Calculus of Variations

Density of Lipschitz functions in energy

Sylvester Eriksson-Bique[1] (ID)

Abstract

In this paper, we show that the density in energy of Lipschitz functions in a Sobolev space $N^{1,p}(X)$ holds for all $p \in [1, \infty)$ whenever the space X is complete and separable and the measure is Radon and positive and finite on balls. Emphatically, $p = 1$ is allowed. We also give a few corollaries and pose questions for future work. The proof is direct and does not involve the usual flow techniques from prior work. It also yields a new approximation technique, which has not appeared in prior work. Notable with all of this work is that we do not use any form of Poincaré inequality or doubling assumption. The techniques are flexible and suggest a unification of a variety of approaches that have appeared in the literature on the topic.

Mathematics Subject Classification Primary 46E36 · Secondary 30L99 · 26B30

1 Introduction

In this paper, we study the density of Lipschitz functions in Sobolev spaces when X is complete and separable, and μ is any Radon measure on X which is positive and finite on balls. We consider the so-called Newton-Sobolev space $N^{1,p}(X)$ defined in [35] (see also [3, 21]), which for $p > 1$ coincides with the one introduced independently in [6]. A function f is in $N^{1,p}(X)$ if $f \in L^p(X)$ and if it has an upper gradient $g \in L^p(X)$; see definition (2.1). Associated to each f there is a *minimal p-weak upper gradient* $g_f \in L^p(X)$, which plays the role of the norm of a gradient. We will give precise definitions of these and the following notation in Sect. 2.

Our main result proves *density in energy* or, rather, produces a sequence of Lipschitz functions which converges in energy. A sequence of functions $(f_i)_{i \in \mathbb{N}}$, with $f_i \in N^{1,p}(X)$ converges to $f \in N^{1,p}(X)$ in energy, if the functions f_i converge to f in $L^p(X)$ and if their minimal p-weak upper gradients g_{f_i} converge to g_f in $L^p(X)$. Our sequences of functions f_i will be Lipschitz functions with bounded support, that is $f_i \in \mathrm{LIP}_b(X) \subset N^{1,p}(X)$. Our argument in fact shows more than the convergence of the minimal p-weak upper gradients. It shows that the asymptotic Lipschitz constants converge;

✉ Sylvester Eriksson-Bique
syerikss@jyu.fi

1 Department of Mathematics, Jyväskylä University, P.O. Box 35 (MaD), FI-40014 Jyväskylä, Finland

$$\mathrm{lip}_a[f](x) := \lim_{r \to 0} \sup_{a \neq b \in B(x,r)} \frac{|f(a) - f(b)|}{d(a, b)},$$

and $\mathrm{lip}_a[f](x) := 0$, if x is an isolated point. Since $\mathrm{lip}_a[f]$ is an upper gradient, and since g_f is minimal, it follows that $\mathrm{lip}_a[f] \geq g_f$ for any $f \in \mathrm{LIP}_b(X)$. Roughly, it is more difficult to ensure that $\mathrm{lip}_a[f_i]$ converges, since it is larger. More precisely, from Lemma 2.9, it follows that convergence of $\mathrm{lip}_a[f_i]$ implies convergence in energy, and is thus a stronger statement.

Theorem 1.1 *Let X be a complete and separable metric space, let $p \in [1, \infty)$ and let μ be a Radon measure which is positive and finite on balls. If $f \in N^{1,p}(X)$, then there exists a sequence $f_i \in \mathrm{LIP}_b(X) \subset N^{1,p}$ so that the following properties hold.*

(1) The functions f_i converge in $L^p(X)$ to f, that is

$$\lim_{i \to \infty} \int_X |f_i - f|^p d\mu = 0.$$

(2) The asymptotic Lipschitz constants $\mathrm{lip}_a[f_i]$ and the minimal p-weak upper gradients g_{f_i} of f_i converge to the minimal p-weak upper gradient g_f of f in $L^p(X)$, that is

$$\lim_{i \to \infty} \int_X |g_{f_i} - g_f|^p d\mu = \lim_{i \to \infty} \int_X |\mathrm{lip}_a[f_i] - g_f|^p d\mu = 0.$$

The conclusion of the theorem, for exponents $p > 1$, is contained in [2] (see also [1, Section 6]). Their methods, however were inexplicit. Further, the present result applies to the case $p = 1$, and gives a conceptually new way of obtaining their result. The case of $p = 1$ is of particular importance in applications that use the co-area inequality, as can be seen from the concurrent work [10].

In fact, the proof yields a slightly stronger conclusion. That this result is stronger, follows from Lemma 2.9.

Theorem 1.2 *Let X be a complete and separable metric space, let $p \in [1, \infty)$ and let μ be a Radon measure which is positive and finite on balls. Let $f \in L^1_{\mathrm{loc}}(X)$ and let $g \in L^p_{\mathrm{loc}}(X)$ be a p-weak upper gradient of f. For every $\epsilon > 0$ and every bounded set $C \subset X$ there exists a function $g_\epsilon \in L^p_{\mathrm{loc}}(X)$ with $\|g_\epsilon - g\|_{L^p(C)} < \epsilon$ and a sequence $f_i \in \mathrm{LIP}_b(X)$, $i \in \mathbb{N}$, so that the following properties hold.*

(1) The functions f_i converge in $L^1_{\mathrm{loc}}(X)$ to f, that is, for any bounded set $A \subset X$,

$$\lim_{i \to \infty} \int_A |f_i - f| d\mu = 0.$$

(2) We have $\mathrm{lip}_a[f_i] \leq g_\epsilon$ for every $i \in \mathbb{N}$.

Remark 1.3 Convergence in energy is *weaker* than convergence in norm. A sequence f_i converges in norm, if for any $\epsilon > 0$ there exists an $N \in \mathbb{N}$ so that for all $i \geq N$ it holds that $\|f - f_i\|_{N^{1,p}} \leq \epsilon$. In particular, this requires that the minimal p-weak upper gradients g_{f-f_i} of the differences $f - f_i$ converge to 0 in $L^p(X)$ with $i \to \infty$. However, if the functions f_i converge in energy to f, then we only know that the differences of minimal p-weak upper gradients, $g_f - g_{f_i}$, converge to 0 in $L^p(X)$. Crucially, (a.e.) we have $g_{f-f_i} \geq |g_f - g_{f_i}|$. This means, that convergence in norm is a stronger statement than convergence in energy. Indeed, by the following example, it is strictly stronger.

Equip $X = [0, 1]^2$ with the metric $d((x_1, y_1), (x_2, y_2)) = |x_1 - y_1| + |x_2 - y_2|$ and the Lebesgue measure. Let $f(x, y) := x$, and $f_n(x, y) := x + n^{-1} \sin(ny)$, and consider any

$p \in [1, \infty)$. For any smooth function $a \in N^{1,p}(X)$, its minimal p-weak upper gradient g_a is given by the ℓ^∞-norm of its gradient vector $\|\nabla a\|_{\ell^\infty}$. (Note that ℓ^∞ is the dual norm of ℓ^1.) While this is classical, its argument is rarely spelled out in detail, and we will do so after explaining the relevance for our example. We get that $g_{f_n} = \max\{|\partial_x f_n|, |\partial_y f_n|\} = 1 = g_f$. In particular, $f_n \to f$ in $L^p([0, 1]^2)$ and $g_{f_n} = g_f = 1$. However, $g_{f_n - f} = |\cos(ny)|$, which does not converge to 0 in $L^p(X)$. Thus, the functions f_n converge to f in energy, but not in norm. If one uses any other uniformly convex norm on $\mathbb{R}^2$ to define the metric, then this phenomenon does not occur. Further, in many settings, such as this, one can fix the problem by taking convex combinations of f_n which converge in norm – even if the sequence $(f_n)_{n \in \mathbb{N}}$ only converges in energy. This final step requires reflexivity of the Sobolev space, which can be obtained, for example, with the assumption of finite Hausdorff dimension, see e.g. [11, Theorem 1.9].

We give a brief explanation for the above identification of the minimal upper gradient which an experienced reader may wish to skip. A proof can be patched together from arguments that are contained in [19, Sections 6-7] and [21, Prop. 6.3.3.]. First, from the triangle inequality and the definition in (2.1) it follows that $\|\nabla a\|_{\ell^\infty}$ is an upper gradient for any smooth function $a \in N^{1,p}(X)$ (since it bounds all directional derivatives). Conversely, it is minimal, for the following reason. If g is any p-integrable upper gradient of a, we have that (2.1) holds for every line segment in any given direction v. By the argument in [21, Prop. 6.3.3.], which consists of Lebesgue differentiation of g and differentiation of a along such line segments, we get that the inequality $g \geq |\partial_v a|$ holds a.e., where $\partial_v a = \langle v, \nabla a \rangle$ is the directional derivative. By taking a supremum over a dense collection of vectors v, and as a consequence of the definition of a dual norm, we obtain $g \geq \|\nabla a\|_{\ell^\infty}$.

Density of energy, albeit weak, suffices for many applications. One application is proving the equivalence of various types of Poincaré inequalities. We say that a pair (u, g) satisfies a p-Poincaré inequality (with constants (C, Λ)) if for each ball $B(x, r) \subset X$

$$\fint_{B(x,r)} |u - u_{B(x,r)}| \, d\mu \leq Cr \left(\fint_{B(x, \Lambda r)} g^p \, d\mu \right)^{1/p}, \tag{1.4}$$

where we define the average by $f_A := \fint_A f \, d\mu := \frac{1}{\mu(A)} \int_A f \, d\mu$, when the final expression is well defined. For the statement of the following corollary, we define:

$$\text{lip}[f](x) := \lim_{r \to 0} \sup_{y \in B(x,r)} \frac{|f(x) - f(y)|}{d(x, y)}.$$

Corollary 1.5 *Let X be a complete and separable metric space, let $p \in [1, \infty)$ and let μ be a Radon measure which is positive and finite on balls. For any fixed constants $(C, \Lambda) \in (0, \infty)^2$ the following three conditions are equivalent.*

(1) For every Lipschitz function $f : X \to \mathbb{R}$ the pair $(f, \text{lip}_a[f])$ satisfies a p-Poincaré inequality with constants (C, Λ).

(2) For every Lipschitz function $f : X \to \mathbb{R}$ the pair $(f, \text{lip}[f])$ satisfies a p-Poincaré inequality with constants (C, Λ).

(3) For every $f \in N^{1,p}(X)$ the pair (f, g_f) satisfies a p-Poincaré inequality with constants (C, Λ).

(4) For every $f \in L^1_{\text{loc}}(X)$ and any upper gradient $g \in L^1_{\text{loc}}(X)$ of f, the pair (f, g) satisfies a p-Poincaré inequality with constants (C, Λ).

The statement does not depend on any doubling or properness assumption, as was for example assumed in [24]. The proof is a simple exercise left to the reader of using Theorems 1.1 and

1.2, cut-off functions (see the proof of Theorem 1.1) and the fact that for Lipschitz functions $g_f \leq \text{lip}[f] \leq \text{lip}_a[f]$ (almost everywhere). (For the last fact, see Example 2.5 and the definition of the minimal p-weak upper gradient in Sect. 2.) The equivalence of (4) is easier for $g \in L^p_{\text{loc}}(X)$, but the case of $g \in L^1_{\text{loc}}(X)$ is obtained by employing a cut-off function and a limiting argument that involves radii $r' \nearrow r$. We note that the assumption of completeness can not be removed, as seen from the examples in [28] (see also [25] for a related self-improvement result).

Remark 1.6 Why is completeness relevant? If $Y \subset X$ is any dense full measure set, then (1) and (2) in Corollary 1.5 remain unchanged. Indeed, Y may be even disconnected! Koskela's example in [28] is of a connected open set $Y \subset \mathbb{R}^2$, but where Y is separated by a set, whose capacity is not bounded from below scale-invariantly. This contradicts (3) and (4) in Corollary 1.5. However, Y is full measure and dense in the plane, so (1) and (2) still hold. In our proof, completeness is used via an application of the Arzelà-Ascoli Theorem.

