) yields

$$\int_Q \frac{|V(Du) - V(Dw)|^2}{(\alpha + |u-w|)^{\xi}} dx dt \leq c \frac{\alpha^{1-\xi}}{\xi-1} |\mu|(Q).$$

This implies, together with Hölder’s inequality and (4.8), recalling also (4.7),

$$\begin{aligned}
 & \int_Q |V(Du) - V(Dw)|^{\frac{2}{p}} dx dt \\
 &= \int_Q \left( \frac{|V(Du) - V(Dw)|^2}{(\alpha + |u - w|)^\xi} \right)^{\frac{1}{p}} (\alpha + |u - w|)^{\frac{\xi}{p}} dx dt \\
 &\leq \left( \int_Q \frac{|V(Du) - V(Dw)|^2}{(\alpha + |u - w|)^\xi} dx dt \right)^{\frac{1}{p}} \left( \int_Q (\alpha + |u - w|)^{\frac{n+1}{n}} dx dt \right)^{\frac{p-1}{p}} \tag{4.11} \\
 &\leq c \left( \frac{|\mu|(Q)}{|Q|} \alpha^{1-\xi} \right)^{\frac{1}{p}} \alpha^{\frac{\xi}{p}} \\
 &\leq c \left( \frac{|\mu|(Q)^{\frac{n+2}{n+1}}}{|Q|} \left( \int_Q |Du - Dw| dx dt \right)^{\frac{n}{n+1}} \right)^{\frac{1}{p}}.
 \end{aligned}$$

We then use (4.10) as follows:

$$\begin{aligned}
 |Du - Dw| &= \left[ (|Du|^2 + |Dw|^2 + s^2)^{\frac{p-2}{2}} |Du - Dw|^2 \right]^{\frac{1}{2}} \\
 &\quad \cdot (|Du|^2 + |Dw|^2 + s^2)^{\frac{2-p}{4}} \tag{4.12} \\
 &\leq c |V(Du) - V(Dw)| (|Du|^2 + |Dw|^2 + s^2)^{\frac{2-p}{4}} \\
 &\leq c |V(Du) - V(Dw)| (|Du - Dw|^2 + |Du|^2 + s^2)^{\frac{2-p}{4}}.
 \end{aligned}$$

Young’s inequality, used when  $p < 2$  with conjugate exponents  $(2/p, 2/(2 - p))$ ,

$$ab^{\frac{2-p}{2}} \leq \gamma b + c(p, \gamma) a^{\frac{2}{p}}, \quad \gamma \in (0, 1),$$

yields

$$\begin{aligned}
 |Du - Dw| &\leq c |V(Du) - V(Dw)|^{\frac{2}{p}} + \frac{1}{2} |Du - Dw| \\
 &\quad + c |V(Du) - V(Dw)| (|Du| + s)^{\frac{2-p}{2}} \tag{4.13}
 \end{aligned}$$

with a constant  $c$  depending only on  $n$  and  $p$ . In turn, we deduce by Hölder’s inequality that

$$\begin{aligned}
 \int_Q |Du - Dw| dx dt &\leq c \int_Q |V(Du) - V(Dw)|^{\frac{2}{p}} dx dt \\
 &\quad + c \left( \int_Q |V(Du) - V(Dw)|^{\frac{2}{p}} dx dt \right)^{\frac{p}{2}} \left( \int_Q (|Du| + s) dx dt \right)^{\frac{2-p}{2}}. \tag{4.14}
 \end{aligned}$$

Now, if on one hand

$$\begin{aligned} & \left( \int_Q |V(Du) - V(Dw)|^{\frac{2}{p}} dx dt \right)^{\frac{p}{2}} \left( \int_Q (|Du| + s) dx dt \right)^{\frac{2-p}{2}} \\ & \leq \int_Q |V(Du) - V(Dw)|^{\frac{2}{p}} dx dt, \end{aligned}$$

implying by (4.14) that

$$\int_Q |Du - Dw| dx dt \leq c \int_Q |V(Du) - V(Dw)|^{\frac{2}{p}} dx dt,$$

then using (4.11) in the above inequality leads to

$$\int_Q |Du - Dw| dx dt \leq c \left[ \frac{|\mu|(Q)}{|Q|^{\frac{n+1}{n+2}}} \right]^{\frac{n+2}{(p-1)n+p}}$$

and the result follows in this case. If on the other hand

$$\begin{aligned} & \int_Q |V(Du) - V(Dw)|^{\frac{2}{p}} dx dt \\ & \leq \left( \int_Q |V(Du) - V(Dw)|^{\frac{2}{p}} dx dt \right)^{\frac{p}{2}} \left( \int_Q (|Du| + s) dx dt \right)^{\frac{2-p}{2}}, \end{aligned}$$

then

$$\begin{aligned} & \int_Q |Du - Dw| dx dt \\ & \leq c \left( \int_Q |V(Du) - V(Dw)|^{\frac{2}{p}} dx dt \right)^{\frac{p}{2}} \left( \int_Q (|Du| + s) dx dt \right)^{\frac{2-p}{2}} \end{aligned}$$

holds by (4.14). Inserting (4.11) into this we obtain

$$\begin{aligned} & \int_Q |Du - Dw| dx dt \\ & \leq c \left( \frac{|\mu|(Q)^{\frac{n+2}{n+1}}}{|Q|^{\frac{n}{n+1}}} \left( \int_Q |Du - Dw| dx dt \right)^{\frac{n}{n+1}} \right)^{\frac{1}{2}} \left( \int_Q (|Du| + s) dx dt \right)^{\frac{2-p}{2}} \end{aligned}$$

and thus also

$$\int_Q |Du - Dw| dx dt \leq c \left[ \frac{|\mu|(Q)}{|Q|^{\frac{n+1}{n+2}}} \right] \left( \int_Q (|Du| + s) dx dt \right)^{(2-p)\frac{n+1}{n+2}}.$$

This concludes the proof.  $\square$

**Corollary 4.4.** *Let  $u$  and  $w$  be as in Lemma 4.3. Suppose that the intrinsic relations*

$$\int_{Q_\varrho^\lambda} (|Du| + s) \, dx \, dt \leq \lambda, \quad \frac{|\mu|(Q_\varrho^\lambda)}{\varrho^{n+1}} \leq \lambda, \tag{4.15}$$

*are satisfied. Then there exists a constant  $c_4 = c_4(n, p, v, L) \geq 1$  such that*

$$\int_{Q_\varrho^\lambda} |Du - Dw| \, dx \, dt \leq c_4 \left[ \frac{|\mu|(Q_\varrho^\lambda)}{\varrho^{N-1}} \right]. \tag{4.16}$$

*Proof.* Simply note that by the first intrinsic inequality in (4.15) we have

$$\begin{aligned} & \left[ \frac{|\mu|(Q_\varrho^\lambda)}{|Q_\varrho^\lambda|^{\frac{n+1}{n+2}}} \right] \left( \int_{Q_\varrho^\lambda} (|Du| + s) \, dx \, dt \right)^{(2-p)\frac{n+1}{n+2}} \\ &= \left[ \frac{|\mu|(Q_\varrho^\lambda)}{\varrho^{n+1}} \right] \lambda^{(p-2)\frac{n+1}{n+2}} \left( \int_{Q_\varrho^\lambda} (|Du| + s) \, dx \, dt \right)^{(2-p)\frac{n+1}{n+2}} \leq \frac{|\mu|(Q_\varrho^\lambda)}{\varrho^{n+1}}, \end{aligned}$$

and, moreover, using the second inequality in (4.15) we also get

$$\left[ \frac{|\mu|(Q_\varrho^\lambda)}{|Q_\varrho^\lambda|^{\frac{n+1}{n+2}}} \right]^{\frac{n+2}{(p-1)n+p}} = \left[ \frac{|\mu|(Q_\varrho^\lambda)}{\varrho^{n+1}} \right]^{\frac{n+2}{(p-1)n+p}} \lambda^{\frac{(n+1)(p-2)}{(p-1)n+p}} \leq \frac{|\mu|(Q_\varrho^\lambda)}{\varrho^{n+1}}.$$

The proof of (4.16) now follows using the last two inequalities together with Lemma 4.3, and finally recalling that  $N - 1 = n + 1$ . □

In the following, in accordance to (3.2), we define

$$E(g, Q) := \int_Q |g - (g)_Q| \, dx \, dt$$

whenever  $g \in L^1(Q, \mathbb{R}^n)$  and  $Q \subset \Omega_T$  is a cylinder. We proceed with another couple of technical lemmas whose general form will make them useful also in the next section, where we will prove the gradient continuity results.

**Lemma 4.5.** *Let  $\delta, \theta \in (0, 1)$ . Suppose that*

$$\int_{Q_\varrho^\lambda} (|Du| + s) \, dx \, dt \leq \lambda \quad \text{and} \quad \frac{|\mu|(Q_\varrho^\lambda)}{\varrho^{N-1}} \leq \frac{\delta^{n+2}}{c_4} \theta \lambda, \tag{4.17}$$

*where  $c_4 \equiv c_4(n, p, v, L)$  is as in Corollary 4.4. Then*

$$s + \sup_{\frac{1}{2}Q_\varrho^\lambda} \|Dw\| \leq 6c_3\lambda, \tag{4.18}$$

where  $c_3 \equiv c_3(n, p, \nu, L)$  is as in Theorem 4.2, and, moreover, the lower bound

$$\int_{\delta Q_\rho^\lambda} |Du| dx dt - \theta\lambda \leq \int_{\delta Q_\rho^\lambda} |Dw| dx dt \leq \sqrt{n} \sup_{\delta Q_\rho^\lambda} \|Dw\| \tag{4.19}$$

holds.

