DIFFEOMORPHIC APPROXIMATION OF PLANAR SOBOLEV HOMEOMORPHISMS IN REARRANGEMENT INVARIANT SPACES

DANIEL CAMPBELL[1,2,*], LUIGI GRECO[3], ROBERTA SCHIATTARELLA[4] AND FILIP SOUDSKÝ[1,5]

Abstract. Let $\Omega \subseteq \mathbb{R}^2$ be a domain, let X be a rearrangement invariant space and let $f \in W^1 X(\Omega, \mathbb{R}^2)$ be a homeomorphism between Ω and $f(\Omega)$. Then there exists a sequence of diffeomorphisms f_k converging to f in the space $W^1 X(\Omega, \mathbb{R}^2)$.

1. INTRODUCTION AND MAIN RESULTS

Recently, motivated by applications in non-linear elasticity and in geometric function theory, a great deal has been devoted in understanding the question of approximating homeomorphisms $f: \Omega \subset \mathbb{R}^n \to f(\Omega) \subset \mathbb{R}^n$ with either diffeomorphisms or piece-wise affine homeomorphisms. This problem is not trivial because the usual approximation techniques like mollification or Lipschitz extension using maximal operator destroy, in general, the injectivity.

In variational models of nonlinear elastic deformations of solid flexible bodies we search for minimisers of energy functionals (often) of the form

$$I(f) = \int_\Omega W(Df) \, \mathrm{d}x \,,$$

where $W : \mathbb{R}^{n \times n} \to \mathbb{R}$ is a stored-energy functional satisfying

$$W(A) \to +\infty \quad \text{as } \det A \to 0 \qquad W(A) = +\infty \quad \text{if } \det A \leq 0 \,. \tag{1.1}$$

[1] Department of Mathematics, University of Hradec Králové, Rokitanského 62, 500 03 Hradec Králové, Czech Republic.

[2] Faculty of Economics, University of South Bohemia, Studentská 13, Ceské Budejovice, Czech Republic.

[3] Dipartimento di Ingegneria Elettrica e delle Tecnologie dell'Informazione, Università degli Studi di Napoli "Federico II", Via Claudio 21, 80125 Napoli, Italy.

[4] Dipartimento di Matematica e Applicazioni "R. Caccioppoli", Università degli Studi di Napoli "Federico II", Via Cintia, 80126 Napoli, Italy.

[5] Department of Mathematics and Didactic of Mathematics, Faculty of Science, Humanities and Education, Technical University of Liberec, Studentská 1402/2, 461 17 Liberec, Czechia.

* Corresponding author: daniel.campbell@uhk.cz

We require that our model respects the law of non-interpenetration of matter and, assuming that the body does not fracture or break, it is therefore natural to look for a minimiser among homeomorphisms. Therefore we minimise the functional over Sobolev homeomorphisms satisfying given boundary values.

Intuition gives the impression that the minimising deformation should be in essence a diffeomorphism (say up to a null set). A naive perception is that a Sobolev homeomorphism is essentially a diffeomorphism. In fact the question of the regularity of minimisers and the question of the behaviour of Sobolev homeomorphisms are somewhat inter-related. A key step to proving the regularity of minimisers (see [1, 2]) is to show that any Sobolev homeomorphism can be approximated arbitrarily well by diffeomorphisms. This is the so-called Ball-Evan's approximation question and is currently a topic of much interest. The initial breakthrough in the planar case were the papers [9, 10] for $f \in W^{1,p}$, $p > 1$. This was followed by [8], planar homeomorphisms in $W^{1,1}$. The latter techniques have further been developed in [12] (bi-Sobolev $W^{1,1}$ case), [13, 14] (BV case) and [4] (Orlicz-Sobolev case). There are still many open questions in this context, especially $W^{1,p}$-bi-Sobolev and dimension $n = 3$.

Naturally, given that one can approximate homeomorphisms by diffeomorphisms in the Orlicz-Sobolev sense (see [4]), the question of approximation in other classes of function spaces such as Lorentz Sobolev spaces, or Grand Sobolev spaces arises. These classes are an important tool in studying the regularity of solutions of certain PDEs and variational problems and usually provide sharper results for existence and regularity of a solution. To provide results for all these important classes at once we will study the question of approximation in general Banach function space. The result from [4] gives us a strong indication that a similar result should hold under a more general context. On the other hand in general r.i. spaces one lacks the explicit norm expression utilised for that result and there are several obstacles that must be overcome.

Let us just recall that by a r.i. space we mean a Banach function space $X(\Omega)$ on the domain Ω, endowed with a norm $\| \cdot \|_{X(\Omega)}$ such that

$$\|u\|_{X(\Omega)} = \|v\|_{X(\Omega)} \text{ whenever } u^* = v^*$$

where u^* and v^* denote the decreasing rearrangements of the functions u, v. The Sobolev space over X is defined as

$$W^1 X(\Omega, \mathbb{R}^2) = \left\{ f \in W^{1,1}_{\text{loc}}(\Omega, \mathbb{R}^2) : \|f\|_{W^1 X(\Omega)} = \|Df\|_{X(\Omega)} + \|f\|_{X(\Omega)} < \infty \right\}.$$

For the full definition see Definition 2.1.

In order to introduce our main result we include the definition of the Lebesgue point property. This property is stronger than the absolute continuity (see (2.8)) of the norm and has been thoroughly characterised in [5].

Definition 1.1 (Lebesgue point property)**.** Let $\Omega \subset \mathbb{R}^n$ be measurable. We say that a function space $X(\Omega)$ satisfies the Lebesgue property if for all $u \in X(\Omega)$ and almost all $x \in \Omega$ one has

$$\lim_{r \to 0+} \frac{\|[u - u(x)]\chi_{B(x,r)}\|_{X(\Omega)}}{|B(x,r)|} = 0. \tag{1.2}$$

We refer to the points x for which (1.2) hold as Lebesgue points of u in X.

By a finitely connected domain $\Omega \subset \mathbb{R}^2$ we refer to a domain $\Omega \subset \mathbb{R}^2$ such that $\mathbb{R}^2 \setminus \Omega$ has a finite number of components.

Theorem 1.2. *Let $\Omega \subseteq \mathbb{R}^2$ be a finitely connected domain. Let $X(\Omega)$ be a rearrangement-invariant Banach function space satisfying the Lebesgue point property (see Def. 1.1). Let $f \in W^1 X(\Omega, \mathbb{R}^2)$ be a homeomorphism. For arbitrary $\varepsilon > 0$ there exists a diffeomorphism $\tilde{f}$ such that*

$$\|Df - D\tilde{f}\|_{X(\Omega)} < \varepsilon \quad \text{and} \quad \|f - \tilde{f}\|_{L^\infty(\Omega)} < \varepsilon.$$

Remark 1.3. Not only can we approximate a Sobolev homeomorphism by diffeomorphisms but by locally finite piece-wise affine homeomorphisms in r.i. spaces satisfying the Lebesgue point property. Moreover, if $\partial\Omega$ is a polygon on which f is piece-wise linear then we can approximate f by finitely piece-wise affine homeomorphisms.

Remark 1.4. Theorem 1.2 provides us with a diffeomorphic approximation of homeomorphisms in customary classes of r.i. spaces. Indeed, besides recovering the case of Sobolev-Orlicz space $W^{1,\Phi}$, where Φ is a Young function satisfying Δ_2-condition, we can consider also the Sobolev-Lorentz spaces, taking $X = L^{p,q}$ for $1 \leq q \leq p < \infty$. Another interesting case is $X = \Lambda_\varphi$, the Lorentz endpoint space associated with a (non-identically vanishing) concave function $\varphi \colon [0, \infty) \to [0, \infty)$ satisfying $\lim_{s \to 0+} \varphi(s) = 0$. For details, see [5].

It is not hard to observe that the absolute continuity of the norm of X is a necessary condition for diffeomorphic approximation of homeomorphisms in $W^1 X$ and by [5] it is also necessary for the Lebesgue point property. Our technique relies heavily on estimates derived directly from the Lebesgue point property. The authors hypothesise that the Lebesgue point property is in fact necessary for the approximation of $f \in W^1 X$ by smooth functions (independent of injectivity).

1.1. A brief description of the proof of Theorem 1.2

In this subsection we outline the basic plan of our proof of Theorem 1.2. As suggested above the general concept is similar to that in [4, 8]. Assume that we have a homeomorphic and locally-finite piece-wise affine approximation of f. We can then approximate these homeomorphisms by diffeomorphisms using [11]. The diffeomorphisms from this result coincide with the original piece-wise affine homeomorphisms up to a tiny set. Further they have the same Lipschitz constant as the approximated map up to a bounded multiplicative constant. The combination of the above two facts with the absolute continuity of the norm of X means that the diffeomorphisms given by [11] also converge to the piece-wise affine homeomorphisms in $W^1 X$. The entire argument is in Lemma 2.9 and thanks to this, the question reduces to approximating by piece-wise affine homeomorphisms.

In Lemma 2.10 we create nested subdomains Ω_k of Ω and a grid of squares of a given size is made in each $\Omega_k \setminus \overline{\Omega}_{k-1}$ so that the following holds

- all the squares have the same size and fill all of $\Omega_k \setminus \overline{\Omega}_{k-1}$ except a set so small that the norm of the restriction of f to this set is bounded by $2^{-k}\varepsilon$,
- thanks to the Lebesgue point property for Df, f is very close to an affine function except for some squares whose union has measure so small that the norm of the restriction of $|Df|$ to this set is bounded by $2^{-k}\varepsilon$.

In general it is necessary to know that the behaviour of f is reasonable on the boundary of the squares. This is not automatically true but by slightly moving the boundaries of the squares it becomes true, which is achieved in Lemma 2.11.

Each of the squares can be split into 2 triangles by dividing along a diagonal (say the southwest-northeast diagonal). On the squares where f is very close to a nice affine map (Jacobian not too small, derivative not too big or too small) the map which is affine on each of the pair of triangles with values coinciding with f at the vertices of the square approximates f well. If the Jacobian is zero we use the result Theorem 2.7 to approximate. Now either the map is close to a constant on the given square or is not close to any linear map. In either case we use Theorem 2.4 to define our piece-wise affine approximation. In the last two cases the smallness either of Df or of the set is enough to make sure that the error is small.

Finally we have Lemma 2.13 to fill the small space around the boundary of Ω_k. Because the size of the set is so small we get that the norm of the map and the approximation is less than $2^{-k}\varepsilon$.

2. Preliminaries

In this section we shortly list the basic notation that will be used throughout the paper. The set $Q(c, r) = \{(x, y) \in \mathbb{R}^2 : |x - c_1| \leq r, |y - c_2| \leq r\}$ will denote the closed square centred at c with side length $2r$. Similarly,

$B(c, r) = \{(x, y) \in \mathbb{R}^2 : |(x, y) - c| < r\}$ is the open ball centred at c with radius r. For the ease of notation, for $t > 0$ we will denote $tQ(c, r) = Q(c, tr)$, and $tB(c, r) = B(c, tr)$.

Sometimes we will work on 1-dimensional objects in $\mathbb{R}^2$, which can be parametrised by a Lipschitz curve $\varphi \colon [0, 1] \to \mathbb{R}^2$, for example segments, and various polygons. We may assume that our φ is one-to-one and $|\varphi'|$ is constant almost everywhere. For almost all $t \in (0, 1)$ there exists a vector $\frac{\varphi'(t)}{|\varphi'(t)|}$ which we call the tangential vector at the point $\varphi(t)$ and denote this vector as $\tau = \tau(\varphi(t))$. If a mapping f is defined on $\varphi([0, 1])$ and $f \circ \varphi$ is absolutely continuous, then we call

$$D_\tau f = \frac{(f \circ \varphi)'}{|\varphi'|}$$

tangential derivative along the curve φ.

We take advantage of standard denotation of average integrals using the symbol $\fint$. Since we integrate with respect to different measures, we emphasise the fact that we divide the integral by the measure of the set we integrated over, *where we measure the set with the same measure used in the integral.*

Through out the whole paper we will assume without loss of generality (see *e.g.* [7], Thm. 5.22) that

$$J_f \geq 0 \text{ a.e.} \tag{2.1}$$

2.1. Rearrangement invariant function spaces

Here we collect all the background material that will be used in the paper.

Let E be a (non-negligibile) Lebesgue measurable subset of $\mathbb{R}^n$ of finite measure. We denote by $\mathcal{L}^n(E)$ its Lebesgue measure. We set

$$L^0(E) = \{f : f \text{ is measurable function on } E \text{ with values in } [-\infty, +\infty]\}$$

and

$$L^0_+(E) = \left\{f \in L^0(E) : f \geq 0\right\}.$$

From now on we shall identify functions f_1, f_2 for which $\mathcal{L}^n(\{f_1 \neq f_2\}) = 0$ in the space L^0. The non-increasing rearrangement $f^* \colon [0, +\infty] \to [0, +\infty]$ of a function $f \in L^0(E)$ is defined by

$$f^*(s) = \inf\left\{t \geq 0 : \mathcal{L}^n\big(\{x \in E : |f(x)| > t\}\big) \leq s\right\} \qquad s \in [0, +\infty)$$

and the *Hardy-Littlewood maximal function* $f^{**} : (0, \mathcal{L}^n(E)) \to [0, +\infty)$ is given by

$$f^{**}(t) := \frac{1}{t} \int_0^t f^*(s) \, \mathrm{d}s.$$

Definition 2.1. Let E be a Lebesgue measurable subset of $\mathbb{R}^n$ and let $\| \cdot \|_{X(E)} : L^0(E) \to [0, +\infty]$ be a functional. Consider the following properties

(P1) $\| \cdot \|_{X(E)}$ is a norm on $L^0(E)$.
(P2) For all $f, g \in L^0_+(E)$ the inequality $f(x) \leq g(x)$ for a.e. x in E implies
     $\|f\|_{X(E)} \leq \|g\|_{X(E)}$.
(P3) $\sup_k \|f_k\|_{X(E)} = \|f\|_{X(E)}$ if $0 \leq f_k(x) \nearrow f(x)$ for a.e. x in E.

(P4) Let $G \subset E$ be a set of a finite measure. Then

$$\|\chi_G\|_{X(E)} < \infty.$$

(P5) Let $G \subset E$ be a set of a finite measure. Then there exists a constant C_G depending only on the choice of the set G for which

$$\|f\chi_G\|_{L^1(E)} \le C_G \|f\chi_G\|_{X(E)}.$$

for all $f \in L^0(E)$.

(P6)

$$\|f\|_{X(E)} = \|g\|_{X(E)} \text{ whenever } f^* = g^* \text{(rearrangement invariance)}$$

If $\| \cdot \|_X$ enjoys the properties (P1)-(P5) we call it a Banach function norm. If it also enjoys (P6) we call it a rearrangement invariant Banach function norm. Let $\| \cdot \|_X$ be (rearrangement-invariant) Banach function norm then we call set

$$X(E) := \{f \in L^0(E) : \|f\|_{X(E)} < \infty\}$$

endowed with the norm $\| \cdot \|_{X(E)}$ a (rearrangement-invariant) Banach function space.

Following the properties (P2),(P4), (P5) one can observe that if $\mathcal{L}^n(E) < \infty$ for arbitrary Banach function space, the following holds true

$$L^\infty(E) \hookrightarrow X(E) \hookrightarrow L^1(E) \tag{2.2}$$

where $\hookrightarrow$ stands for a continuous embedding.

Given a r.i. Banach function space $X(E)$ and $0 \le s < \mathcal{L}^n(E)$ one may define the *fundamental function of* $X(E)$ by

$$\varphi_{X(E)}(s) := \|\chi_G\|_{X(E)} \tag{2.3}$$

where $G \subset E$ is an arbitrary subset of E of measure s.

The properties of r.i. norms guarantee that the fundamental function is well defined. For every r.i. space $X(E)$, its fundamental function φ_X is non-decreasing, $\varphi_{X(E)}(0) = 0$ and $\varphi_X(t)/t$ is non-increasing.

Given a Banach function space $X(E)$ define an associated Banach function space $X'(E)$ as a subspace of measurable functions endowed by associated norm given by

$$\|f\|_{X'(E)} := \sup_{\|g\|_X \le 1} \int_E fg \, \mathrm{d}x.$$

Note, that the associated space of a Banach function space is also a Banach function space and the following Hölder inequality holds

$$\int_E fg \, \mathrm{d}x \le \|f\|_{X(E)} \|g\|_{X'(E)}. \tag{2.4}$$

Let us remind the reader that for arbitrary r.i. Banach function space we have that

$$g^{**}(t) \le f^{**}(t) \quad \forall t \in (0, \mathcal{L}^n(E)) \text{ then } \|g\|_{X(E)} \le \|f\|_{X(E)}. \tag{2.5}$$

The proof of this classical result called Hardy-Littlewood-Polya inequality may be found for instance ([3], Thm. 4.6, Chap. 2, pg. 61).

