J. Eur. Math. Soc. 20, 597–656

DOI 10.4171/JEMS/774

Stanislav Hencl · Aldo Pratelli

Diffeomorphic approximation of $W^{1,1}$ planar Sobolev homeomorphisms

Received February 24, 2015

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.

Keywords. Mapping of finite distortion, approximation

1. Introduction

The general problem of finding suitable approximations of homeomorphisms $f : \mathbb{R}^n \supseteq \Omega \to f(\Omega) \subseteq \mathbb{R}^n$ by piecewise affine homeomorphisms has a long history. As far as we know, in the simplest non-trivial setting (i.e. $n = 2$, approximations in L^∞ norm) the problem was solved by Radó [35]. Due to its fundamental importance in geometric topology, the problem of finding piecewise affine homeomorphic approximations in L^∞ norm and dimensions $n > 2$ was deeply investigated in the '50s and '60s. In particular, it was solved by Moise [25] and Bing [7] in the case $n = 3$ (see also the survey book [26]), while for contractible spaces of dimension $n \geq 5$ the result follows from theorems of Connell [10], Bing [8], Kirby [22] and Kirby, Siebenmann and Wall [23] (for a proof see, e.g., Rushing [36] or Luukkainen [24]). 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, they established the existence of a homeomorphism from the unit ball of $\mathbb{R}^4$ to $\mathbb{R}^4$ which cannot be approximated by bi-Lipschitz homeomorphisms [13, Corollary, p. 183]. Actually, their homeomorphism cannot be approximated even in the quite broader class of quasiconformal homeomorphisms (see their discussion in [13, after Definition, p. 181]).

Once completely solved in the uniform sense, the approximation problem suddenly became of interest again in a completely different context, namely, for variational models

S. Hencl: Department of Mathematical Analysis, Charles University,
Sokolovská 83, 186 00 Praha 8, Czech Republic; e-mail: hencl@karlin.mff.cuni.cz

A. Pratelli: Department of Mathematics, University of Erlangen,
Cauerstrasse 11, 90158 Erlangen, Germany; e-mail: pratelli@math.fau.de

Mathematics Subject Classification (2010): 46E35 in non-linear elasticity. Let us briefly explain why. In the setting of non-linear 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

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

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

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

The former condition in (1.2) prevents 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 non-linear elasticity is *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 [29, 1], or to require that W is *polyconvex* and that some coercivity conditions are satisfied [2, 30]. 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 [4, 5] (who ascribes the question to Evans [14]), 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 generally, a result saying that one can approximate a given homeomorphism by 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 motivated not only by the physical model, but also by the fact that the approximation is false in dimension $n \geq 4$, as shown in the recent paper [18].

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 the maximal operator in general destroy injectivity. And on the other hand, we need of course to approximate our homeomorphism not just by smooth maps, but by smooth homeomorphisms (otherwise the approximating sequence would not even be admissible for the original problem).

A 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 approaches are equivalent: more precisely, it is clear that an approximation by diffeomorphisms easily generates another approximation by piecewise affine homeomorphisms; the converse is not immediate but, at least in the plane, it is anyway known [28]. 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 clearly does not exist.

Let us now describe the results which are known in this direction. The first ones were obtained in 2009 by Mora-Corral [27] (for planar bi-Lipschitz mappings that are smooth outside a finite set) and by Bellido and Mora-Corral [6], who proved 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 [19] 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 $W^{1,p}$ norm (improving the previous result of the same authors [20] for $p = 2$). This was a fundamental breakthrough in the area and enhanced interest in this topic.

Later on, it was shown by Daneri and Pratelli [11, 12] that any planar bi-Lipschitz mapping f can be approximated by diffeomorphisms f_k such that f_k converges to f in $W^{1,p}$ norm and simultaneously f_k^{-1} converges to f^{-1} in $W^{1,p}$, giving the first result in which also the distance of the inverse mappings is controlled.

The goal of the present paper is to prove the approximation of planar $W^{1,1}$ homeomorphism in the $W^{1,1}$ sense, so dealing with the important case $p = 1$ which was left out in [19]. 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, $f_\varepsilon(\Omega) = f(\Omega)$, $f_\varepsilon - f \in W_0^{1,1}(\Omega, \mathbb{R}^2)$, 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 $\partial\Omega$ (Theorem 4.20). If even just one of these two conditions does not hold, then this is clearly impossible (see Remark 4.21). For the applications, it is quite important that every f_ε satisfies $f_\varepsilon(\Omega) = f(\Omega)$, in order to be an admissible competitor for the energy. In fact, all the mappings f_ε satisfy also $f_\varepsilon - f \in W_0^{1,1}(\Omega)$, so whenever f has a boundary value in $\partial\Omega$, the same is true for every f_ε, and $f_\varepsilon = f$ in $\partial\Omega$: this is of course essential in all the situations where the boundary value plays a role. These properties of the approximating maps hold true in all the cases treated in the earlier papers discussed above, for instance [19, 11].

We conclude the introduction with a short comparison of our techniques and those of the other papers discussed above. The proofs in [27, 6] are based on a clever refinement of the supremum norm approximation of Moise [25], while the approach of [19] and of the other contributions of the same authors makes use of the identification $\mathbb{R}^2 \simeq \mathbb{C}$ and involves coordinatewise p-harmonic functions. The techniques of the present paper are completely different; 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 somehow resemble the basic ideas of [11, 12], and we will also use some of the tools introduced there, but there are important differences. More precisely, on the one hand, in [11, 12] one had to approximate Df and Df^{-1} at the same time, while here we only need 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 [11, 12] 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 [11, 12] correspond to much simpler steps here, and vice versa.

1.1. Consequences of our construction

In this section (added during the revision of the paper) we briefly comment on the consequences of our result. In fact, the novelty of the present paper is not only to prove Theorem 1.1, but also—and maybe even more—to describe an approximation technique that appears extremely flexible, somehow following the lines of [11, 12], but in a more precise and well-developed way. And in fact, in the few months since the preprint of this paper appeared, there are already four works which use it in a substantial way, three of which already appeared as preprints, while the last one is still being written.

First of all, in this paper we restrict ourselves to the $W^{1,p}$ case with $p = 1$ because the case $p > 1$ was already known. But the same strategy can also be used to get any $p > 1$, as proved in two distinct papers by Campbell [9] and by Radici [34]. Actually, Campbell also proves that the approximation holds in any Orlicz space $W^{1,\Phi}$, a completely new result which could not be obtained with other techniques.

Moreover, as said in the introduction, the main problem now remaining open is the bi-$W^{1,p}$ case, that is, to approximate a $W^{1,p}$ homeomorphism with $W^{1,p}$ inverse simultaneously in the $W^{1,p}$ sense for the map and for the inverse. And actually two works (one by the second author, and the other by the second author and Radici) show that this can be done for the bi-$W^{1,1}$ case [31] and for the bi-BV case [32, 33]. It is worth pointing out that these are the first general approximation results which also treat the inverse.

1.2. Brief description of the proof

In this section we outline the basic plan of our proof, to underline the main steps and facilitate the reading. We remind the reader that our aim is to find an approximation by piecewise affine homeomorphisms, and then the existence of an approximation by smooth diffeomorphisms will follow by applying the result of [28].

First of all, we will divide our domain into some locally finite grid of small squares, the 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 at some Lebesgue point close to c; in particular, we will need that $⨍_{\mathcal{S}(c,r)} |Df - Df(c)|$ is small enough. Since almost every point of Ω is a Lebesgue point for Df, we will be able to deduce that, up to considering 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

$$⨍_{\partial\mathcal{S}} |Df| \le K ⨍_{\mathcal{S}} |Df|, \tag{1.3}$$

where K is a large 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 just have 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). Of course, this will be much easier for good squares, since on a whole good square, f is already almost affine, and more complicated for bad squares. We will construct our approximation g in such a way that, for any bad square $\mathcal{S}$,

$$\int_{\partial\mathcal{S}} |Dg| \le K \int_{\partial\mathcal{S}} |Df|. \tag{1.4}$$

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

The first tool 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

$$⨍_{\mathcal{S}} |Dh| \le K ⨍_{\partial\mathcal{S}} |Dg|. \tag{1.5}$$

This construction is done by first choosing many points on the boundary of the square; then, for any two of them, say x and y, we select the shortest path joining $g(x)$ to $g(y)$ inside the portion of $\mathbb{R}^2$ having $\varphi(\mathcal{S})$ as boundary. Using these shortest paths in a careful way, we 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

$$\int_{\mathcal{S}} |Dh| \leq K \int_{\mathcal{S}} |Df|; \tag{1.6}$$

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

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

It remains to consider 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 squares corresponding to a matrix M with strictly positive determinant (hence, the parallelogram does not degenerate), extension inside $\mathcal{S}$ is trivial; it is enough to divide the square into two triangles and consider on each triangle the affine map which equals f at 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 large 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 those 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 [16]. Section 3 is devoted to building the extension for this specific case, which is somehow similar to the one with the shortest paths described above. The major difference is, on the one hand, that this time we are in a good square, hence very close to a Lebesgue point for Df, and this helps even in the degenerate case. But on the other hand, this time we are not satisfied with an estimate like (1.5), where a large constant K appears, but we need 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 into several steps, is done in Section 4. Basically, putting together all the ingredients described above, the proof will then be concluded for what concerns the existence of a piecewise affine approximation; the existence of a smooth approximation will then follow thanks to the result of [28], while the claim about the boundary values will be easily deduced from the whole construction.

1.3. Preliminaries and notation

In this section we briefly list the basic notation that will be used throughout. We denote by $\mathcal{S}(c,r)$ the square centered at c, of side length $2r$ and with sides parallel to the coordinate axes, while $\mathcal{S}_0 = \{(x,y) \in \mathbb{R}^2 : |x|+|y| < 1\}$ is the "rotated square", which 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 usually be denoted by capital letters, A, B etc., while points in the image $f(\Omega)$ will always be denoted by bold capital letters, $\boldsymbol{A}$, $\boldsymbol{B}$ etc. To shorten the notation and help the reader, whenever we use the same letter A for a point in the domain and $\boldsymbol{A}$ (in bold) for a point in the target, this always means that $\boldsymbol{A}$ is the image of A under the mapping we are considering. We denote by AB (resp. $\boldsymbol{AB}$) the segment between the points A and B (resp. $\boldsymbol{A}$ and $\boldsymbol{B}$). Its length is denoted as $\mathcal{H}^1(AB)$ or $\overline{AB}$, while $\mathcal{H}^1(\gamma)$ is the length of a curve γ. We will write $\widehat{AB}$ (or $\widehat{\boldsymbol{AB}}$) for a particular path between A and B (or $\boldsymbol{A}$ and $\boldsymbol{B}$), whose length will then be $\mathcal{H}^1(\widehat{AB})$, or $\mathcal{H}^1(\widehat{\boldsymbol{AB}})$; we will use this notation only when it is clear what path we are referring to (often this will be a shortest path between the two points). Given three non-collinear points $A,\ B,\ C$ (or $\boldsymbol{A},\ \boldsymbol{B},\ \boldsymbol{C}$), we will denote by $\widehat{ABC}$ (or $\widehat{\boldsymbol{ABC}}$) the angle in $(0,\pi)$ between them, and by ABC (or $\boldsymbol{ABC}$) the triangle having them as vertices.

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

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

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

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

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 $\int_\gamma |Dg(t)|\, d\mathcal{H}^1(t)$ for the length of the curve $g(\gamma)$; notice that the latter length is always well-defined, possibly ∞, and it actually coincides with $\int_\gamma |Dg(t)|$ as soon as this is defined. Finally, notice that if a function f is affine on a square $\mathcal{S}$, with $Df \equiv M$ for some matrix M, and we let g be 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 increase from line to line. For 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 showing the result below about the extension of a map from the boundary of the square to the whole interior.

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

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

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

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

where $V_1 \equiv (0,-1)$ 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

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

because then we can apply a bi-Lipschitz transformation (with bi-Lipschitz constant independent of P_1 and P_2) which moves the points P_1 and P_2 to 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

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

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^i_1 and with second coordinate A^i_2 increasing, with respect to i, from -1 to 1; on the opposite side, we will take the corresponding points $B^i \equiv (-A^i_1, A^i_2)$, so that the segments A^iB^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^iA^{i+1}, as well as on every

$B^i B^{i+1}$ (in particular, the points $(-1,0)$ and $(1,0)$ must be taken). 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 a few lines, and then also later.

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

Up to adding one more point between A^0 and A^1 (plus the corresponding one on the right edge), we can suppose that either γ^1 is a segment between $\boldsymbol{A}^1$ and $\boldsymbol{B}^1$, and this happens if and only if the angle between $\boldsymbol{A}^1$, $\boldsymbol{A}^0$ and $\boldsymbol{B}^1$ which is inside $\boldsymbol{\mathcal{S}}_0$ is smaller than π, or it is the union of the two segments $\boldsymbol{A}^1\boldsymbol{A}^0$ and $\boldsymbol{A}^0\boldsymbol{B}^1$, thus it entirely lies on $\partial\boldsymbol{\mathcal{S}}_0$. 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 shorten the path.

Step 4. *The path γ^{i+1} is above γ^i; the definition of γ_1^i, γ_2^i, γ_3^i.* For two curves γ and $\tilde{\gamma}$ inside $\boldsymbol{\mathcal{S}}_0$ and with endpoints in $\partial\boldsymbol{\mathcal{S}}_0$, we say that γ *is above* $\tilde{\gamma}$ if γ does not intersect the interior of the (possibly disconnected) subset of $\boldsymbol{\mathcal{S}}_0$ whose boundary is the union of $\tilde{\gamma}$ and the path on $\partial\boldsymbol{\mathcal{S}}_0$ connecting the endpoints of $\tilde{\gamma}$ and containing $\boldsymbol{A}^0 = g(A^0)$. 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 $\boldsymbol{P}$ and $\boldsymbol{Q}$ belong to both γ^i and γ^{i+1}. Then the restrictions of γ^i and γ^{i+1} from $\boldsymbol{P}$ to $\boldsymbol{Q}$ are two shortest paths, and by Step 3 they coincide. As an immediate consequence, γ^{i+1} is above γ^i as claimed.

