PARTIAL REGULARITY FOR PARABOLIC SYSTEMS WITH VMO-COEFFICIENTS

LEON MONS
Department Mathematik, Universität Erlangen-Nürnberg
Cauerstraße 11, 91058 Erlangen, Germany
(Communicated by Enrico Valdinoci)

Abstract. In this article we establish a partial Hölder continuity result for weak solutions of parabolic systems, where the nonlinear vector field $A(\cdot)$ satisfies a standard $p$-growth condition and a non-degenerate ellipticity condition with respect to the gradient variable, while in the space-time variable $z = (x,t)$ it verifies a VMO-type condition. Thus, no continuity in the space-time variable is assumed. The proof is based on the method of $A$-caloric approximation, applied on suitably chosen intrinsic cylinders.

1. Introduction. The subject of this article are parabolic systems

$$u_t - \text{div} A(z, u, Du) = 0 \quad \text{in } \Omega_T = \Omega \times (-T, 0),$$

where $\Omega \subset \mathbb{R}^n$ ($n \geq 2$) is a bounded domain and $u : \Omega_T \to \mathbb{R}^N$ ($N \geq 2$) is the desired function. The coefficients $A : \Omega_T \times \mathbb{R}^N \times \mathbb{R}^{N \times n} \to \mathbb{R}^{N \times n}$ shall satisfy a standard $p$-growth condition and a non-degenerate ellipticity condition with respect to the gradient variable (as will be specified in (1.3) below). We will prove the partial Hölder continuity of weak solutions, i.e. the local Hölder continuity outside a closed subset with zero parabolic measure, under the assumption that the mapping $z \mapsto A(z, u, \xi)$ merely satisfies a VMO-type condition, and thus might not be continuous.

To put this result into a context, let us start with a brief review of previous regularity results for such systems. As is well-known already from the elliptic case, there is a natural gap between the scalar case $N = 1$ and the vectorial case $N \geq 2$. In the scalar case the continuity of the mapping $z \mapsto A(z, u, \xi)$ together with the usual growth and ellipticity assumptions is sufficient to establish the local Hölder continuity of weak solutions everywhere in $\Omega_T$ [12, 22, 24]. Such an everywhere regularity result, however, is not expected to hold in the vectorial case. The first counterexample (for elliptic systems) was constructed by De Giorgi [10]. A counterexample for parabolic systems can be found in [27], where a discontinuous solution to a system with smooth coefficients is constructed. In the vectorial case, however, weak solutions for (1.1) usually still possess a slightly lower degree of regularity, the so-called partial regularity. By partial regularity we mean that the weak solution is

2020 Mathematics Subject Classification. Primary: 35K40; Secondary: 35B65, 35K59.
Key words and phrases. Parabolic systems, partial regularity, vanishing mean oscillation, intrinsic geometry.

The author has been supported by the FWF-Project P31956-N32 “Doubly nonlinear evolution equations”.

1783
continuous in an open subset $\Omega_0 \subseteq \Omega_T$, such that the complement $\Omega_T \setminus \Omega_0$ has zero measure. We will usually refer to $\Omega_0$ as the regular set and to $\Omega_T \setminus \Omega_0$ as the singular set. There have been a lot of investigations into the partial regularity theory for elliptic and parabolic systems, an account of which can be found in the survey paper [23].

Another interesting aspect of the regularity theory for parabolic (and also elliptic) systems is the gap between the regularity of the partial map

$$(z, u) \mapsto A(z, u, \xi) \quad (1.2)$$

and the (partial) regularity of weak solutions. If this map is Hölder continuous, then the gradient of a weak solution is partially Hölder continuous. For the quadratic case $p = 2$ this was proved in [14], and also an upper bound for the Hausdorff dimension of the singular set was given. By the use of similar techniques this result was extended to the superquadratic case $p > 2$ in [16] and to the subquadratic case $\frac{2n}{n+2} < p < 2$ in [25]. Now one might think that if the map (1.2) were continuous, but not Hölder continuous, then the gradient of the weak solution should be partially continuous. This is however not the case, and it turned out that Dini continuity is the weakest requirement on the map (1.2) under which the partial continuity of the gradient can be obtained [1]. If the map (1.2) is merely continuous, the maximum degree of regularity we can expect is the partial Hölder continuity of the solution itself. For the case $p = 2$ this was proved in [4], while the proof for the case $\frac{2n}{n+2} < p < 2$ is in [18]. For the analogous result in the elliptic setting we refer to [19].

The aim of this article is to go one step further and consider coefficients, that might not even be continuous with respect to the space-time variable $z = (x, t)$, but only satisfy the VMO-type condition (1.5) and (1.6) specified below. In the elliptic case, the partial Hölder continuity of weak solutions under such a condition has already been shown in [2]. Thus, we also close this gap between the elliptic and parabolic case. We will consider both the superquadratic and the subquadratic case, thereby unifying the proof for both cases.

Let us now state the assumptions more explicitly. We consider parabolic systems of the type (1.1), where $\Omega_T = \Omega \times (-T, 0)$ is a space-time cylinder, $z = (x, t)$ and $u : \Omega_T \rightarrow \mathbb{R}^N$. We assume that the coefficients $A : \Omega_T \times \mathbb{R}^N \times \mathbb{R}^{N \times n} \rightarrow \mathbb{R}^{N \times n}$ are differentiable with respect to $\xi$ and fulfil a standard $p$-growth condition and a non-degenerate ellipticity condition:

$$
\begin{cases}
|A(z, u, \xi)| + (1 + |\xi|) |D\xi A(z, u, \xi)| \leq L(1 + |\xi|)^{p-1} \\
\langle D\xi A(z, u, \xi) \xi, \xi \rangle \geq \nu(1 + |\xi|)^{p-2} |\xi|^2
\end{cases}
(1.3)
$$

for all $z \in \Omega_T$, $u \in \mathbb{R}^N$ and $\xi, \tilde{\xi} \in \mathbb{R}^{N \times n}$. Here $0 < \nu \leq L$ are some structure constants and $\frac{2n}{n+2} < p < \infty$ is the growth exponent.

**Remark 1.1.** By $\langle \cdot, \cdot \rangle$ we denote the standard scalar product on $\mathbb{R}^{N \times n}$ and by $\cdot$ the standard scalar product on $\mathbb{R}^N$, $\mathbb{R}^n$ or $\mathbb{R}$.

Furthermore, let us assume that $A$ is continuous with respect to $u$ and $D\xi A$ is continuous with respect to $\xi$ with the following estimates:

$$
\begin{cases}
|A(z, u, \xi) - A(z, v, \xi)| \leq L\omega \left( |u - v|^2 \right) (1 + |\xi|)^{p-1} \\
|D\xi A(z, u, \xi) - D\xi A(z, v, \tilde{\xi})| \leq L\mu \left( \frac{|\xi - \tilde{\xi}|}{1 + |\xi| + |\tilde{\xi}|} \right) (1 + |\xi| + |\tilde{\xi}|)^{p-2}
\end{cases}
(1.4)
$$
for all $z \in \Omega_T$, $u, \overline{u} \in \mathbb{R}^N$ and $\xi, \overline{\xi} \in \mathbb{R}^{N \times n}$. Here $\omega, \mu : [0, \infty) \to [0, 1]$ are two moduli of continuity, i.e. concave, non-decreasing and $\omega(0) = \mu(0) = 0$. Finally, we assume that the partial map

$$A(\cdot, u, \xi) : \Omega_T \to \mathbb{R}^N, \ z \mapsto A(z, u, \xi)$$

lies in the space $\text{VMO}(\Omega_T, \mathbb{R}^N)$ of functions with vanishing mean oscillation, for any $u \in \mathbb{R}^N$ and $\xi \in \mathbb{R}^{N \times n}$. More explicitly, we assume that the following estimate holds:

$$|A(z, u, \xi)(z) - A(z, u, \xi)(z_0)| \leq v_{\omega, \mu}(z)(1 + |\xi|)^{p-1} \text{ for all } z \in Q_{\rho, \sigma}(z_0) \cap \Omega_T$$

whenever $z_0 \in \Omega_T$, $0 < \rho, \sqrt{\sigma} \leq \rho_0$, $u \in \mathbb{R}^N$ and $\xi \in \mathbb{R}^{N \times n}$, where $\rho_0 > 0$ is a constant and $v_{\omega, \mu, \sigma} : \Omega_T \to [0, 2L]$ are bounded functions satisfying $\lim_{r \to 0} \frac{v_0}{V}(r) = 0$, where

$$V(r) = \sup \left\{ \int_{Q_{\rho, \sigma}(z_0) \cap \Omega_T} v_{\omega, \mu, \sigma}(z) \, dz \mid \max \{ \rho^2, \sigma \} \leq r \right\}. \quad (1.5)$$

Here we used the notation $Q_{\rho, \sigma}(z_0) := B_\rho(x_0) \times (t_0 - \sigma, t_0)$ for a general parabolic cylinder with centre $z_0 = (x_0, t_0) \in \mathbb{R}^{n+1}$, radius $\rho > 0$ and height $\sigma > 0$, and

$$A(\cdot, u, \xi)(z_0) := \int_{Q_{\rho, \sigma}(z_0) \cap \Omega_T} A(z, u, \xi) \, dz$$

for the integral mean value of $A(\cdot, u, \xi)$ over $Q_{\rho, \sigma}(z_0) \cap \Omega_T$.

**Remark 1.2.** The moduli of continuity $\omega, \mu$ are concave and hence sublinear in the sense that

$$\omega(cs) \leq c \omega(s), \ \mu(cs) \leq c \mu(s) \quad \forall c \geq 1, s \geq 0.$$

We will work with the following notion of weak solution:

**Definition 1.3.** A function $u \in L^p(-T, 0; W^{1,p}(\Omega, \mathbb{R}^N)) \cap C^0(0-T, 0; L^2(\Omega, \mathbb{R}^N))$ is a weak solution of (1.1), if the weak formulation

$$\int_{\Omega_T} u \cdot \varphi_t - \langle A(z, u, Du), D\varphi \rangle \, dz = 0 \quad (1.7)$$

is satisfied for any test function $\varphi \in C_0^\infty(\Omega_T, \mathbb{R}^N)$.

For the formulation of the main theorem we need to define the space of functions that are locally Hölder continuous with respect to the parabolic metric.

**Definition 1.4.** The parabolic distance of two points $z_1 = (x_1, t_1), z_2 = (x_2, t_2) \in \mathbb{R}^{n+1}$ is defined as

$$d_p(z_1, z_2) := \max \left\{ |x_1 - x_2|, \sqrt{|t_1 - t_2|} \right\}.$$ 

Let $\alpha \in (0, 1)$ and $\Omega_0 \subset \mathbb{R}^{n+1}$ be open. By $C_0^{0,\alpha/2}(\Omega_0, \mathbb{R}^N)$ we denote the space of all functions $u : \Omega_0 \to \mathbb{R}^N$ that are locally $\alpha$-Hölder continuous with respect to the parabolic metric. This means that $u \in C_0^{0,\alpha/2}(\Omega_0, \mathbb{R}^N)$, if for any compactly contained subset $K \subset \Omega_0$ there exists some constant $L \in [0, \infty)$, such that for all $z_1, z_2 \in K$ the following estimate holds:

$$|u(z_1) - u(z_2)| \leq L \cdot d_p(z_1, z_2)^\alpha.$$

Our main result reads as follows:
Theorem 1.5. Let \( u \in L^p(-T, 0; W^{1,p}(\Omega, \mathbb{R}^N)) \cap C^0(-T, 0; L^2(\Omega, \mathbb{R}^N)) \) be a weak solution of (1.1) under the structure assumptions (1.3) - (1.6). Let \( \alpha \in (0, 1) \) be arbitrary. Then \( u \) is partially \( \alpha \)-Hölder continuous with respect to the parabolic metric, i.e. there exists an open subset \( \Omega_0 \subseteq \Omega_T \) with \( |\Omega_T \setminus \Omega_0| = 0 \) such that \( u \in C^{0,\alpha/2}(\Omega_0, \mathbb{R}^N) \). Moreover we have that the singular set fulfills \( \Omega_T \setminus \Omega_0 \subseteq \Sigma_1 \cup \Sigma_2 \), where \( \Sigma_1 \) and \( \Sigma_2 \) are defined in the following way:

\[
\Sigma_1 := \left\{ z_0 \in \Omega_T : \liminf_{\epsilon \to 0} \int_{Q_\epsilon(z_0)} \| Du - (Du)_{z_0,\epsilon} \|^p \, dz > 0 \right\},
\]

\[
\Sigma_2 := \left\{ z_0 \in \Omega_T : \limsup_{\epsilon \to 0} \| (Du)_{z_0,\epsilon} \| = \infty \right\}.
\]

Remark 1.6. We note that in particular every Lebesgue point of \( Du \) is contained in the regular set \( \Omega_0 \), and hence \( |\Omega_T \setminus \Omega_0| = 0 \).

We want to conclude this introduction with a very brief outline of the proof of Theorem 1.5. Let \( u \) be a weak solution of (1.1) and \( z_0 \in \Omega_T \) be a regular point, i.e. \( z_0 \in \Omega_0 \). The main idea of the proof is to conduct a local comparison between the weak solution \( u \) and the solution \( v \) of a linearized system

\[
\int_{Q_\epsilon(z_0)} v \cdot \varphi_t - A(Dv, D\varphi) \, dz = 0 \quad \forall \varphi \in C^\infty(Q_\epsilon(z_0), \mathbb{R}^N),
\]

where \( A : \mathbb{R}^{N \times n} \times \mathbb{R}^{N \times n} \to \mathbb{R} \) is an elliptic and bounded bilinear form. For \( A \)-caloric functions \( v \), i.e. solutions of the linearized system (1.10), we know good decay estimates that can be transferred onto \( u \) via the \( A \)-caloric approximation lemma [14, 16, 25]. These decay estimates can then be translated into the preservation of a smallness property of a certain excess functional (as defined in (3.2) below). Once this smallness property is established, one can show that there exists some radius \( R > 0 \), such that for any \( z \in Q_R(z_0) \) and any sufficiently small radius \( 0 < r \ll 1 \) the following estimate holds:

\[
\int_{Q_{r}(z)} \left| u - (u)_{Q_{r}(z)} \right|^2 \, dz \leq C r^{n+2+2\alpha}.
\]

The Hölder continuity of \( u \) then follows from a standard embedding result due to Campanato and Da Prato [8], see Theorem 2.10 below.

The iteration technique that is needed to establish the smallness of the excess functional, however, does not work on the usual parabolic cylinders

\[
Q_\epsilon(z_0) = B_\epsilon(x_0) \times (t_0 - \epsilon^2, t_0).
\]

Instead we will have to work on scaled parabolic cylinders

\[
Q_\epsilon^{(\lambda)}(z_0) = B_\epsilon(x_0) \times \Lambda_\epsilon^{(\lambda)}(t_0),
\]

where \( \Lambda_\epsilon^{(\lambda)}(t_0) \) is defined as

\[
\Lambda_\epsilon^{(\lambda)}(t_0) = (t_0 - \lambda^2 t_0^2, t_0).
\]

In these definitions \( \lambda \geq 1 \) is some scaling parameter. If \( \lambda = 1 \) we will write \( Q_\epsilon(z_0) \) and \( \Lambda_\epsilon(t_0) \) for \( Q_\epsilon^{(1)}(z_0) \) respectively \( \Lambda_\epsilon^{(1)}(t_0) \). Since the scaling parameter \( \lambda \) will be put into relation with the weak solution \( u \), we will refer to these cylinders as intrinsic cylinders. The use of such intrinsic cylinders goes back to the work of DiBenedetto and Friedman [11] and is nowadays a standard tool in the regularity...
theory for parabolic systems. An extensive treatise of this subject can be found in [12].

**Remark 1.7.** For a scaled parabolic cylinder $Q_t^{(\lambda)}(z_0)$, as introduced in (1.12), we will write $v_{z_0,\lambda}^{(\lambda)}$ instead of $v_{z_0,\lambda}^{(\lambda)-r^2\rho^2}$.

This article is structured in the following way: In the second section we have collected some preliminaries that will be needed later on. In the third and fourth section we prove the Caccioppoli and Poincaré type inequalities in the specific form, in which we will need them. In the fifth section we prove a linearization lemma that is needed to make the $A$-caloric approximation technique applicable. In the last section we will prove an excess decay estimate for $u$. We will then iterate this excess decay estimate to obtain an estimate like (1.11) from which we can finally deduce the partial Hölder continuity of $u$.

2. Preliminaries. The following Lemma from [6, Lemma 2.6] will be useful for the proof of the Caccioppoli inequality.

**Lemma 2.1.** Let $\sigma > -1$ and $k \in \mathbb{N}$. There exists a constant $C(\sigma) \geq 1$ such that for every $A, B \in \mathbb{R}^k$ the following inequality holds:

$$C^{-1}(1 + |A| + |B|)^\sigma \leq \int_0^1 (1 + |A + sB|)^\sigma \, ds \leq C(1 + |A| + |B|)^\sigma.$$ 

The following standard iteration lemma [20, Lemma 6.1] will be needed for the proof of the Poincaré inequality in the subquadratic case.

**Lemma 2.2.** Let $Z : [r, R] \to [0, \infty)$ be a bounded non-negative function. Suppose that for any $r \leq t < s \leq R$ we have

$$Z(t) \leq \vartheta Z(s) + \frac{A}{(s-t)^\alpha} + B$$

where $A, B \geq 0$, $\alpha > 0$ and $\vartheta \in (0, 1)$. Then there exists a constant $c(\alpha, \vartheta)$ such that

$$Z(r) \leq c(\alpha, \vartheta) \left[ \frac{A}{(R-r)^\alpha} + B \right].$$

2.1. The $V$-function. In the subquadratic setting it usually becomes necessary to work with the function $V : \mathbb{R}^k \to \mathbb{R}$ defined by

$$V(A) = (1 + |A|^2)^{\frac{p-2}{2}} A.$$ (2.1)

We recall the following properties of $V$ for later reference [7, Lemma 2.1].

**Lemma 2.3.** Let $k \in \mathbb{N}$, $p \in (1, 2)$ and $V : \mathbb{R}^k \to \mathbb{R}$ be the function defined in (2.1). Then we have for any $A, B \in \mathbb{R}^k$ and $r > 0$:

1. $2^{\frac{p-2}{2}} \min \left\{ |A|, |A|^{\frac{p}{2}} \right\} \leq |V(A)| \leq \min \left\{ |A|, |A|^{\frac{p}{2}} \right\}$,
2. $|V(rA)| \leq \max \left\{ r, r^{\frac{p}{2}} \right\} |V(A)|$,
3. $\frac{r^{\frac{p-2}{2}} |A-B|}{(1+|A|^2+|B|^2)^{\frac{p-2}{2}}} \leq |V(A) - V(B)| \leq c(k,p) \frac{|A-B|}{(1+|A|^2+|B|^2)^{\frac{p-2}{2}}}$,
4. $|V(A) - V(B)| \leq c(k,p) |V(A-B)|$,
5. $|V(A-B)| \leq c(p,M) |V(A) - V(B)|$ if $|B| \leq M$. 

2.2. Minimizing affine functions. For \( u \in L^2(Q_{e}^{(t)}(z_0), \mathbb{R}^N) \) we denote by \( \ell_{z_0,e}^{(t)} \) the unique minimizer of the functional
\[
F_2[\ell] := \int_{Q_{e}^{(t)}(z_0)} |u - \ell|^2 \, dz
\]
amongst all affine functions \( \ell : \mathbb{R}^n \to \mathbb{R}^N \) that are independent of \( t \). It is easy to check that the minimizer is given by
\[
(u)^{(t)}_{z_0,e} = \int_{Q_{e}^{(t)}(z_0)} u \, dz,
A^{(t)}_{z_0,e} = \frac{n + 2}{\varrho^2} \int_{Q_{e}^{(t)}(z_0)} u \otimes (x - x_0) \, dz. \tag{2.2}
\]
In the case \( p \neq 2 \) the function \( \ell^{(t)}_{z_0,e} \) is still a quasiminimizer of the functional
\[
F_p[\ell] := \int_{Q_{e}^{(t)}(z_0)} |u - \ell|^p \, dz,
\]
as specified in the following lemma [3, Lemma 2.8].