*Proof.* Corollary 4.4, in view of (4.17), gives the comparison estimate

$$\int_{Q_\rho^\lambda} |Du - Dw| dx dt \leq \delta^{n+2}\theta\lambda. \tag{4.20}$$

This, together with (4.17), further implies the bound

$$\int_{Q_\rho^\lambda} (|Dw| + s) dx dt \leq \int_{Q_\rho^\lambda} (|Du| + s) dx dt + \int_{Q_\rho^\lambda} |Du - Dw| dx dt \leq 2\lambda.$$

Therefore, an application of Theorem 4.2 yields

$$\sup_{\frac{1}{2}Q_\rho^\lambda} \|Dw\| \leq c_3\lambda + c_3 2^{\frac{2}{n(p-2)+2}}\lambda \leq 5c_3\lambda,$$

because  $n(p - 2) + 2 \geq 1$  as we are assuming (1.3). Now (4.18) follows using again (4.17). Applying then (4.20) together with the triangle inequality yields

$$\begin{aligned} \int_{\delta Q_\rho^\lambda} |Du| dx dt &\leq \int_{\delta Q_\rho^\lambda} |Du - Dw| dx dt + \int_{\delta Q_\rho^\lambda} |Dw| dx dt \\ &\leq \delta^{-(n+2)} \int_{Q_\rho^\lambda} |Du - Dw| dx dt + \int_{\delta Q_\rho^\lambda} |Dw| dx dt \\ &\leq \theta\lambda + \int_{\delta Q_\rho^\lambda} |Dw| dx dt, \end{aligned}$$

thereby finishing the proof. □

**Lemma 4.6.** *Let  $\delta \in (0, 1/2)$ . Suppose that  $Dw$  satisfies the decay estimate*

$$E(Dw, \delta Q_\rho^\lambda) \leq \left(\frac{\varepsilon}{2^{n+7}}\right) E(Dw, 2^{-1}Q_\rho^\lambda) \tag{4.21}$$

for some  $\varepsilon \in (0, 1]$ , and that the bounds

$$\int_{Q_\rho^\lambda} (|Du| + s) dx dt \leq \lambda \quad \text{and} \quad \frac{|\mu|(Q_\rho^\lambda)}{\rho^{N-1}} \leq \lambda \tag{4.22}$$

hold. Then we have

$$E(Du, \delta Q_\rho^\lambda) \leq \left(\frac{\varepsilon}{4}\right) E(Du, Q_\rho^\lambda) + 4c_4\delta^{-(n+2)} \left[ \frac{|\mu|(Q_\rho^\lambda)}{\rho^{N-1}} \right],$$

where  $c_4 \equiv c_4(n, p, \nu, L)$  is as in Corollary 4.4.

*Proof.* Applying the triangle inequality and assumption (4.21), we arrive at the following chain of inequalities:

$$\begin{aligned}
 E(Du, \delta Q_\varrho^\lambda) &\leq 2 \int_{\delta Q_\varrho^\lambda} |Du - (Dw)_{\delta Q_\varrho^\lambda}| dx dt \\
 &\leq 2E(Dw, \delta Q_\varrho^\lambda) + 2 \int_{\delta Q_\varrho^\lambda} |Du - Dw| dx dt \\
 &\leq \varepsilon 2^{-(n+6)} E(Dw, 2^{-1} Q_\varrho^\lambda) + 2\delta^{-(n+2)} \int_{Q_\varrho^\lambda} |Du - Dw| dx dt \\
 &\leq \varepsilon 2^{-(n+5)} E(Du, 2^{-1} Q_\varrho^\lambda) + 4\delta^{-(n+2)} \int_{Q_\varrho^\lambda} |Du - Dw| dx dt \\
 &\leq \varepsilon 2^{-(n+5)} E(Du, 2^{-1} Q_\varrho^\lambda) + 4c_4 \delta^{-(n+2)} \left[ \frac{|\mu|(Q_\varrho^\lambda)}{\varrho^{N-1}} \right].
 \end{aligned}$$

The last inequality is a consequence of (4.22) by Corollary 4.4. Observe also that we repeatedly used (3.3). The result follows by observing that

$$E(Du, 2^{-1} Q_\varrho^\lambda) \leq 2 \int_{\frac{1}{2} Q_\varrho^\lambda} |Du - (Du)_{Q_\varrho^\lambda}| dx dt \leq 2^{n+3} E(Du, Q_\varrho^\lambda). \quad \square$$

**4.2. Proof of Theorem 1.2**

The starting point here is a new iteration method, already introduced in [30] in order to treat the case  $p \geq 2$ , that will be here modified to treat the subquadratic case via Lemmas 4.5 and 4.6. We shall use large (de)magnifying constants such as 600, 800, 1200, to clarify the role of certain passages in the proof. Now, define the set  $\mathcal{L}_\lambda$  (of Lebesgue points) as

$$\mathcal{L}_\lambda = \left\{ (x_0, t_0) \in \Omega_T : \lim_{\varrho \rightarrow 0} \int_{Q_\varrho^\lambda(x_0, t_0)} Du dx dt = Du(x_0, t_0) \right\} \quad (4.23)$$

for  $\lambda > 0$ . Basic properties of maximal operators - see for instance [43, Chapter 1, Page 8] - imply that this set is actually independent of  $\lambda$  and, in particular,  $\mathcal{L}_\lambda = \mathcal{L}_1 =: \mathcal{L}$  for all  $0 < \lambda < \infty$ . Moreover,  $\tilde{Q} \setminus \mathcal{L}$  has zero Lebesgue measure. Therefore, in the following, when referring to the statement of Theorem 1.1, we shall prove (1.16) whenever  $(x_0, t_0) \in \mathcal{L}$ .

*Step 1: Setting of the constants and basic inequalities.* With  $(x_0, t_0) \in \mathcal{L}$  being fixed, in the following all the cylinders will have  $(x_0, t_0)$  as vertex, therefore we shall as usual omit denoting the vertex simply writing  $Q_\varrho^\lambda(x_0, t_0) \equiv Q_\varrho^\lambda$ . We start taking  $\lambda$  of the form

$$\lambda \geq H_1 \beta + H_2 \int_0^{2r_\lambda} \frac{|\mu|(Q_\varrho^\lambda)}{\varrho^{N-1}} \frac{d\varrho}{\varrho}, \quad r_\lambda = \lambda^{(p-2)/2r}, \quad (4.24)$$



and fix the constants  $H_1, H_2 \geq 1$  in a few lines, in a way that makes them depending only on  $n, p, v, L$ . In the end, when proving (1.16), we shall simply take  $c := \max\{H_1, H_2\}$ . Then  $\lambda$  defined in (1.16) certainly satisfies (4.24). Taking constant  $c_3$  from Theorem 4.2, we define

$$A := 6c_3, \quad B := 400\sqrt{n}, \quad \gamma = 2^{-(n+7)}. \tag{4.25}$$

We then fix the constant  $\delta_\gamma \equiv \delta_\gamma(n, p, v, L, A, B, \gamma) \in (0, 1/2)$  in Theorem 3.5 corresponding to the choices in (4.25). Since  $A$  depends itself on  $n, p, v, L$ , this ultimately fixes a constant  $\delta_\gamma \in (0, 1/2)$  depending only on  $n, p, v, L$ . Next, we take  $k$  as the smallest integer (larger or equal to 2) satisfying

$$8\sqrt{n}A(\delta_\gamma/2)^{(k-1)\alpha} \leq \frac{(\delta_\gamma/2)^{n+2}}{800}, \tag{4.26}$$

and in this way  $k$  depends only upon  $n, p, v, L$ . Now, define

$$Q_i := Q_{r_i}^\lambda, \quad r_i = \delta_1^i r_\lambda, \quad \delta_1 := \delta_\gamma/2 \tag{4.27}$$

whenever  $i \geq 0$  is an integer; again  $\delta_1 \equiv \delta_1(n, p, v, L) \in (0, 1/4)$ . We also set

$$H_1 := 400\delta_1^{-(n+2)} \quad \text{and} \quad H_2 := 1600c_4\delta_1^{-(k+3)(n+2)}, \tag{4.28}$$

where  $c_4 \equiv c_4(n, p, v, L)$  has been fixed in Corollary 4.4. Notice that the choice of  $H_1$  implies that

$$\int_{Q_0} (|Du| + s) \, dx \, dt + \delta_1^{-(n+2)} E(Du, Q_0) \leq (1 + 2\delta_1^{-(n+2)})\beta \leq \frac{\lambda}{100} \tag{4.29}$$

and

$$s \leq \frac{\lambda}{400} \tag{4.30}$$

hold. Now, recalling again that  $N - 1 = n + 1$ , observe that

$$\begin{aligned} \int_0^{2r_\lambda} \frac{|\mu|(Q_\varrho^\lambda)}{\varrho^{N-1}} \frac{d\varrho}{\varrho} &= \sum_{i=0}^\infty \int_{r_{i+1}}^{r_i} \frac{|\mu|(Q_\varrho^\lambda)}{\varrho^{N-1}} \frac{d\varrho}{\varrho} + \int_{r_\lambda}^{2r_\lambda} \frac{|\mu|(Q_\varrho^\lambda)}{\varrho^{N-1}} \frac{d\varrho}{\varrho} \\ &\geq \sum_{i=0}^\infty \frac{|\mu|(Q_{i+1})}{r_i^{N-1}} \int_{r_{i+1}}^{r_i} \frac{d\varrho}{\varrho} + \frac{|\mu|(Q_0)}{(2r_\lambda)^{N-1}} \int_{r_\lambda}^{2r_\lambda} \frac{d\varrho}{\varrho} \\ &= \delta_1^{n+1} \log\left(\frac{1}{\delta_1}\right) \sum_{i=0}^\infty \frac{|\mu|(Q_{i+1})}{r_{i+1}^{N-1}} + \frac{\log 2}{2^{n+1}} \left[ \frac{|\mu|(Q_0)}{r_\lambda^{N-1}} \right] \\ &\geq \delta_1^{n+2} \sum_{i=0}^\infty \frac{|\mu|(Q_i)}{r_i^{N-1}}. \end{aligned} \tag{4.31}$$

Therefore, by (4.24) and the choice in (4.28) it follows that

$$8c_4\delta_1^{-(k+2)(n+2)} \sum_{i=0}^{\infty} \frac{|\mu|(Q_i)}{r_i^{N-1}} \leq \frac{\lambda}{200}. \quad (4.32)$$

In particular, we have

$$\frac{|\mu|(Q_i)}{r_i^{N-1}} \leq \frac{\delta_1^{(k+2)(n+2)}}{1600c_4} \lambda \leq \lambda \quad \forall i \geq 0. \quad (4.33)$$

*Step 2: Exit time argument.* Next, whenever  $i \geq 0$ , define

$$C_i = \int_{Q_i} (|Du| + s) \, dx \, dt + \delta_1^{-(n+2)} E(Du, Q_i), \quad (4.34)$$

so that (4.29) reads also as  $C_0 \leq \lambda/100$ . Let us show that without loss of generality we may assume there exists an exit index  $i_e \geq 0$  with respect to the previous inequality, that is an integer  $i_e \geq 0$  such that

$$C_{i_e} \leq \frac{\lambda}{100}, \quad C_j > \frac{\lambda}{100} \quad \forall j > i_e. \quad (4.35)$$

Indeed, on the contrary, we could find an increasing subsequence  $\{j_i\}$  such that  $C_{j_i} \leq \lambda/100$ , for every  $i \in \mathbb{N}$ , and then, as  $(x_0, t_0) \in \mathcal{L}$ , obviously

$$|Du(x_0, t_0)| \leq \lim_{i \rightarrow \infty} \int_{Q_{j_i}} |Du| \, dx \, dt \leq \frac{\lambda}{100},$$

and the proof would be finished. Therefore, for the rest of the proof, we shall argue under the additional assumption (4.35).

