Rend. Lincei Mat. Appl. 29 (2018), 511–555
DOI 10.4171/RLM/819

Functional Analysis — *On the planar minimal BV extension problem*, by Aldo Pratelli and Emanuela Radici, communicated on March 9, 2018.

Abstract. — Given a continuous, injective function φ defined on the boundary of a planar open set Ω, we consider the problem of minimizing the total variation among all the BV homeomorphisms on Ω coinciding with φ on the boundary. We find the explicit value of this infimum in the model case when Ω is a rectangle. We also present two important consequences of this result: first, whatever the domain Ω is, the infimum above remains the same also if one restricts himself to consider only $W^{1,1}$ homeomorphisms. Second, any BV homeomorphism can be approximated in the strict BV sense with piecewise affine homeomorphisms and with diffeomorphisms.

Key words: Homeomorphic extension, piecewise affine approximation, functions of bounded variation

Mathematics Subject Classification: 46E35, 26B30, 49J10

1. Introduction

In this paper we consider the problem of minimizing the total variation of BV homeomorphisms on Ω extending a given boundary datum on $\partial\Omega$. More precisely, let $\Omega \subseteq \mathbb{R}^2$ be an open set, let $\varphi : \partial\Omega \to \mathbb{R}^2$ be a continuous, injective function, and let us denote by $\mathrm{Ext}(\varphi)$ the set of the homeomorphisms on Ω which are continuous up to the boundary and coincide with φ there; we are interested in the minimization of $|Du|(\Omega)$ among all the BV homeomorphisms $u \in \mathrm{Ext}(\varphi)$. Our main result is an explicit expression of the infimum in the case when Ω is a rectangle. To state it, we need the following simple definition.

Definition 1.1. Let $\mathcal{R} = [a^-, a^+] \times [b^-, b^+]$ be a rectangle, let $\varphi : \partial\mathcal{R} \to \mathbb{R}^2$ be a continuous, injective curve. We denote by $\mathcal{P} = \mathcal{P}(\varphi)$ its internal part, that is, the bounded closed set whose boundary is the image of φ. For every $x, y \in \mathcal{P}$, we call $d_{\mathcal{P}}(x, y)$ the geodesic distance in $\mathcal{P}$ between x and y, that is, the infimum of the lengths of the paths connecting x and y in the interior of $\mathcal{P}$. We define $\Psi(\varphi) \in \mathbb{R}^+$ as

$$\Psi(\varphi) = \int_{a^-}^{a^+} d_{\mathcal{P}}(\varphi(t, b^-), \varphi(t, b^+))\, dt + \int_{b^-}^{b^+} d_{\mathcal{P}}(\varphi(a^-, t), \varphi(a^+, t))\, dt.$$

Through the whole paper, we will denote by $\|\cdot\|$ the standard Manhattan, or L^1, norm in $\mathbb{R}^2$, that is, $\|(x, y)\| = |x| + |y|$. Accordingly, for a function

$u \in \mathrm{BV}(\Omega)$ we will write $\|Du\|(\Omega) = |D_1u|(\Omega) + |D_2u|(\Omega)$. Our result is then the following.

Theorem A (Explicit infimum of the total variation). *Let $\mathcal{R} = [a^-, a^+] \times [b^-, b^+]$ be a rectangle, and let $\varphi : \partial\mathcal{R} \to \mathbb{R}^2$ be a continuous, injective function. Then*

$$\inf\{\|Du\|(\mathcal{R}) : u \in \mathrm{BV}(\mathcal{R}) \cap \mathrm{Ext}(\varphi)\} = \Psi(\varphi). \tag{1.1}$$

Moreover, for every $\varepsilon > 0$ there exists a piecewise affine homeomorphism $v \in \mathrm{Ext}(\varphi)$ such that

$$\|Dv\|(\mathcal{R}) = \int_{\mathcal{R}} \|Dv\| \leq \Psi(\varphi) + \varepsilon. \tag{1.2}$$

Finally, if φ is piecewise linear then the function v above can be taken finitely piecewise affine.

In this result, as throughout the paper, by "piecewise affine function" we mean a function which is piecewise affine on each triangle of a locally finite decomposition of Ω; in other words, it is possible to write Ω as a countable but locally finite union of triangles, on each of which the function is affine.

Our theorem is not only interesting by itself; in fact, extension results of this kind are always of primary importance in order to show approximation results. The reason is simple to explain: assume that u is a homeomorphism in a given class (for instance, Sobolev or bi-Sobolev homeomorphisms), and that an approximation of u made by "good" homeomorphisms (for instance, piecewise affine ones, or diffeomorphisms) is required. Then, a convenient strategy is to subdivide the domain in small squares, and to look for an approximation inside each one, keeping the boundary values on the boundaries; at least in some "bad" squares, it can be convenient simply to take any homeomorphism having the correct boundary value and not too large energy, hence an extension result is needed.

In the last years, the search for this kind of approximation results is extremely active (see for instance [7, 6, 3, 5, 9, 11, 2]). For the reason just explained, each of these papers uses some extension result in the spirit of Theorem A. The novelty of our result is two-fold: on one hand, we are able to deal with BV homeomorphisms, while in the past only Sobolev ones were considered (hence the total variation is replaced by the L^p norm of the differential). And on the other hand, which is probably more remarkable, we are able to give an explicit expression of the infimum, while in the past only non-sharp estimates were reached.

We will also present two interesting results which follow from our main theorem. The first one asserts that, whatever the domain Ω and the boundary datum φ are, the minimal total variation of homeomorphisms extending φ is the same if one consider all the BV homeomorphisms, or just the $W^{1,1}$ ones. The second one, instead, is an approximation result for BV homeomorphisms with piecewise affine ones, or diffeomorphisms: as already explained above, extension results are always a powerful tool to prove approximation results. Since homeomorphisms are not necessarily continuous up to the boundary, we give the following definition which extends the notion of being equal on the boundary.

Definition 1.2 (Uniformly coincidence at the boundary). Let Ω be an open set, and fix any continuous, strictly increasing function $\delta \mapsto \eta(\delta)$ with $\eta(0) = 0$. Given two homeomorphisms $u, v : \Omega \to \mathbb{R}^2$, we say that *u and v uniformly coincide at* $\partial\Omega$ if, whenever $x \in \Omega$ has distance less than δ from $\mathbb{R}^2 \setminus \Omega$, one has $|u(x) - v(x)| < \eta(\delta)$.

This definition is extremely demanding: in fact, if both u and v are continuous up to the boundary, the property is stronger than having $u = v$ on $\partial\Omega$. Moreover, if both u and v belong to some $W^{1,p}$ space, this property is stronger than $u - v \in W_0^{1,p}(\Omega)$, up to choose $\eta(\delta)$ small enough. Our two consequences of Theorem A are then the following.

Theorem 1.3. *Let $\Omega \subseteq \mathbb{R}^2$ be an open set, and let $\varphi : \partial\Omega \to \mathbb{R}^2$ be a continuous, injective function. Then,*

$$\inf\{|Du|(\Omega) : u \in \mathrm{Ext}(\varphi) \cap \mathrm{BV}(\Omega)\} = \inf\left\{ \int_\Omega |Du| : u \in \mathrm{Ext}(\varphi) \cap W^{1,1}(\Omega) \right\}.$$

Theorem 1.4. *Let Ω be an open set, and let $u \in \mathrm{BV}(\Omega; \mathbb{R}^2)$ be a homeomorphism. Then, there exists a sequence $\{u_j\}$ of piecewise affine homeomorphisms, or of diffeomorphisms, each one uniformly coinciding with u at $\partial\Omega$, that converges to u uniformly and in the strict sense, while also $\{u_j^{-1}\}$ converges uniformly to u^{-1}.*

A few comments about these two results are in order. Concerning Theorem 1.3, it is known (see [5]) that the infimum of $\int_\Omega |Du|$ is the same if one considers homeomorphic extensions of φ which are in $W^{1,1}(\Omega)$, or which are piecewise affine, or which are diffeomorphisms. Hence, Theorem 1.3 says in fact the the infimum in the class of BV homeomorphisms coincides with the infima in the other three classes. Concerning Theorem 1.4 keep in mind that, as shown in [4], a BV homeomorphism is always bi-BV (that is, the inverse is also a BV homeomorphism); in addition, the two total variations coincide. In particular, in Theorem 1.4 the functions u^{-1} are also BV, and the sequence $\{u_j^{-1}\}$ also converges strictly to u^{-1}. We underline that in the proof of Theorem 1.4 the choice of the function $\eta(\delta)$ in Definition 1.2 does not play any role, so basically each element of our approximating sequence approximates u arbitrarily good around the boundary. We also mention that, in the paper [10], which is appearing contemporarily to this one, a stronger, actually sharp, version of Theorem 1.4 is shown, namely, the strict convergence can be replaced by the area-strict one (which is, roughly speaking, the strongest possible convergence in BV which is weaker than the strong one).

The plan of the paper is very simple. In Section 2, which is almost the whole paper, we prove the particular case of Theorem A when φ is piecewise linear. Then, in Section 3 we deduce the general case, and in Section 4 we prove Theorem 1.3 and Theorem 1.4.

2. The proof of Theorem A: the case when φ is piecewise linear

This section is devoted to show Theorem A in the particular case when the boundary datum φ is piecewise linear. First of all, we can start by observing that one inequality in (1.1) is more or less trivial.

Lemma 2.1. *Let $\mathcal{R} = [a^-, a^+] \times [b^-, b^+]$ be a rectangle, let $u \in \mathrm{BV}(\mathcal{R}; \mathbb{R}^2) \cap \mathrm{C}(\bar{\mathcal{R}})$ be a homeomorphism, and let φ be the restriction of u to $\partial\mathcal{R}$. Then $\|Du\|(\mathcal{R}) \geq \Psi(\varphi)$.*

Proof. First of all recall that, being u a BV function, then for almost every x the function $f(y) = u(x, y)$ is a BV function in $[b^-, b^+]$, whose distributional derivative is the measure $D_2u(x, \cdot)$: this is a standard property of BV functions, see for instance [1]. Let us then take any such x; since u is continuous, then so is f, and the total variation $|Df|([b^-, b^+])$ coincides with the length of the curve f; this latter is a curve, contained in $u(\mathcal{R}) = \mathcal{P}(\varphi)$, connecting $f(b^-) = u(x, b^-) = \varphi(x, b^-)$ with $f(b^+) = u(x, b^+) = \varphi(x, b^+)$. We derive then

$$|D_2u(x, \cdot)|([b^-, b^+]) = |Df|([b^-, b^+]) \geq d_{\mathcal{P}}(\varphi(x, b^-), \varphi(x, b^+)).$$

Since this is true for almost every $x \in [a^-, a^+]$, integrating we get

$$|D_2u|(\mathcal{R}) = \int_{x=a^-}^{a^+} |D_2u(x, \cdot)|([b^-, b^+])\, dx \geq \int_{x=a^-}^{a^+} d_{\mathcal{P}}(\varphi(x, b^-), \varphi(x, b^+))\, dx.$$

The analogous argument, done for the horizontal slicing instead of the vertical ones, gives

$$|D_1u|(\mathcal{R}) \geq \int_{y=b^-}^{b^+} d_{\mathcal{P}}(\varphi(a^-, y), \varphi(a^+, y))\, dy.$$

Adding the last two estimates, by Definition 1.1 we get

$$\|Du\|(\mathcal{R}) = |D_1u|(\mathcal{R}) + |D_2u|(\mathcal{R}) \geq \Psi(\varphi),$$

which concludes the proof. □

Thanks to this lemma, in order to get Theorem A it is enough to build a piecewise affine homeomorphism $v \in \mathrm{Ext}(\varphi)$ satisfying (1.2). In particular, in this section we assume that φ is piecewise linear, thus we need the function v to be finitely piecewise affine. In other words, the validity of Theorem A in the case of φ piecewise linear is an obvious consequence of the above Lemma 2.1 and of the following result.

Proposition 2.2. *Let $\mathcal{R} = [a^-, a^+] \times [b^-, b^+]$ be a rectangle, and let $\varphi : \partial\mathcal{R} \to \mathbb{R}^2$ be a piecewise linear and injective function. For every $\varepsilon > 0$ there exists a* *finitely piecewise affine homeomorphism* $v : \mathcal{R} \to \mathbb{R}^2$ *such that* $v = \varphi$ *on* $\partial\mathcal{R}$ *and*

$$\int_{\mathcal{R}} \|Dv\| \leq \Psi(\varphi) + \varepsilon.$$

The remaining of this section is devoted to present the proof of Proposition 2.2, which is quite involved. For the sake of clarity, we will divide our construction into four subsections.

2.1. *Some geometrical definitions and facts*

In this section, we prepare our construction with some technical geometrical definitions and some simple facts. Here, and in the following, we will call "polygons" only the connected and simply connected ones. Then, for every polygon $\mathcal{P} \subseteq \mathbb{R}^2$ there exists some rectangle $\mathcal{R} = [a^-, a^+] \times [b^-, b^+]$ and some continuous, injective and piecewise linear function $\varphi : \partial\mathcal{R} \to \mathbb{R}^2$ such that $\mathcal{P} = \mathcal{P}(\varphi)$, in the sense of Definition 1.1. Through the paper, when some points $\boldsymbol{P}_1, \boldsymbol{P}_2, \ldots, \boldsymbol{P}_H$ are given, we denote by $\boldsymbol{P}_1\boldsymbol{P}_2 \ldots \boldsymbol{P}_H$ the piecewise linear path obtained by joining the points $\boldsymbol{P}_i$ with i ranging from 1 to H.

Definition 2.3 (Geodesics and modified geodesics). Let $\mathcal{P} \subseteq \mathbb{R}^2$ be a polygon, and let $\boldsymbol{A}$ and $\boldsymbol{B}$ be any two distinct points in $\mathcal{P}$. We define $\gamma_{\boldsymbol{AB}}$ the unique geodesic (i.e., curve of minimal length) connecting them in $\mathcal{P}$: notice that $\gamma_{\boldsymbol{AB}}$ is a piecewise linear curve, having as vertices only $\boldsymbol{A}$, $\boldsymbol{B}$ and vertices of $\partial\mathcal{P}$ corresponding to internal angles of width at least π. Assume now that $\boldsymbol{A}, \boldsymbol{B} \in \partial\mathcal{P}$, and let $\boldsymbol{W}_1, \boldsymbol{W}_2, \ldots, \boldsymbol{W}_K$ be all the vertices of $\mathcal{P}$ met by $\gamma_{\boldsymbol{AB}}$, so that $\gamma_{\boldsymbol{AB}} = \boldsymbol{A}\boldsymbol{W}_1\boldsymbol{W}_2 \ldots \boldsymbol{W}_K\boldsymbol{B}$. Fix now any $\delta > 0$: for every $1 \leq i \leq K$, let $\widetilde{\boldsymbol{W}}_i \neq \boldsymbol{W}_i$ be some arbitrary point in the internal bisector of the angle at $\boldsymbol{W}_i$ having distance from $\boldsymbol{W}_i$ smaller than δ. The piecewise linear curve $\tilde{\gamma}_{\boldsymbol{AB}} = \boldsymbol{A}\widetilde{\boldsymbol{W}}_1\widetilde{\boldsymbol{W}}_2 \ldots \widetilde{\boldsymbol{W}}_K\boldsymbol{B}$ is then called a *δ-modification of* $\gamma_{\boldsymbol{AB}}$. Notice that there exists a constant $\bar{\delta}(\mathcal{P})$, depending on $\mathcal{P}$ but not on $\boldsymbol{A}$ and $\boldsymbol{B}$, such that the interior of $\tilde{\gamma}_{\boldsymbol{AB}}$ is contained in the interior of $\mathcal{P}$ if $\delta \leq \bar{\delta}(\mathcal{P})$, unless the segment $\boldsymbol{AB}$ is contained in $\partial\mathcal{P}$, in which case $K = 0$ and then $\tilde{\gamma}_{\boldsymbol{AB}} = \gamma_{\boldsymbol{AB}} \subseteq \partial\mathcal{P}$. Finally, assume that $\boldsymbol{A}$ and $\boldsymbol{B}$ are not vertices of $\mathcal{P}$, and let $\widetilde{\boldsymbol{A}}$ and $\widetilde{\boldsymbol{B}}$ be two points in $\partial\mathcal{P}$ in the same sides as $\boldsymbol{A}$ and $\boldsymbol{B}$, and with distance smaller than δ from $\boldsymbol{A}$ and $\boldsymbol{B}$. Then, the piecewise linear curve $\tilde{\gamma}_{\boldsymbol{AB}} = \widetilde{\boldsymbol{A}}\widetilde{\boldsymbol{W}}_1\widetilde{\boldsymbol{W}}_2 \ldots \widetilde{\boldsymbol{W}}_K\widetilde{\boldsymbol{B}}$ is called a *δ-modification of $\gamma_{\boldsymbol{AB}}$ with variable endpoints.*

Lemma 2.4. *Let $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$ and $\boldsymbol{D}$ be four distinct points in a polygon $\mathcal{P}$. Then the intersection $\gamma_{\boldsymbol{AB}} \cap \gamma_{\boldsymbol{CD}}$ is either empty or connected. Assume now also that the points belong to $\partial\mathcal{P}$, which is then divided in two connected parts by $\boldsymbol{C}$ and $\boldsymbol{D}$: if $\boldsymbol{A}$ and $\boldsymbol{B}$ belong to two different parts, then the intersection between $\gamma_{\boldsymbol{AB}}$ and $\gamma_{\boldsymbol{CD}}$ is surely not empty; if they belong to the same part and the intersection is not empty, then the first and last point of this intersection must be vertices of $\mathcal{P}$, unless they coincide with one of the points $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$ and $\boldsymbol{D}$.*

Proof. The connectedness of the intersection between two geodesics is an immediate consequence of the uniqueness of the geodesics. For the second part, assume first that $\boldsymbol{A}$ and $\boldsymbol{B}$ belong to the two different connected components of $\partial\mathcal{P}$ divided by $\boldsymbol{C}$ and $\boldsymbol{D}$: since any curve in $\mathcal{P}$ connecting $\boldsymbol{C}$ and $\boldsymbol{D}$ divides $\mathcal{P}$ in two or more connected components, and $\boldsymbol{A}$ and $\boldsymbol{B}$ surely belong to two different ones, we can say more in general that any curve in $\mathcal{P}$ between $\boldsymbol{A}$ and $\boldsymbol{B}$ must intersect any curve in $\mathcal{P}$ connecting $\boldsymbol{C}$ and $\boldsymbol{D}$. Suppose now that $\boldsymbol{A}$ and $\boldsymbol{B}$ belong to the same connected component of $\partial\mathcal{P}$: in this case, the two geodesics $\gamma_{\boldsymbol{AB}}$ and $\gamma_{\boldsymbol{CD}}$ may also not intersect. But, if they do, the first intersection point $\boldsymbol{P}$ must be a vertex of $\mathcal{P}$, unless it coincides with one of the points $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$ and $\boldsymbol{D}$: otherwise, $\boldsymbol{P}$ would be internal to both the geodesics, and they would be both linear for a while before and after $\boldsymbol{P}$. This would imply that in a small neighborhood of $\boldsymbol{P}$ there are no other intersection points, and since the intersection must be connected this contradicts the fact that $\boldsymbol{B}$ is on the same side of $\partial\mathcal{P}$ as $\boldsymbol{A}$. The analogous argument works for the last intersection point. □

