# Pointwise differentiability properties of solutions of quasilinear parabolic equations

Paweł STRZELECKI*
(Received December 4, 1991, Revised May 25, 1992)

## Abstract

We prove that weak solutions of the quasilinear equation

$$u_t - \operatorname{div}\mathscr{A}(t, x, u, u_x) = \mathscr{B}(t, x, u, u_x)$$

are at almost every point Lipschitz continuous in the space directions and Hölder continuous with exponent $1/2$ in the time direction. This allows us to conclude that those weak solutions for which $u_t \in L^{\infty,1}_{loc}$ are totally differentiable almost everywhere. In the course of establishing these two theorems we generalize a result of Yu. G. Reshetnyak concerning the so-called $W^{m,p}$-differentiability of Sobolev functions

## 1. Introduction

In this paper, we study the pointwise differentiability (in the classical sense) of weak solutions of a class of quasilinear partial differential equations of parabolic type.

Let $X = (t, x_1, \ldots, x_n)$, $Y = (s, y_1, \ldots, y_n)$, $Z = (r, z_1, \ldots, z_n)$ denote points in $(n+1)$-dimensional Euclidean space $\boldsymbol{R}^{n+1}$. Let $\Omega$ be a bounded open domain in $\boldsymbol{R}^n$ and $G = (0, T) \times \Omega$. For $X \in G$, we consider the second order quasilinear equation

$$u_t - \operatorname{div}\mathscr{A}(t, x, u, u_x) = \mathscr{B}(t, x, u, u_x), \tag{1}$$

where $\mathscr{A} = (\mathscr{A}_1, \ldots, \mathscr{A}_n)$ is a $\boldsymbol{R}^n$-valued function of $(t, x, u, u_x) \in G \times \boldsymbol{R} \times \boldsymbol{R}^n$, $\mathscr{B}$ is a real-valued function of the same variables, $u_x = \left(\dfrac{\partial u}{\partial x_1}, \ldots, \dfrac{\partial u}{\partial x_n}\right)$ denotes the spatial gradient of the function $u$, and $\operatorname{div}\mathscr{A}$ stands for the divergence of the vector $\mathscr{A}(t, x, u(t, x), u_x(t, x))$ with respect to the space variables $x_1, \ldots, x_n$. Moreover, we assume that $\mathscr{A}(X, u(X), \psi(X))$ and $\mathscr{B}(X, u(X), \psi(X))$ are measurable for every choice of measurable func-

tions $u$ and $\psi$, and fulfill the following growth conditions:

$$
\begin{aligned}
|\mathscr{A}(t, x, u, \psi)| &\leq a|\psi| + b_1|u| + b_2, \\
|\mathscr{B}(t, x, u, \psi)| &\leq b_3|\psi| + b_4|u| + b_5, \\
\psi \cdot \mathscr{A}(t, x, u, \psi) &\geq a^{-1}|\psi|^2 - b_6^2 u^2 - b_7^2,
\end{aligned}
\tag{2}
$$

where $a$ is a positive constant, while the coefficients $b_1, b_2, \ldots, b_7$ are non-negative functions of $(t, x)$, each contained in some class $L^{p,q}(G)$. The positive reals $p, q$, which may possibly be different for different coefficients, are subject to the relations:

$$
p > n + 2 \quad \text{and} \quad \frac{n}{p} + \frac{2}{q} \leq 1 - \theta \quad \text{for} \quad b_1, b_2, b_3, b_6, b_7, \tag{3}
$$

$$
p > \frac{n+2}{2} \quad \text{and} \quad \frac{n}{p} + \frac{2}{q} \leq 2 - \theta \quad \text{for} \quad b_4, b_5 \tag{4}
$$

for some $\theta > 0$. We recall that a measurable function $w = w(t, x)$ is said to be of class $L^{p,q}(G)$ if

$$
\|w\|_{L^{p,q}(G)} := \int_0^T \left( \int_\Omega |w(t, x)|^p dx \right)^{q/p} dt < +\infty,
$$

with obvious modifications when $p$ or $q$ is equal to $\infty$.

Let $W^{m,p}(G)$ denote the Sobolev space of functions integrable in the $p$-th power with distributional derivatives up to order $m$ in $L^p(G)$. Then $u \in W^{1,2}(G)$ is called a weak solution of (1) if and only if the integral identity

$$
\iint_G [\varphi u_t + \varphi_x \cdot \mathscr{A}(t, x, u, u_x) - \varphi \mathscr{B}(t, x, u, u_x)] \, dt \, dx = 0 \tag{5}
$$

holds for each infinitely differentiable function $\varphi = \varphi(t, x)$ which, for every fixed $t$, has a compact support as a function of $x_1, \ldots, x_n$ and vanishes on $\Gamma = ([0, T] \times \partial\Omega) \cup (\{0\} \times \Omega)$, the parabolic boundary of the space-time cylinder $G$.

Using refined iterational techniques introduced originally by Moser [Mo1, 2, 3] Aronson and Serrin [AS] proved that weak solutions of (1) are locally Hölder continuous. Interior Hölder continuity has, of course, been proved with use of various methods for different classes of linear and nonlinear parabolic equations with measurable coefficients by a large number of other authors, like Moser [Mo 3], Nash [N], Ivanov [I], Trudinger [T], and recently DiBenedetto [DiB], just to mention a few. However, the value of Hölder exponent obtained by most these authors is usually very close to zero, so that the information about the regularity of solu-

tions provided by these theorems is certainly not of geometric character — the solutions can (at least *a priori*) behave rather wildly.

For elliptic equations with measurable coefficients, apart from the fundamental de Giorgi type theorems ascertaining Hölder continuity of weak solutions, a number of results are known about almost everywhere differentiability of weak solutions, giving additional geometric information about the regularity of such solutions. One example of such a result is provided by a theorem of Mori [M] stating the a.e. differentiability of quasiconformal mappings (in fact, the same theorem has also been proved by Bojarski in his important paper [Bo1] even without the assumption of bijectivity) — the connection with elliptic equations is via the Beltrami equation $f_{\bar z} = \mu(z)f_z$. Other theorems of this type are due to Bojarski [Bo 2] in the case of the linear equation $\mathrm{div}(a(x)u_x)=0$, and to Reshetnyak [R2] in the more general case of the quasilinear equation $\mathrm{div}\mathscr{A}(x, u, u_x)= \mathscr{B}(x, u, u_x)$ (a concise proof of the second result may be found in [HS]).

In this paper, we prove two theorems concerning the weak solutions of (1). First of them, stated below as Theorem 1, roughly speaking ascertains that all the weak solutions of (1) are, at almost every point of $G$, Lipschitz continuous in the space directions and Hölder continuous with exponent equal to 1/2 in the time direction. The second result, stated below as Theorem 2, gives a sufficient (though at first glance artificial) condition for the a.e. differentiability in the classical sence of weak solutions of (1).

THEOREM 1.    *Let* $u \in W^{1,2}(G)$ *be a weak solution of the equation* (1). *Define the parabolic distance of points* $X$, $Y \in \mathbf{R}^{n+1}$ *by the formula*

$$d_p(X, Y)=\max(|t-s|^{1/2}, \max_{1\leq i\leq n}|x_i-y_i|). \tag{6}$$

*Then, for almost every point* $X=(t, x)\in G$, *there exist positive constants* $C$ *and* $\delta$ *such that for any* $h\in(0, \delta)$

$$\underset{\{Y:d_p(X,Y)<h\}}{\mathrm{ess\,max}} |u(X)-u(Y)|\leq C\cdot h. \tag{7}$$

*Moreover, the constant* $C$ *depends only on the following parameters:*

$$C=C(n, a, \theta, \|b_1\|, \|b_2\|, \dots, \|b_7\|, u(X), u_t(X), u_x(X)),$$

*where the norms* $\|b_i\|$ *are taken in the respective* $L^{p,q}$ *spaces.*

THEOREM 2.    *Let* $u$ *be a weak solution of* (1) *such that* $u_t\in L_{loc}^{\infty,1}(G)$. *Then* $u$ *is totally differentiable almost everywhere in* $G$ *with respect to*

*Lebesgue measure.*

Both these theorems have already been proved in [S] for weak solutions of linear uniformly parabolic equation with bounded measurable coefficients:

$$u_t - \sum_{k,l=1}^{n} \frac{\partial}{\partial x_k}\left( a_{kl}(t, x) \cdot \frac{\partial u}{\partial x_l} \right) = \sum_{l=1}^{n} b_l(t, x) \cdot \frac{\partial u}{\partial x_l}.$$

However, to cover the present case, the argument used in [S] has to be modified. The backbone of our reasoning is taken from [HS]: we derive Theorem 1 from the local boundedness of weak solutions (so that it is obtained *before* we even know that weak solutions are Hölder continuous). The main argument is provided by Theorem 3 (see Section 2 for exact formulation and proof) which generalizes a theorem of Reshetnyak [R1] concerning the so-called differentiability in the sense of $W^{m,p}$ of Sobolev functions. To get an idea of how Theorem 3 intervenes, note that the class of differential equations (1) is invariant under the group of transformations

$$X = (t, x) \longmapsto P_h(t, x) + (r, z),   \tag{8}$$

where $P_h : \boldsymbol{R}^{n+1} \longrightarrow \boldsymbol{R}^{n+1}$ is the linear mapping given by

$$P_h X = P_h(t, x) = (h^2 t, hx).   \tag{9}$$

Therefore, the differentiability properties of solutions of (1) crucially rely on the way and sense in which the function $v_h(Y) = u(X + P_h Y)$ can be approximated by the Taylor polynomials of $u$,

$$T^{(m)}(V) \equiv T_{X,u}^{(m)}(V) = \sum_{0 \le |\alpha| \le m} \frac{D^\alpha u(X)}{\alpha !} \cdot V^\alpha,   \tag{10}$$

with $P_h Y$ substituted for $V$. Theorem 3 states that if $u \in W^{m,p}(G)$, then for almost every $X \in G$ and for $h$ tending to zero, $v_h(Y) - T_{X,u}^{(m)}(P_h Y)$ (treated as a function of $Y$ with $Y$ running over some fixed cube centered at 0) tends to zero in the $W^{m,p}$-norm faster than $|h|^m$. In Section 3, we apply this result with $m = 1$, $p = 2$ to establish Theorem 1, and finally we derive Theorem 2 from Theorem 1.