Given a Banach function norm $\|\cdot\|_{X(E)}$ and a normed linear space Y we shall define

$$X(E, Y) := \{f \colon E \to Y : (x \mapsto \|f(x)\|_Y) \in X(E)\}.$$

Let $f \colon \mathbb{R}^n \to \mathbb{R}^k$ be a locally integrable function; we define the maximal operator of such a function by

$$Mf(x) := \sup_Q \frac{1}{\mathcal{L}^n(Q)} \int_Q |f(y)| \, \mathrm{d}y,$$

where the supremum on the right-hand side is taken over all cubes Q containing x. Let us also recall the *Riesz-Herz equivalence* that

$$(Mf)^*(t) \approx f^{**}(t) \quad \forall t \in (0, \infty), \tag{2.6}$$

with constants independent of t and f. For proof of this result see ([3], Thm. 3.8, Chap. 3, pg. 122). Using together Riesz-Herz equivalence (2.6) and Hardy-Littlewood-Polya inequality (2.5) yields the following *norm comparison*

$$Mf \leq Mg \qquad \Rightarrow \qquad \|f\|_{X(E)} \lesssim \|g\|_{X(E)} \tag{2.7}$$

Let $X(E)$ be a Banach function space. We say that $X(E)$ has *locally absolutely continuous norm* if for any finite measure set $M \subset E$ and any function $f \in X(E)$ one has

$$M \supset M_n \to \emptyset \quad \text{implies} \quad \|f\chi_{M_n}\|_X \to 0.$$

Note that if $\mathcal{L}^n(E) < \infty$ this property implies the ε-δ-continuity of the norm. This means that for any $f \in X$ and $\varepsilon > 0$ one can find $\delta > 0$ such that

$$\mathcal{L}^n(M) < \delta \text{ implies } \|f\chi_M\|_X < \varepsilon. \tag{2.8}$$

Let $E \subset \mathbb{R}^n$ be an open set and let $X(E)$ be a Banach function space. We define the Sobolev space over $X(E)$ by

$$W^1 X(E) := \{f \in W^{1,1}_{\mathrm{loc}}(E, \mathbb{R}^m) : \|f\|_{W^1 X(E)} = \|Df\|_{X(E)} + \|f\|_{X(E)}\}$$

where we use the standard operator norm to determine the size of $|Df|$.

Now follows a preparatory lemma.

Lemma 2.2. *Let $G \subset \mathbb{R}^n$ be a set of finite measure and $X(G)$ be a r.i. BFS space satisfying*

$$\lim_{t \to 0+} \varphi_X(t) = 0. \tag{2.9}$$

Then for every $M > 0$ and $\tilde{\varepsilon} > 0$ there exists a $\tilde{\delta} > 0$ such that for all $u \in X(G)$ with $\|u\|_{L^\infty(G)} \leq M$ and $\|u\|_{L^1(G)} < \tilde{\delta}\mathcal{L}^n(G)$ one has $\|u\|_{X(G)} < \tilde{\varepsilon}\mathcal{L}^n(G)$.

Proof. For arbitrary $D > 0$ we have

$$\begin{aligned}
\|u\|_{X(G)} &\leq \|u\chi_{\{|u|>D\}}\|_{X(G)} + \|u\chi_{\{|u|\leq D\}}\|_{X(G)} \\
&\leq M\varphi_X\big(\mathcal{L}^n(\{|u|>D\})\big) + DC_{L^\infty \to X} \\
&\leq M\varphi_X\left(\frac{\tilde{\delta}\mathcal{L}^n(G)}{D}\right) + DC_{L^\infty \to X} \\
&= (1) + (2)
\end{aligned}$$

where $C_{L^\infty \to X}$ stands for the optimal constant of imbedding of $L^\infty(G)$ into $X(G)$. The last inequality follows from

$$D\mathcal{L}^n(\{|u|>D\}) \leq \|u\|_{L^1(G)} \leq \tilde{\delta}\mathcal{L}^n(G)$$

hence

$$\mathcal{L}^n\big(\{|u|>D\}\big) \leq \frac{\tilde{\delta}\mathcal{L}^n(G)}{D}.$$

First choose D such that $(2) < \tilde{\varepsilon}\frac{\mathcal{L}^n(G)}{2}$. Then choose $\tilde{\delta}$ such that $(1) < \tilde{\varepsilon}\frac{\mathcal{L}^n(G)}{2}$. $\qquad\square$

Remark 2.3. If X has the Lebesgue point property, we have by Proposition 3.1 of [5], that the norm is locally absolutely continuous. Therefore one has (2.9) and Lemma 2.2 can be applied for such a space.

2.2. The reformulation of known extension results

In this section our aim is to prove the following extension theorem, which will allow us to construct homeomorphisms from boundary values and gives us a useful control on their Lipschitz constant.

Theorem 2.4. *There exists a $C > 0$ such that for any $r > 0$ and any finitely piece-wise linear and one-to-one function $\varphi : \partial Q(0, r) \to \mathbb{R}^2$ we can find a finitely piece-wise affine homeomorphism $h : Q(0, r) \to \mathbb{R}^2$ such that*

$$\|Dh\|_{L^\infty(Q(0,r))} \leq C_{(2.10)}\|D_\tau\varphi\|_{L^\infty(\partial Q(0,r))}, \tag{2.10}$$

and

$$h|_{\partial Q(0,r)} = \varphi.$$

In (2.10) the L^∞ space on the left is with respect to the two dimensional Lebesgue measure $\mathcal{L}^2$ and the L^∞ space on the right is with respect to the one dimensional Hausdorff measure $\mathcal{H}^1$.

Proof. The construction is precisely that of Hencl and Pratelli from Theorem 2.1 of [8] later expanded upon in [4, 15]. The construction starts by making a mapping that is not injective (only monotone) and then in the end making a small adjustment to make the mapping injective.

Step 1. A square with 'good' corners.

We construct our map on a rotation of the square by 45 degrees, we call this rotated square Q_r. For each $t \in (-\sqrt{2}r, \sqrt{2}r)$ denote the corresponding pair of horizontally opposite points on the rotated square $O = O_t = (|t|, t), P = P_t = (-|t|, t)$. For $t \neq 0$ we call $\widehat{OP}$ the path $[O, (0, \operatorname{sgn}(t)\sqrt{2}r)] \cup [(0, \operatorname{sgn}(t)\sqrt{2}r), P]$ the shorter of the 2 paths from O to P along the boundary of Q_r. It was shown in Theorem 2.1, Step 1 of [8] that (up to a

bi-Lipschitz transformation) one may assume that there exists a C such that for any $t \in (-\sqrt{2}r, \sqrt{2}r)$ it holds that

$$\fint_{\widehat{OP}} |D_\tau \varphi| \, d\mathcal{H}^1 \le C \fint_{\partial Q_r} |D_\tau \varphi| \, d\mathcal{H}^1.$$

For $t = 0$ the claim holds trivially for $C = 2$, for whichever path in ∂Q_r we choose. The details are to be found in Theorem 2.1, Step 1 of [8]. The argument is that given $\mathcal{A}$ the set of points $P \in \partial Q_r$ such that one can find an 'arc' L in ∂Q_r of length $2d$ symmetrically around P such that

$$\fint_L |D_\tau \varphi| \, d\mathcal{H}^1 > 6 \fint_{\partial Q_r} |D_\tau \varphi| \, d\mathcal{H}^1$$

is not very big. In fact using a Vitali covering of $\mathcal{A}$ one can estimate that

$$\mathcal{H}^1(\mathcal{A}) \le \sum_i 6r_i \le \sum_i \frac{\int_{L_i} |D_\tau \varphi|}{\int_{\partial Q_r} |D_\tau \varphi|} < 1 < \tfrac{1}{2}\mathcal{H}^1(Q_r)$$

and since $\mathcal{A}$ covers less than half ∂Q_r one can find a pair of opposing points in $\partial Q_r \setminus \mathcal{A}$. Any such pair can be mapped onto the north and south poles of ∂Q_r by a finitely piece-wise affine map, which preserves arc length on ∂Q and the bi-Lipschitz constant is independent of the choice of the pair.

Step 2. Definition of h and estimate of $|Dh|$.

We define $h = \varphi$ on ∂Q_r. By 'h-vertex' of ∂Q_r we refer to a point $P \in \partial Q_r$ such that $D_\tau \varphi$ does not exist. For every P a h-vertex of ∂Q_r we define h on the segment $[PO]$ (where we denote $O = (-P_1, P_2)]$) in such a way that the image of $[PO]$ in h is the geodesic from $\varphi(P)$ to $\varphi(O)$ inside the closure of the bounded component of $\mathbb{R}^2 \setminus \varphi(\partial Q_r)$ parametrized at constant speed. The geodesic is a polyline and this way h is piecewise linear on the union of ∂Q_r and the added horizontal lines. Using the fact that the geodesic has length bounded by $\mathcal{H}^1(\varphi(\widehat{PO}))$, that h has constant speed on the horizontal segment and that $\mathcal{H}^1([PO]) = \frac{1}{\sqrt{2}}\mathcal{H}^1(\widehat{PO})$ we get that

$$|D_1 h| = \frac{\mathcal{H}^1\big(h([PO])\big)}{\mathcal{H}^1\big([PO]\big)} \le \sqrt{2}\fint_{\widehat{PO}} |D_\tau \varphi| \le C \fint_{\partial Q_r} |D_\tau \varphi| \tag{2.11}$$

on $[PO]$, where the last estimate is from step 1.

Let P_1 and P_2 be a pair of adjacent h-vertices on ∂Q_r and let O_1, O_2 be the corresponding pair of horizontally opposing points in ∂Q_r. Call the strip S the set between a pair of neighbouring horizontal lines, *i.e.* $S = \mathrm{co}\{P_1, P_2, O_1, O_2\}$, where co denotes the convex hull. We have defined h on ∂S. By F_S we denote the union of $h(\partial S)$ with the set disconnected from infinity by $h(\partial S)$.

We simply separate S into triangles $A_1 A_2 A_3$ with $[A_1 A_2]$ lying on one of the horizontal segments and $[A_2 A_3]$ is vertical and all of A_1, A_2, A_3 lie on one of the horizontal segments $[P_1 O_1]$ or $[P_2 O_2]$. Since the images of the corners $A_1 A_2 A_3$ in h have been defined there is exactly one affine map sending A_i onto $h(A_i)$ and we define h as this affine map on each triangle $\mathrm{co}\{A_1, A_2, A_3\}$. Then immediately from (2.11) we get

$$|D_1 h| \le C \fint_{\partial Q_r} |D_\tau \varphi| \le C\|D_\tau \varphi\|_{L^\infty(\partial Q_r)}. \tag{2.12}$$

A careful analysis of the geometry of the geodesics, fully exposed in Theorem 2.1, Step 4,5,6 of [8] proves that these affine images of $A_1 A_2 A_3$ lie in F_S and provides an estimate on $|h(A_2) - h(A_3)|$. We summarize the steps from Theorem 2.1, Step 4,5,6 of [8] in the following lemma. The key length estimate (2.13) is estimate (2.4) of [8].

Lemma 2.5. *Let P_1 and P_2 be a pair of adjacent h-vertices on ∂Q_r and let O_1, O_2 be the neighbouring pair of horizontally opposing points in ∂Q_r. Call $S = \text{co}\{P_1, P_2, O_1, O_2\}$ and F_S is the union of $h(\partial S)$ with the part(s) of the plane it disconnects from infinity. Then $h([A_2A_3]) \subset F_S$ and*

$$\mathcal{H}^1(h([A_2A_3])) \leq \max\{\mathcal{H}^1\big(\varphi([P_1P_2])\big), \mathcal{H}^1\big(\varphi([O_1O_2])\big)\}. \tag{2.13}$$

But then from (2.13), on the triangle $\text{co}\{A_1, A_2, A_3\}$ we have

$$\begin{aligned}
|D_2 h| &= \frac{\mathcal{H}^1(h([A_2A_3]))}{\mathcal{H}^1([A_2A_3])} \\
&\leq \sqrt{2}\frac{\max\{\mathcal{H}^1\big(\varphi([P_1P_2])\big), \mathcal{H}^1\big(\varphi([O_1O_2])\big)\}}{|P_1 - P_2|} \\
&\leq \sqrt{2}\|D_\tau \varphi\|_{L^\infty(\partial Q_r)}
\end{aligned} \tag{2.14}$$

This holds on all triangles $\text{co}\{A_1, A_2, A_3\}$ in all strips S between neighbouring horizontal lines. The triangle at the north and south pole of Q_r is estimated similarly.

Step 3. Injectification of h.

The final step is to replace the image of the lines $[PO]$ with something very close to the geodesics but so that the images never meet. It suffices to shift the image of each $[PO]$ in h slightly inside ∂Q_r at leach corner of $\varphi(\partial Q_r)$. As long as this change is very small then all of the estimates remain intact (see also [8], Thm. 2.1, Step 10).

The claim (2.10) follows from (2.12) and (2.14). $\qquad \square$

The following Corollary is an immediate result of Theorem 2.4. It is the fact that we get an L^∞ bound from the L^1 space that our approach works in the relatively general setting.

Corollary 2.6. *There exists a constant $C > 0$ such that for every $r > 0$ and $\varphi : \partial Q(0, r) \to \mathbb{R}^2$ finitely piece-wise linear and one-to-one function with $|D_\tau \varphi|$ constant on each side of $Q_r = Q(0, r) \subset \mathbb{R}^2$, there exists a piece-wise affine homeomorphism $h : Q_r \to \mathbb{R}^2$ such that*

$$\|Dh\|_{L^\infty(Q_r)} \leq \frac{4C_{(2.10)}}{r} \int_{\partial Q_r} |D_\tau \varphi| \, d\mathcal{H}^1. \tag{2.15}$$

and

$$h|_{\partial Q_r} = \varphi.$$

Proof. If $|D_\tau \varphi|$ is constant on sides, then $\|D_\tau \varphi\|_{L^\infty(\partial Q_r)} \leq \frac{4}{r} \int_{\partial Q_r} |D_\tau \varphi| \, d\mathcal{H}^1$. $\qquad \square$

Further we will reformulate ([4], Thm. 3.7) to fit in with our current setting better. The question is how to approximate a map which is close to a degenerate linear map Φ (up to a rotation in the pre-image we may assume that $\Phi = \begin{pmatrix} d, 0 \\ 0, 0 \end{pmatrix}$).

Theorem 2.7. *Let $d > \delta > 0$, let $r_0 \in (0, 1)$ and let Q be a convex set and the image of $[0, r_0]^2$ in a 2-bi-Lipschitz mapping which is equal to an affine mapping on $\text{co}\{(0,0), (0, r_0), (r_0, 0)\}$ and $\text{co}\{(r_0, r_0), (0, r_0), (r_0, 0)\}$. Then for every $\varphi : \partial Q \to \mathbb{R}^2$ finitely piece-wise linear and one-to-one mapping with*

$$\int_{\partial Q} \left| D_\tau \varphi(t) - \begin{pmatrix} d, 0 \\ 0, 0 \end{pmatrix} \tau \right| \, d\mathcal{H}^1(t) < \delta r_0, \tag{2.16}$$

and $\|D_\tau\varphi\|_{L^\infty(\partial Q)} \leq d + 2\delta$, where τ is the unit tangential vector to ∂Q, there exists a finitely piece-wise affine homeomorphism $g : Q \to \mathbb{R}^2$ and a set $W \subset Q$ such that $\mathcal{L}^2(Q \setminus W) < C\delta r_0^2$, $g = \varphi$ on ∂Q,

$$
\begin{aligned}
\left\| Dg(x) - \begin{pmatrix} d, 0 \\ 0, 0 \end{pmatrix} \right\|_{L^\infty(Q)} &< C(d+1), \\
\left\| Dg(x) - \begin{pmatrix} d, 0 \\ 0, 0 \end{pmatrix} \right\|_{L^1(W)} &< C\delta r_0^2 \text{ and} \\
\left\| Dg(x) - \begin{pmatrix} d, 0 \\ 0, 0 \end{pmatrix} \right\|_{L^\infty(W)} &< 3d.
\end{aligned}
\tag{2.17}
$$

Proof.
Step 1. Definition of W such that $\mathcal{L}^2(Q \setminus W) < C\delta r_0^2$.

Since Q is the convex 2-affine image of a square, it makes sense to talk about its vertices. Calling A_i the vertices of Q and $A = \frac{1}{4}\sum_{i=1}^4 A_i$ the centre of Q, we define

$$
B_i = B_i(\delta) = A_i + 10\delta r_0(A - A_i).
$$

We call $W = W(\delta) = \text{co}\{B_1, B_2, B_3, B_4\}$. Then it follows that $|Q \setminus W(\delta)| \leq C\delta r_0^2$.
Step 2. Definition of a curve γ that goes from close to one end of $\varphi(\partial Q)$ to the other.

We find the point $P \in \partial Q$ such that P is a point where x^1 achieves its minimum for $x \in \partial Q$ and $O \in \partial Q$ is a point where x^1 achieves its maximum for $x \in \partial Q$. We call $\widehat{PO}$ the shortest path in ∂Q connecting P and O. If there are two paths then choose either of them. From (2.16) we have that

$$
\int_{\widehat{PO}} \left| D_\tau\varphi(t) - \begin{pmatrix} d, 0 \\ 0, 0 \end{pmatrix} \tau \right| \, \mathrm{d}\mathcal{H}^1(t) < \delta r_0.
\tag{2.18}
$$

We define a piecewise linear path in the interior of $\varphi(\partial Q)$ following $\varphi(\partial Q)$ from its start very close to $\boldsymbol{P} = \varphi(P)$ to its end very close to $\boldsymbol{O} = \varphi(O)$. Call this path γ. The idea how to do this is simple. At every vertex of $\varphi(\partial Q)$ we bisect the interior angle (the bisector goes inside the bounded component of $\mathbb{R}^2 \setminus \varphi(\partial Q)$) and place a point on the bisector with distance to the vertex as small as required. Then we form a polyline by connecting the points we constructed in the same order as the vertices (see [4], Lem. 3.1 for a precise construction). By placing the points we constructed on the bisectors very close to their respective vertices we ensure that the length of the curve differs from the length along the boundary by a number as small as required.
Step 3. Definition of g on W and estimates.