Lemma 2.5. *Let $\mathcal{P}$ be a polygon, let $\boldsymbol{A}, \boldsymbol{B} \in \partial\mathcal{P}$ be two points such that the segment $\boldsymbol{AB}$ is not contained in $\partial\mathcal{P}$, let $\delta < \bar{\delta}(\mathcal{P})$ and let $\tilde{\gamma}_{\boldsymbol{AB}}$ be a modified geodesic in the sense of Definition* 2.3. *Let also $\mathcal{P}_1$ and $\mathcal{P}_2$ be the two polygons in which $\mathcal{P}$ is divided by $\tilde{\gamma}_{\boldsymbol{AB}}$, and let $\varepsilon > 0$ be a given constant. If δ is small enough, depending only on ε and $\mathcal{P}$, then the following is true:*

(i) *For any two points $\boldsymbol{C}, \boldsymbol{D} \in \mathcal{P}_l$, being $l \in \{1,2\}$, one has*

$$d_{\mathcal{P}_l}(\boldsymbol{C}, \boldsymbol{D}) < d_{\mathcal{P}}(\boldsymbol{C}, \boldsymbol{D}) + \varepsilon. \tag{2.1}$$

(ii) *If $\boldsymbol{C} \in \mathcal{P}_1$, $\boldsymbol{D} \in \mathcal{P}_2$, and $\boldsymbol{E} \in \partial\mathcal{P}_1 \cap \partial\mathcal{P}_2$ is any point with distance at most δ from $\gamma_{\boldsymbol{CD}}$,*

$$d_{\mathcal{P}_1}(\boldsymbol{C}, \boldsymbol{E}) + d_{\mathcal{P}_2}(\boldsymbol{E}, \boldsymbol{D}) < d_{\mathcal{P}}(\boldsymbol{C}, \boldsymbol{D}) + \varepsilon. \tag{2.2}$$

Proof. Let us begin by observing that, for any two points $\boldsymbol{P}, \boldsymbol{Q} \in \tilde{\gamma}_{\boldsymbol{AB}}$, the length of $\tilde{\gamma}_{\boldsymbol{AB}}$ between $\boldsymbol{P}$ and $\boldsymbol{Q}$ is at most the geodesic distance $d_{\mathcal{P}}(\boldsymbol{P}, \boldsymbol{Q})$ plus an error of order δ (more precisely, the error can be bounded by δ times the number of vertices of $\mathcal{P}$). Up to take $\delta \ll \varepsilon$, then, we have that the length of $\tilde{\gamma}_{\boldsymbol{AB}}$ between any two its points $\boldsymbol{P}$ and $\boldsymbol{Q}$ is at most $d_{\mathcal{P}}(\boldsymbol{P}, \boldsymbol{Q}) + \varepsilon/3$.

Take now two points $\boldsymbol{C}$ and $\boldsymbol{D}$ in $\mathcal{P}_l$, for $l \in \{1,2\}$: if $\gamma_{\boldsymbol{CD}}$ does not intersect $\tilde{\gamma}_{\boldsymbol{AB}}$, then it is entirely contained in $\mathcal{P}_l$, so we have $d_{\mathcal{P}_l}(\boldsymbol{C}, \boldsymbol{D}) = d_{\mathcal{P}}(\boldsymbol{C}, \boldsymbol{D})$, which is even more than (2.1). Otherwise, let $\boldsymbol{P}$ and $\boldsymbol{Q}$ be the first and the last point of this intersection. Then, let us call γ' the curve $\gamma_{\boldsymbol{CD}}$ with the piece between $\boldsymbol{P}$ and $\boldsymbol{Q}$ replaced by the part of $\tilde{\gamma}_{\boldsymbol{AB}}$ between $\boldsymbol{P}$ and $\boldsymbol{Q}$. Since γ' is by definition entirely contained in $\mathcal{P}_l$, by the above consideration we get

$$d_{\mathcal{P}_l}(\boldsymbol{C}, \boldsymbol{D}) \leq \mathscr{H}^1(\gamma') < \mathscr{H}^1(\gamma_{\boldsymbol{CD}}) + \frac{\varepsilon}{3} = d_{\mathcal{P}}(\boldsymbol{C}, \boldsymbol{D}) + \frac{\varepsilon}{3}, \tag{2.3}$$

which is stronger than (2.1).

Let us now take $\boldsymbol{C} \in \mathcal{P}_1$ and $\boldsymbol{D} \in \mathcal{P}_2$, and let $\boldsymbol{E} \in \tilde{\gamma}_{\boldsymbol{AB}} = \partial\mathcal{P}_1 \cap \partial\mathcal{P}_2$ be a point with distance at most δ from $\gamma_{\boldsymbol{CD}}$, so in particular there is some point $\boldsymbol{F} \in \gamma_{\boldsymbol{CD}}$

with $|\boldsymbol{F}-\boldsymbol{E}|<\delta$. Applying estimate (2.3) first to $\boldsymbol{C}$ and $\boldsymbol{E}$, which are both in $\mathcal{P}_1$, and then to $\boldsymbol{E}$ and $\boldsymbol{D}$, which are both in $\mathcal{P}_2$, and keeping in mind that $\delta \ll \varepsilon$, we get then

$$\begin{aligned} d_{\mathcal{P}_1}(\boldsymbol{C},\boldsymbol{E})+d_{\mathcal{P}_2}(\boldsymbol{E},\boldsymbol{D}) &\leq d_{\mathcal{P}}(\boldsymbol{C},\boldsymbol{E})+d_{\mathcal{P}}(\boldsymbol{E},\boldsymbol{D})+\frac{2}{3}\varepsilon \\ &\leq d_{\mathcal{P}}(\boldsymbol{C},\boldsymbol{F})+d_{\mathcal{P}}(\boldsymbol{F},\boldsymbol{D})+\frac{2}{3}\varepsilon+2\delta \\ &= d_{\mathcal{P}}(\boldsymbol{C},\boldsymbol{D})+\frac{2}{3}\varepsilon+2\delta < d_{\mathcal{P}}(\boldsymbol{C},\boldsymbol{D})+\varepsilon, \end{aligned}$$

that is, (2.2). □

Definition 2.6. Let $\mathcal{P}$ be a polygon. For every $\boldsymbol{A},\boldsymbol{B}\in\partial\mathcal{P}$, there is a unique ordered set $\mathcal{X}(\boldsymbol{A},\boldsymbol{B})=\{\boldsymbol{X}_1,\boldsymbol{X}_2,\dots,\boldsymbol{X}_N\}$ such that the geodesic in $\mathcal{P}$ between $\boldsymbol{A}$ and $\boldsymbol{B}$ is the piecewise linear curve $\boldsymbol{A}\boldsymbol{X}_1\dots\boldsymbol{X}_N\boldsymbol{B}$, and the points $\boldsymbol{X}_j$ are *all* the vertices of $\mathcal{P}$ met by the geodesic $\gamma_{\boldsymbol{AB}}$ (except $\boldsymbol{A}$ and $\boldsymbol{B}$ themselves, in case they are vertices). The set $\mathcal{X}(\boldsymbol{A},\boldsymbol{B})$ will be called the *set of the vertices of* $\gamma_{\boldsymbol{AB}}$.

The following is a simple but useful geometric property of the geodesics.

Lemma 2.7. *Let $\mathcal{P}$ be a polygon. For every $\xi \ll 1$, there exists a positive $\alpha \ll \xi$ such that, whenever $\boldsymbol{A}$ and $\boldsymbol{B}$ are two points in $\partial\mathcal{P}$, and γ is any curve in $\mathcal{P}$ connecting $\boldsymbol{A}$ and $\boldsymbol{B}$ with length less than $\mathscr{H}^1(\gamma_{\boldsymbol{AB}})+\alpha$, then the Hausdorff distance between $\gamma_{\boldsymbol{AB}}$ and γ is less than ξ. In addition, suppose that γ is a piecewise linear curve $\boldsymbol{A}\boldsymbol{X}'_1\boldsymbol{X}'_2\dots\boldsymbol{X}'_{N'}\boldsymbol{B}$, being all the $\boldsymbol{X}'_j$ vertices of $\mathcal{P}$, and let $\gamma_{\boldsymbol{AB}}=\boldsymbol{A}\boldsymbol{X}_1\dots\boldsymbol{X}_N\boldsymbol{B}$, being the points $\boldsymbol{X}_j$ as in Definition 2.6. If $\gamma\neq\gamma_{\boldsymbol{AB}}$, then there are three consecutive points, either in the set $\{\boldsymbol{A},\boldsymbol{X}'_1,\boldsymbol{X}'_2,\dots,\boldsymbol{X}'_{N'},\boldsymbol{B}\}$ or in $\{\boldsymbol{A},\boldsymbol{X}_1,\dots,\boldsymbol{X}_N,\boldsymbol{B}\}$, which are aligned up to an error ξ.*

Proof. The first property is an immediate consequence of the uniqueness of the geodesics, the continuity of the length, and the compactness of $\partial\mathcal{P}$, so we only have to deal with the second property.

For simplicity of notations, we will write $\boldsymbol{X}_0=\boldsymbol{X}'_0=\boldsymbol{A}$ and $\boldsymbol{X}_{N+1}=\boldsymbol{X}'_{N'+1}=\boldsymbol{B}$. Notice that, without loss of generality, we can assume that the distance between $\boldsymbol{A}$ and $\boldsymbol{X}_1$ is more than ξ, because otherwise the three points $\boldsymbol{A}$, $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are obviously aligned up to an error ξ, and similarly we can assume that the distances between $\boldsymbol{A}$ and $\boldsymbol{X}'_1$, $\boldsymbol{X}_N$ and $\boldsymbol{B}$, and $\boldsymbol{X}'_{N'}$ and $\boldsymbol{B}$ are more than ξ. Since $\gamma\neq\gamma_{\boldsymbol{AB}}$, there must be a "first difference" at some $k\geq 1$, that is, $\boldsymbol{X}_j=\boldsymbol{X}'_j$ for every $j<k$, but $\boldsymbol{X}_k\neq\boldsymbol{X}'_k$. Let us assume that $\mathscr{H}^1(\boldsymbol{X}_{k-1}\boldsymbol{X}_k)\leq\mathscr{H}^1(\boldsymbol{X}_{k-1}\boldsymbol{X}'_k)$: we can do this without loss of generality, since otherwise the very same argument exchanging the two curves applies. If $\boldsymbol{X}_k=\boldsymbol{B}$, then $\boldsymbol{X}'_k$ must be at a distance at most $\alpha\ll\xi$ from $\boldsymbol{B}$, and actually all the points $\boldsymbol{X}'_j$ with $j\geq k$ are within a range α from $\boldsymbol{B}$; in particular, $\boldsymbol{X}'_{k-1}$, $\boldsymbol{X}'_k$ and $\boldsymbol{X}'_{k+1}$ are aligned up to an error ξ. Therefore, we can assume that $\boldsymbol{X}_k\neq\boldsymbol{B}$. Now, the point $\boldsymbol{X}_k$ has distance smaller than ξ from

the segment $X_{k-1}X_k'$, and this implies that the three points X_{k-1}, X_k and X_k' are aligned up to an error ξ. Moreover, also X_k' has distance less than ξ from γ_{AB}; as a consequence, either X_k' is very close to the segment X_kX_{k+1}, or X_{k+1} is very close to the segment $X_{k-1}X_k'$: in both cases, the three points X_{k-1}, X_k and X_{k+1} are aligned up to an error ξ. □

Definition 2.8. Let $\mathcal{R} = [a^-, a^+] \times [b^-, b^+]$ be a rectangle, and let $\varphi : \partial\mathcal{R} \to \mathbb{R}^2$ be a piecewise linear and injective map. For any $b^- \le t \le b^+$ we will call $A_t = (a^-, t)$, $B_t = (a^+, t)$, $\boldsymbol{A}_t = \varphi(A_t)$ and $\boldsymbol{B}_t = \varphi(B_t)$, and we will denote by $\mathcal{X}(t) = \mathcal{X}(\boldsymbol{A}_t, \boldsymbol{B}_t)$ the set of vertices of $\gamma_{\boldsymbol{A}_t\boldsymbol{B}_t}$ according to Definition 2.6. For any maximal interval $I \subseteq \partial\mathcal{R}$ on which φ is linear, we will call *generalised vertex of* $\mathcal{P}$ the image of each of the two endpoints of I; moreover, whenever $\boldsymbol{A}_t$ is a generalised vertex of $\mathcal{P}$, we call generalised vertex also $\boldsymbol{B}_t$, and we do the same if $\boldsymbol{B}_t$, or $\varphi(t, b^-)$, or $\varphi(t, b^+)$ is a generalised vertex. Moreover, we call *generalised side of* $\mathcal{P}$ any interval between two consecutive generalised vertices: notice that every side of $\mathcal{P}$ is a finite union of generalised sides, and each vertex is also a generalised vertex, as well as the image of each of the four vertices of $\mathcal{R}$.

Definition 2.9. Let $\mathcal{R} = [a^-, a^+] \times [b^-, b^+]$ be a rectangle, let $\varphi : \partial\mathcal{R} \to \mathbb{R}^2$ be a piecewise linear and injective function, and let $\mathcal{P}$ be the associated polygon. We say that the map φ is *not aligned* if every three generalised vertices which are not on a same side of $\mathcal{P}$ are not aligned and moreover, for every $b^- \le t \le b^+$, calling $\mathcal{X}(t) = \{X_1, X_2, \dots, X_N\}$, if $\boldsymbol{A}_t$ is aligned with X_1 and X_2, then $\boldsymbol{B}_t$ is not aligned with X_{N-1} and X_N.

Definition 2.10. Let $\mathcal{R}$, φ and $\mathcal{P}$ be as in Definition 2.9, and let $\delta > 0$ be much smaller than the length of any side of $\mathcal{P}$. Assume that φ is linear on $\{a^-\} \times [b^-, b^+]$ and on $\{a^+\} \times [b^-, b^+]$, and that there are points $X_1, X_2, \dots, X_N \in \partial\mathcal{P}$ such that for every $t \in (b^-, b^+)$ one has $\mathcal{X}(t) = \{X_1, X_2, \dots, X_N\}$. We will say that $\mathcal{P}$ is an *upper δ-tube with number N* if it is possible to write $\varphi([a^-, a^+] \times \{b^+\}) = \boldsymbol{A}_{b^+} Y_1 Y_2 \dots Y_N \boldsymbol{B}_{b^+}$, where for any $1 \le j \le N$ the point Y_j lies on the internal bisector of the angle at X_j and $0 \le |Y_j - X_j| < \delta$: notice that it is admissible to choose $Y_1 = \boldsymbol{A}_{b^+}$, as well as $Y_N = \boldsymbol{B}_{b^+}$. The polygon $\mathcal{P}$ will be said a *lower δ-tube with number N* if the analolgous property, with points Z_j instead of Y_j, holds for $\varphi([a^-, a^+] \times \{b^-\})$. If $\mathcal{P}$ is both an upper and a lower δ-tube, we will say that it is a *δ-tube*.

Figure 1 depicts the situation of an upper δ-tube and of a δ-tube. Notice that, if $\mathcal{P}$ is a δ-tube, then for every $1 \le j \le N$ it must be either $X_j = Y_j$, or $X_j = Z_j$.

2.2. *The proof of Lemma 2.11*

This subsection is devoted to show the main brick of our construction. Basically, we are going to take a polygon and subdivide it in several subregions: the top one

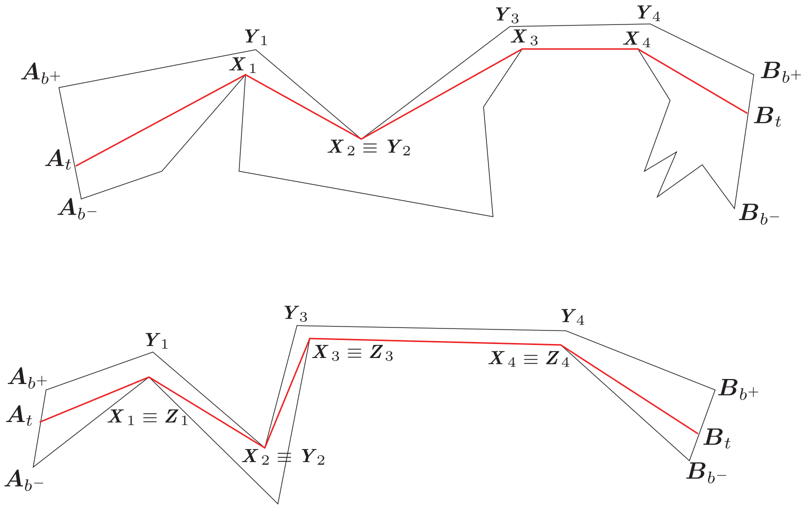

Figure 1. An upper δ-tube and a δ-tube.

will be a lower δ-tube, the bottom one will be an upper δ-tube, and all the internal ones will be δ-tubes.

LEMMA 2.11. *Let $\mathcal{R} = [a^-, a^+] \times [b^-, b^+]$ be a rectangle, let $\varphi : \partial\mathcal{R} \to \mathbb{R}^2$ be a piecewise linear, injective and not aligned function, and let $\mathcal{P} = \mathcal{P}(\varphi)$ be the associated polygon. Then, for every $\eta > 0$ there exist finitely many ordinates $b^- = y_0 < y_1 < \cdots < y_{M-1} < y_M = b^+$, such that $y_{i+1} < y_i + \eta$ for every $0 \leq i < M$, and there exists also a piecewise linear and injective function $\tilde{\varphi} : \bigcup_{i=0}^{M-1} \partial\mathcal{R}_i \to \mathbb{R}^2$, where $\mathcal{R}_i = [a^-, a^+] \times [y_i, y_{i+1}]$, so that $\tilde{\varphi} = \varphi$ on $\partial\mathcal{R}$ and*

$$\sum_{i=0}^{M-1} \Psi(\varphi_i) < \Psi(\varphi) + \eta(b^+ - b^-), \tag{2.4}$$

being φ_i the restriction of $\tilde{\varphi}$ to $\partial\mathcal{R}_i$. Moreover, for each $0 \leq i \leq M-1$ there is a finitely piecewise affine bijection $\Phi_i : \mathbb{R}^2 \to \mathbb{R}^2$, with bi-Lipschitz constant smaller than $1 + \eta$, and there is $\delta_i \ll \eta/M$ much smaller than the length of any side of $\mathcal{P}(\varphi_i)$, such that $\mathcal{P}(\Phi_i \circ \varphi_i)$ is an upper δ_i-tube with number N_i if $i < M-1$, and a lower δ_i-tube with number N_i if $i > 0$, and $N_i \leq T$, being T the number of vertices of $\mathcal{P}(\varphi)$.

