Annali di Matematica (2009) 188:35–59
DOI 10.1007/s10231-008-0065-6

Pointwise behaviour of Orlicz–Sobolev functions

Heli Tuominen

Received: 5 June 2007 / Revised: 9 August 2007 / Published online: 14 March 2008

Abstract We study the regularity of Orlicz–Sobolev functions on metric measure spaces equipped with a doubling measure. We show that each Orlicz–Sobolev function is quasicontinuous and has Lebesgue points outside a set of capacity zero and that the discrete maximal operator is bounded in the Orlicz–Sobolev space. We also show that if the Hardy–Littlewood maximal operator is bounded in the Orlicz space $L^{\Psi}(X)$, then each Orlicz–Sobolev function can be approximated by a Hölder continuous function both in the Lusin sense and in norm.

Keywords Orlicz–Sobolev space · Lebesgue point

Mathematics Subject Classification (2000) 46E30

1 Introduction

In this paper, we consider pointwise properties of Orlicz–Sobolev functions on metric measure spaces equipped with a doubling measure. Recall that, for a domain $\Omega \subset \mathbb{R}^n$ and a Young function Ψ, the Orlicz–Sobolev space $W^{1,\Psi}(\Omega)$ consists of the functions $u \in L^{\Psi}(\Omega)$ all of whose first order weak derivatives belong to the Orlicz space $L^{\Psi}(\Omega)$, see Sect. 2 for the definition of Young function and Orlicz space $L^{\Psi}(\Omega)$. The space $W^{1,\Psi}(\Omega)$ is a Banach space with respect to the norm

$$\|u\|_{W^{1,\Psi}(\Omega)} = \|u\|_{L^{\Psi}(\Omega)} + \||\nabla u|\|_{L^{\Psi}(\Omega)},$$

where $\|\cdot\|_{L^{\Psi}(\Omega)}$ is the Luxemburg norm and ∇u is the weak gradient of u.

The research is supported by the Centre of Excellence Geometric Analysis and Mathematical Physics of the Academy of Finland.

H. Tuominen (✉)
Department of Mathematics and Statistics, University of Jyväskylä,
P. O. Box 35 (MaD), 40014 Jyväskylä, Finland
e-mail: tuheli@maths.jyu.fi

Analysis on metric measure spaces, for example the theory of Sobolev type spaces, has been under active study during the past decade. In a general metric space we cannot speak about weak derivatives; hence there has been a need for characterizations of the classical Sobolev and Orlicz–Sobolev spaces that do not involve derivatives. We use spaces $N^{1,\Psi}(X)$ consisting of $L^{\Psi}(X)$-functions with upper gradients in $L^{\Psi}(X)$. The basic properties of these spaces are studied in [43], see also Sect. 2.

By the Lebesgue differentiation theorem, almost every point $x \in \mathbb{R}^n$ is a Lebesgue point of a locally integrable function u, that is,

$$\lim_{r\to 0} \fint_{B(x,r)} |u(y) - u(x)|\, dx = 0. \tag{1}$$

The proof generalizes to metric spaces with a doubling measure, see [23, Theorem 14.15]. Classical Sobolev and Orlicz–Sobolev functions have Lebesgue points outside a set of capacity zero. The capacities used are often defined in terms of potentials, see for example [2] for the Sobolev case and [3,11] for Orlicz potential spaces. In the interesting paper [36] by Malý, Swanson and Ziemer, Orlicz–Sobolev capacity is defined as an infimum of integrals $\int \Psi(|u|)\, dx + \int \Psi(|\nabla u|)\, dx$ and used for example to show that for Orlicz–Sobolev functions, Lebesgue points exist quasi everywhere. Variational capacity where one takes infimum of the integrals of $\Psi(|\nabla u|)$ is studied by Rudd in [38]. In addition to [36,38], pointwise properties of Orlicz–Sobolev functions in $\mathbb{R}^n$ are recently studied in [6].

In the metric setting, Sobolev functions defined via pointwise inequalities have Lebesgue points quasi everywhere measured by the corresponding Sobolev capacity, see [29,31]. We use Orlicz–Sobolev capacity defined in [43] and show that $N^{1,\Psi}(X)$ functions have Lebesgue points everywhere except on a set of capacity zero. The main step in the proof is to show that the discrete maximal operator, which is smoother than the Hardy–Littlewood maximal operator, is bounded in the Orlicz–Sobolev space $N^{1,\Psi}(X)$.

If $p > n$, then by the Sobolev embedding theorem, (a representative of) each function of $W^{1,p}(\mathbb{R}^n)$ is locally $(1 - n/p)$-Hölder continuous. A corresponding embedding into a space of continuous functions, where the moduli of continuity depends on Ψ, holds also for Orlicz–Sobolev functions; the condition $p > n$ is replaced by an integrability condition $\int^{\infty} \widetilde{\Psi}(t) t^{-(1+n')}\, dt < \infty$, where $\widetilde{\Psi}$ is the conjugate function of Ψ and $n' = n/(n-1)$. Embedding theorems for $W^{1,\Psi}(\mathbb{R}^n)$ were first proved by Donaldson and Trudinger [13] and Adams [1], and improved by Cianchi [9].

We are interested in Lusin type continuity; by the classical Lusin theorem, every measurable function is continuous in a complement of a set of arbitrary small measure. If the function is more regular, then stronger versions of Lusin theorem hold. Namely, Malý showed in [35] that each function $u \in W^{1,p}(\mathbb{R}^n)$ coincides with a Hölder continuous function, that is close to u in Sobolev norm, outside a set with small capacity. The proof uses representation of Sobolev functions by Bessel potentials. In the metric space setting, approximation of Sobolev functions by Hölder continuous functions both in the Lusin sense and in norm, is studied by Hajłasz and Kinnunen [22], and Kinnunen and Tuominen [31]. As in the last two papers, we use maximal function arguments to show that if the Hardy–Littlewood maximal operator is bounded in the Orlicz space, then for a given $0 < \beta \leq 1$, each Orlicz–Sobolev function u coincides with a β-Hölder continuous function outside a set of small $(s - (1 - \beta)$-Hausdorff content, where s is the doubling dimension of μ. The approximating function is close to u in norm, see Theorem 5. This result is new even in the classical Orlicz–Sobolev space $W^{1,\Psi}(\mathbb{R}^n)$.

In the last Section, we also show that the two definitions of Orlicz–Sobolev space, via upper gradients or pointwise inequality, give the same space if X supports a Poincaré inequality and the Hardy–Littlewood maximal operator is bounded in the Orlicz space, see Theorem 4.

Example 1 In the main results, we assume that the Hardy–Littlewood maximal operator is bounded in the Orlicz space. This is true if $\Psi, \widetilde{\Psi}$ is a pair of doubling complementary N-functions. One such example is $\Psi, \widetilde{\Psi}$, where

$$\Psi(t) = t^p \log^\alpha(e+t)$$

with $p > 1$ and $\alpha > 0$, or $p > 1 - \alpha$ and $-1 \le \alpha < 0$. The fact that Ψ is doubling and has a doubling complementary function can be checked by standard tests for N-functions, (cf. [33, Chap. 4], [37, Chap. 2.2.3]). Weakly differentiable functions with gradient in the Orlicz space, Ψ as above, are used in the theory of mappings of finite distortion, see for example [26,27] and the references therein. Orlicz and Orlicz–Sobolev spaces with such Ψ are studied also in [14,18], the list not being exhaustive.

The paper is organized as follows. In Sect. 2, we introduce the notation and the standard assumptions and recall the definitions of Orlicz and Orlicz–Sobolev spaces. In Sect. 3, we discuss capacity and show that each function of $N^{1,\Psi}(X)$ is quasicontinuous. Section 4 contains lemmas. Section 5 deals with Lebesgue points of $N^{1,\Psi}(X)$-functions. The Hölder approximation of Orlicz–Sobolev functions is the content of the last section.

2 Notation and preliminaries

We assume that $X = (X, d, \mu)$ is a metric measure space equipped with a metric d and a Borel regular outer measure μ such that open sets have positive and bounded sets finite measure. We also assume that μ is *doubling*, which means that there is a constant $C_\mu > 0$, called *the doubling constant of* μ, such that

$$\mu(B(x, 2r)) \le C_\mu \mu(B(x, r))$$

for all balls $B(x,r) = \{y \in X : d(y,x) < r\}$. Recall that the doubling condition of μ implies that there exists a constant $C_0 > 0$ such that whenever $B_0 = B(x_0, r_0)$ and $B = B(x, r)$ are balls with $x \in B_0$ and $0 < r \le r_0$, then

$$\frac{\mu(B)}{\mu(B_0)} \ge C_0 \left(\frac{r}{r_0}\right)^s, \tag{2}$$

where $s = \log_2 C_\mu$, (see for example [23, Lemma 14.6]). In this paper, s denotes the smallest exponent for which (2) holds and it is called *the doubling dimension* of μ.

In the main results, we assume that X is proper, that is, closed balls of X are compact. Notice that as a doubling metric space X is proper if and only if it is complete.

The Hardy–Littlewood maximal function of a function $u \in L^1_{\mathrm{loc}}(X)$ is

$$\mathscr{M} u(x) = \sup_{r>0} \fint_{B(x,r)} |u| \, d\mu,$$

where $u_B = \fint_B u \, d\mu = \mu(B)^{-1} \int_B u \, d\mu$ is the integral average of u over B. The local space $L^1_{\mathrm{loc}}(X)$ consists of functions that are integrable in each ball.

Given $0 \le t < \infty$, $0 < \delta \le \infty$ and a set E,

$$\mathscr{H}^t_\delta(E) = \inf\left\{\sum_{i=1}^{\infty} r_i^t : E \subset \bigcup_{i=1}^{\infty} B(x_i, r_i), r_i \le \delta\right\},$$

and $\lim_{\delta\to 0}\mathscr{H}^t_\delta(E)$ is the t-Hausdorff measure of E. The number $0 \le \mathscr{H}^t_\infty(E) \le \infty$ is the t-Hausdorff content of E.

By χ_E, we denote the characteristic function of a set $E \subset X$. If $0 < t < \infty$ and $B = B(x, r)$ is a ball in X, then $tB = B(x, tr)$. In general, C will denote a positive constant whose value is not necessarily the same at each occurrence. By writing $C = C(\tau, \lambda)$, we indicate that the constant depends only on τ and λ.

2.1 Review of Orlicz spaces

We will give a brief review to Orlicz spaces. For more details and proofs, see for example [33,37]. A function $\Psi : [0, \infty) \to [0, \infty]$ is *a Young function* if

$$\Psi(s) = \int_0^s \psi(t)\, dt,$$

where $\psi : [0, \infty) \to [0, \infty]$ is an increasing, left continuous function which is neither identically zero nor identically infinite on $(0, \infty)$, and satisfies $\psi(0) = 0$. A Young function Ψ is convex, increasing, left-continuous, $\Psi(0) = 0$, and $\Psi(t) \to \infty$ as $t \to \infty$. A continuous Young function with properties $\Psi(t) = 0$ only if $t = 0$, $\Psi(t)/t \to \infty$ as $t \to \infty$, and $\Psi(t)/t \to 0$ as $t \to 0$ is called *an N-function*. Below, Ψ is always a Young function. For a given Ψ, the function $\widetilde{\Psi} : [0, \infty) \to [0, \infty]$, $\widetilde{\Psi}(s) = \sup\{st - \Psi(t) : t \ge 0\}$, is *the complementary function* of Ψ.

Convexity and the property $\Psi(0) = 0$ imply that

$$\begin{aligned} \Psi(\alpha t) &\le \alpha\Psi(t), && \text{if } 0 \le \alpha \le 1, \\ \Psi(\beta t) &\ge \beta\Psi(t), && \text{if } \beta \ge 1, \end{aligned} \tag{3}$$

and that the function $t \mapsto \Psi(t)/t$ is increasing. Hence, for a strictly increasing Ψ, the function $t \mapsto \Psi^{-1}(t)/t$ is decreasing.