Another immediate consequence is the following: the intersection of γ^i and γ^{i+1} is always a connected subpath, possibly empty. If it is not empty, and then it is a path $\widehat{\boldsymbol{P}\boldsymbol{Q}}$, we will subdivide γ^i and γ^{i+1} into three parts each, writing

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

where γ_1^i (resp. γ_i^{i+1}) is the first part, from $\boldsymbol{A}^i$ to $\boldsymbol{P}$ (resp. from $\boldsymbol{A}^{i+1}$ to $\boldsymbol{P}$); γ_2^i (resp. γ_2^{i+1}) is the second part, from $\boldsymbol{P}$ to $\boldsymbol{Q}$ (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 $\boldsymbol{Q}$ to $\boldsymbol{B}^i$ (resp. from $\boldsymbol{Q}$ to $\boldsymbol{B}^{i+1}$). 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 γ_2^i, γ_3^i, γ_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 formed by two segments, one on $\partial\boldsymbol{\mathcal{S}}_0$ and the other in the interior of $\boldsymbol{\mathcal{S}}_0$.

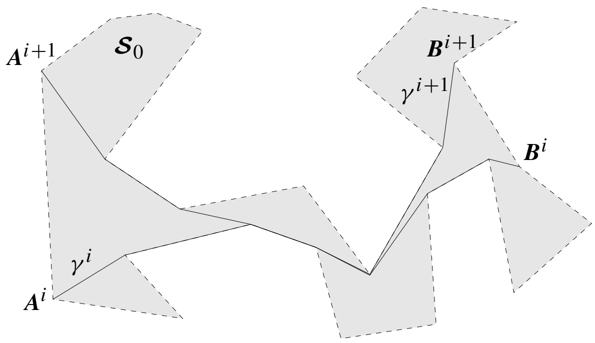

Fig. 1. The paths γ^i and γ^{i+1} in Step 4.

Notice that this subdivision of a path depends not 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 of γ^j and γ^{j+1}, need not coincide with the subdivision of the same path done when $i = j-1$, and then considering the possible common part of γ^j and γ^{j-1}.

Step 5. *Convexity of the polygon having boundary* $\gamma_1^{i+1} \cup \boldsymbol{A}^{i+1}\boldsymbol{P}$. Let $\boldsymbol{P}$ be the last point of γ_1^{i+1}; hence $\boldsymbol{P}$ is the first common point with γ^i if γ^i and γ^{i+1} intersect, and $\boldsymbol{P} = \boldsymbol{B}^{i+1}$ otherwise. We claim that the polygon having $\gamma_1^{i+1} \cup \boldsymbol{A}^{i+1}\boldsymbol{P}$ as boundary is convex (notice that in principle the curve γ_1^{i+1} and the segment $\boldsymbol{A}^{i+1}\boldsymbol{P}$ could have other intersection points besides $\boldsymbol{A}^{i+1}$ and $\boldsymbol{P}$). We start by assuming that $\gamma_2^{i+1} \neq \emptyset$; the other case will be considered at the end of this step.

If γ_1^{i+1} is a single point or just a segment, then the claim is vacuously true, and the convex polygon is degenerate. Assume that γ_1^{i+1} is formed by at least two affine pieces, and also, just to fix ideas, that the direction of the oriented segment $\boldsymbol{A}^i\boldsymbol{A}^{i+1}$ is $\pi/2$, as in Figure 2 (left). Let then $\boldsymbol{\mathcal{D}} \subseteq \boldsymbol{\mathcal{S}}_0$ be the polygon having the Jordan curve $\gamma_1^i \cup \gamma_1^{i+1} \cup \boldsymbol{A}^{i+1}\boldsymbol{A}^i$ as boundary. The same argument as in Step 3 immediately shows that, for any vertex of γ_1^{i+1} (i.e., any junction point between two consecutive linear pieces of γ_1^{i+1}), the angle lying inside $\boldsymbol{\mathcal{D}}$ (hence in particular inside $\boldsymbol{\mathcal{S}}_0$) is greater than π. By construction, and recalling Step 4, we also see that none of these points can belong to the curve in $\partial\boldsymbol{\mathcal{S}}_0$ connecting $\boldsymbol{A}^i$ and $\boldsymbol{B}^i$ and containing $\boldsymbol{A}^0$, since such a point would necessarily also belong to γ^i, contrary to the definition of $\boldsymbol{P}$. Of course, this already suggests that our convexity claim is true, but the proof is not complete yet, since in principle γ_1^{i+1} could be some spiral-like curve connecting $\boldsymbol{A}^{i+1}$ and $\boldsymbol{P}$. To conclude the proof, for any vertex

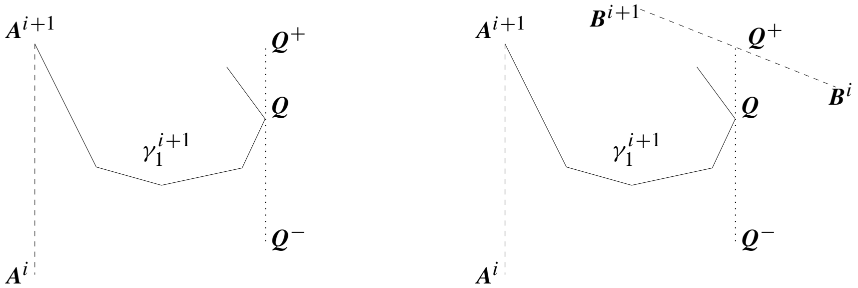

Fig. 2. Construction in Step 5.

of γ_1^{i+1} (except $\boldsymbol{P}$) consider the range of directions pointing toward the interior of $\mathcal{D}$: for instance, the range associated to $\boldsymbol{A}^{i+1}$ in the situation of Figure 2 (left) is formed 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 this is false, and let $\boldsymbol{Q}$ 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 $\boldsymbol{Q}$ is in the interior of $\mathcal{D}$ for a while, both above and below $\boldsymbol{Q}$ itself. As in Figure 2 (left), denote by $\boldsymbol{Q}^-$ and $\boldsymbol{Q}^+$ the first points of this line, respectively below and above $\boldsymbol{Q}$, which are on $\partial\mathcal{D}$. Since the segment $\boldsymbol{Q}^-\boldsymbol{Q}^+$ is parallel to $\boldsymbol{A}^i\boldsymbol{A}^{i+1}$, 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 $\boldsymbol{S}_0$ between its extremes, it is also a shortest path in $\boldsymbol{S}_0$ between any pair of its points. In particular, if the line connecting two points of γ is entirely in the closure of $\boldsymbol{S}_0$, then the part of γ joining them must be a segment. This immediately implies that neither $\boldsymbol{Q}^-$ nor $\boldsymbol{Q}^+$ can belong to γ_1^{i+1}, because otherwise γ_1^{i+1} would be a segment between that point and $\boldsymbol{Q}$; as a consequence, both $\boldsymbol{Q}^-$ and $\boldsymbol{Q}^+$ must belong to γ_1^i, but this is also impossible because then γ_1^i would be a segment. This contradiction proves the claim, yielding the required convexity. Of course, the very same argument works for the polygon having boundary $\gamma_3^{i+1}\cup\boldsymbol{P}\boldsymbol{B}^{i+1}$, where this time $\boldsymbol{P}$ is 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}.

Now consider the case when $\gamma_2^{i+1}=\emptyset$, that is, γ^i and γ^{i+1} are disjoint; this situation is depicted in Figure 2 (right). This time, we let $\mathcal{D}$ be the polygon having the Jordan curve $\gamma^i\cup\boldsymbol{B}^i\boldsymbol{B}^{i+1}\cup\gamma^{i+1}\cup\boldsymbol{A}^{i+1}\boldsymbol{A}^i$ as boundary. The same argument as in the first case shows again that every vertex of γ^{i+1} has angle greater than π in the direction inside $\mathcal{D}$; as a consequence, the required convexity follows as before if the range of no vertex of γ^{i+1} contains the direction $\pi/2$.

However, this time a vertex $\boldsymbol{Q}$ of γ^{i+1} cannot have $\pi/2$ in its range. Indeed, suppose, as before, that $\boldsymbol{Q}$ is the first vertex that does, and let $\boldsymbol{Q}^\pm\in\partial\mathcal{D}$ be as before. As already noticed, none of $\boldsymbol{Q}^\pm$ 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 in $\boldsymbol{B}^i\boldsymbol{B}^{i+1}$. A simple topological argument shows that $\boldsymbol{Q}^-$ must be in γ^i and $\boldsymbol{Q}^+$ in $\boldsymbol{B}^i\boldsymbol{B}^{i+1}$. Indeed, consider the path, contained in $\partial\mathcal{D}$ and not containing $\boldsymbol{A}^i\boldsymbol{A}^{i+1}$, which connects $\boldsymbol{Q}$ and $\boldsymbol{Q}^-$; together with the segment $\boldsymbol{Q}^-\boldsymbol{Q}$, this is a Jordan curve, and so it cannot intersect the other path in $\partial\mathcal{D}$ which contains $\boldsymbol{A}^i\boldsymbol{A}^{i+1}$; in particular, it must contain $\boldsymbol{Q}^+$, and it readily follows, as claimed, that $\boldsymbol{Q}^- \in \gamma^i$ and $\boldsymbol{Q}^+ \in \boldsymbol{B}^i\boldsymbol{B}^{i+1}$. The same topological argument also shows that $\boldsymbol{B}^{i+1}$ is the "left" vertex (that is, the one in the direction $\boldsymbol{A}^i\boldsymbol{A}^{i+1}$) of the segment $\boldsymbol{B}^i\boldsymbol{B}^{i+1}$, and $\boldsymbol{B}^i$ is the "right" one, as in Figure 2 (right).

Now we restrict our attention to the subset $\mathcal{D}_0$ of $\mathcal{D}$ formed by the polygon whose boundary is the part of γ^{i+1} connecting $\boldsymbol{Q}$ to $\boldsymbol{B}^{i+1}$, plus the two segments $\boldsymbol{B}^{i+1}\boldsymbol{Q}^+$ and $\boldsymbol{Q}^+\boldsymbol{Q}$. The same argument of the first half of this step shows 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 $\boldsymbol{B}^i\boldsymbol{B}^{i+1}$, since otherwise a segment parallel to $\boldsymbol{B}^i\boldsymbol{B}^{i+1}$ and contained in $\mathcal{D}_0$ would have both endpoints in $\boldsymbol{Q}\boldsymbol{Q}^+$, which is impossible.

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

Step 6. *Definition of "vertical segments" and their length.* In this step, we associate to any vertex $\boldsymbol{P}$ of the curve γ^{i+1} a point (or many points) $\boldsymbol{Q}$ of the curve γ^i, and vice versa. Every such segment $\boldsymbol{P}\boldsymbol{Q}$, which we will call "vertical", will be contained in the closure of the polygon $\mathcal{D} \subseteq \mathcal{S}_0$ having boundary $\boldsymbol{A}^i\boldsymbol{A}^{i+1} \cup \gamma^{i+1} \cup \boldsymbol{B}^{i+1}\boldsymbol{B}^i \cup \gamma^i$, any two vertical segments will have empty intersection, except possibly for a common endpoint, and the following estimate for the length of the vertical segments will hold:

$$\mathcal{H}^1(\boldsymbol{P}\boldsymbol{Q}) \leq \max\left\{\mathcal{H}^1(\boldsymbol{A}^i\boldsymbol{A}^{i+1}),\ \mathcal{H}^1(\boldsymbol{B}^i\boldsymbol{B}^{i+1})\right\}. \tag{2.4}$$

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} intersect. In the common part $\gamma^i \cap \gamma^{i+1} = \gamma_2^i = \gamma_2^{i+1}$, we will associate to any vertex $\boldsymbol{P}$ of γ^{i+1} the same point $\boldsymbol{Q} \equiv \boldsymbol{P}$, which is also in γ^i by definition. The segment $\boldsymbol{P}\boldsymbol{Q}$ 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 make the following simple definition. To any vertex $\boldsymbol{P} \in \gamma_1^{i+1}$, we associate the point $\boldsymbol{Q} \in \gamma_1^i$ such that the segment $\boldsymbol{P}\boldsymbol{Q}$ is parallel to $\boldsymbol{A}^{i+1}\boldsymbol{A}^i$; the existence and uniqueness of such a point, the validity of (2.4), and the fact that $\boldsymbol{P}\boldsymbol{Q}$ is contained in the closure of $\mathcal{D}$, all come immediately from the convexity obtained in Step 5. We do the same for the vertices of γ_1^i, and we argue similarly for the "right" part of the paths, of course taking segments parallel to $\boldsymbol{B}^{i+1}\boldsymbol{B}^i$ instead of $\boldsymbol{A}^{i+1}\boldsymbol{A}^i$. This completes our definition of the vertical segments, and obviously no two such segments intersect.

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 $\boldsymbol{A}^i\boldsymbol{A}^{i+1}$ is vertical, while the segment $\boldsymbol{B}^i\boldsymbol{B}^{i+1}$ goes "leftwards". Let $\boldsymbol{S} \in \gamma^{i+1}$ and $\boldsymbol{T} \in \gamma^i$ be the two closest points such that the segment $\boldsymbol{T}\boldsymbol{S}$ is vertical; notice that it is possible that $\boldsymbol{S} = \boldsymbol{A}^{i+1}$ or $\boldsymbol{S} = \boldsymbol{B}^{i+1}$ but this makes no difference to our proof, even if the picture shows $\boldsymbol{S}$ in the interior of γ^{i+1}.

