CONTINUITY OF THE MAXIMAL OPERATOR IN SOBOLEV SPACES

HANNES LUIRO

(Communicated by Juha M. Heinonen)

Abstract. We establish the continuity of the Hardy-Littlewood maximal operator on Sobolev spaces $W^{1,p}(\mathbb{R}^n)$, $1 < p < \infty$. As an auxiliary tool we prove an explicit formula for the derivative of the maximal function.

1. Introduction

The classical Hardy-Littlewood maximal operator $M$ is defined on $L^1_{\text{loc}}(\mathbb{R}^n)$ by setting for all $f \in L^1_{\text{loc}}(\mathbb{R}^n)$

\[
Mf(x) = \sup_{r > 0} \frac{1}{m(B_r)} \int_{B(x,r)} |f(y)| \, dy,
\]

for every $x \in \mathbb{R}^n$; here $m$ denotes the Lebesgue measure in $\mathbb{R}^n$ and $B_r = B(0, r)$.

The theorem of Hardy, Littlewood and Wiener asserts that $M$ is bounded on $L^p(\mathbb{R}^n)$ for $1 < p \leq \infty$. This theorem is one of the cornerstones of harmonic analysis. Applications e.g. to the study of Sobolev-functions indicate that it is also useful to know how it preserves differentiability properties of functions. Quite recently, Kinnunen observed [K] that $M$ is bounded on the Sobolev-space $W^{1,p}(\mathbb{R}^n)$, for $1 < p \leq \infty$. Extensions and related results can be found from e.g. [KL], [Ko], [KS], [HO].

Continuity of the maximal operator in $L^p(\mathbb{R}^n)$ follows from its sublinearity and boundedness. Because of boundedness in $W^{1,p}(\mathbb{R}^n)$, it is very natural to ask whether the maximal operator is continuous in $W^{1,p}(\mathbb{R}^n)$, $1 < p < \infty$, or not. This question was posed in [HO, Question 3] where it was attributed to T. Iwaniec. In general, bounded non-sublinear operators need not be continuous. An important example of this kind of phenomenon is the result of Almgren and Lieb [AL] who proved that the (known to be bounded) symmetric rearrangement $R : W^{1,p}(\mathbb{R}^n) \rightarrow W^{1,p}(\mathbb{R}^n)$ is not continuous when $1 < p < n$ and $n > 1$. On Sobolev-spaces, $M$ is not sublinear and the issue of the continuity of $M$ is not trivial even though we know the boundedness.

Our main result (Theorem 4.1 below) is the positive answer to the question of Iwaniec. A central role in our proof is played by a careful analysis of the set $Rf(x)$
(see [2.1] below), which consists of the radii $r$ for which equality is achieved in (1). As a useful auxiliary tool we establish in Theorem 3.1 an explicit formula for the derivative of the maximal function.

### 2. Definitions and auxiliary results

Let us first introduce some notation. If $A \subset \mathbb{R}^n$ and $r \in \mathbb{R}^n$, we define

$$d(r, A) := \inf_{a \in A} |r - a|, \text{ and } A(\lambda) := \{x \in \mathbb{R}^n : d(x, A) \leq \lambda \} \text{ for } \lambda \geq 0.$$  

We endow $W^{1, p}(\mathbb{R}^n)$ with the norm

$$\|f\|_{1,p} = \|f\|_p + \|\nabla f\|_p,$$

where $\nabla f$ is the weak gradient of $f$. Let us also denote by $\|f\|_{p,A}$ the $L^p$-norm of $\chi_A f$ for all measurable sets $A \subset \mathbb{R}^n$.

The following new concept will be central in this work.