A Young function Ψ is *doubling* (satisfies the Δ_2-condition) if there is a constant $C_\Psi > 0$ such that

$$\Psi(2t) \le C_\Psi \Psi(t)$$

for each $t \ge 0$. The smallest such constant is at least 2 by (3), and is called *the doubling constant of* Ψ. The doubling condition implies that

$$\Psi(at) \le C_\Psi a^{\log_2 C_\Psi} \Psi(t) \tag{4}$$

for all $t \ge 0$ and $a \ge 1$, [43, Lemma 2.8], and that Ψ is strictly increasing and continuous. Functions $\Psi_1(t) = at^p$, $a > 0$, $p \ge 1$, and $\Psi_2(t) = (1 + t)\log(1 + t) - t$ are examples of doubling functions, whereas the complementary function of Ψ_2, $\widetilde{\Psi}_2(t) = e^t - t - 1$ is not doubling.

Given Ψ and an open set $\Omega \subset X$, *the Orlicz space* $L^\Psi(\Omega)$ consists of measurable functions $u : \Omega \to [-\infty, \infty]$ for which

$$\int_\Omega \Psi(\alpha|u|)\, d\mu < \infty$$

for some $\alpha > 0$. If Ψ is doubling, then $L^{\Psi}(\Omega)$ coincides with the set of functions u for which $\int_{\Omega} \Psi(|u|)\,d\mu$ is finite. The space $L^{\Psi}(\Omega)$ is a Banach space with *the Luxemburg norm*,

$$\|u\|_{L^{\Psi}(\Omega)} = \inf \left\{ k > 0 : \int_{\Omega} \Psi\left(k^{-1}|u|\right) d\mu \leq 1 \right\}.$$

Using (3), it is easy to see that if $\|u\|_{L^{\Psi}(\Omega)} \leq 1$, then $\int_{\Omega} \Psi(|u|)\,d\mu \leq \|u\|_{L^{\Psi}(\Omega)}$, and if $\|u\|_{L^{\Psi}(\Omega)} \geq 1$, then $\int_{\Omega} \Psi(|u|)\,d\mu \geq \|u\|_{L^{\Psi}(\Omega)}$. Hence $\|u\|_{L^{\Psi}(\Omega)} \leq 1$ if and only if $\int_{\Omega} \Psi(|u|)\,d\mu \leq 1$, see also Lemma 4.

If $\Psi, \widetilde{\Psi}$ is a complementary pair, then

$$t \leq \Psi^{-1}(t)\widetilde{\Psi}^{-1}(t) \leq 2t \tag{5}$$

for all $0 \leq t \leq \infty$, and *the generalized Hölder inequality*

$$\int_{\Omega} |u(x)v(x)|\,d\mu \leq 2\|u\|_{L^{\Psi}(\Omega)}\|v\|_{L^{\tilde{\Psi}}(\Omega)} \tag{6}$$

holds for all $u \in L^{\Psi}(\Omega)$, $v \in L^{\widetilde{\Psi}}(\Omega)$. If Ψ is real-valued, then each $u \in L^{\Psi}(X)$ is locally integrable.

The maximal operator is bounded in $L^{\Psi}(X)$ if Ψ is doubling and $2C\Psi(t) \leq \Psi(Ct)$ for each $t \geq 0$ with a fixed constant $C > 1$, see [32, Theorem 1.2.2], [10,20]. For an N-function Ψ, the inequality above is equivalent to the doubling condition of $\widetilde{\Psi}$. If Ψ is doubling, the weak type estimate

$$\Psi(\lambda)\mu\left(\{x \in X : Mu(x) > \lambda\}\right) \leq C \int_{X} \Psi(|u|)\,d\mu \tag{7}$$

holds for all $u \in L^{\Psi}(X)$ and $\lambda > 0$. The constant $C > 0$ depends only on the doubling constants of the measure μ and the function Ψ, see [19, Theorem 6.2.1], [43, Lemma 6.18].

2.2 Orlicz–Sobolev spaces

We recall the definitions and some properties of Orlicz–Sobolev spaces defined using pairs of integrable functions and upper gradients in metric measure spaces. For proofs, see [43], and for discussion on upper gradients, also [25,41,42]. The spaces $N^{1,\Psi}(X)$ with N-function Ψ are studied also in [5].

A Borel measurable function $g \geq 0$ is *an upper gradient* of a function u *in an open set* $\Omega \subset X$ if

$$|u(x) - u(y)| \leq \int_{\gamma} g\,ds \tag{8}$$

for each pair of points x, y in Ω, and all rectifiable curves γ joining x and y in Ω. We require that if the right-hand side of (8) is finite, then also the left-hand side is finite and well defined.

The Sobolev space $N^{1,\Psi}(\Omega)$ consists of functions $u \in L^{\Psi}(\Omega)$ which have a Ψ-weak upper gradient $g \in L^{\Psi}(\Omega)$ in Ω. Being a Ψ-weak upper gradient in Ω means that inequality (8) holds for u and g except for a family of compact, rectifiable curves in Ω with zero Ψ-modulus. If there is no risk of confusion, Ψ-weak upper gradients are called weak upper gradients. For definition and properties of Ψ-modulus, see Sect. 4 and [43].

The space $N^{1,\Psi}(\Omega)$ is a Banach space with the norm

$$\|u\|_{N^{1,\Psi}(\Omega)} = \|u\|_{L^\Psi(\Omega)} + \inf \|g\|_{L^\Psi(\Omega)}, \tag{9}$$

where the infimum is taken over weak upper gradients, or, equivalently over upper gradients. If Ψ is doubling, then each $u \in N^{1,\Psi}(X)$ has *a minimal weak upper gradient* $g_u \in L^\Psi(X)$ which means that $\|g_u\|_{L^\Psi(X)}$ equals the infimum above.

Note that if $\Psi(t) = t^p$, $p \geq 1$, we obtain the Sobolev space $N^{1,p}(\Omega)$, defined by Shanmugalingam [41].

Density of Lipschitz (or continuous) functions in $N^{1,\Psi}(X)$ is an important property implied by the doubling condition of Ψ and a Poincaré inequality on X. Recall that X *supports a Poincaré inequality* if there exist constants $C_P > 0$ and $\sigma \geq 1$ such that

$$\fint_B |u - u_B|\, d\mu \leq C_P r \fint_{\sigma B} g\, d\mu$$

for each $u \in L^1_{\mathrm{loc}}(X)$ and every upper gradient g of u in each ball $B = B(x, r)$. Note that, if X is proper, the constant σ can be assumed to be 1. Namely, if X is a length space, which means that the distance between any two points is the infimum of the lengths of the curves connecting the points, then the constant σ can be taken to be 1 by enlarging C_P. In a proper space that supports a Poincaré inequality, the metric d can be replaced with a bi-Lipschitz equivalent length metric, and the space with the new metric also supports a Poincaré inequality, see [23, Chapter 9].

3 Capacity and quasicontinuity of $N^{1,\Psi}(X)$-functions

We use a capacity that is based on the norm (9); the Ψ*-capacity* of a set $E \subset X$ is

$$\mathrm{Cap}_\Psi(E) = \inf \left\{ \|u\|_{N^{1,\Psi}(X)} : u \in N^{1,\Psi}(X), u \geq 1 \text{ on } E \right\}. \tag{10}$$

The functions u in (10) are called test functions for $\mathrm{Cap}_\Psi(E)$. If there are no test functions, then we set $\mathrm{Cap}_\Psi(E) = \infty$. The Ψ-capacity is an outer measure, and the infimum in (10) is reached by using test functions which satisfy $0 \leq u \leq 1$, see [43, Chapter 7]. Note the difference between the Ψ-capacity and the p-capacity which is defined as the pth power of the $N^{1,p}(X)$-norm, see [41].

A function $u : X \to \overline{\mathbb{R}}$ is Ψ*-quasicontinuous* if for every $\varepsilon > 0$ there is an open set $U \subset X$ such that $\mathrm{Cap}_\Psi(U) < \varepsilon$ and $u|_{X \setminus U}$ is continuous. Usually, we omit the prefix Ψ.

Next we prove some capacity estimates. The first lemma provides a lower bound for the capacity of an arbitrary set of positive measure. The next one gives an estimate for the capacity of a ball.

Lemma 1 *Assume that Ψ is strictly increasing and $E \subset X$. If $\mathrm{Cap}_\Psi(E) = 0$, then $\mu(E) = 0$. If $\mu(E) > 0$, then*

$$\mu(E) \leq \left(\Psi\Big(\frac{1}{\mathrm{Cap}_\Psi(E)}\Big) \right)^{-1}. \tag{11}$$

Proof The first claim is proved in [43, Proposition 7.4]. Assume that $\mu(E) > 0$, $u \in N^{1,\Psi}(X)$ and $u|_E \geq 1$. Using (3), we have

$$\int_X \Psi\left(|u| \Psi^{-1}\left(\frac{1}{\mu(E)}\right)\right) d\mu \geq \int_E \Psi\left(|u| \Psi^{-1}\left(\frac{1}{\mu(E)}\right)\right) d\mu \geq \int_E \frac{1}{\mu(E)}\, d\mu = 1,$$

and hence $\|u\|_{L^\Psi(X)} \geq (\Psi^{-1}(\mu(E)^{-1}))^{-1}$. The claim follows because $\mathrm{Cap}_\Psi(E) \geq (\Psi^{-1}(\mu(E)^{-1}))^{-1}$ by the definition of Ψ-capacity.

For the next lemma, recall that if $E \subset X$ is of finite and positive measure and Ψ is continuous and $\Psi(t) = 0$ only if $t = 0$, then

$$\|\chi_E\|_{L^\Psi(X)} = \left(\Psi^{-1}\left(\frac{1}{\mu(E)}\right)\right)^{-1}. \tag{12}$$

Moreover, $\|u\|_{L^\Psi(A)} = \|u\chi_A\|_{L^\Psi(X)}$ for all measurable sets $A \subset X$.

Lemma 2 *If Ψ is doubling, then for each ball $B(x,r)$,*

$$\mathrm{Cap}_\Psi(B(x,r)) \leq 2C_\mu \left(\Psi^{-1}\left(\frac{1}{\mu(B(x,r))}\right)\right)^{-1} \max\left\{\frac{1}{r}, 1\right\}. \tag{13}$$

Proof For a ball $B = B(x,r)$, the function

$$u(y) = \begin{cases} (2r - d(y,x))/r, & \text{if } y \in 2B\setminus\overline{B}, \\ 1, & \text{if } y \in \overline{B}, \\ 0, & \text{if } y \in X\setminus 2B \end{cases}$$

is r^{-1}-Lipschitz and has an upper gradient $g_u = r^{-1}\chi_{2B\setminus\overline{B}}$. Since u is a test function for the capacity $\mathrm{Cap}_\Psi(B)$, using (12) and the doubling condition of μ, we have that

$$\begin{aligned} \mathrm{Cap}_\Psi(B) &\leq \|u\|_{N^{1,\Psi}(X)} \leq \|u\|_{L^\Psi(X)} + \|g_u\|_{L^\Psi(X)} \\ &\leq \left(1 + \frac{1}{r}\right)\|\chi_{2B}\|_{L^\Psi(X)} = \left(1 + \frac{1}{r}\right)\left(\Psi^{-1}\left(\frac{1}{\mu(2B)}\right)\right)^{-1} \\ &\leq \left(1 + \frac{1}{r}\right)\left(\Psi^{-1}\left(\frac{1}{C_\mu \mu(B)}\right)\right)^{-1} \leq \left(1 + \frac{1}{r}\right) C_\mu \left(\Psi^{-1}\left(\frac{1}{\mu(B)}\right)\right)^{-1}. \end{aligned} \tag{14}$$

The last inequality holds because $\Psi^{-1}(t)/t$ is decreasing and $C_\mu \geq 1$. The claim follows since the upper bound for $(1 + 1/r)$ is $2/r$ if $0 < r \leq 1$ and 2 if $r > 1$.

Proposition 1 *Assume that Ψ is doubling and that there are constants $C \geq 1$ and $Q > \log_2 C_\Psi$ such that $\mu(B(x,r)) \leq Cr^Q$ for each ball $B(x,r)$, $0 < r \leq 1$. Then*

$$\mathrm{Cap}_\Psi(E) \leq C\mathscr{H}^{(Q/\log_2 C_\Psi - 1)}(E)$$

for all sets $E \subset X$.

