

DIFFEOMORPHIC APPROXIMATION OF $W^{1,1}$ PLANAR SOBOLEV HOMEOMORPHISMS

STANISLAV HENCL AND ALDO PRATELLI

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

1. INTRODUCTION

The general problem of finding suitable approximations of homeomorphisms $f : \mathbb{R}^n \supseteq \Omega \rightarrow f(\Omega) \subseteq \mathbb{R}^n$ with piecewise affine homeomorphisms has a long history. As far as we know, in the simplest non-trivial setting (i.e. $n = 2$, approximations in the L^∞ -norm) the problem was solved by Radó [30]. Due to its fundamental importance in geometric topology, the problem of finding piecewise affine homeomorphic approximations in the L^∞ -norm and dimensions $n > 2$ was deeply investigated in the '50s and '60s. In particular, it was solved by Moise [24] and Bing [7] in the case $n = 3$ (see also the survey book [25]), while for contractible spaces of dimension $n \geq 5$ the result follows from theorems of Connell [9], Bing [8], Kirby [20] and Kirby, Siebenmann and Wall [21] (for a proof see, e.g., Rushing [31] or Luukkainen [22]). Finally, twenty years later, while studying the class of quasi-conformal manifolds, Donaldson and Sullivan proved that the result is false in dimension 4: more precisely, there is a homeomorphism from the unit ball of \mathbb{R}^4 to \mathbb{R}^4 which cannot be approximated by bilipschitz homeomorphisms, see [12, Corollary, page 183].

Once completely solved in the uniform sense, the approximation problem suddenly became of interest again in a completely different context, namely, for variational models in nonlinear elasticity. Let us briefly explain why. In the setting of nonlinear elasticity (see for instance the pioneering work by Ball [3]), one is led to study existence and regularity properties of minimizers of energy functionals of the form

$$I(f) = \int_{\Omega} W(Df) dx, \quad (1.1)$$

where $f : \mathbb{R}^n \supseteq \Omega \rightarrow \Delta \subseteq \mathbb{R}^n$ ($n = 2, 3$) models the deformation of a homogeneous elastic material with respect to a reference configuration Ω and prescribed boundary values, while $W : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}$ is the stored-energy functional. In order for the model to be physically relevant, as pointed out by Ball in [4, 5], one has to require that f is a homeomorphism

2000 *Mathematics Subject Classification.* 46E35.

Key words and phrases. Mapping of finite distortion, approximation.

–this corresponds to the non-impenetrability of the material– and that

$$W(A) \rightarrow +\infty \quad \text{as } \det A \rightarrow 0, \quad W(A) = +\infty \quad \text{if } \det A \leq 0. \quad (1.2)$$

The first condition in (1.2) prevents from too high compressions of the elastic body, while the latter guarantees that the orientation is preserved.

Another property of W that appears naturally in many problems of nonlinear elasticity is the *quasiconvexity* (see for instance [2]). Unfortunately, no general existence result is known under condition (1.2), not even if the quasiconvexity assumption is added: one has either to drop condition (1.2) and impose p -growth conditions on W (see [28, 1]), or to require that W is *polyconvex* and that some coercivity conditions are satisfied (see [2, 29]). Moreover, also in the cases in which the existence of $W^{1,p}$ minimizers is known, very little is known about their regularity.

As pointed out by Ball in [4, 5] (who ascribes the question to Evans [13]), an important issue toward the understanding of the regularity of the minimizers in this setting (i.e., W quasiconvex and satisfying (1.2)) would be to show the existence of minimizing sequences given by piecewise affine homeomorphisms or by diffeomorphisms. In particular, a first step would be to prove that any homeomorphism $u \in W^{1,p}(\Omega; \mathbb{R}^n)$, $p \in [1, +\infty)$, can be approximated in $W^{1,p}$ by piecewise affine ones or smooth ones. One of the main reasons why one should want to do that, is that the usual approach for proving regularity is to test the weak equation or the variation formulation by the solution itself; but unfortunately, in general this makes no sense unless some a priori regularity of the solution is known. Therefore, it would be convenient to test the equation with a smooth test mapping in the given class which is close to the given homeomorphism. More in general, a result saying that one can approximate a given homeomorphism with a sequence of smooth (or piecewise affine) homeomorphisms would be extremely useful, because it would significantly simplify many other known proofs, and it would easily lead to stronger new results. It is important to mention here that the choice of dimension $n = 2, 3$ is not only motivated by the physical model, but also by the fact that the approximation is false in dimension $n \geq 4$, as shown in the very recent paper [17].

However, the finding of diffeomorphisms near a given homeomorphism is not an easy task, as the usual approximation techniques like mollification or Lipschitz extension using maximal operator destroy, in general, the injectivity. And on the other hand, we need of course to approximate our homeomorphism not just with smooth maps, but with smooth homeomorphisms (otherwise the approximating sequence would be not even admissible for the original problem).

Few words have to be said about the choice of the required property for the approximating sequence, namely, either smooth or piecewise affine. Actually, both results would be interesting in different contexts. Luckily, the two things are equivalent: more precisely, it is clear that an approximation with diffeomorphisms easily generates another approximation with piecewise affine homeomorphisms; the converse is not immediate but, at least in the plane, it is anyway known (see [27]). Therefore, one can approximate in either of the two ways, and the other one automatically follows (for instance, in this paper we will look only for piecewise affine approximations). It is important to clarify a point: whenever

we say that a map is piecewise affine, we mean that there is a *locally finite* triangulation of Ω such that the map is affine on every triangle. It is actually possible to find *finite* triangulations whenever this makes sense; but for instance, if Ω is not a polygon, then the triangles must obviously become smaller and smaller near the boundary, so a finite triangulation does clearly not exist.

Let us now describe the results which are known in the literature in this direction. The first ones were obtained in 2009 by Mora-Corral [26] (for planar bi-Lipschitz mappings that are smooth outside a finite set) and by Bellido and Mora-Corral [6], in which they prove that if $u, u^{-1} \in C^{0,\alpha}$ for some $\alpha \in (0, 1]$, then one can find piecewise affine approximations v of u in $C^{0,\beta}$, where $\beta \in (0, \alpha)$ depends only on α .

More recently, Iwaniec, Kovalev and Onninen [18] solved the approximation problem of planar Sobolev homeomorphisms in the case $1 < p < \infty$, proving that any homeomorphism $f \in W^{1,p}(\Omega, \mathbb{R}^2)$ can be approximated by diffeomorphisms f_ε in the $W^{1,p}$ norm (improving the previous result, of the same authors, for the case $p = 2$, see [19]). This was a fundamental breakthrough in the area and strengthened the research interest in this topic.

Later on, it was shown by Daneri and Pratelli in [10] and [11] that any planar bi-Lipschitz mapping f can be approximated by diffeomorphisms f_k such that f_k converge to f in $W^{1,p}$ norm and simultaneously f_k^{-1} converge to f^{-1} in $W^{1,p}$, giving the first result in which also the distance of the inverse mappings is approximated.

The goal of the present paper is to prove the approximation of planar $W^{1,1}$ homeomorphism in the $W^{1,1}$ sense, so basically dealing with the important case $p = 1$ which was left out in [18]. In particular, our main result is the following.

Theorem 1.1. *Let $\Omega \subseteq \mathbb{R}^2$ be an open set and $f \in W^{1,1}(\Omega, \mathbb{R}^2)$ be a homeomorphism. For every $\varepsilon > 0$ there is a smooth diffeomorphism (as well as a countably –but locally finitely– piecewise affine homeomorphism) $f_\varepsilon \in W^{1,1}(\Omega, \mathbb{R}^2)$ such that $\|f_\varepsilon - f\|_{W^{1,1}} + \|f_\varepsilon - f\|_{L^\infty} < \varepsilon$. Moreover, $f_\varepsilon(\Omega) = f(\Omega)$ and in particular, if f is continuous up to the boundary of Ω , then the same holds for f_ε , and $f_\varepsilon = f$ on $\partial\Omega$.*

Actually, our piecewise affine functions f_ε will be globally *finitely* piecewise affine, thus also bi-Lipschitz, as soon as Ω is a polygon and f is piecewise affine on $\partial\Omega$, see Theorem 4.20. If even just one of these two conditions does not hold, then this is clearly impossible (see Remark 4.21). Notice that all the approximating mappings f_ε satisfy $f_\varepsilon(\Omega) = f(\Omega)$, which is essential to be admissible competitors for the energy. In addition, all the mappings f_ε coincide with f on the boundary of Ω whenever the boundary value of f is defined, and this is of course essential in all the situations where the boundary value is important.

We conclude the introduction with a short comparison between the techniques of this paper and those of the other papers discussed above. The proofs in [26, 6] are based on a clever refinement of the supremum norm approximation of Moise [24], while the approach of [18] and of the other contributions of the same authors makes use of the identification $\mathbb{R}^2 \simeq \mathbb{C}$ and involves coordinate-wise p -harmonic functions. The techniques of the present

paper are completely different with respect to them; basically, our proof is constructive, and it is based on an explicit subdivision of the domain f which depends on the Lebesgue points of Df .

Our techniques resembles the basic ideas of [10] and [11], and we will also use some of the tools introduced there, but there are some extremely important differences. More precisely, on one hand in [10] and [11] one had to approximate at the same time Df and Df^{-1} while here we need only to approximate Df , and this is a deep simplification. But on the other hand, in this paper we look for a sharp estimate, that is, f is only in $W^{1,1}$ and we want an approximation exactly in $W^{1,1}$, thus we have not so regular maps and we cannot lose sharpness of the power anywhere, while in [10] and [11] the maps were much better, namely bi-Lipschitz, and in several steps the sharpness of the power was lost. Roughly speaking, we can say that the most difficult steps of [10] and [11] correspond to much simpler steps here, and vice versa.

1.1. Brief description of the proof. In this section we outline the basic plan of our proof, to underline the main steps and help the reading of the construction. We remind the reader that our aim is to find an approximation done with piecewise affine homeomorphisms, and then the existence of an approximation with smooth diffeomorphisms will eventually immediately follow applying the result of [27].

First of all, we will divide our domain into some locally finite grid of small squares, these squares becoming maybe smaller and smaller close to $\partial\Omega$. We will then consider separately the “good” squares, and the “bad” ones. More precisely, a square $\mathcal{S}(c, r)$ in the grid will be called good if f can be well approximated by a linear mapping $f(c) + M(x - c)$ there, where M coincides with Df in some Lebesgue point close to c : in particular, we will need that $\int_{\mathcal{S}(c,r)} |Df - Df(c)|$ is small enough. Since almost every point of Ω is a Lebesgue point for Df , we will be able to deduce that, up to consider a sufficiently fine grid, the area covered by the good squares is as close as we wish to the total area of Ω .

Moreover, up to a slight modification of the value of f on the boundary of the squares, we will reduce ourselves to the case that

$$\int_{\partial\mathcal{S}} |Df| \leq K \int_{\mathcal{S}} |Df|, \quad (1.3)$$

where K is a big, but fixed, constant.

We will then define an approximation of f (which will eventually become f_ε) on the grid: on the boundary of each square, we will find a piecewise linear approximation of f , very close to f , in such a way that these approximations on the whole grid remain one-to-one (we do not have just to take care of the approximation on a single square, but also check that the different approximations coincide on the common sides, and that they do not overlap with each other). Of course this will be much easier for the good squares, since on a whole good square f is already almost affine, and more complicated for the bad squares. We will do our approximation g in such a way that, for any bad square \mathcal{S} ,

$$\int_{\partial\mathcal{S}} |Dg| \leq K \int_{\partial\mathcal{S}} |Df|. \quad (1.4)$$

The next step is then to extend the piecewise linear maps to the interior of each square; a good thing is that, being g already defined on the grid in a one-to-one way, the extension inside each square is completely independent with what happens on the other squares. The rough idea to do so is that on good squares we can obtain very good estimates, while in bad squares we can get only bad estimates; but since the total area of the bad squares is arbitrarily small, in the end everything will work.

The first tool which we will need, presented in Section 2, says that any piecewise linear map g defined on the boundary of a square \mathcal{S} can be extended to a piecewise affine homeomorphism h in the interior of \mathcal{S} in such a way that

$$\int_{\mathcal{S}} |Dh| \leq K \int_{\partial\mathcal{S}} |Dg|. \quad (1.5)$$

This construction is done first by choosing many points on the boundary of the square; then, for any two of these points, say x and y , we will select the shortest path joining $g(x)$ with $g(y)$ remaining inside the portion of \mathbb{R}^2 having $\varphi(\mathcal{S})$ as boundary. Using in a careful way these shortest paths, we will then eventually obtain the definition of h such that (1.5) holds true. This estimate, together with (1.3) and (1.4), readily implies that for every bad square \mathcal{S} one has

$$\int_{\mathcal{S}} |Dh| \leq K \int_{\mathcal{S}} |Df|; \quad (1.6)$$

we have then just to take a very fine grid, so that a very small portion Ω_B of Ω is covered with bad squares, and hence we will have

$$\int_{\Omega_B} |Df - Dh| \leq \int_{\Omega_B} |Df| + |Dh| \leq (K + 1) \int_{\Omega_B} |Df| \leq \varepsilon.$$

It remains then to consider the good squares, and here we will have to be extremely precise. As already said, around every good square \mathcal{S} the map f is very close to being affine, hence the image of \mathcal{S} is very close to a parallelogram; therefore, there is no problem unless this parallelogram degenerates. Let us be more precise: for all the squares corresponding to a matrix M with strictly positive determinant (hence, the parallelogram does not degenerate), the extension inside \mathcal{S} is trivial; it is enough to divide the square in two triangles and consider on each triangle the affine map which equals f on the three vertices. By construction, we will easily see that this works perfectly.

The good squares corresponding to $M = 0$ are a problem only in principle: indeed, we can treat them as bad squares. The estimate (1.6) says that this gives a cost of a big constant K times the total integral of Df on those squares; however, since they are good squares and the corresponding matrix is $M = 0$, by definition the integral of Df will be extremely small, and everything will work.

The hard problem, instead, is for good squares for which $M \neq 0$ but $\det M = 0$: these correspond to degenerate parallelograms, and we have to treat them carefully because these squares can cover a large portion of Ω : in fact, recall that the set $\{x : |Df(x)| \neq 0 \text{ and } J_f(x) = 0\}$ can have positive or even full measure for a Sobolev homeomorphism (see [15]). Section 3 is devoted to build the extension for this specific case, which is somehow similar to the one with the shortest paths described above. The big difference

with that case is, on one hand, that this time we are in a good square, hence very close to a Lebesgue point for Df , and this helps even for the degenerate case. But on the other hand, this time we are not satisfied with an estimate like (1.5), where a big constant K appears, and we need instead an approximation h which is very close to the original f . This extension procedure will be the most delicate step of the construction.

The construction of the proof, divided in its several steps, is done in the last Section 4. Basically, putting together all the ingredients described above, the proof will then be concluded for what concerns the existence of the piecewise affine approximation; the existence of the smooth approximation will then follow thanks to the result of [27], while the claim about the boundary values will be easily deduced by the whole construction.

1.2. Preliminaries and notation. In this section we shortly list the basic notation that will be used throughout the paper. By $\mathcal{S}(c, r)$ we denote the square centered at c , side length $2r$ and sides parallel to the coordinate axis, while $\mathcal{S}_0 = \{(x, y) \in \mathbb{R}^2 : |x| + |y| < 1\}$ is the “rotated square”, that we use only in Section 2. Similarly, $\mathcal{B}(c, r)$ is the ball centered at c with radius r .

The points in the domain Ω will be usually denoted by capital letters, such as A, B and so on, while points in the image $f(\Omega)$ will be always denoted by bold capital letters, such as \mathbf{A}, \mathbf{B} and similar. To shorten the notation and help the reader, whenever we use the same letter A for a point in the domain and \mathbf{A} (in bold) for a point in the target, this always means that \mathbf{A} is the image of A under the mapping that we are considering in that moment. By AB (resp., \mathbf{AB}) we denote the segment between the points A and B (resp. \mathbf{A} and \mathbf{B}). The length of this segment is denoted as $\mathcal{H}^1(AB)$, or as \overline{AB} , while $\mathcal{H}^1(\gamma)$ is the length of a curve γ . With the notation \widehat{AB} (or $\widehat{\mathbf{AB}}$) we will denote a particular path between A and B (or \mathbf{A} and \mathbf{B}), whose length will be then $\mathcal{H}^1(\widehat{AB})$, or $\mathcal{H}^1(\widehat{\mathbf{AB}})$; we will use this notation only when it is clear what is the path we are referring to (often this will be a shortest path between the points). Given three non-aligned points A, B, C (or $\mathbf{A}, \mathbf{B}, \mathbf{C}$), we will denote by \widehat{ABC} (or $\widehat{\mathbf{ABC}}$) the angle in $(0, \pi)$ between them, and by ABC (or \mathbf{ABC}) the triangle having them as vertices.

We will denote the (modulus of the) horizontal and vertical derivatives of any mapping $f = (f_1, f_2) : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ as

$$|D_1 f| = \sqrt{\left(\frac{\partial f_1}{\partial x}\right)^2 + \left(\frac{\partial f_2}{\partial x}\right)^2}, \quad |D_2 f| = \sqrt{\left(\frac{\partial f_1}{\partial y}\right)^2 + \left(\frac{\partial f_2}{\partial y}\right)^2}.$$

Analogously, the derivatives of the components f_1 and f_2 are written as

$$D_1 f_1 = \frac{\partial f_1}{\partial x}, \quad D_2 f_1 = \frac{\partial f_1}{\partial y}, \quad D_1 f_2 = \frac{\partial f_2}{\partial x}, \quad D_2 f_2 = \frac{\partial f_2}{\partial y}.$$

Whenever a continuous function g is defined on some curve γ (usually, γ will simply be the boundary of a square) we will denote by $\tau(t)$ the tangent vector to γ at $t \in \gamma$, and by $Dg(t)$ the derivative of g at t in the direction of $\tau(t)$. With a small abuse of notation, even if the derivative is not necessarily defined, we will write $\int_\gamma |Dg(t)| d\mathcal{H}^1(t)$ to denote the length of the curve $g(\gamma)$: notice that the latter length is always well-defined, possibly $+\infty$, and it actually coincides with $\int_{\partial\mathcal{S}} |Dg(t)|$ as soon as this is defined. Finally, notice

that, if a function f is affine on a square \mathcal{S} , being $Df \equiv M$ for some matrix M , and we call g the restriction of f to $\partial\mathcal{S}$, then $Dg(t) = M \cdot \tau(t)$ for any $t \in \partial\mathcal{S}$.

The letter K will always be used to denote a large purely geometrical constant, not depending on anything; we will not modify the letter, even if the constant may always increase from line to line. For the sake of simplicity (and since the precise value of K does not play any role) we do not explicitly calculate the value of this constant.

2. EXTENSION FROM THE BOUNDARY OF THE SQUARE

This section is entirely devoted to show the result below about the extension of a map from the boundary of the square to the whole interior.

Theorem 2.1. *Let $g : \partial\mathcal{S}_0 \rightarrow \mathbb{R}^2$ be a piecewise linear and one-to-one function. There is a finitely piecewise affine homeomorphism $h : \mathcal{S}_0 \rightarrow \mathbb{R}^2$ such that $h = g$ on $\partial\mathcal{S}_0$, and*

$$\int_{\mathcal{S}_0} |Dh(x)| dx \leq K \int_{\partial\mathcal{S}_0} |Dg(t)| d\mathcal{H}^1(t). \quad (2.1)$$

Proof. The construction of the map h is quite long and technical, and hence we subdivide it in several steps.

Step 1. *Choice of good corners, so that (2.2) holds.*

For our construction, we will need to assume that $\int_{\partial\mathcal{S}_0} |Dg|$ does not concentrate too much around the corners; more precisely, we will need that

$$\int_{\mathcal{B}(V_i, r) \cap \partial\mathcal{S}_0} |Dg| d\mathcal{H}^1 \leq Kr \int_{\partial\mathcal{S}_0} |Dg| d\mathcal{H}^1 \quad \text{for all } r \in (0, 1), i \in \{1, 2\}, \quad (2.2)$$

being $V_1 \equiv (0, -1)$ and $V_2 \equiv (0, 1)$. It is quite easy to achieve that: in fact, it is enough to find two opposite points $P_1, P_2 \in \partial\mathcal{S}_0$ such that

$$\int_{\mathcal{B}(P_i, r) \cap \partial\mathcal{S}_0} |Dg| \leq 6r \int_{\partial\mathcal{S}_0} |Dg| \quad \text{for all } r \in (0, \sqrt{2}), i \in \{1, 2\}, \quad (2.3)$$

because then we can apply a bi-Lipschitz transformation (with bi-Lipschitz constant independent from P_1 and P_2) which moves the points P_1 and P_2 on the vertices V_1 and V_2 , and get (2.2). And in turn, to obtain (2.3), we notice that every point of $\partial\mathcal{S}_0$ is a possible choice for P_1 or P_2 unless it is, or its opposite point is, in the set

$$\mathcal{A} := \left\{ P \in \partial\mathcal{S}_0 : \exists r \in (0, 1) : \int_{\mathcal{B}(P, r) \cap \partial\mathcal{S}_0} |Dg| > 6r \int_{\partial\mathcal{S}_0} |Dg| \right\}.$$

By a Vitali covering argument, we can cover \mathcal{A} with countably many balls $\mathcal{B}(P_i, 3r_i)$ such that every P_i is in \mathcal{A} , and the corresponding sets $\mathcal{B}(P_i, r_i) \cap \partial\mathcal{S}_0$ are as in the definition of \mathcal{A} and are pairwise disjoint. Therefore, we can calculate

$$\mathcal{H}^1(\mathcal{A}) \leq \sum_i 6r_i \leq \sum_i \frac{\int_{\mathcal{B}(P_i, r_i) \cap \partial\mathcal{S}_0} |Dg|}{\int_{\partial\mathcal{S}_0} |Dg|} \leq 1,$$

and since $\mathcal{H}^1(\partial\mathcal{S}_0) = 4\sqrt{2}$ it clearly follows that two opposite points both in $\partial\mathcal{S}_0 \setminus \mathcal{A}$ exist and then satisfy (2.3), as required.

Step 2. *Definition of the grid on $\partial\mathcal{S}_0$, and of the paths γ^i .*

To define our map h , we will make use of a fine grid made by horizontal segments

in \mathcal{S}_0 . More precisely, we will take several (but finitely many) distinct points $A^0 \equiv (0, -1)$, $A^1, A^2, \dots, A^k \equiv (0, 1)$ in $\partial\mathcal{S}_0$, all with non-positive first coordinate A_1^i and with second coordinate A_2^i increasing, with respect to i , from -1 to 1 ; on the opposite side, we will take the corresponding points $B^i \equiv (-A_1^i, A_2^i)$, so that the segments $A^i B^i$ are horizontal.

The way to choose our points is simple: since g is piecewise linear, we can take the points in such a way that g is linear on every segment $A^i A^{i+1}$, as well as in every $B^i B^{i+1}$. Since this property is of course not destroyed if we *add* more points A^i (as long as we also add the corresponding points B^i , of course), we are allowed to add more points during the construction, of course taking care to add only finitely many: we will do this a first time in few lines, and then also later.

From now on, we will call \mathcal{S}_0 the bounded component of $\mathbb{R}^2 \setminus g(\partial\mathcal{S}_0)$, which is a polygon because g is piecewise linear; notice that the map h that we want to construct must be a homeomorphism between \mathcal{S}_0 and \mathcal{S}_0 . Then, for any $0 < i < k$, we define γ^i the shortest path which connects A^i and B^i inside the closure of \mathcal{S}_0 (this shortest path is unique, as we will show in Step 3). Notice that, since \mathcal{S}_0 is a polygon, every γ^i is piecewise linear, and any junction between two consecutive linear pieces is in $\partial\mathcal{S}_0$.

Up to add one more point between A^0 and A^1 (plus the corresponding one on the right part), we can suppose that γ^1 is either a segment between A^1 and B^1 , and this happens if and only if the angle between A^1, A^0 and B^1 which goes inside \mathcal{S}_0 is smaller than π , or it is done by the union of the two segments $A^1 A^0$ and $A^0 B^1$, thus it entirely lies on $\partial\mathcal{S}_0$. We do the same between A^{k-1} and A^k .

Step 3. Uniqueness of the shortest paths.

Let \mathcal{Q} be a simply connected closed planar domain with polygonal boundary. We briefly recall the proof of the well-known fact that, for any two points in \mathcal{Q} –not necessarily on the boundary– there is a unique shortest path inside \mathcal{Q} . Since the existence is obvious, we just have to check the uniqueness.

If the claim were not true, there would be two points $A, B \in \mathcal{Q}$ and two shortest paths τ_1 and τ_2 between A and B inside \mathcal{Q} , such that τ_1 and τ_2 meet only at A and B . The union of the two paths is then a polygon, say with n sides. The sum of the internal angles of this polygon is $\pi(n-2)$, and thus there must be a vertex of the polygon, different from A and B and thus inside one of the shortest paths, corresponding to an angle strictly less than π . Since the interior of the polygon is entirely in the interior of \mathcal{Q} , this is of course impossible, because cutting around that vertex would strictly shorten the length of the path, against the minimality.

Step 4. The path γ^{i+1} is above γ^i , and definition of $\gamma_1^i, \gamma_2^i, \gamma_3^i$.

For two curves γ and $\tilde{\gamma}$ inside \mathcal{S}_0 and with endpoints in $\partial\mathcal{S}_0$, we say that “ γ is above $\tilde{\gamma}$ ” if γ does not intersect the interior of the (possibly disconnected) subset of \mathcal{S}_0 whose boundary is the union between $\tilde{\gamma}$ and the path on $\partial\mathcal{S}_0$ connecting the endpoints of $\tilde{\gamma}$ and containing $A^0 = g(A^0)$. We want to show that, for any $0 < i < k-1$, the path γ^{i+1} is above γ^i .

To show that, assume that two points \mathbf{P} and \mathbf{Q} belong to both the paths γ^i and γ^{i+1} . Then, the two restrictions of γ^i and γ^{i+1} from \mathbf{P} to \mathbf{Q} are two shortest paths, and by Step 3 we derive that γ^i and γ^{i+1} coincide between \mathbf{P} and \mathbf{Q} . As an immediate consequence of this observation, we get that γ^{i+1} is above γ^i as claimed.

Another immediate consequence is the following: the intersection between γ^i and γ^{i+1} is always a connected subpath, possibly empty. If it is not empty, and then it is a path $\widehat{\mathbf{PQ}}$, we will subdivide both γ^i and γ^{i+1} in three parts, writing

$$\gamma^i = \gamma_1^i \cup \gamma_2^i \cup \gamma_3^i, \quad \gamma^{i+1} = \gamma_1^{i+1} \cup \gamma_2^{i+1} \cup \gamma_3^{i+1},$$

where γ_1^i (resp. γ_1^{i+1}) is the first part, from \mathbf{A}^i to \mathbf{P} (resp., from \mathbf{A}^{i+1} to \mathbf{P}); γ_2^i (resp. γ_2^{i+1}) is the second part, from \mathbf{P} to \mathbf{Q} (thus the common part, and $\gamma_2^i = \gamma_2^{i+1}$); and γ_3^i (resp., γ_3^{i+1}) is the third and last part, from \mathbf{Q} to \mathbf{B}^i (resp., from \mathbf{Q} to \mathbf{B}^{i+1}). If γ^i and γ^{i+1} have empty intersection, then we simply set $\gamma_1^i = \gamma^i$ and $\gamma_1^{i+1} = \gamma^{i+1}$, letting $\gamma_2^i, \gamma_3^i, \gamma_2^{i+1}$ and γ_3^{i+1} be empty paths. The situation is depicted in Figure 1, where the common part $\gamma_2^i = \gamma_2^{i+1}$ is done by two segments, one on $\partial\mathcal{S}_0$ and the other one in the interior of \mathcal{S}_0 .

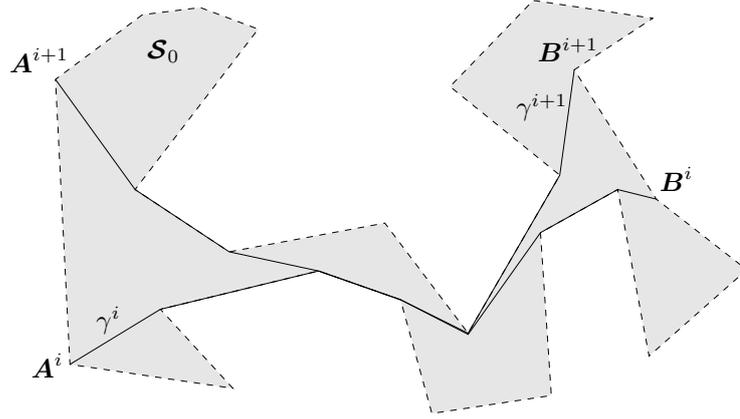


FIGURE 1. The paths γ^i and γ^{i+1} in Step 4.

Notice that this subdivision of a path does not depend only on the path itself, but also on the other path that we are considering; in other words, the subdivision of the path γ^j done when $i = j$, and then considering the possible common part between γ^j and γ^{j+1} , does not need to coincide with the subdivision of the same path done when $i = j - 1$, and then considering the possible common part between γ^j and γ^{j-1} .

Step 5. Convexity of the polygon having boundary $\gamma_1^{i+1} \cup \mathbf{A}^{i+1}\mathbf{P}$.