**Definition 2.1.** Let $f \in L^p(\mathbb{R}^n)$, $1 \leq p < \infty$. The set $\mathcal{R}f(x)$ is defined as

$$\mathcal{R}f(x) = \{r \geq 0 : Mf(x) = \limsup_{r_k \to r} \int_{B(x, r_k)} |f(y)| \, dy, \text{ for some } r_k > 0 \}.$$  

**Remarks.** We comment on the above definition and the properties of the sets $\mathcal{R}f(x)$. First, the definition clearly implies that $\mathcal{R}f(x)$ is always closed. Moreover, for fixed $x \in \mathbb{R}^n$ define $u_x : [0, \infty) \to \mathbb{R}$ by

$$u_x(0) = |f(x)| \quad \text{and} \quad u_x(r) = \int_{B(x, r)} |f(y)| \, dy \quad \text{when} \quad r \in (0, \infty).$$

First of all, the functions $u_x$ are continuous for almost all $x$. The continuity on $(0, \infty)$ is clearly true for all $x$ and at 0 it follows a.e., because almost every point $x \in \mathbb{R}^n$ is a Lebesgue point for $f$. Moreover, by Hölder’s inequality we have

$$u_x(r) \leq \|f\|_p (m(B_x))^\frac{1}{q-1},$$

where $q$ is the conjugate exponent of $p$, and hence $\lim_{r \to 0} u_x(r) = 0$. These facts together imply that, for almost all $x$, the function $u_x$ has at least one maximum point in $[0, \infty)$. Furthermore, they guarantee that for all $x \in \mathbb{R}^n$ the set $\mathcal{R}f(x)$ is nonempty and

$$Mf(x) = \int_{B(x, r)} |f(y)| \, dy \quad \text{if } r \in \mathcal{R}f(x) \text{ and } r > 0, \forall x \in \mathbb{R}^n,$$

$$Mf(x) = |f(x)| \quad \text{for almost every } x \text{ such that } 0 \in \mathcal{R}f(x).$$

Also, it is useful to observe that for every $R > 0$ (assuming $f \not= 0$) it is true that

$$\sup\{r : r \in \mathcal{R}f(x), x \in B(0, R)\} < \infty.$$

The following lemma tells us how the sets $\mathcal{R}f(x)$ and $\mathcal{R}g(x)$ are related to each other, especially when $\|f - g\|_p$ is small.

**Lemma 2.2.** Let $1 \leq p < \infty$ and suppose $f_j \to f$ in $L^p(\mathbb{R}^n)$ when $j \to \infty$. Then for all $R > 0$ and $\lambda > 0$ it holds that

$$m(\{x \in B(0, R) : \mathcal{R}f_j(x) \not\subset \mathcal{R}f(x)(\lambda)\}) \to 0 \quad \text{if } j \to \infty.$$
Proof: First we indicate why the above set is always Lebesgue-measurable when \( f \) and all the functions \( f_j \) are in \( L^p(\mathbb{R}^n) \). The continuity of the average functions \( u_x \), for almost every \( x \), is used as a main tool in the following argument. Let the set \( \mathcal{N} \) consist of those points which are not Lebesgue points of any of the functions \( f_j \) or \( f \), especially, \( m(\mathcal{N}) = 0 \). Moreover we denote by \( Q_+ \) the set of positive rationals. Now we can write
\[
\{ x : R f_j(x) \not\subset R f(x)_\lambda \} \setminus \mathcal{N} = \bigcup_{i=1}^\infty \bigcap_{m=1}^\infty \{ x : \exists r > 0 \text{ s.t. } d(r, R f(x)) > \lambda + \frac{1}{i} \text{ and } M f_j(x) < \int_{B(x,r)} f_j + \frac{1}{m} \} \]
\[
= \bigcup_{i=1}^\infty \bigcap_{m=1}^\infty \bigcup_{q \in Q_+} \left[ \{ x : d(q, R f(x)) > \lambda + \frac{1}{i} \} \cap \{ x : M f_j(x) < \int_{B(x,q)} f_j + \frac{1}{m} \} \right].
\]
From this we conclude that it is enough to prove that the set \( \{ x : d(q, R f(x)) > \lambda \} \) is measurable for arbitrary \( q \) and \( \lambda \). Using the same reasoning as above, especially the continuity of the expression \( \int f \) as a function of \( r \), we write that
\[
\{ x : d(q, R f(x)) > \lambda \} = \bigcup_{k=1}^\infty \bigcap_{k' \in Q_+ \cap [q-\lambda, q+\lambda]} \{ x : M f(x) > \int_{B(x,q')} f + \frac{1}{k} \}.
\]
This implies the measurability.