PROOF. We can assume that every two consecutive generalised vertices in $\mathcal{P}$ have distance smaller than η. This is of course admissible: up to an arbitrarily small reparameterization of φ we can add generalised vertices to $\mathcal{P}$, in particular we can add a finite number of generalised vertices, so that the distance between any two consecutive ones becomes smaller than η, while φ remains piecewise linear, injective and not aligned. We divide our construction in a few steps.

Step I. Definition of the ordinates t_i, $0 \leq i \leq M$.

Assume that, for some $t < t'$, the points $\boldsymbol{A}_t$ and $\boldsymbol{A}_{t'}$ given by Definition 2.8 belong to a same side of $\mathcal{P}$, and the same happens to $\boldsymbol{B}_t$ and $\boldsymbol{B}_{t'}$, and moreover $\mathcal{X}(t) = \mathcal{X}(t')$: then, by Lemma 2.4 one immediately gets that $\mathcal{X}(t'') = \mathcal{X}(t) = \mathcal{X}(t')$ for any $t < t'' < t'$. An immediate consequence of this fact, together with the fact that the generalised sides of $\mathcal{P}$ are finitely many, and so are also the possible values of $\mathcal{X}(t)$, is the following. There exists finitely many ordinates $b^- = t_0 < t_1 < \cdots < t_{M-1} < t_M = b^+$, and corresponding ordered sets $\mathcal{X}_i$ for $0 \leq i < M$ such that, for each $0 \leq i < M$, the points $\boldsymbol{A}_t$ with $t_i < t < t_{i+1}$ belong all to a same generalised side of $\mathcal{P}$, the same happens to the points $\boldsymbol{B}_t$, and the ordered sets $\mathcal{X}(t)$ all equal $\mathcal{X}_i$. In particular, we choose the "minimal" such sequence $\{t_i\}$, in the sense that, for every $1 \leq i < M-1$, either $\boldsymbol{A}_{t_i}$ and $\boldsymbol{B}_{t_i}$ are generalised vertices of $\mathcal{P}$, or $\mathcal{X}_i \neq \mathcal{X}_{i-1}$. Of course, whenever $\boldsymbol{A}_t$ and $\boldsymbol{B}_t$ are generalised vertices, then t is one of the coordinates t_i. Suppose, on the other hand, that for some $1 \leq i < M$ the points $\boldsymbol{A}_{t_i}$ and $\boldsymbol{B}_{t_i}$ are not generalised vertices, so necessarily $\mathcal{X}_i \neq \mathcal{X}_{i-1}$; by continuity of the length, if we call $\mathcal{X}_{i-1} = \{\boldsymbol{X}_1, \boldsymbol{X}_2, \dots, \boldsymbol{X}_N\}$ and $\mathcal{X}_i = \{\boldsymbol{X}'_1, \boldsymbol{X}'_2, \dots, \boldsymbol{X}'_{N'}\}$, then both the paths $\boldsymbol{A}_{t_i}\boldsymbol{X}_1 \dots \boldsymbol{X}_N\boldsymbol{B}_{t_i}$ and $\boldsymbol{A}_{t_i}\boldsymbol{X}'_1 \dots \boldsymbol{X}'_{N'}\boldsymbol{B}_{t_i}$ are geodesics in $\mathcal{P}$ between $\boldsymbol{A}_{t_i}$ and $\boldsymbol{B}_{t_i}$, so by uniqueness they coincide. Since $\mathcal{X}_{i-1} \neq \mathcal{X}_i$, and since $\boldsymbol{A}_{t_i}$ and $\boldsymbol{B}_{t_i}$ are not generalised vertices, the only possibility is that $\boldsymbol{A}_{t_i}$ is aligned with the first two vertices of $\mathcal{X}_{i-1}$ or of $\mathcal{X}_i$, or $\boldsymbol{B}_{t_i}$ is aligned with the last two vertices of $\mathcal{X}_{i-1}$ or of $\mathcal{X}_i$ (and only one of these things can happen, since φ is not aligned). Observe also that $t_{i+1} - t_i < \eta$ for every i, since the maximal distance between two consecutive generalised vertices is less than η. Finally, notice that the set of coordinates $\{t_i\}$ depends on the polygon $\mathcal{P}$ and on the parameterization of φ on the two vertical sides $\{a^-\} \times [b^-, b^+]$ and $\{a^+\} \times [b^-, b^+]$, but *not* on the parameterization of φ on the two horizontal sides. The number M is already the one of the claim, and each ordinate y_i will be very close to the corresponding t_i.

Step II. Definition of y_1 and of $\tilde{\gamma}_1 = \tilde{\varphi}([a^-, a^+] \times \{y_1\})$.

The goal of this step is to define the ordinate y_1 and the curve $\tilde{\gamma}_1$, internal to $\mathcal{P}$, which will be the image of the segment $[a^-, a^+] \times \{y_1\}$ under $\tilde{\varphi}$. The precise parameterization of $\tilde{\varphi}$ on $[a^-, a^+] \times \{y_1\}$ will be presented in the next step, where we will also take care of (2.4): in this step, we only aim to define the curve $\tilde{\gamma}_1 \subseteq \mathbb{R}^2$.

Before doing that, we observe that the curve $\tilde{\gamma}_1$ will divide the polygon $\mathcal{P}$ in two polygons, namely, a polygon $\mathcal{P}_0$ which contains the image of $[a^-, a^+] \times \{b^-\}$ under φ, and the remaining part $\mathcal{P}^+$; in particular, $\mathcal{P}^+$ will be the polygon corresponding to the function $\tilde{\varphi}$ on the boundary of $[a^-, a^+] \times [y_1, b^+]$. As a consequence, as soon as $\tilde{\gamma}_1$ is defined, it will be possible to repeat the definition of the coordinates t_i for the polygon $\mathcal{P}^+$: indeed, as noticed above, this construction does not depend on the precise parameterization of $\tilde{\varphi}$ on the horizontal side $[a^-, a^+] \times \{y_1\}$, which will be presented in the next step. Hence, we will find coordinates $y_1 = t'_0 < t'_1 < \cdots < t'_{M'-1} < t'_{M'} = b^+$. Our definition of y_1 and of $\tilde{\gamma}_1$ will be done in such a way that

$$M' = M - 1, \tag{2.5}$$

and actually every t'_i will be very close to t_{i+1}. The basic idea of the construction is the following: one would like to set $y_1 = t_1$, and to let $\tilde{\gamma}_1$ be a modification, in the sense of Definition 2.3, of the geodesic between $\boldsymbol{A}_{y_1}$ and $\boldsymbol{B}_{y_1}$. This is not always possible, because it could generate a map on the boundary of the rectangle $[a^-, a^+] \times [y_1, b^+]$ which fails to be not aligned, while we want to have a not aligned map, so to be able to argue by recursion.

Let us now recall that the map φ is not aligned: as a consequence, there is some constant $\xi > 0$ such that no three generalised vertices of φ are aligned up to an error 2ξ; we can also assume that ξ is much smaller than the length of any of the segments $\boldsymbol{A}_{t_i}\boldsymbol{A}_{t_{i+1}}$ and $\boldsymbol{B}_{t_i}\boldsymbol{B}_{t_{i+1}}$. We let α be the number corresponding to the constant ξ and the polygon $\mathcal{P}$ in Lemma 2.7.

We divide our construction of y_1 and $\tilde{\gamma}_1$ in two cases.

<u>Step IIa.</u> *The case when $\boldsymbol{A}_{t_1}$ and $\boldsymbol{B}_{t_1}$ are generalised vertices.*

First of all, let us suppose that $\boldsymbol{A}_{t_1}$ and $\boldsymbol{B}_{t_1}$ are generalised vertices. In this case, we let $y_1 = t_1$ and we call $\gamma_1 = \gamma_{\boldsymbol{A}_{y_1}\boldsymbol{B}_{y_1}}$ the geodesic in $\mathcal{P}$ between $\boldsymbol{A}_{y_1}$ and $\boldsymbol{B}_{y_1}$. Let us write $\mathcal{X}_0 = \{\boldsymbol{X}_1, \boldsymbol{X}_2, \dots, \boldsymbol{X}_N\}$, and let us notice that the vertices of γ_1 which are not endpoints are all the points $\boldsymbol{X}_j$, except $\boldsymbol{X}_1$ if $\boldsymbol{A}_{y_1} = \boldsymbol{X}_1$, and except $\boldsymbol{X}_N$ if $\boldsymbol{B}_{y_1} = \boldsymbol{X}_N$. Let now $\tilde{\gamma}_1$ be a δ-modification of γ_1 in the sense of Definition 2.3, where $\delta \ll \eta/M$ will be precised later: we can write $\tilde{\gamma}_1 = \boldsymbol{A}_{y_1}\boldsymbol{Y}_1\boldsymbol{Y}_2\dots\boldsymbol{Y}_N\boldsymbol{B}_{y_1}$ with $|\boldsymbol{Y}_j - \boldsymbol{X}_j| < \delta$ for every $1 \le j \le N$, and with $\boldsymbol{Y}_1 = \boldsymbol{X}_1$ if $\boldsymbol{A}_{y_1} = \boldsymbol{X}_1$, and $\boldsymbol{Y}_N = \boldsymbol{X}_N$ if $\boldsymbol{B}_{y_1} = \boldsymbol{X}_N$. Figure 2 depicts this situation.

In order to check that this curve satisfies our requirements, in particular that (2.5) holds, we need to study the shape of the geodesics between $\boldsymbol{A}_t$ and $\boldsymbol{B}_t$ in $\mathcal{P}_0$ for every $b^- \le t \le y_1$, and in $\mathcal{P}^+$ for $y_1 \le t \le b^+$. We start now with the case of $t \ge y_1$, the other case (which is analogous, yet much simpler) will be done later. Let us then consider the geodesic γ^+ between $\boldsymbol{A}_t$ and $\boldsymbol{B}_t$ in $\mathcal{P}^+$, and call $\mathcal{X}'(t)$ its set of vertices. A first guess could be that $\mathcal{X}'(t) = \mathcal{X}(t)$, but this is false, and what one really needs is actually something a bit different.

We need to be more precise here. Let us write $\mathcal{X}(t) = \{\boldsymbol{V}_1, \boldsymbol{V}_2, \dots, \boldsymbol{V}_K\}$, and notice that the points $\boldsymbol{V}_l$ are all vertices of $\mathcal{P}$, and above γ_1 by Lemma 2.4. Let us now concentrate for a moment on a given point $\boldsymbol{V}_l$. If this point does not equal

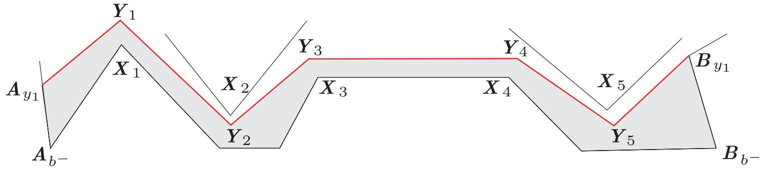

Figure 2. The situation in Step IIa: the coloured region is $\mathcal{P}_0$.

any of the points $\boldsymbol{X}_i$, then it is also a vertex of $\partial\mathcal{P}^+$, and a reasonable guess, that we will investigate later, is that it is also an element of $\mathcal{X}'(t)$.

Suppose instead that $\boldsymbol{V}_l$ coincides with some point $\boldsymbol{X}_j$. In this case, $\boldsymbol{V}_l$ could be a point of $\partial\mathcal{P}^+$, as $\boldsymbol{X}_2$ and $\boldsymbol{X}_5$ in Figure 2, or a point not in $\partial\mathcal{P}^+$, as $\boldsymbol{X}_1$, $\boldsymbol{X}_3$ and $\boldsymbol{X}_4$. If $\boldsymbol{V}_l = \boldsymbol{X}_j$ does not belong to $\partial\mathcal{P}^+$, then of course it cannot be an element of $\mathcal{X}'(t)$, and this shows that trying to prove the equality $\mathcal{X}(t) = \mathcal{X}'(t)$ would be pointless. However, in this case the reasonable guess (to be investigated later) is that $\mathcal{X}'(t)$ contains the corresponding point $\boldsymbol{Y}_j$, which is in fact in $\partial\mathcal{P}^+$. Finally, if $\boldsymbol{V}_l = \boldsymbol{X}_j$ belongs to $\partial\mathcal{P}^+$, then both the points $\boldsymbol{X}_j$ and $\boldsymbol{Y}_j$ are in fact in $\partial\mathcal{P}^+$. Nevertheless, the point $\boldsymbol{Y}_j$ is surely not in $\mathcal{X}'(t)$, since it corresponds to an angle smaller than π, and this time the obvious guess is again that $\mathcal{X}'(t)$ should contain $\boldsymbol{X}_j$.

Summarizing, it seems reasonable to believe that, for each $1 \leq l \leq K$, the set $\mathcal{X}'(t)$ contains the point $\boldsymbol{V}_l$ whenever it belongs to $\partial\mathcal{P}^+$, and the point $\boldsymbol{Y}_j$ whenever $\boldsymbol{V}_l \notin \partial\mathcal{P}^+$, where the index j is uniquely identified by the equality $\boldsymbol{V}_l = \boldsymbol{X}_j$. As a consequence, for every $\boldsymbol{V}_l \in \mathcal{X}(t)$, we set $\widetilde{\boldsymbol{V}}_l = \boldsymbol{V}_l$ whenever $\boldsymbol{V}_l \in \partial\mathcal{P}^+$, while otherwise, if $\boldsymbol{V}_l = \boldsymbol{X}_j \notin \partial\mathcal{P}^+$, we set $\widetilde{\boldsymbol{V}}_l = \boldsymbol{Y}_j$. We will write $\mathcal{X}(t) \approx \mathcal{X}'(t)$ whenever $\mathcal{X}'(t) = \{\widetilde{\boldsymbol{V}}_1, \widetilde{\boldsymbol{V}}_2, \dots, \widetilde{\boldsymbol{V}}_K\}$.

With this notation in mind, we apply Lemma 2.5 to get that, if δ is small enough depending on $\mathcal{P}$ and on α, then $\mathscr{H}^1(\gamma^+) < \mathscr{H}^1(\gamma) + \alpha/2$, denoting for brevity by $\gamma = \gamma_{\boldsymbol{A}_t\boldsymbol{B}_t}$ the geodesic between $\boldsymbol{A}_t$ and $\boldsymbol{B}_t$ in $\mathcal{P}$. Keep in mind that every vertex of γ^+ is a vertex of $\mathcal{P}^+$, so it is either a vertex of $\mathcal{P}$ or a vertex of $\tilde{\gamma}_1$; let us then define γ' the modification of γ^+ made by keeping all its vertices which are on $\partial\mathcal{P}$, while substituting every vertex $\boldsymbol{Y}_j$ in $\tilde{\gamma}_1$ with the corresponding vertex $\boldsymbol{X}_j$ in $\partial\mathcal{P}$. By construction, γ' is a piecewise linear curve in $\mathcal{P}$ between $\boldsymbol{A}_t$ and $\boldsymbol{B}_t$, whose vertices all belong to $\partial\mathcal{P}$, and its length is extremely close to that of γ^+, in particular $\mathscr{H}^1(\gamma') < \mathscr{H}^1(\gamma) + \alpha$. By Lemma 2.7, we deduce that either $\gamma' = \gamma$, or there are three vertices of γ or of γ' which are aligned up to an error ξ. Notice that the equality $\gamma' = \gamma$ is equivalent to say that $\mathcal{X}(t) \approx \mathcal{X}'(t)$.

Let us consider the possibility that $\gamma' \neq \gamma$, so there are three consecutive vertices of γ or of γ' which are aligned up to an error ξ. By construction, this never happens with three vertices of $\mathcal{P}$, and since all the vertices of γ and of γ' are vertices of $\mathcal{P}$ except possibly $\boldsymbol{A}_t$ and $\boldsymbol{B}_t$, we obtain that the only possibility, in order to have $\mathcal{X}(t) \not\approx \mathcal{X}'(t)$, is that $\boldsymbol{A}_t$ is aligned, up to an error ξ, with the first two points of γ, or of γ', or $\boldsymbol{B}_t$ is aligned, up to an error ξ, with the last two points of γ or of γ': in particular, this can happen only if $\boldsymbol{A}_t$ and $\boldsymbol{B}_t$ have at least distance ξ from every generalised vertex of $\mathcal{P}$. Therefore, the equality $\mathcal{X}(t) \approx \mathcal{X}'(t)$ is surely true around every generalised vertex (which is also by definition one of the points t_i). We want to show that the equality is also true if t is not very close to some of the points t_i.

To do so, let us assume that $\mathcal{X}(t) \not\approx \mathcal{X}'(t)$: as already pointed out, this can happen only if $\boldsymbol{A}_t$ is aligned, up to an error ξ, with the first two points of γ (or of γ'), or the analogous property holds for $\boldsymbol{B}_t$. Let us say that $\boldsymbol{A}_t$ is aligned up to an error ξ with $\boldsymbol{P}$ and $\boldsymbol{Q}$, the first two vertices of one of the paths γ and γ': then, $\boldsymbol{A}_t$ belongs to the interior of some generalised side of $\mathcal{P}$, which also contains a point $\boldsymbol{A}_{\hat{t}}$ aligned with $\boldsymbol{P}$ and $\boldsymbol{Q}$. We want to show that there is some t_i between

t and $\hat{t}$ (included): indeed, if there is no t_i in the open interval between t and $\hat{t}$, then by continuity the geodesic $\gamma_{\boldsymbol{A}_i\boldsymbol{B}_i}$ passes through both $\boldsymbol{P}$ and $\boldsymbol{Q}$. And in turn, this implies that $\mathcal{X}$ changes at $\hat{t}$, so $\hat{t}$ itself is one of the points t_i. In other words, every t'_i must be very close to some t_j.