Let us call \mathbf{P} the last point of the path γ_1^{i+1} ; hence \mathbf{P} is the first common point with γ^i , if γ^i and γ^{i+1} have a non-empty intersection, while otherwise $\mathbf{P} = \mathbf{B}^{i+1}$. We claim that the polygon having $\gamma_1^{i+1} \cup \mathbf{A}^{i+1}\mathbf{P}$ as its boundary is convex (notice that in principle the curve γ_1^{i+1} and the segment $\mathbf{A}^{i+1}\mathbf{P}$ could have other intersection points in addition to \mathbf{A}^{i+1} and \mathbf{P}). We start assuming that $\gamma_2^{i+1} \neq \emptyset$, at the end of this step we will then consider the other case.

If γ_1^{i+1} is a single point, or just a segment, then the claim is empty true, and the convex polygon is degenerate. Let us assume then that γ_1^{i+1} is done at least by two affine pieces, and assume also, just to fix the ideas, that the direction of the oriented segment $\mathbf{A}^i \mathbf{A}^{i+1}$ is $\pi/2$, as in Figure 2, left. Call then $\mathcal{D} \subseteq \mathcal{S}_0$ the polygon having, as boundary, the Jordan curve $\gamma_1^i \cup \gamma_1^{i+1} \cup \mathbf{A}^{i+1} \mathbf{A}^i$. The same argument as in Step 3 immediately ensures that, for any vertex of γ^{i+1} (i.e., any junction point between two consecutive linear pieces of γ_1^{i+1}), the angle pointing inside \mathcal{S}_0 (hence in particular inside \mathcal{D}) is bigger than π . By construction, and recalling Step 4, we get also that none of these points can belong to the curve in $\partial \mathcal{S}_0$ connecting \mathbf{A}^i and \mathbf{B}^i and containing \mathbf{A}^0 , since such a point should necessarily belong also to γ^i , against the definition of \mathbf{P} . Of course, this already “suggests” that our convexity claim is true, but observe that the proof is still not over, since in principle γ_1^{i+1} could be some spiral-like curve connecting \mathbf{A}^{i+1} with \mathbf{P} . To conclude the proof, for any vertex of γ_1^{i+1} (except \mathbf{P}) consider the range of directions pointing toward the interior of \mathcal{D} : for instance, the range associated to \mathbf{A}^{i+1} in the situation of Figure 2, left, is done by the angles between $-\pi/2$ and $-\pi/3$. We claim that, for each vertex of the curve γ_1^{i+1} , this range cannot contain the angle $+\pi/2$: observe that this will immediately imply the required convexity.

Assume then by contradiction that this claim is false, and let \mathbf{Q} be the first vertex of γ_1^{i+1} having $\pi/2$ in its range of directions; by a trivial perturbation argument we can assume that $\pi/2$ is in the interior of this range, and then the vertical line passing through \mathbf{Q} is in the interior of \mathcal{D} for a while, both above and below \mathbf{Q} itself. Call then, as in Figure 2, left, \mathbf{Q}^- and \mathbf{Q}^+ the first points of this line, respectively below and above \mathbf{Q} , which are on $\partial \mathcal{D}$. Since the segment $\mathbf{Q}^- \mathbf{Q}^+$ is parallel to $\mathbf{A}^i \mathbf{A}^{i+1}$, each of these points must belong either to γ_1^i or to γ_1^{i+1} . Observe now that, if γ is a shortest path in \mathcal{S}_0 between its extremes, it is also a shortest path in \mathcal{S}_0 between any pair of its points. In particular, if the line connecting two points of γ is entirely in the closure of \mathcal{S}_0 , then γ must be the segment between these two points. This immediately imply that none of the points \mathbf{Q}^- and \mathbf{Q}^+ can belong to γ_1^{i+1} , because otherwise γ_1^{i+1} should be a segment between that point and \mathbf{Q} ; as a consequence, both the points \mathbf{Q}^- and \mathbf{Q}^+ must belong to γ_1^i , but this is also impossible because then γ_1^i should be the segment between them. The contradiction shows the claim, and then we have obtained the required convexity. Of course, the very same argument works for the polygon having boundary $\gamma_3^{i+1} \cup \mathbf{P} \mathbf{B}^{i+1}$, being this time \mathbf{P} the first point of γ_3^{i+1} , and everything also works for the polygons around the path γ^i instead of γ^{i+1} .

Let us then consider the case when $\gamma_2^{i+1} = \emptyset$, that is, the case when γ^i and γ^{i+1} are disjoint: this situation is depicted in Figure 2, right. This time, we define \mathcal{D} the polygon having, as boundary, the Jordan curve $\gamma^i \cup \mathbf{B}^i \mathbf{B}^{i+1} \gamma^{i+1} \cup \mathbf{A}^{i+1} \mathbf{A}^i$. The very same argument as in the first case ensures again that every vertex of γ^{i+1} has angle bigger than π in the direction inside \mathcal{D} ; as a consequence, the required convexity follows, as before, if the range of every vertex of γ^{i+1} does not contain the direction $+\pi/2$.

However, this time it is not impossible that a vertex \mathbf{Q} of γ^{i+1} has $\pi/2$ in its range. Let then, as before, \mathbf{Q} be the first vertex (if any) with this property, and let $\mathbf{Q}^\pm \in \partial \mathcal{D}$ be

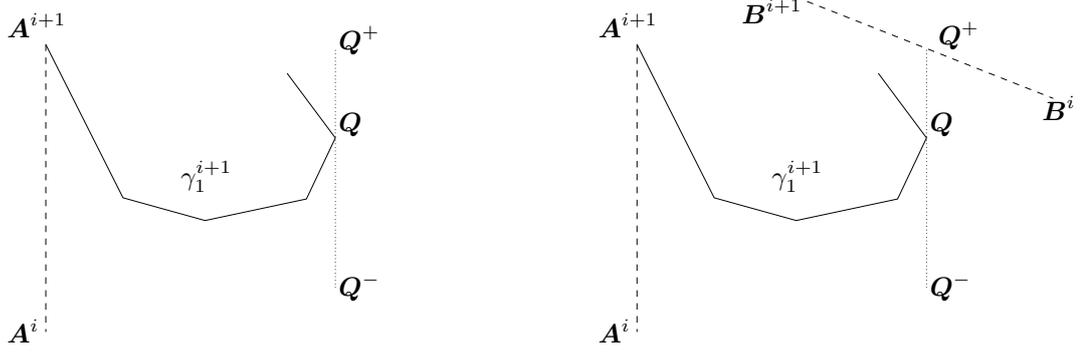


FIGURE 2. Construction in Step 5.

as before. As already noticed, none of the points Q^\pm can be in γ^{i+1} , and at most one in γ^i . Hence, the only possibility is that one point is in γ^i , and the other one in $B^i B^{i+1}$. A simple topological argument ensures that Q^- must be in γ^i and Q^+ in $B^i B^{i+1}$. Indeed, consider the path, contained in $\partial\mathcal{D}$ and not containing $A^i A^{i+1}$, which connects Q and Q^- ; together with the segment $Q^- Q$, this is a Jordan curve, and then it can not intersect the other path in $\partial\mathcal{D}$ which contains $A^i A^{i+1}$: in particular, it must contain Q^+ , and it readily follows, as claimed, that $Q^- \in \gamma^i$, $Q^+ \in B^i B^{i+1}$. The very same topological argument ensures also that B^{i+1} is the “left” vertex (that is, the one in the direction $A^i A^{i+1}$) of the segment $B^i B^{i+1}$, and B^i is the “right” one, as in Figure 2, right.

Let us now restrict our attention to the subset \mathcal{D}_0 of \mathcal{D} made by the polygon whose boundary is the part of γ^{i+1} connecting Q to B^{i+1} , plus the two segments $B^{i+1} Q^+$ and $Q^+ Q$. The same argument of the first half of this step ensures that the range of directions, toward the interior of \mathcal{D}_0 , corresponding to any vertex of γ^{i+1} in $\partial\mathcal{D}_0$, can never contain the direction of the segment $B^i B^{i+1}$, since otherwise a segment parallel to $B^i B^{i+1}$ and contained in \mathcal{D}_0 should have both the endpoints in the segment $Q Q^+$, which is impossible.

Finally, it is immediate to notice that this property of the directions, analogously as before, is enough to ensure the required convexity of the polygon having $\gamma^{i+1} \cup A^{i+1} P$ as boundary.

Step 6. Definition of the “vertical segments” and their length.

In this step, we associate to any vertex P of the curve γ^{i+1} a point (or many points) Q of the curve γ^i , and vice versa. Every such segment PQ , which we will call “vertical”, will be contained in the closure of the polygon $\mathcal{D} \subseteq \mathcal{S}_0$ having boundary $A^i A^{i+1} \cup \gamma^{i+1} \cup B^{i+1} B^i \cup \gamma^i$, any two vertical segments will have empty intersection, except possibly at a common endpoint, and the following estimate for the length of the vertical segments will hold,

$$\mathcal{H}^1(PQ) \leq \max \left\{ \mathcal{H}^1(A^i A^{i+1}), \mathcal{H}^1(B^i B^{i+1}) \right\}. \quad (2.4)$$

Let us give our definition distinguishing the possible cases, as in Step 5.

First of all, consider the situation, depicted in Figure 3, left, when γ^i and γ^{i+1} have a non-empty intersection. In the common part $\gamma^i \cap \gamma^{i+1} = \gamma_2^i = \gamma_2^{i+1}$, we will associate to any vertex P of γ^{i+1} the same point $Q \equiv P$, which is also in γ^i by definition. The

segment PQ is just a point, which is of course in the closure of \mathcal{D} , and the length is 0, so that (2.4) of course holds. In the “left” part of the paths, instead, we will give the following simple definition. To any vertex $P \in \gamma_1^{i+1}$, we associate the point $Q \in \gamma_1^i$ so that the segment PQ is parallel to $A^{i+1}A^i$: the existence and uniqueness of such a point, the validity of (2.4), and the fact that PQ is contained in the closure of \mathcal{D} , all come immediately from the convexity obtained in Step 5. We do the very same thing for the vertices of γ_1^i , and we argue completely similarly for the “right” part of the paths, of course taking segments parallel to $B^{i+1}B^i$, instead of $A^{i+1}A^i$. Then we have already completed our definition of the vertical segments, and the fact that any two such segments do not intersect is obvious from the construction.

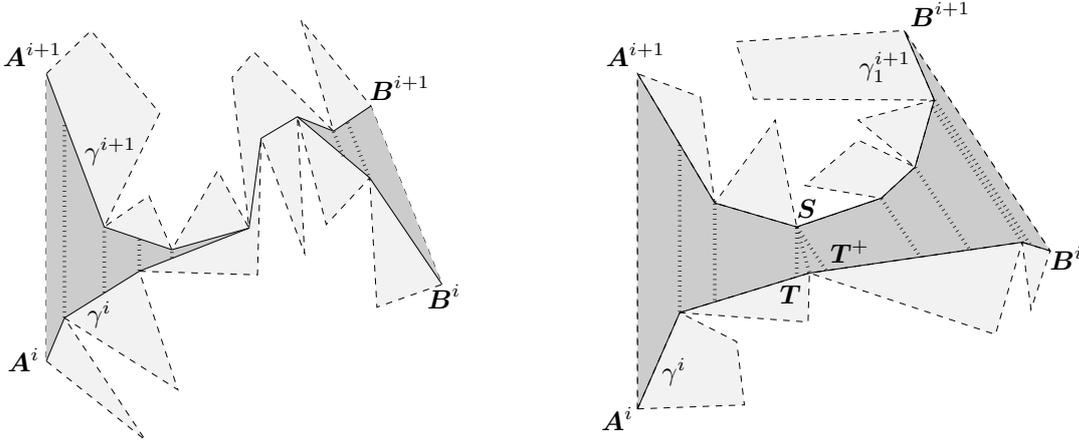


FIGURE 3. Construction in Step 6: the polygon \mathcal{S}_0 (resp., \mathcal{D}) is light (resp., dark) coloured, and the “vertical segments” are dotted.

Consider now the situation when $\gamma^i \cap \gamma^{i+1} = \emptyset$, see Figure 3, right. Without loss of generality we can think that, as in the Figure, the direction of A^iA^{i+1} is vertical, while the segment B^iB^{i+1} goes “toward left”. Let then $S \in \gamma^{i+1}$ and $T \in \gamma^i$ be the two closest points such that the segment TS is vertical; notice that it is possible that $S = A^{i+1}$ or that $S = B^{i+1}$, this makes no difference in our proof, even if the picture shows an example where S is in the interior of the curve γ^{i+1} .

Let us now consider the subset $\mathcal{D}_0 \subseteq \mathcal{D}$ whose boundary is given by the two segments A^iA^{i+1} and ST , together with the parts of γ^i (resp., γ^{i+1}), connecting A^i and T (resp., A^{i+1} and S). Again by the convexity result of Step 5, it is clear that at any point of γ^{i+1} between A^{i+1} and S starts a vertical segment, whose interior is entirely contained in \mathcal{D}_0 , which ends in a point of γ^i between A^i and T . We define then in the obvious way the “vertical segments” inside \mathcal{D}_0 , which are in fact vertical. The validity of (2.4) is as usual obvious from the convexity.

Consider now the half-line starting at S and parallel to $B^{i+1}B^i$. The choice of the points S and T , together with the convexity proved in Step 5, ensure that this half-line remains inside \mathcal{D} for a while, after the point S ; therefore, the intersection of this half-line with \mathcal{D} is a segment ST^+ , and the point T^+ is on γ^i by construction. Observe that T^+ coincides with T in the particular case when $B^{i+1}B^i$ is parallel to $A^{i+1}A^i$, but otherwise

it stays, as in the figure, outside of \mathcal{D}_0 . The construction implies that all the half-lines, parallel to $\mathbf{B}^{i+1}\mathbf{B}^i$ and starting at a point of γ^{i+1} after \mathbf{S} , remain in \mathcal{D} for a while and then intersect γ^i at some point after \mathbf{T}^+ . We use this observation to associate to any vertex of γ^{i+1} after \mathbf{S} a point of γ^i after \mathbf{T}^+ , and we call then “vertical segments” all the corresponding segments, which are actually not vertical but parallel to $\mathbf{B}^{i+1}\mathbf{B}^i$. Finally, to every vertex of γ^i between \mathbf{T} and \mathbf{T}^+ , if any, we associate always the point \mathbf{S} . The validity of (2.4) for all the vertical segments is then again clear by the construction and by Step 5, and any two vertical segments have always empty intersection, unless in the case when they meet at \mathbf{S} . This concludes the step.

From now on, we will always consider as “vertices” the points \mathbf{S} , \mathbf{T} and \mathbf{T}^+ , even if they were not vertices in the sense of the piecewise linear curves. Moreover, for every vertex of γ^i , or of γ^{i+1} , we will consider as “vertex” also the corresponding point in the other curve, which again could be or not be a vertex in the classical sense. Notice that in this way we are adding a *finite number* of new vertices and, as already pointed out before, it is always admissible to regard as “vertices” also finitely many new points in our curves. Summarizing, on the piecewise linear curve γ^i we are considering as “vertices” all the actual vertices, plus some other new points. However, these “new points” have been selected working on the region between γ^i and γ^{i+1} , and then they do not need to coincide with the “new points” selected by working on the region between γ^{i-1} and γ^i .

Step 7. Definition of \tilde{h} on \mathcal{S}_0 .

We are now ready to define a function on \mathcal{S}_0 which extends g ; for simplicity, we start now with the definition of a “temptative” function \tilde{h} , without taking care of the injectivity. The definitive function h will be obtained later.

Recall that we have selected several horizontal segments A^iB^i , $1 \leq i \leq k-1$, in the square \mathcal{S}_0 ; the square is then divided in $k-2$ “horizontal strips”, i.e. the regions between two consecutive horizontal segments, plus two triangles, the “top one” $A^{k-1}A^k B^{k-1}$ and the “bottom one” $A^1A^0B^1$.

We start defining the function \tilde{h} on the “1-dimensional skeleton”, that is, the union of $\partial\mathcal{S}_0$ with all the horizontal segments A^iB^i : more precisely, we set $\tilde{h} = g$ on the boundary $\partial\mathcal{S}_0$, while for every $1 \leq i \leq k-1$ we define \tilde{h} on the horizontal segment A^iB^i as the piecewise linear function, parametrized at constant speed, whose image is the path γ^i . Notice that, with this definition, \tilde{h} is continuous on the 1-skeleton.

To extend \tilde{h} to the whole \mathcal{S}_0 , we can then argue separately on each of the horizontal strips of \mathcal{S}_0 , as well as on the top and bottom triangle. First, let us consider the bottom triangle $A^1A^0B^1$: thanks to the construction of Step 2, we know that the path γ^1 is either the segment $\mathbf{A}^1\mathbf{B}^1$, or the union of the two segments $\mathbf{A}^1\mathbf{A}^0$ and $\mathbf{A}^0\mathbf{B}^1$. In the first case, we define \tilde{h} on the bottom triangle as the affine function extending the values on the boundary; in the second case, let P be the point of the segment A^1B^1 such that $\tilde{h}(P) = \mathbf{A}^0$, let us extend \tilde{h} as constantly \mathbf{A}^0 on the segment PA^0 , and let \tilde{h} be the (degenerate) affine function extending the values on the boundary on each of the two triangles A^1PA^0 and A^0PB^1 . In the top triangle, we give of course the very same definition of \tilde{h} .

Let us now consider the horizontal strip \mathcal{D}_i between $A^i B^i$ and $A^{i+1} B^{i+1}$, and let us call \mathcal{D}_i the bounded region in \mathcal{S}_0 having as boundary the closed curve $\gamma^{i+1} \cup B^{i+1} B^i \cup \gamma^i \cup A^i A^{i+1}$. In Step 6, we have selected a finite number of points on γ^i and on γ^{i+1} , and we have called “vertical segments” the corresponding segments. More precisely, let us denote the points in $A^{i+1} B^{i+1}$ as $P_0 = A^{i+1}, P_1, \dots, P_{M-1}, P_M = B^{i+1}$, and the points in $A^i B^i$ as $Q_0 = A^i, Q_1, \dots, Q_{M-1}, Q_M = B^i$; as always, let us write $\mathbf{P}_j = \tilde{h}(P_j)$, and $\mathbf{Q}_j = \tilde{h}(Q_j)$. Keep in mind that each segment $\mathbf{P}_j \mathbf{Q}_j$, whose interior is entirely contained in \mathcal{D}_i , has been called a “vertical segment”, and notice that the points \mathbf{P}_j and \mathbf{Q}_j are not necessarily all different: for instance, the point \mathbf{S} of Figure 3, right, is the point \mathbf{P}_j for three consecutive indices $0 < j < M$.

We are finally in position to give the definition of \tilde{h} on the interior of each strip \mathcal{D}_i (and, since \tilde{h} has been already defined in the 1-skeleton and on the top and bottom triangle, this will conclude the present step). The strip \mathcal{D}_i is the essentially disjoint union of the triangles $P_j P_{j+1} Q_j$ and $P_{j+1} Q_j Q_{j+1}$ for all $0 \leq j < M$, and \mathcal{D}_i is the essentially disjoint union of the corresponding triangles $\mathbf{P}_j \mathbf{P}_{j+1} \mathbf{Q}_j$ and $\mathbf{P}_{j+1} \mathbf{Q}_j \mathbf{Q}_{j+1}$, where the triangles in \mathcal{D}_i (but not those in \mathcal{D}_i) can be degenerate, in particular they are degenerate for the points in $\gamma_2^{i+1} = \gamma_2^i$. We define then \tilde{h} on \mathcal{D}_i as the function which is affine on each of the above-mentioned triangles. Notice that, by construction, \tilde{h} is linear on each side $P_j P_{j+1}$ and $Q_j Q_{j+1}$, hence this definition on \mathcal{D}_i is a continuous extension of the definition on the 1-skeleton.

Step 8. Estimate for $\int_{A^0 A^1 B^1} |D\tilde{h}|$.

In this and in the following step, we aim to estimate the integral of $|D\tilde{h}|$ on \mathcal{S}_0 ; in particular, in this step we will consider the bottom triangle $A^0 A^1 B^1$ (by symmetry, we will get an estimate valid also for the top triangle, of course), while in the next step we will consider the situation of the horizontal strips \mathcal{D}_i . The aim of this step is to show the validity of the bound

$$\int_{A^0 A^1 B^1} |D\tilde{h}| \leq K \int_{\partial \mathcal{S}_0} |Dg| d\mathcal{H}^1, \quad (2.5)$$

where as usual K denotes a purely geometric constant. By simplicity, let us call

$$\bar{r} := \mathcal{H}^1(A^0 A^1) = \mathcal{H}^1(A^0 B^1).$$

Recall that, on the bottom triangle, the function \tilde{h} has been defined as an affine function, if the angle $\widehat{A^1 A^0 B^1}$, pointing inside \mathcal{S}_0 , is smaller than π –or, equivalently, if the curve γ^1 coincides with the segment $A^1 B^1$ – and as two degenerate affine functions on the two triangles $A^1 P A^0$ and $A^0 P B^1$ (being P as in Step 7) otherwise. Let us then estimate the L^1 norm of $D\tilde{h}$ on the bottom triangle in both cases.

First of all, consider the non-degenerate case when \tilde{h} is a single affine function on the bottom triangle. In particular, the image of the segment $A^0 A^1$ is the segment $A^0 A^1$, while the image of the segment $A^0 B^1$ is the segment $A^0 B^1$; this implies that, on the bottom triangle, one has

$$\frac{\sqrt{2}}{2} |D_1^b \tilde{h} + D_2^b \tilde{h}| = \frac{\mathcal{H}^1(A^0 B^1)}{\mathcal{H}^1(A^0 B^1)}, \quad \frac{\sqrt{2}}{2} |-D_1^b \tilde{h} + D_2^b \tilde{h}| = \frac{\mathcal{H}^1(A^0 A^1)}{\mathcal{H}^1(A^0 A^1)},$$

where by $D_1^b \tilde{h}$ and $D_2^b \tilde{h}$ we denote the constant value of $D_1 \tilde{h}$ and $D_2 \tilde{h}$ on the bottom triangle. This readily implies

$$|D^b \tilde{h}| \leq \frac{\mathcal{H}^1(\mathbf{A}^0 \mathbf{B}^1)}{\mathcal{H}^1(A^0 B^1)} + \frac{\mathcal{H}^1(\mathbf{A}^0 \mathbf{A}^1)}{\mathcal{H}^1(A^0 A^1)} = \frac{\mathcal{H}^1(\mathbf{A}^0 \mathbf{B}^1) + \mathcal{H}^1(\mathbf{A}^0 \mathbf{A}^1)}{\bar{r}}. \quad (2.6)$$

On the other hand,

$$\mathcal{H}^1(\mathbf{A}^0 \mathbf{B}^1) = \int_{A^0}^{B^1} |Dg| d\mathcal{H}^1, \quad \mathcal{H}^1(\mathbf{A}^0 \mathbf{A}^1) = \int_{A^0}^{A^1} |Dg| d\mathcal{H}^1,$$

which inserted in (2.6), and using (2.2) from Step 1, gives

$$|D^b \tilde{h}| \leq \frac{1}{\bar{r}} \int_{\mathcal{B}(V_1, \bar{r}) \cap \partial \mathcal{S}_0} |Dg| d\mathcal{H}^1 \leq K \int_{\partial \mathcal{S}_0} |Dg| d\mathcal{H}^1.$$

Hence, we deduce that

$$\int_{A^0 A^1 B^1} |D\tilde{h}| = \frac{\bar{r}^2}{2} |D^b \tilde{h}| \leq \frac{K \bar{r}^2}{2} \int_{\partial \mathcal{S}_0} |Dg| d\mathcal{H}^1 \leq K \int_{\partial \mathcal{S}_0} |Dg| d\mathcal{H}^1,$$

thus the validity of (2.5) follows.

Let us now consider the degenerate case, where in the bottom triangle the function \tilde{h} is made by two degenerate affine pieces, one on the left triangle $A^1 P A^0$ and the other on the right triangle $A^0 P B^1$; we call $D^l \tilde{h}$ and $D^r \tilde{h}$ the constant values of $D\tilde{h}$ respectively on the left and on the right triangle. Since the image of the segment $A^1 B^1$ through the map \tilde{h} is the path γ^1 (that is, the union of the two segments $\mathbf{A}^1 \mathbf{A}^0$ and $\mathbf{A}^0 \mathbf{B}^1$), parametrized at constant speed, we get that $|D_1^l \tilde{h}| = |D_1^r \tilde{h}|$ (while in general $D_1^l \tilde{h} \neq D_1^r \tilde{h}$); more precisely,

$$|D_1^l \tilde{h}| = |D_1^r \tilde{h}| = \frac{\mathcal{H}^1(\mathbf{A}^1 \mathbf{A}^0) + \mathcal{H}^1(\mathbf{A}^0 \mathbf{B}^1)}{\mathcal{H}^1(A^1 B^1)}. \quad (2.7)$$

Moreover, the affine map in the right triangle transforms the segment $A^0 B^1$ in the segment $\mathbf{A}^0 \mathbf{B}^1$, while the affine map in the left triangle moves $A^0 A^1$ onto $\mathbf{A}^0 \mathbf{A}^1$; this implies

$$\frac{\sqrt{2}}{2} |D_1^r \tilde{h} + D_2^r \tilde{h}| = \frac{\mathcal{H}^1(\mathbf{A}^0 \mathbf{B}^1)}{\mathcal{H}^1(A^0 B^1)}, \quad \frac{\sqrt{2}}{2} |-D_1^l \tilde{h} + D_2^l \tilde{h}| = \frac{\mathcal{H}^1(\mathbf{A}^0 \mathbf{A}^1)}{\mathcal{H}^1(A^0 A^1)},$$

which together with (2.7) gives

$$|D^l \tilde{h}| \leq \frac{3}{\bar{r}} \int_{\mathcal{B}(V_1, \bar{r}) \cap \partial \mathcal{S}_0} |Dg| d\mathcal{H}^1, \quad |D^r \tilde{h}| \leq \frac{3}{\bar{r}} \int_{\mathcal{B}(V_1, \bar{r}) \cap \partial \mathcal{S}_0} |Dg| d\mathcal{H}^1.$$

Arguing exactly as before, again thanks to (2.2) of Step 1, we obtain again the validity of (2.5), possibly with a slightly larger, but still purely geometric, constant K .

Step 9. Estimate for $\int_{\mathcal{D}_i} |D\tilde{h}|$.

In this step, we want again to find an estimate for the integral of $|D\tilde{h}|$, but this time on the generic horizontal strip \mathcal{D}_i , $1 \leq i \leq k-2$. Our goal is to obtain the estimate

$$\int_{\mathcal{D}_i} |D\tilde{h}| \leq K |\mathcal{D}_i| \int_{\partial \mathcal{S}_0} |Dg| d\mathcal{H}^1 + K \int_{A^i A^{i+1} \cup B^i B^{i+1}} |Dg| d\mathcal{H}^1, \quad (2.8)$$

where by $|\mathcal{D}_i|$ we denote the area of the horizontal strip \mathcal{D}_i . Consider the horizontal segment $A^{i+1} B^{i+1}$: by symmetry, it is not restrictive to assume that this segment is below the x -axis, precisely at a distance $0 < r \leq 1$ from the ‘‘south pole’’ $V_1 \equiv (0, -1)$; in other

words, we have that $A^{i+1} \equiv (-r, r-1)$ and $B^{i+1} \equiv (r, r-1)$. Moreover, let us call σ the distance between the segment $A^{i+1}B^{i+1}$ and the segment A^iB^i , and

$$\ell := \max \left\{ \mathcal{H}^1(\mathbf{A}^i \mathbf{A}^{i+1}), \mathcal{H}^1(\mathbf{B}^i \mathbf{B}^{i+1}) \right\} \leq \int_{A^i A^{i+1} \cup B^i B^{i+1}} |Dg| d\mathcal{H}^1. \quad (2.9)$$

Remember now that, in Step 7, we defined \tilde{h} as the function which is affine on each of the triangles $P_j P_{j+1} Q_j$, and $P_{j+1} Q_j Q_{j+1}$, sending each of the points point P_m (resp., Q_m) in \mathcal{S}_0 onto \mathbf{P}_m (resp., \mathbf{Q}_m) in \mathcal{S}_0 . Let us then concentrate ourselves on the generic triangle $\mathcal{T} = P_j P_{j+1} Q_j$ (for the triangles of the form $P_{j+1} Q_j Q_{j+1}$ the very same argument will work); since \tilde{h} is affine on \mathcal{T} , let us for simplicity denote by $D^\tau \tilde{h}$ the constant value of $D\tilde{h}$ on \mathcal{T} .

First of all let us recall that, on the segment $A^{i+1}B^{i+1}$, the function \tilde{h} has been defined as the piecewise linear function whose image is γ^{i+1} , parametrized at constant speed; this ensures that

$$|D_1^\tau \tilde{h}| = \frac{\mathcal{H}^1(\gamma^{i+1})}{\mathcal{H}^1(A^{i+1}B^{i+1})}. \quad (2.10)$$

We observe now that, by definition, γ^{i+1} is the shortest path in the closure of \mathcal{S}_0 connecting \mathbf{A}^{i+1} with \mathbf{B}^{i+1} ; in particular, γ^{i+1} is shorter than the image, through g , of the curve connecting A^{i+1} with B^{i+1} on $\partial\mathcal{S}_0$ passing through the south pole. This implies in particular that

$$\mathcal{H}^1(\gamma^{i+1}) \leq \int_{\mathcal{B}(V_1, \sqrt{2}r) \cap \partial\mathcal{S}_0} |Dg| d\mathcal{H}^1,$$

which inserted in (2.10) and recalling (2.2) gives

$$|D_1^\tau \tilde{h}| \leq \frac{1}{2r} \int_{\mathcal{B}(V_1, \sqrt{2}r) \cap \partial\mathcal{S}_0} |Dg| d\mathcal{H}^1 \leq K \int_{\partial\mathcal{S}_0} |Dg| d\mathcal{H}^1. \quad (2.11)$$

Let us now use the fact that the affine map \tilde{h} on \mathcal{T} sends the segment $P_j Q_j$ onto the segment $\mathbf{P}_j \mathbf{Q}_j$. Calling, as in Figure 4, d and d' the distances between A^{i+1} and P_j , and between A^i and Q_j , we derive that

$$\left| (d' - d + \sigma) D_1^\tau \tilde{h} + \sigma D_2^\tau \tilde{h} \right| = \mathcal{H}^1(\mathbf{P}_j \mathbf{Q}_j) \leq \ell, \quad (2.12)$$

where in the last equality we have used the estimate (2.4) from Step 6, which is valid since $\mathbf{P}_j \mathbf{Q}_j$ is a vertical segment in the sense of Step 6.

