Mathematische Annalen
https://doi.org/10.1007/s00208-024-02972-z

Mathematische Annalen

Mappings of finite distortion on metric surfaces

Damaris Meier[1] · Kai Rajala[2]

Received: 15 May 2024 / Revised: 2 August 2024 / Accepted: 8 August 2024

Abstract

We investigate basic properties of *mappings of finite distortion* $f : X \to \mathbb{R}^2$, where X is any *metric surface*, i.e., metric space homeomorphic to a planar domain with locally finite 2-dimensional Hausdorff measure. We introduce *lower gradients*, which complement the upper gradients of Heinonen and Koskela, to study the distortion of non-homeomorphic maps on metric spaces. We extend the Iwaniec-Šverák theorem to metric surfaces: a non-constant $f : X \to \mathbb{R}^2$ with locally square integrable upper gradient and locally integrable distortion is continuous, open and discrete. We also extend the Hencl-Koskela theorem by showing that if f is moreover injective then f^{-1} is a Sobolev map.

Mathematics Subject Classification Primary 30L10 · 30C65; Secondary 30F10

1 Introduction

1.1 Background

Let $\Omega \subset \mathbb{R}^2$ be a domain. We say that map $f : \Omega \to \mathbb{R}^2$ in the Sobolev space $W^{1,2}_{\mathrm{loc}}(\Omega, \mathbb{R}^2)$ has *finite distortion* if there is a measurable function $K : \Omega \to [1, \infty)$ so that

$$||Df(x)||^2 \leq K(x) J_f(x) \quad \text{for a.e. } x \in \Omega. \tag{1.1}$$

The first-named author is partially supported by UniFr Doc.Mobility Grant DM-22-10.

✉ Damaris Meier
damaris.meier@unifr.ch

Kai Rajala
kai.i.rajala@jyu.fi

1 Department of Mathematics, University of Fribourg, Chemin du Musée 23, 1700 Fribourg, Switzerland

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

Published online: 02 September 2024

Here $||Df(x)||$ and $J_f(x)$ are the operator norm and determinant of $Df(x)$, respectively.

If $K(x) = 1$ for almost every $x \in \Omega$, then (1.1) is valid if and only if f is complex analytic. The basic topological properties of non-constant analytic functions are *continuity*, *openness* and *discreteness* (the preimage of every point is discrete in Ω).

By *Stoïlow factorization* (see [1, Chapter 5.5], [27]) non-constant *quasiregular maps*, i.e., maps f satisfying (1.1) with constant function $K(x) = K \geq 1$, admit a factorization $f = g \circ h$, where h is a quasiconformal homeomorphism and g is analytic. In particular, every such f is also continuous, open and discrete.

In [22] Iwaniec and Šverák showed that boundedness of $K(x)$ may be replaced with local integrability.

Theorem 1.1 (Iwaniec-Šverák theorem) *Suppose $f \in W^{1,2}_{loc}(\Omega, \mathbb{R}^2)$ is non-constant and satisfies (1.1) for some locally integrable $K(x)$. Then f is continuous, open and discrete.*

The assumption on $K(x)$ is essentially best possible (see [2] and [17]). Since the work of Iwaniec and Šverák [22], a rich theory of mappings of finite distortion has been developed (see [1, 12]), with applications to PDE, complex dynamics, inverse problems and non-linear elasticity theory, among other fields.

The theory extends to $W^{1,1}_{\mathrm{loc}}$-maps with exponentially integrable distortion and also to higher dimensions, where continuity, openness and discreteness of quasiregular maps was proved by Reshetnyak already in the 1960s (see [42]). Reshetnyak's theorem has been extended to spatial mappings of finite distortion by several authors (see [13, 20, 21, 25, 26, 37, 38, 45, 46]).

Partially motivated by works of Heinonen and Keith [16], Heinonen-Rickman [17] and Heinonen–Sullivan [18], on BLD- and bi-Lipschitz parametrizations of metric spaces, Kirsilä [24] furthermore extended Reshetnyak's theorem to maps $f : X \to \mathbb{R}^n$, where X is a *generalized n-manifold* satisfying assumptions such as Ahlfors n-regularity and Poincaré inequality.

In this paper we extend the Iwaniec-Šverák theorem to maps $f : X \to \mathbb{R}^2$, where X is a *metric surface*, i.e., a metric space homeomorphic to a domain in $\mathbb{R}^2$ with locally finite 2-dimensional Hausdorff measure. The novelty of our results is that we do not impose any additional conditions on X.

Our research is partially inspired by recent advances on the uniformization problem on metric surfaces (see [5, 20, 30, 34–36, 39]) and the properties of the associated homeomorphisms, such as quasiconformal maps $f : X \to \mathbb{R}^2$. It is desirable to explore the properties of non-homeomorphic maps on metric surfaces. The aim of our paper is to provide the first results in this direction.

1.2 Mappings of finite distortion on metric surfaces

A (euclidean) *metric surface* X is a metric space homeomorphic to a domain $U \subset \mathbb{R}^2$ and with locally finite 2-dimensional Hausdorff measure. Below, $\mathcal{H}^2$ will always be the reference measure on X.

Let X and Y be metric surfaces. We want to establish what it means for a map $f: X \to Y$ to have finite distortion. We first observe that in the euclidean case every mapping of finite distortion is sense-preserving. This follows from inequality (1.1) by applying non-negativity of the Jacobian determinant and integration by parts, a method which is not available in our generality. We call $f : X \to Y$ *sense-preserving* if for any domain Ω compactly contained in X so that $f|_{\partial\Omega}$ is continuous it follows that $\deg(y, f, \Omega) \geq 1$ for any $y \in f(\Omega)\setminus f(\partial\Omega)$. Here deg is the local topological degree of f, see [40, I.4] for a definition in the euclidean setting and note that the concept transfers to our setting as every metric surface is homeomorphic to a domain in $\mathbb{R}^2$.

We apply the theory of Sobolev spaces based on *upper gradients* ([13]). A Borel function $\rho^u : X \to [0, \infty]$ is an *upper gradient* of $f : X \to Y$, if

$$d_Y(f(x), f(y)) \leq \int_\gamma \rho^u \, ds \tag{1.2}$$

for all $x, y \in X$ and every rectifiable curve γ in X joining x and y. We say that f belongs to the Sobolev space $N^{1,2}_{\text{loc}}(X, Y)$ if f has an upper gradient $\rho^u \in L^2_{\text{loc}}(X)$ and if $d_Y(y, f(\cdot)) \in L^2_{\text{loc}}(X)$ for some $y \in Y$ (see Sect. 2.3).

It follows from the proof of [11, Theorem 1.4] that a sense-preserving map $f \in N^{1,2}_{\text{loc}}(X, \mathbb{R}^2)$ is continuous (see Remark 2.3). Such an f also satisfies Lusin's Condition (N): if $E \subset X$ and $\mathcal{H}^2(E) = 0$, then $|f(E)|_2 = 0$ (see Remark 2.8). The converse implication does not hold ([39, Section 17]).

In order to define the distortion of f, we introduce *lower gradients*: a Borel function $\rho^l : X \to [0, \infty]$ is a *lower gradient* of $f \in N^{1,2}_{\text{loc}}(X, Y)$, if $\rho^l \leq \rho^u_f$ almost everywhere and

$$\ell(f \circ \gamma) \geq \int_\gamma \rho^l \, ds \tag{1.3}$$

for every rectifiable curve γ in X with $f \circ \gamma$ being continuous. Our definition is motivated by the observation that the upper gradient inequality (1.2) is equivalent to the reverse inequality of (1.3) for ρ^u (see Sect. 2.3). Every $f \in N^{1,2}_{\text{loc}}(X, Y)$ has an essentially unique *minimal weak upper gradient* ρ^u_f (see Sect. 2.3). Similarly, we prove in Sect. 7 that every such f has an essentially unique *maximal weak lower gradient* ρ^l_f.

We say that a sense-preserving $f \in N^{1,2}_{\text{loc}}(X, Y)$ has *finite distortion (along paths) and denote* $f \in \text{FDP}(X, Y)$, if there is a measurable $K : X \to [1, \infty)$ such that

$$\rho^u_f(x) \leq K(x) \cdot \rho^l_f(x) \quad \text{for almost every } x \in X. \tag{1.4}$$

The *distortion* K_f of f is

$$K_f(x) := \begin{cases} \frac{\rho^u_f(x)}{\rho^l_f(x)}, & \text{if } \rho^l_f(x) \neq 0, \\ 1, & \text{if } \rho^l_f(x) = 0. \end{cases}$$

Our main result is the following extension of the Iwaniec-Šverák theorem. Here X is any metric surface.

Theorem 1.2 *Let $f \in \mathrm{FDP}(X, \mathbb{R}^2)$ be non-constant with $K_f \in L^1_{loc}(X)$. Then f is open and discrete.*

Generalizing the euclidean result by Hencl-Koskela (who assumed $W^{1,1}$-regularity, see [16]), we show that if f is furthermore a homeomorphism, then the inverse is also a Sobolev map. For a related result see [4].

Theorem 1.3 *Let $f \in \mathrm{FDP}(X, \mathbb{R}^2)$ be injective with $K_f \in L^1_{loc}(X)$. Then $f^{-1} \in N^{1,2}_{loc}(f(X), X)$.*

Examples in [2] (f_0 in Proposition 6.1 below, see also [17]) and [16, Example 1.4], respectively, show that condition $K_f \in L^1_{\mathrm{loc}}(X)$ is sharp both in Theorem 1.2 and in Theorem 1.3, even if $X = \mathbb{R}^2$.

We show in Sect. 6 that there are metric surfaces X which do not admit any quasiconformal maps $h : X \to \mathbb{R}^2$ but do admit maps $f : X \to \mathbb{R}^2$ satisfying the assumptions of Theorem 1.2. By [33, Theorem 1.3], such surfaces do not exist if we require K_f to be bounded instead of integrable.

Previous approaches to distortion of maps between metric spaces are mostly based on the *analytic definition*: We say that a sense-preserving $f \in N^{1,2}_{\mathrm{loc}}(X, Y)$ has *finite analytic distortion* and denote $f \in \mathrm{FDA}(X, Y)$, if there is a measurable $C : X \to [1, \infty)$ such that

$$\rho^u_f(x)^2 \leq C(x) \cdot J_f(x) \quad \text{for almost every } x \in X, \tag{1.5}$$

where

$$J_f(x) = \limsup_{r \to 0} \frac{\mathcal{H}^2_Y(f(\overline{B}(x, r)))}{\pi r^2}.$$

Inequality (1.5) is equivalent to (1.4) in the euclidean setting, and also provides a rich theory for homeomorphisms between metric spaces. However, unlike our approach based on lower gradients, the analytic approach is not convenient for treating non-homeomorphic maps between metric surfaces. We nevertheless prove the following in [33].

Theorem 1.4 ([33, Theorem 1.1]) *If $f \in \mathrm{FDA}(X, \mathbb{R}^2)$, then $f \in \mathrm{FDP}(X, \mathbb{R}^2)$. Moreover, for every $C(x)$ in (1.5) we have*

$$K_f(x) \leq 4\sqrt{2}\, C(x) \quad \textit{for almost every } x \in X.$$

Theorem 1.2 can be applied to prove the converse of Theorem 1.4 assuming $K_f \in L^1_{\mathrm{loc}}(X, \mathbb{R}^2)$, see [33]. Combining Theorems 1.2, 1.3 and 1.4 shows that our main results hold under the analytic assumption.

Corollary 1.5 *Let $f \in \mathrm{FDA}(X, \mathbb{R}^2)$ be non-constant with $C(x) \in L^1_{loc}(X)$. Then f is open and discrete. If f is injective, then $f^{-1} \in N^{1,2}_{loc}(f(X), X)$.*

The definition of a metric surface can be relaxed by requiring X to be homeomorphic to an oriented topological surface M instead of a domain in $\mathbb{R}^2$. Our definitions and results are local and remain valid under the relaxed definition. We state them only for euclidean metric surfaces to simplify the presentation.