To conclude that $M' = M - 1$ and that for every $1 \leq i \leq M$ one has $t'_{i-1} \approx t_i$, which is even stronger than (2.5), we have to show that around every t_i with $i \geq 1$ there is *exactly one* t'_j. Since we have already noticed that $\mathcal{X}(t) = \mathcal{X}'(t)$ whenever t is close to a generalised vertex, and on the other hand every generalised vertex must correspond to one of the t_i (as well as one of the t'_i), we immediately obtain what we wanted around every t_i corresponding to a generalised vertex of $\mathcal{P}$ (hence, also of $\mathcal{P}^+$); so, in particular there is no problem around t_1. To conclude, let us restrict our attention to a small neighborhood of some t_i corresponding to points $\boldsymbol{A}_{t_i}$ and $\boldsymbol{B}_{t_i}$ which are not generalised vertices, so in particular $\mathcal{X}_{i-1} \neq \mathcal{X}_i$. As we have already noticed, this implies that either $\boldsymbol{A}_{t_i}$ is aligned with the first two vertices of the geodesic $\gamma_{\boldsymbol{A}_{t_i}\boldsymbol{B}_{t_i}}$, or $\boldsymbol{B}_{t_i}$ with the last two; by symmetry, let us assume that $\boldsymbol{A}_{t_i}$ is aligned with the first two points, call them $\boldsymbol{P}$ and $\boldsymbol{Q}$. The points $\boldsymbol{P}$ and $\boldsymbol{Q}$ need not necessarily to be also vertices of $\mathcal{P}^+$, but we call again $\widetilde{\boldsymbol{P}}$ (resp., $\widetilde{\boldsymbol{Q}}$) the vertex of $\partial\mathcal{P}^+$ which coincides with $\boldsymbol{P}$ (resp., $\boldsymbol{Q}$) or is very close to it: recall that the distance between them is smaller than δ, which is by construction much smaller than ξ. The point $\boldsymbol{A}_{t_i}$ need not to be aligned with $\widetilde{\boldsymbol{P}}$ and $\widetilde{\boldsymbol{Q}}$; nevertheless, since $\boldsymbol{A}_{t_i}$ is not close to a generalised vertex of $\mathcal{P}$, on the same side of $\partial\mathcal{P}$ to which $\boldsymbol{A}_{t_i}$ belongs, there is for sure *exactly one* point, say $\boldsymbol{A}_{\hat{t}}$, which is aligned with $\widetilde{\boldsymbol{P}}$ and $\widetilde{\boldsymbol{Q}}$. We deduce that $\hat{t}$ is one of the coordinates t'_j, and actually the only one near t_i. Then, we have obtained (2.5).

We can now pass to consider the case of $t \leq y_1$: the situation is completely analogous to the one with $t \geq y_1$, yet much simpler because there is no t_j between t_0 and $t_1 = y_1$, and because $\boldsymbol{A}_{y_0}$ can not be aligned with $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ because φ is not aligned. Then, the very same argument as for $t \geq y_1$ implies that the equality $\mathcal{X}(t) \approx \mathcal{X}'(t)$ holds for every $t_0 < t < y_1$. In particular, $\mathcal{X}'(t) = \{\widetilde{\boldsymbol{X}}_1, \widetilde{\boldsymbol{X}}_2, \dots, \widetilde{\boldsymbol{X}}_N\}$, where for every $1 \leq j \leq N$ one has $\widetilde{\boldsymbol{X}}_j = \boldsymbol{X}_j$ if $\boldsymbol{X}_j \in \partial\mathcal{P}_0$, and otherwise $\widetilde{\boldsymbol{X}}_j = \boldsymbol{Y}_j$: for instance, in the situation of Figure 2, one has $\widetilde{\boldsymbol{X}}_j = \boldsymbol{X}_j$ for $j = 1, 3$ and 4, while $\widetilde{\boldsymbol{X}}_j = \boldsymbol{Y}_j$ for $j = 2$ and 5. Thus, we obtain that $\mathcal{P}_0$ is an upper δ-tube in the sense of Definition 2.10: notice that, in this case, Φ_0 is the identity map.

Since we will argue by recursion, we will need to be sure that the map $\tilde{\varphi}$ is not aligned on the boundary of the rectangle $[a^-, a^+] \times [y_1, b^+]$. At this moment, we cannot check the validity of this property, because we still didn't give the precise parameterization of $\tilde{\varphi}$ on the segment $[a^-, a^+] \times \{y_1\}$, we only decided that its image is the curve $\tilde{\gamma}_1$. Nevertheless, we can already check almost everything: more precisely, since φ is not aligned, we know that, for every $y_1 \leq t \leq b^+$, if $\boldsymbol{A}_t$ is aligned with the first two vertices of $\mathcal{X}(t)$, then $\boldsymbol{B}_t$ is not aligned with the last two. Since this alignment can only happen at ordinates t coinciding with some t_j, by the above construction we have the same property also in the polygon $\mathcal{P}^+$ with the sets of vertices $\mathcal{X}'(t)$, as soon as δ is small enough. Then, to obtain that $\tilde{\varphi}$ is not aligned on $[a^-, a^+] \times [y_1, b^+]$, we need to check that every three generalised vertices of $\tilde{\varphi}$ on $\mathcal{P}^+$ are not aligned. Again, this is a property that we already know for φ on $\mathcal{P}$; thus, again up to take δ small enough, the non-alignment is surely true for every three points taken in the set of the generalised vertices of $\mathcal{P}$, plus the vertices of $\mathcal{P}^+$, since every vertex of $\mathcal{P}^+$ has distance at most δ from some vertex of $\mathcal{P}$. To conclude, we will only need to take care also of the points of $\tilde{\gamma}_1$ which will be generalised vertices but not vertices of $\mathcal{P}^+$. We cannot do this now, because the generalised vertices of $\mathcal{P}^+$ depend on the parameterisation of $\tilde{\varphi}$, which will be done in the next step.

Step IIb. *The case when $\boldsymbol{A}_{t_1}$ and $\boldsymbol{B}_{t_1}$ are not generalised vertices (hence $\mathcal{X}_0 \neq \mathcal{X}_1$).*

Let us now consider the second possible case for t_1, namely, that the points $\boldsymbol{A}_{t_1}$ and $\boldsymbol{B}_{t_1}$ are not generalised vertices. As already pointed out, this means that either $\boldsymbol{A}_{t_1}$ is aligned with the first two vertices of $\mathcal{X}(t_1)$, or $\boldsymbol{B}_{t_1}$ with the last two, and the two things cannot happen contemporarily because φ is not aligned. Let us assume without loss of generality that $\boldsymbol{A}_{t_1}$ is aligned with the first two vertices of $\mathcal{X}(t)$, the case for $\boldsymbol{B}_{t_1}$ is of course identical.

What we will do, is again to define γ_1 the geodesic in $\mathcal{P}$ between $\boldsymbol{A}_{t_1}$ and $\boldsymbol{B}_{t_1}$. This time, we cannot choose $y_1 = t_1$, in fact y_1 will be very close to t_1 but different from it. As a consequence, while in Step IIa the curve $\tilde{\gamma}_1$ was a *generic* δ-modification of γ_1, this time $\tilde{\gamma}_1$ will be a *suitable* δ-modification of γ_1 *with variable endpoints*, in the sense of Definition 2.3. In particular, $\tilde{\gamma}_1$ must connect $\boldsymbol{A}_{y_1}$ with $\boldsymbol{B}_{y_1}$.

This definition has to be made in such a way that (2.5) is still valid; moreover, as in Step IIa, we have to check that the polygon $\mathcal{P}_0$ is an upper δ-tube, up to a finitely piecewise affine bijection close to the identity, and that there is no obstruction to the property of $\tilde{\varphi}$ of being not aligned on $[a^-, a^+] \times [y_1, b^+]$. Most of the proof in this step will coincide with the analogous parts in Step IIa, we will only need a few modifications.

Let us be more precise: once y_1 and the curve $\tilde{\gamma}_1$ are defined, we can repeat verbatim everything that we have done in Step IIa, except the parts in which we used that $\boldsymbol{A}_{y_1}$ and $\boldsymbol{B}_{y_1}$ were generalised vertices, which is no more true in this case. In particular, in Step IIa we checked first that the equality $\mathcal{X}(t) \approx \mathcal{X}'(t)$ holds for every t not close to some t_i, and then that it also holds around every t_i corresponding to a generalised vertex: the very same arguments work also in this case. Then, we showed that around every t_i not corresponding to a generalised vertex there was exactly one t'_j. More precisely, we did the following: we took t_i not corresponding to a generalised vertex, so with $\boldsymbol{A}_{t_i}$ aligned with the first two vertices $\boldsymbol{P}$ and $\boldsymbol{Q}$ of $\mathcal{X}(t_i)$ (or $\boldsymbol{B}_{t_i}$ with the last two); then we observed that, in the same generalised side containing $\boldsymbol{A}_{t_i}$, there was surely exactly one point, say $\boldsymbol{A}_{\hat{t}}$, aligned with $\widetilde{\boldsymbol{P}}$ and $\widetilde{\boldsymbol{Q}}$ (the points in $\partial\mathcal{P}^+$ corresponding to $\boldsymbol{P}$ and $\boldsymbol{Q}$); and finally, we deduced that $\hat{t}$ was one of the coordinates t'_j, and in fact the only one near t_i. The very same argument can be done now, except around t_1. Indeed, the generalised side containing $\boldsymbol{A}_{t_1}$ is not completely in $\mathcal{P}^+$, it has a part in $\mathcal{P}^+$ and a part in $\mathcal{P}_0$, so the existence of the coordinate $\hat{t}$ does not work as for the case $i > 1$. Moreover, since by construction we already have $t'_0 = t_1$, we do not want to find any new coordinate t'_j near t_1: such a new coordinate bigger than y_1, so in $\mathcal{P}^+$, would let (2.5) fail, while smaller than y_1, so in $\mathcal{P}_0$, would let the property of $\mathcal{P}_0$ of being an upper δ-tube fail. In other words, there must be no special coordinate near t_1. It is important to observe that we did not have such a problem in Step IIa, because in that case t_1 was corresponding to a generalised vertex, so $\mathcal{X}(t) \approx \mathcal{X}'(t)$ was ensured around t_1. This is the reason why we need now to select a particular modification of γ_1, not simply taking a generic one as in the previous step.

Instead, the last part of Step IIa can be again repeated also in this case. More precisely, we considered the property of $\tilde{\varphi}$ of being not aligned. First we observed that, for every $t \geq y_1$, if $\boldsymbol{A}_t$ is aligned with the first two vertices of $\mathcal{X}'(t)$, then $\boldsymbol{B}_t$ is not aligned with the last two, as soon as δ is small enough; and the very same argument works perfectly also in this case. Then, we noticed that there are no three aligned points among the generalised vertices of $\mathcal{P}$ and the vertices of $\tilde{\gamma}_1$: this was an obvious consequence, again for δ small enough, of the fact that every vertex of $\tilde{\gamma}_1$ was very close to a vertex of $\mathcal{P}$. We have to slightly modify the argument now, since it is no more true that vertices of $\tilde{\gamma}_1$ are close to vertices of $\mathcal{P}$, in particular this fails for $\boldsymbol{A}_{y_1}$ and $\boldsymbol{B}_{y_1}$ (we did not have this problem before because it was $y_1 = t_1$ there, and $\boldsymbol{A}_{t_1}$ and $\boldsymbol{B}_{t_1}$ were generalised vertices). However, as soon as y_1 is very close to t_1 *and not equal to it*, we have that $\boldsymbol{A}_{y_1}$ and $\boldsymbol{B}_{y_1}$ are not aligned to any two generalised vertices of $\mathcal{P}$. In other words, also in this case we have no obstruction to the property of $\tilde{\varphi}$ of being not aligned, and again we will have only to take care of the non alignment for the generalised vertices of $\mathcal{P}^+$. And in turn, this is something we can only do in next step, having defined the precise parameterisation of $\tilde{\varphi}$.

Summarizing, what we have to do in this step is only to select some y_1, close to t_1 but different from it, and to define a δ-modification $\tilde{\gamma}_1$ of γ_1 with variable endpoints, connecting $\boldsymbol{A}_{y_1}$ with $\boldsymbol{B}_{y_1}$, in such a way that $\mathcal{P}_0$ is an upper δ-tube, up to a finitely piecewise affine bijection close to the identity, and that there are no coordinates t'_j close to $t_1 = t'_0$ in $\mathcal{P}^+$ or in $\mathcal{P}_0$ (this means that $\mathcal{X}'(t)$ is constant in an upper neighborhood of y_1, as well as in a lower neighborhood of it).

To do so, we have to further distinguish this case into four possible subcases; let us describe how. As we said, the point $\boldsymbol{A}_{t_1}$ is aligned with $\boldsymbol{P}$ and $\boldsymbol{Q}$, which are the first two points of $\mathcal{X}(t)$. Then, there are two points $P = (p_1, p_2)$ and $Q = (q_1, q_2)$ in $\partial\mathcal{R}$ such that $\varphi(P) = \boldsymbol{P}$ and $\varphi(Q) = \boldsymbol{Q}$. Notice that $p_2 \neq t_1$ and $q_2 \neq t_1$: indeed, the only points in $\partial\mathcal{R}$ with t_1 as second coordinate are A_{t_1} and B_{t_1}, and on the other hand $\boldsymbol{P}$ and $\boldsymbol{Q}$ are surely different from $\boldsymbol{A}_{t_1}$ and $\boldsymbol{B}_{t_1}$. Hence, we have that either $p_2 < t_1$, or $p_2 > t_1$: this means that the point $\boldsymbol{P}$ is respectively below or above the geodesic γ_1, by Lemma 2.4; similarly, also $\boldsymbol{Q}$ can be either above or below γ_1, depending whether q_2 is bigger or smaller than t_1. There are then in general four possible subcases to consider; the first two are depicted in Figure 3.

<u>Subcase 1.</u> *If* $\boldsymbol{P}$ *and* $\boldsymbol{Q}$ *are both below* γ_1.
In this case, we select $y_1 < t_1$, with $t_1 - y_1 \ll \delta$, for some δ extremely small. Then, we let $\tilde{\gamma}_1$ be a δ-modification of γ_1 with variable endpoints. This can be any δ-modification satisfying the following two requirements: first of all, $\tilde{\gamma}_1$ must connect $\boldsymbol{A}_{y_1}$ with $\boldsymbol{B}_{y_1}$; and second, the points $\widetilde{\boldsymbol{P}}$ and $\widetilde{\boldsymbol{Q}}$ corresponding to $\boldsymbol{P}$ and $\boldsymbol{Q}$ must be aligned with $\boldsymbol{A}_{y_1}$, as in Figure 3 left.

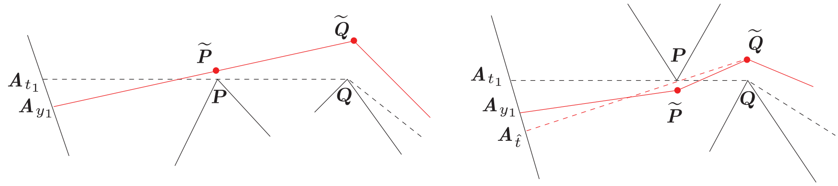

Figure 3. The situation in Step IIb, subcase 1 and subcase 2.

Let us check that $\mathcal{X}'(t)$ is constant for all $t < y_1$ near y_1, as well as for all $t > y_1$ near y_1. Let us write $\mathcal{X}(t_1) = \{\boldsymbol{X}_1, \boldsymbol{X}_2, \dots, \boldsymbol{X}_N\}$, with $\boldsymbol{X}_1 = \boldsymbol{P}$ and $\boldsymbol{X}_2 = \boldsymbol{Q}$. As in Step I, for every $1 \le j \le N$ let us call $\widetilde{\boldsymbol{X}}_j = \boldsymbol{X}_j$ if $\boldsymbol{X}_j \in \partial\mathcal{P}_0$, while otherwise $\widetilde{\boldsymbol{X}}_j$ is the point of $\tilde{\gamma}_1$ very close to $\boldsymbol{X}_j$. It is clear by continuity that, if δ is small enough, then for every $t < y_1$ close to y_1 the geodesic in $\mathcal{P}_0$ connecting $\boldsymbol{A}_t$ and $\boldsymbol{B}_t$ surely has to pass through $\widetilde{\boldsymbol{X}}_j$ for every $j \ge 2$, so for instance through $\boldsymbol{Q} = \boldsymbol{X}_2 = \widetilde{\boldsymbol{X}}_2$, and it cannot pass through other vertices, except possibly $\boldsymbol{P} = \boldsymbol{X}_1 = \widetilde{\boldsymbol{X}}_1$. What is not obvious, is for which $t < y_1$ it also has to pass through $\boldsymbol{P}$: we have to show that this happens for *every* $t < y_1$, so we will get that $\mathcal{X}'(t)$ is constant in $[y_0, y_1]$. In fact, since $\boldsymbol{P}$ and $\boldsymbol{Q}$ are aligned with $\boldsymbol{A}_{t_1}$ and $\boldsymbol{A}_{y_1}$ is below $\boldsymbol{A}_{t_1}$, so also $\boldsymbol{A}_t$ is below $\boldsymbol{A}_{t_1}$, the segment connecting $\boldsymbol{A}_t$ with $\boldsymbol{Q}$ is not entirely contained in $\mathcal{P}$, so the geodesic, in order to start at $\boldsymbol{A}_t$ and to reach $\boldsymbol{Q}$, must necessarily pass through $\boldsymbol{P}$. Similarly, for $t > y_1$ and t close to y_1, the geodesic in $\mathcal{P}^+$ connecting $\boldsymbol{A}_t$ and $\boldsymbol{B}_t$ must necessarily pass through $\widetilde{\boldsymbol{Q}}$, and the question is whether it also passes through $\widetilde{\boldsymbol{P}}$: we have to show that this does not happen. And in fact, it is surely so, because geodesics in a polygon only touch the boundary of the polygon at vertices corresponding to angles strictly larger than π, while $\widetilde{\boldsymbol{P}}$ is not a vertex of $\partial\mathcal{P}^+$.

Finally, the fact that $\mathcal{P}_0$ is an upper δ-tube, again with map Φ_0 coinciding with the identity, is again obvious: indeed, for any $y_0 < t < y_1$ the geodesic in $\mathcal{P}_0$ between $\boldsymbol{A}_t$ and $\boldsymbol{B}_t$ is $\boldsymbol{A}_t\widetilde{\boldsymbol{X}}_1 \dots \widetilde{\boldsymbol{X}}_N\boldsymbol{B}_t$, and $\varphi([a^-, a^+] \times \{y_1\}) = \tilde{\gamma}_1$ is a δ-modification of γ_1 with variable endpoints connecting $\boldsymbol{A}_{y_1}$ and $\boldsymbol{B}_{y_1}$.