Now consider the subset $\mathcal{D}_0 \subseteq \mathcal{D}$ whose boundary is given by the segments $\boldsymbol{A}^i\boldsymbol{A}^{i+1}$ and $\boldsymbol{S}\boldsymbol{T}$, together with the parts of γ^i (resp. γ^{i+1}) connecting $\boldsymbol{A}^i$ and $\boldsymbol{T}$ (resp. $\boldsymbol{A}^{i+1}$ and $\boldsymbol{S}$). Again by the convexity result of Step 5, any point of γ^{i+1} between $\boldsymbol{A}^{i+1}$ and $\boldsymbol{S}$

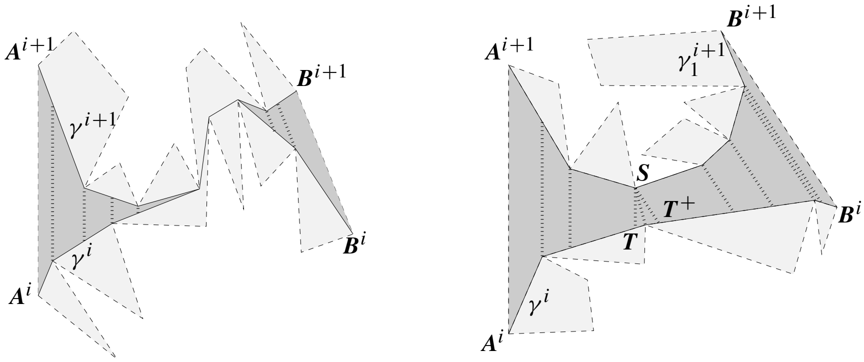

Fig. 3. Construction in Step 6: the polygon $\mathcal{S}_0$ (resp. $\mathcal{D}$) is light (resp. dark), and the "vertical segments" are dotted.

starts a vertical segment whose interior is contained in $\mathcal{D}_0$ and which ends in a point of γ^i between A^i and T. We then define in the obvious way the "vertical segments" inside $\mathcal{D}_0$, which are in fact vertical. The validity of (2.4) is obvious from the convexity as usual.

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, ensures 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 when $B^{i+1}B^i$ is parallel to $A^{i+1}A^i$, but otherwise it stays outside of $\mathcal{D}_0$, as in the figure. The construction implies that all the half-lines parallel to $B^{i+1}B^i$ and starting at points of γ^{i+1} after S remain in $\mathcal{D}$ for a while and then intersect γ^i at some point after T^+. We use this observation to associate to any vertex of γ^{i+1} after S a point of γ^i after T^+, and we then call all the corresponding segments "vertical", although they are not actually vertical but parallel to $B^{i+1}B^i$. Finally, to every vertex of γ^i between T and T^+, if any, we always associate the point 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 empty intersection, unless they meet at S. This concludes the step.

From now on, we will always consider the points S, T and T^+ as "vertices", even if they are not vertices in the sense of piecewise linear curves. Moreover, for every vertex of γ^i or γ^{i+1}, we will also consider as "vertex" the corresponding point in the other curve, which again may or may not be a vertex in the classical sense. Notice that in this way we are adding a *finite number* of new vertices, and as pointed out before, it is always admissible to regard as "vertices" 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 new points. However, these new points have been selected by working on the region between γ^i and γ^{i+1}, and they need not 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 with the definition of a "tentative" function $\tilde{h}$, without taking care of injectivity. The definitive 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 thus divided into $k-2$ *horizontal strips*, lying between two consecutive horizontal segments, plus two triangles, the *top* one $A^{k-1}A^kB^{k-1}$ and the *bottom* one $A^1A^0B^1$.

We start by defining $\tilde{h}$ on the 1-*dimensional skeleton*, that is, the union of $\partial\mathcal{S}_0$ and all the horizontal segments A^iB^i; more precisely, we set $\tilde{h} = g$ on $\partial\mathcal{S}_0$, while for every $1 \leq i \leq k-1$ we define $\tilde{h}$ on A^iB^i as the piecewise linear function, parametrized at constant speed, whose image is γ^i. 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 triangles. First, 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 $\boldsymbol{A}^1\boldsymbol{B}^1$, or the union of the two segments $\boldsymbol{A}^1\boldsymbol{A}^0$ and $\boldsymbol{A}^0\boldsymbol{B}^1$. 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 $\tilde{h}(P) = \boldsymbol{A}^0$, extend $\tilde{h}$ as constantly $\boldsymbol{A}^0$ 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 $\tilde{h}$.

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

We are finally in a position to define $\tilde{h}$ on the interior of each strip $\mathcal{D}_i$ (and since $\tilde{h}$ has already been defined on the 1-skeleton and on the top and bottom triangles, this will conclude the present step). The strip $\mathcal{D}_i$ is the essentially disjoint union of the triangles $P_jP_{j+1}Q_j$ and $P_{j+1}Q_jQ_{j+1}$ for all $0 \leq j < M$, and $\boldsymbol{\mathcal{D}}_i$ is the essentially disjoint union of the corresponding triangles $\boldsymbol{P}_j\boldsymbol{P}_{j+1}\boldsymbol{Q}_j$ and $\boldsymbol{P}_{j+1}\boldsymbol{Q}_j\boldsymbol{Q}_{j+1}$, where the triangles in $\boldsymbol{\mathcal{D}}_i$ (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 then define $\tilde{h}$ on $\mathcal{D}_i$ to be affine on each of the above-mentioned triangles. By construction, $\tilde{h}$ is linear on each side P_jP_{j+1} and Q_jQ_{j+1}, hence this definition on $\mathcal{D}_i$ is a continuous extension of the definition on the 1-skeleton.

Step 8. *Estimate for* $\int_{A^0A^1B^1} |D\tilde{h}|$. 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 $A^0A^1B^1$ (by symmetry, we will get an estimate valid also for the top triangle, of course), while in the next step we will consider the horizontal strips $\mathcal{D}_i$. The aim of this step is to prove the bound

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

where as usual K denotes a purely geometric constant. For simplicity, write

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

Recall that, on the bottom triangle, $\tilde{h}$ has been defined as an affine function, if the angle $\widehat{\boldsymbol{A}^1\boldsymbol{A}^0\boldsymbol{B}^1}$, lying inside $\boldsymbol{\mathcal{S}}_0$, is smaller than π—or, equivalently, if the curve γ^1 coincides with the segment $\boldsymbol{A}^1\boldsymbol{B}^1$—and as two degenerate affine functions on the two triangles A^1PA^0 and A^0PB^1 (with P as in Step 7) otherwise. Let us now 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^0A^1 is $\boldsymbol{A}^0\boldsymbol{A}^1$, while the image of A^0B^1 is $\boldsymbol{A}^0\boldsymbol{B}^1$; this implies that, on the bottom triangle,

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

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

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

On the other hand,

$$\mathcal{H}^1(\boldsymbol{A}^0\boldsymbol{B}^1) = \int_{A^0}^{B^1} |Dg|\, d\mathcal{H}^1, \qquad \mathcal{H}^1(\boldsymbol{A}^0\boldsymbol{A}^1) = \int_{A^0}^{A^1} |Dg|\, d\mathcal{H}^1;$$

inserting this in (2.6), and using (2.2) from Step 1, gives

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

Hence, we deduce that

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

proving (2.5).

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

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

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

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

which together with (2.7) gives

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

Arguing exactly as before, again thanks to (2.2) of Step 1, we again obtain (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

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

where $|\mathcal{D}_i|$ denotes the area of $\mathcal{D}_i$. Consider the horizontal segment $A^{i+1}B^{i+1}$; by symmetry, it is not restrictive to assume that it lies below the x-axis, precisely at a distance $0<r\leq 1$ from the "south pole" $V_1\equiv(0,-1)$; in other words, $A^{i+1}\equiv(-r,r-1)$ and $B^{i+1}\equiv(r,r-1)$. Moreover, let σ be the distance between the segments $A^{i+1}B^{i+1}$ and A^iB^i, and set

$$\ell:=\max\big\{\mathcal{H}^1(\boldsymbol{A}^i\boldsymbol{A}^{i+1}),\ \mathcal{H}^1(\boldsymbol{B}^i\boldsymbol{B}^{i+1})\big\}\leq\int_{A^iA^{i+1}\cup B^iB^{i+1}}|Dg|\,d\mathcal{H}^1. \tag{2.9}$$

Recall now that in Step 7 we defined $\tilde{h}$ to be affine on each of the triangles $P_jP_{j+1}Q_j$ and $P_{j+1}Q_jQ_{j+1}$, sending each of the points P_m (resp. Q_m) in $\mathcal{S}_0$ to $\boldsymbol{P}_m$ (resp. $\boldsymbol{Q}_m$) in $\boldsymbol{\mathcal{S}}_0$. Let us concentrate on the generic triangle $\mathcal{T}=P_jP_{j+1}Q_j$ (for triangles of the form $P_{j+1}Q_jQ_{j+1}$ the same argument will work); since $\tilde{h}$ is affine on $\mathcal{T}$, denote by $D^\tau\tilde{h}$ the constant value of $D\tilde{h}$ on $\mathcal{T}$.

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

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

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

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

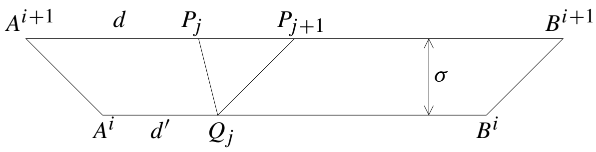

Fig. 4. Position of points and lengths in Step 9.

inserting this in (2.10) and recalling (2.2) gives

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

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

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

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

Since γ^{i+1} is the shortest path between $\boldsymbol{A}^{i+1}$ and $\boldsymbol{B}^{i+1}$ on the closure of $\boldsymbol{\mathcal{S}}_0$, it is in particular shorter than the path obtained as the union of $\boldsymbol{A}^{i+1}\boldsymbol{A}^i$, the part of γ^i between $\boldsymbol{A}^i$ and $\boldsymbol{Q}_j$, the segment $\boldsymbol{Q}_j\boldsymbol{P}_j$, and the part of γ^{i+1} between $\boldsymbol{P}_j$ and $\boldsymbol{B}^{i+1}$; hence,

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

which implies

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

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

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

which further implies

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

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

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

which on recalling also (2.12) finally gives

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

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

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

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

Step 10. *Definition of the modified function h and conclusion of the proof.* We start by observing that summing the estimates (2.5) for the top and bottom triangles and the estimates (2.8) for all the horizontal strips, we directly obtain (2.1) for the function $\tilde{h}$. However, the proof is not complete yet, 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 $\boldsymbol{\mathcal{S}}_0$). 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 $\boldsymbol{\mathcal{S}}_0$ and do not intersect each other. The idea, depicted

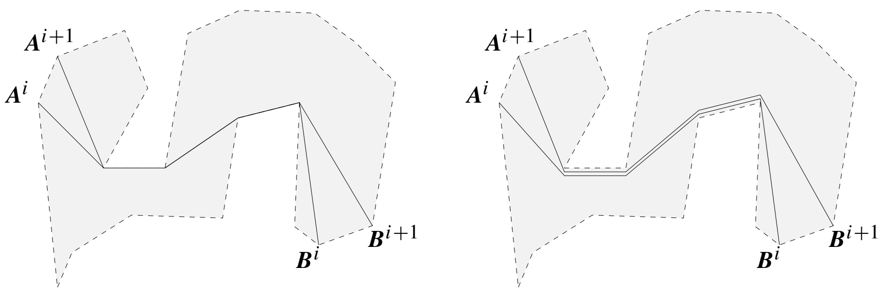

Fig. 5. Modification of the paths γ^i in Step 10.

in Figure 5, is obvious. Notice that since there are only finitely many paths γ^i, and each has only finitely many vertices, it is clear that we can "separate" all the paths as desired, and we can also move each of them a distance which is arbitrarily smaller than all the other distances between extreme points. Then, we define h exactly as $\tilde{h}$, except that we use the modified paths; hence h is now not only finitely piecewise affine and coinciding with g on $\partial\mathcal{S}_0$, but 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. $\square$

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 $g : \partial\mathcal{S} \to \mathbb{R}^2$ is a piecewise linear and one-to-one function, then there exists a piecewise affine extension $h : \mathcal{S} \to \mathbb{R}^2$ of g such that

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

3. Extension in the degenerate case $J_f(c) = 0$ but $|Df(c)| \neq 0$

As already explained in Section 1.2, a crucial difficulty in our proof will be the case when a square $\mathcal{S}$ is "good" (Df is almost constantly equal to some matrix M within $\mathcal{S}$), but $\det M = 0$, while $M \neq 0$. 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 $g : \partial\mathcal{S} \to \mathbb{R}^2$ a piecewise linear and one-to-one function such that*

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

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

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

Proof. We divide this proof into 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, its image $g(\partial\mathcal{S})$ is the boundary of a non-degenerate polygon, which we call $\boldsymbol{\mathcal{S}}$. Moreover, thanks to (3.1), we know that $\boldsymbol{\mathcal{S}}$ is very close to a horizontal segment in $\mathbb{R}^2$. Up to a translation, we can assume that the first coordinates of all points in $\boldsymbol{\mathcal{S}}$ 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 (so with the same first coordinate). In the situation of Figure 6, this happens for σ, but not for σ', since there are four points in $\partial\boldsymbol{\mathcal{S}}$ with first coordinate σ'. We then define any $\sigma \in (0, L)$ to be "good" if exactly two points in $\partial\boldsymbol{\mathcal{S}}$ have first coordinate σ, and additionally $\sigma > 2\delta$ and $\sigma < L - 2\delta$. For any such σ, we denote by $\boldsymbol{P}_\sigma$ and $\boldsymbol{Q}_\sigma$ the above-mentioned two points, with $\boldsymbol{P}_\sigma$ above $\boldsymbol{Q}_\sigma$, and we write $P_\sigma = g^{-1}(\boldsymbol{P}_\sigma)$ and $Q_\sigma = g^{-1}(\boldsymbol{Q}_\sigma)$; up to reversing the orientation of the map g, we

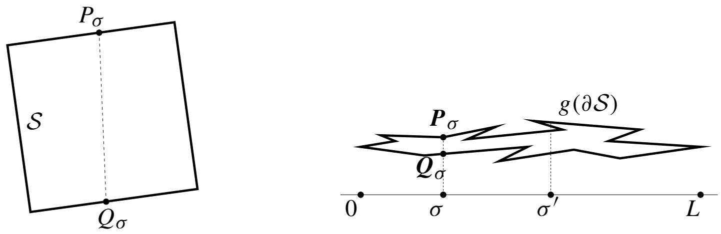

Fig. 6. A good σ and a bad σ' in Step 1.

can assume that every point P_σ is above the corresponding Q_σ, as in our figures. We can immediately show that a large percentage of the points are good, more precisely

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

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

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

As a consequence, if the segment $\boldsymbol{RS}$ is going backward (that is, $\boldsymbol{S}_1 - \boldsymbol{R}_1$ and $S_1 - R_1$ have opposite signs), 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, (3.3) follows.

