# GRADIENT ESTIMATES FOR
# AN ORTHOTROPIC NONLINEAR DIFFUSION EQUATION

PIERRE BOUSQUET, LORENZO BRASCO, CHIARA LEONE, AND ANNA VERDE

*In memory of Emmanuele DiBenedetto, with esteem and admiration*

ABSTRACT. We consider a quasilinear degenerate parabolic equation driven by the orthotropic $p-$Laplacian. We prove that local weak solutions are locally Lipschitz continuous in the spatial variable, uniformly in time.

## CONTENTS

## 1. INTRODUCTION

1.1. **Aim of the paper.** Let $\Omega \subset \mathbb{R}^N$ be an open bounded set and $I \subset \mathbb{R}$ an open bounded interval. We study the gradient regularity of local weak solutions to the following parabolic equation

$$(1.1) \qquad u_t = \sum_{i=1}^{N} \left( |u_{x_i}|^{p-2}\, u_{x_i} \right)_{x_i}, \qquad \text{in } I \times \Omega.$$

Evolution equations of this type have been studied since the 60's of the previous century, especially by the Soviet school, see for example the paper [27] by Vishik. Equation (1.1) also explicitely appears in the monographs [21], [23, Example 4.A, Chapter III] and [29, Example 30.8], among others.

In this paper, we will focus on the case $p \geq 2$. We first observe that (1.1) looks quite similar to the more familiar one

$$(1.2) \qquad u_t = \Delta_p u, \qquad \text{in } I \times \Omega,$$

which involves the $p-$Laplace operator

$$\Delta_p u = \sum_{i=1}^{N} \left( |\nabla u|^{p-2}\, u_{x_i} \right)_{x_i}.$$

Indeed, both parabolic equations are particular instances of equations of the type

$$u_t = \operatorname{div} \nabla F(\nabla u)$$

with $F : \mathbb{R}^N \to \mathbb{R}$ a convex function which satisfies the structural conditions

$$\langle \nabla F(z), z \rangle \geq \frac{1}{C} \, |z|^p \qquad \text{and} \qquad |\nabla F(z)| \leq C \, |z|^{p-1}, \qquad \text{for every } z \in \mathbb{R}^N.$$

Then, the basic regularity theory equally applies to both (1.1) and (1.2). The standard reference in the field is DiBenedetto's monograph [12], where one can find boundedness results for the solution $u$ (see [12, Chapter V]), Hölder continuous estimates for $u$ (see [12, Chapter III]), as well as Harnack inequality for positive solutions (see [12, Chapter VI]). At a technical level, there is no distinction to be made between (1.1) and (1.2).

In contrast, when coming to the regularity of $\nabla u$ (i.e. boundedness and continuity), the situation becomes fairly more complicated. Let us start from (1.2). DiBenedetto and Friedman [14] have proved that the gradients of solutions to this equation are bounded. This is the starting point to obtain the continuity of the gradients for any

$$p > \frac{2N}{N+2}.$$

We refer again to DiBenedetto's book for a comprehensive collection of results on the subject, notably to [12, Chapter VIII]. Since then, there has been a growing literature concerning the regularity for nonlinear, possibly degenerate or singular, parabolic equations (or systems), the main model of which is given by the evolutionary $p-$Laplacian equation (1.2). Without any attempt to completeness, we can just mention some classical references [8, 9, 13, 15, 28], up to the most recent contributions on the subject, given by [2, 18, 19], among others.

However, none of these results apply to our equation (1.1). Indeed, all of them rely on the fact that the loss of ellipticity of the operator $\operatorname{div} \nabla F$ is restricted to a single point, since the Hessian $D^2 F$ behaves as in the model case (1.2)

$$\langle D^2 F(z) \, \xi, \xi \rangle \geq \frac{1}{C} \, |z|^{p-2} \, |\xi|^2,$$

where the elliptic character is lost only for $z = 0$. Such a property dramatically breaks down for our equation (1.1). Indeed, in this case, the function $F$ has the following *orthotropic structure*

$$(1.3) \qquad\qquad F(z) = \frac{1}{p} \sum_{i=1}^{N} |z_i|^p, \qquad \text{for every } z \in \mathbb{R}^N.$$

The Hessian matrix of $F$ now *degenerates on an unbounded set*, namely the set of those $z \in \mathbb{R}^N$ such that one component $z_i$ is 0. As a consequence, the aforementioned references do not provide any regularity results for the gradients of the solutions.

The main goal of the present paper is to prove the $L^\infty$ bound on $\nabla u$ for our equation (1.1), thus extending the result by DiBenedetto and Friedman to this more degenerate setting. In order to do this, we will need to adapt to the parabolic setting the machinery that we developed in [3, 4, 5, 6] and [7], for degenerate *equations with orthotropic structure*. Indeed, the operator

$$\sum_{i=1}^{N} \left( |u_{x_i}|^{p-2} \, u_{x_i} \right)_{x_i},$$

that we called *orthotropic $p-$Laplacian*, is the prominent example of this kind of equations. We also refer to [11] for an approach to the regularity theory for this operator, based on viscosity techniques.

1.2. **Main result.** In this paper, we establish the following regularity result which can be seen as the parabolic counterpart of our previous result [5, Theorem 1.1] for the elliptic case. In the statement below, the notation $\nabla u$ refers to the spatial variables, i.e. $\nabla u = (u_{x_1}, \ldots, u_{x_N})$.

**Main Theorem.** *Let $p > 2$ and let $u \in L^p_{\mathrm{loc}}(I; W^{1,p}_{\mathrm{loc}}(\Omega))$ be a local weak solution of (1.1). Then $\nabla u \in L^\infty_{\mathrm{loc}}(I \times \Omega)$. More precisely, for every parabolic cube*

$$\mathcal{Q}_{\tau,R}(t_0, x_0) := (t_0 - \tau, t_0) \times (x_0 - R, x_0 + R)^N \Subset I \times \Omega,$$

*and every $0 < \sigma < 1$, we have*

$$
(1.4) \qquad
\begin{aligned}
\|\nabla u\|_{L^\infty(\mathcal{Q}_{\sigma\tau,\sigma R}(t_0,x_0))} &\leq C\, \frac{1}{(1-\sigma)^{\frac{N+2}{2}}} \left(\frac{\tau}{R^2}\right)^{\frac{1}{2}} \left(\fint_{\mathcal{Q}_{\tau,R}(t_0,x_0)} |\nabla u|^p \, dt\, dx\right)^{\frac{1}{2}} \\
&\quad + C\left((1-\sigma)\frac{R^2}{\tau}\right)^{\frac{1}{p-2}},
\end{aligned}
$$

*for a constant $C = C(N, p) > 0$.*

**Remark 1.1** (Scalings)**.** We observe that the equation (1.1) is invariant with respect to the "horizontal" and "vertical" scale changes

$$u_{\lambda,\mu}(t, x) = \mu\, u(\mu^{p-2}\, \lambda^p\, t, \lambda\, x),$$

for every $\lambda, \mu > 0$. Then it is easily seen that the a priori estimate (1.4) is invariant with respect to these scale changes. We point out that such estimate is the exact analogue of that for the evolutionary $p-$Laplacian, see [12, Theorem 5.1, Chapter VIII]. Occasionally, in the paper we will work with *anisotropic* parabolic cubes of the type

$$Q_R(t_0, x_0) = (t_0 - R^p, t_0) \times (x_0 - R, x_0 + R)^N.$$

The choice of cubes of this type could be loosely justified by a dimensional analysis of the equation. Indeed, by considering the quantity $u$ as dimensionless and using the family of scalings

$$(t, x) \mapsto (\lambda^p\, t, \lambda\, x), \qquad \text{for every } \lambda > 0,$$

we get the relation

$$\text{time} \sim (\text{length})^p.$$

However, as it is well-known, estimates on cubes of the type $Q_R$ are too restrictive when looking at $C^{0,\alpha}$ estimates for $\nabla u$. Indeed, in light of the so-called *intrinsic geometry*, it is much more important to work with local estimates on cubes $\mathcal{Q}_{\tau,R}$, where the time scale $\tau$ is adapted to the solution itself: roughly speaking, we can take

$$\tau \sim R^2\, |\nabla u|^{p-2}.$$

This explains the importance of having (1.4) with two *independent* scales $R$ and $\tau$. We refer to [12, Chapter VIII] for a description of the method of intrinsic scalings, where these heuristics are clarified.

**Remark 1.2** (Case $1 < p \leq 2$)**.** When $p = 2$, the orthotropic parabolic equation (1.1) boils down to the standard heat equation, for which solutions are well-known to be smooth. For this reason, in our statement we restrict our attention to the case $p > 2$. However, we point out that by making the choice $\tau = R^2$ and taking the limit as $p$ goes to 2 in (1.4), we formally end up with the classical gradient estimate for solutions of the heat equation

$$\|\nabla u\|_{L^\infty(\mathcal{Q}_{\sigma R^2,\sigma R}(t_0,x_0))} \leq \frac{C}{(1-\sigma)^{\frac{N+2}{2}}} \left(\fint_{\mathcal{Q}_{R^2,R}(t_0,x_0)} |\nabla u|^2 \, dt\, dx\right)^{\frac{1}{2}}.$$

In light of the previous remark, in the case $p = 2$ the relation

$$\text{time} \sim (\text{length})^2,$$

is now the natural one.

As for the singular case $p < 2$, this is somehow simpler than its degenerate counterpart. In this case, the local Lipschitz regularity of solutions to (1.1) can be directly inferred from [22, Theorem 1], under the restriction

$$p > \frac{2N}{N+2}.$$

Indeed, the result of [22] covers (among others) the case of parabolic equations of the form

$$u_t = \operatorname{div} \nabla F(\nabla u),$$

under the following assumptions on the convex function $F$

$$|\nabla F(z)| \le C\, |z|^{p-1} \qquad \text{and} \qquad \langle D^2 F(z)\, \xi, \xi \rangle \ge \frac{1}{C}\, |z|^{p-2}\, |\xi|^2, \qquad \text{for every } \xi \in \mathbb{R}^N, z \in \mathbb{R}^N \setminus \{0\}.$$

It is not difficult to see that the orthotropic function (1.3) for $p < 2$ matches both requirements. Indeed, observe that

$$\langle D^2 F(z)\, \xi, \xi \rangle = (p-1) \sum_{i=1}^N |z_i|^{p-2}\, |\xi|^2 \ge (p-1)\, |z|^{p-2}\, |\xi|^2,$$

thanks to the fact that $p - 2 < 0$. Thus in the subquadratic case, the orthotropic structure helps more than it hurts, in a sense.

**Remark 1.3** (Anisotropic diffusion). We conclude this part by observing that, more generally, one could consider the following parabolic equation

$$u_t = \sum_{i=1}^N \left( |u_{x_i}|^{p_i - 2}\, u_{x_i} \right)_{x_i}, \qquad \text{in } \Omega \times I,$$

which still has an orthotropic structure. Now we have a whole set of exponents $1 < p_1 \le p_2 \le \dots p_N$, one for each coordinate direction. We cite the paper [24], where some *global* Lipschitz regularity results are proven for solutions of the relevant Cauchy-Dirichlet problem, under appropriate regularity assumptions on the data. We point out that in light of their global nature, for $p_1 = \dots = p_N = p > 2$ such results are not comparable to ours. We also refer to [10] for a sophisticated Harnack inequality for positive local weak solutions, as well as for some further references on the problem. Finally, the very recent paper [16] contains a thorough study of the Cauchy problem in the case $p_i < 2$, together with some regularity results.

However, as for the counterpart of our Main Theorem for *local solutions* of this equation, this is still an open problem, to the best of our knowledge.

1.3. **Technical aspects of the proof.** The core of the proof of the Main Theorem is an a priori Lipschitz estimate for smooth solutions of the orthotropic parabolic equation, see Proposition 4.1 below. More precisely, we introduce the regularized problem

$$(u_\varepsilon)_t = \operatorname{div} \nabla F_\varepsilon(\nabla u_\varepsilon),$$

where $F_\varepsilon$ is a smooth uniformly convex approximation of the orthotropic function (1.3). By the classical regularity theory, the maps $u_\varepsilon$ are regular enough to justify all the calculations below. The goal is to establish a local uniform Lipschitz estimate on $u_\varepsilon$, which does not depend on the regularization parameter $\varepsilon$. Finally, we let $\varepsilon$ go to 0 and prove that the family $u_\varepsilon$ converges to the original solution $u$. This allows to obtain the Lipschitz estimate for $u$ itself.

In the subsequent part of this section, we emphasize the main difficulties to get such a Lipschitz estimate on $u_\varepsilon$. In order to simplify the presentation, we drop the index $\varepsilon$ both for $F_\varepsilon$ and $u_\varepsilon$. The strategy is apparently quite classical: we rely on a Moser iterative scheme of reverse Hölder's inequalities, resulting from the interplay between Caccioppoli estimates and the Sobolev embeddings.

To be more specific, we first differentiate the equation with respect to a spatial variable $x_j$, so as to get the equation solved by the $j-$th component of the gradient. This is given by

$$(1.5) \qquad \iint_{I \times \Omega} (u_{x_j})_t \, \varphi \, dt \, dx + \iint_{I \times \Omega} \langle (\nabla F(\nabla u))_{x_j}, \nabla \varphi \rangle \, dt \, dx = 0, \qquad \text{for every } \varphi \in C_0^\infty(I \times \Omega).$$

More generally, the composition of the component $u_{x_j}$ with a non-negative convex function $h$ is a subsolution of this equation. Accordingly, the map $h(u_{x_j})$ satisfies the Caccioppoli inequality which is naturally attached to (1.5) (see Lemma 3.1 below). If $I = (T_0, T_1)$ and $\tau \in (T_0, T_1)$, this reads as follows:

$$
(1.6)
\begin{aligned}
\chi(\tau) \int_{\{\tau\} \times \Omega} h^2(u_{x_j}) \, \eta^2 \, dx &+ \iint_{(T_0, \tau) \times \Omega} \langle D^2 F(\nabla u) \nabla h(u_{x_j}), \nabla h(u_{x_j}) \rangle \, \chi \, \eta^2 \, dt \, dx \\
&\lesssim \iint_{(T_0, \tau) \times \Omega} \chi' \, \eta^2 \, h^2(u_{x_j}) \, dt \, dx \\
&+ \iint_{(T_0, \tau) \times \Omega} \langle D^2 F(\nabla u) \nabla \eta, \nabla \eta \rangle \, h^2(u_{x_j}) \, \chi \, dt \, dx.
\end{aligned}
$$

Here, the maps $\chi \in C_0^\infty(I)$ and $\eta \in C_0^\infty(\Omega)$ are non-negative cut-off functions in the time and space variables, respectively. We have used the following expedient notation: given $f \in L^1(I \times \Omega)$,

$$\int_{\{\tau\} \times \Omega} f \, dx := \int_\Omega f(\tau, x) \, dx, \qquad \text{for a. e. } \tau \in I.$$

When $F$ is a uniformly elliptic integrand, in the sense that

$$\frac{1}{C} \, |\xi|^2 \le \langle D^2 F(\nabla u) \, \xi, \xi \rangle \le C \, |\xi|^2, \qquad \text{for every } \xi \in \mathbb{R}^N,$$

one can easily obtain from (1.6) a crucial "unnatural" feature of the subsolution $h(u_{x_j})$; that is, a sort of reverse Poincaré inequality where the Sobolev norm of $h(u_{x_j})$ is controlled by the $L^2$ norm of the subsolution itself. In conjunction with the Sobolev inequality, this is the cornerstone which eventually leads to the classical version of the Moser iterative scheme. It should be noticed that this strategy still works even in the degenerate case, provided the Hessian behaves like

$$\frac{1}{C} \, |\nabla u|^{p-2} \, |\xi|^2 \le \langle D^2 F(\nabla u) \, \xi, \xi \rangle \le C \, |\nabla u|^{p-2} \, |\xi|^2, \qquad \text{for every } \xi \in \mathbb{R}^N,$$

as in the case of the evolutionary $p-$Laplace equation (1.2). It is sufficient to use the "absorption of degeneracy" trick, where the degenerate weight $|\nabla u|^{p-2}$ is recombined with the subsolution $h(u_{x_j})$ by means of simple algebraic manipulations. This still permits to infer from (1.6) a control on the Sobolev norm of a suitable convex function of $u_{x_j}$. This is nowadays a standard technique in the field; for the elliptic case, it goes back to the pioneering works by Ural'tseva [26] and Uhlenbeck [25].

As we explained above, due to the severe degeneracy of $D^2 F$ in our orthotropic situation, *it is not possible* to follow the same path. In order to rely on such an absorption trick, we have to go through a *tour de force* and to introduce a new family of *weird* Caccioppoli inequalities (see Lemma 3.2 below). These are the parabolic counterparts of a corresponding estimate introduced in the elliptic setting in [3] and then fruitfully exploited in [5].