<u>Subcase 2.</u> *If $\boldsymbol{P}$ is above γ_1 and $\boldsymbol{Q}$ is below.*
In this case, we select a point $\widetilde{\boldsymbol{Q}}$ on the internal bisector at $\boldsymbol{Q}$ in $\mathcal{P}$, with distance from $\boldsymbol{Q}$ much smaller than some small δ. Then, as in Figure 3 right, we call $\boldsymbol{A}_{\hat{t}}$ the point in the side of $\mathcal{P}$ containing $\boldsymbol{A}_{t_1}$ such that the points $\boldsymbol{A}_{\hat{t}}$, $\boldsymbol{P}$ and $\widetilde{\boldsymbol{Q}}$ are aligned. Notice that $\hat{t} < t_1$, and $t_1 - \hat{t} \ll \delta$. We fix now $y_1 \in (\hat{t}, t_1)$, with $y_1 - \hat{t} \ll t_1 - y_1$, and finally we take $\widetilde{\boldsymbol{P}}$ on the internal bisector at $\boldsymbol{P}$ in $\mathcal{P}$, with $|\widetilde{\boldsymbol{P}} - \boldsymbol{P}| \ll |\boldsymbol{A}_{y_1} - \boldsymbol{A}_{\hat{t}}|$. Then, we let $\tilde{\gamma}_1$ be a δ-modification of γ_1 with variable endpoints, connecting $\boldsymbol{A}_{y_1}$ with $\boldsymbol{B}_{y_1}$ and passing through $\widetilde{\boldsymbol{P}}$ and $\widetilde{\boldsymbol{Q}}$.

We have again to check that $\mathcal{X}'(t)$ is constant for all $t < y_1$ near y_1, and for all $t > y_1$ near y_1. This means that, for $t < y_1$ close to y_1, we have to check that the geodesic in $\mathcal{P}_0$ between $\boldsymbol{A}_t$ and $\boldsymbol{B}_t$ passes through $\boldsymbol{Q}$ and not through $\widetilde{\boldsymbol{P}}$; instead, for $t > y_1$ close to y_1, we have to check that the geodesic in $\mathcal{P}^+$ between $\boldsymbol{A}_t$ and $\boldsymbol{B}_t$ passes through $\boldsymbol{P}$ and $\widetilde{\boldsymbol{Q}}$. Concerning the case $t < y_1$, the fact that the geodesic passes through $\boldsymbol{Q}$ is obvious, while the fact that it does not pass through $\widetilde{\boldsymbol{P}}$ is true because by construction the line connecting $\boldsymbol{A}_{y_1}$ and $\widetilde{\boldsymbol{P}}$ is above $\boldsymbol{Q}$, hence the segment $\boldsymbol{A}_t\boldsymbol{Q}$ is contained in $\mathcal{P}_0$ and then the geodesic does not pass through $\widetilde{\boldsymbol{P}}$. Instead, if $t > y_1$ is close to y_1, it is again obvious that the geodesic passes through $\widetilde{\boldsymbol{Q}}$; and, again by construction, the line passing through $\boldsymbol{A}_{y_1}$ and $\boldsymbol{P}$ is below $\widetilde{\boldsymbol{Q}}$, so the segment $\boldsymbol{A}_t\widetilde{\boldsymbol{Q}}$ is not contained in $\mathcal{P}^+$ and then the geodesic must pass also through $\boldsymbol{P}$.

Finally, this time $\mathcal{P}_0$ is not an upper δ-tube, because of $\widetilde{\boldsymbol{P}}$. In fact, if we call $\mathcal{X}_0 = \{\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_N\}$, so in particular $\boldsymbol{X}_1 = \boldsymbol{P}$ and $\boldsymbol{X}_2 = \boldsymbol{Q}$, then for any $y_0 < t < y_1$ the set of vertices between $\boldsymbol{A}_t$ and $\boldsymbol{B}_t$ in $\mathcal{P}_0$ is $\mathcal{X}_{\mathcal{P}_0}(t) = \{\widetilde{\boldsymbol{X}}_2, \widetilde{\boldsymbol{X}}_3, \ldots, \widetilde{\boldsymbol{X}}_N\}$ (notice that there is no $\widetilde{\boldsymbol{X}}_1$), with every $\widetilde{\boldsymbol{X}}_j$ with distance smaller than δ from the corresponding $\boldsymbol{X}_j$, and in particular $\widetilde{\boldsymbol{X}}_2 = \boldsymbol{Q}$. Instead, $\varphi([a^-, a^+] \times \{y_1\}) = \tilde{\gamma}_1$ is a piecewise linear curve having one vertex near each $\widetilde{\boldsymbol{X}}_j$ for $j \geq 2$, and also the vertex $\widetilde{\boldsymbol{P}}$ which has no point of $\mathcal{X}_{\mathcal{P}_0}(t)$ around. Nevertheless, it is clear that there is a finitely piecewise affine bijection $\Phi_0 : \mathbb{R}^2 \to \mathbb{R}^2$, with bi-Lipschitz constant of order $1+\delta$, thus smaller than $1+\eta$, that leaves all the sides of $\partial\mathcal{P}_0$ unchanged except $\boldsymbol{A}_{y_1}\widetilde{\boldsymbol{P}}$ and $\widetilde{\boldsymbol{P}}\widetilde{\boldsymbol{Q}}$, while these two sides become on a same line: roughly speaking, Φ_0 moves $\widetilde{\boldsymbol{P}}$ upward until it becomes aligned with $\boldsymbol{A}_{y_1}$ and $\widetilde{\boldsymbol{Q}}$. Obviuously, the bijection Φ_0 transforms $\mathcal{P}_0$ into an upper δ-tube, so also in this subcase we are done.

<u>Subcases 3 and 4.</u> *If* $\boldsymbol{P}$ *and* $\boldsymbol{Q}$ *are both above* γ_1*, or if* $\boldsymbol{P}$ *is below* γ_1 *and* $\boldsymbol{Q}$ *is above.*
Let us now briefly consider the third and fourth possible subcases: we will observe that they are almost identical to the first or the second one. More precisely, let us first assume that $\boldsymbol{P}$ and $\boldsymbol{Q}$ are both above γ_1: then, to define $\tilde{\gamma}_1$ we can argue as in Subcase 1, in a specular way, in particular y_1 must be slightly larger than t_1, and $\boldsymbol{A}_{y_1}$ will remain aligned with $\widetilde{\boldsymbol{P}}$ and $\widetilde{\boldsymbol{Q}}$; so, the situation is still the one depicted in Figure 3 left, the only difference being the orientation of the segments, in particular $\mathcal{P}_0$ is now *above* the curve $\tilde{\gamma}_1$, in the picture. Checking that everything works, in particular that $\mathcal{P}_0$ is an upper δ-tube, can be done exactly as before.

Finally, the case when $\boldsymbol{P}$ is below γ_1 and $\boldsymbol{Q}$ is above, is again very similar to Subcase 2, so also this time we can do the same construction in a specular way, and the situation is again as in Figure 3 right, with reversed orientation. To check that $\mathcal{X}'(t)$ is constant for all $t < y_1$ near y_1, and for all $t > y_1$ near y_1, one can again argue exactly as before. The only difference is when one has to check that $\mathcal{P}_0$ is an upper δ-tube: indeed, in Subcase 2 the point $\boldsymbol{P}$ was missing in $\mathcal{X}_0$, so due to the presence of $\widetilde{\boldsymbol{P}}$ the set $\mathcal{P}_0$ was not an upper δ-tube, without the bijection Φ_0. Instead, in the present case $\boldsymbol{P}$ is missing from $\mathcal{X}_1$ but present in $\mathcal{X}_0$, thus in this case the set $\mathcal{P}_0$ is already an upper δ-tube, without any need of the bijection Φ_0 (that is, we can take again Φ_0 as the identity).

<u>Step III.</u> *Definition of* $\tilde{\varphi}$ *on* $\partial\mathcal{R}_0$ *and* (2.6).

This step is devoted to defining $\tilde{\varphi}$ on $[a^-, a^+] \times \{y_1\}$: this must be a piecewise linear and injective function, having as image the piecewise linear curve $\tilde{\gamma}_1$, already defined in Step II. Extending $\tilde{\varphi} = \varphi$ on $\partial\mathcal{R}$, we will have then a piecewise linear and injective function $\tilde{\varphi} : \partial\mathcal{R}_0 \cup \partial\mathcal{R}^+ \to \mathbb{R}^2$, where $\mathcal{R}_0 = [a^-, a^+] \times [b^-, y_1]$ and $\mathcal{R}^+ = [a^-, a^+] \times [y_1, b^+]$, and then $\partial\mathcal{R}_0 \cup \partial\mathcal{R}^+ = \partial\mathcal{R} \cup [a^-, a^+] \times \{y_1\}$. Our definition must satisfy two requirements: first of all, since in the end we want the validity of the estimate (2.4), we aim to get

$$\Psi(\varphi_0) + \Psi(\varphi^+) < \Psi(\varphi) + \eta(y_1 - b^-), \tag{2.6}$$

where φ_0 and φ^+ are the restrictions of $\tilde{\varphi}$ to $\mathcal{R}_0$ and $\mathcal{R}^+$ respectively. Moreover, we have to check that φ^+ on the rectangle $\mathcal{R}^+$ is not aligned, which will allow us in the next step to perform a recursion. Keep in mind that, as underlined in Step II, to obtain that φ^+ is not aligned it only remains to check that every three generalised vertices of $\mathcal{P}^+$, not on a same side, are not aligned; and in turn, we have already checked this, except for the points in $\tilde{\gamma}_1 \subseteq \partial\mathcal{P}^+$ which are generalised vertices but not vertices of $\mathcal{P}^+$.

We show now that it is enough to satisfy the first requirement, namely, the validity of (2.6). Indeed, assume that we have found a piecewise linear and injective function $\tilde{\varphi}$ such that (2.6) holds true. Then, there are finitely many generalised vertices $\boldsymbol{V}_1, \boldsymbol{V}_2, \dots, \boldsymbol{V}_K$ on $\tilde{\gamma}_1$; so, there are points $V_1, V_2, \dots, V_K$ in $[a^-, a^+] \times \{y_1\}$ such that $\tilde{\varphi}(V_j) = \boldsymbol{V}_j$ and $\tilde{\varphi}$ is linear on every $V_j V_{j+1}$. We claim that it is possible to slightly move the generalised vertices $\boldsymbol{V}_j$ which are not vertices, in such a way that the map φ^+ becomes not aligned. This means that, for every j, we set $\boldsymbol{W}_j = \boldsymbol{V}_j$ if $\boldsymbol{V}_j$ is a vertex of $\tilde{\gamma}_1$; otherwise, if $\boldsymbol{V}_j$ is a generalised vertex but not a vertex, we let $\boldsymbol{W}_j \in \tilde{\gamma}_1$ be a point extremely close to $\boldsymbol{V}_j$ to be specified in a moment. We modify then the function $\tilde{\varphi}$ by leaving it linear on each segment $V_j V_{j+1}$, and setting $\tilde{\varphi}(V_j) = \boldsymbol{W}_j$ instead of $\tilde{\varphi}(V_j) = \boldsymbol{V}_j$: notice that, as soon as every $\boldsymbol{W}_j$ is close enough to $\boldsymbol{V}_j$, the estimate (2.6) remains valid. Let us now see how we define the points $\boldsymbol{W}_j$. If every three generalised vertices of $\mathcal{P}^+$, not on a same side, are not aligned, then there is nothing to do, and we can simply set $\boldsymbol{W}_j = \boldsymbol{V}_j$ for every j. Suppose, instead, that there are three generalised vertices of $\mathcal{P}^+$ which are aligned; as already observed, at least one of the three must be some point $\boldsymbol{V}_j$, in particular a generalised vertex which is not a vertex. In other words, $\boldsymbol{V}_j$ belongs to a line ℓ which contains two other given generalised vertices of $\mathcal{P}^+$; since we have already checked the non alignment of points taken among the generalised vertices of $\mathcal{P}$ and the vertices of $\tilde{\gamma}_1$, the line ℓ does not contain the side of $\mathcal{P}^+$ containing $\boldsymbol{V}_j$. As a consequence, we can take $\boldsymbol{W}_j \neq \boldsymbol{V}_j$ arbitrarily close to $\boldsymbol{V}_j$, and it does not lie on the line ℓ; since there are finitely many generalised vertices in $\mathcal{P}^+$, we can do this in such a way that $\boldsymbol{W}_j$ is not aligned with any other two generalised vertices. By repeating this argument for all the aligned triples, so finitely many times, we end up with the required slight modification of $\tilde{\varphi}$.

Summarizing, to conclude this step we only have to find a piecewise linear and injective parameterisation $\tilde{\varphi}$ of $\tilde{\gamma}_1$ such that (2.6) is satisfied. We can furhter reduce ourselves to simply looking for a parameterization $\tilde{\varphi}$ satisfying (2.6): indeed, once such a function is found, we can uniformly approximate it with a piecewise linear one, then the validity of (2.6) is preserved. Hence, we look now for a bijection $\tilde{\varphi} : [a^-, a^+] \times \{y_1\} \to \tilde{\gamma}_1$ satisfying (2.6).

To do so, for every $a^- \leq t \leq a^+$ we consider the geodesic γ^t in $\mathcal{P}$ between $\boldsymbol{C}_t = \varphi(t, b^-)$ and $\boldsymbol{D}_t = \varphi(t, b^+)$: since the curve $\tilde{\gamma}_1$ divides $\mathcal{P}$ into two connected components, and $\boldsymbol{C}_t$ and $\boldsymbol{D}_t$ are not in the same one, then γ^t must intersect $\tilde{\gamma}_1$. Let us call $\boldsymbol{P}(t)$ the last point (with respect to the order given by the curve $\tilde{\gamma}_1$) in this intersection. By construction and applying Lemma 2.4 to two geodesics γ^s and γ^t, we see that the function $t \mapsto \boldsymbol{P}(t)$ is nondecreasing, that is, if $s > t$ then $\boldsymbol{P}(s)$ is in the closed part of $\tilde{\gamma}_1$ between $\boldsymbol{P}(t)$ and $\boldsymbol{B}_{y_1}$. This function is in general not continuous, nor injective, nor surjective; nevertheless, there is of course a bijection $\tilde{\varphi} : [a^-, a^+] \times \{y_1\} \to \tilde{\gamma}_1$ such that $\tilde{\varphi}(t, y_1)$ has distance less than δ from $\boldsymbol{P}(t)$ for all $t \in [a^-, a^+]$ except those contained in a subset Γ of $[a^-, a^+]$ of measure less than δ.

Since $\tilde{\gamma}_1$ is a δ-modification of γ_1, with variable endpoints, we can apply Lemma 2.5 with some ε to be specified later. Up to decrease δ, then, we find that for every $t \in [a^-, a^+] \setminus \Gamma$

$$d_{\mathcal{P}_0}(\boldsymbol{C}_t, \tilde{\varphi}(t, y_1)) + d_{\mathcal{P}^+}(\tilde{\varphi}(t, y_1), \boldsymbol{D}_t) < d_{\mathcal{P}}(\boldsymbol{C}_t, \boldsymbol{D}_t) + \varepsilon.$$

However, for each t, so in particular for those contain in Γ, we trivially get

$$d_{\mathcal{P}_0}(\boldsymbol{C}_t, \tilde{\varphi}(t, y_1)) + d_{\mathcal{P}^+}(\tilde{\varphi}(t, y_1), \boldsymbol{D}_t) < \mathscr{H}^1(\partial\mathcal{P}) + \mathscr{H}^1(\tilde{\gamma}_1) \leq 3\mathscr{H}^1(\partial\mathcal{P}).$$

Therefore, we derive

$$\begin{aligned} &\int_{t=a^-}^{a^+} d_{\mathcal{P}_0}(\boldsymbol{C}_t, \tilde{\varphi}(t, y_1)) + d_{\mathcal{P}^+}(\tilde{\varphi}(t, y_1), \boldsymbol{D}_t)\, dt \\ &\qquad \leq 3\delta\mathscr{H}^1(\partial\mathcal{P}) + \int_{t=a^-}^{a^+} d_{\mathcal{P}}(\boldsymbol{C}_t, \boldsymbol{D}_t) + \varepsilon\, dt. \end{aligned} \tag{2.7}$$

On the other hand, still by Lemma 2.5 we have for every $b^- \leq t \leq y_1$ that

$$d_{\mathcal{P}_0}(\boldsymbol{A}_t, \boldsymbol{B}_t) < d_{\mathcal{P}}(\boldsymbol{A}_t, \boldsymbol{B}_t) + \varepsilon,$$

while for every $y_1 \leq t \leq b^+$ it is

$$d_{\mathcal{P}^+}(\boldsymbol{A}_t, \boldsymbol{B}_t) < d_{\mathcal{P}}(\boldsymbol{A}_t, \boldsymbol{B}_t) + \varepsilon.$$

Thus,

$$\int_{t=b^-}^{y_1} d_{\mathcal{P}_0}(\boldsymbol{A}_t, \boldsymbol{B}_t)\, dt + \int_{t=y_1}^{b^+} d_{\mathcal{P}^+}(\boldsymbol{A}_t, \boldsymbol{B}_t)\, dt < \int_{t=b^-}^{b^+} d_{\mathcal{P}}(\boldsymbol{A}_t, \boldsymbol{B}_t) + \varepsilon\, dt. \tag{2.8}$$

Keeping in mind Definition 1.1, from (2.7) and (2.8) we obtain

$$\Psi(\varphi_0) + \Psi(\varphi^+) \leq \Psi(\varphi) + 3\delta\mathscr{H}^1(\partial\mathcal{P}) + \varepsilon(a^+ - a^- + b^+ - b^-),$$

from which the validity of (2.6) follows, up to choose ε and δ small enough.

Step IV. Recursion and conclusion.

In this last step, we want to conclude our construction. Applying Steps II and III, we have already divided our rectangle $\mathcal{R}$ into a rectangle $\mathcal{R}_0 = [a^-, a^+] \times [b^-, y_1]$ and the remaining part $\mathcal{R}^+$, and we have defined $\tilde{\varphi}$ on $\partial\mathcal{R}_0 \cup \partial\mathcal{R}$ in such a way that (2.6) holds; in addition, the polygon $\mathcal{P}(\Phi_0 \circ \varphi_0)$ (keep in mind Definition 1.1) is an upper δ-tube with a finitely piecewise affine bijection $\Phi_0 : \mathbb{R}^2 \to \mathbb{R}^2$ with bi-Lipschitz constant smaller than $1 + \eta$, and in addition $\tilde{\varphi}$ is also a not aligned function and the coordinates t_j' in the rectangle $\mathcal{R}^+$ with the function $\tilde{\varphi}$ are exactly $M - 1$ thanks to (2.5).

We can then remain satisfied with the work done on $\mathcal{R}_0$, and argue on $\mathcal{R}^+$: with an obvious recursion, in the end we will have subdivided $\mathcal{R}$ into M rectangles $\mathcal{R}_i$, $0 \le i \le M-1$, and we will have defined a piecewise linear and injective function $\tilde{\varphi} : \bigcup_{i=0}^{M-1} \partial\mathcal{R}_i \to \mathbb{R}^2$, with $\tilde{\varphi} = \varphi$ on $\partial\mathcal{R}$, together with constants δ_i and finitely piecewise affine bijections Φ_i as before. Since (2.4) follows simply adding the inequalities (2.6), we have obtained everything, except the fact that the polygons $\mathcal{P}(\varphi_i)$ with $1 \le i \le M-1$ are also lower δ-tubes, up to suitable finitely piecewise affine bijections Φ_i.