*Step 3: Estimates after the exit index.* The following lemma is the core of the proof:

**Lemma 4.7.** *If  $i \geq i_e$ , then inequality*

$$\int_{Q_i} (|Du| + s) \, dx \, dt \leq \lambda \quad (4.36)$$

*implies*

$$E(Du, Q_{i+1}) \leq \frac{1}{4} E(Du, Q_i) + 4c_4\delta_1^{-(n+2)} \left[ \frac{|\mu|(Q_i)}{r_i^{N-1}} \right]. \quad (4.37)$$

*Proof.* We begin with a comparison estimate. By (4.33) and (4.36) we may apply Corollary 4.4 so that

$$\int_{Q_{i+j}} |Du - Dw_i| dx dt \leq \frac{|Q_i|}{|Q_{i+j}|} \int_{Q_i} |Du - Dw_i| dx dt \leq c_4 \delta_1^{-j(n+2)} \left[ \frac{|\mu|(Q_i)}{r_i^{N-1}} \right]$$

holds for all  $j = 0, 1, \dots, k$  (recall that  $k \equiv k(n, p, v, L) \in \mathbb{N}$  has been defined in Step 1). Using again (4.33) we obtain

$$\int_{Q_{i+j}} |Du - Dw_i| dx dt \leq \frac{\delta_1^{(k+1-j)(n+2)}}{1600} \lambda \tag{4.38}$$

for all  $j = 0, 1, \dots, k$ . Next, conditions (4.36) and (4.33) allow to verify assumption (4.17) in Lemma 4.5 both with  $\delta \equiv \delta_1$  and with  $\delta \equiv \delta_1^k$  (take  $\theta \equiv 1/1600$ ). Thus we obtain

$$s + \sup_{Q_{i+1}} \|Dw_i\| \leq s + \sup_{\frac{1}{2}Q_i} \|Dw_i\| \leq 6c_3 \lambda \equiv A\lambda \tag{4.39}$$

and

$$\int_{Q_{i+k}} |Du| dx dt - \frac{\lambda}{1600} \leq \int_{Q_{i+k}} |Dw_i| dx dt \leq \sqrt{n} \sup_{Q_{i+1}} \|Dw_i\|. \tag{4.40}$$

At this point, as a consequence of Theorem 3.3 (applied with  $Q_\varrho^\lambda \equiv Q_{i+k}$ ) and by (4.39) and (4.26), we have

$$2 \operatorname{osc}_{Q_{i+k}} Dw_i \leq 8\sqrt{n} A \delta_1^{(k-1)\alpha} \lambda \leq \frac{\delta_1^{n+2}}{800} \lambda.$$

This and (4.38) imply

$$\begin{aligned} E(Du, Q_{i+k}) &\leq 2 \int_{Q_{i+k}} |Du - (Dw_i)_{Q_{i+k}}| dx dt \\ &\leq 2E(Dw_i, Q_{i+k}) + 2 \int_{Q_{i+k}} |Du - Dw_i| dx dt \\ &\leq 2 \operatorname{osc}_{Q_{i+k}} Dw_i + \frac{\delta_1^{n+2} \lambda}{800} \\ &\leq \frac{\delta_1^{n+2} \lambda}{400}. \end{aligned} \tag{4.41}$$

Combining (4.30) and (4.41) we obtain

$$\begin{aligned} C_{i+k} &= \int_{Q_{i+k}} (|Du| + s) dx dt + \delta_1^{-(n+2)} E(Du, Q_{i+k}) \\ &\leq \int_{Q_{i+k}} |Du| dx dt + \frac{\lambda}{200}, \end{aligned}$$

and therefore the fact  $C_{i+k} \geq \lambda/100$  and (4.40) imply

$$\frac{\lambda}{B} \equiv \frac{\lambda}{400\sqrt{n}} \leq \sup_{Q_{i+1}} \|Dw_i\|.$$

The lower bound above combined with (4.39) allows to verify assumptions of Theorem 3.5 in  $Q_r^\lambda \equiv 2^{-1}Q_i$ , with the choice of the constants made in (4.25). Hence we get

$$E(Dw_i, Q_{i+1}) = E(Dw_i, (\delta_\gamma/2)Q_i) \leq 2^{-(n+7)} E(Dw_i, 2^{-1}Q_i).$$

Inequality (4.37) now follows by Lemma 4.6, which is in turn applicable with  $\varepsilon = 1$  thanks to the previous inequality, (4.33), and (4.36).

*Step 4: Iteration and conclusion.* Denote in short

$$A_i := E(Du, Q_i), \quad k_i = |(Du)_{Q_i}|.$$

By the definition in (4.34) and (4.35) we have

$$s + k_{i_e} + \delta_1^{-(n+2)} A_{i_e} \leq C_{i_e} \leq \frac{\lambda}{100}. \tag{4.42}$$

We now prove, by induction, that

$$s + k_j + A_j \leq \lambda \tag{4.43}$$

holds whenever  $j \geq i_e$ . Indeed, by (4.42), the case  $j = i_e$  of the previous inequality holds. Then, assume by induction that (4.43) holds whenever  $j \in \{i_e, \dots, i\}$ . This, in particular, implies that

$$\int_{Q_j} (|Du| + s) \, dx \, dt \leq s + k_j + A_j \leq \lambda$$

whenever  $j \in \{i_e, \dots, i\}$ . Lemma 4.7 is hence at our disposal for such  $j$  and estimate (4.37) gives

$$A_{j+1} \leq \frac{1}{4} A_j + 4c_4 \delta_1^{-(n+2)} \left[ \frac{|\mu|(Q_j)}{r_j^{N-1}} \right] \tag{4.44}$$

for all  $j \in \{i_e, \dots, i\}$ . It immediately follows by (4.43) (assumed for all  $j \in \{i_e, \dots, i\}$ ) and (4.33) that

$$A_{i+1} \leq \frac{\lambda}{4} + 4c_4 \delta_1^{-(n+2)} \left[ \frac{|\mu|(Q_i)}{r_i^{N-1}} \right] \leq \frac{\lambda}{4} + \frac{\lambda}{400} \leq \frac{\lambda}{3}. \tag{4.45}$$

Furthermore, summing up (4.44) for  $j \in \{i_e, \dots, i\}$  leads to

$$\sum_{j=i_e}^{i+1} A_j \leq A_{i_e} + \frac{1}{4} \sum_{j=i_e}^i A_j + 4c_4 \delta_1^{-(n+2)} \sum_{j=i_e}^{i+1} \frac{|\mu|(Q_j)}{r_j^{N-1}}, \quad (4.46)$$

yielding

$$\sum_{j=i_e}^{i+1} A_j \leq 2A_{i_e} + 8c_4 \delta_1^{-(n+2)} \sum_{j=i_e}^{i+1} \frac{|\mu|(Q_j)}{r_j^{N-1}}. \quad (4.47)$$

Using the previous inequality we have

$$\begin{aligned} k_{i+1} - k_{i_e} &= \sum_{j=i_e}^i (k_{j+1} - k_j) \leq \sum_{j=i_e}^i \int_{Q_{j+1}} |Du - (Du)_{Q_j}| dx dt \\ &\leq \sum_{j=i_e}^i \frac{|Q_j|}{|Q_{j+1}|} \int_{Q_j} |Du - (Du)_{Q_j}| dx dt \\ &= \delta_1^{-(n+2)} \sum_{j=i_e}^i A_j \\ &\leq 2\delta_1^{-(n+2)} A_{i_e} + 8c_4 \delta_1^{-2(n+2)} \sum_{j=i_e}^i \frac{|\mu|(Q_j)}{r_j^{N-1}} \end{aligned}$$

and thus it follows from (4.32) that

$$k_{i+1} \leq k_{i_e} + 2\delta_1^{-(n+2)} A_{i_e} + 8c_4 \delta_1^{-2(n+2)} \sum_{j=0}^{\infty} \frac{|\mu|(Q_j)}{r_j^{N-1}} \leq 2C_{i_e} + \frac{\lambda}{200}.$$

In turn, by (4.42) the previous estimate yields  $k_{i+1} \leq \lambda/3$ . The last inequality together with (4.30) and (4.45) allows to verify the induction step, *i.e.*

$$s + k_{i+1} + A_{i+1} \leq \frac{\lambda}{400} + \frac{\lambda}{3} + \frac{\lambda}{3} < \lambda.$$

Therefore (4.43) holds for every  $i \geq i_e$ . Estimate (1.16) finally follows with the choice (announced at the beginning)  $c := \max\{H_1, H_2\}$ , since, as  $(x_0, t_0) \in \mathcal{L}$ , it holds that

$$|Du(x_0, t_0)| = \lim_{i \rightarrow \infty} k_i \leq \lambda,$$

finishing the proof of Theorem 1.1.  $\square$

*Proof of Corollary 1.2.* Let us assume that  $\mathbf{I}_1^h(x_0, t_0; 2r) < \infty$ , otherwise the proof trivializes. Next, let us consider the function

$$h(\lambda) := \lambda - c\lambda^{\frac{n(2-p)}{2}} A(\lambda) - c\lambda^{\frac{(n+1)(2-p)}{2}} B(\lambda),$$

where

$$A(\lambda) := \frac{1}{|Q_r|} \int_{Q_{r\lambda}^\lambda} (|Du| + s + 1) \, dx \, dt$$

and

$$B(\lambda) := \int_0^{2r} \frac{|\mu|(Q_{\lambda^{(p-2)/2}\varrho}^\lambda(x_0, t_0)) \, d\varrho}{\varrho^{N-1}} \frac{d\varrho}{\varrho},$$

and  $c$  is again the constant appearing in Theorem 1.2. Clearly both  $A$  and  $B$  are nonincreasing functions of  $\lambda$  in  $(0, \infty)$ , because

$$Q_{\lambda_2^{(p-2)/2}\varrho}^{\lambda_2} \subset Q_{\lambda_1^{(p-2)/2}\varrho}^{\lambda_1} \quad \text{for all } \lambda_2 > \lambda_1 > 0 \quad \text{and } \varrho \in (0, 2r].$$