Proof Denote $\alpha = \log_2 C_\Psi$, where C_Ψ is the doubling constant of Ψ. Let $E \subset X$ and $B_i = B(x_i, r_i)$ be balls such that $E \subset \cup_i B_i$ and $r_i \leq 1$ for all i. By the subadditivity of the Ψ-capacity, Lemma 2 and the upper bound for the measure of balls B_i, we have that

$$\begin{aligned} \mathrm{Cap}_\Psi(E) &\leq \sum_{i=1}^\infty \mathrm{Cap}_\Psi(B_i) \leq 2C_\mu \sum_{i=1}^\infty r_i^{-1}\left(\Psi^{-1}\left(\frac{1}{\mu(B_i)}\right)\right)^{-1} \\ &\leq 2C_\mu \sum_{i=1}^\infty r_i^{-1}\left(\Psi^{-1}\left(\frac{1}{Cr_i^Q}\right)\right)^{-1} \leq C\sum_{i=1}^\infty r_i^{(Q/\alpha - 1)}. \end{aligned}$$

The last inequality follows from (4) and the fact that $\Psi^{-1}(t)/t$ is decreasing. Namely, since $\Psi(t^{1/\alpha}) \le C_\Psi t\Psi(1)$ for all $t \ge 1$ and $\Psi^{-1}(at) \ge a\Psi^{-1}(t)$ for $0 < a \le 1$, we have that

$$\Psi^{-1}(t) \ge \min\left\{1, \frac{1}{C_\Psi \Psi(1)}\right\} t^{1/\alpha}.$$

The claim follows from the definition of $\mathscr{H}^{Q/\alpha-1}(E)$.

Note that Proposition 1 implies, that a set of zero $(Q/\log_2 C_\Psi - 1)$-Hausdorff measure has zero Ψ-capacity.

Variational Orlicz capacity, where the infimum is taken over norms of upper gradients of compactly supported test functions and a connection with a Hausdorff measure on metric spaces are studied in [8]. For a connection between Orlicz capacity and Hausdorff measure in the classical case, see [15,36].

3.1 Quasicontinuity

Recently, in [7], Björn, Björn and Shanmugalingam showed that if X is proper, $\Omega \subset X$ is open and continuous functions are dense in $N^{1,p}(X)$, then each $u \in N^{1,p}(\Omega)$ is p-quasicontinuous. The main tool in the proof of p-quasicontinuity is the outer regularity property of sets of zero p-capacity [7, Proposition 1.4]. We will generalize the quasicontinuity property to spaces $N^{1,\Psi}(X)$. The proofs are modifications of the corresponding results in [7].

Theorem 1 *If X is proper, Ψ is doubling, $\Omega \subset X$ is open and continuous functions are dense in $N^{1,\Psi}(X)$, then each function $u \in N^{1,\Psi}(\Omega)$ is quasicontinuous.*

Lemma 3 *Let X be proper and let Ψ be doubling. If $E \subset X$ with $\operatorname{Cap}_\Psi(E) = 0$, then for each $\varepsilon > 0$ there is an open set U such that $E \subset U$ and $\operatorname{Cap}_\Psi(U) < \varepsilon$.*

Proof Let $E \subset X$ with $\operatorname{Cap}_\Psi(E) = 0$ and let $\varepsilon > 0$, $0 < \delta < 1$. Assume first that E is bounded. Since $\operatorname{Cap}_\Psi(E) = 0$, $\mu(E) = 0$ and the Ψ-modulus of the family of nonconstant, compact, rectifiable curves that intersect E is zero, see [43, Proposition 7.4]. Hence $\chi_E \in N^{1,\Psi}(X)$, (has the zero function as a weak upper gradient), and there is an upper gradient $g \in L^\Psi(X)$ of χ_E. Since Ψ is doubling, the function $\Psi(g)$ is in $L^1(X)$, and by the Vitali-Carathéodory theorem (see, e.g. [39, Theorem 2.25]), there is a lower semicontinuous function $h \in L^1(X)$ for which $0 \le \Psi(g) \le h$.

Now the function

$$\rho = \Psi^{-1}(h+1) \ge \Psi^{-1}(h)$$

belongs to $L^\Psi_{\mathrm{loc}}(X)$ and is lower semicontinuous. For the lower semicontinuity, note that Ψ^{-1} is continuous as the inverse function of the strictly increasing and continuous function. Moreover, the doubling condition of Ψ implies that $\rho \ge \Psi^{-1}(1) > 0$; this is needed in [7, Lemma 3.4].

By the Borel regularity of μ and the absolute continuity of the integral, there is a bounded open set V such that $E \subset V$ and

$$\mu(V) + \int_V \Psi(\rho)\, d\mu < \delta. \tag{15}$$

Define

$$u(x) = 2\min\left\{1, \inf \int_{\gamma} \rho\, ds\right\},$$

where the infimum is taken over all rectifiable curves, including constant ones, that connect x to $X \setminus V$.

Since the function $\inf \int_{\gamma} \rho\, ds$ is lower semicontinuous by [7, Lemma 3.4], u is lower semicontinuous as a minimum of two lower semicontinuous functions, and hence measurable. The function 2ρ is an upper gradient of u by [7, Lemma 3.2]. Moreover, as $u = 0$ on a closed set $X \setminus V$, we may use $2\rho\chi_V$ as an upper gradient of u.

As g is an upper gradient of χ_E and $\Psi(g) \le h$,

$$1 = \chi_E(x) - \chi_E(y) \le \int_{\gamma} g\, ds \le \int_{\gamma} \Psi^{-1}(h)\, ds \le \int_{\gamma} \rho\, ds$$

whenever γ is a curve that connects $x \in E$ and $y \in X \setminus V$. Hence $u = 2$ on E.

The set $U = \{x \in X : u(x) > 1\}$ is open by the lower semicontinuity of u. Since $u \in N^{1,\Psi}(X)$, it is a test function for the Ψ-capacity of U. Using (12), (15) and Lemma 4, we have that

$$\begin{aligned} \mathrm{Cap}_{\Psi}(U) &\le \|u\|_{N^{1,\Psi}(X)} \le \|u\|_{L^{\Psi}(X)} + \|2\rho\chi_V\|_{L^{\Psi}(X)} \\ &\le 2\|\chi_V\|_{L^{\Psi}(X)} + 2\|\rho\chi_V\|_{L^{\Psi}(X)} \\ &\le 2\left(\Psi^{-1}\left(\frac{1}{\mu(V)}\right)\right)^{-1} + 2\|\rho\|_{L^{\Psi}(V)} \\ &\le 2\left(\Psi^{-1}\left(\frac{1}{\delta}\right)\right)^{-1} + 2\,(C_{\Psi}\delta)^{1/\log_2 C_{\Psi}}\,. \end{aligned} \tag{16}$$

Since we also have that $E \subset U$, the claim follows by selecting δ so small that the last row of (16) is less than ε.

The case of unbounded E is proved as in [7] by writing E as a countable union of bounded sets and using the first part of the proof for these bounded sets.

By [7, Proposition 1.2], each function of $N^{1,p}(\Omega)$ has a p-quasicontinuous representative even without the assumption that X is proper. For $\Omega = X$, this result follows from [41]. An essential part of the proof is [41, Proof of Theorem 3.7], which shows that if a sequence of continuous functions converge to $v \in N^{1,p}(X)$ in $N^{1,p}(X)$, then a subsequence converge to a function of $N^{1,p}(X)$, uniformly outside open sets with arbitrarily small p-capacity and the limit function equals v p-quasi everywhere.

Note that, in [43], the result that $N^{1,\Psi}(X)$ is a Banach space is proved without capacity. However, the proof of [41, Theorem 3.7] works for $N^{1,\Psi}(X)$ with obvious modifications, (see also [5, Theorem 3.19]). Since also the remaining part of the proof of [7, Proposition 1.2] holds for $N^{1,\Psi}(\Omega)$, we see that each function $u \in N^{1,\Psi}(\Omega)$ has a quasicontinuous representative whenever Ψ is doubling and continuous functions are dense in $N^{1,\Psi}(X)$.

Proof of Theorem 1 Let $u \in N^{1,\Psi}(\Omega)$ and $\varepsilon > 0$. By the above discussion, u has a quasicontinuous representative u^* and the Ψ-capacity of the set $E = \{x \in \Omega : u(x) \ne u^*(x)\}$ is zero. By Lemma 3, there is an open set U such that $E \subset U$ and $\mathrm{Cap}_{\Psi}(U) < \varepsilon$. Since u^* is quasicontinuous, there is an open set V with $\mathrm{Cap}_{\Psi}(V) < \varepsilon$ such that $u^*|_{X\setminus V}$ is continuous. Since $u = u^*$ on $X \setminus (U \cup V)$ and $\mathrm{Cap}_{\Psi}(U \cup V) < 2\varepsilon$, u is quasicontinuous.

Using Theorem 1, we obtain important properties of Ψ-capacity.

Corollary 1 *Assume that X is proper, Ψ is doubling and continuous functions are dense in $N^{1,\Psi}(X)$. Then*

1. *the Ψ-capacity is an outer capacity;*

$$\mathrm{Cap}_\Psi(E) = \inf\left\{\mathrm{Cap}_\Psi(U) : U \text{ is open}, E \subset U\right\}$$

for all sets $E \subset X$.

2. *if (K_i) is a decreasing sequence of compact sets and $K = \cap_{i=1}^\infty K_i$, then*

$$\mathrm{Cap}_\Psi(K) = \lim_{i\to\infty} \mathrm{Cap}_\Psi(K_i).$$

Proof The proof of the outer regularity is similar to the proof of [7, Corollary 1.3], replace $N^{1,p}(X)$ with $N^{1,\Psi}(X)$ and the p-capacity with the Ψ-capacity.

Let then K_i, $i \in \mathbb{N}$, be compact sets such that $K_{i+1} \subset K_i$ for all i, $K = \cap_{i=1}^\infty K_i$, and let $\varepsilon > 0$. By the outer regularity, there is an open set U containing K such that $\mathrm{Cap}_\Psi(U) < \mathrm{Cap}_\Psi(K) + \varepsilon$. Since K is compact, $K_i \subset U$ for large i, and hence

$$\lim_{i\to\infty} \mathrm{Cap}_\Psi(K_i) \le \mathrm{Cap}_\Psi(U) < \mathrm{Cap}_\Psi(K) + \varepsilon.$$

By letting $\varepsilon \to 0$, we have $\lim_{i\to\infty} \mathrm{Cap}_\Psi(K_i) \le \mathrm{Cap}_\Psi(K)$.

The claim follows because $\mathrm{Cap}_\Psi(K) \le \lim_{i\to\infty} \mathrm{Cap}_\Psi(K_i)$ by the monotonicity of the Ψ-capacity.

4 Lemmas

The first lemma provides an inequality between integrals and Luxemburg norms. By this lemma, the Ψ-capacity and the Orlicz–Sobolev capacity of [36] have same null sets.

Lemma 4 *Let Ψ be doubling, $U \subset X$, $u \in L^\Psi(U)$, and $\|u\|_{L^\Psi(U)} > 0$.*
If $\int_U \Psi(|u|)\,d\mu \le 1$, then

$$\int_U \Psi(|u|)\,d\mu \le \|u\|_{L^\Psi(U)} \le \left(C_\Psi \int_U \Psi(|u|)\,d\mu\right)^{1/\log_2 C_\Psi}. \tag{17}$$

If $\int_U \Psi(|u|)\,d\mu \ge 1$, then

$$\left(\frac{1}{C_\Psi}\int_U \Psi(|u|)\,d\mu\right)^{1/\log_2 C_\Psi} \le \|u\|_{L^\Psi(U)} \le \int_U \Psi(|u|)\,d\mu. \tag{18}$$

Proof The first inequality of (17) and the second of (18) are easy, see Sect. 2.1. Denote $\|u\| = \|u\|_{L^\Psi(U)}$ and assume that $\|u\| \le 1$. Since Ψ is doubling, using (4) we have that

$$\int_U \Psi\left(\frac{|u|}{\|u\|}\right) = 1,$$

see [37, Proposition III.3.4.6]. Using the doubling condition of Ψ, we obtain

$$1 = \int_U \Psi\left(\frac{|u|}{\|u\|}\right) \le C_\Psi \left(\frac{1}{\|u\|}\right)^{\log_2 C_\Psi} \int_U \Psi(|u|)\,d\mu,$$

from which the second inequality of (17) follows.