The crucial idea is to mix together the components of the gradient with respect to 2 orthogonal directions. This compensates the lack of ellipticity of $D^2 F$ and allows to rely on the Sobolev embeddings in the iterative scheme. We do not detail these Caccioppoli-type estimates here, but instead explain the main additional difficulties with respect to the elliptic framework.

Let us come back for one instant to the standard Caccioppoli inequality (1.6). It follows from (1.5) by taking $\varphi = h \, h'(u_{x_j}) \, \chi \, \eta^2$. In particular, the parabolic term is given by

$$\iint_{I \times \Omega} (u_{x_j})_t \, \varphi \, dt \, dx = \frac{1}{2} \iint_{I \times \Omega} (h^2(u_{x_j}))_t \, \chi \, \eta^2 \, dt \, dx.$$

Then an integration by parts yields

$$\iint_{I\times\Omega} (u_{x_j})_t \, \varphi \, dt \, dx = -\frac{1}{2} \iint_{I\times\Omega} h^2(u_{x_j}) \, \chi' \, \eta^2 \, dt \, dx.$$

The latter yields the "time slice" term on the left-hand side of (1.6). This way, the time derivative is transferred to $\chi$ and one can handle the factor $h^2(u_{x_j})$ as in the elliptic framework.

For the weird Caccioppoli inequalities, the test function is now $\varphi = u_{x_j} \, \Phi'(u_{x_j}^2) \, \Psi(u_{x_k}^2) \, \chi \, \eta^2$, for some $1 \le j, k \le N$. The corresponding parabolic term becomes:

$$\iint_{I\times\Omega} (u_{x_j})_t \, \varphi \, dt \, dx = \frac{1}{2} \iint_{I\times\Omega} \left(\Phi(u_{x_j}^2)\right)_t \, \Psi(u_{x_k}^2) \, \chi \, \eta^2 \, dt \, dx.$$

In contrast to the previous situation, we cannot perform an integration by parts to get rid of the time derivative on $\Phi(u_{x_j}^2)$, since it would affect the factor $\Psi(u_{x_k}^2)$. In order to overcome this difficulty, which does not arise in the elliptic setting, we need a new approach, aimed at "symmetrizing" the above quantity containing $u_{x_j}$ and $u_{x_k}$.

Basically, we merge together two weird Caccioppoli inequalities, where the spatial variables $x_j$ and $x_k$ play symmetric roles. More specifically, we insert into (1.5) the test functions:

$$\varphi = u_{x_j} \, \Phi'(u_{x_j}^2) \, \Psi(u_{x_k}^2) \, \chi \, \eta^2, \qquad \widetilde{\varphi} = u_{x_k} \, \Psi'(u_{x_k}^2) \, \Phi(u_{x_j}^2) \, \chi \, \eta^2,$$

and then add the two resulting inequalities. The parabolic term is now replaced by the following quantity:

$$\frac{1}{2} \iint_{I\times\Omega} \left(\Phi(u_{x_j}^2) \, \Psi(u_{x_k}^2)\right)_t \, \chi \, \eta^2 \, dt \, dx.$$

This allows to integrate by parts and transfer the time derivative on the test function. It turns out that by a suitable adaptation of the arguments that we used in the elliptic case, one can incorporate this new term in the iterative Moser scheme. This finally leads to the desired local $L^\infty$ estimate on $\nabla u$.

1.4. **Plan of the paper.** The paper is organized as follows: after collecting the basic terminology and some preliminaries on Steklov averages in Section 2, we present in Section 3 the proofs of the new Caccioppoli inequalities in the parabolic setting. We detail the iterative Moser scheme in Section 4 and finally establish the Main Theorem in Section 5, by transferring to the original solution $u$ the a priori estimates obtained on the approximating solutions $u_\varepsilon$.

## 2. Preliminaries

2.1. **Local solutions.** Let $\Omega \subset \mathbb{R}^N$ be an open bounded set and $I \subset \mathbb{R}$ an open bounded interval. Fix $p > 2$ and take $\mathcal{A} : \mathbb{R}^N \to \mathbb{R}^N$ a continuous function such that

$$\langle \mathcal{A}(z) - \mathcal{A}(w), z - w \rangle \ge 0, \qquad \text{for every } z, w \in \mathbb{R}^N,$$

and

$$\langle \mathcal{A}(z), z \rangle \ge \frac{1}{C} \, |z|^p, \qquad \text{and} \qquad |\mathcal{A}(z)| \le C \, |z|^{p-1}, \qquad \text{for every } z \in \mathbb{R}^N.$$

We say that $u \in L^p_{\mathrm{loc}}(I; W^{1,p}_{\mathrm{loc}}(\Omega))$ is a *local weak solution* of the quasilinear diffusion equation

$$(2.1) \qquad\qquad\qquad u_t = \mathrm{div}\, \mathcal{A}(\nabla u), \qquad \text{in } I \times \Omega,$$

if for every $\varphi \in C_0^\infty(I \times \Omega)$, we have

$$-\iint_{I \times \Omega} u \, \varphi_t \, dt \, dx + \iint_{I \times \Omega} \langle \mathcal{A}(\nabla u), \nabla \varphi \rangle \, dt \, dx = 0.$$

2.2. **Steklov averages.** Throughout the paper, we denote by $T_0 < T_1$ the endpoints of the time interval $I$. Let $v \in L^1_{\text{loc}}(I \times \Omega)$. For every $0 < \sigma < T_1 - T_0$, we define its so-called

- *forward Steklov average*

$$v_\sigma^+(t, x) = \fint_t^{t+\sigma} v(\tau, x) \, d\tau, \qquad \text{for every } (t, x) \in (T_0, T_1 - \sigma) \times \Omega;$$

- *backward Steklov average*

$$v_\sigma^-(t, x) = \fint_{t-\sigma}^t v(\tau, x) \, d\tau, \qquad \text{for every } (t, x) \in (T_0 + \sigma, T_1) \times \Omega.$$

We shall use some standard properties of the Steklov averages. Let $0 < \sigma < T_1 - T_0$ and $\psi \in L^\infty(I \times \Omega)$ such that $\psi$ is compactly supported in $(T_0, T_1 - \sigma) \times \Omega$. We extend $\psi$ by 0 on $(\mathbb{R} \setminus I) \times \Omega$, so that $\psi_\sigma^-$ is well-defined on $I \times \Omega$ and compactly supported therein. By the Fubini theorem, we have

$$(2.2) \qquad \iint_{(T_0, T_1 - \sigma) \times \Omega} v_\sigma^+ \, \psi \, dt \, dx = \iint_{I \times \Omega} v \, \psi_\sigma^- \, dt \, dx.$$

Moreover, if $v \in L^q_{\text{loc}}(I \times \Omega)$ for some $1 \le q < \infty$, then $v_\sigma^+$ converges to $v$ in $L^q_{\text{loc}}(I \times \Omega)$, as $\sigma$ goes to 0, see e.g. [12, Chapter I, Lemma 3.2].

Finally, we can derive from (2.2) the following regularity properties of the Steklov averages:

**Lemma 2.1.** *Let $v \in L^1_{\text{loc}}(I \times \Omega)$. Then for every $0 < \sigma < T_1 - T_0$,*

(1) *the map $v_\sigma^+$ belongs to $W^{1,1}_{\text{loc}}((T_0, T_1 - \sigma); L^1_{\text{loc}}(\Omega))$ and*

$$(2.3) \qquad (v_\sigma^+)_t(t, x) = \frac{v(t + \sigma, x) - v(t, x)}{\sigma}, \qquad \text{for a.e. } (t, x) \in (T_0, T_1 - \sigma) \times \Omega;$$

(2) *if one further assumes that $v \in L^1_{\text{loc}}(I; W^{1,1}_{\text{loc}}(\Omega))$, then $\nabla(v_\sigma^+) \in L^1_{\text{loc}}((T_0, T_1 - \sigma) \times \Omega)$ and*

$$(2.4) \qquad \nabla(v_\sigma^+) = (\nabla v)_\sigma^+.$$

*Proof.* Fix $0 < \sigma < T_1 - T_0$. Let $\psi \in C_0^\infty((T_0, T_1 - \sigma) \times \Omega)$. Then by (2.2),

$$\iint_{(T_0, T_1 - \sigma) \times \Omega} v_\sigma^+ \, \psi_t \, dt \, dx = \iint_{I \times \Omega} v \, (\psi_t)_\sigma^- \, dt \, dx = \iint_{I \times \Omega} v(t, x) \, \frac{\psi(t, x) - \psi(t - \sigma, x)}{\sigma} \, dt \, dx.$$

By an obvious change of variables, this yields

$$\iint_{(T_0, T_1 - \sigma) \times \Omega} v_\sigma^+ \, \psi_t \, dt \, dx = -\iint_{(T_0, T_1 - \sigma) \times \Omega} \frac{v(t + \sigma, x) - v(t, x)}{\sigma} \, \psi(t, x) \, dt \, dx,$$

which gives the desired identity (2.3).

In order to prove (2.4), we rely again on (2.2), this time tested with $\psi_{x_j}$ in place of $\psi$, for some $1 \le j \le N$:

$$\iint_{(T_0, T_1 - \sigma) \times \Omega} v_\sigma^+ \, \psi_{x_j} \, dt \, dx = \iint_{I \times \Omega} v \, (\psi_{x_j})_\sigma^- \, dt \, dx = \iint_{I \times \Omega} v \, (\psi_\sigma^-)_{x_j} \, dt \, dx.$$

In the last equality, we have derived under the integral sign the smooth function $\psi$. Hence, by integrating by parts the last integral and using (2.2) again, one gets

$$\iint_{I \times \Omega} v_\sigma^+ \, \psi_{x_j} \, dt \, dx = -\iint_{(T_0, T_1 - \sigma) \times \Omega} (v_{x_j})_\sigma^+ \psi \, dt \, dx,$$

from which (2.4) follows. $\qquad \square$

## 3. Energy estimates for a regularized equation

### 3.1. An approximating equation. We denote by

$$G(\xi) = \frac{1}{p} \left(1 + |\xi|^2\right)^{\frac{p}{2}}, \qquad \text{for every } \xi \in \mathbb{R}^N,$$

and for every $\varepsilon \in (0,1)$, we consider the convex function

$$(3.1) \qquad F_\varepsilon(\xi) = \frac{1}{p} \sum_{i=1}^{N} |\xi_i|^p + \varepsilon\, G(\xi), \qquad \text{for every } \xi \in \mathbb{R}^N.$$

We consider a local weak solution $u_\varepsilon \in L^p_{\mathrm{loc}}(I; W^{1,p}_{\mathrm{loc}}(\Omega))$ of the equation (2.1) with the choice

$$\mathcal{A}(z) = \nabla F_\varepsilon(z).$$

This means that $u_\varepsilon$ verifies

$$(3.2) \qquad -\iint_{I \times \Omega} u_\varepsilon\, \varphi_t \, dt\, dx + \iint_{I \times \Omega} \langle \nabla F_\varepsilon(\nabla u_\varepsilon), \nabla \varphi \rangle \, dt\, dx = 0,$$

for every $\varphi \in C_0^\infty(I \times \Omega)$. Observe that the map $F_\varepsilon$ belongs to $C^2(\mathbb{R}^N)$ and satisfies

$$\varepsilon \left(1 + |\xi|^2\right)^{\frac{p-2}{2}} |\zeta|^2 \le \langle D^2 F_\varepsilon(\xi)\, \zeta, \zeta \rangle \le (1+\varepsilon)\,(p-1)\left(1 + |\xi|^2\right)^{\frac{p-2}{2}} |\zeta|^2, \qquad \text{for every } \xi, \zeta \in \mathbb{R}^N.$$

Hence, one can rely on the classical regularity theory for quasilinear parabolic equations, see e.g. [12, Theorem 5.1, Chapter VIII] and [1, Lemma 3.1], to get:

$$(3.3) \qquad \nabla u_\varepsilon \in L^\infty_{\mathrm{loc}}(I \times \Omega) \qquad \text{and} \qquad u_\varepsilon \in L^2_{\mathrm{loc}}(I; W^{2,2}_{\mathrm{loc}}(\Omega)).$$

In the following computations, we delete the index $\varepsilon$ both for $u$ and $F$.

### 3.2. An equation for the spatial gradient. In order to establish a Lipschitz bound on our solution $u$, we need to differentiate (3.2) with respect to the spatial variables $x_j$, $1 \le j \le N$.

Fix $0 < \sigma < T_1 - T_0$. Let $\psi \in C_0^\infty((T_0, T_1 - \sigma) \times \Omega)$. As already observed, the backward Steklov average

$$\varphi(t,x) = \psi_\sigma^-(t,x), \qquad \text{for } (t,x) \in I \times \Omega,$$

is compactly supported in $I \times \Omega$. We can thus insert it in (3.2):

$$-\iint_{I \times \Omega} u\,(\psi_\sigma^-)_t \, dt\, dx + \iint_{I \times \Omega} \langle \nabla F(\nabla u), \nabla \psi_\sigma^- \rangle \, dt\, dx = 0.$$

Since $(\psi_\sigma^-)_t = (\psi_t)_\sigma^-$, (2.2) implies that

$$-\iint_{I \times \Omega} u\,(\psi_\sigma^-)_t \, dt\, dx = -\iint_{(T_0, T_1 - \sigma) \times \Omega} u_\sigma^+\, \psi_t \, dt\, dx = \iint_{(T_0, T_1 - \sigma) \times \Omega} (u_\sigma^+)_t\, \psi \, dt\, dx.$$

One thus gets

$$(3.4) \qquad \iint_{(T_0, T_1 - \sigma) \times \Omega} (u_\sigma^+)_t\, \psi \, dt\, dx + \iint_{I \times \Omega} \langle \nabla F(\nabla u), \nabla \psi_\sigma^- \rangle \, dt\, dx = 0,$$

for every $\psi \in C_0^\infty((T_0, T_1 - \sigma) \times \Omega)$.

Let $j \in \{1, \ldots, N\}$ and $\varphi \in C_0^\infty((T_0, T_1 - \sigma) \times \Omega)$. The map

$$(u_\sigma^+)_t(t,x) = \frac{u(t+\sigma, x) - u(t,x)}{\sigma},$$

belongs to $L^p_{\mathrm{loc}}((T_0, T_1 - \sigma); W^{1,p}_{\mathrm{loc}}(\Omega))$ and $((u_\sigma^+)_t)_{x_j} = ((u_{x_j})_\sigma^+)_t$. By derivation under the integral sign, one also has

$$\nabla((\varphi_{x_j})_\sigma^-) = (\nabla(\varphi_\sigma^-))_{x_j}.$$

We insert $\psi = \varphi_{x_j}$ in equation (3.4). An integration by parts in the spatial variable leads to

$$\iint_{(T_0,T_1-\sigma)\times\Omega} \left((u_{x_j})_\sigma^+\right)_t \varphi\, dt\, dx + \iint_{I\times\Omega} \langle(\nabla F(\nabla u))_{x_j}, \nabla\varphi_\sigma^-\rangle\, dt\, dx = 0.$$

Finally, using (2.2) in the second term, one gets

$$(3.5) \qquad \iint_{(T_0,T_1-\sigma)\times\Omega} \left((u_{x_j})_\sigma^+\right)_t \varphi\, dt\, dx + \iint_{(T_0,T_1-\sigma)\times\Omega} \langle((\nabla F(\nabla u))_{x_j})_\sigma^+, \nabla\varphi\rangle\, dt\, dx = 0.$$

We observe that, since $F \in C^2(\mathbb{R}^N)$ and $\nabla u \in L^\infty_{\mathrm{loc}}(I\times\Omega) \cap L^2_{\mathrm{loc}}(I; W^{1,2}_{\mathrm{loc}}(\Omega))$, one has

$$\nabla F(\nabla u) \in L^2_{\mathrm{loc}}((0,T); W^{1,2}_{\mathrm{loc}}(\Omega)).$$

We can thus appeal to a density argument to get that (3.5) remains true for every $\varphi \in L^2(I; W^{1,2}(\Omega))$, with compact support in $(T_0, T_1 - \sigma) \times \Omega$.

3.3. **Caccioppoli-type inequalities.** As explained in the introduction, the first technical tool in the proof of the Lipschitz bound of $u$ is the following Caccioppoli inequality which provides a $W^{1,2}$ estimate on $h(u_{x_j})$, where $h$ is any smooth convex function.