A further application of a technical nature is the following, which was pointed out to us by Elefterios Soultanis.

Corollary 1.7 *Let $p \in [1, \infty)$. Assume that X is complete and separable and equipped with a Radon measure which is positive and finite on balls. If $f \in N^{1,p}(X)$, then there is a Borel function $\tilde{f} \in N^{1,p}(X)$ so that $\tilde{f} = f$ almost everywhere.*

Remark 1.8 If X is additionally measure doubling and satisfies a p-Poincaré inequality, then this result is already known and follows directly from norm-density of Lipschitz functions, see e.g. [21, Theorem 8.2.1].

Note that the definition of $f \in N^{1,p}(X)$ only requires that f is measurable. Further, note that the Newton-Sobolev condition involves a pointwise consideration. Thus, a direct modification using Borel regularity does not yield the result since it may break the property of being a Newton-Sobolev function. The proof of this corollary follows immediately from Theorem 1.1 by considering a subsequence of f_i converging pointwise almost everywhere and their limit together with [15, Proof of Corollary 7.10]. We remark that *a posteriori* also $f = \tilde{f}$ at capacity almost every point, or quasi-everywhere, see [35, Corollary 3.3].

The question of the density of continuous and Lipschitz functions is also crucial in other contexts. For example, quasi-continuity properties of Sobolev functions are implied by the density of continuous functions in norm, see [4] and [33]. While we only get density in energy, it seems our techniques could have something to say in these contexts as well. In conclusion, it appears that density of continuous and Lipschitz functions in energy in Newton-type Sobolev spaces defined using upper gradients is far more generic than it appears from existing literature.

1.1 Approximation scheme

The approximation method that we introduce may seem surprisingly simple and a bit confusing. We thus wish to show how it arises naturally from prior work, explain how it is distinct from the state of the art and to survey existing approximation schemes. Our survey is brief, and undoubtedly not complete. In the following discussion, $f : X \to \mathbb{R}$ will be a Sobolev function (as understoon in the specific context), and $\tilde{f}$ will denote its approximation.

In Euclidean spaces, and Lie groups, the simplest way to approximate an $L^p(X)$-function f is via convolution with $\tilde{f} = f * \phi_n$, where ϕ_n is some approximation of unity. Even without

a group structure, one can mimic this process on manifolds, and even some CAT(0) spaces, using various center of mass constructions. See e.g. [13, 23].

In general metric spaces, such a convolution method is missing. One rather old method is to employ so called discrete convolutions:

$$\tilde{f}(x) = \sum_n \psi_n f_{B_n},$$

where B_n are some balls which cover the space (with bounded overlap), and ψ_n is a partition of unity subordinate to B_n. Such an approximation goes back to Coifman and Weiss [7]. Discrete convolutions have also been independently discovered in various guises, see e.g. [28]. When X is measure doubling (for some $D \in [1, \infty)$ we have $\mu(B(x, 2r)) \leq D\mu(B(x, r))$ for each $x \in X, r > 0$) and satisfies a Poincaré inequality, one can perform a discrete convolution in such a way that $\tilde{f}$ approximates f in $L^p(X)$, and so that the $N^{1,p}(X)$-norm of $\tilde{f}$ is controlled. For such results, see [27]. For applications of discrete convolutions in other contexts and earlier results, see [34, pp. 290–292] and [31].

Discrete convolutions have been quite successful in analysis on metric spaces. In fact, the papers using this technique are too numerous to list here. But, some highlights we found are [26–29]. These papers require strong assumptions: a doubling measure and a Poincaré inequality. It is unclear, if a discrete convolution approach is possible to extend to more general settings. Another issue with discrete convolutions, is that even with the aforementioned strong assumptions, they fail to directly prove Theorem 1.1. Indeed, the approximating functions $\tilde{f}$ have minimal p-weak upper gradients $g_{\tilde{f}}$ with a bound on their $L^p(X)$-norms – but convergence of the minimal p-weak upper gradients may fail.

In order to get convergence of gradients, and to weaken assumptions, we need to look beyond. The seminal paper of Cheeger [6] employs *several* useful approximation schemes, which apply to spaces with a doubling measure and a Poincaré inequality. We summarize them here.

(1) [6, Theorem 4.24]: a sub-level set approach employing an approximation based on the McShane extension in the proof of the existence of a differential for Sobolev functions. As Cheeger notes, this method was standard in Sobolev space theory by the time of his work. We are not certain, where it appeared in the first instance. Hajłasz used the argument earlier for metric measure spaces [14, Theorem 5], and attributes the method to an earlier publication of Liu [30] and Calderón-Zygmund [5, Theorem 13], who proved related results in a Euclidean setting.

(2) [6, Lemma 5.2]: an extension from a net of points using curves similar to Formula (1.10) (see below), which was used to prove a version of Theorem 1.1 and conclude $g_f = \text{lip}[f]$ for length spaces.

(3) [6, Theorem 6.5]: an approximation involving a two-step Lipschitz extension and piece-wise distance function, which was used to prove that $g_f = \text{lip}[f]$ without the length-space assumption.

These get much closer to the approximation method introduced in the present paper – even though the proofs written in [6] involved the strong assumptions of a Poincaré inequality and doubling. These arguments thus do not directly answer our questions. However, with these stronger assumptions, [6, Theorem 6.5] does prove our Theorem 1.1 for $p > 1$.

Each of the methods in Cheeger's paper involved *extending* $f|_A$ out from a subset $A \subset X$ – or interpolating and extrapolating the function while attempting to ensure that $\sup_{a \in A} |\tilde{f}(a) - f(a)|$ is as small as possible. This idea, that approximation and extension problems are

connected, is behind the present paper. We briefly describe the approximation (1) and (2) above in slightly more detail to illustrate the connection to our method.

In the sub-level set approach, one obtains $\tilde{f}$ by taking a sub-level set $A \subset X$ of the Hardy-Littlewood maximal function $M(g_f^p)$, and employing a McShane Lipschitz extension. This is possible, because it follows from a point-wise version of the Poincaré inequality (see e.g. [16, Theorem 3.3]) that $f|_A$ is Lipschitz. This approximation has the remarkable Lusin property: the approximating function $\tilde{f}$ actually agrees with the function f on large measure subsets. Together with locality of the minimal p-weak upper gradient (see e.g. [21, Proposition 6.3.22]), one gets that $\tilde{f} - f$ has small $N^{1,p}(X)$-norm. If one wishes to remove the Poincaré inequality assumption, then one needs substantially different tools. Indeed, without a Poincaré inequality, one must give up the (Lipschitz-)Lusin property because the function $f \in N^{1,p}(X)$ may not, in general, be Lipschitz on any positive measure subset.

Example 1.9 Let $X = [0, 1]$ be a metric measure space equipped with the Lebesgue measure and snow-flake metric $d(x, y) = \sqrt{|x - y|}$. Since X doesn't have any non-constant rectifiable curves, every function $f \in L^p(X)$ is a Sobolev function. Let $f(x) = \sum_{n=1}^{\infty} a^{-n} \sin(2\pi b^n x)$ with $1 < a < \sqrt{b}$ and $1 + 3/2\pi < ab$ and $a, b \in \mathbb{N}$. Then, a classical computation shows that f is $\log_b(a)$-Hölder in the Euclidean metric, and α-Hölder in the metric d, with $\alpha := 2\log_b(a) < 1$. On the other hand, the classical argument of Weierstrass gives [37], that to every $x \in (0, 1)$ and every $m \in \mathbb{N}$ sufficiently large there exists a $y \in [x - b^{-m}, x + b^m] \cap [0, 1]$ with $f(x) - f(y) \gtrsim a^m$. Using this, one can conclude that any set $A \subset (0, 1)$, for which $f|_A$ is Lipschitz with respect to d, must be porous (i.e. there is a constant $c > 0$ so that for every $a \in A$ and any small enough $r > 0$, there exists an $y \in (x - r, x + r)$ so that $B(y, cr) \cap A = \emptyset$). Consequently, there is no positive measure set A for which $f|_A$ is Lipschitz.

The second approximation in [6] takes the form of

$$\tilde{f}(x) := \inf_{\gamma: A \rightsquigarrow x} f(\gamma(0)) + \int_{\gamma} g \, ds, \tag{1.10}$$

where $A \subset X$ is some compact set of points, and the infimum is taken over rectifiable curves $\gamma: [0, 1] \to X$ connecting A to x and the integral $\int_{\gamma} g \, ds$ is the usual curve integral. While the proof, as written, in [6] does use a Poincaré inequality, the full strength of it is not really needed. Indeed, most of the lemmas in [6] do not use this assumption, and the proof of [6, Lemma 5.2] could be rewritten by using the Lusin property and choosing a set A where $f|_A$ is continuous. By such a modification, the proof would apply to any proper length space. To our knowledge, this has not been observed before. However, we omit the details of this claim, since our main theorem contains a stronger result.

The idea in Definition (1.10) is to be close to f on the set A while insisting on g being an upper gradient. Such constructions have arisen in other settings, where one wishes to prescribe a given upper gradient: see e.g. [4, Lemma 3.1]. The main technical problem with Definition (1.10), is that it implicitly insists on the existence of rectifiable curves γ. Without such curves $\tilde{f}$ may even fail to be continuous. Indeed, even the measurability is non-trivial [22].

The issue of a lack of curves has already been identified, and resolved, in limited instances. When proving that a Poincaré inequality implies quasiconvexity, *a priori* one can not assume the existence of any curves. See for example the beautiful discussion in [16, Proposition 4.4] where a version of this fact is proved – or the proof in [6, Theorem 17.1] which is originally due to Semmes. The proof involves "testing" the Poincaré inequality with functions of the

form:

$$f_{\epsilon,x}(y) := \inf_{p_0,\dots,p_n} \sum_{k=0}^{n-1} d(p_k, p_{k+1}),$$

with the infimum is taken over all sequences of points, called discrete paths, $p_0, \dots, p_n$ with $d(p_k, p_{k+1}) \leq \epsilon$ (for $k = 0, \dots, n-1, n \in \mathbb{N}$) and $p_0 = x$, $p_n = y$, and $x \in X$ a fixed point. Indeed, the approximation formula that we use will resemble and generalize this expression.

Our approximation combines the two main aspects from before: discrete paths, and extending from a subset $A \subset X$. First, we explain the approximation for non-negative functions. The approximation is defined using given data: A non-negative function $f : X \to [0, M]$ to approximate which is bounded by $M > 0$, a continuous bounded non-negative function $g : X \to [0, \infty)$ which is our desired upper gradient, a set $A \subset X$ s.t. $f|_A$ is continuous, and a scale parameter $\delta > 0$.

The formula for the approximating function is:

$$\tilde{f}(x) := \min \left\{ \inf_{p_0,\dots,p_n} f(p_0) + \sum_{k=0}^{n-1} g(p_k)d(p_k, p_{k+1}), M \right\} \tag{1.11}$$

where the infimum is taken over all discrete paths $p_0, \dots, p_n$ with $p_0 \in A$, $d(p_k, p_{k+1}) \leq \delta$ and $p_n = x$. There are some key observations, from Lemma 2.13 below, which we highlight here and which guide our definition.

(1) If there are no discrete paths with the given properties, then by the standard convention, the infimum in the definition is ∞, and we get $\tilde{f}(x) = M$. In any case, the cut-off value M ensures that $\tilde{f}(x) \leq M$ for every $x \in X$.

(2) The function $\tilde{f}$ is automatically $\max\{\sup_{x \in X} g(x), M/\delta\}$-Lipschitz, and we have $\mathrm{lip}_a(\tilde{f}) \leq g$. Indeed, if $d(x, y) \leq \delta$ for some $x, y \in X$, by concatenating discrete paths, we obtain a bound for $|\tilde{f}(x) - \tilde{f}(y)| \leq \max\{g(x), g(y)\}d(x, y)$. By boundedness, if $d(x, y) > \delta$, then $|\tilde{f}(x) - \tilde{f}(y)| \leq M \leq \frac{2M}{\delta}d(x, y)$. Further details will be given later.