This paper is organized as follows. In Sect. 2 we recall the background on Analysis in metric spaces needed to prove our main results. In Sect. 3 we prove an area inequality for maps on the rectifiable part of a metric surface which involves lower gradients and may be of independent interest. We prove Theorems 1.2 and 1.3 in Sects. 4 and 5, respectively.

The proofs are based on three main tools: the coarea inequality for Sobolev functions on metric surfaces by Meier- Ntalampekos [32] and Esmayli- Ikonen- Rajala [11], weakly quasiconformal parametrizations of metric surfaces by Ntalampekos- Romney [35, 36] and Meier- Wenger [34], and the area inequality proved in Sect. 3. In addition, to prove Theorem 1.2 we apply estimates inspired by the value distribution theory of quasiregular mappings (see [40]).

In Sect. 6, we discuss connections between our results and the uniformization problem on metric surfaces, as well as different definitions of mappings with controlled distortion. Finally, in Sect. 7 we prove the existence of maximal weak lower gradients.

2 Preliminaries

2.1 Basic definitions and notations

Let (X, d) be a metric space. We denote the *open* and *closed ball* in X of radius $r > 0$ centered at a point $x \in X$ by $B(x, r)$ and $\overline{B}(x, r)$, respectively. When $X = \mathbb{R}^2$ we use notation $\mathbb{D}(x, r)$ instead of $B(x, r)$.

A set $\Omega \subset X$ homeomorphic to the unit disc $\mathbb{D}(0, 1)$ is a *Jordan domain* in X if its boundary $\partial\Omega \subset X$ is a *Jordan curve* in X, i.e., a subset of X homeomorphic to $\mathbb{S}^1$. The *image* of a curve γ in X is indicated by $|\gamma|$ and the *length* by $\ell(\gamma)$.

A curve γ is *rectifiable* if $\ell(\gamma) < \infty$ and *locally rectifiable* if each of its compact subcurves is rectifiable. Moreover, a curve $\gamma\colon [a, b] \to X$ is *geodesic* if $\ell(\gamma) = d(\gamma(a), \gamma(b))$. A curve $\gamma\colon [0, \ell(\gamma)] \to X$ is *parametrized by arclength* if $\ell(\gamma|_I) = |I|_1$ for every interval $I \subset [0, \ell(\gamma)]$. Here, $|\cdot|_n$ denotes the *n-dimensional Lebesgue measure*.

For $s \geq 0$, we denote the *s-dimensional Hausdorff measure* of $A \subset X$ by $\mathcal{H}^s(A)$. The normalizing constant is chosen so that $|V|_n = \mathcal{H}^n(V)$ for open subsets V of $\mathbb{R}^n$.

We equip X with $\mathcal{H}^2$. Let $L^p(X)$ ($L^p_{\mathrm{loc}}(X)$) denote the space of p-integrable (locally p-integrable) Borel functions from X to $\mathbb{R} \cup \{-\infty, \infty\}$. Here locally p-integrable means p-integrable on compact subsets. We say that a subdomain G of X is *compactly contained* in X if the closure $\overline{G}$ is compact.

2.2 Modulus

Let X be a metric space and Γ be a family of curves in X. A Borel function $g\colon X \to [0,\infty]$ is *admissible* for Γ if $\int_\gamma g\,ds \geq 1$ for all locally rectifiable curves $\gamma \in \Gamma$. We define the (2-)*modulus* of Γ as

$$\operatorname{Mod}\Gamma = \inf_g \int_X g^2\,d\mathcal{H}^2,$$

where the infimum is taken over all admissible functions g for Γ. If there are no admissible functions for Γ we set $\operatorname{Mod}\Gamma = \infty$. A property is said to hold for *almost every* curve in Γ if it holds for every curve in $\Gamma \setminus \Gamma_0$ for some family $\Gamma_0 \subset \Gamma$ with $\operatorname{Mod}(\Gamma_0) = 0$. In the definition of $\operatorname{Mod}(\Gamma)$, the infimum can equivalently be taken over all *weakly admissible* functions, i.e., Borel functions $g\colon X \to [0,\infty]$ such that $\int_\gamma g \geq 1$ for almost every locally rectifiable curve $\gamma \in \Gamma$.

2.3 Metric Sobolev spaces

Let $f\colon X \to Y$ be a map between metric spaces. A Borel function $\rho^u\colon X \to [0,\infty]$ is an *upper gradient* of f if

$$d_Y(f(x), f(y)) \leq \int_\gamma \rho^u\,ds \tag{2.1}$$

for all $x, y \in X$ and every rectifiable curve γ in X joining x and y. If the *upper gradient inequality* (2.1) holds for almost every rectifiable curve γ in X joining x and y we call ρ^u *weak upper gradient* of f.

The Sobolev space $N^{1,2}(X,Y)$ is the space of Borel maps $f\colon X \to Y$ with upper gradient $\rho^u \in L^2(X)$ such that $x \mapsto d_Y(y, f(x))$ is in $L^2(X)$ for some and thus any $y \in Y$. The space $N^{1,2}_{\mathrm{loc}}(X,Y)$ is defined in the obvious manner.

Each $f \in N^{1,2}_{\mathrm{loc}}(X,Y)$ has a *minimal* weak upper gradient ρ^u_f, i.e., for any other weak upper gradient ρ^u we have $\rho^u_f \leq \rho^u$ almost everywhere. Moreover, ρ^u_f is unique up to a set of measure zero. See monograph [13] for more background on metric Sobolev spaces.

We apply a notion of "minimal stretching" which compliments the "maximal stretching" represented by upper gradients. To motivate the definition, notice that for continuous maps $f \in N^{1,2}_{\mathrm{loc}}(X,Y)$ the upper gradient inequality (2.1) is equivalent to

$$\ell(f \circ \gamma) \leq \int_\gamma \rho^u\,ds$$

for almost every rectifiable curve γ in X. We call a Borel function $\rho^l\colon X \to [0,\infty]$ a *lower gradient* of $f \in N^{1,2}_{\mathrm{loc}}(X,Y)$, if $\rho^l \leq \rho^u_f$ almost everywhere and

$$\ell(f \circ \gamma) \geq \int_\gamma \rho^l \, ds \tag{2.2}$$

for every rectifiable curve γ in X with $f \circ \gamma$ being continuous. If the *lower gradient inequality* (2.2) holds for almost every rectifiable γ, we call ρ^l *weak lower gradient* of f. Note that 0 is always a lower gradient.

Each $f \in N^{1,2}_{\mathrm{loc}}(X, Y)$ has a *maximal weak lower gradient* ρ^l_f, i.e., for any other weak lower gradient ρ^l we have $\rho^l_f \geq \rho^l$ almost everywhere. Moreover, ρ^l_f is unique up to a set of measure zero. The proof is analogous to the existence of minimal weak upper gradients, see [13, Theorem 6.3.20]. For completeness, we provide a proof in Sect. 7.

2.4 Coarea inequality on metric surfaces

We state the following coarea inequality for Lipschitz functions, which is a consequence of [10, Theorem 1.1] (see [11, Section 5]). Here, $\mathrm{Lip}(u)$ denotes the pointwise Lipschitz constant of a Lipschitz function $u\colon X \to \mathbb{R}$, defined by

$$\mathrm{Lip}(u)(x) = \limsup_{x \neq y \to x} \frac{|u(y) - u(x)|}{d(x, y)}.$$

Theorem 2.1 *(*Lipschitz coarea inequality*) Let X be a metric space and $u\colon X \to \mathbb{R}$ a Lipschitz function. Then*

$$\int^*_{\mathbb{R}} \int_{u^{-1}(t)} g \, d\mathcal{H}^1 dt \leq \frac{4}{\pi} \int_X g \cdot \mathrm{Lip}(u) \, d\mathcal{H}^2$$

for every Borel measurable $g : X \to [0, \infty]$.

Here $\int^*$ denotes the upper integral, which is equal to Lebesgue integral for measurable functions. An important tool throughout this work will be the following coarea inequality for continuous Sobolev functions on metric surfaces.

Theorem 2.2 (Sobolev coarea inequality, [32, Theorem 1.6]) *Let X be a metric surface and $v\colon X \to \mathbb{R}$ be a continuous function in $N^{1,2}_{loc}(X)$.*

(1) If $\mathcal{A}_v$ denotes the union of all non-degenerate components of the level sets $v^{-1}(t)$, $t \in \mathbb{R}$, of v, then $\mathcal{A}_v$ is a Borel set.
(2) For every Borel function $g\colon X \to [0, \infty]$ we have

$$\int^* \int_{v^{-1}(t) \cap \mathcal{A}_v} g \, d\mathcal{H}^1 \, dt \leq \frac{4}{\pi} \int g \cdot \rho^u_v \, d\mathcal{H}^2.$$

Theorem 2.2 generalizes the coarea inequality for monotone Sobolev functions established in [11]. Here $v\colon X \to \mathbb{R}$ is called a *weakly monotone function* if for every open Ω compactly contained in X

$$\sup_{\Omega} v \le \sup_{\partial\Omega} v < \infty \quad \text{and} \quad \inf_{\Omega} v \ge \inf_{\partial\Omega} v > -\infty.$$

A continuous weakly monotone function is *monotone*.

Remark 2.3 In the proof of [11, Theorem 1.4] the coarea inequality for monotone Sobolev functions is used to show that every weakly monotone function $v \in N^{1,2}_{\mathrm{loc}}(X, \mathbb{R})$ is continuous and hence monotone. Continuity of a sense-preserving map $f \in N^{1,2}_{\mathrm{loc}}(X, \mathbb{R}^2)$ now follows by applying the exact same proof strategy while replacing weak monotonicity with sense-preservation and the coarea inequality for monotone Sobolev maps with Theorem 2.2.

2.5 Metric differentiability

Let (Y, d) be a complete metric space and $U \subset \mathbb{R}^n$, $n \ge 1$, a domain. We say that $h\colon U \to Y$ is *approximately metrically differentiable* at $z \in U$ if there exists a seminorm N_z on $\mathbb{R}^2$ for which

$$\operatorname{ap}\lim_{y\to z} \frac{d(h(y), h(z)) - N_z(y-z)}{|y-z|} = 0.$$

Here, ap lim denotes the approximate limit (see [8, Section 1.7.2]). If such a seminorm exists, it is unique and is called *approximate metric derivative* of h at z, denoted $\operatorname{ap\,md} h_z$. The following result follows from [29, Lemma 3.1].

Lemma 2.4 *Let X and Y be metric surfaces and $f \in N^{1,2}_{loc}(X, Y)$. Almost every curve $\gamma\colon [a, b] \to X$ parametrized by arclength satisfies*

$$\int_{f\circ\gamma} g\, ds = \int_a^b g(f(\gamma(t))) \cdot \operatorname{ap\,md}(f\circ\gamma)_t\, dt$$

for all Borel measurable $g : Y \to [0, \infty]$.

Lemma 2.4 leads to the following properties of upper and lower gradients (see [13, Proposition 6.3.3] for a proof involving upper gradients).

Corollary 2.5 *Let X and Y be metric surfaces and $f \in N^{1,2}_{loc}(X, Y)$. Almost every curve $\gamma\colon [a, b] \to X$ parametrized by arclength satisfies the following properties.*

(1) f is absolutely continuous on γ,
(2) $\rho^l_f(\gamma(t)) \le \operatorname{ap\,md}(f\circ\gamma)_t \le \rho^u_f(\gamma(t))$ for almost every $a < t < b$,
(3) if $g : Y \to [0, \infty]$ is Borel measurable, then

$$\int_\gamma \rho^l_f \cdot (g\circ f)\, ds \le \int_{f\circ\gamma} g\, ds \le \int_\gamma \rho^u_f \cdot (g\circ f)\, ds.$$

2.6 Area formula on euclidean domains

Suppose $U \subset \mathbb{R}^2$ is a domain and $h \in N^{1,2}_{\mathrm{loc}}(U, Y)$. Then U can be covered up to a set of measure zero by countably many disjoint measurable sets G_j, $j \in \mathbb{N}$, such that $h|_{G_j}$ is Lipschitz. In particular, outside a set of measure zero $G_0 \subset U$, h satisfies Lusin's condition (N) (see [13, Theorem 8.1.49]).