Assume then that $\|u\| \ge 1$ and denote

$$a = \left(C_\Psi^{-1} \int_U \Psi(|u|)\,d\mu\right)^{1/\log_2 C_\Psi}.$$

If $\int_U \Psi(|u|)\,d\mu \ge C_\Psi$, then (4) gives

$$\int_U \Psi\left(\frac{|u|}{a}\right) \ge \frac{1}{C_\Psi} \frac{C_\Psi}{\int_U \Psi(|u|)\,d\mu} \int_U \Psi(|u|)\,d\mu = 1.$$

If $1 \le \int_U \Psi(|u|)\,d\mu \le C_\Psi$, then by (3)

$$\int_U \Psi\left(\frac{|u|}{a}\right) \ge \frac{1}{a} \int_U \Psi(|u|)\,d\mu \ge \int_U \Psi(|u|)\,d\mu \ge 1.$$

The first inequality of (18) follows from the definition of the Luxemburg norm.

The following lemmas, counterparts to results for generalized gradients in [22,29], hold for weak upper gradients defined either using the p- or Ψ-modulus, and hence for functions of both $N^{1,p}(X)$ and $N^{1,\Psi}(X)$. Recall that the Ψ-modulus of a curve family Γ is

$$\operatorname{Mod}_\Psi(\Gamma) = \inf \|\rho\|_{L^\Psi(X)},$$

where the infimum is taken over all non-negative Borel-functions ρ that satisfy $\int_\gamma \rho\,ds \ge 1$ for all locally rectifiable curves $\gamma \in \Gamma$, see [43]. The first lemma, which is essentially proved in [43, Lemma 6.15] and in [42, Lemma 2.14] for $N^{1,p}(X)$, is a version of the Leibniz differentiation rule.

Lemma 5 *If $u \in N^{1,\Psi}(X)$ and $\varphi : X \to \mathbb{R}$ is a bounded L-Lipschitz function, then $u\varphi \in N^{1,\Psi}(X)$. Moreover, if $g_u \in L^\Psi(X)$ is a weak upper gradient of u, then the function*

$$g_0 = (g_u\|\varphi\|_\infty + 4L|u|)\,\chi_{X\setminus F},$$

where $F = \{x \in X : \varphi(x) = 0\}$, is a weak upper gradient of $u\varphi$.

We will need the next lemma only in the case where the weak upper gradient sequence is a constant sequence.

Lemma 6 *Let $\Omega \subset X$ be an open set, and let (u_i) be a sequence of measurable functions with a corresponding sequence of weak upper gradients (g_i). If $u = \sup_i u_i$ is finite almost everywhere, then $g = \sup_i g_i$ is a weak upper gradient of u in Ω.*

Proof For each $i \in \mathbb{N}$, let Γ_i be the curve family in Ω with zero Ψ-modulus such that (8) holds for the pair u_i, g_i for all curves $\gamma \notin \Gamma_i$. By the subadditivity of Ψ-modulus, the Ψ-modulus of $\Gamma = \cup_i \Gamma_i$ is zero, too.

Let $\varepsilon > 0$ and let $\gamma \notin \Gamma$ be a curve with endpoints x and y. Assume first that $|u(x)|$ and $|u(y)|$ are finite. We may assume that $u(y) \le u(x) < \infty$. Let $i \in \mathbb{N}$ such that $u_i(x) + \varepsilon > u(x)$. Since $u(y) \ge u_i(y)$, g_i is a weak upper gradient of u_i, and $g \ge g_i$, we have that

$$\begin{aligned}|u(x) - u(y)| &= u(x) - u(y) < u_i(x) + \varepsilon - u_i(y)\\ &\le \int_\gamma g_i \, ds + \varepsilon \le \int_\gamma g \, ds + \varepsilon.\end{aligned}$$

The claim in the case where u is finite in the endpoints of γ follows by letting $\varepsilon \to 0$.

For the general case, let Γ_∞ be a family of curves $\gamma \subset \Omega$ for which $|u| = \infty$ on some subcurve of γ, and let

$$A = \{z \in \Omega : |u(z)| = \infty\}, \quad h = \infty \chi_A.$$

Since u is finite almost everywhere in Ω, $\mu(A) = 0$ and $h \in L^\Psi(\Omega)$. As $\int_\gamma h \, ds = \infty$ for all curves $\gamma \in \Gamma_\infty$, Fuglede's characterization [17, Theorem 2], [43, Lemma 3.3], implies that $\mathrm{Mod}_\Psi(\Gamma_\infty) = 0$. By the subadditivity of modulus, $\mathrm{Mod}_\Psi(\Gamma \cup \Gamma_\infty) = 0$, too.

Let now $\gamma \notin \Gamma \cup \Gamma_\infty$ with endpoints x and y. We may assume that $u(x) = \infty$ and $|u(y)| < \infty$, since otherwise, by the selection of Γ_∞, there is a point $z \in \gamma$ such that $|u(z)| < \infty$, and we would estimate subcurves connecting x to z and z to y separately. Let $u_i(x)$ be a subsequence such that $u_i(x) \to u(x)$ as $i \to \infty$. By the selection of the subsequence, there is $i_0 \in \mathbb{N}$ such that $u_i(x) > u(y)$ as $i \ge i_0$. Then, for $i \ge i_0$, we have that

$$|u_i(x) - u(y)| = u_i(x) - u(y) \le u_i(x) - u_i(y) \le \int_\gamma g_i \, ds \le \int_\gamma g \, ds.$$

The claim follows by letting $i \to \infty$.

Lemma 7 *Let u be a measurable function with a weak upper gradient g. If $h\colon X \to [0, \infty]$ is a Borel-function such that $g \le h$ almost everywhere, then h is a weak upper gradient of u.*

Proof Let Γ_0 be the curve family with zero Ψ-modulus such that (8) holds for u and g for every curve $\gamma \notin \Gamma_0$. Moreover, let Γ_1 be a family of non-constant curves γ for which $\int_\gamma g \, ds > \int_\gamma h \, ds$. Then, for each $\gamma \notin \Gamma_0 \cup \Gamma_1$ with endpoints x and y, we have that

$$|u(x) - u(y)| \le \int_\gamma g \, ds \le \int_\gamma h \, ds.$$

Hence, by the subadditivity of modulus, it suffices to show that $\mathrm{Mod}_\Psi(\Gamma_1) = 0$. This follows from [43, Lemma 3.4] because the set $E = \{x : g(x) > h(x)\}$ has zero measure, and $\mathscr{H}^1(\gamma \cap E) > 0$ for each $\gamma \in \Gamma_1$.

5 Lebesgue points

In this section, we will show that the discrete maximal operator, defined by Kinnunen and Latvala [29], is bounded in $N^{1,\Psi}(X)$. This result is then used to show that almost all points, in the Ψ-capacity sense, are Lebesgue points of Orlicz–Sobolev functions.

Let $r > 0$, and let $\{B_i\}_{i=1}^\infty$ be a covering of X by balls of radius r such that $\sum_{i=1}^\infty \chi_{6B_i}(x) \le N$ for all $x \in X$. Let (φ_i) be a partition of unity for the covering $\{B_i\}$ such that $\sum_i \varphi_i(x) = 1$ for all $x \in X$, $0 \le \varphi_i \le 1$ in X, $\varphi_i \ge C$ in $3B_i$, $\operatorname{supp} \varphi_i \subset 6B_i$, and that each φ_i is

L/r-Lipschitz, see for example [12,40]. All constants of $\{B_i\}$ and (φ_i) depend only on the doubling constant of μ. The discrete convolution of u is

$$u_r(x) = \sum_{i=1}^{\infty} \varphi_i(x) u_{3B_i}, \quad x \in X.$$

Below, we will show that if $u \in N^{1,\Psi}(X)$, then both $|u|_r$ and $\mathscr{M}^* u$ (see 19) are in $N^{1,\Psi}(X)$, in particular, the operator $\mathscr{M}^*$ is bounded in the Orlicz–Sobolev space.

Since we are going to use Lemma 6 for functions $|u|_r$, we numerate the positive rationals and choose for each radius r_j a covering $\{B_i^j\}$ consisting of balls of radius r_j, and a corresponding partition of unity as above.

The discrete maximal function of $u \in L^1_{\mathrm{loc}}(X)$ related to coverings $\{B_i^j\}$ is

$$\mathscr{M}^* u(x) = \sup_j |u|_{r_j}(x), \quad x \in X. \tag{19}$$

The maximal operator $\mathscr{M}^*$ depends on the covering but the estimates below are independent on the covering.

By [29, Lemma 3.1], there is a constant $C = C(C_\mu)$ such that

$$C^{-1} \mathscr{M} u(x) \le \mathscr{M}^* u(x) \le C \mathscr{M} u(x) \tag{20}$$

for each $u \in L^1_{\mathrm{loc}}(X)$ and all $x \in X$. Hence, if the maximal operator is bounded in $L^\Psi(X)$, then

$$\|\mathscr{M}^* u\|_{L^\Psi(X)} \le C \|\mathscr{M} u\|_{L^\Psi(X)} \le C \|u\|_{L^\Psi(X)} \tag{21}$$

with $C = C(C_\mu, \Psi)$ for all $u \in L^\Psi(X)$.

The following two proofs are easy modifications of the proofs of [29, Lemma 3.3, Theorem 3.6]. For the readers convenience, we recall the main details of the proof.

Lemma 8 *Assume that X supports a Poincaré inequality, Ψ is doubling and the maximal operator $\mathscr{M}$ is bounded in $L^\Psi(X)$. If $u \in N^{1,\Psi}(X)$, then $|u|_r \in N^{1,\Psi}(X)$ and $C\mathscr{M} g$ is a weak upper gradient of $|u|_r$ whenever $g \in L^\Psi(X)$ is a weak upper gradient of u. Moreover, $\||u|_r\|_{N^{1,\Psi}(X)} \le C\|u\|_{N^{1,\Psi}(X)}$.*

Proof Let $u \in N^{1,\Psi}(X)$ and let $g \in L^\Psi(X)$ be a weak upper gradient of u. Then $|u| \in N^{1,\Psi}(X)$ and g is a weak upper gradient of $|u|$.

A weak upper gradient of $|u|_r$: Since $\sum_i \varphi_i(x) = 1$, we have for each $x \in X$ that

$$|u|_r(x) = \sum_{i=1}^{\infty} \varphi_i(x)|u|_{3B_i} = |u(x)| + \sum_{i=1}^{\infty} \varphi_i(x)(|u|_{3B_i} - |u(x)|),$$

where the sum is over finitely many terms only by the bounded overlap of balls $6B_i$.

Now the function $g + \sum_{i=1}^{\infty} g_i$, where g_i is a weak upper gradient of the function $\varphi_i(|u|_{3B_i} - |u|)$, is a weak upper gradient of $|u|_r$. By Lemma 5 and the properties of the functions φ_i,

$$(g + Cr^{-1}||u| - |u|_{3B_i}|)\chi_{6B_i}$$

is a weak upper gradient of $\varphi_i(|u|_{3B_i} - |u|)$. Hence, by Lemma 7, it suffices to estimate $||u| - |u|_{3B_i}|$. As in [29, Lemma 3.3], a standard telescoping argument, the doubling property of μ and the Poincaré inequality show that

$$||u(x)| - |u|_{3B_i}| \le Cr \mathscr{M} g(x)$$

for almost all $x \in X$, and hence we can select $g_i = (g + C\mathscr{M} g)\chi_{6B_i}$. Using the bounded overlap of the balls $6B_i$, and the fact that $g(x) \le \mathscr{M} g(x)$ for almost every x by the Lebesgue differentiation theorem, we conclude that the function $g_{u_r} = C\mathscr{M} g$ is a weak upper gradient of $|u|_r$.

Norm estimate: Since $|u|_r(x) \le \mathscr{M}^* u(x)$ for all $x \in X$, the boundedness of the maximal operator and (21) imply that $\||u|_r\|_{L^\Psi(X)} \le C\|u\|_{L^\Psi(X)}$. Similarly, $\|g_{u_r}\|_{L^\Psi(X)} \le C\|g\|_{L^\Psi(X)}$ by the boundedness of the maximal operator. The claim follows by choosing g to be the minimal weak upper gradient of u.