(3) We have $\tilde{f}(x) \leq f(x)$ for each $x \in A$. Indeed, in the infimum, we can choose the discrete path $P = (x, x)$. Thus, we have $\tilde{f}(x) \leq f(x) + g(x)d(x, x) = f(x)$.

The difficulty lies then in choosing a function g appropriately so that one can show that $\tilde{f}|_A$ converges to $f|_A$ pointwise as $\delta \to 0$. Here, a refined version of Arzelà-Ascoli is used as part of a compactness and contradiction argument. If the convergence were to fail, then we would get a sequence of discrete paths converging to a curve, which would violate the upper gradient inequality (2.1). This step is classical, and has already appeared in [20, Proposition 2.17] and in [6, Lemma 5.18]. According to Cheeger, this argument has also been used by Rickman and Ziemer [6, Remark 5.26].

For the technically minded, we already mention that properness, which is usually assumed, is avoided by appropriately choosing g to penalize paths that form non-compact families. Indeed, paths that travel "far" away from certain compact sets must have small "modulus", as we will later make precise. This is a new argument and seems to be useful in other settings as well. However, with properness, our proof would be considerably simpler – and this will be indicated in the proof.

Formula (1.11) applies directly only to functions with a bounded support and non-negative functions. For signed functions without bounded support, one first does a truncation and applies a cut-off function. Then, the previous approximation scheme is applied to the positive and negative part individually. We note that modifications of the formula yield approximations

directly for any function $f \in N^{1,p}(X)$. However, the scheme presented here simplifies the proof slightly.

Finally, for the sake of completeness, we wish to mention another enormously successful approximation method which arises via estimates for the gradient flow of convex functionals, cf. [2]. This approximation is implicit, since the method does not directly furnish the approximation, but shows that it exists. Indeed, it proceeds via defining a functional, via relaxation of a Dirichlet energy and needs lower-semicontinuity for it in $L^2(X)$. By showing that two different expressions define the gradient of the flow of this functional, one obtains the existence of approximations for Sobolev functions. While very general and powerful, this approach has two main shortcomings. The first is the inexplicit nature of the approximating function and lack of any pointwise control. Another problem is in the $p = 1$ case for $N^{1,1}(X)$. When $p = 1$, the relaxation approach does not give information for $N^{1,1}$ but instead for functions of bounded variation. To give results for this borderline case, we needed to introduce the methods of the present paper. We remark, that this case is still quite important, since it is connected via the co-area formula to (modulus) estimates on curves and surfaces in the space.

1.2 Further questions

The methods of this paper are likely to apply to a host of other Sobolev type spaces and lead to interesting further questions. A number of works on approximations have appeared in different settings, see e.g. Orlicz-Sobolev spaces [36], Lorenz-Sobolev spaces [8] and variable exponent Sobolev spaces [17, 18]. One can even study these questions with a general Banach function space norm, see e.g. [32, 33]. This list is far from exhaustive. Indeed, a variety of authors have asked for necessary and sufficient conditions for the density of Lipschitz functions in these settings – and we suggest that completeness and separability suffice, with perhaps minimal further assumptions when an upper gradient is used. It is important to note, however, that the situation is quite different for the Sobolev space, often denoted $W^{1,p}(X)$, which is defined using a distributional gradient, see e.g. [18] for such issues in a variable exponent case. The techniques here suggest, that the questions on density in this different setting are equivalent with $N^{1,p}(X) = W^{1,p}(X)$, which is a type of regularity statement.

Another question is when (locally) Lipschitz functions are dense in $N^{1,p}(\Omega)$ when Ω is a domain – i.e open and connected – in a complete and separable space X. We use completeness in our arguments, and additional care is needed close to the boundary of Ω. In some cases, when Ω is say a slit disk in the plane $B(0, 1)\backslash(0, 1) \times \{0\} \subset \mathbb{R}^2$, one would not expect such a density for globally Lipschitz functions. However, it may be that some minimal assumption would guarantee density of locally Lipschitz functions.

A final, and seemingly difficult question, is if Lipschitz functions are always actually dense in $N^{1,p}(X)$ in norm, and not just in energy (when X is complete). If $p > 1$ and if the Sobolev space is reflexive, then the density in energy can be directly upgraded to density in norm. For a space X which is metrically doubling, this follows from [1, Corollary 41]. In a concurrent work with Elefterios Soultanis, we have employed techniques from this paper to get the density result for all $p \in [1, \infty)$ and with a weaker finite dimensionality assumption for an associated p-differentiable structure [11]. This finite dimensionality assumption is satisfied by all spaces X with a finite Hausdorff dimension.

Outline: The proof of Theorem 1.1 will be at the end of Sect. 2. At the beginning of that section, there are three preliminary subsections, which will describe the terminology, basic properties of the approximating functions and some useful lemmas for discrete paths.

2 Proof of Approximation

2.1 Preliminaries

Open balls in a metric space X are defined by $B(x, r) := \{y : d(x, y) < r\}$, for $r > 0$ and $x \in X$. Throughout this section $p \in [1, \infty)$ and X is a complete and separable metric space equipped with a Radon measure which is finite and positive on balls, i.e. $0 < \mu(B(x, r)) < \infty$ for each $x \in X, r > 0$. The spaces $L^p(X)$ consist of p-integrable functions and $L^p_{\text{loc}}(X)$ consist of locally p-integrable functions, i.e. those f so that for every bounded set $A \subset X$ we have $\int_A |f|^p \, d\mu < \infty$. These spaces are equipped with the usual notions of convergence. A curve is a continuous map $\gamma : I \to X$ from a compact interval $I \subset \mathbb{R}$. A curve is rectifiable, if it has finite length $\text{Len}(\gamma)$, see [21, Chapter 5].

Recall that, as introduced by Heinonen and Koskela in [20], a non-negative Borel function $g : X \to [0, \infty]$ is a (true) upper gradient of (or for) $f : X \to [-\infty, \infty]$, if

$$\int_\gamma g \, ds \geq |f(\gamma(1)) - f(\gamma(0))|, \tag{2.1}$$

for any rectifiable curve $\gamma : [0, 1] \to X$. If either $f(\gamma(1))$, or $f(\gamma(0))$, is infinity, we interpret the right hand side as ∞.

The property of being an upper gradient is not closed under $L^p(X)$-convergence. Thus, one introduces the notion of a *p-weak upper gradient*. One simple definition for this is that if $\text{UG}(f) \subset L^p(X)$ is the collection of upper gradients of f that are contained in $L^p(X)$, then g is a p-weak upper gradient if $g \in \overline{\text{UG}(f)}$ (i.e. lies in the closure in $L^p(X)$). Equivalently, this notion can be defined using a notion of "a.e." curves coming from the concept of a modulus of curve families. We say that g is a p-weak upper gradient, if (2.1) holds for p-a.e. rectifiable curve γ: i.e. if there exists a $h \in L^p(X)$, so that for every rectifiable curve γ for which $\int_\gamma h \, ds < \infty$ we have inequality (2.1). We refer the reader to [15, 35] for details on modulus.

Remark 2.2 We prefer not to define modulus of curve families here, since we do not directly need it. Instead, we only need some properties of minimal p-weak upper gradients (which often are proved using modulus techniques). The main property we need is that if g is a p-weak upper gradient for f, then for any $\epsilon > 0$ there is a lower semi-continuous (true) upper gradient $g_\epsilon \geq g$ with $\int_X |g - g_\epsilon|^p \, d\mu \leq \epsilon$. This can be easily seen from $g \in \overline{\text{UG}(f)}$ together with the fact that any function in $L^p(X)$ can be approximated from above by a lower semi-continuous function (by the Vitali-Carathéodory Theorem). For the details we refer to [21, Sections 4.2 and Chapters 5–6].

We define $N^{1,p}(X)$ as the collection of $f \in L^p(X)$ so that there exists a Borel upper gradient $g \in L^p(X)$. The functions in this collections are called Newton, Sobolev, or Newton-Sobolev functions. The space is called either the Newton, Sobolev, or Newton-Sobolev space. We define

$$\|f\|_{N^{1,p}} := \left(\inf_{g \in \overline{\text{UG}(f)}} \|f\|^p_{L^p} + \|g\|^p_{L^p} \right)^{1/p}, \tag{2.3}$$

where the infimum is taken over all p-weak upper gradients of f (or, equivalently, over all true upper gradients g of f). By [15, Theorem 7.16] there always exists a minimal p-weak upper gradient $g_f \in \overline{\text{UG}(f)} \subset L^p(X)$ which attains the infimum in (2.3), and which satisfies (2.1) for p-almost every curve. See also [35] for the $p > 1$ case. Since the set $\overline{\text{UG}(f)}$ of

p-weak upper gradients satisfies the lattice property [21, Corollary 6.3.12], we have that the minimal p-weak upper gradient satisfies the additional property that if $\tilde{g}$ is another p-weak upper gradient for f, then $\tilde{g} \geq g_f$ a.e. Throughout this paper, if $f \in N^{1,p}(X)$, then g_f will denote its minimal p-weak upper gradient.

We also define the Lipschitz and asymptotic Lipschitz constant for a function $f : X \to \mathbb{R}$

$$\text{LIP}[f](A) := \sup_{x,y \in A, x \neq y} \frac{|f(x) - f(y)|}{d(x, y)}, \quad \text{lip}_a[f](x) := \lim_{r \to 0} \text{LIP}[f](B(x, r)). \quad (2.4)$$

When A is a singleton, we interpret $\text{LIP}[f](A) = 0$. A function $f : X \to \mathbb{R}$ is *Lipschitz* if $\text{LIP}[f](X) < \infty$. The collection of Lipschitz functions $f : X \to \mathbb{R}$ is denoted $\text{LIP}[f](X)$ and $\text{LIP}_b(X)$ is the collection of $f \in \text{LIP}[f](X)$, with bounded support (i.e. there exists some ball $B(x_0, R) \subset X$ so that $f(y) = 0$ for each $y \notin B(x_0, R)$). A function f is called L-Lipschitz if $\text{LIP}[f](X) < L$.

Example 2.5 Lipschitz functions yield examples of Sobolev functions. If $f \in \text{LIP}[f]$, then $\text{lip}[f] \leq \text{lip}_a[f] \leq \text{LIP}[f](X)$, and $\text{lip}[f]$ is an upper gradient of f. Indeed, $\text{lip}[f]$ controls the directional derivative along any rectifiable curve and thus (2.1) follows; see [21, Lemma 6.2.6] for a detailed argument. Consequently, $\text{LIP}_b(X) \subset N^{1,p}(X)$ for each $p \in [1, \infty)$ and $g_f \leq \text{lip}_a[f]$ for each $f \in \text{LIP}_b(X)$.

We will need some results from functional analysis. First, we need an argument which replaces the usual application of reflexivity when $p > 1$. This involves the following definition.

Definition 2.6 Let $p \in [1, \infty)$. A collection of functions $\mathcal{F} \subset L^p(X)$ is **strongly** p-equi-integrable if,

a) for every $\epsilon > 0$, there exists, a set $E \subset X$ with $\mu(E) < \infty$ so that $\int_{X \setminus E} |f|^p \, d\mu \leq \epsilon$ for all $f \in \mathcal{F}$;

b) $\sup_{f \in \mathcal{F}} \|f\|_{L^p(X)} < \infty$; and

c) for every $\epsilon > 0$ there exists a $\delta > 0$ so that if $\mu(E) < \delta$, then $\int_E |f|^p \, d\mu \leq \epsilon$ for all $f \in \mathcal{F}$.