We define the function $l : W \to \mathbb{R}^2$ as follows. Let γ denote the constant speed parametrization of the curve γ from $[0, 1]$. Then define

$$
l(x, y) = \gamma\left(\frac{x - P_1}{O_1 - P_1}\right).
$$

Now it is not hard to check using (2.18) that

$$
\int_W \left| Dl - \begin{pmatrix} d, 0 \\ 0, 0 \end{pmatrix} \right| \, \mathrm{d}\mathcal{L}^2 < C\delta r_0^2.
\tag{2.19}
$$

The construction of g on W is a slight modification of l so that the map is injective. This is just a question of decomposing W into quadrilaterals of type $W \cap \{x; a < x < b\}$ and instead of mapping them onto the curve γ we map them onto a tiny tubular neighbourhood of γ. On each of these quadrilaterals we use a bi-affine map.

The details are in Theorem 3.7, step 3 of [4]. Most importantly we get the estimate on the derivative

$$\int_W \left| Dg - \begin{pmatrix} d, 0 \\ 0, 0 \end{pmatrix} \right| \mathrm{d}\mathcal{L}^2 < C\delta r_0^2 \text{ and } \left| Dg - \begin{pmatrix} d, 0 \\ 0, 0 \end{pmatrix} \tau \right| \leq 3d \tag{2.20}$$

immediately from (2.19), the fact that $|D_1 l|$ is constant and g is as close to l in $W^{1,\infty}$ as we like.

Step 4. Extension of g onto $Q \setminus W$ and estimates.

The remaining set, $Q \setminus W$, is a tubular neighbourhood of ∂Q in the direction inside of width approximately δr_0. Therefore it can be divided into small quadrilaterals all 2-biLipschitz equivalent with a square, which we denote by U_i.

On each ∂U_i we define a piecewise linear mapping φ_i as follows. On $\partial U_i \cap \partial W$ put $\varphi_i = g$ and on $\partial U_i \cap \partial Q$ put $\varphi_i = \varphi$. Then we need to define φ on the 'radial' segments. This equates to the following task: define mutually non-intersecting polylines from points on $g(\partial W)$ to corresponding points on $\varphi(\partial Q)$ lying inside the bounded component of $\mathbb{R}^2 \setminus (\varphi(\partial Q) \cup g(W))$. This is obviously a task that has a solution. What is more important is that we have a bound on the length of each polyline, which was proved in Theorem 3.7, step 5, 6 of [4] and which we summarize in the following lemma. The key length estimate (2.21) is (3.27) of [4].

Lemma 2.8. *Let $0 < \delta < \frac{1}{16}$, $w \in \partial W(\delta)$ and $x \in \partial Q$ with $|w - x| < 4\delta r_0$ then there exists a polyline $p_{x,w}$ connecting $\boldsymbol{x} = g(x) = \varphi(x)$ with $\boldsymbol{w} = g(w)$ lying inside the bounded component of $\mathbb{R}^2 \setminus (\varphi(\partial Q) \cup g(W))$ (except for the endpoints) and it holds that*

$$\mathcal{H}^1(p_{x,w}) \leq (d+1)|w - x| + C\delta r_0 \tag{2.21}$$

with C independent on the choice of x and w.

Proof of Lemma 2.8. For simplicity of the argument let us assume that x and w are points on the top of W and Q respectively. In the case that the points lie on the bottom or on a side that is very close to being vertical then the argument is the basically the same and we emphasize the difference when it is relevant.

We start by defining a tentative polyline that we later alter to ensure injectivity. By Lemma 3.5 of [4] it suffices to follow $g(\partial W)$ for a distance of at most $C\delta r_0$ till we find a point $\boldsymbol{w}' \in g(W)$ such that the that the vertical segment going upwards (downwards for points on the bottom of W resp. Q and either case holds for points on sides very close to vertical) from $\boldsymbol{w}'$ does not intersect $g(W)$ but only $g(\partial Q)$ at a point, which we call $\boldsymbol{x}'$. The distance from $\boldsymbol{x}'$ to $\boldsymbol{x}$ along $g(\partial Q)$ is estimated from above by $|\boldsymbol{x} - \boldsymbol{x}'| + \delta r_0$ using (2.16). But $|\boldsymbol{x} - \boldsymbol{x}'| \leq d|w - x| + C\delta r_0$. Therefore there is a polyline from $\boldsymbol{x}$ to $\boldsymbol{w}$ in the closure of the interior of $g(\partial Q)$ not intersecting $g(W^\circ)$ with length $Cd\delta r_0$ with C independent on the choice of $\boldsymbol{x}$ and $\boldsymbol{w}$. The polyline can be moved inside the bounded component of $\mathbb{R}^2 \setminus (\varphi(\partial Q) \cup g(W))$ except for the endpoints making it only ε longer with ε as small as we like. $\qquad\square$

It is a standard technique (used to show the uniqueness of shortest paths in $\mathbb{R}^2$) that allows us to show that if two polylines from Lemma 2.8 intersect then one can redefine them so that they do not touch and the length estimate still holds.

Now we define φ_i as the constant speed parametrization of $p_{x,w}$ on each 'radial' segment of a 'square' U_i. Then we calculate that

$$D_\tau \varphi_i(t) \leq \begin{cases} 2d & t \in \partial Q \\ 2d & t \in \partial W \\ C(d+1) & t \text{ on radial segments of } \partial U_i \end{cases}$$

because $\mathcal{H}^1(p_{x,w}) \leq C(d+1)\delta r_0$ and the length of the radial segments are approximately $C\delta r_0$. Then applying Theorem 2.4 on each U_i we define g on $Q \setminus W$ such that $\|Dg\|_{L^\infty(Q)} \leq C(d+1)$. The fact that $g = \varphi$ on ∂Q is immediate from our definition. $\qquad\square$

Lemma 2.9. *Let $X(\Omega)$ be a r.i. BFS space such that*

$$\lim_{t \to 0} \varphi_{X(\Omega)}(t) = 0.$$

Let $\hat{f} \in W^1 X(\Omega, \mathbb{R}^2)$ be a locally finite piece-wise affine homeomorphism. Then for every $\varepsilon > 0$ there exists a diffeomorphism $\tilde{f}$ such that

$$\|D\tilde{f} - \mathrm{d}\hat{f}\|_{X(\Omega)} < \varepsilon$$

and

$$\|\tilde{f} - \hat{f}\|_{L^\infty(\Omega)} < \varepsilon.$$

Moreover, if $\hat{f}$ is continuous up to the boundary of Ω, then $\tilde{f}$ can be chosen to be continuous up to the boundary of Ω and $\tilde{f} = \hat{f}$ on $\partial\Omega$.

Proof. We denote by $\hat{f}$ the countably and locally-finite piece-wise homeomorphism and by $\tilde{f}$ a diffeomorphism constructed as in [11]. The claim that

$$\|\tilde{f} - \hat{f}\|_{L^\infty(\Omega)} < \varepsilon$$

is part of the claim of Theorem A in [11]. We only need to show that the arguments in [11] extend to handling the norm $\|\cdot\|_{X(\Omega)}$ instead of $\|\cdot\|_{L^p(\Omega)}$ for the derivatives. We do not focus on results for the inverse or the determinant.

The construction from [11] works on a small neighbourhood of the the edges and vertices of each triangle. Around edges of triangles (in [11] referred to as Z_3), but distant from the vertices it suffices to take an appropriate, smooth convex combination of the two affine maps, which meet at the edge. This gives a diffeomorphism on a (small tubular) neighbourhood of the edge, which coincides with $\hat{f}$ away from the edge.

On small disks close to vertices one does two separate steps. One works on an outer annulus (in [11] referred to as Z_2) and a disk inside (in [11] referred to as Z_1). Assume that we have already smoothed close to the edges ending at the vertex. On the outer annulus one does a convex combination (with respect to r) with a smooth map, which expressed in polar coordinates (r, θ) has the same 'argument' θ as the piecewise affine map (after smoothing near the edges) but brings all the points closer to the vertex in such a way that it sends circles onto circles. On the disk inside the outer annulus we smoothly transition to a small multiple of identity. This is possible since circles are sent to circles and it suffices to appropriately rotate the circles and smoothly take the angular speed (of the map expressed in polar coordinates in the preimage and image) to 1. In this case we are equal to a translation plus a small multiple of the identity near each vertex.

In each case the neighbourhoods of the edges (Z_3) and vertices (Z_1, Z_2) can be made as small as we like. On the rest of the space (in [11] referred to as Z_4) we have $\tilde{f} = \hat{f}$. To reiterate [11] we quote the final paragraph of their Proof of Theorem A they say "Finally, the set $[\{y : \tilde{f}(y) \neq \hat{f}(y)\}]$ can be made as small as we wish, by decreasing the constants δ_a, δ_b and δ_c as needed".

The mapping $\hat{f}$ is piecewise affine on Ω, which means there is a set of triangles $\{T_i\}_{i \in I}$ covering Ω and $\hat{f}$ is affine on each T_i. For each T_i, by construction we have that there exists a constant $M_i = 13 \max_{t \in J(x)} |\mathrm{d}\hat{f}(t)|$, where $J(x)$ is the union of T_i containing x with all triangles T_j which intersect T_i (take $q = 1$ in [11], Thm. A) such that

$$\|D\hat{f}\|_{L^\infty(T_i)} \leq M_i, \quad \|D\tilde{f}\|_{L^\infty(T_i)} \leq M_i.$$

Obviously,

$$\|D\hat{f} - D\tilde{f}\|_{X(\Omega)} \leq \sum_{i \in I} \left\| \left(\mathrm{d}\hat{f} - D\tilde{f} \right) \chi_{T_i} \right\|_{X(\Omega)}$$

Moreover, for each $i \in I$

$$|\mathrm{d}\hat{f} - D\tilde{f}|\chi_{T_i} \leq \left(|\mathrm{d}\hat{f}| + |D\tilde{f}| \right) \chi_{T_i \setminus Z_4^i} \leq 2M_i \chi_{T_i \setminus Z_4^i}$$

where we denote by Z_4^i the intersection of the set where $\hat{f} = \tilde{f}$ (in [11] referred to as Z_4) with T_i.

Therefore

$$\left\| \left(\mathrm{d}\hat{f} - D\tilde{f} \right) \chi_{T_i} \right\|_{X(\Omega)} \leq 2M_i \, \varphi_{X(\Omega)} \left(\mathcal{L}^2(T_i \setminus Z_4^i) \right).$$

We can assume that $\mathcal{L}^2(T_i \setminus Z_4^i)$ is so small that

$$\varphi_{X(\Omega)} \left(\mathcal{L}^2(T_i \setminus Z_4^i) \right) < \frac{\varepsilon}{2M_i} 2^{-i}$$

and thus we conclude

$$\|\mathrm{d}\hat{f} - D\tilde{f}\|_{X(\Omega)} \leq \varepsilon.$$

$\square$

2.3. Preliminary results on grids and approximations on grids

Let us remind the reader that by *dyadic squares* we mean the family $\mathcal{D} = \bigcup_{k \in \mathbb{Z}} \mathcal{D}_k$, where

$$\mathcal{D}_k = \{Q((2^k, 2^k) + z2^{k+1}, 2^k); z \in \mathbb{Z}^2\},$$

where the square $Q((a,b), r) = [a-r, a+r] \times [b-r, b+r]$ is closed. Given $v \in \mathbb{R}^2$, by $\mathcal{D}_k^v$ we denote

$$\mathcal{D}_k^v = \{v + Q; Q \in \mathcal{D}_k\}$$

For a domain Ω we say that an open set G *separates components of* $\partial\Omega$ if any continuous curve connecting different components of $\partial\Omega$ has to intersect G. We say that the domain $\Omega \subset \mathbb{R}^2$ is *finitely connected* if $\mathbb{R}^2 \setminus \Omega$ has finite number of components.

Lemma 2.10. *Let $\Omega \subset \mathbb{R}^2$ be a finitely connected bounded domain and let $f \in W_{loc}^{1,1}(\Omega, \mathbb{R}^2)$ be a homeomorphism. Then there exists a strictly increasing sequence of sets $\Omega_k \Subset \Omega_{k+1} \Subset \Omega$ such that $\Omega = \bigcup_k \Omega_k$, also Ω_k separates components of $\partial\Omega$ and $\partial\Omega_k$ is piece-wise linear and parallel to coordinate axes and $\partial\Omega_k$ has the same number of components as $\partial\Omega$.*

Further, for every $k \in \mathbb{N}$ and $\varepsilon_k, \delta_k > 0$ and any set A_k with $\mathcal{L}^2(A_k) < \frac{\delta_k}{32}$ there exists an $m_k \in \mathbb{N}$, and $v_k \in Q(0, 2^{-m_k-1})$ and a collection of K_k shifted dyadic squares $\{Q_i^k = Q(c_i + v_k, 2^{-m_k}) \in \mathcal{D}_{-m_k}^{v_k}\}_{i=1}^{K_k}$ such that

i) $Q_i^k \Subset \Omega_k \setminus \Omega_{k-1}$, *and $\bigcup_{i=1}^{K_k} Q_i^k$ has no holes (i.e. in each component of $\Omega_k \setminus \Omega_{k-1}$ there are precisely 2 components of $\Omega_k \setminus \bigcup_{i=1}^{K_k} Q_i^k$),*

ii) it holds that

$$\mathcal{L}^2\Big(\Omega_k \setminus \Big[\Omega_{k-1} \cup \bigcup_{i=1}^{K_k} Q_i^k\Big]\Big) < \delta_k,$$

iii) for each $1 \le j \le K_k$ the square Q_j^k shares at least two of its sides with other squares of $\{Q_i^k\}_{i=1}^{K_k}$,

iv) the set $\partial\bigcup_{i=1}^{K_k} Q_i^k$ consists of segments and there is a one-to-one correspondence between the endpoints of these segments and vertices of $\partial(\Omega_k \setminus \Omega_{k-1})$ i.e. for every X vertex of $\partial(\Omega_k \setminus \Omega_{k-1})$ there is exactly one vertex of $\partial\bigcup_i Q_i^k$ in $B(X, 2^{4-m_k})$ and vice versa,

v) for every $x \in \partial\bigcup_{i=1}^{K_k} Q_i^k$ we have

$$6 \cdot 2^{-m_k} \le \mathrm{dist}_\infty(x, \partial(\Omega_k \setminus \Omega_{k-1})) \le 8 \cdot 2^{-m_k},$$

vi) there exists a set of indexes B_k such that $\mathcal{L}^2(\bigcup_{i \in B_k} Q_i^k) < \delta_k$ and moreover for every $i \notin B_k$,

$$c_i + v_k \notin A_k$$

and

$$\|f(x) - f(c_i + v_k) - Df(c_i + v_k)(x - c_i - v_k)\|_{L^\infty(2Q_i^k)} < \varepsilon_k 2^{-m_k-1},$$

$$\fint_{2Q_i^k} |Df(x) - Df(c_i + v_k)|\,\mathrm{d}\mathcal{L}^2(x) < \tfrac{\varepsilon_k}{4}, \tag{2.22}$$

$$\tfrac{1}{\mathcal{L}^2(Q_i^k)}\|(Df - Df(c_i + v_k))\chi_{2Q_i^k}\|_{X(\Omega)} < \varepsilon_k$$

vii) for every $i = 1, \ldots, K_k$ we have $\mathrm{diam}(f(2Q_i^k)) \le \varepsilon_k$.

Proof. Our choice of the sets Ω_k is inductive. It is illustrated in Figure 1. We start by choosing Ω_1. After we have done that we assume that we have an Ω_{k-1} and its corresponding parameter l_{k-1} and then choose Ω_k based on the previous set.