**Lemma 3.1** (Standard Caccioppoli inequality). *Let $\eta \in C_0^\infty(\Omega)$ and $\chi \in C_0^\infty((T_0, T_1])$ be two non-negative functions, with $\chi$ non-decreasing. Let $h : \mathbb{R} \to \mathbb{R}$ be a $C^1$ convex non-negative function. Then, for almost every $\tau \in I$ and every $j = 1, ..., N$, we have*

$$\chi(\tau) \int_{\{\tau\}\times\Omega} h^2(u_{x_j})\, \eta^2\, dx + \iint_{(T_0,\tau)\times\Omega} \langle D^2 F(\nabla u)\, \nabla h(u_{x_j}), \nabla h(u_{x_j})\rangle\, \chi\, \eta^2\, dt\, dx$$

$$(3.6) \qquad\qquad\qquad\qquad \leq \iint_{(T_0,\tau)\times\Omega} \chi'\, \eta^2\, h^2(u_{x_j})\, dt\, dx$$

$$\qquad\qquad\qquad\qquad + 4 \iint_{(T_0,\tau)\times\Omega} \langle D^2 F(\nabla u)\, \nabla\eta, \nabla\eta\rangle\, h^2(u_{x_j})\, \chi\, dt\, dx.$$

*Proof.* We first assume that $h$ is a $C^2$ convex non-negative function. Let $\zeta \in C_0^\infty(I)$ and $\eta \in C_0^\infty(\Omega)$. There exists $0 < \sigma_1 < (T_1 - T_0)/2$ such that $\zeta$ is compactly supported in $(T_0 + \sigma_1, T_1 - \sigma_1)$. Given $0 < \sigma < \sigma_1$, Lemma 2.1 and (3.3) imply that $(u_{x_j})_\sigma^+ \in W^{1,1}_{\mathrm{loc}}((T_0, T_1 - \sigma); L^1_{\mathrm{loc}}(\Omega)) \cap L^\infty_{\mathrm{loc}}((T_0, T_1 - \sigma) \times \Omega)$. Hence, the map $h^2((u_{x_j})_\sigma^+)$ belongs to $W^{1,1}_{\mathrm{loc}}((T_0, T_1 - \sigma); L^1_{\mathrm{loc}}(\Omega))$ and we have

$$(3.7) \qquad \frac{1}{2}\left(h^2\left((u_{x_j})_\sigma^+\right)\right)_t = (h\, h')\left((u_{x_j})_\sigma^+\right)\left((u_{x_j})_\sigma^+\right)_t.$$

We insert in (3.5) the test function

$$\varphi = (h\, h')\left((u_{x_j})_\sigma^+\right)\zeta\, \eta^2,$$

which has compact support in $(T_0 + \sigma_1, T_1 - \sigma_1) \times \Omega$ and belongs to $L^\infty((T_0 + \sigma_1, T_1 - \sigma_1) \times \Omega) \cap L^2((T_0 + \sigma_1, T_1 - \sigma_1); W^{1,2}(\Omega))$. By (3.7),

$$\left((u_{x_j})_\sigma^+\right)_t \varphi = \frac{1}{2}\left(h^2\left((u_{x_j})_\sigma^+\right)\right)_t \zeta\, \eta^2.$$

We use the above identity to infer:

$$\frac{1}{2}\iint_{(T_0+\sigma_1,T_1-\sigma_1)\times\Omega} \left(h^2\left((u_{x_j})_\sigma^+\right)\right)_t \zeta\, \eta^2\, dt\, dx + \iint_{(T_0+\sigma_1,T_1-\sigma_1)\times\Omega} \langle((\nabla F(\nabla u))_{x_j})_\sigma^+, \nabla\varphi\rangle\, dt\, dx = 0.$$

We then perform an integration by parts with respect to the time variable in the first term:

$$-\frac{1}{2}\iint_{(T_0+\sigma_1,T_1-\sigma_1)\times\Omega} h^2\left((u_{x_j})_\sigma^+\right)\zeta'\eta^2\, dt\, dx + \iint_{(T_0+\sigma_1,T_1-\sigma_1)\times\Omega} \langle((\nabla F(\nabla u))_{x_j})_\sigma^+, \nabla\varphi\rangle\, dt\, dx = 0.$$

We now want to take the limit as $\sigma$ goes to 0. Let $\Omega_1 \Subset \Omega$ such that $\eta$ is compactly supported in $\Omega_1$. Since $u_{x_j} \in L^2_{\mathrm{loc}}(I \times \Omega)$, we have

$$\lim_{\sigma \to 0^+} \| \left(u_{x_j}\right)^+_\sigma - u_{x_j}\|_{L^2((T_0+\sigma_1, T_1-\sigma_1) \times \Omega_1)} = 0.$$

Moreover, we know that $u_{x_j} \in L^\infty_{\mathrm{loc}}(I \times \Omega)$ which guarantees that there exists $C_1 > 0$ such that for every $\sigma \in (0, \sigma_1)$,

$$\left| \left(u_{x_j}\right)^+_\sigma \right| \le C_1, \qquad \text{a. e. on } (T_0 + \sigma_1, T_1 - \sigma_1) \times \Omega_1.$$

It then follows from the Dominated Convergence Theorem that

(3.8) $$\lim_{\sigma \to 0^+} -\frac{1}{2} \iint_{(T_0+\sigma_1, T_1-\sigma_1) \times \Omega} h^2 \left( \left(u_{x_j}\right)^+_\sigma \right) \zeta' \eta^2 \, dt \, dx = -\frac{1}{2} \iint_{I \times \Omega} h^2 \left(u_{x_j}\right) \zeta' \eta^2 \, dt \, dx.$$

Next, by recalling the choice of $\varphi$ above, we have

$$\nabla \varphi = (h \, h')' \left( (u_{x_j})^+_\sigma \right) \nabla \left( (u_{x_j})^+_\sigma \right) \zeta \, \eta^2 + (h \, h') \left( (u_{x_j})^+_\sigma \right) \zeta \, \nabla(\eta^2).$$

By Lemma 2.1, we know that $\nabla \left( (u_{x_j})^+_\sigma \right) = (\nabla u_{x_j})^+_\sigma$. This implies that $\nabla \left( (u_{x_j})^+_\sigma \right)$ converges to $\nabla u_{x_j}$ in $L^2((T_0 + \sigma_1, T_1 - \sigma_1) \times \Omega)$. Hence, a similar argument to the one leading to (3.8) implies that

$$\lim_{\sigma \to 0^+} \left\| \nabla \varphi - \nabla \left( (h \, h')(u_{x_j}) \, \zeta \, \eta^2 \right) \right\|_{L^2((T_0+\sigma_1, T_1-\sigma_1) \times \Omega)} = 0.$$

Finally, by using that $(\nabla F(\nabla u))_{x_j} \in L^2_{\mathrm{loc}}(I \times \Omega)$, we can infer that

$$\lim_{\sigma \to 0^+} \left\| \left( (\nabla F(\nabla u))_{x_j} \right)^+_\sigma - (\nabla F(\nabla u))_{x_j} \right\|_{L^2((T_0+\sigma_1, T_1-\sigma_1) \times \Omega_1)} = 0.$$

It follows that

$$\lim_{\sigma \to 0^+} \iint_{(T_0+\sigma_1, T_1-\sigma_1) \times \Omega} \left\langle \left( (\nabla F(\nabla u))_{x_j} \right)^+_\sigma, \nabla \varphi \right\rangle dt \, dx = \iint_{I \times \Omega} \left\langle (\nabla F(\nabla u))_{x_j}, \nabla \left( (h \, h')(u_{x_j}) \, \zeta \, \eta^2 \right) \right\rangle dt \, dx.$$

Up to now, we have thus proved:

(3.9) $$-\frac{1}{2} \iint_{I \times \Omega} h^2 \left(u_{x_j}\right) \zeta' \, \eta^2 \, dt \, dx + \iint_{I \times \Omega} \left\langle (\nabla F(\nabla u))_{x_j}, \nabla \left( (h \, h')(u_{x_j}) \, \zeta \, \eta^2 \right) \right\rangle dt \, dx = 0.$$

We now choose $\zeta$ as follows. Let $\chi \in C^\infty_0((T_0, T_1])$ be as in the statement. Given $\tau \in I$ and $\delta > 0$ such that $T_0 < \tau < \tau + \delta < T_1$, we define

$$\widetilde{\chi}_\delta(t) := \begin{cases} 1, & \text{if } t \le \tau, \\ 1 - \dfrac{t - \tau}{\delta}, & \text{if } \tau < t < \tau + \delta, \\ 0, & \text{if } t \ge \tau + \delta. \end{cases}$$

We then insert

(3.10) $$\zeta(t) = \widetilde{\chi}_\delta(t) \, \chi(t),$$

in (3.9). Then, for almost every $\tau \in I$, we can let $\delta$ go to 0 and obtain

(3.11) $$\frac{\chi(\tau)}{2} \int_{\{\tau\} \times \Omega} h^2(u_{x_j}) \, \eta^2 \, dx + \iint_{(T_0, \tau) \times \Omega} \left\langle (\nabla F(\nabla u))_{x_j}, \nabla \left( (h \, h')(u_{x_j}) \, \chi \, \eta^2 \right) \right\rangle dt \, dx$$
$$= \frac{1}{2} \iint_{(T_0, \tau) \times \Omega} \chi' \, \eta^2 \, h^2(u_{x_j}) \, dt \, dx.$$

Since $\chi$ does not depend on the spatial variable, we have

$$\begin{aligned} \left\langle (\nabla F(\nabla u))_{x_j}, \nabla \left( (h h')(u_{x_j}) \, \chi \, \eta^2 \right) \right\rangle &= \left\langle D^2 F(\nabla u) \, \nabla h(u_{x_j}), \nabla h(u_{x_j}) \right\rangle \chi \, \eta^2 \\ &\quad + \left\langle D^2 F(\nabla u) \, \nabla u_{x_j}, \nabla u_{x_j} \right\rangle h''(u_{x_j}) \, h(u_{x_j}) \, \chi \, \eta^2 \\ &\quad + 2 \left\langle D^2 F(\nabla u) \, \nabla u_{x_j}, \nabla \eta \right\rangle (h \, h')(u_{x_j}) \, \chi \, \eta. \end{aligned}$$

Since the second term is non-negative, by dropping it, we get from (3.11)

$$\frac{\chi(\tau)}{2}\int_{\{\tau\}\times\Omega} h^2(u_{x_j})\,\eta^2\,dx + \iint_{(T_0,\tau)\times\Omega}\langle D^2F(\nabla u)\,\nabla h(u_{x_j}),\nabla h(u_{x_j})\rangle\,\chi\,\eta^2\,dt\,dx$$

$$\leq \frac{1}{2}\iint_{(T_0,\tau)\times\Omega}\chi'\,\eta^2\,h^2(u_{x_j})\,dt\,dx$$

$$-2\iint_{(T_0,\tau)\times\Omega}\langle D^2F(\nabla u)\,\nabla u_{x_j},\nabla\eta\rangle\,(h\,h')(u_{x_j})\,\chi\,\eta\,dt\,dx.$$

In order to estimate the last term, we use the Cauchy-Schwarz inequality:

$$\left|\langle D^2F(\nabla u)\,\nabla u_{x_j},\nabla\eta\rangle\right| \leq \left(\langle D^2F(\nabla u)\,\nabla u_{x_j},\nabla u_{x_j}\rangle\right)^{\frac{1}{2}}\left(\langle D^2F(\nabla u)\,\nabla\eta,\nabla\eta\rangle\right)^{\frac{1}{2}}.$$

A further application of Young inequality leads to

$$\left|-2\iint_{(T_0,\tau)\times\Omega}\langle D^2F(\nabla u)\,\nabla u_{x_j},\nabla\eta\rangle\,(h\,h')(u_{x_j})\,\chi\,\eta\,dt\,dx\right|$$

$$\leq \frac{1}{2}\iint_{(T_0,\tau)\times\Omega}\langle D^2F(\nabla u)\nabla u_{x_j},\nabla u_{x_j}\rangle\,h'(u_{x_j})^2\,\chi\,\eta^2\,dt\,dx$$

$$+2\iint_{(T_0,\tau)\times\Omega}\langle D^2F(\nabla u)\nabla\eta,\nabla\eta\rangle\,h^2(u_{x_j})\,\chi\,dt\,dx.$$

In this way, the integral containing $\nabla u_{x_j}$ can be absorbed on the left-hand side. Let us finally observe that we can remove the $C^2$ assumption on the function $h$, by a standard approximation argument. $\qquad\square$

We next establish the key tool for the proof of our main result, namely a Caccioppoli type inequality, where two different partial derivatives $u_{x_j}$ and $u_{x_k}$ come into play.

**Lemma 3.2** (Weird Caccioppoli inequality). *Let $\eta\in C_0^\infty(\Omega)$ and $\chi\in C_0^\infty((T_0,T_1])$ be two non-negative functions, with $\chi$ non-decreasing. Let $\Phi:\mathbb{R}^+\to\mathbb{R}$ and $\Psi:\mathbb{R}^+\to\mathbb{R}$ be two $C^1$ non-decreasing and non-negative convex functions. Then, for almost every $\tau\in I$, every $k,j=1,...,N$ and every $\theta\in[0,1]$, we have*

$$\chi(\tau)\int_{\{\tau\}\times\Omega}\Phi(u_{x_j}^2)\,\Psi(u_{x_k}^2)\,\eta^2\,dx + \iint_{(T_0,\tau)\times\Omega}\langle D^2F(\nabla u)\nabla u_{x_j},\nabla u_{x_j}\rangle\,\Phi'(u_{x_j}^2)\,\Psi(u_{x_k}^2)\,\chi\,\eta^2\,dt\,dx$$

$$\leq \iint_{(T_0,\tau)\times\Omega}\chi'\,\eta^2\,\Phi(u_{x_j}^2)\,\Psi(u_{x_k}^2)\,dt\,dx$$

$$+4\iint_{(T_0,\tau)\times\Omega}\langle D^2F(\nabla u)\nabla\eta,\nabla\eta\rangle\left(u_{x_j}^2\,\Phi'(u_{x_j}^2)\,\Psi(u_{x_k}^2)+u_{x_k}^2\,\Psi'(u_{x_k}^2)\,\Phi(u_{x_j}^2)\right)\chi\,dt\,dx$$

$$+8\left(\iint_{(T_0,\tau)\times\Omega}\langle D^2F(\nabla u)\nabla u_{x_j},\nabla u_{x_j}\rangle\,u_{x_j}^2\left(\Phi'(u_{x_j}^2)\right)^2\left(\Psi'(u_{x_k}^2)\right)^\theta\chi\,\eta^2\,dt\,dx\right)^{\frac{1}{2}}$$

$$\times\left(\frac{1}{4}\iint_{(T_0,\tau)\times\Omega}\chi'\,\eta^2\,|u_{x_k}|^{2\theta}\left(\Psi(u_{x_k}^2)\right)^{2-\theta}dt\,dx\right.$$

$$\left.+\iint_{(T_0,\tau)\times\Omega}\langle D^2F(\nabla u)\nabla\eta,\nabla\eta\rangle\,|u_{x_k}|^{2\theta}\left(\Psi(u_{x_k}^2)\right)^{2-\theta}\chi\,dt\,dx\right)^{\frac{1}{2}}.$$

*Proof.* It is convenient to divide the proof into two steps.