Then we are ready to prove the lemma. It is sufficient to prove the claim in the case where both \( f \) and \( f_j \) are nonnegative, because \( R f(x) = R|f|(x) \). Observe that \( R f(x) \) is \([0, \infty)\) for all \( x \) if \( f \equiv 0 \) a.e., whence this case is trivial. Let \( \lambda > 0, R > 0 \) and \( \varepsilon > 0 \). For almost every \( x \in B(0, R) \) there exists a natural number \( i(x) \in \mathbb{N} \) so that
\[
\int_{B(x,r)} f(y) \, dy < M f(x) - \frac{1}{i(x)} , \text{ when } d(r, R f(x)) > \lambda.
\]
This can be seen in the following way: If the claim is not true there is a sequence of radii \( (r_k)_{k=1}^\infty \) so that
\[
\int_{B(x,r_k)} f(y) \, dy \to M f(x) \text{ and } d(r_k, R f(x)) > \lambda.
\]
By moving to a subsequence, if needed, we may assume that \( r_k \to r \) as \( k \to \infty \), because (2) implies that the sequence \( (r_k)_{k=1}^\infty \) must be bounded. It follows that \( r \in R f(x) \). This is a contradiction, since obviously \( r \) satisfies \( d(r, R f(x)) \geq \lambda \).

From (3) we conclude that there exists \( i \in \mathbb{N} \) so that
\[
B(0, R) \subset \{ x : \int_{B(x,r)} f(y) \, dy < M f(x) - \frac{1}{i} , \text{ if } d(r, R f(x)) > \lambda \} \cup E = : A \cup E,
\]
where \( E \) is a measurable set with \( m(E) < \varepsilon \). The weak type \((1,1)\)-estimate for the maximal operator implies that there exists \( j_0 \in \mathbb{N} \) so that
\[
m\left( \{ x \in B(0, R) : |M(f - f_j)(x)| \geq \frac{1}{4i} \} \right) < \varepsilon \text{ when } j \geq j_0.
\]
For all \( j \) we observe that
\[
A \subset \left\{ x : \int_{B(x,r)} f_j(y) \, dy < M f(x) - \frac{1}{2i} \text{, if } d(r, \mathcal{R} f(x)) > \lambda \right\}
\]
\[
\cup \left\{ x : \int_{B(x,r)} f(y) \, dy - \int_{B(x,r)} f_j(y) \, dy \geq \frac{1}{2i} \text{ for some } r, d(r, \mathcal{R} f(x)) > \lambda \right\}
\]
\[=: A_j \cup B_j.\]
Continuing the same reasoning, and using the fact that \(|M f(x) - M f_j(x)| \leq |M(f - f_j)(x)|\), we get
\[
A_j \subset \left\{ x : \int_{B(x,r)} f_j(y) \, dy < M f_j(x) - \frac{1}{4i} \text{, if } d(r, \mathcal{R} f(x)) > \lambda \right\}
\]
\[
\cup \left\{ x : |M(f - f_j)(x)| \geq \frac{1}{4i} \right\}
\]
\[=: C_j \cup D_j.\]
Now
\[
C_j \subset \left\{ x : \mathcal{R} f_j(x) \subset \mathcal{R} f(x)(\lambda) \right\}.
\]
By combining the above observations, we conclude that for all \( j \)
\[
B(0, R) \subset \left\{ x : \mathcal{R} f_j(x) \subset \mathcal{R} f(x)(\lambda) \right\} \cup E \cup D_j \cup B_j.
\]
Observe finally that \( B_j \subset D_j \) and, by our choice of \( j_0 \) we have \( m(D_j) < \varepsilon \) if \( j \geq j_0 \), and therefore
\[
m(\left\{ x \in B(0, R) : \mathcal{R} f_j(x) \not\subset \mathcal{R} f(x)(\lambda) \right\}) < 2\varepsilon ,
\]
if \( j \geq j_0 \).