Theorem 2 *Assume that X supports a Poincaré inequality, Ψ is doubling and the maximal operator $\mathscr{M}$ is bounded in $L^\Psi(X)$. If $u \in N^{1,\Psi}(X)$, then $\mathscr{M}^* u$ belongs to $N^{1,\Psi}(X)$ and* $\|\mathscr{M}^* u\|_{N^{1,\Psi}(X)} \le C\|u\|_{N^{1,\Psi}(X)}$.

Proof The claim $\mathscr{M}^* u \in L^\Psi(X)$ follows from (21), and hence $\mathscr{M}^* u$ is finite almost everywhere. Since $C\mathscr{M} g \in L^\Psi(X)$ is a weak upper gradient of $|u|_{r_j}$ for all $j \in \mathbb{N}$, $\mathscr{M}^* u \in N^{1,\Psi}(X)$ with a weak upper gradient $C\mathscr{M} g$ by Lemmas 6 and 8.

As in the proof of Lemma 8, we obtain the desired norm estimate by choosing the minimal weak upper gradient g of u and combining L^Ψ-norm estimates for $\mathscr{M}^* u$ and $C\mathscr{M} g$.

Note that similar arguments as the proof of Lemma 8 show that $u_r(x) \to u(x)$ as $r \to 0$ for almost every $x \in X$, and that $u_r \to u$ in $L^\Psi(X)$. Namely,

$$|u_r(x) - u(x)| \le \sum_{i=1}^{\infty} \varphi_i(x)|u(x) - u_{3B_i}| \le \sum_{i'} |u(x) - u_{3B_i}| \le Cr\mathscr{M} g(x), \tag{22}$$

where the last sum is taken over such indices i' for which $x \in 6B_i$. The right-hand side of (22) tends to zero as $r \to 0$ for almost every $x \in X$ because $\mathscr{M} g \in L^\Psi(X)$.

Theorem 3 *Assume that X is proper and supports a Poincaré inequality, Ψ is doubling and the maximal operator is bounded in $L^\Psi(X)$. Then for each $u \in N^{1,\Psi}(X)$ there is a set $E \subset X$ with* $\mathrm{Cap}_\Psi(E) = 0$ *such that*

$$\lim_{r\to 0} \fint_{B(x,r)} |u - u(x)|\, d\mu = 0 \tag{23}$$

for all $x \in X \setminus E$.

We will use the following modification of a part of [24, Theorem 2.11] to prove Theorem 3.

Lemma 9 *Assume that X is proper, Ψ is doubling, continuous functions are dense in $N^{1,\Psi}(X)$, and that there exists a constant $C > 0$ such that*

$$\mathrm{Cap}_\Psi\left(\left\{x \in X : \limsup_{r\to 0} \fint_{B(x,r)} |u|\, d\mu > \lambda\right\}\right) \le C\lambda^{-1}\|u\|_{N^{1,\Psi}(X)} \tag{24}$$

for every $\lambda > 0$ and all $u \in N^{1,\Psi}(X)$. Then for each $u \in N^{1,\Psi}(X)$ there is a set E with $\mathrm{Cap}_\Psi(E) = 0$ *such that (23) holds for all $x \in X \setminus E$.*

Proof Let $u \in N^{1,\Psi}(X)$, $\lambda > 0$ and $\varepsilon > 0$. It suffices to show that

$$\limsup_{r\to 0} \fint_{B(x,r)} |u(y) - u(x)|\, d\mu(y) = 0$$

outside a set of capacity zero. Define

$$E_\lambda = \left\{ x : \limsup_{r\to 0} \fint_{B(x,r)} |u - u(x)|\, d\mu > \lambda \right\}.$$

Let $v \in N^{1,\Psi}(X)$ be a continuous function such that $\|u - v\|_{N^{1,\Psi}(X)} < \varepsilon$. Since u is quasicontinuous by Theorem 1, the function $w = u - v$ belongs to $N^{1,\Psi}(X)$ and is quasicontinuous. Since $u = v + w$, the estimate

$$|u(y) - u(x)| \le |v(x) - v(y)| + |w(x) - w(y)|$$

and the continuity of v imply that

$$\begin{aligned} \limsup_{r\to 0} \fint_{B(x,r)} |u - u(x)|\, d\mu &\le \limsup_{r\to 0} \fint_{B(x,r)} |w - w(x)|\, d\mu \\ &\le \limsup_{r\to 0} \fint_{B(x,r)} |w|\, d\mu + |w(x)| \end{aligned} \tag{25}$$

for all $x \in X$. Hence

$$\begin{aligned} \mathrm{Cap}_\Psi(E_\lambda) \le \mathrm{Cap}_\Psi &\left(\left\{ x : \limsup_{r\to 0} \fint_{B(x,r)} |w|\, d\mu > \lambda/2 \right\} \right) \\ &+ \mathrm{Cap}_\Psi(\{x : |w(x)| > \lambda/2\}), \end{aligned} \tag{26}$$

where assumption (24) and the fact that $\||w|\|_{N^{1,\Psi}(X)} \le \|w\|_{N^{1,\Psi}(X)}$ imply that

$$\mathrm{Cap}_\Psi \left(\left\{ x : \limsup_{r\to 0} \fint_{B(x,r)} |w|\, d\mu > \lambda/2 \right\} \right) \le C\lambda^{-1} \|w\|_{N^{1,\Psi}(X)}.$$

Moreover, since $2\lambda^{-1}|w|$ is a test function for the last capacity of (26), we have

$$\mathrm{Cap}_\Psi(\{x : |w(x)| > \lambda/2\}) \le 2\lambda^{-1} \|w\|_{N^{1,\Psi}(X)},$$

and hence

$$\mathrm{Cap}_\Psi(E_\lambda) \le C\lambda^{-1} \|w\|_{N^{1,\Psi}(X)} \le C\lambda^{-1}\varepsilon.$$

Letting $\varepsilon \to 0$, we conclude that $\mathrm{Cap}_\Psi(E_\lambda) = 0$ for all $\lambda > 0$. By selecting $\lambda = 1/i$ for all $i \in \mathbb{N}$, we have

$$E = \left\{ x : \limsup_{r\to 0} \fint_{B(x,r)} |u - u(x)|\, d\mu > 0 \right\} = \bigcup_{i=1}^{\infty} E_{1/i},$$

where $\operatorname{Cap}_\Psi(E_{1/i}) = 0$ for all i. Hence $\operatorname{Cap}_\Psi(E) = 0$ by the subadditivity of capacity, and the claim follows.

Proof of Theorem 3 By Theorem 9, it suffices to show that there is a constant $C > 0$ such that

$$\operatorname{Cap}_\Psi\left(\left\{x \in X : \limsup_{r\to 0} \fint_{B(x,r)} |u|\, d\mu > \lambda\right\}\right) \le C\lambda^{-1}\|u\|_{N^{1,\Psi}(X)}$$

for all $\lambda > 0$ and every $u \in N^{1,\Psi}(X)$. Since $\limsup_{r\to 0} \fint_{B(x,r)} |u|\, d\mu \le \mathcal{M} u(x)$ for all $x \in X$, it is enough to show that

$$\operatorname{Cap}_\Psi(\{x \in X : \mathcal{M} u(x) > \lambda\}) \le C\lambda^{-1}\|u\|_{N^{1,\Psi}(X)} \tag{27}$$

for all such u and λ.

For weak type estimate (27), let $u \in N^{1,\Psi}(X)$, $\lambda > 0$, and let $\mathcal{M}^* u$ be the discrete maximal function of u. By (20),

$$F = \{x \in X : \mathcal{M} u(x) > \lambda\} \subset \left\{x \in X : C\,\mathcal{M}^* u(x) > \lambda\right\} = E.$$

By Theorem 2, $C\lambda^{-1}\mathcal{M}^* u \in N^{1,\Psi}(X)$ is a test function for $\operatorname{Cap}_\Psi(E)$. Hence

$$\operatorname{Cap}_\Psi(F) \le \operatorname{Cap}_\Psi(E) \le C\lambda^{-1}\|\mathcal{M}^* u\|_{N^{1,\Psi}(X)} \le C\lambda^{-1}\|u\|_{N^{1,\Psi}(X)},$$

from which the theorem follows.

Remark 1 1. With the assumptions of Theorem 3, $\lim_{r\to 0} \fint_{B(x,r)} u\, d\mu = u(x)$ outside of a set of capacity zero.
2. If $W^{1,\Psi}(\mathbb{R}^n)$ is reflexive, then the proofs of the main results of [28,30] generalize to $W^{1,\Psi}(\mathbb{R}^n)$; both the Hardy–Littlewood maximal operator and its local version are bounded in the Orlicz–Sobolev space. This is studied in [16]. Is the Hardy–Littlewood maximal operator bounded also in $N^{1,\Psi}(X)$?
3. The results above imply that functions of $N^{1,p}(X)$ have Lebesgue points outside of a set of zero p-capacity if $p > 1$ and the doubling space X supports a Poincaré inequality. However, this result follows also from [29] since $N^{1,p}(X)$ and the Hajłasz–Sobolev space $M^{1,p}(X)$ (cf. [21]) studied in [29] are isomorphic as Banach spaces by [42].

6 Hölder quasicontinuity

In this section, we show that Hölder continuous functions are dense in $N^{1,\Psi}(X)$ both in norm and the Lusin sense.

We need two maximal functions. Let $0 \le \alpha < \infty$, $0 < \beta < \infty$, $R > 0$, and $u \in L^1_{\mathrm{loc}}(X)$. The (restricted) *fractional maximal function* of u is

$$\mathcal{M}_{\alpha,R}\, u(x) = \sup_{0<r\le R} r^\alpha \fint_{B(x,r)} |u|\, d\mu.$$

If $R = \infty$, we let $\mathcal{M}_\alpha u = \mathcal{M}_{\alpha,\infty} u$. If $\alpha = 0$, the we obtain the usual (restricted) Hardy–Littlewood maximal function. The (restricted) *fractional sharp maximal function* of u is

$$u^\#_{\beta,R}(x) = \sup_{0<r\le R} r^{-\beta} \fint_{B(x,r)} |u - u_{B(x,r)}|\, d\mu,$$

and $u^\#_\beta = u^\#_{\beta,\infty}$.

We cannot prove Theorem 5 directly for the space $N^{1,\Psi}(X)$. Hence we need an other definition of Orlicz–Sobolev spaces on metric spaces. This definition is a generalization of $M^{1,p}(X)$ and is based on a pointwise inequality.

A measurable function $g \geq 0$ is *a generalized gradient* of a measurable function u, $g \in \mathrm{D}(u)$, if there is a set $E \subset X$ with $\mu(E) = 0$ such that

$$|u(x) - u(y)| \leq d(x, y)\,(g(x) + g(y)) \tag{28}$$

for all $x, y \in X \setminus E$. Given a Young function Ψ, the Orlicz–Sobolev space $M^{1,\Psi}(X)$ consists of functions $u \in L^{\Psi}(X)$ for which there exists a function $g \in L^{\Psi}(X) \cap \mathrm{D}(u)$. The space $M^{1,\Psi}(X)$, equipped with the norm

$$\|u\|_{M^{1,\Psi}(X)} = \|u\|_{L^{\Psi}(X)} + \inf \|g\|_{L^{\Psi}(X)}, \tag{29}$$

where the infimum is taken over all functions $g \in L^{\Psi}(X) \cap \mathrm{D}(u)$, is a Banach space, see [4, Theorem 3.6].

Next we show that if Ψ and the space X are nice enough, then the definitions of $N^{1,\Psi}(X)$ and $M^{1,\Psi}(X)$ give the same space.

Theorem 4 *Assume that Ψ is a doubling N-function, the maximal operator is bounded in $L^{\Psi}(X)$ and that X supports a Poincaré inequality. Then $N^{1,\Psi}(X) = M^{1,\Psi}(X)$ and the norms are comparable.*

Proof By [43, Theorem 6.22] (see also [41, Chap. 4]), for each $u \in M^{1,\Psi}(X)$ there is a representative (u itself if continuous) which belongs to $N^{1,\Psi}(X)$. Moreover, if $g \in \mathrm{D}(u) \cap L^{\Psi}(X)$, then $2g$ is a weak upper gradient of u. Hence $M^{1,\Psi}(X) \subset N^{1,\Psi}(X)$ and $\|u\|_{N^{1,\Psi}(X)} \leq 2\|u\|_{M^{1,\Psi}(X)}$.