In the following, we shall use the standard multiindex notation:

$$\alpha = (\alpha_0, \alpha_1, \cdots, \alpha_n) \in \boldsymbol{Z}_+^{n+1},$$
$$|\alpha| = \alpha_0 + \alpha_1 + \cdots + \alpha_n,$$
$$\alpha ! = \alpha_0 ! \, \alpha_1 ! \cdots \alpha_n !$$

Moreover, we write $D_i$ for the partial derivative with respect to $x_i$, $i =$

$1, \ldots, n$, and $D_0$ for the partial derivative with respect to $t$, and let

$$D^\alpha = (D_0)^{\alpha_0}(D_1)^{\alpha_1} \cdots (D_n)^{\alpha_n}.$$

The barred integral $\fint_A f(x)d\mu(x)$ denotes the mean value of a measurable

function $f$ on the set $A$, i.e. $\fint_A f(x)d\mu(x) = (\mu(A))^{-1}\int_A f(x)d\mu(x)$. By

$R(X, h)$ we denote the rectangle with center at $X$ and edges parallel to the coordinate axes in $\boldsymbol{R}^{n+1}$ such that the edge parallel to $t$ is of length $2h^2$ and those parallel to $x_1, \ldots, x_n$ are of length $2h$, that is,

$$
\begin{aligned}
R(X, h) &= \{Y \in \boldsymbol{R}^{n+1} : d_p(Y, Y) < h\} \\
&= \{(s, y) : |t - s| < h^2, \max_{1 \leq i \leq n}|x_i - y_i| < h\}.
\end{aligned}
$$

Rectangles of this kind shall sometimes be referred to as *parabolic*. We abbreviate $R(0, h)$ simply as $R(h)$ when $0$ is the origin of $\boldsymbol{R}^{n+1}$.

## 2. Parabolic Taylor approximations

In this section, we prove the following

THEOREM 3.   *Let* $u \in W^{m,p}(\boldsymbol{R}^{n+1})$. *Then, for almost every point* $X \in \boldsymbol{R}^{n+1}$, *the function*

$$w_{h,X}(Y) := |h|^{-m}(u(X + P_h Y) - T_{X,u}^{(m)}(P_h Y)) \quad (Y \in R(1)) \tag{11}$$

*tends to zero in the space* $W^{m,p}(R(1))$ *as* $h \to 0$.

REMARK.   This theorem generalizes in a natural way the afore-mentioned result of Reshetnyak [R1] (which reads almost exactly as the statement above, the only relevant difference being the definition of the transformation group: in [R1] $P_h x = hx$ for $x \in \boldsymbol{R}^n$). The idea of our proof resembles those of Reshetnyak [R1] and Ziemer [Z, Theorem 3. 4. 2], but some of the details are considerably more delicate; this is mainly due to the fact that in our case $\dfrac{d^2}{dh^2}P_h X \neq 0$. In order to formulate his theorem Reshetnyak has introduced the notion of differentiability in the sense of $W^{m,p}$; this suggests that Theorem 3 could be perhaps referred to as the theorem on parabolic Taylor approximations.

The proof will be split in two steps: at first, we prove a Taylor-like formula for $u(X + P_h Y)$ (differentiating $u(X + P_h Y)$ $m$ times with respect to the parameter $h$) and next we use this formula to prove the theorem. To clarify the reasoning, we divide it into several lemmas. We write $B_h$

$= \dfrac{d}{dh}P_h$, with the convention that $(B_h X)_i = x_i$ for $i = 1, \ldots, n$ and $(B_h X)_0 =$

$2hx_0.$

## 2.1 Preliminary lemmas

LEMMA 1.  *Let*  $u \in C^\infty(\mathbf{R}^{n+1})$,  *fix*  $X=(t, x)=(x_0, x)$,  $Y=(s, y)=(y_0, y)$,  *and let*  $F$  *be a function of one real variable defined by*  $F(h)=u(X+P_hY)$.  *Then*

$$\frac{d^\ell}{dh^\ell}F(h)=\phi_\ell(h, X, Y, ) \tag{12}$$

*where*

$$\phi_\ell(h, X, Y)=\sum_{k=0}^{[\ell/2]}a_{\ell,k}\sum_{|\alpha|=\ell-2k}\frac{|\alpha|!}{\alpha!}D^\alpha(D_0)^k u(X+P_hY)\cdot(B_hY)^\alpha\cdot(2y_0)^k, \tag{13}$$

*where*  $(B_hY)^\alpha=\prod_{i=0}^{n}((B_hY)_i)^{\alpha_i}$,  *and the coefficients*  $a_{\ell,k}$  *fulfill the recurrent formulae*:

$$\begin{aligned} a_{\ell+1,k} &= a_{\ell,k}+(\ell-2k+2)a_{\ell,k-1} \quad \text{for} \quad k=1, \dots, [(\ell+1)/2]; \\ a_{\ell,0} &= 1 \quad \text{for all} \quad \ell \in \mathbf{N}; \\ a_{\ell,k} &= 0 \quad \text{for} \quad k>[\ell/2], \end{aligned} \tag{14}$$

*i.e.*  $a_{\ell,k}=\ell!/((\ell-2k)!(2k)!!)$,  *where*  $(2k)!!=2\cdot4\cdot\dots\cdot2k$.

PROOF.  The proof will proceed by induction on  $\ell$.  Both (13) and (14) obviously hold for  $\ell=1$.  Let  $A(\mu)$  denote the set of all finite sequences in  $\{0, 1, \dots, n\}$  of length  $\mu$,

$$A(\mu)=\underbrace{\{0, 1, \dots, n\}\times\cdots\times\{0, 1, \dots, n\}}_{\mu \text{ times}}.$$

For  $J=(j_1, \dots, j_\mu)\in A(\mu)$, we write

$$\boldsymbol{D}_J=D_{j_1}D_{j_2}.\dots D_{j_\mu}=\frac{\partial}{\partial x_{j_1}}\circ\frac{\partial}{\partial x_{j_2}}\circ\cdots\circ\frac{\partial}{\partial x_{j_\mu}},$$

$$(B_hY)_J=\prod_{\gamma=1}^{\mu}(B_hY)_{j_\gamma}.$$

(If  $\mu=0$, then we take for  $A(0)$  any singleton, and if  $J\in A(0)$, we let  $\boldsymbol{D}_J=$  identity and  $(B_h)_J=1$.)  Summing over ordered sequences, we can rewrite the right hand side of (13) in the form

$$\sum_{k=0}^{[\ell/2]}a_{\ell,k}\sum_{J\in A(\ell-2k)}\boldsymbol{D}_J(D_0)^k u(X+P_hY)\cdot(B_hY)_J\cdot(2y_0)^k.$$

Differentiating the above formula with respect to  $h$, using the chain rule

and the Leibniz formula, and keeping in mind that $\dfrac{d}{dh}(B_h Y)_i \neq 0$ if and only if $i=0$, we conclude

$$\phi_{\ell+1}(h, X, Y)=\frac{d}{dh}\,\phi_\ell(h, X, Y)=\Sigma_1+\Sigma_2,$$

where

$$\Sigma_1=\sum_{k=0}^{[\ell/2]} a_{\ell,k} \sum_{J\in A(\ell-2k+1)} D_J(D_0)^k u(X+P_h Y)\cdot(B_h Y)_J\cdot(2y_0)^k,$$

$$\Sigma_2=\sum_{k=0}^{[\ell/2]} a_{\ell,k}(\ell-2k) \sum_{J\in A(\ell-2k-1)} D_J(D_0)^{k+1} u(X+P_h Y)\cdot(B_h Y)_J\cdot(2y_0)^{k+1}.$$

Hence, in view of $\ell-2k-1=(\ell+1)-2(k+1)$, we obtain (13) and the desired relations between the coefficients $a_{\ell,k}$. The proof is complete.

Using the Taylor formula, we can state the following obvious lemma.