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

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

Moreover, take any good σ and consider all the segments on $\partial\mathcal{S}$ connecting P_σ and Q_σ; again summing (3.4) over all these segments, and recalling that $\boldsymbol{P}_\sigma$ and $\boldsymbol{Q}_\sigma$ have the same first projection, we derive that

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

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 from 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 the same side of $\mathcal{S}$ for some good σ. Adding once again (3.4) over 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 $\partial\boldsymbol{\mathcal{S}}$; 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 the same side of $\mathcal{S}$ if σ is good.

Observe now that since g is piecewise linear, by definition $(0, L)$ is a finite union of intervals, alternately consisting of only bad σ's and of only good ones. However, the endpoints of all these intervals are always bad. Therefore, we slightly shrink the intervals formed by good points and we call these shrinked intervals *good 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 points of any good interval are good points, including the endpoints, while bad intervals may also contain good points. Finally, it is convenient to make the following further slight modification: up to replacing 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 on the segments $P_\sigma P_{\sigma'}$ and $Q_\sigma Q_{\sigma'}$.

Step 2. *Extension onto the segments $P_\sigma Q_\sigma$; definition of good and bad quadrilaterals.* In this step, we extend g—defined on $\partial\mathcal{S}$—to a union of segments in $\mathcal{S}$. More precisely, recall that $(0, L)$ has been divided into 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 $g(P_\sigma) = \boldsymbol{P}_\sigma$ and $g(Q_\sigma) = \boldsymbol{Q}_\sigma$; 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 by construction. We then have many segments $P_\sigma Q_\sigma$ inside $\mathcal{S}$, almost vertical by (3.5), on each of which a linear function g is defined. As a consequence, $\mathcal{S}$ has been divided into several quadrilaterals (actually, the first and the last one are generally triangles), and g is defined on the whole corresponding 1-dimensional grid; also $\boldsymbol{\mathcal{S}}$ has then been subdivided by the images of g into 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 $\boldsymbol{\mathcal{S}}$—respecting of course the boundary data, and with the extension 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 terminology: 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 remainder of the proof, we will first give an estimate for 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 "internal" bad quadrilaterals, which we obtain by considering two subcases.

Step 3. *Extension in good quadrilaterals.* First, consider a good quadrilateral, corresponding to a good interval (σ, σ') in $(0, L)$; for brevity, we will write P, P', $\boldsymbol{P}$, $\boldsymbol{P}'$ in place of P_σ, $P_{\sigma'}$, $\boldsymbol{P}_\sigma$, $\boldsymbol{P}_{\sigma'}$. 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 $\boldsymbol{P}\boldsymbol{P}'\boldsymbol{Q}'\boldsymbol{Q}$, and we have to define the extension h by sending the interior of $PP'Q'Q$ onto

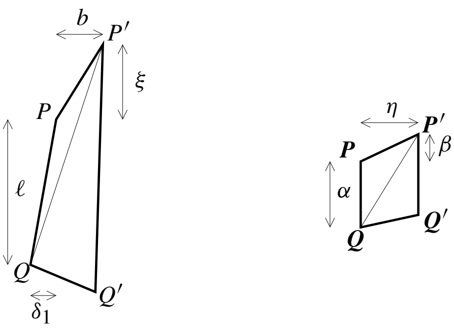

Fig. 7. Approximation in a good quadrilateral, Step 3.

the interior of $\boldsymbol{P}\boldsymbol{P}'\boldsymbol{Q}'\boldsymbol{Q}$. Let h simply be the piecewise affine map sending $PP'Q$ onto $\boldsymbol{P}\boldsymbol{P}'\boldsymbol{Q}$ and $QP'Q'$ onto $\boldsymbol{Q}\boldsymbol{P}'\boldsymbol{Q}'$.

We need to estimate

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

the estimate in the triangle $QP'Q'$ being then of course identical. For brevity, define

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

(see Figure 7). By definition, the constant value of Dh in $PP'Q$ satisfies

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

Let

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

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

$$|\eta - b| \le \varepsilon, \quad |\beta| \le \varepsilon, \quad \alpha \le \delta, \quad |\delta_1| \le \delta/2, \quad \ell > \delta \max\{|\tan\theta|, 1\}. \tag{3.9}$$

The first two estimates can be found by 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'. Notice that $b > 0$ by construction, while ξ could be either positive or negative.

Let us now evaluate D_1h_1. Inserting the third equation of (3.7) into the first one, we get

$$D_1h_1(b - \xi\delta_1/\ell) = \eta,$$

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

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

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

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

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

$$D_2h_2(\ell - \delta_1\xi/b) = \alpha - \delta_1\beta/b,$$

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

$$|D_2h_2| \leq 2\,\frac{\delta}{\ell} + \frac{\delta\varepsilon}{b\ell}, \qquad |D_1h_2| \leq 2\,\frac{\varepsilon}{b} + 2\,\frac{|\xi|\delta}{b\ell}. \tag{3.12}$$

Estimating the integral in (3.6) is then straightforward. Notice that δ is a fixed constant, not depending on the subdivision into 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 into a finite union of good intervals; the area of the triangle $PP'Q$ is then less than $b\ell$, and so from (3.10)–(3.12) 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 the inequalities $\delta \leq 1/2$ and $\ell \leq 3/2$ (the latter follows by straightforward geometrical arguments). Of course, an analogous estimate holds for the integral on the triangle $QP'Q'$, up to suitably modifying the definitions of ε, ξ and b.

To conclude, we need to evaluate the total integral on 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 to each quadrilateral. By definition (3.8) of ε, it is 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,

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

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

Step 4. *Extension in the first and last bad quadrilaterals.* In this and in the next step we are going to consider bad quadrilaterals. Notice that since almost the whole square is covered by good quadrilaterals thanks to (3.3), we can even be satisfied with a rough

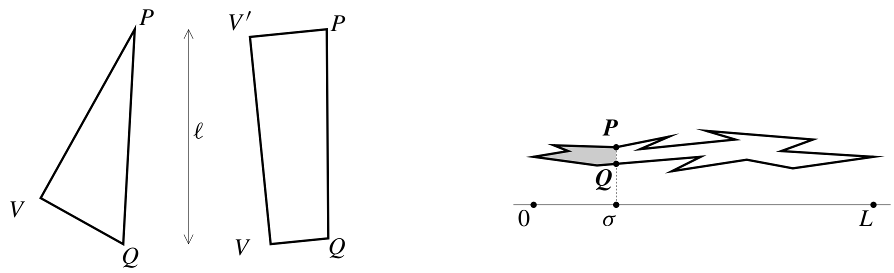

Fig. 8. Approximation in the first bad quadrilateral, Step 4.

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" quadrilaterals, that is, those which correspond to the two intervals having 0 or L as an endpoint. Notice that these "quadrilaterals" are actually triangles, unless some side of the square is very close to being vertical. More precisely, consider just the first bad quadrilateral $\mathcal{C}$, by symmetry; as Figure 8 depicts, it can be either a triangle VPQ, where V is the left vertex of the square, or a quadrilateral $VV'PQ$, if V and V' are the two left vertices of the square, with the side VV' almost vertical, and 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 map $\mathcal{C}$ onto the polygon $\boldsymbol{\mathcal{C}}$ inside $\boldsymbol{\mathcal{S}}$ made up by the points which have first coordinate less than $\sigma = \boldsymbol{P}_1 = \boldsymbol{Q}_1$, 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 (3.4) and (3.5) several times, we know that

$$\boldsymbol{V}_1 \leq \delta/2, \qquad \delta/2 \leq P_1 - V_1 \leq 5\delta, \qquad \delta/2 \leq Q_1 - V_1 \leq 5\delta, \qquad |Q_1 - P_1| \leq \delta/2,$$

and the estimates on V are also valid for V' if the bad quadrilateral $\mathcal{C}$ is actually a quadrilateral. We would like to infer that $\mathcal{C}$ is the bi-Lipschitz image of a square with side δ, with uniformly bounded bi-Lipschitz 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 $\widetilde{\mathcal{C}} = \Phi(\mathcal{C})$; define also $\tilde{g} = g \circ \Phi^{-1}$ on $\partial\widetilde{\mathcal{C}}$, which is admissible since g is defined on the whole $\partial\mathcal{C}$. By construction, $\widetilde{\mathcal{C}}$ is the bi-Lipschitz image of a square of side δ, with bi-Lipschitz constant less than a geometrical constant K. By Theorem 2.1, we find an extension $\tilde{h}$ of $\tilde{g}$ inside $\widetilde{\mathcal{C}}$ such that (2.1) holds, that is,

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

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

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

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

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

We have thus found the estimate we were looking for on the first bad quadrilateral, and by symmetry the same holds in the last bad quadrilateral.

Step 5. *Extension in internal bad quadrilaterals.* To conclude our analysis, we need to concentrate on 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 $\boldsymbol{\mathcal{C}}$. Notice that $\boldsymbol{\mathcal{C}}$ need not be a quadrilateral, since it has two vertical sides, $\boldsymbol{PQ}$ and $\boldsymbol{P'Q'}$, but $\widehat{\boldsymbol{PP'}}$ and $\widehat{\boldsymbol{QQ'}}$ are piecewise linear paths, not necessarily segments. Keeping the notation similar to that of Step 3, we set

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

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

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

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,

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

We now need to further subdivide our analysis into two subcases, depending on whether α is greater 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 $\left(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right)$ as in (3.8). As a consequence, either of the two cases may actually hold.

Step 5a. *The case $\alpha \leq 10\eta$.* In this case we let H be the point in the segment $P'Q'$ satisfying $P_2 = H_2$ (such a point exists by (3.16)), and $\boldsymbol{H} = g(H)$, which is well-defined since g has been defined on the good segment $P'Q'$. We subdivide the quadrilateral $\mathcal{C}$ into the union of the triangle $PP'H$ and the quadrilateral $PHQ'Q$, and we aim to define h separately on these two pieces. First of all, as in the proof of Theorem 2.1, we consider

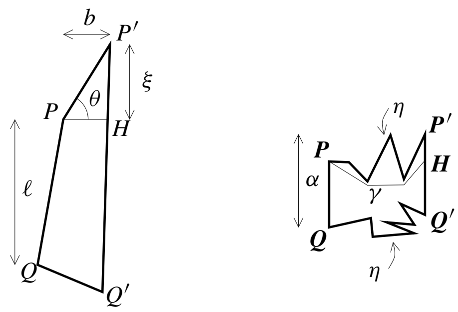

Fig. 9. Approximation in an internal bad quadrilateral: case 1, Step 5a.

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

$$\mathcal{H}^1(\gamma) < \mathcal{H}^1(\widehat{\boldsymbol{P}\boldsymbol{P}'}) + \mathcal{H}^1(\boldsymbol{P}'\boldsymbol{H}). \tag{3.18}$$

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

Let us now focus on the triangle $PP'H$. The segment PH is horizontal by definition, while $P'H$ is "quite vertical"; more precisely, it is contained in $P'Q'$ and by definition we have

$$|P'_1 - Q'_1| \leq \delta/2, \qquad P'_2 - Q'_2 \geq \ell \geq \delta.$$

The triangle would then be a bi-Lipschitz image of a square with side b, with uniformly bounded constant, if ξ is not too much greater than b, or, in other words, if θ is not too large. Since we cannot be sure that this is the case, exactly as in Step 4 we define Φ to be the affine map which does not modify the horizontal segments, and which shrinks the segments parallel to $P'H$ by a factor of ξ/b. Then $\Phi(PP'H)$ is a triangle which is uniformly bi-Lipschitz to 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 an extension of 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

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

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

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 an angle of $\pi/2-\theta \le \pi/4$ with the horizontal direction, this quadrilateral is uniformly bi-Lipschitz to a rectangle. Up to shrinking vertically with a ratio b/ℓ as before, it becomes uniformly bi-Lipschitz to 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

$$\int_{PHQ'Q} |Dh(x)|\,dx \le K\ell\big(\mathcal{H}^1(\boldsymbol{P}\boldsymbol{Q}) + \mathcal{H}^1(\boldsymbol{H}\boldsymbol{Q}') + \mathcal{H}^1(\gamma) + \mathcal{H}^1(\widehat{\boldsymbol{Q}\boldsymbol{Q}'})\big) \le K\eta,$$

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

$$\int_{\mathcal{C}} |Dh(x)|\,dx \le K\eta. \tag{3.19}$$

Step 5b. *The case $\alpha > 10\eta$.* In this case, if we argued as in Step 5a, 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 summing 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 large constant in the end, since bad intervals could be many more than $1/\delta$.

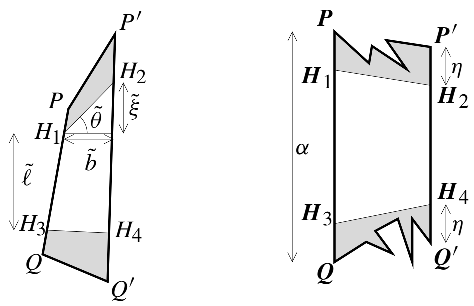

Fig. 10. Approximation in an internal bad quadrilateral: case 2, Step 5b.

