THE CLOSURE OF PLANAR DIFFEOMORPHISMS IN SOBOLEV SPACES

G. DE PHILIPPIS AND A. PRATELLI

ABSTRACT. We characterize the (sequentially) weak and strong closure of planar diffeomorphisms in the Sobolev topology and we show that they always coincide. We also provide some sufficient condition for a planar map to be approximable by diffeomorphisms in terms of the connectedness of its counter-images, in the spirit of Young's characterisation of monotone functions. We finally show that the closure of diffeomorphisms in the Sobolev topology is strictly contained in the class INV introduced by Müller and Spector.

1. INTRODUCTION

1.1. **General overview.** In a series of papers [20, 21, 22] J.W.T. Youngs characterised the closure in the uniform topology of the set of homeomorphisms between two 2-dimensional topological manifolds, by showing that it coincides with the class of monotone maps, i.e. those contiuous maps for which the counter-image of a point is a connected set.

By a trivial gluing argument the same characterisation holds for the uniform closure of the set of homeomorphisms from the two dimensional unit square $\mathcal{Q} := [0,1]^2$ into itself which are equal to the identity on $\partial \mathcal{Q}$ and actually the result can be extended to more general domains, see for instance [14, Section 2].

In recent years, mainly motivated by applications in non-linear elasticity and in geometric function theory, a great deal has been devoted in understanding which is the closure of the class of diffeomorphisms in the weak and/or strong Sobolev topology. In view of applications to non-linear elasticity, it is of particular interest to understand these relations in the case of maps in $W^{1,p}$ for $p \in (1,2)$. Indeed this allows for discontinuous maps which, in principle, may be used to model cavitations, see [2]. In this respect, note that the map $x \mapsto x/|x|$ can be easily obtained as strong $W^{1,p}$ limit of diffeomorphisms if $p \in [1,2)$.

To state the known results and for further use let us first introduce a few notations. For simplicity of notation, we will restrict to functions mapping the closed unit square $\mathcal{Q} = [0,1]^2$ into itself (however, in Theorem 1.5 we give the general claim). For the sake of clarity we will denote by $\mathcal{Q}$ the "domain" unit square and by $\boldsymbol{\mathcal{Q}} = [0,1]^2$ the "target" unit square. In general we will denote by capital letters points and sets in $\mathcal{Q}$ and by bold capital letter points and sets in $\boldsymbol{\mathcal{Q}}$. Hence $u : \mathcal{Q} \to \boldsymbol{\mathcal{Q}}$, and $A, B, \cdots \in \mathcal{Q}$ while $\boldsymbol{A}, \boldsymbol{B}, \cdots \in \boldsymbol{\mathcal{Q}}$. We also set

$$\mathcal{H}_{\mathrm{Id}}(\mathcal{Q}) = \left\{ u \in \mathrm{C}^0(\mathcal{Q}; \boldsymbol{\mathcal{Q}}) \text{ homeomorphism such that } u = \mathrm{Id} \text{ on } \partial \mathcal{Q} \right\},$$

$$\mathcal{D}_{\mathrm{Id}}(\mathcal{Q}) = \left\{ u \in \mathrm{C}^\infty(\mathcal{Q}; \boldsymbol{\mathcal{Q}}) \text{ diffeomorphism such that } u = \mathrm{Id} \text{ on } \partial \mathcal{Q} \right\},$$

$$\mathcal{M}_{\mathrm{Id}}(\mathcal{Q}) = \left\{ u \in \mathrm{C}^0(\mathcal{Q}; \boldsymbol{\mathcal{Q}}) \text{ monotone map such that } u = \mathrm{Id} \text{ on } \partial \mathcal{Q} \right\}.$$

Recall that a continuous map $u : \mathcal{Q} \to \boldsymbol{\mathcal{Q}}$ is said to be *monotone* if for every $\boldsymbol{P} \in \boldsymbol{\mathcal{Q}}$ the set $u^{-1}(\boldsymbol{P}) \subseteq \mathcal{Q}$ is connected. Moreover, for any subset $\mathcal{F} \subseteq W^{1,p}(\mathcal{Q}; \boldsymbol{\mathcal{Q}})$ we denote by

$$\overline{\mathcal{F}}^{\,\mathrm{s}-W^{1,p}} \qquad \text{and} \qquad \overline{\mathcal{F}}^{\,\mathrm{ws}-W^{1,p}}$$

the closure of $\mathcal{F}$ in the *strong* topology and its *weak sequential* closure, i.e., the smallest set containing $\mathcal{F}$ which is sequentially closed[1]. With these notations we can state the following theorem which summarises the work of several authors.

Theorem 1.1. *The following statements hold:*

   (i) *Let $u \in W^{1,p}(\mathcal{Q}) \cap \mathcal{H}_{\mathrm{Id}}(Q)$ and $p \in [1, +\infty)$. Then there exists a sequence $\{u_j\} \subseteq \mathcal{D}_{\mathrm{Id}}(\mathcal{Q})$ such that $\|u_j - u\|_{W^{1,p}(\mathcal{Q})} \to 0$.*

  (ii) *Let $u \in W^{1,p}(\mathcal{Q}) \cap \mathcal{M}_{\mathrm{Id}}(Q)$ and $p \in (1, +\infty)$. Then there exists a sequence $\{u_j\} \subseteq \mathcal{D}_{\mathrm{Id}}(\mathcal{Q})$ such that $\|u_j - u\|_{W^{1,p}(\mathcal{Q})} \to 0$.*

 (iii) *For $p \in [2, +\infty)$,*

$$\overline{\mathcal{D}_{\mathrm{Id}}(\mathcal{Q})}^{\,\mathrm{ws}-W^{1,p}} = \overline{\mathcal{D}_{\mathrm{Id}}(\mathcal{Q})}^{\,\mathrm{s}-W^{1,p}} = \mathcal{M}_{\mathrm{Id}}(\mathcal{Q}) \cap W^{1,p}(\mathcal{Q}) \,.$$

Point (i) has been proved in [12, 13] for $p \in (1, +\infty)$ and [11] for $p = 1$. See also [4] for the case of Orlicz-Sobolev spaces and [7] for the case of bi-Lipschitz maps. Point (ii) as been proved in [14] and point (iii) in [15]. See also the forthcoming paper [5] where point (ii) is extended to the case $p = 1$.

We remark that for the sake of simplicity we have stated the above theorem for the case of maps from $\mathcal{Q}$ to $\boldsymbol{\mathcal{Q}}$ coinciding with the identity on the boundary. However, the quoted results actually work for more general boundary conditions and domains, see in particular [14]. Similarly, also our results directly extend to the case of domains that are bi-Lipschitz equivalent to $\mathcal{Q}$ and of bi-Lipschitz boundary data, see Theorem 1.5.

1.2. **Main results.** The first main result of this paper is a characterisation of $\overline{\mathcal{D}_{\mathrm{Id}}(\mathcal{Q})}^{\,\mathrm{s}-W^{1,p}}$ for *any* $p \in [1 + \infty)$. As a consequence we will obtain the second main result, which states that

$$\overline{\mathcal{D}_{\mathrm{Id}}(\mathcal{Q})}^{\,\mathrm{s}-W^{1,p}} = \overline{\mathcal{D}_{\mathrm{Id}}(\mathcal{Q})}^{\,\mathrm{ws}-W^{1,p}} ,$$

thus extending point (iii) of Theorem 1.1 to $p \in [1, 2)$, see Theorem B below. To state our theorem we first need to give the following definition.

Definition 1.2 (Non-crossing condition)**.** *Let $u : \mathcal{Q} \to \boldsymbol{\mathcal{Q}}$ be a Borel map. We say that u satisfies the non-crossing condition if there is a function $\tilde{u} : \mathcal{Q} \to \boldsymbol{\mathcal{Q}}$ such that $\tilde{u} = u$ $\mathscr{L}^2$ almost everywhere and such that the following three conditions are satisfied:*

  (i) *$\tilde{u} = \mathrm{Id}$ on $\partial \mathcal{Q}$.*

 (ii) *There exists a $\mathscr{H}^1$-negligible set $\mathcal{S}_{\tilde{u}} \subseteq \mathcal{Q}$ such that $\tilde{u}$ is continuous at any point of $\mathcal{Q} \backslash \mathcal{S}_{\tilde{u}}$.*

---

[1] Note that in general the weak sequential closure of $\mathcal{F}$ is strictly smaller than the closure of $\mathcal{F}$ in the weak topology and strictly larger than the set of all weak limit points of sequences in $\mathcal{F}$, see for instance [9, Volume 1, Section 4.1] for a discussion of this fact.

(iii) *Let $N \in \mathbb{N}$, and for any $1 \leq i \leq N$ let $\gamma_i : [0,1] \to \mathcal{Q} \setminus \mathcal{S}_{\tilde{u}}$ be a Lipschitz, injective curve, such that for any $i \neq j$ the intersection $\gamma_i \cap \gamma_j$ is either empty, or a single point $P_{ij} = \gamma_i(t_i) = \gamma_j(t_j)$. In the second case, assume also that both $\gamma_i'(t_i)$ and $\gamma_j'(t_j)$ exist and are not parallel, and $P_{ij} \neq P_{kl}$ whenever $\{i, j\} \neq \{k, l\}$ and both the points are defined. Then, setting $\Gamma = \cup_{i=1}^N \gamma_i([0,1])$, for every $\varepsilon > 0$ there exists a continuous and injective map $\psi_\varepsilon : \Gamma \to \mathbf{\mathcal{Q}}$, coinciding with the identity on $\partial \mathcal{Q} \cap \Gamma$ and such that $\|\tilde{u} - \psi_\varepsilon\|_{L^\infty(\Gamma)} < \varepsilon$.*

We collect the maps satisfying the non-crossing condition in the set

$$\mathcal{NC}_{\mathrm{Id}}(\mathcal{Q}) = \Big\{ u : \mathcal{Q} \to \mathbf{\mathcal{Q}} \ \text{such that } u \text{ satisfies the non-crossing condion} \Big\}.$$

Roughly speaking, a map satisfies the non-crossing condition if every time that we restrict it to a family of possibly intersecting curves, it is possible to make it injective up to a small error in L^∞. We have then the following result.

Theorem A (Characterisation of the closure of diffeomorfisms). *Let $\mathcal{Q} = \mathbf{\mathcal{Q}} = [0,1]^2$, let $p \in [1, +\infty)$ and let $u : \mathcal{Q} \to \mathbf{\mathcal{Q}}$ be a function. Then, there exists a sequence of diffeomorphisms $\{u_j\} \subseteq \mathcal{D}_{\mathrm{Id}}(\mathcal{Q})$ strongly converging in $W^{1,p}$ to u, if and only if $u \in \mathcal{NC}_{\mathrm{Id}}(\mathcal{Q}) \cap W^{1,p}(\mathcal{Q})$. Moreover, the sequence $\{u_j\} \subseteq \mathcal{D}_{\mathrm{Id}}(\mathcal{Q})$ can be chosen such that:*

(i) *If $p \in [1, 2]$, for every j there exists a compact set K_j such that:*

$$\mathscr{H}_\infty^{2-p}(K_j) \leq 2^{-j} \qquad \text{and} \qquad \|u - u_j\|_{L^\infty(\mathcal{Q} \setminus K_j)} \leq 2^{-j}.$$

In particular, defining $K := \bigcap_{\ell \geq 1} \bigcup_{j \geq \ell} K_j$, one has $\mathscr{H}^{2-p}(K) = 0$ and $u_j(x) \to u(x)$ for all $x \in \mathcal{Q} \setminus K$.

(ii) *If $u \in C^0(\mathcal{Q})$ (which is always the case if $p \geq 2$), then $\|u - u_j\|_{L^\infty(\mathcal{Q})} \leq 2^{-j}$.*

Remark 1.3. *Note that in the above theorem we recover the (known) fact that $W^{1,2}$-limits of diffeomorphisms are continuous. By Theorem B below, the same is true for weak limits.*

Since $\overline{\mathcal{D}_{\mathrm{Id}}(\mathcal{Q})}^{\mathrm{ws}-W^{1,p}} \subseteq \mathcal{NC}_{\mathrm{Id}}(\mathcal{Q})$, see Section 3.4, from Theorem A we deduce the following.

Theorem B (Equality of weak and strong closure of diffeomorfisms). *Let $\mathcal{Q} = \mathbf{\mathcal{Q}} = [0,1]^2$. Then, for every $p \in [1, +\infty)$ one has*

$$\overline{\mathcal{D}_{\mathrm{Id}}(\mathcal{Q})}^{\mathrm{s}-W^{1,p}} = \overline{\mathcal{D}_{\mathrm{Id}}(\mathcal{Q})}^{\mathrm{ws}-W^{1,p}}.$$

Remark 1.4. *Note that in the above Theorem we are speaking about* weak *convergence. In particular, when $p = 1$, this requires the* equi-integrability *of the sequence of gradients.*

By the results of Young, a monotone map satisfies the non-crossing condition (hence a corollary of Theorem A is that Theorem 1.1 (ii) is true also for $p = 1$). Since the monotonicity is expressed in terms of the connectedness of $u^{-1}(\mathbf{P})$, it is interesting to understand if the same condition (suitably extended to Sobolev maps) can characterise the weak and/or strong closure of difffeomorphisms. As we shall see with the counterexamples of Section 5 this is actually not the case. However, we can show that a stronger condition is actually sufficient to be a limit of diffeomorphisms.

To understand the condition in Theorem C below, two remarks are in order. First of all, note that for a *continuous monotone* map $u : \mathcal{Q} \to \boldsymbol{\mathcal{Q}}$, it is not hard to show that if $\boldsymbol{C} \subseteq \boldsymbol{\mathcal{Q}}$ is a closed, connected set which does not disconnect $\boldsymbol{\mathcal{Q}}$, i.e. such that $\boldsymbol{\mathcal{Q}} \setminus \boldsymbol{C}$ is connected, then $u^{-1}(\boldsymbol{C}) \subseteq \mathcal{Q}$ is a closed, connected set which does not disconnect $\mathcal{Q}$, see Lemma 4.24.

Second, by the ideas of Šverák and Müller–Spector, [17, 19], to any Sobolev map $u \in W^{1,p}$ we can associate a uniquely defined *multivalued map* $u : \mathcal{Q} \to 2^{\boldsymbol{\mathcal{Q}}}$, see Definition 2.3. This allows to define the counter-image of subsets of $\boldsymbol{\mathcal{Q}}$ as well, as those points whose image intersects the set, that is, for every $\boldsymbol{C} \subseteq \boldsymbol{\mathcal{Q}}$ we let $u^{-1}(\boldsymbol{C}) = \left\{ P \in \mathcal{Q} : u(P) \cap \boldsymbol{C} \neq \emptyset \right\}$. Our sufficient condition reads then as follows.

Theorem C (Sufficiency of the non-disconnecting property). *Let $\mathcal{Q} = \boldsymbol{\mathcal{Q}} = [0,1]^2$, let $1 \leq p < \infty$, and let $u : \mathcal{Q} \to \boldsymbol{\mathcal{Q}}$ be a $W^{1,p}$ function, coinciding with the identity on the boundary. Suppose that, whenever $\boldsymbol{C} \subseteq \boldsymbol{\mathcal{Q}}$ is a closed, connected set which does not disconnect $\boldsymbol{\mathcal{Q}}$, then $u^{-1}(\boldsymbol{C}) \subseteq \mathcal{Q}$ is a closed, connected set which does not disconnect $\mathcal{Q}$. Then, u satisfies the non-crossing condition.*

All the above results are invariant by a bi-Lipschitz change of variables. In particular, we have the following.

Theorem 1.5. *Let Ω be a bounded open set with Lipschitz boundary, and let $\Phi : \mathbb{R}^2 \to \mathbb{R}^2$ be a bi-Lipschitz function. Then, the claims of Theorems A, B and C are valid for maps from Ω to $\Phi(\Omega)$ coinciding with Φ on $\partial\Omega$.*

This is indeed a direct consequence of our main results. To prove it, take a bi-Lipschitz map $\Psi : \mathbb{R}^2 \to \mathbb{R}^2$ such that $\mathcal{Q}_H := \Psi(\Omega)$ is the unit square with finitely many Lipschitz holes. Composing left with $\Psi \circ \Phi^{-1}$ and right with Ψ^{-1} transforms Sobolev maps from Ω to $\Phi(\Omega)$ coinciding with Φ on $\partial\Omega$ onto Sobolev maps from $\mathcal{Q}_H$ on itself equal to the identity on $\partial\mathcal{Q}_H$ and vice versa. Extending all these maps as the identity on $\mathcal{Q} \setminus \mathcal{Q}_H$ allows to directly apply Theorems A, B and C.

It would be of interest to understand how to extend the above results in the more general case of boundary data which extend to a monotone mapping. Our techniques do not seem to straightforwardly apply to this case.

1.3. **Müller and Spector INV condition and counterexamples.** As we already mentioned, in the study of vectorial variational problems, there has been a great interest in understanding weak notions of invertibility which are closed by weak Sobolev limit and which can include discontinuous map.

In [17], Müller and Spector introduced a condition which they called INV (compare also with the weak diffeomorphism introduced by Giaquinta, Modica and Soucek [9, Volume 2, Chapter 2]) and which since then has revealed to be a reasonable substitute of invertibility, in particular INV maps with almost everywhere positive Jacobian share some of reasonable properties one would expect from smooth diffeomorphisms. These made them a natural class of maps on which to settle vectorial variational problems which would be naturally settled on the class of diffeomorphims.

Refering the reader to Section 2.1 for the precise definition and some of the main properties of INV functions, let us just say here that the INV condition does not allow the interpenetration of matter, even after the formation of cavitation.

Since it is not hard to see that the INV condition is closed by (sequentially) $W^{1,p}$ weak limits and since, trivially, diffeomorphisms satisfy this condition, we deduce

$$\overline{\mathcal{D}_{\mathrm{Id}}(\mathcal{Q})}^{\mathrm{s}-W^{1,p}} \subseteq \overline{\mathcal{D}_{\mathrm{Id}}(\mathcal{Q})}^{\mathrm{ws}-W^{1,p}} \subseteq INV(\mathcal{Q}),$$

see Corollary 2.7.

A natural conjecture (to which also the authors have believed for a while) is then that the closure of diffeomorphisms actually coincides with INV maps. This is however not the case, not even for INV maps with almost everywhere positive Jacobian (a class of maps which enjoy better properties, as mentioned above). Indeed, in Section 5.3 we will construct a map which satisfies the INV condition, has positive Jacobian almost everywhere and cannot be approximated by diffeomorphisms. Furthermore, we will show that the INV condition is in general not invariant by reparametrization of the domain, while this invariance holds true under the additional assumption of positive Jacobian almost everywhere, as proved in [17, Section 9]).

Finally, again in Section 5, we will present an example showing that in Theorem C it is not enough to require that the counter-image of any point is a connected set (in particular, not requiring that it does not disconnect) as in the case of continuous maps (see Theorem 1.1 (ii)).

We conclude the introduction by mentioning that we actually do not know if in Theorem C the requirement that $u^{-1}(C)$ is a connected set that does not disconnect $\mathcal{Q}$ for every connected set C which does not disconnect $\mathcal{Q}$ can be weakened by restricting oneself only to the case when C is a point, see also Remark 4.23. Keep in mind that, for continuous maps, these two conditions are easily seen to be equivalent, and in fact they coincide with the monotonicity, i.e., to the mere requirement that the counter-image of a point is connected, see Lemma 4.24.

1.4. **Organisation of the paper and structure of the proof.** The paper is organised as follows. In Section 2 we present some technical definitions and their main properties that we are going to use, in particular we will speak about Sobolev functions, degree, INV condition and Lebesgue squares. In Sections 3 and 4 we prove Theorems A, B and C, and in Section 5 we provide some examples and counterexamples.

We now briefly explain the main ideas behind the proofs of the main results :The main difficulty in the proof of the sufficiency part of Theorem A is that the maps we are dealing with are both discontinuous and non-injective. When dealing with continuous maps, the idea of Young to solve the issue of non injectivity was to perform a fine partition on the target domain $\mathcal{Q}$ with small squares $\boldsymbol{Q}_i$. The map is then approximated separately on each $u^{-1}(\boldsymbol{Q}_i)$ by a bijective map. This same strategy is performed in the case of Sobolev functions in [12, 13] thanks to the p-Laplacian version of the Radó-Knesser theorem proved in [1]. On the other hand, the strategy performed in [11] is based on a suitable decomposition of the domain and to several extension lemmas which allow to extend an injective boundary data to a diffeomorphism in the interior with a suitable control on the Sobolev norms.

The proof of Theorem A mixes both strategies. We start by constructing a "good" grid $\mathscr{G}$ on the domain $\mathcal{Q}$ such that $u \in W^{1,p}(\mathscr{G})$ and there are good estimates on the Sobolev norm of the function when restricted to its edges, see Lemma 3.3. Moreover, the grid can be chosen such that on a good percentage of the squares the function is very close to an affine one. The aim is now to construct a good injective approximation of $u \llcorner \mathscr{G}$ and then to extend it to a piecewise linear homeomorphism on each square of the grid. To this end we start by constructing a grid $\widetilde{\mathscr{G}}$ on the target domain $\boldsymbol{\mathcal{Q}}$, with side length much smaller then the one of $\mathscr{G}$ and such that the image of $u(\mathscr{G})$ intersects $\widetilde{\mathscr{G}}$ transversally, see Lemma 3.6.

Once this second grid has been constructed, we consider $\mathscr{G} \cap u^{-1}(\widetilde{\mathscr{G}})$, which is a finite collection of points dividing each side of $\mathscr{G}$ in finitely many intervals. The idea will be then to substitute on each of these segments the function with an affine one. If we do so by keeping the values at the points of $\mathscr{G} \cap u^{-1}(\widetilde{\mathscr{G}})$, this will preserve the good control on the Sobolev norms, but the resulting function might be non-injective; indeed, there can be two or more points in $\mathscr{G}$ whose image is the same point in $\widetilde{\mathscr{G}}$. To solve this issue, we will make use of the non-crossing condition in order to slightly separate the images of the points in $\mathscr{G} \cap u^{-1}(\widetilde{\mathscr{G}})$, so to get a piecewise linear and injective approximation on $\mathscr{G}$ without destroying the control on the Sobolev norms, see Proposition 3.8. At this point, we will use known extension results to obtain a piecewise affine homeomorphism which approximates u. The smooth approximation is then obtained thanks to the results in [16]. Theorem B will then easily follow from Theorem A.

Concerning Theorem C, the initial strategy is the same as the one described above for the proof of Theorem A. However, we cannot rely anymore on the non-crossing condition in order to define an injective function on the grid. Hence, we need another rule to separate the images of the points in $\mathscr{G} \cap u^{-1}(\widetilde{\mathscr{G}})$. It is here that a key role is played by the assumption that $u^{-1}(\boldsymbol{P})$ is a connected set that does not disconnect $\boldsymbol{\mathcal{Q}}$ for every $\boldsymbol{P} \in \boldsymbol{\mathcal{Q}}$. Indeed, this allows to define the relation of *right* and *left connectibility* among the points in $u^{-1}(\boldsymbol{P}) \cap \mathscr{G}$ and consequently to "blow-up" $\boldsymbol{P} \in u(\mathscr{G}) \cap \widetilde{\mathscr{G}}$, see Section 4.2. Verifying that this rule is actually well defined and provides a injective map is the main content of Chapter 4.

2. SOME PRELIMINARY DEFINITIONS AND FACTS

In this Section we collect some preliminary definitions and some technical facts, mainly known. In particular, in Section 2.1 we discuss the definition of degree and the INV condition, while in Section 2.2 we present the definition of "Lebesgue squares" and we recall some extension results.

2.1. **The definition of degree and the INV condition.** Recall that, given a map $u \in W^{1,1}(\mathcal{Q})$, the limit

$$\lim_{r \to 0} \frac{1}{\pi r^2} \int_{B(P,r)} u(x)\, dx$$

exists for all P outside a $\mathscr{H}^1$-negligible set. We define the *precise representative* of u as the map $u^*(P) = (u_1^*(P), u_2^*(P))$ given by

$$u_i^*(P) = \limsup_{r \to 0} \frac{1}{\pi r^2} \int_{B(P,r)} u_i(x)\, dx\,.$$

Note that, by the 1-dimensional Sobolev embedding, u^* is continuous when restricted to "almost every" curve. In particular we have the following simple remark.

Remark 2.1. *Let $u \in W^{1,1}(\mathcal{Q})$, $D \subset\subset \mathcal{Q}$ be a domain with Lipschitz boundary, and for ε small let*

$$D_\varepsilon = \{x : \mathrm{sdist}(x, \partial D) \leq \varepsilon\} \subset\subset Q$$

where sdist *denotes the signed distance from ∂D. Then, $u^* \llcorner \partial D_\varepsilon$ is continuous for almost every small ε.*

Let us start with the definition of degree for a continuous planar map.

Definition 2.2. *Let $\alpha : \mathbb{S}^1 \to \mathbb{R}^2$ be a continuous curve and let $u : \mathbb{R}^2 \to \mathbb{R}^2$ be a function such that $u \circ \alpha \in C^0(\mathbb{S}^1)$. Then for any $\boldsymbol{P} \in \mathbb{R}^2 \setminus u(\alpha(\mathbb{S}^1))$ we define $\deg(\boldsymbol{P}, u \circ \alpha)$ as the winding number of the curve $u \circ \alpha$ with respect to $\boldsymbol{P}$.*

By Remark 2.1, it is thus possible to define the degree of a $W^{1,1}$ function when restricted to "almost every" curve. If α parametrises the boundary of a domain ∂D on which u^* is continuous, we will sometimes write $\deg(\boldsymbol{P}, u^* \llcorner \partial D)$ to denote the degree of $\boldsymbol{P}$ with respect to $u^* \circ \alpha$.

We now show how, it is possible to associate to every function $u \in W^{1,1}(\mathcal{Q})$ a multivalued function defined at *every* point as follows, see [17, 19].

Definition 2.3. *Let $u : \mathcal{Q} \to \boldsymbol{\mathcal{Q}}$ be a $W^{1,1}$ function. For every Lipschitz open set $D \subseteq \mathcal{Q}$ such that $u^* \llcorner \partial D$ is continuous, we set*

$$F_u(D) = \{\boldsymbol{P} \in \boldsymbol{\mathcal{Q}} : \deg(\boldsymbol{P}, u^* \llcorner \partial D) \neq 0\} \cup u^*(\partial D)\,.$$

Given $P \in \mathcal{Q}$, for $\varepsilon > 0$ we set

$$\mathcal{D}_{P,\varepsilon} = \{D \subseteq \mathcal{Q} : D\ open,\ P \in D \subseteq B(P, \varepsilon)\ and\ u^* \llcorner \partial D\ is\ continuous\}\,,$$

where $B(P, \varepsilon)$ is the ball centered in P with radius ε, and we define

$$u_{\mathrm{multi}}(P) = \{\boldsymbol{P} \in \boldsymbol{\mathcal{Q}} : \exists\, \varepsilon_n \searrow 0,\ D_n \in \mathcal{D}_{P,\varepsilon_n},\ \boldsymbol{P}_n \in F_u(D_n),\ \boldsymbol{P}_n \to \boldsymbol{P}\}\,.$$

Notice that $u_{\mathrm{multi}}(P) : \mathcal{Q} \to 2^{\boldsymbol{\mathcal{Q}}}$ is a pointwise defined multivalued map; moreover, for every $P \in \mathcal{Q}$ the set $u_{\mathrm{multi}}(P)$ is by definition a non-empty closed set. With a small abuse of notation, we will often refer to u_{multi} simply as u and, accordingly, for any $\boldsymbol{C} \subseteq \boldsymbol{\mathcal{Q}}$ we define

$$u^{-1}(\boldsymbol{C}) = \{P \in \mathcal{Q} : u_{\mathrm{multi}}(P) \cap \boldsymbol{C} \neq \emptyset\}\,.$$

Notice that the above definition is consistent; in particular, whenever P is a continuity point for u^* the set $u_{\mathrm{multi}}(P)$ reduces to the sole point $u^*(P)$. Moreover, it is easy to show that the same happens if P is a Lebesgue point for both u and Du. In particular, the multivalued function u_{multi} and the precise representative u^* coincide almost everywhere.

We can now report the definition of INV map as introduced by Müller and Spector in [17]. see also [6].

Definition 2.4 (INV and INV$^+$ conditions)**.** *Let $u : \mathcal{Q} \to \mathcal{Q}$ be a $W^{1,1}$ function. One says that u satisfies the INV condition if for every $P \in \mathrm{Int}(\mathcal{Q})$ and almost every $r < \mathrm{dist}(P, \partial \mathcal{Q})$ such that u^* is continuous on $\partial B(P, r)$ we have*

    (i) *$u(Q) \in F_u(B(P, r))$ for a.e. $Q \in B(P, r)$;*
    (ii) *$u(Q) \in \mathcal{Q} \setminus F_u(B(P, r)) \cup u^*(\partial B(P, r))$ for a.e. $Q \in \mathcal{Q} \setminus B(P, r)$.*

One says that u satisfies the INV^+ condition if $u \circ \Phi$ satisfies the INV condition for every bi-Lipschitz map $\Phi : \mathcal{Q} \to \mathcal{Q}$. We collect all INV (resp., INV^+) maps into the set $INV(\mathcal{Q})$ (resp., $INV^+(\mathcal{Q})$).

Roughly speaking, a map satisfies condition INV if for most balls the image of what is inside (resp., outside) remains inside or on the boundary (resp., outside or on the boundary). Condition INV^+ requires this for more general sets instead than just for balls (this condition was already called INV' in the paper [10]).

Clearly, condition INV^+ is invariant by re-parametrization of the domain, while it is less clear whether the same holds true for condition INV. This is actually the case if $\det Du > 0$ a.e., see [17, Section 9], but as we will show in Section 5.1 it is false in general.

Two are the main properties of INV maps that we are going to need in the sequel: that they are closed under weak convergence, and that they are continuous outside a small set. Both these facts have been established in [17], but since we need them in a slightly different setting we provide here their proofs.