**Lemma 2.4.** Let \( p \geq 1, Q_{e}^{(t)}(z_0) \subseteq \mathbb{R}^{n+1} \) be a scaled parabolic cylinder and \( u \in L^p(Q_{e}^{(t)}(z_0), \mathbb{R}^N) \). There exists a universal constant \( C(n,p) \geq 1 \) such that for every affine map \( \ell : \mathbb{R}^n \to \mathbb{R}^N \) the following inequality holds:
\[
F_p \left[ \ell_{z_0,e}^{(t)} \right] \leq C(n,p) F_p[\ell].
\]

The following lemma [21, Lemma 2.2] will also be useful:

**Lemma 2.5.** For any \( \xi \in \mathbb{R}^N \) and \( w \in \mathbb{R}^{N \times n} \) there holds
\[
\left| A_{z_0,e}^{(t)} - w \right|^2 \leq \frac{n(n + 2)}{\varrho^2} \int_{Q_{e}^{(t)}(z_0)} |u - \xi - w(x - x_0)|^2 \, dz. \tag{2.3}
\]

2.3. A-caloric approximation. We can combine the \( A \)-caloric approximation lemmata for the subquadratic case [18, Lemma 5] and the superquadratic case [16, Lemma 3.2] to obtain the following unified formulation:

**Lemma 2.6.** Given \( \varepsilon > 0, 0 < \nu \leq L \) and \( p > \frac{2n}{n+2} \), there exist positive numbers \( \delta = \delta(N, n, p, \nu, L, \varepsilon) \in (0, 1] \) and \( C = C(n, p) \geq 1 \) with the following property: For any \( \gamma \in (0, 1] \) and any bilinear form \( A \) on \( \mathbb{R}^N \) with ellipticity constant \( \nu \) and upper bound \( L \); whenever \( w \in L^p(\Lambda_{e}(t_0); W^{1,p}(B_{e}(x_0), \mathbb{R}^N)) \cap C^0(\Lambda_{e}(t_0); L^2(B_{e}(x_0), \mathbb{R}^N)) \) satisfying
\[
\begin{align*}
&\left\{ \sup_{t \in \Lambda_{e}} \int_{B_{e}} \frac{w(x,t)}{\varrho} \, dx + \int_{Q_{e}} \frac{w^p}{\varrho} + \frac{1}{p-2} \left| \frac{w}{\varrho} \right|^p + |V(Dw)|^2 \, dz \leq 1, \quad p < 2 \right. \\
&\left. \int_{Q_{e}} \frac{w^2}{\varrho} + |Dw|^2 + \frac{1}{p-2} \left( \frac{w^p}{\varrho} + |Dw|^p \right) \, dz \leq 1, \quad p \geq 2 \right. \tag{2.4}
\end{align*}
\]
is approximately \( A \)-caloric, in the sense that there holds
\[
\left| \int_{Q_{e}^{(t_0)}} w \cdot \varphi_t - A(Dw, D\varphi) \, dz \right| \leq \delta \sup_{Q_{e}^{(t_0)}} |D\varphi| \quad \forall \varphi \in C^\infty_0(Q_{e}(z_0), \mathbb{R}^N), \tag{2.5}
\]
then there exists an \( A \)-caloric function \( h \in L^2(\Lambda_{e/2}(t_0); W^{1,2}(B_{e/2}(x_0), \mathbb{R}^N)) \), i.e.
\[
\int_{Q_{e/2}^{(t_0)}} h \cdot \varphi_t - A(Dh, D\varphi) \, dz = 0 \quad \forall \varphi \in C^\infty_0(Q_{e/2}(z_0), \mathbb{R}^N),
\]
such that

\[
\frac{\int_{Q_{\ell/2}(z_0)} \left| \frac{h}{\ell/2} \right|^2 + \gamma^{p-2} \left| \frac{h}{\ell/2} \right|^p + |V(Dh)|^2}{\ell/2} \, dz \leq C, \quad p < 2
\]

\[
\frac{\int_{Q_{\ell/2}(z_0)} \left| \frac{h}{\ell/2} \right|^2 + |Dh|^2 + \gamma^{p-2} \left( \left| \frac{h}{\ell/2} \right|^p + |Dh|^p \right)}{\ell/2} \, dz \leq C, \quad p \geq 2
\]

and

\[
\frac{\int_{Q_{\ell/2}(z_0)} \left( \frac{w-h}{\ell/2} \right)^2 + \gamma^{p-2} \left| \frac{w-h}{\ell/2} \right|^p}{\ell/2} \, dz \leq \varepsilon.
\]

**Remark 2.7.** We will not reproduce the proof of this lemma. We do note, however, that the requirement \( p > \frac{2n}{n+2} \) is essential for this result to hold. The proof is based on a famous compactness result by Simon [26]. In the parabolic context it requires the compactness of the embedding

\[ W^{1,p}(B_{\varepsilon}, \mathbb{R}^N) \hookrightarrow L^2(B_{\varepsilon}, \mathbb{R}^N) \]

which holds if and only if \( p > \frac{2n}{n+2} \). The \( \mathcal{A} \)-caloric approximation technique is based on a harmonic approximation lemma of De Giorgi [9], which was first adapted to the case of elliptic systems in [13, 17]. A first parabolic version of such a harmonic approximation lemma is due to Duzaar and Mingione [14]. An overview of several such harmonic approximation lemmata, both in the elliptic and in the parabolic setting, can be found in the survey paper [15].

For these \( \mathcal{A} \)-caloric functions there are good a priori estimates available as stated in the following lemma [16, Lemma 4.7], which is based on the work of Campanato [5]:

**Lemma 2.8.** Let \( h \in L^2(\Lambda_{\varphi}(t_0); W^{1,2}(B_{\varepsilon}(x_0), \mathbb{R}^N)) \) be a weak solution in \( Q_{\varepsilon}(z_0) \) of the following linear parabolic system with constant coefficients:

\[
\frac{\int_{Q_{\varepsilon}(z_0)} h \cdot \varphi_t - \mathcal{A}(Dh, D\varphi) \, dz}{\varepsilon} = 0 \quad \forall \varphi \in C_0^\infty(\Omega_{\varepsilon}(z_0), \mathbb{R}^N),
\]

where the coefficients \( \mathcal{A} \) are elliptic with constant \( \nu > 0 \) and bounded with constant \( L \geq \nu \). Then \( h \) is smooth, i.e. \( h \in C^\infty(\Omega_{\varepsilon}(z_0), \mathbb{R}^N) \), and for all \( s \geq 1 \) there exists a constant \( c_{pa} = c_{pa}(N, n, \nu, L, s) \geq 1 \) such that for any affine function \( \ell : \mathbb{R}^n \rightarrow \mathbb{R}^N \) the following estimate holds:

\[
\frac{\int_{Q_{\varepsilon}(z_0)} \left| h - \ell \right|^s}{\varepsilon} \, dz \leq c_{pa} \theta^s \frac{\int_{Q_{\varepsilon}(z_0)} \left| h - \ell \right|^s}{\varepsilon} \, dz,
\]

for every \( \theta \in (0, 1) \).

2.4. **Campanato spaces.** In order to prove the local Hölder continuity of \( u \) in some point \( z_0 \in \Omega_T \) we will show that \( u \) belongs to a certain Campanato space in a neighbourhood of that point. The parabolic Campanato spaces can be defined in the following way [8, Definition 1.2].

**Definition 2.9.** Let \( Q_{R}(z_0) = B_{R}(x_0) \times (t_0 - R^2, t_0) \subseteq \Omega_T \) be a standard parabolic cylinder. For \( z \in Q_{R}(z_0) \) and \( \varrho > 0 \) we define

\[ Q(z, \varrho) := Q_{\varrho}(z) \cap Q_{R}(z_0). \]
Let $p \geq 1$ and $\theta \geq 0$. The parabolic Campanato Space $L^{p,\theta}(Q_R(z_0), \mathbb{R}^N)$ consists of all functions $u \in L^p(Q_R(z_0), \mathbb{R}^N)$ such that the integral seminorm

$$
[u]_{L^{p,\theta}(Q_R(z_0), \mathbb{R}^N)}^p := \sup_{z \in Q_R(z_0), q \geq 0} \left\{ \frac{1}{|Q(z,q)|^{\frac{n}{p}}} \int_{Q(z,q)} |u - (u)_{Q(z,q)}| \, dz \right\}
$$

is finite, where $(u)_{Q(z,q)}$ denotes the integral average of $u$ over $Q(z,q)$.

From [8, Theorem 3.1] we obtain the following characterization of Hölder continuous functions.

**Theorem 2.10.** Let $p \geq 1$ and $\theta > 1$. Define $\alpha := \frac{n+2}{p} (\theta - 1)$ Then we have the following isomorphy:

$$
L^{p,\theta}(Q_R(z_0), \mathbb{R}^N) \cong C^{0,\alpha/2}(\overline{Q_R(z_0)}, \mathbb{R}^N).
$$

**Remark 2.11.** If $\alpha > 1 \iff \theta > 1 + \frac{p}{n+2}$, the space $L^{p,\theta}(Q_R(z_0), \mathbb{R}^N)$ only consists of constant functions.

### 3. Caccioppoli inequality

In order to simplify the notation a bit, for $\lambda \geq 1$ let us define

$$
\lambda_s := \begin{cases} 
\lambda, & p \in \left( \frac{2n}{n+2}, 2 \right) \\
1, & p \in [2, \infty).
\end{cases} 
$$

For $z_0 \in \Omega_T$, $\varrho \in (0,1)$ and an affine function $\ell : \mathbb{R}^n \to \mathbb{R}^N$, we define the excess functionals $\Psi^*_\lambda$ and $\Psi_\lambda$ in the following way:

$$
\Psi^*_\lambda(z_0, \varrho, \ell) := \Psi_\lambda(z_0, \varrho, \ell) + \omega \left( \int_{Q^\lambda(z_0)} |u - \ell(x)|^2 \, dz \right) + V \left( \lambda^2 - p, \varrho^2 \right),
$$

where

$$
\Psi_\lambda(z_0, \varrho, \ell) := \int_{Q^\lambda(z_0)} \frac{|u - \ell|^p}{2^p (1 + |D\ell|^p)^2} + \frac{|u - \ell|^p}{\varrho^p (1 + |D\ell|^p)^p} \, dz.
$$

We can prove the following Caccioppoli inequality for weak solutions on scaled parabolic cylinders:

**Lemma 3.1.** Let $u \in L^p(-T,0; W^{1,p}(\Omega, \mathbb{R}^N)) \cap C^0(-T,0; L^2(\Omega, \mathbb{R}^N))$ be a weak solution of (1.1) and let $M \geq 1$. There exists a positive constant $C_{cacc}(n, p, \nu, L, M)$ such that: If $Q^\lambda(z_0) \subseteq \Omega_T$ is a scaled parabolic cylinder with scaling parameter $\lambda \geq 1$ and radius $\varrho \in (0,1)$, such that $\lambda^2 - p, \varrho^2 \leq \varrho_0$ holds, and if $\ell : \mathbb{R}^n \to \mathbb{R}^N$ is an affine function (independent of $t$) which satisfies the intrinsic coupling

$$
\lambda \leq 1 + |D\ell| \leq M\lambda,
$$

then the following Caccioppoli type inequality holds for any $\sigma \in [\varrho, \varrho_0]$:

$$
\sup_{x \in \Lambda^\lambda_{s_0}(t_0)} \int_{B_{s_0}(x_0)} \frac{|u(x,s) - \ell|^2}{\lambda^{s_0^2 - p} \sigma^2} \, dx + \int_{Q^\lambda(z_0)} (1 + |D\ell| + |D\ell - D\ell|)^{p-2} |D\ell - D\ell|^2 \, dz
\leq C_{cacc} \left( \frac{\varrho}{\varrho - \sigma} \right)^{\max\{p,2\}} \Psi^*_\lambda(z_0, \varrho, \ell).
$$

(3.4)
Proof. In order to simplify the notation we will w.l.o.g. assume that $z_0 = (x_0, t_0) = (0, 0)$ and write $Q_{x}^{(\lambda)}, B_{t}$ and $A_{x}^{(\lambda)}$ instead of $Q_{x}^{(\lambda)}(0), B_{t}(0)$ and $A_{x}^{(\lambda)}(0)$ respectively. With respect to space we choose a cut-off function $\eta \in C_0^\infty (B_\sigma, [0, 1])$ which satisfies $\eta = 1$ in $B_\sigma$ and

$$|D\eta| \leq \frac{C}{\varrho - \sigma}.$$  

With respect to time we define a cut-off function $\zeta \in W_0^{1, \infty} \left( A^{(\lambda)}_{t}, [0, 1] \right)$ via

$$\zeta(t) = \begin{cases} \frac{1}{\lambda^2 - \rho^2} (t + \lambda^2 \rho^2), & \text{on } (-\lambda^2 \rho^2, - \lambda^2 \sigma^2), \\ 1, & \text{on } (-\lambda^2 \sigma^2, \sigma), \\ \frac{1}{2} (s + \varepsilon - t), & \text{on } (s, s + \varepsilon), \\ 0, & \text{on } (s + \varepsilon, 0), \end{cases}$$

for some $s \in A_{t}^{(\lambda)}$ and $\varepsilon \in (0, |s|)$. As a test function in the weak formulation (1.7) we choose

$$\varphi(x, t) := \eta^{2p}(x) \zeta(t)(u(x, t) - \ell(x)).$$

Note that this is actually not an admissible test function since it is not smooth and does not even possess a weak derivative with respect to time. The following formal computations, however, can be made rigorous with a standard smoothing procedure with respect to time, as for instance via Steklov averages. Observe now that

$$\int_{Q_{x}^{(\lambda)}} \ell \cdot \partial_t \varphi \, dz = 0 = \int_{Q_{x}^{(\lambda)}} \left\langle \left( A(\cdot, \ell(0), D\ell) \right)_{Q_{x}^{(\lambda)}}, D\varphi \right\rangle \, dz$$

and

$$D\varphi = \eta^{2p} \zeta(Du - D\ell) + 2p \eta^{2p-1} \zeta((u - \ell) \otimes D\eta).$$

We can thus rewrite the weak formulation (1.7) to obtain

$$I := \int_{Q_{x}^{(\lambda)}} \eta^{2p}(x) \zeta(t)(A(z, u, Du) - A(z, u, D\ell), Du - D\ell) \, dz$$

$$= -2p \int_{Q_{x}^{(\lambda)}} \eta^{2p-1}(x) \zeta(t)(A(z, u, Du) - A(z, u, D\ell), (u - \ell) \otimes D\eta) \, dz$$

$$- \int_{Q_{x}^{(\lambda)}} \left\langle A(z, u, D\ell) - A(z, \ell(0), D\ell), D\varphi \right\rangle \, dz$$

$$- \int_{Q_{x}^{(\lambda)}} \left\langle A(z, \ell(0), D\ell) - (A(\cdot, \ell(0), D\ell))_{Q_{x}^{(\lambda)}}, D\varphi \right\rangle \, dz$$

$$+ \int_{Q_{x}^{(\lambda)}} (u - \ell) \cdot \partial_t \varphi \, dz =: II + III + IV + V.$$  

We will now derive some estimates for these five terms. An application of Lemma 2.1 and the ellipticity condition (1.3) yields the following estimate for the left-hand side:

$$I = \int_{Q_{x}^{(\lambda)}} \eta^{2p}(x) \zeta(t) \int_{0}^{1} (D_x A(z, u, D\ell + s(Du - D\ell))(Du - D\ell), Du - D\ell) \, ds \, dz$$

$$\geq \nu \int_{Q_{x}^{(\lambda)}} \eta^{2p}(x) \zeta(t) \int_{0}^{1} (1 + |D\ell + s(Du - D\ell)|)^{p-2} \, ds |Du - D\ell|^2 \, dz$$

$$\geq \frac{1}{C(p, \nu)} \int_{Q_{x}^{(\lambda)}} \eta^{2p}(x) \zeta(t)(1 + |D\ell| + |Du - D\ell|)^{p-2} |Du - D\ell|^2 \, dz.$$
For the first term on the right-hand side we can again apply Lemma 2.1 and the growth condition \((1.3)_1\) to obtain

\[ |II| \leq C(p, L) \int_{Q_{\delta}^{(\lambda)}} \eta^{2p-1}(x)\zeta(t)(1 + |D\ell| + |Du - D\ell|)^{p-2} |Du - D\ell| \left| \frac{u - \ell}{\varrho - \sigma} \right|^2 \, dz. \]

An application of Young’s inequality yields that

\[ |II| \leq \delta \int_{Q_{\delta}^{(\lambda)}} \eta^{2p}(x)\zeta(t)(1 + |D\ell| + |Du - D\ell|)^{p-2} |Du - D\ell|^2 \, dz \]

\[ + C(\delta, p, L) \int_{Q_{\delta}^{(\lambda)}} \eta^{2p-2}(x)\zeta(t)(1 + |D\ell| + |Du - D\ell|)^{p-2} |Du - D\ell|^2 \, dz. \]

for some \(\delta \in (0, 1)\) to be chosen later. In the subquadratic case \(p < 2\) we use the fact that \(\eta, \zeta \leq 1, 2p - 2 \geq 0\) to conclude with

\[ |II| \leq \delta \int_{Q_{\delta}^{(\lambda)}} \eta^{2p}(x)\zeta(t)(1 + |D\ell| + |Du - D\ell|)^{p-2} |Du - D\ell|^2 \, dz \]

\[ + C(\delta, p, L)(1 + |D\ell|)^{p-2} \int_{Q_{\delta}^{(\lambda)}} \left| \frac{u - \ell}{\varrho - \sigma} \right|^2 \, dz. \]

In the case \(p \geq 2\) we apply Young’s inequality for a second time and obtain

\[ \int_{Q_{\delta}^{(\lambda)}} \eta^{2p-2}(x)\zeta(t)(1 + |D\ell| + |Du - D\ell|)^{p-2} \left| \frac{u - \ell}{\varrho - \sigma} \right|^2 \, dz \]

\[ \leq 2^{p-3}(1 + |D\ell|)^2 \int_{Q_{\delta}^{(\lambda)}} \eta^{2p-2}(x)\zeta(t) \left| \frac{u - \ell}{\varrho - \sigma} \right|^2 \, dz \]

\[ + 2^{p-3} \int_{Q_{\delta}^{(\lambda)}} \eta^{2p-2}(x)\zeta(t) |Du - D\ell|^{p-2} \left| \frac{u - \ell}{\varrho - \sigma} \right|^2 \, dz \]

\[ \leq 2^{p-3}(1 + |D\ell|)^{p-2} \int_{Q_{\delta}^{(\lambda)}} \left| \frac{u - \ell}{\varrho - \sigma} \right|^2 \, dz \]

\[ + \delta \int_{Q_{\delta}^{(\lambda)}} \eta^{2p}(x)\zeta(t) |Du - D\ell|^p \, dz + C(\delta, p) \int_{Q_{\delta}^{(\lambda)}} \eta^p(x)\zeta(t) \left| \frac{u - \ell}{\varrho - \sigma} \right|^p \, dz, \]

where we again used the fact that \(\eta, \zeta \leq 1, 2p - 2 \geq 0\). Due to

\[ |Du - D\ell|^p \leq (1 + |D\ell| + |Du - D\ell|)^{p-2} |Du - D\ell|^2 \]

we conclude with

\[ |II| \leq \delta \int_{Q_{\delta}^{(\lambda)}} \eta^{2p}(x)\zeta(t)(1 + |D\ell| + |Du - D\ell|)^{p-2} |Du - D\ell|^2 \, dz \]

\[ + C(\delta, p, L)(1 + |D\ell|)^{p-2} \int_{Q_{\delta}^{(\lambda)}} \left| \frac{u - \ell}{\varrho - \sigma} \right|^2 \, dz + C(\delta, p, L) \int_{Q_{\delta}^{(\lambda)}} \left| \frac{u - \ell}{\varrho - \sigma} \right|^p \, dz. \]