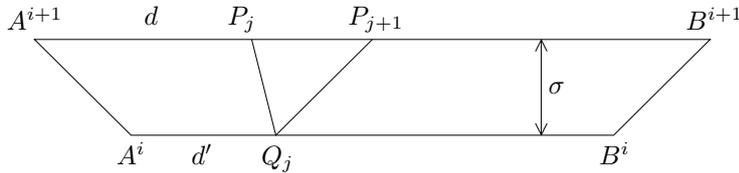


FIGURE 4. Position of points and lengths in Step 9.

Let us now use once again the fact that γ^{i+1} is the shortest path between \mathbf{A}^{i+1} and \mathbf{B}^{i+1} on the closure of \mathcal{S}_0 : in particular, γ^{i+1} is shorter than the path obtained as the

union of $\mathbf{A}^{i+1}\mathbf{A}^i$, the part of γ^i between \mathbf{A}^i and \mathbf{Q}_j , the segment $\mathbf{Q}_j\mathbf{P}_j$, and the part of γ^{i+1} between \mathbf{P}_j and \mathbf{B}^{i+1} ; namely,

$$\begin{aligned} \mathcal{H}^1(\gamma^{i+1}) &\leq \mathcal{H}^1(\mathbf{A}^{i+1}\mathbf{A}^i) + \frac{d'\mathcal{H}^1(\gamma^i)}{\mathcal{H}^1(\mathbf{A}^i\mathbf{B}^i)} + \mathcal{H}^1(\mathbf{Q}_j\mathbf{P}_j) + \mathcal{H}^1(\gamma^{i+1})\left(1 - \frac{d}{\mathcal{H}^1(\mathbf{A}^{i+1}\mathbf{B}^{i+1})}\right) \\ &\leq 2\ell + \frac{d'\mathcal{H}^1(\gamma^i)}{\mathcal{H}^1(\mathbf{A}^i\mathbf{B}^i)} + \mathcal{H}^1(\gamma^{i+1})\left(1 - \frac{d}{\mathcal{H}^1(\mathbf{A}^{i+1}\mathbf{B}^{i+1})}\right), \end{aligned}$$

which implies

$$d \frac{\mathcal{H}^1(\gamma^{i+1})}{\mathcal{H}^1(\mathbf{A}^{i+1}\mathbf{B}^{i+1})} - d' \frac{\mathcal{H}^1(\gamma^i)}{\mathcal{H}^1(\mathbf{A}^i\mathbf{B}^i)} \leq 2\ell.$$

The completely symmetric argument, using that γ^i is the shortest path between \mathbf{A}^i and \mathbf{B}^i , thus shorter than the union of $\mathbf{A}^i\mathbf{A}^{i+1}$, the part of γ^{i+1} between \mathbf{A}^{i+1} and \mathbf{P}_j , the segment $\mathbf{P}_j\mathbf{Q}_j$, and the part of γ^i between \mathbf{Q}_j and \mathbf{B}^i , gives the opposite inequality, hence we get

$$\left| d \frac{\mathcal{H}^1(\gamma^{i+1})}{\mathcal{H}^1(\mathbf{A}^{i+1}\mathbf{B}^{i+1})} - d' \frac{\mathcal{H}^1(\gamma^i)}{\mathcal{H}^1(\mathbf{A}^i\mathbf{B}^i)} \right| \leq 2\ell,$$

which further implies

$$\begin{aligned} |d - d'| \frac{\mathcal{H}^1(\gamma^{i+1})}{\mathcal{H}^1(\mathbf{A}^{i+1}\mathbf{B}^{i+1})} &\leq 2\ell + d' \left| \frac{\mathcal{H}^1(\gamma^{i+1})}{\mathcal{H}^1(\mathbf{A}^{i+1}\mathbf{B}^{i+1})} - \frac{\mathcal{H}^1(\gamma^i)}{\mathcal{H}^1(\mathbf{A}^i\mathbf{B}^i)} \right| \\ &\leq 2\ell + \left| \frac{\mathcal{H}^1(\mathbf{A}^i\mathbf{B}^i)}{\mathcal{H}^1(\mathbf{A}^{i+1}\mathbf{B}^{i+1})} \mathcal{H}^1(\gamma^{i+1}) - \mathcal{H}^1(\gamma^i) \right| \\ &\leq 2\ell + \left| \mathcal{H}^1(\gamma^{i+1}) - \mathcal{H}^1(\gamma^i) \right| + 2\sigma \frac{\mathcal{H}^1(\gamma^{i+1})}{\mathcal{H}^1(\mathbf{A}^{i+1}\mathbf{B}^{i+1})} \leq 4\ell + 2\sigma \frac{\mathcal{H}^1(\gamma^{i+1})}{\mathcal{H}^1(\mathbf{A}^{i+1}\mathbf{B}^{i+1})}. \end{aligned}$$

Using now (2.10), we can rewrite the above estimate as

$$|d - d'| |D_1^\tau \tilde{h}| \leq 4\ell + 2\sigma |D_1^\tau \tilde{h}|,$$

which recalling also (2.12) finally gives

$$\sigma |D_2^\tau \tilde{h}| \leq 5\ell + 3\sigma |D_1^\tau \tilde{h}|.$$

We can then easily evaluate the integral of $|D\tilde{h}|$ on \mathcal{T} , also by (2.11), as

$$\int_{\mathcal{T}} |D\tilde{h}| = \int_{\mathcal{T}} |D^\tau \tilde{h}| \leq \int_{\mathcal{T}} |D_1^\tau \tilde{h}| + |D_2^\tau \tilde{h}| \leq |\mathcal{T}| \left(4K \int_{\partial\mathcal{S}_0} |Dg| d\mathcal{H}^1 + 5 \frac{\ell}{\sigma} \right).$$

Adding now the above estimate over all the triangles \mathcal{T} forming \mathcal{D}_i , and recalling (2.9) and the fact that $|\mathcal{D}_i| \leq \sigma$, we directly obtain (2.8).

Step 10. Definition of the modified function h and conclusion of the proof.

We start observing that, adding the estimates (2.5) for the top and for the bottom triangle together with the estimates (2.8) for all the horizontal strips, we directly obtain the validity of (2.1) for the function \tilde{h} . However, the proof is still not over, because \tilde{h} satisfies (2.1), coincides with g on $\partial\mathcal{S}_0$, and it is finitely piecewise affine, but it is not a homeomorphism (unless all the paths γ^i lie in the interior of \mathcal{S}_0). However, we can easily obtain this with a simple modification of \tilde{h} : more precisely, let us slightly modify all the paths γ^i , so that they remain piecewise linear but they live in the interior of \mathcal{S}_0 and they do not intersect

each other. The idea, depicted in Figure 5, is obvious. Notice that, since there are only

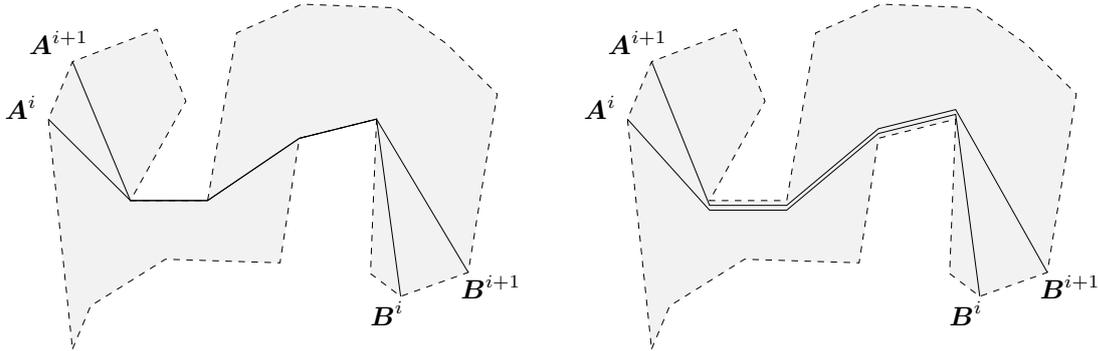


FIGURE 5. Modification of the paths γ^i in Step 10.

finitely many paths γ^i , and each of them has only finitely many vertices, it is clear that we can “separate” as desired all the paths, and we can also move each of them of a distance which is arbitrarily smaller than all the other distances between extreme points. Then, we define the function h exactly in the same way as we defined \tilde{h} , except that we use not the original paths γ^i but the modified ones; therefore, the function h is now not only finitely piecewise affine and coinciding with g on $\partial\mathcal{S}_0$, but it is also a homeomorphism. Moreover, the estimate (2.1) is still valid, with a geometric constant K which is as close as we wish to the one found above. The proof of the theorem is then now concluded. \square

Remark 2.2. *A trivial rotation and dilation argument proves the following generalization of Theorem 2.1. If \mathcal{S} is a square of side $2r$ and $g : \partial\mathcal{S} \rightarrow \mathbb{R}^2$ is a piecewise linear and one-to-one function, there exists a piecewise affine extension $h : \mathcal{S} \rightarrow \mathbb{R}^2$ of g such that*

$$\int_{\mathcal{S}} |Dh(x)| dx \leq Kr \int_{\partial\mathcal{S}} |Dg(t)| d\mathcal{H}^1(t). \quad (2.13)$$

3. EXTENSION IN THE DEGENERATE CASE $|Df(c)| = 0$ BUT $J_f(c) \neq 0$

As already explained in the description of Section 1.1, a crucial difficulty in our proof will be the case when a square \mathcal{S} is “good” (this means that Df is almost constantly equal to some matrix M within \mathcal{S}), but $\det M = 0$, while being $M \neq 0$. It will be important to handle this case with care, because the map f on \mathcal{S} is then very close to an affine map, but this affine map is degenerate. The goal of this section is to prove a single result, which will solve this difficulty. Recall that, whenever a map g is defined on $\partial\mathcal{S}$, for any $t \in \partial\mathcal{S}$ we denote by $\tau(t)$ the tangent vector to $\partial\mathcal{S}$ at t , by $Dg(t)$ the derivative of g in the direction $\tau(t)$ (whenever it exists), and by $\int_{\partial\mathcal{S}} |Dg| d\mathcal{H}^1$ the length of the curve $g(\partial\mathcal{S})$.

Theorem 3.1. *Let \mathcal{S} be a square of unit side and $g : \partial\mathcal{S} \rightarrow \mathbb{R}^2$ a piecewise linear and one-to-one function such that*

$$\int_{\partial\mathcal{S}} \left| Dg(t) - \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \cdot \tau(t) \right| d\mathcal{H}^1(t) < \delta \quad (3.1)$$

for $\delta \leq \delta_{\text{MAX}}$, where $\delta_{\text{MAX}} \ll 1$ is a geometric quantity. Then there is a finitely piecewise affine homeomorphism $h : \mathcal{S} \rightarrow \mathbb{R}^2$ such that $h = g$ on $\partial\mathcal{S}$ and

$$\int_{\mathcal{S}} \left| Dh(x) - \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \right| dx \leq K\delta. \quad (3.2)$$

Proof. We divide this proof in several steps, to make it as clear as possible.

Step 1. Definition of good and bad intervals.

First of all we notice that, since g is a one-to-one piecewise linear function, then its image $g(\partial\mathcal{S})$ is the boundary of a nondegenerate polygon, that we call \mathcal{S} . Moreover, thanks to (3.1), we know that this polygon is very close to a horizontal segment in \mathbb{R}^2 . Up to a translation, we can assume that the first coordinates of the points in \mathcal{S} are between 0 and L .

Fix now any $0 < \sigma < L$: it is reasonable to expect that there are exactly two points in $g(\partial\mathcal{S})$ having first coordinate σ , and that the two counterimages in $\partial\mathcal{S}$ are more or less one above the other (this means, with the same first coordinate). In the situation of Figure 6, this happens with σ , but not with σ' , since the points in $\partial\mathcal{S}$ with first coordinate σ' are four. We define then “good” any $\sigma \in (0, L)$ with the property that exactly two points in $\partial\mathcal{S}$ have first coordinate σ , and with the additional requirement that $\sigma > 2\delta$ and $\sigma < L - 2\delta$. For any such σ , we call \mathbf{P}_σ and \mathbf{Q}_σ the two above-mentioned points,

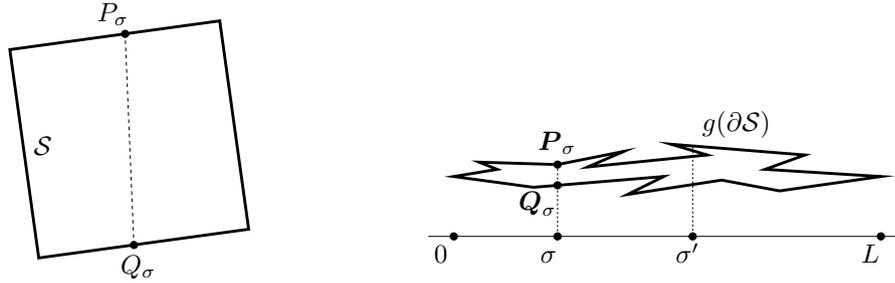


FIGURE 6. A good σ and a bad σ' in Step 1.

being \mathbf{P}_σ above \mathbf{Q}_σ , and we call $P_\sigma = g^{-1}(\mathbf{P}_\sigma)$ and $Q_\sigma = g^{-1}(\mathbf{Q}_\sigma)$. We can immediately show that a big percentage of the points are good, more precisely

$$\mathcal{H}^1\left(\left\{\sigma \in (0, L) : \sigma \text{ is not good}\right\}\right) \leq 5\delta. \quad (3.3)$$

Indeed, take any segment RS in $\partial\mathcal{S}$ on which g is linear, and call as usual $\mathbf{R} = g(R)$ and $\mathbf{S} = g(S)$: by definition, we have that

$$\begin{aligned} \int_{RS} \left| Dg(t) - \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \cdot \tau(t) \right| d\mathcal{H}^1(t) &\geq \left| \int_{RS} Dg(t) d\mathcal{H}^1(t) - \int_{RS} \begin{pmatrix} \tau_1(t) \\ 0 \end{pmatrix} d\mathcal{H}^1(t) \right| \\ &\geq \left| (\mathbf{S}_1 - \mathbf{R}_1) - (S_1 - R_1) \right|. \end{aligned} \quad (3.4)$$

As a consequence, if the segment \mathbf{RS} is going backward (that is, $\mathbf{S}_1 - \mathbf{R}_1$ and $S_1 - R_1$ have opposite sign), then its horizontal spread is bounded by the above integral on the interval RS . Recalling (3.1), this means that all the backward segments have a projection

on $(0, L)$ with total length less than δ . Since of course any $\sigma \in (2\delta, L - 2\delta)$ which is not good must belong to this projection, the validity of (3.3) follows.

By adding the inequality (3.4) for all the segments of $\partial\mathcal{S}$, we find also that L equals the horizontal width of \mathcal{S} up to an error $\delta/2$, thus in particular

$$1 - \frac{\delta}{2} \leq L \leq \sqrt{2} + \frac{\delta}{2}.$$

Moreover, take any good σ and consider all the segments on $\partial\mathcal{S}$ connecting P_σ and Q_σ : again adding (3.4) on all these segments, and recalling that \mathbf{P}_σ and \mathbf{Q}_σ have the same first projection, we derive that

$$|(P_\sigma)_1 - (Q_\sigma)_1| \leq \frac{\delta}{2}, \quad (3.5)$$

that is, the points P_σ and Q_σ are always exactly one above the other up to an error $\delta/2$: the factor $1/2$ comes by the possibility of choosing either of the two paths in \mathcal{S} from P_σ to Q_σ .

Finally, assume that P_σ and Q_σ lie on a same side of \mathcal{S} for some good σ . Adding once again (3.4) among all the segments where g is linear connecting P_σ and Q_σ , we derive that, up to an error δ , the sum of all the horizontal spreads of these segments coincides with the corresponding sum of the horizontal spreads in $\partial\mathcal{S}$; however, the first sum is smaller than $\delta/2$ by (3.5), while the second is at least the minimum between 2σ and $2(L - \sigma)$, which is impossible by the definition of good σ . In other words, we have proved that P_σ and Q_σ never lie on a same side of \mathcal{S} if σ is good.

Observe now that, since g is piecewise linear, then by definition $(0, L)$ is a finite union of intervals, alternately done entirely by bad σ and entirely by good ones. However, the endpoints of all these intervals are always bad. Therefore, we slightly shrink the intervals made by good points and we call *good intervals* these shrunked intervals. Thanks to (3.3), we can do this in such a way that the union of the good intervals covers the whole $(0, L)$ up to a length of 6δ : notice that all the points of any good interval are good points, also the endpoints, while the bad intervals may also contain good points. Finally, it is convenient to make the following further slight modification: up to replace a good interval with a finite union of good intervals, we can also assume that whenever (σ, σ') is a good interval, the map g is linear in the segments $P_\sigma P_{\sigma'}$ and $Q_\sigma Q_{\sigma'}$.

Step 2. The extension on the segments $P_\sigma Q_\sigma$, definition of good and bad quadrilaterals.

In this step, we extend g –which is defined on $\partial\mathcal{S}$ – to a union of segments in \mathcal{S} . More precisely, recall that $(0, L)$ has been divided in intervals, which can be either bad or good. Moreover, the extremes of these intervals are all good points except 0 and L . Take then any other of these extremes, say σ , and consider the points P_σ and Q_σ in $\partial\mathcal{S}$. We define g on the segment $P_\sigma Q_\sigma$ as the linear function such that $g(P_\sigma) = \mathbf{P}_\sigma$ and $g(Q_\sigma) = \mathbf{Q}_\sigma$; notice that all the different open segments $P_\sigma Q_\sigma$ are contained in the interior of \mathcal{S} by Step 1, and they do not intersect with each other by construction. We have then many segments $P_\sigma Q_\sigma$ inside \mathcal{S} , almost vertical by (3.5), on each of which a linear function g is defined. Observe that, as a consequence, \mathcal{S} has been divided in several quadrilaterals (actually, the first

and the last one are generally triangles), and g is defined in the whole corresponding 1-dimensional grid; also \mathcal{S} has then been subdivided by the images of g in a union of several polygons. A positive consequence of this fact is that we can now define the extension h of g in an independent way from each quadrilateral in \mathcal{S} to the corresponding polygon in \mathcal{S} —respecting of course the boundary data, and being a piecewise affine homeomorphism: then, the resulting function h will automatically be a piecewise affine homeomorphism.

Let us conclude this short step with another piece of notation: any quadrilateral in \mathcal{S} will be called a *good quadrilateral* if it corresponds to a good interval in $(0, L)$, and a *bad quadrilateral* otherwise. In the remaining of the proof, we will first give an estimate for the good quadrilaterals; then, we will give one for the first and the last quadrilateral, that is, the one starting at 0 and the one ending at L : notice that these quadrilaterals are always bad by definition, and actually they are usually triangles. Finally, we will give an estimate for the “internal” bad quadrilaterals, which we obtain by considering two subcases.

Step 3. *The extension in good quadrilaterals.*

Let us first start by considering a good quadrilateral, corresponding to the good interval (σ, σ') in $(0, L)$; for brevity, we will write $P, P', \mathbf{P}, \mathbf{P}'$ in place of $P_\sigma, P_{\sigma'}, \mathbf{P}_\sigma, \mathbf{P}_{\sigma'}$. Recall that the map g is linear between P and P' , as well as between Q and Q' , thanks to the construction in Step 1. As a consequence, the image under g of the boundary of the quadrilateral $PP'Q'Q$ is the boundary of the quadrilateral $\mathbf{P}\mathbf{P}'\mathbf{Q}'\mathbf{Q}$, and we have to define the extension h by sending the interior of $PP'Q'Q$ onto the interior of $\mathbf{P}\mathbf{P}'\mathbf{Q}'\mathbf{Q}$. Let h simply be the piecewise affine map sending $PP'Q$ onto $\mathbf{P}\mathbf{P}'\mathbf{Q}$ and $QP'Q'$ onto $\mathbf{Q}\mathbf{P}'\mathbf{Q}'$.

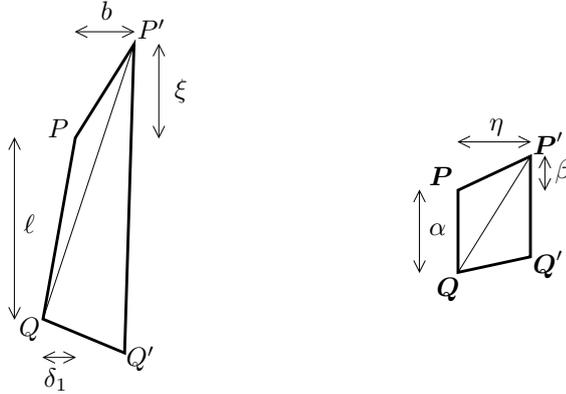


FIGURE 7. The approximation in a good quadrilateral, Step 3.

We need to estimate

$$\int_{PP'Q} \left| Dh - \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \right| dx, \quad (3.6)$$

the estimate in the triangle $QP'Q'$ being then of course identical. Let us define for shortness

$$\begin{aligned} \ell &= P_2 - Q_2, & b &= P'_1 - P_1, & \xi &= P'_2 - P_2, & \theta &= \arctan(\xi/b), \\ \delta_1 &= P_1 - Q_1, & \alpha &= \mathbf{P}_2 - \mathbf{Q}_2, & \eta &= \mathbf{P}'_1 - \mathbf{P}_1, & \beta &= \mathbf{P}'_2 - \mathbf{P}_2, \end{aligned}$$

we refer to Figure 7 for help with the notation. By definition, the constant value of Dh in $PP'Q$ satisfies

$$\begin{aligned} bD_1h_1 + \xi D_2h_1 &= \eta, & bD_1h_2 + \xi D_2h_2 &= \beta, \\ \delta_1 D_1h_1 + \ell D_2h_1 &= 0, & \delta_1 D_1h_2 + \ell D_2h_2 &= \alpha. \end{aligned} \quad (3.7)$$

Let us start by defining

$$\varepsilon = \int_{PP'} \left| Dg(t) - \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \cdot \tau(t) \right| d\mathcal{H}^1(t), \quad (3.8)$$

so that adding the values of ε on the different segments we will get less than δ by (3.1). We claim now the validity of the following estimates, all obtained again arguing as in (3.4):

$$|\eta - b| \leq \varepsilon, \quad |\beta| \leq \varepsilon, \quad \alpha \leq \delta, \quad |\delta_1| \leq \frac{\delta}{2}, \quad \ell > \delta \max\{\tan \theta, 1\}. \quad (3.9)$$

The first two estimates can be found just integrating on the segment PP' , so they are valid with the small constant ε ; instead, to get the third estimate we have to integrate on all the segments connecting P and Q on $\partial\mathcal{S}$, so we can only estimate with δ ; the fourth estimate is given by (3.5). Finally, the evaluation of ℓ follows by a simple geometric argument, just recalling that $\sigma > 2\delta$, (3.5) and that we have defined θ as the direction of the side containing PP' .

Let us now start by evaluating D_1h_1 : inserting the third equation of (3.7) into the first one, we get

$$D_1h_1 \left(b - \frac{\xi\delta_1}{\ell} \right) = \eta,$$

from which we readily obtain, by using the estimates (3.9) and recalling that $\xi/b = \tan \theta$,

$$|D_1h_1 - 1| \leq 2 \left(\frac{\varepsilon}{b} + \frac{\xi\delta}{b\ell} \right). \quad (3.10)$$

Substituting the value of D_1h_1 again in the third equation of (3.7), we get then

$$|D_2h_1| = \frac{\delta_1}{\ell} |D_1h_1| \leq \frac{\delta}{2\ell} + \frac{\varepsilon}{b} + \frac{\xi\delta}{b\ell}. \quad (3.11)$$

We control now the derivatives of h_2 : inserting the second equation of (3.7) into the fourth, we get

$$D_2h_2 \left(\ell - \delta_1 \frac{\xi}{b} \right) = \alpha - \frac{\delta_1\beta}{b}$$

so that, again using (3.9) and again recalling that $\xi/b = \tan \theta$, we deduce

$$|D_2h_2| \leq 2 \frac{\delta}{\ell} + \frac{\delta\varepsilon}{b\ell}, \quad |D_1h_2| \leq 2 \frac{\varepsilon}{b} + 2 \frac{\xi\delta}{b\ell}. \quad (3.12)$$

Estimating the integral in (3.6) is then straightforward. Notice that δ is a fixed constant, not depending on the subdivision in intervals: as a consequence, we can assume without loss of generality that $\xi \leq \delta < \ell$, otherwise it is enough to subdivide a good interval in a finite union of good intervals; the area of the triangle $PP'Q$ is then less than $b\ell$, and so from (3.10), (3.11) and (3.12) we obtain

$$\int_{PP'Q} \left| Dh - \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \right| dx \leq \left(5 \frac{\varepsilon}{b} + 5 \frac{\xi\delta}{b\ell} + 3 \frac{\delta}{\ell} + \frac{\delta\varepsilon}{b\ell} \right) \cdot b\ell \leq 8\varepsilon + \delta(5\xi + 3b),$$

where we have also used that $\delta \leq 1/2$ and $\ell \leq 3/2$ (the latter follows by straightforward geometrical arguments). Of course, the fully analogous estimate holds for the integral in the triangle $QP'Q'$, up to replace the segment PP' by QQ' in the definition (3.8) of ε .

To conclude, we need to evaluate the total integral in the union of the good quadrilaterals; this is simply achieved by summing the above estimates over all the different quadrilaterals. Notice that the constant δ is fixed and does not depend on the quadrilateral, while the constants ε , b and ξ are specific of each quadrilateral. By definition (3.8) of ε , it appears clear that the sum of all the different ε 's is less than δ , while by definition of the lengths in the square it is clear that the sum of the different ξ , as well as of the different b , is bounded by 4. As a consequence, we deduce that

$$\int_G \left| Dh(x) - \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \right| dx \leq K\delta, \quad (3.13)$$

where G denotes the union of all the good quadrilaterals in \mathcal{S} , while K is as usual a purely geometric constant.

Step 4. The extension in the first and last bad quadrilaterals.

In this and in the next step we are going to consider the bad quadrilaterals. Notice that, since almost the whole square is done by good quadrilaterals thanks to (3.3), we can even be satisfied by a rough estimate here, while we needed a precise one in the preceding step: what is important, is that we can define a piecewise affine homeomorphism h on each of the bad quadrilaterals.