Lemma 2.5. *Let $u \in W^{1,p}(\mathcal{Q}) \cap INV(\mathcal{Q})$, $1 \leq p \leq 2$. Then, there exists a set $\mathcal{S}_u \subseteq \mathcal{Q}$ such that $\mathscr{H}^{2-p}(\mathcal{S}_u) = 0$ and $u \llcorner (\mathcal{Q} \setminus \mathcal{S}_u)$ is continuous.*

Proof. We note that, according to the INV condition, given $P \in \mathcal{Q}$, then for almost every r small

$$\mathrm{osc}_{B(P,r)}\, u \leq \mathrm{osc}_{\partial B(P,r)}\, u \leq \int_{\partial B(P,r)} |Du|. \tag{2.1}$$

Here by $\mathrm{osc}_{B(P,r)}\, u$ we mean the oscillation of u intended as a multivalued function, cf. Definition 2.3; that is,

$$\mathrm{osc}_{B(P,r)}\, u = \mathrm{diam} \bigcup_{Q \in B(P,r)} u(Q).$$

By integrating (2.1) with respect to $r \in (\rho, 2\rho)$ and using the monotonicity of $r \to \mathrm{osc}_{B(P,r)}\, u$ we get

$$\mathrm{osc}_{B(P,\rho)}\, u \leq \frac{1}{\rho} \int_{B(P,2\rho)} |Du| \leq (4\pi)^{1/p'} \rho^{1-2/p} \Big(\int_{B(P,2\rho)} |Du|^p \Big)^{1/p}. \tag{2.2}$$

In particular,

$$\big\{ P : u \text{ is not continuous at } P \big\} = \big\{ P : \liminf_{r \to 0} \mathrm{osc}_{B(P,r)}\, u > 0 \big\}$$

$$\subseteq \Big\{ P : \limsup_{r \to 0} r^{p-2} \int_{B(P,r)} |Du|^p > 0 \Big\}.$$

Since the latter set is $\mathscr{H}^{2-p}$-negligible (see for instance [8, Section 2.4.3]), we conclude the proof. $\qquad\square$

Lemma 2.6. *Let $\{u_k\} \subseteq W^{1,p}(\mathcal{Q}) \cap INV(\mathcal{Q})$ (resp., $\{u_k\} \subseteq W^{1,p}(\mathcal{Q}) \cap INV^+(\mathcal{Q})$), be such that $u_k \rightharpoonup u$ in $W^{1,p}$. Then, $u \in W^{1,p}(\mathcal{Q}) \cap INV(\mathcal{Q})$ (resp., $u \in W^{1,p}(\mathcal{Q}) \cap INV^+(\mathcal{Q})$).*

Proof. We will prove the Lemma only in the case $p = 1$ and for the INV condition, the other cases being analogous. Recall that in this case, by the Dunford-Pettis Theorem, there exists a super-linear function g such that

$$\sup_k \int_{\mathcal{Q}} g(|Du_k|) < +\infty.$$

We need to show that that for every $P \in \mathcal{Q}$ and almost every radius r

$$u(Q) \in F_u(B(P,r)) \qquad \text{for almost all } Q \in B(P,r).$$

To see this, note that for almost every $r > 0$ there is a (not re-labeled) subsequence (possibly depending on r) such that

$$\limsup_k \int_{\partial B(P,r)} g(|Du_k|) < +\infty \text{ and } \|u_k^* - u^*\|_{L^\infty(\partial B(P,r))} \to 0 \,.$$

In particular, $u^* \llcorner \partial B(P,r)$ is continuous. Let now $Q \in B(P,r)$ be such that $u^*(Q)$ exists as the limit of averages and such that $u_k^*(Q) \to u^*(Q)$. If $u^*(Q) \in u^*(\partial B(P,r))$ there is nothing to prove, hence we can assume that $u^*(Q) \notin u^*(\partial B(P,r))$ and thus, for k large, $u_k^*(Q) \notin u_k^*(\partial B(P,r))$. Since u_k satisfies the INV condition and $Q \in B(P,r)$ we have that

$$\deg(u_k^*(Q), u_k^* \llcorner \partial B(P,r)) \neq 0.$$

By the continuity of the degree we deduce that $\deg(u^*(Q), u^* \llcorner \partial B(P,r)) \neq 0$. By the same argument one shows that for almost all $Q \in \mathcal{Q} \setminus B(P,r)$ either $u^*(Q) \in u^*(\partial B(P,r))$ or $\deg(u^*(Q), u^* \llcorner \partial B(P,r)) = 0$. $\qquad\square$

Note that since every diffeomorphism is clearly an INV^+ map we have the following result.

Corollary 2.7. *One has*

$$\overline{\mathcal{D}_{\mathrm{Id}}(\mathcal{Q})}^{\mathrm{ws}-W^{1,p}} \subseteq INV^+(\mathcal{Q}) \,.$$

We now show that a map which equals the identity on $\partial \mathcal{Q}$ and for which the counter-image of any point is connected satisfies condition INV^+.

Lemma 2.8. *Let $u : \mathcal{Q} \to \mathcal{Q}$ be a $W^{1,p}$ function such that $u = \mathrm{Id}$ on $\partial \mathcal{Q}$ and for every $\boldsymbol{P} \in \mathrm{Int}(\boldsymbol{\mathcal{Q}})$ the set $u^{-1}(\boldsymbol{P})$ is connected. Then u satisfies the INV^+ condition. Moreover, whenever $\gamma : \mathbb{S}^1 \to \mathcal{Q}$ is an injective, Lipschitz continuous, counterclockwise oriented curve on which u^* is continuous, one has $\deg(\boldsymbol{P}, u^* \circ \gamma) \in \{0, 1\}$ for every $\boldsymbol{P} \in \boldsymbol{\mathcal{Q}} \setminus u^*(\gamma)$.*

Proof. Up to a re-parametrization (i.e., by pre-composing u with a bi-Lipschitz map Φ) it is enough to show that the conditions in Definition 2.4 are satisfied for a fixed square $S \subseteq \mathcal{Q}$ with sides parallel to $\partial \mathcal{Q}$ and such that $u^* \llcorner S$ is continuous.

Step I. Internal points have internal counter-images.

Let $\boldsymbol{P} \in \mathcal{Q}$ be such that $\deg(\boldsymbol{P}, u^* \llcorner \partial S) \neq 0$. We want to show that there exists $P \in S$ such $\boldsymbol{P} \in u(P)$. To do so, let us divide divide S in four rectangles S^i, $i = 1, \ldots, 4$, such that $u^* \llcorner \partial S^i$ is continuous; notice that, since u^* is continuous on most of the vertical and horizontal segments, we can assume that $\mathrm{diam}(S^i) \leq 3 \, \mathrm{diam}(S)/4$. Then, either $\boldsymbol{P} \in u^*(\partial S^i)$ for some i or $\deg(\boldsymbol{P}, u^* \llcorner \partial S^i)$ is well defined for all $i = 1, \ldots, 4$. In the first case we have that $\boldsymbol{P} \in u(P)$ for some $P \in S$, so we are already done, while in the latter case we have, by the additivity of the degree, that

$$0 \neq \deg(\boldsymbol{P}, u^* \llcorner \partial S) = \sum_{i=1}^4 \deg(\boldsymbol{P}, u^* \llcorner \partial S^i).$$

In particular $\deg(\boldsymbol{P}, u^* \llcorner \partial S^i) \neq 0$ for some i. Iterating this argument, either we prove our claim, or we find a sequence of encapsulated rectangles S_k with $\mathrm{diam}(S_k) \to 0$ such that $0 \neq \deg(\boldsymbol{P}, u^* \llcorner \partial S_k)$. If $P = \cap_k S_k$, one has then that $\boldsymbol{P} \in u(P)$.

Step II. External points have external counter-images.

Let $\boldsymbol{P} \in \mathcal{Q}$ be such that $\deg(\boldsymbol{P}, u^* \llcorner \partial S) = 0$, so in particular $\boldsymbol{P} \notin u^*(\partial S)$: we want to show that there exists a point $P \in \mathcal{Q} \setminus S$ such that $P \in u^{-1}(\boldsymbol{P})$. To this end note that, up to a bi-Lipschitz change of variable, we can assume that S is a dyadic square, i.e., of the form $(i2^{-k}, (i+1)2^{-k}) \times (j2^{-k}, (j+1)2^{-k})$ for some $k \in \mathbb{N}$ and $0 \leq i, j < 2^k$, and that u^* is continuous on the boundary of every dyadic square with side length 2^{-k}. Let S_m, for $1 \leq m \leq 2^{2k}$, be the collection of these squares and let us assume that they are numbered so that $S_1 = S$. Now, either $\boldsymbol{P} \in \bigcup_m u^*(\partial S_m)$, and in this case we are done, or by the additivity of the degree:

$$1 = \deg(\boldsymbol{P}, u^* \llcorner \partial \mathcal{Q}) = \sum_m \deg(\boldsymbol{P}, u^* \llcorner \partial S_m),$$

where the first equality follows form the fact that $u = \mathrm{Id}$ on $\partial \mathcal{Q}$. Since $\deg(\boldsymbol{P}, u^* \llcorner \partial S_1) = 0$ by assumption, there exists $m \geq 2$ such that $\deg(\boldsymbol{P}, u^* \llcorner \partial S_m) \neq 0$. By the previous step there exists $P \in S_m \subseteq \mathcal{Q} \setminus S$ such that $P \in u^{-1}(\boldsymbol{P})$, and this concludes the proof of the claim.

Step III. Conclusion.

The conclusion is now simple. Indeed, let $Q \in S$ be such that $u_{\mathrm{multi}}(Q) = u^*(Q) = \boldsymbol{Q}$ is uni-valued, and assume by contradiction that

$$\deg(\boldsymbol{Q}, u^* \llcorner S) = 0,$$

so that in particular $\boldsymbol{Q} \notin u^*(\partial S)$. By the second step, there exists $P \in \mathcal{Q} \setminus \overline{S}$ such that $P \in u^{-1}(\boldsymbol{Q})$. Then, $u^{-1}(\boldsymbol{Q})$ contains $Q \in S$, $P \in \mathcal{Q} \setminus \overline{S}$, and no point in ∂S, and this is a contradiction with the fact that $u^{-1}(\boldsymbol{Q})$ is connected by assumption; thus, the first condition in Definition 2.4 is satisfied. In the very same way one checks the second one.

It only remains to show the last claim, namely, that the degree can only assume the values 0 and 1. To prove this, let S be the domain bounded by γ; notice that, since we can extend u as the identity in an external neighborhood of $\mathcal{Q}$, it is not restrictive to assume that the image of γ lies in the interior of $\mathcal{Q}$. Hence, up to a bi-Lipschitz change of coordinates, we can assume that S is a dyadic square as in Step II.

Let $\boldsymbol{P} \in \mathcal{Q} \setminus u^*(\partial S)$ be such that $\deg(\boldsymbol{P}, u^* \llcorner \partial S) \neq 0$: we need to show that $\deg(\boldsymbol{P}, u^* \llcorner \partial S) = 1$. Note that, by Step 1, there exists a point $P \in S$ such that $P \in u^{-1}(\boldsymbol{P})$. This and the fact that $u^{-1}(\boldsymbol{P})$ is a connected set disjoint from ∂S imply that $u^{-1}(\boldsymbol{P}) \subseteq S$. Let now S_m be the partition of $\mathcal{Q}$ in dyadic cube as in Step II, and assume again that $S = S_1$. Again by Step I, $\deg(\boldsymbol{P}, u^* \llcorner S_m)$ is well defined and equal to 0 for every $m \geq 2$ since otherwise $u^{-1}(\boldsymbol{P}) \cap \overline{S}_m \neq \emptyset$, a contradiction. By the additivity of the degree and since u equals the identity on $\partial \mathcal{Q}$ we get

$$1 = \deg(\boldsymbol{P}, u^* \llcorner \partial \mathcal{Q}) = \sum_m \deg(\boldsymbol{P}, u^* \llcorner \partial S_m) = \deg(\boldsymbol{P}, u^* \llcorner \partial S_1),$$

so the proof is concluded. $\qquad \square$

Remark 2.9. *Notice that, thanks to steps I and II of the above proof, if u is a function for which $u^{-1}(\boldsymbol{P})$ is connected for every $\boldsymbol{P} \in \mathcal{Q}$ then, whenever D and E are two Lipschitz sets such that u^* is continuous on ∂D and ∂E, if $D \subseteq E$ then $F_u(D) \subseteq F_u(E)$ by Definition 2.3. As a consequence, under this assumption the definition of $u_{\mathrm{multi}}(P)$ can be simplified as*

$$u_{\mathrm{multi}}(P) = \bigcap_{\mathcal{D}_P} F_u(D),$$

where $\mathcal{D}_P = \cup_{\varepsilon > 0} \mathcal{D}_{P,\varepsilon}$. It is actually possible to show that, if $u \in INV^+$, then for every $\boldsymbol{P} \in \mathcal{Q}$ the set $u^{-1}(\boldsymbol{P})$ is necessarily connected, thus the condition of the above Lemma is actually a characterisation of INV^+, see [3, Proposition 5.14]. In particular, by Corollary 2.7, connectedness of any $u^{-1}(\boldsymbol{P})$ is a necessary condition to be approximable by diffeomorphisms. However, as we will show in Section 5.2, it is not sufficient.

2.2. **Definition of Lebesgue squares and extension results.** In this short section, we recall the definition and some properties of the "Lebesgue squares", which have been already used several times in the last years to get approximation results, see for instance [7, 11, 4, 18]. Moreover, we list three extension results, which will be used in the sequel.

Definition 2.10 (Lebesgue squares). *Let $u \in W^{1,p}(\mathcal{Q})$. We say that a square $S \subseteq \mathcal{Q}$ is a Lebesgue square with matrix M and constant κ if*

$$\frac{1}{|S|} \int_S |Du - M|^p \leq \kappa.$$

Lemma 2.11. *Let $u \in W^{1,p}(\mathcal{Q})$, let $\mathcal{A} \subseteq \mathbb{R}^{2 \times 2}$ be a set of matrixes, and let $\kappa > 0$ be fixed. For every $K \geq 1$ let $\{S_m^K\}$, for $1 \leq m \leq K^2$, be the standard partition of $\mathcal{Q}$ in K^2 squares of side $1/K$, and let us call $\mathrm{Leb}_K(\kappa, \mathcal{A})$ the set of squares of this partition which are Lebesgue square with some matrix $M \in \mathcal{A}$ and constant κ. If K is big enough, depending on u, κ and $\mathcal{A}$, then*

$$\sum_{S_m^K \notin \mathrm{Leb}_K(\kappa, \mathcal{A})} |S_m^K| \leq 2|\{x \in \mathcal{Q} : Du(x) \notin \mathcal{A}\}|.$$

Proof. Let $x \in \mathcal{Q}$ be a Lebesgue point for Du such that $Du(x) \in \mathcal{A}$. By definition of Lebesgue point, there exists $\bar{r} = \bar{r}(x, \kappa)$ such that, whenever $r < \bar{r}$, one has

$$\fint_{B(x,r)} |Du - Du(x)|^p \leq \frac{\kappa}{2} r^2.$$

Assume now that x is contained in some square $\mathcal{S}$ with side ℓ for some $\ell\sqrt{2} < \bar{r}$; then, since $\mathcal{S} \subseteq B(x, \ell\sqrt{2})$, we obtain

$$\frac{1}{|\mathcal{S}|} \int_{\mathcal{S}} |Du - Du(x)|^p \leq \frac{1}{\ell^2} \int_{B(x,\ell\sqrt{2})} |Du - Du(x)|^p \leq \kappa \,,$$

that is, $\mathcal{S}$ is a Lebesgue square with matrix $Du(x) \in \mathcal{A}$ and constant κ. As a consequence, if a square S_m^K of the partition is not contained in $\mathrm{Leb}_K(\kappa, \mathcal{A})$, then the whole square is contained in the set of points x which are either not Lebesgue points for Du, or Lebesgue points for Du with $Du(x) \notin \mathcal{A}$, or Lebesgue points for Du with $Du(x) \in \mathcal{A}$ but $\bar{r}(x, \kappa) \leq \sqrt{2}/K$. Since the set of points for which the first condition holds is negligible, and the set of those for which the third one holds has mass which goes to 0 when $K \to \infty$, the lemma is proved. $\qquad\square$

Lemma 2.12. *Let $\alpha \in W^{1,1}(PQ; \mathbb{R}^2)$ where PQ is a segment, and let β be the linear function such that $\beta(P) = \alpha(P)$ and $\beta(Q) = \alpha(Q)$. Then, for every matrix $M \in \mathbb{R}^{2\times2}$,*

$$\int_{PQ} |\beta' - M\tau|^p \leq \int_{PQ} |\alpha' - M\tau|^p \,,$$

where τ is the unit vector in direction of PQ.

Proof. This is just a trivial application of Jensen inequality. $\qquad\square$

We present now three extension results. The first two have been proved in the case $p = 1$ in [11, Theorem 2.1, Theorem 3.1], and in the case $p > 1$ in [18, Theorem 1.1], [4, Theorem 2.1, Theorem 3.7]. Actually, the second one was proved with a slightly different claim, hence we add a very short proof to show that our claim follows from the original one.

Proposition 2.13. *For every $p \geq 1$ there exists a purely geometrical constant $H_1 = H_1(p) \geq 1$ such that the following holds. For every square $\mathcal{R} \subseteq \mathbb{R}^2$ with side ℓ, and for every injective, piecewise linear function $\varphi : \partial\mathcal{R} \to \mathbb{R}^2$, there exists a finitely piecewise affine homeomorphism $\omega : \mathcal{R} \to \mathbb{R}^2$, coinciding with φ on $\partial\mathcal{R}$, such that*

$$\int_{\mathcal{R}} |D\omega|^p \leq H_1 \ell \int_{\partial\mathcal{R}} |\varphi'|^p \,.$$

Proposition 2.14. *For every $p \geq 1$ there exists a purely geometrical constant $H_2 = H_2(p)$ such that the following holds. For every non-zero 2×2 matrix M with $\det M = 0$, every square $\mathcal{R} \subseteq \mathbb{R}^2$ with side ℓ, and every injective, piecewise linear function $\varphi : \partial\mathcal{R} \to \mathbb{R}^2$, there exists a finitely piecewise affine homeomorphism $\omega : \mathcal{R} \to \mathbb{R}^2$, coinciding with φ on $\partial\mathcal{R}$, such that*

$$\int_{\mathcal{R}} |D\omega - M|^p \leq H_2 \ell \int_{\partial\mathcal{R}} |\varphi' - M\tau|^p \,,$$

where τ is the tangential vector on $\partial\mathcal{R}$.

Proof. In [4, Theorem 3.7], this result was proved (actually, in a quite more general situation) with the additional assumption that

$$\int_{\partial\mathcal{R}} |\varphi' - M\tau|^p < \delta_{\max} \ell |M|^p \,,$$

for some geometrical constant $\delta_{\max}$ only depending on p. However, if this assumption is not satisfied, we can define the extension ω from Proposition 2.13, and we get

$$\int_{\mathcal{R}} |D\omega - M|^p \leq 2^p \int_{\mathcal{R}} |D\omega|^p + 2^p |M|^p \ell^2 \leq 2^p H_1 \ell \int_{\partial\mathcal{R}} |\varphi'|^p + 2^p |M|^p \ell^2$$

$$\leq 4^p H_1 \ell \int_{\partial\mathcal{R}} |\varphi' - M\tau|^p + (4^{p+2} H_1 + 2^p)|M|^p \ell^2$$

$$\leq \left(4^p H_1 + \frac{(4^{p+2} H_1 + 2^p)}{\delta_{\max}} \right) \ell \int_{\partial\mathcal{R}} |\varphi' - M\tau|^p \,,$$

so that the result is still true, up to possibly enlarge the constant H_2. $\square$

Proposition 2.15. *For every $p \geq 1$ and $\sigma < 1$ there exists a constant $H_3 = H_3(\sigma, p)$ such that the following holds. For every square $\mathcal{R} \subseteq \mathbb{R}^2$ with side ℓ, every matrix $M \in \mathbb{R}^{2\times 2}$ with $|M|, \det M \in (\sigma, 1/\sigma)$, and every injective, piecewise linear function $\varphi : \partial\mathcal{R} \to \mathbb{R}^2$, there exists a finitely piecewise affine homeomorphisms $\omega : \mathcal{R} \to \mathbb{R}^2$, coinciding with φ on $\partial\mathcal{R}$, such that*

$$\int_{\mathcal{R}} |D\omega - M|^p \leq H_3 \ell^{2-\frac{1}{p}} \left(\int_{\partial\mathcal{R}} |\varphi' - M\tau|^p \right)^{\frac{1}{p}} , \tag{2.3}$$

where τ is the tangent vector $\partial\mathcal{R}$.

Proof. By a change of variables, we can assume that M is the identity matrix; it is actually not difficult to see how H_3 depends on σ, but we are not interested in the precise dependance. Moreover, it is enough to consider the case when $\mathcal{R} = \mathcal{Q} = [0,1]^2$ is the unit square, and we will do so for simplicity. By the same argument as in Proposition 2.14, we can assume that

$$\delta = \left(\int_{\partial\mathcal{Q}} |\varphi' - \tau|^p \right)^{\frac{1}{p}} ,$$

is as small as we wish, up to increase the constant H_3.

Let us first show that $\|\varphi - Id\|_{L^\infty(\partial\mathcal{Q})} \leq 2\delta$, assuming up to a translation that $\varphi(O) = (O)$, where $O = (0,0)$. In fact, for any $0 \leq t \leq 1$, calling $P = (t,0)$ we have

$$|\varphi(P) - P| \leq \int_{s=0}^{t} |\varphi'(t) - \mathrm{e}_1| \leq \|\varphi' - \tau\|_{L^1([0,1]\times\{0\})} \leq \delta \,,$$

and arguing similarly on the other sides of $\partial\mathcal{Q}$ we obtain the L^∞ estimate.

Let now $C = C(p, \sigma)$ be a large, purely geometric constant, and let $N \approx 1/(C\delta)$ be a large integer. First of all, we set ω as the identity on the square $[1/N, 1-1/N]^2$, and of course $\omega = \varphi$ on $\partial\mathcal{Q}$. Let now $P_i = \boldsymbol{P}_i = (i/N, 1/N)$ for some $1 \leq i \leq N-1$, and let us call $\boldsymbol{P}_i'$ a point on $\varphi([0,1] \times \{0\})$ minimizing the distance with $\boldsymbol{P}_i$. Notice that, since $\|\varphi - \mathrm{Id}\|_{L^\infty} < 2\delta$ and C is a large constant, then the distance between $\boldsymbol{P}_i'$ and $(i/N, 0)$ is much smaller than $1/N$, and in turn also the distance between $P_i' = \varphi^{-1}(\boldsymbol{P}_i')$ and $(i/N, 0)$. We define in the analogous way the points $Q_i = \boldsymbol{Q}_i = (i/N, 1-1/N)$ with $\boldsymbol{Q}_i' \in \varphi([0,1] \times \{1\})$ and $Q_i' = \varphi^{-1}(\boldsymbol{Q}_i)$ near $(i/N, 1)$, the points $R_i = \boldsymbol{R}_i = (1/N, i/N)$ with the points $\boldsymbol{R}_i'$ and R_i' near $(0, i/N)$ and the points $S_i = \boldsymbol{S}_i = (1 - 1/N, i/N)$ with $\boldsymbol{S}_i'$ and S_i' near $(1, 1/N)$. Note that for $i \neq j$ we always have that $|\boldsymbol{P}_j - \boldsymbol{P}_j| = 1/N(1 + O(\delta))$.

Let us consider each segment $P_i P_i'$: since ω is already defined on both points, we extend ω on the whole segment $P_i P_i'$ as a linear function; we do the same on all the segments $Q_i Q_i'$, $R_i R_i'$ and $S_i S_i'$. Notice that by construction ω is injective, where it is already defined.

Observe now that the region $\mathcal{Q} \setminus [1/N, 1 - 1/N]^2$ is subdivided in $4N - 4$ quadrilaterals, on the boundary of each of which ω is already defined; moreover, each quadrilateral is the 2-bi-Lipschitz copy of a square of side $1/N$: as a consequence, by Proposition 2.13 we can extend ω on the interior of any such quadrilateral, call it $\mathcal{R}$, so that

$$\int_{\mathcal{R}} |D\omega|^p \leq \frac{4H_1}{N} \int_{\partial \mathcal{R}} |D_\tau \omega|^p \,.$$

Moreover, of course

$$\int_{\partial \mathcal{R} \cap \partial \mathcal{Q}} |D_\tau \omega|^p = \int_{\partial \mathcal{R} \cap \partial \mathcal{Q}} |\varphi'|^p \,.$$

On the other hand, on each of the (either two or three) sides of $\partial \mathcal{R}$ which are not contained in $\partial \mathcal{Q}$, the function ω is a linear function which maps a segment of length almost precisely $1/N$ onto a segment of length almost precisely $1/N$; therefore, $\int_{\partial \mathcal{R} \setminus \partial \mathcal{Q}} |D_\tau \omega|^p \leq \frac{4}{N}$. Summarizing,

$$\int_{\mathcal{R}} |D\omega|^p \leq \frac{4H_1}{N} \left(\int_{\partial \mathcal{R} \cap \partial \mathcal{Q}} |\varphi'|^p + \frac{4}{N} \right) ,$$

which adding on the $4N - 4$ quadrilaterals gives

$$\int_{\mathcal{Q} \setminus [1/N, 1-1/N]^2} |D\omega|^p \leq \frac{4H_1}{N} \int_{\partial \mathcal{Q}} |\varphi'|^p + \frac{64H_1}{N} \leq \frac{2^{p+2} H_1}{N} \left(\delta^p + 4 \right) + \frac{64 H_1}{N} \leq \frac{2^{p+6} H_1}{N} \,.$$

As a consequence, we can calculate

$$\int_{\mathcal{Q}} |D\omega - M|^p = \int_{Q \setminus [1/N, 1-1/N]^2} |D\omega - M|^p \leq 2^p \int_{Q \setminus [1/N, 1-1/N]^2} |D\omega|^p + |M|^p$$

$$\leq 2^p \left(\frac{2^{p+6} H_1}{N} + \frac{4}{N\sigma^p} \right) \leq \frac{2^{2p+8} H_1 C}{\sigma^p} \, \delta =: H_3(p, \sigma) \, \delta \,,$$

which is (2.3) for the case $M = \mathrm{Id}$ and $\ell = 1$ that we are considering. $\qquad\square$

It is easy to guess that a stronger version of last extension result should actually be true, namely, without the p-th root and with ℓ in place of $\ell^{2 - \frac{1}{p}}$ in (2.3). However, this weaker result is enough for our needs.

3. Proof of Theorem A and B

In this Section we will prove Theorems A and B. Before going in the details of the proof we will make a couple of simplifying assumption that will be in force for all the Section:

- We will always work with the representatives defined in Definition 1.2.
- We will always assume that the functions coincide with the identity in a neighbourhood of the boundary (we can do this without loss of generality, up to extend all the maps as the identity outside $\mathcal{Q}$).

3.1. **The two grids.** This short section is devoted to define suitable grids $\mathscr{G}$ and $\widetilde{\mathscr{G}}$ on the squares $\mathcal{Q}$ and $\boldsymbol{\mathcal{Q}}$.

Definition 3.1 (Admissible curve). *Let X be either $[0,1]$ or $\mathbb{S}^1$, and let $\gamma : X \to \mathcal{Q}$ be a piecewise C^1, injective curve. We say that γ is an admissible curve if $\gamma(t)$ is a Lebesgue point for Du for $\mathscr{H}^1$-a.e. $t \in X$, and $u \circ \gamma$ belongs to $W^{1,p}(X)$.*

Definition 3.2. *Let $K \geq 1$ be an integer. Then, we call $\mathcal{R}_{i,j} = [i/K, (i+1)/K] \times [j/K, (j+1)/K] \subseteq \mathcal{Q}$, for all $0 \leq i, j < K$, the elements of the standard partition of $\mathcal{Q}$ in K^2 squares. Moreover, we call $\mathscr{G} = \mathscr{G}(K) := \bigcup_{0 \leq i,j < K} \mathcal{R}_{i,j} \subseteq \mathcal{Q}$.*

Lemma 3.3. *Let $u \in W^{1,p}(\mathcal{Q}; \boldsymbol{\mathcal{Q}})$, with $u = \mathrm{Id}$ in a small neighborhood of $\partial \mathcal{Q}$, and let $\delta \ll 1$. Then, for every integer $K > 1/\delta$ and any choice of matrices $M_{i,j} \in \mathbb{R}^{2 \times 2}$, with $0 \leq i, j < K$, there exists a finitely piecewise affine, $(1+4\delta)$-biLipschitz homeomorphism $\Phi : \mathcal{Q} \to \mathcal{Q}$ coinciding with the identity on $\partial \mathcal{Q}$, such that $\|\Phi - \mathrm{Id}\|_{L^\infty(\mathcal{Q})} < 2\delta/K$ and that, for every square $\mathcal{R}_{i,j} \in \mathscr{G}(K)$, the curve $\partial(\Phi(\mathcal{R}_{i,j}))$ is an admissible curve and one has*

$$\int_{\Phi(\partial \mathcal{R}_{i,j})} |Du - M_{i,j}|^p \, d\mathscr{H}^1 \leq \frac{204K}{\delta} \int_{\mathcal{R}_{i,j}^+} |Du - M_{i,j}|^p \, d\mathscr{H}^2 . \tag{3.1}$$

Here $\mathcal{R}_{i,j}^+$ the union of the squares $\mathcal{R}_{i',j'}$ with $\max\{|i - i'|, |j - j'|\} \leq 1$. In addition, for every $0 \leq i, j < K$, Φ is affine on each of the two triangles in which the square $\mathcal{R}_{i,j}$ is divided by the northeast-southwest diagonal. Finally, the points $\Phi(i/K, j/K)$ are all continuity points for u.

Proof. The proof is a simple variation of the argument of Lemmas 4.9 and 4.13 in [11], see also step II in the proof of [4, Theorem 4.1]. Let us fix $\delta > 0$, so small that $u = \mathrm{Id}$ in a 2δ-neighborhood of $\partial \mathcal{Q}$, and an integer $K > 1/\delta$. For every choice of $1 \leq l, m \leq K - 1$, we consider the following short segment near $A = (l/K, m/K)$, with a slope of $45°$,

$$I_A = \left\{ \left(\frac{l+t}{K}, \frac{m+t+\eta}{K} \right), \ -\delta < t < \delta \right\} .$$

The constant $\eta \ll \delta$ is chosen in such a way that $\mathscr{H}^1$-a.e. point of each I_A is a continuity point of u.

Let us now consider two adjacent points $A = (l/K, m/K)$ and $B = (l'/K, m'/K)$, where by "adjacent" we mean that $|l - l'| + |m - m'| = 1$. Let us call $\mathcal{R}^+(A, B)$ the union of the six squares $\mathcal{R}_{i,j}$ having either A or B or both as vertices, and let M_1 and M_2 be the matrices corresponding to the two squares having AB as a side. Defining then the set

$$\Gamma(A, B) = \left\{ (x, y) \in I_A \times I_B : \exists n \in \{1, 2\}, \int_{xy} |Du - M_n|^p > \frac{51K}{\delta} \int_{\mathcal{R}^+(A,B)} |Du - M_n|^p \right\},$$

an immediate change of variable argument ensures that

$$\mathscr{H}^2\big(\Gamma(A, B)\big) < \frac{1}{25} \mathscr{H}^1(I_A) \times \mathscr{H}^1(I_B) .$$

In particular

$$\mathscr{H}^1\left(\left\{ x \in I_A : \mathscr{H}^1\left(\left\{ y \in I_B : (x, y) \in \Gamma(A, B) \right\} \right) > \frac{\mathscr{H}^1(I_B)}{5} \right\} \right) < \frac{\mathscr{H}^1(I_A)}{5} .$$

An easy recursion argument, identical to the one done in Lemma 4.13 of [11], ensures that it is possible to pick a point $V_A \in I_A$ for each $A = (l/K, m/K)$ in such a way that, whenever AB is a side of some square $\mathcal{R}_{i,j}$, the pair (V_A, V_B) does not belong to $\Gamma(A, B)$, almost every point in the segment $V_A V_B$ is a Lebesgue point for Du, and V_A and V_B are continuity points of u. We let then $\Phi : \mathcal{Q} \to \mathcal{Q}$ be the finitely piecewise affine bijection which sends each vertex A in the corresponding point V_A, and which is affine on each triangle obtained subdividing any of the squares with the northeast-southwest diagonal. The facts that $\|\Phi - \mathrm{Id}\|_{L^\infty} < 2\delta/K$, that Φ is $(1+4\delta)$-biLipschitz, and that the points $\Phi(i/K, j/K)$ are all continuity points for u, are true by construction. Moreover, since Du is an L^p function in the segment $V_A V_B$ for each pair such that AB is the side of some square, so in particular each segment $V_A V_B = \Phi(AB)$ is an admissible curve, and then so is each $\partial(\Phi(\mathcal{R}_{i,j}))$.