For the second term on the right-hand side we use \((1.4)_1\) to obtain

\[ |III| \leq L \int_{Q_{\delta}^{(\lambda)}} \eta^{2p}(x)\zeta(t) \omega(|u - \ell(0)|^2)(1 + |D\ell|)^{p-1} |Du - D\ell| \, dz \]

\[ + 2pL \int_{Q_{\delta}^{(\lambda)}} \eta^{2p-1}(x)\zeta(t) \omega(|u - \ell(0)|^2)(1 + |D\ell|)^{p-1} \left| \frac{u - \ell}{\varrho - \sigma} \right| \, dz \]

\[ =: III_1 + III_2. \]
An application of Young’s inequality yields

$$II_2 \leq 2pL \int_{Q_{\varrho}^{(c)}} \omega(|u - \ell(0)|^2) \left(1 + |D\varrho|\right)^{p-1} \left|\frac{u - \ell}{\varrho - \sigma}\right| \, dz$$

$$\leq C(p, L)(1 + |D\varrho|)^p \int_{Q_{\varrho}^{(c)}} \omega(\varrho^p (|u - \ell(0)|^2)) \, dz + C(p, L) \int_{Q_{\varrho}^{(c)}} \left|\frac{u - \ell}{\varrho - \sigma}\right|^p \, dz$$

$$\leq C(p, L)(1 + |D\varrho|)^p \int_{Q_{\varrho}^{(c)}} \omega(|u - \ell(0)|^2) \, dz + C(p, L) \int_{Q_{\varrho}^{(c)}} \left|\frac{u - \ell}{\varrho - \sigma}\right|^p \, dz,$$

where we also used the fact that $\omega \leq 1$. Another application of Young’s inequality yields

$$III_1 \leq \delta \int_{Q_{\varrho}^{(c)}} \eta^{2p}(x)\zeta(t) u(|u - \ell(0)|^2) \, dz$$

$$+ C(\delta, p, L)(1 + |D\varrho|)^p \int_{Q_{\varrho}^{(c)}} \omega(|u - \ell(0)|^2) \, dz$$

In the case $p \geq 2$ we use the fact that $\omega \leq 1$ to conclude with

$$III_1 \leq \delta \int_{Q_{\varrho}^{(c)}} \eta^{2p}(x)\zeta(t)(1 + |D\varrho| + |Du - D\varrho|)^{p-2} |Du - D\varrho|^2 \, dz$$

$$+ C(\delta, p, L)(1 + |D\varrho|)^p \int_{Q_{\varrho}^{(c)}} \omega(|u - \ell(0)|^2) \, dz$$

In the case $p < 2$ we need a more subtle argument to obtain an estimate for the term $III_1$. Let us first define the set

$$G := \left\{ z \in Q_{\varrho}^{(c)} : |Du| \geq 2 |D\varrho| \text{ and } |Du - D\varrho| \geq 1 \right\}. \quad (3.5)$$

On $G$ we have

$$|Du| + |D\varrho| \leq |Du - D\varrho| + 2 |D\varrho| \leq |Du - D\varrho| + |Du| \leq \frac{3}{2} |Du - D\varrho| + \frac{1}{2} |Du| + \frac{1}{2} |D\varrho|,$$

and hence $|Du| + |D\varrho| \leq 3 |Du - D\varrho|$. Since $|Du - D\varrho| \geq 1$ by definition, we obtain

$$\frac{1}{2} (1 + |D\varrho| + |Du - D\varrho|) \leq 1 + |Du| + |D\varrho| \leq 4 |Du - D\varrho| \text{ on } G. \quad (3.6)$$

On $Q_{\varrho}^{(c)} \setminus G$, on the other hand, we have $|Du| < 2 |D\varrho|$ or $|Du - D\varrho| < 1$, which yields the estimate

$$|Du - D\varrho| \leq 3(1 + |D\varrho|) \text{ on } Q_{\varrho}^{(c)} \setminus G. \quad (3.7)$$

Splitting the integration domain into $G$ and $Q_{\varrho}^{(c)} \setminus G$ and applying (3.6), (3.7) yields

$$\int_{Q_{\varrho}^{(c)}} \eta^{2p}(x)\zeta(t) \omega(|u - \ell(0)|^2) |Du - D\varrho|^p \, dz$$

$$\leq 8^{2-p} \int_{G} \eta^{2p}(x)\zeta(t)(1 + |D\varrho| + |Du - D\varrho|)^{p-2} |Du - D\varrho|^2 \, dz$$

$$+ 3^p(1 + |D\varrho|)^p \int_{Q_{\varrho}^{(c)} \setminus G} \omega(|u - \ell(0)|^2) \, dz,$$

where we also used the fact that $\eta, \zeta, \omega \leq 1$. Combining the estimates for $III_1$ and $II_2$, we see that the following estimate holds (both for $p < 2$ and for $p \geq 2$):

$$|III| \leq \delta \int_{Q_{\varrho}^{(c)}} \eta^{2p}(x)\zeta(t)(1 + |D\varrho| + |Du - D\varrho|)^{p-2} |Du - D\varrho|^2 \, dz$$
For the third term on the right-hand side we use the VMO condition (1.5) to obtain
\[
|IV| \leq \int_{Q_{\rho}^\delta} \eta^{2p}(x) \zeta_\ell(t) v_{0,\delta}^{(\lambda)}(z) (1 + |D\ell|)^{p-1} |Du - D\ell| \, dz
\]
\[+ 2p \int_{Q_{\rho}^\delta} \eta^{2p-1}(x) \zeta_\ell(t) v_{0,\delta}^{(\lambda)}(z) (1 + |D\ell|)^{p-1} \left| \frac{u - \ell}{\rho - \sigma} \right|^p \, dz. \]

Since \(v_{0,\delta}^{(\lambda)}\) is bounded from above by \(2L\), we can argue in the same way as for \(III\) to obtain the following estimate for \(IV\):
\[
|IV| \leq \delta \int_{Q_{\rho}^\delta} \eta^{2p}(x) \zeta_\ell(t) (1 + |D\ell| + |Du - D\ell|)^{p-2} |Du - D\ell| \, dz
\]
\[+ C(\delta, p, L)(1 + |D\ell|)^p \int_{Q_{\rho}^\delta} v_{0,\delta}^{(\lambda)}(z) \, dz + C(p, L) \int_{Q_{\rho}^\delta} \left| \frac{u - \ell}{\rho - \sigma} \right|^p \, dz. \]

Finally we can estimate the last term on the right-hand side by using the definition of \(\zeta_\ell\):
\[
V = \int_{Q_{\rho}^\delta} \eta^{2p}(x) \partial_t \zeta_\ell(t) \, |u - \ell|^2 \, dz + \frac{1}{2} \int_{Q_{\rho}^\delta} \eta^{2p}(x) \zeta_\ell(t) \partial_t (|u - \ell|^2) \, dz
\]
\[= \frac{1}{2} \int_{Q_{\rho}^\delta} \eta^{2p}(x) \partial_t \zeta_\ell(t) \, |u - \ell|^2 \, dz
\]
\[\leq \lambda^{p-2} \int_{Q_{\rho}^\delta} \eta^{2p}(x) \, dx \, dt - \frac{1}{\varepsilon} \int_{s}^{s+\varepsilon} \int_{B_\rho} |u - \ell|^2 \eta^{2p}(x) \, dx \, dt
\]
\[\leq C(p, M)(1 + |D\ell|)^{p-2} \int_{Q_{\rho}^\delta} \left| \frac{u - \ell}{\rho - \sigma} \right|^2 \, dz - \frac{1}{\varepsilon} \int_{s}^{s+\varepsilon} \int_{B_\rho} |u - \ell|^2 \, dx \, dt,
\]

where we also used (3.3) and \(\eta = 1\) on \(B_\sigma\) in the last step. Combining the estimates for \(I - V\), we arrive at
\[
\int_{Q_{\rho}^\delta} \frac{1}{C(p, \nu)} \eta^{2p}(x) \zeta_\ell(t) (1 + |D\ell| + |Du - D\ell|)^{p-2} |Du - D\ell| \, dz
\]
\[+ \frac{1}{\varepsilon} \int_{s}^{s+\varepsilon} \int_{B_\rho} |u - \ell|^2 \, dx \, dt
\]
\[\leq 3\delta \int_{Q_{\rho}^\delta} \eta^{2p}(x) \zeta_\ell(t) (1 + |D\ell| + |Du - D\ell|)^{p-2} |Du - D\ell| \, dz
\]
\[+ C(\delta, p, L)(1 + |D\ell|)^p \int_{Q_{\rho}^\delta} \left| \frac{u - \ell}{\rho - \sigma} \right|^p \, dz + C(\delta, p, L) \int_{Q_{\rho}^\delta} \left| \frac{u - \ell}{\rho - \sigma} \right|^p \, dz
\]
\[+ C(\delta, p, L)(1 + |D\ell|)^p \int_{Q_{\rho}^\delta} \omega(|u - \ell(0)|^2) + v_{0,\delta}^{(\lambda)}(z) \, dz.
\]

Choosing \(\delta = \frac{1}{6C(p, \nu)}\), we can absorb the first term from the right-hand side on the left-hand side. Now letting \(\varepsilon \to 0\), we obtain the following inequality:
\[
\int_{B_\sigma} |u(x, s) - \ell|^2 \, dx + \int_{Q_{\rho}^\delta} \eta^{2p}(x) \zeta_\ell(t) (1 + |D\ell| + |Du - D\ell|)^{p-2} |Du - D\ell| \, dz
\]
\[\begin{align*}
&\leq C(p, \nu, L, M) \left[ (1 + |D\ell|)^{p-2} \int_{Q_v^{(s)}} \left| \frac{u - \ell}{\varrho - \sigma} \right|^2 dz + \int_{Q_v^{(s)}} \left| \frac{u - \ell}{\varrho - \sigma} \right|^p dz \right] \\
&+ C(p, \nu, L) (1 + |D\ell|)^p \int_{Q_v^{(s)}} \omega(|u - \ell(0)|^2) + \nu_{\ell, e}^{(s)}(z) dz,
\end{align*}\]

where

\[\zeta_0(t) = \begin{cases}
\frac{1}{\lambda^2 - \nu^2 - \sigma^2} (t + \lambda^2 - \nu^2), & \text{on } (-\lambda^2 - \nu^2, -\lambda^2 - \nu^2), \\
1, & \text{on } (-\lambda^2 - \nu^2, \sigma), \\
0, & \text{on } (s, 0)
\end{cases}\]

is the limit of \( \zeta_\varepsilon \) as \( \varepsilon \to 0 \). We use this inequality in two different ways. By taking the supremum over all \( s \in \Lambda^{(s)}_\sigma \) in the first term on the left-hand side we obtain

\[\sup_{s \in \Lambda^{(s)}_\sigma} \int_{B_s} |u(x, s) - \ell|^2 \, dx \leq C(p, \nu, L, M) \left[ (1 + |D\ell|)^{p-2} \int_{Q_v^{(s)}} \left| \frac{u - \ell}{\varrho - \sigma} \right|^2 dz + \int_{Q_v^{(s)}} \left| \frac{u - \ell}{\varrho - \sigma} \right|^p dz \right] + C(p, \nu, L) (1 + |D\ell|)^p \int_{Q_v^{(s)}} \omega(|u - \ell(0)|^2) + \nu_{\ell, e}^{(s)}(z) dz.\]

On the other hand, we let \( s \to 0 \) in the second term on the right-hand side to obtain

\[\int_{Q_v^{(s)}} (1 + |D\ell| + |Du - D\ell|)^{p-2} |Du - D\ell|^2 dz \leq C(p, \nu, L, M) \left[ (1 + |D\ell|)^{p-2} \int_{Q_v^{(s)}} \left| \frac{u - \ell}{\varrho - \sigma} \right|^2 dz + \int_{Q_v^{(s)}} \left| \frac{u - \ell}{\varrho - \sigma} \right|^p dz \right] + C(p, \nu, L) (1 + |D\ell|)^p \int_{Q_v^{(s)}} \omega(|u - \ell(0)|^2) + \nu_{\ell, e}^{(s)}(z) dz.\]

By adding the previous two inequalities we infer that

\[\sup_{s \in \Lambda^{(s)}_\sigma} \int_{B_s} |u(x, s) - \ell|^2 \, dx + \int_{Q_v^{(s)}} (1 + |D\ell| + |Du - D\ell|)^{p-2} |Du - D\ell|^2 dz \leq C(p, \nu, L, M) \left[ (1 + |D\ell|)^{p-2} \int_{Q_v^{(s)}} \left| \frac{u - \ell}{\varrho - \sigma} \right|^2 dz + \int_{Q_v^{(s)}} \left| \frac{u - \ell}{\varrho - \sigma} \right|^p dz \right] + C(p, \nu, L) (1 + |D\ell|)^p \int_{Q_v^{(s)}} \omega(|u - \ell(0)|^2) + \nu_{\ell, e}^{(s)}(z) dz.\]

Dividing this inequality by \( \left| Q_v^{(s)} \right| \), using Jensen’s inequality in the form

\[\int_{Q_v^{(s)}} \omega \left( |u - \ell(0)|^2 \right) dz \leq \omega \left( \int_{Q_v^{(s)}} |u - \ell(0)|^2 \, dz \right)\]

and recalling that \( \lambda^{2-\nu^2}_0 \varrho \leq \varrho_0 \), and hence

\[\int_{Q_v^{(s)}} \nu_{\ell, e}^{(s)}(z) dz \leq V(\lambda^{2-\nu^2}_0 \varrho^2)\]
due to (1.6), we obtain
\[
\left( \frac{\sigma}{\varrho} \right)^{n+2} \sup_{s \in \Lambda_{s}^{(\lambda)}, B_{r}} \int_{\Omega} \frac{|u(x, s) - \ell|^{2}}{\lambda^{2} - \sigma_{s}^{2}} \, dx
\]
\[+ \left( \frac{\sigma}{\varrho} \right)^{n+2} \int_{Q_{\varrho}^{(\lambda)}} (1 + |D\ell| + |Du - D\ell|)^{p-2} |Du - D\ell|^{2} \, dz
\]
\[\leq C \left( \frac{\varrho}{\theta} \right)^{\max(p,2)} \left[ (1 + |D\ell|)^{p-2} \int_{Q_{\varrho}^{(\lambda)}} \left| \frac{u - \ell}{\varrho} \right|^{2} \, dz + \int_{Q_{\varrho}^{(\lambda)}} \left| \frac{u - \ell}{\varrho} \right|^{p} \, dz \right]
\]
\[+ C(1 + |D\ell|)^{p} \left[ \omega \left( \int_{Q_{\varrho}^{(\lambda)}} (u - \ell(0))^2 \, dz \right) + V \left( \lambda^{2} - \sigma^{2} \right) \right],
\]
where \( C = C(p, \nu, L, M) \). Noting that \( \frac{\sigma}{\varrho} \geq \frac{1}{2} \), we conclude with the desired Caccioppoli inequality (3.4). \( \square \)

**Remark 3.2.** In the following, we will only apply the Caccioppoli inequality (3.4) for the two special cases \( \sigma = \varrho/2 \) respectively \( \lambda = 1 \).

### 4. Poincaré inequality

In this chapter we will prove a Poincaré type inequality, see Theorem 4.2 below. We will apply this inequality to show that in any regular point \( z_{0} \in \Omega_{r} \) some smallness conditions (which are necessary for the iteration of the excess decay estimate) hold, see Remark 4.3 and Step 4 in the proof of Theorem 1.5 below. But first, we will prove the following auxiliary lemma:

**Lemma 4.1.** Let \( u \in L^{p}(-T, 0; W^{1,p}(\Omega, \mathbb{R}^{N})) \cap C^{0}(-T, 0; L^{2}(\Omega, \mathbb{R}^{N})) \) be a weak solution of (1.1). Let \( Q_{\varrho}(z_{0}) \subseteq \Omega_{r} \) be a standard parabolic cylinder with radius \( \varrho \in (0,1) \). Furthermore let \( \ell : \mathbb{R}^{n} \to \mathbb{R}^{N} \) be an affine function. There exists a constant \( C(p, L) \) such that for any \( r, s \in (t_{0} - \sigma^{2}, t_{0}) \), any test function \( \psi \in C_{0}^{\infty}(B_{r}(x_{0})) \) and any \( i \in \{1, ..., N\} \) the following estimate holds:

\[
\left| \int_{B_{r}(x_{0})} (u_{i}(\cdot, r) - u_{i}(\cdot, s)) \psi \, dx \right|
\]
\[\leq C |r - s|^{\frac{2}{p}} \|D\psi\|_{L^{p}} |Q_{\varrho}|^{\frac{p-1}{p}} \left( \int_{Q_{\varrho}(z_{0})} |Du - D\ell|^{p} \, dz \right)^{\frac{1}{p}}
\]
\[+ C \chi_{\{p > 2\}} |r - s|^{\frac{2}{p}} \|D\psi\|_{L^{p}} |Q_{\varrho}|^{\frac{p-1}{p}} (1 + |D\ell|)^{p-2} \left( \int_{Q_{\varrho}(z_{0})} |Du - D\ell|^{p} \, dz \right)^{\frac{1}{p}}
\]
\[+ C |r - s|^{\frac{2}{p}} \|D\psi\|_{L^{p}} |Q_{\varrho}|^{\frac{p-1}{p}} (1 + |D\ell|)^{p-1} \omega^{\frac{p-1}{p}} \left( \int_{Q_{\varrho}(z_{0})} |u - \ell(x_{0})|^{2} \, dz \right)
\]
\[+ C |r - s|^{\frac{2}{p}} \|D\psi\|_{L^{p}} |Q_{\varrho}|^{\frac{p-1}{p}} (1 + |D\ell|)^{p-1} \mathbf{V}^{\frac{p-1}{p}}(\sigma^{2}),
\]
where

\[
\chi_{\{p > 2\}} := \begin{cases} 1, & p > 2 \\ 0, & p \leq 2. \end{cases}
\]

**Proof.** For notational convenience we will assume without loss of generality that \( z_{0} = 0 \) and that \( s < r \). Let \( i \in \{1, ..., N\} \) and \( 0 < \varepsilon < \frac{1}{\lambda^{2}} \). We choose the test
function \( \varphi_h(x,t) := \zeta_h(t)\psi(x)e_i \), where \( \zeta_h \in W^{1,\infty}_0(-\sigma^2,0) \) is given by

\[
\zeta_h(t) = \begin{cases} 
0, & t \leq s \\
\frac{1}{\sigma}(t-s), & t \in (s,s+h) \\
1, & t \in [s+h,r-h] \\
\frac{1}{\sigma}(t-r), & t \in (r-h,r) \\
0, & t \geq r.
\end{cases}
\]

Note that this function is not smooth and thus not actually an admissible test function. However, the following formal computations can be made rigorous via a smoothing procedure with respect to time, as for instance via Steklov averages.