LEMMA 2. *Let* $u\in C^\infty(\boldsymbol{R}^{n+1})$. *Then, with the notation of Lemma 1,*

$$u(X+P_h Y)=u(X)+\sum_{\ell=1}^{m-1} \phi_\ell(0, X, Y)\cdot\frac{h^\ell}{\ell!}+\int_0^h \phi_m(\lambda, X, Y)\frac{(h-\lambda)^{m-1}}{(m-1)!}\,d\lambda.$$

$$(15)$$

We shall need one more result concerning the *parabolic* rectangles. Though the family $\mathscr{R}$ of all parabolic rectangles $R(X, h)$ is not regular according to the classical definition, which one may find for instance in Stein's monograph [St], nevertheless it is possible to prove the following.

LEMMA 3. *If* $1\leq p<\infty$, $G$ *is an open set in* $\boldsymbol{R}^{n+1}$ *and* $f\in L^p_{loc}(G)$, *then for almost every* $X\in G$;

$$\lim_{h\to 0}\fint_{R(X,h)} |f(Y)-f(X)|^p dY=0.$$

This fact can be obtained as a corollary to rather general theorems concerning the covering properties of different families of measurable subsets of Euclidean space (the key fact here is that, given any two parabolic rectangles in $\boldsymbol{R}^{n+1}$, one can always translate one of them into another; for details see De Guzmán's book [DeG]). Related topics — but in a far more general setting — have been considered by Calderón and Torchinsky in their paper [CT]. A simple straightforward proof of Lemma 3 can be obtained as follows. Mimicking Banach's argument, one first establishes a covering lemma of Vitali type. Next one introduces the maximal function $\mathscr{M}_{\mathscr{R}}(f)$ by setting

$$\mathcal{M}_{\mathscr{R}}(f)(X) = \sup_{h>0} \int_{R(X,h)} |f(Y)| \, dY.$$

With the help of the covering lemma it can be shown in a standard way that $\mathcal{M}_{\mathscr{R}}(f)$ is of weak type 1-1, whence the desired result follows easily (the argument imitates the one used e.g. in [St], Chapter 1).

## 2.2. Proof of Theorem 3

We begin by reordering the sum which defines the Taylor polynomial $T_{X,u}^{(m)}(P_h Y)$ with respect to the increasing powers of $h$. In what follows, $\hat{a}$ denotes the multiindex with $n$ integer coordinates $a_1, \ldots, a_n$, so that

$$a = (a_0, 0, \ldots, 0) + (0, \hat{a}),$$

and $\lceil r \rceil$ denotes the *ceiling* of the real number $r$, i.e., the smallest integer greater or equal to $r$. Remembering also that $Y = (s, y) = (y_0, y) = (y_0, y_1, \ldots, y_n)$ and changing the summation order several times, we find that

$$T_{X,u}^{(m)}(P_h Y) = \sum_{0 \le |a| \le m} \frac{D^a u(X)}{a!} \cdot (P_h Y)^a$$

$$= \sum_{0 \le |a| \le m} \frac{D^{\hat{a}}(D_0)^{a_0} u(X)}{\hat{a}! a_0!} \cdot h^{|\hat{a}|+2a_0} \cdot Y^a$$

$$= \sum_{k=0}^{m} \sum_{a_0=0}^{k} \sum_{|\hat{a}|=k-a_0} \frac{D^{\hat{a}}(D_0)^{a_0} u(X)}{\hat{a}! a_0!} \cdot y^{\hat{a}} \cdot s^{a_0} \cdot h^{k+a_0}$$

$$= \sum_{k=0}^{m} \sum_{\mu=k}^{2k} \left( \sum_{|\hat{a}|=2k-\mu} \frac{D^{\hat{a}}(D_0)^{\mu-k} u(X)}{\hat{a}! (\mu-k)!} \cdot y^{\hat{a}} \cdot s^{\mu-k} \right) \cdot h^{\mu}$$

$$= \sum_{\mu=0}^{2m} \left( \sum_{k=\lceil \mu/2 \rceil}^{\min(\mu,m)} \sum_{|\hat{a}|=2k-\mu} \frac{D^{\hat{a}}(D_0)^{\mu-k} u(X)}{\hat{a}! (\mu-k)!} \cdot y^{\hat{a}} \cdot s^{\mu-k} \right) \cdot h^{\mu}.$$

Write now the outer sum $\sum_{\mu=0}^{2m}(\ldots) \cdot h^{\mu}$ in the form $\sum_{\mu=0}^{m}(\ldots) \cdot h^{\mu} + \sum_{\mu=m+1}^{2m}(\ldots) \cdot h^{\mu}$ and let $R_m(h, X, Y) := \sum_{\mu=m+1}^{2m}(\ldots) \cdot h^{\mu}$. Then, changing the summation index inside the first sum from $k$ to $i = \mu - k$ leads almost immediately[1] to

$$T_{X,u}^{(m)}(P_h Y) = \sum_{\mu=0}^{m} \frac{\phi_\mu(0, X, Y) h^{\mu}}{\mu!} + R_m(h, X, Y), \tag{16}$$

where

$$R_m(h, X, Y) = \sum_{\mu=m+1}^{2m} h^{\mu} \left( \sum_{k=\lceil \mu/2 \rceil}^{m} \sum_{|\hat{a}|=2k-\mu} \frac{D^{\hat{a}}(D_0)^{\mu-k} u(X)}{\hat{a}! (\mu-k)!} \cdot y^{\hat{a}} \cdot s^{\mu-k} \right). \tag{17}$$

Using Lemma 2 and the identity

---

[1] Observe that in order to determine $\phi_\mu(0, X, Y)$, it suffices to take the sum over those multiindices $a$ for which $a_0 = 0$ — the remaining terms vanish for $h = 0$.

$$\frac{h^m}{m}=\int_0^h (h-\lambda)^{m-1}d\lambda,$$

we see that for any $u \in W^{m,p}(\boldsymbol{R}^{n+1}) \cap C^\infty(\boldsymbol{R}^{n+1})$

$$w_{h,x}(Y)=|h|^{-m}(u(X+P_h Y)-T_{X,u}^{(m)}(P_h Y))$$

$$=\int_0^h [\phi_m(\lambda,\ X,\ Y)-\phi_m(0,X,\mathrm{Y})]\frac{(h-\lambda)^{m-1}}{|h|^m(m-1)!}\,d\lambda-\frac{R_m(h,X,Y)}{|h|^m}.$$

$$(18)$$

### 2.2.1.  Convergence of $w_{h,x}(Y)$ in the $L^p$-norm

Assume for a moment that $u \in W^{m,p}(\boldsymbol{R}^{n+1}) \cap C^\infty(\boldsymbol{R}^{n+1})$.  Let

$$\tilde{w}_{h,x}(Y)=\frac{|h|^{-m}}{(m-1)!}\int_0^h [\phi_m(\lambda,\ X,\ Y)-\phi_m(0,\ X,\ Y)]\,(h-\lambda)^{m-1}d\lambda. \qquad (19)$$

According to the footnote on the page 8,

$$\phi_m(\lambda,\ X,\ Y)-\phi_m(0,\ X,\ Y)$$

$$=\sum_{k=0}^{[m/2]} a_{m,k}\sum_{\substack{|\alpha|=m-2k\\ \alpha_0=0}}\frac{|\alpha|!}{\alpha!}(D^\alpha (D_0)^k u(X+P_h Y)-D^\alpha (D_0)^k u(X))\,Y^\alpha \cdot (2y_0)^k$$

$$+\sum_{k=0}^{[m/2]} a_{m,k}\sum_{\substack{|\alpha|=m-2k\\ \alpha_0>0}}\frac{|\alpha|!}{\alpha!}D^\alpha (D_0)^k u(X+P_\lambda Y)\cdot Y^\alpha \cdot (2y_0)^k \cdot (2\lambda)^{\alpha_0}.$$

Hence, combining the equations (17) and (18) with Jensen's inequality and Fubini's theorem, we obtain after simple calculations (involving the elementary estimate $|\sum_{i=1}^N a_i|^p \leq N^{p-1}\sum_{i=1}^N |a_i|^p$)

$$\int_{R(1)}|w_{h,x}(Y)|^p dY \leq 2^{p-1}\left(\int_{R(1)}\frac{|R_m(h,X,Y)|^p}{|h|^{pm}}\,dY+\int_{R(1)}|\tilde{w}_{h,x}(Y)|^p dY\right)$$

$$\leq C\sum_{|\beta|\leq m}\left\{\int_0^h\int_{R(1)}|D^\beta u(X+P_\lambda Y)-D^\beta u(X)|^p dY\,d\lambda \qquad (20)\right.$$

$$\left.+\int_0^h\int_{R(1)}|D^\beta u(X+P_\lambda Y)|^p dY\,d\lambda+|h\cdot D^\beta u(X)|^p\right\}$$

for $h \in (0,1)$.  We now show that (20) is valid not only for smooth $u$ but also for any $u \in W^{m,p}(\boldsymbol{R}^{n+1})$; to be more precise, if $u \in W^{m,p}(\boldsymbol{R}^{n+1})$, then, for almost all $X \in \boldsymbol{R}^{n+1}$, (20) holds for all $h \in (0,1)$.  To this end, let $u_\varepsilon = u*\psi_\varepsilon$ denote the regularizers of $u$, where $\psi_\varepsilon$ is the approximative identity of the form

$$\psi_\varepsilon(X) := \varepsilon^{-\nu}\psi(P_{1/\varepsilon}(X)),\quad \nu=n+2=\mathrm{trace}\left(\frac{d}{dh}P_h\right)\Big|_{h=1},$$

and $\psi$ is a positive function of class $C_0^\infty(R(1))$ with $\int_{\boldsymbol{R}^{n+1}}\psi dx=1$.  The fol-

lowing lemma lists a number of useful properties of $u_\varepsilon$.