For every $l \in \mathbb{N}$ we define $\mathcal{W}_{-l}$ as the set of squares $Q \in \mathcal{D}_{-l}$ such that $16Q \subset \Omega$. As l tends to infinity these sets fill Ω and because Ω is finitely connected there exists an l_0 such that the set

$$\tilde{\Omega}_1 = \bigcup_{Q \in \mathcal{W}_{-l_0}} Q$$

separates components of $\partial\Omega$ (*i.e.* any continuous curve connecting different components of $\partial\Omega$ intersect $\tilde{\Omega}_1$). The set $\tilde{\Omega}_1$ is not necessarily connected. If $\tilde{\Omega}_1$ is connected, we set $\Omega_1 = \tilde{\Omega}_1$. In the opposite case, we argue as follows. Assume that there are N_1 components, we choose a point y_n^1 in each of the components. Each pair y_1^1 and y_n^1, $n = 2, \ldots, N_1$ are path-wise connected in Ω. Call γ_n^1 a path connecting y_n^1 with y_1^1 inside Ω. Now set

$$l_1 := \min\Big\{l \in \mathbb{N} \ : \ l > l_0 \text{ and } \gamma_j^1 \subset \bigcup_{\mathcal{W}_{-l}} Q \quad \text{for } 2 \le j \le N_1\Big\}.$$

Call $\tilde{C}_1$ the union of all Q such that

$$Q \in \Big\{\tilde{Q} \in \mathcal{D}_{-l_1} \ : \ 4\tilde{Q} \cap \gamma_j^1 \ne \emptyset \quad \text{for } 2 \le j \le N_1\Big\}$$
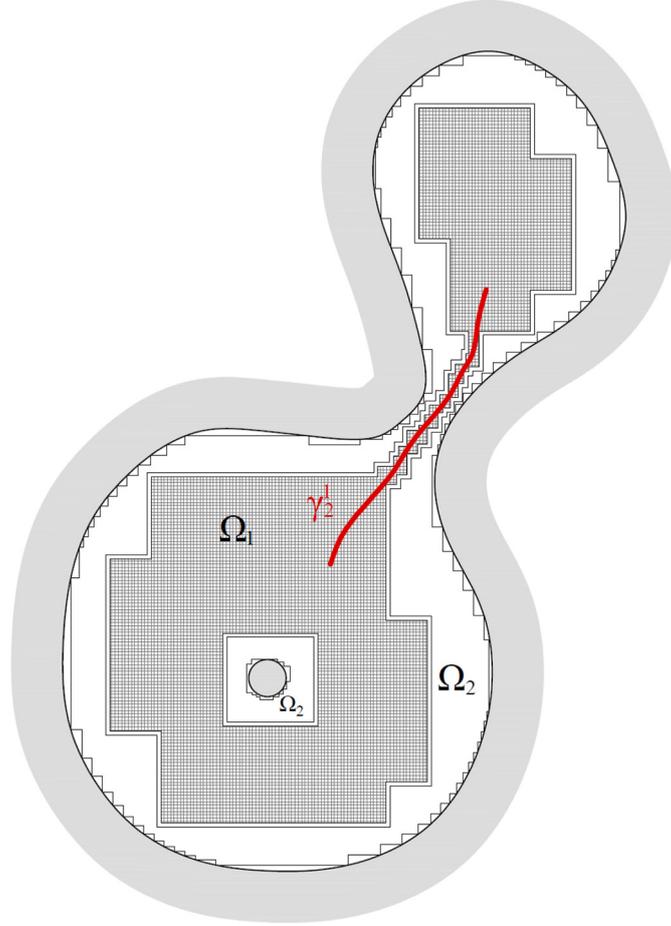
FIGURE 1. The sets Ω_1 and Ω_2 inside Ω a doubly connected domain. The squares Q_i^1 are also depicted and the curve γ_2^1.

then we define the set

$$C_1 = \tilde{\Omega}_1 \cup \tilde{C}_1.$$

Also we call D_1 the union of the squares Q such that

$$Q \in \left\{ \tilde{Q} \ : \ \tilde{Q} \in \mathcal{D}_{-l_1} \text{ and } \tilde{Q} \text{ is disconnected from } \partial\Omega \text{ by } C_1 \right\}.$$

By A° we denote the topological interior of A. Note that the set

$$\Omega_1 := (C_1 \cup D_1)^\circ$$

is a finitely connected domain and every component of $\mathbb{R}^2 \setminus \Omega_1$ contains exactly one component of $\partial\Omega$.

Continuing to the induction step, let us suppose that we have defined l_{k-1} and Ω_{k-1}. Set

$$\tilde{\Omega}_k = \bigcup_{Q \in \mathcal{W}_{-l_{k-1}-5}} Q.$$

If $\tilde{\Omega}_k$ is not connected then continue as before. Call N_k the number of components of $\tilde{\Omega}_k$ and let γ_n^k, $n = 2, 3, \ldots, N_k$ be paths in Ω connecting each component of $\tilde{\Omega}_k$ to one of the given components. We find an $l_k > l_{k-1} + 5$ such that we can cover each γ_n^k with dyadic squares from $\mathcal{W}_{-l_k}$. We call $\tilde{C}_k$ the union of squares Q such that

$$Q \in \left\{ \tilde{Q} \in \mathcal{D}_{-l_k} \ : \ 4\tilde{Q} \cap \gamma_j^k \neq \emptyset \quad \text{for } 2 \leq j \leq N_k \right\}$$

and use this to define C_k as follows

$$C_k = \tilde{\Omega}_k \cup \tilde{C}_k.$$

Further we define D_k as the union of the squares Q such that

$$Q \in \left\{ Q \in \mathcal{D}_{-l_k} \ : \ Q \text{ is disconnected from } \partial\Omega \text{ by } C_k \right\}$$

for squares that were surrounded by C_k in order to define

$$\Omega_k := C_k \cup D_k.$$

Thus we define Ω_k, $k \in \mathbb{N}$ inductively. The set Ω_k is a union of some squares in $\mathcal{D}_{-l_k}$ with distance to the boundary of at least $2 \cdot 2^{-l_k}$. On the other hand $\tilde{\Omega}_{k+1}$ contains the union of all squares in $\mathcal{W}_{-l_k-5}$ and this contains all points with distance to the boundary of 2^{-l_k} and therefore is a strictly larger set than $\mathcal{W}_{-l_k}$. Hence it is easy to deduce that $\Omega_k \Subset \tilde{\Omega}_{k+1} \subset \Omega_{k+1}$ and so this decomposition satisfies our requirements for Ω_k.

Now let us choose $k \in \mathbb{N}$ and construct the collection of cubes $\{Q_i^k\}$ in the claim. Either Ω is simply connected and Ω_k has one component, or Ω is multiply connected and Ω_1 has one component and for all $k \geq 2$ the number of components of $\Omega_k \setminus \Omega_{k-1}$ equals the number of components of $\partial\Omega$. In the latter case it suffices to consider each component separately and so with respect to this fact we may assume that $\Omega_k \setminus \Omega_{k-1}$ is connected.

We have k fixed; for each $m \in \mathbb{N}$ we call

$$U_m^k = \{Q \in \mathcal{D}_{-m} \ : \ Q \subset \Omega_k \setminus \Omega_{k-1}\}$$

where $m \geq l_k + 8$. Firstly notice that the map f is uniformly continuous on Ω_k and so for ε_k there exists an m' such that $\operatorname{diam} f(Q) < \varepsilon_k$ for all $m \geq m'$ and $Q \in U_m^k$. By this condition we get point $vii)$.

Calling

$$D = \mathcal{H}^1(\partial(\Omega_k \setminus \Omega_{k-1})) + \#\{\text{vertices of } \partial(\Omega_k \setminus \Omega_{k-1})\}$$

we get the existence of an m'' such that

$$2^{3-m} D \leq \delta_k \tag{2.23}$$

for all $m \geq m''$. This will be crucial for getting $ii)$.

Our next step will be to shift the squares to guarantee that $vi)$ holds, then we exclude the squares too close to $\partial(\Omega_k \setminus \Omega_{k-1})$ to give $iv)$ and $v)$. The other properties will follow quickly.

Almost every $a \in \Omega_k \setminus \Omega_{k-1}$ is a point of differentiability of f (for example see [7], Lem. A.28) and a Lebesgue point of the derivative of f in X (it follows from (2.2) that it is also a Lebesgue point of Df in classical sense). We have

$$
\begin{aligned}
I_m(a) := {}& 2^m \| f(x) - f(a) - Df(a)(x-a) \|_{L^\infty(Q(a,2^{1-m}))} \\
& + \fint_{Q(a,2^{1-m})} |Df(x) - Df(a)| \, d\mathcal{L}^2(x) \\
& + \frac{1}{\mathcal{L}^2(Q(a,2^{1-m}))} \| (Df - Df(a)) \chi_{2Q(a,2^{1-m})} \|_{X(\Omega)} \longrightarrow 0 \quad \text{as } m \to \infty
\end{aligned}
\tag{2.24}
$$

for almost every $a \in \Omega_k \setminus \Omega_{k-1}$ (because $\Omega_k \Subset \Omega$ we may assume that $Q(a, 2^{1-m}) \subset \Omega$ for all m large enough). Therefore there exist an m''' and a set E_k such that for all $m \geq m'''$

$$
\left\{ I_m \geq \tfrac{\varepsilon_k}{4} \right\} \subset E_k \quad \text{and} \quad \mathcal{L}^2(E_k) < \tfrac{\delta_k}{32}.
$$

Now we choose $m_k = \max\{l_k + 8, m', m'', m'''\}$ and denote $U_k = U_{m_k}^k = \{\tilde{Q}_1^k, \tilde{Q}_2^k, \ldots, \tilde{Q}_{O_k}^k\}$, where $\tilde{Q}_i^k = Q(c_i, 2^{-m_k})$. Define

$$
\psi(x) = \sum_{i=1}^{O_k} (\chi_{E_k}(x + c_i) + \chi_{A_k}(x + c_i)) \quad \text{for } x \in Q(0, 2^{-m_k-1})
$$

and then

$$
\begin{aligned}
\int_{Q(0,2^{-m_k-1})} \psi \, d\mathcal{L}^2 &= \sum_{i=1}^{O_k} \int_{Q(c_i, 2^{-m_k-1})} (\chi_{E_k} + \chi_{A_k}) \, d\mathcal{L}^2 \\
&\leq \int_\Omega (\chi_{E_k} + \chi_{A_k}) \, d\mathcal{L}^2 \\
&\leq \mathcal{L}^2(E_k) + \mathcal{L}^2(A_k) \\
&\leq \frac{\delta_k}{16}.
\end{aligned}
$$

Therefore

$$
\fint_{Q(0,2^{-m_k-1})} \psi \, d\mathcal{L}^2 \leq \frac{\delta_k 2^{2m_k}}{16} = \delta_k 2^{2m_k - 4}
\tag{2.25}
$$

and therefore we can find a $v_k \in Q(0, 2^{-m_k-1})$ such that

$$
\psi(v_k) \leq \delta_k 2^{2m_k - 4}.
\tag{2.26}
$$

Now, set

$$
a_i = c_i + v_k \quad \text{and} \quad Q_i^k = Q(a_i, 2^{-m_k}) = v_k + \tilde{Q}_i^k.
$$

Let

$$
B_k := \left\{ i : I_{m_k}(a_i) \geq \frac{\varepsilon_k}{4} \right\} \cup \{ i : a_i \in A_k \}
$$

be the set of indices of the bad squares. Now, note that since (2.26) holds, we have $\mathrm{card}(B_k) \le \delta_k 2^{2m_k-4}$ and thus

$$\mathcal{L}^2\left(\bigcup_{i\in B_k} Q_k^i\right) \le \sum_{i\in B_k}\mathcal{L}^2(Q_k^i) \le 2^{2-2m_k}\delta_k 2^{2m_k-4} = \tfrac{1}{4}\delta_k.$$

In summary we have the following, each square $Q(a_i, 2^{-m_k})$ either has $i \in B_k$ and then calculated above is $\mathcal{L}^2(\bigcup_{i\in B_k} Q_i^k) < \delta_k$, otherwise $i \notin B_k$ and in this case $a_i \notin A_k$ and $I_{m_k}(a_i) < \frac{\varepsilon_k}{4}$. From $I_{m_k}(a_i) < \frac{\varepsilon_k}{4}$ we easily get the inequalities in $vi)$.

It holds that

$$\bigcup_{i=1}^{O_k} \tilde{Q}_i^k = \Omega_k \setminus \Omega_{k-1}, \qquad \bigcup_{i=1}^{O_k} Q_i^k = v_k + (\Omega_k \setminus \Omega_{k-1}).$$

As $v_k \in Q(0, 2^{-m_k-1})$, we have $\|v_k\|_\infty \le 2^{-m_k-1}$ and only squares Q_i^k in the most external layer may intersect $\partial(\Omega_k \setminus \Omega_{k-1})$. Now we exclude some of the outer squares to leave a small gap to $\partial(\Omega_k \setminus \Omega_{k-1})$. If necessary we change the indexing to get the set $\{Q_i^k; 1 \le i \le \tilde{O}_k\}$ and none of the squares Q_i^k intersects $\partial(\Omega_k \setminus \Omega_{k-1})$ for $1 \le i \le \tilde{O}_k$. Moreover, (because $\mathrm{diam}_\infty Q_i^k = 2^{1-m_k}$) for arbitrary x in the outer squares of $\{Q_i^k; 1 \le i \le \tilde{O}_k\}$ we have that

$$0 \le \mathrm{dist}_\infty(x, \partial(\Omega_k \setminus \Omega_{k-1})) < 2 \cdot 2^{-m_k}.$$

We exclude also this layer of outer squares. Now for the elements of the outer squares of the remaining system $\{Q_i^k; 1 \le i \le \tilde{O}_k\}$ one has

$$2 \cdot 2^{-m_k} \le \mathrm{dist}_\infty(x, \partial(\Omega_k \setminus \Omega_{k-1})) < 4 \cdot 2^{-m_k}.$$

We repeat this excluding process two more times and we get the final set $\{Q_i^k; 1 \le i \le K_k\}$ with

$$6 \cdot 2^{-m_k} \le \mathrm{dist}_\infty(x, \partial(\Omega_k \setminus \Omega_{k-1})) < 8 \cdot 2^{-m_k}$$

and so satisfies $v)$. There is only a corner of $\partial\bigcup_i^{K_k} Q_i^k$ near a corner of $\partial(\Omega_k \setminus \Omega_{k-1})$ because $m_k \ge l_k + 8$ (which is $iv)$). Simultaneously, because the corners of $\partial\bigcup_i^{K_k} Q_i^k$ have distance greater than $32 \cdot 2^{1-m_k}$, any single square of $\{Q_i^k\}$ contains at most one corner of $\partial\bigcup_i^{K_k} Q_i^k$, therefore no square has more than two external sides and so shares at least two sides with other squares of $\{Q_i^k\}$, which is $iii)$.

It is obvious that $i)$ holds; the union of the squares in U_k has no holes, the shifted squares are the same and no holes inside can be formed by removing squares from the edge. The set $\bigcup_{i=1}^{K_k} Q_i^k$ consists of all of the squares.

The last point to check is $ii)$. By the choice of $m_k \ge m''$ we can estimate

$$\mathcal{L}^2\left(\Omega_k \setminus \left[\Omega_{k-1} \cup \bigcup_{i=1}^{K_k} Q_i^k\right]\right) < \delta_k$$

by (2.23) and so we get $ii)$. $\square$

The next lemma is needed to describe how to slightly move the vertices of a grid to obtain better boundary values. This result is proved in [4], but we give the proof for the reader's convenience.

Lemma 2.11. *Let $f \in W^{1,1}_{loc}(\Omega; \mathbb{R}^2)$ be a homeomorphism and $k, \varepsilon_k, \delta_k$ be given numbers. Let $\left\{ Q_i^k = Q_i^k(a_i, 2^{m_k}) \right\}$ be a grid of squares in $\Omega_k \setminus \Omega_{k-1}$ determined by given numbers ε_k, δ_k in Lemma 2.10. Then there exists $\left\{ \mathcal{Q}_i^k \right\}$ a grid of quadrilaterals verifying $\frac{3}{4} Q_i^k \subset \mathcal{Q}_i^k \subset \frac{5}{4} Q_i^k$ and such that f is absolutely continuous on each side of $\partial \mathcal{Q}_i^k$ and for all i it holds that*

$$\fint_{\partial \mathcal{Q}_i^k} |D_\tau f| \, d\mathcal{H}^1 \leq C \fint_{2Q_i^k} |Df| \, d\mathcal{L}^2. \tag{2.27}$$

Further for those $i \notin B_k$ it holds that

$$\fint_{\partial \mathcal{Q}_i^k} |D_\tau f - Df(a_i)\tau| \, d\mathcal{H}^1 \leq C\varepsilon_k. \tag{2.28}$$

In both cases above, $C > 0$ is an absolute constant.

Proof. For each vertex $V = (v_1, v_2)$ of the grid, we let S_V be the segment

$$S_V = \left\{ (x, y) : x \in \left[v_1 - \frac{r_k}{8}, v_1 + \frac{r_k}{8} \right], y - v_2 = x - v_1 \right\}.$$

For each set of two neighbouring vertices V_1 and V_2 of the grid (*i.e.*, V_1 and V_2 are endpoints of the same side of a square Q of the grid), the following estimates hold

$$\int_{S_{V_1} \times S_{V_2}} \left(\int_{[X_1, X_2]} |D_\tau f(t)| \, d\mathcal{H}^1(t) \right) d(\mathcal{H}^1 \times \mathcal{H}^1)(X_1, X_2) \leq C r_k \int_{2Q} |Df| \, d\mathcal{L}^2, \tag{2.29}$$

$$\int_{S_{V_1} \times S_{V_2}} \left(\int_{[X_1, X_2]} |D_\tau f(t) - Df(a)\tau| \, d\mathcal{H}^1(t) \right) d(\mathcal{H}^1 \times \mathcal{H}^1)(X_1, X_2) \\ \leq C r_k \int_{2Q} |Df - Df(a)| \, d\mathcal{L}^2. \tag{2.30}$$

Above, for $X_i \in S_{V_i}$, $i = 1, 2$, $[X_1, X_2]$ denotes the segment whose endpoints are X_1 and X_2. The integrals in left-hand side are meaningful, as a consequence of the well-known property of the Sobolev mapping f of being a.c. on almost all lines. Moreover, inequalities (2.29) and (2.30) hold because $\mathcal{H}^1(S_{V_1}) \approx r_k$ and $\mathrm{co}(S_{V_1} \cup S_{V_2}) \subset 2Q$. Furthermore, Q is any square of the grid such that $V_1, V_2 \in \partial Q$.

For simplicity, we only estimate (2.27). It will be clear that we shall be able to guarantee also (2.28) for the 'good' quadrilaterals.