Inserting \( \varphi_h \) into the weak formulation (1.7) yields

\[
\int_{-\sigma^2}^0 \int_{B_{s}} u_i \partial_t \zeta_h \psi \, dx \, dt = \int_{-\sigma^2}^0 \int_{B_{s}} A^i(z,u,Du) \cdot D\psi \zeta_h \, dx \, dt, \tag{4.2}
\]

where \( A = (A^1,\ldots,A^N) \). Using the definition of \( \zeta_h \) and letting \( h \to 0 \), we get

\[
\int_{-\sigma^2}^0 \int_{B_{s}} u_i \partial_t \zeta_h \psi \, dx \, dt = \frac{1}{h} \int_s^{s+h} \int_{B_{s}} u_i \psi \, dx \, dt - \frac{1}{h} \int_r^{r-h} \int_{B_r} u_i \psi \, dx \, dt \to \int_{B_r} (u_i(\cdot,s) - u_i(\cdot,r)) \psi \, dx.
\]

Also letting \( h \to 0 \) on the right-hand side of (4.2), we arrive at

\[
\int_{B_{s}} (u_i(\cdot,s) - u_i(\cdot,r)) \psi \, dx = \int_{s}^{r} \int_{B_{s}} A^i(z,u,Du) \cdot D\psi \, dx \, dt. \tag{4.3}
\]

We will now estimate the integral w.r.t. space from the right-hand side of this equation by rewriting it as follows:

\[
\int_{B_{s} \times \{t\}} A^i(z,u,Du) \cdot D\psi \, dx
= \int_{B_{s} \times \{t\}} [A^i(z,u,Du) - A^i(z,u,D\ell)] \cdot D\psi \, dx
+ \int_{B_{s} \times \{t\}} [A^i(z,u,D\ell) - A^i(z,\ell(0),D\ell)] \cdot D\psi \, dx
+ \int_{B_{s} \times \{t\}} [A^i(z,\ell(0),D\ell) - (A^i(\cdot,\ell(0),D\ell))_{Q_s}] \cdot D\psi \, dx =: I + II + III,
\]

with the obvious meaning of \( I - III \). Arguing in the same way as in the estimate for the term \( II \) in the proof of Lemma 3.1 we obtain

\[
|I| \leq C(p,L) \int_{B_{s} \times \{t\}} (1 + |D\ell| + |Du - D\ell|^{p-2}) |Du - D\ell| |D\psi| \, dx.
\]

In the case \( p < 2 \) an application of Hölder’s inequality yields

\[
|I| \leq C \int_{B_{s} \times \{t\}} |Du - D\ell|^{p-1} |D\psi| \, dx
\leq C \|D\psi\|_{L^p} \left( \int_{B_{s} \times \{t\}} |Du - D\ell|^p \, dx \right)^{\frac{p-1}{p}}.
\]
In the case \( p \geq 2 \) we also apply Hölder’s inequality to obtain

\[
|I| \leq C (1 + |D\ell|)^{p-2} \int_{B_{r} \times \{t\}} |Du - D\ell| |D\psi| \, dx
\]

\[
+ C \int_{B_{r} \times \{t\}} |Du - D\ell|^{p-1} |D\psi| \, dx
\]

\[
\leq C \|D\psi\|_{L^p} \left( \int_{B_{r} \times \{t\}} |Du - D\ell|^{p} \, dx \right)^{\frac{p-1}{p}}
\]

\[
+ C (1 + |D\ell|)^{p-2} \|D\psi\|_{L^p} \left( \int_{B_{r} \times \{t\}} |Du - D\ell|^{\frac{p}{p-1}} \, dx \right)^{\frac{p-1}{p}}.
\]

For the second term we use the continuity assumption \((1.4)_1\), Hölder’s inequality and the fact that \( \omega \leq 1 \) to derive the following estimate:

\[
|II| \leq L (1 + |D\ell|)^{p-1} \int_{B_{r} \times \{t\}} \omega \left( |u - \ell(0)|^2 \right) |D\psi| \, dx
\]

\[
\leq L (1 + |D\ell|)^{p-1} \|D\psi\|_{L^p} \left( \int_{B_{r} \times \{t\}} \omega^{\frac{p}{p-1}} \left( |u - \ell(0)|^2 \right) \, dx \right)^{\frac{p-1}{p}}
\]

\[
\leq L (1 + |D\ell|)^{p-1} \|D\psi\|_{L^p} \left( \int_{B_{r} \times \{t\}} \omega \left( |u - \ell(0)|^2 \right) \, dx \right)^{\frac{p-1}{p}}.
\]

Using the VMO condition \((1.5)\) and arguing as before, we obtain the following estimate for the third term:

\[
|III| \leq L (1 + |D\ell|)^{p-1} \int_{B_{r} \times \{t\}} v_{0,\sigma}(z) |D\psi| \, dx
\]

\[
\leq L (1 + |D\ell|)^{p-1} \|D\psi\|_{L^p} \left( \int_{B_{r} \times \{t\}} v_{0,\sigma}^{\frac{p}{p-1}} (z) \, dx \right)^{\frac{p-1}{p}}
\]

\[
\leq C(p, L) (1 + |D\ell|)^{p-1} \|D\psi\|_{L^p} \left( \int_{B_{r} \times \{t\}} v_{0,\sigma}(z) \, dx \right)^{\frac{p-1}{p}},
\]

where we also used the fact that \( v_{0,\sigma} \leq 2L \). Combining the estimates for \( I - III \), then integrating over \((s, r)\) and finally applying Hölder’s inequality, we arrive at

\[
\left| \int_{s}^{r} \int_{B_{r}} A^{i}(z, u, Du) \cdot D\psi \, dx \, dt \right|
\]

\[
\leq C (r - s)^{\frac{1}{2}} \|D\psi\|_{L^p} \left( \int_{Q_{r}} |Du - D\ell|^{p} \, dz \right)^{\frac{p-1}{p}}
\]

\[
+ C \chi_{(p>2)} (r - s)^{\frac{1}{2}} \|D\psi\|_{L^p} (1 + |D\ell|)^{p-2} \left( \int_{Q_{r}} |Du - D\ell|^{\frac{p}{p-1}} \, dz \right)^{\frac{p-1}{p}}
\]

\[
+ C (r - s)^{\frac{1}{2}} \|D\psi\|_{L^p} (1 + |D\ell|)^{p-1} \left( \int_{Q_{r}} \omega \left( |u - \ell(0)|^2 \right) \, dz \right)^{\frac{p-1}{p}}.
\]
Keeping in mind the equation (4.3), applying Hölder’s inequality in the form
\[
\left( \int_{Q_r} |Du - D\ell|^{p^*} \, dz \right)^{\frac{p}{p^*}} \leq \left( \int_{Q_r} |Du - D\ell|^p \, dz \right)^{\frac{1}{p}}
\]
and Jensen’s inequality in the form
\[
f_{Q_r} \omega \left( |u - \ell(0)|^2 \right) \, dz \leq \omega \left( \int_{Q_r} |u - \ell(0)|^2 \, dz \right)
\]
and using
\[
f_{Q_r} v_{0,\sigma}(z) \, dz \leq V(\sigma^2),
\]
we finally obtain the desired estimate (4.1).

In order to simplify the notation for the following theorem and its proof, let us define
\[
m := \max \{p, 2\}.
\]

**Theorem 4.2.** Let \( u \in L^p(-T, 0; W^{1-p}(\Omega, \mathbb{R}^N)) \cap C^0(-T, 0; L^2(\Omega, \mathbb{R}^N)) \) be a weak solution of (1.1). Let \( \varrho \in (0, \frac{1}{2}) \) be a radius such that \( Q_{2\varrho}(z_0) \subseteq \Omega_T \). Let \( E \in \mathbb{R}^{N \times n} \) be a constant matrix. There exists a positive constant \( C_{\text{poin}}(N, n, p, \nu, L) \) such that the following Poincaré type estimate holds:
\[
f_{Q_{\varrho}(z_0)} \left| u - (u)_{z_0} - E(x-x_0) \right|^p \, dz \leq C_{\text{poin}}(1 + |E|)^{p(n-2)} \left( \int_{Q_{2\varrho}(z_0)} |Du - E|^p \, dz \right)^{p-1}
+ C_{\text{poin}}(1 + |E|)^{p(n-2)} \left( \int_{Q_{2\varrho}(z_0)} |Du - E|^p \, dz \right)^\frac{2}{p}
+ C_{\text{poin}}(1 + |E|)^{p(n-1)} \omega \left( \int_{Q_{\varrho}(z_0)} |u - (u)_{z_0, 2\varrho}|^2 \, dz \right)^\frac{2(n-1)}{m}
+ C_{\text{poin}}(1 + |E|)^{p(n-1)} \omega \left( \int_{Q_{\varrho}(z_0)} |u - (u)_{z_0, 2\varrho}|^2 \, dz \right)\omega \left( \int_{Q_{\varrho}(z_0)} |u - (u)_{z_0, 2\varrho}|^2 \, dz \right)\omega \left( \int_{Q_{\varrho}(z_0)} |u - (u)_{z_0, 2\varrho}|^2 \, dz \right)
\]

**Proof.** For the sake of notational simplicity let us assume that \( z_0 = 0 \). Let \( \sigma \) and \( \alpha \) be two radii such that
\[
\varrho \leq \sigma < \alpha \leq 2\varrho.
\]
We choose a symmetric smoothing kernel \( \psi \in C_0^\infty(B_1) \) with
\[
\int_{B_1} \psi(x) \, dx = 1 \quad \text{and} \quad \|\psi\|_{L^\infty} + \|D\psi\|_{L^\infty} \leq \frac{2(n+2)}{|B_1|}.
\]
The rescaled function
\[
\psi_\varrho(x) := \varrho^{-n} \cdot \psi \left( \frac{x}{\varrho} \right)
\]
satisfies
\[
\psi_\varrho \in C_0^\infty(B_\varrho), \quad \int_{B_\varrho} \psi_\varrho(x) \, dx = 1
\]
as well as the following estimates:

\[ \|D\psi_\varphi\|_{L^p} \leq C(n)g^{-n-1+n/p} \leq C(n)\sigma^{-n-1+n/p} \tag{4.6} \]
\[ \|\psi_\varphi\|_{L^\infty} \leq C(n)g^{-n}. \tag{4.7} \]

In the following, we will use several different notions of means of \( u \). We define the means and \( \psi \)-means over a time slice for \( t \in (-2g)^2, 0 \) as

\[ \bar{u}_\varphi(t) := \int_{B_\varphi} u(x,t) \, dx \quad \text{and} \quad \bar{u}\psi(t) := \int_{B_\varphi} u(x,t)\psi_\varphi(x) \, dx. \]

For the means and \( \psi \)-means over a cylinder \( Q_\varphi \), we write

\[ (u)_\varphi := \int_{Q_\varphi} u(z) \, dz = \int_{-g^2}^0 \bar{u}_\varphi(t) \, dt \quad \text{and} \quad (u)^\psi := \int_{-g^2}^0 \bar{u}\psi(t) \, dt. \]

For notational clarity let us define the following functions that we will use repeatedly during the course of the proof:

\[ \Psi(r,s) := \int_{Q_r} \left| u(z) - (u)_r - Ex \right|^2 + \left| u(z) - (u)_s - Ex \right|^p \, dz, \]
\[ \Phi^q(r) := \left( \int_{Q_r} |Du - E|^p \, dz \right)^{q(p-1)} + \left( \int_{Q_r} |Du - E|^p \, dz \right)^q, \]
\[ \Upsilon^q(r,s) := (1 + |E|)^p \left[ \omega^q \left( \int_{Q_r} |u - (u)_s|^2 \, dz \right) + \nabla^q \left( r^2 \right) \right], \]

where \( r, s \in [g, 2g] \) and \( q \geq 0 \). We will also write \( \Phi(r) \) and \( \Upsilon(r,s) \) for \( \Phi^1(r) \) and \( \Upsilon^1(r,s) \) respectively. Let us first compare the different notions of means by deriving some estimates that will be useful later on. Noting that

\[
|t - s|^{1/p} \|D\psi_\varphi\|_{L^p}[Q_{s\varphi}]^{1-p} \leq C\sigma^{2/p} \sigma^{-n-1+n/p} (\sigma^{n+2})^{1-p} = C\sigma
\]

for any \( s, t \in (-\sigma^2, 0) \), where we have used (4.6), an application of Lemma 4.1 yields the following estimate for any \( t \in (-\sigma^2, 0) \):

\[
\left| \bar{u}\psi(t) - (u)^\psi \right| \\
\leq \int_{-g^2}^0 \left| \bar{u}\psi(t) - \bar{u}\psi(s) \right| \, ds \\
\leq 4 \int_{-\sigma^2}^0 \left| \bar{u}\psi(t) - \bar{u}\psi(s) \right| \, ds \\
\leq C \sum_{i=1}^N \int_{-\sigma^2}^0 \left| \int_{B_{s\varphi}} (u_i(x,t) - u_i(x,s)) \psi_\varphi(x) \, dx \right| \, ds \\
\leq C\sigma \left( \int_{Q_{s\varphi}(z_0)} |Du - E|^p \, dz \right)^{\frac{p-1}{p}} + C\sigma(1 + |E|)^{m-2} \left( \int_{Q_{s\varphi}(z_0)} |Du - E|^p \, dz \right)^{\frac{1}{p}} \\
+ C\sigma(1 + |E|)^{p-1} \omega^q \left( \int_{Q_{s\varphi}(z_0)} |u - (u)_s|^2 \, dz \right) + C\sigma(1 + |E|)^{p-1} \nabla \omega^q(\sigma^2) \\
\leq C\sigma \left[ (1 + |E|)^{m-2} \Phi^1(\sigma) + \Upsilon^{\frac{p-1}{p}}(\sigma, \varphi) \right],
\]

Raising this inequality to the power \( p \), we arrive at

\[
\left| \bar{u}\psi(t) - (u)^\psi \right|^p \leq C\sigma^p \left[ (1 + |E|)^{p(m-2)} \Phi(\sigma) + \Upsilon^{p-1}(\sigma, \varphi) \right] \tag{4.8}
\]
for any \( t \in (-\sigma^2, 0) \). Furthermore, a slicewise application of Poincaré’s inequality for functions with vanishing mean value yields

\[
\left| \bar{u}_\varphi(t) - \bar{\psi}(t) \right|^p = \left| \int_{B_\varphi} u(x, t) \varphi(x) \, dx - \bar{u}_\varphi(t) \right|^p = \left| \int_{B_\varphi} (u(x, t) - \bar{u}_\varphi(t) - Ex) \psi(x) \, dx \right|^p
\]

\[
\leq C(n) \left( \int_{B_\varphi} |u(x, t) - \bar{u}_\varphi(t) - Ex| \, dx \right)^p \leq C(n) \int_{B_\varphi} |u(x, t) - \bar{u}_\varphi(t) - Ex|^p \, dx
\]

\[
\leq C(n, p) p^p \int_{B_\varphi \times \{ t \}} |Du - E|^p \, dx \leq C(n, p) p^p \int_{B_{2\varphi} \times \{ t \}} |Du - E|^p \, dx.
\] (4.9)

In these calculations we have also used that \( \int_{B_\varphi} \psi(x) \, dx = 1 \) and \( \int_{B_\varphi} Ex \psi(x) \, dx = 0 \) due to the symmetry of \( \psi \). Integrating this inequality with respect to time yields

\[
|(u)_\varphi - (u)_\varphi^\psi| = \left| \int_{-\sigma^2}^0 \bar{u}_\varphi(t) - \bar{\psi}(t) \, dt \right|^p \leq \left| \int_{-\sigma^2}^0 \bar{u}_\varphi(t) - \bar{\psi}(t) \, dt \right|^p \leq 4 \left( \int_{-2\sigma}^0 \bar{u}_\varphi(t) - \bar{\psi}(t) \, dt \right)^p \leq C(n, p) \sigma^p \int_{-2\sigma}^0 |Du - E|^p \, dx \, dt
\]

\[
= C(n, p) \sigma^p \int_{Q_{2\sigma}} |Du - E|^p \, dz.
\] (4.10)

Finally, we derive an estimate for the \( L^p \)-norm of the difference of two means with different radii. To this end, we use Hölder’s inequality and the fact that \( \sigma \in [\varphi, 2\varphi] \) to obtain

\[
\int_{-\sigma^2}^0 |\bar{u}_\varphi(t) - \bar{u}_\sigma(t)|^p \, dt
\]

\[
= \int_{-\sigma^2}^0 \left| \int_{B_\varphi} (u(x, t) - \bar{u}_\sigma(t) - Ex) \, dx \right|^p \leq \int_{-\sigma^2}^0 \left( \int_{B_\varphi} |u(x, t) - \bar{u}_\sigma(t) - Ex|^m \, dx \right)^{p/m} \, dt
\]

\[
\leq C \int_{-\sigma^2}^0 \left( \int_{B_\varphi} |u(x, t) - \bar{u}_\sigma(t) - Ex|^m \, dx \right)^{p/m} \, dt.
\]

Combining this with the Minkowski inequality, we infer that

\[
\int_{-\sigma^2}^0 \left( \int_{B_\varphi} |u(x, t) - \bar{u}_\varphi(t) - Ex|^m \, dx \right)^{p/m} \, dt
\]

\[
\leq 2^{p-1} \int_{-\sigma^2}^0 \left( \int_{B_\varphi} |u(x, t) - \bar{u}_\varphi(t) - Ex|^m \, dx \right)^{p/m} + |\bar{u}_\varphi(t) - \bar{u}_\sigma(t)|^p \, dt
\]

\[
\leq C \int_{-\sigma^2}^0 \left( \int_{B_\varphi} |u(x, t) - \bar{u}_\sigma(t) - Ex|^m \, dx \right)^{p/m} \, dt.
\]

To the right-hand side of this inequality we can now slicewise apply the Sobolev-Poincaré inequality for functions with vanishing mean value. Since \( \frac{2n}{n+2} < p \) and \( \frac{pn}{n+p} < p \), a further application of Hölder’s inequality yields

\[
\int_{-\sigma^2}^0 \left( \int_{B_\varphi} |u(x, t) - \bar{u}_\sigma(t) - Ex|^m \, dx \right)^{p/m} \, dt
\]
\[ \leq C \sigma^p \int_{-\sigma^2}^0 \left( \int_{B_\sigma} |Du - E|^\frac{m}{\min(p,m)} \, dx \right)^{\frac{p(n+m)}{m}} \, dt \]
\[ \leq C \sigma^p \int_{-\sigma^2}^0 \int_{B_\sigma} |Du - E|^p \, dx \, dt \leq C \sigma^p \int_{Q_{2\sigma}} |Du - E|^p \, dz, \]

where we used \( \sigma \in [\varrho, 2\varrho] \) in the last step. Combining the last two inequalities, we finally conclude with
\[ \int_{-\sigma^2}^0 \left( \int_{B_\sigma} \left| \frac{u(x,t) - \tilde{u}_\varrho(t) - Ex}{\sigma} \right|^m \, dx \right)^{p/m} \, dt \leq C \int_{Q_{2\sigma}} |Du - E|^p \, dz. \quad (4.11) \]

In the next step we will combine the estimates (4.8) - (4.11) and apply Minkowski's inequality to obtain
\[ \int_{-\sigma^2}^0 \left( \int_{B_\sigma} \left| \frac{u(x,t) - \tilde{u}_\varrho(t) - Ex}{\sigma} \right|^m \, dx \right)^{p/m} \, dt \leq 4^{p-1} \int_{-\sigma^2}^0 \left[ \left( \int_{B_\sigma} \left| \frac{u(x,t) - \tilde{u}_\varrho(t) - Ex}{\sigma} \right|^m \, dx \right)^{p/m} \right. \]
\[ + \sigma^{-p} |\tilde{u}_\varrho(t) - \tilde{v}^\varrho(t)|^p + \sigma^{-p} |\tilde{v}^\varrho(t) - (u)^\varrho|^p + \sigma^{-p} \, (|u|^\varrho - (u)\varrho)^p \left] \right. \, dt \]
\[ \leq C \int_{Q_{2\sigma}} |Du - E|^p \, dz + C \left[ (1 + |E|)^{p(m-2)} \Phi(\sigma) + \Upsilon^{p-1}(\sigma, q) \right] \]
\[ \leq C \left[ (1 + |E|)^{p(m-2)} \Phi(2\varrho) + \Upsilon^{p-1}(\sigma, q) \right]. \quad (4.12) \]

In the last step we have used that \( \sigma \in [\varrho, 2\varrho] \) and hence
\[ \Phi(\sigma) \leq C(n,p) \Phi(2\varrho). \]

**In the superquadratic case** \( p \geq 2 \) we obtain
\[ \Psi(\sigma, q) = \int_{Q_{\sigma}} \left| \frac{u(z) - (u)_{\varrho} - Ex}{\sigma} \right|^2 + \left| \frac{u(z) - (u)_{\varrho} - Ex}{\sigma} \right|^p \, dz \]
\[ \leq \int_{-\sigma^2}^0 \int_{B_\sigma} \left| \frac{u(z) - (u)_{\varrho} - Ex}{\sigma} \right|^p \, dx \, dt + \left( \int_{-\sigma^2}^0 \int_{B_\sigma} \left| \frac{u(z) - (u)_{\varrho} - Ex}{\sigma} \right|^p \, dx \, dt \right)^{\frac{2}{p}} \]
\[ \leq C \left[ (1 + |E|)^{p(p-2)} \Phi(2\varrho) + \Upsilon^{p-1}(\sigma, q) \right] + C \left[ (1 + |E|)^{p(p-2)} \Phi(2\varrho) + \Upsilon^{p-1}(\sigma, q) \right] \]
\[ \leq C \left[ (1 + |E|)^{p(p-2)} \left( \Phi(2\varrho) + \Phi^{\varphi}(2\varrho) \right) + \Upsilon^{p-1}(\sigma, q) + \Upsilon^{2(p-1)}(\sigma, q) \right]. \]

For arbitrary \( r, s \in [\varrho, 2\varrho] \), we have
\[ \int_{Q_r} |u - (u)_s|^2 \, dz \leq C \int_{Q_{2\varrho}} |u - (u)_{2\varrho}|^2 \, dz + C \int_{Q_{2\varrho}} |(u)_{2\varrho} - (u)_s|^2 \, dz \leq C \int_{Q_{2\varrho}} |u - (u)_{2\varrho}|^2 \, dz, \]
and hence \( \Upsilon(r, s) \leq C \Upsilon(2\varrho, 2\varrho) \). Thus, for the choice \( \sigma = \varrho \), we obtain
\[ \Psi(\varrho, q) \leq C \left[ (1 + |E|)^{p(p-2)} \left( \Phi(2\varrho) + \Phi^{\varphi}(2\varrho) \right) + \Upsilon^{p-1}(2\varrho, 2\varrho) + \Upsilon^{2(p-1)}(2\varrho, 2\varrho) \right]. \]
Taking into account that \( \frac{2}{p} \leq 1 < \frac{2(p-1)}{p} \leq p - 1 \), \( \omega \leq 1 \) and \( V \leq 2L \), we obtain
the desired estimate \( (4.4) \).