As a consequence, in this substep we make a different definition of the extension h for $\alpha > 10\eta$. More precisely, as in Figure 10, we take four points H_1, H_2, H_3, H_4 in the segments PQ and $P'Q'$, and the corresponding points $\boldsymbol{H}_1$, $\boldsymbol{H}_2$, $\boldsymbol{H}_3$, $\boldsymbol{H}_4$ so that $\boldsymbol{H}_i = g(H_i)$ for $i = 1,2,3,4$; the points are chosen in such a way that the segments $\boldsymbol{P}\boldsymbol{H}_1$ and $\boldsymbol{Q}'\boldsymbol{H}_4$ have length η, while H_1H_2 and H_3H_4 are parallel to PP' and QQ' respectively. By definition of η, and keeping in mind that the length of $\boldsymbol{P}'\boldsymbol{Q}'$ is at most α, and at least $\alpha-\eta$, we find that the open segments $\boldsymbol{H}_1\boldsymbol{H}_2$ and $\boldsymbol{H}_3\boldsymbol{H}_4$ are contained in the interior of $\boldsymbol{\mathcal{C}}$, and by construction the same holds for the segments H_1H_2 and H_3H_4 in $\mathcal{C}$. We then regard both $\mathcal{C}$ and $\boldsymbol{\mathcal{C}}$ as the union of three pieces: the internal quadrilaterals $H_1H_2H_4H_3$ and $\boldsymbol{H}_1\boldsymbol{H}_2\boldsymbol{H}_4\boldsymbol{H}_3$, and the "top" and "bottom" remaining parts, shaded in Figure 10. We aim to define the piecewise affine function h so as to send each part of $\mathcal{C}$ onto the corresponding one in $\boldsymbol{\mathcal{C}}$.

For the "top" and "bottom" parts, we can argue more or less as in the last steps: each of the quadrilaterals $PP'H_2H_1$ 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 $\boldsymbol{\mathcal{C}}$ is now at most 4η, the same estimates as in Step 5a can be repeated, so that similarly to (3.19) we now get

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

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

$$H_1 - H_3 = (\tilde{\delta}_1, \tilde{\ell}), \qquad H_2 - H_1 = (\tilde{b}, \tilde{\xi});$$

notice that

$$\frac{|P-H_1|}{|P-Q|} = \frac{|\boldsymbol{P}-\boldsymbol{H}_1|}{|\boldsymbol{P}-\boldsymbol{Q}|} = \frac{\eta}{\alpha} < \frac{1}{10},$$

which implies that $\tilde{b} \geq \frac{9}{10}b$ and $\tilde{\xi} \geq \frac{9}{10}\xi$. Hence, the constant matrix Dh in $H_1H_2H_3$ satisfies

$$\begin{aligned} &|Dh(\tilde{\delta}_1, \tilde{\ell})| \leq \alpha \leq \delta, \\ &|Dh(\tilde{b}, \tilde{b}\tan\theta)| = |Dh(\tilde{b}, \tilde{\xi})| = |\boldsymbol{H}_2 - \boldsymbol{H}_1| \leq 4\eta, \end{aligned} \tag{3.21}$$

since by construction we readily get $|\boldsymbol{P}' - \boldsymbol{H}_2| \leq 2\eta$. As a consequence, we get first

$$|Dh(\tilde{\delta}_1, \tilde{\delta}_1\tan\theta)| \leq 4\,\frac{\tilde{\delta}_1}{\tilde{b}}\,\eta \leq 3\delta\,\frac{\eta}{b},$$

and then

$$|Dh(0, \tilde{\ell} - \tilde{\delta}_1\tan\theta)| \leq \delta(1 + 3\eta/b). \tag{3.22}$$

Recall now that the estimates (3.9) ensure $\ell > 2|\delta_1\tan\theta|$, hence

$$\tilde{\ell} > 2|\tilde{\delta}_1\tan\theta|;$$

indeed, 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 on ℓ and δ_1 are still valid. As a consequence, by (3.22) we get

$$|D_2h| \leq 2\,\frac{\delta}{\tilde{\ell}}\left(1 + 3\,\frac{\eta}{b}\right) \leq 4\,\frac{\delta}{\ell}\left(1 + 3\,\frac{\eta}{b}\right) \leq 4 + 12\,\frac{\eta}{b}, \tag{3.23}$$

and substituting this in (3.21) we also have

$$|Dh(\tilde{b}, 0)| \leq 4\eta + \tilde{b}|\tan\theta|\,|D_2h| \leq 4\eta + \tilde{b}\,\frac{\ell}{\delta}|D_2h| \leq 4\eta + 4\tilde{b} + 12\tilde{b}\,\frac{\eta}{b},$$

from which we derive

$$|D_1 h| \leq 4\eta/\tilde{b} + 4 + 12\eta/b \leq 17\eta/b + 4. \tag{3.24}$$

Since the area of the triangle $H_1 H_2 H_3$ is bounded by b, from (3.23) and (3.24) we get

$$\int_{H_1 H_2 H_3} |Dh(x)|\, dx \leq b(8 + 29\eta/b) \leq K(b+\eta),$$

so that repeating the same estimate in the triangle $H_3 H_2 H_4$, and adding (3.20), we conclude that in a bad quadrilateral $\mathcal{C}$ where $\alpha > 10\eta$ we have

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

Step 6. *Conclusion.* We can now put together all the estimates of the last steps to conclude the proof. Consider first 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), the integral of $|Dh|$ on 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, summing the different η's, as already noticed after (3.17), we get something smaller than 3δ. 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 quadrilaterals, 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 to 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 the proof of Theorem 3.1 is complete. □

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 $2r$, $g : \partial\mathcal{S} \to \mathbb{R}^2$ 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 $h : \mathcal{S} \to \mathbb{R}^2$ 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

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

4. Proof of Theorem 1.1

The proof is still quite involved, but as already explained, the overall idea is simple: to divide the whole Ω into 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 into 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 for f to become piecewise linear. Then, in 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 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 former, we will use Theorem 3.1, while for the latter 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

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

Definition 4.2. Let $\mathcal{S}(x, r) \subseteq \Omega$ be a square, and denote by T_1 and T_2 the two triangles into which $\mathcal{S}$ is subdivided by the diagonal connecting $(x_1 - r, x_2 + r)$ and $(x_1 + r, x_2 - r)$. We let $\varphi_{\mathcal{S}(x,r)}$ be the piecewise affine function which is affine on T_1 and on T_2, and which coincides with f at 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, of 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 if* $\mathcal{S}(c, r)$ *is a Lebesgue square with matrix* M *and constant* $\delta \le \bar\delta$*, then*

$$\|f - \varphi\|_{L^\infty(\mathcal{S}(c,r))} \le r\varepsilon, \qquad \|Df - D\varphi\|_{L^1(\mathcal{S}(c,r))} \le r^2\varepsilon, \qquad \|f - \psi\|_{L^\infty(\mathcal{S}(c,2r))} \le r\varepsilon/10, \tag{4.1}$$

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

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

Proof. We assume for simplicity of notation that c is the origin, and we write $\mathcal{S}(r)$ instead of $\mathcal{S}(0, r)$. Let R be a large constant, depending only on M and ε, to be specified later, and define

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

By definition of Lebesgue square, we immediately get

$$|A| \leq 6r/R, \qquad |B| \leq 6r/R. \tag{4.3}$$

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

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

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 $x_2 - x_1 \leq 7r/R$, $y_2 - y_1 \leq 7r/R$. Hence, w is inside the small rectangle $\mathcal{R}$ having sides with coordinates x_i and y_i, and by (4.4) we know that

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

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

$$\psi(\partial\mathcal{R}) \subseteq \mathcal{B}\left(\psi(w), \frac{7r\sqrt{2}}{R}\|M\|\right),$$

where $\|M\| := \max |M(v)|/|v|$. This estimate, together with (4.5), ensures that the whole curve $f(\partial\mathcal{R})$ consists of points less than $12rR\delta + 7r\sqrt{2}\,\|M\|/R$ away from $\psi(w)$. Since f is a homeomorphism, the point $f(w)$ is inside this curve, hence we finally deduce

$$|f(w) - \psi(w)| = |g(w)| \leq 12rR\delta + \frac{7r\sqrt{2}}{R}\|M\| < r\,\frac{\varepsilon}{10}, \tag{4.6}$$

where the last inequality holds as soon as R has been chosen large 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).

Now consider the function $\varphi = \varphi_{\mathcal{S}(r)}$, and let

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

be 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 T_2) the affine functions φ and ψ satisfy $|\varphi - \psi| = |f - \psi|$ at every vertex, thus (4.6) gives

$$\|\varphi - \psi\|_{L^\infty(\mathcal{S}(r))} \leq \|f - \psi\|_{L^\infty(\mathcal{S}(r))} \leq \|f - \psi\|_{L^\infty(\mathcal{S}(2r))} \leq 12rR\delta + \frac{7r\sqrt{2}}{R}\|M\| < r\,\frac{\varepsilon}{10}, \tag{4.7}$$

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

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

$$|(M_1 - M)(\mathrm{e}_1)| = \left| \frac{(\varphi(V_2) - \varphi(V_1)) - (\psi(V_2) - \psi(V_1))}{2r} \right| < \frac{\varepsilon}{10},$$

and the same holds for $|(M_1 - M)(\mathrm{e}_2)|$ simply by checking the vertices V_1 and V_4. As a consequence, the constant value of $|D\varphi - D\psi|$ in T_1 is less than $\varepsilon/5$, and the same holds in T_2. In other words, $\|D\varphi - D\psi\|_{L^\infty(\mathcal{S}(r))} \leq \varepsilon/5$. Since $D\psi$ is constantly M in $\mathcal{S}(r)$, by the definition of Lebesgue square we get

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

where the last inequality is true after possibly decreasing $\bar{\delta}$; this gives the second estimate in (4.1).

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 (4.2) is obvious for $\|\psi - f\|_{L^\infty(\mathcal{S}(2r))}/r$ small enough, depending on M and ε. Since (4.6) is valid for every $w \in \mathcal{S}(2r)$, we get (4.2) up to further increasing R and decreasing $\bar{\delta}$, again depending only on M and ε. □

Remark 4.4. In the last estimate of the above proof, the final values of $1/R$ and δ behave more or less like 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 $\det M = 0$; 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 shows 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 depends 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 $\det M \to 0$. As a consequence, $\bar{\delta}$ is bounded if ε is bounded from below and $\|M\|$ from above, and if $\det M$ is either 0 or 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 $\delta > 0$ 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

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

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

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

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

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 by the average of $|Df|$ on the boundary of the square. To do so, we first introduce the following notion.

Definition 4.7. We say that the domain Ω is an *r-set* if it is a 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 a *side of type A* if both the endpoints are in the interior of Ω, *of type B* if at least one endpoint is in $\partial\Omega$, but the interior of the side is inside Ω, and *of type C* if the whole side is in $\partial\Omega$. Any vertex of one of the squares will then be called *of type A* if it belongs to the interior of Ω, *of type B* if it belongs to $\partial\Omega$ but it is an endpoint of at least one side of type B, and *of type C* otherwise.

For any small constant $\varepsilon > 0$, we will define a short segment or curve around each vertex of type A or B. We start with 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 let $I_\varepsilon(V)$ be the segment of length $2\sqrt{2}\,\varepsilon r$ centered at V and with direction $\pi/4$, that is,

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

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

The main result that we prove ensures that the average of $|Df|$ in a side of a square can always be estimated by the average of $|Df|$ in the whole square, up to moving the two vertices in the corresponding segments I_ε. Actually, in order to be able to also 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 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. Denote by $\mathcal{R} \subseteq \Omega$ the union of the six squares of the grid having either A or B (or both) as a vertex, and set*

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

Then

$$\mathcal{H}^1\Big(\big\{x \in I_\varepsilon(A) : \mathcal{H}^1(\{y \in I_\varepsilon(B) : (x, y) \in \Gamma\}) > \mathcal{H}^1(I_\varepsilon(B))/5\big\}\Big) < \mathcal{H}^1(I_\varepsilon(A))/5. \tag{4.10}$$

Proof. Suppose, just to fix ideas, that 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 the fact that

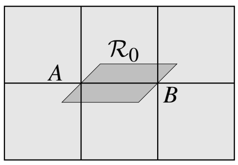

Fig. 11. The rectangle $\mathcal{R}$, the side AB, and the two segments $I_\varepsilon(A)$ and $I_\varepsilon(B)$.

$\mathcal{R}_0 \subseteq \mathcal{R}$, yields

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

On the other hand, writing $\Gamma = \Gamma(A, B, 0)$ for brevity, 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

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

from which (4.10) immediately follows.

To handle 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 $\Gamma(A, B, M)$ coincides with $\Gamma(A, B, 0)$ corresponding to $\tilde{f}$, hence (4.10) follows also in the general case. $\square$

The above result will be useful in treating internal squares, but we also have to take care of boundary squares. We now extend the definition of the segment $I_\varepsilon(V)$ to 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 a vertex of exactly two squares of the decomposition, and these two squares are adjacent (as for V_1 in Figure 12, left), then let $I_\varepsilon(V)$ be the segment of length $2\varepsilon r$ on $\partial\Omega$ centered at V. If V is a vertex of three squares of the decomposition (as for V_2 in the figure), then let $I_\varepsilon(V)$ be the union of the two segments contained in $\partial\Omega$, of length εr, having V as an endpoint.