Here we begin with the “first” and the “last” quadrilateral, that is, with the quadrilaterals which correspond to the two intervals having 0 or L as one endpoint. Notice that these “quadrilaterals” are actually triangles, unless some side of the square is very close to being vertical. More precisely, let us consider just the first bad quadrilateral \mathcal{C} , by symmetry: as Figure 8 depicts, it can be either a triangle VPQ , being V the left vertex of the square, or a quadrilateral $VV'PQ$, if V and V' are the two left vertices of the square, being the side VV' almost vertical, and then the sides $V'P$ and VQ almost horizontal. Notice that all the sides of \mathcal{C} belong to $\partial\mathcal{S}$ except PQ . We need to send \mathcal{C} on the polygon \mathcal{C} inside \mathcal{S} made by the points which have first coordinate less than $\sigma = \mathbf{P}_1 = \mathbf{Q}_1$, shaded in the right of the figure. Keep in mind that by construction (recall Step 1) the coordinate σ is good, and the bad intervals cover only a portion less than 2δ of $(2\delta, L - 2\delta)$: this means that $2\delta \leq \sigma \leq 4\delta$. As a consequence, again by using several times (3.4) and (3.5),

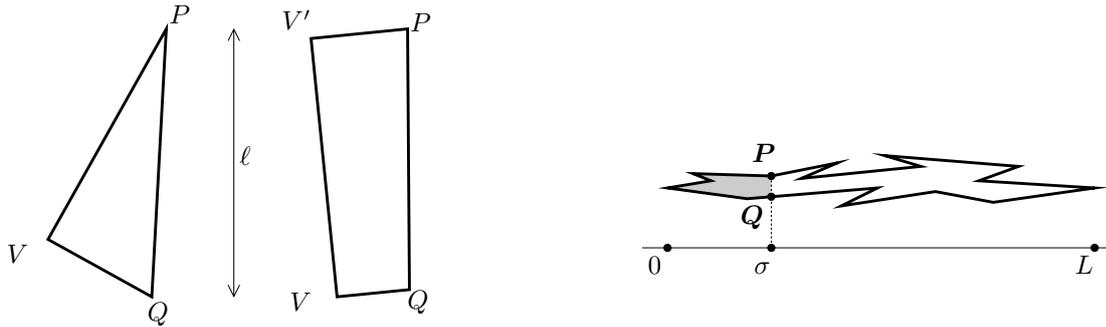


FIGURE 8. The approximation in the first bad quadrilateral, Step 4.

we know that

$$\mathbf{V}_1 \leq \frac{\delta}{2}, \quad \frac{\delta}{2} \leq P_1 - V_1 \leq 5\delta, \quad \frac{\delta}{2} \leq Q_1 - V_1 \leq 5\delta, \quad |Q_1 - P_1| \leq \frac{\delta}{2},$$

and the estimates on V are valid also for V' in the case when the bad quadrilateral \mathcal{C} is actually a quadrilateral. We would like to infer that \mathcal{C} is the biLipschitz image of a square with side δ , with uniformly bounded biLipschitz constant; however, this is true only if ℓ is comparable to δ , while we only know by Step 3 that $\ell \geq \delta$ —this was established in (3.9). Let us then consider the affine map $\Phi(x_1, x_2) = (x_1, x_2\delta/\ell)$, and let $\tilde{\mathcal{C}} = \Phi(\mathcal{C})$; define also $\tilde{g} = g \circ \Phi^{-1}$ on $\partial\tilde{\mathcal{C}}$, which is admissible since g is defined on the whole $\partial\mathcal{C}$. By construction, the set $\tilde{\mathcal{C}}$ is the biLipschitz image of a square of side δ , with biLipschitz constant less than a geometrical constant K . We can then apply Theorem 2.1 to the map \tilde{g} , and we find an extension \tilde{h} of \tilde{g} inside $\tilde{\mathcal{C}}$ such that (2.1) holds, that is,

$$\int_{\tilde{\mathcal{C}}} |D\tilde{h}(y)| dy \leq K\delta \int_{\partial\tilde{\mathcal{C}}} |D\tilde{g}(t)| d\mathcal{H}^1(t)$$

(notice that the multiplication by δ comes from the argument of Remark 2.2). Observe now that the integral in the right side of the above inequality is simply the perimeter of \mathcal{C} , which is less than $K\delta$ by (3.1) and again by (3.4). Thus, we infer

$$\int_{\tilde{\mathcal{C}}} |D\tilde{h}(y)| dy \leq K\delta^2. \quad (3.14)$$

Finally, we define $h = \tilde{h} \circ \Phi$ on \mathcal{C} : this is a piecewise affine homeomorphism from \mathcal{C} to \mathcal{C} , and by definition it extends the map h already defined on $\partial\mathcal{C}$. We have then to show that Dh is not too big on \mathcal{C} , and to do so it is enough to observe that

$$|Dh(\Phi^{-1}(y))| \leq |D\tilde{h}(y)|,$$

which by (3.14) finally implies

$$\int_{\mathcal{C}} |Dh(x)| dx \leq \int_{\tilde{\mathcal{C}}} |Dh(\Phi^{-1}(y))| \frac{\ell}{\delta} dy \leq \frac{2}{\delta} \int_{\tilde{\mathcal{C}}} |D\tilde{h}(y)| dy \leq K\delta. \quad (3.15)$$

We have then found the estimate we were looking for related to the first bad quadrilateral, and by symmetry the same holds also in the last bad quadrilateral.

Step 5. The extension in the internal bad quadrilaterals.

To conclude our analysis, we need to concentrate in the internal bad quadrilaterals. Let

\mathcal{C} be a bad quadrilateral, and let us call its vertices, as usual, P, Q, P' and Q' ; the image of $\partial\mathcal{C}$ under g is then the boundary of a polygon \mathcal{C} . Notice that \mathcal{C} needs not to be a quadrilateral, since it has two vertical sides, PQ and $P'Q'$, but $\widehat{PP'}$ and $\widehat{QQ'}$ are piecewise linear paths, not necessarily segments. Keeping a notation similar to that of Step 3, we set

$$\begin{aligned} \ell &= P_2 - Q_2, & b &= P'_1 - P_1, & \alpha &= \max\{P_2, P'_2\} - \min\{Q_2, Q'_2\}, \\ \xi &= P'_2 - P_2, & \theta &= \arctan(\xi/b), & \eta &= \mathcal{H}^1(\widehat{PP'}) + \mathcal{H}^1(\widehat{QQ'}), \end{aligned}$$

see Figures 9 and 10. By a simple symmetry argument, we can assume without loss of generality that

$$\theta \geq 0, \text{ and either } PP' \text{ and } QQ' \text{ are parallel, or } \theta \geq \pi/4. \quad (3.16)$$

Observe that this is possible because, if PP' and QQ' are not parallel, then they belong to two consecutive sides of the square, hence if $\theta \leq \pi/4$ we just have to exchange P with Q . Notice that, by definition,

$$\eta = \int_{PP' \cup QQ'} |Dg(t)| d\mathcal{H}^1(t). \quad (3.17)$$

We need now to further subdivide our analysis in two subcases, depending whether α is bigger or smaller than 10η . Notice that α is bounded by δ , while η could be even much smaller than δ , since the sum of all the different η 's corresponding to bad intervals is smaller than 3δ : indeed, the total length of the internal bad intervals is less than 2δ , so we do not even need to subtract the matrix $\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ as in (3.8). As a consequence, either of the two cases may actually hold.

Step 5a. The case $\alpha \leq 10\eta$.

Let us start with the case when $\alpha \leq 10\eta$. We let H be the point in the segment $P'Q'$ satisfying $P_2 = H_2$ (such a point exists by (3.16)), and $\mathbf{H} = g(H)$, which is well defined since g has been defined on the good segment $P'Q'$. We subdivide the quadrilateral \mathcal{C}

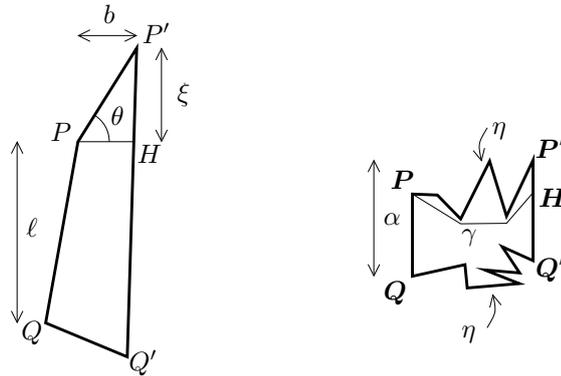


FIGURE 9. The approximation in an internal bad quadrilateral: case 1, Step 5a.

into the union of the triangle $PP'H$ and the quadrilateral $PHQ'Q$, and we aim to define the function h separately on these two pieces. First of all, similarly as in the proof of

Theorem 2.1, we consider the shortest path between \mathbf{P} and \mathbf{H} in \mathcal{C} , which is a piecewise affine path, possibly intersecting $\partial\mathcal{C}$ in other points than \mathbf{P} and \mathbf{H} , and we call γ a slight modification of this path, which is still piecewise affine, but which is entirely in the interior of \mathcal{C} except for the two extremes \mathbf{P} and \mathbf{H} . By minimality, we can of course take the modified γ satisfying

$$\mathcal{H}^1(\gamma) < \mathcal{H}^1(\widehat{\mathbf{PP}'}) + \mathcal{H}^1(\mathbf{P}'\mathbf{H}). \quad (3.18)$$

We extend then g to the segment PH as the piecewise affine function sending the segment PH onto the path γ at constant speed.

Let us now point our attention on the triangle $PP'H$: the segment PH is horizontal by definition, while the segment $P'H$ is “quite vertical”; more precisely, it is contained in the segment $P'Q'$ and by definition we have

$$|P'_1 - Q'_1| \leq \frac{\delta}{2}, \quad P'_2 - Q'_2 \geq \ell \geq \delta.$$

The triangle would then be a biLipschitz image of a square with side b , with uniformly bounded constant, if ξ were not too much bigger than b , or, in other words, if θ is not too big. Since we cannot be sure that this is the case, exactly as in Step 4 we define Φ the affine map which does not modify the horizontal segments, and which shrinks of a ratio ξ/b the segments parallel to $P'H$. Then, $\Phi(PP'H)$ is a triangle which is uniformly biLipschitz with a square of side b , so exactly as in Step 4 we apply Theorem 2.1 to the map $\tilde{g} = g \circ \Phi^{-1}$ finding an extension \tilde{h} on $\Phi(PP'H)$, and we finally obtain the extension to g in $PP'H$ as $h = \tilde{h} \circ \Phi$. Estimating the derivatives of h , \tilde{h} , g and \tilde{g} exactly as in Step 4, we get then the estimate

$$\begin{aligned} \int_{PP'H} |Dh(x)| dx &\leq K \frac{\xi}{b} \int_{\Phi(PP'H)} |D\tilde{h}(y)| dy \leq K\xi \int_{\partial(\Phi(PP'H))} |D\tilde{g}(t)| dt \\ &= K\xi \mathcal{H}^1(\partial(g(PP'H))) = K\xi \left(\mathcal{H}^1(\widehat{\mathbf{PP}'}) + \mathcal{H}^1(\gamma) + \mathcal{H}^1(\mathbf{P}'\mathbf{H}) \right) \\ &\leq K \left(\mathcal{H}^1(\widehat{\mathbf{PP}'}) + \mathcal{H}^1(\mathbf{P}'\mathbf{H}) \right) \leq K(\eta + \alpha) \leq K\eta, \end{aligned}$$

where we have also used (3.18) and the assumption $\alpha \leq 10\eta$.

Let us now consider the quadrilateral $PHQ'Q$. Since we have already seen that PQ and HQ' are “quite vertical”, while PH is exactly horizontal and QQ' is “quite horizontal” since it makes with the horizontal direction an angle equal to $\pi/2 - \theta \leq \pi/4$, this quadrilateral is uniformly biLipschitz with a rectangle. Up to shrinking vertically of a ratio b/ℓ as before, it becomes uniformly biLipschitz with a square of side b , hence by arguing as before by shrinking, applying Theorem 2.1 and then stretching back, we define an extension h of g inside the quadrilateral $PHQ'Q$ which satisfies

$$\int_{PHQ'Q} |Dh(x)| dx \leq K\ell \left(\mathcal{H}^1(\mathbf{PQ}) + \mathcal{H}^1(\mathbf{HQ}') + \mathcal{H}^1(\gamma) + \mathcal{H}^1(\widehat{\mathbf{QQ}'}) \right) \leq K\eta,$$

which put together with the estimate above for the triangle $PP'H$ gives

$$\int_{\mathcal{C}} |Dh(x)| dx \leq K\eta. \quad (3.19)$$

Step 5b. The case $\alpha > 10\eta$.

The last case that we have to consider is that of an internal bad quadrilateral corresponding to $\alpha > 10\eta$: in this case, if we argued as in Step 5a, then we would find the same estimates as in (3.19), but with $K\delta$ in place of $K\eta$; and in turn, this would not be acceptable, because adding all the different η 's for the bad quadrilaterals we get something smaller than δ , while adding a term δ in each of the bad quadrilaterals we could get any big constant in the end, since bad intervals could be many more than $1/\delta$.

As a consequence, in this substep we present a different definition of the extension h for the case $\alpha > 10\eta$. More precisely, take four points $\mathbf{H}_1, \mathbf{H}_2, \mathbf{H}_3, \mathbf{H}_4$, as in Figure 10, in

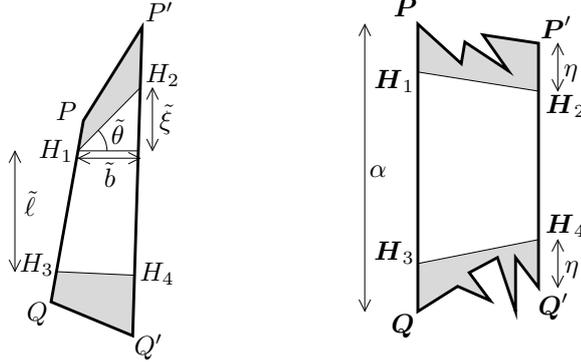


FIGURE 10. The approximation in an internal bad quadrilateral: case 2, Step 5b.

the segments PQ and $P'Q'$, at a distance η from the four vertices; let also $H_i = g^{-1}(\mathbf{H}_i)$ for $i = 1, 2, 3, 4$. By definition of η , the open segments H_1H_2 and H_3H_4 are entirely contained in the interior of \mathcal{C} , and by construction the same happens for the segments H_1H_2 and H_3H_4 in \mathcal{C} . We regard then both \mathcal{C} and \mathcal{C} as the union of three pieces: the internal quadrilaterals $H_1H_2H_4H_3$ and $\mathbf{H}_1\mathbf{H}_2\mathbf{H}_4\mathbf{H}_3$, and the “top” and “bottom” remaining parts, shaded in Figure 10. We aim to define the piecewise affine function h so to send each part of \mathcal{C} onto the corresponding one in \mathcal{C} .

For the “top” and “bottom” part, we can argue more or less exactly as in the last steps: each of the quadrilateral $PP'H_2H_1$ and $H_3H_4Q'Q$ can be transformed into a square, then one applies Theorem 2.1 and then goes back to the quadrilateral; since the perimeter of each of the shaded regions in \mathcal{C} is now at most 4η , the same estimates as in Step 5a can be repeated, so that similarly to (3.19) we get now

$$\int_{PP'H_2H_1 \cup H_3H_4Q'Q} |Dh(x)| dx \leq K\eta. \quad (3.20)$$

To conclude, we have to define the extension h so to send the internal quadrilateral of \mathcal{C} onto the internal quadrilateral of \mathcal{C} , and we will do that again by sending in an affine way the triangle $H_1H_2H_3$ (resp. $H_3H_2H_4$) onto the triangle $\mathbf{H}_1\mathbf{H}_2\mathbf{H}_3$ (resp. $\mathbf{H}_3\mathbf{H}_2\mathbf{H}_4$). We need thus only to check the value of $|Dh|$ on the triangle $H_1H_2H_3$, being the estimate for the triangle $H_3H_2H_4$ completely similar. To get the estimate, as in Figure 10 we set

$$H_1 - H_3 = (\tilde{\delta}_1, \tilde{\ell}), \quad H_2 - H_1 = (\tilde{b}, \tilde{\xi}), \quad \tilde{\theta} = \arctan \frac{\tilde{\xi}}{\tilde{b}};$$

notice that $\tilde{\delta}_1, \tilde{\ell}, \tilde{b}, \tilde{\xi}$ and $\tilde{\theta}$ are very similar to δ_1, ℓ, b, ξ and θ , since the assumption $\alpha > 10\eta$ implies that the points H_i are very close to the vertices of \mathcal{C} . Hence, the constant matrix Dh in $H_1H_2H_3$ satisfies

$$\begin{aligned} |Dh(\tilde{\delta}_1, \tilde{\ell})| &= |(\alpha - 2\eta, 0)| \leq \delta, \\ |Dh(\tilde{b}, \tilde{b} \tan \theta)| &= |Dh(\tilde{b}, \tilde{\xi})| = |\mathbf{H}_2 - \mathbf{H}_1| \leq \eta. \end{aligned} \quad (3.21)$$

As a consequence, we get first that

$$|Dh(\tilde{\delta}_1, \tilde{\delta}_1 \tan \theta)| \leq \frac{\tilde{\delta}_1}{\tilde{b}} \eta \leq \delta \frac{\eta}{b},$$

and then that

$$|Dh(0, \tilde{\ell} - \tilde{\delta}_1 \tan \theta)| \leq \delta \left(1 + \frac{\eta}{b}\right). \quad (3.22)$$

Recall now that the estimates (3.9) ensure

$$\frac{\tilde{\ell}}{\tilde{\delta}_1} = \frac{\ell}{\delta_1} \geq 2 \tan \theta \geq 2 \tan \tilde{\theta};$$

notice carefully that the estimates (3.9) were obtained in a good quadrilateral, so they are not valid now, but since the segment PQ corresponds to a good σ , and in particular it is in the boundary of the good quadrilateral immediately preceding \mathcal{C} , the estimates about ℓ and δ_1 are still valid and we can use them now. As a consequence, by (3.22) we get

$$|D_2h| \leq 2 \frac{\delta}{\tilde{\ell}} \left(1 + \frac{\eta}{b}\right) \leq 3 \frac{\delta}{\ell} + 3 \frac{\delta\eta}{b\ell}, \quad (3.23)$$

and substituting this in (3.21) we have also

$$|Dh(\tilde{b}, 0)| \leq \eta + \tilde{b} \tan \theta |D_2h| \leq \eta + 2\tilde{b} \tan \theta \frac{\delta}{\tilde{\ell}} \left(1 + \frac{\eta}{b}\right) \leq \eta + 3\tilde{b} \left(1 + \frac{\eta}{b}\right) \leq 4\eta + 3\tilde{b},$$

from which we derive

$$|D_1h| \leq 5 \frac{\eta}{b} + 3. \quad (3.24)$$

Since the area of the triangle $H_1H_2H_3$ is bounded by $b(\ell + b \tan \theta)$, by (3.23) and (3.24), and recalling again $\delta \tan \theta \leq \ell$, we get

$$\begin{aligned} \int_{H_1H_2H_3} |Dh(x)| dx &\leq (b\ell + b^2 \tan \theta) \left(3 \frac{\delta}{\ell} + 3 \frac{\delta\eta}{b\ell} + 5 \frac{\eta}{b} + 3\right) \\ &\leq 3\delta b + 3b^2 + 3\delta\eta + 3\eta b + 5\eta\ell + 5\eta\xi + 3b\ell + 3b\xi \leq K(b + \eta), \end{aligned}$$

so that repeating the same estimate in the triangle $H_3H_2H_4$, and adding (3.20), we obtain that in a bad quadrilateral \mathcal{C} where $\alpha > 10\eta$ the estimate

$$\int_{\mathcal{C}} |Dh(x)| dx \leq Kb + K\eta \quad (3.25)$$

holds.

Step 6. Conclusion.

We can now put together all the estimates of the last steps to conclude the proof. Let us start by considering the bad quadrilaterals, whose union is $\mathcal{S} \setminus G$, since in Step 3 we

have defined G as the union of the good quadrilaterals. Thanks to (3.19) and (3.25), we know that the integral of $|Dh|$ in any internal bad quadrilateral can always be estimated by $b + \eta$. If we add the different b 's corresponding to the bad quadrilaterals, up to an error δ we find the sum of the lengths of the internal bad intervals, which is at most 2δ by construction. On the other hand, adding the different η 's and recalling (3.17), we get something smaller than (twice) the sum of the lengths of the bad intervals. As a consequence, putting together the estimates for all the internal bad quadrilaterals, and also adding the estimate (3.15) for the first and the last bad quadrilateral, we obtain

$$\int_{\mathcal{S} \setminus G} |Dh(x)| dx \leq K\delta.$$

Since the total area of the bad quadrilaterals can be estimated by (twice) the total length of their horizontal projections, which in turn corresponds with the total length of the bad intervals up to an error δ , and so it is less than 7δ , we can now insert (3.13) to get

$$\begin{aligned} \int_{\mathcal{S}} \left| Dh(x) - \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \right| dx &= \int_G \left| Dh(x) - \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \right| dx + \int_{\mathcal{S} \setminus G} \left| Dh(x) - \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \right| dx \\ &\leq K\delta + \int_{\mathcal{S} \setminus G} |Dh(x)| dx + |\mathcal{S} \setminus G| \leq K\delta, \end{aligned}$$

which is (3.2), and then the proof is concluded. \square

Remark 3.2. *A trivial rotation and dilation argument proves the following generalization of Theorem 3.1: whenever \mathcal{S} is a square of side $2r$, $g : \partial\mathcal{S} \rightarrow \mathbb{R}^2$ is a piecewise linear and one-to-one function, and M is a matrix with $\det M = 0$, there is a finitely piecewise affine extension $h : \mathcal{S} \rightarrow \mathbb{R}^2$ of g such that*

$$\int_{\mathcal{S}} |Dh(x) - M| dx \leq Kr \int_{\partial\mathcal{S}} |Dg(t) - M \cdot \tau(t)| d\mathcal{H}^1(t),$$

as soon as

$$\int_{\partial\mathcal{S}} |Dg(t) - M \cdot \tau(t)| d\mathcal{H}^1(t) < r\delta_{\text{MAX}} \|M\|.$$

4. PROOF OF THEOREM 1.1

This last section is devoted to give the proof of Theorem 1.1. This proof is still quite involved, but the overall idea is simple after the preceding sections. As already explained in the introduction, the idea is to divide the whole Ω in squares, and then treat them in three different ways: roughly speaking, the “good” squares, where the function is very close to an affine map (and this group will be further divided in two subgroups), and the “bad” ones, where this is not true. Moreover, we will have to slightly change the value of f on the boundaries of all these squares, in order to become piecewise linear. Then, in the bad squares we will simply use Theorem 2.1 to get an extension, and the constant K in (2.1) will not be a problem because the bad squares will cover only a small portion of Ω . Instead, we have to perform a very precise approximation of f in the good squares; to do so, we will treat differently the squares where the affine map close to f has zero determinant, and those where the determinant is strictly positive. For the first ones, we

will use Theorem 3.1, while for the second ones it will be enough to interpolate the values of f on the boundary, as we show in Section 4.1.

Before starting with the proof, let us give a couple of definitions.

Definition 4.1. *We say that $\mathcal{S}(c, r)$ is a Lebesgue square with matrix $M \in \mathbb{R}^{2 \times 2}$ and constant $\delta > 0$ if $\mathcal{S}(c, 3r) \subseteq \Omega$ and*

$$\int_{\mathcal{S}(c, 3r)} |Df(z) - M| dz \leq \delta.$$

Definition 4.2. *Let $\mathcal{S}(x, r) \subseteq \Omega$ be a square, and denote by T_1 and T_2 the two triangles on which \mathcal{S} is subdivided by the diagonal connecting $(x_1 - r, x_2 + r)$ and $(x_1 + r, x_2 - r)$. We call $\varphi_{\mathcal{S}(x, r)}$ the piecewise affine function which is affine on the two triangles T_1 and T_2 , and which coincides with f on the four vertices of $\mathcal{S}(x, r)$.*

4.1. The Lebesgue squares. In this first subsection, we consider the situation in the best possible case, namely, a Lebesgue square. It is rather easy to show the following uniform estimate.

Lemma 4.3. *For every $\varepsilon > 0$ and every matrix M , there exists $\bar{\delta} = \bar{\delta}(M, \varepsilon) \ll \varepsilon$ such that the following holds: if $\mathcal{S}(c, r)$ is a Lebesgue square with matrix M and constant $\delta \leq \bar{\delta}$,*

$$\|f - \varphi\|_{L^\infty(\mathcal{S}(c, r))} \leq r\varepsilon, \quad \|Df - D\varphi\|_{L^1(\mathcal{S}(c, r))} \leq r^2\varepsilon, \quad \|f - \psi\|_{L^\infty(\mathcal{S}(c, 2r))} \leq \frac{r\varepsilon}{10}, \quad (4.1)$$

where $\varphi = \varphi_{\mathcal{S}(c, r)}$ is the piecewise affine map of Definition 4.2 and $\psi : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ is an affine function satisfying $D\psi = M$. If in addition $\det M > 0$, then φ is injective and

$$f(\mathcal{S}(c, (1 - \varepsilon)r)) \subseteq \varphi(\mathcal{S}(c, r)) \subseteq f(\mathcal{S}(c, (1 + \varepsilon)r)). \quad (4.2)$$

Proof. We assume for simplicity of notation that the point c is the origin of \mathbb{R}^2 , and we write $\mathcal{S}(r)$ instead of $\mathcal{S}(0, r)$. Let R be a big constant, depending only on M and ε , to be specified later, and let us define the two sets A and B as

$$A = \left\{ x \in (-3r, 3r) : \int_{-3r}^{3r} |Df(x, t) - M| dt \geq R\delta \right\},$$

$$B = \left\{ y \in (-3r, 3r) : \int_{-3r}^{3r} |Df(t, y) - M| dt \geq R\delta \right\}.$$

By definition of Lebesgue square, we immediately get that

$$|A| \leq \frac{6r}{R}, \quad |B| \leq \frac{6r}{R}. \quad (4.3)$$

Let us now arbitrarily fix a point $z = (\bar{x}, \bar{y}) \in \mathcal{S}(r)$ with $\bar{x} \notin A$, $\bar{y} \notin B$, and let us define $\psi : \mathcal{S}(c, 3r) \rightarrow \mathbb{R}^2$ as $\psi(w) = f(z) + M(w - z)$: it is clear that ψ is an affine map with $D\psi \equiv M$, and calling $g = f - \psi$ we have by definition $g(z) = 0$. We claim that

$$|g(w)| \leq 12rR\delta \quad \forall w = (x, y) \in \mathcal{S}(3r) \setminus (A \times B). \quad (4.4)$$

Indeed, assume that $x \notin A$ (if $y \notin B$ the obvious modification of the argument works). Recalling that $g(\bar{x}, \bar{y}) = 0$ and that $x \notin A$ and $\bar{y} \notin B$, we can evaluate

$$\begin{aligned} |g(w)| &= |g(x, y) - g(\bar{x}, \bar{y})| \leq |g(x, y) - g(x, \bar{y})| + |g(x, \bar{y}) - g(\bar{x}, \bar{y})| \\ &\leq \int_{\bar{y}}^y |Df(x, t) - M| dt + \int_{\bar{x}}^x |Df(t, \bar{y}) - M| dt \leq 12rR\delta, \end{aligned}$$

and (4.4) is proved.

Let now $w = (x, y) \in \mathcal{S}(2r)$ be a generic point. By (4.3), we can find $x_1 < x < x_2$ and $y_1 < y < y_2$ such that for $i = 1, 2$ we have $(x_i, y_i) \in \mathcal{S}(3r)$, $x_i \notin A$, $y_i \notin B$ and $x_2 - x_1 \leq 6r/R$, $y_2 - y_1 \leq 6r/R$. Hence, w is inside the small rectangle \mathcal{R} having sides with coordinate x_i and y_i , and by (4.4) we know that

$$|g(P)| \leq 12rR\delta \quad \forall P \in \partial\mathcal{R}. \quad (4.5)$$

By definition $\psi(\partial\mathcal{R})$ is a small parallelogram around $\psi(w)$, with

$$\psi(\partial\mathcal{R}) \subseteq \mathcal{B}\left(\psi(w), \frac{6r\sqrt{2}}{R} \|M\|\right),$$

having defined $\|M\| = \max |M(v)|/|v|$. This estimate, together with (4.5), ensures that the whole curve $f(\partial\mathcal{R})$ is done by points with distance less than $12rR\delta + 6r\sqrt{2}\|M\|/R$ from $\psi(w)$. Since f is a homeomorphism, the point $f(w)$ is inside this curve, hence we finally deduce

$$|f(w) - \psi(w)| = |g(w)| \leq 12rR\delta + \frac{6r\sqrt{2}}{R} \|M\| < r \frac{\varepsilon}{10}, \quad (4.6)$$

where the last inequality holds as soon as R has been chosen big enough, depending on M and on ε , and then $\bar{\delta}$ has been chosen small enough, depending on R and ε , and thus ultimately only on M and ε . Hence, we have obtained the third estimate in (4.1).

Let us now consider the function $\varphi = \varphi_{\mathcal{S}(r)}$, and let us call

$$V_1 = (c_1 - r, c_2 + r), \quad V_2 = (c_1 + r, c_2 + r), \quad V_3 = (c_1 + r, c_2 - r), \quad V_4 = (c_1 - r, c_2 - r)$$

the four vertices of the square $\mathcal{S}(r)$. By definition, $\varphi = f$ at every vertex V_i : then, in the triangle T_1 (and the same holds true in the triangle T_2) we have that the two affine functions φ and ψ satisfy $|\varphi - \psi| = |f - \psi|$ at every vertex, thus (4.6) gives

$$\|\varphi - \psi\|_{L^\infty(\mathcal{S}(r))} \leq \|f - \psi\|_{L^\infty(\mathcal{S}(r))} \leq \|f - \psi\|_{L^\infty(\mathcal{S}(2r))} \leq 12rR\delta + \frac{6r\sqrt{2}}{R} \|M\| < r \frac{\varepsilon}{10}, \quad (4.7)$$

and this, together with (4.6), implies the first estimate in (4.1).

Concerning the second one, let us call M_1 the constant value of $D\varphi$ in T_1 , and notice that by (4.7) we get

$$\left| (M_1 - M)(e_1) \right| = \left| \frac{(\varphi(V_2) - \varphi(V_1)) - (\psi(V_2) - \psi(V_1))}{2r} \right| < \frac{\varepsilon}{10},$$

and the same estimate holds true for $\left| (M_1 - M)(e_2) \right|$ simply by checking the vertices V_1 and V_4 . As a consequence, we have that the constant value of $|D\varphi - D\psi|$ in T_1 is less than $\varepsilon/5$, and the same holds true of course also in T_2 . In other words, $\|D\varphi - D\psi\|_{L^\infty(\mathcal{S}(r))} \leq \varepsilon/5$.

Since $D\psi$ is constantly equal to M in the square $\mathcal{S}(r)$, by the definition of Lebesgue square we get

$$\begin{aligned} \|Df - D\varphi\|_{L^1(\mathcal{S}(r))} &= \int_{\mathcal{S}(r)} |Df(z) - D\varphi(z)| dz \leq \int_{\mathcal{S}(r)} |Df - M| + \int_{\mathcal{S}(r)} |M - D\varphi| \\ &\leq 36r^2\delta + \frac{4}{5}r^2\varepsilon < r^2\varepsilon, \end{aligned}$$

where the last inequality is true up to possibly decrease the value of $\bar{\delta}$: this gives the second estimate in (4.1).