To conclude, we have then only to establish (3.1). To do so, let us fix some square $\mathcal{R}_{i,j}$, and let AB be one of its sides. Since $M_{i,j}$ is one of the two matrices M_1 and M_2 corresponding to the side AB, and $\mathcal{R}^+(A, B) \subseteq \mathcal{R}^+_{i,j}$, the fact that $(V_A, V_B) \notin \Gamma(A, B)$ implies

$$\int_{V_A V_B} |Du - M_{i,j}|^p \leq \frac{51K}{\delta} \int_{\mathcal{R}^+(A,B)} |Du - M_{i,j}|^p \leq \frac{51K}{\delta} \int_{\mathcal{R}^+_{i,j}} |Du - M_{i,j}|^p.$$

Since $\Phi(\partial \mathcal{R}_{i,j})$ is the union of four sides of the form $V_A V_B$, the last estimate yields (3.1).   □

Relying on this lemma, we give the following definition.

Definition 3.4 (Good starting grid)**.** *Let K be any large integer. The set $\mathscr{G} = \mathscr{G}(K)$ is called a* good starting grid *if every vertex of each square $\mathcal{R}_{i,j}$ is a continuity point for u, and the boundary of any square $\mathcal{R}_{i,j}$ is an admissible curve in the sense of Definition 3.1.*

Concerning the arrival grid, what we need is the following.

Definition 3.5 (Good arrival grid)**.** *Let $\mathscr{G} \subseteq \mathcal{Q}$ be a good starting grid. Let $\eta \ll 1/K$ and let $0 = x_0 < x_1 < x_2 < \cdots < x_N = 1$ and $0 = y_0 < y_1 < y_2 < \cdots < y_M = 1$ be given coordinates, such that $\eta < x_{n+1} - x_n < 2\eta$ and $\eta < y_{m+1} - y_m < 2\eta$ for every $0 \leq n < N$ and every $0 \leq m < M$. The set $\widetilde{\mathscr{G}} \subseteq \boldsymbol{Q}$ given by*

$$\widetilde{\mathscr{G}} = \bigcup_{0 \leq n \leq N} x_n \times [0,1] \cup \bigcup_{0 \leq m \leq M} [0,1] \times y_m \tag{3.2}$$

is called a good arrival grid *associated with $\mathscr{G}$ and with side-length η if*

(1) *$u^{-1}(\widetilde{\mathscr{G}}) \cap \mathscr{G}$ is a finite subset of $\mathscr{G}$;*

(2) *for any $P \in u^{-1}(\widetilde{\mathscr{G}}) \cap \mathscr{G} \setminus \partial\mathcal{Q}$, $u(P)$ is not a vertex of some rectangle of $\widetilde{\mathscr{G}}$, P is not a vertex of some square of $\mathscr{G}$ and is a Lebesgue point for Du and, calling τ the direction of the open side of $\mathscr{G}$ containing P, the tangential derivative $D_\tau u(P)$ has a non-zero component in the direction orthogonal to the side of $\widetilde{\mathscr{G}}$ containing $u(P)$;*

(3) *for every vertex $V \notin \partial\mathcal{Q}$ of some square $\mathcal{R}_{i,j} \subseteq \mathscr{G}$, calling $u(V) = \boldsymbol{V} = (\boldsymbol{V}_1, \boldsymbol{V}_2)$ and letting n and m be such that $x_n < \boldsymbol{V}_1 < x_{n+1}$ and $y_m < \boldsymbol{V}_2 < y_{m+1}$, one has*

$$\min\left\{ |x_n - \boldsymbol{V}_1|, |x_{n+1} - \boldsymbol{V}_1|, |y_m - \boldsymbol{V}_2|, |y_{m+1} - \boldsymbol{V}_2| \right\} > \frac{\eta}{2}.$$

The existence of good arrival grids is ensured by the following result.

Lemma 3.6. *Let $\mathscr{G}$ be a good starting grid, and let $\bar{\eta} \ll 1/K$ be fixed. Then, there exists a good arrival grid $\widetilde{\mathscr{G}}$ associated with $\mathscr{G}$ with side-length $\eta < \bar{\eta}$.*

Proof. Let $A \subseteq \mathcal{Q}$ be the set of the vertices of $\mathscr{G}$, and let $\boldsymbol{A}^1 = \pi_1\big(u(A)\big)$ and $\boldsymbol{A}^2 = \pi_2\big(u(A)\big)$, where π_1 and π_2 are, respectively, the projections on the first and second coordinate. Recall that u is uniquely defined at any point of A, hence $\boldsymbol{A}^1$ (resp., $\boldsymbol{A}^2$) is a finite set of abscissae (resp., ordinates). Let now Γ be a horizontal or vertical segment, contained in $\mathscr{G}$ and with endpoints in $\partial\mathcal{Q}$. Since by construction Γ is an admissible curve, then the set B_Γ of the points of Γ which are not Lebesgue points for Du is $\mathscr{H}^1$-negligible, thus so are the sets $\boldsymbol{B}_\Gamma^1 = \pi_1\big(u(B_\Gamma)\big)$ and $\boldsymbol{B}_\Gamma^2 = \pi_2\big(u(B_\Gamma)\big)$. Let us now call $C_\Gamma^1 \subseteq \Gamma$ (resp., C_Γ^2) the set of points which are Lebesgue points for $D_\tau u$, but for which the first (resp., second) component of $D_\tau u$ is zero. Let $\delta > 0$ be any small positive number. By definition, for any $P \in C_\Gamma^1$ there exists $\varepsilon = \varepsilon(P, \delta)$ such that, calling $I(P, \varepsilon) = \{Q \in \Gamma : |Q - P| < \varepsilon\}$, it is

$$\mathscr{H}^1\Big(\pi_1\big(u(I(P, \varepsilon))\big)\Big) < \delta\varepsilon\,.$$

With a standard covering argument, we can cover C_Γ^1 with countably many intervals $I(P_i, \varepsilon_i)$ satisfying the above estimate, in such a way that the intervals $I(P_i, \varepsilon_i/4)$ are disjoint. As a consequence, calling $\boldsymbol{C}_\Gamma^1 = \pi_1\big(u(C_\Gamma^1)\big) \subseteq [0, 1]$, we have

$$\mathscr{H}^1\big(\boldsymbol{C}_\Gamma^1\big) \leq \sum_{i \in \mathbb{N}} \mathscr{H}^1\Big(\pi_1\big(u(I(P_i, \varepsilon_i))\big)\Big) \leq \sum_{i \in \mathbb{N}} \delta\varepsilon_i \leq 4\delta\mathscr{H}^1(\Gamma)\,.$$

Since δ was arbitrary we get $\mathscr{H}^1(\boldsymbol{C}_\Gamma^1) = 0$. The very same argument ensures also $\mathscr{H}^1(\boldsymbol{C}_\Gamma^2) = 0$. Calling now $\boldsymbol{B}^1$ and $\boldsymbol{C}^1$ (resp., $\boldsymbol{B}^2$ and $\boldsymbol{C}^2$) the union of all the sets $\boldsymbol{B}_\Gamma^1$ and $\boldsymbol{C}_\Gamma^1$ (resp., $\boldsymbol{B}_\Gamma^2$ and $\boldsymbol{C}_\Gamma^2$) for all the horizontal and vertical segments Γ contained in $\mathscr{G}$, we have $\mathscr{H}^1(\boldsymbol{A}^1 \cup \boldsymbol{B}^1 \cup \boldsymbol{C}^1) = \mathscr{H}^1(\boldsymbol{A}^2 \cup \boldsymbol{B}^2 \cup \boldsymbol{C}^2) = 0$.

Let us now select $s \notin \boldsymbol{A}^1 \cup \boldsymbol{B}^1 \cup \boldsymbol{C}^1$, let Γ be a segment contained in $\mathscr{G}$, and let P be a point in $u^{-1}\big(\{s\} \times [0, 1]\big) \cap \Gamma$. Since $s \notin \boldsymbol{A}^1$ we have that P is not a vertex of $\mathscr{G}$, and since $s \notin \boldsymbol{B}^1 \cup \boldsymbol{C}^1$ we have that P is a Lebesgue point for Du, and $Du(P)$ has a non-zero horizontal component. In particular there are no other points of $u^{-1}\big(\{s\} \times [0, 1]\big)$ in a small neighborhood of P in Γ. As a consequence, the set $u^{-1}\big(\{s\} \times [0, 1]\big) \cap \Gamma$ is finite. Repeating the same argument for every segment in $\mathscr{G}$, and arguing in the very same way for $t \notin \boldsymbol{A}^2 \cup \boldsymbol{B}^2 \cup \boldsymbol{C}^2$, we obtain that for every $s \notin \boldsymbol{A}^1 \cup \boldsymbol{B}^1 \cup \boldsymbol{C}^1$ (resp., $t \notin \boldsymbol{A}^2 \cup \boldsymbol{B}^2 \cup \boldsymbol{C}^2$), the set $u^{-1}\big(\{s\} \times [0, 1]\big) \cap \mathscr{G}$ (resp., $u^{-1}\big([0, 1] \times \{t\}\big) \cap \mathscr{G}$) is a finite set, which does not contain any vertex of $\mathscr{G}$, and containing only Lebesgue points for Du, at which Du has a non-zero horizontal (resp., vertical) component.

Let now $\eta < \bar{\eta}$ be any number much smaller than the difference between any two different elements of $\boldsymbol{A}^1$ or $\boldsymbol{A}^2$. It is then possible to select numbers $0 = x_0 < x_1 < x_2 < \cdots < x_N = 1$, none of which (except x_0 and x_N) belonging to $\boldsymbol{A}^1 \cup \boldsymbol{B}^1 \cup \boldsymbol{C}^1$, in such a way that $\eta < x_{n+1} - x_n < 2\eta$ for every $0 \leq n < N$, and also that $\min\big\{|x_n - \boldsymbol{V}_1|, |x_{n+1} - \boldsymbol{V}_1|\big\} > \eta/2$ for every vertex $V \in \mathscr{G} \setminus \partial\mathcal{Q}$, being $\boldsymbol{V} = u(V)$, and being n such that $x_n < \boldsymbol{V}_1 < x_{n+1}$.

By construction, the set $\mathcal{F} = u^{-1}\big(\{x_n,\, 1 \le n < N\} \times [0,1]\big) \cap \mathscr{G}$ is a finite set of points, and so it is the set $\boldsymbol{D}^2 = \pi_2\big(u(\mathcal{F})\big)$. As a consequence, it is possible to choose numbers $0 = y_0 < y_1 < y_2 < \cdots < y_M = 1$ not belonging to $\boldsymbol{A}^2 \cup \boldsymbol{B}^2 \cup \boldsymbol{C}^2 \cup \boldsymbol{D}^2$ (except y_0 and y_M), and so that $\eta < y_{m+1} - y_m < 2\eta$ for every $0 \le m < M$, and that $\min\big\{|y_m - \boldsymbol{V}_2|,\, |y_{m+1} - \boldsymbol{V}_2|\big\} > \eta/2$ for every vertex $V \in \mathscr{G} \backslash \partial \mathcal{Q}$, again calling $\boldsymbol{V} = u(V)$ and letting m be such that $y_m < \boldsymbol{V}_2 < y_{m+1}$. It is then clear that the set $\widetilde{\mathscr{G}}$ given by (3.2) is a good arrival grid with side-length $\eta' < \eta$. $\qquad\square$

3.2. **The piecewise linear function on the grids.** This section is devoted to define the piecewise linear modification of a non-crossing function u on a good starting grid. We will show the following key result. It is here that the non-crossing condition is used, see Step II in the proof.

We start with the following useful definition:

Definition 3.7 (generalized segment). *Let $\boldsymbol{\mathcal{R}}$ be a closed rectangle of the grid $\widetilde{\mathscr{G}}$, and let $\boldsymbol{A}$, $\boldsymbol{B}$ be any two points in $\partial \boldsymbol{\mathcal{R}}$. Given ξ small, the generalized segment $[\boldsymbol{AB}]$ between $\boldsymbol{A}$ and $\boldsymbol{B}$ in $\boldsymbol{\mathcal{R}}$ is defined as the segment $\boldsymbol{AB}$ if the two points are not on a same side of $\partial \boldsymbol{\mathcal{R}}$; otherwise, $[\boldsymbol{AB}]$ is the union of the segments $\boldsymbol{AM}$ and $\boldsymbol{MB}$, where $\boldsymbol{M}$ is the point in $\boldsymbol{\mathcal{R}}$ whose distance from the side containing $\boldsymbol{A}$ and $\boldsymbol{B}$ is ξ times the length of $\boldsymbol{AB}$, and whose projection there is the mid-point of the segment $\boldsymbol{AB}$.*

Notice that the generalized segment $[\boldsymbol{AB}]$ is entirely contained in the interior of $\boldsymbol{\mathcal{R}}$ except for $\boldsymbol{A}$ and $\boldsymbol{B}$. Moreover, taken four distinct ordered points $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$ and $\boldsymbol{D}$ on a same side of $\partial \boldsymbol{\mathcal{R}}$, the generalized segments $[\boldsymbol{AB}]$ and $[\boldsymbol{CD}]$ have empty intersection, as well as $[\boldsymbol{AD}]$ and $[\boldsymbol{BC}]$, while $[\boldsymbol{AC}]$ and $[\boldsymbol{BD}]$ have exactly an intersection point.

Proposition 3.8. *Let $v \in \mathcal{NC}_{\mathrm{Id}}(\mathcal{Q}) \cap W^{1,p}$, assume that $\mathscr{G} = \mathscr{G}(K)$ is a good starting grid for v, let $M_{i,j}$ for $0 \le i,\, j < K$ be given 2×2 matrices, and let $\sigma \ll 1$ be fixed. Then, there exists an injective, piecewise linear function $\varphi : \mathscr{G} \to \mathcal{Q}$, coinciding with the identity on $\partial \mathcal{Q}$, such that, for every $0 \le i,\, j < K$, one has*

$$\int_{\partial \mathcal{R}_{i,j}} |\varphi' - M_{i,j}\tau|^p \le \frac{\sigma}{K} + 2\int_{\partial \mathcal{R}_{i,j}} |Dv - M_{i,j}|^p\,, \tag{3.3}$$

where τ is the unit tangent vector to $\partial \mathcal{R}_{i,j}$.

Proof. We divide the proof in three steps.

Step I. Preliminaries.

First of all, we fix a length $\xi \ll 1/K$ with the property that, for every segment $AB \subseteq \mathscr{G}$ with length $\mathscr{H}^1(AB) \le \xi$, one has

$$\int_{AB} |Dv|^p < \frac{(\sqrt[p]{2} - 1)^p}{16 \cdot 7^p K}\,\sigma\,, \tag{3.4}$$

which is clearly possible since $v \in W^{1,p}(\mathscr{G})$. Then, by Lemma 3.6 we can take a good arrival grid $\widetilde{\mathscr{G}}$ associated with $\mathscr{G}$ and with side-length equal to some $\eta \ll \xi \ll 1/K$ satisfying

$$\frac{(6\eta)^p}{\xi^{p-1}} < \frac{\sigma}{16K}\,(\sqrt[p]{2} - 1)^p\,. \tag{3.5}$$

Let us now call P_1, P_2, ..., P_N the points of $v^{-1}(\widetilde{\mathscr{G}}) \cap \mathscr{G}$, and $\boldsymbol{P}_j = v(P_j)$ for every $1 \leq j \leq N$. Notice that the points $\boldsymbol{P}_j$ are all on $\widetilde{\mathscr{G}}$, but they are not necessarily all distinct. Let us also notice that, by definition of good arrival grid, each $D_\tau v(P_i)$ makes a non-zero angle θ_i with respect to the horizontal (resp., vertical) direction if $\boldsymbol{P}_i$ belongs to a horizontal (resp., vertical) side of $\widetilde{\mathscr{G}}$; let then C be any constant bigger than $\max_{1 \leq i \leq N} 1/\sin\theta_i$.

Let now ε be a constant such that $C\varepsilon$ is much smaller than η, than the distance between any two $\boldsymbol{P}_i$ and $\boldsymbol{P}_j$ which do not coincide, and than the distance between any $\boldsymbol{P}_i$ and the closest vertex of $\widetilde{\mathscr{G}}$. By definition of good starting and arrival grid, up to decrease ε if necessary we have that, for any $P \in \mathscr{G}$, the distance between $v(P)$ and $\widetilde{\mathscr{G}}$ can be smaller than ε only if P is close to some of the P_i, and in this case $|v(P) - \boldsymbol{P}_i| < C\varepsilon$.

Since any P_i is in the interior of a side of $\mathscr{G}$, again up to decrease ε if necessary we can find points $P_i^\pm$ on $\mathscr{G}$ for any $1 \leq i \leq N$, in such a way that the following holds. First of all, P_i^+ and P_i^- are on the same side of $\mathscr{G}$ to which P_i belongs, P_i is in the segment between them, and these segments, varying $1 \leq i \leq N$, are all disjoint. Moreover, the curve $v(P_i^- P_i^+)$ lies within a distance $C\varepsilon$ from $\boldsymbol{P}_i$, in particular the ε-neighborhood of this curve is contained in the union of the two squares of $\widetilde{\mathscr{G}}$ having $\boldsymbol{P}_i$ on the boundary, and the points $v(P_i^-)$ and $v(P_i^+)$ do not belong to the same square. Finally, for any $P \in \mathscr{G}$ one has

$$P \notin \cup_{i=1}^N P_i^- P_i^+ \implies \operatorname{dist}(v(P), \widetilde{\mathscr{G}}) > 2\varepsilon \,,$$

$$\operatorname{dist}(v(P), \widetilde{\mathscr{G}}) < \varepsilon \implies \exists! \, 1 \leq i \leq N \text{ s.t. } P \in P_i^- P_i^+, \text{ and } |v(P) - \boldsymbol{P}_i| < C \operatorname{dist}(v(P), \widetilde{\mathscr{G}}) \,.$$

Step II. Definition of the function $\varphi : \mathscr{G} \to \boldsymbol{\mathcal{Q}}$.
For every $1 \leq i \leq K - 1$, let us now call γ_i the horizontal segment connecting $(0, i/K)$ with $(1, i/K)$, and γ_{K-1+i} the vertical segment connecting $(i/K, 0)$ with $(i/K, 1)$. By construction, these curves satisfy the assumption of Definition 1.2. As a consequence, there exists an injective function $\psi_1 : \mathscr{G} \to \boldsymbol{\mathcal{Q}}$, coinciding with the identity on $\mathscr{G} \cap \partial \boldsymbol{\mathcal{Q}}$, such that $\|v - \psi_1\|_{L^\infty(\mathscr{G})} < \varepsilon/C$.

By construction, $\psi_1(P)$ has distance larger than ε from $\widetilde{\mathscr{G}}$ for any $P \notin \cup_{i=1}^N P_i^- P_i^+$. Moreover, for any $1 \leq i \leq N$, the whole curve $\psi_1(P_i^- P_i^+)$ is contained in the union of the two squares containing $\boldsymbol{P}_i$ in the boundary, and $\psi_1(P_i^+)$ and $\psi_1(P_i^-)$ belong to the two different squares. Let us then call Q_i^- and Q_i^+ the first and the last point of the interval $P_i^- P_i^+$ such that $|\psi_1(Q_i^\pm) - \boldsymbol{P}_i| = \varepsilon$, which exist since $|\psi_1(P_i) - \boldsymbol{P}_i| < \varepsilon/C$ while $|\psi_1(P_i^\pm) - \boldsymbol{P}_i| > \varepsilon$. Let us then call $\psi_2 : \mathscr{G} \to \boldsymbol{\mathcal{Q}}$ the function which coincides with ψ_1 outside of the intervals $Q_i^- Q_i^+$, and which is linear on each segment $Q_i^- Q_i^+$: a standard geometrical argument ensures that also ψ_2 is injective, and moreover by construction $\|\psi_2 - v\|_{L^\infty(\mathscr{G})} \leq (C+3)\varepsilon$, and the segment connecting $\psi_2(Q_i^-)$ and $\psi_2(Q_i^+)$ intersects $\widetilde{\mathscr{G}}$ in a single point $\boldsymbol{Q}_i$, having distance at most ε from $\boldsymbol{P}_i$.

Let us now consider any $1 \leq i < j \leq N$: we will say that $P_i P_j$ is a "principal segment" if it lies on $\mathscr{G}$ and it does not contain other points P_l. Notice that $\mathscr{G}$ is a finite union of principal segments, and the intresection between any two such segments can be only empty, or a common endpoint, or a common internal point, in such a case the intersection must be a vertex $V_{ab} = (a/K, b/K)$ for some $1 \leq a, b \leq K - 1$.

We are now ready to define the function $\varphi : \mathscr{G} \to \boldsymbol{\mathcal{Q}}$. To this end we start noticing that if in the definition of generalized segment, Defintion 3.7, ξ is taken sufficiently small, then the following holds: whenever P_iP_j and $P_i'P_j'$ are two principal segments, and the curves $\psi_2(P_iP_j)$ and $\psi_2(P_i'P_j')$ are on a same rectangle $\boldsymbol{\mathcal{R}}$ of $\widetilde{\mathscr{G}}$, then the two generalized segments $[\boldsymbol{Q}_i\boldsymbol{Q}_j]$ and $[\boldsymbol{Q}_i'\boldsymbol{Q}_j']$ have a non-empty intersection if and only if the points $\boldsymbol{Q}_i$ and $\boldsymbol{Q}_j$ are on the two different parts in which $\partial\boldsymbol{\mathcal{R}}$ is divided by $\boldsymbol{Q}_i'$ and $\boldsymbol{Q}_j'$.

For every principal segment $P_iP_j \subseteq \mathscr{G}$ we decide that $\varphi(P_iP_j)$ has to be $[\boldsymbol{Q}_i\boldsymbol{Q}_j]$, with an injective parametrization still to be precised. Notice that, by construction, the function φ will be injective regardless of the parametrization on each principal segment P_iP_j. Notice also that, for any vertex $V_{ab} = (a/K, b/K)$ of the grid $\mathscr{G}$, the point $\varphi(V_{ab})$ is already defined. Indeed, the point V_{ab} is contained in the interior of exactly two principal segments P_iP_j and $P_i'P_j'$, one horizontal and one vertical, and by construction the generalized segments $[\boldsymbol{Q}_i\boldsymbol{Q}_j]$ and $[\boldsymbol{Q}_i'\boldsymbol{Q}_j']$ have exactly one intersection point: this point must obviously be $\varphi(V_{ab})$. The definition of φ is then easy: we let φ be linear on each segment of $\mathscr{G}$ whose both endpoints, and no interior point, are contained in $\{P_i\}_{1\leq i\leq N} \cup \{V_{ab}\}_{1\leq a,\, b\leq K-1}$.

Step III. The validity of (3.3).

To conclude, we only have to check the validity of the estimate (3.3). Let $\mathcal{R} = \mathcal{R}_{i,j}$ be a given square of the grid $\mathscr{G}$, and let $M = M_{i,j}$ be the corresponding given matrix. Notice that $\partial\mathcal{R}$ is the union of finitely many segments on which φ is linear, which are all principal segments except those having at least one vertex as endpoint. Let us consider separately the possible cases. First of all, let $P_iP_j \subseteq \partial\mathcal{R}$ be a principal segment. We claim that

$$\int_{P_iP_j} |\varphi' - M\tau|^p \leq 2 \int_{P_iP_j} |Dv - M|^p. \tag{3.6}$$

Indeed, since φ is linear on P_iP_j, this inequality would be given by Lemma 2.12, with constant 1 instead of 2, if $\boldsymbol{Q}_i = v(P_i)$ and $\boldsymbol{Q}_j = v(P_j)$. But then, since $|\boldsymbol{Q}_i - v(P_i)| < \varepsilon$ and $|\boldsymbol{Q}_j - v(P_j)| < \varepsilon$, the inequality with constant 2 follows as soon as ε is small enough: in fact, keep in mind that the principal segments are finitely many and the constant ε depends on them, and moreover by construction $\int_{P_iP_j} |Dv - M|^p > 0$ even if $v(P_i) = v(P_j)$.

Let now $AB \subseteq \partial\mathcal{R}$ be another of the segments on which φ is linear, with the property that at least one between A and B is a vertex of $\mathcal{R}$. Moreover, let us call $\beta : AB \to \boldsymbol{\mathcal{Q}}$ the linear function such that $\beta(A) = v(A)$ and $\beta(B) = v(B)$. Since for every two positive numbers a, b one always has

$$(a+b)^p \leq 2a^p + \frac{2}{(\sqrt[p]{2}-1)^p}\, b^p,$$

keeping in mind again Lemma 2.12 we have

$$\begin{aligned}
\int_{AB} |\varphi' - M\tau|^p &\leq 2 \int_{AB} |\beta' - M\tau|^p + \frac{2}{(\sqrt[p]{2}-1)^p} \int_{AB} |\varphi' - \beta'|^p \\
&\leq 2 \int_{AB} |Dv - M|^p + \frac{2}{(\sqrt[p]{2}-1)^p} \int_{AB} |\varphi' - \beta'|^p.
\end{aligned} \tag{3.7}$$

We have now to distinguish two possible sub-cases, namely, whether $\mathscr{H}^1(AB)$ is larger or smaller than ξ.

Suppose first that $\mathscr{H}^1(AB) \leq \xi$. Since $\xi < 1/K$, this ensures that exactly one between A and B is a vertex of $\mathscr{G}$, and the other one is a point of the form P_i for some $1 \leq i \leq N$. By definition of good arrival grid, this means that $|\beta(A) - \beta(B)| > \eta/2$; on the other hand, of course $|\varphi(A) - \varphi(B)| < 2\sqrt{2}\eta$, and since both φ and β are linear in AB this gives $|\varphi'| < 6|\beta'|$, so again by Lemma 2.12

$$\int_{AB} |\varphi' - \beta'|^p < 7^p \int_{AB} |\beta'|^p \leq 7^p \int_{AB} |Dv|^p \,,$$

which inserted in (3.7) and keeping in mind (3.4) gives

$$\int_{AB} |\varphi' - M\tau|^p \leq 2 \int_{AB} |Dv - M|^p + \frac{\sigma}{8K} \,. \tag{3.8}$$

Let us now instead assume that $\mathscr{H}^1(AB) > \xi$. In this case, it might be possible that both A and B are vertices of $\mathscr{G}$ (so AB is a whole side of $\mathcal{R}$), and that v is constant on the whole segment AB, hence we cannot estimate $|\varphi'| < 6|\beta'|$ as before. Nevertheless, $\varphi - \beta$ is a linear function on AB, and since the four points $\varphi(A)$, $\varphi(B)$, $\beta(A)$ and $\beta(B)$ lie in a same rectangle of the grid $\widetilde{\mathscr{G}}$, we have $|(\varphi(A) - \beta(A)) - (\varphi(B) - \beta(B))| < 6\eta$, hence

$$\int_{AB} |\varphi' - \beta'|^p < \frac{(6\eta)^p}{\mathscr{H}^1(AB)^{p-1}} < \frac{(6\eta)^p}{\xi^{p-1}} \,,$$

which inserted in (3.7) and keeping in mind (3.5) gives again (3.8). Summarizing, $\partial\mathcal{R}$ is a finite union of segments, and estimate (3.6) holds for all segments except those having at least a vertex of $\mathcal{R}$ as endpoint, for which in turn whe weaker estimate (3.8) holds. Since the latter segments are at most 8, adding over all the segments immediately yields (3.3). $\square$

3.3. **Proof of Theorem A (sufficiency).** This section is devoted to find a sequence of diffeomorphisms which strongly converge to a given Sobolev function satisfying the non-crossing condition. In fact, we will look for a sequence of finitely piecewise affine homeomorphisms; this is easier to achieve with our construction, and the two things are in fact equivalent. Indeed, as proved in [16], given any finitely piecewise affine homeomorphism $v : \mathcal{Q} \to \mathcal{Q}$, coinciding with the identity on the boundary, there is a sequence of diffeomorphisms $v_j : \mathcal{Q} \to \mathcal{Q}$, also coinciding with the identity on the boundary, which converge to v strongly in $W^{1,p}$ and uniformly.

Proof of Theorem A, sufficiency part. Let $u : \mathcal{Q} \to \mathcal{Q}$ be a $W^{1,p}$ function, coinciding with the identity on $\partial\mathcal{Q}$ and satisfying the non-crossing condition, and let $\varepsilon > 0$ be given. To obtain the thesis, we will find a finitely piecewise affine homeomorphism $\omega : \mathcal{Q} \to \mathcal{Q}$, coinciding with the identity on $\partial\mathcal{Q}$, and such that

$$\|\omega - u\|_{W^{1,p}(\mathcal{Q})} < \varepsilon \,; \tag{3.9}$$

moreover, we will have to establish the "improved convergence" stated in the claim. The proof will be divided in several steps for the sake of clarity.

Step I. Definition of δ, ξ, σ and K and subdivision in squares.
Let $\delta = \delta(p, u, \varepsilon)$ be a very small constant, to be specified later, and let H_1, H_2 and H_3 be the

constants of Propositions 2.13, 2.14 and 2.15. Since $u \in W^{1,p}$, we can select $\xi = \xi(p, u, \varepsilon, \delta) \ll \delta$ such that for every set $E \subseteq \mathcal{Q}$ one has

$$|E| < \xi \qquad \Longrightarrow \qquad \int_E |Du|^p < \frac{\delta^2}{4100 H_1} \,. \tag{3.10}$$

Let now $\sigma = \sigma(p, u, \varepsilon, \delta, \xi) \ll \xi$ be a constant such that, calling

$$\mathcal{A}_\sigma = \left\{ M \in \mathbb{R}^{2 \times 2} : \text{ either } |M| \in (0, \sigma) \cup (1/\sigma, +\infty), \text{ or } \det M \in (0, \sigma) \cup (1/\sigma, +\infty) \right\},$$

one has

$$\left| \left\{ x : Du(x) \in \mathcal{A}_\sigma \right\} \right| < \frac{\xi}{2} \,, \qquad\qquad 615 \cdot 2^p \sigma \frac{H_1 + H_2}{\delta} < \delta \,. \tag{3.11}$$

Let now $K = K(p, u, \varepsilon, \delta, \xi, \sigma) \gg 1/\sigma$ be a large integer, to be fixed in a moment. We will subdivide the K^2 squares of the grid $\mathscr{G} = \mathscr{G}(K)$ in four groups, as follows. A square $\mathcal{R}_{i,j}$, with $1 \le i,\, j < K$, will be called a "Lebesgue square adapted to σ" if there is a matrix $M_{i,j} \notin \mathcal{A}_\sigma$ such that $\mathcal{R}_{i,j}$ is a Lebesgue square with matrix $M_{i,j}$ and constant $\kappa = \sigma^{3p}$. All the other squares will be called "bad squares", and we let $\mathcal{Q}_1$ be their union, and set $M_{i,j} = 0$ for any bad square $\mathcal{R}_{i,j}$. Thanks to Lemma 2.11 and to the first estimate in (3.11), up to take $K = K(p, u, \varepsilon, \delta, \xi, \sigma) \gg 1$ big enough, we have

$$\left| \mathcal{Q}_1 \right| < \xi \,. \tag{3.12}$$

Let us now take a Lebesgue square $\mathcal{R}_{i,j}$ adapted to σ, and let us consider the matrix $M_{i,j}$. If $M_{i,j} = 0$, we will say that $\mathcal{R}_{i,j}$ is a "Lebesgue zero square", if $M_{i,j} \neq 0$ but $\det M_{i,j} = 0$ we will say that $\mathcal{R}_{i,j}$ is a "Lebesgue square with zero determinant", and otherwise we will say that $\mathcal{R}_{i,j}$ is a "general Lebesgue square". We will call $\mathcal{Q}_2$, $\mathcal{Q}_3$ and $\mathcal{Q}_4$ the union of all the Lebesgue zero squares, the Lebesgue squares with zero determinant, and the general Lebesgue squares respectively.