In the literature, the condition p-equi-integrability usually only includes condition c). In order to state more compact theorems and proofs below, we add the conditions a) and b) and add the modifier "strongly" to the term. If the measure is non-atomic, then by decomposing the measure to sufficiently small parts, one can show that a) and c) imply b).

The following theorem is classical - although often stated only for finite measure spaces. See [9, Theorem IV.8.9] or [12, Theorem 2.54].

Theorem 2.7 *(Dunford-Pettis theorem) A collection $\mathcal{F} \subset L^1(X)$ is pre-compact if and only if it is strongly* 1-*equi-integrable.*

We will further need the Vitali-Convergence Theorem. A sequence of functions $f_n \in L^p(X)$ converges in measure to a function $f \in L^p(X)$ if and only if for every $\epsilon > 0$ it holds that $\lim_{n \to \infty} \mu(\{x : |f_n(x) - f(x)| \geq \epsilon\}) = 0$. For a proof, see [12, Theorem 2.24].

Theorem 2.8 *(Vitali-Convergence Theorem) A sequence $(f_n)_{n \in \mathbb{N}}$ of functions $f_n \in L^p(X)$ converges to $f \in L^p(X)$ if and only if it converges in measure to f and is strongly p-equi-integrable.*

We apply these classical results to show that minimal p-weak upper gradients converge, if they possess some upper bounds which converge.

Lemma 2.9 *Suppose that f_i, $f \in N^{1,p}(X)$ are functions with minimal p-weak upper gradients g_{f_i}, $g_f \in L^p(X)$ and that $f_i \to f$ in $L^p(X)$. Suppose further that $\tilde{g}_i \in L^p(X)$ are functions which converge $\tilde{g}_i \to g_f$ in $L^p(X)$. If $g_{f_i} \leq \tilde{g}_i$, then $g_{f_i} \to g_f$ in $L^p(X)$.*

Proof First, we will show that the sequence g_{f_i} converges to g_f weakly in $L^p(X)$. To this end, it suffices to show that every subsequence of g_{f_i} has a further subsequence that converges to g_f. After relabelling indices, without loss of generality, it suffices to find such a subsequence for g_{f_i}.

First, we show that g_{f_i} has a weakly convergent subsequence in $L^p(X)$, which converges to some function $\tilde{g}$. If $p > 1$, this follows from reflexivity of $L^p(X)$. If $p = 1$, then the inequality $g_{f_i} \leq \tilde{g}_i$ and the fact that $\tilde{g}_i$ converge in $L^1(X)$ easily imply that the sequence of functions g_{f_i} is strongly 1-equi-integrable. Therefore, by the Dunford-Pettis Theorem 2.7, the sequence has a weakly convergent subsequence. We will now denote this subsequence by g_{f_i}, i.e. $g_{f_i} \to \tilde{g}$ weakly in $L^p(X)$.

It follows from [15, Lemma 7.8] that $\tilde{g}$ is a p-weak upper gradient of f and hence $\tilde{g} \geq g_f$ a.e. On the other hand, the weak lower semicontinuity of L^p-norms gives

$$\int_X \tilde{g}^p d\mu \leq \liminf_{i \to \infty} \int_X g_{f_i}^p d\mu \leq \liminf_{i \to \infty} \int_X \tilde{g}_i^p d\mu = \int_X g_f^p d\mu, \qquad (2.10)$$

which together with $\tilde{g} \geq g_f$ yields that $\tilde{g} = g_f$ a.e. Thus $g_{f_i} \to g_f$ weakly in $L^p(X)$.

Since the sequence $(g_{f_i})_{i \in \mathbb{N}}$ is strongly p-equi-integrable, by the Vitali-Convergence Theorem 2.8, it will converge to g_f if it converges in measure. Further, it follows from part c) of Definition 2.6, and the strong p-equi-integrability of the sequence g_{f_i}, that it suffices to prove convergence in measure on all sets of finite measure, i.e.

$$\limsup_{i \to \infty} \mu(\{x \in A : |g_{f_i}(x) - g_f(x)| > \epsilon\}) = 0$$

for every $\epsilon > 0$ and every $A \subset X$ with $\mu(A) < \infty$. Fix such a set $A \subset X$ and $\epsilon > 0$. We have,

$$\mu(\{x \in A : |g_{f_i}(x) - g_f(x)| > \epsilon\}) \leq \frac{1}{\epsilon} \int_A |g_f - g_{f_i}| d\mu$$

$$= \frac{1}{\epsilon} \left(2 \int_{A \cap \{g_{f_i} > g_f\}} g_{f_i} - g_f d\mu + \int_A g_f - g_{f_i} d\mu \right).$$

Since g_{f_i} converge weakly to g_f, we get $\lim_{i \to \infty} \int_A g_f - g_{f_i} d\mu = 0$. In order to estimate the first term, recall that $g_{f_i} \leq \tilde{g}_i$ and $\tilde{g}_i \to g_f$ in $L^p(X)$. Then,

$$\limsup_{i \to \infty} \int_{A \cap \{g_{f_i} > g_f\}} g_{f_i} - g_f d\mu \leq \limsup_{i \to \infty} \int_{A \cap \{g_{f_i} > g_f\}} \tilde{g}_i - g_f d\mu$$

$$\leq \limsup_{i \to \infty} \int_A |\tilde{g}_i - g_f| d\mu = 0.$$

By combining the two previous limits, we get

$$\limsup_{i \to \infty} \mu(\{x \in A : |g_{f_i}(x) - g_f(x)| > \epsilon\}) = 0.$$

□

We need a version of the Arzelà-Ascoli Theorem, which is easy to prove using standard techniques. We state this lemma to highlight the fact that our theorems apply to general complete and separable metric spaces, whereas most of the existing literature uses an assumption of properness. Indeed, we avoid properness by adding the assumption that the sets A_t that appear in the statement of the lemma are pre-compact. Consequently, we will be able to apply the lemma with $Y = \ell^\infty(\mathbb{N})$ which is not proper.

A curve $\gamma : [0, 1] \to X$ is L-Lipschitz if $d(\gamma(s), \gamma(t)) \leq L|s - t|$ for every $s, t \in [0, 1]$.

Lemma 2.11 *Let $L \in [0, \infty)$. Suppose that Y is a complete metric space. Let $\gamma_k : [0, 1] \to Y$ be a sequence of L-Lipschitz curves so that for every $t \in [0, 1]$ the set $A_t = \{\gamma_k(t) : k \in \mathbb{N}\}$ is pre-compact in Y. There exists a subsequence of γ_k which converges uniformly to a L-Lipschitz curve γ.*

2.2 Approximating function

We introduce in this subsection the formula of our approximation and prove its main properties.

A *discrete path P* (or simply "a path" P) is a sequence of points $P = (p_0, \ldots, p_n)$ with $p_k \in X$ for each $k = 0, \ldots, n$, and $n \in \mathbb{N}$ with $n \geq 1$. We define the mesh of P by $\mathrm{Mesh}(P) := \max_{k=0,\ldots,n-1} d(p_k, p_{k+1})$, the diameter of P by $\mathrm{diam}(P) := \max_{k,l} d(p_k, p_l)$ and the length of P by $\mathrm{Len}(P) := \sum_{k=0}^{n-1} d(p_k, p_{k+1})$. By a slight abuse of notation, we will write $p \in P$ if there is a $k = 0, \ldots, n$ so that $p_k = p$. Further, we write $P \subset U$ for a subset $U \subset X$, if $p_k \in U$ for each $k = 0, \ldots, n$. We write $P \subset Q$ if the sequence of points in P forms a subsequence of the points in Q without gaps: that is $P = (p_0, \ldots, p_m)$ and $Q = (q_0, \ldots, q_n), m \leq n$ and there exists a non-negative $0 \leq s \leq n - m$ so that $q_{s+k} = p_k$ for each $k = 0, \ldots, m$. Such a path P is called a *sub-path* of Q.

For $\delta > 0$ and $A \subset X$ a fixed closed set, we say that a discrete path $P = (p_0, \ldots, p_n)$ is (δ, A, x)-admissible, if $\mathrm{Mesh}(P) \leq \delta$, $p_0 \in A$ and $p_n = x$. The collection of all (δ, A, x) admissible discrete paths is denoted $\mathcal{P}(\delta, A, x)$.

Suppose that $f : X \to [0, M]$ is a bounded function for some $M > 0$, $g : X \to [0, \infty)$ is a continuous bounded function, and $A \subset X$ is a closed subset. Then, for $\delta > 0$ we define an *approximating* function $\tilde{f}$ with data (f, g, A, M, δ) as

$$\tilde{f}(x) := \min\{M, \inf_{(p_0,\ldots,p_n)\in\mathcal{P}(\delta,A,x)} f(p_0) + \sum_{k=0}^{n-1} g(p_k)d(p_k, p_{k+1})\}. \tag{2.12}$$

Lemma 2.13 *Let $M, \delta > 0$, $f : X \to [0, M]$, $A \subset X$ be closed, and let $g : X \to [0, \infty)$ be continuous and bounded. The function $\tilde{f}$ in Formula (2.12) satisfies the following properties.*

A) $\tilde{f} : X \to [0, M]$.
B) For each $x \in A$, $0 \leq \tilde{f}(x) \leq f(x)$.
C) If $x \in A$ and $f(x) = 0$, then $\tilde{f}(x) = 0$.
D) For each $x, y \in X$ with $d(x, y) \leq \delta$, we have

$$|\tilde{f}(x) - \tilde{f}(y)| \leq \max\{g(x), g(y)\}d(x, y). \tag{2.14}$$

E) $\mathrm{lip}_a[\tilde{f}](x) \leq g(x)$ for every $x \in X$.
F) $\tilde{f}$ is $\max\{M\delta^{-1}, \sup_{x\in X} g(x)\}$-Lipschitz.

Proof A) First, since each term in the infimum in (2.12) is non-negative, $\tilde{f}$ is also non-negative. Further, $\tilde{f}(x) \leq M$ for each $x \in X$ follows immediately from the definition.

B) Let $x \in A$. By noting that the discrete path $P = (x, x)$ is (δ, A, x)-admissible, we get
   $\tilde{f}(x) \leq f(x) + g(x)d(x, x) = f(x)$.
C) Let $x \in A$. By B) and A) we get $0 \leq \tilde{f}(x) \leq f(x) = 0$.
D) Let $x, y \in X$ be arbitrary with $d(x, y) \leq \delta$. Consider an arbitrary discrete (δ, A, x)-admissible path $P = (p_0, \ldots, p_n)$. We form a (δ, A, y)-admissible path $Q = (q_0, \ldots, q_{n+1})$ by adjoining $q_{n+1} = y$ and setting $q_i = p_i$ for $i \in \{0, \ldots, n\}$. With such choices

$$\tilde{f}(y) \leq f(q_0) + \sum_{k=0}^{n} g(q_k)d(q_k, q_{k+1}) = f(p_0) + \sum_{k=0}^{n-1} g(p_k)d(p_k, p_{k+1}) + g(x)d(x, y)$$

Taking an infimum over all (n, A, x)-admissible paths and a minimum with M on both sides of the inequality, we get $\tilde{f}(y) \leq \tilde{f}(x) + g(x)d(x, y)$. By switching the roles of x and y, we obtain Inequality (2.14).
E) Take points $x, a, b \in X$. Then, by sending $a, b \to x$ and applying (2.14) to the pair of points a, b together with the continuity of g, we get

$$\mathrm{lip}_a[\tilde{f}](x) \leq g(x).$$

F) Let $L := \max\{M\delta^{-1}, \sup_{x \in X} g(x)\}$. If $x, y \in X$, with $d(x, y) \leq \delta$, then from D) we get $|\tilde{f}(x) - \tilde{f}(y)| \leq Ld(x, y)$. On the other hand, if $x, y \in X$ with $d(x, y) > \delta$, then by A) we have $|\tilde{f}(x) - \tilde{f}(y)| \leq M \leq M\delta^{-1}d(x, y) \leq Ld(x, y)$. □

2.3 A compactness result for discrete curves

The convergence of our approximating functions to f is obtained via a contradiction argument, where we are given a sequence of discrete paths P^i, and then extract a subsequence converging to a curve $\gamma : [0, 1] \to X$. In this subsection, we give the definition of convergence that we use and the main results for it.