Let us now suppose that $\det M > 0$. As a consequence, the image of $\mathcal{S}(r)$ under ψ is a non-degenerate parallelogram, and then the image of $\mathcal{S}(r)$ under φ is the disjoint union of two non-degenerate triangles, as soon as $\|\psi - \varphi\|_{L^\infty(\mathcal{S}(r))}/r$ is small enough, depending on M ; moreover, also the validity of (4.2) is obvious as soon as $\|\psi - f\|_{L^\infty(\mathcal{S}(2r))}/r$ is small enough, depending on M and on ε . Since (4.6) is valid for every $w \in \mathcal{S}(2r)$, we get (4.2) up to further increase the value of R and decrease the value of $\bar{\delta}$, again depending only on M and on ε ; thus, the proof is concluded. \square

Remark 4.4. *Notice that, in the last estimate of the above proof, the final values of $1/R$ and of δ behave more or less as ε multiplied by $\min |M(v)|/|v|$, and the latter number is strictly positive exactly when $\det M > 0$. This clarifies the need of the assumption $\det M > 0$ in order to get (4.2). We can come to the same conclusion also directly by considering the claim of (4.2): we cannot hope it to be valid for the case when $\det M = 0$; indeed, it is true that f is as close as we wish to an affine function, but this affine function is degenerate, hence the image of a small square around c is close to a degenerate parallelogram, which is a segment (or even a point if $M = 0$). And of course, knowing that the four vertices of a small square are sent very close to the vertices of a parallelogram does not even imply that the piecewise affine function $\varphi_{\mathcal{S}(x,r)}$ is injective, if this parallelogram is degenerate!*

Remark 4.5. *A quick look at the proof of the above lemma ensures that the constant $\bar{\delta}(M, \varepsilon)$ actually depends only on ε , $\|M\|$, and $\det M$: more precisely, $\bar{\delta}(M, \varepsilon)$ is the minimum between a constant which continuously depend on ε , $\|M\|$, and $\det M$ for $\det M \geq 0$ (found in the first part of the proof), and another constant which also depends continuously on ε , $\|M\|$, and $\det M$, but in the range $\det M > 0$ (found at the end of the proof): this second constant explodes when $\det M \rightarrow 0$. As a consequence, $\bar{\delta}$ is bounded if ε is bounded from below, $\|M\|$ from above, and if $\det M$ is either 0 or it is bounded both from above and below (with a strictly positive constant).*

The crucial importance of the above general lemma comes from the fact that we can always apply it for small squares “almost centered” at Lebesgue points \bar{x} .

Lemma 4.6. *Let $\delta > 0$ be given and let \bar{x} be a Lebesgue point for Df . Then, there exists $\bar{r} = \bar{r}(\bar{x}, \delta)$ such that, for any $r < \bar{r}$ and any $x \in \mathcal{S}(\bar{x}, r/2)$, the square $\mathcal{S}(x, r)$ is a Lebesgue square with matrix $M = Df(\bar{x})$ and constant δ .*

Proof. Let $\bar{x} \in \Omega$, $M \in \mathbb{R}^{2 \times 2}$ and $\delta > 0$ be as in the claim. Since \bar{x} is a Lebesgue point, there exists $\bar{r} = \bar{r}(\bar{x}, \delta)$ such that, for any $r < \bar{r}$, one has $\mathcal{B}(\bar{x}, 5r) \subseteq \Omega$ and

$$\int_{\mathcal{B}(\bar{x}, 5r)} |Df(z) - M| dz \leq \frac{\delta}{3}. \quad (4.8)$$

Let now $x \in \mathcal{S}(\bar{x}, r/2)$. We have $\mathcal{S}(x, 3r) \subseteq \mathcal{B}(\bar{x}, 5r) \subseteq \Omega$, and moreover (4.8) gives

$$\begin{aligned} \int_{\mathcal{S}(x, 3r)} |Df(z) - M| dz &= \frac{1}{36r^2} \int_{\mathcal{S}(x, 3r)} |Df - M| \leq \frac{1}{36r^2} \int_{\mathcal{B}(\bar{x}, 5r)} |Df - M| \\ &= \frac{25\pi}{36} \int_{\mathcal{B}(\bar{x}, 5r)} |Df - M| \leq \delta, \end{aligned}$$

and by Definition 4.1 this means that $\mathcal{S}(x, r)$ is a Lebesgue square with matrix M and constant δ . Thus, the proof is concluded. \square

4.2. How to “move the vertices” of a grid. In this section we describe how to “move the vertices” of a grid in order to be able to control the average of $|Df|$ inside a square with the average of $|Df|$ on the boundary of the same square. To do so, we first introduce the following notation.

Definition 4.7. *We say that the domain Ω is an r -set if it is the finite union of essentially disjoint squares, all having side $2r$ and sides parallel to the coordinate axes. Any side of one of these squares will be called side of type A if both the endpoints are in the interior of Ω , side of type B if at least one endpoint is in $\partial\Omega$, but the interior of the side is inside Ω , and side of type C if the whole side is in $\partial\Omega$. Any vertex of one of the squares will be then called vertex of type A if it belongs to the interior of Ω , vertex of type B if it belongs to $\partial\Omega$ but it is endpoint of at least one side of type B, and vertex of type C otherwise.*

For any small constant $\varepsilon > 0$, we will define a short segment or curve around each vertex of type A and B. We start with the vertices of type A.

Definition 4.8. *Let Ω be an r -set, $\varepsilon \ll 1$ and let $V = (V_1, V_2)$ be a vertex of type A. We call $I_\varepsilon(V)$ the segment of length $2\sqrt{2}\varepsilon r$ centered at V and with direction $\pi/4$, namely*

$$I_\varepsilon(V) = \{(V_1 + t, V_2 + t) : |t| \leq \varepsilon r\}. \quad (4.9)$$

Notice that all the segments $I_\varepsilon(V)$ lie inside Ω and they do not intersect with each other.

The main result that we prove ensures that the average of $|Df|$ in a side of a square can be always estimated with the average of $|Df|$ in the whole square, up to move the two vertices in the corresponding segments I_ε . Actually, in order to be able also to treat Lebesgue squares with $\det M = 0$, we will in fact estimate $|Df - M|$ instead of $|Df|$ for some matrix M : then, we will apply this result with $M = 0$ for all the non-Lebesgue squares, and with $M = Df(\bar{x})$ for Lebesgue squares “almost centered” at a Lebesgue point \bar{x} . Unfortunately (but the reason is quite evident) the estimate must explode as $1/\varepsilon$; however, this will not be a problem for our construction.

Lemma 4.9. *Let Ω be an r -set, let AB be a side of type A , and let $M \in \mathbb{R}^{2 \times 2}$ be a matrix. Calling $\mathcal{R} \subseteq \Omega$ the union of the 6 squares of the grid having either A or B (or both) as one vertex and*

$$\Gamma(A, B, M) = \left\{ (x, y) \in I_\varepsilon(A) \times I_\varepsilon(B) : \int_{xy} |Df - M| d\mathcal{H}^1 > \frac{25}{\varepsilon r} \int_{\mathcal{R}} |Df - M| d\mathcal{H}^2 \right\},$$

one has

$$\mathcal{H}^1 \left(\left\{ x \in I_\varepsilon(A) : \mathcal{H}^1 \left(\left\{ y \in I_\varepsilon(B) : (x, y) \in \Gamma \right\} \right) > \frac{\mathcal{H}^1(I_\varepsilon(B))}{5} \right\} \right) < \frac{\mathcal{H}^1(I_\varepsilon(A))}{5}. \quad (4.10)$$

Proof. Let us suppose, just to fix the ideas, that the side AB is horizontal, as in Figure 11; suppose also, for the moment, that $M = 0$. Let then $\mathcal{R}_0 \subseteq \mathcal{R}$ be the small parallelogram –dark shaded in the figure, while \mathcal{R} is light shaded– whose four vertices are the endpoints of the segments $I_\varepsilon(A)$ and $I_\varepsilon(B)$. A simple change of variable argument, together with

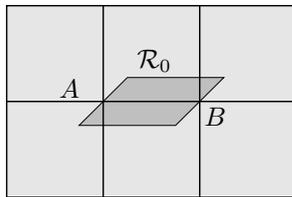


FIGURE 11. The rectangle \mathcal{R} , the side AB , and the two segments $I_\varepsilon(A)$ and $I_\varepsilon(B)$.

the fact that $\mathcal{R}_0 \subseteq \mathcal{R}$, ensures that

$$\int_{x \in I_\varepsilon(A)} \int_{y \in I_\varepsilon(B)} \int_{xy} |Df| d\mathcal{H}^1 dy dx \leq 8\varepsilon r \int_{\mathcal{R}_0} |Df(z)| d\mathcal{H}^2(z) \leq 8\varepsilon r \int_{\mathcal{R}} |Df(z)| d\mathcal{H}^2(z).$$

On the other hand, writing $\Gamma = \Gamma(A, B, 0)$ for shortness, we also have

$$\int_{x \in I_\varepsilon(A)} \int_{y \in I_\varepsilon(B)} \int_{xy} |Df| d\mathcal{H}^1 dy dx \geq \int_{(x,y) \in \Gamma} \int_{xy} |Df| d\mathcal{H}^1 dy dx \geq \mathcal{H}^2(\Gamma) \frac{25}{\varepsilon r} \int_{\mathcal{R}} |Df| d\mathcal{H}^2,$$

hence we deduce

$$\mathcal{H}^2(\Gamma) \leq \frac{8}{25} \varepsilon^2 r^2 = \frac{1}{25} \mathcal{H}^1(I_\varepsilon(A)) \cdot \mathcal{H}^1(I_\varepsilon(B)),$$

from which (4.10) immediately follows.

To show the general case with $M \neq 0$, it is enough to apply the above argument to the function $\tilde{f}(x) = f(x) - Mx$: of course the set $\Gamma(A, B, M)$ coincides with the set $\Gamma(A, B, 0)$ corresponding to the function \tilde{f} , hence (4.10) follows also in the general case. \square

The above result will be useful in order to treat the internal squares, but we have to take care also of the boundary squares. Let us then extend the definition of the segment $I_\varepsilon(V)$ to the vertices of type B .

Definition 4.10. *Let Ω be an r -set and let $V \in \partial\Omega$ be a vertex of type B . If V is vertex of exactly two squares of the decomposition, and these two squares are adjacent (as for V_1 in Figure 12, left), then we call $I_\varepsilon(V)$ the segment of length $2\varepsilon r$ on $\partial\Omega$ centered at V . If V is vertex of three squares of the decomposition (as for V_2 in the figure), then we call $I_\varepsilon(V)$ the union of the two segments contained in $\partial\Omega$, of length εr , having V as one endpoint.*

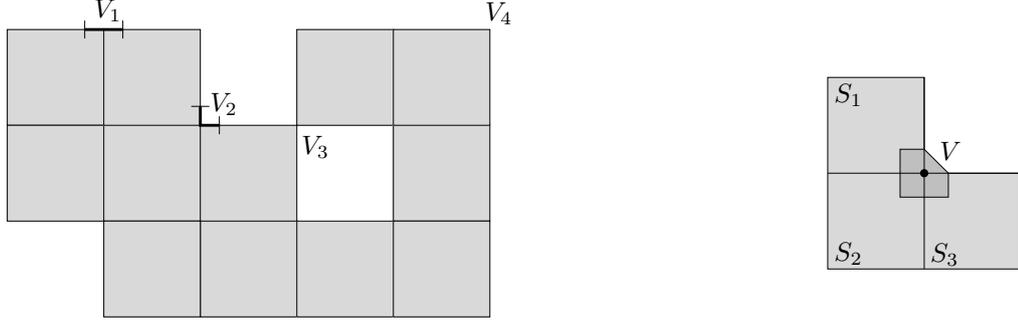


FIGURE 12. Left: some squares and vertices V_i near the boundary of an r -set Ω , and the corresponding $I_\varepsilon(V_i)$. Right: definition of \mathcal{T}_V and Ψ_V .

Notice that we have defined $I_\varepsilon(V)$ only for the vertices of type A and B, thus for instance not for points as V_3 or V_4 in Figure 12. We want now to extend the validity of Lemma 4.9 for sides of type B. To do so near points like V_2 in Figure 12 left, we need a last simple definition.

Definition 4.11. *Let Ω be an r -set and let $V \in \partial\Omega$ be a vertex of three squares of the decomposition, say $\mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3$. We call \mathcal{T}_V the right triangle having right angle at V , two sides of length $r/2$, one horizontal and one vertical, and being not contained in Ω , and for $i = 1, 2, 3$ we call \mathcal{S}_i^- the square contained in \mathcal{S}_i , having one vertex at V , and side $r/2$. Then, we let Ψ_V be the obvious piecewise affine homeomorphism between $\mathcal{S}_1^- \cup \mathcal{S}_2^- \cup \mathcal{S}_3^-$ and $\mathcal{S}_1^- \cup \mathcal{S}_2^- \cup \mathcal{S}_3^- \cup \mathcal{T}_V$, which is bi-Lipschitz with constant 2. Finally, we call Ω^+ the union of Ω with all the triangles \mathcal{T}_V for vertices V as above, and we call $\Psi : \Omega \rightarrow \Omega^+$ the piecewise affine homeomorphism which coincides with Ψ_V around every vertex V , and which is the identity outside (see Figure 12, right). Notice that Ψ is the identity in a r -neighborhood of all the internal squares of the decomposition, and it is globally 2-biLipschitz. Finally, for every $x, y \in \bar{\Omega}$ such that the segment $\Psi(x)\Psi(y)$ is contained in $\bar{\Omega}^+$, we call $\tilde{x}\tilde{y}$ the counterimage, under Ψ , of this segment (which is of course a piecewise linear path).*

We can finally generalize Lemma 4.9 for all the sides of type B; it will be enough to limit ourselves to consider the simpler case $M = 0$.

Lemma 4.12. *Let Ω be an r -set and let $AB \subseteq \Omega$ be a side of type B. Calling $\mathcal{R} \subseteq \Omega$ the union of the squares of the grid having either A or B (or both) as one vertex and defining*

$$\Gamma(A, B) = \left\{ (x, y) \in I_\varepsilon(A) \times I_\varepsilon(B) : \int_{\tilde{x}\tilde{y}} |Df| d\mathcal{H}^1 > \frac{100}{\varepsilon r} \int_{\mathcal{R}} |Df| d\mathcal{H}^2 \right\},$$

one has

$$\mathcal{H}^1 \left(\left\{ x \in I_\varepsilon(A) : \mathcal{H}^1 \left(\left\{ y \in I_\varepsilon(B) : (x, y) \in \Gamma \right\} \right) > \frac{\mathcal{H}^1(I_\varepsilon(B))}{5} \right\} \right) < \frac{\mathcal{H}^1(I_\varepsilon(A))}{5}. \quad (4.11)$$

Proof. This is a very simple generalization of Lemma 4.9, we just have to consider a few possible cases, all depicted in Figure 13. Without loss of generality, we can assume that $B \in \partial\Omega$ and that the segment AB is horizontal. If B belongs to three squares of the decomposition, then there are three possible subcases. First of all, A can be inside Ω

(this is the first case depicted in the figure); second, A can also belong to three squares of the decomposition (depending on how these squares are, this is the second or the third case in the figure); last, A can belong to two squares, which are then necessarily the two squares having AB as a side (this is the fourth case in the figure). Otherwise, B can belong to exactly two adjacent squares of the decomposition, and then again A can either be inside Ω (fifth case in the figure) or in the boundary of Ω : in this latter case (the sixth and last one in the figure) A must also belong only to the same two squares to which B belong, since otherwise we fall back into an already considered case. To clarify the situation, Figure 13 does not show the situation in Ω , but directly in Ω^+ .

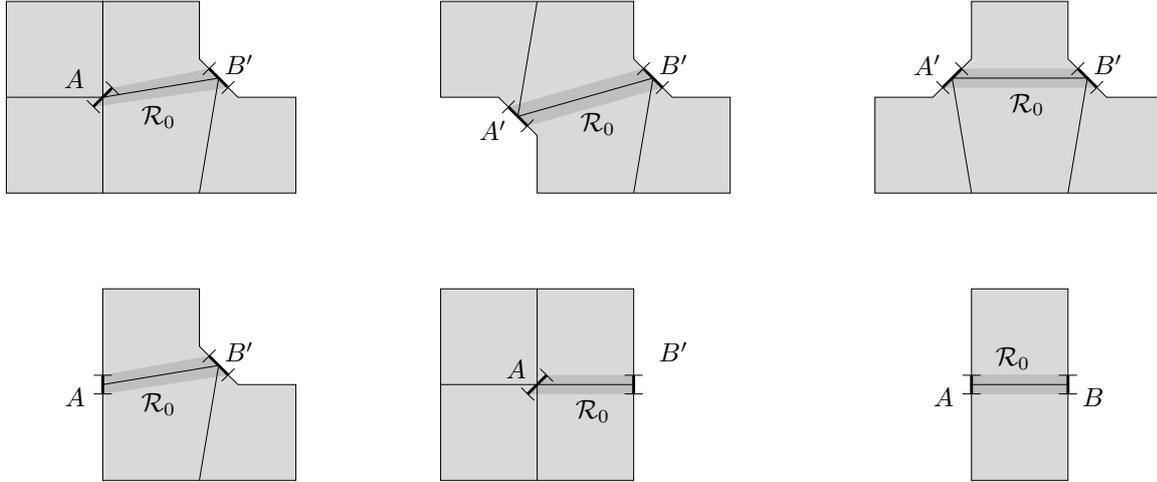


FIGURE 13. The six possibilities in Lemma 4.12.

The proof is now almost identical to the proof already done in Lemma 4.10. Let us call again \mathcal{R}_0 the quadrilateral having as vertices the endpoints of $\Psi(I_\varepsilon(A))$ and $\Psi(I_\varepsilon(B))$; this quadrilateral, depicted in the figure for all the possible cases, belongs to Ω^+ . Notice that, depending on the case, \mathcal{R} can be done by 2, or 3, or 4, or 5 squares, and the figure always shows only (the image under Ψ of) these squares. To conclude the proof we only have to keep in mind that we are interested in what happens in the real domain Ω , not in the simplified domain Ω^+ . However, we can use the map Ψ to move the situation from Ω to Ω^+ ; then, we notice that the very simple calculation done in Lemma 4.10 works perfectly in the new situation; and finally, we use Ψ^{-1} to get back to the case of Ω . The only detail which changes, since Ψ is 2-biLipschitz, is that the constant 25 in the old definition of Γ for internal sides becomes 100 for the new definition of Γ for sides touching the boundary. \square

Let us now fix a matrix $M = M(A, B)$ for any side AB of type A, and write for brevity $\Gamma(A, B) = \Gamma(A, B, M(A, B))$. Thanks to the above results, we have defined a set of “bad pairs” $(x, y) \in \Gamma(A, B)$, where “bad” means that in the segment xy (or in the curve \widehat{xy}) too much derivative is concentrated. We can now find a selection of points $x_V \in I_\varepsilon(V)$ for any vertex V , so that for any side AB the pair (x_A, x_B) does not belong to $\Gamma(A, B)$.

Lemma 4.13. *Let Ω be an r -set, and let us fix a matrix $M(A, B)$ for any side of type A . It is possible to select a point $x_V \in I_\varepsilon(V)$ for any vertex V of type A or B in such a way that, for every side AB of type A or B , one has $(x_A, x_B) \notin \Gamma(A, B)$.*

Proof. We will argue recursively. First of all, let us enumerate all the vertices of type A or B as V_1, V_2, \dots, V_N , for some $N \in \mathbb{N}$. Then, we aim to show that it is possible to select by recursion points $x_i = x_{V_i}$ in every $I_\varepsilon(V_i)$ in such a way that, whenever $V_i V_m$ is a side of type A or B , the point x_i is chosen in such a way that

$$\begin{cases} \mathcal{H}^1\left(\left\{y \in I_\varepsilon(V_m) : (x_i, y) \in \Gamma(V_i, V_m)\right\}\right) \leq \frac{\mathcal{H}^1(I_\varepsilon(V_m))}{5} & \text{if } m > i, \\ (x_i, x_m) \notin \Gamma(V_i, V_m) & \text{if } m < i. \end{cases} \quad (4.12)$$

Notice that, since we will define the points recursively, then the above requests make sense: in particular, if $m < i$ then the point x_m has been already chosen when we have to choose x_i . Of course, if we can select all the points x_i according to (4.12), then we are done: the thesis will be simply given by the second property in (4.12), but the first one is essential to let the recursion work.

Let then $1 \leq i \leq N$, and let us suppose that the points x_j for $j < i$ have been already chosen according to (4.12); let n^- (resp., n^+) be the number of the indices $j < i$ (resp., $j > i$) such that $V_i V_j$ is a side of type A or B . By (4.12) applied to the indices $j < i$, we know that the points $x \in I_\varepsilon(V_i)$ such that $(x, x_j) \in \Gamma(V_i, V_j)$ for some $j < i$ corresponding to a side $V_i V_j$ cover a portion at most $n^-/5$ of $I_\varepsilon(V_i)$. On the other hand, by Lemma 4.12, the points $x \in I_\varepsilon(V_i)$ such that

$$\mathcal{H}^1\left(\left\{y \in I_\varepsilon(V_j) : (x_i, y) \in \Gamma(V_i, V_j)\right\}\right) > \frac{\mathcal{H}^1(I_\varepsilon(V_j))}{5}$$

for some $j > i$ for which $V_i V_j$ is a side of type A or B cover a portion at most $n^+/5$ of $I_\varepsilon(V_i)$. Since of course $n^- + n^+ \leq 4$, we can pick a point $x_i \in I_\varepsilon(V_i)$ for which none of the above problems occur, hence by definition this choice fulfills (4.12). The recursion argument is then proved, and the proof is concluded. \square

The last goal of this section is to define an approximating function \tilde{f} of f on the grid given by the boundaries of the squares. First of all, let us give the definition of “grid” and “modified grid”.

Definition 4.14. *Let Ω be an r -set, and for any vertex V of type B let V' be a given point x_V in $I_\varepsilon(V)$; instead, let $V' = V$ for any vertex V of type A or C . We call grid the union \mathcal{G} of all the sides AB of the squares of the decomposition, while the modified grid is the union $\tilde{\mathcal{G}}$ of all the “modified sides”, that is, the piecewise linear curves $\widetilde{A'B'}$. Notice that, if $AB \subseteq \partial\Omega$, it might happen that the curve $\widetilde{A'B'}$ has not been defined in Definition 4.11; if this is the case, we simply denote by $\widetilde{A'B'}$ the shortest curve in $\overline{\Omega}$ connecting A' and B' : notice that this shortest curve lies entirely inside $\partial\Omega$, and that actually this minimizing property for $\widetilde{A'B'}$ is true also for the sides $AB \subseteq \partial\Omega$ where $\widetilde{A'B'}$ was already defined in Definition 4.11. For every square \mathcal{S} of the grid, we call $\tilde{\mathcal{S}}$ the union of its modified sides.*

Observe that the grid $\widetilde{\mathcal{G}}$ coincides with the grid \mathcal{G} , except near the boundary of Ω ; analogously, the piecewise linear curve $\widetilde{A'B'}$ is nothing else than the segment AB , if it is a side of type A. Notice that both the grid and the modified grid contain the boundary of Ω . We give now our last definition of a map \tilde{f} on $\widetilde{\mathcal{G}}$.

Definition 4.15. *Let Ω be an r -set, and for any side AB of type A fix a matrix $M = M(A, B)$. Let the points $x_V \in I_\varepsilon(V)$ for the different vertices V of type A or B be as in Lemma 4.13. We define the function $g : \widetilde{\mathcal{G}} \rightarrow \mathbb{R}^2$ as follows. For any side AB of type A or B, we define g on the curve $\widetilde{A'B'}$ as the reparameterization, at constant speed, of the function f on the curve $\overline{x_A x_B}$; moreover, we let $g = f$ on $\partial\Omega \subseteq \widetilde{\mathcal{G}}$.*

In the above definition, it is important not to confuse the points x_V with the points V' : according with Definition 4.14, $V' = x_V$ if V is a vertex of type B, while $V' = V$ if V is a vertex of type A or C. In particular, if AB is a side of type A, then $\widetilde{A'B'}$ is simply the segment AB , hence g on the segment AB is the reparameterized copy of f on the segment $x_A x_B = \overline{x_A x_B}$. We conclude this section with an estimate for the function g .

Lemma 4.16. *Let Ω be an r -set, and let the matrices $M = M(A, B)$, the points x_V and the function $g : \widetilde{\mathcal{G}} \rightarrow \mathbb{R}^2$ be as in Definition 4.15. Then, for any side AB of type A, calling ν the unit vector with direction AB and \mathcal{R} again the union of the squares of the grid having either A or B as one vertex, we have*

$$\int_{AB} |Dg(t) - M \cdot \nu| dt \leq \frac{25}{\varepsilon r} \int_{\mathcal{R}} |Df - M| d\mathcal{H}^2 + 11\|M\|\varepsilon r. \quad (4.13)$$

Instead, for any side AB of type B, we have

$$\int_{\widetilde{A'B'}} |Dg(t)| dt \leq \frac{100}{\varepsilon r} \int_{\mathcal{R}} |Df| d\mathcal{H}^2. \quad (4.14)$$

Proof. Let us start with a side AB of type A, and call for brevity $x = x_A$ and $y = x_B$. By Lemma 4.9, we already know that

$$\int_{xy} |Df(s) - M| ds \leq \frac{25}{\varepsilon r} \int_{\mathcal{R}} |Df - M| d\mathcal{H}^2. \quad (4.15)$$

Recall now that, by definition, the function g on the segment AB is simply the reparameterization of the function f on the segment xy . Define then $\lambda = \overline{AB}/\overline{xy}$, and call $\tilde{\nu}$ the unit vector with direction xy : notice that by construction

$$1 - 2\varepsilon \leq \lambda \leq 1 + 3\varepsilon, \quad |\tilde{\nu} - \nu| \leq 2\varepsilon.$$

As a consequence, by a change of variable we obtain

$$\begin{aligned} \int_{AB} |Dg(t) - M \cdot \nu| dt &= \lambda \int_{xy} \left| \frac{Df(s) \cdot \tilde{\nu}}{\lambda} - M \cdot \nu \right| ds = \int_{xy} |Df(s) \cdot \tilde{\nu} - M \cdot \lambda \nu| ds \\ &\leq \int_{xy} |(Df(s) - M) \cdot \tilde{\nu}| ds + \int_{xy} \|M\| |\lambda \nu - \tilde{\nu}| ds \\ &\leq \int_{xy} |Df(s) - M| ds + 5\|M\|\varepsilon \overline{xy}, \end{aligned}$$

thus recalling (4.15) we get (4.13).

Let now AB be a side of type B, and call again for brevity $x = x_A$ and $y = y_B$. In this case, by Lemma 4.12 we already know that

$$\int_{\widetilde{xy}} |Df| d\mathcal{H}^1 \leq \frac{100}{\varepsilon r} \int_{\mathcal{R}} |Df| d\mathcal{H}^2. \quad (4.16)$$

Now, by definition of g we have that the image of $\widetilde{A'B'}$ under g coincides with the image of \widetilde{xy} under f , hence the lengths of the two curves coincide, which means

$$\int_{\widetilde{xy}} |Df| d\mathcal{H}^1 = \int_{\widetilde{A'B'}} |Dg(t)| dt.$$

Hence, (4.14) directly follows from (4.16), and the proof is concluded. \square

4.3. How to modify f on a grid. In this section, we show how to modify a function on a one-dimensional grid; more precisely, we take a generic function defined on a grid, and we modify it in order to become piecewise linear. We have to do so because both our big results, namely Theorems 2.1 and 3.1, need a function which is piecewise linear on the boundary of a square. We start with a rather simple modification, which we will eventually apply to the “bad” squares and to the “good” squares corresponding to a matrix with $\det M = 0$; this construction is reminiscent to the one made in [10], where the situation was more complicated because also the inverse should be approximated.

Proposition 4.17. *Let Ω be an r -set, let \mathcal{G} and $\widetilde{\mathcal{G}}$ be a grid and a modified grid according to Definition 4.14, and let $g : \widetilde{\mathcal{G}} \rightarrow \mathbb{R}^2$ be a continuous, injective function which is piecewise linear on $\partial\Omega$. Then, there exists a piecewise linear and injective function $\hat{g} : \widetilde{\mathcal{G}} \rightarrow \mathbb{R}^2$ such that $\hat{g} = g$ on $\partial\Omega$, $\hat{g}(V') = g(V')$ for every vertex V of the grid, and for every side AB of type A and every matrix M one has*

$$\int_{AB} |D\hat{g}(t) - M \cdot \nu| dt \leq \int_{AB} |Dg(t) - M \cdot \nu| dt, \quad (4.17)$$

while for every side AB of type B one has

$$\int_{\widetilde{A'B'}} |D\hat{g}(t)| dt \leq \int_{\widetilde{A'B'}} |Dg(t)| dt. \quad (4.18)$$

Moreover, on each curve $\widetilde{A'B'}$ the function \hat{g} is an interpolation of finitely many points of the curve $g(\widetilde{A'B'})$.

Proof. We define the map \hat{g} in two steps, namely, we work first around the vertices, and then in the interior of the sides. Figure 14 depicts how the construction works.

Step I. Definition of \hat{g} around the vertices.