**In the subquadratic case** \( p < 2 \) we apply \((4.12)\), Hölder’s inequality and the
Caccioppoli inequality \( (3.4) \) (on \( Q_{\alpha} \) with \( \lambda = 1 \), where \( \alpha \) is the radius from \((4.5)\))
to derive a similar estimate for \( \Psi(\sigma, \varrho) \):

\[
\Psi(\sigma, \varrho) = \int_{Q_T} \left| \frac{u(z) - (u)_{E}}{\sigma} - E \right|^2 + \left| \frac{u(z) - (u)_{E} - E}{\sigma} \right|^p \, dz \\
\leq \int_{-\sigma^2}^0 \left( \int_{B_s} \left| \frac{u(z) - (u)_{E} - E}{\sigma} \right|^2 \, dx \right)^{1-\frac{p}{2}} \left( \int_{B_s} \left| \frac{u(z) - (u)_{E} - E}{\sigma} \right|^2 \, dx \right)^{\frac{p}{2}} \, dt \\
+ \int_{-\sigma^2}^0 \left( \int_{B_s} \left| \frac{u(z) - (u)_{E} - E}{\sigma} \right|^2 \, dx \right)^{\frac{p}{2}} \, dt \\
\leq \int_{-\sigma^2}^0 \left( \int_{B_s} \left| \frac{u(z) - (u)_{E} - E}{\sigma} \right|^2 \, dx \right)^{\frac{p}{2}} \, dt \\
\times \left[ \sup_{s \in (-\sigma^2, 0)} \left( \int_{B_s} \left| \frac{u(x, s) - (u)_{E} - E}{\sigma} \right|^2 \, dx \right)^{1-\frac{p}{2}} + 1 \right] \\
\leq C(C_{\text{cacc}}) \left[ \Phi(2\varrho) + \Upsilon^{p-1}(\sigma, \varrho) \right] \left( \frac{\alpha}{\alpha - \sigma} \right)^2 \left( \Psi(\alpha, \varrho) + \Upsilon(\alpha, \varrho) \right) \left( \frac{\alpha}{\alpha - \sigma} \right)^{1-\frac{p}{2}} \\
+ C \left[ \Phi(2\varrho) + \Upsilon^{p-1}(\sigma, \varrho) \right].
\]

An application of Young’s inequality with exponents \( \frac{2}{p} \) and \( \frac{2}{2-p} \) yields

\[
\Psi(\sigma, \varrho) \leq \frac{1}{2} \Psi(\alpha, \varrho) + \Upsilon(\alpha, \varrho) + C \left( \frac{\alpha}{\alpha - \sigma} \right)^{\frac{2(p-1)}{p}} \left[ \Phi(2\varrho) + \Upsilon^{\frac{2(p-1)}{p}}(\sigma, \varrho) \right] \\
+ C \left[ \Phi(2\varrho) + \Upsilon^{p-1}(\sigma, \varrho) \right] \\
\leq \frac{1}{2} \Psi(\alpha, \varrho) + \Upsilon(2\varrho, 2\varrho) + C \left( \frac{2\varrho}{\alpha - \sigma} \right)^{\frac{2(p-1)}{p}} \left[ \Phi(2\varrho) + \Upsilon^{\frac{2(p-1)}{p}}(2\varrho, 2\varrho) \right] \\
+ C \left[ \Phi(2\varrho) + \Upsilon^{p-1}(2\varrho, 2\varrho) \right].
\]

We are now in the position to apply Lemma 2.2 and thus conclude with

\[
\Psi(\varrho, \varrho) \leq C \left[ \Upsilon(2\varrho, 2\varrho) + \Phi(2\varrho) + \Upsilon^{\frac{2(p-1)}{p}}(2\varrho, 2\varrho) + \Phi(2\varrho) + \Upsilon^{p-1}(2\varrho, 2\varrho) \right].
\]

Taking into account that \( p - 1 < \frac{2(p-1)}{p} < 1 < \frac{2}{p} \), \( \omega \leq 1 \) and \( V \leq 2L \), we obtain
the desired estimate \((4.4)\). This concludes the proof of the theorem. \( \square \)

**Remark 4.3.** We will later use this Poincaré type inequality in two different ways:
On the one hand, let us suppose that on some parabolic cylinder \( Q_{4\varrho}(z_0) \subseteq \Omega_T \) the
following smallness conditions hold:

\[
\int_{Q_{4\varrho}(z_0)} |Du - (Du)_{z_0,4\varrho}|^p \, dz \leq \varepsilon_0 \quad \text{and} \quad |(Du)_{z_0,4\varrho}| \leq M_0,
\]

\((4.13)\).
with some constants $\varepsilon_0 \in (0, 1]$ and $M_0 \geq 1$. For sufficiently small values of $\varrho$, the first smallness condition is satisfied for all $z_0 \in \Omega_T \setminus \Sigma_1$, while the second smallness condition is satisfied for all $z_0 \in \Omega_T \setminus \Sigma_2$, where $\Sigma_1, \Sigma_2$ are the sets from the characterization of the singular set (1.8) and (1.9). Let $\ell_{z_0, \varrho}$ be the affine function defined in (2.2) (for $\lambda = 1$). An application of (4.4) with $E \equiv 0$ yields

$$|Df_{z_0, \varrho}| = \frac{n + 2}{q^2} \left( \int_{Q_{c}(z_0)} (u - (u)_{z_0, \varrho}) \otimes (x - x_0) \, dz \right)$$

$$\leq (n + 2) \left( \int_{Q_{c}(z_0)} \left| \frac{u - (u)_{z_0, \varrho}}{\varrho} \right| \, dz \right) \leq (n + 2) \left( \int_{Q_{c}(z_0)} \left| \frac{u - (u)_{z_0, \varrho}}{\varrho} \right|^2 \, dz \right)^{\frac{1}{2}}$$

$$\leq C_{\text{point}} \left[ \left( \int_{Q_{2c}(z_0)} |Du|^p \, dz \right)^{\frac{1}{p-1}} + \left( \int_{Q_{2c}(z_0)} |Du|^p \, dz \right)^{\frac{1}{2}} \right]$$

$$+ C_{\text{point}} \omega^{\frac{p-1}{2}} \int_{Q_{2c}(z_0)} |u - (u)_{z_0, 2\varrho}|^2 \, dz + V^{\frac{p-1}{2}} (4\varrho^2).$$

Due to (4.13) we have

$$\int_{Q_{c}(z_0)} |Du|^p \, dz \leq 2^{n+2} \int_{Q_{4c}(z_0)} |Du|^p \, dz$$

$$\leq 2^{n+2+p-1} \left( \int_{Q_{2c}(z_0)} |Du - (Du)_{z_0, 4\varrho}|^p \, dz + |(Du)_{z_0, 4\varrho}|^p \right)$$

$$\leq 2^{n+p+1} (\varepsilon_0 + M_0^p) \leq CM_0^p,$$

where we used the assumption that $\varepsilon_0 \leq 1 \leq M_0$ in the last step. Using the fact that $\omega \leq 1$ and $V \leq 2L$, we obtain

$$|D\ell_{z_0, \varrho}| \leq C \left[ M_0^{\frac{p-1}{2}} + M_0 + 1 \right] \leq CM_0^{\frac{m(n-1)}{2}}$$

(4.14)

for a constant $C = C(N, n, p, \nu, L)$.

On the other hand, for the excess iteration in the partial regularity proof we will also need an estimate for the excess functional $\Psi(z_0, \varrho, \ell_{z_0, \varrho})$ from (3.2). An application of (4.4) with $E = (Du)_{z_0, 4\varrho}$ combined with the quasi-minimizing property of $\ell_{z_0, \varrho}$ yields

$$\Psi(z_0, \varrho, \ell_{z_0, \varrho}) = \int_{Q_c(z_0)} \left| \frac{u - \ell_{z_0, \varrho}}{\varrho(1 + |D\ell_{z_0, \varrho}|)} \right|^2 + \left| \frac{u - \ell_{z_0, \varrho}}{\varrho(1 + |D\ell_{z_0, \varrho}|)} \right|^p \, dz$$

$$\leq C \int_{Q_c(z_0)} \left| \frac{u - (u)_{z_0, \varrho} - (Du)_{z_0, 4\varrho}(x - x_0)}{\varrho} \right|^2 \, dz$$

$$+ C \int_{Q_c(z_0)} \left| \frac{u - (u)_{z_0, \varrho} - (Du)_{z_0, 4\varrho}(x - x_0)}{\varrho} \right|^p \, dz$$

$$\leq C(1 + |(Du)_{z_0, 4\varrho}|)^{p(m-2)} \left( \int_{Q_{2c}(z_0)} |Du - (Du)_{z_0, 4\varrho}|^p \, dz \right)^{p-1}$$

$$+ C(1 + |(Du)_{z_0, 4\varrho}|)^{p(m-2)} \left( \int_{Q_{2c}(z_0)} |Du - (Du)_{z_0, 4\varrho}|^p \, dz \right)^{\frac{2}{p}}.$$
Thus, we obtain the following estimate for the excess functional:

\[ + C(1 + |(Du)_{z_0, A_0}|)^{p(m-1)} \omega^{2(p-1)} \left( \int_{Q_2(z_0)} |u - (u)_{z_0, 2g}|^2 \, dz \right) \]

\[ + C(1 + |(Du)_{z_0, A_0}|)^{p(m-1)} V^{2(p-1)} \frac{\omega}{m} (4g^2) \]

\[ \leq C(1 + M_0)^{p(m-1)} \left[ \epsilon_0^{p-1} + \epsilon_0^{2/p} + V^{2(p-1)} \frac{\omega}{m} (4g^2) \right] \]

\[ + C(1 + M_0)^{p(m-1)} \omega^{2(p-1)} \frac{\omega}{m} \left( \int_{Q_2(z_0)} |u - (u)_{z_0, 2g}|^2 \, dz \right). \]

In these calculations we also used that

\[ \int_{Q_{2\epsilon}(z_0)} |Du - (Du)_{z_0, 4g}|^p \, dz \leq 2^{n+2} \int_{Q_4(z_0)} |Du - (Du)_{z_0, 4g}|^p \, dz \leq 2^{n+2} \epsilon_0. \]

Applying (4.4) with \( E \equiv 0 \) and arguing in the same way as in the estimate for \( |D\ell_{z_0, g}| \), we obtain

\[ \int_{Q_{2\epsilon}(z_0)} |u - (u)_{z_0, 2g}|^2 \, dz = (2g)^2 \int_{Q_{2\epsilon}(z_0)} \left| \frac{u - (u)_{z_0, 2g}}{2g} \right|^2 \, dz \]

\[ \leq C \left( M_0^{p(p-1)} + M_0^2 + 1 \right) (2g)^2 \leq C M_0^{m(m-1)} (2g)^2. \]

Due to the sublinearity of \( \omega \) we have

\[ \omega^{2(p-1)} \frac{\omega}{m} \left( \int_{Q_{2\epsilon}(z_0)} |u - (u)_{z_0, 2g}|^2 \, dz \right) \leq C M_0^{2(p-1)(m-1)} \omega^{2(p-1)} (4g^2). \]

Thus, we obtain the following estimate for the excess functional:

\[ \Psi(z_0, g, \ell_{z_0, g}) \leq C(1 + M_0)^{p(m-1)} \left[ \epsilon_0^{p-1} + \epsilon_0^{2/p} + V^{2(p-1)} \frac{\omega}{m} (4g^2) \right] \]

\[ + C(1 + M_0)^{(3p-2)(m-1)} \omega^{2(p-1)} (4g^2). \]

(4.15)

The two inequalities (4.14), (4.15) will later enable us to iterate the excess decay estimate, see Steps 2-4 in the proof of Theorem 1.5 below.

5. Linearization. The following lemma indicates that \( u - \ell \) “almost solves a linear parabolic system with constant coefficients” (for some suitably chosen affine function \( \ell : \mathbb{R}^n \to \mathbb{R}^N \)), which will be necessary for the application of the \( A \)-caloric approximation lemma in the proof of Lemma 6.1.

**Lemma 5.1.** Let \( u \in L^p(-T, 0; W^{1,p}(\Omega, \mathbb{R}^N)) \cap C^0(-T, 0; L^2(\Omega, \mathbb{R}^N)) \) be a weak solution of (1.1) and let \( M \geq 1 \). There exists a positive constant \( C_{lin}(\nu, p, \nu, L, M) \) such that the following holds: Let \( Q_{\nu/2}^{(\lambda)}(z_0) \subseteq \Omega_T \) be a scaled parabolic cylinder with center \( z_0 = (x_0, t_0) \), radius \( \nu \in (0, 1) \) and scaling factor \( \lambda \geq 1 \). Furthermore let \( \ell : \mathbb{R}^n \to \mathbb{R}^N \) be an affine function that is independent of \( t \) and satisfies the intrinsic coupling (3.3). Then the following inequality holds true:

\[ \left| \int_{Q_{\nu/2}^{(\lambda)}(z_0)} (u - \ell) \cdot \varphi - \langle (Du - \ell)(\xi(x_0), \ell_0) \rangle Q_{\nu/2}^{(\lambda)}(Du - \ell, D\varphi) \, dz \right| \]

\[ \leq C_{lin}(1 + |D\ell|)^{p-1} \left[ \mu^{1/2} \left( \frac{\Psi^\lambda_\nu(\varphi)}{\Psi^\lambda_\nu(g)} \right)^2 + \Psi^\lambda_\nu(g) \right] \sup_{Q_{\nu/2}^{(\lambda)}(z_0)} |D\varphi|, \]

(5.1)

for all test functions \( \varphi \in C_0^\infty(Q_{\nu/2}^{(\lambda)}(z_0), \mathbb{R}^N) \).
Proof. To simplify the notation we will again assume that \( z_0 = 0 \). Furthermore, we assume w.l.o.g. that \( \sup_{Q^{(\lambda)}_{e/2}} |D\varphi| = 1 \). Using the identities

\[
\int_{Q^{(\lambda)}_{e/2}} \ell \cdot \varphi_t \, dz = 0 \quad \text{and} \quad \int_{Q^{(\lambda)}_{e/2}} \langle (A(\cdot, \ell(0), D\ell)), Q^{(\lambda)}_{e/2}, D\varphi \rangle \, dz = 0
\]

we can rewrite the weak formulation (1.7) in the following way:

\[
\int_{Q^{(\lambda)}_{e/2}} (u - \ell) \cdot \varphi_t - \langle (D\xi A(\cdot, \ell(0), D\ell)), Q^{(\lambda)}_{e/2}, D\varphi \rangle \, dz \\
= \int_{Q^{(\lambda)}_{e/2}} \langle (A(z, u, Du) - (A(\cdot, u, Du)), Q^{(\lambda)}_{e/2}, D\varphi) \rangle \\
+ \int_{Q^{(\lambda)}_{e/2}} \langle (A(\cdot, u, Du) - (A(\cdot, \ell(0), Du)), Q^{(\lambda)}_{e/2}, D\varphi) \rangle \\
+ \int_{Q^{(\lambda)}_{e/2}} \left[ \langle (A(\cdot, \ell(0), Du) - (A(\cdot, \ell(0), D\ell)), Q^{(\lambda)}_{e/2}, D\varphi) \rangle \\
- \langle (D\xi A(\cdot, \ell(0), D\ell)), Q^{(\lambda)}_{e/2}, (Du - D\ell), D\varphi \rangle \right] dz =: I + II + III.
\]

We will now derive estimates for the terms \( I - III \). For the first term we can use the VMO condition (1.5) to obtain:

\[
|I| \leq \int_{Q^{(\lambda)}_{e/2}} \left| A(z, u, Du) - (A(\cdot, u, Du))_{Q^{(\lambda)}_{e/2}} \right| |D\varphi| \, dz \\
\leq \int_{Q^{(\lambda)}_{e/2}} v_{0, e/2}^{(\lambda)}(z)(1 + |Du|)^{p-1} \, dz \\
\leq C(p)(1 + |D\ell|)^{p-1} \int_{Q^{(\lambda)}_{e/2}} v_{0, e/2}^{(\lambda)}(z) \, dz + C(p) \int_{Q^{(\lambda)}_{e/2}} v_{0, e/2}^{(\lambda)}(z) |Du - D\ell|^{p-1} \, dz \\
=: I_1 + I_2.
\]

For the term \( I_1 \) we have

\[
I_1 \leq C(p)(1 + |D\ell|)^{p-1} V \left( \lambda^{2-p} \phi(2) \right) \\
\leq C(p)(1 + |D\ell|)^{p-1} V \left( \lambda^{2-p} \phi^2 \right) \leq C(p)(1 + |D\ell|)^{p-1} \Psi^*_\lambda(\phi).
\]

Let us now consider the term \( I_2 \). In the superquadratic case \( p \geq 2 \) we apply Young’s inequality to obtain

\[
|I_2| \leq C(p)(1 + |D\ell|)^{p-1} \int_{Q^{(\lambda)}_{e/2}} \frac{|Du - D\ell|^{p-1}}{(1 + |D\ell|)^{p-1}} v_{0, e/2}^{(\lambda)}(z) \, dz \\
\leq C(p)(1 + |D\ell|)^{p-1} \left[ \int_{Q^{(\lambda)}_{e/2}} \frac{|Du - D\ell|^{p}}{(1 + |D\ell|)^{p}} \, dz + \int_{Q^{(\lambda)}_{e/2}} (v_{0, e/2}^{(\lambda)}(z))^p \, dz \right] \\
\leq C(p)(1 + |D\ell|)^{p-1} \frac{1 + |D\ell| + |Du - D\ell|^{p-2} |Du - D\ell|^2}{(1 + |D\ell|)^p} \, dz \\
+ C(p, L)(1 + |D\ell|)^{p-1} V(\phi^2) \\
\leq C(n, p, \nu, L, M)(1 + |D\ell|)^{p-1} \Psi^*_\lambda(\phi),
\]

where \( \Psi^*_\lambda(\phi) \) is the best \( \lambda \)-approximation of \( \phi \) with respect to the VMO space.
where we used the Caccioppoli inequality (3.4) (with \( \sigma = 9/2 \)) in the last step and \( v_{0, 9/2} \leq 2L \) in the penultimate step. In the subquadratic case \( p < 2 \) we split the integration domain into \( G \) and \( Q_{9/2}^\lambda \setminus G \), with \( G \) from (3.5). As in the superquadratic case, this also yields

\[
|I_2| \leq C(p) (1 + |D\ell|)^{p-1} \int_{Q_{9/2}^\lambda \setminus G} v_{0, 9/2}^{(\lambda)}(z) \, dz
\]

\[
+ C(p, L) \int_G (1 + |D\ell| + |Du - D\ell|)^{p-3} |Du - D\ell|^2 \, dz
\]

\[
\leq C(p) (1 + |D\ell|)^{p-1} \mathbf{V}(\lambda^{2-p} \rho^2)
\]

\[
+ C(p, L)(1 + |D\ell|)^{-1} \int_{Q_{9/2}^\lambda} (1 + |D\ell| + |Du - D\ell|)^{p-2} |Du - D\ell|^2 \, dz
\]

\[
\leq C(n, p, \nu, L, M)(1 + |D\ell|)^{p-1} \Psi_\lambda^*(\rho),
\]

where we used the Caccioppoli inequality (3.4) (with \( \sigma = 9/2 \)) in the last step and \( v_{0, 9/2} \leq 2L \) in the first step. Adding up the estimates for \( I_1 \) and \( I_2 \) yields the following estimate for \( I \):

\[
|I| \leq C(n, p, \nu, L, M)(1 + |D\ell|)^{p-1} \Psi_\lambda^*(\rho).
\]