Fix $\lambda > 4$. By Chebyshev's inequality, from (2.29) we deduce that there exist a subset $S(V_1, V_2)$ of S_{V_1} and a subset $S(V_2, V_1)$ of S_{V_2} such that

$$\mathcal{H}^1\left(S(V_1, V_2) \right) \geq \left(1 - \frac{1}{\lambda} \right) \mathcal{H}^1(S_{V_1}),$$

$$\mathcal{H}^1\left(S(V_2, V_1) \right) \geq \left(1 - \frac{1}{\lambda} \right) \mathcal{H}^1(S_{V_2})$$

and for every $X_1 \in S(V_1, V_2)$ and $X_2 \in S(V_2, V_1)$ the following estimate holds

$$\int_{[X_1, X_2]} |D_\tau f(t)| \, d\mathcal{H}^1(t) \leq \frac{\lambda^2}{r_k} C \int_{2Q} |Df| \, d\mathcal{L}^2.$$

Now, let V be a vertex of the grid. There are (at most) four vertices V_1, V_2, V_3, V_4 of the grid which are neighbouring vertices of V and therefore, by the above construction we find four subsets $S(V, V_i)$, $i = 1, 2, 3, 4$ of S_V such that

$$\mathcal{S} := \bigcap_{i=1}^{4} S(V, V_i)$$

has positive $\mathcal{H}^1$-measure, since $\lambda > 4$ and thus it is not empty. We replace V with a $\tilde{V} \in \mathcal{S}$. To conclude, each square $Q = \mathrm{co}\,\{V_1, V_2, V_3, V_4\}$ of the grid will be replaced by the quadrilateral $\mathcal{Q} = \mathrm{co}\,\{\tilde{V}_1, \tilde{V}_2, \tilde{V}_3, \tilde{V}_4\}$.
□

For the following lemma recall that we assume (2.1).

Lemma 2.12. *Let $\Omega \subset \mathbb{R}^2$ be a finitely connected bounded domain and let $f \in W^1 X(\Omega, \mathbb{R}^2)$ be a homeomorphism. Let $k \in \mathbb{N}$ and $\varepsilon_k, \delta_k > 0$ be given numbers. Let Ω_k be the set chosen in Lemma 2.10 containing squares of type $Q_i^k = Q(a_i, 2^{-m_k})$ determined by the given numbers ε_k, δ_k. Let $\mathcal{Q}_i^k$ be the quadrilaterals in $\Omega_k \setminus \Omega_{k-1}$ derived from Q_i^k by Lemma 2.11. Then there exists a piece-wise linear injective function h defined on $\Gamma_k = \bigcup_i \partial \mathcal{Q}_i^k$ such that:*

i) it holds that

$$\|f - h\|_{L^\infty(\Gamma_k)} < 4\varepsilon_k,$$

ii) at each vertex x of each $\mathcal{Q}_i^k$ it holds that $h(x) = f(x)$,
iii) for any S a side of a $\mathcal{Q}_i^k$ and for $\mathcal{H}^1$ a.e. $x \in S$ we have

$$|D_\tau h(x)| \leq \fint_S |D_\tau f| \, d\mathcal{H}^1, \tag{2.31}$$

iv) for any $i \notin B_k$ it holds that

$$\fint_{\partial \mathcal{Q}_i^k} |Df(a_i)\tau - D_\tau h| \, d\mathcal{H}^1 \leq C \varepsilon_k.$$

v) for any $i \notin B_k$ with $\sqrt{\varepsilon_k} < |Df(a_i)| < \frac{1}{\sqrt{\varepsilon_k}}$ and $J_f(a_i) > 4\sqrt{\varepsilon_k}$ it holds that h is linear on each S side of $\mathcal{Q}_i^k$.

Proof. The first step is to define a piece-wise linear approximation of f, (call it $\tilde{f}_n$) on each side S of every $\mathcal{Q}_i^k$. We call $x_1, \ldots, x_n$ n evenly spaced points along S with x_1 and x_n being the two endpoints of S. Then we put $\tilde{f}_n(x_j) = f(x_j)$ and $\tilde{f}_n$ is linear on each segment between x_j and x_{j+1}. Of course $\tilde{f}_n$ converges uniformly to f on S and for n large enough is injective. The argument is rather simple and a detailed version can be found in [8], [4] or an alternative approach can be found in [6]. We define h so that $h(S) = \tilde{f}(S)$ for each S, but h parametrizes its image from S at constant speed.

Point *i)* holds because of Lemma 2.10 point *vii)* and because the oscillation of h on any $\partial \mathcal{Q}_i^k$ is bounded by the oscillation of f on the $\partial \mathcal{Q}_i^k$ which is bounded by the oscillation of f on the $2Q_i^k$.

Point ii) is obvious. Point iii) holds because the piece-wise linear curve is not longer than the original curve $i.e.$

$$\fint_S |D_\tau h| \, \mathrm{d}\mathcal{H}^1 \leq \fint_S |D_\tau f| \, \mathrm{d}\mathcal{H}^1,$$

and because the derivative has constant size, $i.e.$

$$|D_\tau h(x)| = \fint_S |D_\tau h| \, \mathrm{d}\mathcal{H}^1$$

almost everywhere.

Let x_m, x_{m+1} be a pair of adjacent points on S such that the derivative of h is constant on the segment $L = [x_m x_{m+1}]$ and $h(x_m) = f(x_m)$ and $h(x_{m+1}) = f(x_{m+1})$. Then we can calculate for all $t \in L$ that

$$D_\tau h(t) = \fint_L D_\tau f \, \mathrm{d}\mathcal{H}^1.$$

When we denote $L(t)$ as the segment $L \subset \partial \mathcal{Q}_i^k$ described above that contains the point $t \in \partial \mathcal{Q}_i^k$ we can easily calculate

$$\fint_{\partial \mathcal{Q}_i^k} \big| \fint_{L(t)} D_\tau f \, \mathrm{d}\mathcal{H}^1 - Df(a_i)\tau \big| \, \mathrm{d}\mathcal{H}^1(t) \leq \fint_{\partial \mathcal{Q}_i^k} \big| \fint_{L(t)} D_\tau f - Df(a_i)\tau \, \mathrm{d}\mathcal{H}^1 \big| \, \mathrm{d}\mathcal{H}^1(t)$$

$$\leq \fint_{\partial \mathcal{Q}_i^k} \fint_{L(t)} |D_\tau f - Df(a_i)\tau| \, \mathrm{d}\mathcal{H}^1 \, \mathrm{d}\mathcal{H}^1(t)$$

$$\leq \frac{1}{\mathcal{H}^1(\partial \mathcal{Q}_i^k)} \sum_{L \subset \partial \mathcal{Q}_i^k} \mathcal{H}^1(L) \fint_L |D_\tau f - Df(a_i)\tau| \, \mathrm{d}\mathcal{H}^1$$

$$\leq \fint_{\partial \mathcal{Q}_i^k} |D_\tau f - Df(a_i)\tau| \, \mathrm{d}\mathcal{H}^1.$$

Now for any $i \notin B_k$ we can see by (2.28) of Lemma 2.11 that the above can further be estimated by $C\varepsilon_k$, which is exactly point iv).

We now prove point v). At any point x where $J_f(x) > 0$ it holds that $J_f(x) = \lambda_1 \lambda_2$, where $\lambda_1 = \max\{|Df(x)v|; |v| = 1\}$ and $\lambda_2 = \min\{|Df(x)v|; |v| = 1\}$. Given that $\lambda_1 < \frac{1}{\sqrt{\varepsilon_k}}$ and $J_f(x) > 4\sqrt{\varepsilon_k}$ it follows that $\lambda_2 > 4\varepsilon_k$ (in the following we will use this fact repeatedly to estimate distance between images). In Lemma 2.11 we made $\mathcal{Q}_i^k$ from Q_i^k by moving its vertexes along a southwest-northeast diagonal (we will call it now the SW-NE diagonal). Therefore the distance of the northwest (we will call it now the NW) and southeast (SE) vertexes to the SW-NE diagonal is not changed in the process. This distance is $\sqrt{2} \cdot 2^{-m_k}$. Therefore the distance of the images in $Df(a_i)$ of the NW and SE vertexes of $\mathcal{Q}_i^k$ to the image of the SW-NE diagonal is at least $4\sqrt{2} \cdot 2^{-m_k}\varepsilon_k$. This means that the balls of radius $2^{-m_k}\varepsilon_k$ centered at the image of the NW and SE vertexes of $\mathcal{Q}_i^k$ both lie entirely on different sides of the line containing the image of the SW-NE diagonal. In fact they have

$$\text{distance from the image of the diagonal of more than } 2^{-m_k}\varepsilon_k. \tag{2.32}$$

The SW and NE vertexes of $\mathcal{Q}_i^k$ lie on a SW-NE line through a_i and so the images of a_i and the SW and NE vertices lie on the line containing the image of the SW-NE diagonal. It holds that $\mathcal{Q}_i^k \supset \frac{3}{4}Q_i^k$ and so the distance of the SW (NE) vertex from a_i is at least $\frac{3\sqrt{2}}{4}2^{-m_k}$ (and the images of the vertexes lie on opposite direction

from sides of the image of a_i). Thus the distance of the image of the SW (NE) vertex in $Df(a_i)$ to the image of a_i in $Df(a_i)$ is at least $3\sqrt{2} \cdot 2^{-m_k}\varepsilon_k$. This means that the ball of radius $2^{-m_k}\varepsilon_k$ centered at the image of the SW (or NE) vertex lies entirely on different sides of the line containing the point $Df(a_i)a_i$ and perpendicular to the image of the SW-NE diagonal. Therefore, and thanks to (2.32), any pair of balls of radius $2^{-m_k}\varepsilon_k$ centered at the image of a pair of vertexes do not intersect. Now any triangle having a vertex in each of the 3 balls of radius $2^{-m_k}\varepsilon_k$ centered at the image in $Df(a_i)$ of a the SW, NW and NE vertex (or similarly for SW, NE, SW vertex) of $\mathcal{Q}_i^k$ is a positively oriented Jordan curve (there is a simple homotopy with $Df(a_i)$ on SW NW NE). Because the image of the vertexes of $\mathcal{Q}_i^k$ in f each lie in a ball of radius $2^{-m_k}\varepsilon_k$ around the image in $Df(a_i)$ of the same vertex the 2-piece-wise affine map (divide $\mathcal{Q}_i^k$ by the SW-NE diagonal) coinciding with f on the vertices of $\mathcal{Q}_i^k$ has positive orientation and is injective on $\mathcal{Q}_i^k$.   □

Lemma 2.13. *Let $\Omega \subset \mathbb{R}^2$ be a finitely connected bounded domain and let $f \in W^1X(\Omega, \mathbb{R}^2)$ be a homeomorphism. Let $k \in \mathbb{N}_0$ and $\varepsilon_k, \delta_k > 0$. Let $\Omega_k \subset \Omega_{k+1}$ be a pair of sets chosen in Lemma 2.10 which we apply with the given $\varepsilon_k, \varepsilon_{k+1}$ and δ_k, δ_{k+1} giving squares $\{Q_i^k\}_i$, resp $\{Q_i^{k+1}\}_i$. Let $\mathcal{Q}_i^k$ be the quadrilaterals from Lemma 2.11 in $\Omega_k \setminus \Omega_{k-1}$ and $\mathcal{Q}_i^{k+1}$ be the quadrilaterals from Lemma 2.11 in $\Omega_{k+1} \setminus \Omega_k$. Call S_k the space between $\mathcal{Q}_i^k$ and $\mathcal{Q}_i^{k+1}$, i.e. the union of all components of $\Omega \setminus [\bigcup_i \mathcal{Q}_i^k \cup \bigcup_i \mathcal{Q}_i^{k+1}]$ which intersects $\partial\Omega_k$. Finally let h be the piece-wise linear function determined by Lemma 2.12. Then there exists a finite collection of quadrilaterals $\{\mathcal{Q}_i^k\}_{i=K_k+1}^{N_k}$ such that $\bigcup_{i=K_k+1}^{N_k} \mathcal{Q}_i^k = S_k$ and a finitely piece-wise affine homeomorphism g defined on S_k such that $g = h$ on ∂S_k and*

$$\|Dg\chi_{S_k}\|_{X(\Omega)} \le C\|Df\chi_{S_k}\|_{X(\Omega)}. \tag{2.33}$$

Proof. We start by assuming that Ω is simply connected and $\Omega_{k+1} \setminus \Omega_k$ has exactly one component. If this were not so, then we could deal with each component seperately in the same way.

Step 1. Go from squares of size 2^{-m_k} to squares of size $2^{-m_{k+1}}$. Call $K_k^1 = K_k$. We have the quadrilaterals $\mathcal{Q}_i^k$, $1 \le i \le K_k^1$, from Lemma 2.11 which were constructed from the squares $Q_i^k \in \mathcal{D}_{-m_k}^{v_k}$ from Lemma 2.10. Now we name the neighbours of Q_i^k. We index the finite set of neighbouring squares

$$\left\{ Q \in \mathcal{D}_{-m_k}^{v_k}; Q^\circ \cap \left(\bigcup_{i=1}^{K_k^1} Q_i^k \right) = \emptyset, \overline{Q} \cap \left(\overline{\bigcup_{i=1}^{K_k^1} Q_i^k} \right) \ne \emptyset \right\} = \{Q_i^k\}_{i=K_k+1}^{\tilde{K}_k^1}.$$

That is to say the squares Q_i^k, $K_k^1 + 1 \le i \le \tilde{K}_k^1$ are those shifted dyadic squares that are not contained in the set $\bigcup_{i=1}^{K_k^1} Q_i^k$ but share at least one common vertex with them (see Fig. 2). By $v)$ of Lemma 2.10 and by $\operatorname{diam}_\infty Q(x, 2^{-m_k}) = 2^{1-m_k}$ we have that

$$4 \cdot 2^{-m_k} \le \operatorname{dist}_\infty \left(x, \partial \left[\bigcup_{i=1}^{\tilde{K}_k^1} Q_i^k \right] \right)$$

for all $x \in \partial\Omega_k$. Similarly call the neighbours of the previously added squares

$$\left\{ Q \in \mathcal{D}_{-m_k}^{v_k}; Q^\circ \cap \left(\bigcup_{i=1}^{\tilde{K}_k^1} Q_i^k \right) = \emptyset, \overline{Q} \cap \left(\overline{\bigcup_{i=1}^{\tilde{K}_k^1} Q_i^k} \right) \ne \emptyset \right\} = \{\tilde{Q}_i^k\}_{i=\tilde{K}_k^1+1}^{\tilde{\tilde{K}}_k^1}.$$
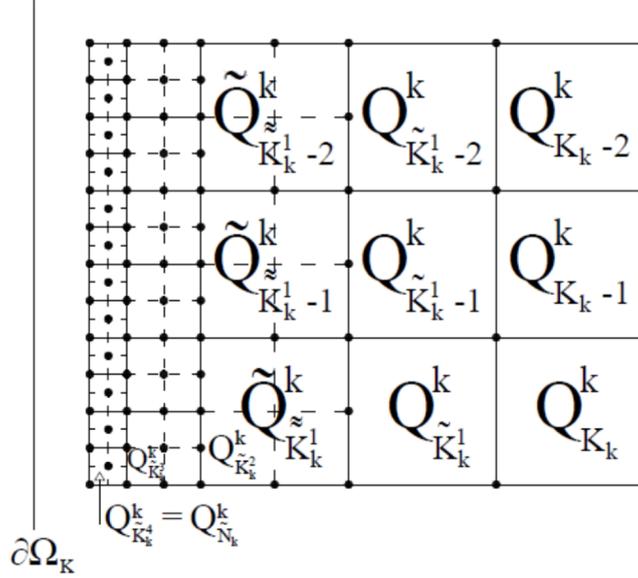
FIGURE 2. The gradual halving of squares untill we have squares with the same sidelength as in the neighbouring set.

Then

$$2 \cdot 2^{-m_k} \le \operatorname{dist}_\infty \left(x, \partial \Big[\bigcup_{i=1}^{\bar{K}_k^1} Q_i^k \cup \bigcup_{i=\tilde{K}_k^1+1}^{\tilde{\bar{K}}_k^1} \tilde{Q}_i^k \Big] \right).$$

for all $x \in \partial \Omega_k$.

We divide each square $\tilde{Q}_i^k = Q(\tilde{a}_i, 2^{-m_k})$ for $\tilde{K}_k^1 + 1 \le i \le \tilde{\bar{K}}_k^1$ into its four quarters $Q(\tilde{a}_i + 2^{-m_k-1}z, 2^{-m_k-1})$, for each z a corner of $Q(0,1)$. We number the squares we get by dividing $\tilde{Q}_i^k$, $\tilde{K}_k^1 + 1 \le i \le \tilde{\bar{K}}_k^1$ into quarters as $Q_i^k = Q(a_i, 2^{-m_k-1})$ for $\tilde{K}_k^1 + 1 \le i \le \tilde{K}_k^2$. Then we have

$$4 \cdot 2^{-m_k-1} \le \operatorname{dist}_\infty \left(x, \partial \Big[\bigcup_{i=1}^{\tilde{K}_k^2} Q_i^k \Big] \right)$$

for all $x \in \partial \Omega_k$.

We now repeat this last operation. We call the neighbours of the previously added squares

$$\Big\{ Q \in \mathcal{D}_{-m_k-1}^{v_k}; Q^\circ \cap \Big(\bigcup_{i=1}^{\tilde{K}_k^2} Q_i^k \Big) = \emptyset, \overline{Q} \cap \overline{\Big(\bigcup_{i=1}^{\tilde{K}_k^2} Q_i^k \Big)} \neq \emptyset \Big\} = \{\tilde{Q}_i^k\}_{i=\tilde{K}_k^2+1}^{\tilde{\bar{K}}_k^2}.$$

We divide the squares $\{\tilde{Q}_i^k = Q(\tilde{a}_i, 2^{-m_k-1})\}_{i=\tilde{K}_k^2+1}^{\tilde{\bar{K}}_k^2}$ into its four quarters $Q(\tilde{a}_i + 2^{-m_k-2}z, 2^{-m_k-2}) \in \mathcal{D}_{-m_k-2}^{v_k}$, for each z a corner of $Q(0,1)$ and call these squares $Q_i^k = Q(a_i, 2^{-m_k-2})$, $\tilde{K}_k^2 + 1 \le i \le \tilde{K}_k^3$. As before we have

that

$$4 \cdot 2^{-m_k-2} \leq \text{dist}_\infty \left(x, \partial \bigcup_{i=1}^{\tilde{K}_k^3} Q_i^k \right)$$

for all $x \in \partial\Omega_k$.