Let us introduce more notation. Assume that \( f \in L^p(\mathbb{R}^n) \), \( 1 \leq p < \infty \). Let \( e_i \) be one of the standard basevectors of \( \mathbb{R}^n \). For all \( h \in \mathbb{R} \), \( |h| > 0 \), we define the functions \( f^i_h \) and \( f^{i(h)}_r \) by setting
\[
f^i_h(x) = \frac{f(x + he_i) - f(x)}{|h|} \quad \text{and} \quad f^{i(h)}_r(x) = f(x + he_i).
\]
Now we know that \( f^{i(h)}_r \to f \) in \( L^p(\mathbb{R}^n) \) when \( |h| \to 0 \) and, if \( p > 1 \), for functions \( f \in W^{1,p}(\mathbb{R}^n) \) we have (see [G! 7.11]) that \( f^i_h \to D_i f \) in \( L^p(\mathbb{R}^n) \) when \( |h| \to 0 \).

**Corollary 2.3.** Let \( f \in L^p(\mathbb{R}^n) \), \( 1 \leq p < \infty \). Then for all \( i, 1 \leq i \leq n, R > 0, \lambda > 0 \) one has
\[
m(\left\{ x \in B(0, R) : \mathcal{R} f(x) \not\subset \mathcal{R} f^{i(h)}_r(x)(\lambda) \text{ or } \mathcal{R} f^{i(h)}_r(x) \not\subset \mathcal{R} f(x)(\lambda) \right\}) \to 0 .
\]

**Proof.** Now \( f^{i(h)}_r \to f \) and as a consequence of Lemma 2.2 it is clearly sufficient to prove that
\[
m(\left\{ x \in B(0, R) : \mathcal{R} f(x) \not\subset \mathcal{R} f^{i(h)}_r(x)(\lambda) \right\}) \to 0 \text{ as } h \to 0 .
\]
But this also follows easily from Lemma 2.2 because for \( |h| < 1 \) one has that
\[
\left\{ x \in B_R : \mathcal{R} f(x) \not\subset \mathcal{R} f^{i(h)}_r(x)(\lambda) \right\}
\]
\[
= \left\{ x \in B_R : \mathcal{R} f^{i(-h)}_r(x + he_i) \not\subset \mathcal{R} f(x + he_i)(\lambda) \right\}
\]
\[
\subset \left\{ y \in B_{R+1} : \mathcal{R} f^{i(-h)}_r(y) \not\subset \mathcal{R} f(y)(\lambda) \right\} - he_i.
\]
\[\quad \quad \square\]
Remark. The previous corollary will became useful after the following observation. Let us denote by
\[ \pi(A, B) := \inf \{ \delta > 0 : A \subset B(\delta) \text{ and } B \subset A(\delta) \} \]
the Hausdorff distance of the sets \( A \) and \( B \). Let \( f \) be in \( L^p(\mathbb{R}^n) \). With the new notation, the corollary says that
\[ m(\{ x \in B_R : \pi(Rf(x), Rf(x + he_i)) > \lambda \}) \to 0 \text{ when } h \to 0. \]
Therefore we easily infer that there is a sequence \((h_k)_{k=1}^{\infty}, h_k > 0 \) with \( h_k \to 0 \), and such that \( \pi(Rf(x), Rf(x + h_k e_i)) \to 0 \) as \( k \to \infty \) for almost every \( x \in B_R \). This is the decisive fact needed in the following section.

3. A formula for the derivative of the maximal function

Let us denote by \( D_i f(x) \) the partial derivative \( \frac{\partial f}{\partial x_i} \).

**Theorem 3.1.** Let \( f \in W^{1,p}(\mathbb{R}^n), 1 < p < \infty \). Then we have for almost all \( x \in \mathbb{R}^n \) that

1. \( D_i Mf(x) = \int_{B(x,r)} D_i f(y) \, dy \) for all \( r \in Rf(x), r > 0 \), and
2. \( D_i Mf(x) = D_i |f|(x) \) if \( 0 \in Rf(x) \).