**LEMMA 3.**   *Let* $u \in W^{m,p}(\boldsymbol{R}^{n+1})$. *If* $u_\varepsilon$ *is defined as above, then, for all* $\beta$ *with* $|\beta| \in [0, m]$,

*   $D^\beta u_\varepsilon(X) \to D^\beta u(X)$ *as* $\varepsilon \downarrow 0$ *for almost all* $X \in \boldsymbol{R}^{n-1}$, *in particular for those* $X$ *for which*

$$\lim_{h \to 0} \int_{R(X,h)} |D^\beta u(Y) - D^\beta u(X)|^p dY = 0, \tag{21}$$

*   $D^\beta u_\varepsilon \to D^\beta u$ *as* $\varepsilon \downarrow 0$ *in* $L^p(\boldsymbol{R}^{n+1})$.

Let $E$ be the set of those $X \in \boldsymbol{R}^{n+1}$ for which the condition (21) holds for all $\beta$ with $0 \le |\beta| \le m$. Then $|\boldsymbol{R}^{n+1} \backslash E| = 0$ and, for every $X \in E$, there is a constant $M = M(X)$ such that

$$\int_{R(X,\delta)} |D^\alpha u(Y)|^p dY \le M \tag{22}$$

for all $\alpha$ with $0 \le |\alpha| \le m$ and all $\delta > 0$. If we set $Z = X + P_h Y$, then $dZ = h^\nu \cdot dY$, and $Z \in R(X, h)$ if and only if $Y \in R(1)$. Therefore (20) leads directly to

$$\int_{R(X,h)} \frac{|u_\varepsilon(Z) - T_{X,u_\varepsilon}^{(m)}(Z - X)|^p}{|h|^{pm}} dZ$$

$$\le C \sum_{0 \le |\beta| \le m} \left( \int_0^h \int_{R(X,\lambda)} |D^\beta u_\varepsilon(Z) - D^\beta u_\varepsilon(X)|^p dZ \, d\lambda \right. \tag{23}$$

$$\left. + \int_0^h \int_{R(X,\lambda)} |D^\beta u_\varepsilon(Z)|^p dZ \, d\lambda + |h D^\beta u_\varepsilon(X)|^p \right).$$

Consider now $X \in E$, fix $h > 0$, and let $\varepsilon \to 0$. By Fatou's lemma and Lemma 4, the lower limit of the left-hand side of (23) is greater than or equal to the left-hand side of (20). The desired result will follow from Lemmas 3, 4, and the Dominated Convergence theorem of Legesgue once we prove the following

**Claim.**   There exists a positive constant $K$ such that

$$\int_{R(X,\lambda)} |D^\beta u_\varepsilon(Z)|^p dZ \le K \tag{24}$$

for any $\varepsilon > 0$, $\lambda \in [0, h]$ and $\beta$ with $|\beta| \le m$.

We now proceed to establish (24). In view of Jensen's inequality, for a measurable $A \subset \boldsymbol{R}^{n+1}$,

$$\int_A |D^\beta u_\varepsilon(Z)|^p dZ = \int_A \left| \int_{R^{n+1}} \psi_\varepsilon(Z-Y) D^\beta u(Y) dY \right|^p dZ$$

$$\leq \sup \psi \cdot (\varepsilon^{-\nu})^p \int_A \left( \int_{R(Z,\varepsilon)} |D^\beta u(Y)| dY \right)^p dZ$$

$$\leq 2^{(n+1)p} \sup \psi \cdot \varepsilon^{-\nu} \int_A \int_{R(Z,\varepsilon)} |D^\beta u(Y)|^p dY \, dZ.$$

Hence, for $A = R(X, \lambda)$,

$$\fint_{R(X,\lambda)} |D^\beta u_\varepsilon(Z)|^p dZ \leq C\varepsilon^{-\nu} \fint_{R(X,\lambda)} \int_{R(Z,\varepsilon)} |D^\beta u(Y)|^p dY \, dZ.$$

Case 1. If $\lambda \leq 3\varepsilon$, then $R(Z, \varepsilon) \subset R(X, 4\varepsilon)$ for $Z \in R(X, \lambda)$ and by (22),

$$\fint_{R(X,\lambda)} |D^\beta u_\varepsilon(Z)|^p dZ \leq C\varepsilon^{-\nu} \fint_{R(X,\lambda)} \int_{R(X,4\varepsilon)} |D^\beta u(Y)|^p dY \, dZ$$

$$= C\varepsilon^{-\nu} \int_{R(X,4\varepsilon)} |D^\beta u(Y)|^p dY$$

$$\leq \text{constant}.$$

Case 2. If $\lambda > 3\varepsilon$, then $\lambda + \varepsilon < \lambda + 3\varepsilon < 2\lambda$ and

$$\int_{R(x,\lambda)} |D^\beta u_\varepsilon(Z)|^p dZ = \int_{R(X,3\varepsilon)} |D^\beta u_\varepsilon(Z)|^p dZ + \int_{3\varepsilon < d_p(Z,X) < \lambda} |D^\beta u_\varepsilon(Z)|^p dZ$$

$$\leq C\varepsilon^{-\nu} \int_{R(X,3\varepsilon)} \int_{R(X,4\varepsilon)} |D^\beta u(Y)|^p dY \, dZ$$

$$+ C\varepsilon^{-\nu} \int_{3\varepsilon < d_p(Z,X) < \lambda} \int_{R(Z,\varepsilon)} |D^\beta u(Y)|^p dY \, dZ$$

$$= I_1 + I_2.$$

From (22) it is clear that $I_1$ does not exceed $C\lambda^\nu$. To obtain a similar estimate for $I_2$ write $Y = Z + W$, whereupon

$$Y \in R(Z, \varepsilon) \quad \text{if and only if} \quad W \in R(\varepsilon)$$

and

$$2\varepsilon < d_p(Z + W, X) < \lambda + \varepsilon,$$

and hence, by Fubini's theorem,

$$I_2 \leq C\varepsilon^{-\nu} \int_{R(\varepsilon)} \int_{2\varepsilon < d_p(Z+W,X) < \lambda + \varepsilon} |D^\beta u(Z + W)|^p dZ \, dW$$

$$\leq C\varepsilon^{-\nu} \int_{R(\varepsilon)} \int_{d_p(Y,X) < 2\lambda} |D^\beta u(Y)|^p dY \, dW$$

$$\leq \text{constant} \cdot \lambda^\nu.$$

Thus (24) is established. Consequently, we obtain (20) for $u \in W^{m,p}(R^{n+1})$ and almost all[2] $X$. Hence, by Lemma 3,

---

[2] The set to which $X$ belngs depends on $u$, of course!

$$\|w_{h,X}(\cdot)\|_{L^p(R(1))} \to 0 \quad \text{as } h \downarrow 0, \text{ for a.e. } X \in \mathbf{R}^{n+1}. \tag{25}$$

### 2.2.2. Convergence of $D_Y^\beta w_{h,X}(Y)$ with $|\beta| = m$ in the $L^p$-norm

Let $\beta$ be a fixed multiindex with $|\beta| = m$. If $u \in C^\infty(\mathbf{R}^{n+1})$, then clearly
- if $\beta_0 = 0$, then $D_Y^\beta \phi_m(0, X, Y) = m! D^\beta u(X)$,
- if $\beta_0 > 0$, then $D_Y^\beta \phi_m(0, X, Y) = 0$,

and

$$D_Y^\beta\left(\int_0^h \phi_m(\lambda, X, Y)\frac{(h-\lambda)^{m-1}}{(m-1)!} d\lambda\right) = D_Y^\beta(u(X + P_h Y))$$
$$= D^\beta u(X + P_h Y)\cdot h^{m+\beta_0}.$$

Hence, for the function $\tilde{w}_{h,X}(Y)$ defined by (19) the following equality holds

$$D_Y^\beta(\tilde{w}_{h,X}(Y)) = \begin{cases} D^\beta u(X + P_h Y) - D^\beta u(X) & \text{if } \beta_0 = 0, \\ h^{\beta_0}\cdot D^\beta u(X + P_h Y) & \text{if } \beta_0 > 0. \end{cases} \tag{26}$$

It is also clear that for small $h$

$$\left|D_Y^\beta\left(\frac{R_m(h, X, Y)}{|h|^m}\right)\right| \leq C\cdot|hD^\beta u(X)|. \tag{27}$$