**Step 1: an identity involving $u_{x_j}$ and $u_{x_k}$.** We first assume that $\Phi$ and $\Psi$ are two $C^2$ non-decreasing and non-negative convex functions. We fix $k,j\in\{1,\dots,N\}$. Given $0<\sigma_1<(T_1-T_0)/2$ and $\zeta\in$

$C_0^\infty(T_0 + \sigma_1, T_1 - \sigma_1)$, we consider (3.5) with the index $j$ and for every $0 < \sigma < \sigma_1$, we insert the test function

$$\varphi = \left(u_\sigma^+\right)_{x_j} \Phi' \left(\left(\left(u_\sigma^+\right)_{x_j}\right)^2\right) \Psi \left(\left(\left(u_\sigma^+\right)_{x_k}\right)^2\right) \zeta \eta^2.$$

Symmetrically, we consider (3.5) with the index $k$ and insert the test function

$$\widetilde\varphi = \left(u_\sigma^+\right)_{x_k} \Psi' \left(\left(\left(u_\sigma^+\right)_{x_k}\right)^2\right) \Phi \left(\left(\left(u_\sigma^+\right)_{x_j}\right)^2\right) \zeta \eta^2.$$

The functions $\varphi$ and $\widetilde\varphi$ are compactly supported in $(T_0 + \sigma_1, T_1 - \sigma_1) \times \Omega$ and belong to $L^\infty((T_0 + \sigma_1, T_1 - \sigma_1) \times \Omega) \cap L^2((T_0 + \sigma_1, T_1 - \sigma_1); W^{1,2}(\Omega))$. Thus they are admissible test functions. We observe that

$$\left((u_\sigma^+)_{x_j}\right)_t \varphi = \frac{1}{2} \left(\Phi\left(\left(\left(u_\sigma^+\right)_{x_j}\right)^2\right)\right)_t \Psi\left(\left(\left(u_\sigma^+\right)_{x_k}\right)^2\right) \zeta \eta^2,$$

and similarly

$$\left((u_\sigma^+)_{x_k}\right)_t \widetilde\varphi = \frac{1}{2} \left(\Psi\left(\left(\left(u_\sigma^+\right)_{x_k}\right)^2\right)\right)_t \Phi\left(\left(\left(u_\sigma^+\right)_{x_j}\right)^2\right) \zeta \eta^2.$$

Thus we obtain

$$(3.12) \qquad \left((u_\sigma^+)_{x_j}\right)_t \varphi + \left((u_\sigma^+)_{x_k}\right)_t \widetilde\varphi = \frac{1}{2} \left[\Phi\left(\left(\left(u_\sigma^+\right)_{x_j}\right)^2\right) \Psi\left(\left(\left(u_\sigma^+\right)_{x_k}\right)^2\right)\right]_t \zeta \eta^2.$$

By summing the two equations obtained from (3.5) as described above, and using the identity (3.12), we get

$$\frac{1}{2} \iint_{(T_0+\sigma_1, T_1-\sigma_1)\times\Omega} \left[\Phi\left(\left(\left(u_\sigma^+\right)_{x_j}\right)^2\right) \Psi\left(\left(\left(u_\sigma^+\right)_{x_k}\right)^2\right)\right]_t \zeta \eta^2 \, dt \, dx$$
$$+ \iint_{(T_0+\sigma_1, T_1-\sigma_1)\times\Omega} \langle \left((\nabla F(\nabla u))_{x_j}\right)_\sigma^+, \nabla\varphi \rangle \, dt \, dx$$
$$+ \iint_{(T_0+\sigma_1, T_1-\sigma_1)\times\Omega} \langle \left((\nabla F(\nabla u))_{x_k}\right)_\sigma^+, \nabla\widetilde\varphi \rangle \, dt \, dx = 0.$$

We then perform an integration by parts in the time variable in the first integral, which gives

$$-\frac{1}{2} \iint_{(T_0+\sigma_1, T_1-\sigma_1)\times\Omega} \Phi\left(\left(\left(u_\sigma^+\right)_{x_j}\right)^2\right) \Psi\left(\left(\left(u_\sigma^+\right)_{x_k}\right)^2\right) \zeta' \eta^2 \, dt \, dx$$
$$+ \iint_{(T_0+\sigma_1, T_1-\sigma_1)\times\Omega} \langle \left((\nabla F(\nabla u))_{x_j}\right)_\sigma^+, \nabla\varphi \rangle \, dt \, dx + \iint_{(T_0+\sigma_1, T_1-\sigma_1)\times\Omega} \langle \left((\nabla F(\nabla u))_{x_k}\right)_\sigma^+, \nabla\widetilde\varphi \rangle \, dt \, dx = 0.$$

Finally, we let $\sigma$ go to 0. In the same vein as in the proof of (3.9), the Dominated Convergence Theorem implies that

$$(3.13) \qquad \begin{aligned} &-\frac{1}{2} \iint_{I\times\Omega} \Phi\left(\left(u_{x_j}\right)^2\right) \Psi\left(\left(u_{x_k}\right)^2\right) \zeta' \eta^2 \, dt \, dx \\ &+ \iint_{I\times\Omega} \langle (\nabla F(\nabla u))_{x_j}, \nabla\varphi_0 \rangle \, dt \, dx + \iint_{I\times\Omega} \langle (\nabla F(\nabla u))_{x_k}, \nabla\widetilde\varphi_0 \rangle \, dt \, dx = 0, \end{aligned}$$

where

$$\varphi_0 = u_{x_j} \Phi'\left(u_{x_j}^2\right) \Psi\left(u_{x_k}^2\right) \zeta \eta^2,$$

and

$$\widetilde\varphi_0 = u_{x_k} \Psi'\left(u_{x_k}^2\right) \Phi\left(u_{x_j}^2\right) \zeta \eta^2.$$

We now choose $\zeta$ as in (3.10) and insert it in (3.13). By letting $\delta$ go to 0, we obtain for almost every $\tau \in I$

$$
\begin{aligned}
(3.14) \qquad & \frac{\chi(\tau)}{2} \int_{\{\tau\}\times\Omega} \Phi\left(u_{x_j}^2\right) \Psi\left(u_{x_k}^2\right) \eta^2 \, dx \\
& + \iint_{(T_0,\tau)\times\Omega} \langle (\nabla F(\nabla u))_{x_j}, \nabla \psi_0 \rangle \, dt \, dx + \iint_{(T_0,\tau)\times\Omega} \langle (\nabla F(\nabla u))_{x_k}, \nabla \widetilde{\psi}_0 \rangle \, dt \, dx \\
& = \frac{1}{2} \iint_{(T_0,\tau)\times\Omega} \chi' \eta^2 \Phi(u_{x_j}^2) \Psi(u_{x_k}^2) \, dt \, dx,
\end{aligned}
$$

where $\psi_0$ and $\widetilde{\psi}_0$ are defined as $\varphi_0$ and $\widetilde{\varphi}_0$, except that $\zeta$ is now replaced by $\chi$.

**Step 2: completion of the proof.** We first observe that

$$
\langle (\nabla F(\nabla u))_{x_j}, \nabla \psi_0 \rangle = \langle D^2 F(\nabla u) \nabla u_{x_j}, \nabla \psi_0 \rangle, \qquad \langle (\nabla F(\nabla u))_{x_k}, \nabla \widetilde{\psi}_0 \rangle = \langle D^2 F(\nabla u) \nabla u_{x_k}, \nabla \widetilde{\psi}_0 \rangle.
$$

Taking into account the definition of $\psi_0$ and $\widetilde{\psi}_0$, we thus get from (3.14)

$$
\begin{aligned}
(3.15) \qquad & \frac{\chi(\tau)}{2} \int_{\{\tau\}\times\Omega} \Phi\left(u_{x_j}^2\right) \Psi\left(u_{x_k}^2\right) \eta^2 \, dx \\
& + \iint_{(T_0,\tau)\times\Omega} \langle D^2 F(\nabla u) \nabla u_{x_j}, \nabla u_{x_j} \rangle \left( \Phi'(u_{x_j}^2) + 2u_{x_j}^2 \Phi''(u_{x_j}^2) \right) \Psi(u_{x_k}^2) \chi \, \eta^2 \, dt \, dx \\
& + \iint_{(T_0,\tau)\times\Omega} \langle D^2 F(\nabla u) \nabla u_{x_k}, \nabla u_{x_k} \rangle \left( \Psi'(u_{x_k}^2) + 2u_{x_k}^2 \Psi''(u_{x_k}^2) \right) \Phi(u_{x_j}^2) \chi \, \eta^2 \, dt \, dx \\
& = \frac{1}{2} \iint_{(T_0,\tau)\times\Omega} \chi' \, \eta^2 \, \Phi(u_{x_j}^2) \Psi(u_{x_k}^2) \, dt \, dx \\
& - 4 \iint_{(T_0,\tau)\times\Omega} \langle D^2 F(\nabla u) \nabla u_{x_j}, \nabla u_{x_k} \rangle \, u_{x_j} \, \Phi'(u_{x_j}^2) \, u_{x_k} \, \Psi'(u_{x_k}^2) \chi \, \eta^2 \, dt \, dx \\
& - 2 \iint_{(T_0,\tau)\times\Omega} \langle D^2 F(\nabla u) \nabla u_{x_j}, \nabla \eta \rangle \, u_{x_j} \, \Phi'(u_{x_j}^2) \, \Psi(u_{x_k}^2) \chi \, \eta \, dt \, dx \\
& - 2 \iint_{(T_0,\tau)\times\Omega} \langle D^2 F(\nabla u) \nabla u_{x_k}, \nabla \eta \rangle \, u_{x_k} \, \Phi(u_{x_j}^2) \, \Psi'(u_{x_k}^2) \chi \, \eta \, dt \, dx.
\end{aligned}
$$

We first estimate the two last terms. We use the Cauchy-Schwarz inequality:

$$
|\langle D^2 F(\nabla u) \nabla u_{x_j}, \nabla \eta \rangle| \le \left( \langle D^2 F(\nabla u) \nabla u_{x_j}, \nabla u_{x_j} \rangle \right)^{\frac{1}{2}} \left( \langle D^2 F(\nabla u) \nabla \eta, \nabla \eta \rangle \right)^{\frac{1}{2}}.
$$

By the Young inequality, this implies

$$
\begin{aligned}
& \left| -2 \iint_{(T_0,\tau)\times\Omega} \langle D^2 F(\nabla u) \nabla u_{x_j}, \nabla \eta \rangle \, u_{x_j} \, \Phi'(u_{x_j}^2) \, \Psi(u_{x_k}^2) \chi \, \eta \, dt \, dx \right| \\
& \le \frac{1}{2} \iint_{(T_0,\tau)\times\Omega} \langle D^2 F(\nabla u) \nabla u_{x_j}, \nabla u_{x_j} \rangle \, \Phi'(u_{x_j}^2) \, \Psi(u_{x_k}^2) \chi \, \eta^2 \, dt \, dx \\
& \quad + 2 \iint_{(T_0,\tau)\times\Omega} \langle D^2 F(\nabla u) \nabla \eta, \nabla \eta \rangle \, u_{x_j}^2 \, \Phi'(u_{x_j}^2) \, \Psi(u_{x_k}^2) \chi \, dt \, dx.
\end{aligned}
$$

A similar inequality holds true for the last term in (3.15). Hence,

$$
\begin{aligned}
(3.16) \quad & \frac{\chi(\tau)}{2} \int_{\{\tau\} \times \Omega} \Phi\left(u_{x_j}^2\right) \Psi\left(u_{x_k}^2\right) \eta^2 \, dx \\
& + \iint_{(T_0,\tau) \times \Omega} \langle D^2 F(\nabla u) \nabla u_{x_j}, \nabla u_{x_j} \rangle \left( \frac{1}{2} \Phi'(u_{x_j}^2) + 2 \, u_{x_j}^2 \, \Phi''(u_{x_j}^2) \right) \Psi(u_{x_k}^2) \, \chi \, \eta^2 \, dt \, dx \\
& + \iint_{(T_0,\tau) \times \Omega} \langle D^2 F(\nabla u) \, \nabla u_{x_k}, \nabla u_{x_k} \rangle \left( \frac{1}{2} \Psi'(u_{x_k}^2) + 2 \, u_{x_k}^2 \, \Psi''(u_{x_k}^2) \right) \Phi(u_{x_j}^2) \, \chi \, \eta^2 \, dt \, dx \\
& \leq \frac{1}{2} \iint_{(T_0,\tau) \times \Omega} \chi' \, \eta^2 \, \Phi(u_{x_j}^2) \, \Psi(u_{x_k}^2) \, dt \, dx \\
& - 4 \iint_{(T_0,\tau) \times \Omega} \langle D^2 F(\nabla u) \, \nabla u_{x_j}, \nabla u_{x_k} \rangle \, u_{x_j} \, \Phi'(u_{x_j}^2) \, u_{x_k} \Psi'(u_{x_k}^2) \, \chi \, \eta^2 \, dt \, dx \\
& + 2 \iint_{(T_0,\tau) \times \Omega} \langle D^2 F(\nabla u) \nabla \eta, \nabla \eta \rangle \left( u_{x_j}^2 \, \Phi'(u_{x_j}^2) \Psi(u_{x_k}^2) + u_{x_k}^2 \, \Psi'(u_{x_k}^2) \Phi(u_{x_j}^2) \right) \chi \, dt \, dx.
\end{aligned}
$$

In the left-hand side of (3.16), in the second term, we drop $2 \, u_{x_j}^2 \, \Phi''(u_{x_j}^2)$ which is non-negative. We also drop the whole last term of the left-hand side for the same reason. This yields

$$
\begin{aligned}
(3.17) \quad & \frac{\chi(\tau)}{2} \int_{\{\tau\} \times \Omega} \Phi(u_{x_j}^2) \, \Psi(u_{x_k}^2) \, \eta^2 \, dx \\
& + \frac{1}{2} \iint_{(T_0,\tau) \times \Omega} \langle D^2 F(\nabla u) \nabla u_{x_j}, \nabla u_{x_j} \rangle \Phi'(u_{x_j}^2) \, \Psi(u_{x_k}^2) \, \chi \, \eta^2 \, dt \, dx \\
& \leq \frac{1}{2} \iint_{(T_0,\tau) \times \Omega} \chi' \, \eta^2 \, \Phi(u_{x_j}^2) \, \Psi(u_{x_k}^2) \, dt \, dx \\
& - 4 \iint_{(T_0,\tau) \times \Omega} \langle D^2 F(\nabla u) \, \nabla u_{x_j}, \nabla u_{x_k} \rangle \, u_{x_j} \, \Phi'(u_{x_j}^2) \, u_{x_k} \, \Psi'(u_{x_k}^2) \, \chi \, \eta^2 \, dt \, dx \\
& + 2 \iint_{(T_0,\tau) \times \Omega} \langle D^2 F(\nabla u) \nabla \eta, \nabla \eta \rangle \left( u_{x_j}^2 \, \Phi'(u_{x_j}^2) \Psi(u_{x_k}^2) + u_{x_k}^2 \, \Psi'(u_{x_k}^2) \Phi(u_{x_j}^2) \right) \chi \, dt \, dx.
\end{aligned}
$$

We next estimate the second term of the right-hand side of (3.17), that we denote by

$$
\mathcal{A} := \iint_{(T_0,\tau) \times \Omega} \langle D^2 F(\nabla u) \, \nabla u_{x_j}, \nabla u_{x_k} \rangle \, u_{x_j} \, \Phi'(u_{x_j}^2) \, u_{x_k} \, \Psi'(u_{x_k}^2) \, \chi \, \eta^2 \, dt \, dx.
$$

We first use the Cauchy-Schwarz inequality to get

$$
|\langle D^2 F(\nabla u) \, \nabla u_{x_j}, \nabla u_{x_k} \rangle| \leq \left( \langle D^2 F(\nabla u) \nabla u_{x_j}, \nabla u_{x_j} \rangle \right)^{\frac{1}{2}} \left( \langle D^2 F(\nabla u) \, \nabla u_{x_k}, \nabla u_{x_k} \rangle \right)^{\frac{1}{2}}.
$$

We then introduce a parameter $\theta \in [0,1]$. By writing $\Psi'(u_{x_k}^2) = \left( \Psi'(u_{x_k}^2) \right)^{\frac{\theta}{2}} \left( \Psi'(u_{x_k}^2) \right)^{1-\frac{\theta}{2}}$, one gets by the Cauchy-Schwarz inequality again:

$$
\begin{aligned}
(3.18) \quad \mathcal{A} \leq & \left( \iint_{(T_0,\tau) \times \Omega} \langle D^2 F(\nabla u) \nabla u_{x_j}, \nabla u_{x_j} \rangle \, u_{x_j}^2 \left( \Phi'(u_{x_j}^2) \right)^2 \left( \Psi'(u_{x_k}^2) \right)^{\theta} \chi \, \eta^2 \, dt \, dx \right)^{\frac{1}{2}} \\
& \times \left( \iint_{(T_0,\tau) \times \Omega} \langle D^2 F(\nabla u) \, \nabla u_{x_k}, \nabla u_{x_k} \rangle \left( \Psi'(u_{x_k}^2) \right)^{2-\theta} u_{x_k}^2 \, \chi \, \eta^2 \, dt \, dx \right)^{\frac{1}{2}}.
\end{aligned}
$$

We define

$$
G(t) = \int_0^{t^2} \Psi'(s)^{1-\frac{\theta}{2}} \, ds.
$$

Then $G$ is a $C^1$ non-negative convex function. Moreover, by its definition

$$\nabla G(u_{x_k}) = 2\, u_{x_k} \left(\Psi'(u_{x_k}^2)\right)^{1-\frac{\theta}{2}} \nabla u_{x_k}.$$

Hence, by the standard Caccioppoli inequality (3.6) with $h = G$ and $k$ in place of $j$, this yields:

$$\iint_{(T_0,\tau)\times\Omega} \langle D^2 F(\nabla u)\, \nabla u_{x_k}, \nabla u_{x_k}\rangle \left(\Psi'(u_{x_k}^2)\right)^{2-\theta} u_{x_k}^2\, \chi\, \eta^2\, dt\, dx$$

$$= \frac{1}{4} \iint_{(T_0,\tau)\times\Omega} \langle D^2 F(\nabla u)\nabla G(u_{x_k}), \nabla G(u_{x_k})\rangle\, \chi\, \eta^2\, dt\, dx$$

$$\leq \frac{1}{4} \iint_{(T_0,\tau)\times\Omega} \chi'\, \eta^2\, G^2(u_{x_k})\, dt\, dx + \iint_{(T_0,\tau)\times\Omega} \langle D^2 F(\nabla u)\nabla\eta, \nabla\eta\rangle\, G^2(u_{x_k})\, \chi\, dt\, dx.$$