Let us start by selecting a vertex V of type A, that is, V belongs to the interior of Ω . There are then four sides of the grid for which V is an endpoint, and we can call V_i , $1 \leq i \leq 4$ the four other endpoints of these sides. Since g is continuous and injective, the quantity

$$\inf \left\{ \overline{g(x)g(V)} : x \in \widetilde{\mathcal{G}} \setminus \bigcup_{i=1}^4 VV_i \right\}$$

is strictly positive: let us then select a small radius $\rho = \rho(V) > 0$, much smaller than this quantity. Hence, by definition, the ball $\mathcal{B}(g(V), \rho)$ intersects the image of the grid

$\tilde{\mathcal{G}}$ under g only in points of the form $g(x)$ for x belonging to one of the four sides VV_i . On the other hand, for each $i = 1, \dots, 4$ there is at least a point $x \in VV_i$ such that $g(x)$ belongs to $\partial\mathcal{B}(g(V), \rho)$. Let us then call V_i^+ the last point x of the segment VV_i for which $g(x) \in \partial\mathcal{B}(g(V), \rho)$, where “last” means “the farthest from V ”. We define then the function \hat{g} , on each of the four segments VV_i^+ , simply as the linear function connecting $g(V)$ with $g(V_i^+)$.

Let us now consider a vertex V of type B, that is, V belongs to $\partial\Omega$ but there is some side of the grid, contained in the interior of Ω , of which V is an endpoint: let us call j the number of such sides, and notice that by construction j is either one or two (keep in mind Figure 13). We argue then similarly as before: we call V_i , for $1 \leq i \leq j$ the other endpoints of these internal sides, and we consider the strictly positive quantity

$$\inf \left\{ \overline{g(x)g(V)} : x \in \tilde{\mathcal{G}} \setminus \left(\bigcup_{i=1}^j \widetilde{V'V'_i} \cup \partial\Omega \right) \right\}.$$

This time, we will define $\rho = \rho(V)$ not only much smaller than the above quantity, but also so small that $g(\partial\Omega) \cap \mathcal{B}(g(V'), \rho)$ is the union of two segments (this is surely true as soon as ρ is small enough, since g is piecewise linear on $\partial\Omega$). Exactly as before, for $1 \leq i \leq j$ we call V_i^+ the last point $x \in \widetilde{V'V'_i}$ such that $g(x) \in \partial\mathcal{B}(g(V'), \rho)$: up to further decrease $\rho(V)$, we can also assume that the portion of the piecewise linear curve $\widetilde{V'V'_i}$ connecting V' and V_i^+ is simply a segment. Then, as before we define \hat{g} on each of the j segments $V'V_i^+$ as the linear function connecting $g(V')$ with $g(V_i^+)$.

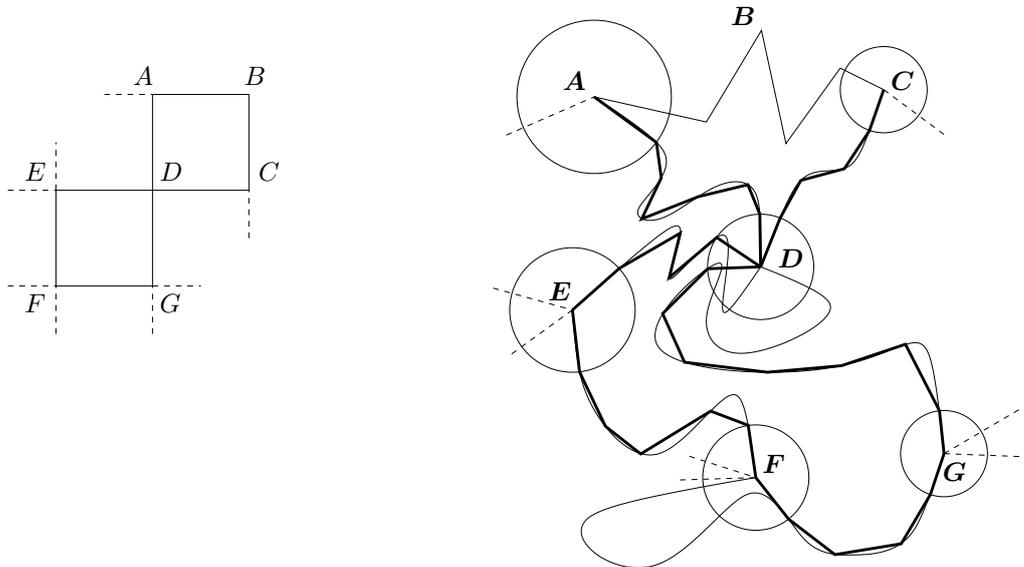


FIGURE 14. The construction in Proposition 4.17: the points A , B and C are in $\partial\Omega$, while D , E , F and G are inside Ω ; the image of \hat{g} inside Ω is thicker.

Step II. Definition of \hat{g} inside the sides.

Up to now, we have defined \hat{g} on a neighborhood of each vertex of type A and B; moreover, \hat{g} is already automatically defined on $\partial\Omega$, since it must be $\hat{g} = g$ on $\partial\Omega$. Therefore, to conclude we have to define \hat{g} on the remaining portion of $\tilde{\mathcal{G}}$. By construction, this

remaining portion is a finite and disjoint union of internal parts of sides of type A or B: more precisely, for every side AB of type A there is a segment $A^+B^- \subset\subset AB$ where \hat{g} has to be defined, while for every side AB of type B \hat{g} has still to be defined in some piecewise linear curve $\widetilde{A^+B^-} \subset\subset \widetilde{A'B'}$.

Let us first consider the case of a side AB of type A: the function \hat{g} has been already defined in the segment AA^+ (resp., B^-B) as the linear function connecting $g(A)$ and $g(A^+)$ (resp., $g(B^-)$ and $g(B)$), and moreover the points $\hat{g}(A^+) = g(A^+)$ and $\hat{g}(B^-) = g(B^-)$ are in the boundary of the disks $\mathcal{B}(g(A), r(A))$ and $\mathcal{B}(g(B), r(B))$ respectively. We have to define \hat{g} in the segment A^+B^- , and this must be a piecewise linear curve connecting $g(A^+)$ and $g(B^-)$. Observe that g , in the segment A^+B^- , is already a curve connecting $g(A^+)$ and $g(B^-)$, the only problem being that it is not necessarily piecewise linear. We can then select many points $P_0 = A^+, P_1, P_2, \dots, P_N = B^-$ in the segment A^+B^- , and define \hat{g} in A^+B^- as the piecewise affine interpolation of these values (that is, $\hat{g}(P_i) = g(P_i)$ and \hat{g} is linear on each P_iP_{i+1}). A very simple geometric argument (similar to [10, Lemma 5.5], but much easier) shows that, by choosing carefully many points, the map \hat{g} in A^+B^- is injective, it never crosses the two disks $\mathcal{B}(g(A), \rho(A))$ and $\mathcal{B}(g(B), \rho(B))$, and its L^∞ distance with g is much smaller than

$$\inf \left\{ \overline{g(x)g(y)} : x \in A^+B^-, y \in \widetilde{\mathcal{G}} \setminus AB \right\}.$$

Let us now consider a side AB of type B: in this case, \hat{g} is piecewise linear in the two segments $A'A^+$ and B^-B' , and we have to extend the definition to the piecewise linear curve $\widetilde{A^+B^-}$. This can be done exactly as we just did for a side of type A, the only difference is that, by doing the interpolation, the points P_i must include all the extremes of the segments forming the curve $\widetilde{A^+B^-}$. Apart from that, nothing else changes, thus the definition of \hat{g} is finally concluded.

Step III. The properties of \hat{g} .

To conclude the proof, we just need to check that \hat{g} fulfills all the required properties. The fact that $\hat{g} = g$ on $\partial\Omega$ and at every vertex is true by construction, as well as the fact that \hat{g} is an interpolation of finitely many points of the curve $g(\widetilde{A'B'})$ on every side $\widetilde{A'B'}$ of $\widetilde{\mathcal{G}}$. To check (4.17) and (4.18), we just have to keep in mind that $\int_{AB} |Dg|$ is the length of the curve g on the segment AB , while $\int_{AB} |D\hat{g}|$ is the length of the (piecewise affine) curve \hat{g} on AB . But any interpolation of points of a curve is shorter than the curve itself, thus (4.17) follows immediately for the case $M = 0$, and the very same argument with $\widetilde{A'B'}$ in place of AB ensures also (4.18).

To show (4.17) when $M \neq 0$, let $CD \subseteq AB$ be a segment where \hat{g} is linear and satisfies $\hat{g}(C) = g(C)$, $\hat{g}(D) = g(D)$: then

$$\begin{aligned} \int_C^D |D\hat{g}(t) - M \cdot \nu| dt &= \left| \int_C^D D\hat{g}(t) - M \cdot \nu dt \right| = \left| \int_C^D Dg(t) - M \cdot \nu dt \right| \\ &\leq \int_C^D |Dg(t) - M \cdot \nu| dt, \end{aligned}$$

thus adding over the segments where \hat{g} is linear we get (4.17) and the proof is finished. \square

We conclude this section with the following generalization of the above result, rather technical but very useful to obtain Theorem 1.1, and whose proof is actually nothing else than a straightforward modification of the previous one.

Proposition 4.18. *Let $\mathcal{C} = \cup_{i=1}^N A_i B_i$ be a finite union of closed segments in \mathbb{R}^2 , and let $\mathcal{C}_0 = \cup_{i=1}^{N_0} A_i B_i$, with $N_0 \leq N$, be a selection of some of them. Let $g : \mathcal{C} \rightarrow \mathbb{R}^2$ be a continuous one-to-one function, and let $\eta > 0$ be given. Then, there exists another continuous and one-to-one function $\hat{g} : \mathcal{C} \rightarrow \mathbb{R}^2$ such that $\hat{g} = g$ at every endpoint of each segment, $\{\hat{g} \neq g\}$ is contained in the η -neighborhood of \mathcal{C}_0 , \hat{g} is piecewise linear on \mathcal{C}_0 , and for every side of \mathcal{C} the estimates (4.17) and (4.18) hold.*

Proof. The proof can be done almost exactly as in Proposition 4.17. First of all, up to a subdivision of some of the segments, we can assume that every two segments are either disjoint, or they meet at a common endpoint. Then, we start defining $\hat{g} = g$ on all the sides both whose endpoints are not contained in \mathcal{C}_0 . Further, we consider any of the remaining vertices, say V . We can select a small ρ such that the ball $\mathcal{B}(g(V), \rho)$ contains only points of the form $g(x)$ for x belonging to one of the (finitely many) segments VV_i having V as an endpoint; up to decrease ρ , we can also ensure that $|x - V| < \eta$ for any such x . We define then the points V_i^+ exactly as in Step I of Proposition 4.17, and we let \hat{g} be linear on each segment VV_i^+ : the continuity and injectivity of \hat{g} up to now is then clear. In the portions of the segments where \hat{g} has not yet been defined, we can then define it in two different ways: inside the segments which belong to \mathcal{C}_0 , we define a piecewise linear \hat{g} exactly as in Step II of Proposition 4.17; inside the other segments, we simply let $\hat{g} = g$.

By construction and arguing as in Step III of Proposition 4.17, we can then immediately observe that the function \hat{g} is as required. \square

4.4. The proof of Theorem 1.1. We are finally in position to give the proof of Theorem 1.1, which will come by putting together all the results that we got up to now. For the reader's convenience, we split the proof in some parts. The first one is a very peculiar case, namely, when Ω is an r -set and the function f is already piecewise linear on the boundary; nevertheless, most of the difficulties are contained in this part.

Proposition 4.19. *Under the assumption of Theorem 1.1, assume that in addition Ω is an r -set, and that f is continuous up to $\partial\Omega$ and piecewise linear there. Then, for every $\varepsilon > 0$ there exists a finitely piecewise affine homeomorphism $f_\varepsilon : \Omega \rightarrow \mathbb{R}^2$ such that*

$$\|f_\varepsilon - f\|_{W^{1,1}} + \|f_\varepsilon - f\|_{L^\infty} < \varepsilon, \quad f_\varepsilon = f \text{ on } \partial\Omega. \quad (4.19)$$

Proof. The idea of the construction is rather simple: we divide the squares in four groups, namely, the Lebesgue squares with positive determinant, the Lebesgue squares with $M \neq 0$ but zero determinant, the Lebesgue squares with $M = 0$, and the other ones. Inside the first squares we can substitute f with $\varphi_{\mathcal{S}(c,r)}$ and rely on Lemma 4.3, for the second ones we will use Theorem 3.1, and for the third and fourth ones Theorem 2.1. However, to treat each square separately, we need to take care of the values on the boundaries of the squares: on one hand, they must be piecewise linear, in order to allow us to apply Theorems 3.1

and 2.1, and this will be obtained thanks to Proposition 4.17; but on the other hand, any two adjacent squares must have the same boundary values on the common side, and this will require same care. Let us then start with the proof, dividing it in several steps.

Step I. Definition of the constants ε_i and of the sets \mathcal{A}_1 , \mathcal{A}_2 and \mathcal{A}_3 .

First of all, we have to take five small constant ε_i for $1 \leq i \leq 5$. More precisely, ε_1 is a small geometric constant (for instance, $\varepsilon_1 = 1/10$ is enough); instead, the constants $\varepsilon_5 \ll \varepsilon_4 \ll \varepsilon_3 \ll \varepsilon_2 \ll \varepsilon$ will depend on the data, that is, Ω , f and ε . More precisely, since $f \in W^{1,1}(\Omega)$, we can select $\varepsilon_2 \ll 1$ so small that

$$\int_A |Df| \leq \frac{\varepsilon_1 \varepsilon}{54K} \quad \forall A \subseteq \Omega : |A| \leq \varepsilon_2, \quad (4.20)$$

where K is a purely geometric constant, which we will make explicit during the proof. Then, we define $\varepsilon_3 \ll \varepsilon_2$ in such a way that

$$\left| \left\{ x \in \Omega : 0 < |Df(x)| < \varepsilon_3 \text{ or } |Df(x)| > \frac{1}{\varepsilon_3} \text{ or } 0 < \det(Df(x)) < \varepsilon_3 \right\} \right| < \frac{\varepsilon_2}{45}. \quad (4.21)$$

This estimate is surely true as soon as ε_3 is small enough, depending on ε_2 , on Ω , and on f . Now, we define the following two subsets of the matrices $M \in \mathbb{R}^{2 \times 2}$,

$$\mathcal{M}^+ = \left\{ \varepsilon_3 < \|M\| < \frac{1}{\varepsilon_3}, \det M > \varepsilon_3 \right\}, \quad \mathcal{M}^0 = \left\{ \varepsilon_3 < \|M\| < \frac{1}{\varepsilon_3}, \det M = 0 \right\},$$

which of course depend only on ε_3 . Finally, we let $\varepsilon_5 \ll \varepsilon_4 \ll \varepsilon_3$ be so that

$$\frac{\varepsilon_5}{\varepsilon_4} + \frac{\varepsilon_4}{\varepsilon_3} + \frac{\varepsilon_5}{\varepsilon_1} \leq \frac{\varepsilon}{6K|\Omega|}, \quad \varepsilon_5 \ll \varepsilon_4 \varepsilon_3, \quad \varepsilon_4 \ll \varepsilon_3^2, \quad (4.22)$$

and we set $\hat{\delta} = \hat{\delta}(\varepsilon_3, \varepsilon_5)$ as

$$\hat{\delta} = \min \{ \bar{\delta}(M, \varepsilon_5), M \in \mathcal{M}^+ \cup \mathcal{M}^0 \}, \quad (4.23)$$

where $\bar{\delta}$ are the constants of Lemma 4.3. Observe that $\hat{\delta}$ really depends only on ε_3 and ε_5 by construction, as observed in Remark 4.5. The last constant to select is r : indeed, Ω is an r -set, but then we can regard it as an r/H -set for every $H \in \mathbb{N}$; as a consequence, we can now change the value of r , making it as small as we need: in particular, we let r be so small that

$$rP(\Omega) + \left| \{x \in \Omega : \bar{r}(x, \hat{\delta}) \leq r\} \right| < \frac{\varepsilon_2}{180}, \quad rP(f(\Omega)) \leq \frac{\varepsilon}{66K}, \quad (4.24)$$

where the constants $\bar{r}(x, \hat{\delta})$ have been defined in Lemma 4.6 for every $x \in \Omega$ which is a Lebesgue points for Df (so, for almost every point of Ω), and where $P(A)$ is as usual the perimeter of A , that is, $\mathcal{H}^1(\partial A)$.

Having fixed all the constants ε_i , and having also chosen the final value of r , we can now enumerate the squares of the grid as \mathcal{S}_i , $1 \leq i \leq N$, and we subdivide these squares

in four groups, namely,

$$\begin{aligned}\mathcal{A}_1 &= \left\{ \mathcal{S}_i \subset\subset \Omega : \mathcal{S}_i \text{ is a Lebesgue square with matrix } M_i \in \mathcal{M}^+ \text{ and constant } \hat{\delta} \right\}, \\ \mathcal{A}_2 &= \left\{ \mathcal{S}_i \subset\subset \Omega : \mathcal{S}_i \text{ is a Lebesgue square with matrix } M_i \in \mathcal{M}^0 \text{ and constant } \hat{\delta} \right\}, \\ \mathcal{A}_3 &= \left\{ \mathcal{S}_i \subset\subset \Omega : \mathcal{S}_i \text{ is a Lebesgue square with matrix } M_i = 0 \text{ and constant } \hat{\delta} \right\}, \\ \mathcal{A}_4 &= \left\{ \mathcal{S}_i : \mathcal{S}_i \notin \mathcal{A}_1 \cup \mathcal{A}_2 \cup \mathcal{A}_3 \right\}.\end{aligned}$$

We aim now to show that most of the squares belong to the first three groups. In fact, let us consider the total area of the squares contained in \mathcal{A}_4 . The union of those which touch the boundary of Ω has of course an area smaller than $rP(\Omega)$. Let us instead consider a square $\mathcal{S}(c, r) \in \mathcal{A}_4$ compactly contained in Ω : this means that, for every point $x \in \mathcal{S}(c, r/2)$, either we cannot apply Lemma 4.6 with constant $\hat{\delta}$ to x (thus, $\bar{r}(x, \hat{\delta}) \leq r$), or x belongs to the set in (4.21). Since the area of $\mathcal{S}(c, r)$ is four times greater than the area of $\mathcal{S}(c, r/2)$, by (4.21) and (4.24) we deduce

$$\left| \bigcup \{ \mathcal{S}_i \in \mathcal{A}_4 \} \right| \leq rP(\Omega) + 4 \left(\left| \{ x \in \Omega : \bar{r}(x, \hat{\delta}) \leq r \} \right| + \frac{\varepsilon_2}{45} \right) < \frac{\varepsilon_2}{9}. \quad (4.25)$$

Step II. Squares in \mathcal{A}_1 and \mathcal{A}_2 never meet.

Let us now show that squares in \mathcal{A}_1 and \mathcal{A}_2 never meet, that is, no vertex of a square in \mathcal{A}_1 can be vertex also of a square in \mathcal{A}_2 : this will come as a simple consequence of the L^∞ estimate (4.1) in Lemma 4.3. Indeed, assume for simplicity of notations that $\mathcal{S}_1 = \mathcal{S}(c_1, r)$ and $\mathcal{S}_2 = \mathcal{S}(c_2, r)$ have some common vertex, and that $\mathcal{S}_1 \in \mathcal{A}_1$, $\mathcal{S}_2 \in \mathcal{A}_2$. Then, Lemma 4.3 provides us with two affine functions ψ_1, ψ_2 satisfying $D\psi_1 = M_1$, $D\psi_2 = M_2$, each \mathcal{S}_i being a Lebesgue square with matrix M_i . Let us now call

$$\mathcal{S}_{3/2} := \mathcal{S}\left(\frac{c_1 + c_2}{2}, r\right) \subseteq \mathcal{S}(c_1, 2r) \cap \mathcal{S}(c_2, 2r)$$

so that by (4.1) and recalling (4.23), we have

$$\|\psi_1 - \psi_2\|_{L^\infty(\mathcal{S}_{3/2})} \leq \|\psi_1 - f\|_{L^\infty(\mathcal{S}(c_1, 2r))} + \|\psi_2 - f\|_{L^\infty(\mathcal{S}(c_2, 2r))} < \frac{r\varepsilon_5}{5}.$$

Since $\det M_2 = 0$, by construction we can find two points $x, y \in \mathcal{S}_{3/2}$ such that $|y - x| = r$ and $\psi_2(y - x) = 0$. The last inequality, keeping in mind the definition of \mathcal{A}_1 , yields then

$$\frac{r\varepsilon_5}{5} > |\psi_1(y - x)| \geq \frac{\det M_1}{\|M_1\|} |y - x| > \varepsilon_3^2 r.$$

Since this is in contradiction with (4.22), we have concluded the proof of this step. For later use we underline that, more in general, we have proved what follows:

$$\forall \mathcal{S}_a, \mathcal{S}_b \in \mathcal{A}_1 \cup \mathcal{A}_2 \text{ adjacent, one has } \|M_a - M_b\| < \frac{\varepsilon_5}{5}. \quad (4.26)$$

Step III. A tentative modified grid $\tilde{\mathcal{G}}_1$ and of a tentative function $g_1 : \tilde{\mathcal{G}}_1 \rightarrow \mathbb{R}^2$.

In this step, we define a modified grid $\tilde{\mathcal{G}}_1$ and a function g_1 on it. To do so, we simply have to choose a matrix $M(A, B)$ for every side AB of type A of the grid \mathcal{G} ; once done

this, we get first the points $x_V \in I_{\varepsilon_4}(V)$ from Lemma 4.13 applied with ε_4 in place of ε , then the modified grid $\widetilde{\mathcal{G}}_1$ from Definition 4.14, and finally the function $g_1 : \widetilde{\mathcal{G}}_1 \rightarrow \mathbb{R}^2$ from Definition 4.15.

The matrices $M(A, B)$ will be defined as follows. For any side AB of type A, let for a moment \mathcal{S}_a and \mathcal{S}_b be the two squares of the grid having AB as a side: then, if neither \mathcal{S}_a nor \mathcal{S}_b belong to \mathcal{A}_2 , we let $M(A, B) = 0$; if $\mathcal{S}_a \in \mathcal{A}_2$ but $\mathcal{S}_b \notin \mathcal{A}_2$, then we let $M(A, B) = M_a$, and analogously if $\mathcal{S}_a \notin \mathcal{A}_2$ and $\mathcal{S}_b \in \mathcal{A}_2$ we let $M(A, B) = M_b$; finally, if both \mathcal{S}_a and \mathcal{S}_b belong to \mathcal{A}_2 , then we set $M(A, B)$ to be arbitrarily one between M_a and M_b : it makes no difference which one we choose, since by (4.26) we know that $M_a \approx M_b$.

We want now to evaluate $\int_{\partial\widetilde{\mathcal{S}}} |Dg_1|$ for some of the modified squares $\widetilde{\mathcal{S}}$ (recall Definition 4.14). Let us start by taking a square $\mathcal{S} = \mathcal{S}(c, r) \in \mathcal{A}_3 \cup \mathcal{A}_4$, and let us call $\mathcal{S}^+ = \mathcal{S}(c, 3r) \cap \Omega$ the union of the nine squares around it (more precisely, of those which are inside Ω). Take any side $AB \subseteq \partial\mathcal{S}$ of type A or B, and observe that the union \mathcal{R}_{AB} of the squares touching A or B is contained in \mathcal{S}^+ . If AB is of type A but $M(A, B) = 0$, or if AB is of type B, we can apply Lemma 4.16 (using (4.13) or (4.14) if AB is of type A or B respectively) and get

$$\int_{\widetilde{A'B'}} |Dg_1| d\mathcal{H}^1 \leq \frac{100}{\varepsilon_4 r} \int_{\mathcal{R}_{AB}} |Df| d\mathcal{H}^2 \leq \frac{100}{\varepsilon_4 r} \int_{\mathcal{S}^+} |Df| d\mathcal{H}^2.$$

Instead, if $AB \subseteq \partial\mathcal{S} \cap \partial\Omega$, then of course

$$\int_{\widetilde{A'B'}} |Dg_1| d\mathcal{H}^1 = \int_{x_A x_B} |Df| d\mathcal{H}^1 = \int_{x_A x_B \cap \partial\Omega} |Df| d\mathcal{H}^1.$$

Since the boundary of $\widetilde{\mathcal{S}}$ is nothing else than the union of its four modified sides $\widetilde{A'B'}$, adding the last two estimates for the four sides of $\partial\mathcal{S}$ we get

$$\int_{\partial\widetilde{\mathcal{S}}} |Dg_1| d\mathcal{H}^1 \leq \frac{400}{\varepsilon_4 r} \int_{\mathcal{S}^+} |Df| d\mathcal{H}^2 + \int_{\partial\mathcal{S}^+ \cap \partial\Omega} |Df| d\mathcal{H}^1 \quad \forall \mathcal{S} \in \mathcal{A}_{3,4}^-, \quad (4.27)$$

being

$$\mathcal{A}_{3,4}^- = \left\{ \mathcal{S} \in \mathcal{A}_3 \cup \mathcal{A}_4 : M(A, B) = 0 \text{ for each side } AB \text{ of } \mathcal{S} \right\}.$$

Let us now consider a square $\mathcal{S} = \mathcal{S}(c, r) \in \mathcal{A}_2$, and notice that by definition it is compactly contained in Ω , so $\widetilde{\mathcal{S}} = \mathcal{S}$ and $\widetilde{A'B'} = AB$ for any of its sides. Let AB be one of those sides, and notice that $\|M(A, B) - M\| \leq \varepsilon_5/5$, being \mathcal{S} a Lebesgue square with matrix M and constant $\hat{\delta}$: indeed if the other square having AB as a side is not in \mathcal{A}_2 , then $M(A, B) = M$, and otherwise the inequality is given by (4.26). As a consequence,

again (4.13) of Lemma 4.16 and the definition of \mathcal{M}^0 give

$$\begin{aligned}
\int_{AB} |Dg_1 - M \cdot \nu| d\mathcal{H}^1 &\leq \int_{AB} |Dg_1 - M(A, B) \cdot \nu| d\mathcal{H}^1 + \frac{2}{5} r\varepsilon_5 \\
&\leq \frac{25}{\varepsilon_4 r} \int_{\mathcal{R}} |Df - M(A, B)| d\mathcal{H}^2 + 11 \|M\| \varepsilon_4 r + \frac{2}{5} r\varepsilon_5 \\
&\leq \frac{25}{\varepsilon_4 r} \int_{\mathcal{R}} |Df - M| d\mathcal{H}^2 + 120r \frac{\varepsilon_5}{\varepsilon_4} + 11r \frac{\varepsilon_4}{\varepsilon_3} + \frac{2}{5} r\varepsilon_5 \quad (4.28) \\
&\leq \frac{25}{\varepsilon_4 r} \int_{\mathcal{S}^+} |Df - M| d\mathcal{H}^2 + 120r \frac{\varepsilon_5}{\varepsilon_4} + 11r \frac{\varepsilon_4}{\varepsilon_3} + \frac{2}{5} r\varepsilon_5 \\
&\leq 1020r \frac{\varepsilon_5}{\varepsilon_4} + 11r \frac{\varepsilon_4}{\varepsilon_3} + \frac{2}{5} r\varepsilon_5,
\end{aligned}$$

where in the last inequality we have used Definition 4.1 together with the fact that $\hat{\delta} \leq \bar{\delta}(M, \varepsilon_5) \leq \varepsilon_5$.

Step IV. The “correct” modified grid $\tilde{\mathcal{G}}$ and a second tentative function $g_2 : \tilde{\mathcal{G}} \rightarrow \mathbb{R}^2$.

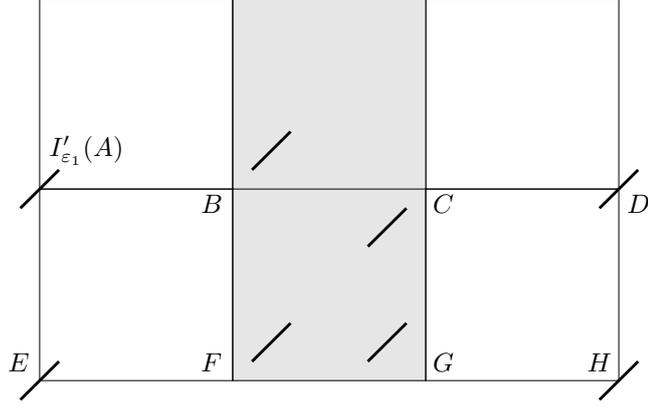
In this step, we define a second modified grid and a second tentative function; the idea is to repeat almost exactly the procedure of Step III, but using the neighborhoods $I_{\varepsilon_1}(V)$ instead of those $I_{\varepsilon_4}(V)$. In fact, the presence of ε_4 is perfect for the squares in \mathcal{A}_2 , since in (4.28) we only have small terms like $\varepsilon_4/\varepsilon_3$ or $\varepsilon_5/\varepsilon_4$; instead, for the squares in $\mathcal{A}_3 \cup \mathcal{A}_4$, the constant ε_4 in (4.27) is too small, and we would need something much larger than ε_2 . Since there is no constant which is at the same time much larger than ε_2 and much smaller than ε_3 , we are forced to repeat the procedure.