For the other direction, let $u \in N^{1,\Psi}(X)$ with a weak upper gradient $g \in L^{\Psi}(X)$. By [22, Lemma 3.6], the inequality

$$|u(x) - u(y)| \leq C d(x, y) \left(u_1^{\#}(x) + u_1^{\#}(y)\right)$$

holds for almost every $x, y \in X$. Since the Poincaré inequality implies that

$$u_1^{\#}(x) \leq C \mathscr{M} g(x)$$

for all $x \in X$, we have that $C \mathscr{M} g \in \mathrm{D}(u)$. The claim follows from the boundedness of $\mathscr{M}$.

In this section, we use the representative $\tilde{u}$,

$$\tilde{u}(x) = \limsup_{r \to 0} \fint_{B(x,r)} u \, d\mu \tag{30}$$

for $u \in N^{1,\Psi}(X)$ and denote it by u. By Theorem 3, the limit of the right-hand side of (30) exists and equals $u(x)$, except on a set of Ψ-capacity zero. Moreover, since the inequality

$$|u(x) - u(y)| \leq C d(x, y)^{\beta} \left(u^{\#}_{\beta,4d(x,y)}(x) + u^{\#}_{\beta,4d(x,y)}(y)\right) \tag{31}$$

holds for every $x, y \in X$ and for all $0 < \beta \leq 1$ by [22, Lemma 3.6], we see that u is Hölder continuous with exponent β if $\|u^{\#}_{\beta}\|_{\infty} < \infty$.

Theorem 5 *Assume that X is proper, Ψ is a doubling N-function, the maximal operator $\mathscr{M}$ is bounded in $L^{\Psi}(X)$ and that X supports a Poincaré inequality. If $u \in N^{1,\Psi}(X)$, then for any $0 < \beta \leq 1$ and each $\varepsilon > 0$, there is an open set Ω and a function $v \in N^{1,\Psi}(X)$ such that*

1. *$u = v$ in $X \setminus \Omega$,*
2. *v is Hölder continuous with exponent β on every bounded set of X,*
3. $\|u - v\|_{N^{1,\Psi}(X)} < \varepsilon$,
4. $\mathscr{H}_\infty^{s-(1-\beta)}(\Omega) < \varepsilon$*, where s is the doubling dimension of μ.*

In the proof, we first assume that u vanishes outside a ball. The general case follows by using a localization argument. We will correct the function in "the bad set", where the fractional sharp maximal function is big, using a discrete convolution. That kind of smoothing technique is used to prove corresponding approximation results for Sobolev functions on metric measure spaces in [22, Theorem 5.3] and [31, Theorem 5].

For the bad set, we will use a Whitney type covering from [12, Theorem III.1.3], [34, Lemma 2.9]. For an open set $U \subset X$, then there are balls $B_i = B(x_i, r_i)$, $i \in \mathbb{N}$, where $r_i = \operatorname{dist}(x_i, X \setminus U)/10$, such that

1. the balls $B(x_i, r_i/5)$ are pairwise disjoint,
2. $U = \cup_i B(x_i, r_i)$,
3. $B(x_i, 5r_i) \subset U$,
4. if $x \in B(x_i, 5r_i)$, then $5r_i \le \operatorname{dist}(x, X \setminus U) \le 15r_i$,
5. there is $x_i^* \in X \setminus U$ such that $d(x_i, x_i^*) < 15r_i$, and
6. $\sum_{i=1}^{\infty} \chi_{B(x_i,5r_i)}(x) \le M$ for all $x \in U$.

We need the following technical lemma for the Whitney covering. We omit the proof which consists of simple calculations using the properties of the Whitney covering and the doubling property of μ. All constants depend only on the constants of the Whitney covering, or on the doubling constant of μ.

Lemma 10 *Let $\mathscr{B} = \{B_i\}$ be a Whitney covering of an open set U. Let $x \in B_{i_0}$, $y \in B_{i_1}$, where $B_{i_0}, B_{i_1} \in \mathscr{B}$, and $\delta = 1/4 \max\{\operatorname{dist}(x, X \setminus U), \operatorname{dist}(y, X \setminus U)\}$.*

1. *If $x \in 2B_i$, then $2/3 r_i \le r_{i_0} \le 3/2 r_i$.*
2. *Let $y \in 2B_i$ and $d(x, y) \le \delta$. If $\operatorname{dist}(y, X \setminus U) \le \operatorname{dist}(x, X \setminus U)$, then $y \in 5B_{i_0}$ and $1/2 r_i \le r_{i_0} \le 3r_i$. Otherwise $x \in 5B_{i_1}$ and $2/3 r_i \le r_{i_1} \le 3/2 r_i$. In both cases, $r_i \approx r_{i_0} \approx \operatorname{dist}(x, X \setminus U)$.*
3. *If x or y is in $2B_i$ and $d(x, y) \le \delta$, then*

$$2B_i \subset B(x, C_1 r_i) \subset B(x, C_2 r_{i_0}) \subset B(x, C_3 \operatorname{dist}(x, X \setminus U))$$

and $d(x, y) \le C_4 r_i$. Moreover, $2B_i \subset B(x_{i_0}^*, C_5 r_i)$.

Proof of Theorem 5 Let $u \in N^{1,\Psi}(X)$ with a weak upper gradient $g \in L^\Psi(x)$.
Step 1 Suppose that the support of u is in $B(x_0, 1)$ for some $x_0 \in X$. Let $\lambda > 0$, and denote

$$E_\lambda = \left\{x \in X : u_\beta^\#(x) > \lambda\right\}.$$

It is easy to show that E_λ is open. By (31), u is Hölder continuous with exponent β in $X \setminus E_\lambda$. We will correct the values of u in the bad set E_λ. For that, let $\mathscr{B} = \{B_i\}$ be a Whitney covering of E_λ, and let (φ_i) be a partition of unity for which $\operatorname{supp} \varphi_i \subset 2B_i$,$0 \le \varphi_i \le 1$, each φ_i is K/r_i-Lipschitz, and $\sum_{i=1}^{\infty} \varphi_i(x) = \chi_{E_\lambda}(x)$, see for example [34, Lemma 2.16]. For each x_i, let x_i^* be the "closest" point in $X \setminus E_\lambda$ given by 5.

We begin with the properties of the set E_λ.

Claim 1 There is $\lambda_0 > 0$ such that $E_\lambda \subset B(x_0, 2)$ for each $\lambda > \lambda_0$.

Proof of Claim 1 Since $\operatorname{supp} u \subset B(x_0, 1)$, it is enough to show that there is $\lambda_0 > 0$ such that

$$r^{-\beta} \fint_{B(x,r)} |u - u_{B(x,r)}| \, d\mu < \lambda_0 \tag{32}$$

for all $x \in X$ and $r > 1$. If $B = B(x,r)$, $r > 1$ and $r^{-\beta} \fint_B |u - u_B| \, d\mu = a > 0$, then $B \cap B(x_0, 1) \neq \emptyset$ because $\operatorname{supp} u \subset B(x_0, 1)$. Using the doubling property of μ, the Hölder inequality (6) and inequality (12), we obtain

$$\begin{aligned} r^{-\beta} \fint_B |u - u_B| \, d\mu &\le 2C_\mu^2 \mu(B(x_0, 1))^{-1} \int_{B \cap B(x_0,1)} |u| \, d\mu \\ &\le C\mu(B(x_0, 1))^{-1} \|u\|_{L^\Psi(X)} \left(\widetilde{\Psi}^{-1}(\mu(B(x_0, 1))^{-1}) \right)^{-1}, \end{aligned}$$

from which Claim 1 follows. Note that $\widetilde{\Psi}^{-1}(t) > 0$ for $t > 0$ by (5) and the doubling property of Ψ.

Claim 2 $\mu(E_\lambda) \to 0$ as $\lambda \to \infty$.

Proof It easily follows from the Poincaré inequality that

$$u^\#_{\beta,R}(x) \le C \mathscr{M}_{1-\beta,R}\, g(x)$$

for all $x \in X$ and $R > 0$. Moreover, if $\alpha \le 1$ and $R \ge 1$, then $\mathscr{M}_{\alpha,R}\, g(x) \le R \mathscr{M} g(x)$. Hence, if $x \in E_\lambda$ and $\lambda > \lambda_0$, then by (32),

$$u^\#_\beta(x) = u^\#_{\beta,1}(x) \le C \mathscr{M}_{(1-\beta),1}\, g(x) \le C \mathscr{M} g(x). \tag{33}$$

Now the weak type estimate (7) for the maximal operator implies that

$$\mu(E_\lambda) \le \mu\left(\{x \in B(x_0, 2) : \mathscr{M} g(x) > C\lambda\}\right) \le C(\Psi(\lambda))^{-1} \int_X \Psi(g) \, d\mu.$$

Claim 2 follows because $g \in L^\Psi(X)$ and Ψ is doubling.

We define the function v as a Whitney type extension of u to the set E_λ by setting

$$v(x) = \begin{cases} u(x), & \text{if } x \in X \setminus E_\lambda, \\ \sum_{i=1}^\infty \varphi_i(x) u_{2B_i}, & \text{if } x \in E_\lambda. \end{cases}$$

We will select the open set Ω to be E_λ for sufficiently large $\lambda > \lambda_0$. Hence claim 1 of Theorem 5 follows from the definition of v. Since $\operatorname{supp} u \subset B(x_0, 1)$ and $E_\lambda \subset B(x_0, 2)$ for $\lambda > \lambda_0$, the support of v is in $B(x_0, 2)$.

Proof (Proof of 2—the Hölder continuity of v) We begin with an estimate for $|v(x) - v(\bar{x})|$, where $x \in E_\lambda$ and $\bar{x} \in X \setminus E_\lambda$ is such that $d(x, \bar{x}) \le 2 \operatorname{dist}(x, X \setminus E_\lambda)$. Denote

$$\mathscr{B}_x = \{B_i \in \mathscr{B} : x \in 2B_i\},$$

and note that, by the bounded overlap of the balls $2B_i$, there is a bounded number of balls in $\mathscr{B}_x$. Using the properties of the functions φ_i and the Whitney covering, we have that

$$|v(x) - v(\bar{x})| = \left| \sum_{i=1}^\infty \varphi_i(x)(u(\bar{x}) - u_{2B_i}) \right| \le \sum_{\mathscr{B}_x} |u(\bar{x}) - u_{2B_i}|, \tag{34}$$

and that $2B_i \subset B(\bar{x}, Cr_i)$. Now

$$|u(\bar{x}) - u_{2B_i}| \le |u(\bar{x}) - u_{B(\bar{x},Cr_i)}| + |u_{B(\bar{x},Cr_i)} - u_{2B_i}|, \tag{35}$$

where, for the first term on the right-hand side a telescoping argument shows that

$$|u(\bar{x}) - u_{B(\bar{x},Cr_i)}| \le Cr_i^{\beta} u^{\#}_{\beta,Cr_i}(\bar{x}). \tag{36}$$

For the second term, the fact that $B(\bar{x}, Cr_i) \subset CB_i$ and the doubling property of μ imply that

$$|u_{B(\bar{x},Cr_i)} - u_{2B_i}| \le C \fint_{B(\bar{x},Cr_i)} |u - u_{B(\bar{x},Cr_i)}|\, d\mu \le Cr_i^{\beta} u^{\#}_{\beta,Cr_i}(\bar{x}). \tag{37}$$

Since $r_i \approx \operatorname{dist}(x, X \setminus E_\lambda)$ by the properties of the Whitney covering, estimates (34)–(37) show that

$$|v(x) - v(\bar{x})| \le C \operatorname{dist}(x, X \setminus E_\lambda)^{\beta} u^{\#}_{\beta}(\bar{x}) \le C \operatorname{dist}(x, X \setminus E_\lambda)^{\beta} \lambda, \tag{38}$$

where the last inequality follows because $\bar{x} \in X \setminus E_\lambda$.