Step II. Definition of the functions $v : \mathcal{Q} \to \mathbf{Q}$ and $\varphi : \mathscr{G} \to \mathbf{Q}$.

Let us now apply Lemma 3.3 to the function u, with the constants δ and K and the matrices $M_{i,j}$ introduced in Step I.

We obtain then a $(1 + 4\delta)$-biLipschitz homeomorphism $\Phi : \mathcal{Q} \to \mathcal{Q}$, coinciding with the identity on $\partial\mathcal{Q}$, such that $\|\Phi - \mathrm{Id}\|_{L^\infty} < 2\delta/K$ and that (3.1) holds for every $1 \le i,\, j < K$. We let then $v = u \circ \Phi : \mathcal{Q} \to \mathbf{Q}$. Notice that v is also a $W^{1,p}$ function coinciding with the identity on a neighbourhood of $\partial\mathcal{Q}$ and satisfying the non-crossing condition, and in addition $\mathscr{G}$ is a good starting grid for v. As a consequence, we can apply Proposition 3.8 to find an injective, piecewise linear function $\varphi : \mathscr{G} \to \mathbf{Q}$ satisfying (3.3). We will construct our function $\omega : \mathcal{Q} \to \mathbf{Q}$ satisfying (3.9) so that $\omega = \varphi$ on $\mathscr{G}$. Notice that, since φ is piecewise linear, injective and coincides with the identity on $\partial\mathcal{Q}$, we can define the extension ω independently on every square of the grid $\mathscr{G}$: if each extension is finitely piecewise affine and injective, then the resulting ω is a finitely piecewise affine homeomorphism.

Step III. Extension on the "bad squares" $\mathcal{R}_{i,j} \subseteq \mathcal{Q}_1$.

Let us start by defining the extension ω on a given bad square $\mathcal{R} = \mathcal{R}_{i,j} \in \mathcal{Q}_1$. Since by

construction $M_{i,j} = 0$, the estimates (3.3) of Proposition 3.8 and (3.1) of Lemma 3.3 ensure that

$$\int_{\partial \mathcal{R}} |\varphi'|^p \leq \frac{\sigma}{K} + 2 \int_{\partial \mathcal{R}} |Dv|^p = \frac{\sigma}{K} + 2 \int_{\partial \mathcal{R}} |D(u \circ \Phi)|^p$$

$$\leq \frac{\sigma}{K} + 2(1 + 4\delta)^{p+1} \int_{\Phi(\partial \mathcal{R})} |Du|^p \leq \frac{\sigma}{K} + \frac{410K}{\delta} \int_{\mathcal{R}^+} |Du|^p \,.$$

By Proposition 2.13 we obtain then a finitely piecewise affine homeomorphism $\omega : \mathcal{R} \to \mathbb{R}^2$, coinciding with φ on $\partial \mathcal{R}$, and such that

$$\int_{\mathcal{R}} |D\omega|^p \leq 410 \frac{H_1}{\delta} \int_{\mathcal{R}^+} |Du|^p + \sigma H_1 |\mathcal{R}| \,.$$

Keeping in mind (3.12) and (3.10), we obtain then that

$$\int_{\mathcal{Q}_1} |D\omega|^p \leq \frac{9}{10} \delta + \sigma H_1 \xi \,,$$

hence again by (3.10) and recalling that $\xi \ll \sigma \ll \delta$

$$\int_{\mathcal{Q}_1} |D\omega - Du|^p \leq 2^p \int_{\mathcal{Q}_1} |D\omega|^p + |Du|^p \leq 2^p \delta \,. \tag{3.13}$$

Step IV. Extension on the "Lebesgue zero squares" $\mathcal{R}_{i,j} \subseteq \mathcal{Q}_2$.

Let us now consider a Lebesgue zero square $\mathcal{R} = \mathcal{R}_{i,j}$. Since $M_{i,j}$ is again 0, exactly as in Step III we can define the extension ω by Proposition 2.13 and notice that

$$\int_{\mathcal{R}} |D\omega|^p \leq 410 \frac{H_1}{\delta} \int_{\mathcal{R}^+} |Du|^p + \sigma H_1 |\mathcal{R}| \,. \tag{3.14}$$

Then, we cannot continue the argument of Step III, because the area of $\mathcal{Q}_2$ needs not to be small. However, since $\mathcal{R}$ is a Lebesgue square with matrix $M = 0$ and constant σ^{3p}, by definition

$$\int_{\mathcal{R}^+} |Du|^p \leq \sigma^{3p} |\mathcal{R}| < \sigma |\mathcal{R}| \,. \tag{3.15}$$

Hence, by (3.14) we obtain

$$\int_{\mathcal{R}} |D\omega|^p \leq \left(\frac{410}{\delta} + 1 \right) H_1 \sigma |\mathcal{R}| \leq \frac{411}{\delta} H_1 \sigma |\mathcal{R}| \,,$$

which again by (3.15) and keeping in mind the second estimate in (3.11) gives

$$\int_{\mathcal{Q}_2} |D\omega - Du|^p \leq 2^p \int_{\mathcal{Q}_2} |D\omega|^p + |Du|^p \leq 2^p \sigma \left(411 \frac{H_1}{\delta} + 1 \right) |\mathcal{Q}_2| \leq \delta \,. \tag{3.16}$$

Step V. Extension on the "Lebesgue squares with zero determinant" $\mathcal{R}_{i,j} \subseteq \mathcal{Q}_3$.

Let now $\mathcal{R} = \mathcal{R}_{i,j}$ be a Lebesgue square with zero determinant, and let us call for brevity $M = M_{i,j}$ the corresponding matrix. First of all, since $\mathcal{R}$ is a Lebesgue square with area $1/K^2$ and constant σ^{3p}, and $|M| > \sigma$, we have

$$\|Du - M\|^p_{L^p(\mathcal{R})} < \sigma^{3p} |\mathcal{R}| < \sigma^{2p} |M|^p |\mathcal{R}| = (\sigma^2 \|M\|_{L^p(\mathcal{R})})^p \,,$$

hence $\|Du\|_{L^p(\mathcal{R})} \geq \|M\|_{L^p(\mathcal{R})} (1 - \sigma^2)$, thus

$$\int_{\mathcal{R}} |Du|^p > (1 - \sigma^2)^p |M|^p |\mathcal{R}| > (1 - \sigma) |M|^p |\mathcal{R}| \,. \tag{3.17}$$

Now, by the estimates (3.3) of Proposition 3.8 and (3.1) of Lemma 3.3, and recalling that for every $a, b \in \mathbb{R}$ it is $(a + \delta b)^p \leq a^p + 2p\delta(a^p + b^p)$, we have

$$\int_{\partial \mathcal{R}} |\varphi' - M\tau|^p \leq \frac{\sigma}{K} + 2 \int_{\partial \mathcal{R}} |Dv - M|^p = \frac{\sigma}{K} + 2 \int_{\partial \mathcal{R}} |D(u \circ \Phi) - M|^p$$

$$\leq \frac{\sigma}{K} + 2(1 + 4\delta)^{p+1} \int_{\Phi(\partial \mathcal{R})} \left(|Du - M| + 5\delta|M| \right)^p$$

$$\leq \frac{\sigma}{K} + 3 \int_{\Phi(\partial \mathcal{R})} (1 + 10p\delta)|Du - M|^p + 10p\delta|M|^p$$

$$\leq \frac{\sigma}{K} + \frac{121p\delta}{K} |M|^p + \frac{613K}{\delta} \int_{\mathcal{R}^+} |Du - M|^p \leq \frac{121p\delta}{K} |M|^p + \frac{614}{K\delta} \sigma \,,$$

where the last inequality comes recalling that $\mathcal{R}$ is a Lebesgue square with matrix M and constant $\sigma^{3p} < \sigma$. Since M has zero determinant, we can apply Proposition 2.14 to get a finitely piecewise affine homeomorphism ω on $\mathcal{R}$, coinciding with φ on $\partial \mathcal{R}$, such that

$$\int_{\mathcal{R}} |D\omega - M|^p \leq \frac{H_2}{K^2} \left(121p\delta \, |M|^p + \frac{614}{\delta} \sigma \right).$$

So, from (3.17) we deduce

$$\int_{\mathcal{R}} |D\omega - Du|^p \leq 2^p \left(\int_{\mathcal{R}} |D\omega - M|^p + \int_{\mathcal{R}} |Du - M|^p \right) \leq 2^{p+7} p\delta H_2 \int_{\mathcal{R}} |Du|^p + 615 \, \frac{2^p H_2}{\delta} \sigma |\mathcal{R}| \,,$$

and again by the second estimate in (3.11), adding over all the squares contained in $\mathcal{Q}_3$ we get

$$\int_{\mathcal{Q}_3} |D\omega - Du|^p \leq \delta \left(1 + H_2 2^{p+7} p \int_{\mathcal{Q}_3} |Du|^p \right). \tag{3.18}$$

Step VI. Extension on the "general Lebesgue squares" $\mathcal{R}_{i,j} \subseteq \mathcal{Q}_4$.

Let us finally consider a general Lebesgue square $\mathcal{R} = \mathcal{R}_{i,j}$, with corresponding matrix $M = M_{i,j}$. The calculation already done in Step V, together with the fact that $614\sigma < \delta^2$ by (3.11), ensures

$$\int_{\partial \mathcal{R}} |\varphi' - M\tau|^p \leq \frac{121p\delta}{K} |M|^p + \frac{614}{K\delta} \sigma < (121p\delta|M|^p + 1) \frac{\delta}{K} \,.$$

As a consequence, by Proposition 2.15 we get a finitely piecewise affine homeomorphism ω on $\mathcal{R}$, coinciding with φ on $\partial \mathcal{R}$, such that also by (3.17) we have

$$\int_{\mathcal{R}} |D\omega - M|^p \leq \frac{H_3}{K^{2-\frac{1}{p}}} \left(\int_{\partial \mathcal{R}} |\varphi' - M\tau|^p \right)^{\frac{1}{p}} \leq \frac{H_3}{K^2} (121p\delta|M|^p + 1)\delta^{1/p}$$

$$\leq H_3 \delta^{1/p} |\mathcal{R}| + 122p\delta H_3 \int_{\mathcal{R}} |Du|^p \,.$$

Arguing as usual, we get then

$$\int_{\mathcal{R}} |D\omega - Du|^p \leq 2^p \left(\int_{\mathcal{R}} |D\omega - M|^p + \int_{\mathcal{R}} |Du - M|^p \right)$$

$$\leq 2^p |\mathcal{R}| \left(H_3 \delta^{1/p} + \sigma \right) + 2^{p+7} p\delta H_3 \int_{\mathcal{R}} |Du|^p \,,$$

which adding on the squares contained in $\mathcal{Q}_4$ gives

$$\int_{\mathcal{Q}_4} |D\omega - Du|^p \leq \delta^{1/p} \left(2^{p+1} H_3 + 2^{p+7} p H_3 \int_{\mathcal{Q}_4} |Du|^p \right). \tag{3.19}$$

Step VII. Existence of the approximation.

In Steps III–VI we have defined the function ω on each of the finitely many squares $\mathcal{R}_{i,j}$. By construction, the resulting function $\omega : \mathcal{Q} \to \mathcal{Q}$ is a finitely piecewise affine homeomorphism coinciding with the identity on $\partial \mathcal{Q}$. Moreover, by (3.13), (3.16), (3.18) and (3.19), and since $\omega - u \in W_0^{1,p}$, we get the validity of (3.9) as soon as δ has been chosen small enough, only depending on p, u and ε. The proof of the existence of a sequence strongly converging to u is then concluded.

Step VIII. Proof of points (i) and (ii).

To conclude the sufficiency part of Theorem A, we need to show the validity of points (i) and (ii) in the statement. To this end we first note that, on each square $\mathcal{R}$ of the starting grid $\mathscr{G}$, the map ω constructed in the above steps satisfies

$$\|u - \omega\|_{L^\infty(\mathbb{R})} \leq C\Big(\operatorname{osc}_{\mathcal{R}} u + \operatorname{osc}_{\partial \mathcal{R}} u + \eta \Big),$$

where η is the size of the grid $\widetilde{\mathscr{G}}$ and C is a purely geometric constant. Hence, to obtain the validity of points (i) and (ii), we need to show that given ε small and setting

$$\mathcal{S} = \bigcup_{\mathcal{R} : \operatorname{osc}_{\partial \mathcal{R}} u + \operatorname{osc}_{\mathcal{R}} u \geq 2\varepsilon} \mathcal{R},$$

then, as soon as the starting and arrival grids $\mathscr{G}$ and $\widetilde{\mathscr{G}}$ are fine enough, we have $\mathcal{S} = \emptyset$ if u is continuous, and $\mathscr{H}_\infty^{2-p}(\mathcal{S}) \leq \varepsilon$ if $p \in [1, 2]$.

Being the first part clear by uniform continuity, we will just prove the second one. For, note that by Corollary 2.7 $u \in INV^+(\mathcal{Q})$ and thus, by (2.2), for every square $\mathcal{R}$ of the starting grid $\mathscr{G}$ one has

$$\big(\operatorname{osc}_{\mathcal{R}} u \big)^p \leq \big(\operatorname{osc}_{\partial \mathcal{R}} u \big)^p \leq C\ell(\mathcal{R})^{p-2} \int_{\mathcal{R}^+} |Du|^p,$$

where $\mathcal{R}^+$ is the union of $\mathcal{R}$ with the squares of $\mathscr{G}$ which are adjacent to it, and where $\ell(\mathcal{R}) = 1/K$ is the side of $\mathcal{R}$. In particular, if $\operatorname{osc}_{\partial \mathcal{R}} u + \operatorname{osc}_{\mathcal{R}} u \geq 2\varepsilon$, then $\operatorname{osc}_{\mathcal{R}} u \geq \varepsilon$ hence

$$|\mathcal{R}| = \ell(\mathcal{R})^2 \leq \frac{C\ell(\mathcal{R})^p}{\varepsilon^p} \int_{\mathcal{R}^+} |Du|^p = \frac{CK^{-p}}{\varepsilon^p} \int_{\mathcal{R}^+} |Du|^p,$$

and then, up to multiply C by 9, we have

$$|\mathcal{S}| \leq \frac{CK^{-p}}{\varepsilon^p} \int_{\mathcal{S}^+} |Du|^p, \tag{3.20}$$

being $\mathcal{S}^+$ the union of all the squares which either belong to $\mathcal{S}$, or are adjacent to a square of $\mathcal{S}$. Since $Du \in L^p(\mathcal{Q})$, as soon as K is big enough depending on ε we have by the above estimate that $\mathcal{S}^+$ is as small as we wish, hence by the absolute continuity of the integral we deduce

$$\int_{\mathcal{S}^+} |Du|^p \leq \frac{\varepsilon^{p+1}}{C},$$

so by (3.20) we get that

$$\mathscr{H}_\infty^{2-p}(\mathcal{S}) \leq \sum_{\mathcal{R} : \operatorname{osc}_{\partial R} u + \operatorname{osc}_R u \geq 2\varepsilon} \ell(\mathcal{R})^{2-p} = |\mathcal{S}| K^p \leq \frac{C}{\varepsilon^p} \int_{\mathcal{S}^+} |Du|^p \leq \varepsilon,$$

and this concludes the proof of the final claim in Theorem A by suitably choosing ε. $\qquad \square$

3.4. **Proof of Theorem A (necessity) and of Theorem B.** In this Section we show that the non-crossing condition is necessary, for a function, to be approximable by diffeomorphisms, thus completing the proof of Theorem A and of Theorem B.

Lemma 3.9. *Let $\{u_k\} \subseteq \mathcal{D}_{\mathrm{id}}(\mathcal{Q}) \cap W^{1,p}(\mathcal{Q})$ be a sequence, weakly converging to u in $W^{1,p}(\mathcal{Q})$. Then u satisfies the non-crossing condition.*

Proof. We start noticing that by Corollary 2.7 and Lemma 2.5 u is continuous outside a set $\mathcal{S}_u$ of zero $\mathscr{H}^1$-measure. Moreover, by extending each u_k as the identity outside $\mathcal{Q}$ and by performing a small dilation, we can assume that all the maps u_k are equal to the identity on a neighbourhood of $\partial \mathcal{Q}$.

Let us now assume that $p = 1$, the other case being analogous. By Dunford-Pettis Theorem there exists a superlinear function g such that

$$\sup_k \int_{\mathcal{Q}} g(|Du_k|) < +\infty.$$

Let $\varepsilon > 0$ and be $\Gamma \subseteq \mathcal{Q} \setminus \mathcal{S}_u$ be as in the Definition 1.2. Note that, by pre-composing all the maps with a Bi-Lipischitz transformation, we can assume without loss of generality that Γ is composed by piecewise linear curves. In particular, by making small translations of all the segments belonging to Γ and by slightly extending them, we can find, for almost every small δ, a set Γ_δ with the following properties,

- $\Gamma_\delta = \Phi_\delta(\Gamma)$ where Φ_δ is a bi-Lipschitz transformation equal to the identity on $\partial \mathcal{Q}$ and such that $\|\Phi_\delta - \mathrm{Id}\|_{L^\infty} \leq \delta$;
- there exists a (not relabed) subsequence such that $u_k \rightharpoonup u$ in $W^{1,1}(\Gamma_\delta)$.

By the continuity of u on Γ, we can first chose δ sufficiently small such that

$$\|u - u \circ \Phi_\delta\|_{L^\infty(\Gamma)} \leq \varepsilon/2$$

and then, by the 1-dimensional Sobolev embedding Theorem, $\bar{k}$ sufficiently big such that

$$\|u_{\bar{k}} - u\|_{L^\infty(\Gamma_\delta)} \leq \varepsilon/2.$$

The map $\psi = u_{\bar{k}} \circ \Phi_\delta : \Gamma \to \mathcal{Q}$ is then the injective modification of u. $\square$

Proof of Theorem A. The sufficiency part has been showed in Section 3.3 while the necessary one follows immediately from Lemma 3.9. $\square$

Proof of Theorem B. We divide the proof in two short steps.

Step I. The set of maps satisfying the non-crossing condition is (sequentially) weakly closed.
Let $\{u_j\}$ be a sequence of maps satisfying the non-crossing condition and such that $u_j \rightharpoonup u$. We claim that u satisfies the non-crossing condition as well. Indeed by Theorem A for every j there is a diffeomorphism v_j such that $\|v_j - u_j\|_{W^{1,p}} \leq 2^{-j}$, hence $v_j \rightharpoonup u$ and by Lemma 3.9, u satisfies the non-crossing condition.

Step II. Conclusion.
By the previous step, if $u \in \overline{\mathcal{D}_{\mathrm{Id}}(\Omega)}^{\mathrm{ws}-W^{1,p}}$ then u satisfies the non-crossing condition. Hence, by Theorem A, $u \in \overline{\mathcal{D}_{\mathrm{Id}}(\Omega)}^{\mathrm{s}-W^{1,p}}$, and this concludes the proof. $\square$

4. Proof of Theorem C

This section is devoted to prove Theorem C. Through the section, u will be a given $W^{1,p}$ function, coinciding with the identity on the boundary, such that the counter-image (in the sense of Definition 2.3) of any closed, connected subset of $\mathcal{Q}$ which does not disconnect $\mathcal{Q}$ is a closed, connected subset of $\mathcal{Q}$ which does not disconnect it. In particular, by Lemma 2.8 we know that u satisfies the INV^+ condition, thus by Lemma 2.5 we find a $\mathscr{H}^1$-negligible set $\mathcal{S}_u \subseteq \mathcal{Q}$ such that u is continuous on $\mathcal{Q} \setminus \mathcal{S}_u$. The proof is divided in some subsections for clarity.

4.1. **Preliminary geometrical properties.**
Let us start by listing some simple geometrical properties of curves and connected sets.

Lemma 4.1. *Let $\Gamma \subseteq \mathrm{Int}(\mathcal{Q})$ be a closed, connected set which does not disconnect $\mathcal{Q}$ (that is, $\mathcal{Q} \setminus \Gamma$ is connected). Then, for every $\varepsilon > 0$ there exists a C^1, closed, injective curve $\theta : \mathbb{S}^1 \to \mathrm{Int}(\mathcal{Q})$, such that Γ is contained in the internal part of θ, and for every $s \in \mathbb{S}^1$ one has $\mathrm{dist}(\theta(s), \Gamma) < \varepsilon$.*

Proof. Let $\varphi : \mathcal{Q} \to \mathbb{R}^+$ be a regularised distance function from Γ, that is, φ is a regular function and for every $x \in \mathcal{Q}$ one has

$$\mathrm{dist}(x, \Gamma) \leq \varphi(x) < 2\mathrm{dist}(x, \Gamma).$$

By Sard's Lemma, there is some $0 < \varepsilon' \leq \varepsilon$ such that $\varphi > \varepsilon'$ on $\partial \mathcal{Q}$ and $\varphi^{-1}(\varepsilon')$ is the union of finitely many C^1 curves. Since the level set $\{\varphi = \varepsilon'\}$ disconnects Γ and $\partial \mathcal{Q}$, one of those curves must contain Γ in its internal part. $\qquad\square$

Lemma 4.2. *Let $\mathcal{C} \subseteq \mathrm{Int}(\mathcal{Q})$ be a closed set which does not disconnect $\mathcal{Q}$, and let $\gamma : [0,1] \to \mathrm{Int}(\mathcal{Q})$ be an admissible curve in the sense of Definition 3.1. If γ intersects $\mathcal{C}$ exactly once, then $\mathcal{C} \cup \gamma$ does not disconnect $\mathcal{Q}$; if γ intersects $\mathcal{C}$ exactly twice, then $\mathcal{Q} \setminus (\mathcal{C} \cup \gamma)$ has at most two connected components, in particular surely two if $\mathcal{C}$ is connected.*

Proof. For any number $\delta > 0$, much smaller than the diameter of γ and the distance between γ and $\partial \mathcal{Q}$, we call $\mathcal{A}_\delta$ the set of the points $x \in \mathcal{Q} \setminus (\mathcal{C} \cup \gamma)$ with distance larger than δ from $\mathcal{C} \cup \gamma$, and which can be connected with $\partial \mathcal{Q}$ with a path having distance at least δ from $\mathcal{C}$. Notice that, since $\mathcal{C}$ does not disconnect $\mathcal{Q}$, the sets $\mathcal{A}_\delta$ fill the whole $\mathcal{Q} \setminus (\mathcal{C} \cup \gamma)$ when $\delta \searrow 0$.

Let us now assume that γ intersects $\mathcal{C}$ only once, and precisely at $\gamma(0)$. For every $\delta \ll \mathrm{diam}(\gamma)$ we can find $\eta \ll \delta$ such that, if for some $\bar{t} \in (0,1)$ one has $|\gamma(0) - \gamma(\bar{t})| = \eta$, then for all $0 < t < \bar{t}$ one has $|\gamma(0) - \gamma(t)| < \delta$. Applying now Lemma 4.1 with $\varepsilon \ll \eta$ to the image of γ, which is a closed, connected set which does not disconnect $\mathcal{Q}$, we find a C^1, closed, injective curve θ, containing γ in its internal part, and so that $\mathrm{dist}(\theta(s), \gamma) < \varepsilon$ for every $s \in \mathbb{S}^1$. If ε is small enough, depending on $\mathcal{C}$ and η, then the intersection between θ and $\mathcal{C}$ contains only points which have distance strictly less than η from $\gamma \cap \mathcal{C} = \gamma(0)$. There is a unique subpath θ_1 of θ, with $\mathrm{diam}(\theta_1) \approx \mathrm{diam}(\theta)$, whose both endpoints have distance η from $\gamma(0)$, and whose internal points have all distance strictly larger than η from $\gamma(0)$; by construction, $\theta_1 \cap (\mathcal{C} \cup \gamma) = \emptyset$. We can extend θ_1 to an injective, closed curve $\tilde{\theta}$, in such a way that $\tilde{\theta} \setminus \theta_1$ is an arc of the circle

$\{z \in \mathcal{Q} : |z - \gamma(0)| = \eta\}$; it is possible to do this in two distinct ways, we just pick one of them. Notice that γ is not necessarily contained in the internal part of $\tilde{\theta}$; nevertheless, let us call $\bar{t} \in [0, 1]$ the smallest number such that the restriction of γ to $[\bar{t}, 1]$ is contained in the internal part of $\tilde{\theta}$. Since γ does not intersect θ_1, then either $\bar{t} = 0$ or $\gamma(\bar{t})$ has distance exactly η from $\gamma(0)$: in both cases, the curve γ in $[0, \bar{t}]$ only contains points within distance $\eta \ll \delta$ from $\gamma(0)$.

Let now $x \in \mathcal{A}_\delta$, and let σ be a path connecting x with $\partial \mathcal{Q}$ having always distance at least δ from $\mathcal{C}$. If $\sigma \cap \tilde{\theta} = \emptyset$, then σ is entirely contained in the external part of $\tilde{\theta}$ (we can exclude that σ is entirely in the internal part because x has distance larger than δ from γ). As a consequence, σ does not contain any point of the restriction of γ to $[\bar{t}, 1]$; on the other hand, σ does also not contain points with distance less than δ from $\gamma(0) \in \mathcal{C}$, hence it cannot intersect the restriction of γ to $[0, \bar{t}]$. As a consequence, the whole path σ is contained in $\mathcal{Q} \setminus (\mathcal{C} \cup \gamma)$, so x is in the same connected component of $\mathcal{Q} \setminus (\mathcal{C} \cup \gamma)$ as $\partial \mathcal{Q}$. Assume instead that $\sigma \cap \tilde{\theta} \neq \emptyset$, and let x^+ and y^- be the first and the last point of σ which intersect $\tilde{\theta}$; arguing as before, the curve σ cannot intersect γ between x and x^+, nor between y^+ and the last point of σ, and moreover x^+ and y^- belong to θ_1; indeed, points of σ have distance at least $\delta \gg \eta$ from $\mathcal{C}$, while points of $\tilde{\theta} \setminus \theta_1$ have distance η from $\gamma(0) \in \mathcal{C}$. Then, we can find a path between x and $\partial \mathcal{Q}$ putting together σ from x to x^+, then θ_1 between x^+ and y^-, and finally again σ between y^- and $\partial \mathcal{Q}$. By construction, this path does not intersect $\mathcal{C} \cup \gamma$, so again x is in the same connected component of $\mathcal{Q} \setminus (\mathcal{C} \cup \gamma)$ as $\partial \mathcal{Q}$. Summarizing, $\mathcal{A}_\delta$ belongs to a connected component of $\mathcal{Q} \setminus (\mathcal{C} \cup \gamma)$, and since this holds for every $\delta \ll 1$ we deduce that $\mathcal{C} \cup \gamma$ does not disconnect $\mathcal{Q}$. We have then obtained the thesis in the case that γ intersect $\mathcal{C}$ only at $\gamma(0)$, and of course the same argument works if the only intersection point is at $\gamma(1)$.

Assume now that γ intersects $\mathcal{C}$ only once, and precisely at $\gamma(t)$ for some $0 < t < 1$. Then, we can write $\gamma = \gamma_1 \cup \gamma_2$, where γ_1 and γ_2 are the restrictions of γ to $[0, t]$ and $[t, 1]$ respectively. The above argument applied to $\mathcal{C}$ and γ_1 ensures that $\mathcal{C} \cup \gamma_1$ is a closed set which does not disconnect $\mathcal{Q}$. And then, the above argument again applied to $\mathcal{C} \cup \gamma_1$ and γ_2 ensures that also $(\mathcal{C} \cup \gamma_1) \cup \gamma_2 = \mathcal{C} \cup \gamma$ does not disconnect $\mathcal{Q}$.

Let us now pass to consider the case that γ intersects $\mathcal{C}$ exactly twice; the very same argument of the last sentence ensures that it is not restrictive to assume that the intersection points are $\gamma(0)$ and $\gamma(1)$.

The argument is similar as before: for any δ much smaller than the diameter of γ and the distance between γ and $\partial \mathcal{Q}$ we select $\eta \ll \delta$ such that if for some $\bar{t} \in (0, 1)$ one has $|\gamma(0) - \gamma(\bar{t})| = \eta$ (resp., $|\gamma(1) - \gamma(\bar{t})| = \eta$), then for all $0 < t < \bar{t}$ (resp., all $\bar{t} < t < 1$) one has $|\gamma(0) - \gamma(t)| < \delta$ (resp., $|\gamma(1) - \gamma(t)| < \delta$). Let then again θ be given by Lemma 4.1, with $\varepsilon \ll \eta$ in such a way that $\theta \cap \mathcal{C}$ only contains points with distance strictly less than η from $\gamma(0)$ or $\gamma(1)$. This time, we find two distinct subpaths θ_1 and θ_2 of θ, each of them with the property that one endpoint has distance η from $\gamma(0)$, the other one has distance η from $\gamma(1)$, and all the interior points (that is, all the points which are not endpoints) have distance strictly larger than η from both $\gamma(0)$ and $\gamma(1)$. Again, by construction θ_1 and θ_2 do not intersect $\mathcal{C} \cup \gamma$. We define now a closed, injective curve $\tilde{\theta}$ by putting together θ_1, θ_2 and two arcs of circle with radius η, one with

center at $\gamma(0)$ and the other at $\gamma(1)$. Up to decrease δ if necessary, we can assume the existence of a path between $\gamma(1/2)$ and $\partial\mathcal{Q}$ having distance larger than δ from $\mathcal{C}$, so in particular from $\gamma(0) \cup \gamma(1)$; since $\gamma(1/2)$ is in the internal part of $\tilde{\theta}$, there exists a last point of this path in $\tilde{\theta}$, and we assume without loss of generality that this point belongs to θ_1. In other words, all the points of θ_1 are in the same connected component as $\partial\mathcal{Q}$ inside $\mathcal{Q} \setminus (\mathcal{C} \cup \gamma)$. Let us now take any point $x \in \mathcal{A}_\delta$, and let us consider a path σ between x and $\partial\mathcal{Q}$ having distance larger than δ from $\mathcal{C}$. If σ does not intersect $\tilde{\theta}$, then exactly as before we deduce that it also does not intersect γ, so x is in the same connected component of $\mathcal{Q} \setminus (\mathcal{C} \cup \gamma)$ as $\partial\mathcal{Q}$, hence also the same as θ_1. Otherwise, let x^+ be the first point of σ which intersects $\tilde{\gamma}$: as before, we know that x^+ belongs either to θ_1 or to θ_2. Summarizing, every point of $\mathcal{A}_\delta$ is connected, in $\mathcal{Q} \setminus (\mathcal{C} \cup \gamma)$, either with θ_1 or with θ_2; then, every $\mathcal{A}_\delta$ intersects at most two connected components of $\mathcal{Q} \setminus (\mathcal{C} \cup \gamma)$, hence $\mathcal{Q} \setminus (\mathcal{C} \cup \gamma)$ has at most two connected components.