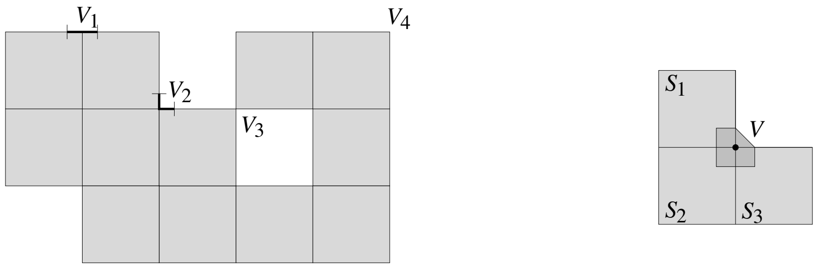

Fig. 12. Left: some squares and vertices V_i near the boundary of an r-set Ω, and the corresponding $I_\varepsilon(V_i)$. Right: definition of $\mathcal{T}_V$ and Ψ_V.

Notice that we have defined $I_\varepsilon(V)$ only for vertices of type A and B, thus for instance not for points like V_3 or V_4 in Figure 12. We now want to extend Lemma 4.9 to 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 denote by $\mathcal{T}_V$ the right triangle having right angle at V, two sides of length $r/2$, one horizontal and one vertical, and not contained in Ω, and for $i = 1, 2, 3$ we let $\mathcal{S}_i^-$ be 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 write Ω^+ for the union of Ω with all the triangles $\mathcal{T}_V$ for vertices V as above, and we let $\Psi : \Omega \to \Omega^+$ be 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 the r-neighborhood of all the internal squares of the decomposition, and it is globally 2-bi-Lipschitz. Finally, for every $x, y \in \overline{\Omega}$ such that the segment $\Psi(x)\Psi(y)$ is contained in $\overline{\Omega^+}$, we denote by $\widetilde{xy}$ the counterimage, under Ψ, of this segment (which is of course a piecewise linear path).

We can finally generalize Lemma 4.9 to all sides of type B; it will be enough to limit ourselves to 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. Denote by $\mathcal{R} \subseteq \Omega$ the union of the squares of the grid having either A or B (or both) as one vertex and define*

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

Then

$$\mathcal{H}^1\Big(\big\{x \in I_\varepsilon(A) : \mathcal{H}^1(\{y \in I_\varepsilon(B) : (x, y) \in \Gamma\}) > \mathcal{H}^1(I_\varepsilon(B))/5\big\}\Big) < \mathcal{H}^1(I_\varepsilon(A))/5. \tag{4.11}$$

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 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 what these squares are like, this is the second or third case in the figure); last, A can belong to two squares, which are then necessarily the two squares having AB as a side (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 (the fifth case in the figure) or in the boundary of Ω: in this latter case (the sixth and last in the figure), A must also belong only to the same two squares to which B belongs, since otherwise we fall into a case already considered. To clarify the situation, Figure 13 does not show the situation in Ω, but directly in Ω^+.

The proof is now almost identical to that of Lemma 4.10. Let again $\mathcal{R}_0$ be 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, lies in Ω^+. Notice that, depending on the case, $\mathcal{R}$ can be made up by 2, 3, 4, or 5 squares, and the figure always shows only (the image under Ψ of) these squares. To conclude the proof we 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 for 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-bi-Lipschitz, 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. $\square$

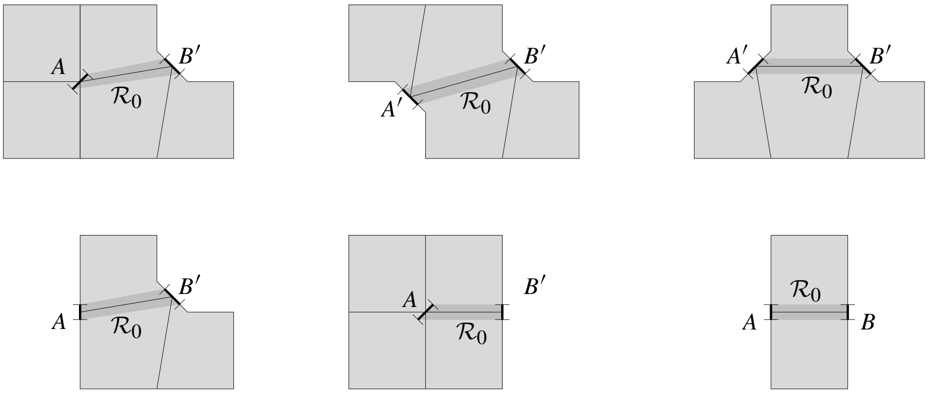

Fig. 13. The six possibilities in Lemma 4.12.

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 $\widetilde{xy}$) too much derivative is concentrated. We can now find a selection of points $x_V \in I_\varepsilon(V)$ for any vertex V such 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 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, we enumerate all the vertices of type A or B as $V_1, \dots, V_N$, 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 satisfies

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

Notice that since we will define the points recursively, the above requirements make sense: in particular, if $m < i$ then the point x_m has already been 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 conclusion 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 suppose that the points x_j for $j < i$ have been chosen according to (4.12); let n^- (resp. n^+) be the number of 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

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

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 conditions holds, hence by definition this choice fulfills (4.12). The recursion argument is then complete. □

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 $I_\varepsilon(V)$; moreover, let $V' = V$ for any vertex V of type A or C. We define the *grid* to be the union $\mathcal{G}$ of all the sides AB of the squares of the decomposition, while the *modified grid* is the union $\widetilde{\mathcal{G}}$ 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; we then 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 actually this minimizing property of $\widetilde{A'B'}$ is also true 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 write $\widetilde{\mathcal{S}}$ for 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 other 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 now give 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 $x_V \in I_\varepsilon(V)$ for vertices V of type A or B be as in Lemma 4.13. We define $g : \widetilde{\mathcal{G}} \to \mathbb{R}^2$ as follows. For any side AB of type A or B, we define g on $\widetilde{A'B'}$ as the reparametrization, at constant speed, of the function f on $\widetilde{x_A x_B}$; 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 to 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 AB is the reparametrized copy of f on the segment $x_A x_B = \widetilde{x_A x_B}$.

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 $g : \widetilde{\mathcal{G}} \to \mathbb{R}^2$ be as in Definition 4.15. Then, for any side AB of type A, letting ν be the unit vector with direction AB and $\mathcal{R}$ again the union of the squares of the grid having either A or B as a vertex, we have*

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

On the other hand, for any side AB of type B, we have

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

Proof. For a side AB of type A, set for brevity $x = x_A$ and $y = x_B$. By Lemma 4.9, we already know that

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

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

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

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 set again $x = x_A$ and $y = y_B$. In this case, by Lemma 4.12 we already know that

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

Now, by definition of g 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). $\square$

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 make it piecewise linear. We have to do so because both our major results, 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 "bad" squares and to "good" squares corresponding to a matrix with $\det M = 0$; this construction is reminiscent of the one in [11], where the situation was more complicated because also the inverse had to 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 $g : \widetilde{\mathcal{G}} \to \mathbb{R}^2$ be a continuous, injective function which is piecewise linear on $\partial\Omega$. Then there exists a piecewise linear and injective function $\hat{g} : \widetilde{\mathcal{G}} \to \mathbb{R}^2$ 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*

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

while for every side AB of type B one has

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

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: 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.* Select 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 denote by V_i, $1 \leq i \leq 4$, the other four endpoints of these sides. Since g is continuous and injective, the quantity

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

is strictly positive; 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 $\widetilde{\mathcal{G}}$ 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 $i = 1, \ldots, 4$ there is a point $x \in VV_i$ such that $g(x) \in \partial\mathcal{B}(g(V), \rho)$. Let V_i^+ be the last such point, where "last" means "farthest from V". We then define the function $\hat{g}$, on each of the four segments VV_i^+, simply as the linear function connecting $g(V)$ and $g(V_i^+)$.

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 j be the number of such sides, and notice that by construction, j is either 1 or 2 (keep in mind Figure 13). We then argue much as before: we denote by V_i for $1 \leq i \leq j$ the other endpoints of these internal sides, and we consider the strictly positive quantity

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

This time, we take $\rho = \rho(V)$ 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 write V_i^+ for the last point $x \in \widetilde{V'V_i'}$ such that $g(x) \in \partial\mathcal{B}(g(V'), \rho)$; up to further

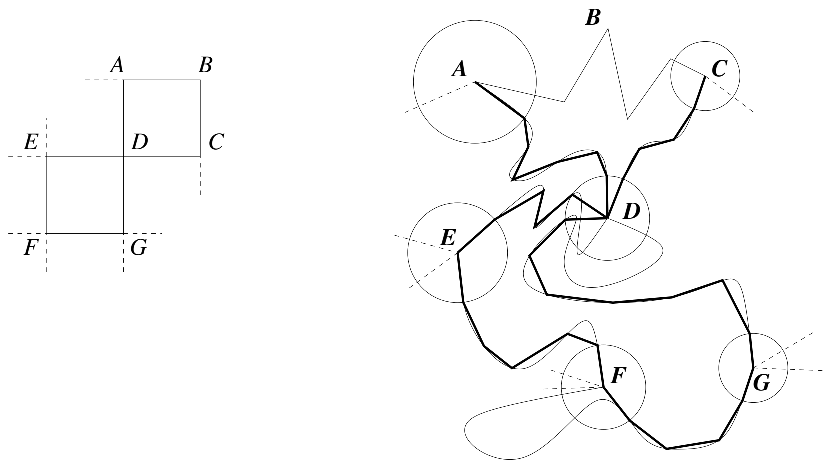

Fig. 14. The construction in Proposition 4.17: the points A, B and C are in $\partial\Omega$, while D, E, F and G are inside Ω; the image of $\hat{g}$ inside Ω is thicker.

decreasing $\rho(V)$, 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 $g(V')$ and $g(V_i^+)$.

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 or B; moreover, $\hat{g}$ is already automatically defined on $\partial\Omega$, since we must have $\hat{g} = g$ on $\partial\Omega$. Therefore, to conclude we have to define $\hat{g}$ on the remaining part of $\widetilde{\mathcal{G}}$. By construction, this part 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 on some piecewise linear curve $\widetilde{A^+B^-} \subset\subset \widetilde{A'B'}$.

First consider a side AB of type A. The function $\hat{g}$ has already been defined on 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 $\mathcal{B}(g(A), \rho(A))$ and $\mathcal{B}(g(B), \rho(B))$ respectively. We have to define $\hat{g}$ on the segment A^+B^-, and this must be a piecewise linear curve connecting $g(A^+)$ and $g(B^-)$. Observe that g, on 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 $\hat{g}$ on 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 simple geometric argument (similar to [11, Lemma 5.5], but much easier) shows that, by carefully choosing 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 to g is much smaller than

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

Now consider a side AB of type B. In this case, $\hat{g}$ is piecewise linear on 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 being 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, and thus the definition of $\hat{g}$ is complete.

Step III. *The properties of* $\hat{g}$. To conclude the proof, we just need to check that $\hat{g}$ has all the required properties. The fact that $\hat{g} = g$ on $\partial\Omega$ and at every vertex is true by construction, as is 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, so (4.17) follows immediately for $M = 0$, and the same argument with $\widetilde{A'B'}$ in place of AB also yields (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 summing over the segments where $\hat{g}$ is linear we get (4.17). □

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 but a straightforward modification of the previous one.

Proposition 4.18. *Let $\mathcal{C} = \bigcup_{i=1}^{N} A_iB_i$ be a finite union of closed segments in $\mathbb{R}^2$, and let $\mathcal{C}_0 = \bigcup_{i=1}^{N_0} A_iB_i$, with $N_0 \leq N$, be a selection of some of them. Let $g : \mathcal{C} \to \mathbb{R}^2$ be a continuous one-to-one function, and let $\eta > 0$. Then there exists another continuous one-to-one function $\hat{g} : \mathcal{C} \to \mathbb{R}^2$ 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 goes almost exactly as that of Proposition 4.17. First of all, up to a subdivision of some of the segments, we can assume that any two segments are either disjoint, or meet in a common endpoint. Then, we start by defining $\hat{g} = g$ on all the sides both of whose endpoints are off $\mathcal{C}_0$. 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; by decreasing ρ, we can also ensure that $|x - V| < \eta$ for any such x. We then define the points V_i^+ exactly as in Step I of the proof 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 been defined yet, we can then define it in two different ways: inside the segments which form $\mathcal{C}_0$, 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 the proof of Proposition 4.17, we can then immediately observe that the function $\hat{g}$ is as required. □

4.4. *The proof of Theorem 1.1*

We are finally in a position to prove Theorem 1.1, which will be done by putting together all the results that we got up to now. For the reader's convenience, we divide the proof into several parts. The first one is a very peculiar case, when Ω is an r-set and 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 in addition that 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 $f_\varepsilon : \Omega \to \mathbb{R}^2$ such that*

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

Proof. The idea of the construction is rather simple: we divide the squares into four groups: 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 replace 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 the 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 some care.

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 take five small constants 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

$$\int_A |Df| \leq \frac{\varepsilon_1\varepsilon}{54K} \qquad \forall A \subseteq \Omega,\ |A| \leq \varepsilon_2, \tag{4.20}$$

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

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

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

$$\mathcal{M}^+ = \{\varepsilon_3 < \|M\| < 1/\varepsilon_3,\ \det M > \varepsilon_3\}, \qquad \mathcal{M}^0 = \{\varepsilon_3 < \|M\| < 1/\varepsilon_3,\ \det M = 0\},$$

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

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

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

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

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

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

where the constants $\bar{r}(x, \hat{\delta})$ have been defined in Lemma 4.6 for every $x \in \Omega$ which is a Lebesgue point 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 into four groups:

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

We now aim to show that most of the squares belong to the first three groups. In fact, consider the total area of the squares in $\mathcal{A}_4$. The union of those which touch the boundary of Ω has of course an area smaller than $rP(\Omega)$. Let $\mathcal{S}(c, r) \in \mathcal{A}_4$ be compactly contained in Ω; this means that, for every $x \in \mathcal{S}(c, r/2)$, 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 that of $\mathcal{S}(c, r/2)$, by (4.21) and (4.24) we deduce

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

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 also be a vertex 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 notation that $\mathcal{S}_1 = \mathcal{S}(c_1, r)$ and $\mathcal{S}_2 = \mathcal{S}(c_2, r)$ have a common vertex, and $\mathcal{S}_1 \in \mathcal{A}_1$, $\mathcal{S}_2 \in \mathcal{A}_2$. Then Lemma 4.3 provides 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. Set

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

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

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

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 $\mathcal{A}_1$, then yields

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

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

$$\forall \mathcal{S}_a,\ \mathcal{S}_b \in \mathcal{A}_1 \cup \mathcal{A}_2 \text{ adjacent}, \qquad \|M_a - M_b\| < \varepsilon_5/5. \tag{4.26}$$

Step III. *A tentative modified grid $\widetilde{\mathcal{G}}_1$ and a tentative function $g_1 : \widetilde{\mathcal{G}}_1 \to \mathbb{R}^2$.* In this step, we define a modified grid $\widetilde{\mathcal{G}}_1$ 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 this is done 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 $g_1 : \widetilde{\mathcal{G}}_1 \to \mathbb{R}^2$ from Definition 4.15.