After $m_{k+1} - m_k + 1$ steps we start adding squares of type $Q(a_i, 2^{-m_{k+1}})$. We call those Q_i^k, $\tilde{K}_k^{m_{k+1}-m_k+1} \leq i \leq \tilde{N}_k$ the dyadic squares of type $Q(a_i, 2^{-m_{k+1}}) \subset \Omega_k$ which are not in $\bigcup_{i=1}^{\tilde{K}_k^{m_{k+1}-m_k+1}} Q_i^k$. Therefore

$$\text{the distance from } \partial \bigcup_{i=1}^{\tilde{N}_k} Q_i^k \text{ to } \partial\Omega_k \text{ is between 2 and 4 times } 2^{-m_{k+1}}. \tag{2.34}$$

By adding shifted dyadic squares of type $Q \in \mathcal{D}_{-m_{k+1}}^{v_{k+1}}$ to the grid in $\Omega_{k+1} \setminus \Omega_k$ we can guarantee that the distance from $\partial \bigcup_i Q_i^{k+1}$ to $\partial\Omega_k$ is also between 2 and 4 times $2^{-m_{k+1}}$. Thus the distance between $\partial \bigcup_{i=1}^{\tilde{N}_k} Q_i^k$ and $\partial \bigcup_{i=1}^{\tilde{N}_{k+1}} Q_i^{k+1}$ is between 4 and 8 times $2^{-m_{k+1}}$. Call $\tilde{G}_k$ the component of $\mathbb{R}^2 \setminus (\bigcup_{i=1}^{\tilde{N}_k} Q_i^k \cup \bigcup_{i=1}^{\tilde{N}_{k+1}} Q_i^{k+1})$ containing $\partial\Omega_k$.

In our construction we always have two layers of squares of side length 2^{1-m_k-j} before we start adding squares of side length $2^{1-m_k-(j+1)}$. Therefore

$$2Q_i^k \text{ intersects at most its neighbours and its neighbours' neighbours but no}$$
$$\text{other squares further away.} \tag{2.35}$$

Further the square $2Q_i^k$ is contained in the neighbours of Q_i^k. Therefore any square $Q_{i'}^k$ such that $2Q_{i'}^k$ intersects $2Q_i^k$ then it is at most a third neighbour (*i.e.* a neighbour of Q_i^k a neighbour's neighbour or a neighbour of a neighbour of a neighbour of Q_i^k). But the number of third neighbours is bounded. Therefore there exists an L such that

$$\sum_{i=1}^{\tilde{N}_k} \chi_{2Q_i^k}(x) \leq L \tag{2.36}$$

for almost every $x \in S_k$.

Step 2. Fill the rest of S_k with quadrilaterals. What is left over is a 'tube' around $\partial\Omega_k$ approximately $2^{-m_{k+1}}$ wide, which can be divided into quadrilaterals all uniformly bi-Lipschitz equivalent with a square of side length $2^{1-m_{k+1}}$. Although this fact is obvious we describe one way how to do this in detail here.

We call those vertices X of squares Q_i^k such that $X \in \partial \bigcup_{i=1}^{\tilde{N}_k} Q_i^k$ outer vertices. Similarly we also call those vertices X of squares Q_i^{k+1} such that $X \in \partial \bigcup_{i=1}^{\tilde{N}_{k+1}} Q_i^{k+1}$ outer vertices. From $iv)$ of Lemma 2.10 we have that $\partial \bigcup_{i=1}^{\tilde{N}_k} Q_i^k$ and $\partial \bigcup_{i=1}^{\tilde{N}_{k+1}} Q_i^{k+1}$ are piece-wise linear parallel to coordinate axes and each of their sides corresponds to a side of $\partial\Omega_k$. We call a vertex of $\partial \bigcup_{i=1}^{\tilde{N}_k} Q_i^k$ (or $\partial \bigcup_{i=1}^{\tilde{N}_{k+1}} Q_i^{k+1}$) a reflex corner of $\partial \bigcup_{i=1}^{\tilde{N}_k} Q_i^k$ (or $\partial \bigcup_{i=1}^{\tilde{N}_{k+1}} Q_i^{k+1}$) if $270°$ of a small circular arc centered at the vertex lies inside $\bigcup_{i=1}^{\tilde{N}_k} Q_i^k$ (or $\bigcup_{i=1}^{\tilde{N}_{k+1}} Q_i^{k+1}$) for example A_{-1} in Figure 3. In the other case we call a vertex of $\partial \bigcup_{i=1}^{\tilde{N}_k} Q_i^k$ (or $\partial \bigcup_{i=1}^{\tilde{N}_{k+1}} Q_i^{k+1}$) an acute corner of $\partial \bigcup_{i=1}^{\tilde{N}_k} Q_i^k$ (or $\partial \bigcup_{i=1}^{\tilde{N}_{k+1}} Q_i^{k+1}$) if $90°$ of a small circular arc centered at the vertex lies inside $\bigcup_{i=1}^{\tilde{N}_k} Q_i^k$ (or $\bigcup_{i=1}^{\tilde{N}_{k+1}} Q_i^{k+1}$) for example C in Figure 3. We describe an outer vertex as a line vertex if it is not a corner of $\partial \bigcup_{i=1}^{\tilde{N}_k} Q_i^k$ neither is it the vertex neighbouring a relfex corner of $\partial \bigcup_{i=1}^{\tilde{N}_k} Q_i^k$ (similarly for $\bigcup_{i=1}^{\tilde{N}_{k+1}} Q_i^{k+1}$) for example X_i or Y_i in Figure 3.
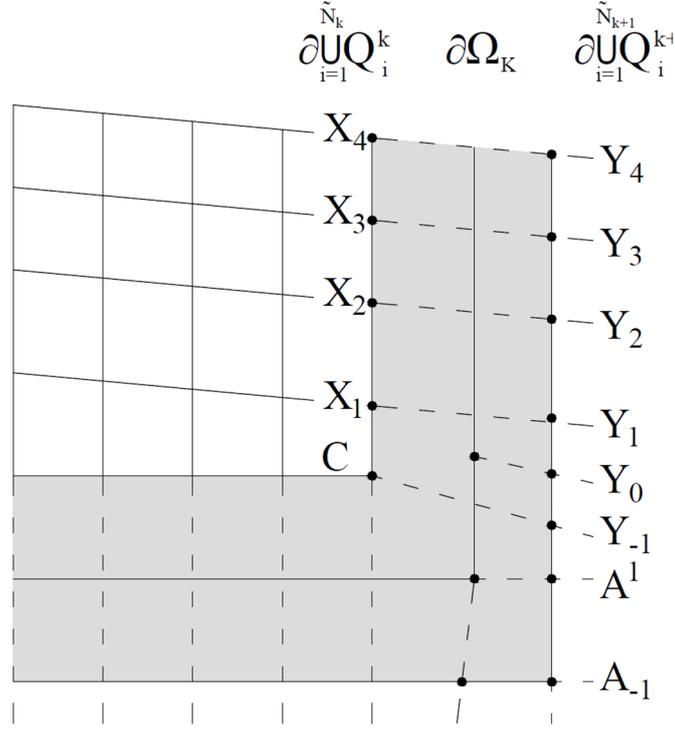
FIGURE 3. Filling in the left over space (in the picture shaded) between two neighbouring grids with quadrilaterals uniformly piece-wise affine bi-Lipschitz equivalent with $Q(0, 2^{-m_{k+1}})$. The line vertices X_i are partnered with the line vertices Y_i. The vertex A_{-1} is a reflex vertex and C is an acute vertex. We the two vertices adjacent to A_{-1} (one of them is denoted as A^1 the other is not denoted) to the near vertex of $\partial\Omega_k$ with a segment. The vertex C is connected with the next available neighbour of A (in this case Y_{-1}). The vertex Y_0 is connected with $\partial\Omega_k$.

Firstly we deal with the area around corners of $\partial\Omega_k$. We add no new segments going from a reflex corner A. We add a segment from the vertices A', A'' neighbouring A to the nearby corner of $\partial\Omega_k$, this creates a quadrilateral (see Fig. 3). For an acute corner C with adjacent sides S_1 and S_2 (whose corresponding opposite sides are $\tilde{S}_1$ and $\tilde{S}_2$) we create new quadrilaterals by adding a segment from C to both of the second neighbours of A.

Now choose any side of $\partial\Omega_k$ and consider the two corresponding sides $S, \tilde{S}$, with S a side of $\partial\bigcup_{i=1}^{\tilde{N}_k} Q_i^k$ and $\tilde{S}$ a side of $\partial\bigcup_{i=1}^{\tilde{N}_{k+1}} Q_i^{k+1}$. Both S and $\tilde{S}$ have endpoints very close to the same corners of $\partial\Omega_k$, see (2.34). This means that in the extreme scenario the distance between the length of S and $\tilde{S}$ is at less than $8 \cdot 2^{-m_{k+1}}$. Because the side length of the squares is $2 \cdot 2^{-m_{k+1}}$ the difference in number of line vertices on S and $\tilde{S}$ varies by at most 4, at most 2 at each corner. Thus it is possible to pair the line vertices of S and the line vertices of $\tilde{S}$ so that at most two line vertices of S (resp. $\tilde{S}$) close to a given corner of S, (resp. $\tilde{S}$) do not have a partner in $\tilde{S}$, (resp S) moreover the segment between a pair of partnered vertexes is nearly perpendicular to the corresponding segment of $\partial\Omega_k$. Let X_1, X_2 be a pair of line vertices on S with partners Y_1 and Y_2, their partners in $\tilde{S}$. Each of the segments X_1Y_1 and X_2Y_2 intersect $\partial\Omega_k$ exactly once, call the points Z_1 and Z_2 respectively. Any of the quadrilaterals $X_1Z_1Z_2X_2$ and $Y_1Z_1Z_2Y_2$ formed by any two pairs of neighbouring line vertexes are uniformly bi-Lipschitz equivalent with $Q(0, 2^{-m_{k+1}})$ by a 2-piece-wise affine map. Permitting a small bastardisation of the notation we call these quadrilaterals Q_i^k for $\tilde{N}_k + 1 \le i \le \tilde{\tilde{N}}_k$.

If we have a left over line vertex X with no partner we create a new quadrilateral with a vertex at X by adding a segment from X to the point on $\partial\Omega_k$ half way between the neighbouring vertices on $\partial\Omega_k$ (see Y_0 in Fig. 3). We call the entire collection of these quadrilaterals Q_i^k, $1 \leq i \leq N_k$. Up to maybe increase L the equation (2.36) is extended as

$$\sum_{i=1}^{N_k} \chi_{2Q_i^k}(x) \leq L \tag{2.37}$$

Step 3. Move the corners. It is possible by moving the corners of the quadrilaterals of Q_i^k $K_k \leq i \leq N_k$ to get quadrilaterals $\mathcal{Q}_i^k$ which satisfy the estimate (2.27). The idea is exactly that of the proof of Lemma 2.11, the only difference here is that neighbouring squares may not have exactly the same side length, but the ratio is bounded by 2. This case is dealt with in detail in Theorem 4.1, step 2 of [4] and the interested reader can check the details there. We move corners by at most 2^{-2-m_k-j}, for Q_i^k in the j-th generation of added squares.
Step 4. Define g on the grid and extend. We define g on $\partial\mathcal{Q}_i^k$ for $K_k + 1 \leq i \leq N_k$ exactly the same way as we defined h in Lemma 2.12 for $1 \leq i \leq K_k + 1$. We get a piece-wise linear and injective function g on $\bigcup_{i=K_k+1}^{N_k} \partial\mathcal{Q}_i^k$. Because each $\mathcal{Q}_i^k$ is uniformly bi-Lipschitz equivalent with a square by (uniformly bounded number of pieces) piece-wise affine maps we can apply Corollary 2.6 and get a finitely piece-wise affine homeomorphism g on S_k, extending the original mapping and satisfying

$$\|Dg\|_{L^\infty(\mathcal{Q}_i^k)} \leq \frac{C}{r} \int_{\partial\mathcal{Q}_i^k} |D_\tau f| \, d\mathcal{H}^1.$$

Further by applying the (2.27) type estimate we got from step 3, we get

$$\|Dg\|_{L^\infty(\mathcal{Q}_i^k)} < C \fint_{2Q_i^k} |Df| \, d\mathcal{L}^2 \tag{2.38}$$

on each Q_i^k, for $K_k + 1 \leq i \leq N_k$.

Let $K_k + 1 \leq i \leq N_k$, let $y \in \mathcal{Q}_i^k$ and let $r > 0$. For the simplicity of notation call

$$A_{k,y,r} = \left\{ l \in \mathbb{N} : K_k + 1 \leq l \leq N_k; \mathcal{L}^2(\mathcal{Q}_l^k \cap Q(y,r)) > 0 \right\},$$

then it is obvious that

$$\fint_{Q(y,r)} |Dg \chi_{\bigcup_{i=K_k+1}^{N_k} \mathcal{Q}_i^k}| \, d\mathcal{L}^2 \leq \max_{l \in A_{k,y,r}} \|Dg\|_{L^\infty(\mathcal{Q}_l^k)}. \tag{2.39}$$

But in step 3 we moved corners by at most 2^{-2-m_k-j} for squares in the j-th generation. Therefore, for our $K_k + 1 \leq i \leq N_k$ (say $Q_i^k = Q(a_i, 2^{-m_k-j})$ is a j-th generation square) we have that $\mathcal{Q}_i^k \subset Q(a_i, \frac{5}{4}2^{-m_k-j})$ and this implies that for each $y \in \mathcal{Q}_i^k$ and $0 < r < 2^{-2-m_k-j}$ it holds that $Q(y,r) \subset 2Q_i^k$. Therefore, for any $l \in A_{k,y,r}$ we have that $\mathcal{Q}_l^k \cap 2Q_i^k \neq \emptyset$. On the other hand any 'square' intersecting $2Q_i^k$ must be contained in $6Q_i^k$, because neighbours (and second neighbours) have side lengths bounded by a factor of 2. But then for $l \in A_{k,y,r}$

$$2Q_l^k \subset 12Q_i^k. \tag{2.40}$$

Further, thanks to (2.35) we have that

$$\mathcal{L}^2(Q_l^k) \geq C\mathcal{L}^2(Q_i^k), \tag{2.41}$$

with C independent of i, j, k and l. We continue (2.39) using (2.38) and then (2.41) together with (2.40) and get

$$
\begin{aligned}
\fint_{Q(y,r)} |Dg\chi_{\bigcup_{l=K_k+1}^{N_k} \mathcal{Q}_l^k}| \, \mathrm{d}\mathcal{L}^2 &\leq \max_{l\in A_{k,y,r}} \|Dg\|_{L^\infty(\mathcal{Q}_l^k)} \\
&\leq C \max_{l\in A_{k,y,r}} \fint_{2Q_l^k} |Df| \, \mathrm{d}\mathcal{L}^2 \\
&\leq C \fint_{12Q_i^k} |Df| \, \mathrm{d}\mathcal{L}^2 \leq CM(|Df|)(y)
\end{aligned}
\tag{2.42}
$$

where M denotes the maximal operator.

Now we consider a $y \in \mathcal{Q}_i^k$ and $r > 2^{-2-m_k-j}$, then, using (2.10) and (2.27) we have

$$
\begin{aligned}
\int_{Q(y,r)} |Dg\chi_{\bigcup_{l=K_k+1}^{N_k} \mathcal{Q}_l^k}| \, \mathrm{d}\mathcal{L}^2 &= \sum_{l=K_k+1}^{N_k} \int_{Q(y,r)\cap \mathcal{Q}_l^k} |Dg| \, \mathrm{d}\mathcal{L}^2 \\
&\leq C \sum_{l\in A_{k,y,r}} \int_{2Q_l^k} |Df| \, \mathrm{d}\mathcal{L}^2.
\end{aligned}
$$

We use (2.40), the fact that $r > 2^{-2-m_k-j}$ to get that $Q(y, 52r) \supset \bigcup_{l\in A_{k,y,r}} 2Q_l^k$. Then in the previous estimate, using (2.37), we get

$$
\sum_{l\in A_{k,y,r}} \int_{2Q_l^k} |Df| \leq CL \int_{Q(y,52r)} |Df\chi_{\bigcup_{i=K_k+1}^{N_k} \mathcal{Q}_i^k}|
$$

for $r > 2^{-2-m_k-j}$. Therefore

$$
\begin{aligned}
\fint_{Q(y,r)} |Dg\chi_{\bigcup_{i=K_k+1}^{N_k} \mathcal{Q}_i^k}| \, \mathrm{d}\mathcal{L}^2 &\leq C \fint_{Q(y,52r)} |Df\chi_{\bigcup_{i=K_k+1}^{N_k} \mathcal{Q}_i^k}| \, \mathrm{d}\mathcal{L}^2 \\
&\leq CM(|Df|\chi_{\bigcup_{i=K_k+1}^{N_k} \mathcal{Q}_i^k})(y).
\end{aligned}
\tag{2.43}
$$

By taking supremum over all $0 < r < \infty$ in (2.42) and (2.43) on the left-hand side we obtain that

$$
M(|Dg|\chi_{\bigcup_{i=K_k+1}^{N_k} \mathcal{Q}_i^k})(x) \leq CM(|Df|\chi_{\bigcup_{i=K_k+1}^{N_k} \mathcal{Q}_i^k})(x).
\tag{2.44}
$$

By the norm comparison (2.7) we obtain that

$$
\left\| Dg\chi_{\bigcup_{i=K_k+1}^{N_k} \mathcal{Q}_i^k} \right\|_X \leq C \left\| Df\chi_{\bigcup_{i=K_k+1}^{N_k} \mathcal{Q}_i^k} \right\|_X
$$

which is (2.33). $\qquad\square$

3. Proof of Theorem 1.2

We now prove Theorem 1.2.