By using the Jensen inequality with the concave function $y \mapsto y^{1-\theta/2}$, we obtain

$$0 \leq G(u_{x_k}) \leq |u_{x_k}|^\theta \left(\Psi(u_{x_k}^2)\right)^{1-\frac{\theta}{2}}.$$

This implies

$$\iint_{(T_0,\tau)\times\Omega} \langle D^2 F(\nabla u)\, \nabla u_{x_k}, \nabla u_{x_k}\rangle \left(\Psi'(u_{x_k}^2)\right)^{2-\theta} u_{x_k}^2\, \chi\, \eta^2\, dt\, dx$$

$$\leq \frac{1}{4} \iint_{(T_0,\tau)\times\Omega} \chi'\, \eta^2\, |u_{x_k}|^{2\theta} \left(\Psi(u_{x_k}^2)\right)^{2-\theta}\, dt\, dx$$

$$+ \iint_{(T_0,\tau)\times\Omega} \langle D^2 F(\nabla u)\, \nabla\eta, \nabla\eta\rangle\, |u_{x_k}|^{2\theta} \left(\Psi(u_{x_k}^2)\right)^{2-\theta}\, \chi\, dt\, dx.$$

Coming back to (3.18), it follows that

$$\mathcal{A} \leq \left(\iint_{(T_0,\tau)\times\Omega} \langle D^2 F(\nabla u)\nabla u_{x_j}, \nabla u_{x_j}\rangle\, u_{x_j}^2 \left(\Phi'(u_{x_j}^2)\right)^2 \left(\Psi'(u_{x_k}^2)\right)^\theta \chi\, \eta^2\, dt\, dx\right)^{\frac{1}{2}}$$

$$\times \left(\frac{1}{4} \iint_{(T_0,\tau)\times\Omega} \chi'\, \eta^2\, |u_{x_k}|^{2\theta} \left(\Psi(u_{x_k}^2)\right)^{2-\theta}\, dt\, dx\right.$$

$$\left.+ \iint_{(T_0,\tau)\times\Omega} \langle D^2 F(\nabla u)\nabla\eta, \nabla\eta\rangle\, |u_{x_k}|^{2\theta} \left(\Psi(u_{x_k}^2)\right)^{2-\theta} \chi\, dt\, dx\right)^{\frac{1}{2}}.$$

Together with (3.17), this yields the desired inequality. Finally, the $C^2$ assumption on $\Phi$ and $\Psi$ can be removed by a standard approximation argument. $\qquad\square$

## 4. Uniform Lipschitz estimate for the regularized equation

This section is devoted to the proof of the following uniform estimate. For simplicity, we will work with anisotropic parabolic cubes of the form

$$Q_R(t_0, x_0) = (t_0 - R^p, t_0) \times (x_0 - R, x_0 + R)^N.$$

As in the previous section, we drop the index $\varepsilon$ and simply write $u$ and $F$, in place of $u_\varepsilon$ and $F_\varepsilon$.

**Proposition 4.1.** *There exist $\alpha = \alpha(N) > 2$ and $C = C(N, p) > 0$ such that for every $\varepsilon > 0$ and for every $Q_r(x_0, t_0) \subset Q_R(t_0, x_0) \Subset I \times \Omega$ with $R \leq 1$, one has*

$$(4.1) \qquad \|\nabla u\|_{L^\infty(Q_r(x_0,t_0))} \leq \frac{C}{(R-r)^{\alpha p}} \left[\left(\iint_{Q_R(t_0,x_0)} |\nabla u|^p\, dt\, dx\right)^{\frac{1}{2}} + 1\right].$$

*Proof.* We will limit ourselves for simplicity to the case $N \geq 3$. This allows to use the Sobolev inequality valid for every $f \in W_0^{1,2}(\Omega)$

$$\|f\|_{L^{2^*}(\Omega)} \leq C_N \|\nabla f\|_{L^2(\Omega)}, \qquad \text{with } 2^* = \frac{2N}{N-2}.$$

Here $C_N$ is a constant which depends only on $N$. The case $N = 2$ follows with minor modifications and we omit the details.

The proof is quite involved and for ease of readability, we divide it into several steps.

**Step 1: the choices of $\Phi$ and $\Psi$.** We apply Lemma 3.2 with the following choices

$$\Phi(t) = t^s \qquad \text{and} \qquad \Psi(t) = t^m, \qquad \text{for } t \geq 0,$$

with $1 \leq s \leq m$. We also take

$$\theta = \begin{cases} \dfrac{m-s}{m-1} \in [0,1] & \text{if } m > 1, \\ 1 & \text{if } m = 1. \end{cases}$$

This gives

$$\chi(\tau) \int_{\{\tau\}\times\Omega} |u_{x_j}|^{2s} |u_{x_k}|^{2m} \eta^2 \, dx$$

$$+ s \iint_{(T_0,\tau)\times\Omega} \langle D^2 F(\nabla u)\nabla u_{x_j}, \nabla u_{x_j}\rangle |u_{x_j}|^{2s-2} |u_{x_k}|^{2m} \chi \eta^2 \, dt \, dx$$

$$\leq \iint_{(T_0,\tau)\times\Omega} \chi' \eta^2 |u_{x_j}|^{2s} |u_{x_k}|^{2m} \, dt \, dx$$

$$+ 4 \iint_{(T_0,\tau)\times\Omega} \langle D^2 F(\nabla u)\nabla\eta, \nabla\eta\rangle \left(s |u_{x_j}|^{2s} |u_{x_k}|^{2m} + m |u_{x_k}|^{2m} |u_{x_j}|^{2s}\right) \chi \, dt \, dx$$

$$+ 8 \, s \, m^{\frac{\theta}{2}} \left(\iint_{(T_0,\tau)\times\Omega} \langle D^2 F(\nabla u)\nabla u_{x_j}, \nabla u_{x_j}\rangle |u_{x_j}|^{4s-2} |u_{x_k}|^{2m-2s} \chi\eta^2 \, dt \, dx\right)^{\frac{1}{2}}$$

$$\times \left(\frac{1}{4} \iint_{(T_0,\tau)\times\Omega} \chi'\eta^2 |u_{x_k}|^{2(s+m)} \, dt \, dx + \iint_{(T_0,\tau)\times\Omega} \langle D^2 F(\nabla u)\nabla\eta, \nabla\eta\rangle |u_{x_k}|^{2(s+m)} \chi \, dt \, dx\right)^{\frac{1}{2}}.$$

On the product of the two last integrals, we use the Young inequality in the form

$$a \, b \leq a^2 + \frac{b^2}{4}.$$

This gives

$$\chi(\tau) \int_{\{\tau\}\times\Omega} |u_{x_j}|^{2s} |u_{x_k}|^{2m} \eta^2 \, dx + s \iint_{(T_0,\tau)\times\Omega} \langle D^2 F(\nabla u)\nabla u_{x_j}, \nabla u_{x_j}\rangle |u_{x_j}|^{2s-2} |u_{x_k}|^{2m} \chi \eta^2 \, dt \, dx$$

$$\leq \iint_{(T_0,\tau)\times\Omega} \chi' \eta^2 |u_{x_j}|^{2s} |u_{x_k}|^{2m} \, dt \, dx + 4 (s+m) \iint_{(T_0,\tau)\times\Omega} \langle D^2 F(\nabla u)\nabla\eta, \nabla\eta\rangle |u_{x_j}|^{2s} |u_{x_k}|^{2m} \chi \, dt \, dx$$

$$+ 16 \, s^2 \, m^\theta \left(\frac{1}{4} \iint_{(T_0,\tau)\times\Omega} \chi' \eta^2 |u_{x_k}|^{2(s+m)} \, dt \, dx + \iint_{(T_0,\tau)\times\Omega} \langle D^2 F(\nabla u)\nabla\eta, \nabla\eta\rangle |u_{x_k}|^{2(s+m)} \chi \, dt \, dx\right)$$

$$+ \iint_{(T_0,\tau)\times\Omega} \langle D^2 F(\nabla u)\nabla u_{x_j}, \nabla u_{x_j}\rangle |u_{x_j}|^{4s-2} |u_{x_k}|^{2m-2s} \chi \eta^2 \, dt \, dx.$$

By the Young inequality again, we can estimate

$$|u_{x_j}|^{2s} |u_{x_k}|^{2m} \leq |u_{x_j}|^{2m+2s} + |u_{x_k}|^{2m+2s}.$$

Using that $s \geq 1$ in the left-hand side and $m^\theta \leq m$ in the right-hand side, we thus obtain

$$\chi(\tau) \int_{\{\tau\} \times \Omega} |u_{x_j}|^{2s} |u_{x_k}|^{2m} \eta^2 \, dx + \iint_{(T_0,\tau) \times \Omega} \langle D^2 F(\nabla u)\nabla u_{x_j}, \nabla u_{x_j}\rangle \, |u_{x_j}|^{2s-2} |u_{x_k}|^{2m} \chi \, \eta^2 \, dt \, dx$$

$$\leq \iint_{(T_0,\tau) \times \Omega} \chi' \, \eta^2 \left( |u_{x_j}|^{2(s+m)} + |u_{x_k}|^{2(s+m)} \right) dt \, dx$$

$$+ 4(s+m) \iint_{(T_0,\tau) \times \Omega} \langle D^2 F(\nabla u)\nabla \eta, \nabla \eta \rangle \left( |u_{x_j}|^{2(s+m)} + |u_{x_k}|^{2(s+m)} \right) \chi \, dt \, dx$$

$$+ 16 \, s^2 \, m \left( \frac{1}{4} \iint_{(T_0,\tau) \times \Omega} \chi' \, \eta^2 \, |u_{x_k}|^{2(s+m)} \, dt \, dx + \iint_{(T_0,\tau) \times \Omega} \langle D^2 F(\nabla u)\nabla \eta, \nabla \eta \rangle |u_{x_k}|^{2(s+m)} \chi \, dt \, dx \right)$$

$$+ \iint_{(T_0,\tau) \times \Omega} \langle D^2 F(\nabla u)\nabla u_{x_j}, \nabla u_{x_j}\rangle |u_{x_j}|^{4s-2} |u_{x_k}|^{2m-2s} \chi \, \eta^2 \, dt \, dx.$$

This finally implies

$$\chi(\tau) \int_{\{\tau\} \times \Omega} |u_{x_j}|^{2s} |u_{x_k}|^{2m} \eta^2 \, dx + \iint_{(T_0,\tau) \times \Omega} \langle D^2 F(\nabla u)\nabla u_{x_j}, \nabla u_{x_j}\rangle \, |u_{x_j}|^{2s-2} |u_{x_k}|^{2m} \chi \, dt \, dx$$

$$(4.2) \qquad \leq 16(s+m+s^2 m) \iint_{(T_0,\tau) \times \Omega} \left( \chi' \, \eta^2 + \chi \, \langle D^2 F(\nabla u)\nabla \eta, \nabla \eta \rangle \right) \left( |u_{x_j}|^{2(s+m)} + |u_{x_k}|^{2(s+m)} \right) dt \, dx$$

$$+ \iint_{(T_0,\tau) \times \Omega} \langle D^2 F(\nabla u)\nabla u_{x_j}, \nabla u_{x_j}\rangle \, |u_{x_j}|^{4s-2} |u_{x_k}|^{2m-2s} \chi \, \eta^2 \, dt \, dx.$$

**Step 2: the staircase.** Let $\ell_0 \in \mathbb{N} \setminus \{0\}$ and set $q = 2^{\ell_0} - 1$. We define the two families of indices

$$s_\ell = 2^\ell \qquad \text{and} \qquad m_\ell = q + 1 - 2^\ell, \qquad \text{for } \ell \in \{0, \dots, \ell_0\}.$$

We observe that by construction, for every $0 \leq \ell \leq \ell_0 - 1$,

$$s_\ell + m_\ell = q + 1, \qquad 4s_\ell - 2 = 2s_{\ell+1} - 2 \qquad \text{and} \qquad 2m_\ell - 2s_\ell = 2m_{\ell+1}.$$

We also use that $s_\ell + m_\ell + s_\ell^2 \, m_\ell \leq 2\,(q+1)^3$. Then the above inequality (4.2) written for $s = s_\ell$ and $m = m_\ell$ with $0 \leq \ell \leq \ell_0 - 1$ gives

$$\chi(\tau) \int_{\{\tau\} \times \Omega} |u_{x_j}|^{2s_\ell} |u_{x_k}|^{2m_\ell} \eta^2 \, dx$$

$$+ \iint_{(T_0,\tau) \times \Omega} \langle D^2 F(\nabla u)\nabla u_{x_j}, \nabla u_{x_j}\rangle \, |u_{x_j}|^{2s_\ell-2} |u_{x_k}|^{2m_\ell} \chi \, \eta^2 \, dt \, dx$$

$$\leq 32\,(q+1)^3 \iint_{(T_0,\tau) \times \Omega} \left( \chi' \, \eta^2 + \chi \, \langle D^2 F(\nabla u)\nabla \eta, \nabla \eta \rangle \right) \left( |u_{x_j}|^{2(q+1)} + |u_{x_k}|^{2(q+1)} \right) dt \, dx$$

$$+ \iint_{(T_0,\tau) \times \Omega} \langle D^2 F(\nabla u)\nabla u_{x_j}, \nabla u_{x_j}\rangle \, |u_{x_j}|^{2s_{\ell+1}-2} |u_{x_k}|^{2m_{\ell+1}} \chi \, \eta^2 \, dt \, dx.$$

Observe that we used that $2\,s_\ell + 2\,m_\ell = 2\,(q+1)$ on the first term on the right-hand side. By summing from $\ell = 0$ up to $\ell = \ell_0 - 1$ and erasing the common terms on both sides, one gets

$$\chi(\tau) \sum_{\ell=0}^{\ell_0-1} \int_{\{\tau\} \times \Omega} |u_{x_j}|^{2s_\ell} |u_{x_k}|^{2m_\ell} \eta^2 \, dx + \iint_{(T_0,\tau) \times \Omega} \langle D^2 F(\nabla u)\nabla u_{x_j}, \nabla u_{x_j}\rangle \, |u_{x_k}|^{2q} \chi \, \eta^2 \, dt \, dx$$

$$\leq C \, q^3 \, \ell_0 \iint_{(T_0,\tau) \times \Omega} \left( \chi' \eta^2 + \chi \, \langle D^2 F(\nabla u)\nabla \eta, \nabla \eta \rangle \right) \left( |u_{x_j}|^{2(q+1)} + |u_{x_k}|^{2(q+1)} \right) dt \, dx$$

$$+ \iint_{(T_0,\tau) \times \Omega} \langle D^2 F(\nabla u)\nabla u_{x_j}, \nabla u_{x_j}\rangle \, |u_{x_j}|^{2q} \chi \, \eta^2 \, dt \, dx.$$

For the last term, we apply Lemma 3.1 with the choice

$$h(t) = \frac{|t|^{q+1}}{q+1}, \qquad t \in \mathbb{R}.$$

We thus get

$$\chi(\tau) \sum_{\ell=0}^{\ell_0-1} \int_{\{\tau\}\times\Omega} |u_{x_j}|^{2s_\ell} |u_{x_k}|^{2m_\ell} \eta^2 \, dx + \iint_{(T_0,\tau)\times\Omega} \langle D^2 F(\nabla u)\,\nabla u_{x_j}, \nabla u_{x_j}\rangle\, |u_{x_k}|^{2q}\, \chi\, \eta^2 \, dt\, dx$$

$$\leq C\, q^4 \iint_{(T_0,\tau)\times\Omega} \left(\chi'\, \eta^2 + \chi\, \langle D^2 F(\nabla u)\nabla\eta, \nabla\eta\rangle\right) \left(|u_{x_j}|^{2\,(q+1)} + |u_{x_k}|^{2\,(q+1)}\right)\, dt\, dx$$

$$+ \frac{1}{(q+1)^2} \iint_{(T_0,\tau)\times\Omega} \chi'\, \eta^2\, |u_{x_j}|^{2(q+1)}\, dt\, dx$$

$$+ \frac{4}{(q+1)^2} \iint_{(T_0,\tau)\times\Omega} \langle D^2 F(\nabla u)\nabla\eta, \nabla\eta\rangle\, |u_{x_j}|^{2(q+1)}\, \chi\, dt\, dx.$$

This is turn implies that

$$(4.3) \quad \begin{aligned} &\chi(\tau) \sum_{\ell=0}^{\ell_0-1} \int_{\{\tau\}\times\Omega} |u_{x_j}|^{2s_\ell} |u_{x_k}|^{2m_\ell} \eta^2 \, dx + \iint_{(T_0,\tau)\times\Omega} \langle D^2 F(\nabla u)\nabla u_{x_j}, \nabla u_{x_j}\rangle |u_{x_k}|^{2q} \chi\eta^2 \, dt\, dx \\ &\leq C\, q^4 \iint_{(T_0,\tau)\times\Omega} \left(\chi'\eta^2 + \chi\, \langle D^2 F(\nabla u)\nabla\eta, \nabla\eta\rangle\right) \left(|u_{x_j}|^{2(q+1)} + |u_{x_k}|^{2(q+1)}\right)\, dt\, dx, \end{aligned}$$

possibly for a different constant $C > 0$.