First, to define our notion of convergence we fix an isometric Kuratowski embedding $\iota : X \to \ell^\infty(\mathbb{N})$, and identify X with its image under this embedding. *A priori* the notions that follow may depend on the choice of such an embedding. For our ultimate argument, such a dependence will play no role, and thus we do not analyse it much further. However, we will later indicate, how to prove that the definitions, in fact, are independent of the choice of such an embedding.

For a subset $A \subset X$ and $x \in A$, the distance from a to A is given by $d(x, A) := \inf_{a \in A} d(a, x)$. If $P = (p_0, \ldots, p_n)$ is a discrete path, we define its *linearly interpolating curve* as follows. If $\mathrm{Len}(P) = 0$, we define $\gamma_P : [0, 1] \to \ell^\infty(\mathbb{N})$ by $\gamma(t) = p_0$ for each $t \in [0, 1]$. If $\mathrm{Len}(P) > 0$ we define the *sequence of interpolating times* $T_P = (t_0, \ldots, t_n)$ by $t_0 = 0$ and $t_k = \sum_{i=0}^{k-1} d(p_i, p_{i+1})/\mathrm{Len}(P)$ for $k = 1, \ldots, n$. Then, we define $\gamma_P : [0, 1] \to \ell^\infty(\mathbb{N})$ piecewise by linear interpolation in $\ell^\infty(\mathbb{N})$: we set $\gamma_P(t_k) = p_k$, and when $t \in [t_k, t_{k+1}]$ and $t_{k+1} > t_k$, we set $\gamma(t) = ((t - t_k)p_k + (t_{k+1} - t)p_{k+1})(t_{k+1} - t_k)^{-1}$. The following lemma is elementary to verify.

Lemma 2.15 *If P is a discrete path, and γ_P is its linearly interpolating curve, then γ_P is $\mathrm{Len}(P)$-Lipschitz, parametrized by constant speed, $\mathrm{Len}(P) = \mathrm{Len}(\gamma_P)$ and for each $t \in [0, 1]$ there exists a point $p \in P \subset X$ so that $d(\gamma_P(t), p) \leq \mathrm{Mesh}(P)$.*

We say that a sequence of discrete paths P^i *converges* to a curve $\gamma : [0, 1] \to X$, if γ_{P^i} converges uniformly to γ and if $\lim_{i \to \infty} \mathrm{Mesh}(P^i) = 0$. While we do not need this, we

note that this notion of convergence does not depend on the embedding to $\ell^\infty(\mathbb{N})$ and could be defined intrinsically. Indeed, we could also define paths with jumps $\gamma'_P : [0, 1] \to X$ piecewise: if $t \in [t_k, t_{k+1})$ we set $\gamma_P(t_k) = p_k$. Then, we say that P^i converges to γ if γ'_{P^i} converges uniformly to γ. It is straightforward to show that these two definitions are equivalent. However, by using the linearly interpolating curves we avoid some unfortunate technical issues later.

We need compactness results for discrete paths. First, we give a simpler compactness statement in order to illustrate the main idea in the setting where X is compact.

Lemma 2.16 *If X is a compact metric space, and if $\{P^i\}_{i \in \mathbb{N}}$ is a sequence of discrete paths with $\lim_{i \to \infty} \mathrm{Mesh}(P^i) = 0$ and $\sup_{i \in \mathbb{N}} \mathrm{Len}(P^i) < \infty$, then there exists a subsequence i_k so that P^{i_k} converges to a curve $\gamma : [0, 1] \to X$.*

Proof Let $\gamma_{P^i} : [0, 1] \to \ell^\infty(\mathbb{N})$ be the linearly interpolating curves and $L := \sup_{i \in \mathbb{N}} \mathrm{Len}(P^i) < \infty$. To prove the claim, we first need to show that the sequence of linearly interpolating curves γ_{P^i} satisfies the assumptions of Lemma 2.11. First, Lemma 2.15 gives that each γ_{P^i} is L-Lipschitz. Second, let $t \in [0, 1]$, and consider $A_t := \{\gamma_{P^i}(t) : i \in \mathbb{N}\}$. We will show that A_t is precompact in $\ell^\infty(\mathbb{N})$ by showing that A_t is totally bounded.

Fix $\eta > 0$, and $N \in \mathbb{N}$ so that for $i \geq N$ we have $\mathrm{Mesh}(P^i) \leq \eta$. Let $K := X \cup \{\gamma_{P^1}(t), \ldots, \gamma_{P^N}(t)\}$. We claim that $d(a, K) \leq \eta$ for each $a = \gamma_{P^i}(t) \in A_t$. For each $i \leq N$ this claim is trivial. For each $i > N$, by Lemma 2.15, there exists a $p^i \in P^i$ so that $d(\gamma_{P^i}(t), p^i) \leq \mathrm{Mesh}(P^i) \leq \eta$. Since $p^i \in X$, we have $d(\gamma_{P^i}(t), X) \leq \eta$.

The set K is compact, and thus totally bounded. Thus, we can find points $x_1, \ldots, x_M \in K$ so that $K \subset \bigcup_{j=1}^M B(x_j, \eta)$. Combining with the previous paragraph, we get $A_t \subset \bigcup_{j=1}^M B(x_j, 2\eta)$, which proves the totally boundedness since $\eta > 0$ is arbitrary.

Now, by Lemma 2.15, we have that a subsequence of γ_{P^i} converges to some curve $\gamma : [0, 1] \to \ell^\infty(\mathbb{N})$. By Lemma 2.15, for each $t \in [0, 1]$, we have $d(\gamma(t), X) \leq \limsup_{i \to \infty} d(\gamma_{P^i}(t), X) \leq \lim_{i \to \infty} \mathrm{Mesh}(P^i) = 0$, and thus the image of γ is contained in X. $\qquad\square$

If we were to work only in proper metric spaces, the previous lemma would be quite sufficient for the proof of Theorem 1.1. However, for non-proper spaces, we need to force the discrete curves to "not pass far" from a sequence of compact sets.

First, let $\{K_n\}_{n \in \mathbb{N}}$ be an increasing sequence of non-empty compact sets $K_n \subset X$, with $K_n \subset K_m$ for $n \leq m$. One should imagine these sets coming from the tightness of the measure μ on the complete and separable metric space X. We call a sequence of continuous bounded functions $h_n : \ell^\infty(\mathbb{N}) \to [0, \infty]$ defined by

$$h_n(x) := \sum_{k=1}^n \min\{nd(x, K_k), 1\}. \tag{2.17}$$

a good sequence of functions for $\{K_k\}_{k \in \mathbb{N}}$.

These functions penalize paths that travel far away from the compact sets K_n. Indeed, if $d(x, K_n) \geq \eta$, then $h_n(x) \geq n \min\{n\eta, 1\}$. If we *assume* that a sum involving h_n over a discrete path P is controlled, then we can use this bound effectively to conclude that P is contained within an η-neighborhood of K_n (for some $n \in \mathbb{N}$).

Since the proof of the following Lemma is nearly identical to the previous Lemma, we will slightly abbreviate the proof, and utilize the same notation.

Lemma 2.18 *Let $K_n \subset X$ be an increasing sequence of compact sets and let $M, L, \Delta > 0$ be constants. Let further h_n be a good sequence of functions for $\{K_k\}_{k \in \mathbb{N}}$. If $\{P^i =$*

$(p_0^i, \ldots, p_{n(i)}^i)\}_{i \in \mathbb{N}}$ *is a sequence of discrete paths, with* $\lim_{i \to \infty} \text{Mesh}(P^i) = 0,$ $\sup_{i \in \mathbb{N}} \text{Len}(P^i) \leq L,$ $\inf_{i \in \mathbb{N}} \text{diam}(P^i) \geq \Delta$ *and*

$$\sum_{k=0}^{n(i)-1} h_i(p_k^i) d(p_k^i, p_{k+1}^i) \leq M$$

for each $i \in \mathbb{N}$*, then there exists a subsequence of* P^i *which converges to a curve* $\gamma : [0, 1] \to X$.

Proof The proof proceeds in the same way as Lemma 2.16. First, for every $i \in \mathbb{N}$ the linearly interpolating curve γ_{P^i} is L−Lipschitz. Second, if a subsequence of γ_{P^i} converges to some curve $\gamma : [0, 1] \to \ell^\infty(\mathbb{N})$, then by Lemma 2.15, we have $d(\gamma(t), X) \leq \lim \sup_{i \to \infty} d(\gamma_{P^i}(t), X) \leq \lim_{i \to \infty} \text{Mesh}(P^i) = 0$ for each $t \in [0, 1]$, and thus the image of γ is contained in X.

Thus, we only need to verify the second assumption of Lemma 2.11, i.e. that the set $A_t = \{\gamma_{P^k}(t) : k \in \mathbb{N}\}$ is precompact in $\ell^\infty(\mathbb{N})$ for each $t \in [0, 1]$. Equivalently, we need to show that it is totally bounded. This in turn is equivalent to the fact that for every $\eta > 0$ we can find a compact set K so that $d(a, K) \leq \eta$ for each $a \in A_t$.

Fix $\eta > 0$ and $t \in [0, 1]$. Fix $N \in \mathbb{N}$ so that $\text{Mesh}(P^i) \leq \eta/8$ for each $i \geq N$. Then, fix $T := 8\lceil \eta^{-1} \rceil + \lceil 2M \min\{\eta/8, \Delta\}^{-1} \rceil$, and set $\tilde{N} := \max\{N, T\}$. Set $K := K_{\tilde{N}} \cup \{\gamma_{P^i}(t) : i = 1, \ldots, \tilde{N}\}$.

Take any $i \in \mathbb{N}$ and consider $a = \gamma_{P^i}(t) \in A_t$. If $i \leq \tilde{N}$, we have $d(\gamma_{P^i}(t), K) = 0 \leq \eta$. Next, assume $i \geq \tilde{N}$. By Lemma 2.15, there exists a $p^i \in P^i$ so that $d(p^i, \gamma_{P^i}(t)) \leq \text{Mesh}(P^i) \leq \eta/8$. If $d(p^i, K_{\tilde{N}}) \leq \eta/2$ for each $i \geq \tilde{N}$, then $d(\gamma_{P^i}(t), K) \leq \eta$, and we are done. Suppose therefore that we have some index $i \geq \tilde{N}$, with $d(p^i, K_{\tilde{N}}) \geq \eta/2$. We will show that this leads to a contradiction and thus completes the proof.

Let $Q^i = (q_0^i, \ldots, q_{m(i)}^i) \subset P^i$ be the largest sub-path of P^i which contains p^i and which is contained in $B(p^i, \eta/4)$. Since $\text{Mesh}(P^i) \leq \eta/8$, and $\text{diam}(P^i) \geq \Delta$, we have $\text{diam}(Q^i) \geq \min\{\eta/8, \Delta\}$. Then, $d(q, K_{\tilde{N}}) \geq \eta/4 \geq \tilde{N}^{-1}$ for each $q \in Q^i$. Thus, $d(q, K_j) \geq \tilde{N}^{-1}$ for each $q \in Q^i$ and $j \leq \tilde{N}$. In particular $h_i(q) \geq h_{\tilde{N}}(q) \geq \tilde{N} \geq \lceil 2M \min\{\eta/8, \Delta\}^{-1} \rceil$, where we used Formula (2.17). Thus,

$$\sum_{k=0}^{m(i)-1} h_i(q_k^i) d(q_k^i, q_{k+1}^i) \geq \lceil 2M \min\{\eta/8, \Delta\}^{-1} \rceil \text{Len}(Q^i) \geq 2M.$$

Since $Q \subset P$, we have

$$2M \leq \sum_{k=0}^{m(i)-1} h_i(q_k^i) d(q_k^i, q_{k+1}^i) \leq \sum_{k=0}^{n(i)-1} h_i(p_k^i) d(p_k^i, p_{k+1}^i) \leq M,$$

which is a contradiction.   □

We will also need a slightly technical lower semicontinuity statement reminiscent of [24, Proposition 4]. The sums that appear in the statement, and in Lemma 2.18, should be thought of as discrete Riemann sums.