Step 1. A grid of 'squares'.

Let $\varepsilon^* > 0$ be any fixed positive number, we want to find a piece-wise affine homeomorphic approximation $\hat{f}$ such that

$$\|Df - \mathrm{d}\hat{f}\|_{X(\Omega)} < C\varepsilon^*$$

and

$$\|f - \hat{f}\|_{L^\infty(\Omega)} < \varepsilon^*.$$

The first stage of this is to separate Ω into nested sets Ω_k. This separation is the subject of Lemma 2.10. We want to apply Lemma 2.10 and get a grid of squares Q_i^k but in order to do this we need to choose $\varepsilon_k, \delta_k > 0$ and a set A_k. Using the 'ε-δ' continuity of the norm of X (recall that all spaces supporting the Lebesgue Point Property have this property) we choose δ_k be a number so small that

$$\|Df\chi_E\|_{X(\Omega)} < 2^{-k}\varepsilon^* \text{ for any } E \subset \Omega_k \text{ such that } \mathcal{L}^2(E) < 5\delta_k. \tag{3.1}$$

We find a number $T_k \geq 1$ such that

$$\begin{aligned}
\mathcal{L}^2\Big(\{x : |Df(x)| > T_k\}\Big) &< \frac{\delta_k}{96}, \\
\mathcal{L}^2\Big(\{x : 0 < |Df(x)| < T_k^{-1}\}\Big) &< \frac{\delta_k}{96} \text{ and} \\
\mathcal{L}^2\Big(\{x : 0 < |J_f(x)| < 4T_k^{-1}\}\Big) &< \frac{\delta_k}{96}.
\end{aligned} \tag{3.2}$$

Then we call A_k the union of these sets, *i.e.*

$$A_k = \{x : |Df(x)| > T_k\} \cup \{x : 0 < |Df(x)| < T_k^{-1}\} \cup \{x : 0 < |J_f(x)| < 4T_k^{-1}\}$$

and $\mathcal{L}^2(A_k) < \frac{\delta_k}{32}$. Now we apply Lemma 2.2 with $G = \Omega_k \setminus \Omega_{k-1}$, $M = (2 + C_{(2.10)})T_k$, $\tilde{\varepsilon}_k = \frac{2^{-k}\varepsilon^*}{\mathcal{L}^2(\Omega_k \setminus \Omega_{k-1})}$. This gives us a number $\tilde{\delta}_k$. We require

$$\begin{aligned}
\varepsilon_k &< \frac{\tilde{\delta}_k}{1 + C_{(2.10)} + 4T_k} \text{ and} \\
\varepsilon_k &< \frac{2^{-4-k}\varepsilon^*}{C_{(2.27)}C_{(2.10)} + \varphi_X(\mathcal{L}^2(\Omega_k \setminus \Omega_{k-1})) + (1 + C_{(2.10)} + 4T_k^2)\mathcal{L}^2(\Omega_k \setminus \Omega_{k-1})},
\end{aligned} \tag{3.3}$$

where $\varphi_X(\cdot)$ is the fundamental function of X. In each $\Omega_k \setminus \Omega_{k-1}$, Lemma 2.10 gives a grid of squares Q_i^k which cover most (up to a set of measure δ_k) of $\Omega_k \setminus \Omega_{k-1}$. Now we focuss on calculations for a fixed k. Finally at the end of the proof we sum over k.

We apply Lemma 2.11 to slightly alter the squares Q_i^k and call the resulting quadrilaterals $\mathcal{Q}_i^k$. We deal with the set S_k (the set from Lemma 2.13) later. Recall $\mathcal{L}^2(S_k) < \delta_k$.

By Lemma 2.10 point $vi)$ we get that if $i \notin B_k$ then also $a_i \notin A_k$. Then for all $i \notin B_k$

$$(Df(a_i) = 0 \text{ or } T_k^{-1} < |Df(a_i)| < T_k) \quad \text{and} \quad (J_f(a_i) = 0 \text{ or } J_f(a_i) > 4T_k^{-1}).$$

On the other hand by Lemma 2.10 point vi)

$$\mathcal{L}^2\Big(\bigcup_{i\in B_k}\mathcal{Q}_i^k\Big) \leq \sum_{i\in B_k}\mathcal{L}^2(\mathcal{Q}_i^k) \leq \sum_{i\in B_k}\mathcal{L}^2(2Q_i^k) \leq 4\mathcal{L}^2\Big(\bigcup_{i\in B_k}Q_i^k\Big) \leq 4\delta_k \tag{3.4}$$

because $\mathcal{Q}_i^k$ and Q_i^k are pair-wise essentially disjoint.

On all the quadrilaterals $\mathcal{Q}_i^k$ we have the estimates (2.27) and if $i \notin B_k$ also the estimate (2.28). We make the following categorisation of the quadrilaterals $\mathcal{Q}_i^k$. We put

$$\begin{aligned}
\mathcal{G}_k &= \big\{\mathcal{Q}_i^k; 1 \leq i \leq K_k, i \notin B_k, T_k^{-1} < |Df(a_i)| < T_k, J_f(a_i) > 4T_k^{-1}\big\}\\
\mathcal{N}_k &= \big\{\mathcal{Q}_i^k; 1 \leq i \leq K_k, i \notin B_k, T_k^{-1} < |Df(a_i)| < T_k, J_f(a_i) = 0\big\}\\
\mathcal{Z}_k &= \big\{\mathcal{Q}_i^k; 1 \leq i \leq K_k, i \notin B_k, |Df(a_i)| = 0\big\}\\
\mathcal{B}_k &= \big\{\mathcal{Q}_i^k; 1 \leq i \leq K_k, i \in B_k\big\}.
\end{aligned}$$

Then every quadrilateral $\mathcal{Q}_i^k$, $1 \leq i \leq K_k$ belongs exactly to one of $\mathcal{G}_k$, $\mathcal{N}_k$, $\mathcal{Z}_k$ or $\mathcal{B}_k$. As calculated in (3.4)

$$\mathcal{L}^2\Big(\bigcup_{\mathcal{Q}\in\mathcal{B}_k}\mathcal{Q}\Big) < 4\delta_k. \tag{3.5}$$

Strictly speaking to define $\hat{f}$ on $\Omega_k \setminus \Omega_{k-1}$ we need to use Lemma 2.13 on both S_k (to define $\hat{f}$ on $S_k \cap \Omega_k$) and on S_{k-1} (to define $\hat{f}$ on $S_{k-1} \setminus \Omega_{k-1}$). To make the following easier to read we redefine each Ω_k as $\Omega_k \cup S_k$ and then to define $\hat{f}$ on $\Omega_k \setminus \Omega_{k-1}$ we only have to apply Lemma 2.13 on S_k.

By the choice of the grid (*i.e.* all the squares in the grid in $\Omega_k \setminus \Omega_{k-1}$ have the same size) then each square in $\Omega_k \setminus \Omega_{k-1}$ has at most 8 neighbours (where a neighbour is a square that shares at least one vertex). The exception is when the square is at the edge of our grid in $\Omega_k \setminus \Omega_{k-1}$ because these squares may have neighbours of half their side length but the number of neighbours is still bounded by 12. Because, for any $x \in (Q_i^k)^\circ$ and any $Q_j^k \neq Q_i^k$, the only way for $x \in 2Q_j^k$ is if Q_j^k is a neighbour of Q_i^k. Therefore

$$\sum_{i=1}^{K_k}\chi_{2Q_i^k}(x) \leq 13 \tag{3.6}$$

for almost every $x \in \bigcup_i Q_i^k$.

Step 2. Defining a piece-wise linear map on each $\partial\mathcal{Q}_i^k$. We define an injective piece-wise linear function $\hat{f}$ using Lemma 2.12 for each $\mathcal{Q}_i^k$, $1 \leq i \leq K_k$ we put $\hat{f}(x) = f(x)$ at each vertex x of $\mathcal{Q}_i^k$. Especially we note that $\hat{f}$ is linear on each side of each $\mathcal{Q}_i^k \in \mathcal{G}_k$ by point $v)$ of Lemma 2.12.

Step 3. Defining $\hat{f}$, the piece-wise affine approximation of f. In each case $\mathcal{Q} \in \mathcal{G}_k$, $\mathcal{Q} \in \mathcal{N}_k$, $\mathcal{Q} \in \mathcal{Z}_k$ or $\mathcal{Q} \in \mathcal{B}_k$ we define an injective piece-wise affine extension of $\hat{f}$ from $\partial\mathcal{Q}$. The quadrilateral $\mathcal{Q}$ is the union of 2 triangles (divided by the SW-NE diagonal) and, as was proved in Lemma 2.12 point $v)$, for $\mathcal{Q} \in \mathcal{G}_k$ we define $\hat{f}$ as linear on each side of each of these triangles and this definition is injective. In that case $\hat{f}$ extends as an affine map onto each triangle.

If $\mathcal{Q} \in \mathcal{B}_k$ or if $\mathcal{Q} \in \mathcal{Z}_k$ then we apply a 2-piece-wise affine 2-bi-Lipschitz mapping Ψ, which maps $\mathcal{Q}$ onto $Q(0, 2^{-m_k})$ and there we use the extension Theorem 2.4, which gives us g. We define $\hat{f} = g \circ \Psi$ on $\mathcal{Q}$. On $\mathcal{Q} \in \mathcal{N}_k$ we define $\hat{f}$ using Theorem 2.7.

Notice that in each case we have defined $\hat{f}$ (see the use of Lemma 2.12 in step 3) on each side of $\partial\mathcal{Q}$ so that

$$|D_\tau\hat{f}| \leq C|Df(a_i)| \text{ is constant on each side of } \partial\mathcal{Q}. \tag{3.7}$$

The last definition that needs to be made in $\Omega_k \setminus \Omega_{k-1}$ is the application of Lemma 2.13 to get $\hat{f}$ on S_k.

Step 4. Uniform convergence estimates. It suffices to combine the estimates from Lemma 2.10 point *vii*) and point *i*) from Lemma 2.12 to get that $\|\hat{f} - f\|_{L^\infty(\bigcup_i Q_i^k)} < 5\varepsilon_k$. The 'squares' $\mathcal{Q}_i^k$, $i = K_k + 1, \ldots, N_k$ are even smaller than the squares $\mathcal{Q}_i^k$ for $i = 1 \ldots K_k$ and so we may assume that the oscillation there has the same bound.

Step 5. Estimating the distance of $\mathrm{d}\hat{f}$ from Df in X.

In the following we refer to the 'centre' of the quadrilateral $\mathcal{Q}$ as $a_\mathcal{Q}$. That is if $\mathcal{Q} = \mathcal{Q}_i^k$ then $a_\mathcal{Q} = a_i^k$ the centre of Q_i^k. We estimate

$$
\begin{aligned}
\|(Df - \mathrm{d}\hat{f})\chi_{\Omega_k \setminus \Omega_{k-1}}\|_{X(\Omega)} &\leq \|(Df - \mathrm{d}\hat{f})(\sum_{i=1}^{K_k} \chi_{\mathcal{Q}_i^k} + \chi_{S_k})\|_{X(\Omega)} \\
&\leq \|(\sum_{\mathcal{Q} \in \mathcal{B}_k} \chi_\mathcal{Q} + \chi_{S_k})Df\|_{X(\Omega)} + \|\sum_{\mathcal{Q} \in \mathcal{B}_k} \chi_\mathcal{Q} \mathrm{d}\hat{f}\|_{X(\Omega)} \\
&\quad + \|\chi_{S_k} \mathrm{d}\hat{f}\|_{X(\Omega)} \\
&\quad + \|\sum_{\mathcal{Q} \in \mathcal{G}_k} \chi_\mathcal{Q}(Df - Df(a_\mathcal{Q}))\|_{X(\Omega)} + \|\sum_{\mathcal{Q} \in \mathcal{G}_k} \chi_\mathcal{Q}(\mathrm{d}\hat{f} - Df(a_\mathcal{Q}))\|_{X(\Omega)} \\
&\quad + \|\sum_{\mathcal{Q} \in \mathcal{N}_k} \chi_\mathcal{Q}(Df - Df(a_\mathcal{Q}))\|_{X(\Omega)} + \|\sum_{\mathcal{Q} \in \mathcal{N}_k} \chi_\mathcal{Q}(\mathrm{d}\hat{f} - Df(a_\mathcal{Q}))\|_{X(\Omega)} \\
&\quad + \|\sum_{\mathcal{Q} \in \mathcal{Z}_k} \chi_\mathcal{Q} Df\|_{X(\Omega)} + \|\sum_{\mathcal{Q} \in \mathcal{Z}_k} \chi_\mathcal{Q} \mathrm{d}\hat{f}\|_{X(\Omega)}.
\end{aligned}
\tag{3.8}
$$

By (3.1), (3.5) and Lemma 2.10 point *ii*) ($\mathcal{L}^2(S_k) < \delta_k$) we have that

$$
\|(\sum_{\mathcal{Q} \in \mathcal{B}_k} \chi_\mathcal{Q} + \chi_{S_k})Df\|_{X(\Omega)} < 2^{-k}\varepsilon^*.
\tag{3.9}
$$

The sum of the terms

$$
\begin{aligned}
w_k := &\|\sum_{\mathcal{Q} \in \mathcal{G}_k} \chi_\mathcal{Q}(Df - Df(a_\mathcal{Q}))\|_{X(\Omega)} + \|\sum_{\mathcal{Q} \in \mathcal{Z}_k} \chi_\mathcal{Q} Df\|_{X(\Omega)} \\
&+ \|\sum_{\mathcal{Q} \in \mathcal{N}_k} \chi_\mathcal{Q}(Df - Df(a_\mathcal{Q}))\|_{X(\Omega)}
\end{aligned}
\tag{3.10}
$$

is immediately estimated by Lemma 2.10 *vi*), (that is (2.22)), the finite overlap property (3.6) and (3.3) as follows

$$
\begin{aligned}
w_k &\leq \sum_{\mathcal{Q} \in \mathcal{G}_k \cup \mathcal{N}_k \cup \mathcal{Z}_k} \|\chi_\mathcal{Q}(Df - Df(a_\mathcal{Q}))\|_{X(\Omega)} \\
&\leq \sum_{\mathcal{Q} \in \mathcal{G}_k \cup \mathcal{N}_k \cup \mathcal{Z}_k} \varepsilon_k \mathcal{L}^2(2Q_i^k) \\
&\leq 13\varepsilon_k \mathcal{L}^2(\Omega_k \setminus \Omega_{k-1}) \\
&< 2^{-k}\varepsilon^*.
\end{aligned}
\tag{3.11}
$$

Having estimated the 'f'-terms in (3.8), we proceed with the '$\hat{f}$'-terms. Because every square in S_k has side length at most 2^{1-m_k} it holds that

$$\bigcup_{i=K_k+1}^{N_k} 2Q_i^k \subset S_k + Q(0, 2^{-m_k}).$$

By Lemma 2.10 point v) we have that

$$\partial\Omega_k + Q(0, 6 \cdot 2^{-m_k}) \subset S_k \subset \partial\Omega_k + Q(0, 8 \cdot 2^{-m_k}).$$

Therefore the tube around $\partial\Omega_k$ which has double the measure of S_k is a superset of $\partial\Omega_k + Q(0, 12 \cdot 2^{-m_k})$. Therefore the inclusion

$$\bigcup_{i=K_k+1}^{N_k} 2Q_i^k \subset \partial\Omega_k + Q(0, 9 \cdot 2^{-m_k}) \subset \partial\Omega_k + Q(0, 12 \cdot 2^{-m_k})$$

implies that

$$\mathcal{L}^2\Big(\bigcup_{i=K_k+1}^{N_k} 2Q_i^k \Big) \leq \mathcal{L}^2\big(\partial\Omega_k + Q(0, 9 \cdot 2^{-m_k})\big) \leq 2\mathcal{L}^2(S_k) < 2\delta_k.$$

Therefore defining $\hat{f} = g$ on S_k by Lemma 2.13 we get, using (2.33), (3.1) and the previous estimate, the following

$$\|\mathrm{d}\hat{f}\chi_{S_k}\|_{X(\Omega)} \leq C\|Df\chi_{\bigcup_{i=K_k+1}^{N_k} 2Q_i^k}\|_{X(\Omega)} < 2^{-k}C\varepsilon^*. \tag{3.12}$$