To conclude, we assume that $\mathcal{C} \cap \gamma = \gamma(0) \cup \gamma(1)$ and that $\mathcal{Q} \setminus (\mathcal{C} \cup \gamma)$ is connected, and we have to prove that $\mathcal{C}$ is not connected. Let the paths θ_1 and θ_2 be as before, and let $x \in \theta_1$ and $y \in \theta_2$ be two points very close to $\gamma(1/2)$. Since $\mathcal{Q} \setminus (\mathcal{C} \cup \gamma)$ is connected, there is a path σ between x and y which does not intersect $\gamma \cup \mathcal{C}$, and we can easily assume that this path does not intersect the open segment xy. As a consequence, the union between σ and the segment xy is an injective, closed curve, which does not intersect $\mathcal{C}$; by construction, one between $\gamma(0)$ and $\gamma(1)$ is in the internal part of this curve, and the other one is in the external part, hence $\mathcal{C}$ is not connected and the proof is concluded. $\qquad\square$

Lemma 4.3. *Let $\gamma : [0,1] \to \mathcal{Q} \setminus \mathcal{S}_u$ be an admissible curve with both endpoints in $\partial\mathcal{Q}$, let $\boldsymbol{P} \in \text{Int}(\mathcal{Q})$ and assume that $(u \circ \gamma)^{-1}(\boldsymbol{P})$ consists of exactly two numbers t_1 and t_2 and that $v_1 = D(u \circ \gamma)(t_1)$ and $v_2 = D(u \circ \gamma)(t_2)$ are both well-defined and non-zero. Then, the two vectors v_1 and v_2 are parallel and with opposite directions.*

Proof. Since $u = \text{Id}$ in a neighborhood of $\partial\mathcal{Q}$, we can assume without loss of generality that γ intersects $\partial\mathcal{Q}$ only at $\gamma(0)$ and $\gamma(1)$, and we can extend γ to an injective, closed curve $\tilde{\gamma}$ with $\tilde{\gamma} \setminus \gamma \subseteq \partial\mathcal{Q}$. Since u is continuous also on $\tilde{\gamma} \subseteq \mathcal{Q} \setminus \mathcal{S}_u$, by Lemma 2.8 we know that u satisfies the INV condition and, in particular, that the points in $\mathcal{Q} \setminus u(\tilde{\gamma})$ have all degree 0 or 1 with respect to $u(\tilde{\gamma})$. Since $u^{-1}(\boldsymbol{P}) \cap \gamma = \{t_1, t_2\}$, $u \circ \gamma(t)$ is close to $\boldsymbol{P}$ only if t is close to t_1 or t_2; as a consequence, the points near $\boldsymbol{P}$ which belong to the image of $u \circ \gamma$ must have direction with respect to $\boldsymbol{P}$ very close to v_1 or v_2. Call ω_1 and ω_2 the unit vectors obtained by rotating clockwise of $90°$ the directions of v_1 and v_2. Assume first that v_1 and v_2 are not parallel, hence so are neiter ω_1 and ω_2: then, there is a short segment σ near $\boldsymbol{P}$, with direction ν, which intersects $u \circ \gamma$ exactly twice, with $\nu \cdot \omega_1 > 0$, $\nu \cdot \omega_2 > 0$. As a consequence, the degree of the last point of σ equals the degree of the first one plus two, hence the two degrees cannot both be in $\{0, 1\}$ and we have a contradiction. Assume instead that v_1 and v_2 are parallel: then, we can find again a segment σ near $\boldsymbol{P}$, with direction ν, which intersects $u \circ \gamma$ exactly twice, and we can assume that $\nu \cdot \omega_1 > 0$. If the direction of v_2 is the same as that of v_1, then also $\nu \cdot \omega_2 > 0$ and we have a contradiction as before. $\qquad\square$

Lemma 4.4. *Let $P \in \mathrm{Int}(\mathcal{Q})$, and let the segments A^-A^+ and B^-B^+ be two admissible curves in $\mathrm{Int}(\mathcal{Q}) \setminus \mathcal{S}_u$ which intersect $u^{-1}(P)$ exactly in two points $A \in A^-A^+$ and $B \in B^-B^+$. Then, there exists an admissible curve $\gamma : [0,1] \to \mathcal{Q} \setminus \mathcal{S}_u$, with $\gamma(0)$ and $\gamma(1)$ in $\partial\mathcal{Q}$, which intersects $u^{-1}(P)$ exactly at A and B, and which contains the segment A^-A^+ and either B^-B^+ or B^+B^-. In particular, only one of these two possibilities can occur if the derivatives of u at A in the direction of A^-A^+ and at B in the direction of B^-B^+ are both well-defined and non-zero.*

Proof. Let us call for brevity $\mathcal{C} = u^{-1}(P)$, which is a closed, connected set which does not disconnect $\mathcal{Q}$. By Lemma 4.2, we know that $\mathcal{C} \cup A^-A^+ \cup B^-B^+$ does not disconnect $\mathcal{Q}$, hence we find an injective curve γ_0 between $\partial\mathcal{Q}$ and A^- whose interior does not intersect $\mathcal{C} \cup A^-A^+ \cup B^-B^+$. Since this curve has a strictly positive Hausdorff distance from $\mathcal{C}$, we can assume without loss of generality that it is an admissible curve; moreover, since $\mathscr{H}^1(\mathcal{S}_u) = 0$, we can also assume without loss of generality that $\gamma_0 \cap \mathcal{S}_u = \emptyset$. Again by Lemma 4.2, $\mathcal{C} \cup \gamma_0 \cup A^-A^+ \cup B^-B^+$ does not disconnect $\mathcal{Q}$, so we can find an admissible curve $\gamma_1 \subseteq \mathcal{Q} \setminus \mathcal{S}_u$ between A^+ and B^-, and an admissible curve $\gamma_2 \subseteq \mathcal{Q} \setminus \mathcal{S}_u$ between B^+ and $\partial\mathcal{Q}$, whose interiors do not intersect $\mathcal{C} \cup \gamma_0 \cup A^-A^+ \cup B^-B^+$. Up to a trivial modification of the curves γ_1 and γ_2, we can assume that $\gamma_1 \cap \gamma_2$ is either empty or consists of exactly a point.

If $\gamma_1 \cap \gamma_2 = \emptyset$, then the curve γ_0, followed by the segment A^-A^+, then by the curve γ_1, the segment B^-B^+, and the curve γ_2, is an admissible curve in $\mathcal{Q} \setminus \mathcal{S}_u$ starting and ending at $\partial\mathcal{Q}$, which intersects $\mathcal{C}$ exactly at A and B and which contains the segments A^-A^+ and B^-B^+. In particular, if the derivatives of u at A in the direction of A^-A^+ and at B in the direction of B^-B^+ are both well-defined and non-zero, then by Lemma 4.3 these two derivatives are parallel and with opposite direction.
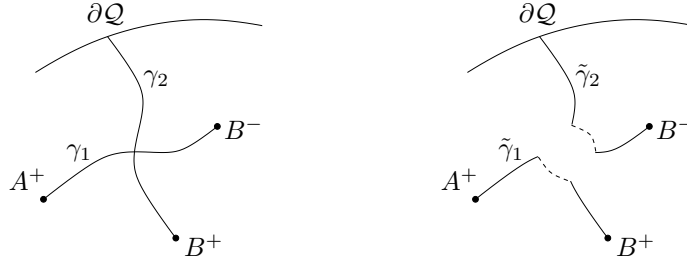

FIGURE 1. Definition of $\tilde{\gamma}_1$ and $\tilde{\gamma}_2$ starting from γ_1 and γ_2 in Lemma 4.4.

Instead, suppose that $\gamma_1 \cap \gamma_2$ consist of a point; then, we can trivially build an admissible curve $\tilde{\gamma}_1$ between A^+ and B^+, and an admissible curve $\tilde{\gamma}_2$ between B^- and $\partial\mathcal{Q}$, with empty intersection and with interiors not intersecting $\mathcal{C} \cup \gamma_0 \cup A^-A^+ \cup B^-B^+$ (see Figure 1). In this case, the searched curve is γ_0, followed by A^-A^+, by $\tilde{\gamma}_1$, by B^+B^-, and by $\tilde{\gamma}_2$. Notice that, in this second case, if the derivatives of u at A and B in the direction of A^-A^+ and B^-B^+ are both well-defined and non-zero, then they must be parallel and with the same direction, since γ contains the oriented segment B^+B^-, and not B^-B^+. This concludes the proof. $\square$

4.2. **Left and right connectibility.** In this section we present the definition of the left and right connectibility, and we show their main properties.

Definition 4.5 (Compatible curve)**.** *Let $\sigma : [0,1] \to \mathcal{Q} \setminus \mathcal{S}_u$ be an admissible curve such that $\sigma(0), \sigma(1) \in \partial\mathcal{Q}$ while $\sigma(t) \in \mathrm{Int}(\mathcal{Q})$ for every $0 < t < 1$, and let $P \in \mathrm{Int}(\mathcal{Q})$. We say that σ is compatible with P if $u^{-1}(P) \cap \sigma$ consists of finitely many points $\{\sigma(t_i),\, 1 \leq i \leq N\}$, σ contains a short segment around each $\sigma(t_i)$, each point $\sigma(t_i)$ is a Lebesgue point for Du, $D(u \circ \sigma)$ is non-zero at each t_i and*

$$D(u \circ \sigma)(t) = Du(\sigma(t))\sigma'(t).$$

We will call σ^- and σ^+ the closures of the left and the right connected components of $\mathcal{Q} \setminus \sigma$.

In the following Lemma we show that if σ is compatible with P and $\mathcal{H}^0(u^{-1}(P)) > 1$, then Du has rank 1 at each point of $\sigma \cap u^{-1}(P)$. This will be useful in the proof of Lemma 4.20.

Lemma 4.6. *Let $P \in \mathcal{Q}$ be such that $u^{-1}(P)$ consists of more than one element. Then, for every $Q \in u^{-1}(P)$, there exists exactly one direction $\nu \in \mathbb{S}^1$ such that, whenever σ is a path compatible with P, then $Du(Q)\nu = 0$. Moreover, for every sequence $\{Q_j\} \subseteq u^{-1}(P)$ converging to Q, one has*

$$\nu(Q) = \lim_{j \to \infty} \pm \frac{Q_j - Q}{|Q_j - Q|}. \tag{4.1}$$

Proof. To prove the first part of the Lemma, it is enough to assume the existence of a path σ compatible with P and containing Q, and to deduce that $\mathrm{rank}(Du(Q)) = 1$. Indeed, by definition $Du(Q) \neq 0$, so we only have to exclude that $Du(Q)$ is invertible; to do so, we will show that there exists $0 < r \ll 1$ such that

$$u^{-1}(P) \cap \partial B(Q, r) = \emptyset, \tag{4.2}$$

which gives a contradiction with the fact that $u^{-1}(P)$ is a connected set with more than one point.

To prove (4.2), by contradicion, we can assume without loss of generality that $Q = P = (0,0)$ and that $Du(Q) = \mathrm{Id}$. Using that Q is a Lebesgue point of the gradient it is then well known that the maps

$$u_r(y) = \frac{1}{r} u(ry)$$

converge in $W^{1,1}_{\mathrm{loc}}(\mathbb{R}^2)$ to $u_\infty = \mathrm{Id}$. By the one dimensional Sobolev embedding and a simple slicing argument, we can find a radius $R \in (1/2, 1)$ and a subsequence $r_j \searrow 0$ such that

$$\|u_{r_j} - \mathrm{Id}\|_{L^\infty(\partial B(0,R))} \to 0$$

Choosing j big enough and scaling back, we deduce (4.2).

Let us now show (4.1); we can assume without loss of generality that $Q = P = (0,0)$ and that $Du(0) = \mathrm{e}_1 \otimes \mathrm{e}_1$, so that $\nu = \mathrm{e}_2$. Let $\varepsilon > 0$; arguing as before, thanks to the 1-dimensional Sobolev embedding we obtain that, for almost every $r \ll 1$,

$$|u(x) - (x_1, 0)| < \varepsilon |x|$$

for every x with $|x| = r$. This means that, for almost every $r \ll 1$, one has

$$\left\{\theta \in \mathbb{S}^1 : r\theta \in u^{-1}(P)\right\} \subseteq \left(\frac{\pi}{2} - 2\varepsilon, \frac{\pi}{2} + 2\varepsilon\right) \cup \left(-\frac{\pi}{2} - 2\varepsilon, -\frac{\pi}{2} + 2\varepsilon\right).$$

In the very same way, for almost every $|y| \ll 1$ we obtain that

$$\left\{\theta \in \mathbb{S}^1 : (y/\tan\theta, y) \in u^{-1}(\boldsymbol{P})\right\} \subseteq \left(\frac{\pi}{2} - 2\varepsilon, \frac{\pi}{2} + 2\varepsilon\right) \cup \left(-\frac{\pi}{2} - 2\varepsilon, -\frac{\pi}{2} + 2\varepsilon\right).$$

The last two inclusions, together with the fact that $u^{-1}(\boldsymbol{P})$ is connected, immediately imply the validity of (4.1). $\qquad\square$

Definition 4.7 (Connectibility). *Let $\boldsymbol{P} \in \mathrm{Int}(\boldsymbol{\mathcal{Q}})$, let $\sigma : [0,1] \to \mathcal{Q}$ be compatible with $\boldsymbol{P}$, and let A and B be any two points of $u^{-1}(\boldsymbol{P}) \cap \sigma$. We say that A and B are left-connectible if there is a connected component of $u^{-1}(\boldsymbol{P}) \cap \sigma^-$ containing both A and B. The definition of the right-connectibility is the same one, replacing σ^- with σ^+.*

The following characterization of the connectibility will be useful in the sequel. Here, and in the rest of the section, whenever two points A and B belong to a same given curve, we write $\widehat{AB}$ to denote the part of the curve between them, and we call it "arc between A and B".

Lemma 4.8. *Let $\boldsymbol{P} \in \mathrm{Int}(\boldsymbol{\mathcal{Q}})$, let $\sigma : [0,1] \to \mathcal{Q}$ be compatible with $\boldsymbol{P}$, and let A and B be any two points of $u^{-1}(\boldsymbol{P}) \cap \sigma$. Then, A and B are left-connectible if and only if, for every two points C, D in $\sigma \setminus \{A, B\}$, exactly one of those in the arc $\widehat{AB} \subseteq \sigma$, it is impossible to find a path in $\sigma^- \setminus u^{-1}(\boldsymbol{P})$ connecting C and D.*

Proof. Let us first assume the existence of two points C, $D \in \sigma \setminus \{A, B\}$, only one of which in the arc $\widehat{AB}$, and of a path $\tau_1 \subseteq \sigma^- \setminus u^{-1}(\boldsymbol{P})$ connecting C and D. We have to show that A and B are not left-connectible. Let τ_2 be any arc connecting C and D in σ^+, in particular we can assume that $\tau_2 \setminus \{C, D\}$ is contained in the internal part of σ^+. Putting together τ_1 and τ_2 (we can assume without loss of generality that they are both injective), we have a closed curve which does not intersect $u^{-1}(\boldsymbol{P}) \cap \sigma^-$, and by construction exactly one between A and B belongs to the internal part of this curve. As a consequence, A and B do not belong to the same connected component of $u^{-1}(\boldsymbol{P}) \cap \sigma^-$.

Conversely, assume that A and B are not left-connectible: we have to find two points C and D in $\sigma \setminus \{A, B\}$, only one of which in $\widehat{AB}$, and a path connecting C and D in $\sigma^- \setminus u^{-1}(\boldsymbol{P})$. For the sake of clarity, and without loss of generality, we can think that σ is a vertical segment. First of all notice that, since $u^{-1}(\boldsymbol{P})$ is connected, then every connected component of $u^{-1}(\boldsymbol{P}) \cap \sigma^-$ must intersect some of the finitely many points of $u^{-1}(\boldsymbol{P}) \cap \sigma$. Hence, there are finitely many connected components of the set $u^{-1}(\boldsymbol{P}) \cap \sigma$, and since $u^{-1}(\boldsymbol{P})$ is closed then every component is closed, and any two are a strictly positive distance apart. Let then Γ_A and Γ_B be the connected components of $u^{-1}(\boldsymbol{P}) \cap \sigma^-$ containing A and B respectively. Let us call A^- and A^+ the first and the last point of σ which belong to Γ_A (they could both coincide with A), and B^- and B^+ those in Γ_B: a possible situation is depicted in Figure 2. Notice that the intersection between the two curves $\widehat{A^-A^+}$ and $\widehat{B^-B^+}$ must be either one of the two curves (as in the figure), or empty. Indeed, otherwise Γ_A and Γ_B could clearly not be both connected and with no intersection.

As a consequence, without loss of generality we can assume that, as in the figure, the point B is not contained in the arc $\widehat{A^-A^+}$. Let us then apply Lemma 4.1 to Γ_A, with ε smaller than the distance between Γ_A and any other connected component of $u^{-1}(\boldsymbol{P}) \cap \sigma^-$, so to find
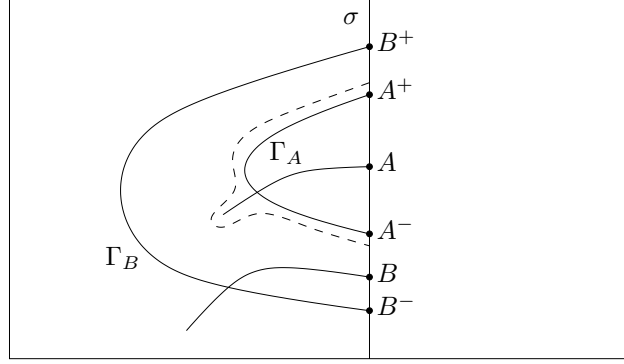
FIGURE 2. Construction for the proof of Lemma 4.8; the dashed curve is $\tilde{\theta}$.

a closed, injective curve θ, which has to intersect both σ^- and σ^+ by construction. Let P be a point of θ, in the interior of σ^-, with ordinate higher than that of any point of Γ_A. Then, let $\tilde{\theta}$ be the shortest subpath of θ containing P and having both endpoints on σ, and call C and D its endpoints, being C below D; a simple geometric argument, using the fact that $\Gamma_A \subseteq \sigma^-$ is connected, ensures that C is below A^- and D is above A^+, however the distance between C and A^-, as well as that between D and A^+, is less than ε. As a consequence, C belongs to $\widehat{AB}$ while D does not, and $\tilde{\theta}$ is a path connecting C and D in $\sigma^- \setminus u^{-1}(\boldsymbol{P})$, since $\tilde{\theta}$ does not intersect Γ_A by construction and it does not intersect any other component of $\sigma^- \cap u^{-1}(P)$, since they have all distance larger than ε from Γ_A, while each point of $\tilde{\theta}$ has distance less than σ from it. $\qquad\square$

The first use of the definition of connectibility is to decide which of the two possibilities in Lemma 4.4 holds.

Lemma 4.9. *Let $\boldsymbol{P} \in \mathrm{Int}(\boldsymbol{\mathcal{Q}})$, let $\sigma : [0,1] \to \mathcal{Q}$ be compatible with $\boldsymbol{P}$, let A and B be two connectible points of $u^{-1}(\boldsymbol{P}) \cap \sigma$, and let A^-A^+ and B^-B^+ be two short segments around A and B, contained in σ and which do not contain other points of $u^{-1}(\boldsymbol{P})$ apart from A and B. Then, there exists a curve $\gamma : [0,1] \to \mathcal{Q}$ compatible with $\boldsymbol{P}$, which intersects $u^{-1}(\boldsymbol{P})$ exactly at A and B, and which contains the segments A^-A^+ and B^-B^+.*

Proof. By Lemma 4.4, we know that the conclusion is true, up to replace the segment B^-B^+ with the segment B^+B^-. Let us then assume by contradiction the existence of a curve γ as in the claim, but with the wrongly oriented segment B^+B^-. In particular, the curve γ is done by a path γ_0 between $\partial\mathcal{Q}$ and A^-, then the segment A^-A^+, then a path γ_1 between A^+ and B^+, then the segment B^+B^-, and finally a last path γ_2 between B^- and $\partial\mathcal{Q}$. Let us assume, without loss of generality, that A and B are left-connectible, the analogous argument works if they are right-connectible. Let then Γ be the connected component of $u^{-1}(\boldsymbol{P}) \cap \sigma^-$ containing both A and B. Since it is a closed, connected set which does not disconnect $\mathcal{Q}$, by Lemma 4.2 its union Γ^+ with the segment AA^+, the curve γ_1 and the segment B^+B disconnects $\mathcal{Q}$ in two connected components $\mathcal{Q}_{\mathrm{int}}$ and $\mathcal{Q}_{\mathrm{ext}}$, where we call $\mathcal{Q}_{\mathrm{ext}}$ the one containing $\partial\mathcal{Q}$. Since the curve γ_0 does not intersect Γ^+, we deduce that $A^- \in \mathcal{Q}_{\mathrm{ext}}$, and since $\Gamma \subseteq \sigma^-$ we deduce that the right part of the curve γ_1 (that is, the points arbitrarily close to γ_1 on the right part with respect to

the vector γ_1') is done by points in $\mathcal{Q}_{\mathrm{ext}}$. In particular, the point B^- belongs to $\mathcal{Q}_{\mathrm{int}}$, which is against the existence of the curve γ_2 connecting B^- with $\partial\mathcal{Q}$ without intersecting Γ^+. $\qquad\square$

The following corollary is immediate.

Corollary 4.10. *Let $\boldsymbol{P} \in \mathrm{Int}(\boldsymbol{\mathcal{Q}})$, let $\sigma : [0,1] \to \mathcal{Q}$ be compatible with $\boldsymbol{P}$, let A and B be two connectible points of $u^{-1}(\boldsymbol{P}) \cap \sigma$. Then, the vectors $D(u \circ \sigma)(A)$ and $D(u \circ \sigma)(B)$ are parallel and with opposite directions.*

Proof. Let γ be a curve given by Lemma 4.9, and notice that since both σ and γ contain the segments A^-A^+ and B^-B^+ then

$$D(u \circ \gamma)\big(\gamma^{-1}(A)\big) = D(u \circ \sigma)\big(\sigma^{-1}(A)\big), \qquad D(u \circ \gamma)\big(\gamma^{-1}(B)\big) = D(u \circ \sigma)\big(\sigma^{-1}(B)\big).$$

Moreover, both the vectors are non-zero by definition of compatible curves, so the thesis follows by applying Lemma 4.3 to the curve γ. $\qquad\square$

Corollary 4.11. *Let $\boldsymbol{P} \in \mathrm{Int}(\boldsymbol{\mathcal{Q}})$, let $\sigma : [0,1] \to \mathcal{Q}$ be compatible with $\boldsymbol{P}$. Then, every connected component of $\sigma^- \cap u^{-1}(\boldsymbol{P})$ contains at most two points in σ. Hence, any point of $\sigma \cap u^{-1}(\boldsymbol{P})$ is left-connectible with at most one other point.*

Proof. Assume that A, B and C are three distinct points in $u^{-1}(\boldsymbol{P}) \cap \sigma$, contained in the same connected component of $\sigma^- \cap u^{-1}(\boldsymbol{P})$; then, any two of these three points are left-connectible. As a consequence, the derivatives of u in the direction of σ at the points A, B and C are three non-zero vectors, and by Corollary 4.10 any two of them have opposite orientation, which is clearly absurd. $\qquad\square$

Lemma 4.12. *Let $\boldsymbol{P} \in \mathrm{Int}(\boldsymbol{\mathcal{Q}})$, let $\sigma : [0,1] \to \mathcal{Q}$ be compatible with $\boldsymbol{P}$; then, two points in $\sigma \cap u^{-1}(\boldsymbol{P})$ cannot be both left-connectible and right-connectible.*

Proof. Assume that A and B are both left-connectible and right-connectible. Let then $C, D \in \sigma \setminus u^{-1}(\boldsymbol{P})$ be any two points, exactly one of which in the arc $\widehat{AB}$. Since $u^{-1}(\boldsymbol{P})$ does not disconnect $\mathcal{Q}$, there exists a path connecting C and D in $\mathcal{Q} \setminus u^{-1}(\boldsymbol{P})$; without loss of generality, we can assume that this path intersects σ in finitely many points $P_1 = C, P_2, \ldots, P_N = D$. The subpath between any P_i and P_{i+1} is then either entirely in σ^- or entirely in σ^+. By Lemma 4.8, keeping in mind that A and B are both left- and right-connectible, we deduce that P_{i+1} belongs to the arc $\widehat{AB}$ if and only if so does P_i. By recursion, $D = P_N$ belongs to $\widehat{AB}$ if and only if so does $C = P_1$, which is a contradiction. $\qquad\square$

Lemma 4.13. *Let $\boldsymbol{P} \in \mathrm{Int}(\boldsymbol{\mathcal{Q}})$, let $\sigma : [0,1] \to \mathcal{Q}$ be compatible with $\boldsymbol{P}$. Then, we can write $\sigma \cap u^{-1}(\boldsymbol{P}) = \{P_1, P_2, \ldots, P_N\}$, where each P_i is either left- or right-connectible with P_{i+1}, and P_1 and P_N are not connectible.*

Proof. We can assume without loss of generality that $u^{-1}(\boldsymbol{P}) \cap \sigma^-$ contains at least two points, otherwise the claim is emptily true. Let us consider the connected components of $u^{-1}(\boldsymbol{P}) \cap \sigma^-$ and of $u^{-1}(\boldsymbol{P}) \cap \sigma^+$; since $u^{-1}(\boldsymbol{P})$ is connected, each of these connected components must have

at least a point in σ. Moreover, any two of these connected components have strictly positive distance unless they have a common point in σ. Again since $u^{-1}(\boldsymbol{P})$ is connected, we deduce that any connected component can reach any other one in finitely many "steps", where any step means passing from a connected component to another one which has a common point in σ. In other words, for any two points A and B in $u^{-1}(\boldsymbol{P}) \cap \sigma$, there is a finite "chain" of points $Q_1 = A$, Q_2, ..., $Q_M = B$, where every Q_i is left- or right-connectible with Q_{i+1}.

Take now any point in $u^{-1}(\boldsymbol{P}) \cap \sigma$: for what we just said, it must be left- or right-connectible with some other point, but by Corollary 4.11 it can be left-connectible, as well as right-connectible, with at most one other point. Assume for a moment the existence of a point $P_1 \in u^{-1}(\boldsymbol{P}) \cap \sigma$ which is connectible with only one other point; let us call P_1 this point, and P_2 the point with which P_1 is connected, and let us assume that P_1 and P_2 are right-connectible (if they are left-connectible the obvious modification of the argument works). If $u^{-1}(\boldsymbol{P}) \cap \sigma = \{P_1, P_2\}$ we are done; otherwise, since P_1 is not connectible with any other point, P_2 must be connectible with some point P_3, in particular it must be left-connectible, because it is already right-connectible with P_1. Again, if we have taken all the points of $u^{-1}(\boldsymbol{P}) \cap \sigma$ we are done, otherwise P_3 must be right-connectible with some P_4 (since P_1 and P_2 already found all their connectibility points). With an obvious induction, we are done.

To conclude the proof, we must then exclude that all the points have a left-connectible and a right-connectible other point. If that was the case, we could still write $\sigma \cap u^{-1}(\boldsymbol{P}) = \{P_1, P_2, \dots, P_N\}$, where every P_i is left- (resp., right-)connectible with P_{i+1} if i is even (resp., odd); moreover, N would be surely even and P_N and P_1 would be left-connectible. For any even (resp., odd) number $1 \leq i < N$, let us call Γ_i the connected component of $u^{-1}(\boldsymbol{P}) \cap \sigma^-$ (resp., $u^{-1}(\boldsymbol{P}) \cap \sigma^+$) which contains P_i and P_{i+1}, while Γ_N is the connected component of $u^{-1}(\boldsymbol{P}) \cap \sigma^-$ containing P_N and P_1. Applying Lemma 4.9 to P_1 and P_2, we find a compatible curve γ, starting and ending in $\partial \mathcal{Q}$, which contains exactly P_1 and P_2 in $u^{-1}(\boldsymbol{P})$, and which coincides with σ in two short segments around P_1 and around P_2. Since Γ_1 is contained in σ^+ and intersects γ only at P_1 and P_2, we deduce that Γ_1 is also contained in γ^+; then, P_1 and P_2 are right-connectible also with respect to the curve γ. Similarly, since Γ_2 and Γ_N are contained in σ^-, we deduce that they are also contained in γ^-. Let us now consider any Γ_i with $3 \leq i \leq N - 1$: it has no intersection with γ, hence it is entirely contained in the interior of γ^- or in the interior of γ^+. However, since Γ_3 has one point in common with Γ_2 (that is, P_3), we deduce that Γ_3 is entirely in the interior of γ^-, and by obvious recursion the same holds true for every Γ_i with $4 \leq i \leq N - 1$. The union of all the sets Γ_i for $i \neq 1$ is then a connected set contained in γ^-: this means that P_1 and P_2 are also left-connectible with respect to γ. This is a contradiction with Lemma 4.12, so we conclude. $\qquad\square$

Notice that the order of the points P_i given by Lemma 4.13 is not necessarily the order in which points are met by σ, see Figure 3 left for an example, where again σ is depicted as a vertical segment just for the sake of clarity. Notice also that, if P_1 and P_2 are right-connectible, as in the figure, then every P_i is right-connectible with P_{i+1} when i is odd and left-connectible when i is even. The opposite clearly holds if P_1 and P_2 are left-connectible.
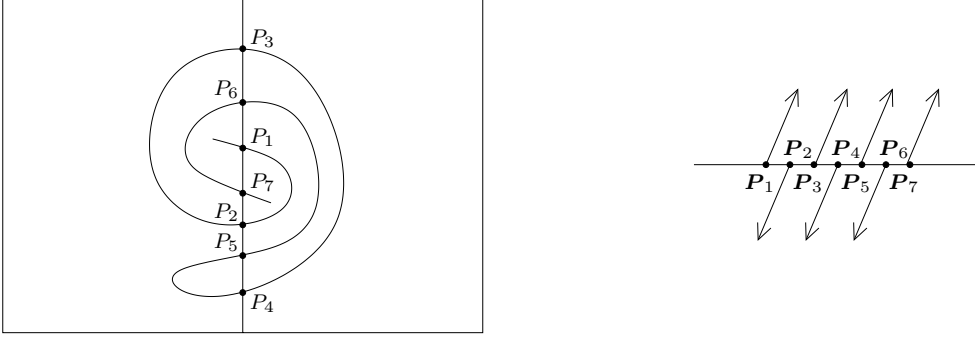
FIGURE 3. Left: the points P_1, P_2, ..., P_N in Lemma 4.13. Right: the corresponding points $\boldsymbol{P}_1$, $\boldsymbol{P}_2$, ..., $\boldsymbol{P}_N$ in Definition 4.17.

Notice that, by Lemma 4.13 and Corollary 4.10, the vectors $D(u \circ \sigma)(\sigma^{-1}(P_i))$ are all parallel. We can then conclude the section with a last definition.