By [28, Proposition 4.3], every $h \in N^{1,2}_{\mathrm{loc}}(U, Y)$ is approximately metrically differentiable at a.e. $z \in U$. The following area formula follows from [23, Theorem 3.2]. Here, the *Jacobian* $J(N_z)$ of a seminorm N_z on $\mathbb{R}^2$ is zero if N_z is not a norm and $J(N_z) = \pi/|\{y \in \mathbb{R}^2 : N_z(y) \leq 1\}|_2$ otherwise.

Theorem 2.6 (Area formula) *If $h \in N^{1,2}_{loc}(U, Y)$, then there exists $G_0 \subset U$ with $\mathcal{H}^2(G_0) = 0$ such that for every measurable set $A \subset U \setminus G_0$ we have*

$$\int_A J(\operatorname{ap} \operatorname{md} h_z)\, d\mathcal{H}^2 = \int_Y N(y, h, A)\, d\mathcal{H}^2. \tag{2.3}$$

Here, $N(y, h, A)$ denotes the *multiplicity* of $y \in Y$ with respect to h in A:

$$N(y, h, A) := \#\{z \in A : h(z) = y\}. \tag{2.4}$$

2.7 Weakly quasiconformal parametrizations

A map $h\colon X \to Y$ between metric surfaces is *cell-like* if the preimage of each point is a continuum that is contractible in each of its open neighborhoods. A continuous, surjective, proper and cell-like map $h\colon X \to Y$ is *weakly C-quasiconformal* if

$$\operatorname{Mod} \Gamma \leq C \operatorname{Mod} h(\Gamma)$$

holds for every family of curves Γ in X. It follows from [47, Theorem 1.1] that every weakly quasiconformal map $h\colon X \to Y$ is contained in $N^{1,2}_{\mathrm{loc}}(X, Y)$.

It was shown in [35] that any metric surface admits a weakly quasiconformal parametrization, see also [30, 34, 36].

Theorem 2.7 ([35, Theorem 1.2]) *Let X be any metric surface. There is a weakly $(4/\pi)$-quasiconformal $u\colon U \to X$, where $U \subset \mathbb{R}^2$ is a domain.*

Remark 2.8 Condition (N) for sense-preserving maps $f \in N^{1,2}_{\mathrm{loc}}(X, \mathbb{R}^2)$ can be proved using the area formula and Theorem 2.7 as follows: suppose $E \subset X$ and $\mathcal{H}^2(E) = 0$, and let $u\colon U \to X$ be a (sense-preserving) weakly $(4/\pi)$-quasiconformal parametrization of X provided by Theorem 2.7. Define $h\colon U \to \mathbb{R}^2$ by $h := f \circ u$. Then $u \in N^{1,2}_{\mathrm{loc}}(U, X)$ and $h \in N^{1,2}_{\mathrm{loc}}(U, \mathbb{R}^2)$, see [33, Theorem 2.5].

By Theorem 2.6 there exists $G_0 \subset U$ with $|G_0|_2 = 0$ and such that (2.3) holds for u and h and every measurable set $A \subset U \setminus G_0$. We set $X_0 := u(G_0)$. Now h is sense-preserving and thus monotone. Therefore, h satisfies Condition (N) by [31]. In particular, with the above notation,

$$|f(E)|_2 \le \int_{u^{-1}(E)} J(\operatorname{ap}\operatorname{md} h_z)\, dz.$$

On the other hand, applying Theorem 2.6 to u shows that

$$\int_{u^{-1}(E)} J(\operatorname{ap}\operatorname{md} u_z)\, dz \le \mathcal{H}^2(E) = 0,$$

and so $J(\operatorname{ap}\operatorname{md} u_z) = 0$ almost everywhere in $u^{-1}(E)$. Since u is weakly quasiconformal, it moreover follows that $\operatorname{ap}\operatorname{md} u_z = 0$. Then, by Lemmas 2.9 and 2.10 below, $J(\operatorname{ap}\operatorname{md} h_z) = 0$ almost everywhere in $u^{-1}(E)$ as well. We conclude that $|f(E)|_2 = 0$.

2.8 Distortion of Sobolev maps

Let $U \subset \mathbb{R}^2$ be a domain. We define the *maximal and minimal stretches* of $h \in N^{1,2}_{\mathrm{loc}}(U, Y)$ at points of approximate differentiability by

$$L_h(z) = \max\{\operatorname{ap}\operatorname{md} h_z(v) : |v| = 1\}, \quad l_h(z) = \min\{\operatorname{ap}\operatorname{md} h_z(v) : |v| = 1\}.$$

Recall that maps $h \in N^{1,2}_{\mathrm{loc}}(U, Y)$ are approximately differentiable almost everywhere.

Lemma 2.9 *Let $h \in N^{1,2}_{loc}(U, Y)$. Then L_h and l_h are representatives of the minimal weak upper gradient and the maximal weak lower gradient of h, respectively. Moreover,*

$$2^{-1} L_h(z) l_h(z) \le J(\operatorname{ap}\operatorname{md} h_z) \le 2 L_h(z) l_h(z) \tag{2.5}$$

at points of approximate differentiability.

Proof The first claim concerning upper gradients is [32, Lemma 2.14]. A slight modification of the proof gives the claim concerning lower gradients.

Towards (2.5), we may assume that $\operatorname{ap}\operatorname{md} h_z$ is a norm. Then the unit ball B_z of $\operatorname{ap}\operatorname{md} h_z(v)$ contains a unique ellipse of maximal area E_z, called the *John ellipse* of B_z, which satisfies

$$E_z \subset B_z \subset \sqrt{2} E_z, \tag{2.6}$$

see [3, Theorem 3.1]. Let N_z be the norm whose unit ball is E_z, and

$$M_z = \max\{N_z(v) : |v| = 1\}, \quad m_z = \min\{N_z(v) : |v| = 1\}.$$

Then $J(N_z) = \pi/|E_z|_2 = M_z m_z$, and (2.6) gives

$$L_h(z) l_h(z) \le M_z m_z = J(N_z) = 2\pi/|\sqrt{2} E_z|_2 \le 2\pi/|B_z|_2 = 2J(\operatorname{ap}\operatorname{md} h_z).$$

On the other hand, (2.6) also gives

$$J(\operatorname{ap}\operatorname{md} h_z) \leq J(N_z) = M_z m_z \leq 2L_h(z) l_h(z).$$

The proof is complete. □

We will apply distortion estimates on composed mappings.

Lemma 2.10 *Let X and Y be metric surfaces and $U \subset \mathbb{R}^2$ a domain, $u : U \to X$ weakly quasiconformal, and $f \in N^{1,2}_{loc}(X, Y)$. Then*

$$l_{f\circ u}(z) \geq \rho^l_f(u(z)) \cdot l_u(z) \quad \textit{and} \quad L_{f\circ u}(z) \leq \rho^u_f(u(z)) \cdot L_u(z)$$

for almost every $z \in U$.

Proof Let Γ_0 be the family of paths γ in U so that l_u does not satisfy the lower gradient inequality (2.2) for u on some subcurve of γ or ρ^l_f does not satisfy the lower gradient inequality for f on some subcurve of $u \circ \gamma$. Then, since u is weakly quasiconformal and l_u, ρ^l_f are weak lower gradients (Lemma 2.9), we conclude that $\operatorname{Mod}(\Gamma_0) = 0$. Applying Corollary 2.5, we have

$$\ell(f \circ u \circ \gamma) \geq \int_{u\circ\gamma} \rho^l_f \, ds \geq \int_\gamma (\rho^l_f \circ u) \cdot l_u \, ds$$

for every $\gamma \notin \Gamma_0$ parametrized by arclength. We conclude that $(\rho^l_f \circ u) \cdot l_u$ is a weak lower gradient of $f \circ u$. But $l_{f\circ u}$ is a maximal weak lower gradient of $f \circ u$ by Lemma 2.9. The first inequality follows. The second inequality is proved in a similar way. □

3 Area inequality on Metric surfaces

Let X and Y be metric surfaces. In this section we establish Theorem 3.1, an area inequality for Sobolev maps in $N^{1,2}_{\mathrm{loc}}(X, Y)$ on measurable subsets of the rectifiable part of X. We apply Theorem 3.1 in Sects. 4 and 5 below to prove our main results, Theorems 1.2 and 1.3.

As in Remark 2.8, let $u : U \to X$ be a weakly $(4/\pi)$-quasiconformal parametrization of X provided by Theorem 2.7, and $h : U \to Y$, $h := f \circ u$. Then $u \in N^{1,2}_{\mathrm{loc}}(U, X)$ and $h \in N^{1,2}_{\mathrm{loc}}(U, Y)$. By Theorem 2.6, there exists $G_0 \subset U$ with $|G_0|_2 = 0$ and such that (2.3) holds for both u and h and every measurable set $A \subset U \setminus G_0$. We set $X_0 := u(G_0)$.

Theorem 3.1 *(*Area inequality*) If $g : Y \to [0, \infty]$ and $E \subset X \setminus X_0$ are Borel measurable, then*

$$\int_E g(f(x)) \cdot \rho^u_f(x) \rho^l_f(x) \, d\mathcal{H}^2 \leq 4\sqrt{2} \int_Y g(y) \cdot N(y, f, E) \, dy.$$

If in addition, the map f satisfies Lusin's condition (N), then

$$\int_E g(f(x)) \cdot \rho_f^u(x)\rho_f^l(x)\, d\mathcal{H}^2 \geq \frac{1}{4\sqrt{2}} \int_Y g(y) \cdot N(y, f, E)\, dy.$$

In order to establish Theorem 3.1, we make use of the following proposition which can be seen as a counterpart to Lemma 2.10.

Proposition 3.2 *Let f, u and $h = f \circ u$ be as above. Then*

$$\rho_f^u(u(z)) \cdot l_u(z) \leq L_h(z) \quad \textit{and} \quad l_h(z) \leq \rho_f^l(u(z)) \cdot L_u(z) \tag{3.1}$$

for almost every $z \in U \setminus G_0$.

Proof Fix Borel representatives of the maps $z \mapsto \operatorname{ap\,md} u_z$ and $z \mapsto \operatorname{ap\,md} h_z$. Towards the first inequality in (3.1), we denote

$$G_0' = G_0 \cup \{z \in U : l_u(z) = 0\},$$

and notice that it suffices to prove the inequality for almost every $z \in U \setminus G_0'$. By [28, Proposition 4.3], there are pairwise disjoint Borel sets $K_i \subset U \setminus G_0'$, $i \in \mathbb{N}$, so that

$$|U \setminus (G_0' \cup (\cup_i K_i))|_2 = 0 \tag{3.2}$$

and so that for every $i \in \mathbb{N}$ we have

(i) $\operatorname{ap\,md} u_z$ and $\operatorname{ap\,md} h_z$ exist for every $z \in K_i$ and
(ii) for every $\varepsilon > 0$ there is $r_i(\varepsilon) > 0$ so that

$$|d_X(u(z+v), u(z+w)) - \operatorname{ap\,md} u_z(v-w)| \leq \varepsilon |v-w| \quad \text{and}$$
$$|d_Y(h(z+v), h(z+w)) - \operatorname{ap\,md} h_z(v-w)| \leq \varepsilon |v-w|$$

for every $z \in K_i$ and all $v, w \in \mathbb{R}^2$ with $|v|, |w| \leq r_i(\varepsilon)$ and such that $z+v, z+w \in K_i$.