Now we deal with $\|\mathrm{d}\hat{f}\chi_{\bigcup_{\mathcal{B}_k} \mathcal{Q}_i^k}\|_{X(\Omega)}$. For all $\mathcal{Q}_i^k$ it holds that $\mathcal{Q}_i^k \subset Q(a_i, \frac{5}{4}2^{-m_k})$ and therefore, for each $y \in \mathcal{Q}_i^k$ and each $0 < r < 2^{-2-m_k}$ we get $Q(y, r) \subset 2Q_i^k$. Recall that each $\mathcal{Q}_l^k$ satisfies $\frac{5}{4}Q_l^k \supset \mathcal{Q}_l^k \supset \frac{3}{4}Q_l^k$ and that $2Q_i^k$ is the square intersecting the centres of all its neighbours Q_l^k. Therefore any 'square' $\mathcal{Q}_l^k$ intersecting $2Q_i^k$ must be a neighbour of $\mathcal{Q}_i^k$ and further, because Q_i^k and all Q_l^k neighbouring squares have the same side length $i.e.$ $|a_i - a_l| = \mathrm{diam}_\infty Q_i^k = \mathrm{diam}_\infty Q_l^k$ we have

$$2\mathcal{Q}_l^k \subset \tfrac{5}{2}Q_l^k \subset (\tfrac{5}{2} + |a_i - a_l|)Q_i^k \subset 5Q_i^k. \tag{3.13}$$

Call $A_{k,y,r} = \{1 \leq l \leq N_k; \mathcal{Q}_l^k \cap Q(y, r) \neq \emptyset\}$, then similar to (2.39) we get

$$\fint_{Q(y,r)} |\mathrm{d}\hat{f}| \, \mathrm{d}\mathcal{L}^2 \leq \max_{l \in A_{k,y,r}} \fint_{\mathcal{Q}_l^k \cap Q(y,r)} |\mathrm{d}\hat{f}| \, \mathrm{d}\mathcal{L}^2 \leq \max_{l \in A_{k,y,r}} \|\mathrm{d}\hat{f}\|_{L^\infty(\mathcal{Q}_l^k)}. \tag{3.14}$$

The combination of (2.31), (2.15) and (2.27) gives that for each $\mathcal{Q}_l^k \in \mathcal{B}_k$ and $\mathcal{Q}_l^k \in \mathcal{Z}_k$

$$\|\mathrm{d}\hat{f}\|_{L^\infty(\mathcal{Q}_l^k)} \leq C \max_{l \in A_{k,y,r}} \fint_{2Q_l^k} |Df| \, \mathrm{d}\mathcal{L}^2 \tag{3.15}$$

and the details are already in step 4 of the proof of Lemma 2.13. Considering (2.17) (2.22) and (2.27) we easily see that the above holds also for $\mathcal{Q}_l^k \in \mathcal{N}_k$. For $\mathcal{Q}_l^k \in \mathcal{G}_k$ it is immediate from (2.22). We continue (3.14) using

(3.15) and (3.13) to get

$$
\begin{aligned}
\fint_{Q(y,r)} |\mathrm{d}\hat{f}|\,\mathrm{d}\mathcal{L}^2 &\leq \max_{l \in A_{k,y,r}} \|\mathrm{d}\hat{f}\|_{L^\infty(\mathcal{Q}_l^k)} \\
&\leq \max_{l \in A_{k,y,r}} C \fint_{2Q_l^k} |Df|\,\mathrm{d}\mathcal{L}^2 \\
&\leq C \fint_{5Q_i^k} |Df|\,\mathrm{d}\mathcal{L}^2 \leq CM(|Df|)(y).
\end{aligned}
\tag{3.16}
$$

Now we consider a $y \in \mathcal{Q}_i^k$ and $r > 2^{-2-m_k}$. Using (3.15) we have

$$
\begin{aligned}
\int_{Q(y,r)} |\mathrm{d}\hat{f}|\,\mathrm{d}\mathcal{L}^2 &= \sum_{l \in A_{k,y,r}} \int_{Q(y,r) \cap \mathcal{Q}_l^k} |\mathrm{d}\hat{f}|\,\mathrm{d}\mathcal{L}^2 \\
&\leq \sum_{l \in A_{k,y,r}} \mathcal{L}^2(Q(y,r) \cap \mathcal{Q}_l^k) \|\mathrm{d}\hat{f}\|_\infty \\
&\leq C \sum_{l \in A_{k,y,r}} \mathcal{L}^2(Q(y,r) \cap \mathcal{Q}_l^k) \fint_{2Q_i^k} |Df|\,\mathrm{d}\mathcal{L}^2 \\
&\leq C \sum_{l \in A_{k,y,r}} \int_{2Q_l^k} |Df|\,\mathrm{d}\mathcal{L}^2.
\end{aligned}
$$

Now we use the fact that $8r > 2^{1-m_k}$ and $\mathcal{Q}_l^k \subset 2Q_l^k$ to see that if $l \in A_{k,y,r}$ then $Q(y,r)$ intersects $2Q_l^k$ and so $Q(y, 9r) \supset 2Q_l^k \supset \mathcal{Q}_l^k$. Therefore, using (3.6) we get

$$
\sum_{l \in A_{k,y,r}} \int_{2Q_l^k} |Df|\,\mathrm{d}\mathcal{L}^2 \leq 13 \int_{Q(y,9r)} |Df|\,\mathrm{d}\mathcal{L}^2
$$

for $r > 2^{-2-m_k}$. Thus

$$
\begin{aligned}
\fint_{Q(y,r)} |\mathrm{d}\hat{f}|\,\mathrm{d}\mathcal{L}^2 &\leq C \fint_{Q(y,9r)} |Df|\,\mathrm{d}\mathcal{L}^2 \\
&\leq CM(|Df|)(y).
\end{aligned}
\tag{3.17}
$$

By taking supremum over all $0 < r < \infty$ in (3.16) and (3.17) on the left-hand side we obtain that

$$
M(|\mathrm{d}\hat{f}|)(x) \leq CM(|Df|)(x)
$$

for $x \in \Omega_k \setminus (\Omega_{k-1} \cup S_k)$.

By the norm comparison (2.7) and Lemma 2.10 *vi)* and the ε-δ continuity of the norm we obtain that

$$
\left\| \mathrm{d}\hat{f} \chi_{\bigcup_{\mathcal{B}_k} Q} \right\|_X \leq C \left\| Df \chi_{\bigcup_{\mathcal{B}_k} Q} \right\|_X \leq C 2^{-k} \varepsilon^*.
\tag{3.18}
$$

We estimate the term $\| \sum_{\mathcal{Q} \in \mathcal{G}_k} \chi_{\mathcal{Q}} (\mathrm{d}\hat{f} - Df(a_{\mathcal{Q}})) \|_{X(\Omega)}$ in (3.8) as follows. For each $\mathcal{Q} \in \mathcal{G}_k$ we have from Lemma 2.10 point *iv)* the estimate

$$
2^{1+m_k} \|f(x) - f(a_{\mathcal{Q}}) - Df(a_{\mathcal{Q}})(x - a_{\mathcal{Q}})\|_\infty < \varepsilon_k
$$

and $\hat{f}(x) = f(x)$ at each x, vertex of $\mathcal{Q}$. Therefore on both of the triangles of $\mathcal{Q}$ we have $|\mathrm{d}\hat{f} - Df(a_{\mathcal{Q}})| < 4\varepsilon_k$. Therefore on all $\mathcal{Q} \in \mathcal{G}_k$ we have $\|\mathrm{d}\hat{f} - Df(a_{\mathcal{Q}})\|_{L^\infty(\mathcal{Q})} < 4\varepsilon_k$. This means, by (3.3), we can estimate

$$\begin{aligned}
\Big\|[\mathrm{d}\hat{f} - Df(a_{\mathcal{Q}})] \sum_{\mathcal{Q} \in \mathcal{G}_k} \chi_{\mathcal{Q}}\Big\|_{X(\Omega)} &\leq \Big\|4\varepsilon_k \sum_{\mathcal{Q} \in \mathcal{G}_k} \chi_{\mathcal{Q}}\Big\|_{X(\Omega)} \\
&< 4\varepsilon_k \varphi(\mathcal{L}^2(\Omega_k \setminus \Omega_{k-1})) \\
&< 2^{-k}\varepsilon^*.
\end{aligned}$$
(3.19)

We estimate the term $\|\sum_{\mathcal{Q} \in \mathcal{Z}_k} \chi_{\mathcal{Q}}\mathrm{d}\hat{f}\|_{X(\Omega)}$ as follows. For each $\mathcal{Q} \in \mathcal{Z}_k$ we have from Corollary 2.6, (2.28) that

$$\begin{aligned}
\|\mathrm{d}\hat{f}\|_{L^\infty(\mathcal{Q})} &< C_{(2.10)} \fint_{\partial \mathcal{Q}} |D_\tau \hat{f}| \, \mathrm{d}\mathcal{H}^1 \\
&< C_{(2.10)} C_{(2.27)} \varepsilon_k \fint_{2Q} |Df| \, \mathrm{d}\mathcal{L}^2 \\
&< C_{(2.10)} C_{(2.27)} \varepsilon_k^2 \\
&< \varepsilon_k.
\end{aligned}$$

Therefore using (3.3)

$$\|\mathrm{d}\hat{f} \sum_{\mathcal{Q} \in \mathcal{Z}_k} \chi_{\mathcal{Q}}\|_{X(\Omega)} \leq \|\varepsilon_k \sum_{\mathcal{Q} \in \mathcal{Z}_k} \chi_{\mathcal{Q}}\|_{X(\Omega)} < \varepsilon_k \varphi_X(\mathcal{L}^2(\Omega_k \setminus \Omega_{k-1})) < 2^{-k}\varepsilon^*.$$
(3.20)

Now let us estimate the $\|\sum_{\mathcal{Q} \in \mathcal{N}_k} \chi_{\mathcal{Q}}(\mathrm{d}\hat{f} - Df(a_{\mathcal{Q}}))\|_{X(\Omega)}$ term. We have defined $\hat{f}$ using Theorem 2.7 on each $\mathcal{Q} \in \mathcal{N}_k$. The extension on each $\mathcal{Q}$ has two parts, the $W_{\mathcal{Q}}$ part and the $\mathcal{Q} \setminus W_{\mathcal{Q}}$ part. Now let us estimate

$$\begin{aligned}
\Big\| \sum_{\mathcal{Q} \in \mathcal{N}_k} (\mathrm{d}\hat{f} &- Df(a_{\mathcal{Q}}))\chi_{\mathcal{Q}}\Big\|_{L^1(\bigcup_{\mathcal{N}_k} \mathcal{Q})} \\
&\leq \sum_{\mathcal{Q} \in \mathcal{N}_k} \Big(\|\mathrm{d}\hat{f} - Df(a_{\mathcal{Q}})\|_{L^1(W_{\mathcal{Q}})} + \|\mathrm{d}\hat{f} - Df(a_{\mathcal{Q}})\|_{L^1(\mathcal{Q} \setminus W_{\mathcal{Q}})}\Big) \\
&\leq \sum_{\mathcal{Q} \in \mathcal{N}_k} \Big(C\mathcal{L}^2(\mathcal{Q})\varepsilon_k + \mathcal{L}^2(\mathcal{Q} \setminus W_{\mathcal{Q}})\big(\|D\hat{f}\|_{L^\infty(\mathcal{Q})} + |Df(a_{\mathcal{Q}})|\big)\Big) \\
&\leq \sum_{\mathcal{Q} \in \mathcal{N}_k} \Big((1 + C_{(2.10)})\mathcal{L}^2(\mathcal{Q})\varepsilon_k + \varepsilon_k \mathcal{L}^2(\mathcal{Q}) \cdot 4T_k\Big) \\
&\leq (1 + C_{(2.10)} + 4T_k)\mathcal{L}^2(\Omega_k \setminus \Omega_{k-1})\varepsilon_k \\
&< \tilde{\delta}_k \mathcal{L}^2(\Omega_k \setminus \Omega_{k-1})
\end{aligned}$$

by (3.3). Of course however we have that for all $\mathcal{Q} \in \mathcal{N}_k$ that $|Df(a_{\mathcal{Q}})| < T_k$ and by (2.17) that $\|\mathrm{d}\hat{f}\|_{L^\infty(\mathcal{Q})} \leq (1 + C_{(2.10)})T_k$. Therefore

$$\Big\| \sum_{\mathcal{Q} \in \mathcal{N}_k} (\mathrm{d}\hat{f} - Df(a_{\mathcal{Q}}))\chi_{\mathcal{Q}}\Big\|_{L^\infty(\bigcup_{\mathcal{N}_k} \mathcal{Q})} \leq (2 + C_{(2.10)})T_k.$$
(3.21)

But by the choice of $\tilde{\delta}_k$ from Lemma 2.2 (see paragraph just before (3.3)) we have that because $\left\|\sum_{\mathcal{Q}\in\mathcal{N}_k}(\mathrm{d}\hat{f} - Df(a_\mathcal{Q}))\chi_\mathcal{Q}\right\|_{L^1(\bigcup_{\mathcal{N}_k}\mathcal{Q})} < \tilde{\delta}_k\mathcal{L}^2(\Omega_k\setminus\Omega_{k-1})$ and (3.21) that

$$\left\|\sum_{\mathcal{Q}\in\mathcal{N}_k}(\mathrm{d}\hat{f} - Df(a_\mathcal{Q}))\chi_\mathcal{Q}\right\|_{X(\Omega)} \leq \tilde{\varepsilon}\mathcal{L}^2(\Omega_k\setminus\Omega_{k-1}) = 2^{-k}\varepsilon^*. \tag{3.22}$$

From (3.8) and summing (3.9), (3.11) (considering (3.10)), (3.18), (3.12), (3.19), (3.20) and (3.22) we get that

$$\|(Df - \mathrm{d}\hat{f})\chi_{\Omega_k\setminus\Omega_{k-1}}\|_{X(\Omega)} < C2^{-k}\varepsilon^*.$$

Summing this over k we get that

$$\|Df - \mathrm{d}\hat{f}\|_{X(\Omega)} \leq \sum_k \|(Df - \mathrm{d}\hat{f})\chi_{\Omega_k\setminus\Omega_{k-1}}\|_{X(\Omega)} < C\varepsilon^*\sum_k 2^{-k} = C\varepsilon^*.$$

From step 4 and the fact that $\varepsilon_k < 2^{-k}\varepsilon^*$ (see (3.3)) we know that

$$\|f - \hat{f}\|_{L^\infty(\Omega)} < \varepsilon^*.$$

Step 6. Smoothing piece-wise affine maps.
From the above we may assume that we have a $\hat{f}$ satisfying

$$\|Df - \mathrm{d}\hat{f}\|_{X(\Omega)} < \tfrac{\varepsilon}{2}$$

and

$$\|f - \hat{f}\|_{L^\infty(\Omega)} < \tfrac{\varepsilon}{2}.$$

Now it suffices to apply Lemma 2.9 to prove Theorem 1.2.
Step 7. Finite-triangulation.
The finite-triangulation part of Theorem 1.2 follows immediately from the calculations above and Section 4.2 of [4]. □

References

[1] J.M. Ball, Singularities and computation of minimizers for variational problems. Foundations of computational mathematics. London Math. Soc. Lecture Note Ser., 284, Cambridge Univ. Press, Cambridge (2001) 1–20.
[2] J.M. Ball, Progress and puzzles in Nonlinear Elasticity. Poly-, Quasi- and Rank-One Convexity in Applied Mechanics. Springer (2010).
[3] C. Bennet and R. Sharpley, Interpolation of Operators. Academic Press (1988) 129.
[4] D. Campbell, Diffeomorphic approximation of Planar Sobolev Homeomorphisms in Orlicz-Sobolev spaces. *J. Funct. Anal.* **273** (2017) 125–205
[5] P. Cavaliere, A. Cianchi, L. Pick and L. Slavíková, Norms supporting the Lebesgue differentiation theorem. *Commun. Contemp. Math.* **20** (2018) 1–33

[6] G. De Philippis and A. Pratelli, The closure of planar diffeomorphisms in Sobolev spaces. *Ann. Inst. H. Poincaré Anal. Non Linéaire* **37** (2020) 181–224

[7] S. Hencl and P. Koskela, Lectures on Mappings of finite distortion. Vol. 2096 of: *Lecture Notes in Mathematics*, Springer (2014).

[8] S. Hencl and A. Pratelli, Diffeomorphic approximation of $W^{1,1}$ planar Sobolev homeomorphisms. *J. Eur. Math. Soc. (JEMS)* **20** (2018) 597–656.

[9] T. Iwaniec, L.V. Kovalev and J. Onninen, Diffeomorphic Approximation of Sobolev Homeomorphisms. Arch. Ratl. Mech. Anal. **201** (2011) 1047–1067.

[10] T. Iwaniec, L.V. Kovalev and J. Onninen, Hopf Differentials and Smoothing Sobolev Homeomorphisms. *Int. Math. Res. Notices* **14** (2012) 3256–3277.

[11] C. Mora-Corral and A. Pratelli, Approximation of piece-wise Affine Homeomorphisms by Diffeomorphisms. *J. Geom. Anal.* **24** (2014) 1398–1424.

[12] A. Pratelli, On the bi-Sobolev planar homeomorphisms and their approximation. *Nonlinear Anal.* **154** (2017) 258–268.

[13] A. Pratelli and E. Radici, Approximation of planar BV homeomorphisms by diffeomorphisms. *J. Funct. Anal.* **276** (2019) 659–686.

[14] A. Pratelli and E. Radici, On the planar minimal BV extension problems. *Rend. Lincei Mat. Appl.* **29** (2018) 511–555.

[15] E. Radici, A planar Sobolev extension theorem for piece-wise linear homeomorphisms. *Pacific J. Math.* **239** (2016) 405–418.