**Proof.** It is sufficient to prove the claim for nonnegative functions, because \( Mf = |f| \) and \( |f| \in W^{1,p}(\mathbb{R}^n) \) if \( f \in W^{1,p}(\mathbb{R}^n) \). Let \( R > 0 \). We start by choosing a sequence \((h_k)_{k=1}^{\infty}, h_k > 0 \) and \( h_k \to 0 \), so that \( \pi(Rf(x), Rf(x + h_k e_i)) \to 0 \) as \( k \to \infty \) for almost all \( x \in B_R \) (see the Remark after Corollary 2.3). Then we have

(i) \[ \| D_i Mf - (Mf) \|_{p,B_R} \to 0 \text{ as } k \to \infty, \]

(ii) \[ \| D_i f - f_{h_k} \|_{p,B_R} \to 0 \text{ as } k \to \infty, \]

(iii) \[ \| M(D_i f - f_{h_k}^i) \|_{p,B_R} \to 0 \text{ as } k \to \infty. \]

Now, by extracting a subsequence if needed, we may assume that the convergences above are true pointwise almost everywhere as well. Moreover, we recall that the set
\[ \{ x \in \mathbb{R}^n : \exists k \in \mathbb{N} \text{ s.t. } 0 \in Rf(x + h_k e_i) \text{ with } Mf(x + h_k e_i) \neq f(x + h_k e_i) \} \]
has measure zero as a countable union of the sets having measure zero. Let \( x \in B_R \) be a Lebesgue point of both \( f \) and \( D_i f \) outside the union of all these unwanted sets of measure zero (in particular, the pointwise analogies of (i)–(iii) hold at \( x \)) and let \( r \in Rf(x) \).
Now, because \( \pi(\mathcal{R}f(x), \mathcal{R}(x + h_k e_i)) \to 0 \), we find \( r_k \in \mathcal{R}f(x + h_k e_i) \) so that \( r_k \to r \) when \( k \to \infty \). If \( r > 0 \) we can estimate:

\[
D_i Mf(x) = \lim_{k \to \infty} \frac{1}{h_k} (Mf(x + h_k e_i) - Mf(x)) \\
\leq \lim_{k \to \infty} \frac{1}{h_k} \left( \int_{B(x + h_k e_i, r_k)} f(y) \, dy - \int_{B(x, r_k)} f(y) \, dy \right) \\
= \lim_{k \to \infty} \frac{1}{m(B(x, r_k))} \int_{B(x, r_k)} \left( \frac{f(y + h_k e_i) - f(y)}{h_k} \right) \, dy \\
= \int_{B(x, r)} D_i f(y) \, dy.
\]

The last equation holds, because \( m(B_{r_k}) \to m(B_r) \) and

\[
\chi_{B(x, r_k)} f_{h_k}^i \to \chi_{B(x, r)} D_i f \quad \text{in} \quad L^1(\mathbb{R}^n) \quad \text{as} \quad k \to \infty.
\]

On the other hand, we get that

\[
D_i Mf(x) \geq \lim_{k \to \infty} \frac{1}{h_k} \left( \int_{B(x + h_k e_i, r_k)} f(y) \, dy - \int_{B(x, r_k)} f(y) \, dy \right) \\
= \lim_{k \to \infty} \frac{1}{m(B(x, r_k))} \int_{B(x, r_k)} \left( \frac{f(y + h_k e_i) - f(y)}{h_k} \right) \, dy = \int_{B(x, r)} D_i f(y) \, dy.
\]