We will show that if $i \in \mathbb{N}$ then almost every curve γ in X parametrized by arclength has the following property: almost every $t \in \gamma^{-1}(u(K_i))$ satisfies

$$\operatorname{ap\,md}(f \circ \gamma)_t \leq \frac{L_h(z)}{l_u(z)} \quad \text{for all } z \in u^{-1}(\gamma(t)) \cap K_i. \tag{3.3}$$

We show how to conclude the first inequality in (3.1) from (3.3). By Lemma 2.4, Corollary 2.5 and (3.3), $\rho \colon X \to [0, \infty]$ is a weak upper gradient of f, where $\rho(x) = \rho_f^u(x)$ for $x \in X \setminus u(K_i)$ and

$$\rho(x) = \inf_{z \in K_i,\, u(z) = x} \frac{L_h(z)}{l_u(z)}$$

when $x \in u(K_i)$. By the definition of minimal weak upper gradients, we then have that

$$\rho_f^u(x) \le \rho(x) \quad \text{for almost every } x \in u(K_i). \tag{3.4}$$

Since $K_i \subset U \setminus G_0'$, we have $l_u > 0$ and thus $J(\operatorname{ap\,md} u_z) > 0$ in K_i. Combining (3.4) with the Area formula (Theorem 2.6) for u now yields

$$\rho_f^u(u(z)) \cdot l_u(z) \le L_h(z)$$

for almost every $z \in K_i$. The first inequality in (3.1) follows from (3.2).

We now prove (3.3). Denote by $\widehat{X} \subset X$ the set of points x for which $N(x, u, U) = 1$. By [36, Remark 7.2], $\mathcal{H}^2(X \setminus \widehat{X}) = 0$. In particular, almost every rectifiable curve $\gamma : [0, \ell(\gamma)] \to X$ parametrized by arclength satisfies $\gamma(t) \in \widehat{X}$ for $\mathcal{H}^1$-almost every $0 < t < \ell(\gamma)$.

We fix such a γ and a density point $t_0 \in \gamma^{-1}(u(K_i) \cap \widehat{X}) =: T$ of T. By Corollary 2.5, we may moreover assume that $f \circ \gamma$ is approximately metrically differentiable at t_0. It suffices to show that (3.3) holds for t_0 and the unique $z_0 = u^{-1}(\gamma(t_0)) \in K_i$.

Fix a sequence (t_j) of points in T converging to t. Then $x_j := \gamma(t_j) \to \gamma(t_0) =: x_0$. Moreover, since $x_0 \in \widehat{X}$, we have $z_j := u^{-1}(x_j) \to z_0$. We are now in position to apply Property (ii) above. Denoting $y_j = f(x_j)$ for $j = 0, 1, \ldots$, (ii) and triangle inequality yield

$$\begin{aligned}
\frac{d_X(x_j, x_0)}{|z_j - z_0|} &\ge \operatorname{ap\,md} u_{z_0}\left(\frac{z_j - z_0}{|z_j - z_0|}\right) - o(|z_j - z_0|) \ge l_u(z_0) - o(|z_j - z_0|), \\
\frac{d_Y(y_j, y_0)}{|z_j - z_0|} &\le \operatorname{ap\,md} h_{z_0}\left(\frac{z_j - z_0}{|z_j - z_0|}\right) + o(|z_j - z_0|) \le L_h(z_0) + o(|z_j - z_0|).
\end{aligned}$$

Combining the inequalities, we have

$$\frac{d_Y(y_j, y_0)}{d_X(x_j, x_0)} = \frac{d_Y(y_j, y_0) \cdot |z_j - z_0|}{|z_j - z_0| \cdot d_X(x_j, x_0)} \le \frac{L_h(z_0)}{l_u(z_0)} + o(|z_j - z_0|). \tag{3.5}$$

Since γ is parametrized by arclength, (3.5) gives (3.3). The first inequality in (3.1) follows. The second inequality follows in a similar way, namely showing that instead of (3.3) we have

$$\operatorname{ap\,md}(f \circ \gamma)_t \ge \frac{l_h(z)}{L_u(z)}$$

outside suitable exceptional sets. We leave the details to the reader. □

Proof of Theorem 3.1 We may approximate g with simple functions and replace E with appropriate subsets to see that it suffices to show the claim for $g \equiv 1$. We set

$E' = E \cap \widehat{X}$, where $\widehat{X}$ is as in the proof of Proposition 3.2, and obtain

$$N(y, h, u^{-1}(E')) = \sum_{x \in f^{-1}(y)} N(x, u, u^{-1}(E')) = N(y, f, E') \tag{3.6}$$

for every $y \in f(E')$.

The area formula (Theorem 2.6) implies

$$\begin{aligned}\int_E \rho_f^u(x)\rho_f^l(x)\, d\mathcal{H}^2 &= \int_{E'} \rho_f^u(x)\rho_f^l(x) N(x, u, u^{-1}(E'))\, d\mathcal{H}^2 \\ &= \int_{u^{-1}(E')} \rho_f^u(u(z))\rho_f^l(u(z)) J(\operatorname{ap}\operatorname{md} u_z)\, dz.\end{aligned}$$

By Lemma 2.9, $J(\operatorname{ap}\operatorname{md} u_z) \leq 2L_u(z) \cdot l_u(z)$ for almost every $z \in u^{-1}(E')$. Moreover, it follows from the proof of Theorem 2.7 given in [35] that we can choose u so that the John ellipse of $\operatorname{ap}\operatorname{md} u_z$ (see (2.6)) is a disk. Then $L_u(z) \leq \sqrt{2} l_u(z)$, which leads to

$$J(\operatorname{ap}\operatorname{md} u_z) \leq 2L_u(z) \cdot l_u(z) \leq 2\sqrt{2} \cdot l_u(z)^2 \quad \text{for almost every } z \in u^{-1}(E').$$

Combining with Lemma 2.10 and Proposition 3.2, we conclude that

$$\int_E \rho_f^u(x)\rho_f^l(x)\, d\mathcal{H}^2 \leq 2\sqrt{2} \int_{u^{-1}(E')} L_h(z) l_h(z)\, dz.$$

Applying Lemma 2.9 and the area formula (Theorem 2.6) to h, we finally obtain

$$\begin{aligned}\int_E \rho_f^u(x)\rho_f^l(x)\, d\mathcal{H}^2 &\leq 4\sqrt{2} \int_{u^{-1}(E')} J(\operatorname{ap}\operatorname{md} h_z)\, dz \\ &= 4\sqrt{2} \int_{f(E')} N(y, h, u^{-1}(E'))\, dy.\end{aligned}$$

The theorem follows by combining with (3.6).

For the second statement we note that f satisfying Lusin's condition (N) implies $\mathcal{H}^2(f(E \setminus E')) = 0$ as, by [36, Remark 7.2], $\mathcal{H}^2(E \setminus E') = 0$. The rest of the proof is analogous to the arguments above. □

4 Openness and discreteness

Throughout this section let f be as in Theorem 1.2, i.e., $f \in N^{1,2}_{\mathrm{loc}}(X, \mathbb{R}^2)$ is non-constant, sense-preserving and satisfies $K_f \in L^1_{\mathrm{loc}}(X)$. Recall that f is continuous by Remark 2.3.

A map $f \colon X \to \mathbb{R}^2$ is *light* if $f^{-1}(y)$ is totally disconnected for every $y \in \mathbb{R}^2$. It is well-known that if f is continuous, sense-preserving and light, then f is open and discrete [43], [40, Lemma VI.5.6]. Thus, in order to prove Theorem 1.2 it suffices to show that f is in fact light. The proof of this fact relies on the following two propositions involving estimates on the multiplicity of f (recall notation $N(y, h, A)$ for multiplicity in (2.4)).

Proposition 4.1 *Suppose that there are $s, r_0 > 0$ and $C > 0$ such that*

$$\int_0^{2\pi} N(f(x_0) + re^{i\theta}, f, B(x_0, s))\, d\theta \le C \log \frac{1}{r} \tag{4.1}$$

for all $r < r_0$. Then the x_0-component of $f^{-1}(f(x_0))$ either is $\{x_0\}$ or contains an open neighborhood of x_0.

Recall that X is homeomorphic to a planar domain. In particular, for every $x_0 \in X$ there is $s > 0$ so that $\overline{B}(x_0, 2s)$ is a compact subset of X.

Proposition 4.2 *Let $x_0 \in X$ and $s > 0$ so that $\overline{B}(x_0, 2s) \subset X$ is compact. Then Condition (4.1) holds with some $r_0, C > 0$.*

Theorem 1.2 follows by combining Propositions 4.1 and 4.2: since f is not constant, for every $y_0 \in f(X)$ every component F of $f^{-1}(y_0)$ contains a point $x_0 \in X$ which is a boundary point of F. Combining Propositions 4.1 and 4.2, we see that $F = \{x_0\}$. We conclude that f is light and therefore open and discrete.

4.1 Proof of Proposition 4.1

Let $f : X \to \mathbb{R}^2$ be a map of finite distortion and Γ a curve family in X. We define the weighted modulus

$$\mathrm{Mod}_{K^{-1}} \Gamma = \inf_g \int_X \frac{g(x)^2}{K_f(x)}\, d\mathcal{H}^2,$$

where the infimum is taken over all weakly admissible functions g for Γ.

Let $u : U \to X$ be a weakly $(4/\pi)$-quasiconformal parametrization of X as in Theorem 2.7. Let $G_0 \subset U$ and $X_0 = u(G_0) \subset X$ be as in the paragraph preceding Theorem 3.1. Recall that $|G_0|_2 = 0$. We set $X' := X \setminus X_0$.

Lemma 4.3 *Let Γ' be a family of curves in $\Omega \subset X$ with $\mathcal{H}^1(|\gamma| \cap X_0) = 0$ for every $\gamma \in \Gamma'$. Then*

$$\mathrm{Mod}_{K^{-1}} \Gamma' \le 4\sqrt{2} \int_{\mathbb{R}^2} g(y)^2 N(y, f, \Omega)\, dy,$$

whenever g is admissible for $\Gamma = f(\Gamma')$.