Lemma 2.19 *Let* $g : X \to [0, \infty]$ *be a lower semicontinuous function, and assume that* $\{g_i : X \to [0, \infty]\}_{i \in \mathbb{N}}$ *is an increasing sequence of continuous functions which converges to* g *pointwise, with* $g_i(x) \leq g_j(x)$ *for each* $x \in X$ *and each* $i \leq j$*. If* $\{P^i = (p_0^i, \ldots, p_{n(i)}^i)\}_{i \in \mathbb{N}}$

is a sequence of discrete paths, with $\sup_{i \in \mathbb{N}} \mathrm{Len}(P^i) < \infty$ and which converges to a curve $\gamma : [0, 1] \to X$, then

$$\int_\gamma g \, ds \leq \liminf_{i \to \infty} \sum_{k=0}^{n(i)-1} g_i(p_k^i) d(p_k^i, p_{k+1}^i).$$

Proof Let γ_{P^i} be the linearly interpolating curves to P^i and let $L := \sup_{i \in \mathbb{N}} \mathrm{Len}(P^i)$. Use the Tietze Extension Theorem to extend each $g_i : X \to \mathbb{R}$ to be a continuous function on $\ell^\infty(\mathbb{N})$. By constructing the extensions recursively and taking maxima, we can ensure $g_i \leq g_j$ for $i \leq j$. Extend the function g to all of $\ell^\infty(\mathbb{N})$ by the formula $g(x) := \lim_{i \to \infty} g_i(x)$ where $x \in \ell^\infty(\mathbb{N})$. Since $(g_i)_{i \in \mathbb{N}}$ is an increasing sequence of continuous functions, it follows that g is lower semicontinuous. Denote the extensions still by the same letter.

Fix for the moment an index $i \in \mathbb{N}$ and consider the function g_i. Since γ_{P^j} converges uniformly to γ, when $j \to \infty$, we have that the set $K \subset \ell^\infty(\mathbb{N})$ formed by the union of the images of the curves γ_{P^j} ($j \in \mathbb{N}$) and γ is compact. On K, the function g_i is uniformly continuous.

Fix an $\epsilon > 0$. Since $g_i|_K$ is uniformly continuous, there exists a $\delta > 0$ so that if $x, y \in K$ and $d(x, y) \leq \delta$, then $|g_i(x) - g_i(y)| \leq \epsilon/L$. Choose then N so large that $\mathrm{Mesh}(P^j) \leq \delta$ for each $j \geq N$. If $T_{P^j} = (t_0^j, \ldots, t_{n(j)}^j)$ is the sequence of interpolating times for P^j, then $d(\gamma_{P^j}(t), p_k^j) \leq \mathrm{Mesh}(P^j) \leq \delta$ for each $t \in [t_k^j, t_{k+1}^j]$ ($k = 0, \ldots, n(j)-1$). Consequently, for each $j \geq N$, we get

$$\left| \int_{\gamma_{P^j}} g_i \, ds - \sum_{k=0}^{n(j)-1} g_i(p_k^j) d(p_k^j, p_{k+1}^j) \right|$$

$$= \left| \sum_{k=0}^{n(j)-1} \int_{\gamma_{P^j}|_{[t_k^j, t_{k+1}^j]}} g_i \, ds - \sum_{k=0}^{n(j)-1} g_i(p_k^j) d(p_k^j, p_{k+1}^j) \right|$$

$$= \left| \sum_{k=0}^{n(j)-1} \int_{\gamma_{P^j}|_{[t_k^j, t_{k+1}^j]}} (g_i(\cdot) - g_i(p_k^j)) \, ds \right|$$

$$\leq \mathrm{Len}(P^j) \epsilon / L \leq \epsilon.$$

On the last line we used the fact that $\mathrm{Len}(P^j) = \mathrm{Len}(\gamma_{P^j})$ by Lemma 2.15. Since $\epsilon > 0$ was arbitrary, by sending $j \to \infty$ we get

$$\lim_{j \to \infty} \left| \int_{\gamma_{P^j}} g_i \, ds - \sum_{k=0}^{n(j)-1} g_i(p_k^j) d(p_k^j, p_{k+1}^j) \right| = 0. \tag{2.20}$$

By combining this with the lower semi-continuity of curve integrals (see e.g. the argument in [24, Proposition 4]) and the fact that $g_i \leq g_j$ for $i \leq j$, we have for each $i \in \mathbb{N}$ that

$$\int_\gamma g_i \, ds \leq \liminf_{j \to \infty} \int_{\gamma_{P^j}} g_i \, ds \stackrel{(2.20)}{=} \liminf_{j \to \infty} \sum_{k=0}^{n(j)-1} g_i(p_k^j) d(p_k^j, p_{k+1}^j)$$

$$\leq \liminf_{j \to \infty} \sum_{k=0}^{n(j)-1} g_j(p_k^j) d(p_k^j, p_{k+1}^j).$$

Now, by letting $i \to \infty$ and by using monotone convergence on the left hand side, we obtain the statement of the lemma.　　　　　　　　　　　　　　　　　　　　　　　□

2.4 Proof of Theorem 1.1

With the stage set, we are now able to prove the main theorem of the paper. The same proof, up to cosmetic changes, also gives Theorem 1.2. The main change, is replacing the convergence of f_n in $L^p(X)$ with convergence in L^1_{loc}, and instead of $N^{1,p}(X)$ one employs the Dirichlet space defined in [21, Section 7.1] (alternatively, see [3]).

The argument will first use cut-off functions and truncation to reduce the problem to approximating non-negative and bounded functions with bounded support. This is based on the following lemma.

Lemma 2.21 *If Theorem 1.1 is true for every $f \in N^{1,p}(X)$ with $f : X \to [0, M]$ for some $M > 0$ and with $f|_{X \setminus B(x_0, R)} = 0$ for some $x_0 \in X$, $R > 0$, then Theorem 1.1 is true for all $f \in N^{1,p}(X)$.*

Proof Let $f \in N^{1,p}(X)$ be arbitrary. We will reduce the claim of Theorem 1.1 to first non-negative functions, and then to bounded functions, and finally to ones with bounded support. *Reduction to non-negative:* We can write $f = f_+ - f_-$, where $f_+ = \max\{f, 0\}$ and $f_- = \max\{-f, 0\}$ with $f_\pm \in N^{1,p}(X)$. Then $g_f = g_{f_+} + g_{f_-}$ by [21, Proposition 6.3.22][1]. If Theorem 1.1 is true for $f_\pm$, then we can find a sequence of functions $f^n_\pm \in \text{LIP}_b(X)$ with $\text{lip}_a[f^n_\pm] \to g_{f_\pm}$ in $L^p(X)$. Let $f_n = f^n_+ - f^n_-$ and note that $g_{f_n} \leq \text{lip}_a[f_n] \leq \text{lip}_a[f^n_+] + \text{lip}_a[f^n_-]$. Since $\text{lip}_a[f^n_+] + \text{lip}_a[f^n_-] \to g_{f_+} + g_{f_-} = g_f$ in $L^p(X)$, Lemma 2.9 gives that $\text{lip}_a[f_n] \to g_f$ and $g_{f_n} \to g_f$ in $L^p(X)$. In other words, $f_n \in \text{LIP}_b(X)$ approximates f in energy. Thus, without loss of generality it suffices to prove the claim for all non-negative functions f. Let $f : X \to [0, \infty)$ be an arbitrary non-negative function in $N^{1,p}(X)$. *Reduction to a bounded case:* Consider functions $f_M = \min\{f, M\}$ for all $M > 0$. Then, we have $f_M \to f$ in $L^p(X)$ as $M \to \infty$. By [21, Proposition 6.3.22], the function $g_M := g_f \cdot 1_{X \setminus f^{-1}[0,M]}$ is a minimal p-weak upper gradient for $f - f_M$. One directly observes that $g_M \to 0$ in $L^p(X)$. Thus, f_M converges to f in norm in $N^{1,p}(X)$. In particular, $g_{f_M} \to g_f$ in $L^p(X)$. If Theorem 1.1 is true for each f_M, then there exist sequences $\{f^n_M\}_{n \in \mathbb{N}}$ which converge in energy to f_M and with $\text{lip}_a[f^n_M] \to g_{f_M}$ in $L^p(X)$ as $n \to \infty$. By a diagonal argument, with $M \to \infty$ together with $n \to \infty$, we get the claim of Theorem 1.1 for f.

Finally, we can assume that $f \in N^{1,p}(X)$ and that for some $M > 0$ we have $0 \leq f \leq M$. *Reduction to bounded support:* Let $x_0 \in X$ be any fixed point, and consider the functions $f_R(x) = f \psi_R(x)$, where $\psi_R(x) = \max\{0, \min\{1, R - d(x_0, x)\}\}$ and $R \in \mathbb{N}$, $R > 0$. The functions ψ_R are 1-Lipschitz and $0 \leq \psi_R \leq 1$. Further, $f_R \to f$ pointwise and in $L^p(X)$. Further, the function $f - f_R$ has a weak upper gradient $g_R = 1_{X \setminus B(x_0, R-1)} g_f + f 1_{B(x_0, R+1) \setminus B(x_0, R-1)}$ as follows from the Leibniz rule in [21, Proposition 6.3.28 and Proposition 6.3.22]. Thus $g_R \to 0$ and we get that $f_R \to f$ in the norm of $N^{1,p}(X)$. Therefore, functions with bounded support are norm dense, and we can repeat the argument in the previous step to obtain the lemma. Indeed, by assumption each f_R admits a sequence $\{f^n_R\}_{n \in \mathbb{N}}$ of functions in $\text{LIP}_b(X)$ with $f^n_R \to f_R$ in $L^p(X)$ and with $g_{f^n_R} \to g_{f_R}$ as $n \to \infty$. Then, a diagonal argument concludes the proof.　　　　　　　　□

---

[1] This proposition states that if $u, v \in N^{1,p}(X)$ and $u = v$ on a set $A \subset X$, then for μ-almost every $x \in A$ we have $g_u(x) = g_v(x)$ for their minimal p-weak upper gradients.

With this lemma, we are left to consider bounded non-negative functions with bounded support. For such functions we construct the approximation by making appropriate choices of g, A, δ in Formula (2.12). In the proof, we first state a goal for the proof that the choices will guarantee, and show that reaching this goal suffices for the claim. After stating that goal, we construct a sequence of approximations f_n that reaches this goal for all large enough $n \in N$. (The actual approximating sequence of f would be obtained by a diagonal sequence involving sending $\epsilon > 0$, which is introduced in the proof, to zero together with $n \to \infty$.) The tricky bit of the proof is showing a pointwise convergence result for the approximating sequence f_n, which is done using the lemmas above and a contradiction argument.

Proof of Theorem 1.1 By Lemma 2.21, we may assume that $f \in N^{1,p}(X)$ satisfies $f : X \to [0, M]$ for some $M > 0$ and $f|_{X \setminus B(x_0, R)} = 0$ for some $x_0 \in X, R > 2$. Further, let $g_f \in L^p(X)$ be the minimal p-weak upper gradient of f.

Goals of proof: Let $\epsilon \in (0, 1)$ be fixed. We will show that we can find functions $g_\epsilon \in L^p(X)$ and $f_\epsilon \in \mathrm{LIP}_b(X)$ so that $g_\epsilon \geq \mathrm{lip}_a[f_\epsilon]$,

$$\int_X |g_\epsilon - g_f|^p d\mu \leq \epsilon, \tag{2.22}$$

and

$$\int_X |f - f_\epsilon|^p d\mu \leq \epsilon. \tag{2.23}$$

The claim of the theorem then follows directly from Lemma 2.9 by using the inequality $\mathrm{lip}_a[f_\epsilon] \geq g_{f_\epsilon}$ and by choosing a sequence of ϵ with $\epsilon \searrow 0$.