Again arguing by recursion, it is enough to check that $\mathcal{P}(\varphi_1)$ is a lower δ-tube: this is quite simple, one has only to consider three possible cases; we call again $\mathcal{X}_0 = \{\boldsymbol{X}_1, \boldsymbol{X}_2, \dots, \boldsymbol{X}_N\}$.

Subcase 1. If $\mathcal{X}_1 = \mathcal{X}_0$.

Let us consider first the case when $\mathcal{X}_1 = \mathcal{X}_0$, depicted in Figure 4 left. This means that the points $\boldsymbol{A}_{y_1}$ and $\boldsymbol{B}_{y_1}$ are necessarily generalised vertices of $\mathcal{P}$, hence the curve $\tilde{\varphi}([a^-, a^+] \times \{y_1\}) = \tilde{\gamma}_1$ has been defined in Step IIa. In particular, $\tilde{\gamma}_1 = \boldsymbol{A}_{y_1}\widetilde{\boldsymbol{X}}_1 \dots \widetilde{\boldsymbol{X}}_N\boldsymbol{B}_{y_1}$, where each $\widetilde{\boldsymbol{X}}_j$ has distance at most δ_0 from the corresponding $\boldsymbol{X}_j$. For $y_1 < t < y_2$, by Step II we know that $\mathcal{X}'(t) \approx \mathcal{X}(t)$: this means that the geodesic in $\mathcal{P}^+$ between $\boldsymbol{A}_t$ and $\boldsymbol{B}_t$ is $\boldsymbol{A}_t\boldsymbol{Y}_1 \dots \boldsymbol{Y}_N\boldsymbol{B}_t$, where for each $1 \le j \le N$ we have $\boldsymbol{Y}_j = \boldsymbol{X}_j$ if the latter belongs to $\partial\mathcal{P}^+$, and otherwise $\boldsymbol{Y}_j = \widetilde{\boldsymbol{X}}_j$. If $M = 2$, then $\mathcal{P}(\varphi_1) = \mathcal{P}^+$ and we have already the fact that $\mathcal{P}(\varphi_1)$ is a lower δ_0-tube, again Φ_1 being the identity. Instead, if $M \ge 3$, then there is also a curve $\tilde{\gamma}_2$, which equals $\tilde{\varphi}([a^-, a^+] \times \{y_2\})$, and for $y_1 < t < y_2$ we have again by Step II that $\mathcal{X}'(t) \approx \mathcal{X}''(t)$, where $\mathcal{X}'(t)$ and $\mathcal{X}''(t)$ are the sets of vertices of the geodesic between $\boldsymbol{A}_t$ and $\boldsymbol{B}_t$ in $\mathcal{P}^+$ and in $\mathcal{P}(\varphi_1)$ respectively. This means that the geodesic in $\mathcal{P}(\varphi_1)$ between $\boldsymbol{A}_t$ and $\boldsymbol{B}_t$ is the curve $\boldsymbol{A}_t\boldsymbol{Z}_1 \dots \boldsymbol{Z}_N\boldsymbol{B}_t$, where $\boldsymbol{Z}_j$ equals $\boldsymbol{Y}_j$ if

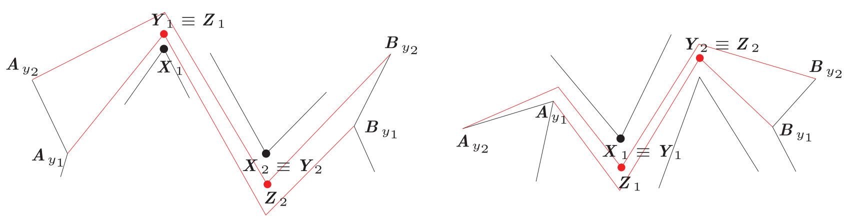

Figure 4. The situation in Step IV, subcase 1 and subcase 2.

the latter belongs to $\partial\mathcal{P}(\varphi_1)$, while otherwise $\boldsymbol{Z}_j$ is the point of $\tilde{\gamma}_2$ which is very close to $\boldsymbol{Y}_j$, in particular $|\boldsymbol{Y}_j - \boldsymbol{Z}_j| < \delta_1$. Then, every $\boldsymbol{Z}_j$ is very close to the corresponding $\widetilde{\boldsymbol{X}}_j$, so we have that $\mathcal{P}(\varphi_1)$ is a lower $(\delta_0 + \delta_1)$-tube.

Subcase 2. If $\boldsymbol{A}_{y_1}$ and $\boldsymbol{B}_{y_1}$ are generalised vertices and $\mathcal{X}_0 \neq \mathcal{X}_1$.
Now, let us consider the case when $\boldsymbol{A}_{y_1}$ and $\boldsymbol{B}_{y_1}$ are generalised vertices of $\mathcal{P}$, but $\mathcal{X}_0 \neq \mathcal{X}_1$. Since immediate geometrical considerations ensure that $\mathcal{X}_0 \Delta \mathcal{X}_1 \subseteq \{\boldsymbol{A}_{y_1}, \boldsymbol{B}_{y_1}\}$, the only possibility is that at least one between $\boldsymbol{A}_{y_1}$ and $\boldsymbol{B}_{y_1}$ is also a vertex of $\mathcal{P}$, not only a generalised one, and that it belongs to exactly one between $\mathcal{X}_0$ and $\mathcal{X}_1$ (this can also happen contemporarily to both $\boldsymbol{A}_{y_1}$ and $\boldsymbol{B}_{y_1}$). We are going to describe what happens if $\boldsymbol{A}_{y_1} \in \mathcal{X}_1 \setminus \mathcal{X}_0$, so $\mathcal{X}_1$ is the set $\{\boldsymbol{A}_{y_1}, \boldsymbol{X}_1, \ldots, \boldsymbol{X}_N\}$: the situation when $\boldsymbol{A}_{y_1} \in \mathcal{X}_0 \setminus \mathcal{X}_1$ is completely similar, so as the analogous cases with $\boldsymbol{B}_{y_1}$ in place of $\boldsymbol{A}_{y_1}$. The situation is depicted in Figure 4 right.

Since, as showed in Step II, for $y_1 < t < y_2$ one has $\mathcal{X}(t) \approx \mathcal{X}'(t)$, being $\mathcal{X}'(t)$ the set of vertices of the geodesic between $\boldsymbol{A}_t$ and $\boldsymbol{B}_t$ in $\mathcal{P}^+$, we have $\mathcal{X}'(t) = \{\boldsymbol{A}_{y_1}, \boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_N\}$, where each $\boldsymbol{Y}_j$ coincides with $\boldsymbol{X}_j$ if $\boldsymbol{X}_j \in \partial\mathcal{P}^+$, while otherwise $\boldsymbol{Y}_j$ is the point of $\tilde{\gamma}_1$ very close to $\boldsymbol{X}_j$. As before, if $M = 2$ we already have that $\mathcal{P}(\varphi_1) = \mathcal{P}^+$ is a lower δ_0-tube, since we can write $\tilde{\gamma}_1 = \boldsymbol{A}_{y_1}\boldsymbol{A}_{y_1}\widetilde{\boldsymbol{X}}_1 \ldots \widetilde{\boldsymbol{X}}_N\boldsymbol{B}_{y_1}$ (notice that $\boldsymbol{A}_{y_1}$ appears twice, according with Definition 2.10). And again, if $M > 2$ then the set of vertices $\mathcal{X}''(t)$ in $\mathcal{P}(\varphi_1)$ corresponding to $y_1 < t < y_2$ will be simply $\mathcal{X}''(t) = \{\boldsymbol{A}_{y_1}, \boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_N\}$, again with either $\boldsymbol{Z}_j = \boldsymbol{Y}_j$, or $|\boldsymbol{Z}_j - \boldsymbol{Y}_j| < \delta_1$. Thus, also for $M > 2$ we have proved that $\mathcal{P}(\varphi_1)$ is a lower $(\delta_0 + \delta_1)$-tube, still with Φ_0 being the identity.

Subcase 3. If $\boldsymbol{A}_{y_1}$ and $\boldsymbol{B}_{y_1}$ are not generalised vertices.
The last case to consider is when $\boldsymbol{A}_{y_1}$ and $\boldsymbol{B}_{y_1}$ are not generalised vertices, so necessarily $\mathcal{X}_0 \neq \mathcal{X}_1$ and the curve $\tilde{\gamma}_1$ has been defined in Step IIb. Remember that, in this case, the point $\boldsymbol{A}_{t_1} \neq \boldsymbol{A}_{y_1}$ was aligned with two points $\boldsymbol{P}$ and $\boldsymbol{Q}$ in $\partial\mathcal{P}$, and the only difference between $\mathcal{X}_0$ and $\mathcal{X}_1$ was that only one of them contained $\boldsymbol{P}$. Keep in mind that the definition of $\tilde{\gamma}_1$ was split into four subcases, depending if $\boldsymbol{P}$ and $\boldsymbol{Q}$ were below, or above γ_1.

Suppose first that $\boldsymbol{P}$ and $\boldsymbol{Q}$ are both below γ_1, as in Figure 3 left. The situation now is then the one depicted in Figure 5 left: it is evident how, repeating the very same arguments as above, the fact that $\mathcal{P}(\varphi_1)$ is a lower δ-tube follows, once again Φ_1 being the identity. The case when both $\boldsymbol{P}$ and $\boldsymbol{Q}$ are above γ_1 is identical.

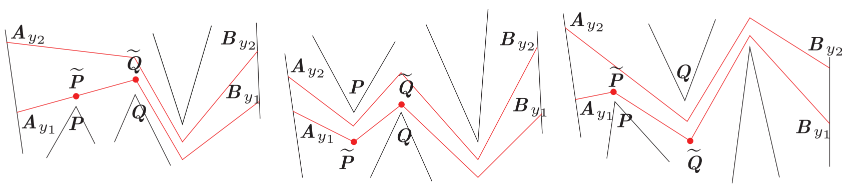

Figure 5. The situation in Step IV, subcase 3.

Let us now consider what happens if $\boldsymbol{P}$ is above γ_1 and $\boldsymbol{Q}$ below, as in Figure 3 right, or if $\boldsymbol{P}$ is below and $\boldsymbol{Q}$ above: the situation now is shown respectively in Figure 5 center and in Figure 5 right. Observe that the point $\boldsymbol{P}$ is contained in $\mathcal{X}_1$ in the first case, while in the second case the first vertex of $\mathcal{X}_1$ is $\boldsymbol{Q}$. Then, once again the same arguments as before show that $\mathcal{P}(\varphi_1)$ is a lower δ-tube in the first case, because $\widetilde{\boldsymbol{P}}$ is a vertex of $\tilde{\gamma}_1$ and $\boldsymbol{P}$ belongs to $\mathcal{X}_1$, so we get once more the thesis with Φ_1 being the identity. Instead, in the second case $\mathcal{P}(\varphi_1)$ is not a lower δ-tube, because the set of vertices $\mathcal{X}_1$ starts with $\boldsymbol{Q}$, while $\tilde{\gamma}_1$ contains the vertex $\widetilde{\boldsymbol{P}}$. Nevertheless, exactly as already did in Step II, we can easily notice that the set $\mathcal{P}(\Phi_1 \circ \varphi_1)$ is a lower δ-tube, where Φ_1 is a finitely piecewise affine bijection which does not move any side of $\mathcal{P}(\varphi_1)$ except the sides $\boldsymbol{A}_{y_1}\widetilde{\boldsymbol{P}}$ and $\widetilde{\boldsymbol{P}}\widetilde{\boldsymbol{Q}}$, and these two sides become on a same line (that is, $\boldsymbol{A}_{y_1}$, $\widetilde{\boldsymbol{P}}$ and $\widetilde{\boldsymbol{Q}}$ are transformed into three aligned points). Notice that the bi-Lipschitz constant of Φ_1 is of order $1+\delta$, so much smaller than $1+\eta$; hence, also this final case is concluded.

In order to conclude the proof, we need just the following simple observation. Strictly speaking, for every $1 \leq i < M-1$ we have found *two* finitely piecewise affine bijections, and we have called both them Φ_i: one of them, call it Φ_i^+, was needed to make $\mathcal{P}(\varphi_i)$ an upper δ-tube, and the other one, call it Φ_i^-, to make it a lower δ-tube. This if of course not admissible, we need a single bijection. But in fact, as we observed during the proof, the bijection Φ_i^+ can fail to be the identity only in an arbitrarily small neighborhood of the curve $\tilde{\gamma}_{i+1}$, while Φ_i^- can fail to be the identity only in an arbitrarily small neighborhood of $\tilde{\gamma}_i$. As a consequence, it is enough to define Φ_i the finitely piecewise affine bijection which coincides with Φ_i^+ and Φ_i^- near $\tilde{\gamma}_{i+1}$ and $\tilde{\gamma}_i$ respectively, and which is the identity otherwise: by construction, this bijection satisfies our requirements. □

2.3. The extension in the δ-tubes and in the upper or lower δ-tubes

Thanks to Lemma 2.11, in order to build the extension v required by Proposition 2.2 we can consider separately each of the rectangles $\mathcal{R}_i$. And in turn, there are two kinds of rectangles to consider: the "internal" ones, corresponding to $0 < i < M-1$, which are related to δ-tubes, and the two "external" ones, corresponding to $i = 0$ and $i = M-1$, which are related to polygons which are only upper or lower δ-tubes. We present immediately the result that we will use to deal with the δ-tubes, while for the upper or lower δ-tubes we will use Lemma 2.15.

LEMMA 2.12. *Let $\mathcal{R} = [a^-, a^+] \times [b^-, b^+]$ be a rectangle, and let $\varphi : \partial\mathcal{R} \to \mathbb{R}^2$ be a piecewise linear and injective function such that the associated polygon $\mathcal{P} = \mathcal{P}(\varphi)$ is a δ-tube with number N. Then, there exists an injective, finitely piecewise affine function $v : \mathcal{R} \to \mathbb{R}^2$ such that $v = \varphi$ on $\partial\mathcal{R}$ and*

$$\int_{\mathcal{R}} \|Dv\| \leq \Psi(\varphi) + (a^+ - a^- + b^+ - b^-)(N+2)\delta. \tag{2.9}$$

To show this lemma, we need some simple preliminary observations.

LEMMA 2.13. *Let $2 \leq K \in \mathbb{N}$, and let $\boldsymbol{A}_i$ and $\boldsymbol{B}_i$, for $1 \leq i \leq K$, be points in $\mathbb{R}^N$. For every $0 \leq t \leq 1$, let $\boldsymbol{P}_i(t) = (1-t)\boldsymbol{A}_i + t\boldsymbol{B}_i$, and let $\gamma(t)$ be the piecewise linear curve $\boldsymbol{P}_1(t)\boldsymbol{P}_2(t)\dots\boldsymbol{P}_K(t)$. Then, one has $\mathscr{H}^1(\gamma(t)) \leq (1-t)\mathscr{H}^1(\gamma(0)) + t\mathscr{H}^1(\gamma(1))$.*

PROOF. This is very simple: by obvious recursion, it is sufficient to consider the case $K = 2$, which is in turn equivalent to show that, given two vectors $v, w \in \mathbb{R}^N$, the function $t \mapsto |v + tw|$ is convex. And this latter fact is obvious. □

COROLLARY 2.14. *Let $\mathcal{R} = [a^-, a^+] \times [b^-, b^+]$ be a rectangle, and let $\varphi : \partial\mathcal{R} \to \mathbb{R}^2$ be an injective function which is linear on each of the four sides of $\mathcal{R}$. Then, there exists an injective, finitely piecewise affine function $v : \mathcal{R} \to \mathbb{R}^2$ such that $v = \varphi$ on $\partial\mathcal{R}$ and for every $0 \leq t \leq 1$ one has*

$$\begin{aligned} \mathscr{H}^1(\gamma_t) &\leq (1-t)\mathscr{H}^1(\gamma_0) + t\mathscr{H}^1(\gamma_1), \\ \mathscr{H}^1(\gamma^t) &\leq (1-t)\mathscr{H}^1(\gamma^0) + t\mathscr{H}^1(\gamma^1), \end{aligned} \tag{2.10}$$

where for each $0 \leq t \leq 1$ we denote by $\gamma_t : [b^-, b^+] \to \mathbb{R}^2$ and $\gamma^t : [a^-, a^+] \to \mathbb{R}^2$ the curves

$$\gamma_t(s) = v((1-t)a^- + ta^+, s), \quad \gamma^t(s) = v(s, (1-t)b^- + tb^+).$$

In particular, one has

$$\int_{\mathcal{R}} \|Dv\| \leq (a^+ - a^-)\frac{\mathscr{H}^1(\gamma_0) + \mathscr{H}^1(\gamma_1)}{2} + (b^+ - b^-)\frac{\mathscr{H}^1(\gamma^0) + \mathscr{H}^1(\gamma^1)}{2}. \tag{2.11}$$

PROOF. The image of φ is the boundary of a non self-intersecting quadrilateral $\mathcal{P} \subseteq \mathbb{R}^2$, whose vertices are

$$\boldsymbol{C} = \varphi(a^-, b^-), \quad \boldsymbol{D} = \varphi(a^+, b^-), \quad \boldsymbol{E} = \varphi(a^+, b^+), \quad \boldsymbol{F} = \varphi(a^-, b^+).$$

At least one of the two diagonals of $\mathcal{P}$ is necessarily contained in $\mathcal{P}$ itself, without loss of generality we assume that the segment $\boldsymbol{CE}$ is contained in $\mathcal{P}$. Hence, we define v as the function on $\mathcal{R}$ which coincides with φ on $\partial\mathcal{R}$, and which is affine on the two triangles $\boldsymbol{CDE}$ and $\boldsymbol{CEF}$.

Notice that, if we choose $K = 3$ and we set the pairs $(\boldsymbol{A}_1, \boldsymbol{B}_1) = (\boldsymbol{C}, \boldsymbol{D})$, $(\boldsymbol{A}_2, \boldsymbol{B}_2) = (\boldsymbol{C}, \boldsymbol{E})$, $(\boldsymbol{A}_3, \boldsymbol{B}_3) = (\boldsymbol{F}, \boldsymbol{E})$, then our curves γ_t coincide with the curves $\gamma(t)$ of Lemma 2.13. Thus, the first estimate in (2.10) directly follows from Lemma 2.13; the second estimate is completely symmetric.