We will show that v is β-Hölder continuous, that is,

$$|v(x) - v(y)| \le C\lambda d(x, y)^{\beta} \quad \text{for all } x, y \in X. \tag{39}$$

(i) If $x, y \in X \setminus E_\lambda$, then (39) follows from (31) and the definition of E_λ.
(ii) Let $x, y \in E_\lambda$ and $d(x, y) \ge \delta$, where

$$\delta = \frac{1}{4} \max\{\operatorname{dist}(x, X \setminus E_\lambda), \operatorname{dist}(y, X \setminus E_\lambda)\},$$

and let $\bar{x}, \bar{y} \in X \setminus E_\lambda$ be as above. Then, by (38),

$$\begin{aligned} |v(x) - v(y)| &\le |v(x) - v(\bar{x})| + |v(\bar{x}) - v(\bar{y})| + |v(y) - v(\bar{y})| \\ &\le C\lambda \operatorname{dist}(x, X \setminus E_\lambda)^{\beta} + |v(\bar{x}) - v(\bar{y})| + C\lambda \operatorname{dist}(y, X \setminus E_\lambda)^{\beta}, \end{aligned}$$

where $|v(\bar{x}) - v(\bar{y})| \le C\lambda d(\bar{x}, \bar{y})^{\beta}$ by (31) and the fact that $\bar{x}, \bar{y} \in X \setminus E_\lambda$. Since $d(x, y) \ge \delta$ and

$$d(\bar{x}, \bar{y}) \le d(\bar{x}, x) + d(x, y) + d(\bar{y}, y) \le 17 d(x, y),$$

we have that $|v(x) - v(y)| \le C\lambda d(x, y)^{\beta}$.
(iii) Assume then that $x, y \in E_\lambda$ and $d(x, y) \le \delta$. Similarly as $\mathscr{B}_x$ above, we let

$$\mathscr{B}_y = \{B_i \in \mathscr{B} : y \in 2B_i\}.$$

Let B_{i_0} be a Whitney ball for which $x \in B_{i_0}$, and let $x^*_{i_0}$ be the closest point of x_{i_0} in $X \setminus E_\lambda$. By the properties of the functions φ_i, we have that

$$\begin{aligned} |v(x) - v(y)| &= \left| \sum_{i=1}^{\infty} (\varphi_i(x) - \varphi_i(y)) \left(u(x^*_{i_0}) - u_{2B_i}\right) \right| \\ &\le C d(x, y) \sum_{\mathscr{B}_x \cup \mathscr{B}_y} r_i^{-1} |u(x^*_{i_0}) - u_{2B_i}|. \end{aligned} \tag{40}$$

We continue as in (35)–(37); by Lemma 10, we have that $r_i \approx r_{i_0}$ and that $2B_i \subset B(x^*_{i_0}, C_5 r_i)$, and obtain

$$|u(x^*_{i_0}) - u_{2B_i}| \le Cr_i^{\beta} u^{\#}_{\beta}(x^*_{i_0}) \le Cr_i^{\beta} \lambda. \tag{41}$$

Now (40) and (41) show that

$$|v(x)-v(y)| \le C\lambda d(x,y)^{\beta} \sum_{\mathscr{B}_x\cup\mathscr{B}_y} \frac{d(x,y)^{1-\beta}}{r_i^{1-\beta}}.$$

The desired estimate follows because $d(x,y) \le C_4 r_i$ by Lemma 10.
(iv) Finally, let $x \in E_\lambda$ with $\bar{x}$ as above, and let $y \in X\setminus E_\lambda$. Then $v(y) = u(y)$, and using (38) and (31) we have that

$$\begin{aligned}|v(x)-v(y)| &= |v(x)-u(y)| \le |v(x)-v(\bar{x})| + |u(\bar{x})-u(y)| \\ &\le C\lambda\,\mathrm{dist}(x, X\setminus E_\lambda)^{\beta} + C\lambda d(y,\bar{x})^{\beta} \le C\lambda d(x,y)^{\beta}.\end{aligned}$$

The Hölder continuity of v with estimate (39) follows from the four cases above.

Next we should show that $v \in N^{1,\Psi}(X)$. We have to show that $v \in L^{\Psi}(X)$, that it has a weak upper gradient g_v, and that the norms of v and g_v are controlled by the norms of u and g.

We begin with the integrability of v. By the properties of the Whitney covering, we have for each $x \in E_\lambda$ that

$$|v(x)| \le \sum_{i=1}^{\infty} \varphi_i(x)|u|_{2B_i} \le \sum_{i=1}^{\infty} \varphi_i(x)\,\mathscr{M}(u\chi_{E_\lambda})(x) \le C\,\mathscr{M}(u\chi_{E_\lambda})(x),$$

and by the boundedness of the maximal operator, that

$$\begin{aligned}\|v\|_{L^{\Psi}(X)} &\le \|u\chi_{X\setminus E_\lambda}\|_{L^{\Psi}(X)} + C\|\mathscr{M}(u\chi_{E_\lambda})\|_{L^{\Psi}(X)} \\ &\le \|u\|_{L^{\Psi}(X)} + C\|u\|_{L^{\Psi}(E_\lambda)} \le C\|u\|_{L^{\Psi}(X)}.\end{aligned} \tag{42}$$

To find a weak upper gradient for v, we use the equivalence of the spaces $N^{1,\Psi}(X)$ and $M^{1,\Psi}(X)$. By Theorem 4 and the boundedness of the maximal operator, the function $C\,\mathscr{M}\,g$ belongs to $\mathrm{D}(u)\cap L^{\Psi}(X)$. We will show that $C\,\mathscr{M}\,g \in \mathrm{D}(v)\cap L^{\Psi}(X)$. As in the proof of the Hölder continuity, we will consider four cases.
(i) By Theorem 4 and the choice of the representative of u, $C\,\mathscr{M}\,g \in \mathrm{D}(u)$ and

$$|v(x)-v(y)| = |u(x)-u(y)| \le d(x,y)(C\,\mathscr{M}\,g(x) + C\,\mathscr{M}\,g(y))$$

for all $x, y \in X\setminus E_\lambda$.
(ii) If $x, y \in E_\lambda$ and $d(x,y) \le \delta$, then similar calculation as in (40) shows that

$$|v(x)-v(y)| \le C d(x,y) \sum_{\mathscr{B}_x\cup\mathscr{B}_y} r_i^{-1}|u(x)-u_{2B_i}|, \tag{43}$$

where

$$|u(x)-u_{2B_i}| \le |u(x)-u_{B(x,C_1r_i)}| + |u_{B(x,C_1r_i)}-u_{2B_i}|, \tag{44}$$

and $2B_i \subset B(x, C_1r_i)$ by Lemma 10. Using the Poincaré inequality and the doubling property of μ, we have

$$|u_{B(x,C_1r_i)}-u_{2B_i}| \le C \fint_{B(x,C_1r_i)} |u-u_{B(x,C_1r_i)}|\,d\mu \le Cr_i \fint_{B(x,C_1r_i)} g\,d\mu.$$

This together with a telescoping argument for $|u(x)-u_{B(x,C_1r_i)}|$ shows that

$$|u(x)-u_{2B_i}| \le Cr_i\,\mathscr{M}\,g(x) \tag{45}$$

for almost all x. Since the number of balls in $\mathscr{B}_x \cup \mathscr{B}_y$ is bounded, the estimates (43)–(45) show that

$$|v(x) - v(y)| \le Cd(x, y)\,\mathscr{M}\,g(x)$$

for almost all $x, y \in E_\lambda$ with $d(x, y) \le \delta$.
(iii) Let $x, y \in E_\lambda$ with $d(x, y) \ge \delta$. Using the properties of the functions φ_i, the fact that $C\,\mathscr{M}\,g \in \mathrm{D}(u)$, similar estimates for $|u(x) - u_{2B_i}|$ and $|u(y) - u_{2B_i}|$ as in the previous case, and Lemma 10 to conclude that $r_i \approx \operatorname{dist}(x, X \setminus E_\lambda)$ for all $B_i \in \mathscr{B}_x$ (and similarly for $\mathscr{B}_y$), we have that

$$\begin{aligned} |v(x) - v(y)| &\le \sum_{\mathscr{B}_x} |u(x) - u_{2B_i}| + \sum_{\mathscr{B}_y} |u(y) - u_{2B_i}| + |u(x) - u(y)|, \\ &\le C \operatorname{dist}(x, X \setminus E_\lambda)\,\mathscr{M}\,g(x) + C \operatorname{dist}(y, X \setminus E_\lambda)\,\mathscr{M}\,g(y) \\ &\quad + d(x, y)\,(C\,\mathscr{M}\,g(x) + C\,\mathscr{M}\,g(y)) \\ &\le Cd(x, y)\,(\mathscr{M}\,g(x) + \mathscr{M}\,g(y))\,. \end{aligned}$$

(iv) If $y \in E_\lambda$ and $x \in X \setminus E_\lambda$, then

$$|v(x) - v(y)| = |u(x) - v(y)| = \left| \sum_{i=1}^{\infty} \varphi_i(y)(u(x) - u_{2B_i}) \right| \le \sum_{\mathscr{B}_y} |u(x) - u_{2B_i}|,$$

where, by the fact that $C\,\mathscr{M}\,g \in \mathrm{D}(u)$, and by a similar calculation as for (44),

$$\begin{aligned} |u(x) - u_{2B_i}| &\le |u(x) - u(y)| + |u(y) - u_{2B_i}| \\ &\le d(x, y)\,(C\,\mathscr{M}\,g(x) + C\,\mathscr{M}\,g(y)) + Cr_i\,\mathscr{M}\,g(y). \end{aligned}$$

Since for $B_i \in \mathscr{B}_y$, $r_i \approx \operatorname{dist}(y, X \setminus E_\lambda)$ and $\operatorname{dist}(y, X \setminus E_\lambda) \le d(x, y)$, we obtain

$$|v(x) - v(y)| \le Cd(x, y)\,(\mathscr{M}\,g(x) + \mathscr{M}\,g(y))\,.$$

By the above calculation and the boundedness of the maximal operator, we conclude that $C\,\mathscr{M}\,g \in L^\Psi(X)$ and $\|\mathscr{M}\,g\|_{L^\Psi(X)} \le C\|g\|_{L^\Psi(X)}$. Lemma 4 together with the continuity of v shows that $v \in N^{1,\Psi}(X)$ and that $C\,\mathscr{M}\,g$ is a weak upper gradient of v.

Proof (Proof of 3—Approximation in norm) Using the fact that $v = u$ in $X \setminus E_\lambda$ and (42), we have that

$$\|u - v\|_{L^\Psi(X)} = \|u - v\|_{L^\Psi(E_\lambda)} \le C\|u\|_{L^\Psi(E_\lambda)},$$

which tends to 0 as $\lambda \to \infty$ because $\mu(E_\lambda) \to 0$ as $\lambda \to \infty$, (see Lemma 4).

We have to find a weak upper gradient g_λ of $u - v$ for which $\|g_\lambda\|_{L^\Psi(X)} \to 0$ as $\lambda \to \infty$. Since g and $C\,\mathscr{M}\,g$ are weak upper gradients of u and v, $C\,\mathscr{M}\,g$ is a weak upper gradient of $u - v$. As $u - v$ vanishes outside an open set E_λ, the function $g_\lambda = C(\mathscr{M}\,g)\chi_{E_\lambda}$ is a weak upper gradient of $u - v$. As above, the boundedness of the maximal operator implies that

$$\|g_\lambda\|_{L^\Psi(X)} \le C\|g\chi_{E_\lambda}\|_{L^\Psi(X)} = C\|g\|_{L^\Psi(E_\lambda)},$$

and hence $\|u - v\|_{N^{1,\Psi}(X)} \to 0$ as $\lambda \to \infty$.