Choice of g_ϵ and g_n: First, by Remark 2.2, we can choose a lower semicontinuous upper gradient g_1 of f so that $g_1 \geq g_f$ and

$$\int_X |g_1 - g_f|^p d\mu \leq \epsilon 4^{-2p}. \tag{2.24}$$

Further, since 0 is an upper gradient for f when restricted to the set $X \setminus B(x_0, R)$, we can set $g_1(x) = g_f(x) = 0$ for all $x \in X \setminus B(x_0, R)$. Indeed, if g_1 is any upper gradient, which does not satisfy this property, we may modify it to satisfy this equation. This modification preserves the lower semi-continuity property and the property of being an upper gradient. The latter of these is seen directly from the definition of being an upper gradient as follows. Decompose the curve appearing in Definition (2.1) to the open set $\gamma^{-1}(B(x_0, R))$, which is an at most countable union of open intervals, and to the compact set $\gamma^{-1}(X \setminus B(x_0, R))$, where $f = 0$. By applying (2.1) to each of the open intervals comprising $\gamma^{-1}(B(x_0, R))$, and by summing over them, we obtain (2.1) for the curve γ.

By Lusin's theorem and inner regularity of μ, we can choose an increasing sequence of compact sets $K_n \subset B(x_0, 2R)$ so that $\mu(B(x_0, 2R) \setminus K_n) \leq \epsilon n^{-p} 4^{-n-2p}$ and so that $f|_{K_n}$ is continuous. Since K_n is an increasing sequence of sets, we get

$$\int_X \left(\sum_{n=1}^{\infty} 1_{B(x_0, 2R) \setminus K_n} \right)^p d\mu \leq \sum_{m=2}^{\infty} \int_{K_m \setminus K_{m-1}} \left(\sum_{n=1}^{\infty} 1_{B(x_0, 2R) \setminus K_n} \right)^p d\mu$$

$$\leq \sum_{m=2}^{\infty} \mu(K_m \setminus K_{m-1})(m-1)^p \leq \epsilon 4^{-2p-2}. \tag{2.25}$$

Choose a $\sigma \in (0, 1)$ so that $\mu(B(x_0, 2R))\sigma^p \leq \epsilon 4^{-2p}$, and define again $\psi_{2R}(x) = \max\{0, \min\{1, 2R - d(x_0, x)\}\}$. Since $R > 2$, we have $\psi_{2R}|_{B(x_0, 3R/2)} = 1$ and $\psi_{2R}|_{X \setminus B(x_0, 2R)} = 0$.

Define

$$g_\epsilon(x) := g_1(x) + \sigma\psi_{2R}(x) + \sum_{n=1}^{\infty} 1_{B(x_0,2R)\setminus K_n}(x). \tag{2.26}$$

Inequality (2.22) follows from (2.25), (2.26) and (2.24). Also, g_ϵ is lower-semicontinuous.

By Baire's theorem for lower semicontinuous functions, cf. [21, Proposition 4.2.2], we can find an increasing sequence of bounded continuous functions $\{\tilde{g}_n\}_{n\in\mathbb{N}}$ converging to g_1 with $0 \le \tilde{g}_m \le \tilde{g}_n \le g_1$ for $m \le n$. Define

$$g_n(x) = \tilde{g}_n(x) + \sigma\psi_{2R}(x) + \sum_{k=1}^{n} \min\{nd(x, K_k), 1_{B(x_0,2R)}(x)\}. \tag{2.27}$$

From this definition, we get that $g_n \le g_m$ for $n \le m$, that $0 \le g_n \le g_\epsilon$ and that g_n converges pointwise to g_ϵ as $n \to \infty$. Finally, choose $L \in \mathbb{N}$ so that $\mu(B(x_0, 2R) \setminus K_L) \le \epsilon(2M)^{-p}$, and define the closed set $A := K_L \cup X \setminus B(x_0, R)$. Since $f|_{K_L}$ and $f|_{X\setminus B(x_0,R)}$ are continuous, the function $f|_A$ is continuous.

We remark, that in definitions (2.26) and (2.27), we could avoid adding the summation term to g_ϵ and g_n if the space X was proper. This is one place where properness would yield a simplification. In this case, later in the proof we would use the simpler Lemma 2.16 (with a bounded subset $\overline{B(x_0, 2R)}$ replacing X) instead of Lemma 2.18.

Approximating sequence f_n: Define for each $n \in \mathbb{N}$ the approximating function f_n with data (f, g_n, A, M, n^{-1}) by the formula

$$f_n(x) := \min\left\{M, \inf_{p_0,\dots,p_N} f(p_0) + \sum_{k=0}^{N-1} g_n(p_k)d(p_k, p_{k+1})\right\}, \tag{2.28}$$

where the infimum is taken over all (n^{-1}, A, x) admissible discrete paths $(p_0, \dots, p_N)$.

Since g_n is bounded and continuous for each $n \in \mathbb{N}$, Properties A, B, E and F of Lemma 2.13 yields that that $f_n : X \to [0, M]$ is a Lipschitz functions with

$$\text{lip}_a[f_n] \le g_n \le g_\epsilon \quad \text{and} \quad 0 \le f_n|_A \le f|_A.$$

Further, Property C of Lemma 2.13 gives that $f_n(x) = 0$ for each $x \notin X \setminus B(x_0, R)$. Thus $f_n \in \text{LIP}_b(X)$.

The main step remaining is to show that for each $x \in K_L$ we have $\lim_{n\to\infty} f_n(x) = f(x)$, that is the task is to prove pointwise convergence on K_L. Indeed, suppose that we have shown this. Note that $f_n(x) = f(x) = 0$ on $A \setminus K_L$, since $A \setminus K_L \subset X \setminus B(x_0, R)$. By Lebesgue dominated convergence, since $f_n, f : X \to [0, M]$, we get for n large enough that

$$\int_A |f_n - f|^p \, d\mu = \int_{K_L} |f_n - f|^p \, d\mu \le \epsilon/2.$$

We have $f_n(x) = f(x) = 0$ for $x \in X\setminus B(x_0, R)$ and $f_n(x), f(x) \in [0, M]$ for $x \in B(x_0, R)\setminus K_L$. Thus, from the choice of L and the fact $X = A \cup (B(x_0, R) \setminus K_L)$, we get for $n \in \mathbb{N}$ large enough

$$\int_X |f - f_n|^p \, d\mu = \int_A |f - f_n|^p \, d\mu + \int_{B(x_0,R)\setminus K_L} |f - f_n|^p \, d\mu$$
$$\le \epsilon/2 + M^p\mu(B(x_0, R) \setminus K_L) \le \epsilon.$$

Thus, the function $f_\epsilon = f_n$ for n large enough and g_ϵ realize all the aspects of the goal of the proof. We are left to show that $f_n|_{K_L}$ converges to $f|_{K_L}$ pointwise.[2]

Pointwise convergence $\lim_{n\to\infty} f_n(x) = f(x)$ **for** $x \in K_L$: For the sake of a contradiction, assume that there exists some $x \in K_L \subset B(x_0, R)$ for which pointwise convergence fails.

For $n \geq m$ we have $g_n \geq g_m$. Thus $f_n \geq f_m$, since for each $x \in X$ any (n^{-1}, A, x)-admissible path is also (m^{-1}, A, x)-admissible. Thus, the sequence $\{f_n(x)\}_{n\in\mathbb{N}}$ is increasing in $n \in \mathbb{N}$. By Lemma 2.13 we have $f_n(x) \leq f(x)$ for each $n \in \mathbb{N}$. Thus, the limit $\lim_{n\to\infty} f_n(x)$ exists. Since the limit is not equal to $f(x)$, there exists a constant $\delta > 0$, so that $\lim_{n\to\infty} f_n(x) \leq f(x) - \delta$.

Because $f(x) \leq M$, we get $f_n(x) \leq M - \delta$ for each $n \in \mathbb{N}$. From the definition of $f_n(x)$ in Formula (2.28), we obtain discrete paths $P^n = (p_0^n, \ldots, p_{N_n}^n)$ which are (n^{-1}, A, x)-admissible, and with

$$f(p_0^n) + \sum_{k=0}^{N_n-1} g_n(p_k^n)d(p_k^n, p_{k+1}^n) < f(x) - \delta/2 \leq M. \tag{2.29}$$

Our contradiction will be obtained by finding a curve in the limit of these discrete paths. Recall, that (n^{-1}, A, x)-admissibility means that $p_0^n \in A$, $\mathrm{Mesh}(P^n) \leq n^{-1}$ and $p_{N_n}^n = x$.

First, we restrict to a sub-path. Let $Q^n = (q_0^n, \ldots, q_{M_n}^n)$ be the largest sub-path of P^n with $q_{M_n}^n = x$ and with $Q^n \subset B(x_0, 3R/2)$. We show that Q^n is (n^{-1}, A, x) admissible and satisfies Estimate (2.30). Since $Q^n \subset P^n$, it is clear that $\mathrm{Mesh}(Q^n) \leq n^{-1}$. Further, by construction $q_{M_n}^n = x$. Thus, we need to check that $q_0^n \in A$.

If $Q^n = P^n$, then clearly Q^n is (n^{-1}, A, x)-admissible and $f(q_0^n) = f(p_0^n)$ and (2.30) is immediate. Next, assume that $Q^n \subsetneq P^n$. Recall that at the beginning of the proof we assumed $R > 2$. Since $\mathrm{Mesh}(P^n) \leq 1 < R/2$ and since Q^n is the largest sub-path with the given inclusion, we must have $q_0^n \in B(x_0, 3R/2) \setminus B(x_0, R)$. Thus, $q_0^n \in A$, and Q^n is still (n^{-1}, A, x)-admissible. Also, $0 = f(q_0^n) \leq f(p_0^n)$. These give that

$$f(q_0^n) + \sum_{k=0}^{M_n-1} g_n(q_k^n)d(q_k^n, q_{k+1}^n) \leq f(p_0^n) + \sum_{k=0}^{N_n-1} g_n(p_k^n)d(p_k^n, p_{k+1}^n)$$
$$< f(x) - \delta/2 \leq M. \tag{2.30}$$

We will show that $\{Q^n\}_{n\in\mathbb{N}}$ satisfies the assumptions of Lemma 2.18. This consists of verifying the mesh, length, diameter and h-sum conditions for a good sequence of functions. Recall the definition of a good sequence of functions for $\{K_n\}_{n\in\mathbb{N}}$:

$$h_n(x) := \sum_{k=1}^{n} \min\{nd(x, K_k), 1\}. \tag{2.31}$$

Here is another point, where properness would simplify the proof. If X was proper, we would only need to verify the mesh- and length- conditions and use Lemma 2.16. No h-sum condition or diameter condition or good sequence of functions would be involved.

(1) *Mesh condition* $\lim_{n\to\infty} \mathrm{Mesh}(Q^n) = 0$: We have

$$\lim_{n\to\infty} \mathrm{Mesh}(Q^n) \leq \lim_{n\to\infty} n^{-1} = 0,$$

---

[2] After showing this, Dini's Theorem with $f_n \leq f_m$ for $n \leq m$, implies that $f_n|_{K_L}$ converges uniformly to $f|_{K_L}$. We do not need this claim here.

since Q^n is (n^{-1}, A, x)-admissible.