We consider the function  $h(\cdot)$  defined for all those  $\lambda$  such that  $Q_r^\lambda \subset \Omega_T$ ; observe that the domain of definition of  $h(\cdot)$  includes  $[1, \infty)$  as  $Q_{\lambda^{(p-2)/2}r}^\lambda \subset Q_r \subset \Omega_T$  when  $\lambda \geq 1$ . Again, observe that  $h(\cdot)$  is a continuous function and moreover  $h(1) \leq 0$  as  $c \geq 1$  and  $A(1) \geq 1$ . On the other hand, observe that

$$\lim_{\lambda \rightarrow \infty} h(\lambda) \geq \lim_{\lambda \rightarrow \infty} \left[ \lambda - c\lambda^{\frac{n(2-p)}{2}} A(1) - c\lambda^{\frac{(n+1)(2-p)}{2}} B(1) \right] = \infty,$$

because  $(n + 1)(2 - p)/2 < 1$  for  $p > 2 - 1/(n + 1)$ . It follows that there exists a number  $\lambda \geq 1$  such that  $h(\lambda) = 0$  and this means that  $\lambda$  solves (1.16) with

$$\beta = \beta(\lambda) = \int_{Q_{r\lambda}^\lambda} (|Du| + s + 1) \, dx \, dt = \lambda^{\frac{n(2-p)}{2}} A(\lambda).$$

Therefore we can apply (1.16) to have

$$\lambda + |Du(x_0, t_0)| \leq 2c\lambda^{\frac{n(2-p)}{2}} A(\lambda) + 2c\lambda^{\frac{(n+1)(2-p)}{2}} B(\lambda). \tag{4.48}$$

On the other hand, observe that by Young’s inequality with conjugate exponents

$$\left( \frac{2}{n(2-p)}, \frac{2}{2-n(2-p)} \right)$$

and

$$\left( \frac{2}{(n+1)(2-p)}, \frac{2}{(n+1)p-2n} \right)$$

we have

$$\begin{aligned} 2c\lambda^{\frac{n(2-p)}{2}} A(\lambda) &\leq \lambda/4 + [2cA(1)]^{\frac{2}{2-n(2-p)}} \\ &\leq \lambda/4 + \tilde{c} \left( \int_{Q_r} (|Du| + s + 1) \, dx \, dt \right)^{\frac{2}{2-n(2-p)}} \end{aligned}$$

and

$$\begin{aligned}
 2c\lambda^{\frac{(n+1)(2-p)}{2}} B(\lambda) &\leq \lambda/4 + [2cB(1)]^{\frac{2}{(n+1)p-2n}} \\
 &\leq \lambda/4 + \tilde{c} \left( \int_0^{2r} \left[ \frac{|\mu|(Q_\varrho(x_0, t_0))}{\varrho^{n+1}} \right] \frac{d\varrho}{\varrho} \right)^{\frac{2}{(n+1)p-2n}}. \tag{4.49}
 \end{aligned}$$

Substituting the last inequalities into (4.48) readily gives (1.17). □

*Proof of Corollary 1.3.* We adopt the notation from the proof of Theorem 1.3 and replace (4.49) by a different estimate. Indeed, the integrand of  $B(\lambda)$  can be estimated as follows:

$$\begin{aligned}
 \frac{|\mu| \left( Q_{\lambda^{(p-2)/2}\varrho}^\lambda \right)}{\varrho^{N-1}} &\leq \|f\|_{L^\infty} \frac{|\mu_0| \left( B_{\lambda^{(p-2)/2}\varrho} \right)}{\varrho^{n-1}} \\
 &= \|f\|_{L^\infty} \lambda^{\frac{(n-1)(p-2)}{2}} \frac{|\mu_0| \left( B_{\lambda^{(p-2)/2}\varrho} \right)}{(\lambda^{(p-2)/2}\varrho)^{n-1}}.
 \end{aligned}$$

Changing variables eventually leads to

$$\begin{aligned}
 c\lambda^{\frac{(n+1)(2-p)}{2}} B(\lambda) &\leq c\lambda^{2-p} \|f\|_{L^\infty} \mathbf{I}_1^{\mu_0}(x_0, 2\lambda^{(p-2)/2}r) \\
 &\leq c\lambda^{2-p} \|f\|_{L^\infty} \mathbf{I}_1^{\mu_0}(x_0, 2r),
 \end{aligned}$$

where we used that  $\lambda \geq 1$  to derive the last inequality. In turn applying Young’s inequality with conjugate exponents  $(1/(2-p), 1/(p-1))$  gives

$$c\lambda^{2-p} \|f\|_{L^\infty} \mathbf{I}_1^{\mu_0}(x_0, 2r) \leq \lambda/4 + c\|f\|_{L^\infty}^{1/(p-1)} [\mathbf{I}_1^{\mu_0}(x_0, 2r)]^{1/(p-1)}.$$

The rest of the proof is analogous to the one of Theorem 1.3. □

## 5. Continuity of the gradient via potentials

### 5.1. Preliminary choice of the geometry

Let us fix an open subcylinder  $\tilde{Q} \Subset \Omega_T$  such that  $\tilde{Q} = \tilde{\Omega} \times (t_1, t_2)$  where  $\tilde{\Omega} \Subset \Omega$  is a smooth subdomain, and let us take an intermediate cylinder  $\tilde{Q}'$  such that

$$\tilde{Q} \Subset \tilde{Q}' \Subset \Omega_T. \tag{5.1}$$

Let us finally set

$$\bar{R}_0 := \text{dist}_{\text{par}}(\tilde{Q}, \partial_{\text{par}} \tilde{Q}') \approx \text{dist}_{\text{par}}(\tilde{Q}', \partial_{\text{par}} \Omega_T) > 0. \tag{5.2}$$

Under the assumptions of Theorems 1.5-1.9, since  $(x, t) \mapsto \mathbf{I}_1^\mu(x, t; r)$  is locally bounded in  $\Omega_T$ , we can apply Theorem 1.3 so that the gradient is locally bounded  $\Omega_T$ ; in particular  $Du$  is bounded in  $\tilde{Q}'$ . Consequently, we denote

$$M := 1 + s + \sup_{\tilde{Q}'} |Du| < \infty. \tag{5.3}$$

Notice that by estimate (1.17) and (5.2) the number  $M$  depends only on the quantities  $n, N, p, \nu, L, \|Du\|_{L^p}$  and  $R_0$ . We now distinguish two cases, the first is when  $p \leq 2$  (and of course the lower bound in (1.9) holds). In this case we consider cylinders of the type  $Q_{r,\lambda}^\lambda(x_0, t_0)$  defined in (1.14); notice that whenever  $\lambda$  increases the cylinder does not increase as  $p \leq 2$ . Let us fix

$$\lambda_M := M \quad \text{and} \quad R_0 := \lambda_M^{(p-2)/2} \bar{R}_0/4. \tag{5.4}$$

It follows that  $Q_r^{\lambda_M}(x_0, t_0) \subset \tilde{Q}'$  whenever  $(x_0, t_0) \in \tilde{Q}$  and  $r \leq R_0$ , and in particular

$$s + \sup_{Q_r^{\lambda_M}} \|Du\| \leq \lambda_M \quad \text{whenever } r \leq R_0. \tag{5.5}$$

In the proof of the continuity results given in the next section we shall now solely consider stretched cylinders of the type appearing in the previous display.

In the case  $p > 2$  (where the proofs can be obtained combining the methods used here with those already explained in [30]) the intrinsic cylinders taken are a bit different. Indeed, with the definition in (5.3) we consider intrinsic cylinders of the type  $Q_r^{\lambda_M}(x_0, t_0)$  with  $r \leq \bar{R}_0$ . Since this time  $p \geq 2$  we have  $Q_r^{\lambda_M}(x_0, t_0) \subset \tilde{Q}'$  when  $(x_0, t_0) \in \tilde{Q}$  and  $r \leq \bar{R}_0$ , that is when  $Q_r(x_0, t_0) \subset \tilde{Q}'$ . As stated above, we shall not deal with the case  $p \geq 2$  for brevity, if not to provide the description of the necessary modifications. Anyway we again remark that in the following we shall give full details only for the case  $p \leq 2$ , which is indeed more delicate.

**5.2. Proof of the gradient continuity results**

The plan of the section is the following. We start proving the VMO-regularity of  $Du$  by showing Theorem 1.6; this in turn opens the way to the Hölder continuity of  $Du$  under additional assumptions on  $\mu(\cdot)$ , and this is the content of Theorem 1.9. Then we upgrade the arguments for proving Theorem 1.6 to prove the continuity of results stated in Theorem 1.5. Finally, we report the necessary remarks for Theorem 1.7 and Corollaries 1.8-1.10.