Obtaining (2.11) is then immediate: for every $0 \leq t \leq 1$, by (2.10) one has

$$\int_{b^-}^{b^+} |D_2 v((1-t)a^- + ta^+, s)|\, ds = \mathscr{H}^1(\gamma_t) \leq (1-t)\mathscr{H}^1(\gamma_0) + t\mathscr{H}^1(\gamma_1),$$

thus

$$\begin{aligned}\int_{\mathcal{R}} |D_2 v| &= \int_{x=a^-}^{a^+} \int_{s=b^-}^{b^+} |D_2 v(x,s)|\,ds\,dx \\ &= (a^+ - a^-) \int_{t=0}^{1} \int_{s=b^-}^{b^+} |D_2 v((1-t)a^- + ta^+, s)|\,ds\,dt \\ &\leq (a^+ - a^-) \frac{\mathscr{H}^1(\gamma_0) + \mathscr{H}^1(\gamma_1)}{2},\end{aligned}$$

and similarly

$$\int_{\mathcal{R}} |D_1 v| \leq (b^+ - b^-) \frac{\mathscr{H}^1(\gamma^0) + \mathscr{H}^1(\gamma^1)}{2},$$

hence (2.11) directly follows. □

While the above corollary is needed for the proof of Lemma 2.12, hence in turn with the situation of the "internal" rectangles $\mathcal{R}_i$ in the proof of Proposition 2.2, to deal with the two "external" ones we will need the following other consequence of Lemma 2.13.

LEMMA 2.15. *Let $\mathcal{R} = [a^-, a^+] \times [b^-, b^+]$ be a rectangle, and let $\varphi : \partial\mathcal{R} \to \mathbb{R}^2$ be an injective and piecewise linear function. Assume that φ is linear on the segments $[a^-, a^+] \times \{b^-\}$, $\{a^-\} \times [b^-, b^+]$ and $\{a^+\} \times [b^-, b^+]$, and that $\mathcal{P}(\varphi)$ is an upper δ-tube for some δ much smaller than the length of any side of $\mathcal{P}(\varphi)$. Then, there exists an injective, finitely piecewise affine function $v : \mathcal{R} \to \mathbb{R}^2$ such that $v = \varphi$ on $\partial\mathcal{R}$ and*

$$\int_{\mathcal{R}} \|Dv\| \leq \frac{3}{2}(a^+ - a^- + b^+ - b^-)\mathscr{H}^1(\partial\mathcal{P}(\varphi)). \tag{2.12}$$

PROOF. For brevity, let us call

$$\begin{aligned}&A = (a^-, b^-), \quad &&B = (a^+, b^-), \quad &&C = (a^+, b^+), \quad &&D = (a^-, b^+), \\ &\boldsymbol{A} = \varphi(A), &&\boldsymbol{B} = \varphi(B), &&\boldsymbol{C} = \varphi(C), &&\boldsymbol{D} = \varphi(D).\end{aligned}$$

Then, the image of the segments AB, BC and AD through φ are the segments $\boldsymbol{AB}$, $\boldsymbol{BC}$ and $\boldsymbol{AD}$, while the image of the segment CD is some piecewise linear curve between $\boldsymbol{C}$ and $\boldsymbol{D}$. By definition of upper δ-tubes, there is some $\mathcal{X} = \{\boldsymbol{X}_1, \boldsymbol{X}_2, \dots, \boldsymbol{X}_N\}$ such that $\mathcal{X}(t) = \mathcal{X}$ for every $b^- < t < b^+$. In particular, the geodesic in $\mathcal{P} = \mathcal{P}(\varphi)$ between $\boldsymbol{A}$ and $\boldsymbol{B}$ is $\boldsymbol{A}\boldsymbol{X}_1\boldsymbol{X}_2 \dots \boldsymbol{X}_N\boldsymbol{B}$; on the other hand, of course the geodesic in $\mathcal{P}$ between $\boldsymbol{A}$ and $\boldsymbol{B}$ is the segment $\boldsymbol{AB}$ itself, since it is contained in $\mathcal{P}$ by construction. As a consequence, the only points that $\mathcal{X}$ could contain are $\boldsymbol{A}$ and $\boldsymbol{B}$: in other words, either $\mathcal{X} = \emptyset$, or $\mathcal{X}$ contains exactly one between $\boldsymbol{A}$ and $\boldsymbol{B}$, or $\mathcal{X} = \{\boldsymbol{A}, \boldsymbol{B}\}$. Let us consider separately the three possibilities.

If $\mathcal{X} = \emptyset$, keeping in mind that $\mathcal{P}$ is an upper δ-tube we get that the image of CD under φ is the segment $\boldsymbol{CD}$ itself, so $\mathcal{P}(\varphi)$ is the quadrilateral $\boldsymbol{ABCD}$. Moreover, we have that the quadrilateral is convex, because otherwise there would be some t for which $\mathcal{X}(t) \neq \emptyset$. As a consequence, we can easily reduce ourselves to the case when φ is linear on the four sides of $\mathcal{R}$. More precisely, there exists finitely many coordinates $a^- = x_0 < x_1 < \cdots < x_{M-1} < x_M = a^+$ such that φ is linear on each segment Q_iQ_{i+1}, writing $Q_i = (x_i, b^+) \in \partial\mathcal{R}$. We can then call $P_i = (x_i, b^-)$ the corresponding points in the bottom side of $\mathcal{R}$, and extend φ linearly to each segment P_iQ_i: by convexity of $\mathcal{P}(\varphi)$, the extension φ is still piecewise linear and injective. The rectangle $\mathcal{R}$ has then been subdivided in the rectangles $\mathcal{R}_i = [x_i, x_{i+1}] \times [b^-, b^+]$, and the function φ is linear on the boundary of each of these rectangles. We can then define the function $v : \mathcal{R} \to \mathbb{R}^2$ by using Corollary 2.14 on each $\mathcal{R}_i$: we will get a finitely piecewise affine and injective function, and (2.11) on each $\mathcal{R}_i$ gives

$$
\begin{aligned}
\int_{\mathcal{R}} \|Dv\| \leq{}& \sum_{i=0}^{M-1} (x_{i+1} - x_i) \frac{\mathscr{H}^1(\boldsymbol{P}_i\boldsymbol{Q}_i) + \mathscr{H}^1(\boldsymbol{P}_{i+1}\boldsymbol{Q}_{i+1})}{2} \\
&+ \sum_{i=0}^{M-1} (b^+ - b^-) \frac{\mathscr{H}^1(\boldsymbol{P}_i\boldsymbol{P}_{i+1}) + \mathscr{H}^1(\boldsymbol{Q}_i\boldsymbol{Q}_{i+1})}{2} \\
\leq{}& \frac{a^+ - a^-}{2} \max\{\mathscr{H}^1(\boldsymbol{AD}), \mathscr{H}^1(\boldsymbol{BC})\} \\
&+ \frac{b^+ - b^-}{2} (\mathscr{H}^1(\boldsymbol{AB}) + \mathscr{H}^1(\boldsymbol{CD})) \\
\leq{}& \frac{a^+ - a^- + b^+ - b^-}{2} \mathscr{H}^1(\partial\mathcal{P}(\varphi)),
\end{aligned}
\tag{2.13}
$$

thus (2.12) is established.

Let us now consider the case when $\mathcal{X}$ consists of exactly one element, without loss of generality let us think that $\mathcal{X} = \{\boldsymbol{B}\}$. Then, since $\mathcal{P}(\varphi)$ is an upper δ-tube we know that the image of DC through φ is the piecewise linear curve $\boldsymbol{DQC}$ for some point $\boldsymbol{Q}$ on the internal bisector to the angle in $\boldsymbol{B}$, with distance from $\boldsymbol{B}$ less than δ: Figure 6 (left) depicts the situation in this case. As in the figure, we call $Q = \varphi^{-1}(\boldsymbol{Q}) \in \partial\mathcal{R}$, and we let $P \in \partial\mathcal{R}$ be the point in the bottom side with the same abscissa $x_Q \in (a^-, a^+)$ as Q. Then, $\boldsymbol{P} = \varphi(P)$ is some point in the segment $\boldsymbol{AB}$. We extend now the function φ as a linear function on the segment PQ, as well as on the segment BQ, and we define v separately on the rectangle $APQD$ and on the triangles PBQ and BCQ. First of all, the restriction of φ to the rectangle $APQD$ is still injective and piecewise linear, and also linear on the left, right, and bottom side, and moreover the corresponding polygon is the convex quadrilateral $\boldsymbol{APQD}$, so it is again a δ-cube corresponding to $\mathcal{X} = \emptyset$. Since we have already considered this case, proving in particular the estimate (2.13), we get a finitely piecewise affine function v on the rectangle

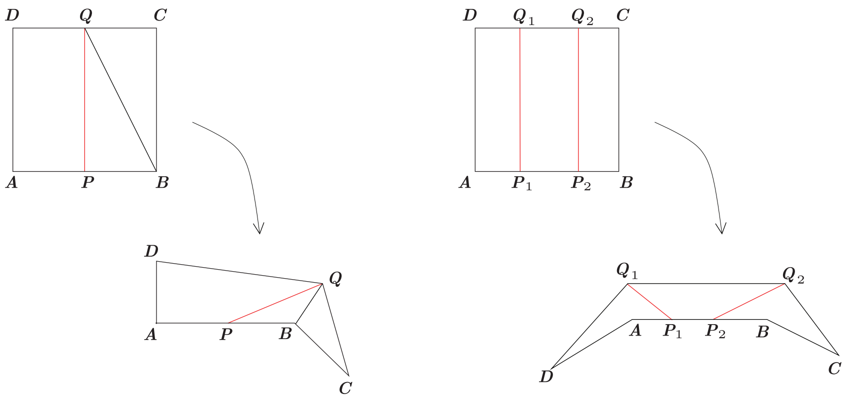

Figure 6. The situation when $\mathcal{X} = \{\boldsymbol{B}\}$ and when $\mathcal{X} = \{\boldsymbol{A}, \boldsymbol{B}\}$ in Lemma 2.15.

$APQD$ satisfying

$$\int_{APQD} \|Dv\| \leq \frac{x_Q - a^- + b^+ - b^-}{2} \mathscr{H}^1(\partial \boldsymbol{APQD}) \leq \frac{x_Q - a^- + b^+ - b^-}{2} \mathscr{H}^1(\partial \mathcal{P}(\varphi)). \tag{2.14}$$

On the triangle PBQ, we let v be then the affine function extending φ, which makes sense since φ is linear on each side of the triangle. Then, we readily get

$$\int_{PBQ} \|Dv\| = \int_{PBQ} |D_1 v| + \int_{PBQ} |D_2 v| = \frac{b^+ - b^-}{2} \mathscr{H}^1(\boldsymbol{PB}) + \frac{a^+ - x_Q}{2} \mathscr{H}^1(\boldsymbol{PQ}).$$

Then, concerning the triangle BQC, we keep in mind that φ is linear on the sides BC and BQ, and piecewise linear on the side QC, and that the image under φ of the segment QC is the segment $\boldsymbol{QC}$. Thus, we can define v on the triangle BQC the finitely piecewise affine function which is affine on each triangle RSB, being R and S any two points on QC such that φ is linear on RS. The resulting function $v : \mathcal{R} \to \mathbb{R}^2$ is finitely piecewise affine and injective, and we get the estimate

$$\int_{BQC} \|Dv\| = \int_{BQC} |D_1 v| + \int_{BQC} |D_2 v| \leq \frac{b^+ - b^-}{2} \mathscr{H}^1(\boldsymbol{QC}) + \frac{a^+ - x_Q}{2} (\mathscr{H}^1(\boldsymbol{BC}) + \mathscr{H}^1(\boldsymbol{QC})),$$

which added to the previous one gives, for the rectangle $PBCQ$,

$$\int_{PBCQ} \|Dv\| \leq \frac{b^+ - b^- + a^+ - x_Q}{2} \mathscr{H}^1(\partial \mathcal{P}(\varphi)). \tag{2.15}$$

Together with the estimate (2.14) for the rectangle $APQD$, we get then the validity of (2.12) also in this case.

To conclude, let us consider the situation when $\mathcal{X} = \{\boldsymbol{A}, \boldsymbol{B}\}$, depicted in Figure 6 (right). In this case, the fact that $\mathcal{P}(\varphi)$ is an upper δ-tube implies that the image of DC through φ is the piecewise linear curve $\boldsymbol{D}\boldsymbol{Q}_1\boldsymbol{Q}_2\boldsymbol{C}$, with $\boldsymbol{Q}_1$ and $\boldsymbol{Q}_2$ very close to $\boldsymbol{A}$ and $\boldsymbol{B}$ respectively, and on the two related bisectors. Let us now call again $Q_1 = \varphi^{-1}(\boldsymbol{Q}_1)$, $Q_2 = \varphi^{-1}(\boldsymbol{Q}_2)$, and let P_1 and P_2 be the points on $\partial\mathcal{R}$ below Q_1 and Q_2 as before, with abscissae x_1 and x_2. We extend now the function φ as the linear function on the segments P_1Q_1 and P_2Q_2: φ remains then piecewise linear and injective because the two segments $\boldsymbol{P}_1\boldsymbol{Q}_1$ and $\boldsymbol{P}_2\boldsymbol{Q}_2$ are in the interior of $\mathcal{P}(\varphi)$. Moreover, $\mathcal{R}$ has been subdivided in three rectangles; in the central rectangle $P_1P_2Q_2Q_1$ the function φ corresponds to a δ-tube with $\mathcal{X} = \emptyset$, so the first considered case provides us with an extension v inside this rectangle, satisfying (2.13), which now reads as

$$\int_{P_1P_2Q_2Q_1} \|Dv\| \leq \frac{x_2 - x_1 + b^+ - b^-}{2} \mathscr{H}^1(\partial \mathcal{P}(\varphi)),$$

where we have also used the fact that the perimeter of the polygon $\boldsymbol{P}_1\boldsymbol{P}_2\boldsymbol{Q}_2\boldsymbol{Q}_1$ is less than $\mathscr{H}^1(\partial\mathcal{P}(\varphi))$. Instead, the situation in the right rectangle P_2BCQ_2 is exactly the same as in the previous case for the rectangle $PBCQ$, hence the same argument as before gives us an extension v inside this rectangle satisfying (2.15), which this time reads as

$$\int_{P_2BCQ_2} \|Dv\| \leq \frac{b^+ - b^- + a^+ - x_2}{2} \mathscr{H}^1(\partial \mathcal{P}(\varphi)).$$

In the very same way, in the left rectangle AP_1Q_1B we have an extension v with

$$\int_{AP_1Q_1B} \|Dv\| \leq \frac{b^+ - b^- + x_1 - a^-}{2} \mathscr{H}^1(\partial \mathcal{P}(\varphi)).$$

Adding the last three estimates, we finally find the validity of (2.12) also in this last case. □

We are now in position to show Lemma 2.12.

Proof (of Lemma 2.12). We divide the proof in few steps.

Step I. Definition of the points $\boldsymbol{C}_i$, $\boldsymbol{D}_i$, $\boldsymbol{P}_j$, $\boldsymbol{Q}_j$ and $\boldsymbol{V}_j^\alpha$.

Since $\mathcal{P}(\varphi)$ is a δ-tube, we can call

$$\boldsymbol{P}_0 = \varphi(a^-, b^-), \quad \boldsymbol{Q}_0 = \varphi(a^-, b^+), \quad \boldsymbol{P}_{N+1} = \varphi(a^+, b^-), \quad \boldsymbol{Q}_{N+1} = \varphi(a^+, b^+),$$

and there are points $\boldsymbol{P}_j$ and $\boldsymbol{Q}_j$, for $1 \leq j \leq N$, such that the curve $t \mapsto \varphi(t, b^-)$ is the piecewise linear curve $\boldsymbol{P}_0\boldsymbol{P}_1 \dots \boldsymbol{P}_N\boldsymbol{P}_{N+1}$, while $t \mapsto \varphi(t, b^+)$ coincides with $\boldsymbol{Q}_0\boldsymbol{Q}_1 \dots \boldsymbol{Q}_N\boldsymbol{Q}_{N+1}$. Notice that $\partial\mathcal{P}$ only consists of the points $\boldsymbol{P}_j$ and $\boldsymbol{Q}_j$, hence for each $1 \leq j \leq N$ the point $\boldsymbol{X}_j$ in the Definition 2.10 of the δ-tubes coincides either with $\boldsymbol{P}_j$ or with $\boldsymbol{Q}_j$, and in particular $|\boldsymbol{P}_j - \boldsymbol{Q}_j| < \delta$ (this is in general false for $j = 0$ and $j = N + 1$!). Recall that the points $\boldsymbol{P}_0$ and $\boldsymbol{P}_1$ are not necessarily different, as well as $\boldsymbol{P}_N$ and $\boldsymbol{P}_{N+1}$, $\boldsymbol{Q}_0$ and $\boldsymbol{Q}_1$, and $\boldsymbol{Q}_N$ and $\boldsymbol{Q}_{N+1}$. However, this does not make any substantial difference in the present proof.

We select several numbers $a^- = s_0 < s_1 < \cdots < s_K < s_{K+1} = a^+$, in such a way that, calling $C_i = (s_i, b^-)$ and $D_i = (s_i, b^+)$, φ is linear on each segment C_iC_{i+1} and D_iD_{i+1}; moreover, for each $1 \leq i \leq K$ we define $\boldsymbol{C}_i = \varphi(C_i)$ and $\boldsymbol{D}_i = \varphi(D_i)$. By continuity of the length, we can take points in such a way that, for each $0 \leq i \leq K$ and for each $0 \leq t \leq 1$, one has

$$|d_{\mathcal{P}}((1-t)\boldsymbol{C}_i + t\boldsymbol{C}_{i+1}, (1-t)\boldsymbol{D}_i + t\boldsymbol{D}_{i+1}) - d_{\mathcal{P}}(\boldsymbol{C}_i, \boldsymbol{D}_i)| < \delta. \tag{2.16}$$

For any $1 \leq j \leq N$, let us now divide the segment $\boldsymbol{P}_j\boldsymbol{Q}_j$ in $K + N$ equal parts, calling $\boldsymbol{V}_j^\alpha$, for $0 \leq \alpha \leq K + N$, the corresponding points, in particular $\boldsymbol{V}_j^0 = \boldsymbol{Q}_j$ while $\boldsymbol{V}_j^{K+N} = \boldsymbol{P}_j$. We do the same also for $j = 0$ and $j = N + 1$, calling $\boldsymbol{A}^\alpha$ and $\boldsymbol{B}^\alpha$ (instead of $\boldsymbol{V}_0^\alpha$ and $\boldsymbol{V}_{N+1}^\alpha$) the corresponding points. Similarly, we divide in $K + N$ parts the segments $\{a^-\} \times [b^-, b^+]$ and $\{a^+\} \times [b^-, b^+]$, defining the points A^α and B^α for $0 \leq \alpha \leq K + N$. Notice that $\varphi(A^\alpha) = \boldsymbol{A}^\alpha$ and $\varphi(B^\alpha) = \boldsymbol{B}^\alpha$ since φ is linear on the vertical sides of $\mathcal{R}$. Again by continuity of the length, and again adding new coordinates s_i if necessary, we can assume that for every $0 \leq \alpha \leq N + K - 1$ and every $0 \leq t \leq 1$ one has

$$|d_{\mathcal{P}}((1-t)\boldsymbol{A}^\alpha + t\boldsymbol{A}^{\alpha+1}, (1-t)\boldsymbol{B}^\alpha + t\boldsymbol{B}^{\alpha+1}) - d_{\mathcal{P}}(\boldsymbol{A}^\alpha, \boldsymbol{B}^\alpha)| < \delta. \tag{2.17}$$

Step II. Definition of the curves γ_i and γ^α.