Proof Fix an admissible g for Γ, and let $g' : X \to \mathbb{R}$,

$$g'(x) := g(f(x)) \cdot \rho_f^u(x) \cdot \chi_{\Omega \cap X'}(x).$$

Here, χ_E denotes the indicator function on a set $E \subset X$, i.e., $\chi_E(x) = 1$ if $x \in E$ and $\chi_E(x) = 0$ else. For almost every $\gamma \in \Gamma'$ we have that f is absolutely continuous on γ, $\mathcal{H}^1(|\gamma| \cap X_0) = 0$, and

$$\int_\gamma g' \, ds = \int_\gamma (g \circ f) \cdot \rho_f^u \, ds \geq \int_{f \circ \gamma} g \, ds,$$

see Corollary 2.5. Since g is admissible for $\Gamma = f(\Gamma')$, it follows that g' is weakly admissible for Γ'. Moreover,

$$\begin{aligned}\operatorname{Mod}_{K^{-1}} \Gamma' &\leq \int_X \frac{g'(x)^2}{K_f(x)} \, d\mathcal{H}^2 = \int_{\Omega \cap X'} g(f(x))^2 \cdot \rho_f^u(x) \rho_f^l(x) \, d\mathcal{H}^2 \\ &\leq 4\sqrt{2} \int_{\mathbb{R}^2} g(y)^2 \cdot N(y, f, \Omega) \, dy,\end{aligned}$$

where the last inequality follows from the area inequality, Theorem 3.1. $\square$

Lemma 4.4 *Let $\varphi \in N^{1,2}_{loc}(X, \mathbb{R})$, and consider $E \subset \mathbb{R}$ with $|E|_1 > 0$ and so that each level set $\varphi^{-1}(t)$, $t \in E$, contains a non-degenerate continuum η_t. Then $\mathcal{H}^1(\eta_t \cap X_0) = 0$ for almost every $t \in E$.*

Proof Note that $\widehat{\varphi} = \varphi \circ u$ is in $N^{1,2}_{\mathrm{loc}}(U, \mathbb{R})$. For every $t \in E$, let $\widehat{\eta_t} = u^{-1}(\eta_t)$. Then, since u is continuous and proper, $\widehat{\eta_t}$ is a non-degenerate continuum for every $t \in E$. Moreover, the coarea inequality for Sobolev functions (Theorem 2.2) shows that $\mathcal{H}^1(\widehat{\eta_t}) < \infty$ for almost every $t \in E$. For every such t, there is a surjective two-to-one 1-Lipschitz curve

$$\widehat{\gamma_t} : [0, 2\mathcal{H}^1(\widehat{\eta_t})] \to \widehat{\eta_t},$$

cf. [41, Proposition 5.1]. Let $\widehat{\Gamma}$ be the family of the curves $\widehat{\gamma_t}$, and let $g : U \to [0, \infty]$ be admissible for $\widehat{\Gamma}$. We apply the coarea inequality for Sobolev functions (Theorem 2.2) and Hölder's inequality to obtain

$$\begin{aligned}|E|_1 &\leq \int_E^* \int_{\widehat{\gamma_t}} g \, ds \, dt \leq 2 \int_E^* \int_{\widehat{\eta_t}} g \, d\mathcal{H}^1 \, dt \leq \frac{8}{\pi} \int_{\widehat{\varphi}^{-1}(E)} g \cdot \rho_{\widehat{\varphi}}^u \, d\mathcal{H}^2 \\ &\leq \frac{8}{\pi} \left(\int_{\widehat{\varphi}^{-1}(E)} g^2 \, d\mathcal{H}^2 \right)^{1/2} \left(\int_{\widehat{\varphi}^{-1}(E)} (\rho_{\widehat{\varphi}}^u)^2 \, d\mathcal{H}^2 \right)^{1/2}.\end{aligned}$$

Since $\rho_{\widehat{\varphi}}^u \in L^2_{\mathrm{loc}}(U)$ and $|E|_1 > 0$ it follows that $\operatorname{Mod}(\widehat{\Gamma}) > 0$. As a Sobolev function, u is therefore absolutely continuous along $\widehat{\gamma_t}$ for almost every $t \in E$, see e.g. [13, Lemma 6.3.1]. Moreover, for almost every $t \in E$ we have that $\mathcal{H}^1(\widehat{\eta_t} \cap G_0) = 0$, since $|G_0|_2 = 0$. Combining these two facts shows that $\mathcal{H}^1(\eta_t \cap X_0) = 0$ for almost every $t \in E$. $\square$

Lemma 4.5 *Let $V \subset X$ be open and connected, and $I, J \subset V$ disjoint non-trivial continua. There are $E \subset \mathbb{R}$, $|E|_1 > 0$, and a family $\Gamma' = \{\gamma_t : t \in E\}$ satisfying*

(1) every $\gamma_t \in \Gamma'$ is a non-degenerate curve connecting I and J in V,
(2) there exists $\varphi \in N^{1,2}_{loc}(V, \mathbb{R})$ such that for every $t \in E$ the curve $\gamma_t \in \Gamma'$ has image in the level set $\varphi^{-1}(t)$, and
(3) $\mathrm{Mod}_{K^{-1}} \Gamma' > 0$.

Proof Replacing V with a compactly connected subdomain if necessary, we may assume that

$$\int_V K_f(x)\, d\mathcal{H}^2(x) = K < \infty. \tag{4.2}$$

Fix points $a \in I$ and $b \in J$ and a continuous curve η joining a and b in V. Define $\varphi\colon X \to \mathbb{R}$ by $\varphi(x) = \mathrm{dist}(x, |\eta|)$. As described in the proof of [39, Proposition 3.5], we find $\varepsilon' > 0$, a set $E_0 \subset (0, \varepsilon')$ with $\mathcal{H}^1(E_0) = 0$, and for every $t \in E = (0, \varepsilon') \setminus E_0$ a rectifiable injective curve γ_t joining I and J in V, with image in the level set $\varphi^{-1}(t)$. We set $\Gamma' = \{\gamma_t : t \in E\}$.

Let $g\colon V \to [0, \infty]$ be admissible for Γ'. We apply the coarea inequality for Lipschitz maps (Theorem 2.1) and Hölder's inequality to obtain

$$\begin{aligned}\varepsilon' &\leq \int_0^{\varepsilon'} \int_{\gamma_t} g\, ds\, dt \leq \frac{4}{\pi} \int_V g(x) K_f(x)^{-1/2} K_f(x)^{1/2}\, d\mathcal{H}^2(x) \\ &\leq \frac{4}{\pi} \left(\int_V K_f(x)\, d\mathcal{H}^2(x) \right)^{1/2} \left(\int_V \frac{g(x)^2}{K_f(x)}\, d\mathcal{H}^2(x) \right)^{1/2}.\end{aligned}$$

Combining with (4.2) gives

$$\mathrm{Mod}_{K^{-1}} \Gamma' \geq \left(\frac{\pi \varepsilon'}{4K} \right)^2 > 0,$$

where we used that the estimate above holds for all admissible functions. □

If Z is a metric surface, $G \subset Z$ a domain, and $E, F \subset \overline{G}$ disjoint sets, we denote by $\Gamma(E, F; G)$ the family of curves joining E and F in $\overline{G}$.

Lemma 4.6 *For any $\varepsilon > 0$ the function $g_\varepsilon\colon \mathbb{R}^2 \to [0, \infty)$ defined by*

$$g_\varepsilon(y) = \varepsilon \left(|y| \log \frac{1}{|y|} \log\log \frac{1}{|y|} \right)^{-1} \chi_{\mathbb{D}(0, e^{-2})}$$

is admissible for $\Gamma(\{0\}, \partial\mathbb{D}(0, e^{-2}); \mathbb{R}^2)$ and

$$\int_{\mathbb{R}^2} g_\varepsilon(y)^2 \log \frac{1}{|y|}\, dy \to 0$$

as $\varepsilon \to 0$.

Proof Fix $\gamma \in \Gamma(\{0\}, \partial\mathbb{D}(0, e^{-2}); \mathbb{R}^2)$. We may assume that $\gamma\colon [0, \ell(\gamma)] \to \mathbb{R}^2$ is parametrized by arclength and $\gamma(0) = 0$. Then $\ell(\gamma) \geq e^{-2}$ and $|\gamma(t)| \leq t$ for every $0 \leq t \leq \ell(\gamma)$. We compute

$$\begin{aligned}\int_\gamma g_1\, ds &= \int_0^{\ell(\gamma)} g_1(\gamma(t))\, dt \\ &= \int_0^{\ell(\gamma)} \left(|\gamma(t)| \log \frac{1}{|\gamma(t)|} \log\log \frac{1}{|\gamma(t)|}\right)^{-1} dt \\ &\geq \int_0^{e^{-2}} \left(t \log \frac{1}{t} \log\log \frac{1}{t}\right)^{-1} dt = \infty,\end{aligned}$$

where the last equality follows since

$$\frac{d}{ds} \log\log\log \frac{1}{s} = -\left(s \log \frac{1}{s} \log\log \frac{1}{s}\right)^{-1}.$$

Thus, $g_\varepsilon = \varepsilon \cdot g_1$ is admissible for $\Gamma(\{0\}, \partial\mathbb{D}(0, e^{-2}); \mathbb{R}^2)$ for any $\varepsilon > 0$.

In order to prove the second claim we use polar coordinates and compute

$$\begin{aligned}\int_{\mathbb{R}^2} g_\varepsilon(y)^2 \log \frac{1}{|y|}\, dy &= \varepsilon^2 \int_{\mathbb{R}^2} \left(|y|^2 \log \frac{1}{|y|} \left(\log\log \frac{1}{|y|}\right)^2\right)^{-1} \chi_{\mathbb{D}(0,e^{-2})}\, dy \\ &= \varepsilon^2 \int_0^{2\pi} \int_0^{e^{-2}} \left(r \log \frac{1}{r} \left(\log\log \frac{1}{r}\right)^2\right)^{-1} dr\, d\varphi.\end{aligned}$$

The last term converges to 0 as $\varepsilon \to 0$ since

$$\frac{d}{ds} \left(\log\log \frac{1}{s}\right)^{-1} = \left(s \log \frac{1}{s} \left(\log\log \frac{1}{s}\right)^2\right)^{-1}.$$

The second claim follows. □

We are now able to prove Proposition 4.1. Let V_0 be the x_0-component of $B(x_0, s)$. Denote the x_0-component of $f^{-1}(f(x_0)) \cap V_0$ by J. We may assume that $V_0 \setminus f^{-1}(f(x_0)) \neq \emptyset$, since otherwise there is nothing to prove. Towards contradiction, assume that J is a non-trivial continuum. Fix another non-trivial continuum $I \subset V_0 \setminus f^{-1}(f(x_0))$.

By scaling and translating the target we may assume that $f(x_0) = 0$, $f(I) \cap \mathbb{D}(0, e^{-2}) = \emptyset$, and that the constant r_0 in Condition (4.1) satisfies $r_0 \geq e^{-2}$. Let Γ' be the curve family from Lemma 4.5. Note that $\Gamma = f(\Gamma')$ is a subfamily of $\Gamma(\{0\}, \partial\mathbb{D}(0, e^{-2}); \mathbb{R}^2)$. Hence, we know from Lemma 4.6 that for any $\varepsilon > 0$ the function g_ε is admissible for Γ. Lemma 4.4 implies that Lemma 4.3 can be applied to our setting and thus

$$\operatorname{Mod}_{K^{-1}} \Gamma' \leq 4\sqrt{2} \int_{\mathbb{R}^2} g_\varepsilon(y)^2 N(y, f, B(x_0, s))\, dy.$$

Since g_ε is symmetric with respect to the origin, combining Assumption (4.1) with polar coordinates yields

$$\begin{aligned}\int_{\mathbb{R}^2} g_\varepsilon(y)^2 N(y, f, B(x_0, s))\, dy &= \int_0^{e^{-2}} r g_\varepsilon(r)^2 \int_0^{2\pi} N(re^{i\theta}, f, B(x_0, s))\, d\theta\, dr \\ &\leq C \int_0^{e^{-2}} r g_\varepsilon(r)^2 \log \frac{1}{r}\, dr = C \int_{\mathbb{R}^2} g_\varepsilon(y)^2 \log \frac{1}{|y|}\, dy.\end{aligned}$$

By the second part of Lemma 4.6, the right hand integral converges to 0 as ε goes to 0. Thus, $\operatorname{Mod}_{K^{-1}} \Gamma' = 0$, contradicting Lemma 4.5. The proof is complete.

4.2 Proof of Proposition 4.2

Let x_0 and s be as in the statement. We may assume that $f(x_0) = 0$. We first show that $f^{-1}(y)$ is totally disconnected for most points $y \in f(X)$ around 0.