Let $\sigma = 1 - \mathrm{sgn}\beta_0$ (i.e. $\sigma = 1$ if $\beta_0 = 0$ and $\sigma = 0$ if $\beta_0 > 0$). The formulae (26) and (27) lead — as in the last section — to the estimate:

$$|h|^{-mp}\int_{R(1)}|D_Y^\beta(u_\varepsilon(X + P_h Y) - T_{X,u_\varepsilon}^{(m)}(P_h Y))|^p dY$$
$$\leq C\left(\int_{R(1)}|h|^{p\beta_0}|D^\beta u_\varepsilon(X + P_h Y) - \sigma D^\beta u_\varepsilon(X)|^p dY + |h\cdot D^\beta u_\varepsilon(X)|^p\right) \tag{28}$$
$$= C\left(\fint_{R(X,h)}|h|^{p\beta_0}|D^\beta u_\varepsilon(Z) - \sigma D^\beta u_\varepsilon(X)|^p dZ + |h\cdot D^\beta u_\varepsilon(X)|^p\right).$$

Passage to the limit $\varepsilon \to 0$ (which is legitimate according to the arguments of Section 2.2.1) shows that, for any $X$ in a set $E$ of full measure in $\mathbf{R}^{n+1}$, the following inequality holds:

$$|h|^{-mp}\int_{R(1)}|D_Y^\beta(u(X + P_h Y) - T_{X,u}^{(m)}(P_h Y))|^p dY$$
$$\leq C\left(\fint_{R(X,h)}|h|^{p\beta_0}|D^\beta u(Z) - \sigma D^\beta u(X)|^p dZ + |h\cdot D^\beta u(X)|^p\right). \tag{29}$$

Thanks to Lemma 3, the right-hand side of (29) tends to zero as $h$ tends to zero. Hence, for $|\beta| = m$,

$$\|D_Y^\beta w_{h,X}(\cdot)\|_{L^p(R(1))} \to 0 \quad \text{as } h \downarrow 0 \text{ for a.e. } X \in \boldsymbol{R}^{n+1}. \tag{30}$$

Since the norm in $W^{m,p}$ given by

$$\|u\|_1 := \sum_{|\alpha|=m} \|D^\alpha u\|_{L^p} + \|u\|_{L^p}$$

is equivalent to the standard one defined as

$$\|u\| = \sum_{0 \le |\alpha| \le m} \|D^\alpha u\|_{L^p},$$

the theorem follows from (25) and (30).

REMARK. Theorem 3 holds true for functions in $W^{m,p}(G)$, where $G$ is an open subset of $\boldsymbol{R}^{n+1}$. Indeed, let $G = \bigcup_{i=1}^\infty \Omega_i$, where the $\Omega_i$ are open sets such that $\overline{\Omega_i} \subset \Omega_{i+1}$ for $i = 1, 2, \ldots$, and choose $\varphi_i \in C_0^\infty(\Omega_{i+1})$ so that $\varphi_i \equiv 1$ on some open neighbourhood of $\overline{\Omega_i}$. If $u \in W^{m,p}(G)$, then $u \cdot \varphi_i \in W^{m,p}(\boldsymbol{R}^{n+1})$ (we extend $u \cdot \varphi_i$ to the whole of $\boldsymbol{R}^{n+1}$ by putting $u \cdot \varphi_i \equiv 0$ outside the support of $\varphi_i$). Applying Theorem 3, we see that

$$|h|^{-m}(u(X + P_h Y) - T_{X,u}^{(m)}(P_h Y)) \to 0 \quad \text{as } h \downarrow 0, \text{ in } W^{m,p}(R(1))$$

for a. e. $X \in \Omega_i$. The desired conclusion follows upon noting that $G = \bigcup_{i=1}^\infty \Omega_i$.

## 3. Proofs of Theorems 1 and 2

In this section, we establish the results alluded to in the Introduction concerning the weak solutions of (1). With Theorem 3 at hand, their proofs are quite easy. In the next section, we briefly indicate possible generalizations of our theorems to other classes of equations (invariant under the action of other multiplicative groups of linear mappings) and ways of getting somewhat more precise information about the non -differentiability set might be obtained.

### 3. 1. The case $b_i \in L^p$

For the sake of simplicity, we assume throughout this subsection that $p = q$ in (3) and (4), so that the coefficients $b_i$ are in the suitable $L^p$ spaces. We will show in the next subsection how to cope with the general case.

In the proof of Theorem 1, we shall need the following result concerning the local boundedness of weak solutions.

THEOREM 4. (Aronson & Serrin). *If $u \in W^{1,2}(R(2))$ is a solution to*

(1), *then*

$$\|u\|_{L^{\infty}(R(1))} \leq C \cdot (\|u\|_{L^2(R(2))} + K), \qquad (31)$$

*where*

$$C = C(n, a, \theta, \|b_1\|, \|b_3\|, \|b_4\|, \|b_6\|),$$

$$K = \|b_2\| + \|b_5\| + \|b_7\|.$$

*Moreover, C is a non-decreasing function of the last four arguments.*

PROOF OF THEOREM 1.  Our first goal is to show that the parabolic difference quotient

$$v_h(Y) := \frac{u(X + P_h Y) - u(X)}{h}$$

satisfies a parabolic equation resembling (1), with slightly changed coefficient functions.  It is quite clear that, for $2h < d_p(X, \partial G)$, $v_h$ is a well defined function of $Y \in R(2)$ of class $W^{1,2}$.  Let $\psi$ be a function of class $C^{\infty}(R(2))$ with compact support.  Extend $\psi$ to the whole of $\mathbf{R}^{n+1}$ by putting $\psi \equiv 0$ outside $R(2)$ and let, for $(r, z) = Z = X + P_h Y \in G$,

$$\varphi(Z) = \psi(P_{1/h}(Z - X)).$$

Since $u$ is a weak solution of (1), we have

$$\iint_G [\varphi u_r + \varphi_z \cdot \mathscr{A}(r, z, u, u_z) - \varphi \mathscr{B}(r, z, u, u_z)] \, dr \, dz = 0. \qquad (32)$$

If we change the integration variables from $Z \in G$ to $Y \in R(2)$, then $dr \, dz = h^{n+2} ds \, dy$ and

$$\nabla_z u(r, z) = \nabla_z [u(X) + h v_h(P_{1/h}(Z - X))]$$
$$= \nabla_y v_h(s, y),$$
$$\nabla_z \varphi(r, z) = h^{-1} \nabla_y \psi(s, y),$$
$$\frac{\partial}{\partial r} u(r, z) = \frac{\partial}{\partial r} [u(X) + h v_h(P_{1/h}(Z - X))]$$
$$= h^{-1} \frac{\partial}{\partial s} v_h(s, y).$$

Thus, in view of (32), we see that $v_h = v_h(s, y)$ is a weak solution of the parabolic equation

$$\frac{\partial v}{\partial s} - \mathrm{div}_y \mathscr{A}_h(s, y, v, v_y) = \mathscr{B}_h(s, y, v, v_y),$$

where $Y = (s, y) \in R(2)$ and

$$\mathscr{A}_h : R(2) \times \mathbf{R} \times \mathbf{R}^n \to \mathbf{R}^n,$$
$$\mathscr{B}_h : R(2) \times \mathbf{R} \times \mathbf{R}^n \to \mathbf{R}$$

are defined by

$$\mathscr{A}_h = \mathscr{A}_h(Y, v, W) = \mathscr{A}(X + P_h Y, u(X) + hv, W),$$
$$\mathscr{B}_h = \mathscr{B}_h(Y, v, W) = h \mathscr{B}(X + P_h Y, u(X) + hv, W).$$

Theorem 4 implies that

$$\operatorname*{ess\,max}_{Y \in R(1)} \frac{u(X + P_h Y) - u(X)}{h} \le C_h \cdot (\|v_h\|_{L^2(R(2))} + K_h). \tag{33}$$

Notice that by changing $u$ on a set of null measure we can replace the *essential maximum* by the *maximum* in (33). In fact, it suffices to let

$$u(X) := \limsup_{h \to 0} \int_{R(X,h)} u(Y) dY.$$

To complete the proof, we now only have to show that the right hand side of (33) remains bounded when $h$ tends to zero.

STEP 1. Using the properties of $\mathscr{A}$ and $\mathscr{B}$ one can easily check that $\mathscr{A}_h$, $\mathscr{B}_h$ fulfill the growth condition (2) with the constant $a$ unchanged the $b_i$ being replaced by $b_{i,h}$, $i = 1, \ldots, 7$ :

$$b_{1,h}(Y) = |h| \cdot b_1(X + P_h Y),$$
$$b_{2,h}(Y) = |u(X)| \cdot b_1(X + P_h Y) + b_2(X + P_h Y),$$
$$b_{3,h}(Y) = |h| \cdot b_3(X + P_h Y),$$
$$b_{4,h}(Y) = |h|^2 \cdot b_4(X + P_h Y),$$
$$b_{5,h}(Y) = |h|(|u(X)| \cdot b_4(X + P_h Y) + b_5(X + P_h Y)),$$
$$b_{6,h}(Y) = 2|h| \cdot b_6(X + P_h Y),$$
$$b_{7,h}(Y) = 2|u(X)| \cdot b_6(X + P_h Y) + b_7(X + P_h Y).$$

Lemma 3 implies now that for almost every $X$ the norms of $b_i$, $1 = 1, \ldots, 7$, in the respective $L^p(R(2))$ spaces (for each $b_i$ the exponent $p$ being taken accordingly to (3) or (4)) are bounded as $h$ tends to zero. For instance,

$$\|b_{1,h}\|_p = \left( \int_{R(2)} [b_{1,h}(Y)]^p dY \right)^{1/p}$$
$$\le C \cdot \left( \int_{R(X,2h)} [b_1(Z)]^p dZ \right)^{1/p}$$
$$\to C \cdot b_1(X) \quad \text{for } h \downarrow 0,$$

and obviously the remaining cases can be treated in the same way. Hence, the constants $C_h$ and $K_h$ on the right hand side of (33) remain bounded (by a constant independent of $h$) for sufficiently small positive

$|h|.$

STEP 2.    Theorem 3, applied with $m=1$ and $p=2$, readily implies that for $h \to 0$ the function of $Y=(s, y)$ given by

$$\frac{1}{h}(u(X+P_h Y)-T_{X,u}^{(1)}(P_h Y))=$$

$$=v_h(s, y)-\sum_{i=1}^{n}\frac{\partial u}{\partial x_i}(X)\cdot y_i-h\frac{\partial u}{\partial t}(X)\cdot s$$

tends to zero in $W^{1,2}(R(2))$ as $h \to 0$. Hence, the $L^2$-norm of $v_h$ is bounded for small $|h|$.