Proof (Proof of 4—Hausdorff content of E_λ) By (33) and Claim 1,

$$E_\lambda \subset \left\{x \in B(x_0, 2) : \mathscr{M}_{(1-\beta),1}\,g(x) > C\lambda\right\}$$

for each $\lambda > \lambda_0$. Moreover, (3), the Jensen inequality and the continuity of Ψ imply that

$$\Psi(\mathscr{M}_{(1-\beta),1}\, g)(x) \leq \mathscr{M}_{(1-\beta),1}\, \Psi(g)(x).$$

Hence for $\lambda > \lambda_0$,

$$E_\lambda \subset \big\{x \in B(x_0, 2) : \mathscr{M}_{(1-\beta),1}\, \Psi(g)(x) > C\Psi(\lambda)\big\}.$$

Since $\Psi(g) \in L^1(X)$, a weak type estimate for the Hausdorff content of the fractional maximal function (see, for example [22, Lemma 2.6]), implies that

$$\mathscr{H}_\infty^{s-(1-\beta)}(E_\lambda) \leq C\Psi(\lambda)^{-1} \int_X \Psi(g)\, d\mu, \tag{46}$$

where $C \leq 5^{s-(1-\beta)}(2\,\mathrm{diam}(B(x_0, 2)))^s \mu(B(x_0, 2))^{-1}$, the claim 4 follows because the right-hand side of (46) tends to 0 as $\lambda \to \infty$.

Step 2 General case.
Let $\varepsilon > 0$. We cover X by balls of radius $1/10$, and use the $5r$-covering theorem to obtain pairwise disjoint balls $B(a_j, 1/10)$ such that $X \subset \cup_{j=1}^\infty B(a_j, 1/2)$ and that the balls $2B_j$, where $B_j = B(a_j, 1)$ have bounded overlap. Let (ψ_j) be a partition of unity for this covering such that $\sum_{j=1}^\infty \psi_j(x) = 1$ for all $x \in X$, each ψ_j is L-Lipschitz, $0 \leq \psi_j \leq 1$, and $\mathrm{supp}\, \psi_j \subset B_j$ for all $j \in \mathbb{N}$.

Let $u \in N^{1,\Psi}(X)$ with a weak upper gradient $g \in L^\Psi(X)$, and let $u_j = u\psi_j$. Then

$$u(x) = \sum_{j=1}^\infty u_j(x), \tag{47}$$

and the sum is finite for all $x \in X$. By Lemma 5, each u_j is in $N^{1,\Psi}(X)$ and

$$g_j = (g + C|u|)\chi_{B_j}$$

is a weak upper gradient of u_j. Since $\mathrm{supp}\, u_j \subset B_j$, the first step of the proof shows there are functions $v_j \in N^{1,\Psi}(X)$ and open sets $\Omega_j \subset 2B_j$ such that

(i) $v_j = u_j$ in $X \setminus \Omega_j$, $\mathrm{supp}\, v_j \subset 2B_j$,
(ii) $v_j \in N^{1,\Psi}(X)$ is Hölder continuous with exponent β,
(iii) $\|u_j - v_j\|_{N^{1,\Psi}(X)} < 2^{-j}\varepsilon$,
(iv) $\mathscr{H}_\infty^{s-(1-\beta)}(\Omega_j) < 2^{-j}\varepsilon$,
(v) $h_j = C\,\mathscr{M}\, g_j$ is a weak upper gradient of v_j.

We define $\Omega = \cup_{j=1}^\infty \Omega_j$, and claim that the function $v = \sum_{j=1}^\infty v_j$ has properties 1–4. The first claim follows from (i) and (47). The Hausdorff content estimate for Ω follows from (iv) using the subadditivity of $\mathscr{H}_\infty^{s-(1-\beta)}$. By (39), we have

$$|v_j(x) - v_j(y)| \leq C\lambda_j d(x, y)^\beta$$

for all $x, y \in X$. Since, by the proof above, the constant λ_j depends on ε and on j, the Hölder continuity of the functions v_j and the fact that $\mathrm{supp}\, v_j \subset 2B_j$ give Hölder continuity of v only in bounded subsets of X.

To complete the proof of Theorem 5, we have to show that $v \in N^{1,\Psi}(X)$ and that the norm estimate holds. By (iii), we have

$$\sum_{j=1}^\infty \|u_j - v_j\|_{N^{1,\Psi}(X)} < \sum_{j=1}^\infty 2^{-j}\varepsilon = \varepsilon, \tag{48}$$

that is, the series $\sum_{j=1}^{\infty}(u_j - v_j)$ convergences absolutely, and hence converges in the Banach space $N^{1,\Psi}(X)$. Since $u = \sum_{j=1}^{\infty} u_j$ is in $N^{1,\Psi}(X)$, also $\sum_{j=1}^{\infty} v_j$ converges in $N^{1,\Psi}(X)$. Moreover, by (48) we obtain

$$\|u - v\|_{N^{1,\Psi}(X)} \leq \sum_{j=1}^{\infty} \|u_j - v_j\|_{N^{1,\Psi}(X)} < \varepsilon.$$

Remark 2 For the classical Orlicz–Sobolev space $W^{1,\Psi}(\mathbb{R}^n)$ the proof of Theorem 5 is shorter because the property that $v \in W^{1,\Psi}(\mathbb{R}^n)$ follows easily using absolute continuity on lines. Moreover, using the uniqueness of weak gradients, we obtain a better norm estimate $\|u - v\|_{W^{1,\Psi}(\mathbb{R}^n)} \leq \|u\|_{W^{1,\Psi}(\Omega)} < \varepsilon$.

Acknowledgments I would like to thank the referee for comments and suggestions that improved the paper.

References

1. Adams, R.A.: On the Orlicz–Sobolev imbedding theorem. J. Funct. Anal. **24**(3), 241–257 (1977)
2. Adams, R.A., Hedberg, L.I.: Function spaces and potential theory. Grundlehren der Mathematischen Wissenschaften, 314. Springer, Berlin (1996)
3. Aïssaoui, N.: Maximal operators, Lebesgue points and quasicontinuity in strongly nonlinear potential theory. Acta Math. Univ. Comenian. (N.S.) **71**(1), 35–50 (2002)
4. Aïssaoui, N.: Strongly nonlinear potential theory of metric spaces. Abstr. Appl. Anal. **7**(7), 357–374 (2002)
5. Aïssaoui, N.: Another extension of Orlicz–Sobolev spaces to metric spaces. Abstr. Appl. Anal. **1**, 1–26 (2004)
6. Alberico, A., Cianchi, A.: Differentiability properties of Orlicz–Sobolev functions. Ark. Mat. **43**(1), 1–28 (2005)
7. Björn, A., Björn, J., Shanmugalingam, N.: Quasicontinuity of Newton–Sobolev functions and density of Lipschitz functions in metric measure spaces. To appear in Houston J. Math.
8. Björn, J., Onninen, J.: Orlicz capacities and Hausdorff measures on metric spaces. Math. Z. **251**(1), 131–146 (2005)
9. Cianchi, A.: Continuity properties of functions from Orlicz–Sobolev spaces and embedding theorems. Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4) **23**(3), 575–608 (1996)
10. Cianchi, A.: Strong and weak type inequalities for some classical operators in Orlicz spaces. J. Lond. Math. Soc. (2) **60**(1), 187–202 (1999)
11. Cianchi, A., Stroffolini, B.: An extension of Hedberg's convolution inequality and applications. J. Math. Anal. Appl. **227**(1), 166–186 (1998)
12. Coifman, R.R., Weiss, G.: Analyse harmonique non-commutative sur certains espaces homogènes. Lecture Notes in Mathematics, vol. 242. Springer, Berlin (1971)
13. Donaldson, T.K., Trudinger, N.S.: Orlicz–Sobolev spaces and imbedding theorems. J. Funct. Anal. **8**, 52–75 (1971)
14. Edmunds, D.E., Gurka, P., Opic, P.: Double exponential integrability of convolution operators in generalized Lorentz-Zygmund spaces. Indiana Univ. Math. J. **44**(1), 19–43 (1995)
15. Fiorenza, A., Prignet, A.: Orlicz capacities and applications to some existence questions for elliptic PDEs having measure data. ESAIM Control Optim. Calc. Var. **9**, 317–341 (2003)
16. Fu, Y.Q., Wang, J., Yang, F.L.: Boundedness of Hardy–Littlewood maximal functions in Orlicz–Sobolev spaces. (Chinese) J. Harbin Inst. Tech. **38**(1), 85–87 (2006)
17. Fuglede, B.: Extremal length and functional completion. Acta Math. **98**, 171–219 (1957)
18. Fusco, N., Lions, P.-L., Sbordone, C.: Sobolev imbedding theorems in borderline cases. Proc. Amer. Math. Soc. **124**(2), 561–565 (1996)
19. Genebashvili, I., Gogatishvili, A., Kokilashvili, V., Krbec, M.: Weight theory for integral transforms on spaces of homogeneous type. Pitman Monographs and Surveys in Pure and Applied Mathematics, vol. 92. Longman, Harlow (1998)
20. Gogatishvili, A., Kokilashvili, V., Krbec, M.: Maximal functions, $\phi(L)$ classes and Carleson measures. Proc. A. Razmadze Math. Inst. **102**, 85–97 (1993)
21. Hajłasz, P.: Sobolev spaces on an arbitrary metric space. Potential Anal. **5**(4), 403–415 (1996)

22. Hajłasz, P., Kinnunen, J.: Hölder quasicontinuity of Sobolev functions on metric spaces. Rev. Mat. Iberoamericana **14**(3), 601–622 (1998)
23. Hajłasz, P., Koskela, P.: Sobolev met Poincaré. Mem. Amer. Math. Soc. **145**, 688 (2000)
24. Harjulehto, P., Kinnunen, J.: Differentiation bases for Sobolev functions on metric spaces. Publ. Mat. **48**(2), 381–395 (2004)
25. Heinonen, J., Koskela, P.: Quasiconformal maps on metric spaces with controlled geometry. Acta Math. **181**, 1–61 (1998)
26. Iwaniec, T., Koskela, P., Onninen, J.: Mappings of finite distortion: monotonicity and continuity. Invent. Math. **144**(3), 507–531 (2001)
27. Kauhanen, J., Koskela, P., Malý, J., Onninen, J., Zhong, X.: Mappings of finite distortion: sharp Orlicz–conditions. Rev. Mat. Iberoamericana **19**(3), 857–872 (2003)
28. Kinnunen, J.: The Hardy–Littlewood maximal function of a Sobolev function. Israel J. Math. **100**, 117–124 (1997)
29. Kinnunen, J., Latvala, V.: Lebesgue points for Sobolev functions on metric spaces. Rev. Mat. Iberoamericana **18**, 685–700 (2002)
30. Kinnunen, J., Lindqvist, P.: The derivative of the maximal function. J. Reine Angew. Math. **503**, 161–167 (1998)
31. Kinnunen, J., Tuominen, H.: Pointwise behaviour of $M^{1,1}$-Sobolev functions. Math. Z. **257**(3), 613–630 (2007)
32. Kokilashvili, V., Krbec, M.: Weighted Inequalities in Lorentz and Orlicz spaces. World Scientific Publishing Co, River Edge, NJ (1991)
33. Krasnosel'skiĭ, M.A., Rutickiĭ, J.B.: Convex functions and Orlicz spaces. Noordhoff Ltd., Groningen (1961)
34. Macías, R.A., Segovia, C.: A decomposition into atoms of distributions on spaces of homogeneous type. Adv. in Math. **33**(3), 271–309 (1979)
35. Malý, L.: Hölder type quasicontinuity. Potential Anal. **2**(3), 249–254 (1993)
36. Malý, J., Swanson, D., Ziemer, W.P.: Fine behavior of functions whose gradients are in an Orlicz space. Preprint MATH-KMA-2005/178. Charles University, Prague
37. Rao, M.M., Ren, Z.D.: Theory of Orlicz spaces. Monographs and Textbooks in Pure and Applied Mathematics 146, Marcel Dekker, New York (1991)
38. Rudd, M.: A direct approach to Orlicz–Sobolev capacity. Nonlinear Anal. **60**(1), 129–147 (2005)
39. Rudin, W.: Real and Complex Analysis, 3rd edn. McGraw-Hill, New York (1987)
40. Semmes, S.: Finding curves on general spaces through quantitative topology with applications to Sobolev and Poincaré inequalities. Selecta Math. (N.S.) **2**, 155–295 (1996)
41. Shanmugalingam, N.: Newtonian spaces: An extension of Sobolev spaces to metric measure spaces. Rev. Mat. Iberoamericana **16**, 243–279 (2000)
42. Shanmugalingam, N.: Harmonic functions on metric spaces. Illinois J. Math. **45**(3), 1021–1050 (2001)
43. Tuominen, H.: Orlicz–Sobolev spaces on metric measure spaces. Ann. Acad. Sci. Fenn. Math. Dissertation No. 135 (2004)