*Proof of Theorem 1.6.* According to the definition of VMO-regularity of  $Du$  given in (1.22), let us fix an open subcylinder  $\tilde{Q} \Subset \Omega_T$  as in Section 5.1. We shall show that for every  $\varepsilon \in (0, 1)$  and  $(x_0, t_0) \in \tilde{Q}$  there exists a radius  $\bar{r}_\varepsilon < \bar{R}_0$ , which is independent of  $(x_0, t_0) \in \tilde{Q}$ , such that

$$E(Du, Q_\varepsilon(x_0, t_0)) = \int_{Q_\varepsilon(x_0, t_0)} |Du - (Du)_{Q_\varepsilon(x_0, t_0)}| dx dt < \varepsilon \tag{5.6}$$



holds whenever  $\varrho \in (0, \bar{r}_\varepsilon]$ . Preliminary to this fact, we are going to prove the following lemma:

**Lemma 5.1.** *Under the assumptions of Theorem 1.6, and with the notation of Section 5.1, for every  $\varepsilon > 0$ , there exists a radius of the type*

$$r_\varepsilon = \frac{\varepsilon^{1/\alpha_1}}{c_5} R(\varepsilon), \quad \text{with } \alpha_1 \in (0, 1), \quad c_5 \geq 1, \quad R(\varepsilon) \in (0, R_0] \quad (5.7)$$

such that

$$\int_{Q_\varrho^{\lambda_M}(x_0, t_0)} |Du - (Du)_{Q_\varrho^{\lambda_M}(x_0, t_0)}| dx dt < \varepsilon \lambda_M \quad (5.8)$$

holds whenever  $\varrho \in (0, r_\varepsilon]$ . Here  $c_5, \alpha_1 \equiv c_5, \alpha_1(n, p, \nu, L)$ , and  $R(\varepsilon)$  denotes another radius such that  $R(\varepsilon) \equiv R(n, p, \nu, L, M, \mu(\cdot), \varepsilon)$ . The radius  $R(\varepsilon)$  is determined in (5.12) below.

The previous lemma will eventually be used in the proof of Theorem 1.9.

*Step 1: Preliminaries.* With the choices made in (5.1)-(5.5) it follows that

$$\int_{Q_\varrho^{\lambda_M}(x_0, t_0)} (|Du| + s) dx dt \leq M = \lambda_M \quad (5.9)$$

holds whenever  $(x_0, t_0) \in \tilde{Q}$  and  $\varrho < R_0$ . With  $\varepsilon > 0$ , we choose the number  $\delta_\gamma \equiv \delta_\gamma(n, p, \nu, L, \varepsilon) \in (0, 1/2)$  in Theorem 3.5 corresponding to the choice of parameters

$$\lambda \equiv \lambda_M, \quad A \equiv 6c_3, \quad B \equiv \frac{100\sqrt{n}}{\varepsilon}, \quad \gamma \equiv \frac{\varepsilon}{2n+7}, \quad q = 1, \quad (5.10)$$

where  $c_3 \equiv c_3(n, p, \nu, L)$  is the constant fixed in Theorem 4.2. Set  $\delta_1 := \delta_\gamma/2$ , while in the following  $c_4 \equiv c_4(n, p, \nu, L)$  is the constant introduced in Corollary 4.4. In particular, by taking (3.24) into account we have

$$\delta_1 = \frac{\varepsilon^{2/\alpha}}{c_5}, \quad \alpha \in (0, 1), \quad c_5 \geq 1, \quad (5.11)$$

where  $\alpha$  and  $c_5$  depend only on  $n, p, \nu, L$ . We then choose  $R \in (0, R_0]$  such that

$$\sup_{0 < \varrho \leq \lambda_M^{(2-p)/2}} \sup_R \frac{|\mu|(Q_\varrho(x_0, t_0))}{\varrho^{N-1}} \leq \frac{\delta_1^{n+2} \varepsilon}{400c_4 \lambda_M^{(N-1)(2-p)/2}}, \quad (5.12)$$

which is again possible by assumption (1.21). Notice that at this stage  $R$  depends only on  $n, p, \nu, L, M$  and  $\varepsilon$  (and of course on the measure  $\mu(\cdot)$ ). As a consequence

we obtain

$$\begin{aligned} \sup_{0 < \varrho \leq R} \sup_{(x_0, t_0)} \frac{|\mu|(Q_{\varrho}^{\lambda_M}(x_0, t_0))}{\varrho^{N-1}} &\leq \sup_{0 < \varrho \leq R} \sup_{(x_0, t_0)} \frac{|\mu|(Q_{\lambda_M^{(2-p)/2} \varrho}(x_0, t_0))}{\varrho^{N-1}} \\ &\leq \frac{\delta_1^{n+2} \varepsilon}{400c_4} \leq \frac{\delta_1^{n+2} \varepsilon}{400c_4} \lambda_M. \end{aligned} \tag{5.13}$$

We remark that standard parabolic cylinders appear in the second term of the above lines. Finally, with  $(x_0, t_0) \in \tilde{Q}$ , we define the chain of shrinking intrinsic cylinders

$$Q_i \equiv Q_{r_i}^{\lambda_M}(x_0, t_0), \quad r_i = \delta_1^i r, \quad r \in (\delta_1 R, R]. \tag{5.14}$$

*Step 2: Proof of Lemma 5.1.* With  $\varepsilon > 0$  being fixed as in the statement of Lemma 5.1, we shall prove that

$$E(Du, Q_h) < \varepsilon \lambda_M \quad \forall h \in \mathbb{N} \cap [1, \infty). \tag{5.15}$$

Let  $h \geq 1$  and let us distinguish two cases; the first is when

$$\int_{Q_h} |Du| dx dt < \frac{\varepsilon \lambda_M}{50},$$

so that we have

$$E(Du, Q_h) \leq 2 \int_{Q_h} |Du| dx dt < \frac{\varepsilon \lambda_M}{25} \leq \varepsilon \lambda_M, \tag{5.16}$$

and (5.15) follows. The other case is when

$$\int_{Q_h} |Du| dx dt \geq \frac{\varepsilon \lambda_M}{50}. \tag{5.17}$$

Let  $w_{h-1}$  be the comparison solution introduced at the beginning of the Section 4.1 corresponding to the cylinder  $Q_{h-1}$ , i.e.  $w_{h-1}$  solves the Cauchy-Dirichlet problem

$$\begin{cases} (w_{h-1})_t - \operatorname{div} a(Dw_{h-1}) = 0 & \text{in } Q_{h-1} \\ w_{h-1} = u & \text{on } \partial_{\text{par}} Q_{h-1}. \end{cases}$$

Since (5.9) and (5.13) hold, Lemma 4.5 is at our disposal with choices of parameters  $\lambda \equiv \lambda_M \geq 1$ ,  $\delta \equiv \delta_1$ ,  $\theta \equiv \varepsilon/400$ ,  $Q_{\varrho}^{\lambda} \equiv Q_{h-1}$ , and  $\delta Q_{\varrho}^{\lambda} \equiv Q_h$ . Using also (5.17) we then have

$$\frac{\lambda_M}{B} \equiv \frac{\varepsilon \lambda_M}{100\sqrt{n}} \leq \sup_{Q_h} \|Dw_{h-1}\|, \quad s + \sup_{\frac{1}{2}Q_{h-1}} \|Dw_{h-1}\| \leq 6c_3 \lambda_M \equiv A \lambda_M.$$

Theorem 3.5 then gives

$$E(Dw_{h-1}, Q_h) = E(Dw_{h-1}, (\delta_\gamma/2)Q_{h-1}) \leq \left(\frac{\varepsilon}{2^{n+7}}\right) E(Dw_{h-1}, 2^{-1}Q_{h-1})$$

and hence Lemma 4.6, together with (5.9) and (5.13), implies

$$\begin{aligned} E(Du, Q_h) &\leq \frac{\varepsilon}{4} E(Du, Q_{h-1}) + 4c_4\delta_1^{-(n+2)} \left[ \frac{|\mu|(Q_{h-1})}{r_{h-1}^{N-1}} \right] \\ &\leq \frac{\varepsilon}{2}\lambda_M + \frac{\varepsilon\lambda_M}{100} \leq \varepsilon\lambda_M. \end{aligned} \tag{5.18}$$

This completes the proof of (5.15). Now, since the reasoning is independent of the choice of  $(x_0, t_0) \in \tilde{Q}$  and of the initial radius  $r \in (\delta_1 R, R]$  chosen to build the chain in (5.14), we obtain (5.8) with

$$r_\varepsilon = \delta_1 R = \delta_\gamma R/2. \tag{5.19}$$

Indeed, let  $\varrho \leq \delta_1 R$ ; this means there exists an integer  $m \geq 1$  such that  $\delta_1^{m+1} R < \varrho \leq \delta_1^m R$ . Therefore we have  $\varrho = \delta_1^m r$  for some  $r \in (\delta_1 R, R]$  and (5.8) follows from (5.15). In order to obtain the required form in (5.7), it is sufficient to recall (5.19) together with (5.11) and the choice of  $B$  and  $\gamma$  made in (5.10). The proof of Lemma 5.1 is complete.

*Step 3: Proof of VMO-regularity.* Here we are going to finally prove (5.6). By the previous step we can find a new radius  $\tilde{r}_\varepsilon$  as in (5.7), depending, for the choice made here, only on  $n, p, \nu, L, R_0, \varepsilon$  and  $M$ , such that (5.8) holds with  $\varepsilon$  replaced by  $\lambda_M^{p-3} \varepsilon/2$ . Therefore, as  $Q_\varrho(x_0, t_0) \subset Q_\varrho^{\lambda_M}(x_0, t_0)$ , and using (3.3), we have that

$$\begin{aligned} &\int_{Q_\varrho(x_0, t_0)} |Du - (Du)_{Q_\varrho(x_0, t_0)}| dx dt \\ &\leq 2\lambda_M^{2-p} \int_{Q_\varrho^{\lambda_M}(x_0, t_0)} |Du - (Du)_{Q_\varrho^{\lambda_M}(x_0, t_0)}| dx dt \leq \varepsilon \end{aligned} \tag{5.20}$$

holds provided  $\varrho \leq \tilde{r}_\varepsilon$ . This finishes the proof. □