Suppose instead that \( r = 0 \). The proof of the lower bound of \( D_i Mf(x) \) applies now, too, and we get that \( D_i Mf(x) \geq D_i f(x) \). If we have \( r_k = 0 \) for infinitely many \( k \), we can decide straightforwardly that \( D_i Mf(x) = D_i f(x) \). If \( r_k > 0 \) starting from some \( k_0 \), we get by the same way as when studying the upper bound of \( D_i Mf(x) \) in the case \( r > 0 \) that

\[
D_i Mf(x) \leq \lim_{k \to \infty} \int_{B(x, r_k)} f_{h_k}^i(y) \, dy = D_i f(x),
\]

because

\[
\lim_{k \to \infty} \left| \int_{B(x, r_k)} f_{h_k}^i(y) \, dy - D_i f(x) \right| = \lim_{k \to \infty} \left| \int_{B(x, r_k)} \left( f_{h_k}^i(y) - D_i f(y) \right) \, dy \right| \\
\leq \lim_{k \to \infty} M(f_{h_k}^i - D_i f)(x) = 0.
\]

Now we have shown the claim in the ball \( B(0, R) \). Since \( R \) was arbitrary, this completes the proof. \( \square \)

4. **CONTINUITY OF THE MAXIMAL OPERATOR IN \( W^{1,p}(\mathbb{R}^n) \)**

By using Theorem \[3.1\] and Lemma \[2.2\] we can establish quite easily our main result which verifies the continuity of the maximal operator in \( W^{1,p}(\mathbb{R}^n) \).

**Theorem 4.1.** \( M : W^{1,p}(\mathbb{R}^n) \to W^{1,p}(\mathbb{R}^n) \) is continuous for all \( 1 < p < \infty \).

**Proof.** Let \( f_j \to f \) in \( W^{1,p}(\mathbb{R}^n) \) when \( j \to \infty \). We have to show that \( \|Mf_j - Mf\|_{1,p} \to 0 \). Because we know the continuity of \( M \) in \( L^p(\mathbb{R}^n) \), it is sufficient to prove that
\[ \|D_i M f_j - D_i M f\|_p \to 0 \text{ for all } i, \ 1 \leq i \leq n. \] Also it is clear that we may assume the functions \( f_j \) and \( f \) to be nonnegative.

Let \( \varepsilon > 0 \) be fixed but arbitrary. We start by choosing \( R > 0 \) so that \( \|2 M D_i f\|_{p,C_1} < \varepsilon \), where \( C_1 = \mathbb{R}^n \setminus B(0, R) \). By absolute continuity we choose \( \alpha > 0 \) so that \( \|2 M D_i f\|_{p,A} < \varepsilon \) always when \( m(A) < \alpha \) and \( A \) is a measurable subset of \( B(0, R) \).

We let \( \{r \} \) (compare with the remark after Definition 2.1) \( u_x(r) \) stand for the average of \( D_i f \) in the ball \( B(x, r) \) and \( u_x(0) = D_i f(x) \). As already observed, for almost every \( x \in \mathbb{R}^n \) the functions \( u_x \) are continuous on \( [0, \infty) \) and converge to 0 when \( r \to \infty \). Consequently for almost every \( x \) the function \( u_x \) is uniformly continuous on \( [0, \infty) \) and therefore we can find \( \delta(x) > 0 \) such that \( |u_x(r_1) - u_x(r_2)| < \frac{\varepsilon}{(m(B))^{\frac{1}{p}}} \) when \( |r_1 - r_2| < \delta(x) \). Now we write that

\[
B_R = \left( \bigcup_{i=1}^{\infty} \{x \in B_R : \delta(x) > \frac{1}{i}\} \right) \cup N,
\]

where \( m(N) = 0 \). From that we infer that there exists \( \delta > 0 \) such that

\[
m(\{x \in B_R : |u_x(r_1) - u_x(r_2)| > \frac{\varepsilon}{(m(B))^{\frac{1}{p}}} \text{ for some } r_1, r_2, |r_1 - r_2| < \delta\})
=: m(C_2) < \frac{\alpha}{2}.
\]

The set \( C_2 \) is easily shown to be measurable. Furthermore, Lemma 3.1 says that we can find \( j_0 \) so that

\[
m(\{x : R f_j(x) \nsubseteq R f(x)(\delta)\}) =: m(C^j) < \frac{\alpha}{2} \text{ when } j \geq j_0.
\]