Now the combination of both steps implies the desired result.

According to our remark about the possibility of replacing the *essential maximum* by the *maximum* in (33), Theorem 1 can be restated in the form of the following

COROLLARY 1.    *Let $u$ be a weak solution of* (1). *Then, for almost every $X \in G$, there exist a constant $C$ and a positive $\delta$ such that*

$$|u(Y)-u(X)| \leq C \cdot d_p(X, Y)$$

*for $d_p(X, Y) < \delta$. In other words, at almost every point in G, $u$ is Lipschitz continuous with respect to the parabolic distance $d_p$ (or, to be more precise, there exists a representative of $u \in W^{1,2}$, which is, at almost every point of G, Lipschitz continuous with respect to $d_p$).*

Recall the classical

THEOREM 5 (Stepanoff differentiability criterion).    *Let $f : G \to R$ be an arbitrary function defined on an open set $G \subset R^m$. Define*

$$E=\left\{a \in G : \limsup_{x \to a}\frac{|f(x)-f(a)|}{|x-a)} < +\infty\right\},$$

*then $E$ is Lebesque measurable and $f$ is differentiable a.e. in E.*

Combining this theorem (see [Sff] or [F], Theorem 3.1.9) with Theorem 1 and the local Hölder continuity of weak solution of (1) implies Theorem 2.

PROOF[3] OF THEOREM 2.    Let $A=(a_0, a_1, \ldots, a_n) \in G$ be a point at which $u$ is Lipschitz continuous with respect to $d_p$, so that (7) holds with $X$ replaced by $A$. Fix a positive, sufficiently small number $\delta$ and take $K$

---

[3]The proof is exactly the same as in the case of linear uniformly parabolic equation with bounded coefficients [S]. To render our exposition self-contained we repeat it here.

$=(k_0, k_1, \ldots, k_n)$ with $|K|=\delta$. Choose also an additional point $B$ with time coordinate equal to that of $A$ and with spatial coordinates equal to those of $X=(t, x)=A+K$, $B=(a_0, x_1, \ldots, x_n)$. Let $Q(x, h)$ denote a standard cube in $\boldsymbol{R}^n$, $Q(x, h)=\{y : \max_{1 \leq i \leq n}|x_i - y_i| < h\}$. By the triangle inequality,

$$|u(A)-u(X)| \leq$$

$$\leq |u(A)-u(B)|+\left|u(B)-\fint_{Q(x,\rho)}u(a_0, y)dy\right| \tag{34}$$

$$+\left|\fint_{Q(x,\rho)}[u(a_0, y)-u(t, y)]dy\right|+\left|\fint_{Q(x,\rho)}u(t, y)dy-u(X)\right|$$

$$=S_1+S_2+S_3+S_4.$$

To establish the theorem, it suffices to prove that each of the terms $S_i$ on the right-hand side of the above inequality is bounded by a constant times $\delta$. For $S_1$ this is true by virtue of Theorem 1. To estimate $S_2$ and $S_4$, we use the local Hölder continuity of weak solutions of (1) ([AS], Theorem 4, p. 110). Namely, in (34) we perform the integrations over an $n$-dimensional cube $Q(x, \rho)$ contained in the space $\boldsymbol{R}^n$ of $x$ variables. Select $\rho = \delta^\gamma = |K|^\gamma$, where $\gamma$ is the reciprocal of the Hölder exponent $\alpha$ of the function $u$, i.e., $\gamma = 1/\alpha$. Then, for $y \in Q(x, \rho)$, one has

$$|u(B)-u(a_0, y)| \leq C \cdot (|K|^\gamma)^\alpha = C \cdot \delta,$$

and in a similar manner

$$|u(X)|-u(t, y)| \leq C \cdot \delta,$$

whence the desired estimates of $S_2$ and $S_4$ follow.

Finally, to estimate $S_3$, we exploit the assumption $u_t \in L^{\infty,1}_{loc}$. If $I$ denotes the interval $(a_0, t)$, then $|I| \leq \delta$ and

$$\left|\fint_{Q(x,\rho)}[u(a_0, y)-u(t, y)]\, dy\right| \leq |K| \cdot \fint_{Q(x,\rho)}\fint_I|u_t(s, y)|ds\, dy$$

$$\leq |K| \cdot \fint_I M(s)ds,$$

where

$$M(s)=\operatorname*{ess\,sup}_{y}|u_t(s, y)|$$

is a locally integrable function of one real variable $s$. The classical differentiation theorem of Lebesgue implies now that for $\boldsymbol{R}^{n+1} \ni K \to 0$, i.e., for $|I| \to 0$, we have

$$\fint_I M(s)ds \to M(a_0) \quad \text{for a.e. } a_0.$$

Putting all these estimates together, we conclude that, for sufficiently small $|K|$,

$$|u(A+K)-u(A)| \le C \cdot |K| \quad \text{for a. e. } A, \tag{35}$$

where the constant $C$ depends on $n$, the structure of (1), the values of $u$ and its distributional derivatives at $A$, the local Hölder norm of $u$, and $M(a_0)$. This allows us to apply Stepanoff's criterion and obtain the a. e. differentiability of $u$ in the classical sense. The proof is complete.

## 3. 2  The general case

We now indicate how the proof of Theorem 1 has to be modified when the functions $b_i$ are in the $L^{p,q}$ spaces.

LEMMA 5 (Minkowski's inequality for integrals).  *Let $X$ and $Y$ be measurable spaces with $\sigma$-finite measures $\mu$ and $\nu$, respectively. Then, for $r > 1$ and for any nonnegative real-valued function $w = w(x, y)$ which is measurable on the product space $(X \otimes Y, \mu \otimes \nu)$,*

$$\left( \int_Y \left( \int_X w(x, y) d\mu_x \right)^r d\nu_y \right)^{1/r} \le \int_X \left( \int_Y w(x, y)^r d\nu_y \right)^{1/r} d\mu_x.$$

PROOF.  The proof is based on a simple argument using the duality between $L^r$ and $L^s$ for $\frac{1}{r} + \frac{1}{s} = 1$. Let $\Phi(y) = \int_X w(x, y) d\mu_x$ and define, for $g \in L^s(Y)$, the mapping

$$L^s(Y) \ni g \mapsto \Psi(g) = \int_Y g(y) \cdot \Phi(y) d\nu_y \in \mathbf{R}.$$

With the help of the Fubini theorem and the Hölder inequality, one finds that $\Psi$ is a continuous linear functional on $L^s$ and that $\|\Psi\|_{(L^s)*} = \|\Phi\|_{L^r}$ — which is equal to the expression on the left-hand side of Minkowski's inequality — does not exceed the integral standing on the right hand side.

LEMMA 6.  *Let $u$ be a positive function of class $L^{p,q}(G)$, where $p, q \in [1, \infty]$ fulfill one of the following conditions:*

- $p > \dfrac{n+2}{2}$ *and* $\dfrac{n}{p} + \dfrac{2}{q} < 2$,

- $p > n+2$ *and* $\dfrac{n}{p} + \dfrac{2}{q} < 1$.

*It is then possible to find $\bar{p}, \bar{q} \ge 1$ such that:*

- $\tilde{p} \leq p$, $\tilde{q} \leq q$ and $u \in L^{\tilde{p}, \tilde{q}}(G)$,
- $\tilde{p}$ and $\tilde{q}$ *fulfill the same inequalities as $p$ and $q$, respectively,*
- *for almost every $X = (t, x) \in G = (0, T) \times \Omega$ the following holds*:

$$\limsup_{h \to 0} \int_{t-h^2}^{t+h^2} \left( \int_{Q(x,h)} u(s, y)^{\tilde{p}} dy \right)^{\tilde{q}/\tilde{p}} ds < +\infty.$$

In the proof of this fact we shall need some well known results concerning the Hardy–Littlewood maximal functions. For $f \in L^p(\Omega)$ define

$$\mathcal{M}_p f(x) = \sup_{Q(x,h) \subset \Omega} \left( \int_{Q(x,h)} |f(y)|^p dy \right)^{1/p}.$$

From the covering lemma of Vitali, it follows that $\mathcal{M}_p$ is of weak type $p$-$p$. More precisely, for every positive $t$ and every $\sigma \in [0, 1]$,

$$|\{x \in \Omega : \mathcal{M}_p f(x) > t\}| \leq \frac{5^n}{(1-\sigma)^p t^p} \cdot \int_{\{y : |f| \geq \sigma t\}} |f(y)|^p dy. \tag{36}$$

Let $\lambda_g(t) := |\{x : |g(x)| > t\}|$ denote the distribution function of $g$. Then the familiar formula (*the Bonaventura Cavalieri principle*)

$$\int_{\Omega} |g(x)|^r dx = r \int_0^{+\infty} t^{r-1} \lambda_g(t) dt$$

combined with the weak type property of $\mathcal{M}_p$ implies that for $f \in L^r(\Omega)$, where $r > p$, $\mathcal{M}_p f$ also belongs to $L^r(\Omega)$ and

$$\|\mathcal{M}_p f\|_{L^r(\Omega)} \leq 2 \left( \frac{5^n r}{r-p} \right)^{1/r} \cdot \|f\|_{L^r(\Omega)}. \tag{37}$$

REMARK. In the inequalities (36) and (37), $5^n$ can be replaced by $3^n$, but it is, of course, irrelevant here.