The matrices $M(A, B)$ will be defined as follows. For any side AB of type A, let for the 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 $\mathcal{S}_b$ belongs to $\mathcal{A}_2$, we let $M(A, B) = 0$; if $\mathcal{S}_a \in \mathcal{A}_2$ but $\mathcal{S}_b \notin \mathcal{A}_2$,, 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 let $M(A, B)$ be M_a or M_b—it makes no difference which one we choose, since $M_a \approx M_b$ by (4.26).

We now want to evaluate $\int_{\partial\widetilde{\mathcal{S}}} |Dg_1|$ for some of the modified squares $\widetilde{\mathcal{S}}$ (recall Definition 4.14). Take a square $\mathcal{S} = \mathcal{S}(c, r) \in \mathcal{A}_3 \cup \mathcal{A}_4$, and let $\mathcal{S}^+ = \mathcal{S}(c, 3r) \cap \Omega$ be 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) to get

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

If instead $AB \subseteq \partial\mathcal{S} \cap \partial\Omega$, then of course

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

Since the boundary of $\widetilde{\mathcal{S}}$ is just 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

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

where

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

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 $\|M(A,B)-M\|\leq\varepsilon_5/5$, since $\mathcal{S}$ is 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) and the definition of $\mathcal{M}^0$ give

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

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 $\widetilde{\mathcal{G}}$ and a second tentative function $g_2:\widetilde{\mathcal{G}}\to\mathbb{R}^2$.* 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 $I_{\varepsilon_4}(V)$. 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, we define the matrices $M'(A,B)=0$ for all the sides AB of type A, and 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 a square in $\mathcal{A}_1$ or $\mathcal{A}_2$ (these two things cannot happen simultaneously, thanks to Step II), define $I'_{\varepsilon_1}(V)$ as a translation of $I_{\varepsilon_1}(V)$ by $(\pm 2\varepsilon_1 r, \pm 2\varepsilon_1 r)$, 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 a 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 upper square and the second interval in the lower one.

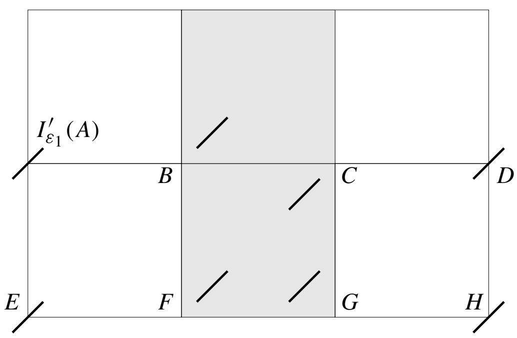

Fig. 15. The intervals $I'_{\varepsilon_1}(V)$ in Step IV.

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 $I_{\varepsilon_1}(V)$: indeed, in that simple proof we have just used the fact that the internal intervals are all of length $2\sqrt{2}\,\varepsilon r$, with direction 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 $\widetilde{\mathcal{G}}$ from Definition 4.14, and the function $g_2 : \widetilde{\mathcal{G}} \to \mathbb{R}^2$ from Definition 4.15. The modified grid that we get 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 same calculations of the last step work also for this new case, just by replacing 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

$$\int_{\partial \widetilde{\mathcal{S}}} |Dg_2|\, d\mathcal{H}^1 \leq \frac{400}{\varepsilon_1 r} \int_{\mathcal{S}^+} |Df|\, d\mathcal{H}^2 + \int_{\partial \mathcal{S}^+ \cap \partial\Omega} |Df|\, d\mathcal{H}^1 \quad \forall \mathcal{S} \in \mathcal{A}_3 \cup \mathcal{A}_4. \tag{4.29}$$

Step V. *Definition of* $g_3 : \widetilde{\mathcal{G}} \to \mathbb{R}^2$. We are now in a position to define a function g_3 on the grid $\widetilde{\mathcal{G}}$ introduced in Step IV. This function will behave almost correctly on the whole grid, its only fault (which will be solved in the next step) being not to be piecewise linear. As 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 g_3 would not then 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 as 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 given by Step IV. Notice that, up to now, the function g_3 is continuous and injective: this is a straightforward consequence of the L^∞ estimate around squares in $\mathcal{A}_1$, and of the fact that g_3 is a reparametrization of f on different segments around squares in $\mathcal{A}_2$ or $\mathcal{A}_3 \cup \mathcal{A}_4$. However, g_3 has not yet been defined in the whole $\widetilde{\mathcal{G}}$.

Consider a side AB such that g_3 has not been defined on $\widetilde{A'B'}$ yet; by construction, this means that A or B is a 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 just 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.

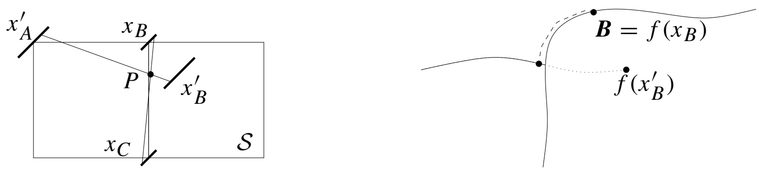

Fig. 16. The definition of g_3 in Step V if B is in some square of $\mathcal{A}_2$.

For simplicity, assume first that A is not a vertex of squares in $\mathcal{A}_1$ or $\mathcal{A}_2$, and B is a vertex of at least a square in $\mathcal{A}_2$ (by construction and recalling Step II, B is then a 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 $C \neq B$ be 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 of the part of $\widetilde{x'_A x'_B}$ from x_A to P and the segment Px_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 $\boldsymbol{B} = f(x_B)$. Of course, we need to modify $\tilde{\gamma}$ only between $f(P)$ and $\boldsymbol{B}$; 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

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

Thanks to (4.28), letting M be 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

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

Now assume that, instead, A is still not a vertex of squares in $\mathcal{A}_1$ or $\mathcal{A}_2$, and B is a 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 $\boldsymbol{P}$ is the first intersection point (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 $\boldsymbol{P}$ and the segment $\boldsymbol{PB} = \boldsymbol{P}f(B)$. 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

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

also in this case we get (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 segment AB: both 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 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 with γ_0 in a large central part, 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 made two modifications, instead of (4.31) we will have

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

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 on the whole grid $\widetilde{\mathcal{G}}$, and by construction it is clear that 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 write again $\mathcal{S}^+$ for 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 the fact 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 (by summing over the four sides of $\mathcal{S}$) gives

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

Finally, if $\mathcal{S} \in \mathcal{A}_3 \cup \mathcal{A}_4$, then $\int_{\partial\widetilde{\mathcal{S}}} |Dg_3|$ 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

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

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

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

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

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

Step VI. *The piecewise linear function* $g_4 : \widetilde{\mathcal{G}} \to \mathbb{R}^2$. In this step, we want to define a piecewise linear function $g_4 : \widetilde{\mathcal{G}} \to \mathbb{R}^2$; this will finally be the correct map on the grid $\widetilde{\mathcal{G}}$, in the sense that our approximating function f_ε will coincide with g_4 on $\widetilde{\mathcal{G}}$. To do so, it is enough to apply Proposition 4.17 to the map g_3 and denote by $g_4 = \hat{g}$ the resulting map. As a consequence, g_4 is a piecewise linear map on $\widetilde{\mathcal{G}}$, 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 also valid with g_4 in place of g_3.

Step VII. *Definition of the approximating function $f_\varepsilon : \Omega \to \mathbb{R}^2$.* 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 $\widetilde{\mathcal{G}}$, hence in particular $f_\varepsilon = f$ on $\partial\Omega$; by construction, f_ε is injective and piecewise linear on $\widetilde{\mathcal{G}}$. To keep the injectivity, it is then enough to extend f_ε in the interior of any square $\widetilde{\mathcal{S}}$ 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) we know that

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

Now consider $\mathcal{S} \in \mathcal{A}_2$; 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 letting f_ε be the resulting extension on $\mathcal{S}$, we get

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

as soon as

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

Thanks to (4.33), which also holds 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 the definition of $\bar{\delta}$, and also that $\varepsilon_4 \leq 1$, from (4.38) and (4.33) we get

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

where K is as always a purely geometric constant.

Finally, let $\mathcal{S} \in \mathcal{A}_3 \cup \mathcal{A}_4$. This time, since g_4 is piecewise linear on $\partial\widetilde{\mathcal{S}}$ and $\widetilde{\mathcal{S}}$ is (a 2-bi-Lipschitz copy of) a square of side $2r$, we let f_ε on $\widetilde{\mathcal{S}}$ 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 also valid with g_4 in place of g_3 by Step VI, gives

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

Since

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

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

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

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

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

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

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

Step VIII. *Conclusion.* By construction, f_ε is a finitely piecewise affine homeomorphism which coincides with f on $\partial\Omega$. Moreover, it is immediate that $\|f_\varepsilon - f\|_{L^\infty}$ and $\|f_\varepsilon - f\|_{L^1}$ are as small as we wish (it is enough to choose r small enough at the beginning); as a consequence, we can assume that they are smaller than $\varepsilon/4$ each. Hence, to get (4.19) and conclude, we only have to check that

$$\|Df_\varepsilon - Df\|_{L^1} < \varepsilon/2. \tag{4.42}$$

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

$$\Omega_j = \bigcup \{\widetilde{\mathcal{S}}_i : \mathcal{S}_i \in \mathcal{A}_j\},$$

we have

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

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, those which lie in Ω). As a consequence, setting for brevity $\mathcal{A}_4^+ = \bigcup_{i:\,\mathcal{S}_i \in \mathcal{A}_4} \mathcal{S}_i^+$, also recalling (4.25), we have $|\mathcal{A}_4^+| \leq 9 |\bigcup_{i:\,\mathcal{S}_i \in \mathcal{A}_4} \mathcal{S}_i| < \varepsilon_2$. 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 lie in $\partial\mathcal{S}_i^+$ 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

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

Finally, by (4.22) we have

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

Inserting the last three estimates into (4.43) we get (4.42). □

The above proposition shows that, under stronger assumptions than in Theorem 1.1, we can obtain something better than what is claimed in Theorem 1.1. Indeed, if Ω is an r-set and f is piecewise linear on $\partial\Omega$, then we get not just 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 under 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 $f_\varepsilon : \Omega \to \mathbb{R}^2$ 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 $\Phi : \Omega \to \widehat{\Omega}$. There exists then some constant $H = H(\Omega)$ such that

$$|D\Phi(x)| \le H, \qquad \det D\Phi(x) \ge 1/H,$$

for almost every $x \in \Omega$. Define $\hat f : \widehat{\Omega} \to \mathbb{R}^2$ 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 $\hat f_\varepsilon : \widehat{\Omega} \to \mathbb{R}^2$ which coincides with $\hat f$ on $\partial\widehat{\Omega}$ and satisfies

$$\|\hat f_\varepsilon - \hat f\|_{W^{1,1}(\widehat{\Omega})} + \|\hat f_\varepsilon - \hat f\|_{L^\infty(\widehat{\Omega})} \le \varepsilon/H^2. \tag{4.44}$$

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

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

By a simple change of variable argument, we obtain

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

and similarly

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

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

Remark 4.21. The assumptions 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 of f.

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

$$\int_A |Df| \le \frac{\varepsilon\varepsilon_1}{72K}\qquad \forall A\subseteq\Omega,\ |A|\le\varepsilon_2. \tag{4.45}$$

The next step is to write Ω as a countable union of r_n-sets. More precisely, we can take a sequence of constants $r_n\to 0$ 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 into two disjoint parts, $\partial\Omega_n = \partial^-\Omega_n\cup\partial^+\Omega_n$, with

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

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

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

and we can also take Ω_1 large enough that

$$\int_{\Omega\setminus\Omega_1} |Df| \le \frac{\varepsilon\varepsilon_1}{72K}. \tag{4.47}$$

Naively speaking, the idea is to try to work on each Ω_n separately. However, since f is not necessarily piecewise linear on the boundary of the sets Ω_n, we cannot simply rely on Proposition 4.19 for each Ω_n; moreover, since Ω may not have 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 summing.

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 as in the proof of Proposition 4.19, we first let ε_3 be a constant such that

$$|\{x \in \Omega_1 : 0 < |Df(x)| < \varepsilon_3 \text{ or } |Df(x)| > 1/\varepsilon_3 \text{ or } 0 < \det(Df(x)) < \varepsilon_3\}| < \varepsilon_2/45; \tag{4.48}$$

then we let again

$$\mathcal{M}^+ = \{\varepsilon_3 < \|M\| < 1/\varepsilon_3,\ \det M > \varepsilon_3\}, \qquad \mathcal{M}^0 = \{\varepsilon_3 < \|M\| < 1/\varepsilon_3,\ \det M = 0\};$$

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

$$\frac{\varepsilon_5}{\varepsilon_4} + \frac{\varepsilon_4}{\varepsilon_3} + \frac{\varepsilon_5}{\varepsilon_1} \leq \frac{\varepsilon}{8K|\Omega_1|}, \qquad \varepsilon_5 \ll \varepsilon_4\varepsilon_3, \qquad \varepsilon_4 \ll \varepsilon_3^2; \tag{4.49}$$

and finally we define $\hat{\delta} = \hat{\delta}(\varepsilon_3, \varepsilon_5)$ 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 an r_n/H_n-set for any constant $H_n \in \mathbb{N}$. As a consequence, we can now decrease r_n (without changing Ω_n, of course); in particular, also thanks to (4.46), we can assume that r_1 is so small that

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

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

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

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, we see that any Ω_n is divided into a finite union of squares, all of side $2r_n$. We 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 into four groups:

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

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

$$\Big|\bigcup\{\mathcal{S}_i \in \mathcal{A}_4^1\}\Big| \le r_1 P(\Omega_1) + 4\big(|\{x \in \Omega_1 : \bar r(x, \hat\delta) \le r_1\}| + \varepsilon_2/45\big) < \varepsilon_2/9. \tag{4.52}$$

For any $n \ge 2$, 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.