*Proof of Theorem 1.9.* The proof essentially revisits the one of Theorem 1.6, and makes essential use of Lemma 5.1; for this reason we shall adopt the notation introduced in the proof of Theorem 1.6. Our aim is to show that, for every cylinder  $\tilde{Q} \Subset \Omega_T$  as in Section 5.1, there exists a radius  $R_0 > 0$ , depending on  $n, p, \nu, L, \delta, M, c_D$ , an exponent  $\beta \in (0, 1)$ , depending only on  $n, p, \nu, L, \delta$ , but independent of  $M$ , and finally a constant  $c$ , depending on  $n, p, \nu, L, M, \delta, c_D$ , such that the decay estimate

$$E(Du, Q_\rho(x_0, t_0)) = \int_{Q_\rho(x_0, t_0)} |Du - (Du)_{Q_\rho(x_0, t_0)}| dx dt < c\rho^\beta \tag{5.21}$$

holds whenever  $\rho \leq R_0$  and  $(x_0, t_0) \in \tilde{Q}$ . In turn, the local Hölder continuity of  $Du$  in  $\Omega_T$  follows from a classical Campanato-type integral characterization of Hölder continuity due to Da Prato [7]. As mentioned, the main tool here is Lemma 5.1, and the dependence in (5.7), where we are now going to disclose the exact identity of  $R(\varepsilon)$  for a certain choice of  $\varepsilon$ . More precisely, we start taking  $\varepsilon = \rho$  with  $\rho \leq R_0$ , where  $R_0$  has been initially determined in (5.4). By recalling (5.11), verifying (5.12) amounts to take  $R$  such that

$$\sup_{0 < \varrho \leq \lambda_M^{(2-p)/2}} \sup_R \frac{|\mu|(Q_\varrho(x_0, t_0))}{\varrho^{N-1}} \leq \frac{\rho^{\frac{2n+4}{\alpha}+1}}{400c_4c_5^{n+2}\lambda_M^{(N-1)(2-p)/2}},$$

In turn, using (1.24) it is sufficient to verify

$$R^\delta \leq \frac{\rho^{\frac{2n+4}{\alpha}+1}}{400c_4c_5^{n+2}c_D\lambda_M^{N(2-p)/2}} \iff R \leq \frac{\rho^{1/\alpha_2}}{c_6}$$

with  $c_6 \geq 1$  which depends only on  $n, p, v, L, M, \delta, c_D$  while  $\alpha_2 \in (0, 1)$  is depending only on  $n, p, v, L, \delta$ , but neither on  $M$  or  $c_D$ . By further reducing the size of  $R_0$ , in a way that makes it depending only on  $n, p, v, L, M, \delta, c_D$ , we may assume that  $R \leq \rho^{1/\alpha_3}$  for  $\alpha_3 = 2\alpha_2$ . Using this relation in (5.8) we have that

$$\int_{Q_{\rho^{1/\alpha_3}}^{\lambda_M}(x_0, t_0)} |Du - (Du)_{Q_{\rho^{1/\alpha_3}}^{\lambda_M}(x_0, t_0)}| dx dt < \lambda_M \rho$$

whenever  $\rho \leq R_0 \equiv R_0(n, v, L, M, \delta, c_D)$  and  $\alpha_3$  depends only on  $n, p, v, L, \delta$ . Proceeding as in (5.20) we obtain

$$\int_{Q_{\rho^{1/\alpha_3}}(x_0, t_0)} |Du - (Du)_{Q_{\rho^{1/\alpha_3}}(x_0, t_0)}| dx dt < c\rho$$

for a new constant  $c$ , from which (5.21) follows, taking  $\beta = \alpha_3$  and yet a new constant  $c$ . □

*Proof of Theorem 1.5.* The proof upgrades the one given for Theorem 1.6. The idea is to prove the continuity of the gradient by showing that this is the uniform limit of a sequence of continuous maps.

*Step 1: The basic function sequence.* We shall keep here the notation introduced in Section 5.1. With  $(x_0, t_0) \in \tilde{Q}$ , consider the maps

$$(x_0, t_0) \rightarrow (Du)_{Q_\varrho^{\lambda_M}(x_0, t_0)}, \quad \varrho \leq R_0,$$

which are obviously continuous; the radius  $R_0$  has been determined in (5.4). The proof breaks now in several steps: first, we show that there exists a continuous map  $g: \tilde{Q} \rightarrow \mathbb{R}^n$  such that

$$(Du)_{Q_\varrho^{\lambda_M}(x_0, t_0)} \rightarrow g(x_0, t_0) \quad \text{as } \varrho \rightarrow 0. \tag{5.22}$$

Then, as on the other hand the convergence

$$(Du)_{Q_\varrho^{\lambda_M}(x_0, t_0)} \rightarrow Du(x_0, t_0) \quad \text{as } \varrho \rightarrow 0$$

holds at every point  $(x_0, t_0) \in \tilde{Q} \cap \mathcal{L}_{\lambda_M}$ , this will imply that (the precise representative of)  $Du$  is continuous in  $\tilde{Q}$ . Since  $\tilde{Q} \Subset \Omega_T$  is arbitrary, this will finally prove the theorem. In turn, in order to prove, *simultaneously* that (5.22) holds pointwise and then uniformly, it is sufficient to prove that, for every  $\varepsilon > 0$ , there exists a radius  $r_\varepsilon \leq R_0$ , independent of the point  $(x_0, t_0)$  considered, such that

$$|(Du)_{Q_\varrho^{\lambda_M}(x_0, t_0)} - (Du)_{Q_\rho^{\lambda_M}(x_0, t_0)}| \leq \varepsilon \lambda_M \quad \forall \varrho, \rho \in (0, r_\varepsilon] \quad (5.23)$$

and  $(x_0, t_0) \in \tilde{Q}$ . In fact, for every fixed  $(x_0, t_0) \in \tilde{Q}$ , the net

$$\varrho \rightarrow (Du)_{Q_\rho^{\lambda_M}(x_0, t_0)}$$

is a Cauchy one, and this allows to define the function  $g(\cdot)$  appearing in (5.22). Then, keeping  $\varrho$  fixed and letting  $\rho \rightarrow 0$  in (5.23) gives

$$|(Du)_{Q_\varrho^{\lambda_M}(x_0, t_0)} - g(x_0, t_0)| \leq \varepsilon \lambda_M \quad \forall \varrho \in (0, r_\varepsilon]$$

and this means that the convergence in (5.22) is uniform. The rest of the proof is now dedicated to show the validity of (5.23).

*Step 2: Smallness conditions.* With  $\varepsilon > 0$  fixed in (5.23), we choose the number  $\delta_\gamma \equiv \delta_\gamma(n, p, \nu, L, \varepsilon) \in (0, 1/2)$  in Theorem 3.5 corresponding to the choice of parameters

$$\lambda \equiv \lambda_M, \quad A \equiv 6c_3, \quad B \equiv \frac{100\sqrt{n}}{\varepsilon}, \quad \gamma \equiv \frac{1}{2^{n+7}}, \quad q = 1,$$

where  $c_3 \equiv c_3(n, p, \nu, L)$  is the constant fixed in Theorem 4.2. Set  $\delta_1 := \delta_\gamma/2$ , while in the following  $c_4 \equiv c_4(n, p, \nu, L)$  is the constant introduced in Corollary 4.4. Next, we take a positive radius  $R \leq R_0$  such that

$$\sup_{(x_0, t_0)} \int_0^{2\lambda_M^{(2-p)/2}R} \frac{|\mu|(Q_\varrho(x_0, t_0))}{\varrho^{N-1}} \frac{d\varrho}{\varrho} \leq \frac{\delta_1^{4(n+2)} \varepsilon}{800c_4\lambda_M^{(2-p)(N-1)/2}}, \quad (5.24)$$

$$\sup_{0 < \varrho \leq \lambda_M^{(2-p)/2}} \sup_{(x_0, t_0)} \frac{|\mu|(Q_\varrho(x_0, t_0))}{\varrho^{N-1}} \leq \frac{\delta_1^{n+2} \varepsilon}{800c_4\lambda_M^{(N-1)(2-p)/2}}, \quad (5.25)$$

and

$$\sup_{0 < \varrho \leq R} \sup_{(x_0, t_0) \in \tilde{Q}} E(Du, Q_\varrho^{\lambda_M}(x_0, t_0)) \leq \frac{\delta_1^{4(n+2)} \varepsilon}{800}. \quad (5.26)$$

Let us observe that (5.24)-(5.25) are possible since now (1.20) is in force. Finally, (5.26) is allowed by Theorem 1.6 as its assumptions are in turn again verified by (1.20); in particular observe that (5.26) follows from Step 3 of the proof of Theorem 1.6. Notice that exactly as in (5.13) from (5.25) it follows that

$$\sup_{0 < \varrho \leq R} \sup_{(x_0, t_0)} \frac{|\mu|(Q_\varrho^{\lambda_M}(x_0, t_0))}{\varrho^{N-1}} \leq \frac{\delta_1^{n+2} \varepsilon}{800c_4} \lambda_M. \tag{5.27}$$

We shall eventually show that the radius  $R$  determined by the smallness conditions (5.24)-(5.26) will work as  $r_\varepsilon$  in (5.23).

*Step 3: Preliminary iteration step.* With  $(x_0, t_0) \in \tilde{Q}$ , we again define the chain of shrinking intrinsic cylinders

$$Q_i \equiv Q_{r_i^M}^{\lambda_M}(x_0, t_0), \quad r_i = \delta_1^i R, \quad i \geq 0. \tag{5.28}$$

We then have the following:

**Lemma 5.2.** *Assume that*

$$\int_{Q_{i+1}} |Du| dx dt \geq \frac{\varepsilon \lambda_M}{50} \quad \text{and} \quad \frac{|\mu|(Q_i)}{r_i^{N-1}} \leq \frac{\delta_1^{n+2} \varepsilon}{400c_4} \lambda_M. \tag{5.29}$$

*Then it holds that*

$$E(Du, Q_{i+1}) \leq \frac{1}{2} E(Du, Q_i) + 4c_4 \delta_1^{-(n+2)} \left[ \frac{|\mu|(Q_i)}{r_i^{N-1}} \right]. \tag{5.30}$$

*Proof.* Let  $w_i$  be the comparison solution introduced at the beginning of the Section 4.1 corresponding to the cylinder  $Q_i$ , i.e.  $w_i$  solves the Cauchy-Dirichlet problem

$$\begin{cases} (w_i)_t - \operatorname{div} a(Dw_i) = 0 & \text{in } Q_i \\ w_i = u & \text{on } \partial_{\text{par}} Q_i. \end{cases}$$

Since (5.9) and (5.13) hold, Lemma 4.5 is at our disposal with choices of parameters  $\lambda \equiv \lambda_M \geq 1$ ,  $\delta \equiv \delta_1$ ,  $\theta \equiv \varepsilon/400$ ,  $Q_\varrho^\lambda \equiv Q_i$ , and  $\delta Q_\varrho^\lambda \equiv Q_{i+1}$ . Using also the first inequality in (5.29) we then have

$$\frac{\lambda_M}{B} \equiv \frac{\varepsilon \lambda_M}{100\sqrt{n}} \leq \sup_{Q_{i+1}} \|Dw_i\|, \quad s + \sup_{\frac{1}{2}Q_i} \|Dw_i\| \leq 6c_3 \lambda_M \equiv A \lambda_M.$$

Theorem 3.5 then gives

$$E(Dw_i, Q_{i+1}) = E(Dw_i, (\delta_\gamma/2)Q_i) \leq 2^{-(n+7)} E(Dw_i, 2^{-1}Q_i)$$

and hence Lemma 4.6 - used this time with  $\varepsilon = 1$  - together with (5.9) and (5.25), implies (5.30). This completes the proof of the lemma.  $\square$

*Step 4: Exit times set and iteration chains.* The main fact here is the following:

**Lemma 5.3.** *It holds that*

$$|(Du)_{Q_h} - (Du)_{Q_k}| \leq \frac{\varepsilon\lambda_M}{10} \tag{5.31}$$

whenever  $0 \leq k \leq h$ .