Definition 4.14 (Characteristic vector). *Let $\boldsymbol{P} \in \mathrm{Int}(\boldsymbol{\mathcal{Q}})$ and let σ be a path compatible with $\boldsymbol{P}$. We call* characteristic vector of σ at $\boldsymbol{P}$ *the vector $v \in \mathbb{S}^1$, defined up to a sign, such that $D(u \circ \sigma)(t_i)$ is parallel to v for every $t_i \in (u \circ \sigma)^{-1}(\boldsymbol{P})$.*

4.3. **Separation of points on a curve.** This subsection is devoted to present a rule to "separate" the image of the points $u^{-1}(\boldsymbol{P}) \cap \sigma$, where σ is some given curve compatible with $\boldsymbol{P}$. Let us be more precise.

Definition 4.15 (Adapted curve). *Let $\widetilde{\mathscr{G}} \subseteq \boldsymbol{\mathcal{Q}}$ be a grid as in (3.2), and let $\sigma : [0,1] \to \mathcal{Q} \setminus \mathcal{S}_u$ be an admissible curve with $\sigma(0)$, $\sigma(1) \in \partial\mathcal{Q}$ and $\sigma(t) \in \mathrm{Int}(\mathcal{Q})$ for $0 < t < 1$. We say that σ is adapted to $\widetilde{\mathscr{G}}$ if $u(\sigma) \cap \widetilde{\mathscr{G}}$ is a finite set not containing any vertex of $\widetilde{\mathscr{G}}$, and for any $\boldsymbol{P} \in u(\sigma) \cap \widetilde{\mathscr{G}}$ the curve σ is compatible with $\boldsymbol{P}$, and the characteristic vector of σ at $\boldsymbol{P}$, according to Definition 4.14, is not parallel to the side of $\widetilde{\mathscr{G}}$ containing $\boldsymbol{P}$. In this case, a number $\xi > 0$ will be said small for σ if it is much smaller than the distance between any two elements of $u(\sigma) \cap \widetilde{\mathscr{G}}$, as well as between any such element and any vertex of $\widetilde{\mathscr{G}}$.*

Remark 4.16. *An obvious consequence of the above definition, together with Definitions 3.4 and 3.5, is the following. If $\widetilde{\mathscr{G}}$ is a good arrival grid associated with the good starting grid $\mathscr{G}$, then every path contained in $\mathscr{G}$ is adapted to $\widetilde{\mathscr{G}}$.*

Our aim is to define a modification of u on σ which is piecewise linear an injective. We start by defining the images of the finitely many points of $\sigma \cap u^{-1}(\widetilde{\mathscr{G}})$.

Definition 4.17. *Let $\widetilde{\mathscr{G}} \subseteq \boldsymbol{\mathcal{Q}}$ be a grid, let σ be a curve adapted to $\widetilde{\mathscr{G}}$, and let ξ be small for σ. Consider any point $\boldsymbol{P} \in u(\sigma) \cap \widetilde{\mathscr{G}}$: being σ compatible with $\boldsymbol{P}$, we write $\sigma \cap u^{-1}(\boldsymbol{P}) = \{P_1, P_2, \ldots, P_N\}$ as in Lemma 4.13. We let then $\boldsymbol{P}_1$, $\boldsymbol{P}_2$, ..., $\boldsymbol{P}_N$ be N distinct points on $\widetilde{\mathscr{G}}$, all contained within a distance ξ from $\boldsymbol{P}$ (hence they are all in the same side of $\widetilde{\mathscr{G}}$ as $\boldsymbol{P}$), and ordered as follows. Assume first that P_1 and P_2 are right-connected. If the side containing $\boldsymbol{P}$ is horizontal, then the points $\boldsymbol{P}_i$ are ordered from left to right if $D(u \circ \sigma)(\sigma^{-1}(P_1))$ has a positive vertical component (as in Figure 3 right), and from right to left otherwise. Similarly, if*

the side of P is vertical, the points P_i are ordered from top to bottom if $D(u \circ \sigma)(\sigma^{-1}(P_1))$ has a positive horizontal component, and from bottom to top otherwise. If, instead, P_1 and P_2 are left-connected, then the orientation of the points P_i is done in the opposite way.

It is important to understand the meaning of the orientation of the points P_i just defined, again a look to Figure 3 right can also help. Roughly speaking, if the point P_i is right-connected with the point P_{i+1}, then an observer sitting at P_i, looking in the direction of the curve $u \circ \sigma$ at $\sigma^{-1}(P_i)$, will have the point P_{i+1} at his right. Notice that, at the same time, also P_{i+1} is right-connected with P_i; and in fact, the point P_i is also at the right of the point P_{i+1}, from the point of view of P_{i+1}, at which the direction of $u \circ \sigma$ is opposite.

We are now ready to define the modification of u on σ. Before doing so note that, if ξ is taken sufficiently small, then the following simple geometric fact about the generalized segments introduced in Definiton 3.7 holds: let $\mathcal{R}$ be a rectangle of the grid $\widetilde{\mathcal{G}}$ and A, B, C and D be four distinct ordered points on a same side of $\partial\mathcal{R}$, then $[AB]$ and $[CD]$ have empty intersection, as well as $[AD]$ and $[BC]$, while $[AC]$ and $[BD]$ have exactly an intersection point.

Definition 4.18 (The function v_σ). *Let $\widetilde{\mathcal{G}} \subseteq \mathcal{Q}$ be a grid, let $\sigma : [0,1] \to \mathcal{Q}$ be adapted to $\widetilde{\mathcal{G}}$, and let ξ be small for σ. The function $v_\sigma : \sigma \to \mathcal{Q}$ is defined as follows. For any point $P \in u(\sigma) \cap \widetilde{\mathcal{G}}$, we define the points P_i as in Lemma 4.13 and the points P_i as in Definition 4.17, and we set $v_\sigma(P_i) = P_i$ for any i; we also set $v_\sigma = u$ at $\sigma(0)$ and $\sigma(1)$. We have then defined v_σ at all the points of $\sigma \cap u^{-1}(\widetilde{\mathcal{G}})$, and at $\sigma(0)$ and $\sigma(1)$. Let now A, B be any two consecutive points on σ on which v_σ is already defined. Notice that, by definition, the function u in the open curve $\widehat{AB}$ is contained in the interior of a square $\mathcal{R}$ of $\widetilde{\mathcal{G}}$, and $u(A)$ and $u(B)$, as well as $v_\sigma(A)$ and $v_\sigma(B)$, lie on $\partial\mathcal{R}$. We extend then v_σ on the curve $\widehat{AB}$ as the generalized segment $[AB]$, being $A = v_\sigma(s)$ and $B = v_\sigma(t)$, parametrized with constant speed.*

The main goal of this section is to show that the function v_σ, which is uniformly very close to u, is injective.

Proposition 4.19. *Let $\widetilde{\mathcal{G}}$, σ and v_σ be as in Definition 4.18. Then, the piecewise linear function $v_\sigma : \sigma \to \mathcal{Q}$ is injective.*

Since the proof of this result is quite involved, we divide it in several intermediate results. First of all, we show that the order given by Lemma 4.13 and Definition 4.17 is "well-defined", that is, if we take two different adapted curves containing the same two points in $u^{-1}(P)$ for some $P \in \widetilde{\mathcal{G}}$, the order between the images of these two points in the two corresponding functions in Definition 4.18 is the same.

Lemma 4.20. *Let $\widetilde{\mathcal{G}} \subseteq \mathcal{Q}$ be a grid as in (3.2), let $P \in \widetilde{\mathcal{G}}$, let A, $B \in u^{-1}(P)$, and let σ and $\tilde{\sigma}$ be two curves adapted to $\widetilde{\mathcal{G}}$, both containing A and B. Then, the order of the points $v_\sigma(A)$ and $v_\sigma(B)$ in the segment of $\widetilde{\mathcal{G}}$ containing P is the same as that of $v_{\tilde{\sigma}}(A)$ and $v_{\tilde{\sigma}}(B)$, where v_σ and $v_{\tilde{\sigma}}$ are as in Definition 4.18.*

Proof. Since σ is adapted to $\widetilde{\mathcal{G}}$, it is compatible with P and we can apply Lemma 4.13 to σ so to write $u^{-1}(P) = \{P_1, P_2, \ldots, P_N\}$. Up to exchange A and B, there are $1 \le i < j \le N$ such

that $A = P_i$ and $B = P_j$. Just to fix the ideas, we assume that P_i is right connected with P_{i+1} (of course, the symmetric argument works in the opposite case). Then, by Definition 4.17 we know that $v_\sigma(B)$ is on the "right" of $v_\sigma(A)$, where the right, in the segment of $\widetilde{\mathscr{G}}$ containing $\boldsymbol{P}$, is intended with respect to $\omega = D(u \circ \sigma)(\sigma^{-1}(A))$. For every $1 \le l \le N - 1$ such that $i - l$ is even (resp., odd), let us call Γ_l the connected component of $u^{-1}(\boldsymbol{P}) \cap \sigma^+$ (resp., $u^{-1}(\boldsymbol{P}) \cap \sigma^-$) which contains P_l and P_{l+1}.

Let now $\gamma : [0, 1] \to \mathcal{Q}$ be a curve, compatible with $\boldsymbol{P}$, which intersects $u^{-1}(\boldsymbol{P})$ exactly at A and B (such a curve surely exists by Lemma 4.4). Up to reverse the orientation of γ, we can assume that $\Gamma_i \subseteq \gamma^+$. But then, the set $\mathcal{U} = \Gamma_i \cup \Gamma_{i+1} \cup \cdots \cup \Gamma_{j-1}$ is a connected subset of $u^{-1}(\boldsymbol{P})$, which intersects γ exactly at A and B; since Γ_i is contained in γ^+, then the same is true for the whole $\mathcal{U}$, which proves that A and B are right-connectible with respect to γ. Then, defining v_γ according to Definition 4.18, we get that $v_\gamma(B)$ is on the right of $v_\gamma(A)$: notice that, this time, the "right" on the segment of $\widetilde{\mathscr{G}}$ containing $\boldsymbol{P}$ must be intended with respect to $\omega' = D(u \circ \gamma)(\gamma^{-1}(A))$, not with respect to ω. Nevertheless, we can readily observe that the two things are equivalent; in fact, we can show that the non-zero vectors ω and ω' are parallel and with the same direction. Indeed, by Lemma 4.6, there is a direction $\nu \in \mathbb{S}^1$ such that $Du(A)\nu = 0$, and points of $u^{-1}(\boldsymbol{P})$ close to A have direction, with respect to A, converging to $\pm\nu$. As a consequence, calling $\nu^\perp$ a vector orthogonal to ν, and calling $\alpha = Du(A)\nu^\perp \neq 0$, we have that

$$\omega = D(u \circ \sigma)(\sigma^{-1}(A)) = Du(A)\sigma'(\sigma^{-1}(A)) = \alpha \sigma'(\sigma^{-1}(A)) \cdot \nu^\perp, \quad \omega' = \alpha \gamma'(\gamma^{-1}(A)) \cdot \nu^\perp.$$

And in turn, the scalar products of $\sigma'(\sigma^{-1}(A))$ and $\gamma'(\gamma^{-1}(A))$ with $\nu^\perp$ have the same sign, since Γ_i' is both in σ^+ and in γ^+.

Summarizing, we have proved that the order of the images of A and B under v_σ is the same as under v_γ, where γ is any curve, compatible with $\boldsymbol{P}$, intersecting $u^{-1}(\boldsymbol{P})$ exactly at A and B. But then, the very same argument shows that also the order of the images of A and B under $v_{\widetilde{\sigma}}$ is the same as under v_γ, hence in turn also as under v_σ. The proof is then concluded. $\qquad \square$

We can now show that the orientation of points on the boundary of squares of $\widetilde{\mathscr{G}}$ is consistent, in the following sense.

Lemma 4.21. *Let $\widetilde{\mathscr{G}} \subseteq \mathcal{Q}$ be a grid, let $\sigma : [0, 1] \to \mathcal{Q}$ be adapted to $\widetilde{\mathscr{G}}$, let $\mathcal{R}$ be a rectangle of the grid $\widetilde{\mathscr{G}}$, and let A, B, C, D be four distinct points on σ such that the arcs $\widehat{AB}$ and $\widehat{CD}$ are disjoint, and their images under u are in $\mathrm{Int}(\mathcal{R})$ except the four points $\boldsymbol{A} = u(A)$, $\boldsymbol{B} = u(B)$, $\boldsymbol{C} = u(C)$, $\boldsymbol{D} = u(D)$ on $\partial\mathcal{R}$. If $\boldsymbol{A} \neq \boldsymbol{B}$, then the points $\boldsymbol{C}$ and $\boldsymbol{D}$ cannot belong to the two different open paths in which $\partial\mathcal{R}$ is divided by $\boldsymbol{A}$ and $\boldsymbol{B}$.*

Proof. Since σ is a segment near the four points A, B, C and D, then it contains four small disjoint segments $A_{\mathrm{int}}A_{\mathrm{ext}}$, $B_{\mathrm{int}}B_{\mathrm{ext}}$, $C_{\mathrm{int}}C_{\mathrm{ext}}$ and $D_{\mathrm{int}}D_{\mathrm{ext}}$ around them, and the image of these segments under u crosses $\partial\mathcal{R}$ at $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$ and $\boldsymbol{D}$ respectively, being $u(X_{\mathrm{int}})$ inside $\mathcal{R}$ and $u(X_{\mathrm{ext}})$ outside $\mathcal{R}$ for each $X \in \{A, B, C, D\}$. By assumption, the set $\mathcal{D} = u^{-1}(\mathcal{R})$ is a closed, connected subset of $\mathcal{Q}$ which does not disconnect $\mathcal{Q}$, hence we can find two injective

and continuous paths γ_1, $\gamma_2 : [0,1] \to \mathcal{Q} \setminus \mathcal{D}$ connecting B_{ext} with C_{ext}, and D_{ext} with A_{ext} respectively. Arguing as in Lemma 4.4, we can assume that γ_1 and γ_2 are disjoint, up to exchange C with D; we can clearly also assume without loss of generality that γ_1 and γ_2 do not intersect the segments $X_{\text{ext}}X$ for $X \in \{A, B, C, D\}$, nor the set $\mathcal{S}_u$ (keep in mind that $\mathcal{S}_u$ is $\mathcal{H}^1$-negligible). Since the arcs $\widehat{AB}$ and $\widehat{CD}$ are disjoint, and by assumption they belong to $\mathcal{D}$, then the closed curve $\gamma : \mathbb{S}^1 \to \mathcal{Q} \setminus \mathcal{S}_u$ obtained joining $\widehat{A_{\text{ext}}B_{\text{ext}}}$ with γ_1, then $\widehat{C_{\text{ext}}D_{\text{ext}}}$ and then γ_2 is injective. As a consequence, by Lemma 2.8, for every point of $\mathcal{Q} \setminus u(\gamma)$ the degree with respect to $u(\gamma)$ must be either 0 or 1.

Notice now that $u(\gamma) \cap \partial\mathcal{R}$ consists precisely of the points A, B, C and D. Therefore, since $u \circ \gamma$ is exiting from $\partial\mathcal{R}$ at B, we have that

$$\deg(B^+, u(\gamma)) = \deg(B^-, u(\gamma)) - 1\,, \tag{4.3}$$

where B^+ and B^- are two points on $\partial\mathcal{R}$, very close to B, and in such a way that B^+ and B^- are respectively on the right and of the left of B, with respect to the direction of $u \circ \gamma$ at B. Similarly, since $u \circ \gamma$ is exiting from $\partial\mathcal{R}$ also at D, calling D^+ and D^- two points on $\partial\mathcal{R}$, close to D and respectively at its right and at its left, we have that

$$\deg(D^+, u(\gamma)) = \deg(D^-, u(\gamma)) - 1\,. \tag{4.4}$$

Let us now assume that C and D belong to the two different open paths in which $\partial\mathcal{R}$ is divided by A and B. As a consequence, there is a path θ in $\partial\mathcal{R}$ connecting B and D without intersecting A and D; then, θ must contain either both B^+ and D^-, or both B^- and D^+. In the first case, the part of θ between B^+ and D^- is entirely contained in $\mathcal{Q} \setminus u(\gamma)$, hence $\deg(B^+, u(\gamma)) = \deg(D^-, u(\gamma))$, so by (4.3) and (4.4) we find $\deg(B^-, u(\gamma)) = \deg(D^+, u(\gamma)) + 2$, which is absurd because the degrees can only be 0 or 1 (or only 0 and -1) by the INV condition. In the second case, the analogous argument shows $\deg(B^+, u(\gamma)) = \deg(D^-, u(\gamma)) - 2$, so again a contradiction. $\qquad\square$

The above lemma says, roughly speaking, that the orientation of the points on the boundaries of the squares of $\widetilde{\mathscr{G}}$ is "correct"; nevertheless, the lemma is useless if the images of some of the points coincide, which is in fact the case that we have to deal with more care. The next result can also treat the case when two points have the same image, if they are connectible.

Lemma 4.22. *Let $\widetilde{\mathscr{G}} \subseteq \mathcal{Q}$ be a grid, $\sigma : [0,1] \to \mathcal{Q}$ be adapted to $\widetilde{\mathscr{G}}$, $R \in \widetilde{\mathscr{G}}$, $P, Q \in u^{-1}(R) \cap \sigma$, and assume that P and Q are left-connectible. Since $D(u \circ \sigma)$ is opposite at $\sigma^{-1}(P)$ and $\sigma^{-1}(Q)$, let $\mathcal{R}$ be the rectangle of $\widetilde{\mathscr{G}}$ in which $u \circ \sigma$ is entering at $\sigma^{-1}(P)$ and from which it is exiting at $\sigma^{-1}(Q)$, and let P^+ (resp., Q^-) be the first point on σ after P (resp., the last point before Q) at which $u \circ \sigma$ is exiting from $\mathcal{R}$ (resp., entering in $\mathcal{R}$). If the three points $P^+ = u(P^+)$, $Q^- = u(Q^-)$ and R on $\partial\mathcal{R}$ are distinct, then the path on $\partial\mathcal{R}$ connecting R and P^+ without intersecting Q^- starts at R towards right, with respect to $D(u \circ \sigma)(\sigma^{-1}(P))$.*

Proof. The geometrical meaning of the claim might seem a bit obscure, but it becomes quite evident with the help of a figure. Figure 4 (right) shows the rectangle $\mathcal{R}$, with the points R, P^+ and Q^-. Notice that, since $u \circ \sigma$ is entering in $\mathcal{R}$ at $\sigma^{-1}(P)$, then the vector $D(u \circ \sigma)(\sigma^{-1}(P))$

has a positive vertical component if, as in the figure, we assume just to fix the ideas that $\boldsymbol{R}$ is on the bottom side of $\boldsymbol{\mathcal{R}}$. In the situation of the figure, the path connecting $\boldsymbol{R}$ with $\boldsymbol{P^+}$ on $\partial\boldsymbol{\mathcal{R}}$ without intersecting $\boldsymbol{Q^-}$ starts at $\boldsymbol{R}$ towards *left*: then, the claim of the lemma says that the situation of the figure is impossible, because the order of the points $\boldsymbol{P^+}$ and $\boldsymbol{Q^-}$ should be the opposite. Notice that, since P and Q are left-connectible, then the function v_σ of Definition 4.18 will have $v_\sigma(Q)$ on the left of $v_\sigma(P)$, then in the situation of the figure the map v_σ would not be injective: this perfectly clarifies the importance of the present result. For the sake of clarity, we divide the proof in two steps.
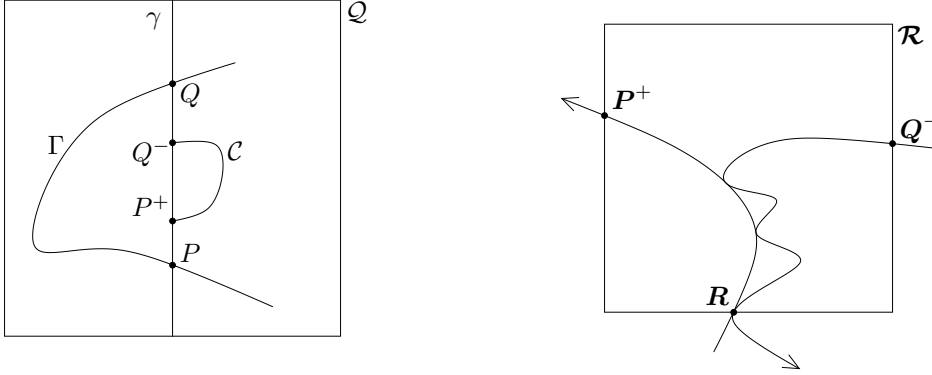

FIGURE 4. The situation in Lemma 4.22.

<u>*Step I. There is an admissible, counterclockwise curve $\gamma : \mathbb{S}^1 \to \mathcal{Q} \setminus \mathcal{S}_u$, containing the arcs $\widehat{PP^+}$*</u> *and $\widehat{Q^-Q}$ of σ, being a segment around P, P^+, Q^- and Q, and for which $\gamma^{-1}(\mathcal{R}) = \widehat{PP^+} \cup \widehat{Q^-Q}$.* Since, by assumption, $u^{-1}(\mathcal{R})$ is a closed, connected set which does not disconnect $\mathcal{Q}$, as usual we can find two disjoint and admissible paths γ_0 and γ_2 which connect $\partial\mathcal{Q}$ with P and Q with $\partial\mathcal{Q}$ respectively, which are contained in $\mathrm{Int}(\mathcal{Q}) \setminus u^{-1}(\mathcal{R})$ except at their endpoints, which do not intersect $\mathcal{S}_u$, and which coincide with the two segments of σ near P and Q. We claim that there exists also a path γ_1 connecting P^+ and Q^-, contained in $\mathcal{Q} \setminus u^{-1}(\mathcal{R})$ except at P^+ and Q^-, not intersecting γ_0 and γ_2, and coinciding with the segments of σ near P^+ and Q^-. The claim will immediately conclude the step, being γ simply the curve obtained putting together γ_0, $\widehat{PP^+}$, γ_1, $\widehat{Q^-Q}$, γ_2, and a part of $\partial\mathcal{Q}$ connecting the last point of γ_2 with the first point of γ_0, since by definition the arcs $\widehat{PP^+}$ and $\widehat{Q^-Q}$ are contained in $u^{-1}(\mathcal{R})$. Notice that there are two paths in $\partial\mathcal{Q}$ which connect the last point of γ_2 with the first one of γ_0, we have to choose the "left" one, so that γ is percurred in the counterclockwise sense.

Let us now show the claim: if it were false, then arguing again as in Lemma 4.4 we could find two disjoint paths $\tilde{\gamma}_1$ and $\tilde{\gamma}_2$, connecting respectively P^+ and Q, and Q^- and $\partial\mathcal{Q}$, being in $\mathcal{Q} \setminus (u^{-1}(\mathcal{R}) \cup \gamma_0)$ except at P^+, Q^- and Q. In this case, we could define $\tilde{\gamma}$ putting together γ_0, $\widehat{PP^+}$, $\tilde{\gamma}_1$, $\widehat{QQ^-}$ and $\tilde{\gamma}_2$. This path would intersect $u^{-1}(\boldsymbol{R})$ exactly at P and Q, which would then be connectible for $\tilde{\gamma}$, so the vectors $D(u \circ \tilde{\gamma})$ at $\tilde{\gamma}^{-1}(P)$ and $\tilde{\gamma}^{-1}(Q)$ would be parallel and with opposite directions by Lemma 4.10. On the other hand, also the vectors $D(u \circ \sigma)$ at $\sigma^{-1}(P)$ and $\sigma^{-1}(Q)$ must be parallel and with opposite directions for the same reason, and this brings to a contradiction because the directions of $D(u \circ \tilde{\gamma})(\tilde{\gamma}^{-1}(P))$ and $D(u \circ \sigma)(\sigma^{-1}(P))$ are the same,

while the directions of $D(u \circ \tilde{\gamma})(\tilde{\gamma}^{-1}(Q))$ and $D(u \circ \sigma)(\sigma^{-1}(Q))$ are opposite because $\tilde{\gamma}$ contains the curve $\widehat{QQ^-}$ instead of $\widehat{Q^-Q}$.

Step II. Conclusion.

We will work with the curve γ given by Step I, instead of the original curve σ; this curve is depicted in Figure 4 (left): as usual, just for the sake of clarity, we draw $\gamma \cap \mathrm{Int}(\mathcal{Q})$ as a vertical segment. Let us suppose by contradiction that the claim is false, hence the relative position of the points $\boldsymbol{P}^+$ and $\boldsymbol{Q}^-$ is as in Figure 4 (right). Then, we call $\boldsymbol{\mathcal{C}}$ the closed arc of $\partial \boldsymbol{\mathcal{R}}$ connecting $\boldsymbol{P}^+$ and $\boldsymbol{Q}^-$ without containing $\boldsymbol{R}$. By assumption, the set $\mathcal{C} = u^{-1}(\boldsymbol{\mathcal{C}})$ is a closed, connected subset of $\mathcal{Q}$ which does not disconnect it. Moreover, by construction $\mathcal{C}$ intersects γ precisely at P^+ and Q^-. Since $u \circ \gamma$ is exiting from $\boldsymbol{\mathcal{R}}$ at P^+ and entering in $\boldsymbol{\mathcal{R}}$ at Q^-, the points of $\boldsymbol{\mathcal{C}} \setminus \{\boldsymbol{P}^+, \boldsymbol{Q}^-\}$ have all degree 0 with respect to the curve γ, hence the set $\mathcal{C}$ is contained in γ^+. Summarizing, $\mathcal{C}$ is a connected set in γ^+ which connects P^+ and Q^-, while by assumption there is a connected component Γ of $u^{-1}(\boldsymbol{R}) \cap \gamma^-$ which connects P and Q. Let us now consider $u^{-1}(\boldsymbol{\mathcal{R}})$: this is a connected subset of $\mathcal{Q}$, which does not disconnect it, and by construction it contains $\Gamma \subseteq \gamma^-$, $\mathcal{C} \subseteq \gamma^+$, and the two arcs $\widehat{PP^+}$ and $\widehat{Q^-Q}$. Therefore, it must contain the whole arc $\widehat{P^+Q^-}$ of γ, which is absurd since the curve γ is outside $\boldsymbol{\mathcal{R}}$ between P^+ and Q^-. $\square$

We are now ready to show Proposition 4.19.

Proof of Proposition 4.19. The function v_σ is piecewise linear; in particular, each linear piece connects two points of the grid $\widetilde{\mathscr{G}}$ (with a small abouse of notation, we refer to each generalized segment as a "linear piece", even if it can be actually either one or two linear pieces). Moreover, by construction the points of v_σ which are on $\widetilde{\mathscr{G}}$ are finitely many and disjoint. As a consequence, assuming by contradiction that v_σ is not injective, there must exist two disjoint arcs $\widehat{AB}$ and $\widehat{CD}$ on σ such that $v_\sigma(\widehat{AB})$ and $v_\sigma(\widehat{CD})$ have a non-empty intersection. In particular, there exists a rectangle $\boldsymbol{\mathcal{R}}$ of the grid $\widetilde{\mathscr{G}}$ which contains the whole linear pieces $v_\sigma(\widehat{AB})$ and $v_\sigma(\widehat{CD})$. Let us call $\boldsymbol{A} = u(A)$, $\boldsymbol{B} = u(B)$, $\boldsymbol{C} = u(C)$ and $\boldsymbol{D} = u(D)$: notice that these points do not necessarily coincide with $v_\sigma(A)$, $v_\sigma(B)$, $v_\sigma(C)$ and $v_\sigma(D)$, but they are very close to them.

Since $u^{-1}(\boldsymbol{\mathcal{R}})$ is a closed, connected set which does not disconnect $\mathcal{Q}$, as already done several times we can find an admissible path $\gamma : [0,1] \to \mathcal{Q}$ which contains the arc $\widehat{AB}$ and either the arc $\widehat{CD}$ or the arc $\widehat{DC}$, which coincides with σ on four segments around the points A, B, C and D, and such that $u^{-1}(\boldsymbol{\mathcal{R}})$ consists precisely of the arcs $\widehat{AB}$ and $\widehat{CD}$ (or $\widehat{DC}$). Notice that it is admissible to assume that γ is also adapted to $\widetilde{\mathscr{G}}$; in fact, we are only interested in the behaviour of σ, or γ, around $\boldsymbol{\mathcal{R}}$. Thanks to Lemma 4.20, the two generalized segments $v_\sigma(\widehat{AB})$ and $v_\sigma(\widehat{CD})$ have a non-empty intersection if and only if the same happens to the two generalized segments $v_\gamma(\widehat{AB})$ and $v_\gamma(\widehat{CD})$ (and the fact whether γ contains the arc $\widehat{CD}$ or $\widehat{DC}$ has no influence at all). Hence, it is not restrictive to assume directly that $u^{-1}(\boldsymbol{\mathcal{R}}) \cap \sigma$ consists precisely of the arcs $\widehat{AB}$ and $\widehat{CD}$, and we will do so. We will consider separately the possible cases of the relative positions of the points $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$ and $\boldsymbol{D}$.

Case I. One has $\{\boldsymbol{A}, \boldsymbol{B}\} \cap \{\boldsymbol{C}, \boldsymbol{D}\} = \emptyset$.

We start by considering the case in which each of the points $\boldsymbol{A}$ and $\boldsymbol{B}$ is different from both

C and D. If the four points A, B, C and D are different, then by Lemma 4.21 there are two disjoint paths on $\partial\mathcal{R}$ which connect A with B, and C and D respectively; since the four points are distinct, by construction the same is true for the points $v_\sigma(A)$, $v_\sigma(B)$, $v_\sigma(C)$ and $v_\sigma(D)$. Keeping in mind Definition 3.7, we obtain that the two generalized segments $v_\sigma(\widehat{AB})$ and $v_\sigma(\widehat{CD})$ have no intersection, against the assumption. To conclude this case, we assume then that the four points are not distinct. In view of the assumption $\{A, B\} \cap \{C, D\} = \emptyset$, this means that either $A = B$, or $C = D$, or both. Suppose for instance that $A = B$; then, the generalized segment $v_\sigma(\widehat{AB})$ is extremely close to the point $A = B$; instead, the points C and D are both different from A, hence the generalized segment $v_\sigma(\widehat{CD})$ is away from the point A. This shows again that the two generalized segments are disjoint, against the assumption; hence, this case is concluded.

As a consequence, we can assume that $\{A, B\} \cap \{C, D\} \neq \emptyset$. Keep in mind that, since σ intersects $\mathcal{R}$ only in the two arcs $\widehat{AB}$ and $\widehat{CD}$, then the curve $u \circ \sigma$ enters in $\mathcal{R}$ at A and exits at B, then enters again at C and exits at D; this shows that necessarily $A = D$ or $B = C$ (or both). In fact, the case $A = C \notin \{B, D\}$ would contradict Lemma 4.3, as well as $B = D \notin \{A, D\}$. Therefore, from now on we will assume that $A = D$, being the case $B = C$ completely analogous.

Case II. One has $A = D$, and the points B and C are different, and different from A.
In this case, the points $v_\sigma(A)$ and $v_\sigma(D)$ are very close to A and to each other, while $v_\sigma(B)$ and $v_\sigma(C)$ are close to B and C, so away from each other and from A. Since $u^{-1}(A) \cap \sigma$ consists precisely of A and D, the two points must be connectible; let us assume, just to fix the ideas, that they are left-connectible. We have then to show that the path in $\mathcal{R}$ connecting A and B without containing C starts at A towards right, with respect to the direction of $D(u \circ \sigma)$ at $\sigma^{-1}(A)$. And in fact, this is precisely given by Lemma 4.22.