We will now estimate the term \( II \). By (1.4) we have

\[
\left| (A(\cdot, u(z), Du(x)))_{Q_{\rho}^\lambda} - (A(\cdot, \ell(0), Du(x)))_{Q_{\rho}^\lambda} \right|
\]

\[
\leq \int_{Q_{\rho}^\lambda} |A(y, u(z), Du(x)) - A(y, \ell(0), Du(x))| \, dy
\]

\[
\leq L\omega \left( |u(z) - \ell(0)|^2 \right) (1 + |Du(x)|)^{p-1}.
\]

This yields

\[
|II| \leq L \int_{Q_{\rho}^\lambda} \omega \left( |u - \ell(0)|^2 \right) (1 + |Du|)^{p-1} \, dz.
\]

Arguing in the same way as for \( I \), we see that

\[
|II| \leq C(n, p, \nu, L, M)(1 + |D\ell|)^{p-1} \Psi_\lambda^*(\rho).
\]

We can use the fundamental theorem of calculus and the structure assumption (1.4) to estimate the third term in the following way:

\[
III = \int_{Q_{\rho}^\lambda} \int_0^1 \left( D_\xi A(\cdot, \ell(0), D\ell + s(Du - D\ell)) - D_\xi A(\cdot, \ell(0), D\ell) \right)_{Q_{\rho}^\lambda} \times (Du - D\ell, D\varphi) \, ds \, dz
\]

\[
\leq L \int_{Q_{\rho}^\lambda} \int_0^1 \mu \left( \frac{s |Du - D\ell|}{1 + |D\ell| + |D\ell + s(Du - D\ell)|} \right) \times (1 + |D\ell| + |D\ell + s(Du - D\ell)|)^{p-2} |Du - D\ell| \, ds \, dz.
\]

Note that we have

\[
1 + |D\ell| + |D\ell + s(Du - D\ell)| \leq 2(1 + |D\ell| + |Du - D\ell|)
\]

(5.2)
and
\[1 + |D\ell| + |D\ell + s(Du - D\ell)| \geq 1 + |D\ell| + s|Du| - (1 - s)|D\ell| \geq s(1 + |Du| + |D\ell|) \geq \frac{s}{2}(1 + |D\ell| + |Du - D\ell|). \tag{5.3}\]

Let us now consider the case \( p < 2 \). Using the auxiliary inequality (5.3) and the sublinearity of \( \mu \), we get the following estimate for the third term:
\[
|III| \leq L \int_{Q^\lambda_{e/2}} \int_0^1 \mu \left( \frac{2s |Du - D\ell|}{s(1 + |D\ell| + |Du - D\ell|)} \right) \\
\times \left[ \frac{s}{2}(1 + |D\ell| + |Du - D\ell|) \right]^{p-2} ds |Du - D\ell| \ dz \\
\leq C \int_{Q^\lambda_{e/2}} \mu \left( \frac{|Du - D\ell|}{1 + |D\ell| + |Du - D\ell|} \right) \left( 1 + |D\ell| + |Du - D\ell| \right)^{p-2} |Du - D\ell| \ dz \\
\leq C(1 + |D\ell|)^{p-2} \int_{Q^\lambda_{e/2}} \mu \left( \frac{|Du - D\ell|}{1 + |D\ell| + |Du - D\ell|} \right) \\
\times (1 + |D\ell| + |Du - D\ell|)^{p-2} |Du - D\ell| \ dz,
\]
where \( C = C(p) \). Now we apply Hölder’s inequality, the Caccioppoli inequality from Lemma 3.1 (with \( \sigma = \theta/2 \)) and use the fact that \( \mu \) is concave and bounded from above by 1 to obtain
\[
|III| \leq C(p) (1 + |D\ell|)^{p-2} \left( \int_{Q^\lambda_{e/2}} (1 + |D\ell| + |Du - D\ell|)^{p-2} |Du - D\ell|^2 \ dz \right)^{1/2} \\
\times \left( \int_{Q^\lambda_{e/2}} \mu^2 \left( \frac{|Du - D\ell|}{1 + |D\ell| + |Du - D\ell|} \right) \ dz \right)^{1/2} \\
\leq C(n, p, \nu, L, M) (1 + |D\ell|)^{p-1} \sqrt{\Psi^\lambda_{\gamma}(\theta)} \left( \int_{Q^\lambda_{e/2}} \mu \left( \frac{|Du - D\ell|}{1 + |D\ell| + |Du - D\ell|} \right) \ dz \right)^{1/2} \\
\leq C(n, p, \nu, L, M) (1 + |D\ell|)^{p-1} \sqrt{\Psi^\lambda_{\gamma}(\theta) \mu^{1/2}} \left( \int_{Q^\lambda_{e/2}} \frac{|Du - D\ell|}{1 + |D\ell| + |Du - D\ell|} \ dz \right)^{1/2} \\
\leq C(n, p, \nu, L, M) (1 + |D\ell|)^{p-1} \sqrt{\Psi^\lambda_{\gamma}(\theta)} \\
\times \mu^{1/2} \left[ (1 + |D\ell|)^{-\frac{p}{2}} \left( \int_{Q^\lambda_{e/2}} (1 + |D\ell| + |Du - D\ell|)^{p-2} |Du - D\ell|^2 \ dz \right)^{1/2} \right] \\
\leq C(n, p, \nu, L, M) (1 + |D\ell|)^{p-1} \sqrt{\Psi^\lambda_{\gamma}(\theta) \mu^{1/2}} \sqrt{\Psi^\lambda_{\gamma}(\theta)}.
\]

Let us now turn our attention to the case \( p \geq 2 \). Using the auxiliary inequality (5.2), we get the following estimate for the third term:
\[
|III| \leq L \int_{Q^\lambda_{e/2}} \int_0^1 \mu \left( \frac{s |Du - D\ell|}{s + s |D\ell|} \right) (1 + |D\ell| + |Du - D\ell|)^{p-2} ds |Du - D\ell| \ dz
\]
\[ \begin{align*}
&= L \int_{Q_{n/2}^{(n)}} \mu \left( \frac{|Du - D\ell|}{1 + |D\ell|} \right) (1 + |D\ell| + |Du - D\ell|)^{p-2} |Du - D\ell| \, dz \\
&\leq C (1 + |D\ell|)^{p-1} \int_{Q_{n/2}^{(n)}} \mu \left( \frac{|Du - D\ell|}{1 + |D\ell|} \right) \left[ |Du - D\ell| + |Du - D\ell|^{p-1} \right] (1 + |D\ell|)^{p-1} \, dz,
\end{align*}\]

where \( C = C(p, L) \). An application of Hölder’s inequality yields

\[ |III| \leq C (1 + |D\ell|)^{p-1} \left( \int_{Q_{n/2}^{(n)}} \mu^2 \left( \frac{|Du - D\ell|}{1 + |D\ell|} \right) \, dz \right)^{\frac{1}{2}} \left( \int_{Q_{n/2}^{(n)}} \frac{|Du - D\ell|^2}{(1 + |D\ell|)^2} \, dz \right)^{\frac{1}{2}} \]

\[ + C (1 + |D\ell|)^{p-1} \left( \int_{Q_{n/2}^{(n)}} \mu^p \left( \frac{|Du - D\ell|}{1 + |D\ell|} \right) \, dz \right)^{\frac{1}{2}} \left( \int_{Q_{n/2}^{(n)}} \frac{|Du - D\ell|^p}{(1 + |D\ell|)^p} \, dz \right)^{\frac{p-1}{p}}, \]

again with \( C = C(p, L) \). Due to the Caccioppoli inequality (3.4) we have

\[ \int_{Q_{n/2}^{(n)}} \frac{|Du - D\ell|^q}{(1 + |D\ell|)^q} \, dz \leq (1 + |D\ell|)^{-p} \int_{Q_{n/2}^{(n)}} (1 + |D\ell| + |Du - D\ell|)^{p-2} |Du - D\ell|^2 \, dz \]

\[ \leq C(n, p, \nu, L, M) \Psi_\lambda^*(q) \quad (5.4) \]

for \( q = 2 \) and \( q = p \). Under consideration of \( \mu \leq 1 \) and the concavity of \( \mu \) it follows that

\[ |III| \leq C (1 + |D\ell|)^{p-1} \mu^{1/2} \left( \int_{Q_{n/2}^{(n)}} \frac{|Du - D\ell|}{1 + |D\ell|} \, dz \right) \sqrt{\Psi_\lambda^*(q)} \]

\[ + C (1 + |D\ell|)^{p-1} \mu^{1/p} \left( \int_{Q_{n/2}^{(n)}} \frac{|Du - D\ell|}{1 + |D\ell|} \, dz \right) \Psi_\lambda^*(q)^{\frac{p-1}{p}} \]

\[ \leq C (1 + |D\ell|)^{p-1} \mu^{1/2} \left[ \left( \int_{Q_{n/2}^{(n)}} \frac{|Du - D\ell|^2}{(1 + |D\ell|)^2} \, dz \right)^{\frac{1}{2}} \sqrt{\Psi_\lambda^*(q)} \right] \]

\[ + C (1 + |D\ell|)^{p-1} \mu^{1/p} \left[ \left( \int_{Q_{n/2}^{(n)}} \frac{|Du - D\ell|^2}{(1 + |D\ell|)^2} \, dz \right)^{\frac{1}{2}} \Psi_\lambda^*(q)^{\frac{p-1}{p}} \right] \]

\[ \leq C (1 + |D\ell|)^{p-1} \left[ \mu^{1/2} \left( \sqrt{\Psi_\lambda^*(q)} \right)^2 \Psi_\lambda^*(q) + \mu^{1/p} \left( \sqrt{\Psi_\lambda^*(q)} \right)^{\frac{p-1}{p}} \right], \]

again with \( C = C(n, p, \nu, L, M) \). Now we apply Young’s inequality in the form

\[ a^{\frac{1}{r}} b^{\frac{r-1}{r}} = a^{\frac{1}{r}} b^{\frac{r-1}{r}} \leq a^{\frac{1}{r} \frac{1}{2} \frac{r-2}{r}} b^{\frac{1}{r} \frac{r-2}{r}} a b \]

\[ \mu^{1/2} \left( \sqrt{\Psi_\lambda^*(q)} \right)^2 \Psi_\lambda^*(q) + \mu^{1/p} \left( \sqrt{\Psi_\lambda^*(q)} \right)^{\frac{p-1}{p}} \]

Combining the estimates for the terms \( I - III \) yields the desired inequality (5.1).

6. Proof of the partial regularity result. The following lemma is the central argument in the proof of the main theorem. We show an excess decay estimate for weak solutions from one scaled cylinder to another scaled cylinder with a smaller radius and a different scaling factor. The precise formulation reads as follows:
Lemma 6.1. Let $u \in L^p(-T, 0; W^{1,p}(\Omega, \mathbb{R}^N)) \cap C^0(-T, 0; L^2(\Omega, \mathbb{R}^N))$ be a weak solution of (1.1) and let $M \geq 1$ and

$$0 < \vartheta < \frac{1}{4} \min \left\{ (2M)^{\frac{p-2}{2}}, 2^{\frac{p-2}{2}} \right\} =: \vartheta_*$$

There exist constants $\varepsilon_1(N, n, p, \nu, L, M, \vartheta, \mu(\cdot)) \in (0, 1]$ and $C_\ast(N, n, p, \nu, L, M) \geq 1$ such that the following holds: If $Q_*^{(\lambda)}(z_0) \subseteq \Omega_T$ is a scaled parabolic cylinder on which the intrinsic coupling

$$\lambda \leq 1 + |D\ell_0^{(\lambda)}| \leq M\lambda$$

holds for the affine function $\ell_0^{(\lambda)}$ from (2.2), and moreover the smallness condition

$$\Psi_\ast(z_0, \vartheta_0, \ell_0^{(\lambda)}) \leq \varepsilon_1 \quad (6.1)$$

is satisfied, then there exists a new scaling factor $\lambda_1 \in \left[ \frac{1}{2}, 2M\lambda \right]$ such that

$$1 + |D\ell_0^{(\lambda_1)}| = \lambda_1 \quad (6.2)$$

and the following mixed excess decay estimate holds:

$$\Psi_{\lambda_1}(z_0, \vartheta_0, \ell_0^{(\lambda_1)}) \leq C_{\ast}^{\text{min}(p, 2)} \Psi_{\lambda_1}^{\ast}(z_0, \vartheta_0, \ell_0^{(\lambda)}) \quad (6.3)$$

Proof. We will assume w.l.o.g. that $z_0 = (x_0, t_0) = (0, 0)$. In the following we will usually omit the reference point and write e.g. $Q_*^{(\lambda)}$ instead of $Q_*^{(\lambda)}(0, \vartheta)$ or $\ell_0^{(\lambda)}$ instead of $\ell_0^{(\lambda)}(0, \vartheta)$ and $\Psi_\ast(\vartheta)$ instead of $\Psi_\ast(0, \vartheta, \ell)$. We only need to consider the case $\Psi_\ast(\vartheta) > 0$, since (6.3) holds trivially if $\Psi_\ast(\vartheta) = 0$. We define an auxiliary function $v$ via

$$v(x, t) := \frac{u(x, \lambda^{2-p}t) - \ell_0^{(\lambda)}(x)}{\gamma_1 \left(1 + |D\ell_0^{(\lambda)}|\right)} \sqrt{\Psi_\ast(\vartheta)}$$

for $(x, t) \in Q_\vartheta \equiv Q_1^{(1)}$

and $\gamma := \sqrt{\Psi_\ast(\vartheta)} \leq 1$. The constant $\gamma_1 \geq 1$ will be chosen later, only depending on $N, n, p, \nu, L$ and $M$. We want to apply the $A$-caloric approximation lemma for $v$. First of all we need to check that the requirement (2.4) is satisfied. We start by observing that

$$\frac{1}{Q_{\vartheta/2}} \int_{Q_{\vartheta/2}} \left| \frac{v}{\vartheta/2} \right|^2 + \gamma^{p-2} \left| \frac{v}{\vartheta/2} \right|^p \, dz \leq \frac{2^{\text{max}(p, 2)} \gamma^{2\text{min}(p, 2)}}{\gamma_1^{\text{min}(p, 2)}} \int_{Q_{\vartheta/2}^{(\lambda)}} \left| \frac{u - \ell_0^{(\lambda)}}{\vartheta(1 + |D\ell_0^{(\lambda)}|)} \right|^2 + \left| \frac{u - \ell_0^{(\lambda)}}{\vartheta(1 + |D\ell_0^{(\lambda)}|)} \right|^p \, dz \leq \frac{\gamma^{2\text{min}(p, 2)} \gamma^{2\text{min}(p, 2)} + 2^{\text{max}(p, 2)} \gamma^{2\text{min}(p, 2)}}{\gamma_1^{\text{min}(p, 2)}} \,$$

where we increased the domain of integration from $Q_{\vartheta/2}^{(\lambda)}$ to $Q_\vartheta^{(\lambda)}$ and used the identity $\gamma = \sqrt{\Psi_\ast(\vartheta)}$.

We will now turn our attention to the case $\frac{2n}{n+2} < p < 2$. Under consideration of $\gamma_1 \geq 1, \gamma \leq 1$ and $|D\ell_0^{(\lambda)}| \leq 1$, an application of Lemma 2.3 (ii), (v), (iii) yields

$$|V(Dv)|^2 = \left| V \left( \frac{Du - D\ell_0^{(\lambda)}}{c_1(1 + |D\ell_0^{(\lambda)}|)} \right) \right|^2 \leq \max \left\{ \frac{1}{(c_1\gamma)^2}, \frac{1}{(c_1\gamma)^p} \right\} \left| V \left( \frac{Du - D\ell_0^{(\lambda)}}{1 + |D\ell_0^{(\lambda)}|} \right) \right|^2 \,$$
\[ \leq \frac{c(p)}{c_1^\gamma} \left| V \left( \frac{Du}{1 + |D\ell|} \right) - V \left( \frac{D\ell}{1 + |D\ell|} \right) \right|^2 \]
\[ \leq \frac{c(N, n, p)}{c_1^\gamma} \left( 1 + \frac{|Du|}{1 + |D\ell|} \right)^2 + \left( 1 + \frac{|D\ell|}{1 + |D\ell|} \right)^2 \left| \frac{Du}{1 + |D\ell|} - \frac{D\ell}{1 + |D\ell|} \right|^2 \]
\[ \leq \frac{c(N, n, p)}{c_1^\gamma} (1 + |D\ell|)^{-p} (1 + |Du| + |D\ell|)^{p-2} |Du - D\ell|^2 \]
\[ \leq \frac{c(N, n, p)}{c_1^\gamma} (1 + |D\ell|)^{-p} (1 + |D\ell| + |Du - D\ell|)^{p-2} |Du - D\ell|^2. \]

From the definition of \( v \) and the Caccioppoli inequality (3.4) we therefore infer that
\[ \sup_{s \in \Lambda_{x,s/2}} \int_{B_{x,s/2}} \left| \frac{v(x, s)}{\rho/2} \right|^2 dx + \int_{Q_{x,s/2}} |V(Dv)|^2 dz \]
\[ \leq \frac{\lambda^{2-p}}{c_1^\gamma (1 + |D\ell|)^2} \sup_{s \in \Lambda_{x,s/2}} \int_{B_{x,s/2}} \left| u(x, s) - \ell \right|^2 dx + \frac{c(N, n, p)}{c_1^\gamma (1 + |D\ell|)^p} \int_{Q_{x,s/2}} (1 + |D\ell| + |Du - D\ell|)^{p-2} |Du - D\ell|^2 dz \]
\[ \leq \frac{C_{cacc} \lambda^{2-p}}{c_1^2 (1 + |D\ell|)^{2-p}} + \frac{C(N, n, p, C_{cacc})}{c_1^2} \leq \frac{C}{c_1^2}, \]
where \( C \) also depends on the constant from the Caccioppoli inequality. Hence we obtain
\[ \sup_{s \in \Lambda_{x,s/2}} \int_{B_{x,s/2}} \left| \frac{v(x, s)}{\rho/2} \right|^2 dx + \int_{Q_{x,s/2}} \left| \frac{v}{\rho/2} \right|^2 + \gamma^{p-2} \left| \frac{v}{\rho/2} \right|^p + |V(Dv)|^2 dz \leq \frac{2^{n+p} + C}{c_1^p} \leq 1, \]
if \( c_1 \) is chosen sufficiently large (depending on \( N, n, p, \nu, L \) and \( M \)), and therefore (2.4)_1 is satisfied. Let us now consider the case \( p \geq 2 \). By the Caccioppoli inequality (3.4) we obtain
\[ \int_{Q_{x,s/2}} |Dv|^2 + \gamma^{p-2} |Dv|^p dz = \frac{1}{\gamma} \int_{Q_{x,s/2}} \left| \frac{Du - D\ell}{1 + |D\ell|} \right|^2 + \left| \frac{Du - D\ell}{1 + |D\ell|} \right|^p dz \leq \frac{C_{cacc}}{c_1^2}. \]
Hence we obtain
\[ \int_{Q_{x,s/2}} \left| \frac{v}{\rho/2} \right|^2 + |Dv|^2 dz + \gamma^{p-2} \int_{Q_{x,s/2}} \left| \frac{v}{\rho/2} \right|^p + |Dv|^p dz \leq \frac{2^{n+p+2} + C_{cacc}}{c_1^2} \leq 1, \]
if \( c_1 \gg 1 \) is chosen sufficiently large (depending on \( N, n, p, \nu, L \) and \( M \)), and thus (2.4)_2 is satisfied.