PROOF OF LEMMA 6. For the sake of simplicity assume that $p > n+2$ and $\frac{n}{p} + \frac{2}{q} < 1$ (in the other case the proof is literally the same). If $q \geq p$, then take $\tilde{q} = p = \tilde{p}$. It is obvious that $\tilde{p}$ and $\tilde{q}$ fulfill the desired inequalities. The last condition follows directly from Lemma 3.

The case $p = \infty$ present no difficulties.

Suppose now that $\infty > p > q$. Take $\tilde{q} = q$ and $\tilde{p} \in (\max(\tilde{q}, n+2), p)$ such that $\frac{n}{\tilde{p}} + \frac{2}{\tilde{q}} < 1$ (it is of course possible because all the inequalities are strict). We shall prove slightly more than required, namely

$$\limsup_{h_1, h_2 \to 0} \int_{t-h_1}^{t+h_1} \left( \int_{Q(x,h_2)} u(s, y)^{\tilde{p}} dy \right)^{\tilde{q}/\tilde{p}} ds < \infty \quad \text{for a.e. } X \in G. \tag{38}$$

It is obvious that

$$\int_{t-h_1}^{t+h_1}\left(\int_{Q(x,h_2)}u(s,y)^{\tilde{p}}dy\right)^{\tilde{q}/\tilde{p}}ds\leq\int_{t-h_1}^{t+h_1}[\mathscr{M}_{\tilde{p}}u(s,\cdot)(x)]^{\tilde{q}}ds.$$

Hence, in order to prove (38), it suffices to show that, for a. e. $x\in\Omega$, the function $\mathscr{M}_{\tilde{p}}u(s,\cdot)(x)$ — treated as a function of $s\in(0,T)$ — belongs to the space $L^{\tilde{q}}((0,T);ds)$ and next to apply the classical Lebesgue differentiation theorem. The Minkowski inequality for integrals and the estimate (37) imply in an obvious way that

$$\left[\int_{\Omega}\left(\int_0^T[\mathscr{M}_{\tilde{p}}u(s,\cdot)(x)]^{\tilde{q}}ds)^{p/\tilde{q}}dx\right]^{\tilde{q}/p}\leq\int_0^T\left(\int_{\Omega}[\mathscr{M}_{\tilde{p}}u(s,\cdot)(x)]^p dx\right)^{\tilde{q}/p}ds$$

$$\leq C\cdot\int_0^T(\|u(s,\cdot)\|_{L^p(\Omega)})^{\tilde{q}}ds$$

$$=C\cdot(\|u\|_{L^{p,\tilde{q}}(G)})^{\tilde{q}}$$

$$<+\infty,$$

whence the desired result follows. The proof of Lemma 6 is complete.

Using the above lemma, we can treat each of the functions $b_i$ as elements of $L^{\tilde{p},\tilde{q}}$ rather than $L^{\tilde{p},q}$. The proof of Theorem 1 goes almost without changes, the only exception being that (cf. STEP 2 of the proof in Section 3) in order to obtain the boundedness of the norms of $b_{i,h}$ in the respective $L^{\tilde{p},\tilde{q}}$ spaces one has to apply Lemma 6 instead of Lemma 3. This simple observation enables one to complete the proof of Theorem 1 (and hence also Theorem 2) in the general case.

## 4. Conjectures

In this section, we sketch briefly how the results of the previous sections, which are still far from being conclusive, could possibly be generalized. It is just a plan of what remains to be done.

Let $A_h$ be a group of linear dilations in $\mathbf{R}^m$ defined for positive $h$ by

$$A_h(x)=(h^{\alpha_1}x_1,h^{\alpha_2}x_2,\dots,h^{\alpha_m}x_m)\tag{39}$$

for some multiindex $\alpha\in(\mathbf{R}_+)^m$. Set

$$\rho=\min_{1\leq i\leq m}\alpha_i\quad\text{and}\quad\sigma=\max_{1\leq i\leq m}\alpha_i.$$

Then

$$h^{\rho}|x|\leq|A_h(x)|\leq h^{\sigma}|x|,$$

for any $x\in\mathbf{R}^m$ and any $h>1$, and

$$h^{\sigma}|x|\leq|A_h(x)|\leq h^{\rho}|x|,$$

for any $x \in \mathbf{R}^m$ and any $h \in (0, 1)$. One can also define a homogeneous norm $r$ associated with the group $\{A_h : h > 0\}$, $r(A_h x) = h \cdot r(x)$, and an associated (pseudo) distance $d(x, y) = r(x - y)$. The choice of $r$ is not unique. One possibility is the following norm

$$r(x) := \max_{1 \le i \le m}\left(|x_i|^{\frac{1}{\alpha_i}}\right).$$

The balls in the associated metric $d \equiv d_A$ are rectangles of the form

$$R_A(x, h) = \{y : \forall_{1 \le i \le m} |y_i - x_i| < h^{\alpha_i}\}.$$

The homogeneous norm $r$ can always be chosen to be of class $C^\infty(\mathbf{R}^m \setminus \{0\})$ (*see* [L], [CT]). It is possible to prove (see [CT]) that, for the family $\mathscr{F}$ of all the rectangles $R_A$, a covering lemma of Vitali type holds. This in turn allows to obtain in a standard way an analogue of Lemma 3.

Our proof of Theorem 3 relies much more on general properties of the group $P_h : \mathbf{R}^{n+1} \to \mathbf{R}^{n+1}$, $P_h(\mathrm{t}, x) = (h^2 t, hx)$ than on specific definitions involving expressions like $r$. Accordingly, it seems that the substantiation of the conjecture stated below is a matter of overcoming certain difficulties of merely *technical* charactor.

CONJECTURE 1.    Let $u \in W^{\ell,p}(\Omega)$, where $\Omega \subset \mathbf{R}^m$ is an open domain. Then, for almost every $x \in \Omega$, the following function of $y \in R_A(0, 1)$:

$$w_h(y) := |h|^{-\rho\ell} \cdot (u(x + A_h y) - T_{x,u}^{(\ell)}(A_h y))$$

tends to zero in $W^{\ell,p}(R_A(0, 1))$ as h tends to zero.

REMARK 1.    Assume now that a class of partial differential equations (in $m$ real variables) with measurable coefficients is invariant under the action of the group $A_h$ and that it is possible to prove a local boundedness result à la Aronson & Serrin. If Conjecture 1 were true, it would then be possible to prove that weak solutions of the class of equations in question are at almost every point of their domain of definition Lipschitz continuous with respect to the metric $d_A$, or, in other words, Hölder continuous with exponent $1/\alpha_i$ in the direction of the $x_i$-axis. An analogue of Theorem 2 would also be true.

REMARK 2.    Hajłasz [H] proved that in the case of the elliptic equation

$$\mathrm{div} \, \mathscr{A}(x, u, u_x) = \mathscr{B}(x, u, u_x)$$

the weak solutions which belong to $W^{\ell,p}$, $\ell > 1$, are in fact differentiable outside a set of suitable Riesz capacity equal to zero, if we assume that,

for the functions which appear in the growth conditions analogous to (2), the complements of sets of $p$-Lebesgue points are not only of zero Lebesgue measure but also of zero Riesz capacity. His proof uses the well known fact that for Sobolev functions the set of exceptional points (for which the limit of integral averages over the Euclidean balls does not exist) is of Riesz (and Bessel) capacity zero.

Introduce the space $L^{a,p}$ of anisotropic Bessel potentials,

$$L^{a,p}(\pmb{R}^m) := \{g : g = G_a * f \text{ for some } f \in L^p(\pmb{R}^m)\},$$

where $G_a$ is the anisotropic Bessel kernel defined by

$$\hat{G}_a(\xi) = \int_{R^m} G_a(x) e^{-2\pi i x \cdot \xi} dx = \frac{1}{(1 + r^2(\xi))^{a/2}}$$

($r$ denotes the $A_h$-homogeneous norm smooth outside zero). One can prove that $G_a \in L^1(R^m)$ is analytic in $\pmb{R}^m \backslash \{0\}$ and vanishes exponentially at $\infty$. The anisotropic Bessel capacity is defined in the following standard way (cf. [Vo1], [Vo2])

$$B_{a,p}(E) = \inf\{\|f\|_p^p : f \geq 0, \ G_a * f \geq 1 \text{ on } E\}.$$

It is expected that the following conjecture holds true.