Case III. One has $A = D$, $C = B$ and $A \neq B$.
By construction, $u^{-1}(A) \cap \sigma$ consists only of A and D, while $u^{-1}(B)$ only of B and C. As a consequence, A and D are connectible in $u^{-1}(A)$, as well as B and C in $u^{-1}(B)$; just to fix the ideas, assume that A and D are right-connectible. Notice that the non-emptiness of the intersection between the two generalized segments $v_\sigma(\widehat{AB})$ and $v_\sigma(\widehat{CD})$ implies that B and C are left-connectible. As a consequence, there is a connected set $\Gamma_1 \subseteq u^{-1}(A) \cap \sigma^+$ which contains both A and D, and a connected set $\Gamma_2 \subseteq u^{-1}(B) \cap \sigma^-$ which contains both B and C. Let us now consider the connected set $u^{-1}(\mathcal{R}) \subseteq \mathcal{Q}$, which does not disconnect $\mathcal{Q}$: by construction, it contains Γ_1 and Γ_2, as well as both the arcs $\widehat{AB}$ and $\widehat{CD}$. As a consequence, it must also contain the whole arc $\widehat{BC}$, which is absurd because σ exits from $\mathcal{R}$ at B.

Case IV. One has $A = D = B \neq C$, or $A = D = C \neq B$.
Assume that $A = D = B \neq C$, the other case being fully analogous. The non-emptiness of the intersection between the two generalized segments $v_\sigma(\widehat{AB})$ and $v_\sigma(\widehat{CD})$ means that $v_\sigma(D)$ is between $v_\sigma(A)$ and $v_\sigma(B)$; since $u^{-1}(A) \cap \sigma$ consists precisely of A, B and D, this means that D is left connected, in $u^{-1}(A)$, with one between A and B, and right connected with the other one. Again, this means that there exists a connected set $\Gamma_1 \subseteq u^{-1}(A) \cap \sigma^-$ which contains D

and one between A and B, and a connected set $\Gamma_2 \subseteq u^{-1}(\boldsymbol{A}) \cap \sigma^+$ which contains D and the other one. Since $u^{-1}(\boldsymbol{\mathcal{R}})$ does not disconnect $\mathcal{Q}$ and contains both Γ_1 and Γ_2, as well as the arc $\widehat{AB}$, this means again that the arc $\widehat{BC}$ is contained in $u^{-1}(\boldsymbol{\mathcal{R}})$, a contradiction.

<u>Case V.</u> *One has* $\boldsymbol{A} = \boldsymbol{D} = \boldsymbol{B} = \boldsymbol{C}$.

This final case is very similar to the last ones. The non-emptiness of $v_\sigma(\widehat{AB}) \cap v_\sigma(\widehat{CD})$ implies, this time, that the two segments $v_\sigma(A)v_\sigma(B)$ and $v_\sigma(C)v_\sigma(D)$ are not disjoint neither contained one into the other. Hence, up to exchange the roles of the points, we can assume that $v_\sigma(D)$ is between $v_\sigma(A)$ and $v_\sigma(B)$, while $v_\sigma(C)$ is not between them. By definition, this means that D is left-connectible with one between A and B, and right-connectible with the other one: thus, we find a contradiction exactly as in Case IV. $\qquad\square$
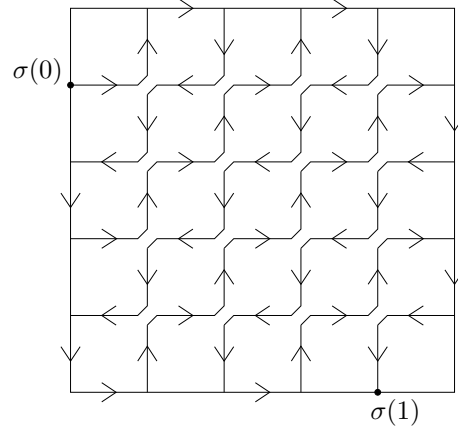
4.4. **Proof of Theorem C.**

This last subsection is devoted to show Theorem C, which now comes as a consequence of Proposition 4.19.

Proof of Theorem C. Let us call $\mathcal{S}_u$ the set of the points at which u is not continuous. Since u is INV by Lemma 2.8, then we know that $\mathcal{S}_u$ is $\mathcal{H}^1$-negligible. Let then γ_i be curves as in Definition 1.2. By a trivial induction argument, up to a diffeomorphisms of $\mathcal{Q}$ onto itself, coinciding with the identity on the boundary, we can assume that the union of the curves γ_i is contained in some grid $\mathscr{G} = \mathscr{G}(K)$. As a consequence, to show the result it is enough to show that, if the grid $\mathscr{G}$ is contained in $\mathcal{Q} \setminus \mathcal{S}_u$, then there exists an injective function $v : \mathscr{G} \to \mathcal{Q}$ such that $\|u - v\|_{L^\infty(\mathscr{G})} < \varepsilon$. Since u is continuous on $\mathscr{G}$, it is not restrictive to assume that $\mathscr{G}$ is a good starting grid.

We can now apply Lemma 3.6 to obtain a good arrival grid $\widetilde{\mathscr{G}}$ with side-length $\eta < \varepsilon/(2\sqrt{2})$; keep in mind that, as observed in Remark 4.16, any path contained in $\mathscr{G}$ is adapted to $\widetilde{\mathscr{G}}$. Since for any vertex $V_{ij} = (i/K, j/K)$ of the grid $\mathscr{G}$ we have that $u(V_{ij})$ is in the interior of some square of $\widetilde{\mathscr{G}}$, and since u is continuous at V_{ij}, we can define an injective path $\sigma : [0,1] \to \mathcal{Q}$, adapted to $\widetilde{\mathscr{G}}$, as in Figure 5. More precisely, we select very small balls B_{ij} around the vertices V_{ij}, so that $u(B_{ij})$ is contained in the interior of the same square as $u(V_{ij})$, and the image of σ consists of the whole grid $\mathscr{G}$ outside the union of the balls B_{ij}. Inside each ball, instead, σ is the union of two disconnected piecewise linear paths joining the west and the north pole, and the south and the east pole respectively.

Let us now call P_l, with $1 \le l \le N$, the finitely many points of $\mathscr{G}$ contained in $u^{-1}(\widetilde{\mathscr{G}})$: by construction, P_l are exactly the points of $\sigma \cap u^{-1}(\widetilde{\mathscr{G}})$. We apply Proposition 4.19, finding the points $\boldsymbol{P}_l$ on $\widetilde{\mathscr{G}}$ such that the function v_σ of Definition 4.18 is injective. Our function $v : \mathscr{G} \to \mathcal{Q}$ will be such that $v(P_l) = \boldsymbol{P}_l$ for every $1 \le l \le N$.

Let us be more precise. We introduce the finite set $\mathcal{A} = \bigcup_{l=1}^N P_l \cup \bigcup_{i,j=0}^K V_{ij}$. Notice that $\mathscr{G}$ is a finite union of segments, each having both endpoints in $\mathcal{A}$ and no interior point in $\mathcal{A}$. Our aim is to select suitable points $\boldsymbol{V}_{ij} \in \mathcal{Q}$, for $0 \le i, j \le K$, and to define $v : \mathscr{G} \to \mathcal{Q}$ as the function which is linear on each of the segments defined above, and such that $v(P_l) = \boldsymbol{P}_l$ and $v(V_{ij}) = \boldsymbol{V}_{ij}$ for every admissible l, i, j.

FIGURE 5. The function σ in the proof of Theorem C.

Of course, we only have to define the image of the points V_{ij} for $1 \leq i, j \leq K-1$, because for the other points it must necessarily be $\boldsymbol{V}_{ij} = (i/K, j/K)$. Let then V_{ij} be any of these points, and notice that $u(V_{ij})$ is in the interior of some rectangle $\boldsymbol{\mathcal{R}}$ of the grid $\widetilde{\mathscr{G}}$. We will define $\boldsymbol{V}_{ij}$ as a suitable point in the interior of the same rectangle: notice that, as a consequence, for every $P \in \mathscr{G}$ the point $v(P)$ will be in the same rectangle of $\widetilde{\mathscr{G}}$ as $u(P)$: hence, $\|v - u\|_{L^\infty(\mathscr{G})} \leq 2\sqrt{2}\eta < \varepsilon$. Therefore, to conclude we only have to find points $\boldsymbol{V}_{ij}$ so that the resulting function v is injective.

Let us now fix $1 \leq i, j \leq K - 1$. Let AB (resp., CD) be the smallest horizontal (resp., vertical) segment containing V_{ij}, and such that A and B (resp., C and D) are of the form P_l. Since the injective function v_σ does not intersect $\widetilde{\mathscr{G}}$ in the union of the open segments AB and CD, then $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$ and $\boldsymbol{D}$ are four distinct points on the boundary of the same rectangle $\boldsymbol{\mathcal{R}}$ of the grid $\widetilde{\mathscr{G}}$ containing $u(V_{ij})$. We state what follows.

Claim: *The points $\boldsymbol{C}$ and $\boldsymbol{D}$ belong to the two different parts in which $\partial\boldsymbol{\mathcal{R}}$ is subdivided by $\boldsymbol{A}$ and $\boldsymbol{B}$.*

Indeed, there exists a path $\tilde{\sigma}$, adapted to $\widetilde{\mathscr{G}}$, which contains the four points A, B, C and D, and for which A and C, as well as B and D, are consecutive points of $\tilde{\sigma} \cap u^{-1}(\widetilde{\mathscr{G}})$: notice that we can simply take $\tilde{\sigma} = \sigma$ if the four points A, B, C and D belong to the four sides of $\mathscr{G}$ starting at V_{ij}, but also otherwise the existence of $\tilde{\sigma}$ is obvious. As a consequence, we have an injective function $v_{\tilde{\sigma}} : \tilde{\sigma} \to \boldsymbol{\mathcal{Q}}$ from Proposition 4.19: if we call $\widetilde{\boldsymbol{A}}$, $\widetilde{\boldsymbol{B}}$, $\widetilde{\boldsymbol{C}}$ and $\widetilde{\boldsymbol{D}}$ the images of A, B, C and D under $v_{\tilde{\sigma}}$, by injectivity we know that the generalized segments $[\widetilde{\boldsymbol{A}}\widetilde{\boldsymbol{C}}]$ and $[\widetilde{\boldsymbol{B}}\widetilde{\boldsymbol{D}}]$ do not intersect themselves, so by Lemma 4.20 we have that also the generalized segments $[\boldsymbol{A}\boldsymbol{C}]$ and $[\boldsymbol{B}\boldsymbol{D}]$ do not intersect themselves. Analogously, there exists another path adapted to $\widetilde{\mathscr{G}}$ containing the four points, and such that A and D, as well as B and C, are consecutive points on the path in $u^{-1}(\widetilde{\mathscr{G}})$. Arguing as before, this implies that also $[\boldsymbol{A}\boldsymbol{D}]$ and $[\boldsymbol{B}\boldsymbol{C}]$ have no intersection, and finally, this concludes the claim.

Thanks to the above claim, we know that the generalized segments $[\boldsymbol{A}\boldsymbol{B}]$ and $[\boldsymbol{C}\boldsymbol{D}]$ have exactly a point of intersection, which is in the interior of $\boldsymbol{\mathcal{R}}$. We let $\boldsymbol{V}_{ij}$ be this point, hence the function $v : \mathscr{G} \to \boldsymbol{\mathcal{Q}}$ is now defined. Notice that, by construction, if $AB \subseteq \mathscr{G}$ is a segment

with both endpoints and no interior point of the form P_l, then the image of the open segment AB under v coincides with the open generalized segment $[\boldsymbol{AB}]$; up to now, we still do not know that v is injective on AB. To conclude the proof, we have only to check that v is injective.

Keep in mind that $\mathscr{G}$ is a finite union of essentially disjoint segments having both endpoints and no interior point in $\mathcal{A}$, and that v is linear on each of these pieces. We take now any two of these segments which are disjoint, call them XY and ZW: we have to show that $v(XY) \cap v(ZW) = \emptyset$. Similarly to what already done above, let us call AB the smallest segment containing XY with both endpoints of the form P_l, and CD the one corresponding to ZW. Notice that AB (resp., CD) can coincide with XY (resp., ZW) or be larger, and even if XY and ZW have no intersection, the segments AB and CD could have a non-empty intersection, and even coincide. Let us consider the four different possible cases for AB and CD.

Case I. The segments AB and CD are disjoint.

Suppore first that the two segments AB and CD have empty intersection; in this case, we will prove that $v(AB) \cap v(CD) = \emptyset$, which of course implies $v(XY) \cap v(ZW) = \emptyset$. As noticed above, $v(AB)$ and $v(CD)$ are the (generalized) segments connecting $\boldsymbol{A} = v(A)$ and $\boldsymbol{B} = v(B)$, and $\boldsymbol{C} = v(C)$ and $\boldsymbol{D} = v(D)$ respectively. We can assume that these two segments are in the same rectangle $\boldsymbol{\mathcal{R}}$ of the grid $\widetilde{\mathscr{G}}$, since otherwise there is nothing to prove; we have only to check that $\boldsymbol{C}$ and $\boldsymbol{D}$ belong to the same of the two parts in which $\partial\boldsymbol{\mathcal{R}}$ is divided by $\boldsymbol{A}$ and $\boldsymbol{B}$. Again by Lemma 4.20, this follows by the existence of a path $\tilde{\sigma}$, adapted to $\widetilde{\mathscr{G}}$, which contains the four points A, B, C and D, and for which A and B are two consecutive points on σ of $u^{-1}(\widetilde{\mathscr{G}})$, as well as C and D; and in turn, the existence of such a path is again obvious.

Case II. The segments AB and CD coincide.

Suppose now that $AB = CD$, and assume just to fix the ideas that this is a horizontal segment. Since XY and ZW are two disjoint subsegments of AB, it is enough to prove that v is injective on AB. By construction, there exist $1 \le i^-$, i^+, $j \le K-1$, with $i^- < i^+$, such that a vertex V_{im} is contained in AB if and only if $i^- \le i \le i^+$ and $m = j$. Let then $i^- \le i \le i^+$, and let us call $C_i D_i$ the shortest vertical segment containing V_{ij} such that C_i and D_i are of the form P_l. By definition, $\boldsymbol{V}_{ij}$ is the unique intersection point of the generalized segments $[\boldsymbol{AB}]$ and $[\boldsymbol{C_i D_i}]$: all we have to do, then, is to show that the points $\boldsymbol{V}_{ij}$ are ordered from $\boldsymbol{A}$ to $\boldsymbol{B}$, on $[\boldsymbol{AB}]$, when i ranges from i^- to i^+. Once again by Lemma 4.20, this immediately follows by taking the path $\tilde{\sigma}$ which is the union of the segments $V_{i^-0} V_{i^-K}$, then $V_{i^-K} V_{(i^-+1)K}$, then $V_{(i^-+1)K} V_{(i^-+1)0}$, then $V_{(i^-+1)0} V_{(i^-+2)0}$, and so on up to $V_{i^+0} V_{i^+K}$ or $V_{i^+K} V_{i^+0}$, depending whether the difference $i^+ - i^-$ is even or odd.

Case III. The intersection between AB and CD is a common endpoint.

Suppose now that $AB \cap CD$ is done by a single point, which is a common endpoint. Up to a change in the name of the points, we can then assume that $B = C$. Keep again in mind that $v(AB)$ and $v(CD)$ are the two generalized segments between $\boldsymbol{A}$ and $\boldsymbol{B}$, and between $\boldsymbol{C}$ and $\boldsymbol{D}$ respectively; since $B = C$, hence $\boldsymbol{B} = \boldsymbol{C}$, these two generalized segments are contained in the two different rectangles of the grid $\widetilde{\mathscr{G}}$ having $\boldsymbol{B} = \boldsymbol{C}$ on the boundary. As a consequence, the only intersection between $v(AB)$ and $v(CD)$ is the point $\boldsymbol{B} = \boldsymbol{C}$; moreover, as noticed above,

the only point of AB (resp., CD) having $\boldsymbol{B}$ as image is B (resp., C) itself. As a consequence, $v(XY) = v(ZW)$ can be non-empty only if $B = C$ belongs both to XY and to ZW, which is in turn absurd since $XY \cap ZW = \emptyset$. Also this case is then concluded.

Case IV. The intersection between AB and CD is a point, which is not a common endpoint. The last possible case is that the two segments AB and CD have a single intersection point, which is internal to at least one of them, hence automatically to both. This means that necessarily $AB \cap CD = V_{ij}$ for a suitable $1 \leq i, j \leq K-1$, and AB and CD are the shortest horizontal and vertical segment containing V_{ij} and with both endpoints of the form P_l. We know then that the intersection $v(AB) \cap v(CD)$ consists of a single point, namely $\boldsymbol{V}_{ij}$. Moreover, since by Case II we know that v is injective both on AB and on CD, then V_{ij} is the unique point of AB, as well as of CD, having $\boldsymbol{V}_{ij}$ as image under v. Therefore, the only possibility for the intersection $v(XY) \cap v(ZW)$ not to be empty, is that V_{ij} is contained both in XY and in ZW, which is in turn impossible since XY and ZW have empty intersection. We have then concluded also this final case.   □

Remark 4.23. *A quick look to the proof of Theorem C ensures that what we really use is the fact that $u^{-1}(\mathcal{C})$ is a closed, connected set which does not disconnect $\mathcal{Q}$ whenever $\mathcal{C} \subseteq \boldsymbol{Q}$ is a point, or a square, or a connected piece of the boundary of a square. Actually, we only need the assumption about the points in order to introduce the function v_σ in Definition 4.18, while the assumption about squares and parts of the boundary of squares is only used –several times– in order to prove Proposition 4.19 and its preparatory Lemmas 4.21 and 4.22. An interesting question is then whether given $u \in W^{1,p}$ it suffices, in order to satisfy the non-crossing condition, that the counterimage of every point in $\boldsymbol{Q}$ is a connected set which does not disconnect $\mathcal{Q}$.*

4.5. **Monotone continuous maps satisfy the conclusion of Theorem C.** In this short final section we show that a continuous, $W^{1,p}$ map is monotone if and only if it satisfies the assumptions of Theorem C, hence Theorem C is somehow a reasonable extension to the non-continuous case of the results of [14]. In particular, thanks to Theorem A, this extends Theorem 1.1 (ii) to the case when $p = 1$, see also [5] for further extensions. Notice that, by the results of Young, it is clear that a monotone map satisfies the non-crossing condition, so the fact that it can be approximated by diffeomorphisms follows directly from Theorem A, even without using Theorem C.

Lemma 4.24. *Let $u : \mathcal{Q} \to \boldsymbol{Q}$ be a continuous, monotone function which is equal to the identity on $\partial \mathcal{Q}$. Then, for every closed, connected set $\boldsymbol{C} \subseteq \boldsymbol{Q}$ such that $\boldsymbol{Q} \setminus \boldsymbol{C}$ is connected, the set $C = u^{-1}(\boldsymbol{C}) \subseteq \mathcal{Q}$ is also a closed, connected set such that $\mathcal{Q} \setminus C$ is connected.*

Proof. Let $\boldsymbol{C}$ be a closed connected subset of $\boldsymbol{Q}$ such that $\boldsymbol{Q} \setminus \boldsymbol{C}$ is connected. Clearly, C is closed. Assume that there are two disjoint closed sets A and B such that $u^{-1}(\boldsymbol{C}) = A \cup B$ and let us define

$$\boldsymbol{A} = \{\boldsymbol{P} \in \boldsymbol{C} \text{ such that } u^{-1}(\boldsymbol{P}) \subseteq A\} \qquad \text{and} \qquad \boldsymbol{B} = \{\boldsymbol{P} \in \boldsymbol{C} \text{ such that } u^{-1}(\boldsymbol{P}) \subseteq B\}.$$

For every $\boldsymbol{P} \in \boldsymbol{\mathcal{C}}$, by monotonicity the set $u^{-1}(\boldsymbol{P})$ is connected, hence it must be entirely containd either in A or in B. Thus, $\boldsymbol{A}$ and $\boldsymbol{B}$ are a closed partition of $\boldsymbol{C}$ and then, by connectedness, either $\boldsymbol{A}$ or $\boldsymbol{B}$ is empty, which in turn implies that either A or B is empty and thus that $u^{-1}(\boldsymbol{C})$ is connected.

Let now $P, Q \in \mathcal{Q} \setminus u^{-1}(\boldsymbol{C})$ and let $\boldsymbol{P} = u(P)$ and $\boldsymbol{Q} = u(Q)$. Since $\boldsymbol{P}, \boldsymbol{Q} \in \boldsymbol{\mathcal{Q}} \setminus \boldsymbol{C}$ and the last is set pathwise connected (since $\boldsymbol{C}$ is closed), we can find a continuous path γ in $\boldsymbol{\mathcal{Q}}$ connecting $\boldsymbol{P}$ and $\boldsymbol{Q}$ and not intersecting $\boldsymbol{C}$. From the first part $u^{-1}(\gamma)$ is a connected set containing P and Q but not intersecting C, which shows that P and Q belong to the same connected component of $\mathcal{Q} \setminus u^{-1}(\boldsymbol{C})$. Being P and Q arbitrary points, this concludes the proof. $\square$

5. Counterexamples

This last section is devoted to show four counterexamples, which help to clarify the contrast between the non-crossing condition, the INV condition, and the sufficient condition that we use in Theorem C. In particuar, we will obtain the following situations.

In the first example, Section 5.1, we show that the INV condition is very poor without the additional assumption that $\det Du > 0$ almost everywhere. Let us be more precise: in the paper [17], which introduced the INV condition, many of the results for INV functions were proved under this additional assumption, which is physically meaningful, since a deformation not satisfying it has infinite energy in most of the models. In particular, in [17, Theorem 9.1] it is proved that the INV condition with the additional assumption $\det Du > 0$ a.e. is stable under a regular change of variables; using the notation of this paper, we can equivalently say that the INV condition implies the INV$^+$ one. Our example of Section 5.1 shows an INV function u, not satisfying the assumption $\det Du > 0$ a.e., which fails to be INV$^+$; our function is actually INV and Lipschitz continuous but not monotone, and the counter-image of some points is disconnected. In particular, it is an INV function which is not approximable by diffeomorphisms.

In the second example, Section 5.2, we show a function which is not approximable by diffeomorphisms, but which satisfies the INV$^+$ condition, and for which the counter-image of any point is connected. This example shows that even the condition INV$^+$ is too weak to guarantee that a function is limit of diffeomorphisms. In addition, also the assumption that the counter-image of any point is connected is not sufficient to be limit of diffeomorphisms, while it is necessary as noted in Remark 2.9. Keep in mind that, in the particular case of continuous functions, the condition that the counter-image of points is connected is the monotonicity condition, hence it is both necessary and sufficient, as discussed in the Introduction.

In the third example, Section 5.3, we show that the INV condition is not sufficient even with the additional assumption that $\det Du > 0$ almost everywhere. In fact, we present a function u for which $\det Du > 0$ a.e., which is INV (hence also INV$^+$), but which is not limit of diffeomorphisms.

In the last example, Section 5.4, we show that the sufficient condition of Theorem C is far from being necessary. In fact, it is very simple to present examples of limits of diffeomorphisms

for which the counter-image of some points disconnects $\mathcal{Q}$. However, a quick inspection to the proof of the Theorem clearly shows that the assumption on the counter-image of points is actually needed only for points which are in the arrival grid $\widetilde{\mathscr{G}}$; as a consequence, if for a function the set of points with counter-image which disconnects $\mathcal{Q}$ is $\mathscr{H}^1$-negligible, the proof works exactly in the same way, up to chose an arrival grid which does not meet any of those points. In our counterexample, instead, we show a function which is limit of diffeomorphisms, but for which the set of points whose counter-image disconnects $\mathcal{Q}$ is a whole segment.

5.1. **The first counterexample.** Let us construct our first example. We consider the rectangles $\mathcal{R} = [0,4] \times [-10,10]$ and $\mathcal{R}^- = [1,3] \times [-9,9]$, and the segment $\mathcal{S} = [1,3] \times \{0\}$. Let $v : \mathcal{R} \to \mathcal{R}$ be a smooth function for which $v(x,y) = (x,0)$ on $\mathcal{R}^-$, and which is a diffeomorphism between $\mathcal{R} \setminus \mathcal{R}^-$ and $\mathcal{R} \setminus \mathcal{S}$, coinciding with the identity on $\partial\mathcal{R}$. It is obvious that this function is a limit of diffeomorphisms, and that it satisfies the INV condition. Our "special" function is instead $u : \mathcal{R} \to \mathcal{R}$ given by

$$u(x,y) = \begin{cases} v(x,y) & \text{if } (x,y) \in \mathcal{R} \setminus \mathcal{R}^-, \\ (\varphi(x,y),0) & \text{if } (x,y) \in \mathcal{R}^-, \end{cases}$$

where $\varphi : \mathcal{R}^- \to [1,3]$ is a continuous function such that $\varphi(x,y) = x$ whenever $(x,y) \in \partial\mathcal{R}^-$, so that u is continuous. To define φ, we divide the rectangle $\mathcal{R}^-$ in the three essentially disjoint parts $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ as in Figure 6. More precisely, $\mathcal{A}$ is given by all the points $(x,y) \in \mathcal{R}^-$ with
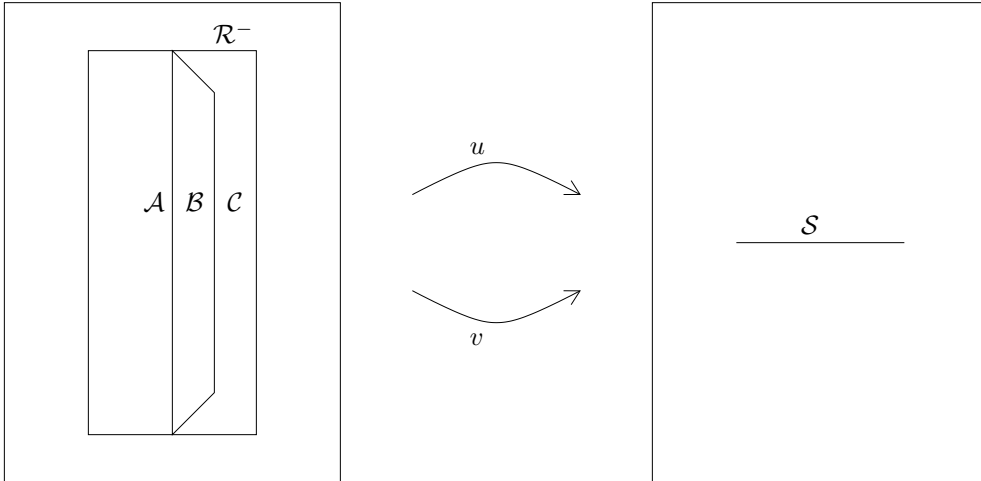

FIGURE 6. The function u for the example of Section 5.1.

$1 \leq x \leq 2$, $\mathcal{B}$ by the points $(x,y) \in \mathcal{R}^-$ with $2 \leq x \leq \min\{11 - |y|, 2.5\}$, and $\mathcal{C}$ by the other points. The function φ is identified by the properties that $\varphi(x,y) = x$ in $\mathcal{A}$, that $\partial\varphi/\partial x = -1$ in $\mathcal{B}$, and that $\partial\varphi/\partial x$ is constant on each horizontal segment contained in $\mathcal{C}$. Hence, on any horizontal segment $[1,3] \times \{y\} \subseteq \mathcal{R}^-$ one has that φ is linear with slope 1 between 1 and 2, then linear with slope -1 between 2 and $\min\{11 - |y|, 2.5\}$, and then again linear with some strictly positive slope depending on $|y|$ between $\min\{11 - |y|, 2.5\}$ and 3.

It is easy to check that $u \in W^{1,p}(\mathcal{R})$ is not a limit of diffeomorphisms (for instance, since it does not satisfy the non-crossing condition). In addition, there are points in $\mathcal{R}$ whose counter-image is not even connected, for instance the point $(1.5, 0)$. Nevertheless, we will show in a moment that u satisfies the INV condition. It is easy to observe that u does not satisfy the INV$^+$ condition: just take a curve γ contained in the interior of the union of the two regions $\mathcal{B}$ and $\mathcal{C}$, and which contains the whole vertical segment between $\mathcal{B}$ and $\mathcal{C}$. Then, any point of this vertical segment is in the interior of the curve γ, but its image is $(1.5, 0)$, which is external to $u(\gamma)$, since $u(\gamma)$ is a segment $[a, b] \times \{0\}$ with $1.5 < a < b < 3$.

Lemma 5.1. *The function $u : \mathcal{R} \to \mathcal{R}$ defined above satisfies the INV condition.*

Proof. Let Γ be a circle contained in $\mathcal{R}$. Notice that Γ cannot intersect both the upper and the lower segment of $\partial \mathcal{R}^-$; as a consequence, also keeping in mind that u and v are continuous, we obtain that the points which are internal to $u(\Gamma)$ and to $v(\Gamma)$ are the same, and in particular no such point belongs to the segment $\mathcal{S}$. Let then $P \notin \mathcal{R}^-$: since v satisfies the INV condition and $u(P) = v(P)$ we have that, if P is internal to Γ, then $u(P)$ is not external to $u(\Gamma)$, while if P is external to Γ, then $u(P)$ is not internal to $u(\Gamma)$.

To conclude checking the INV condition, let us then take $P = (x, y) \in \mathcal{R}^-$. Since $u(P)$ belongs to the segment $\mathcal{S}$, then it cannot be internal to $u(\Gamma)$. We can then assume that P is internal to Γ, and we have to exclude that $u(P)$ is external to $u(\Gamma)$: since φ is continuous, to do so it is enough to find two points $A', B' \in \Gamma$ such that $\varphi(A') \leq \varphi(P) \leq \varphi(B')$. Let AB be the horizontal segment containing P with $A, B \in \Gamma$.

Assume first that $x \leq 2$: in this case, if $A \in \mathcal{R}^-$ then $\varphi(A) < \varphi(P)$, and if $A \notin \mathcal{R}^-$ then there is another $A' \in \Gamma$ belonging to the left side of $\mathcal{R}^-$, so $\varphi(A') < \varphi(P)$. Thus, if $\varphi(B) \geq \varphi(P)$ we are done. On the other hand, if $\varphi(B) < \varphi(P)$, then the segment AB crosses the vertical segment $\partial \mathcal{A} \cap \partial \mathcal{B}$, thus Γ must cross the same segment; hence, there is a point $B' \in \Gamma \cap \partial \mathcal{A} \cap \partial \mathcal{B}$, which means $\varphi(B') = 2$, and since $x \leq 2$ implies $\varphi(P) \leq 2$ we are done.

Assume now that $2 < x < \min\{11 - |y|, 2.5\}$. The existence of a point $B' \in \Gamma$ with $\varphi(B') \geq \varphi(P)$ is clear: indeed, either $\varphi(A) > \varphi(P)$, or the segment AB crosses the vertical segment $\partial \mathcal{A} \cap \partial \mathcal{B}$, so the argument above again applies. If $\varphi(B) < \varphi(P)$ we are done; otherwise, the segment AB crosses $\partial \mathcal{B} \cap \partial \mathcal{C}$ in the point $(\tilde{x}, y)$, with $\tilde{x} = \min\{11 - |y|, 2.5\}$. As a consequence, we find another point $A' = (x', y') \in \Gamma \cap \partial \mathcal{B} \cap \partial \mathcal{C}$, and in particular there is one with $x' \geq \tilde{x}$: this implies $\varphi(A') < \varphi(P)$, so we are again done.