Lemma 4.7 *Let β' be the set of those $0 \leq \theta < 2\pi$ for which there is $R_\theta > 0$ so that $f^{-1}(R_\theta e^{i\theta})$ contains a non-degenerate continuum. Then $|\beta'|_1 = 0$.*

Proof We define

$$\varphi \colon X \setminus f^{-1}(0) \to \mathbb{S}^1, \quad \varphi(x) = \frac{f(x)}{|f(x)|},$$

and note that $\rho_f^u/|f|$ is a weak upper gradient of φ. Towards a contradiction we assume that $|\beta'|_1 > 0$. Then there are $\delta, \varepsilon > 0$ and a set $\beta'_\delta \subset \beta'$, $|\beta'_\delta|_1 > 0$, such that for every $\theta \in \beta'_\delta$ there exists $R_\theta \in [\varepsilon, 1]$ for which $f^{-1}(R_\theta e^{i\theta})$ contains a continuum E_θ with $\mathcal{H}^1(E_\theta) \geq \delta$. As in the proof of Lemma 4.4, we see that almost every $\theta \in \beta'_\delta$ the continuum E_θ is the image of a rectifiable curve γ_θ, and the modulus of the family of such curves is positive. By the definition of lower gradients and since $f \circ \gamma_\theta$ is constant by construction, we then have that $\rho_f^l = 0$ almost everywhere in

$$E = \bigcup_{\theta \in \beta'_\delta} E_\theta.$$

Furthermore, since f has finite distortion, also $\rho_f^u = 0$ almost everywhere in E. Let

$$F = \{x \in X : |f(x)| \geq \varepsilon,\ \rho_f^u(x) = 0\} \supset E.$$

We apply the Sobolev coarea inequality (Theorem 2.2) to compute

$$0 < \delta|\beta'_\delta|_1 \leq \int^*_{\beta'_\delta} \mathcal{H}^1(E_\theta)\, d\theta \leq \frac{4}{\pi} \int_F \frac{\rho_f^u}{|f|}\, d\mathcal{H}^2 = 0,$$

a contradiction. The proof is complete. □

Lemma 4.8 *Let β' be the set in Lemma 4.7. There exists $\beta \supset \beta'$ with $|\beta|_1 = 0$, and an open $\Omega' \subset X$, such that*

(1) $f|_{\Omega'}$ is a local homeomorphism, and
(2) if $V = \{te^{i\theta} : t > 0, \theta \in \beta\}$, then $\Omega' \supset X \setminus f^{-1}(V)$.

Proof Set $V' = \{te^{i\theta} : \theta \in \beta', t > 0\}$. Let $y \in f(X) \setminus V'$ and $x \in f^{-1}(y)$. Then, since $\{x\}$ is a component of $f^{-1}(y)$, there is a Jordan domain $\widetilde{U}_x$ in X such that $x \in \widetilde{U}_x$ and $y \notin f(\partial \widetilde{U}_x)$. Let W_x be the y-component of $\mathbb{R}^2 \setminus f(\partial \widetilde{U}_x)$ and U_x the x-component of $f^{-1}(W_x)$. It follows that $f(\partial U_x) \subset \partial W_x$. Indeed, otherwise there is a point $a \in \partial U_x$ with $f(a) \in W_x$ and therefore there exists a neighbourhood Y of $f(a)$ in W_x, but the a-component of $f^{-1}(Y)$ is not contained in U_x, which is a contradiction.

The assumption that f is sense-preserving now implies $f(\partial U_x) = \partial W_x$. Using basic degree theory, we conclude that $f^{-1}(z)$ has at most $\deg(y, f, U_x)$ components in U_x for every $z \in W_x$. Furthermore, arguing as in the proof of Lemma 4.7 we see that for almost every such z all of these components are points. In other words,

$$N(z, f, U_x) \leq \deg(y, f, U_x) < \infty$$

for almost every $z \in W_x$. In particular, every $x \in U_x$ satisfies the conditions in Proposition 4.1, and therefore $f|_{U_x}$ is open and discrete.

We have established the following.

(i) If $y \in f(X) \setminus V'$ and $x \in f^{-1}(y)$, then x has a neighbourhood U_x such that $f|_{U_x}$ is open and discrete.

We define

$$\widehat{\Omega} = \{x \in X : x\text{-component of } f^{-1}(f(x)) \text{ is } \{x\}\}.$$

Note that if $x \in \widehat{\Omega}$, then there exists a neighbourhood Y of $f(x)$ such that the closure of the x-component of $f^{-1}(Y)$ is compact. As above, we find a neighbourhood U_x of x such that $f|_{U_x}$ is open and discrete. In particular, $\widehat{\Omega}$ is open. Moreover, it follows from (i) that $\widehat{\Omega} \supset X \setminus f^{-1}(V')$. We have shown that

(ii) $\widehat{\Omega}$ is open, $f|_{\widehat{\Omega}}$ is open and discrete, and $\widehat{\Omega} \supset X \setminus f^{-1}(V')$.

Denote by $\mathcal{B}_f$ the branch set of $f|_{\widehat{\Omega}}$, i.e., the set of points where $f|_{\widehat{\Omega}}$ fails to be locally invertible, and define

$$\beta'' = \{0 \leq \theta < 2\pi : Re^{i\theta} \in f(\mathcal{B}_f) \text{ for some } R > 0\}.$$

Recall that $\mathcal{B}_f$ is closed and countable, see [6, 7, 44], thus β'' is countable. It follows from Lemma 4.7 and (ii) that the sets $\Omega' = \widehat{\Omega} \setminus \mathcal{B}_f$ and $\beta = \beta' \cup \beta''$ possess the desired properties. $\square$

Lemma 4.9 *Let $m \in \mathbb{N}$, $0 < r < e^{-2}$, and assume that $\overline{B}(x_0, 2s)$ is compact and satisfies $f(\overline{B}(x_0, 2s)) \subset \mathbb{D}(0, 1)$. If*

$$E_m = \{0 \leq \theta < 2\pi : N(re^{i\theta}, f, B(x_0, s)) = m\},$$

then

$$m|E_m|_1 \leq \frac{64\sqrt{2}}{\pi s^2} \int_{F_m} K_f \, d\mathcal{H}^2 \cdot \log \frac{1}{r},$$

where $F_m = \{x \in X : \arg(f(x)) \in E_m\}$.

Proof We assume $|E_m|_1 > 0$, otherwise there is nothing to show. Let β and Ω' be as in Lemma 4.8. We set $E'_m = E_m \setminus \beta$ and note that $|E'_m|_1 = |E_m|_1$ since $|\beta|_1 = 0$. We also denote

$$F'_m = \{x \in X : \arg(f(x)) \in E'_m\} \subset F_m.$$

Fix $\theta \in E'_m$, then

$$f^{-1}(\{te^{i\theta} : t \geq r\}) \subset \Omega'.$$

We can therefore apply path lifting of local homeomorphisms to curves $I_\theta = \{te^{i\theta} : r \leq t \leq 1\}$ as follows: if $\{x_1, ..., x_m\} = f^{-1}(re^{i\theta}) \cap B(x, s)$ then for every $j \in \{1, ..., m\}$ there exists a maximal lift γ_θ^j of I_θ starting at x_j, see [40, Theorem II.3.2]. Note that if $\varphi \colon X \to [0, 2\pi)$ is defined by $\varphi(x) = \arg(f(x))$, then the image of each γ_θ^j is contained in the level set $\varphi^{-1}(\theta)$.

Since $\overline{B}(x, 2s)$ is compact and $f(\overline{B}(x, 2s)) \subset \mathbb{D}(0, 1)$, every curve γ_θ^j connects $B(x, s)$ and $X \setminus B(x, 2s)$, and so $\mathcal{H}^1(|\gamma_\theta^j|) \geq s$. Moreover, $f|_{|\gamma_\theta^j|}$ is injective. It follows that

$$s \cdot m \leq \sum_{j=1}^{m} \mathcal{H}^1(|\gamma_\theta^j|) \leq \mathcal{H}^1(\{x \in X : \arg(f(x)) = \theta\}) \tag{4.3}$$

for every $\theta \in E'_m$.

We combine (4.3) with the Sobolev coarea inequality (Theorem 2.2) and Hölder's inequality to compute

$$\begin{aligned} s \cdot m \cdot |E_m|_1 &= s \cdot m \cdot |E'_m|_1 \\ &\leq \int_{E'_m} \mathcal{H}^1(\{x \in X : \arg(f(x)) = \theta\}) \, d\theta \end{aligned}$$

$$\leq \frac{4}{\pi}\int_{F_m}\frac{\rho_f^u}{|f|}\,d\mathcal{H}^2 \leq \frac{4}{\pi}\int_{F_m} K_f^{1/2}\cdot\frac{(\rho_f^u\cdot\rho_f^l)^{1/2}}{|f|}\,d\mathcal{H}^2$$
$$\leq \frac{4}{\pi}\left(\int_{F_m} K_f\,d\mathcal{H}^2\right)^{1/2}\underbrace{\left(\int_{F_m'}\frac{\rho_f^u\cdot\rho_f^l}{|f|^2}\,d\mathcal{H}^2\right)^{1/2}}_{=:I}.$$

For each $j\in\{1,...,m\}$ we define the curve family

$$\Gamma_j' = \{\gamma_\theta^j : t\in E_m'\}.$$

Lemma 4.4 applied to Γ_j' shows that $\mathcal{H}^1(|\gamma_\theta^j|\cap X_0)=0$ for almost every $\theta\in E_m'$ and every $j\in\{1,...,m\}$, where X_0 is as in Theorem 3.1. Hence, if

$$F_m'' = \{x\in X : x\in|\gamma_\theta^j| \text{ for some } \theta\in E_m' \text{ and } 1\leq j\leq m\}\supset F_m',$$

then $\mathcal{H}^2(F_m''\cap X_0)=0$ and $N(y,f,F_m'')\leq m$ for every $y\in\mathbb{R}^2$. By the area inequality (Theorem 3.1) and polar coordinates,

$$I\leq 4\sqrt{2}\int_{E_m}\int_r^1\frac{N(se^{i\theta},f,F_m'')}{s}\,ds\,d\theta\leq 4\sqrt{2}\cdot|E_m|_1\cdot m\cdot\log\frac{1}{r}.$$

The lemma follows by combining the estimates. □

Proposition 4.2 follows from Lemma 4.9: notice that by scaling we may assume that $f(\overline{B}(x_0,2s))\subset\mathbb{D}(0,1)$, so that the conditions of Lemma 4.9 are satisfied. Recall that the sets F_m are pairwise disjoint. Therefore, summing the estimate in Lemma 4.9 over m gives

$$\int_0^{2\pi} N(re^{i\theta},f,B(x_0,s))\,d\theta = \sum_{m=1}^\infty m|E_m|_1$$
$$\leq C\log\frac{1}{r}\sum_{m=1}^\infty\int_{F_m}K_f(x)\,d\mathcal{H}^2$$
$$\leq C\log\frac{1}{r}\int_X K_f(x)\,d\mathcal{H}^2.$$

We may replace X with a compactly contained subdomain if necessary to guarantee that K_f is integrable. Proposition 4.2 follows.

5 Regularity of the inverse

In this section we study the regularity of the inverse of a mapping of finite distortion and prove Theorem 1.3. Let $f\in N^{1,2}_{\mathrm{loc}}(X,\Omega')$ be a homeomorphism with $K_f\in L^1_{\mathrm{loc}}(X)$, where $\Omega' \subset \mathbb{R}^2$. We set $\phi = f^{-1} \colon \Omega' \to X$ and define $\psi \colon \Omega' \to [0,\infty]$ by

$$\psi(y) = \frac{1}{\rho_f^l(\phi(y))}.$$

Lemma 5.1 *We have*

$$\int_E \psi(y)^2\,dy \le 2\int_{\phi(E)} K_f(x)\,d\mathcal{H}^2(x)$$

for every Borel set $E \subset \Omega'$. In particular, $\psi \in L^2_{loc}(\Omega')$.