**Step 3: weak ellipticity and boundedness of $D^2 F$.** We now use the explicit expression of $F$. We recall that

$$F(\xi) = \frac{1}{p} \sum_{i=1}^{N} |\xi_i|^p + \varepsilon\, G(\xi), \qquad \text{for every } \xi \in \mathbb{R}^N,$$

where

$$G(\xi) = \frac{1}{p} \left(1 + |\xi|^2\right)^{\frac{p}{2}}, \qquad \text{for every } \xi \in \mathbb{R}^N.$$

If follows that for every $\xi, \lambda \in \mathbb{R}^N$, we have

$$(p-1) \sum_{i=1}^{N} |\xi_i|^{p-2}\, \lambda_i^2 \leq \langle D^2 F(\xi)\lambda, \lambda\rangle \leq C \left(|\xi|^{p-2} + 1\right)\, |\lambda|^2.$$

By inserting these estimates into (4.3), one gets

$$\chi(\tau) \sum_{\ell=0}^{\ell_0-1} \int_{\{\tau\}\times\Omega} |u_{x_j}|^{2s_\ell} |u_{x_k}|^{2m_\ell} \eta^2 \, dx$$

$$+ (p-1) \sum_{i=1}^{N} \iint_{(T_0,\tau)\times\Omega} |u_{x_i}|^{p-2} u_{x_i x_j}^2 |u_{x_k}|^{2q} \chi\, \eta^2 \, dt\, dx$$

$$\leq C\, q^4 \iint_{(T_0,\tau)\times\Omega} \left(\chi'\, \eta^2 + \chi\, \left(|\nabla u|^{p-2} + 1\right) |\nabla\eta|^2\right) \left(|u_{x_j}|^{2(q+1)} + |u_{x_k}|^{2(q+1)}\right)\, dt\, dx.$$

We consider the second term on the left-hand side: observe that by keeping in the sum only the term with $i = k$ and dropping the others, we get

$$\sum_{i=1}^{N} \iint_{(T_0,\tau)\times\Omega} |u_{x_i}|^{p-2}\, u_{x_i x_j}^2\, |u_{x_k}|^{2q}\, \chi\, \eta^2\, dt\, dx \geq \iint_{(T_0,\tau)\times\Omega} |u_{x_k}|^{p-2}\, u_{x_k x_j}^2\, |u_{x_k}|^{2q}\, \chi\, \eta^2\, dt\, dx$$

$$= \frac{1}{\left(q+\dfrac{p}{2}\right)^2} \iint_{(T_0,\tau)\times\Omega} \left| \left( |u_{x_k}|^{q+\frac{p-2}{2}} u_{x_k} \right)_{x_j} \right|^2 \chi\, \eta^2\, dt\, dx.$$

When we sum over $j = 1, \ldots, N$ the resulting estimate, we thus get

$$\chi(\tau) \sum_{\ell=0}^{\ell_0-1} \int_{\{\tau\}\times\Omega} \sum_{j=1}^{N} |u_{x_j}|^{2s_\ell}\, |u_{x_k}|^{2m_\ell}\, \eta^2\, dx + \frac{p-1}{\left(q+\dfrac{p}{2}\right)^2} \iint_{(T_0,\tau)\times\Omega} \left| \nabla\left( |u_{x_k}|^{q+\frac{p-2}{2}} u_{x_k} \right) \right|^2 \chi\, \eta^2\, dt\, dx$$

$$\leq C q^4 \iint_{(T_0,\tau)\times\Omega} \left( \chi'\eta^2 + \chi\left(|\nabla u|^{p-2}+1\right)|\nabla\eta|^2 \right) \left( \sum_{j=1}^{N} |u_{x_j}|^{2(q+1)} + N\,|u_{x_k}|^{2(q+1)} \right) dt\, dx,$$

which in turn implies

$$\chi(\tau) \sum_{\ell=0}^{\ell_0-1} \int_{\{\tau\}\times\Omega} \sum_{j=1}^{N} |u_{x_j}|^{2s_\ell}\, |u_{x_k}|^{2m_\ell}\, \eta^2\, dx + \iint_{(T_0,\tau)\times\Omega} \left| \nabla\left( |u_{x_k}|^{q+\frac{p-2}{2}} u_{x_k} \right) \right|^2 \chi\, \eta^2\, dt\, dx$$

$$\leq C q^6 \iint_{(T_0,\tau)\times\Omega} \left( \chi'\eta^2 + \chi\left(|\nabla u|^{p-2}+1\right)|\nabla\eta|^2 \right) |\nabla u|^{2(q+1)}\, dt\, dx,$$

up to redefine the constant $C > 0$. We now add the term

$$\iint_{(T_0,\tau)\times\Omega} |u_{x_k}|^{2q+p}\, \chi\, |\nabla\eta|^2\, dt\, dx,$$

on both sides of the above inequality. With some algebraic manipulations, this gives

$$\chi(\tau) \sum_{\ell=0}^{\ell_0-1} \int_{\{\tau\}\times\Omega} \sum_{j=1}^{N} |u_{x_j}|^{2s_\ell}\, |u_{x_k}|^{2m_\ell}\, \eta^2\, dx + \iint_{(T_0,\tau)\times\Omega} \left| \nabla\left( |u_{x_k}|^{q+\frac{p-2}{2}} u_{x_k}\, \eta \right) \right|^2 \chi\, dt\, dx$$

$$\leq C q^6 \iint_{(T_0,\tau)\times\Omega} \left( \chi'\,\eta^2 + \chi\left(|\nabla u|^{p-2}+1\right)|\nabla\eta|^2 \right) |\nabla u|^{2(q+1)}\, dt\, dx + \iint_{(T_0,\tau)\times\Omega} |u_{x_k}|^{2q+p}\chi|\nabla\eta|^2\, dt\, dx$$

$$\leq C q^6 \iint_{(T_0,\tau)\times\Omega} \left( \chi'\eta^2 + \chi\left(|\nabla u|^{p-2}+1\right)|\nabla\eta|^2 \right) |\nabla u|^{2(q+1)}\, dt\, dx.$$

By using the Sobolev inequality in the spatial variable for the second term of the left-hand side, one obtains

$$\chi(\tau) \sum_{\ell=0}^{\ell_0-1} \int_{\{\tau\}\times\Omega} \sum_{j=1}^{N} |u_{x_j}|^{2s_\ell}\, |u_{x_k}|^{2m_\ell}\, \eta^2\, dx + \int_{T_0}^{\tau} \chi \left( \int_{\Omega} \left( |u_{x_k}|^{2q+p}\eta^2 \right)^{\frac{2^*}{2}} dx \right)^{\frac{2}{2^*}} dt$$

$$\leq C q^6 \iint_{(T_0,\tau)\times\Omega} \left( \chi'\,\eta^2 + \chi\left(|\nabla u|^{p-2}+1\right)|\nabla\eta|^2 \right) |\nabla u|^{2(q+1)}\, dt\, dx.$$

We now finally sum over $k = 1, \ldots, N$ and use the Minkowski inequality for the second term of the left-hand side. This gives

$$\chi(\tau) \sum_{\ell=0}^{\ell_0 - 1} \int_{\{\tau\} \times \Omega} \sum_{j=1}^{N} |u_{x_j}|^{2s_\ell} \sum_{k=1}^{N} |u_{x_k}|^{2m_\ell} \eta^2 \, dx + \int_{T_0}^{\tau} \chi \left( \int_\Omega \left( \sum_{k=1}^{N} |u_{x_k}|^{2q+p} \eta^2 \right)^{\frac{2^*}{2}} dx \right)^{\frac{2}{2^*}} dt$$

$$\leq C \, q^6 \iint_{(T_0, \tau) \times \Omega} \left( \chi' \eta^2 + \chi \left( |\nabla u|^{p-2} + 1 \right) |\nabla \eta|^2 \right) |\nabla u|^{2(q+1)} \, dt \, dx.$$

We introduce the expedient function

$$\mathcal{U} = \max_{1 \leq k \leq N} |u_{x_k}|.$$

We observe that for every $s \geq 0$, one has

$$\mathcal{U}^s \leq \sum_{k=1}^{N} |u_{x_k}|^s \leq N \mathcal{U}^s.$$

In particular for $s = 2$, we get $\mathcal{U} \leq |\nabla u| \leq \sqrt{N} \, \mathcal{U}$. Hence, we get

$$\chi(\tau) \int_{\{\tau\} \times \Omega} \mathcal{U}^{2(q+1)} \eta^2 \, dx + \int_{T_0}^{\tau} \chi \left( \int_\Omega \mathcal{U}^{(2q+p) \frac{2^*}{2}} \eta^{2^*} \, dx \right)^{\frac{2}{2^*}} dt$$

(4.4)
$$\leq C \, q^6 \iint_{(T_0, \tau) \times \Omega} \left( \chi' \, \eta^2 + \chi \left( \mathcal{U}^{p-2} + 1 \right) |\nabla \eta|^2 \right) \mathcal{U}^{2(q+1)} \, dt \, dx$$

$$\leq C \, q^6 \iint_{(T_0, \tau) \times \Omega} \left( \chi' \, \eta^2 + \chi |\nabla \eta|^2 \right) \left( 1 + \mathcal{U}^{2q+p} \right) \, dt \, dx.$$

**Step 4: choice of the cut-off functions.** Let $(t_0, x_0) \in I \times \Omega$ and $0 < r < R \leq 1$ such that the cube $Q_R(x_0) = (x_0 - R, x_0 + R)^N$ is compactly contained in $\Omega$. We further require that

$$(t_0 - R^p, t_0) \Subset I,$$

so that we must have $T_0 < t_0 < T_1$ and $R^p < t_0 - T_0$. Let $\chi : [T_0, T_1] \to \mathbb{R}$ be a non-decreasing Lipschitz function such that

$$\chi(t) = 0 \ \text{ on } [T_0, t_0 - R^p], \qquad \chi(t) = 1 \ \text{ on } [t_0 - r^p, t_0] \qquad \text{and} \qquad |\chi'(t)| \leq \frac{C}{(R - r)^p}.$$

Let $\eta \in C_0^\infty(Q_R(x_0))$ be such that

$$0 \leq \eta \leq 1, \qquad \eta = 1 \ \text{ on } Q_r(x_0) \qquad \text{and} \qquad |\nabla \eta| \leq \frac{C}{R - r}.$$

We recall the notation for the anisotropic parabolic cube

$$Q_\rho(t_0, x_0) = (t_0 - \rho^p, t_0) \times Q_\rho(x_0).$$

With such a choice of $\chi$ and $\eta$, we use (4.4) twice:
- firstly, by dropping the second term in the left-hand side, and taking the supremum in $\tau$ over the interval $(t_0 - r^p, t_0)$;
- secondly, by dropping the first term in the left-hand side and taking $\tau = t_0$.

By summing up the two resulting contributions, this gives

$$\sup_{\tau \in (t_0 - r^p, t_0)} \int_{\{\tau\} \times Q_r(x_0)} \mathcal{U}^{2(q+1)} \, dx + \int_{t_0 - r^p}^{t_0} \left( \int_{Q_r(x_0)} \mathcal{U}^{(2q+p)\frac{2^*}{2}} \, dx \right)^{\frac{2}{2^*}} dt$$

(4.5)
$$\leq C \frac{q^6}{(R-r)^p} \iint_{Q_R(t_0, x_0)} \left( 1 + \mathcal{U}^{2q+p} \right) \, dt \, dx.$$

Observe that we also used that $(R-r)^p \leq (R-r)^2$, since $R \leq 1$ and $p > 2$. By the Hölder inequality, one has

$$\iint_{Q_r(t_0, x_0)} \mathcal{U}^{2q+p+\frac{4(q+1)}{N}} \, dt \, dx \leq \int_{t_0-r^p}^{t_0} \left( \int_{Q_r(x_0)} \mathcal{U}^{(2q+p)\frac{2^*}{2}} \, dx \right)^{\frac{2}{2^*}} \left( \int_{Q_r(x_0)} \mathcal{U}^{2(q+1)} \, dx \right)^{\frac{2}{N}} dt$$

$$\leq \left( \sup_{\tau \in (t_0 - r^p, t_0)} \int_{\{\tau\} \times Q_r(x_0)} \mathcal{U}^{2(q+1)} \, dx \right)^{\frac{2}{N}} \int_{t_0-r^p}^{t_0} \left( \int_{Q_r(x_0)} \mathcal{U}^{(2q+p)\frac{2^*}{2}} \, dx \right)^{\frac{2}{2^*}} dt.$$

Using (4.5), this implies

(4.6)
$$\iint_{Q_r(t_0, x_0)} \mathcal{U}^{2q+p+\frac{4(q+1)}{N}} \, dt \, dx \leq \left( C \frac{q^6}{(R-r)^p} \iint_{Q_R(t_0, x_0)} \left( 1 + \mathcal{U}^{2q+p} \right) \, dt \, dx \right)^{\frac{2}{N}+1}.$$

**Step 5: the local $L^\infty$ estimate on $\nabla u$.** Take $q = q_j = 2^{j+1} - 1$ with $j \in \mathbb{N}$. We set

$$\gamma_j := 2\,q_j + p = 2^{j+2} - 2 + p, \qquad \delta_j := 2\,q_j + p + \frac{4}{N}\,(q_j + 1) = 2^{j+2} - 2 + p + \frac{4}{N}\,2^{j+1},$$

and

$$\tau_j = \frac{\delta_j - \gamma_j}{\delta_j - \gamma_{j-1}} \frac{\gamma_{j-1}}{\gamma_j}.$$

We observe that $\gamma_{j-1} < \gamma_j < \delta_j$ and $\tau_j \in (0, 1)$ is defined in a such a way that

$$\frac{1}{\gamma_j} = \frac{\tau_j}{\gamma_{j-1}} + \frac{1 - \tau_j}{\delta_j}.$$

The estimate (4.6) can be rewritten as

$$\iint_{Q_r(t_0, x_0)} \mathcal{U}^{\delta_j} \, dt \, dx \leq \left( C \frac{q_j^6}{(R-r)^p} \iint_{Q_R(t_0, x_0)} \left( 1 + \mathcal{U}^{\gamma_j} \right) \, dt \, dx \right)^{\frac{2}{N}+1}$$

By interpolation in $L^p$ spaces, we obtain

$$\iint_{Q_r(t_0, x_0)} \mathcal{U}^{\gamma_j} \, dt \, dx \leq \left( \iint_{Q_r(t_0, x_0)} \mathcal{U}^{\gamma_{j-1}} \, dt \, dx \right)^{\tau_j \frac{\gamma_j}{\gamma_{j-1}}} \left( \iint_{Q_r(t_0, x_0)} \mathcal{U}^{\delta_j} \, dt \, dx \right)^{(1-\tau_j) \frac{\gamma_j}{\delta_j}}.$$

By lengthy but elementary computations, we see that

$$(1 - \tau_j) \frac{\gamma_j}{\delta_j} = \frac{N}{N+4} \qquad \text{and} \qquad \tau_j \frac{\gamma_j}{\gamma_{j-1}} = \frac{4}{N+4},$$

thus the combination of the two previous inequalities leads to

$$\iint_{Q_r(t_0, x_0)} \mathcal{U}^{\gamma_j} \, dt \, dx \leq \left( \iint_{Q_r(t_0, x_0)} \mathcal{U}^{\gamma_{j-1}} \, dt \, dx \right)^{\frac{4}{N+4}} \left( C \frac{q_j^6}{(R-r)^p} \iint_{Q_R(t_0, x_0)} \left( 1 + \mathcal{U}^{\gamma_j} \right) \, dt \, dx \right)^{\frac{N+2}{N+4}}.$$