Let us now define the bilinear form
\[ A(w, \overline{w}) := \langle (D\xi A(\cdot, \ell(0), D\ell)) Q_{x,s/2} w, \overline{w} \rangle \quad \text{for } w, \overline{w} \in \mathbb{R}^{N \times n}. \]
We start by checking that \( A \) is elliptic and bounded:
\[ A(w, w) = \frac{1}{\lambda^{p-2}} \int_{Q_{x,s/2}} \langle D\xi A(z, \ell(0), D\ell)w, w \rangle dz \]
\[ \geq \nu \frac{(1 + |D\ell|)^{p-2}}{\lambda^{p-2}} |w|^2 \geq \nu M^{\min(p,2)-2} |w|^2 \]
for all $w \in \mathbb{R}^{N \times n}$, where we used the ellipticity condition $(1.3)_2$ and the intrinsic coupling $(3.3)$. On the other hand, using the growth condition $(1.3)_1$, we obtain

$$|A(w, \overline{w})| \leq \frac{1}{\lambda_{p-2}} \int_{Q_{\epsilon/2}} |(D_\xi A(z, \ell(0), D\ell)w, \overline{w})| \, dz$$

$$\leq L \left(1 + \frac{|D\ell|^p - 2}{\lambda_{p-2}}\right) |w| \, |\overline{w}| \leq LM^{\max(p,2) - 2} |w| \, |\overline{w}|.$$  

It only remains to verify that the condition $(2.5)$ from the $A$-caloric approximation lemma is also satisfied. Let $\varphi \in C_0^\infty(Q_{\epsilon/2}, \mathbb{R}^N)$ be an arbitrary test function. We can define a rescaled test function $\tilde{\varphi} \in C_0^\infty(Q_{\epsilon/2}(x, \lambda^2 t))$ by $\tilde{\varphi}(x, \lambda^2 t) = \varphi(x, t)$. By chain rule we have $D\varphi = D\tilde{\varphi}$. The application of Lemma 5.1 yields after a standard transformation in the time variable:

$$\left| \int_{Q_{\epsilon/2}} v \cdot \varphi_t - A(Dv, D\varphi) \, dz \right|$$

$$= \frac{1}{c_1} \left(1 + |D\ell|\right) \frac{1}{\gamma \lambda_2} \left| \int_{Q_{\epsilon/2}} (u - \ell) \cdot \varphi_t - ((D_\xi A(\cdot, \ell(0), D\ell))_{Q_{\epsilon/2}} (Du - D\ell), D\tilde{\varphi}) \, dz \right|$$

$$\leq C_{\lim} \left(1 + |D\ell|\right) \frac{1}{c_1 \gamma \lambda_2} \left[ \Psi_1^\lambda(\varphi) + \sqrt{\Psi_1^\lambda(\varphi)} \mu^{1/2} \left(\sqrt{\Psi_1^\lambda(\varphi)}\right) \right] \sup_{Q_{\epsilon/2}} |D\tilde{\varphi}|$$

$$\leq C_{\lim} M^{\max(p,2) - 2} \frac{1}{c_1} \left[ \sqrt{\Psi_1^\lambda(\varphi)} + \mu^{1/2} \left(\sqrt{\Psi_1^\lambda(\varphi)}\right) \right] \sup_{Q_{\epsilon/2}} |D\varphi| \leq \delta \sup_{Q_{\epsilon/2}} |D\varphi|,$$

if $c_1 = c_1(N, n, p, \nu, L, M)$ is chosen sufficiently large and if the smallness condition

$$\mu^{1/2} \left(\sqrt{\Psi_1^\lambda(\varphi)}\right) + \sqrt{\Psi_1^\lambda(\varphi)} \leq \delta \quad (6.4)$$

is satisfied, where $\delta \in (0, 1]$ is the constant from the $A$-caloric approximation lemma, which depends on some $\varepsilon > 0$ that remains to be chosen. Now we can finally apply the $A$-caloric approximation lemma which yields an $A$-caloric function $h \in C^\infty(Q_{\epsilon/4}, \mathbb{R}^N)$ satisfying

$$\left\{ \begin{aligned} &\int_{Q_{\epsilon/4}} \left| \frac{h}{\theta/4} \right|^2 + \gamma^{p-2} \left| \frac{h}{\theta/4} \right|^p + |V(Dh)|^2 \, dz \leq C(n, p), \quad p < 2 \\
&\int_{Q_{\epsilon/4}} \left| \frac{h}{\theta/4} \right|^2 + |Dh|^2 + \gamma^{p-2} \left( \left| \frac{h}{\theta/4} \right|^p + |Dh|^p \right) \, dz \leq C(n, p), \quad p \geq 2 \end{aligned} \right. \quad (6.5)$$

and

$$\int_{Q_{\epsilon/4}} \left| \frac{v - h}{\theta/4} \right|^2 + \gamma^{p-2} \left| \frac{v - h}{\theta/4} \right|^p \, dz \leq \varepsilon. \quad (6.6)$$

In the next step we want to transfer the estimate $(6.5)$ for $h$ to the auxiliary function $v$. For this we recall the definition

$$m := \max\{p, 2\}.$$
Furthermore, let
\[
\theta := \begin{cases} 
\frac{4\theta}{(2M)^{\frac{4\theta}{2+\gamma}}}, & 2n < p < 2 \\
\frac{4\theta}{2^{\frac{4\theta}{2+\gamma}}}, & p \geq 2.
\end{cases}
\tag{6.7}
\]
Due to \( \theta < \theta_\ast \) we have \( \theta \in (0, 1) \). From Lemma 2.8 it follows that the following decay estimate for \( h \) holds true for \( s = 2 \) and \( s = p \):
\[
\gamma^{s-2} \left( \frac{\theta}{4} \right)^{-s} \int_{Q_{2x}^4} |h - (h)_{\theta/4} - \gamma \frac{m-s}{s} |(Dh)_{\theta/4}|^{\frac{p-s}{p}} (Dh)_{\theta/4}x|^{s} \, dz
\]
\[
\leq c_p \gamma^{s-2} \theta^s \left( \frac{\theta}{4} \right)^{-s} \int_{Q_{\theta/4}^4} |h - (h)_{\theta/4} - \gamma \frac{m-s}{s} |(Dh)_{\theta/4}|^{\frac{p-s}{p}} (Dh)_{\theta/4}x|^{s} \, dz
\]
\[
\leq 3^{s-1} c_p \gamma^{s-2} \theta^s \left( \frac{\theta}{4} \right)^{-s} \left[ \int_{Q_{\theta/4}^4} |h|^s + |(h)_{\theta/4}|^s + \gamma^{m-s} |(Dh)_{\theta/4}|^p |x|^s \, dz \right]
\]
\[
\leq 2 \cdot 3^{\max\{p,2\}-1} c_p \theta^s \left[ \gamma^{s-2} \int_{Q_{\theta/4}^4} \left| \frac{h}{\theta/4} \right|^s \, dz + \gamma^{m-2} \int_{Q_{\theta/4}^4} |V(Dh)|^p \, dz \right].
\]
**In the case** \( p \geq 2 \) **we have** \( m = p \) **and can use (6.5) to conclude with**
\[
\gamma^{s-2} \left( \frac{\theta}{4} \right)^{-s} \int_{Q_{2x}^4} |h - (h)_{\theta/4} - \gamma \frac{m-s}{s} |(Dh)_{\theta/4}|^{\frac{p-s}{p}} (Dh)_{\theta/4}x|^{s} \, dz
\]
\[
\leq C(N, n, p, \nu, L) \theta^s.
\tag{6.8}
\]
**In the case** \( \frac{2n}{n+2} < p < 2 \) **we use Lemma 2.3 (i) to obtain that**
\[
|Dh|^p \leq 1 + 2^{\frac{2n}{n+2}} |V(Dh)|^2.
\]
This leads to
\[
\int_{Q_{\theta/4}^4} |Dh|^p \, dz \leq 2^{\frac{2n}{n+2}} \int_{Q_{\theta/4}^4} |V(Dh)|^2 \, dz + 1.
\]
Taking into account that \( m = 2 \) and using (6.5) \( 1 \) we also arrive at
\[
\gamma^{s-2} \left( \frac{\theta}{4} \right)^{-s} \int_{Q_{2x}^4} |v - (h)_{\theta/4} - \gamma \frac{m-s}{s} |(Dh)_{\theta/4}|^{\frac{p-s}{p}} (Dh)_{\theta/4}x|^{s} \, dz
\]
\[
\leq C \theta^s \left[ \gamma^{s-2} \int_{Q_{\theta/4}^4} \left| \frac{h}{\theta/4} \right|^s \, dz + \int_{Q_{\theta/4}^4} |V(Dh)|^2 \, dz + 1 \right] \leq C(N, n, p, \nu, L) \theta^s.
\]
Thus we have proved that the estimate (6.8) for \( h \) holds both in the superquadratic and the subquadratic case. Combining (6.8) with (6.6) we obtain
\[
\gamma^{s-2} \left( \frac{\theta}{4} \right)^{-s} \int_{Q_{2x}^4} |v - (h)_{\theta/4} - \gamma \frac{m-s}{s} |(Dh)_{\theta/4}|^{\frac{p-s}{p}} (Dh)_{\theta/4}x|^{s} \, dz
\]
\[
\leq 2^{s-1} \gamma^{s-2} \left( \frac{\theta}{4} \right)^{-s} \int_{Q_{2x}^4} |v - h|^s \, dz
\]
\[
+ 2^{s-1} \gamma^{s-2} \left( \frac{\theta}{4} \right)^{-s} \int_{Q_{2x}^4} |h - (h)_{\theta/4} - \gamma \frac{m-s}{s} |(Dh)_{\theta/4}|^{\frac{p-s}{p}} (Dh)_{\theta/4}x|^{s} \, dz
\]
\[
\leq 2^{\max\{p,2\}-1} \left[ \gamma^{s-2} \theta^{-(n+2+s)} \left( \frac{\theta}{4} \right)^{-s} \int_{Q_{\ell/4}} \left| v - w \right|^s \, dz + C \theta^s \right] \leq C \left[ \theta^{-(n+2+s)} \varepsilon + \theta^s \right]
\]
with a constant \( C = C(N, n, p, \nu, L) \). By choosing \( \varepsilon := \theta^{n+2+2\max\{p,2\}} \), we can simplify the above estimate to obtain
\[
\frac{\theta}{4} \int_{Q_{\ell/4}} \left| u - (h)_{\ell/4} - \gamma \frac{m_{\ell/4}}{\theta} \right| \left| (Dh)_{\ell/4} \right|^s \, dz 
\leq C \left[ \theta^{2\max\{p,2\}-s} + \theta^s \right] \gamma^{2-s} \leq C \theta^s \gamma^{2-s},
\]
where we have used the fact that \( \theta < 1 \). Using the definition of the auxiliary function \( v \) we can scale back to \( u \) which yields
\[
\frac{\theta}{4} \int_{Q_{s/4}} \left| u - \ell - c_1 \left( 1 + |D\ell| \right) \gamma \right| \left| (h)_{s/4} + \gamma \frac{m_{s/4}}{\theta} \right| \left| (Dh)_{s/4} \right|^s \, dz 
\leq C c_1 \left( 1 + |D\ell| \right) \theta^s \gamma^2 \leq C c_1 M^s \lambda^s \theta^s \gamma^2 = C(N, n, p, \nu, L, M) \lambda^s \theta^s \Psi_\lambda(\ell).
\]
Since \( \lambda(\ell) \) is a minimizer (if \( s = 2 \)) resp. a quasiminimizer (if \( s = p \)) of the functional \( F_s[\ell] = \int_{Q_{s/4}} \left| u - \ell \right|^s \, dz \) we also obtain
\[
\frac{\theta}{4} \int_{Q_{s/4}} \left| u - \lambda(\ell) \right|^s \, dz \leq C \lambda^s \theta^s \Psi_\lambda(\ell) \quad (6.9)
\]
We will now determine the new scaling factor \( \lambda_1 \in \left[ \frac{\lambda}{2}, 2M \lambda \right] \) such that (6.2) holds.
First of all we note that \( Q^{(\lambda_1)}_{\theta^2} \subseteq Q^\lambda_{\theta^2} \) holds for every \( \lambda_1 \in \left[ \frac{\lambda}{2}, 2M \lambda \right] \) because of \( \theta < \frac{\theta}{4} \) and
\[
\left\{ \begin{array}{l}
(\theta \theta^2)^2 \lambda^{2-p} \leq (2M)^{p-2} \left( \frac{\theta}{\theta^2} \right)^2 (2M \lambda)^{2-p} = \left( \frac{\theta}{\theta^2} \right)^2 \lambda^{2-p}, \quad \frac{2n}{n+2} < p < 2 \\
(\theta \theta^2)^2 \lambda^{2-p} \leq 2^{2-p} \left( \frac{\theta}{\theta^2} \right)^2 (\lambda/2)^{2-p} \lambda^{2-p} = \left( \frac{\theta}{\theta^2} \right)^2 \lambda^{2-p}, \quad p \geq 2,
\end{array} \right.
\]
where we used (6.7). Next we apply (2.3) with \( \xi = \ell(0) \) and \( w = D\ell \), enlarge the domain of integration from \( Q^{(\lambda_1)}_{\theta^2} \) to \( Q^{(\lambda_1)}_{\theta} \) and use the definition of \( \Psi_\lambda(\ell) \) to obtain:
\[
\left| D\ell^{(\lambda_1)} - D\ell \right|^2 \leq \frac{n(n+2)}{\theta^2} \int_{Q^{(\lambda_1)}_{\theta}} \left| u - \ell(0) - D\ell x \right|^2 \, dz = n(n+2) \int_{Q^{(\lambda_1)}_{\theta}} \left| u - \ell \right|^2 \, dz
\]
\[
\leq n(n+2) \theta^{-(n+4)} \left( \frac{\lambda}{\lambda_1} \right)^{2-p} \int_{Q^{(\lambda_1)}_{\theta}} \left| u - \ell \right|^2 \, dz
\]
\[
\leq n(n+2) \theta^{-(n+4)} \left( 2M \right)^{p-2} \left( \frac{\lambda}{\lambda_1} \right)^{2-p} \left( 1 + |D\ell| \right)^2 \Psi_\lambda(\ell)
\]
\[
\leq n(n+2) \theta^{-(n+4)} \left( 2M \right)^{p-2} \left( 2^{2-p} \right) \lambda^2 \Psi_\lambda(\ell)
\]
\[
= C(n, p, M, \vartheta) \lambda^2 \Psi_\lambda(\ell) \leq \frac{\lambda^2}{4},
\]
if the smallness condition
\[
C(n, p, M, \vartheta) \Psi_\lambda(\ell) \leq \frac{1}{4} \quad (6.10)
\]
is satisfied. Here we have denoted
\[ C(n,p,M,\vartheta) = n(n+2)\vartheta^{-(n+4)} \max \{(2M)^{p-2}, 2^{2-p}\} M^2. \]

Taking square roots on both sides of the preceding inequality yields \(|D\ell^{(\lambda_\vartheta)} - D\ell| \leq \frac{\lambda}{2}\), which ensures that
\[ 1 + |D\ell^{(\lambda_\vartheta)}| \leq 1 + |D\ell| + |D\ell^{(\lambda_\vartheta)} - D\ell| \leq M\lambda + \frac{\lambda}{2} \leq 2M\lambda \]
and on the other hand
\[ 1 + |D\ell^{(\lambda_\vartheta)}| \geq 1 + |D\ell| - |D\ell^{(\lambda_\vartheta)} - D\ell| \geq \lambda - \frac{\lambda}{2} = \frac{\lambda}{2}. \]

We can now define the function
\[ f = \begin{cases} \frac{1}{2nM} & \text{for } n \geq 2, \\ \frac{1}{2} & \text{for } n = 1. \end{cases} \]
which ensures that
\[ f(\lambda) := 1 + |D\ell^{(\lambda_\vartheta)}| - \lambda_\vartheta, \]
which is continuous due to the absolute continuity of the integral. By using the above inequalities we infer that \( f(\lambda) \geq 0, f(2M\lambda) \leq 0 \). Hence we can apply the intermediate value theorem for continuous functions to find some \( \lambda_1 \in \left[ \frac{1}{2}, 2M\lambda \right] \) which satisfies \( f(\lambda_1) = 0 \). It only remains to prove that the mixed excess decay estimate (6.3) holds. To this end we use the energy estimate (6.9) and the intrinsic coupling (6.2) to obtain (again for \( s = 2 \) and \( s = p \)):

\[
\int_{Q_{\vartheta}^{(\lambda_1)}} \left| \frac{u - \ell^{(\lambda_\vartheta)}}{\vartheta} \right|^s \, dz \leq C(n,p) \int_{Q_{\vartheta}^{(\lambda_1)}} \left| \frac{u - \ell^{(\lambda_\vartheta)}}{\vartheta} \right|^s \, dz \leq C \left( \frac{\theta}{4\vartheta} \right)^{n+2+s} \left( \frac{\lambda}{\lambda_1} \right)^2 \left( \frac{\theta}{4} \right)^{-s} \left( \frac{\vartheta}{\lambda_1} \right)^{s} \int_{Q_{\vartheta}^{(\lambda_1)}} \left| u - \ell^{(\lambda_\vartheta)} \right|^s \, dz
\]
\[
= C \left( \frac{\theta}{4\vartheta} \right)^{n+2+s} \left( \frac{\lambda}{\lambda_1} \right)^2 \left( \frac{\vartheta}{\lambda_1} \right)^{s} \int_{Q_{\vartheta}^{(\lambda_1)}} \left| u - \ell^{(\lambda_\vartheta)} \right|^s \, dz
\]
\[
\leq C_s \min(p,2) \left( 1 + |D\ell^{(\lambda_\vartheta)}| \right)^s \Psi_\lambda^*(\vartheta)
\]
with a constant \( C_s = C_s(N,n,p,v,L,M) \), where we also used the quasi-minimizing property of \( \ell^{(\lambda_1)} \) in the first step.

Dividing both sides of this inequality by \((1 + |D\ell^{(\lambda_\vartheta)}|)^s\) and then adding the respective terms for \( s = 2 \) and \( s = p \) yields the desired mixed excess decay estimate (6.3). We finally have to check that (6.4) and (6.10) hold if the initial smallness condition (6.1) is satisfied. If we choose \( \varepsilon_1 \) sufficiently small in dependence on \( \mu(\cdot) \) and \( \delta = \delta(N,n,p,v,L,\varepsilon = \theta^{n+2+2p}) \) the first smallness condition (6.4) will be satisfied. The second smallness condition (6.10) will also be satisfied if \( \varepsilon \) is chosen sufficiently small in dependence on \( n,p,M \) and \( \vartheta \). Recalling that \( \theta = \theta(p,M,\vartheta) \) we conclude that both smallness conditions are satisfied if the initial smallness condition (6.1) is satisfied with a sufficiently small constant \( \varepsilon_1 = \varepsilon_1(N,n,p,v,L,M,\vartheta,\mu(\cdot)) \). This concludes the proof of the lemma.

Now we can finally prove the main result about the partial Hölder continuity of weak solutions:

**Proof of Theorem 1.5.** The proof is divided into several steps.
Step 1: Choice of the constants Let $\alpha \in (0,1)$ and $M \geq 1$ be given. Let $\partial_* (p, M)$ and $C_* (N, n, p, \nu, L, M)$ be the constants from Lemma 6.1. Choose

$$
\vartheta := \min \left\{ \vartheta_*, \left( \frac{1}{3C_*} \right)^{\frac{1}{\min\{p, p\}}}, \left( \frac{1}{2M} \right)^{\frac{4-p}{2(1-\nu)}}, \left( \frac{1}{2M} \right)^{\frac{4+\nu}{2(1-\nu)}} \right\}
$$

and observe that $\vartheta$ depends on $N, n, p, \nu, L, M$ and $\alpha$. We also note that the definition of $\vartheta$ implies that $\vartheta \leq (2M)^{-1}$.

Subsequently fix the constant $\varepsilon_1 (N, n, p, \nu, L, \vartheta, \mu(\cdot)) \in (0,1]$ from Lemma 6.1. This implies that $\varepsilon_1 = \varepsilon_1 (N, n, p, \nu, L, M, \alpha, \mu(\cdot))$. Now we set

$$
\varepsilon_2 := \frac{\varepsilon_1}{3},
$$

such that $\varepsilon_2 = \varepsilon_2 (N, n, p, \nu, L, M, \alpha, \mu(\cdot))$, and choose $\varrho_* \in (0,1]$ sufficiently small such that

$$
\begin{aligned}
&\omega \left( (2M \varrho^*)^2 \right) \leq \varepsilon_2, \\
&V(\varrho^2) \leq \varepsilon_2, \\
&\varrho_* \leq \varrho_0.
\end{aligned}
$$

This means that $\varrho_* = \varrho_*(N, n, p, \nu, L, M, \alpha, \varrho_0, \omega(\cdot), \mu(\cdot), V(\cdot))$, where $\varrho_0$ is the constant from the structure assumption (1.5).