Proof Again, let $u : U \to X$, $U \subset \mathbb{R}^2$, be a weakly $(4/\pi)$-quasiconformal parametrization and $h = f \circ u$. Then h is locally in $N^{1,2}(U, \mathbb{R}^2)$ and monotone. Therefore, h satisfies Condition (N) and consequently the euclidean area formula, see [30]. Combining the area formula with distortion estimates established in previous sections, we have

$$\begin{aligned}
\int_E \psi(y)^2\,dy &= \int_{h^{-1}(E)} \frac{J(\operatorname{ap}\operatorname{md} h_z)}{\rho_f^l(u(z))^2}\,dz = \int_{h^{-1}(E)} \frac{L_h(z)\cdot l_h(z)}{\rho_f^l(u(z))^2}\,dz \\
&\le \int_{h^{-1}(E)} \frac{\rho_f^u(u(z))\cdot \rho_f^l(u(z))}{\rho_f^l(u(z))^2} L_u(z)^2\,dz \\
&\le 2\int_{h^{-1}(E)} K_f(u(z))\cdot J(\operatorname{ap}\operatorname{md} u_z)\,dz.
\end{aligned}$$

Here the second equality holds since both the domain and target of h are euclidean domains and the first inequality holds by Lemma 2.10 and Proposition 3.2. The second inequality holds by (2.6) and recalling that we can choose u so that the John ellipses of $\operatorname{ap}\operatorname{md} u_z$ are disks for almost every z. The claim now follows from the area formula for u (Theorem 2.6). □

Lemma 5.2 *Suppose $\alpha : X \to \mathbb{R}$ is 1-Lipschitz. Then $v = \alpha \circ \phi$ is absolutely continuous on almost every line parallel to coordinate axes, and $|\partial_j v| \le \frac{16\sqrt{2}}{\pi}\cdot\psi$ almost everywhere for $j = 1, 2$.*

Proof It suffices to consider horizontal lines. Fix a square $Q \subset \Omega'$ with sides parallel to coordinate axes. By scaling and translating, we may assume that $Q = [0,1]^2$.

By Lebesgue's theorem, there exists a set $\Phi \subset (0,1)$ of full measure so that if $s_0 \in \Phi$ then

$$\frac{1}{2\varepsilon}\int_{F_\varepsilon} \psi(y)\,dy = \frac{1}{2\varepsilon}\int_{s_0-\varepsilon}^{s_0+\varepsilon}\int_{t_1}^{t_2} \psi(t,s)\,dt\,ds \to \int_{t_1}^{t_2} \psi(t,s_0)\,dt \tag{5.1}$$

as $\varepsilon \to 0$ for every $0 \le t_1 < t_2 \le 1$, where $F_\varepsilon = [t_1,t_2]\times[s_0-\varepsilon, s_0+\varepsilon]$.

Fix $s_0 \in \Phi$. The claim now follows from Lemma 5.1 if we can show that

$$|\phi(t_2, s_0) - \phi(t_1, s_0)| \leq \frac{16\sqrt{2}}{\pi} \int_{t_1}^{t_2} \psi(t, s_0)\, dt \tag{5.2}$$

for every $0 \leq t_1 < t_2 \leq 1$.

Given $0 < \varepsilon < \min\{s_0, 1 - s_0\}$ we set $E_\varepsilon = \phi(F_\varepsilon)$. Let $\varphi = \pi_2 \circ f|_{E_\varepsilon}$, where π_2 denotes projection to the s-axis on the (t, s)-plane. By continuity of φ, Lemma 4.4, and the Sobolev coarea inequality (Theorem 2.2) applied to φ, we have

$$\begin{aligned} |\phi(t_2, s_0) - \phi(t_1, s_0)| &\leq \delta(\varepsilon) + \frac{1}{2\varepsilon} \int_{s_0-\varepsilon}^{s_0+\varepsilon} \mathcal{H}^1(\varphi^{-1}(s) \setminus X_0)\, ds \\ &\leq \delta(\varepsilon) + \frac{2}{\pi\varepsilon} \int_{E_\varepsilon \setminus X_0} \frac{\rho_f^u \cdot \rho_f^l}{\rho_f^l} \chi_{\rho_f^l \neq 0}\, d\mathcal{H}^2, \end{aligned}$$

where X_0 is the set in the Area inequality (Theorem 3.1) and $\delta(\varepsilon) \to 0$ as $\varepsilon \to 0$. Combining with Theorem 3.1, we obtain

$$|\phi(t_2, s_0) - \phi(t_1, s_0)| \leq \delta(\varepsilon) + \frac{8\sqrt{2}}{\pi\varepsilon} \int_{F_\varepsilon} \psi(y)\, dy. \tag{5.3}$$

Now (5.2) follows by combining (5.3) and (5.1). □

We are ready to prove Theorem 1.3. Since ϕ is continuous, $d_X(\phi(\cdot), x_0) \in L^2_{\mathrm{loc}}(\Omega')$ for every $x_0 \in X$. By Lemma 5.1 and the ACL-characterization of Sobolev functions (see [13, Theorem 6.1.17]), we see that every v in Lemma 5.2 belongs to $W^{1,2}_{\mathrm{loc}}(\Omega')$ and satisfies $|\nabla v| \leq \frac{32\psi}{\pi}$ almost everywhere. Furthermore, the characterization of Sobolev maps in terms of post-compositions with 1-Lipschitz functions, i.e., in terms of the functions v above (see [13, Theorem 7.1.20 and Proposition 7.1.36]), shows that $\phi \in N^{1,2}_{\mathrm{loc}}(\Omega', X)$. The proof is complete.

Remark 5.3 When $X \subset \mathbb{R}^2$, the $N^{1,2}_{\mathrm{loc}}(X, \mathbb{R}^2)$-regularity assumption in Theorem 1.3 may be replaced with $f \in N^{1,1}_{\mathrm{loc}}(X, \mathbb{R}^2)$. Moreover, the conclusion on the regularity of f^{-1} is more precise, see [16]. While our results only concern $N^{1,2}_{\mathrm{loc}}$-maps, it would be interesting to extend the definition of finite distortion to $N^{1,1}_{\mathrm{loc}}$-maps between metric surfaces and develop basic properties including improvements of Theorem 1.3. One cannot expect the conclusions of Remarks 2.3 and 2.8 to hold in the $N^{1,1}$-setting without additional assumptions; maps $f \in N^{1,1}_{\mathrm{loc}}(X, \mathbb{R}^2)$ of finite distortion need not be continuous or satisfy Condition (N) even when $X \subset \mathbb{R}^2$ (see e.g. [12]).

6 Reciprocal surfaces

Recall the *geometric definition of quasiconformality*: a homeomorphism $f\colon X \to Y$ is *quasiconformal* if there exists $C \geq 1$ such that

$$C^{-1} \operatorname{Mod} f(\Gamma) \leq \operatorname{Mod} \Gamma \leq C \operatorname{Mod} f(\Gamma) \tag{6.1}$$

for each curve family Γ in X.

We say that metric surface X is *reciprocal* if there exists $\kappa > 0$ such that for every topological quadrilateral $Q \subset X$ and for the families $\Gamma(Q)$ and $\Gamma^*(Q)$ of curves joining opposite sides of Q we have

$$\operatorname{Mod} \Gamma(Q) \cdot \operatorname{Mod} \Gamma^*(Q) \leq \kappa.$$

If X is reciprocal, $x \in X$ and $R > 0$ so that $X \backslash B(x, R) \neq \emptyset$, then by [35, Theorem 1.8] we have

$$\lim_{r\to 0} \operatorname{Mod} \Gamma(B(x,r), X \setminus B(x,R); X) = 0. \tag{6.2}$$

Recall that $\Gamma(E, F; G)$ is the family of curves joining E and F in $\overline{G}$.

Reciprocal surfaces are the metric surfaces that admit quasiconformal parametrizations by euclidean domains, see [20, 35, 39]. See [9, 34–36, 39, 41] for further properties of reciprocal surfaces.

It is desirable to find non-trivial conditions which imply reciprocality. For instance, one could hope that the existence of maps satisfying the conditions of Theorem 1.2 forces X to be reciprocal. However, this is not the case.

Proposition 6.1 *Given an increasing $\phi : [1, \infty) \to [1, \infty)$ so that $\phi(t) \to \infty$ as $t \to \infty$, there is a non-reciprocal metric surface X and a homeomorphism $f\colon X \to \mathbb{R}^2$ so that $f \in N^{1,2}_{loc}(X, \mathbb{R}^2)$ and $\phi(K_f)$ is locally integrable.*

The map f_0 defined in the proof below is known as Ball's map ([2]) and illustrates that the integrability condition in Theorem 1.2 is sharp.

Proof Let $f_0\colon \mathbb{R}^2 \to \mathbb{R}^2$ be defined by $f_0(x, y) = (x, \eta(x, y))$, where

$$\eta(x,y) = \begin{cases} |x|y, & 0 \leq |x| \leq 1,\ 0 \leq |y| \leq 1, \\ (2(|y|-1) + |x|(2-|y|))\frac{y}{|y|}, & 0 \leq |x| \leq 1,\ 1 \leq |y| \leq 2, \\ y, & \text{otherwise.} \end{cases}$$

Note that f_0 is not open and discrete since it maps the segment $I = \{0\} \times [-1, 1]$ to the origin. Also, f_0 is the identity outside $(-1, 1) \times (-2, 2)$. Calculating the Jacobian matrix shows that f_0 is sense-preserving and Lipschitz, K_{f_0} is bounded outside $(-1, 1) \times (-1, 1)$, and

$$K_{f_0}(x, y) = \frac{1}{|x|} \quad \text{for all } (x, y) \in (-1, 1) \times (-1, 1). \tag{6.3}$$

It follows that K_{f_0} is not in $L^1_{\mathrm{loc}}(\mathbb{R}^2)$ but $K_{f_0} \in L^p_{\mathrm{loc}}(\mathbb{R}^2)$ for every $0 < p < 1$.

We change the metric on $\mathbb{R}^2$ to obtain the desired metric surface X and $f : X \to \mathbb{R}^2$. Define $\omega : \mathbb{R}^2 \to [0, 1]$ by $\omega(z) = 1$ when $\operatorname{dist}(z, I) \geq 1$ and by

$$\omega(z) = \frac{1}{\phi(\operatorname{dist}(z, I)^{-1})} \tag{6.4}$$

otherwise, where $I = \{0\} \times [-1, 1]$. Moreover, let

$$d_\omega(x, y) := \inf_\gamma \int_\gamma \omega \, ds,$$

where the infimum is taken over all rectifiable curves γ connecting $x, y \in \mathbb{R}^2$.

Now $X = (\mathbb{R}^2/I, d_\omega)$ is homeomorphic to $\mathbb{R}^2$ and has locally finite $\mathcal{H}^2$-measure. Let $\pi : \mathbb{R}^2 \to \mathbb{R}^2/I$ be the projection map, $\mathrm{id}_\omega : \mathbb{R}^2/I \to X$ the identity, and $\pi_\omega : \mathbb{R}^2 \to X$, $\pi_\omega = \mathrm{id}_\omega \circ \pi$.

Since modulus is conformally invariant and ω is a conformal change of metric outside I, the family of curves joining any non-trivial continuum and the point $p := \pi_\omega(I)$ in X has positive modulus. By (6.2), it follows that X is non-reciprocal.

We define $f : X \to \mathbb{R}^2$ by $f := f_0 \circ \pi_\omega^{-1}$. Then f is absolutely continuous on almost every rectifiable curve in X, and $\rho_f^u(z) \leq (\omega(z))^{-1} \cdot L$ for almost every $z \in X$, where L is the Lipschitz constant of f_0. Therefore,

$$\int_E (\rho_f^u)^2 \, d\mathcal{H}^2 \leq L^2 |\pi_\omega^{-1}(E)|_2$$

for every Borel set $E \subset X$. We conclude that $f \in N^{1,2}_{\mathrm{loc}}(X, \mathbb{R}^2)$.