*Proof.* For the proof we need some terminology. Given a chain  $\{Q_i\}$  of geometrically shrinking intrinsic cylinders as in (5.14), we consider the set  $L$  defined by

$$L := \left\{ i \in \mathbb{N} : \int_{Q_i} |Du| dx dt < \frac{\varepsilon\lambda_M}{50} \right\}. \tag{5.32}$$

Accordingly, we define the “left and right boundaries” of  $L$  as follows:

$$\partial_{\text{left}}L := \{i \in L : i - 1 \notin L\}, \quad \partial_{\text{right}}L := \{i \in L : i + 1 \notin L\} \cup \{\infty\}. \tag{5.33}$$

We then define a set of the type

$$C_i^m = \{j \in \mathbb{N} : i \leq j \leq i + m, i \in \partial_{\text{right}}L, i + m + 1 \in \partial_{\text{left}}L, j \notin L \text{ if } j > i\}$$

as the *maximal iteration chain* of length  $m$ , starting at  $i$ . In other words we have  $C_i^m = \{i, \dots, i + m\}$  and each element of  $C_i^m$  but  $i$  lies outside of  $L$ ; moreover  $C_i^m$  is maximal in the sense that there cannot be another set of the same type properly containing it. Obviously, such sets do not exist when  $L = \mathbb{N}$ . In the same way we define the *the infinite maximal chain* starting at  $i$  as

$$C_i^\infty = \{j \in \mathbb{N} : i \leq j < \infty, i \in \partial_{\text{right}}L, j \notin L \text{ if } j > i\}.$$

Notice that, in every case, the smallest element of such a chain always belongs to  $L$ , being then the only one of the chain to have such a property. Moreover, we define

$$i_e := \min L. \tag{5.34}$$

Note that we set  $i_e = \infty$  if  $L = \emptyset$ . We are now ready for the proof of (5.31); for this we need to distinguish three cases. We shall of course assume  $0 \leq k < h$ .

*Case 1:*  $k < h \leq i_e$ . Keeping (5.27) in mind, notice that if  $h - 1 > k$ , then we can apply Lemma 5.2 repeatedly, and this yields

$$E(Du, Q_{i+1}) \leq \frac{1}{2}E(Du, Q_i) + 4c_4\delta_1^{-(n+2)} \left[ \frac{|\mu|(Q_i)}{r_i^{N-1}} \right] \tag{5.35}$$

for every  $i \in \{k, \dots, h - 2\}$ . Summing up the previous inequalities, and making obvious manipulations – see (4.46)-(4.47) – we have

$$\begin{aligned} \sum_{i=k}^{h-1} E(Du, Q_i) &\leq 2E(Du, Q_k) + 8c_4\delta_1^{-(n+2)} \sum_{i=k}^{h-2} \frac{|\mu|(Q_i)}{r_i^{N-1}} \\ &\leq \frac{\delta_1^{4(n+2)}\varepsilon}{400} + 8c_4\delta_1^{-(n+2)} \sum_{i=k}^{h-2} \frac{|\mu|(Q_i)}{r_i^{N-1}}, \end{aligned} \tag{5.36}$$

where we have used (5.26) in the last line. By recalling that

$$\begin{aligned} \sum_{i=0}^{\infty} \frac{|\mu|(Q_i(x_0, t_0))}{r_i^{N-1}} &\leq \sum_{i=0}^{\infty} \frac{|\mu|\left(Q_{\lambda_M^{(2-p)/2}r_i}(x_0, t_0)\right)}{r_i^{N-1}} \\ &\leq \lambda_M^{(2-p)(N-1)/2} \delta_1^{-(n+2)} \int_0^{2\lambda_M^{(2-p)/2}R} \frac{|\mu|(Q_\varrho(x_0, t_0))}{\varrho^{N-1}} \frac{d\varrho}{\varrho} \\ &\leq \frac{\delta_1^{3(n+2)}\varepsilon}{800c_4}, \end{aligned}$$

see for instance (4.31) above, and using directly (5.26) for the case  $h - 1 = k$ , we conclude that in any case (*i.e.*  $h - 1 \geq k$ ) it holds

$$\sum_{i=k}^{h-1} E(Du, Q_i) \leq \frac{\delta_1^{2(n+2)}\varepsilon}{50}.$$

In turn, (5.31) follows since

$$\begin{aligned} |(Du)_{Q_h} - (Du)_{Q_k}| &\leq \sum_{i=k}^{h-1} |(Du)_{Q_{i+1}} - (Du)_{Q_i}| \\ &\leq \sum_{i=k}^{h-1} \int_{Q_{i+1}} |Du - (Du)_{Q_i}| dx dt \\ &\leq \sum_{i=k}^{h-1} \frac{|Q_i|}{|Q_{i+1}|} E(Du, Q_i) \\ &= \delta_1^{-(n+2)} \sum_{i=k}^{h-1} E(Du, Q_i) \leq \frac{\varepsilon}{50} \leq \frac{\varepsilon\lambda_M}{50}. \end{aligned} \tag{5.37}$$

Notice that the case analyzed here includes the one when the index  $i_e$  defined in (5.34) is infinite, *i.e.* the set  $L$  is empty.

Case 2:  $i_e \leq k < h$ . Let us prove that in this case we have

$$|(Du)_{Q_h}| \leq \frac{\varepsilon\lambda_M}{25} \quad \text{and} \quad |(Du)_{Q_k}| \leq \frac{\varepsilon\lambda_M}{25}. \tag{5.38}$$



We prove the former inequality in (5.38), the proof of the latter being the same. If  $h \in L$ , the first inequality in (5.38) follows immediately from the definition of  $L$ . On the other hand, if  $h \notin L$ , then, as  $h \geq i_e$ , it is possible to consider the maximal iteration chain  $\mathcal{C}_{i_h}^{m_h}$  such that  $h \in \mathcal{C}_{i_h}^{m_h}$ ; notice that  $h > i_h$  as  $h \notin L \ni i_h$ . Then iterating Lemma 5.2 as done after (5.35) - *i.e.* replacing  $k$  by  $i_h$  - we gain the analogue of (5.37), that is  $|(Du)_{Q_h} - (Du)_{Q_{i_h}}| \leq \varepsilon/50$ . In turn using that  $|(Du)_{Q_{i_h}}| \leq \varepsilon\lambda_M/50$  as  $i_h \in L$ , we again obtain the first inequality in (5.38) and in any case (5.38) follows. Estimating as

$$|(Du)_{Q_h} - (Du)_{Q_k}| \leq |(Du)_{Q_h}| + |(Du)_{Q_k}| \leq \frac{\varepsilon\lambda_M}{25} + \frac{\varepsilon\lambda_M}{25} \leq \frac{\varepsilon\lambda_M}{10}$$

we have that (5.31) holds in the second case too.

*Case 3:*  $k < i_e < h$ . Here we prove that (5.38) still holds and then we conclude as in Step 2. Indeed, the first inequality in (5.38) follows as in Case 2. As for the second estimate in (5.38), let us remark that, as  $i_e \in L$ , we have that

$$|(Du)_{Q_{i_e}}| \leq \frac{\varepsilon\lambda_M}{50}. \tag{5.39}$$

On the other hand, we can argue exactly as in Case 1, *i.e.* this time replacing  $h$  by  $i_e$ , thereby obtaining  $|(Du)_{Q_{i_e}} - (Du)_{Q_k}| \leq \varepsilon/50$  that together with (5.39) gives the second inequality in (5.38). In turn, (5.31) follows also in this case. The proof of the lemma is complete.  $\square$

*Step 5: Proof of (5.23).* The proof of (5.23) follows using Lemma 5.3 together with the already proved VMO-regularity of the gradient, that is (5.26). We actually take  $r_\varepsilon = R$  and fix  $0 < \rho < \varrho \leq R$ . This means there exists two integers,  $0 \leq k \leq h$  such that

$$\delta_1^{k+1} R < \varrho \leq \delta_1^k R \quad \text{and} \quad \delta_1^{h+1} R < \rho \leq \delta_1^h R. \tag{5.40}$$

Observe that

$$\begin{aligned} |(Du)_{Q_\varrho^{\lambda_M}(x_0,t_0)} - (Du)_{Q_{k+1}}| &\leq \int_{Q_{k+1}} |Du - (Du)_{Q_\varrho^{\lambda_M}(x_0,t_0)}| dx dt \\ &\leq \frac{|Q_\varrho^{\lambda_M}(x_0, t_0)|}{|Q_{k+1}|} \int_{Q_\varrho^{\lambda_M}(x_0,t_0)} |Du - (Du)_{Q_\varrho^{\lambda_M}(x_0,t_0)}| dx dt \\ &\leq \delta_1^{-(n+2)} E(Du, Q_\varrho^{\lambda_M}(x_0, t_0)) \leq \frac{\varepsilon}{10} \end{aligned}$$

where in the last line we have used (5.26) and in second-last one we have used (5.40). In the same way we also obtain

$$|(Du)_{Q_\rho^{\lambda_M}(x_0,t_0)} - (Du)_{Q_{h+1}}| \leq \frac{\varepsilon}{10}.$$

Using the last two inequalities together with Lemma 5.3 we conclude with (5.23), and the proof is complete.  $\square$