Step 2: Iteration of the excess decay estimate Suppose that for some $z_0 \in \Omega_T$ and some radius $\varrho \in (0, \varrho_*]$, such that $Q_{\varrho, \varrho}(z_0) \subseteq \Omega_T$, there holds

$$
\begin{aligned}
&1 + |D\ell_{z_0, \varrho}| \leq M, \\
&\Psi(z_0, \varrho, \ell_{z_0, \varrho}) \leq \varepsilon_2.
\end{aligned}
$$

We will prove by induction that there exists a sequence of numbers $(\lambda_j)_{j \in \mathbb{N}_0}$ such that for all $j \in \mathbb{N}_0$ we have

$$
\begin{aligned}
&1 \leq \lambda_j \leq (2M)^j, \\
&\lambda_j \leq 1 + |D\ell_{z_0, \varrho}(\lambda_j)| \leq M \lambda_j, \\
&\Psi_{\lambda_j} (z_0, \varrho, \ell_{z_0, \varrho}) \leq \varepsilon_2.
\end{aligned}
$$

For $j = 0$ we set $\lambda_0 := 1$ and immediately spot that (6.13) is satisfied due to (6.12). Now assume that (6.13) holds for some $j \in \mathbb{N}_0$. In order to prove that (6.13) also holds for $j + 1$ we will apply Lemma 6.1 on the scaled intrinsic cylinder $Q(\lambda_j, \varrho, \varrho)$. But first we have to verify that all the prerequisites from Lemma 6.1 are fulfilled. First of all we note that $Q(\lambda_j, \varrho, \varrho) \subseteq Q_{\varrho, \varrho}(z_0) \subseteq \Omega_T$. In the case $p \geq 2$ this follows directly from the fact that $\lambda_j \geq 1$. In the case $p < 2$ this follows from the following chain of inequalities (note that $\vartheta_*(p) = \frac{1}{4}(2M)^{\frac{p-2}{2}}$ for $p < 2$):

$$
(\vartheta^2)^{2-p} \leq \vartheta^{2j} \varrho^2 (2M)^j (2-p) \leq \left( \frac{1}{4} \right)^{2j} (2M)^{j(p-2)} \varrho^2 (2M)^j (2-p) \leq \varrho^2,
$$

where we used (6.13) and (6.11). The intrinsic coupling $\lambda_j \leq 1 + |D\ell_{z_0, \varrho}(\lambda_j)| \leq M \lambda_j$ from (3.3) is also satisfied since we presumed that (6.13) holds for $j$. It remains to prove that the smallness condition

$$
\Psi_{\lambda_j} (z_0, \varrho, \ell_{z_0, \varrho}) \leq \varepsilon_1
$$

(6.14)
from (6.1) is also satisfied. Our assumption that (6.13) holds for \( j \) immediately yields \( \Psi_{\lambda_j}(z_0, \partial^j \varrho, \ell_j(z_0, \partial^j \varrho, \varrho)) \leq \varepsilon_2 \). On the other hand the choice of \( \varrho_* \) implies
\[
V(\lambda_j^{2-p}(\partial^j \varrho)^2) \leq V((2M)^{j(2-p)}(\partial^j \varrho)^2) \leq V(4^{-2j}\varrho^2) \leq V(\varrho^2) \leq \varepsilon_2,
\]
if \( \frac{2n}{n+2} < p < 2 \). In the complementary case \( p \geq 2 \) we also get
\[
V((\partial^j \varrho)^2) \leq V(\varrho^2) \leq \varepsilon_2.
\]
Treating the remaining term from the excess functional \( \Psi^*_{\lambda_j}(z_0, \partial^j \varrho, \ell_j(z_0, \partial^j \varrho, \varrho)) \) is a bit tricky. We first have to prove the following inequality:
\[
\int_{Q_{\ell_j}(z_0)} |u - \ell_j(x_0)|^2 \, dz \leq (2M)^{2(j+1)}(\partial^j \varrho)^2.
\]
(6.15)
Here we have abbreviated \( \ell_j = \ell_j(z_0, \partial^j \varrho) \). This can be seen as follows:
\[
\int_{Q_{\ell_j}(z_0)} |u - \ell_j(x_0)|^2 \, dz \\
= \int_{Q_{\ell_j}(z_0)} |u - \ell_j + D\ell_j(x - x_0)|^2 \, dz \\
\leq 2 \int_{Q_{\ell_j}(z_0)} |u - \ell_j|^2 \, dz + 2 \int_{Q_{\ell_j}(z_0)} |D\ell_j|^2 |x - x_0|^2 \, dz \\
\leq 2(\partial^j \varrho)^2(1 + |D\ell_j|)^2 \Psi_{\lambda_j}(z_0, \partial^j \varrho, \ell_j) + 2(\partial^j \varrho)^2 |D\ell_j|^2 \\
\leq 2(\partial^j \varrho)^2(1 + |D\ell_j|)^2 \varepsilon_2 + 2(\partial^j \varrho)^2 |D\ell_j|^2 \\
\leq 4(\partial^j \varrho)^2(1 + |D\ell_j|)^2 \leq 4(\partial^j \varrho)^2 M^2 \lambda_j^2 \leq 4(\partial^j \varrho)^2 M^2 (2M)^2.
\]
Here we have again used that (6.13) holds for \( j \) and also that \( \varepsilon_2 \leq 1 \). At this point the choices of \( \varrho \) and \( \varrho_* \) come into play as we can now estimate:
\[
\omega \left( \int_{Q_{\ell_j}(z_0)} |u - \ell_j(x_0)|^2 \, dz \right) \\
\leq \omega \left( (2M)^{2(j+1)}(\partial^j \varrho)^2 \right) \leq \omega (2M\varrho_*^2) \leq \omega (2M\varrho_*^2) \leq \varepsilon_2.
\]
Here we have used \( \varrho \leq (2M)^{-1} \) and the monotonicity of \( \omega(\cdot) \). Adding up all these estimates yields the desired inequality (6.14). At this point we can finally apply Lemma 6.1 which yields some \( \lambda_{j+1} \in \left[ \frac{\lambda_j}{2}, 2M\lambda_j \right] \) such that \( \lambda_{j+1} = 1 + |D\ell_{j+1}| \) and
\[
\Psi_{\lambda_{j+1}}(z_0, \partial^{j+1} \varrho, \ell_{j+1}) \leq C_* \varrho^{\min(p,2)} \Psi_{\lambda_j}(z_0, \partial^j \varrho, \ell_j) \leq 3C_* \varrho^{\min(p,2)} \varepsilon_2 \leq \varepsilon_2
\]
holds, where we have used the definition of \( \varrho \) again. Moreover we have \( \lambda_{j+1} = 1 + |D\ell_{j+1}| \geq 1 \) and on the other hand \( \lambda_{j+1} \leq 2M\lambda_j \leq 2M(2M)^j = (2M)^{j+1} \). This concludes the proof of (6.13) for \( j + 1 \). By induction it follows that (6.13) and (6.15) hold for every \( j \in \mathbb{N}_0 \).

**Step 3: Obtaining a Campanato type estimate**
We still assume that for \( z_0 \in \Omega_T \) and \( \varrho \in (0, \varrho_*] \) the smallness conditions (6.12) are met. We have just shown that in this case (6.15) holds for any \( j \in \mathbb{N}_0 \). From this we will derive the following Campanato type estimate:
\[
\int_{Q_r(z_0)} |u - (u)_{z_0,r}|^2 \, dz \leq C r^{n+2+2\alpha} \quad \forall r \in (0, \varrho]
\]
(6.16)
with some constant $C = C(N, n, p, \nu, L, M, \alpha)$. To this end, we first multiply (6.15) with $|Q_{\vartheta^j \varrho}(z_0)|$ to obtain
\[
\int_{Q_{\vartheta^j \varrho}(z_0)} |u - (u)_{z_0, \vartheta^j \varrho}|^2 \, dz \\
\leq (2M)^{2(j+1)}(\vartheta^j \varrho)^2|Q_{\vartheta^j \varrho}(z_0)| = (2M)^{2(j+1)}(\vartheta^j \varrho)^{n+4}\lambda_j^{2-p}\alpha_n,
\]
where $\alpha_n$ denotes the Lebesgue measure of the unit ball in $\mathbb{R}^n$. Let us now define
\[
\theta := \begin{cases} \vartheta, & p < 2 \\ (2M)^{\frac{2-p}{2}} \varrho, & p \geq 2. \end{cases}
\]
Since $1 \leq \lambda_j \leq (2M)^j$ holds due to (6.13), it is easy to verify that $Q_{\vartheta^j \varrho}(z_0) \subseteq Q_{\vartheta^j \varrho}(z_0)$. Using this and the minimizing property of $(u)_{z_0, \vartheta^j \varrho}$ we deduce that
\[
\int_{Q_{\vartheta^j \varrho}(z_0)} |u - (u)_{z_0, \vartheta^j \varrho}|^2 \, dz \leq \int_{Q_{\vartheta^j \varrho}(z_0)} |u - (u)_{z_0, \vartheta^j \varrho}|^2 \, dz \\
\leq \int_{Q_{\vartheta^j \varrho}(z_0)} |u - (u)_{z_0, \vartheta^j \varrho}|^2 \, dz \\
\leq (2M)^{2(j+1)}(\vartheta^j \varrho)^{n+4}\lambda_j^{2-p}\alpha_n.
\]

The case $\frac{2n}{n+2} < p < 2$ : We use the fact that $\lambda_j \leq (2M)^j$ to deduce that
\[
\int_{Q_{\vartheta^j \varrho}(z_0)} |u - (u)_{z_0, \vartheta^j \varrho}|^2 \, dz \\
\leq (2M)^2(2M)^j(\vartheta^j \varrho)^{n+4}\alpha_n = \alpha_n(2M)^2\varrho^{n+4} ((2M)^4-p\varrho^{n+2+2\alpha}\varrho^{2-2\alpha})^j \\
= \alpha_n(2M)^2\varrho^{n+4}\varrho^j(n+2+2\alpha) [(2M)^4-p\varrho^{2-2\alpha}]^j \leq \alpha_n(2M)^2\varrho^j(n+2+2\alpha)\varrho^{n+4} \\
= \alpha_n(2M)^2\varrho^j(n+2+2\alpha)\varrho^{n+4},
\]
where we used the choice of $\vartheta$ from (6.11) in the penultimate step.

The case $p \geq 2$ : We use the fact that $\lambda_j \geq 1$ to deduce that
\[
\int_{Q_{\vartheta^j \varrho}(z_0)} |u - (u)_{z_0, \vartheta^j \varrho}|^2 \, dz \\
\leq (2M)^2(2M)^j(\vartheta^j \varrho)^{n+4}\alpha_n = \alpha_n(2M)^2\varrho^{n+4} ((2M)^2\varrho^{n+2+2\alpha}\varrho^{2-2\alpha})^j \\
= \alpha_n(2M)^2\varrho^{n+4}\varrho^j(n+2+2\alpha) \left[ (2M)^{4(p-2)(n+2+2\alpha)}\varrho^{2-2\alpha} \right]^j \leq \alpha_n(2M)^2\varrho^j(n+2+2\alpha)\varrho^{n+4}
\]
where we used the choice of $\vartheta$ from (6.11) in the last step.

Now let $r \in (0, \varrho]$ be arbitrary. There exists some $j \in \mathbb{N}_0$ such that $\vartheta^j \varrho < r \leq \vartheta^{j+1} \varrho$ and hence $\vartheta^j < \frac{r}{\vartheta}$, and $\vartheta^j \varrho < r < \frac{r}{\vartheta}$. Using the minimizing property of $(u)_{z_0, r}$ then leads to the following estimate:
\[
\int_{Q_{r}(z_0)} |u - (u)_{z_0, r}|^2 \, dz \\
\leq \int_{Q_{r}(z_0)} |u - (u)_{z_0, \vartheta^j \varrho}|^2 \, dz \leq \int_{Q_{\vartheta^j \varrho}(z_0)} |u - (u)_{z_0, \vartheta^j \varrho}|^2 \, dz \leq \alpha_n(2M)^2\varrho^j(n+2+2\alpha)\varrho^{n+4} \\
\leq \alpha_n(2M)^2\varrho^{-(n+2+2\alpha)}\varrho^{n+4} \left( \frac{1}{\varrho} \right)^{n+2+2\alpha} = \alpha_n(2M)^2\varrho^{-(n+2+2\alpha)}\varrho^{2-2\alpha(n+2+2\alpha)}$
with a constant $C = C(n, M, \alpha, \theta) = C(n, p, M, \alpha, \vartheta) = C(N, n, p, \nu, L, M, \alpha)$. In the last step we used that $\varrho \leq \varrho_*$ $\leq 1$. This concludes the proof of (6.16).

Step 4: Proof of the partial Hölder continuity Let $\alpha \in (0, 1)$ be arbitrary. For given $M \geq 1$ we denote by $\varepsilon_2(M) = \varepsilon_2(N, n, p, \nu, L, M, \alpha, \mu(\cdot))$ and $\varrho_*(M) = \varrho_*(N, n, p, \nu, L, M, \alpha, \varrho_0(\cdot), \mu(\cdot), \Psi(\cdot))$ the constants from Step 1. Let $z_0 \in \Omega_T \setminus (\Sigma_1 \cup \Sigma_2)$. By the definition of the sets $\Sigma_1$ and $\Sigma_2$ and by (4.14) and (4.15) we infer the existence of $M \geq 1$ and $\varrho \in (0, \varrho_*(M)]$ such that $Q_{2\varrho}(z_0) \subseteq \Omega_T$ and

\[
\left\{ \begin{array}{l}
1 + |D\ell_{z_0, \varrho}| < M, \\
|\Psi(z_0, \varrho, \ell_{z, \varrho})| < \varepsilon_2(M).
\end{array} \right.
\]

Now, by the continuity of the mappings $z \mapsto D\ell_{z, \varrho}$ and $z \mapsto \Psi(z, \varrho, \ell_{z, \varrho})$ we see that there exists some radius $R \in (0, \varrho]$ such that there holds:

\[1 + |D\ell_{z, \varrho}| \leq M \quad \text{and} \quad \Psi(z, \varrho, \ell_{z, \varrho}) \leq \varepsilon_2(M) \quad \forall z \in Q_R(z_0).
\]

This means that the smallness assumptions (6.12) from Step 2 hold for any $z \in Q_R(z_0)$. Moreover we have $Q_\varrho(z) \subseteq Q_{2\varrho}(z_0) \subseteq \Omega_T$ for any $z \in Q_R(z_0)$. By an application of Step 3 we see that

\[
\int_{Q_R(z_0)} |u - (u)_{z, r}|^2 \, dz \leq C r^{n+2+2\alpha} \quad \forall r \in (0, \varrho], z \in Q_R(z_0)
\]

holds. Hence we can deduce that $u$ belongs to the parabolic Campanato space $L^{2,1+\frac{2\alpha}{p}}(Q_R(z_0), \mathbb{R}^N)$. By Theorem 2.10 we obtain $u \in C^{0,\alpha/2}(Q_R(z_0), \mathbb{R}^N)$. This concludes the proof of the theorem.

REFERENCES

[1] P. Baroni, Regularity in parabolic Dini continuous systems, Forum Math., 23 (2011), 1281–1322.
[2] V. Bögelein, F. Duzaar, J. Habermann and C. Scheven, Partial Hölder continuity for discontinuous elliptic problems with VMO-coefficients, Proc. Lond. Math. Soc., 103 (2011), 371–404.
[3] V. Bögelein, F. Duzaar and G. Mingione, The regularity of general parabolic systems with degenerate diffusion, Mem. Amer. Math. Soc., 221 (2013), no. 1041.
[4] V. Bögelein, M. Foss and G. Mingione, Regularity in parabolic systems with continuous coefficients, Math. Z., 270 (2012), 903–938.
[5] S. Campanato, Equazioni paraboliche del secondo ordine e spazi $L^{2,\delta}(\Omega, \delta)$, Ann. Mat. Pura Appl., 73 (1966), 55–102.
[6] S. Campanato, Differentiability of the solutions of nonlinear elliptic systems with natural growth, Ann. Mat. Pura Appl., 131 (1982), 75–106.
[7] M. Carozza, N. Fusco and G. Mingione, Partial regularity of minimizers of quasiconvex integrals with subquadratic growth, Ann. Mat. Pura Appl., 175 (1998), 141–164.
[8] G. Da Prato, Spazi $\mathcal{L}^{p,q}(\Omega, \delta)$ e loro proprietà, Ann. Mat. Pura Appl., 69 (1965), 383–392.
[9] E. De Giorgi, Frontiere orientate di misura minima, Seminario di Matematica della Scuola Normale Superiore di Pisa, 1960-61.
[10] E. De Giorgi, Un esempio di extremali discontinue per un problema variazionale di tipo ellittico, Boll. Un. Mat. Ital. (4), 1 (1968), 135–137.
[11] E. DiBenedetto and A. Friedman, Regularity of solutions of nonlinear degenerate parabolic systems, J. Reine Angew. Math., 349 (1984), 83–128.
[12] E. DiBenedetto, Degenerate Parabolic Equations, Springer-Verlag, New York, 1993.
[13] F. Duzaar and J. F. Grotowski, Optimal interior partial regularity for nonlinear elliptic systems: the method of A-harmonic approximation, Manuscripta Math., 103 (2000), 267–298.
[14] F. Duzaar and G. Mingione, Second order parabolic systems, optimal regularity, and singular sets of solutions, Ann. Inst. H. Poincaré Anal. Non Linéaire, 22 (2005), 705–751.
[15] F. Duzaar and G. Mingione, Harmonic type approximation lemmas, *J. Math. Anal. Appl.*, 352 (2009), 301–335.
[16] F. Duzaar, G. Mingione and K. Steffen, Parabolic systems with polynomial growth and regularity, *Mem. Amer. Math. Soc.*, 214 (2011), no. 1005.
[17] F. Duzaar and K. Steffen, Optimal interior and boundary regularity for almost minimizers to elliptic variational integrals, *J. Reine Angew. Math.*, 546 (2002), 73–138.
[18] M. Foss and J. Geisbauer, Partial regularity for subquadratic parabolic systems with continuous coefficients, *Manuscripta Math.*, 139 (2012), 1–47.
[19] M. Foss and G. Mingione, Partial continuity for elliptic problems, *Ann. Inst. H. Poincaré Anal. Non Linéaire*, 25 (2008), 471–503.
[20] E. Giusti, *Direct Methods in the Calculus of Variations*, World Scientific, Singapore, 2003.
[21] M. Kronz, Partial regularity results for minimizers of quasiconvex functionals of higher order, *Ann. Inst. H. Poincaré Anal. Non Linéaire*, 19 (2002), 81–112.
[22] O. A. Ladyzhenskaya, V. A. Solonnikov and N. N. Ural’tseva, *Linear and quasilinear equations of parabolic type*, Translations of Mathematical Monographs Vol. 23, American Mathematical Society, Providence R.I., 1968.
[23] G. Mingione, Regularity of minima: an invitation to the dark side of the calculus of variations, *Appl. Math.*, 51 (2006), 355–425.
[24] J. Nash, Continuity of solutions of parabolic and elliptic equations, *Amer. J. Math.*, 80 (1958), 931–954.
[25] C. Scheven, Partial regularity for subquadratic parabolic systems by A-caloric approximation, *Rev. Mat. Iberoam.*, 27 (2011), 751–801.
[26] J. Simon, Compact sets in the space $L^p(0,T;B)$, *Ann. Mat. Pura Appl.*, 146 (1987), 65–96.
[27] J. Stará and O. John, Some (new) counterexamples of parabolic systems, *Comment. Math. Univ. Carolin.*, 36 (1995), 503–510.

Received October 2020; revised January 2021.

E-mail address: leon.mons@fau.de