(2) *Length bound* $\sup_{n \in \mathbb{N}} \text{Len}(Q^n) < \infty$: Fix $n \in \mathbb{N}$. Recall (2.27), $Q^n \subset B(x_0, 3R/2)$ and $R \geq 2$. The definition of ψ_{2R} guarantees that $\psi_{2R}|_{B(x_0, 3R/2)} = 1$. Thus, we get $g_n(q) \geq \sigma$ for every $q \in Q^n$. By (2.30) and by setting $L := \sigma^{-1} M$ we get

$$\text{Len}(Q^n) = \sigma^{-1} \sum_{k=0}^{M_n - 1} \sigma d(q_k^n, q_{k+1}^n)$$

$$\leq \sigma^{-1} \left(f(q_0^n) + \sum_{k=0}^{M_n - 1} g_n(q_k^n) d(q_k^n, q_{k+1}^n) \right) \leq L, \qquad (2.32)$$

for all $n \in \mathbb{N}$.

(3) *Diameter bound* $\inf_{n \in \mathbb{N}} \text{diam}(Q^n) > 0$: Fix $n \in \mathbb{N}$. We stated earlier in the proof that $f|_A$ is continuous. Thus, we can find a constant $\Delta > 0$ so that $y \in A$ and $d(x, y) \leq \Delta$ imply that $|f(x) - f(y)| \leq \delta/4$. We get $f(q_0^n) < f(x) - \delta/2$ from Inequality (2.30). Since $q_0^n \in A$, we get $d(q_0^n, x) = d(q_0^n, q_{M_n}^n) \geq \Delta$. In particular, $\text{diam}(Q^n) \geq \Delta$.

(4) *h-sum bound:* Let h_n be the good sequence of functions for $\{K_n\}_{n \in \mathbb{N}}$. Note that, by definition (2.31), $h_n|_{B(x_0, 2R)} \leq g_n|_{B(x_0, 2R)}$. Inequality (2.30) together with $Q_n \subset B(x_0, 2R)$ thus gives

$$\sum_{k=0}^{M_n - 1} h_n(q_k^n) d(q_k^n, q_{k+1}^n) \leq \sum_{k=0}^{M_n - 1} g_n(q_k^n) d(q_k^n, q_{k+1}^n) \leq M.$$

As a consequence, Lemma 2.18 shows that a subsequence of Q^n converges to a curve $\gamma : [0, 1] \to X$. We now pass to this subsequence.

Since $q_0^n \in A$, A is closed and $\lim_{n \to \infty} q_0^n = \gamma(0)$, we get that $\gamma(0) \in A$. Similarly, $\gamma(1) = x$. Recall that $f|_A$ is continuous. Therefore, $f(\gamma(0)) = \lim_{n \to \infty} f(q_0^n)$. Recall that the increasing sequence of functions g_n converges pointwise to g_ϵ. From Lemma 2.19 applied to Q^n, γ, g_n and g_ϵ, we obtain

$$f(\gamma(0)) + \int_\gamma g_\epsilon \, ds \overset{2.19}{\leq} \liminf_{n \to \infty} f(q_0^n) + \sum_{k=0}^{M_n - 1} g_n(q_k^n) d(q_k^n, q_{k+1}^n) \overset{(2.30)}{\leq} f(\gamma(1)) - \delta/2.$$

We obtain

$$f(\gamma(1)) - f(\gamma(0)) > \int_\gamma g_\epsilon \, ds,$$

which contradicts g_ϵ being a true upper gradient. □

Remark 2.33 The proof above shows a more technical statement, which we highlight for purposes of future work. Suppose that f is non-negative, boundedly supported in $B(x_0, R)$ and bounded by M and has a lower semi-continuous upper gradient g_1. Then, let $\sigma > 0$ be arbitrary and take any increasing sequence of compact sets K_n as in the statement and define the function g_ϵ as in (2.26). By the proof, for every $L \in \mathbb{N}$, there is a sequence of Lipschitz functions f_m with $\text{lip}_a[f_m] \leq g_\epsilon$ and $f_m \to f$ pointwise on each compact set K_L. Further, if X is proper we do not need to modify g_ϵ with the summation term. Also, if g_1 is bounded below on $B(x_0, 2R)$, we do not even need to add $\sigma \psi_{2R}$, which is only added to ensure the length bound for Q^n.

Additionally, the proof shows the following: if f is a continuous function with bounded support, then we do not need to consider the exhaustion by compact sets K_n and get convergence on all of $A = X$. (This may be useful, for example, if f is continuous but not

Lipschitz. On the other hand, even if f is Lipschitz, one could have $\mathrm{lip}_a[f] > g_f$, and wish to construct approximations.)

A natural follow up work would consider these techniques in other Banach function spaces and associated Newton-Sobolev spaces, such as authors have done in [8, 17, 18, 36]. See also the versions in general Banach function spaces in [33]. In these other function settings, one would need to first ensure a lower semi-continuous upper gradient g_1 which is close in norm to the minimal one (by a version of Vitali-Carathéodory as in Remark 2.2 and [32]). Then, check an appropriate version of Lemma 2.9. Finally, one would need to argue that the choices of K_n and σ can be made so that g_ϵ and g_1 are close in norm – which relies on some absolute continuity and monotone convergence in the applicable Banach function space. If one wishes, in proper metric spaces this should be slightly easier. For this argument, some form of Vitali-Carathéodory theorem holding for the Banach function space seems necessary, see [33]. Further ideas or techniques, such as some form of differential structure, would be needed to upgrade the density in energy to density in norm. For such ideas, see [11].

Data availibity No datasets were generated or analysed in the present study.

References

1. Ambrosio, L., Colombo, M., Di Marino, S.: Sobolev spaces in metric measure spaces: reflexivity and lower semicontinuity of slope. In: Variational methods for evolving objects. Adv. Stud. Pure Math. Math. Soc. Jpn. **67**, 1–58 (2015)
2. Ambrosio, L., Gigli, N., Savaré, G.: Density of Lipschitz functions and equivalence of weak gradients in metric measure spaces. Rev. Mat. Iberoam. **29**(3), 969–996 (2013)
3. Björn, A., Björn, J.: Nonlinear potential theory on metric spaces, vol. 17 of EMS Tracts in Mathematics. European Mathematical Society (EMS), Zürich (2011)
4. Björn, A., Björn, J., Shanmugalingam, N.: Quasicontinuity of Newton–Sobolev functions and density of Lipschitz functions on metric spaces. Houston J. Math. **34**(4), 1197–1211 (2008)
5. Calderón, A.-P., Zygmund, A.: Local properties of solutions of elliptic partial differential equations. Studia Math. **20**, 171–225 (1961)
6. Cheeger, J.: Differentiability of Lipschitz functions on metric measure spaces. Geom. Funct. Anal. **9**(3), 428–517 (1999)
7. Coifman, R.R., Weiss, G.: Analyse Harmonique Non-commutative sur Certains Espaces Homogènes. Lecture Notes in Mathematics, vol. 242. Springer-Verlag, Berlin-New York (1971)
8. Costea, S., Miranda, M., Jr.: Newtonian Lorentz metric spaces. Illinois J. Math. **56**(2), 579–616 (2012)
9. Dunford, N., Schwartz, J.T.: Linear operators. Part I. General Theory. John Wiley & Sons Inc, New York (1988). With the assistance of Bade, W.G. and Bartle, R.G.: Reprint of the 1958 original

10. Eriksson-Bique, S., Poggi-Corradini, P.: On the sharp lower bound for duality of modulus. Proc. Amer. Math. Soc. **150**(7), 2955–2968 (2022)
11. Eriksson-Bique, S., and Soultanis, E. Curvewise characterizations of minimal upper gradients and the construction of a Sobolev differential. arXiv:2102.08097
12. Fonseca, I., Leoni, G.: Modern Methods in the Calculus of Variations: L^p Spaces. Springer Monographs in Mathematics, Springer, New York (2007)
13. Grove, K., Karcher, H.: How to conjugate C^1-close group actions. Math. Z. **132**, 11–20 (1973)
14. Hajłasz, P.: Sobolev spaces on an arbitrary metric space. Potential Anal. **5**(4), 403–415 (1996)
15. Hajłasz, P.: Sobolev spaces on metric-measure spaces. In: Heat Kernels and Analysis on Manifolds, Graphs, and Metric Spaces (Paris, 2002). Contemp. Math., vol. 338, pp. 173–218. Amer. Math. Soc., Providence, RI (2003)
16. Hajłasz, P., Koskela, P.: Sobolev met Poincaré. Mem. Amer. Math. Soc. **145**, 688 (2000)
17. Harjulehto, P., Hästö, P., Pere, M.: Variable exponent Sobolev spaces on metric measure spaces. Funct. Approx. Comment. Math. **36**, 79–94 (2006)
18. Hästö, P.A.: On the density of continuous functions in variable exponent Sobolev space. Rev. Mat. Iberoam. **23**(1), 213–234 (2007)
19. Heinonen, J.: Nonsmooth Calculus. Bull. Amer. Math. Soc. **44**(2), 163–232 (2007)
20. Heinonen, J., Koskela, P.: Quasiconformal maps in metric spaces with controlled geometry. Acta Math. **181**(1), 1–61 (1998)
21. Heinonen, J., Koskela, P., Shanmugalingam, N., Tyson, J.T.: Sobolev spaces on metric measure. In: New, Mathematical Monographs Spaces, vol. 27. Cambridge University Press, Cambridge (2015)
22. Järvenpää, E., Järvenpää, M., Rogovin, K., Rogovin, S., Shanmugalingam, N.: Measurability of equivalence classes and MEC_p-property in metric spaces. Rev. Mat. Iberoam. **23**(3), 811–830 (2007)
23. Karcher, H.: Riemannian center of mass and mollifier smoothing. Commun. Pure Appl. Math. **30**(5), 509–541 (1977)
24. Keith, S.: Modulus and the Poincaré inequality on metric measure spaces. Math. Z. **245**, 255–292 (2003)
25. Keith, S., Zhong, X.: The Poincaré inequality is an open ended condition. Annals Math. **167**(2), 575–599 (2008)
26. Kinnunen, J., Korte, R., Shanmugalingam, N., Tuominen, H.: Pointwise properties of functions of bounded variation in metric spaces. Rev. Mat. Comput. **27**(1), 41–67 (2014)
27. Kinnunen, J., Latvala, V.: Lebesgue points for Sobolev functions on metric spaces. Rev. Mat. Iberoam. **18**(3), 685–700 (2002)
28. Koskela, P.: Removable sets for Sobolev spaces. Ark. Mat. **37**(2), 291–304 (1999)
29. Lahti, P., Shanmugalingam, N.: Fine properties and a notion of quasicontinuity for BV functions on metric spaces. J. Math. Pures Appl. **107**(2), 150–182 (2017)
30. Liu, F.C.: A Luzin type property of Sobolev functions. Indiana Univ. Math. J. **26**(4), 645–651 (1977)
31. Macías, R.A., Segovia, C.: A decomposition into atoms of distributions on spaces of homogeneous type. Adv. Math. **33**(3), 271–309 (1979)
32. Malý, L.: Minimal weak upper gradients in Newtonian spaces based on quasi-Banach function lattices. Ann. Acad. Sci. Fenn. Math. **38**(2), 727–745 (2013)
33. Malý, L.: Fine properties of Newtonian functions and the Sobolev capacity on metric measure spaces. Rev. Mat. Iberoam. **32**(1), 219–255 (2016)
34. Semmes, S.: Finding curves on general spaces through quantitative topology, with applications to Sobolev and Poincaré inequalities. Selecta Math. **2**(2), 155–295 (1996)
35. Shanmugalingam, N.: Newtonian spaces: an extension of Sobolev spaces to metric measure spaces. Rev. Mat. Iberoam. **16**(2), 243–279 (2000)
36. Tuominen, H.: Pointwise behaviour of Orlicz–Sobolev functions. Ann. Mat. Pura Appl. **188**(1), 35–59 (2009)
37. Weierstraß, K.: Über Continuirliche Functionen Eines Reellen Arguments, die für Keinen Werth des Letzteren Einen Bestimmten Differentialquotienten Besitzen, pp. 190–193. Springer Vienna, Vienna (1988)