Let us finally assume that $x \geq \min\{11 - |y|, 2.5\}$. This time, it is clear that $\varphi(B) > \varphi(P)$ (or, if $B \notin \mathcal{R}^-$, that there is some $B' \in \Gamma$ belonging to the right side of $\mathcal{R}^-$, thus with $\varphi(B') > \varphi(P)$); moreover, either $\varphi(A) < \varphi(P)$, or the segment AB crosses $\partial \mathcal{B} \cap \partial \mathcal{C}$, so the existence of a point A' with $\varphi(A') \leq \varphi(P)$ follows by the above argument. The proof is then concluded. $\square$

5.2. The second counterexample.

Let us now construct our second example. We consider the domain $\Omega = \{|x| < 2\}$, the balls $\mathcal{B} = \{|x| < 1\}$ and $\mathcal{B}^- = \{|x| < 1/2\}$, and the unit segment $\mathcal{S} = [0, 1] \times \{0\}$, and we aim to define a function $u : \Omega \to \Omega$ which equals the identity on $\partial \Omega$. First of all, we define u outside $\mathcal{B}$. For ε small, we let the image under u of the circle with

radius $1 + \varepsilon$ be the outer boundary Γ_ε of an ε-neighborhood of the segment $\mathcal{S}$, parametrized with two different speeds. More precisely, as Figure 7 (left) shows, the part of length ε around the point $(-1-\varepsilon, 0)$ is sent, with constant speed, to the whole Γ_ε except an ε-neighborhood of the point $(1+\varepsilon, 0)$, and the remaining part of the circle is sent, again with constant speed, to the remaining part of Γ_ε. It is clearly possible to extend the definition of u to the rest of $\Omega \setminus \mathcal{B}$ in such a way that u is a $W^{1,p}$ function on $\Omega \setminus \mathcal{B}$ coinciding with the identity on $\partial\Omega$ for every $p < 2$. Notice that u is a homeomorphism of $\Omega \setminus \mathcal{B}$ onto $\Omega \setminus \mathcal{S}$, and that the whole boundary of $\mathcal{B}$ except the "west pole" $W = (-1, 0)$ is sent on $(1, 0)$, while the generalized image of W is the whole segment $\mathcal{S}$.

Let us now define u in the annulus between $\partial\mathcal{B}$ and $\partial\mathcal{B}^-$. First of all, the image of the circle $\partial\mathcal{B}^-$ is the segment $\mathcal{S}$ done four times with constant speed: more precisely, for every $\theta \in \mathbb{S}^1$ we let $u(\frac{1}{2}\cos\theta, \frac{1}{2}\sin\theta) = (\varphi(\theta), 0)$, where $\varphi : \mathbb{S}^1 \to [0, 1]$ is the Lipschitz function such that

$$\varphi(0) = \varphi(\pi) = 1, \qquad\qquad \varphi(\pi/2) = \varphi(-\pi/2) = 0,$$

and $|\varphi'| = 2/\pi$. Then, let us consider the circle $\{|x| = 1 - \varepsilon\}$ for any $0 < \varepsilon < 1/2$: in this case, for every $\theta \in \mathbb{S}^1$ we let $u\big((1-\varepsilon)(\cos\theta, \sin\theta)\big) = \varphi(\tau_\varepsilon(\theta))$, where $\tau_\varepsilon : \mathbb{S}^1 \to \mathbb{S}^1$ is the Lipschitz homeomorphism which maps at constant speed the arc L_ε centered at π of width $2\pi\varepsilon$ on the whole $\mathbb{S}^1$ except the arc R_ε centered at 0 of width $2\pi\varepsilon$, while $\mathbb{S}^1 \setminus L_\varepsilon$ is mapped, again at constant speed, onto R_ε. Notice that $\tau_{1/2}$ is the identity, so the function u is continuous on $\partial\mathcal{B}^-$, and moreover it is still in $W^{1,p}$ for every $p < 2$.
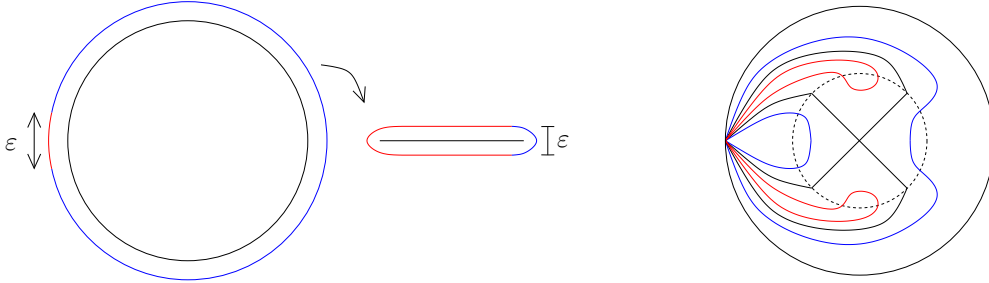

FIGURE 7. Definition of u outside the circle $\partial\mathcal{B}$ (left); counter-image of $(x, 0)$ for $0 < x < 1/2$ in red, for $x = 1/2$ in black, for $1/2 < x < 1$ in blue (right). The dashed circle is $\partial\mathcal{B}^-$.

To conclude, we have to define u inside $\mathcal{B}^-$, and we simply shrink everything onto the point $(1/2, 0)$. More precisely, for every $0 < \rho < 1/2$ and every $\theta \in \mathbb{S}^1$ we set

$$u(\rho\cos\theta, \rho\sin\theta) = \frac{1}{2} + \left(\varphi(\theta) - \frac{1}{2}\right)2\rho.$$

The resulting function $u : \Omega \to \Omega$ is then a $W^{1,p}$ function, coinciding with the identity on the boundary; it is smooth except at the point $W = (-1, 0)$. This function is not a limit of diffeomorphisms, since it does not satisfy the non-crossing condition. To see this, it is enough to take a grid made by a horizontal segment S_1 and a vertical segment S_2, both contained in $\mathcal{B}^-$. More precisely, we set $S_1 = [-1/2, 1/2] \times \{0\}$ and $S_2 = \{a\} \times [-b, b]$ for some $0 < a < \sqrt{2}/4$ and

$b = \sqrt{1 - 4a^2}/2$. The restriction of u to S_1 is the segment $[1/2, 1] \times \{0\}$ done first from right to left and then from left to right. The restriction of u to S_2 is the segment $[\alpha, \beta] \times \{0\}$ with some $\alpha < 1/2 < \beta < 1$, done first from left to right and then from right to left. It is clear that u cannot be uniformly approximated with an injective function on $S_1 \cup S_2$. Nevertheless, we can prove the following.

Lemma 5.2. *The function u defined above satisfies the INV$^+$ condition. Moreover, the counter-image of every point is connected.*

Proof. Let us first consider the counter-image of a point $\boldsymbol{P} = (x, y) \in \Omega$. If $\boldsymbol{P} \notin \mathcal{S}$, then $u^{-1}(\boldsymbol{P})$ is a single point in $\Omega \setminus \mathcal{B}$, hence connected. Assume instead that $\boldsymbol{P} \in \mathcal{S}$, so $\boldsymbol{P} = (x, 0)$ for some $0 \leq x \leq 1$. Let us for a moment consider the case $0 < x < 1$: in this case, $\boldsymbol{P}$ has exactly four counter-images on the circle of radius $1/2$; therefore, for any radius ρ between $1/2$ and 1, the point $\boldsymbol{P}$ has still four counter-images on the circle of radius ρ, and these are four points which all converge to W when $\rho \to 1$. When the radius ρ goes from $1/2$ to 0, instead, the counter-images remain four for a while, in particular until ρ becomes equal to $\rho_{\min} = |x - 1/2|$; then, on the boundary of the $\rho_{\min}$, the four points have become only the two points $(0, \pm\rho_{\min})$ if $0 < x < 1/2$, only the two points $(\pm\rho_{\min}, 0)$ if $1/2 < x < 1$, and only the origin if $\rho_{\min} = 0$, that is, $x = 1/2$. The counter-image does not have points in the open ball with radius $\rho_{\min}$. Figure 7 (right) depicts the counter-image of $\boldsymbol{P}$ in these three cases. In the limiting case $x = 0$, the counter-image is done only by two points on each circle $\{|x| = \rho\}$ for every $1/2 \leq \rho < 1$, which again converge to W when $\rho \to 1$; in the limiting case $x = 1$, instead, the counter-image is the whole circle $\partial\mathcal{B}$ together with the two segment WA and BC with $A = (-1/2, 0)$, $B = (1/2, 0)$ and $C = (1, 0)$.

As a consequence of the above characterization, we have that the counter-image of any point of Ω is a connected set. The validity of the INV$^+$ condition follows then by an obvious geometric argument. $\qquad\square$

5.3. The third counterexample.

We can now present our third counterexample, which is an INV map $u \in W^{1,p}([-2, 2]^2; [-2, 2]^2)$ such that $\det Du > 0$ almost everywhere, and which does not satisfy the non-crossing condition.

First of all, we need to present the auxiliary function $v : [-2, 2]^2 \to [-2, 2]^2$, coinciding with the identity on the boundary. For any $0 \leq t \leq 1$, we give four paths α_t, β_t, γ_t and σ_t, see Figure 8 for an illustration. The path α_t is simply the segment between $(-2, t)$ and $(0, 0)$, while σ_t is the segment between $(0, 0)$ and $(2, t)$. Instead, the paths β_t and γ_t are two loops, starting and ending at $(0, 0)$, lying on the half-space $\{y \geq 0\}$, percurred in the clockwise sense. In addition, these paths depend smoothly on t, each path β_t (resp., γ_t) is contained in the internal part of β_s (resp., γ_s) when $t < s$, the paths β_0 and γ_0 consist of the sole point $(0, 0)$, and the intersection between any two paths of the form α_t, β_t, γ_t and/or σ_t is always the sole point $(0, 0)$. Finally, each path β_t is on the left of each path γ_s, as in the figure.

For any $0 \leq t \leq 1$, we call now $x_t = \min\{t, 1 - t\}$, and we introduce the points $A_t = (-2, t)$, $B_t = (-x_t, t)$, $C_t = (0, t)$, $D_t = (x_t, t)$ and $E_t = (2, t)$; notice that these points are all distinct
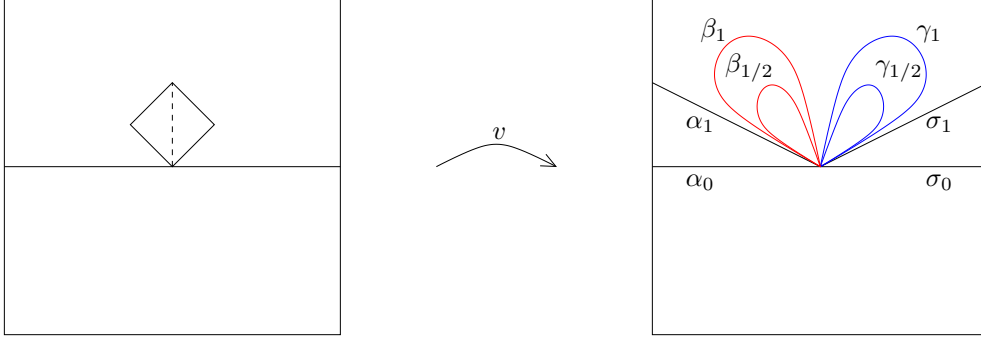
FIGURE 8. Definition of v in the example of Section 5.3.

except $B_0 = C_0 = D_0 = (0,0)$ and $B_1 = C_1 = D_1 = (0,1)$. We can then define v on the horizontal segment $A_t E_t$ as follows: on the segment $A_t B_t$ (resp., $B_t C_t$, $C_t D_t$, $D_t E_t$), the image of v is the curve α_t (resp., β_t, γ_t, σ_t), parametrized at constant speed. Notice that, up to now, the function v is smooth except at the point $(0,1)$, where it is discontinuous. Moreover, by construction the function v belongs to $W^{1,p}([-2,2] \times [0,1])$ for any $p < 2$. We can now extend v to the whole square $[-2,2]^2$: on the rectangle $[-2,2] \times [-2,0)$ we can simply set $v = \mathrm{Id}$, and it is simple to observe the existence also of a $W^{1,p}$ extension of v on the rectangle $[-2,2] \times (1,2]$ which is a homeomorphism onto the image. The resulting function v belongs to $W^{1,p}([-2,2]^2)$, it is smooth except at the point $(0,1)$, where it is discontinuous, it is injective except at $v^{-1}(0,0)$, and it is clearly a strong $W^{1,p}$ limit of diffeomorphisms, in particular it satisfies the INV and the INV$^+$ conditions, which are equivalent since by construction $\det Dv > 0$ almost everywhere.

Let us now call $\mathcal{T}^\pm$ the two triangles having vertices in $(0,0)$, $(0,1)$ and $(\pm 1/2, 1/2)$, and let $\psi : [-2,2]^2 \to [-2,2]^2$ the function defined as

$$\psi(s,t) = \begin{cases} (s + x_t, t) & \text{if } (s,t) \in \mathcal{T}^-, \\ (s - x_t, t) & \text{if } (s,t) \in \mathcal{T}^+, \\ (s,t) & \text{if } (s,t) \notin \mathcal{T}^+ \cup \mathcal{T}^-. \end{cases}$$

Notice that ψ is a bijection moving $\mathcal{T}^-$ on $\mathcal{T}^+$ and vice-versa. We can then define our function $u : [-2,2]^2 \to [-2,2]^2$ simply as $u = v \circ \psi$. Since $v(x) = (0,0)$ for every point $x \in \partial \mathcal{T}^- \cup \partial \mathcal{T}^+$ except at the discontinuity point $(0,1)$, also u belongs to $W^{1,p}([-2,2]^2)$; moreover, u is also continuous except at $(0,1)$, injective except at $\partial \mathcal{T}^- \cup \partial \mathcal{T}^+$, and coincides with the identity on the boundary, and we have that also $\det Du > 0$ almost everywhere (so, the INV and the INV$^+$ conditions are equivalent for u, as well as for v). On the other hand, u does not satisfy the non-crossing condition, so it is not a limit of diffeomorphisms. To observe the failure of the non-crossing condition, it is enough to consider the image of an horizontal segment joining $(-2,t)$ with $(2,t)$ for any $0 < t < 1$: it is a continuous mappings, namely, the union of the four paths α_t, γ_t, β_t and σ_t, and it is clearly impossible to make it injective with a small uniform variation because β_t has been postponed after γ_t. Nevertheless, we can show the validity of the INV condition for u.

Lemma 5.3. *The function u defined above satisfies the INV condition.*

Proof. Let $\tau : \mathbb{S}^1 \to [-2,2]^2 \setminus (0,1)$ be a circle: we have to prove that all the points internal (resp., external) to τ have degree 1 (resp., 0) with respect to $u(\tau)$, unless they are contained on $u(\tau)$ itself. We check this in few possible cases.

Case I. If τ does not intersect $\partial \mathcal{T}^- \cup \partial \mathcal{T}^+$.

In this case, the property comes immediately from the analogous property of u. Indeed, any point $x \in [-2,2]^2$ is internal (resp., external) to τ if and only if $\psi(x)$ is internal (resp., external) to $\psi(\tau)$, which in this case is also an injective closed path. Then, the degree of $u(x) = v(\tau(x))$ with respect to $u(\tau) = v(\psi(\tau))$ behaves correctly because u satisfies the INV condition.

Case II. If τ intersects only $\partial \mathcal{T}^-$, or only $\partial \tau^+$.

Let us now assume that τ intersects only $\partial \mathcal{T}^-$, the case when τ intersects only $\partial \mathcal{T}^+$ is clearly identical. Notice that $u(\tau) = v(\psi(\tau))$. However, this time $\psi(\tau)$ is done by two disconnected open curves, one outside $\mathcal{T}^- \cup \mathcal{T}^+$ and the other one inside $\mathcal{T}^-$; let us call them τ_a and τ_b. As in Figure 9 (above), we denote by τ_1 and τ_2 the two injective, closed paths obtained adding to τ_a (resp., τ_b) a segment contained in $\partial \mathcal{T}^-$ (resp., $\partial \mathcal{T}^- \cap \partial \mathcal{T}^+$). Notice that u is constantly $(0,0)$ on both segments, hence $u(\tau) = v(\tau_a \cup \tau_b) = v(\tau_1) \cup v(\tau_2)$. Moreover, $v(\tau_1)$ and $v(\tau_2)$ are two injective curves, whose only intersection is the point $(0,0)$. In particular, for any $y \in [-2,2]^2 \setminus u(\tau)$ one has that

$$\deg(y, u(\tau)) = \deg(y, v(\tau_1)) + \deg(y, v(\tau_2)). \tag{5.1}$$

Let now $x \in [-2,2]^2$ be any point. If it belongs to $\partial \mathcal{T}^- \cup \partial \mathcal{T}^+$, then $u(x) = (0,0) \in u(\tau)$, hence there is nothing to prove. If x is internal to τ and not on $\partial \mathcal{T}^-$, then $y = \psi(x)$ is internal to one between τ_1 and τ_2, and external to the other one. Since v satisfies the INV^+ condition, we can apply it to $\psi(x)$ with respect to τ_1 and τ_2, so we find that one between $\deg(u(x), v(\tau_1)) = \deg(v(y), v(\tau_1))$ and $\deg(u(x), v(\tau_2)) = \deg(v(y), v(\tau_2))$ is 0 and the other one is 1, thus by (5.1) we get $\deg(u(x), u(\tau)) = 1$, as required since x is internal to τ.
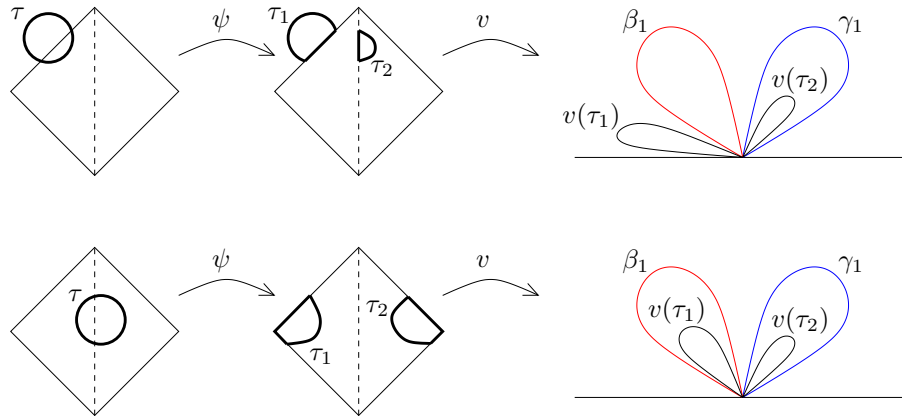

FIGURE 9. The paths τ, τ_1 and τ_2 in Case II (above) and in Case III (below).

Finally, if x is external to τ and not on $\partial \mathcal{T}^- \cup \partial \mathcal{T}^+$, then $\psi(x)$ is external to both τ_1 and τ_2, so an analogous argument ensures that $\deg(u(x), u(\tau)) = 0$, so we are done.
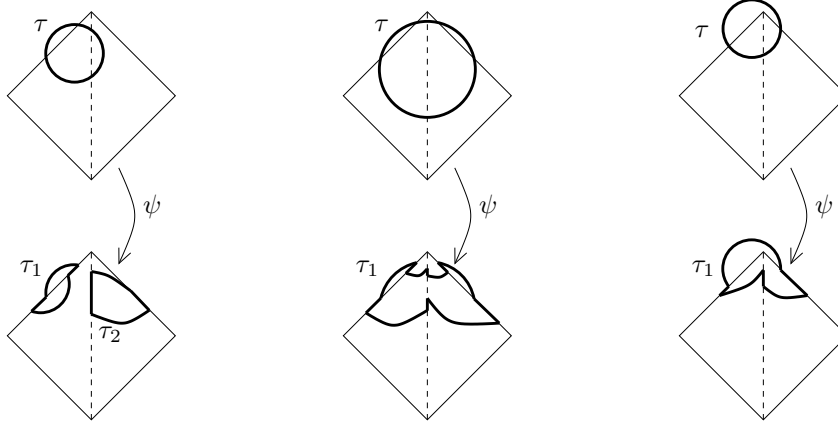
FIGURE 10. The paths τ, τ_1 and τ_2 in Case IV (left) and in Case V (center and right).

Case III. _If τ intersects $\partial\mathcal{T}^- \cap \partial\mathcal{T}^+$ but not $\partial\mathcal{T}^- \setminus \partial\mathcal{T}^+$ nor $\partial\mathcal{T}^+ \setminus \partial\mathcal{T}^-$._

This case is completely analogous to Case II, as a quick look to Figure 9 (below) shows.

Case IV. _If τ intersects $\partial\mathcal{T}^- \cap \partial\mathcal{T}^+$ and one between $\partial\mathcal{T}^- \setminus \partial\mathcal{T}^+$ and $\partial\mathcal{T}^+ \setminus \partial\mathcal{T}^-$._

This case is again very similar to that of Case II, see Figure 10 (left); let us assume, just to fix the ideas, that γ intersects $\partial\mathcal{T}^- \setminus \partial\mathcal{T}^+$ and not $\partial\mathcal{T}^+ \setminus \partial\mathcal{T}^-$. This time, $\psi(\tau)$ is done by four disconnected open arcs, namely, an arc τ_a outside $\mathcal{T}^- \cup \mathcal{T}^+$, two arcs τ_b and τ_d inside $\mathcal{T}^+$ and an arc τ_c inside $\mathcal{T}^-$. As in the Figure, we denote by τ_1 the closed, injective curve obtained adding to $\tau_a \cup \tau_c$ two segments on $\partial\mathcal{T}^- \setminus \partial\mathcal{T}^+$. Similarly, we denote by τ_2 the closed, injective curve obtained adding to $\tau_b \cup \tau_d$ a segment on $\partial\mathcal{T}^- \cap \partial\mathcal{T}^+$ and another one on $\partial\mathcal{T}^+ \setminus \partial\mathcal{T}^-$. It is then enough to ovserve that τ_1 and τ_2 are disjoint and that (5.1) holds as before, and then to argue as in the preceding steps.

Case V. _If τ intersects $\partial\mathcal{T}^- \cap \partial\mathcal{T}^+$, $\partial\mathcal{T}^- \setminus \partial\mathcal{T}^+$ and $\partial\mathcal{T}^+ \setminus \partial\mathcal{T}^-$._

This case is now very simple to handle, it is enough to argue more or less exactly as in the last cases. As Figure 10 (center and right) shows, there are two possible subcases, namely, if the point $(0,1)$ is internal or external to τ. In the first case, $\psi(\tau)$ is done by six different pieces, and we can consider a single path τ_1 by adding six intervals to it; in the second case, $\psi(\tau)$ is done by three pieces, and we obtain a single path τ_1 by adding three intervals. $\square$

5.4. **The fourth counterexample.** We give now our last counterexample, which is the simplest one. It ensures that the sufficient condition in Theorem C is not necessary, not even up to $\mathscr{H}^1$-negligible sets. More precisely, we exhibit a function $u : [-2,2]^2 \to [-2,2]^2$, coinciding with the identity on the boundary, being the $W^{1,p}$ limit of diffeomorphisms for $p < 2$, but for which $u^{-1}(\boldsymbol{P})$ is a connected set which disconnects the domain for every $\boldsymbol{P}$ in a segment.

For every $t \in [0, 1/2]$, we consider the 2-dimensional quadrilateral $\mathcal{Q}_t$ having as vertices the points $(0,0)$, $(2t,0)$ and $(t, \pm t \tan t)$. Notice that, for every $s < t$, $\mathcal{Q}_s$ is contained in $\mathcal{Q}_t$, having the origin as only common point. We start defining u inside the biggest quadrilateral $\mathcal{Q}_{1/2}$ as the function which, for every $0 < t < 1/2$, maps the whole $\partial\mathcal{Q}_t$ on the point $(2t, 0)$; Figure 11 depicts the function. Notice that this function is continuous except at $(0,0)$, whose image is

the whole segment $[0,1] \times \{0\}$. Exactly as in the example of Section 5.2, it is possible extend u as a homeomorphism of $[-2,2]^2 \setminus \mathcal{Q}_{1/2}$ onto $[-2,2]^2 \setminus [0,1] \times \{0\}$, belonging to the Sobolev space $W^{1,p}([-2,2]^2 \setminus \mathcal{Q}_{1/2})$ for $p < 2$ and coinciding with the identity on $\partial([-2,2]^2)$. Notice that the whole u is a Borel function, continuous up to the point $(0,0)$, and which is easily seen to satisfy the non-crossing condition; nevertheless, for every $0 < t < 1$, the counter-image of $(t,0)$ is the polygon $\partial \mathcal{Q}_{t/2}$, which disconnects $[-2,2]^2$. By Theorem A, the function u is then limit of diffeomorhisms if and only if it belongs to $W^{1,p}([-2,2]^2)$, and in turn this property is true, as we prove now.
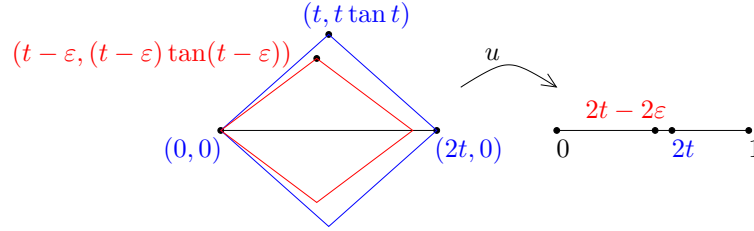

FIGURE 11. Definition of u and two quadrilaterals $\mathcal{Q}_t$ and $\mathcal{Q}_{t+\varepsilon}$.

Lemma 5.4. *The function u defined above belongs to $W^{1,p}([-2,2]^2)$.*

Proof. Since u is continuous except at $(0,0)$, and we already know that it is of class $W^{1,p}$ on $[-2,2]^2 \setminus \mathcal{Q}_{1/2}$, we have only to show that $u \in W^{1,p}$ inside $\mathcal{Q}_{1/2}$; in fact, it is clear that u is smooth in the interior of $\mathcal{Q}_{1/2}$, so the possible problems are only close to $(0,0)$. Let P be a point on $\partial \mathcal{Q}_t$, having distance σ from the origin, and assume for a moment that P belongs to one of the two sides originating at $(0,0)$: then, for every $\varepsilon \ll t$ the point P has distance $\sigma \tan \varepsilon$ from $\partial \mathcal{Q}_{t+\varepsilon}$, hence $|Du(P)| = 1/\sigma$; a trivial geometrical argument shows that, if P belongs to one of the two other sides of $\partial \mathcal{Q}_t$, then we have the stronger estimate $|Du(P)| < 1/\sigma$, hence in general $|Du(P)| \le 1/\sigma$ at every $P \in \partial \mathcal{Q}_t$ having distance σ from the origin. Observe that this estimate does not depend on t. As a consequence, to check that $Du \in L^p(\mathcal{Q}_{1/2})$ it is enough to keep in mind that $p < 2$ and to calculate

$$\int_{\mathcal{Q}_{1/2}} |Du|^p \, dx \le \int_{\theta=-1/2}^{1/2} \int_{\sigma=0}^{1} \frac{1}{\sigma^p} \, \sigma \, d\theta \, d\sigma = \int_{\sigma=0}^{1} \frac{1}{\sigma^{p-1}} \, d\sigma < +\infty \,.$$

$\square$

REFERENCES

[1] G. Alessandrini & M. Sigalotti, Geometric properties of solutions to the anisotropic p-Laplace equation in dimension two, Ann. Acad. Sci. Fenn. Math. **26** (2001), no. 1, 249–266.

[2] J. M. Ball, Discontinuous equilibrium solutions and cavitation in nonlinear elasticity. Phil. Trans. Roy. Soc. London, A **306** (1982), 557–611.

[3] M. Barchiesi & D. Henao & C. Mora-Corral, Local invertibility in Sobolev spaces with applications to nematic elastomers and magnetoelasticity, Arch. Ration. Mech. Anal. **224** (2017), no. 2, 743–816.

[4] D. Campbell, Diffeomorphic approximation of Planar Sobolev Homeomorphisms in Orlicz-Sobolev spaces, Journal of Functional Analysis **273** (2017), no. 1, 125–205.

[5] D. Campbell, J. Onninen, A. Rabina & V. Tengvall, Diffeomorphic Approximation of $W^{1,1}$ Planar Sobolev Monotone Maps, in preparation (2017).

[6] S. Conti & C. De Lellis, Some remarks on the theory of elasticity for compressible Neohookean materials, Ann. Scuola Norm. Sup. Cl. Sci. **2** (2003), 521–549.

[7] S. Daneri & A. Pratelli, Smooth approximation of bi-Lipschitz orientation-preserving homeomorphisms, Ann. Inst. H. Poincaré Anal. Non Linéaire **31** (2014), 567–589.

[8] L.C. Evans, R.F. Gariepy, Measure theory and fine properties of functions. Studies in Advanced Mathematics. CRC Press, Boca Raton, FL, (1992).

[9] M. Giaquinta & G. Modica & J. Souček, Cartesian Currents in Calculus of Variations I and II, Springer (1998).

[10] D. Henao, Cavitation, invertibility, and convergence of regularized minimizers in nonlinear elasticity, J. Elasticity **94** (2009), no. 1, 55–68.

[11] S. Hencl & A. Pratelli, Diffeomorphic Approximation of $W^{1,1}$ Planar Sobolev Homeomorphisms, to appear on JEMS (2016).

[12] T. Iwaniec, L. Kovalev & J. Onninen, Diffeomorphic Approximation of Sobolev Homeomorphisms, Arch. Rational Mech. Anal. **201** (2011), no. 3, 1047–1067.

[13] T. Iwaniec, L. Kovalev & J. Onninen, Hopf Differentials and Smoothing Sobolev Homeomorphisms, International Mathematics Research Notices **14** (2012), 3256–3277.

[14] T. Iwaniec & J. Onninen, Monotone Sobolev mappings of planar domains and surfaces, Arch. Ration. Mech. Anal. **219** (2016), no. 1, 159–181.

[15] T. Iwaniec & J. Onninen, Limits of Sobolev homeomorphisms, J. Eur. Math. Soc. **19** (2017), no. 2, 473–505.

[16] C. Mora Corral & A. Pratelli, Approximation of piecewise affine homeomorphisms by diffeomorphisms, J. Geom. Anal. **24** (2014), no. 3, 1398–1424.

[17] S. Müller & S.J. Spector, An existence theory for nonlinear elasticity that allows for cavitation, Arch. Rat. Mech. Anal. **131** (1995), 1–66.

[18] E. Radici, A planar Sobolev extension theorem for piecewise linear homeomorphisms, Pacific J. Math. **283** (2016), no. 2, 405–418.

[19] V. Šverák, Regularity properties of deformations with finite energy. Arch. Rational Mech. Anal. **100** (1988), 105–127.

[20] J.W.T. Youngs, Homeomorphic approximations to monotone mappings, Duke Math. J. **15** (1948), 87–94.

[21] J.W.T. Youngs, The topological theory of Fréchet surfaces, Annals of Mathematics **45** (1944), 753–785.

[22] J.W.T. Youngs, A reduction theorem concerning the representation problem for Fréchet varieties, Proceedings of the National Academy of Sciences of the United States of America **32** (1946), 328–330.