It remains to estimate the integral of $\phi(K_f)$. To this end, notice that since ω is a conformal change of metric, we have

$$K_f(z) = K_{f_0}(\pi_\omega^{-1}(z))$$

for almost every $z \in X$. Therefore, it suffices to check that $\phi(K_f)$ is integrable over $E = \pi_\omega((-1, 1) \times (-1, 1))$. By (6.3) and (6.4), we have

$$\int_E \phi(K_f(z)) \, d\mathcal{H}^2 = \int_{(-1,1)^2} \phi(K_{f_0}) \cdot \omega^2 \, dx \, dy \leq \int_{(-1,1)^2} \frac{1}{\phi(|x|^{-1})} \, dx \, dy < \infty.$$

The proof is complete. □

We prove in [33, Theorem 1.3] that if there is a non-constant $f \in \mathrm{FDP}(X, \mathbb{R}^2)$ (not necessarily a homeomorphism) with *bounded* distortion, then X is reciprocal. We also show (see [33, Corollary 1.2]) that the geometric definition (6.1) is quantitatively equivalent with the path definition (requiring K_f to be bounded) of quasiconformality for homeomorphisms $f : X \to \mathbb{R}^2$. By Williams' theorem [47], the equivalence between the analytic (requiring $C(x)$ to be bounded in (1.5)) and geometric definitions of quasiconformality for homeomorphisms holds in even greater generality.

7 Existence of maximal weak lower gradients

Let X and Y be metric surfaces. We now complete the discussion in Sect. 2.3 by proving that each $f \in N^{1,2}_{\mathrm{loc}}(X, Y)$ has a *maximal weak lower gradient*. Precisely, we claim that there is a weak lower gradient ρ^l_f of f so that if ρ^l is another weak lower gradient of f then

$$\rho^l_f(x) \geq \rho^l(x) \quad \text{for almost every } x \in X.$$

Moreover, ρ^l_f is unique up to a set of measure zero. The proof of these facts is analogous to the existence of minimal weak upper gradients, see [13, Theorem 6.3.20].

First, recall that f is absolutely continuous along almost every curve [13, Lemma 6.3.1]. It follows from [13, Lemma 5.2.16] that if ρ is a weak lower gradient of f and $\sigma\colon X \to [0,\infty]$ is a Borel function such that $\sigma = \rho$ almost everywhere in X, then σ is a weak lower gradient of f. In particular, if $E \subset X$ is Borel and satisfies $\mathcal{H}^2(E) = 0$ then $\rho\chi_{X\setminus E}$ is a weak lower gradient of u, compare to [13, Lemma 6.2.8]. We conclude that if there exists a maximal weak lower gradient ρ^l_f of f, it has to be unique up to sets of measure zero.

To prove existence of ρ^l_f, we may assume without loss of generality that $\mathcal{H}^2(X) < \infty$. Arguing exactly as in the proof of [13, Lemma 6.3.8], we can show that if $\sigma, \tau \in L^2(X)$ are weak lower gradients of a map $f\colon X \to Y$ that is absolutely continuous along almost every curve in X and if E is a measurable subset of X then the function

$$\rho = \sigma \cdot \chi_E + \tau \cdot \chi_{X\setminus E}$$

is a weak lower gradient of f. Now, by choosing $E = \{x \in X : \sigma > \tau\}$, it follows that $\rho\colon X \to [0,\infty]$ defined by

$$\rho(x) = \max\{\sigma(x), \tau(x)\}$$

is a 2-integrable weak lower gradient of f. After applying Fuglede's lemma, see e.g. [13, Section 5.1], we established the following lemma.

Lemma 7.1 *If $f\colon X \to Y$ is absolutely continuous along almost every curve, then the collection $\mathcal{L}$ of 2-integrable weak lower gradients of f is closed under pointwise maximum operations.*

Let $(\rho_i) \subset \mathcal{L}$ be a sequence such that

$$\lim_{i\to\infty} ||\rho_i||_{L^2} = \sup\{||\rho||_{L^2} : \rho \in \mathcal{L}\}.$$

By Lemma 7.1, the sequence (ρ'_i) given by $\rho'_i(x) = \max_{1\leq j\leq i} \rho_j(x)$ is in $\mathcal{L}$. Note that (ρ'_i) is pointwise increasing. The limit function

$$\rho^l_f := \lim_{i\to\infty} \rho'_i$$

is Borel by [13, Proposition 3.3.22]. The monotone convergence theorem implies that $\rho_i' \to \rho_f^l$ in $L^2(X)$ and by Fuglede's lemma $\rho_f^l \in \mathcal{L}$, see e.g. [16, Section 5.1]. By construction, ρ_f^l is a maximal weak lower gradient of f. The proof is complete.

Funding Open access funding provided by University of Fribourg

Data availability The paper has no associated data.

References

1. Astala, K., Iwaniec, T., Martin, G.: Elliptic Partial Differential Equations and Quasiconformal Mappings in the Plane, Princeton Mathematical Series, vol. 48. Princeton University Press, Princeton (2009)
2. Ball, J.M.: Global invertibility of Sobolev functions and the interpenetration of matter. Proc. Roy. Soc. Edinb. Sect. A **88**(3–4), 315–328 (1981)
3. Ball, K.: An Elementary Introduction to Modern Convex Geometry, Flavors of Geometry, pp. 1-58 (1997)
4. Brena, C., Campbell, D.: BV and Sobolev homeomorphisms between metric measure spaces and the plane. Adv. Calc. Var. **16**(2), 363–377 (2023)
5. Bonk, M., Kleiner, B.: Quasisymmetric parametrizations of two-dimensional metric spheres. Invent. Math. **150**(1), 127–183 (2002)
6. Černavskiĭ, A.V.: Finite-to-one open mappings of manifolds. Mat. Sb. (N.S.) **65**(107), 357–369 (1964)
7. Černavskiĭ, A.V.: Addendum to the paper Finite-to-one open mappings of manifolds. Mat. Sb. (N.S.) **66**(108), 471–472 (1965)
8. Evans, L.C., Gariepy, R.F.: Measure Theory and Fine Properties of Functions. Studies in Advanced Mathematics. CRC Press, Boca Raton (1992)
9. Eriksson-Bique, S., Poggi-Corradini, P.: On the sharp lower bound for duality of modulus. Proc. Am. Math. Soc. **150**(7), 2955–2968 (2022)
10. Esmayli, B., Hajłasz, P.: The coarea inequality. Ann. Fenn. Math. **46**(2), 965–991 (2021)
11. Esmayli, B., Ikonen, T., Rajala, K.: Coarea inequality for monotone functions on metric surfaces. Trans. Am. Math. Soc. **376**(10), 7377–7406 (2023)
12. Hencl, S., Koskela, P.: Regularity of the inverse of a planar Sobolev homeomorphism. Arch. Ration. Mech. Anal. **180**(1), 75–95 (2006)
13. Hencl, S., Rajala, K.: Optimal assumptions for discreteness. Arch. Ration. Mech. Anal. **207**(3), 775–783 (2013)
14. Hencl, S., Koskela, P.: Lectures on Mappings of Finite Distortion. Lecture Notes in Mathematics, vol. 2096. Springer, Cham (2014)
15. Heinonen, J., Keith, S.: Flat forms, bi-Lipschitz parameterizations, and smoothability of manifolds. Publ. Math. Inst. Hautes Études Sci. **113**, 1–37 (2011)
16. Heinonen, J., Koskela, P., Shanmugalingam, N., Tyson, J.T.: Sobolev Spaces on Metric Measure Spaces: An Approach Based on Upper Gradients, New Mathematical Monographs, vol. 27. Cambridge University Press, Cambridge (2015)
17. Heinonen, J., Rickman, S.: Geometric branched covers between generalized manifolds. Duke Math. J. **113**(3), 465–529 (2002)

18. Heinonen, J., Sullivan, D.: On the locally branched Euclidean metric gauge. Duke Math. J. **114**(1), 15–41 (2002)
19. Ikonen, T.: Uniformization of metric surfaces using isothermal coordinates. Ann. Fenn. Math. **47**(1), 155–180 (2022)
20. Iwaniec, T., Koskela, P., Onninen, J.: Mappings of finite distortion: monotonicity and continuity. Invent. Math. **144**(3), 507–531 (2001)
21. Iwaniec, T., Martin, G.: Geometric Function Theory and Non-linear Analysis,Oxford Mathematical Monographs. The Clarendon Press, New York (2001)
22. Iwaniec, T., Šverák, V.: On mappings with integrable dilatation. Proc. Am. Math. Soc. **118**(1), 181–188 (1993)
23. Karmanova, M.B.: Area and co-area formulas for mappings of the Sobolev classes with values in a metric space. Sibirsk. Mat. Zh. **48**(4), 778–788 (2007)
24. Kirsilä, V.: Integration by parts on generalized manifolds and applications on quasiregular maps. Ann. Acad. Sci. Fenn. Math. **41**(1), 321–341 (2016)
25. Kauhanen, J., Koskela, P., Malý, J.: Mappings of finite distortion: discreteness and openness. Arch. Ration. Mech. Anal. **160**(2), 135–151 (2001)
26. 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)
27. Luisto, R., Pankka, P.: Stoïlow's theorem revisited. Expo. Math. **38**(3), 303–318 (2020)
28. Lytchak, A., Wenger, S.: Area minimizing discs in metric spaces. Arch. Ration. Mech. Anal. **223**(3), 1123–1182 (2017)
29. Lytchak, A., Wenger, S.: Intrinsic structure of minimal discs in metric spaces. Geom. Topol. **22**(1), 591–644 (2018)
30. Malý, J., Martio, O.: Lusin's condition (N) and mappings of the class W1, n. J. Reine Angew. Math. **458**, 19–36 (1995)
31. Meier, D.: Quasiconformal uniformization of metric surfaces of higher topology. Indiana Univ. Math. J. (2024) **(To appear)**
32. Meier, D., Ntalampekos, D.: Lipschitz-volume rigidity and Sobolev coarea inequality for metric surfaces. J. Geom. Anal. **34**(5), 128–30 (2024)
33. Meier, D., Rajala, K.:Definitions of quasiconformality on metric surfaces. arXiv:2405.07476 (2024)
34. Meier, D., Wenger, S.: Quasiconformal almost parametrizations of metric surfaces. J. Eur. Math. Soc. published online first (2024)
35. Ntalampekos, D., Romney, M.: Polyhedral approximation and uniformization for non-length surfaces. J. Eur. Math. Soc. (2024) (To appear)
36. Ntalampekos, D., Romney, M.: Polyhedral approximation of metric surfaces and applications to uniformization. Duke Math. J. **172**(9), 1673–1734 (2023)
37. Onninen, J., Zhong, X.: Mappings of finite distortion: a new proof for discreteness and openness. Proc. R. Soc. Edinb. Sect. A **138**(5), 1097–1102 (2008)
38. Rajala, K.: Remarks on the Iwaniec-Šverák conjecture. Indiana Univ. Math. J. **59**(6), 2027–2039 (2010)
39. Rajala, K.: Uniformization of two-dimensional metric surfaces. Invent. Math. **207**(3), 1301–1375 (2017)
40. Rajala, K., Romney, M.: Reciprocal lower bound on modulus of curve families in metric surfaces. Ann. Acad. Sci. Fenn. Math. **44**(2), 681–692 (2019)
41. Rešetnjak, Ju. G.: Spatial mappings with bounded distortion. Sibirsk. Mat. Z **8**, 629–658 (1967)
42. Rickman, S.: Quasiregular mappings, Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)], vol. 26. Springer, Berlin (1993)
43. Titus, C.J., Young, G.S.: The extension of interiority, with some applications. Trans. Am. Math. Soc. **103**, 329–340 (1962)
44. Väisälä, J.: Discrete open mappings on manifolds. Ann. Acad. Sci. Fenn. Ser. A **I**(392), 10 (1966)
45. Vodop'yanov, S.K., Gol'dshtein, V.M.: Quasiconformal mappings, and spaces of functions with first generalized derivatives. Sibirsk. Mat. Ž. **17**(3), 515–531 (1976). (**715**)
46. Villamor, E., Manfredi, J.J.: An extension of Reshetnyak's theorem. Indiana Univ. Math. J. **47**(3), 1131–1145 (1998)
47. Williams, M.: Geometric and analytic quasiconformality in metric measure spaces. Proc. Am. Math. Soc. **140**(4), 1251–1266 (2012)