Notice that the only assumption which is true for Ω in Proposition 4.19 and may now fail for the generic Ω_n is the following: f is assumed to be piecewise linear on $\partial\Omega$ in Proposition 4.19, while f need not be piecewise linear on $\partial\Omega_n$; 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–V of that proof for Ω_1; so, we discover first that squares in $\mathcal{A}_1^1$ and $\mathcal{A}_2^1$ can never touch, then we define a tentative modified grid $\widetilde{\mathcal{G}}_1^1$ with a function $g_1^1 : \widetilde{\mathcal{G}}_1^n \to \mathbb{R}^2$, then the correct modified grid $\widetilde{\mathcal{G}}^1$ with the function $g_2^1 : \widetilde{\mathcal{G}}^1 \to \mathbb{R}^2$, and finally the function $g_3^1 : \widetilde{\mathcal{G}}^1 \to \mathbb{R}^2$. By definition, 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

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

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

Now consider Ω_n for any $n \ge 2$. 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 analogue of Step IV, which immediately gives us a function $g_2^n : \widetilde{\mathcal{G}}_n \to \mathbb{R}^2$ satisfying the analogue of (4.29):

$$\int_{\partial\widetilde{\mathcal{S}}} |Dg_2^n|\, d\mathcal{H}^1 \le \frac{400}{\varepsilon_1 r_n}\int_{\mathcal{S}^+} |Df|\, d\mathcal{H}^2 + \int_{\partial\mathcal{S}^+\cap\partial\Omega_n} |Df|\, d\mathcal{H}^1 \quad \forall \mathcal{S} \in \mathcal{A}_4^n. \tag{4.54}$$

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 $g_3^n = g_2^n$.

Now observe that every function g_3^n coincides with f on $\partial\Omega_n$, by construction and by Definition 4.15. As a consequence, if $\widetilde{\mathcal{G}}$ is the union of all the grids $\widetilde{\mathcal{G}}^n$ and we define $g_3 : \widetilde{\mathcal{G}} \to \mathbb{R}^2$ as $g_3 = g_3^n$ on each $\widetilde{\mathcal{G}}^n$, then 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 as to make it piecewise linear on each $\partial\Omega_n$; we will do that by applying Proposition 4.18. More precisely, for every $j \ge 2$ we define $\mathcal{C}^j$ as the union of all the segments of $\widetilde{\mathcal{G}}^{j-1}$ and $\widetilde{\mathcal{G}}^{j}$, and $\mathcal{C}_0^j = \partial^-\Omega_j$. We then apply 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 $g = g_3$; thus, we get a function $\hat{g}_j$, piecewise linear on $\partial^-\Omega_j$, which coincides with g (hence with g_3) off the η-neighborhood of $\partial^-\Omega_j$. As a consequence, $\hat{g}_j$ is different from g_3 only on the boundaries of squares which meet $\partial^-\Omega_j$. Then, define $\hat{g}_3 : \widetilde{\mathcal{G}} \to \mathbb{R}^2$ as $\hat{g}_3 = \hat{g}_j$ on the boundaries 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, $\hat{g}_3$ is injective, piecewise linear on each $\partial\Omega_j$, and (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 $\widetilde{\mathcal{G}}^n$, and we get a new function g_4^n, piecewise linear on $\widetilde{\mathcal{G}}^n$ and coinciding with $\hat{g}_3$ on $\partial\Omega_n$. Finally, we define $g_4 : \widetilde{\mathcal{G}} \to \mathbb{R}^2$ as $g_4 = g_4^n$ on every $\widetilde{\mathcal{G}}^n$; 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 $\widetilde{\mathcal{G}}^n$ is injective and piecewise linear on the boundary. Exactly as in Step VII of the proof of Proposition 4.19, we can 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 $f_\varepsilon : \Omega \to \mathbb{R}^2$ to coincide with f_ε^n on every Ω_n. Now 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 are smaller than $\varepsilon/4$. In addition, by construction we have

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

and then we immediately see that $f_\varepsilon(\Omega) = f(\Omega)$, $f_\varepsilon - f \in W_0^{1,1}(\Omega)$, and $f_\varepsilon = f$ on $\partial\Omega$ whenever f is continuous up to $\partial\Omega$. Hence, to conclude the proof of Theorem 1.1 for what concerns piecewise affine approximation, we just have to check that

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

This will be obtained by arguing almost exactly as in Steps VII and VIII of the proof of Proposition 4.19. More precisely, we start with Ω_1; repeating *verbatim* the arguments leading to (4.37), (4.39), (4.40) and (4.41), this time from (4.53) we get

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

For every $n \geq 2$, we instead apply Theorem 2.1—recalling also Remark 2.2—to the generic $\widetilde{\mathcal{S}}$, which is a 2-bi-Lipschitz copy of a square of side $2r$; then, also from (4.54), we get

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

Notice that this estimate is better than the corresponding one for Proposition 4.19, namely, (4.40): indeed, there we also had 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 came 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 same argument as for Proposition 4.19 implies again that every square can lie in $\mathcal{S}^+$ for at most nine different squares $\mathcal{S}$, and every side of some $\partial\Omega_n$ can lie in $\partial\mathcal{S}^+ \cap \partial\Omega_n$ for at most 11 different squares of the grid of Ω_n; as a consequence, summing (4.56) and (4.57) for all the squares of the different grids, 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 $|\mathcal{A}_4^{1,+}| \leq 9|\bigcup\{\mathcal{S}_i \in \mathcal{A}_4^1\}| < \varepsilon_2$, and where we have used (4.49), (4.45), (4.47) and (4.51). As a consequence, we have established (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 [28, Theorem A], so we have finished our proof. □

Acknowledgments. Part of this research was done when the first author was visiting University of Erlangen, and 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] Acerbi, E., Fusco, N.: Semicontinuity problems in the calculus of variations. Arch. Ration. Mech. Anal. **86**, 125–145 (1984) Zbl 0565.49010 MR 0751305

[2] Ball, J. M.: Convexity conditions and existence theorems in nonlinear elasticity. Arch. Ration. Mech. Anal. **63**, 337–403 (1977) Zbl 0368.73040 MR 0475169

[3] Ball, J. M.: Discontinuous equilibrium solutions and cavitation in nonlinear elasticity. Philos. Trans. Roy. Soc. London A **306**, 557–611 (1982) Zbl 0513.73020 MR 0703623

[4] Ball, J. M.: Singularities and computation of minimizers for variational problems. In: Foundations of Computational Mathematics (Oxford, 1999), London Math. Soc. Lecture Note Ser. 284, Cambridge Univ. Press, Cambridge, 1–20 (2001) Zbl 0978.65053 MR 1836612

[5] Ball, J. M.: Progress and puzzles in nonlinear elasticity. In: Poly-, Quasi- and Rank-One Convexity in Applied Mechanics, CISM 516, Springer, 1–16 (2010)

[6] Bellido, J. C., Mora-Corral, C.: Approximation of Hölder continuous homeomorphisms by piecewise affine homeomorphisms. Houston J. Math. **37**, 449–500 (2011) Zbl 1228.57009 MR 2794559

[7] Bing, R. H.: Locally tame sets are tame. Ann. of Math. **59**, 145–158 (1954) Zbl 0055.16802 MR 0061377

[8] Bing, R. H.: Stable homeomorphisms on E^5 can be approximated by piecewise linear ones. Notices Amer. Math. Soc. **10**, 607–616 (1963)

[9] Campbell, D.: Diffeomorphic approximation of planar Sobolev homeomorphisms in Orlicz–Sobolev spaces. J. Funct. Anal. **273**, 125–205 (2017) Zbl 06715578 MR 3646299

[10] Connell, E. H.: Approximating stable homeomorphisms by piecewise linear ones. Ann. of Math. **78**, 326–338 (1963) Zbl 0116.14802 MR 0154289

[11] Daneri, S., Pratelli, A.: Smooth approximation of bi-Lipschitz orientation-preserving homeomorphisms. Ann. Inst. H. Poincaré Anal. Non Linéaire **31**, 567–589 (2014) Zbl 1348.37071 MR 3208455

[12] Daneri, S., Pratelli, A.: A planar bi-Lipschitz extension theorem. Adv. Calc. Var. **8**, 221–266 (2015) Zbl 1331.26020 MR 3365742

[13] Donaldson, S. K., Sullivan, D. P.: Quasiconformal 4-manifolds. Acta Math. **163**, 181–252 (1989) Zbl 0704.57008 MR 1032074

[14] Evans, L. C.: Quasiconvexity and partial regularity in the calculus of variations. Ann. of Math. **95**, 227–252 (1986) Zbl 0627.49006 MR 0853966

[15] Gehring, F. W., Lehto, O.: On the total differentiability of functions of a complex variable. Ann. Acad. Sci. Fenn. Ser. A I **272**, 1–9 (1959) Zbl 0090.05302 MR 0124487

[16] Hencl, S.: Sobolev homeomorphism with zero Jacobian almost everywhere. J. Math. Pures Appl. **95**, 444–458 (2011) Zbl 1222.26018 MR 2776377

[17] Hencl, S., Koskela, P.: Lectures on Mappings of Finite Distortion. Lecture Notes in Math. 2096, Springer (2014) Zbl 1293.30051 MR 3184742

[18] Hencl, S., Vejnar, B.: Sobolev homeomorphism that cannot be approximated by diffeomorphisms in $W^{1,1}$. Arch. Ration. Mech. Anal. **219**, 183–202 (2016) Zbl 06545482 MR 3437850

[19] Iwaniec, T., Kovalev, L. V., Onninen, J.: Diffeomorphic approximation of Sobolev homeomorphisms. Arch. Ration. Mech. Anal. **201**, 1047–1067 (2011) Zbl 1260.46023 MR 2824471

[20] Iwaniec, T., Kovalev, L. V., Onninen, J.: Hopf differentials and smoothing Sobolev homeomorphisms. Int. Math. Res. Notices **2012**, 3256–3277 Zbl 1248.49052 MR 2946225

[21] Iwaniec, T., Martin, G.: Geometric Function Theory and Nonlinear Analysis. Oxford Math. Monogr., Clarendon Press, Oxford (2001) Zbl 1045.30011 MR 1859913

[22] Kirby, R. C.: Stable homeomorphisms and the annulus conjecture. Ann. of Math. **89**, 575–582 (1969) Zbl 0176.22004 MR 0242165

[23] Kirby, R. C., Siebenmann, L. C., Wall, C. T. C.: The annulus conjecture and triangulation. Notices Amer. Math. Soc. **16**, abstract 69T-G27, p. 432 (1969)

[24] Luukkainen, J.: Lipschitz and quasiconformal approximation of homeomorphism pairs. Topology Appl. **109**, 1–40 (2001) Zbl 0964.57023 MR 1804561

[25] Moise, E. E.: Affine structures in 3-manifolds. IV. Piecewise linear approximations of homeomorphisms. Ann. of Math. **55**, 215–222 (1952) Zbl 0047.16804 MR 0046644

[26] Moise, E. E.: Geometric Topology in Dimensions 2 and 3. Grad. Texts in Math. 47, Springer, New York (1977) Zbl 0349.57001 MR 0488059

[27] Mora-Corral, C.: Approximation by piecewise affine homeomorphisms of Sobolev homeomorphisms that are smooth outside a point. Houston J. Math. **35**, 515–539 (2009) Zbl 1182.57019 MR 2519545

[28] Mora-Corral, C., Pratelli, A.: Approximation of piecewise affine homeomorphisms by diffeomorphisms. J. Geom. Anal. **24**, 1398–1424 (2014) Zbl 1300.41014 MR 3223559

[29] Morrey, C. B.: Quasi-convexity and the semicontinuity of multiple integrals. Pacific J. Math. **2**, 5–53 (1952) Zbl 0046.10803 MR 0054865

[30] Müller, S., Tang, Q., Yan, B. S.: On a new class of elastic deformations not allowing for cavitation. Ann. Inst. H. Poincaré Anal. Non Linéaire **11**, 1994 (217–243) Zbl 0863.49002 MR 1267368

[31] Pratelli, A.: On the bi-Sobolev planar homeomorphisms and their approximation. Nonlinear Anal. **154**, 2017 (258–268) Zbl 1372.46029 MR 3614654

[32] Pratelli, A., Radici, E.: Approximation of planar BV homeomorphisms by diffeomorphisms. http://cvgmt.sns.it/paper/3521/ (2017)

[33] Pratelli, A., Radici, E.: On the planar minimal BV extension problem. http://cvgmt.sns.it/paper/3520/ (2017)

[34] Radici, E.: A planar Sobolev extension theorem for piecewise linear homeomorphisms. Pacific J. Math. **283**, 405–418 (2016) Zbl 1359.46036 MR 3519110

[35] Radó, T.: Über den Begriff der Riemannschen Fläche. Acta Sci. Math. (Szeged) **2**, 101–121 (1925) JFM 51.0273.01

[36] Rushing, T. B.: Topological Embeddings. Pure Appl. Math. 52, Academic Press, New York (1973) Zbl 0295.57003 MR 0348752