We now use the Young inequality

$$\left(\iint_{Q_R(t_0,x_0)} \mathcal{U}^{\gamma_{j-1}}\,dt\,dx\right)^{\frac{4}{N+4}}\left(C\,\frac{q_j^6}{(R-r)^p}\iint_{Q_R(t_0,x_0)}\left(1+\mathcal{U}^{\gamma_j}\right)dt\,dx\right)^{\frac{N+2}{N+4}}$$

$$\leq \frac{N+2}{N+4}\iint_{Q_R(t_0,x_0)}\left(1+\mathcal{U}^{\gamma_j}\right)dt\,dx$$

$$+\frac{2}{N+4}\left(C\,\frac{q_j^6}{(R-r)^p}\right)^{\frac{N+2}{2}}\left(\iint_{Q_R(t_0,x_0)}\mathcal{U}^{\gamma_{j-1}}\,dt\,dx\right)^2.$$

Thus we have proved the estimate

$$(4.7)\qquad
\begin{aligned}
\iint_{Q_r(t_0,x_0)}\mathcal{U}^{\gamma_j}\,dt\,dx &\leq \frac{N+2}{N+4}\iint_{Q_R(t_0,x_0)}\left(1+\mathcal{U}^{\gamma_j}\right)dt\,dx\\
&+\frac{2}{N+4}\left(C\,\frac{q_j^6}{(R-r)^p}\right)^{\frac{N+2}{2}}\left(\iint_{Q_R(t_0,x_0)}\mathcal{U}^{\gamma_{j-1}}\,dt\,dx\right)^2.
\end{aligned}$$

Next, we observe that

$$\frac{|Q_r(t_0,x_0)|}{|Q_R(t_0,x_0)|^2}\leq C\left(\frac{r}{R}\right)^{N+p}\frac{1}{R^{N+p}}\leq \frac{C}{(R-r)^{N+p}}\leq \frac{C}{(R-r)^{p\,\frac{N+2}{2}}},$$

where the last inequality relies on the two facts: $R\leq 1$ and $p>2$. Hence, using that $q_j\geq 1$, one gets

$$\begin{aligned}
|Q_r(t_0,x_0)| &\leq C\left(\frac{q_j^6}{(R-r)^p}\right)^{\frac{N+2}{2}}|Q_R(t_0,x_0)|^2\\
&\leq C\left(\frac{q_j^6}{(R-r)^p}\right)^{\frac{N+2}{2}}\left(\iint_{Q_R(t_0,x_0)}\left(1+\mathcal{U}^{\gamma_{j-1}}\right)dt\,dx\right)^2.
\end{aligned}$$

By summing $|Q_r(t_0,x_0)|$ on both sides of (4.7), this implies

$$\begin{aligned}
\iint_{Q_r(t_0,x_0)}\left(1+\mathcal{U}^{\gamma_j}\right)dt\,dx &\leq \frac{N+2}{N+4}\iint_{Q_R(t_0,x_0)}\left(1+\mathcal{U}^{\gamma_j}\right)dt\,dx\\
&+C\left(\frac{q_j^6}{(R-r)^p}\right)^{\frac{N+2}{2}}\left(\iint_{Q_R(t_0,x_0)}\left(1+\mathcal{U}^{\gamma_{j-1}}\right)dt\,dx\right)^2.
\end{aligned}$$

We can now appeal to [17, Lemma 6.1] and absorb the term on the right-hand side containing $1+\mathcal{U}^{\gamma_j}$, in a standard way. By using the definition of $q_j$, this leads to

$$(4.8)\qquad \iint_{Q_r(t_0,x_0)}\left(1+\mathcal{U}^{\gamma_j}\right)dt\,dx \leq C\,\frac{2^{3(N+2)j}}{(R-r)^{p\,\frac{N+2}{2}}}\left(\iint_{Q_R(t_0,x_0)}\left(1+\mathcal{U}^{\gamma_{j-1}}\right)dt\,dx\right)^2.$$

We want to iterate the previous estimate on a sequence of shrinking parabolic cylinders. We fix two radii $0<r<R\leq 1$, then we consider the sequence

$$R_j = r+\frac{R-r}{2^{j-1}},\qquad j\in\mathbb{N}\setminus\{0\},$$

and we apply (4.8) with $R_{j+1}<R_j$ in place of $r<R$. We introduce the notation

$$Y_j = \iint_{Q_{R_j}(t_0,x_0)}\left(1+\mathcal{U}^{\gamma_{j-1}}\right)dt\,dx.$$

Thus we get

$$Y_{j+1} \leq C\, 2^{2\,p\,(N+2)\,j}\, (R-r)^{-\frac{p}{2}\,(N+2)}\, Y_j^2.$$

By iterating the previous estimate starting from $j = 1$, we obtain for every $n \in \mathbb{N} \setminus \{0\}$,

$$Y_{n+1} \leq \left( C\, 2^{2\,p\,(N+2)}\, (R-r)^{-\frac{p}{2}\,(N+2)} \right)^{\sum\limits_{j=0}^{n-1}(n-j)\,2^j}\, Y_1^{2^n}.$$

possibly for a different constant $C > 1$. We now take the power $2^{-n}$ on both sides:

$$Y_{n+1}^{2^{-n}} \leq \left( C\, 2^{2\,p\,(N+2)}\, (R-r)^{-\frac{p}{2}\,(N+2)} \right)^{\sum\limits_{j=0}^{n-1}(n-j)\,2^{j-n}}\, Y_1 \leq C\,(R-r)^{-p\,(N+2)}\, Y_1,$$

where the last inequality relies on the fact that

$$\sum_{j=0}^{n-1}(n-j)\,2^{j-n} \leq \sum_{j=1}^{\infty} \frac{j}{2^j} = 2.$$

We thus get

$$
\begin{aligned}
\|\mathcal{U}\|_{L^\infty(Q_r(t_0,x_0))} &= \lim_{n\to\infty}\left(\iint_{Q_{R_{n+1}}(t_0,x_0)} \mathcal{U}^{\gamma_n}\, dt\, dx\right)^{\frac{1}{\gamma_n}} \leq \limsup_{n\to+\infty}\left(Y_{n+1}^{2^{-n}}\right)^{\frac{2^n}{\gamma_n}} \\
&\leq \limsup_{n\to+\infty}\left(C\,(R-r)^{-p\,(N+2)}\, Y_1\right)^{\frac{2^n}{\gamma_n}} \\
&\leq C\,(R-r)^{-p\,\frac{N+2}{4}}\left(\iint_{Q_R(t_0,x_0)}\left(1+\mathcal{U}^{p+2}\right) dt\, dx\right)^{\frac{1}{4}}.
\end{aligned}
$$

(4.9)

Here, we have also used that $\gamma_0 = p + 2$ and $\gamma_n \sim 2^{n+2}$, for $n$ going to $\infty$. Finally, in order to remove the dependence on the $L^{p+2}$ norm of the gradient, we use a standard interpolation trick. We write

$$
\begin{aligned}
\left(\iint_{Q_R(t_0,x_0)}\left(1+\mathcal{U}^{p+2}\right) dt\, dx\right)^{\frac{1}{4}} &\leq \left(\|\mathcal{U}\|_{L^\infty(Q_R(t_0,x_0))}^2 \iint_{Q_R(t_0,x_0)} \mathcal{U}^p\, dt\, dx + |Q_R(t_0,x_0)|\right)^{\frac{1}{4}} \\
&\leq \|\mathcal{U}\|_{L^\infty(Q_R(t_0,x_0))}^{\frac{1}{2}}\left(\iint_{Q_R(t_0,x_0)} \mathcal{U}^p\, dt\, dx\right)^{\frac{1}{4}} + C\,R^{\frac{N+p}{4}}.
\end{aligned}
$$

Inserting this estimate into (4.9), we get

$$
\begin{aligned}
\|\mathcal{U}\|_{L^\infty(Q_r(t_0,x_0))} &\leq \frac{C}{(R-r)^{p\,\frac{N+2}{4}}}\left[\|\mathcal{U}\|_{L^\infty(Q_R(t_0,x_0))}^{\frac{1}{2}}\left(\iint_{Q_R(t_0,x_0)} \mathcal{U}^p\, dt\, dx\right)^{\frac{1}{4}} + R^{\frac{N+p}{4}}\right] \\
&\leq \frac{1}{2}\,\|\mathcal{U}\|_{L^\infty(Q_R(t_0,x_0))} + \frac{C}{(R-r)^{p\,\frac{N+2}{2}}}\left(\iint_{Q_R(t_0,x_0)} \mathcal{U}^p\, dt\, dx\right)^{\frac{1}{2}} + \frac{C\,R^{\frac{N+p}{4}}}{(R-r)^{p\,\frac{N+2}{2}}} \\
&\leq \frac{1}{2}\,\|\mathcal{U}\|_{L^\infty(Q_R(t_0,x_0))} + \frac{C}{(R-r)^{p\,\frac{N+2}{2}}}\left[\left(\iint_{Q_R(t_0,x_0)} \mathcal{U}^p\, dt\, dx\right)^{\frac{1}{2}} + 1\right],
\end{aligned}
$$

where in the last line, we have used that $R \leq 1$. By [17, Lemma 6.1] again, we get

$$\|\mathcal{U}\|_{L^\infty(Q_r(t_0,x_0))} \leq \frac{C}{(R-r)^{p\,\frac{N+2}{2}}}\left[\left(\iint_{Q_R(t_0,x_0)} \mathcal{U}^p\, dt\, dx\right)^{\frac{1}{2}} + 1\right].$$

Since $\mathcal{U} \leq |\nabla u| \leq \sqrt{N}\, \mathcal{U}$, this completes the proof. $\qquad\qquad\qquad\qquad\qquad\qquad\square$

**Remark 4.2.** An inspection of the proof reveals that the exponent $\alpha$ in (4.1) can be taken to be

$$\alpha = \begin{cases} \dfrac{N+2}{2}, & \text{if } N \geq 3, \\ \text{any number} > 2, & \text{if } N = 2. \end{cases}$$

In the case $N = 2$, the constant $C$ blows-up as $\alpha$ tends to 2.

## 5. Proof of the Main Theorem

We take $u \in L^p_{\mathrm{loc}}(I; W^{1,p}_{\mathrm{loc}}(\Omega))$ to be a local weak solution of equation (1.1), i.e. it satisfies

$$(5.1) \qquad\qquad -\iint_{I \times \Omega} u\, \varphi_t \, dt\, dx + \iint_{I \times \Omega} \langle \nabla F_0(\nabla u), \nabla \varphi \rangle \, dt\, dx = 0,$$

for every $\varphi \in C^\infty_0(I \times \Omega)$. We recall that $F_0$ indicates the convex function

$$F_0(\xi) = \frac{1}{p} \sum_{i=1}^{N} |\xi_i|^p, \qquad \text{for every } \xi \in \mathbb{R}^N.$$

Our program is as follows:

- we approximate $u$ by solutions $u_\varepsilon$ of the regularized equation;

- we transfer the Lipschitz estimate of Proposition 4.1 from $u_\varepsilon$ to $u$;

- we use a scaling argument to "rectify" the local estimate and obtain (1.4).

We start by recalling that $u$ has the following additional properties: for every subinterval $J \Subset I$ and every open set $\mathcal{O} \Subset \Omega$, we have

$$(5.2) \qquad\qquad u_t \in L^{p'}(J; W^{-1,p'}(\mathcal{O})) \qquad \text{and} \qquad u \in C(\overline{J}; L^2(\mathcal{O})).$$

Here $W^{-1,p'}(\mathcal{O})$ is the topological dual space of $W^{1,p}_0(\mathcal{O})$ and the latter is the completion of $C^\infty_0(\mathcal{O})$ with respect to the $L^p$ norm of the gradient.

We briefly recall the argument to get (5.2), for completeness. Fix $J$ and $\mathcal{O}$ as above, by using equation (5.1) and the fact that $\nabla F_0(\nabla u) \in L^{p'}_{\mathrm{loc}}(I \times \Omega)$, we get

$$\left| \iint_{J \times \mathcal{O}} u\, \psi_t \, dt\, dx \right| = \left| \iint_{J \times \mathcal{O}} \langle \nabla F_0(\nabla u), \nabla \psi \rangle \, dt\, dx \right|$$

$$\leq C \, \|\nabla \psi\|_{L^p(J \times \mathcal{O})} = \|\psi\|_{L^p(J; W^{1,p}_0(\mathcal{O}))}, \qquad \text{for every } \psi \in C^\infty_0(J \times \mathcal{O}).$$

By density, we can extend the linear functional

$$\Lambda : \psi \mapsto \iint_{J \times \mathcal{O}} u\, \psi_t \, dt\, dx,$$

to the whole space $L^p(J; W^{1,p}_0(\mathcal{O}))$. This implies that (see for example [23, Theorem 1.5, Chapter III])

$$\Lambda \in \left( L^p(J; W^{1,p}_0(\mathcal{O})) \right)^* = L^{p'}(J; W^{-1,p'}(\mathcal{O})).$$

By definition of $\Lambda$ and of weak derivative, we get the first property in (5.2).

The second property in (5.2) follows by recalling that for every open set $E \subset \mathbb{R}^N$, we have (see [23, Proposition 1.2, Chapter III])

$$\mathcal{W}_p(J \times E) := \left\{ \varphi \in L^p(J; W^{1,p}_0(E)) \, : \, \varphi_t \in L^{p'}(J; W^{-1,p'}(E)) \right\} \subset C(\overline{J}; L^2(E)).$$

Indeed, it is sufficient to take $\mathcal{O} \Subset \mathcal{O}' \Subset \Omega$ and use the previous inclusion for the function $\theta\, u$, where $\theta \in C_0^\infty(\mathcal{O}')$ is such that $\theta \equiv 1$ on $\mathcal{O}$. Thanks to first property in (5.2) (used with $\mathcal{O}'$ in place of $\mathcal{O}$) and the properties of $\theta$, we have that

$$\theta\, u \in \mathcal{W}_p(J \times \mathcal{O}') \subset C(\overline{J}; L^2(\mathcal{O}')).$$

Since $\theta \equiv 1$ on $\mathcal{O}$, the last fact entails that $u \in C(\overline{J}; L^2(\mathcal{O}))$, as claimed.

In light of (5.2) and since we are only interested in a local result, it is not restrictive to assume from the beginning that

$$u \in L^p(I; W^{1,p}(\Omega)) \cap C(\overline{I}; L^2(\Omega)) \qquad \text{and} \qquad u_t \in L^{p'}(I; W^{-1,p'}(\Omega)).$$

**Part 1: convergence of the approximation scheme**. We remember the definition (3.1) of $F_\varepsilon$ for every $\varepsilon \geq 0$. We then consider the approximating initial boundary value problem parametrized by $\varepsilon > 0$

$$\begin{cases} v_t & = & \mathrm{div}\nabla F_\varepsilon(\nabla v), & \text{in } I \times \Omega, \\ v & = & u, & \text{on } I \times \partial\Omega, \\ v(T_0, \cdot) & = & u(T_0, \cdot), & \text{in } \Omega. \end{cases}$$

By [23, Propositon 4.1, Chapter III], there exists a unique weak solution $u_\varepsilon \in L^p(I; W^{1,p}(\Omega))$ to this problem, such that

$$(u_\varepsilon)_t \in L^{p'}(I; W^{-1,p'}(\Omega)) \qquad \text{and thus } u_\varepsilon \in C(\overline{I}; L^2(\Omega)).$$

The boundary value condition is taken in the sense that

$$u_\varepsilon - u \in L^p(I; W_0^{1,p}(\Omega)),$$

and the initial condition is taken in the $L^2$ sense, which is feasible thanks to the continuity properties of both $u_\varepsilon$ and $u$. The function $u_\varepsilon$ verifies

$$(5.3) \qquad\qquad -\iint_{I \times \Omega} u_\varepsilon\, \varphi_t \, dt\, dx + \iint_{I \times \Omega} \langle \nabla F_\varepsilon(\nabla u_\varepsilon), \nabla\varphi \rangle\, dt\, dx = 0,$$

for every $\varphi \in C_0^\infty(I \times \Omega)$. An integration by parts in (5.3) leads to

$$(5.4) \qquad\qquad \int_I ((u_\varepsilon)_t, \varphi)_{(W^{-1,p'}, W_0^{1,p})}\, dt + \iint_{I \times \Omega} \langle \nabla F_\varepsilon(\nabla u_\varepsilon), \nabla\varphi \rangle\, dt\, dx = 0.$$

By density, the above identity remains true for every $\varphi \in L^p(I; W_0^{1,p}(\Omega))$. Then the choice $\varphi = u_\varepsilon - u$ yields

$$\int_I ((u_\varepsilon)_t, u_\varepsilon - u)_{(W^{-1,p'}, W_0^{1,p})}\, dt + \iint_{I \times \Omega} \langle \nabla F_\varepsilon(\nabla u_\varepsilon), \nabla u_\varepsilon - \nabla u \rangle\, dt\, dx = 0.$$

By recalling the expression (3.1) of $F_\varepsilon$, the previous integral identity can be rewritten as

$$\int_I ((u_\varepsilon)_t, u_\varepsilon - u)_{(W^{-1,p'}, W_0^{1,p})}\, dt + \iint_{I \times \Omega} \langle \nabla F_0(\nabla u_\varepsilon), \nabla u_\varepsilon - \nabla u \rangle\, dt\, dx$$
$$+ \varepsilon \iint_{I \times \Omega} \langle \nabla G(\nabla u_\varepsilon), \nabla u_\varepsilon - \nabla u \rangle\, dt\, dx = 0.$$

Starting from (5.1), we have similarly

$$\int_I (u_t, u_\varepsilon - u)_{(W^{-1,p'}, W_0^{1,p})}\, dt + \iint_{I \times \Omega} \langle \nabla F_0(\nabla u), \nabla u_\varepsilon - \nabla u \rangle\, dt\, dx = 0.$$

Upon subtracting the two identities above, we get

$$(5.5) \qquad \iint_{I \times \Omega} ((u_\varepsilon)_t - u_t, u_\varepsilon - u)_{(W^{-1,p'}, W_0^{1,p})}\, dt + \iint_{I \times \Omega} \langle \nabla F_0(\nabla u_\varepsilon) - \nabla F_0(\nabla u), \nabla u_\varepsilon - \nabla u \rangle\, dt\, dx$$
$$+ \varepsilon \iint_{I \times \Omega} \langle \nabla G(\nabla u_\varepsilon), \nabla u_\varepsilon - \nabla u \rangle\, dt\, dx = 0.$$

For the first term, we rely on

$$\int_I ((u_\varepsilon)_t - u_t, u_\varepsilon - u)_{(W^{-1,p'}, W_0^{1,p})} \, dt = \frac{1}{2} \int_{\{T_1\} \times \Omega} |u_\varepsilon - u|^2 \, dx,$$

which follows from the fact that

$$t \mapsto \frac{1}{2} \int_\Omega |u_\varepsilon(t,x) - u(t,x)|^2 \, dx,$$

is absolutely continuous on $I$, with derivative given exactly by

$$\left( (u_\varepsilon)_t - u_t, u_\varepsilon - u \right)_{(W^{-1,p'}, W_0^{1,p})}, \qquad \text{for a. e. } t \in I,$$

see [23, Proposition 1.2, Chapter III].