CONJECTURE 2. For any function $f$ belonging to the space of anisotropic Bessel potentials $L^{a,p}(\pmb{R}^m)$, the set $E_f$ of those points $x \in \pmb{R}^m$ for which

$$\limsup_{\delta \to 0} \int_{\{y \,:\, r(x-y) < \delta\}} |f(y) - f(x)|^p dy \neq 0$$

is of zero $B_{a,p}$-capacity, i. e. $B_{a,p}(E_f) = 0$.

This conjecture (if valid) combined with theorems describing the mutual embedding relations of classical Sobolev spaces and anisotropic potential spaces (see the papers [L], [DT]) would allow one to obtain in Theorem 1 additional information about the set of points at which the solution is not Lipschitz continuous with respect to the parabolic distance $d_p$.

To be more precise, let $1 < \ell < \frac{n+1}{2}$, and assume that a weak solution $u$ of equation (1) fulfills

$$u \in W^{\ell,2} \simeq H^\ell = \{w \in L^2(\pmb{R}^{n+1}) : \int (1 + |\xi|^2)^{\ell/2} |\hat{w}| \xi|^2 d\xi < +\infty\} \qquad (40)$$

(the statement of Theorem 1 is purely local so that there is no less of generality in assuming that $u$ can be extended to the whole of $\boldsymbol{R}^{n+1}$). If $a = \dfrac{(\ell-1)(n+1)}{n+2}$, then (see [DT], Theorem 10)

$$\operatorname{grad}\ u \in W^{\ell-1,2} \subset L^{a,2},$$

where $L^{a,2}$ denotes the space of anisotropic Bessel potentials associated with the group of linear transformations

$$A_h X = \left( h^{\frac{2n+2}{n+2}} x_0,\ h^{\frac{n+1}{n+2}} x_1, \ldots, h^{\frac{n+1}{n+2}} x_n \right), \quad X \in \boldsymbol{R}^{n+1},\ h > 0. \tag{41}$$

Assume moreover, that, for each of the functions $b_i$, the set $E_{b_i}$,

$$E_{b_i} = \{ X = (t, x) \in G:\ \limsup_{h \to 0} \int_{t-h^2}^{t+h^2} \left( \fint_{Q(x,h)} b_i(s, y)^p dy \right)^{q/p} ds < +\infty \},$$

is of full $B_{a,2}$ − capacity:

$$B_{a,2}(G \backslash E_{b_i}) = 0, \quad i = 1, \ldots, 7 \tag{42}$$

(this is always true e. g. if the $b_i$ are continuous except at isolated points of $G$). Finally, let $\dim E$ denote the Hausdorff dimension of a set $E$ with respect to the homogeneous norm $r$ associated with the group of 'dilations' (41).

CONJECTURE 3.   Let $u$ be a weak solution of (1) and denote by $L_u$ the set of those points in $G$ at which $u$ is Lipschitz continuous with respect to the parabolic distance $d_p$, that is, $X \in G$ belongs to the set $L_u$ if and only if there exist positive $C$ and $\delta$ such that for $h \in (0, \delta)$

$$\max_{Y \in R(X,h)} |u(X) - u(Y)| \leq C \cdot h.$$

If $a = \dfrac{(\ell+1)(n+1}{n+2}$ and the conditions (40) and (42) are fulfilled, then

$$B_{a,2}(G \backslash L_u) = 0, \tag{43}$$
$$\dim(G \backslash L_u) \leq n + 1 - 2a. \tag{44}$$

In particular, $\dim(G \backslash L_u) < n$ for $\ell = 2$.

REMARK.   This Conjecture follows from Conjecture 2, Theorem 10 of [DT] and the results of Section 4 of [Vol].

However, even in the elliptic case, when it is possible to prove the a.e. differentiability of *all* the weak solutions using only their local boundedness (see [HS]), this additional information about the capacity (and the

Hausdorff dimension) of the non-differentiability set cannot be obtained independently of the Hölder continuity of weak solutions (all the subtle information about the capacity is lost when Stepanoff's criterion is applied). Therefore, to establish an analogue of Theorem 2 containing information about the capacity and Hausdorff dimension of the exceptional set, it seems to be necessary to dispose of some more delicate from of Stepanoff's criterion.

Acknowledgement. The author would like to express his sincere thanks to Professor Bogdan Bojarski for introducing him to this subject. He is also much indebted to Dr. Wojciech Chojnacki for his valuable help and encouragement, and to Piotr Hajłasz for endless conversations without which this paper would have never been written.

## References

[AS]  D. G. ARONSON & J. SERRIN, Local behavior of solutions of quasilinear parabolic equations, Arch. Rat. Mech. Anal. **25** (1967), 81-122.

[Bo1]  B. BOJARSKI, The generalized solutions of a system of elliptic equations of first order with discontinuous coefficients, Mat. Sb. **43** (1957), 451-503 (in Russian).

[Bo2]  B. BOJARSKI, Pointwise differentiability of weak solutions of elliptic divergence type equations, Bull. Polish Acad. Sci. Math. **33** (1985), 1-6.

[CT]  A. P. CALDERÓN & A. TORCHINSKY, Parabolic maximal functions associated with a distribution, Advances in Math. **16** (1975), 1-63.

[DT]  H. DAPPA & W. TREBELS, On $L^1$-criteria for quasi-radial Fourier multipliers with applications to some anisotropic function spaces, Anal. Math. **9** (1983), 275-289.

[DeG]  M. DE GUZMÁN, Differentiation of integrals in $R^n$, Lecture Notes in Math. vol. 481, Springer 1975.

[DiB]  E. DiBENEDETTO, Local behavior of solutions of degenerate parabolic equations with measurable coefficients, Ann. Scuola Norm. Sup. Pisa, ser. IV, **13** (1986), 487 -535.

[F]  H. FEDERER, Geometric measure theory, Springer-Verlag, 1969.

[H]  P. HAJŁASZ, Sobolev functions, pointwise differentiability, and elliptic equations (in preparation).

[HS]  P. HAJŁASZ & P. STRZELECKI, A new proof of Reshetnyak's theorem concerning the pointwise differentiability of solutions of quasilinear equations, Warsaw University, preprint no. 1/1990.

[I]  A. V. IVANOV, A priori estimates for solutions of linear equations of second order of elliptic and parabolic type, Dokl, Akad. Nauk SSSR **161** (1965), 1270-1273 (in Russian).

[L]  P. I. LIZORKIN, Generalized Liouville differentiation and the method of multipliers in the theory of embeddings of classes of differentiable functions, Trudy Math. Inst. Steklov., **105** (1969), 89-167 (in Russian).

[M]  A. MORI, On quasiconformality and pseudoanalyticity, Trans. Amer. Math. Soc. **84** (1957), 56-77.

[Mo1]  J. MOSER, A new proof of de Giorgi's theorem, Comm. Pure Appl. Math. **13** (1960), 457-468.

[Mo2] J. MOSER, On Harnack's theorem for elliptic differential equations, Comm. Pure Appl. Math. **14** (1961), 577-591.

[Mo3] J. MOSER, A Harnack inequality for parabolic equations, Comm. Pure Appl. Math. **17** (1964), 101-134; Correction, ibidem **20** (1967), 231-236.

[N] J. NASH, Continuity of solutions of parabolic and elliptic equations, Amer. J. Math. **80** (1958), 931-954.

[R1] Yu. G. RESHETNYAK, The generalized derivatives and the a. e. differentiability, Mat. Sb. **75** (1968), no. 3, 323-334 (in Russian).

[R2] Yu. G. RESHETNYAK, Almost everywhere differentiability of solutions of elliptic equations, Sibirsk. Mat. Zh. **28** (1987), no. 4, 193-196 (in Russian).

[St] E. M. STEIN, Singular integrals and differentiability properties of functions, Princeton University Press, Princeton 1970.

[Sff] V. STEPANOFF, Sur les conditions de l'existence de la différentielle totale, Mat. Sb. **32** (1925), 511-526.

[S] P. STRZELECKI, Pointwise differentiability of weak solutions of parabolic equtions with measurable coeffients, Ann. Acad. Sci. Fenn., in press.

[T] N. S. TRUDINGER, Pointwise estimates and quasilinear parabolic equations, Comm. Pure Appl. Math. **21** (1968), 205-226.

[Vo1] S. K. VODOP'YANOV, Potential theory on homogeneous groups, Mat. Sb. **180**, no. 1 (1989), 57-77 (in Russian).

[Vo2] S. K. VODOP'YANOV, $L^p$-potential theory for generalized kernels and its applications, Institue of Mathematics, AN SSSR (Siberian Branch), Novosibirsk, preprint no. 6/1990.

[Z] W. P. ZIEMER, Weakly differentiable functions, Graduate texts in Mathematics, Springer 1989.

INSTYTUT MATEMATYKI,
UNIWERSYTET WARSZAWSKI
UL. BANACHA 2, 02-097 WARSZAWA
POLSKA (POLAND)
E-MAIL: PAWELST@MIMUW.EDU.PL