This time, let us define the matrices $M'(A, B) = 0$ for all the sides AB of type A, and let us consider a slightly modified version of the intervals $I_{\varepsilon_1}(V)$. More precisely, we let $I'_{\varepsilon_1}(V) = I_{\varepsilon_1}(V)$ for all the vertices V which are not in the boundary of some square of \mathcal{A}_1 or \mathcal{A}_2 . Instead, if a vertex $V = (V_1, V_2)$ belongs to the boundary of at least a square in \mathcal{A}_1 or \mathcal{A}_2 (these two things cannot happen simultaneously, thanks to Step II), let us define $I'_{\varepsilon_1}(V)$ as a translation of $I_{\varepsilon_1}(V)$ of $(\pm 2\varepsilon_1 r, \pm 2\varepsilon_1 r)$, where the two choices of the sign \pm are done in such a way that the whole interval is inside a square of \mathcal{A}_1 or \mathcal{A}_2 ; for instance, if V is the lower-left corner of a square in \mathcal{A}_1 (or \mathcal{A}_2), we can set

$$I'_{\varepsilon_1}(V) = \{(V_1 + t, V_2 + t) : \varepsilon_1 r \leq t \leq 3\varepsilon_1 r\},$$

compare with (4.9). If V is corner of more than one square in \mathcal{A}_1 or \mathcal{A}_2 , then we let $I'_{\varepsilon_1}(V)$ be inside one of them arbitrarily, this will not make any difference. Figure 15 shows an example of a portion of an r -set Ω , where eight intervals $I'_{\varepsilon_1}(V)$ are depicted and the shaded squares are those in \mathcal{A}_1 (or \mathcal{A}_2). Notice that the intervals $I'_{\varepsilon_1}(B)$ and $I'_{\varepsilon_1}(C)$ could be inside each of the two shaded squares, in this example we have put the first interval inside the above square and the second interval in the below one.

After a quick look at the proof of Lemma 4.9, it is evident that it works perfectly even with the intervals $I'_{\varepsilon_1}(V)$ in place of the $I_{\varepsilon_1}(V)$: indeed, in that simple proof we just used that the internal intervals are all of length $2\sqrt{2}\varepsilon r$, with direction at 45° , and placed very close to the vertices. As a consequence, we obtain the points $x'_V \in I'_{\varepsilon_1}(V)$ from Lemma 4.13, the modified grid $\tilde{\mathcal{G}}$ from Definition 4.14, and the function $g_2 : \tilde{\mathcal{G}} \rightarrow \mathbb{R}^2$


 FIGURE 15. The intervals $I'_{\epsilon_1}(V)$ in Step IV.

from Definition 4.15. The modified grid that we obtained now is the “correct” one, and we will use the function g_1 (resp., g_2) around squares in \mathcal{A}_2 (resp., \mathcal{A}_3 and \mathcal{A}_4). The very same calculations of last step work also for this new case, just substituting the constant ϵ_4 with ϵ_1 . In particular, since this time $M'(A, B) = 0$ for all the sides, the estimate (4.27) is true for every square in $\mathcal{A}_3 \cup \mathcal{A}_4$, so we can rewrite it (with ϵ_1 in place of ϵ_4) as

$$\int_{\partial \tilde{\mathcal{S}}} |Dg_2| d\mathcal{H}^1 \leq \frac{400}{\epsilon_1 r} \int_{\mathcal{S}^+} |Df| d\mathcal{H}^2 + \int_{\partial \mathcal{S}^+ \cap \partial \Omega} |Df| d\mathcal{H}^1 \quad \forall \mathcal{S} \in \mathcal{A}_3 \cup \mathcal{A}_4. \quad (4.29)$$

Step V. Definition of the function $g_3 : \tilde{\mathcal{G}} \rightarrow \mathbb{R}^2$.

We are now in position to define a function g_3 on the grid $\tilde{\mathcal{G}}$ introduced in Step IV. This function will behave almost correctly on the whole grid, its only fault (which will be solved in next step) being not to be piecewise linear. As we already observed, we would like to set $g_3 = \varphi_{\mathcal{S}}$ on the boundary of any square $\mathcal{S} \in \mathcal{A}_1$, $g_3 = g_1$ on the boundary of the squares in \mathcal{A}_2 , and $g_3 = g_2$ on the boundary of squares in \mathcal{A}_3 or \mathcal{A}_4 ; of course, this is impossible because the function g_3 would then not be continuous and injective. As a consequence, we use the above overall strategy to define g_3 , but with some *ad hoc* modification where squares of different types meet, so to get continuity and injectivity.

Let us start with the easy part of this definition. For every side AB which is in the boundary of some square of \mathcal{A}_1 , we define g_3 on AB as the linear interpolation which satisfies $g_3(A) = f(A)$ and $g_3(B) = f(B)$: as a consequence, $g_3 = \varphi_{\mathcal{S}}$ on $\partial \mathcal{S}$ for every $\mathcal{S} \in \mathcal{A}_1$, where $\varphi_{\mathcal{S}}$ is given by Definition 4.2. Second, for every side AB which is in the boundary of some square in \mathcal{A}_2 , we let $g_3 = g_1$ on AB : recall that vertices of squares in \mathcal{A}_1 and vertices of squares in \mathcal{A}_2 are distinct by Step II. Finally, for every side AB such that neither A nor B are vertices of squares of \mathcal{A}_1 or of \mathcal{A}_2 , we let $g_3 = g_2$ on $\widetilde{A'B'}$, where the points A' and B' are those given by Step IV. Notice that, up to now, the function g_3 is continuous and injective: this comes as a straightforward consequence of the L^∞ estimate around squares in \mathcal{A}_1 , and of the fact that g_3 is a reparameterization of f on different segments around squares in \mathcal{A}_2 or $\mathcal{A}_3 \cup \mathcal{A}_4$. Nevertheless, g_3 has still not been defined in the whole $\tilde{\mathcal{G}}$.

Let us then consider a side AB such that g_3 has not yet been defined on $\widetilde{A'B'}$: by construction, this means that at least one between A and B is vertex of a square in \mathcal{A}_1 or \mathcal{A}_2 , thus in particular AB is not in the boundary of Ω , and both the squares of the grid having AB in the boundary belong to $\mathcal{A}_3 \cup \mathcal{A}_4$. We aim now to define g_3 on $\widetilde{A'B'}$. To do so, let us keep in mind that we would like to set $g_3 = g_2$, and observe also that g_2 , on $\widetilde{A'B'}$, is nothing else than the reparametrization, at constant speed, of the image γ_0 of some piecewise linear curve $\widetilde{x'_A x'_B}$ under f . Our idea is to define g_3 , on $\widetilde{A'B'}$, again as the reparametrization at constant speed of some modification γ of γ_0 . In particular, γ and γ_0 will coincide in their big “internal” parts, the difference being only near the endpoints of these curves.

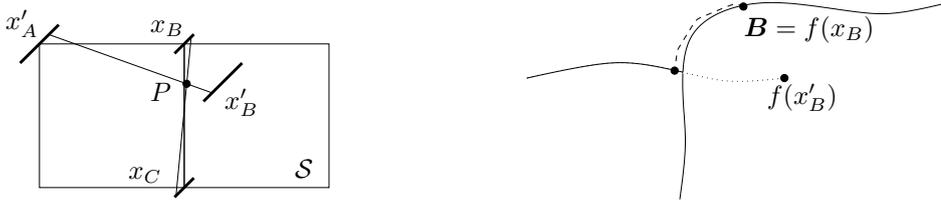


FIGURE 16. The definition of g_3 in Step V if B is in some square of \mathcal{A}_2 .

For simplicity, let us start assuming that A is not vertex of squares in \mathcal{A}_1 or \mathcal{A}_2 , and B is vertex of at least a square in \mathcal{A}_2 (by construction and recalling Step II, B is then vertex of either one or two squares in \mathcal{A}_2 , and of no square in \mathcal{A}_1). By Step IV, we know that the interval $I'_{\varepsilon_1}(B)$ is entirely inside a square \mathcal{S} of \mathcal{A}_2 which has B as a vertex. As in Figure 16 (left), let us call $C \neq B$ the vertex of \mathcal{S} such that there is a square having both A and C as vertices. The function g_3 has already been defined, on BC , as the reparametrization of the image of a segment $x_B x_C$ under f ; by definition and by construction, the segment $x_B x_C$ and the curve $\widetilde{x'_A x'_B}$ meet in some point P near B ; notice that $B \in \Omega$, and then the curve $\widetilde{x'_A x'_B}$ is actually a segment, except possibly in a small neighborhood of x'_A . We are now ready to define the curve γ : first, we define $\tilde{\gamma}$ as the image, under f , of the union between the part of $\widetilde{x'_A x'_B}$ from x_A to P and the segment $P x_B$. Then, since $\tilde{\gamma}$ and $g_3(x_B x_C)$ have of course a part in common, we let γ be a slight modification of $\tilde{\gamma}$ which intersects $g_3(x_B x_C)$ only at $\mathbf{B} = f(x_B)$. Of course, we need to modify $\tilde{\gamma}$ only between $f(P)$ and \mathbf{B} : this modification can be done as the enlargement of Figure 16 (right) shows, and it works exactly as in Step 10 of the proof of Theorem 2.1; in particular, the length of γ is as close as we wish to the length of $\tilde{\gamma}$. By definition, we have

$$\mathcal{H}^1(\tilde{\gamma}) \leq \int_{\widetilde{x'_A x'_B}} |Df| + \int_{x_B x_C} |Df| = \int_{\widetilde{A'B'}} |Dg_2| + \int_{BC} |Dg_1|. \quad (4.30)$$

Thanks to (4.28), calling for a moment M the matrix associated to \mathcal{S} , we know that

$$\int_{BC} |Dg_1| \leq \int_{BC} |Dg_1 - M \cdot \nu| d\mathcal{H}^1 + 2r|M| \leq \left(1020 \frac{\varepsilon_5}{\varepsilon_4} + 11 \frac{\varepsilon_4}{\varepsilon_3} + \frac{2}{5} \varepsilon_5\right) r + \frac{2}{r} \int_{\mathcal{S}} |Df|,$$

and then (4.30) becomes

$$\mathcal{H}^1(\tilde{\gamma}) \leq \int_{\widetilde{A'B'}} |Dg_2| + \left(1020 \frac{\varepsilon_5}{\varepsilon_4} + 11 \frac{\varepsilon_4}{\varepsilon_3} + \frac{2}{5} \varepsilon_5 \right) r + \frac{2}{r} \int_{\mathcal{S}} |Df|. \quad (4.31)$$

Let us now assume, instead, that A is still not vertex of squares in \mathcal{A}_1 or \mathcal{A}_2 , and that B is vertex of some square in \mathcal{A}_1 . We can then define \mathcal{S} and C as before; this time, g_3 in the segment BC is not defined as the reparametrization of the image under f of some segment $x_B x_C$, but as the affine interpolation satisfying $g_3(B) = f(B)$ and $g_3(C) = f(C)$. However, the L^∞ estimate (4.1) and the property (4.2) immediately imply that, exactly as before, the curve γ_0 intersects the image of BC under g_3 (which is the segment $f(B)f(C)$). If we call \mathbf{P} the first point of intersection (starting from $f(x'_A)$), we can argue exactly as before: we define γ as a slight modification of $\tilde{\gamma}$, which is this time the union of the curve γ_0 from $f(x'_A)$ to \mathbf{P} with the segment $\mathbf{PB} = \mathbf{P}f(B)$. In this case, instead of (4.30) we get the estimate

$$\mathcal{H}^1(\tilde{\gamma}) \leq \int_{\widetilde{x'_A x'_B}} |Df| + \overline{f(B)f(C)} = \int_{\widetilde{A'B'}} |Dg_2| + \overline{f(B)f(C)},$$

and since by the L^∞ estimate of Lemma 4.3 we have of course

$$\overline{f(B)f(C)} \leq \frac{2}{r} \int_{\mathcal{S}} |Df|,$$

then also in this case we get the validity of (4.31): of course even the better estimate without the big term in parentheses is true, but it is simpler to consider the same estimate (4.31) in both cases.

Let us finally consider the general case for the segment AB : both the points A and B can be vertices of some square in \mathcal{A}_1 or in \mathcal{A}_2 . Nevertheless, as noticed above, in the cases already considered the path $\tilde{\gamma}$ coincides with γ_0 from the starting point x'_A to almost the final point x'_B , and only a small part near the end has been modified. As a consequence, it is obvious how to deal with the case when both the points A and B are vertices of squares in \mathcal{A}_1 or in \mathcal{A}_2 : we let $\tilde{\gamma}$ be the path which coincides in a large central part with γ_0 , and we apply one of the above described modifications both near the starting point, and near the endpoint. Of course, for a side AB where we have done two modifications, instead of (4.31) we will have

$$\mathcal{H}^1(\tilde{\gamma}) \leq \int_{\widetilde{A'B'}} |Dg_2| + \left(2040 \frac{\varepsilon_5}{\varepsilon_4} + 22 \frac{\varepsilon_4}{\varepsilon_3} + \frac{4}{5} \varepsilon_5 \right) r + \frac{2}{r} \int_{\mathcal{R}_{AB}} |Df|, \quad (4.32)$$

where as usual \mathcal{R}_{AB} is the union of the squares having either A or B as a vertex, which contains both \mathcal{S} and the corresponding square around A . In this way, we have finally defined g_3 in the whole grid $\tilde{\mathcal{G}}$, and by construction it is clear that the map g_3 is injective and coincides with f on $\partial\Omega$. We conclude this step by evaluating the integral of $|Dg_3|$ on the boundary of the different squares: for any square \mathcal{S} of the grid, we call again \mathcal{S}^+ the intersection with Ω of the nine squares around \mathcal{S} .

If $\mathcal{S} \in \mathcal{A}_1$ we do not need any particular estimate, in the next steps we will only need to use that g_3 coincides with $\varphi_{\mathcal{S}}$ on $\partial\mathcal{S}$. If $\mathcal{S} \in \mathcal{A}_2$, instead, we know that g_3 coincides

with g_1 on $\partial\mathcal{S}$, hence we only need to keep in mind the estimate (4.28) already found in Step III, which (adding on the four sides of \mathcal{S}) gives

$$\int_{\partial\mathcal{S}} |Dg_3 - M \cdot \nu| d\mathcal{H}^1 \leq \left(4080 \frac{\varepsilon_5}{\varepsilon_4} + 44 \frac{\varepsilon_4}{\varepsilon_3} + \frac{8}{5} \varepsilon_5 \right) r \leq K \left(\frac{\varepsilon_5}{\varepsilon_4} + \frac{\varepsilon_4}{\varepsilon_3} \right) r \quad \forall \mathcal{S} \in \mathcal{A}_2, \quad (4.33)$$

Finally, if $\mathcal{S} \in \mathcal{A}_3 \cup \mathcal{A}_4$, then $\int_{\partial\tilde{\mathcal{S}}} |Dg_3|$ is the sum of the integrals on the four sides; for each side $\widetilde{A'B'}$, either $g_3 = g_2$, and then of course

$$\int_{\widetilde{A'B'}} |Dg_3| d\mathcal{H}^1 = \int_{\widetilde{A'B'}} |Dg_2| d\mathcal{H}^1, \quad (4.34)$$

or $\int_{\widetilde{A'B'}} |Dg_3| = \mathcal{H}^1(\gamma)$ for some curve $\gamma = \gamma(A, B)$ defined as above. Since the length of γ can be taken as close as we wish to the length of $\tilde{\gamma}$, from (4.32) we derive

$$\int_{\widetilde{A'B'}} |Dg_3| \leq \int_{\widetilde{A'B'}} |Dg_2| + \left(2050 \frac{\varepsilon_5}{\varepsilon_4} + 23 \frac{\varepsilon_4}{\varepsilon_3} + \varepsilon_5 \right) r + \frac{3}{r} \int_{\mathcal{S}^+} |Df|. \quad (4.35)$$

Since (4.34) is stronger than (4.35), we get the validity of (4.35) for any side AB of any square $\mathcal{S} \in \mathcal{A}_3 \cup \mathcal{A}_4$. As a consequence, adding (4.35) on the four sides and keeping in mind (4.29) and the fact that $\varepsilon_1 \ll 1$, we get

$$\begin{aligned} \int_{\partial\tilde{\mathcal{S}}} |Dg_3| d\mathcal{H}^1 &\leq \int_{\partial\tilde{\mathcal{S}}} |Dg_2| d\mathcal{H}^1 + \left(8200 \frac{\varepsilon_5}{\varepsilon_4} + 92 \frac{\varepsilon_4}{\varepsilon_3} + 4\varepsilon_5 \right) r + \frac{12}{r} \int_{\mathcal{S}^+} |Df| d\mathcal{H}^2 \\ &\leq \frac{412}{\varepsilon_1 r} \int_{\mathcal{S}^+} |Df| + \int_{\partial\mathcal{S}^+ \cap \partial\Omega} |Df| + 8200 \left(\frac{\varepsilon_5}{\varepsilon_4} + \frac{\varepsilon_4}{\varepsilon_3} + \varepsilon_5 \right) r \quad \forall \mathcal{S} \in \mathcal{A}_3 \cup \mathcal{A}_4. \end{aligned} \quad (4.36)$$

Step VI. The piecewise linear function $g_4 : \tilde{\mathcal{G}} \rightarrow \mathbb{R}^2$.

In this step, we want to define a piecewise linear function $g_4 : \tilde{\mathcal{G}} \rightarrow \mathbb{R}^2$: this one will finally be the correct map on the grid $\tilde{\mathcal{G}}$, in the sense that our approximating function f_ε will coincide with g_4 on $\tilde{\mathcal{G}}$. To do so, it is enough to apply Proposition 4.17 to the map g_3 and call $g_4 = \hat{g}$ the resulting map. As a consequence, g_4 is a piecewise linear map on $\tilde{\mathcal{G}}$, which coincides with f on $\partial\Omega$. Moreover, since for every side AB of the grid the map g_4 on $\widetilde{A'B'}$ is an interpolation of values of g_3 on $\widetilde{A'B'}$, of course we still have $g_4 = \varphi_{\mathcal{S}}$ on the boundary of every square $\mathcal{S} \in \mathcal{A}_1$. Moreover, (4.17) and (4.18) imply that the estimates (4.33) and (4.36) are valid also with g_4 in place of g_3 .

Step VII. Definition of the approximating function $f_\varepsilon : \Omega \rightarrow \mathbb{R}^2$.

We are almost at the end of the proof, since we can finally define the required function f_ε . We set $f_\varepsilon = g_4$ on the grid $\tilde{\mathcal{G}}$, hence in particular $f_\varepsilon = f$ on $\partial\Omega$; by construction, f_ε is injective and piecewise linear on $\tilde{\mathcal{G}}$. To keep the injectivity, it is then enough to extend f_ε in the interior of any square $\tilde{\mathcal{S}}$, in such a way that f_ε remains continuous and injective on it. We will argue differently on the different squares.

If $\mathcal{S} \in \mathcal{A}_1$, then we know that $f_\varepsilon = \varphi_{\mathcal{S}}$ on $\partial\mathcal{S}$, so we extend $f_\varepsilon = \varphi_{\mathcal{S}}$ on the whole square \mathcal{S} : this map is continuous and injective by construction, and by (4.1) of Lemma 4.3 we know that

$$\int_{\mathcal{S}} |Df_\varepsilon - Df| \leq r^2 \varepsilon_5 \quad \forall \mathcal{S} \in \mathcal{A}_1. \quad (4.37)$$

Let us now consider a square $\mathcal{S} \in \mathcal{A}_2$: in this case, we want to apply Theorem 3.1 to the function $\varphi = g_4$ on $\partial\mathcal{S}$. Keeping in mind the generalization of Theorem 3.1 observed in Remark 3.2, and calling f_ε the obtained extension on \mathcal{S} , we get

$$\int_{\mathcal{S}} |Df_\varepsilon(x) - M| dx \leq Kr \int_{\partial\mathcal{S}} |Dg_4(t) - M \cdot \tau(t)| d\mathcal{H}^1(t) \quad (4.38)$$

as soon as the estimate

$$\int_{\partial\mathcal{S}} |Dg_4(t) - M \cdot \tau(t)| d\mathcal{H}^1(t) < r\delta_{\text{MAX}}\|M\|$$

holds. Thanks to (4.33), which holds also with g_4 in place of g_3 as noticed in Step VI, and recalling that $\|M\| > \varepsilon_3$ by definition of \mathcal{A}_2 and that δ_{MAX} is a small purely geometric constant, the latter estimate is true thanks to (4.22). As a consequence, recalling that \mathcal{S} is a Lebesgue square with matrix M and constant $\hat{\delta}$, that $\hat{\delta} \ll \varepsilon_5$ by (4.23) and by definition of $\bar{\delta}$, and using also that $\varepsilon_4 \leq 1$, from (4.38) and (4.33) we get

$$\int_{\mathcal{S}} |Df_\varepsilon - Df| \leq \int_{\mathcal{S}} |Df_\varepsilon - M| + \int_{\mathcal{S}} |Df - M| \leq Kr^2 \left(\frac{\varepsilon_5}{\varepsilon_4} + \frac{\varepsilon_4}{\varepsilon_3} \right) \quad \forall \mathcal{S} \in \mathcal{A}_2, \quad (4.39)$$

where K is as always a purely geometric constant.

Finally, let \mathcal{S} be a square in $\mathcal{A}_3 \cup \mathcal{A}_4$. This time, since g_4 is piecewise linear on $\partial\tilde{\mathcal{S}}$ and $\tilde{\mathcal{S}}$ is (a 2-biLipschitz copy of) a square of side $2r$, we let f_ε on $\tilde{\mathcal{S}}$ be the extension of g_4 given by Theorem 2.1, keeping in mind also the generalization of Remark 2.2. The estimate (2.13), together with (4.36), which is valid also with g_4 in place of g_3 by Step VI, gives then

$$\int_{\tilde{\mathcal{S}}} |Df_\varepsilon| \leq Kr \int_{\partial\tilde{\mathcal{S}}} |Dg_4| d\mathcal{H}^1 \leq \frac{K}{\varepsilon_1} \int_{\mathcal{S}^+} |Df| + Kr \int_{\partial\mathcal{S}^+ \cap \partial\Omega} |Df| d\mathcal{H}^1 + K \left(\frac{\varepsilon_5}{\varepsilon_4} + \frac{\varepsilon_4}{\varepsilon_3} + \varepsilon_5 \right) r^2.$$

Since

$$\int_{\tilde{\mathcal{S}}} |Df_\varepsilon - Df| \leq \int_{\tilde{\mathcal{S}}} |Df_\varepsilon| + \int_{\tilde{\mathcal{S}}} |Df| \leq \int_{\tilde{\mathcal{S}}} |Df_\varepsilon| + \int_{\mathcal{S}^+} |Df|,$$

since $\varepsilon_4 \leq 1$, and since K is a purely geometric constant while $\varepsilon_1 \leq 1$, we deduce

$$\int_{\tilde{\mathcal{S}}} |Df_\varepsilon - Df| \leq \frac{K}{\varepsilon_1} \int_{\mathcal{S}^+} |Df| + Kr \int_{\partial\mathcal{S}^+ \cap \partial\Omega} |Df| d\mathcal{H}^1 + K \left(\frac{\varepsilon_5}{\varepsilon_4} + \frac{\varepsilon_4}{\varepsilon_3} \right) r^2 \quad \forall \mathcal{S} \in \mathcal{A}_4. \quad (4.40)$$

Notice that the same estimate holds true also for $\mathcal{S} \in \mathcal{A}_3$; nevertheless, in this case we can say even something more. Indeed, since \mathcal{S} is a Lebesgue square with matrix $M = 0$, by Definition 4.1 we know

$$\int_{\mathcal{S}^+} |Df| = \int_{\mathcal{S}^+} |Df - M| \leq 36r^2\hat{\delta} \leq 36r^2\varepsilon_5.$$

As a consequence, for squares in \mathcal{A}_3 we can deduce from (4.40)

$$\int_{\tilde{\mathcal{S}}} |Df_\varepsilon - Df| \leq Kr \int_{\partial\mathcal{S}^+ \cap \partial\Omega} |Df| d\mathcal{H}^1 + K \left(\frac{\varepsilon_5}{\varepsilon_4} + \frac{\varepsilon_4}{\varepsilon_3} + \frac{\varepsilon_5}{\varepsilon_1} \right) r^2 \quad \forall \mathcal{S} \in \mathcal{A}_3. \quad (4.41)$$

Step VIII. Conclusion.

We are now ready to conclude. Indeed, by construction f_ε is a finitely piecewise affine homeomorphism which coincides with f on $\partial\Omega$. Moreover, it is immediate by construction that $\|f_\varepsilon - f\|_{L^\infty}$ and $\|f_\varepsilon - f\|_{L^1}$ are as small as we wish (it is enough to have chosen at

the beginning r small enough); as a consequence, we can assume that they are smaller than $\varepsilon/4$ each. As a consequence, to get (4.19) and conclude, we only have to check that

$$\|Df_\varepsilon - Df\|_{L^1} < \frac{\varepsilon}{2}. \quad (4.42)$$

Thanks to (4.37), (4.39), (4.40) and (4.41), and calling, for $j = 1, 2, 3, 4$,

$$\Omega_j = \cup \left\{ \tilde{\mathcal{S}}_i : \mathcal{S}_i \in \mathcal{A}_j \right\},$$

we have

$$\begin{aligned} \int_{\Omega} |Df_\varepsilon - Df| &= \int_{\Omega_1} |Df_\varepsilon - Df| + \int_{\Omega_2} |Df_\varepsilon - Df| + \int_{\Omega_3} |Df_\varepsilon - Df| + \int_{\Omega_4} |Df_\varepsilon - Df| \\ &\leq K \left(\frac{\varepsilon_5}{\varepsilon_4} + \frac{\varepsilon_4}{\varepsilon_3} + \frac{\varepsilon_5}{\varepsilon_1} \right) |\Omega| + \sum_{i: \mathcal{S}_i \in \mathcal{A}_4} \frac{K}{\varepsilon_1} \int_{\mathcal{S}_i^+} |Df| + \sum_{i: \mathcal{S}_i \in \mathcal{A}_3 \cup \mathcal{A}_4} Kr \int_{\partial \mathcal{S}_i^+ \cap \partial \Omega} |Df|. \end{aligned} \quad (4.43)$$

Let us now recall that, for each square \mathcal{S}_i of the grid, the set \mathcal{S}_i^+ is the union of the nine squares around it (to be precise, to those which belong to Ω). As a consequence, calling for brevity $\mathcal{A}_4^+ = \cup_{i: \mathcal{S}_i \in \mathcal{A}_4} \mathcal{S}_i^+$, also recalling (4.25) we have $|\mathcal{A}_4^+| \leq 9 |\cup_{i: \mathcal{S}_i \in \mathcal{A}_4} \mathcal{S}_i| < \varepsilon_2$. Thus, by (4.20) we can write

$$\sum_{i: \mathcal{S}_i \in \mathcal{A}_4} \frac{K}{\varepsilon_1} \int_{\mathcal{S}_i^+} |Df| \leq 9 \frac{K}{\varepsilon_1} \int_{\mathcal{A}_4^+} |Df| \leq 9 \frac{K}{\varepsilon_1} \frac{\varepsilon_1 \varepsilon}{54K} = \frac{\varepsilon}{6}.$$

Analogously, each side in $\partial \Omega$ can belong to $\partial \mathcal{S}_i^+$ for at most eleven different indices i with $\mathcal{S}_i \in \mathcal{A}_3 \cup \mathcal{A}_4$, hence by (4.24) we get

$$\sum_{i: \mathcal{S}_i \in \mathcal{A}_3 \cup \mathcal{A}_4} Kr \int_{\partial \mathcal{S}_i^+ \cap \partial \Omega} |Df| \leq 11Kr \int_{\partial \Omega} |Df| = 11KrP(f(\Omega)) \leq \frac{\varepsilon}{6}.$$

Finally, by (4.22) we have

$$K \left(\frac{\varepsilon_5}{\varepsilon_4} + \frac{\varepsilon_4}{\varepsilon_3} + \frac{\varepsilon_5}{\varepsilon_1} \right) |\Omega| \leq \frac{\varepsilon}{6}.$$

Inserting the last three estimates inside (4.43) we get (4.42), so the proof is concluded. \square

The above proposition shows that, under stronger assumptions than in Theorem 1.1, we can obtain something better than what claimed in Theorem 1.1. Indeed, if Ω is an r -set and f is piecewise linear on $\partial \Omega$, then we do not only get simply a *countably* piecewise affine approximation, but a much better *finitely* piecewise affine one. We can now give the sharpest possible result of this approximation, that is, we can prove the existence of a finitely piecewise affine approximation with the weakest possible assumptions.