For the second term in (5.5), we can use that for every $\xi, \zeta \in \mathbb{R}^N$, we have

$$\langle \nabla F_0(\xi) - \nabla F_0(\zeta), \xi - \zeta \rangle = \sum_{i=1}^N \left( |\xi_i|^{p-2} \xi_i - |\zeta_i|^{p-2} \zeta_i \right) (\xi_i - \zeta_i)$$

$$\geq 2^{2-p} \sum_{i=1}^N |\xi_i - \zeta_i|^p \geq \frac{1}{C} |\xi - \zeta|^p,$$

for some $C = C(N, p) > 0$. The first inequality can be found in [20, Section 10]. On the other hand, the convexity of $G$ implies that $\nabla G$ is a monotone map and thus

$$\langle \nabla G \left( \nabla u_\varepsilon \right), \nabla u_\varepsilon - \nabla u \rangle \geq \langle \nabla G(\nabla u), \nabla u_\varepsilon - \nabla u \rangle.$$

By using these two pointwise estimates in (5.5), we thus get

$$\frac{1}{2} \int_{\{T_1\} \times \Omega} |u_\varepsilon - u|^2 \, dx + \frac{1}{C} \iint_{I \times \Omega} |\nabla u_\varepsilon - \nabla u|^p \, dt \, dx \leq \varepsilon \iint_{I \times \Omega} |\langle \nabla G \left( \nabla u \right), \nabla u_\varepsilon - \nabla u \rangle| \, dt \, dx.$$

By the Cauchy-Schwarz inequality and the inequality $|\nabla G(\xi)| \leq (1 + |\xi|)^{p-1}$, the right-hand side is not larger than

$$\varepsilon \iint_{I \times \Omega} (1 + |\nabla u|)^{p-1} |\nabla u_\varepsilon - u| \, dt \, dx,$$

which, by the Young inequality, in turn can be bounded from above by

$$\frac{p-1}{p} \varepsilon \iint_{I \times \Omega} |\nabla u_\varepsilon - \nabla u|^p \, dt \, dx + \frac{\varepsilon}{p} \iint_{I \times \Omega} (1 + |\nabla u|)^p \, dt \, dx.$$

Then for every $\varepsilon > 0$, we have

$$\frac{1}{2} \int_{\{T_1\} \times \Omega} |u_\varepsilon - u|^2 \, dx + \left( \frac{1}{C} - \frac{p-1}{p} \varepsilon \right) \iint_{I \times \Omega} |\nabla u_\varepsilon - \nabla u|^p \, dt \, dx \leq \frac{\varepsilon}{p} \iint_{I \times \Omega} (1 + |\nabla u|)^p \, dt \, dx.$$

This estimate guarantees that the family $\{\nabla u_\varepsilon\}_{\varepsilon>0}$ is strongly convergent to $\nabla u$ in $L^p(I \times \Omega)$. This is now sufficient to pass to the limit in the uniform Lipschitz estimate (4.1), which will be still valid for $u$, as well. By a covering argument, this in turn implies

$$\nabla u \in L^\infty_{\text{loc}}(I \times \Omega),$$

as claimed.

**Part 2: scale invariant estimate.** We finally focus on obtaining the estimate (1.4). We fix

$$\mathcal{Q}_{\tau,R}(t_0, x_0) = (t_0 - \tau, t_0) \times (x_0 - R, x_0 + R)^N \Subset I \times \Omega,$$

as in the statement of the Main Theorem. By recalling Remark 1.1, we know that for every $R > 0$ and $\mu > 0$ the function

$$U_R(t, x) = \mu \, u(t_0 + \mu^{p-2} R^p \, t, x_0 + R \, x),$$

is still a local weak solution of our parabolic equation, this time in the rescaled set

$$\widetilde{I} \times \widetilde{\Omega} := \left( \frac{T_0 - t_0}{\mu^{p-2} R^p}, \frac{T_1 - t_0}{\mu^{p-2} R^p} \right) \times \frac{\Omega - x_0}{R}.$$

We consider the compactly contained cube $\mathcal{Q}_{\frac{\tau}{\mu^{p-2} R^p}, 1}(0,0) \Subset \widetilde{I} \times \widetilde{\Omega}$ and make the choice

$$\mu^{p-2} R^p = \tau, \qquad \text{that is,} \qquad \mu = \left( \frac{\tau}{R^p} \right)^{\frac{1}{p-2}},$$

so that

$$\mathcal{Q}_{\frac{\tau}{\mu^{p-2} R^p}, 1}(0,0) = \mathcal{Q}_{1,1}(0,0) = Q_1(0,0).$$

From **Part 1**, we know that we can use the a priori estimate (4.1) for $U_R$ on the anisotropic parabolic cubes

$$Q_{\frac{1}{2}}(0,0) = \left( -\frac{1}{2^p}, 0 \right) \times \left( -\frac{1}{2}, \frac{1}{2} \right)^N \qquad \text{and} \qquad Q_1(0,0) = (-1,0) \times (-1,1)^N.$$

This gives

$$\|\nabla U_R\|_{L^\infty(Q_{\frac{1}{2}}(0,0))} \le C \left( \iint_{Q_1(0,0)} |\nabla U_R|^p \, dt \, dx \right)^{\frac{1}{2}} + C.$$

If we scale back, we get

(5.6) $$\|\nabla u\|_{L^\infty(\mathcal{Q}_{\frac{\tau}{2^p}, \frac{R}{2}}(t_0, x_0))} \le C \left( \frac{\tau}{R^2} \right)^{\frac{1}{2}} \left( \fint_{\mathcal{Q}_{\tau, R}(t_0, x_0)} |\nabla u|^p \, dt \, dx \right)^{\frac{1}{2}} + C \left( \frac{R^2}{\tau} \right)^{\frac{1}{p-2}}.$$

This already proves the claimed a priori estimate (1.4) for $0 < \sigma \le 1/2^p$.

For any $\sigma \in (0,1)$, we can take $(t_1, x_1) \in \mathcal{Q}_{\sigma\tau, \sigma R}(t_0, x_0)$ such that

$$\|\nabla u\|_{L^\infty(\mathcal{Q}_{\sigma\tau, \sigma R}(t_0, x_0))} \le \|\nabla u\|_{L^\infty(\mathcal{Q}_{\frac{1-\sigma}{2^p}\tau, \frac{1-\sigma}{2} R}(t_1, x_1))}.$$

Observe that for every $(s, y) \in \mathcal{Q}_{\sigma\tau, \sigma R}(t_0, x_0)$, we have

$$\mathcal{Q}_{\frac{1-\sigma}{2^p}\tau, \frac{1-\sigma}{2} R}(s, y) \subset \mathcal{Q}_{(1-\sigma)\tau, (1-\sigma) R}(s, y) \Subset \mathcal{Q}_{\tau, R}(t_0, x_0).$$

By applying (5.6) on this parabolic cylinder, we get

$$\|\nabla u\|_{L^\infty(\mathcal{Q}_{\sigma\tau, \sigma R}(t_0, x_0))} \le \|\nabla u\|_{L^\infty(\mathcal{Q}_{\frac{1-\sigma}{2^p}\tau, \frac{1-\sigma}{2} R}(t_1, x_1))}$$

$$\le \frac{C}{\sqrt{1-\sigma}} \left( \frac{\tau}{R^2} \right)^{\frac{1}{2}} \left( \fint_{\mathcal{Q}_{(1-\sigma)\tau, (1-\sigma) R}(t_1, x_1)} |\nabla u|^p \, dt \, dx \right)^{\frac{1}{2}}$$

$$+ C (1-\sigma)^{\frac{1}{p-2}} \left( \frac{R^2}{\tau} \right)^{\frac{1}{p-2}}.$$

By observing that

$$\fint_{\mathcal{Q}_{(1-\sigma)\tau, (1-\sigma) R}(t_1, x_1)} |\nabla u|^p \, dt \, dx = \frac{1}{(1-\sigma)^{N+1} (2 R)^N \tau} \iint_{\mathcal{Q}_{(1-\sigma)\tau, (1-\sigma) R}(t_1, x_1)} |\nabla u|^p \, dt \, dx$$

$$\le \frac{1}{(1-\sigma)^{N+1} (2 R)^N \tau} \iint_{\mathcal{Q}_{\tau, R}(t_0, x_0)} |\nabla u|^p \, dt \, dx$$

$$= \frac{1}{(1-\sigma)^{N+1}} \fint_{\mathcal{Q}_{\tau, R}(t_0, x_0)} |\nabla u|^p \, dt \, dx,$$

we eventually reach the desired conclusion.

# References

[1] V. Bögelein, F. Duzaar, P. Marcellini, Parabolic equations with $p, q-$growth, J. Math. Pures Appl., **100** (2013), 535–563. 8

[2] V. Bögelein, F. Duzaar, G. Mingione, The regularity of general parabolic systems with degenerate diffusion, Mem. Amer. Math. Soc., **221** (2013). 2

[3] P. Bousquet, L. Brasco, $C^1$ regularity of orthotropic $p-$harmonic functions in the plane, Anal. PDE, **11** (2018), 813–854. 2, 5

[4] P. Bousquet, L. Brasco, V. Julin, Lipschitz regularity for local minimizers of some widely degenerate problems, Ann. Scuola Norm. Sup. Pisa Cl. Sci. (5), **16** (2016), 1235–1274. 2

[5] P. Bousquet, L. Brasco, C. Leone, A. Verde, On the Lipschitz character of orthotropic $p-$harmonic functions, Calc. Var. Partial Differential Equations, **57** (2018), 57–88. 2, 3, 5

[6] L. Brasco, G. Carlier, On certain anisotropic elliptic equations arising in congested optimal transport: local gradient bounds, Adv. Calc. Var., **7** (2014), 379–407. 2

[7] L. Brasco, C. Leone, G. Pisante, A. Verde, Sobolev and Lipschitz regularity for local minimizers of widely degenerate anisotropic functionals, Nonlinear Anal., **153** (2017), 169–199. 2

[8] Y. Z. Chen, E. DiBenedetto, On the local behavior of solutions of singular parabolic equations, Arch. Rational Mech. Anal., **103** (1988), 319–345. 2

[9] H. J. Choe, Hölder regularity for the gradient of solutions of certain singular parabolic equations, Comm. Partial Differential Equations, **16** (1991), 1709–1732. 2

[10] S. Ciani, S. Mosconi, V. Vespri, Parabolic Harnack estimates for anisotropic slow diffusion, preprint (2020), available at https://arxiv.org/abs/2012.09685 4

[11] F. Demengel, Lipschitz interior regularity for the viscosity and weak solutions of the pseudo $p-$Laplacian equation, Adv. Differential Equations, **21** (2016), 373–400. 2

[12] E. DiBenedetto, *Degenerate Parabolic Equations.* Universitext, Springer-Verlag, New York, 1993. 2, 3, 7, 8

[13] E. DiBenedetto, On the local behavior of solutions of degenerate parabolic equations with measurable coefficients, Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4), **13** (1986), 487–535. 2

[14] E. DiBenedetto, A. Friedman, Regularity of solutions of nonlinear degenerate parabolic systems, J. Reine Angew. Math., **349** (1984), 83–128. 2

[15] E. DiBenedetto, A. Friedman, Hölder estimates for nonlinear degenerate parabolic systems, J. Reine Angew. Math., **357** (1985), 1–22. 2

[16] F. Feo, J. L. Vazquez, B. Volzone, Anisotropic $p-$Laplacian evolution of fast diffusion type, Adv. Nonlinear Stud., **21** (2021), 523–555. 4

[17] E. Giusti, Direct methods in the calculus of variations. World Scientific Publishing Co., Inc., River Edge, NJ, 2003. 22, 23

[18] T. Kuusi, G. Mingione, Riesz potentials and nonlinear parabolic equations, Arch. Ration. Mech. Anal., **212** (2014), 727–780. 2

[19] T. Kuusi, G. Mingione, Potential estimates and gradient boundedness for nonlinear parabolic systems, Rev. Mat. Iberoam., **28** (2012), 535–576. 2

[20] P. Lindqvist, *Notes on the $p-$Laplace equation.* Report. University of Jyväskylä Department of Mathematics and Statistics, **102**. University of Jyväskylä, Jyväskylä, 2006. 26

[21] J.-L. Lions, *Quelques méthodes de résolution des problèmes aux limites non linéaires*, Paris, Dunod, 1969. 1

[22] G. Pisante, A. Verde, Regularity results for non smooth parabolic problems, Adv. Differential Equations, **13** (2008), 367–398. 4

[23] R. E. Showalter, *Monotone Operators in Banach Spaces and Nonlinear Partial Differential Equations.* Mathematical Surveys and Monographs, **49**. American Mathematical Society, Providence, RI, 1997. 1, 24, 25, 26

[24] Al. S. Tersenov, Ar. S. Tersenov, Existence of Lipschitz continuous solutions to the Cauchy-Dirichlet problem for anisotropic parabolic equations, J. Funct. Anal., **272** (2017), 3965–3986 4, 6

[25] K. Uhlenbeck, Regularity for a class of non-linear elliptic systems, Acta Math., **138** (1977), 219–240. 5

[26] N. N. Ural'ceva, Degenerate quasilinear elliptic systems, Zap. Naučn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI), **7** (1968), 184–222. 5

[27] M. I. Vishik, Solubility of boundary-value problems for quasi-linear parabolic equations of higher orders, Mat. Sb. (N.S.), **59 (101)** (1962) suppl., 289–325. 1

[28] M. Wiegner, On $C_\alpha$-regularity of the gradient of solutions of degenerate parabolic systems, Ann. Mat. Pura Appl., **145** (1986), 385–405. 2

[29] E. Zeidler, *Nonlinear functional analysis and its applications II/A – Linear monotone operators*, Springer-Verlag, New York, 1989 1

(P. Bousquet) Institut de Mathématiques de Toulouse, CNRS UMR 5219
Université de Toulouse
F-31062 Toulouse Cedex 9, France.
*E-mail address*: pierre.bousquet@math.univ-toulouse.fr

(L. Brasco) Dipartimento di Matematica e Informatica
Università degli Studi di Ferrara
Via Machiavelli 35, 44121 Ferrara, Italy
*E-mail address*: lorenzo.brasco@unife.it

(C. Leone & A. Verde) Dipartimento di Matematica "R. Caccioppoli"
Università degli Studi di Napoli "Federico II"
Via Cinthia, Complesso Universitario di Monte S. Angelo, 80126 Napoli, Italy
*E-mail address*: chiara.leone@unina.it
*E-mail address*: anna.verde@unina.it