*Proof of Theorem 1.7.* The proof of this theorem can be now obtained by adapting the methods of proof of Theorems 1.5-1.6 to the case  $p \geq 2$  according to the scheme introduced in [30]. In particular, notice that in the case  $p \geq 2$ , Theorems 3.5 and 3.3 – in slightly different versions – have been proved in [30].  $\square$

*Proof of Corollary 1.8.* By Lemma 2.1, and in particular by (2.9) it follows that

$$\lim_{r \rightarrow 0} \sup_{(x,t) \in Q} \mathbf{W}_q^\mu(x, t; r) = 0, \quad \text{where } q = \min\{1, p/[2(p-1)]\}.$$

Recall that the quantity  $\mathbf{W}_q^\mu$  has been defined in (2.6). The proof now follows appealing to Theorems 1.5 and 1.7 and recalling (2.7).  $\square$

*Proof of Corollary 1.10.* By (1.25) it follows that

$$\lim_{r \rightarrow 0} \sup_{(x,t) \in Q} \mathbf{W}_q^\mu(x, t; r) \leq c \lim_{r \rightarrow 0} \int_0^r [h(\varrho)]^q \frac{d\varrho}{\varrho} = 0,$$

with  $q = \min\{1, p/[2(p-1)]\}$ . Once again the proof follows appealing to Theorems 1.5 and 1.7.  $\square$

## References

- [1] E. ACERBI and G. MINGIONE, *Gradient estimates for a class of parabolic systems*, Duke Math. J. **136** (2007), 285–320.
- [2] G. I. BARENBLATT, *On self-similar motions of a compressible fluid in a porous medium*, Akad. Nauk SSSR. Prikl. Mat. Meh. **16** (1952), 679–698.
- [3] P. BARONI, *Regularity in parabolic Dini-continuous systems*, Forum Math. **23** (2011), 1281–1322.
- [4] L. BOCCARDO, *Elliptic and parabolic differential problems with measure data*, Boll. Un. Mat. Ital. A (7) **11** (1997), 439–461.
- [5] L. BOCCARDO and T. GALLOUËT, *Nonlinear elliptic and parabolic equations involving measure data*, J. Funct. Anal. **87** (1989), 149–169.
- [6] L. BOCCARDO, A. DALL’AGLIO, T. GALLOUËT and L. ORSINA, *Nonlinear parabolic equations with measure data*, J. Funct. Anal. **147** (1997), 237–258.
- [7] G. DA PRATO, *Spazi  $\mathcal{L}^{p,\theta}(\Omega, \delta)$  e loro proprietà*, Ann. Mat. Pura Appl. (4) **69** (1965), 383–392.
- [8] E. DiBENEDETTO,  *$C^{1+\alpha}$  local regularity of weak solutions of degenerate elliptic equations*, Nonlinear Anal. **7** (1983), 827–850.
- [9] E. DiBENEDETTO, *On the local behaviour of solutions of degenerate parabolic equations with measurable coefficients*, Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4) **13** (1986), 487–535.
- [10] E. DiBENEDETTO, *Intrinsic Harnack type inequalities for solutions of certain degenerate parabolic equations*, Arch. Ration. Mech. Anal. **100** (1988), 129–147.
- [11] E. DiBENEDETTO, “Degenerate Parabolic Equations”, Universitext, Springer-Verlag, New York, 1993.
- [12] E. DiBENEDETTO and A. FRIEDMAN, *Hölder estimates for nonlinear degenerate parabolic systems*, J. Reine Angew. Math. **357** (1985), 1–22.

- [13] E. DiBENEDETTO, U. GIANAZZA and V. VESPRI, *Alternative forms of the Harnack inequality for non-negative solutions to certain degenerate and singular parabolic equations*, Atti Accad. Naz. Lincei, Cl. Sci. Fis. Mat. Natur., Rend. Lincei (9) Mat. Appl. **20** (2009), 369–377.
- [14] E. DiBENEDETTO, U. GIANAZZA and V. VESPRI, *Forward, backward and elliptic Harnack inequalities for non-negative solutions to certain singular parabolic partial differential equations*, Ann. Sc. Norm. Super. Pisa Cl. Sci. (5) **9** (2010), 385–422.
- [15] F. DUZAAR and G. MINGIONE, *Gradient estimates via non-linear potentials*, Amer. J. Math. **133** (2011), 1093–1149.
- [16] F. DUZAAR and G. MINGIONE, *Gradient estimates via linear and nonlinear potentials*, J. Funct. Anal. **259** (2010), 2961–2998.
- [17] F. DUZAAR and G. MINGIONE, *Gradient continuity estimates*, Calc. Var. Partial Differential Equations **39** (2010), 379–418.
- [18] R. A. HUNT, *On  $L(p, q)$  spaces*, Einsegnement Math. (II) **12** (1966), 249–276.
- [19] T. KILPELÄINEN, *Singular solutions to  $p$ -Laplacian type equations*, Ark. Mat. **37** (1999), 275–289.
- [20] T. KILPELÄINEN and J. MALÝ, *The Wiener test and potential estimates for quasilinear elliptic equations*, Acta Math. **172** (1994), 137–161.
- [21] J. KINNUNEN and P. LINDQVIST, *Summability of semicontinuous supersolutions to a quasilinear parabolic equation*, Ann. Sc. Norm. Super. Pisa Cl. Sci. (5) **4** (2005), 59–78.
- [22] J. KINNUNEN and P. LINDQVIST, *Definition and properties of supersolutions to the porous medium equation*, J. Reine Angew. Math. **618** (2008), 135–168.
- [23] J. KINNUNEN, T. LUKKARI and M. PARVIAINEN, *An existence result for superparabolic functions*, J. Funct. Anal. **258** (2010), 713–728.
- [24] J. KINNUNEN and M. PARVIAINEN, *Stability for degenerate parabolic equations*, Adv. Calc. Var. **3** (2010), 29–48.
- [25] R. KORTE and T. KUUSI, *A note on the Wolff potential estimate for solutions to elliptic equations involving measures*, Adv. Calc. Var. **3** (2010), 99–113.
- [26] T. KUUSI, *Harnack estimates for weak supersolutions to nonlinear degenerate parabolic equations*, Ann. Sc. Norm. Super. Pisa Cl. Sci. (5) **7** (2008), 673–716.
- [27] T. KUUSI and G. MINGIONE, *Potential estimates and gradient boundedness for nonlinear parabolic systems*, Rev. Mat. Iberoam. **28** (2012), 535–576.
- [28] T. KUUSI and G. MINGIONE, *Nonlinear potential estimates in parabolic problems*, Atti Accad. Naz. Lincei, Cl. Sci. Fis. Mat. Natur., Rend. Lincei (9) Mat. Appl. **22** (2011), 161–174.
- [29] T. KUUSI and G. MINGIONE, *A surprising linear type estimate for nonlinear elliptic equations*, C. R. Math. Acad. Sci. Paris **349** (2011), 889–892.
- [30] T. KUUSI and G. MINGIONE, *The Wolff gradient bound for degenerate parabolic equations*, J. Eur. Math. Soc. (JENS), to appear.
- [31] T. KUUSI and G. MINGIONE, *New perturbation methods for nonlinear parabolic problems*, J. Math. Pures Appl.(9) **98** (2012), 390–427.
- [32] T. KUUSI and G. MINGIONE, *Linear potentials in nonlinear potential theory*, Arch. Ration. Mech. Anal. **207** (2013), 215–246.
- [33] T. KUUSI and M. PARVIAINEN, *Existence for a degenerate Cauchy problem*, Manuscripta Math. **128** (2009), 213–249.
- [34] G. M. LIEBERMAN, *Boundary regularity for solutions of degenerate parabolic equations*, Nonlinear Anal. **14** (1990), 501–524.
- [35] G. M. LIEBERMAN, *Sharp forms of estimates for subsolutions and supersolutions of quasilinear elliptic equations involving measures*, Comm. Partial Differential Equations **18** (1993), 1191–1212.
- [36] G. M. LIEBERMAN, “Second Order Parabolic Differential Equations”, World Scientific press, River Edge, 1996.

- [37] P. LINDQVIST, *On the definition and properties of  $p$ -superharmonic functions*, J. Reine Angew. Math. **365** (1986), 67–79.
- [38] P. LINDQVIST, *Notes on the  $p$ -Laplace equation*, Univ. Jyväskylä, Report **102**, 2006.
- [39] G. MINGIONE, *The Calderón-Zygmund theory for elliptic problems with measure data*, Ann. Sc. Norm. Super. Pisa Cl. Sci. (5) **6** (2007), 195–261.
- [40] G. MINGIONE, *Gradient estimates below the duality exponent*, Math. Ann. **346** (2010), 571–627.
- [41] G. MINGIONE, *Gradient potential estimates*, J. Eur. Math. Soc. (JENS) **13** (2011), 459–486.
- [42] J. MOSER, *A Harnack inequality for parabolic differential equations*, Comm. Pure Appl. Math. **17** (1964), 101–134. Corrections in: Comm. Pure Appl. Math. **20** (1967), 231–236.
- [43] E. M. STEIN, “Harmonic Analysis: Real-Variable Methods, Orthogonality, and Oscillatory Integrals”. Princeton Math. Series, 43, Princeton University Press, Princeton, NJ, 1993.
- [44] E. M. STEIN and G. WEISS, “Introduction to Fourier Analysis on Euclidean Spaces”, Princeton Math. Ser., 32, Princeton Univ. Press, Princeton, NJ 1971.
- [45] N. S. TRUDINGER and X. J. WANG, *On the weak continuity of elliptic operators and applications to potential theory*, Amer. J. Math. **124** (2002), 369–410.
- [46] N. S. TRUDINGER and X. J. WANG, *Quasilinear elliptic equations with signed measure data*, Discrete Contin. Dyn. Syst. **23** (2009), 477–494.
- [47] J. L. VÁZQUEZ, “Smoothing and Decay Estimates for Nonlinear Diffusion Equations, Equations of Porous Medium Type”, Oxford Lecture Series in Math. Appl., 33. Oxford University Press, Oxford, 2006, xiv+234.

Aalto University  
Institute of Mathematics  
P.O. Box 11100  
FI-00076 Aalto, Finland  
tuomo.kuusi@tkk.fi

Dipartimento di Matematica e Informatica  
Università di Parma  
Parco Area delle Scienze 53/a  
Campus  
43124 Parma, Italia  
giuseppe.mingione@math.unipr.it