Then, let \( j \geq j_0 \) be fixed. It follows from Theorem 3.1 that almost everywhere in \( \mathbb{R}^n \)

\[
|D_i M f_j(x) - D_i M f(x)| = \left| \int_{B(x,r_1)} D_i f_j(y) \, dy - \int_{B(x,r_2)} D_i f(y) \, dy \right|
\leq \left| \int_{B(x,r_1)} D_i f_j(y) \, dy - \int_{B(x,r_1)} D_i f(y) \, dy \right|
+ \left| \int_{B(x,r_1)} D_i f(y) \, dy - \int_{B(x,r_2)} D_i f(y) \, dy \right|
\leq M |D_i f_j - D_i f|(x) + \left| \int_{B(x,r_1)} D_i f(y) \, dy - \int_{B(x,r_2)} D_i f(y) \, dy \right|
\]

for all \( r_1 \in R f_j(x), r_2 \in R f(x) \). This inequality applies also to the cases \( r_1 = 0 \) or \( r_2 = 0 \) when we agree that

\[
\int_{B(x,0)} D_i f(y) \, dy := D_i f(x).
\]

This is obvious because for almost every \( x \) it is true that \( M f(x) \geq f(x) \), and by Theorem 3.1 \( D_i M f(x) = D_i f(x) \) if \( 0 \in R f(x) \).
Now, if \( x \notin C_1 \cup C_2 \cup C_j \), we can pick \( r_1 \in \mathcal{R}_{f_j}(x) \) and \( r_2 \in \mathcal{R}_f(x) \) so that \( |r_1 - r_2| < \delta \). Our choice of \( \delta \) implies that

\[
s := \left| \int_{B(x,r_1)} D_i f(y) \, dy - \int_{B(x,r_2)} D_i f(y) \, dy \right| < \frac{\varepsilon}{(m(B_R))^\frac{1}{p}}.
\]

If \( x \in C_1 \cup C_2 \cup C_j \), we estimate that \( s \leq 2MD_i f(x) \). Observe also that \( m(C_2 \cup C_j) < \alpha \).

Combining the above estimates it follows that

\[
\|D_i Mf_j - D_i Mf\|_{p,\mathbb{R}^n} \leq \|M(D_i f_j - D_i f)\|_{p,\mathbb{R}^n} + \left\| \frac{\varepsilon}{(m(B_R))^\frac{1}{p}} \right\|_{p,B_R}^p + \|2MD_i f\|_{p,C_1} + \|2MD_i f\|_{p,C_2 \cup C_j}.
\]

The first term in the right-hand side of the inequality converges to zero when \( j \to \infty \). The rest of the terms are less than \( \varepsilon \), because of the choices of \( R \) and \( \alpha \). As \( \varepsilon \) was arbitrary we conclude that \( \|D_i Mf_j - D_i Mf\|_p \to 0 \) as \( j \to \infty \). The proof is complete. \( \square \)

**Remark.** One may ask, what kind of estimates we can find for the modulus of continuity of \( M \). Quite surprisingly, it turns out that there does not exist a function \( F : (0,\infty) \to (0,\infty) \) such that

\[
\|Mf - Mg\|_{1,p} \leq F(\|f - g\|_{1,p}) \quad \text{for all } f, g \in W^{1,p}(\mathbb{R}^n).
\]

This is a consequence of the following two facts. First, \( M \) is not Lipschitz-continuous in \( W^{1,p}(\mathbb{R}^n) \), because this would imply that \( M \) is bounded in \( W^{2,p}(\mathbb{R}^n) \) which is not true (see for example [Ko]). The philosophy of this phenomenon is that even the maximal function of a smooth positive function usually has angles in its graph. Second, the maximal operator is scale-invariant, thus \( M(cf) = cMf \) for all \( c > 0 \).

We thank Jani Onninen for pointing out the first fact.

**Acknowledgements**

I would like to thank Eero Saksman for the advice and support he has given me. I also thank Juha Kinnunen and Jani Onninen for their valuable comments on the manuscript.

**References**


Department of Mathematics and Statistics, University of Jyväskylä, P.O. Box 35 (MaD), 40014 University of Jyväskylä, Finland

E-mail address: haluio@maths.jyu.fi