Theorem 4.20. *Let $\Omega \subseteq \mathbb{R}^2$ be a polygon and let $f \in W^{1,1}(\Omega, \mathbb{R}^2)$ be a homeomorphism, continuous up to the boundary and such that f is piecewise linear on $\partial \Omega$. Then, for every $\varepsilon > 0$ there exists a finitely piecewise affine homeomorphism $f_\varepsilon : \Omega \rightarrow \mathbb{R}^2$ such that $\|f_\varepsilon - f\|_{W^{1,1}} + \|f_\varepsilon - f\|_{L^\infty} < \varepsilon$, and $f_\varepsilon = f$ on $\partial \Omega$.*

Proof. Since Ω is a polygon, there exists an r -set $\widehat{\Omega}$ and a finitely piecewise affine homeomorphism $\Phi : \Omega \rightarrow \widehat{\Omega}$. There exists then some constant $H = H(\Omega)$ such that

$$|D\Phi(x)| \leq H, \quad \det D\Phi(x) \geq \frac{1}{H},$$

for almost every $x \in \Omega$. Let us then define $\hat{f} : \widehat{\Omega} \rightarrow \mathbb{R}^2$ as $\hat{f} = f \circ \Phi^{-1}$; by construction, \hat{f} belongs to $W^{1,1}(\widehat{\Omega}, \mathbb{R}^2)$, and it is continuous up to $\partial\widehat{\Omega}$ and piecewise linear there. As a consequence, we can apply Proposition 4.19 to \hat{f} in $\widehat{\Omega}$, finding a finitely piecewise affine homeomorphism $\hat{f}_\varepsilon : \widehat{\Omega} \rightarrow \mathbb{R}^2$ which coincides with \hat{f} on $\partial\widehat{\Omega}$ and such that

$$\|\hat{f}_\varepsilon - \hat{f}\|_{W^{1,1}(\widehat{\Omega})} + \|\hat{f}_\varepsilon - \hat{f}\|_{L^\infty(\widehat{\Omega})} \leq \frac{\varepsilon}{H^2}. \quad (4.44)$$

We can then define $f_\varepsilon = \hat{f}_\varepsilon \circ \Phi$: this is a finitely piecewise affine homeomorphism, of course it coincides with g on ∂f , and we have

$$\|f_\varepsilon - f\|_{L^\infty(\Omega)} = \|\hat{f}_\varepsilon - \hat{f}\|_{L^\infty(\widehat{\Omega})}.$$

By a simple change of variable argument, we obtain then

$$\begin{aligned} \|f_\varepsilon - f\|_{L^1(\Omega)} &= \int_{\Omega} |f_\varepsilon(x) - f(x)| dx = \int_{\Omega} |\hat{f}_\varepsilon(\Phi(x)) - \hat{f}(\Phi(x))| dx \\ &= \int_{\widehat{\Omega}} \frac{|\hat{f}_\varepsilon(y) - \hat{f}(y)|}{|\det D\Phi(\Phi^{-1}(y))|} dy \leq H \| \hat{f}_\varepsilon - \hat{f} \|_{L^1(\widehat{\Omega})}, \end{aligned}$$

and similarly

$$\begin{aligned} \|Df_\varepsilon - Df\|_{L^1(\Omega)} &= \int_{\Omega} |D(\hat{f}_\varepsilon \circ \Phi)(x) - D(\hat{f} \circ \Phi)(x)| dx \\ &= \int_{\Omega} \left| \left(D\hat{f}_\varepsilon(\Phi(x)) - D\hat{f}(\Phi(x)) \right) \cdot D\Phi(x) \right| dx \leq H^2 \|D\hat{f}_\varepsilon - D\hat{f}\|_{L^1(\widehat{\Omega})}. \end{aligned}$$

Inserting the last three estimates in (4.44), we conclude that f_ε is the desired approximation. \square

Remark 4.21. *It is immediate to observe that the assumption of Theorem 4.20 are sharp. Indeed, assume that a homeomorphism $f \in W^{1,1}(\Omega, \mathbb{R}^2)$ admits a finitely piecewise affine approximation f_ε . Since f_ε is finitely piecewise affine, it is defined on a polygon, hence Ω must be a polygon. Similarly, f_ε is piecewise linear on $\partial\Omega$ by definition, and since $f_\varepsilon = f$ on $\partial\Omega$ the same must be true for f .*

We are now in position to give the proof of our main result, Theorem 1.1.

Proof of Theorem 1.1. We will argue in a way quite similar to Proposition 4.19, we only need some additional care to reach the boundary of Ω . First of all, we look for a piecewise affine approximation, the smooth one will be found at the end.

We again start by selecting the small constants ε_i : first of all, we let ε_1 be a small geometric constant, say $\varepsilon_1 = 1/10$. Then, since $f \in W^{1,1}(\Omega)$, we can select a constant ε_2 such that

$$\int_A |Df| \leq \frac{\varepsilon\varepsilon_1}{72K} \quad \forall A \subseteq \Omega : |A| \leq \varepsilon_2. \quad (4.45)$$

The next step is to write Ω as a countable union of r_n -sets. More precisely, we can take a sequence of constants $r_n \rightarrow 0$ and a sequence of disjoint open sets $\Omega_n \subset \subset \Omega$, in such a way that each Ω_n is an r_n -set, the union of the closures $\overline{\Omega_n}$ is the whole Ω , and for each $n \in \mathbb{N}$ we can divide the boundary of Ω_n in two disjoint parts $\partial\Omega_n = \partial^-\Omega_n \cup \partial^+\Omega_n$, being

$$\partial^-\Omega_1 = \emptyset, \quad \partial^+\Omega_n = \partial^-\Omega_{n+1} \quad \forall n \in \mathbb{N}.$$

Since $f \in W^{1,1}(\Omega)$, we can select these sequences in such a way that

$$\int_{\partial\Omega_n} |Df| = P(f(\Omega_n)) < +\infty \quad \forall n \in \mathbb{N}, \quad (4.46)$$

and we can also take Ω_1 large enough so that that

$$\int_{\Omega \setminus \Omega_1} |Df| \leq \frac{\varepsilon \varepsilon_1}{72K}. \quad (4.47)$$

Naively speaking, the idea now is to try to work on each set Ω_n separately. However, since f is not necessarily piecewise linear on the boundaries of the sets Ω_n , we cannot simply rely on Proposition 4.19 for each Ω_n ; moreover, since Ω has not necessarily finite area, estimates like (4.37) or (4.39), where the area of a square appears, are not acceptable because they could give an infinite contribution after adding.

Let us now concentrate on Ω_1 in order to select the constants ε_3 , ε_4 and ε_5 : indeed, we will use these constants only inside Ω_1 . Arguing exactly as in the proof of Proposition 4.19, we first let ε_3 be a constant such that

$$\left| \left\{ x \in \Omega_1 : 0 < |Df(x)| < \varepsilon_3 \text{ or } |Df(x)| > \frac{1}{\varepsilon_3} \text{ or } 0 < \det(Df(x)) < \varepsilon_3 \right\} \right| < \frac{\varepsilon_2}{45}, \quad (4.48)$$

then we let again

$$\mathcal{M}^+ = \left\{ \varepsilon_3 < \|M\| < \frac{1}{\varepsilon_3}, \det M > \varepsilon_3 \right\}, \quad \mathcal{M}^0 = \left\{ \varepsilon_3 < \|M\| < \frac{1}{\varepsilon_3}, \det M = 0 \right\},$$

then we let $\varepsilon_5 \ll \varepsilon_4 \ll \varepsilon_3$ be such that

$$\frac{\varepsilon_5}{\varepsilon_4} + \frac{\varepsilon_4}{\varepsilon_3} + \frac{\varepsilon_5}{\varepsilon_1} \leq \frac{\varepsilon}{8K|\Omega_1|}, \quad \varepsilon_5 \ll \varepsilon_4 \varepsilon_3, \quad \varepsilon_4 \ll \varepsilon_3^2, \quad (4.49)$$

and finally we let $\hat{\delta} = \hat{\delta}(\varepsilon_3, \varepsilon_5)$ be given by

$$\hat{\delta} = \min \{ \bar{\delta}(M, \varepsilon_5), M \in \mathcal{M}^+ \cup \mathcal{M}^0 \},$$

where $\bar{\delta}$ are the constants of Lemma 4.3. The last thing we have to fix, is the final value of the constants r_n : indeed, each Ω_n is an r_n -set, but then it can be regarded as a r_n/H_n -set for every constants $H_n \in \mathbb{N}$. As a consequence, we can now decrease the values of r_n (without changing the sets Ω_n , of course); in particular, also thanks to (4.46), we can assume that r_1 is so small that

$$r_1 P(\Omega_1) + \left| \left\{ x \in \Omega_1 : \bar{r}(x, \hat{\delta}) \leq r_1 \right\} \right| < \frac{\varepsilon_2}{180}, \quad r_1 P(f(\Omega_1)) \leq \frac{\varepsilon}{66K}, \quad (4.50)$$

being $\bar{r}(x, \delta)$ the constants of Lemma 4.6, while any other r_n is so small that

$$r_n \leq r_{n-1}, \quad r_n \int_{\partial\Omega_n} |Df| \leq \frac{\varepsilon}{88K \cdot 2^n}, \quad r_n \ll \text{dist}(\partial^-\Omega_n, \partial^+\Omega_n), \quad \forall n \geq 2; \quad (4.51)$$

notice that the last requirement basically means that the “thickness” of any Ω_n is of several squares.

Having fixed the sets Ω_n and the corresponding r_n , any Ω_n is divided in a finite union of squares, all with side $2r_n$. Let us enumerate them by saying that the squares of the grid of Ω_n are \mathcal{S}_i^n with $1 \leq i \leq N(n)$; then, we subdivide the squares of Ω_1 in four groups, namely, $\mathcal{A}_1^1, \mathcal{A}_2^1, \mathcal{A}_3^1, \mathcal{A}_4^1$, as follows,

$$\begin{aligned} \mathcal{A}_1^1 &= \left\{ \mathcal{S}_i^1 \subset\subset \Omega_1 : \mathcal{S}_i^1 \text{ is a Lebesgue square with matrix } M_i^1 \in \mathcal{M}^+ \text{ and constant } \hat{\delta} \right\}, \\ \mathcal{A}_2^1 &= \left\{ \mathcal{S}_i^1 \subset\subset \Omega_1 : \mathcal{S}_i^1 \text{ is a Lebesgue square with matrix } M_i^1 \in \mathcal{M}^0 \text{ and constant } \hat{\delta} \right\}, \\ \mathcal{A}_3^1 &= \left\{ \mathcal{S}_i^1 \subset\subset \Omega_1 : \mathcal{S}_i^1 \text{ is a Lebesgue square with matrix } M_i^1 = 0 \text{ and constant } \hat{\delta} \right\}, \\ \mathcal{A}_4^1 &= \left\{ \mathcal{S}_i^1 : \mathcal{S}_i^1 \notin \mathcal{A}_1^1 \cup \mathcal{A}_2^1 \cup \mathcal{A}_3^1 \right\}. \end{aligned}$$

We immediately record that, exactly as in Step I of the proof of Proposition 4.19, from (4.48) and (4.50) it follows that

$$\left| \bigcup \{ \mathcal{S}_i \in \mathcal{A}_4^1 \} \right| \leq r_1 P(\Omega_1) + 4 \left(\left| \{ x \in \Omega_1 : \bar{r}(x, \hat{\delta}) \leq r_1 \} \right| + \frac{\varepsilon_2}{45} \right) < \frac{\varepsilon_2}{9}. \quad (4.52)$$

For any $n \geq 2$, instead, we simply let \mathcal{A}_4^n be the collection of all the squares \mathcal{S}_i^n of the grid of Ω_n , while $\mathcal{A}_1^n, \mathcal{A}_2^n$ and \mathcal{A}_3^n are empty.

Let us notice that the only assumption which is true for Ω in Proposition 4.19 and may fail now for the generic Ω_n is the following one: f is assumed to be piecewise linear on $\partial\Omega$ in Proposition 4.19, while f needs now not to be piecewise linear on the boundaries $\partial\Omega_n$; on the other hand, this assumption has been used only in Step VI of the proof of Proposition 4.19. As a consequence, we can repeat *verbatim* all the arguments of Steps II, III, IV and V of that proof for Ω_1 : so, we discovery first that squares in \mathcal{A}_1^1 and \mathcal{A}_2^1 can never touch, then we define a temptative modified grid $\tilde{\mathcal{G}}_1^1$ with a function $g_1^1 : \tilde{\mathcal{G}}_1^1 \rightarrow \mathbb{R}^2$, then the correct modified grid $\tilde{\mathcal{G}}^1$ with the function $g_2^1 : \tilde{\mathcal{G}}^1 \rightarrow \mathbb{R}^2$, and finally the function $g_3^1 : \tilde{\mathcal{G}}^1 \rightarrow \mathbb{R}^2$. By definition, the function g_3^1 is injective and coincides with f on $\partial\Omega_1$ (this was explicitly decided in Definition 4.15), and moreover we have the estimates

$$\begin{aligned} g_3^1 &= \varphi_{\mathcal{S}} \text{ on } \partial\mathcal{S} && \forall \mathcal{S} \in \mathcal{A}_1^1, \\ \int_{\partial\mathcal{S}} |Dg_3^1 - M \cdot \nu| d\mathcal{H}^1 &\leq K \left(\frac{\varepsilon_5}{\varepsilon_4} + \frac{\varepsilon_4}{\varepsilon_3} \right) r_1 && \forall \mathcal{S} \in \mathcal{A}_2^1, \\ \int_{\partial\tilde{\mathcal{S}}} |Dg_3^1| d\mathcal{H}^1 &\leq \int_{\partial\mathcal{S} + \partial\Omega_1} |Df| + K \left(\frac{\varepsilon_5}{\varepsilon_4} + \frac{\varepsilon_4}{\varepsilon_3} + \frac{\varepsilon_5}{\varepsilon_1} \right) r_1 && \forall \mathcal{S} \in \mathcal{A}_3^1, \\ \int_{\partial\tilde{\mathcal{S}}} |Dg_3^1| d\mathcal{H}^1 &\leq \frac{K}{\varepsilon_1 r_1} \int_{\mathcal{S}^+} |Df| + \int_{\partial\mathcal{S} + \partial\Omega_1} |Df| + K \left(\frac{\varepsilon_5}{\varepsilon_4} + \frac{\varepsilon_4}{\varepsilon_3} \right) r_1 && \forall \mathcal{S} \in \mathcal{A}_4^1, \end{aligned} \quad (4.53)$$

where K is a purely geometric constant (it suffices to take $K = 9000$ here).

Let us now consider Ω_n for any $n \geq 2$. In this case, the situation is much simpler than in Proposition 4.19: indeed, by definition we only have squares in \mathcal{A}_4^n , so we do not need the arguments of Steps II and III and we can directly start with the analogous of Step IV, which immediately gives us a function $g_2^n : \tilde{\mathcal{G}}_n \rightarrow \mathbb{R}^2$ satisfying the analogous

of (4.29), that is,

$$\int_{\partial\tilde{\mathcal{S}}} |Dg_2^n| d\mathcal{H}^1 \leq \frac{400}{\varepsilon_1 r_n} \int_{\mathcal{S}^+} |Df| d\mathcal{H}^2 + \int_{\partial\mathcal{S}^+ \cap \partial\Omega_n} |Df| d\mathcal{H}^1 \quad \forall \mathcal{S} \in \mathcal{A}_4^n. \quad (4.54)$$

Again since there are no squares in \mathcal{A}_1^n , \mathcal{A}_2^n and \mathcal{A}_3^n , we do not even need the argument of Step V, and we can simply set the function $g_3^n = g_2^n$.

Let us now observe that every function g_3^n coincides with f , by construction and by Definition 4.15, on $\partial\Omega_n$. As a consequence, if we call $\tilde{\mathcal{G}}$ the union of all the grids $\tilde{\mathcal{G}}^n$ and we set $g_3 : \tilde{\mathcal{G}} \rightarrow \mathbb{R}^2$ as $g_3 = g_3^n$ on each $\tilde{\mathcal{G}}^n$, then the resulting function g_3 is also injective.

The last thing we have to do, before having the right of treating each Ω_n separately, is to modify g_3 so to become piecewise linear on the boundary of each Ω_n ; we will do that by applying Proposition 4.18. More precisely, for every $j \geq 2$ we define \mathcal{C}^j the union of all the segments of $\tilde{\mathcal{G}}^{j-1}$ and $\tilde{\mathcal{G}}^j$, and $\mathcal{C}_0^j = \partial^-\Omega_j$. We apply then Proposition 4.18 with $\mathcal{C} = \mathcal{C}^j$, $\mathcal{C}_0 = \mathcal{C}_0^j$, $\eta \ll r_j \leq r_{j-1}$, and with the function $g = g_3$; thus, we get a function \hat{g}_j , piecewise linear on $\partial^-\Omega_j$, which coincides with g (hence, with g_3) out of a η -neighborhood of $\partial^-\Omega_j$. As a consequence, the function \hat{g}_j is different from g_3 only on the boundary of squares which meet $\partial^-\Omega_j$. Then, let us define the function $\hat{g}_3 : \tilde{\mathcal{G}} \rightarrow \mathbb{R}^2$ as $\hat{g}_3 = \hat{g}_j$ on the boundary of squares touching $\partial^-\Omega_j$, and $\hat{g}_3 = g_3$ on the boundaries of all the other squares. By construction and by Proposition 4.18, the function \hat{g}_3 is injective, piecewise linear on each $\partial\Omega_j$, and the estimates (4.53) and (4.54) are valid with \hat{g}_3 in place of g_3 .

Now, for every $n \in \mathbb{N}$ we apply Proposition 4.17 to the set Ω_n with the function \hat{g}_3 on $\tilde{\mathcal{G}}^n$, and we get a new function g_4^n , piecewise linear on $\tilde{\mathcal{G}}^n$ and coinciding with \hat{g}_3 on $\partial\Omega_n$. Finally, we define $g_4 : \tilde{\mathcal{G}} \rightarrow \mathbb{R}^2$ as $g_4 = g_4^n$ on every $\tilde{\mathcal{G}}^n$: also this function satisfies (4.53) and (4.54), and it is piecewise linear on the boundary of each square of any grid.

We are now ready to define the piecewise affine approximation f_ε : indeed, for every $n \in \mathbb{N}$, the function g_4 on $\tilde{\mathcal{G}}^n$ is injective and piecewise linear on the boundary. Exactly as in Step VII of the proof of Proposition 4.19, we can then define f_ε^n on Ω_n , which is a finitely piecewise affine function coinciding with g_4 on $\partial\Omega_n$, and then set $f_\varepsilon : \Omega \rightarrow \mathbb{R}^2$ as the function which coincides with f_ε^n on every Ω_n . This function f_ε is by construction a countably piecewise affine homeomorphism, and also locally finitely piecewise affine; moreover, it is clear that $\|f - f_\varepsilon\|_{L^\infty(\Omega_n)}$ and $\|f - f_\varepsilon\|_{L^1(\Omega_n)}$ are as small as we wish as soon as the constants r_n have been chosen small enough: in particular, we can think that both them are smaller than $\varepsilon/4$. In addition, by construction we have

$$f\left(\bigcup_{j=1}^{n-1} \Omega_j\right) \subseteq f_\varepsilon\left(\bigcup_{j=1}^n \Omega_j\right) \subseteq f\left(\bigcup_{j=1}^{n+1} \Omega_j\right),$$

and then we immediately get that $f_\varepsilon(\Omega) = f(\Omega)$, and also that $f_\varepsilon = f$ on $\partial\Omega$ whenever f is continuous up to $\partial\Omega$. As a consequence, to conclude the proof of Theorem 1.1 for what concerns the piecewise affine approximation, we just have to check that

$$\|Df_\varepsilon - Df\|_{L^1(\Omega)} \leq \frac{\varepsilon}{2}. \quad (4.55)$$

This will be obtained arguing almost exactly as in Steps VII and VIII of the proof of Proposition 4.19. More precisely, let us start with Ω_1 : repeating *verbatim* the arguments

bringing to (4.37), (4.39), (4.40) and (4.41), this time from (4.53) we get

$$\begin{aligned}
 \int_{\mathcal{S}} |Df_\varepsilon - Df| &\leq r_1^2 \varepsilon_5 & \forall \mathcal{S} \in \mathcal{A}_1^1, \\
 \int_{\mathcal{S}} |Df_\varepsilon - Df| &\leq Kr_1^2 \left(\frac{\varepsilon_5}{\varepsilon_4} + \frac{\varepsilon_4}{\varepsilon_3} \right) & \forall \mathcal{S} \in \mathcal{A}_2^1, \\
 \int_{\tilde{\mathcal{S}}} |Df_\varepsilon - Df| &\leq Kr_1 \int_{\partial\mathcal{S}^+ \cap \partial\Omega_1} |Df| d\mathcal{H}^1 + K \left(\frac{\varepsilon_5}{\varepsilon_4} + \frac{\varepsilon_4}{\varepsilon_3} + \frac{\varepsilon_5}{\varepsilon_1} \right) r_1^2 & \forall \mathcal{S} \in \mathcal{A}_3^1, \\
 \int_{\tilde{\mathcal{S}}} |Df_\varepsilon - Df| &\leq \frac{K}{\varepsilon_1} \int_{\mathcal{S}^+} |Df| + Kr_1 \int_{\partial\mathcal{S}^+ \cap \partial\Omega_1} |Df| d\mathcal{H}^1 + K \left(\frac{\varepsilon_5}{\varepsilon_4} + \frac{\varepsilon_4}{\varepsilon_3} \right) r_1^2 & \forall \mathcal{S} \in \mathcal{A}_4^1.
 \end{aligned} \tag{4.56}$$

Instead, for every $n \geq 2$, we apply Theorem 2.1 –recalling also Remark 2.2– to the generic $\tilde{\mathcal{S}}$, which is a 2-biLipschitz copy of a square of side $2r$: then, also from (4.54), we get

$$\begin{aligned}
 \int_{\tilde{\mathcal{S}}} |Df_\varepsilon - Df| &\leq \int_{\tilde{\mathcal{S}}} |Df| + \int_{\tilde{\mathcal{S}}} |Df_\varepsilon| \leq \int_{\mathcal{S}^+} |Df| + Kr_n \int_{\partial\tilde{\mathcal{S}}} |Dg_4| \\
 &\leq \frac{K}{\varepsilon_1} \int_{\mathcal{S}^+} |Df| + Kr_n \int_{\partial\mathcal{S}^+ \cap \partial\Omega_n} |Df| & \forall \mathcal{S} \in \mathcal{A}_4^n.
 \end{aligned} \tag{4.57}$$

Notice that this estimate is better than the corresponding one for Proposition 4.19, namely, (4.40): indeed, there we had also the additional term $K(\varepsilon_5/\varepsilon_4 + \varepsilon_4/\varepsilon_3)r^2$, which now would be quite a problem since in principle Ω may have infinite area. The reason why we do not have this term now, is that it was coming from the interaction between squares in \mathcal{A}_4 touching squares in \mathcal{A}_1 or \mathcal{A}_2 , while now in Ω_n we only have squares in \mathcal{A}_4^n .

The very same argument as in Proposition 4.19 implies again that every square can belong to \mathcal{S}^+ at most for 9 different squares \mathcal{S} , and that every side of some $\partial\Omega_n$ can belong to $\partial\mathcal{S}^+ \cap \partial\Omega_n$ at most for 11 different squares of the grid of Ω_n ; as a consequence, adding (4.56) and (4.57) for all the squares of the different grid, we find

$$\begin{aligned}
 \int_{\Omega} |Df_\varepsilon - Df| &\leq K \left(\frac{\varepsilon_5}{\varepsilon_4} + \frac{\varepsilon_4}{\varepsilon_3} + \frac{\varepsilon_5}{\varepsilon_1} \right) |\Omega_1| + 9 \frac{K}{\varepsilon_1} \int_{\mathcal{A}_4^{1,+}} |Df| \\
 &\quad + 9 \frac{K}{\varepsilon_1} \int_{\Omega \setminus \Omega_1} |Df| + 11K \sum_{n \in \mathbb{N}} r_n \int_{\partial\Omega_n} |Df| \\
 &\leq \frac{\varepsilon}{8} + \frac{\varepsilon}{8} + \frac{\varepsilon}{8} + 11K \sum_{n \in \mathbb{N}} \frac{\varepsilon}{88K \cdot 2^n} \leq \frac{\varepsilon}{2},
 \end{aligned}$$

where $\mathcal{A}_4^{1,+}$ is the union of the sets \mathcal{S}^+ for all the squares $\mathcal{S} \in \mathcal{A}_4^1$, so that by (4.52) we have $|\mathcal{A}_4^{1,+}| \leq 9 \cup \{\mathcal{S}_i \in \mathcal{A}_4^1\} < \varepsilon_2$, and where we have used (4.49), (4.45), (4.47) and (4.51). As a consequence, we have established the validity of (4.55), so the proof of the existence of the required piecewise affine approximation is concluded.

As already remarked, once the piecewise affine approximation is found, the existence of the required approximating diffeomorphisms is exactly the content of [27, Theorem A], so we have finished our proof. \square

Acknowledgment. Part of this research was done when the first author was visiting University of Erlangen, and while the second author was visiting Charles University. They wish to thank both departments for hospitality. Both authors were supported through the ERC CZ grant LL1203 of the Czech Ministry of Education and the ERC St.G. AnOptSetCon of the European Community.

REFERENCES

- [1] E. Acerbi and N. Fusco, Semicontinuity problems in the calculus of variations, *Arch. Rat. Mech. Anal.* **86** (1984), 125–145.
- [2] J.M. Ball, Convexity conditions and existence theorems in nonlinear elasticity, *Arch. Rat. Mech. Anal.* **63** (1977), 337–403.
- [3] J.M. Ball, Discontinuous equilibrium solutions and cavitation in nonlinear elasticity, *Philos. Trans. R. Soc. Lond. A* **306** n.1496 (1982), 557–611.
- [4] J.M. Ball, *Singularities and computation of minimizers for variational problems*, Foundations of computational mathematics (Oxford, 1999), 1–20, London Math. Soc. Lecture Note Ser., 284, Cambridge Univ. Press, Cambridge, 2001.
- [5] J.M. Ball, *Progress and puzzles in Nonlinear Elasticity*, Proceedings of course on Poly-, Quasi- and Rank-One Convexity in Applied Mechanics, CISM, Udine, to appear.
- [6] J.C. Bellido and C. Mora-Corral, Approximation of Hölder continuous homeomorphisms by piecewise affine homeomorphisms, *Houston J. Math.* **37** no.2 (2011), 449–500.
- [7] R.H. Bing, Locally tame sets are tame, *Ann. of Math.* **59** (1954), 145–158.
- [8] R.H. Bing, Stable homeomorphisms on E^5 can be approximated by piecewise linear ones, *Notices Amer. Math. Soc.* **10** (1963), 607–616.
- [9] E.H. Connell, Approximating stable homeomorphisms by piecewise linear ones, *Ann. of Math.* **78** (1963), 326–338.
- [10] S. Daneri and A. Pratelli, Smooth approximation of bi-Lipschitz orientation-preserving homeomorphisms, *Ann. Inst. H. Poincaré Anal. Non Linéaire* **31** (2014), 567–589.
- [11] S. Daneri and A. Pratelli, *A planar bi-Lipschitz extension theorem*, to appear in *Adv. Calc. Var.*
- [12] S.K. Donaldson and D.P. Sullivan, Quasiconformal 4-manifolds, *Acta Math.* **163** (1989), 181–252.
- [13] L.C. Evans, Quasiconvexity and partial regularity in the calculus of variations, *Ann. of Math.* **95** no.3 (1986), 227–252.
- [14] F. W. Gehring and O. Lehto, On the total differentiability of functions of a complex variable, *Ann. Acad. Sci. Fenn. Ser. A I* **272** (1959), 1–9.
- [15] S. Hencl, Sobolev homeomorphism with zero Jacobian almost everywhere, *J. Math. Pures Appl.* **95** (2011), 444–458.
- [16] S. Hencl and P. Koskela, *Lectures on Mappings of finite distortion*, Lecture Notes in Mathematics 2096, Springer, 2014, 176pp.
- [17] S. Hencl and B. Vejnar, *Sobolev homeomorphism that cannot be approximated by diffeomorphisms in $W^{1,1}$* , preprint (2014).
- [18] T. Iwaniec, L.V. Kovalev and J. Onninen, Diffeomorphic Approximation of Sobolev Homeomorphisms, *Arch. Rational Mech. Anal.* **201** no. 3 (2011), 1047–1067.
- [19] T. Iwaniec, L.V. Kovalev and J. Onninen, Hopf Differentials and Smoothing Sobolev Homeomorphisms, *International Mathematics Research Notices* **14** (2012), 3256–3277.
- [20] R.C. Kirby, Stable homeomorphisms and the annulus conjecture, *Ann. of Math.* **89** (1969), 575–582.
- [21] R.C. Kirby, L.C. Siebenmann and C.T.C. Wall, The annulus conjecture and triangulation, *Notices Amer. Math. Soc.* **16** 432 (1969), abstract 69T-G27.
- [22] J. Luukkainen, Lipschitz and quasiconformal approximation of homeomorphism pairs, *Topology Appl.* **109** (2001), 1–40.

- [23] T. Iwaniec and G. Martin, *Geometric function theory and nonlinear analysis*, Oxford Mathematical Monographs, Clarendon Press, Oxford 2001.
- [24] E.E. Moise, Affine structures in 3-manifolds. IV. Piecewise linear approximations of homeomorphisms, *Ann. of Math.* **55** (1952), 215–222.
- [25] E.E. Moise, *Geometric topology in dimensions 2 and 3*, Graduate Texts in Mathematics 47. Springer. New York-Heidelberg 1977.
- [26] C. Mora-Corral, Approximation by piecewise affine homeomorphisms of Sobolev homeomorphisms that are smooth outside a point, *Houston J. Math.* **35** (2009), 515–539.
- [27] C. Mora-Corral and A. Pratelli, Approximation of Piecewise Affine Homeomorphisms by Diffeomorphisms, *J Geom. Anal* **24** (2014), 1398–1424.
- [28] C.B. Morrey, Quasi-convexity and the semicontinuity of multiple integrals, *Pacific J. Math.* **2** (1952), 25–53.
- [29] S. Müller, Q. Tang and B. S. Yan, On a new class of elastic deformations not allowing for cavitation, *Ann. Inst. Henri Poincaré* **11** (1994), 217–243.
- [30] T. Radó, Über den Begriff Riemannschen Fläche, *Acta. Math. Szeged* **2** (1925), 101–121.
- [31] T.B. Rushing, *Topological embeddings*, Pure and Applied Mathematics 52. Academic Press, New York-London, 1973.

DEPARTMENT OF MATHEMATICAL ANALYSIS, CHARLES UNIVERSITY, SOKOLOVSKÁ 83, 186 00 PRAGUE 8, CZECH REPUBLIC

E-mail address: `henc1@karlin.mff.cuni.cz`

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF ERLANGEN, CAUERSTRASSE 11, 90158 ERLANGEN, GERMANY

E-mail address: `pratelli@math.fau.de`