Our next aim is to define suitable piecewise linear curves γ^α and γ_i in $\mathcal{P}$, for $1 \leq \alpha \leq K + N - 1$ and $1 \leq i \leq K$. Each curve γ^α will connect $\boldsymbol{A}^\alpha$ with $\boldsymbol{B}^\alpha$, and it will be eventually the image of the segment $A^\alpha B^\alpha$ under v; similarly, each curve γ_i will connect $\boldsymbol{C}_i$ with $\boldsymbol{D}_i$, and it will be eventually the image of the segment C_iD_i under v.

The curves γ^α are easy to define; namely, for each $1 \leq \alpha \leq K + N - 1$ we let γ^α be the piecewise linear curve $\boldsymbol{A}^\alpha\boldsymbol{V}_1^\alpha\boldsymbol{V}_2^\alpha \dots \boldsymbol{V}_N^\alpha\boldsymbol{B}^\alpha$: each curve γ^α is entirely contained in the interior of $\mathcal{P}$, except its two endpoints $\boldsymbol{A}^\alpha$ and $\boldsymbol{B}^\alpha$, and each two such curves have empty intersection. Notice that, by definition of δ-tubes, the geodesic in $\mathcal{P}$ between any $\boldsymbol{A}^\alpha$ and the corresponding $\boldsymbol{B}^\alpha$ is the curve $\boldsymbol{A}^\alpha\boldsymbol{X}_1\boldsymbol{X}_2 \dots \boldsymbol{X}_N\boldsymbol{B}^\alpha$; then, since for every $1 \leq i \leq N$ one has $|\boldsymbol{V}_i^\alpha - \boldsymbol{X}_i| < \delta$, we immediately derive

$$\mathscr{H}^1(\gamma^\alpha) < d_{\mathcal{P}}(\boldsymbol{A}^\alpha, \boldsymbol{B}^\alpha) + N\delta. \tag{2.18}$$

In order to define the curves γ_i, we first need to consider a generic segment $\boldsymbol{P}_j\boldsymbol{Q}_j$, with $1 \leq j \leq N$: since φ is linear on the segments C_iC_{i+1} and D_iD_{i+1}, $\boldsymbol{P}_j$ must necessarily be one of the points $\boldsymbol{C}_i$, and $\boldsymbol{Q}_j$ one of the points $\boldsymbol{D}_i$. So, we have $i_1(j)$ and $i_2(j)$ in $\{1, 2, \dots, K\}$ with the property that $\boldsymbol{C}_{i_1(j)} = \boldsymbol{P}_j$ and $\boldsymbol{C}_{i_2(j)} = \boldsymbol{Q}_j$. If $i_1(j) > i_2(j)$, we will say that *geodesics are arriving from left at* $\boldsymbol{P}_j\boldsymbol{Q}_j$, while if $i_1(j) < i_2(j)$ we say that *geodesics are arriving from right at* $\boldsymbol{P}_j\boldsymbol{Q}_j$. The reason for the name is clear: if we assume that $i_1(j) > i_2(j)$, then the geodesic between $\boldsymbol{C}_i$ and $\boldsymbol{D}_i$ in $\mathcal{P}$ crosses the segment $\boldsymbol{P}_j\boldsymbol{Q}_j$ if and only if $i_2(j) < i < i_i(j)$: in this case, $\boldsymbol{C}_i$ is "on the left" of the segment $\boldsymbol{P}_j\boldsymbol{Q}_j$ (that is, $C_i \in \partial\mathcal{R}$ is on the left of the segment $\varphi^{-1}(\boldsymbol{P}_j)\varphi^{-1}(\boldsymbol{Q}_j) = C_{i_1(j)}D_{i_2(j)}$), while $\boldsymbol{D}_i$ is on the right. Notice that, at every segment $\boldsymbol{P}_j\boldsymbol{Q}_j$, either geodesics are arriving from left, or they are arriving from right, or $i_1(j) = i_2(j)$.

Let us now fix $1 \le i \le K$. Any curve in $\mathcal{P}$ between $\boldsymbol{C}_i$ and $\boldsymbol{D}_i$ must intersect the closed segment $\boldsymbol{P}_j\boldsymbol{Q}_j$ for every j such that

$$\min\{i_1(j), i_2(j)\} \le i \le \max\{i_1(j), i_2(j)\}. \tag{2.19}$$

Assume that this inequality is false for every $1 \le j \le N$: then, there must exist some j such that $\boldsymbol{C}_i$ is in the interior of the segment $\boldsymbol{P}_j\boldsymbol{P}_{j+1}$, while $\boldsymbol{D}_i$ is in the interior of $\boldsymbol{Q}_j\boldsymbol{Q}_{j+1}$. By construction, the open segment $\boldsymbol{C}_i\boldsymbol{D}_i$ is entirely contained in $\mathcal{P}$, and we define $\gamma_i = \boldsymbol{C}_i\boldsymbol{D}_i$.

Assume, instead, that there are some $1 \le j \le N$ such that the inequality (2.19) holds, and let $j_{\min}$ and $j_{\max}$ the minimal and the maximal j for which this happens (of course they depend on i). It is immediate to notice that (2.19) holds true for all $j_{\min} \le j \le j_{\max}$. If $i_1(j_{\min}) = i_2(j_{\min})$, then of course $j_{\max} = j_{\min}$ and the points $\boldsymbol{C}_i$ and $\boldsymbol{D}_i$ coincide with $\boldsymbol{P}_{j_{\min}}$ and $\boldsymbol{Q}_{j_{\min}}$ respectively: also in this case we set $\gamma_i = \boldsymbol{C}_i\boldsymbol{D}_i$.

Otherwise, either $i_1(j_{\min}) < i_2(j_{\min})$ or $i_1(j_{\min}) > i_2(j_{\min})$, so geodesics arrive either from right or from left at $\boldsymbol{P}_{j_{\min}}\boldsymbol{Q}_{j_{\min}}$. It readily follows from the definition that in the first case geodesics arrive from right at every $\boldsymbol{P}_j\boldsymbol{Q}_j$ with $j_{\min} \le j \le j_{\max}$, while in the second case they all arrive from left. So, we define the curve γ_i as

$$\gamma_i = \boldsymbol{T}_i^- \boldsymbol{V}_{j_{\min}}^{\alpha(i, j_{\min})} \boldsymbol{V}_{j_{\min}+1}^{\alpha(i, j_{\min}+1)} \dots \boldsymbol{V}_{j_{\max}}^{\alpha(i, j_{\max})} \boldsymbol{T}_i^+, \tag{2.20}$$

where $\boldsymbol{T}_i^- = \boldsymbol{C}_i$ and $\boldsymbol{T}_i^+ = \boldsymbol{D}_i$ if geodesics arrive from left at $\boldsymbol{P}_j\boldsymbol{Q}_j$, while otherwise $\boldsymbol{T}_i^- = \boldsymbol{D}_i$ and $\boldsymbol{T}_i^+ = \boldsymbol{C}_i$, and where for every $1 \le i \le K$ and $1 \le j \le N$ the number $\alpha(i, j)$ is defined as

$$\alpha(i, j) = \begin{cases} i + N - j & \text{if geodesics arrive from left at } \boldsymbol{P}_j\boldsymbol{Q}_j, \\ j + K - i & \text{if geodesics arrive from right at } \boldsymbol{P}_j\boldsymbol{Q}_j. \end{cases} \tag{2.21}$$

Notice that we are defining $\alpha(i, j)$ for every i and j, except if j is so that $i_1(j) = i_2(j)$. However, in this case the definition of γ_i does not require the definition of $\alpha(i, j)$: indeed, if $i \ne i_1(j)$, then j does not belong to the interval $[j_{\min}, j_{\max}]$, so the value of $\alpha(i, j)$ is not needed in (2.20); instead, if $i = i_1(j) = i_2(j)$, then γ_i is not defined through (2.20), it is simply the segment $\boldsymbol{P}_j\boldsymbol{Q}_j$, and so also in this case the definition of $\alpha(i, j)$ is not used.

An immediate consequence of our definition is that all the curves γ_i lie in the interior of $\mathcal{P}$, except their two endpoints, and two different such curves are disjoint: notice that the disjointness is true because of the different definitions of $\alpha(i, j)$ given in the two possible cases in (2.21).

To conclude this step, we claim that

$$\mathscr{H}^1(\gamma_i) < d_{\mathcal{P}}(\boldsymbol{C}_i, \boldsymbol{D}_i) + N\delta. \tag{2.22}$$

In fact, depending on the position of the points we have defined γ_i either as the segment between $\boldsymbol{C}_i$ and $\boldsymbol{D}_i$, or through (2.20). In the first case, γ_i is of course the geodesic between $\boldsymbol{C}_i$ and $\boldsymbol{D}_i$ in $\mathcal{P}$, so in fact $\mathscr{H}^1(\gamma_i) = d_{\mathcal{P}}(\boldsymbol{C}_i, \boldsymbol{D}_i)$ and then (2.22) clearly holds. Suppose, instead, that γ_i has been defined through (2.20); then, γ_i is a piecewise linear path between $\boldsymbol{C}_i$ and $\boldsymbol{D}_i$, and its vertices are $\boldsymbol{C}_i$, $\boldsymbol{D}_i$, and one single point in every segment $\boldsymbol{P}_j\boldsymbol{Q}_j$ such that j satisfies (2.19). Let us instead consider the geodesic in $\mathcal{P}$ between $\boldsymbol{C}_i$ and $\boldsymbol{D}_i$: as noticed above, it must also intersect the segments $\boldsymbol{P}_j\boldsymbol{Q}_j$ for all the j satisfying (2.19). Since the total number of segments $\boldsymbol{P}_j\boldsymbol{Q}_j$ is N and each one has length smaller than δ by construction, the validity of (2.22) simply comes by triangular inequality.

<u>Step III.</u> *Uniqueness of the intersection $\gamma_i \cap \gamma^\alpha$.*

Let us now fix any $1 \leq i \leq K$, and any $1 \leq \alpha \leq N + K - 1$. We aim to show that the intersecion $\gamma_i \cap \gamma^\alpha$ consists of exactly one point; notice that this intersection is surely not empty, because A^α and B^α are on the two different parts in which $\partial\mathcal{R}$ is divided by the points C_i and D_i, so we only have to exclude multiple intersections.

Recall that the construction of γ_i has been done, in Step II, in three different cases: we will consider them separately. The first possibility was when (2.19) was false for every $1 \leq j \leq N$: in this case, there exists some $0 \leq \bar{j} \leq N$ such that $\boldsymbol{C}_i$ is in the open segment $\boldsymbol{P}_{\bar{j}}\boldsymbol{P}_{\bar{j}+1}$ and $\boldsymbol{D}_i$ in the open segment $\boldsymbol{Q}_{\bar{j}}\boldsymbol{Q}_{\bar{j}+1}$, and γ_i was simply defined as the segment $\boldsymbol{C}_i\boldsymbol{D}_i$. Since the whole segment γ_i lies in the interior of the quadrilateral $\boldsymbol{P}_{\bar{j}}\boldsymbol{P}_{\bar{j}+1}\boldsymbol{Q}_{\bar{j}+1}\boldsymbol{Q}_{\bar{j}}$, and in this quadrilateral also γ^α is a segment, then the intersection consists of exactly a point, so the step is concluded in this first case.

The second case was if (2.19) holds true for some j, then in particular for all $j_{\min} \leq j \leq j_{\max}$, and $i_1(j_{\min}) = i_2(j_{\min})$: in this case, we had noticed that $\boldsymbol{C}_i = \boldsymbol{P}_{j_{\min}}$ and $\boldsymbol{D}_i = \boldsymbol{Q}_{j_{\min}}$, and defined γ_i as the segment $\boldsymbol{P}_{j_{\min}}\boldsymbol{Q}_{j_{\min}}$; the unique intersection of γ_i with γ^α is then again obvious.

Let us finally consider the third and last possible case studied in Step II, namely, (2.19) holds true for all $j_{\min} \leq j_{\max}$ and $i_1(j_{\min}) \neq i_2(j_{\min})$. In this case, γ_i has been defined through (2.20), and we have also noticed that geodesics are arriving from left at $\boldsymbol{P}_j\boldsymbol{Q}_j$ for every $j_{\min} \leq j \leq j_{\max}$, or they are arriving from right for every $j_{\min} \leq j \leq j_{\max}$. Let us suppose that they arrive all from left, the other case is completely similar. Then, the numbers $\alpha(i, j)$ defined in (2.21) are all given by $\alpha(i, j) = i + N - j$, so they range from $i + N - j_{\max}$ to $i + N - j_{\min}$. Suppose first that $\alpha > i + N - j_{\min}$: in this case, the point $\boldsymbol{V}_{j_{\min}}^{\alpha(i, j_{\min})} \in \gamma_i$ is above the point $\boldsymbol{V}_{j_{\min}}^{\alpha} \in \gamma^\alpha$, on the segment $\boldsymbol{P}_{j_{\min}}\boldsymbol{Q}_{j_{\min}}$, and more in general the point

$V_j^{\alpha(i,j)} \in \gamma_i$ is above $V_j^\alpha \in \gamma^\alpha$ for every $j_{\min} \le j \le j_{\max}$. So, there is no intersection between γ_i and γ^α after the segment $P_{j_{\min}} Q_{j_{\min}}$. On the other hand, in the closed quadrilateral $P_{j_{\min}-1} P_{j_{\min}} Q_{j_{\min}} Q_{j_{\min}-1}$ both the curves γ_i and γ^α are segments, so they have at most one point of intersection; and, since of course there is no intersection before $P_{j_{\min}-1} Q_{j_{\min}-1}$ and the intersection cannot be empty, then $\gamma_i \cap \gamma^\alpha$ consist of exactly a point. So, we have proved the claim under the assumption that $\alpha > i + N - j_{\min}$. If, on the contrary, $\alpha < i + N - j_{\max}$, then a completely symmetric argument shows that $\gamma_i \cap \gamma^\alpha$ consists of exactly a point, lying in the quadrilateral $P_{j_{\max}} P_{j_{\max}+1} Q_{j_{\max}+1} Q_{j_{\max}}$. Finally, if $i + N - j_{\max} \le \alpha \le i + N - j_{\min}$, then there exists exactly one $j_{\min} \le \bar{j} \le j_{\max}$ such that $\alpha(i, \bar{j}) = \alpha$. And then, the intersection between γ_i and γ^α consists of the sole point $V_{\bar{j}}^\alpha$.

Step IV. The curves γ_i and γ^α are segments between two consecutive intersections.

Thanks to Step III, for each $1 \le i \le K$ and $1 \le \alpha \le N + K - 1$ there is exactly one point of intersection between γ_i and γ^α, call it S_i^α. We can extend the definition of the points S_i^α also to $i = 0$ or $i = K + 1$, as well as to $\alpha = 0$ or $\alpha = K + N$, of course setting

$$S_i^0 = C_i, \quad S_i^{K+N} = D_i, \quad S_0^\alpha = A^\alpha, \quad S_{K+1}^\alpha = B^\alpha.$$

The aim of this step is to show that γ_i is a segment between any two consecutive points S_i^α, and the same happens to γ^α. Since all the curves γ_i and γ^α are piecewise linear, the claim is equivalent to say that every vertex of each curve γ_i is the intersection point with some γ^α, and analogously every vertex of each curve γ^α is the intersection point with some γ_i. Since, by construction, every vertex of any curve γ_i or γ^α is necessarily one of the points V_j^α, then it is enough to show that, for every $1 \le \bar{j} \le N$ and $1 \le \bar{\alpha} \le N + K - 1$, the point $V_{\bar{j}}^{\bar{\alpha}}$ belongs to some curve γ_i, as well as to some curve γ^α. Moreover, since of course $V_{\bar{j}}^{\bar{\alpha}}$ belongs to the curve $\gamma^{\bar{\alpha}}$, we only have to show that it belongs to some curve γ_i.

Let us then consider the segment $P_{\bar{j}} Q_{\bar{j}}$, and keep in mind that either geodesics are arriving there from left, or from right, or none of the two, the last possibility being true if and only if $i_1(\bar{j}) = i_2(\bar{j})$. In this last case, the curve $\gamma_{i_1(\bar{j})}$ coincides by definition with the whole segment $P_{\bar{j}} Q_{\bar{j}}$, hence in particular it contains $V_{\bar{j}}^{\bar{\alpha}}$. Let us then suppose that geodesics are arriving at $P_{\bar{j}} Q_{\bar{j}}$ from left, if they are arriving from right the completely symmetric argument can be done. Then, the geodesic γ_i has some intersection with $P_{\bar{j}} Q_{\bar{j}}$ if and only if $i_2(\bar{j}) \le i \le i_1(\bar{j})$. In particular, keeping in mind (2.21), we know that for any such i the geodesic γ_i passes through the point $V_{\bar{j}}^{i+N-\bar{j}}$. If $i_2(\bar{j}) + N - \bar{j} \le \bar{\alpha} \le i_1(\bar{j}) + N - \bar{j}$, then we are clearly done. Suppose, instead, that $\bar{\alpha} > i_1(\bar{j}) + N - \bar{j}$, again a completely symmetric argument would work if $\bar{\alpha} < i_2(\bar{j}) + N - \bar{j}$. Let us then set $\bar{i} = i_1(\bar{j})$: notice that by definition $C_{\bar{i}} = C_{i_1(\bar{j})} = P_{\bar{j}}$, and $j_{\min}(\bar{i}) = \bar{j}$. Hence, keeping in mind (2.20), we know that the first two points of the curve $\gamma_{\bar{i}}$ are $P_{\bar{j}}$ and $V_{\bar{j}}^{i_1(\bar{j})+N-\bar{j}}$. We conclude simply by observing that, since $\bar{\alpha} > i_1(\bar{j}) + N - \bar{j}$, then the point $V_{\bar{j}}^{\bar{\alpha}}$ is contained in the open segment $P_{\bar{j}} V_{\bar{j}}^{i_1(\bar{j})+N-\bar{j}}$, which is a part of $\gamma_{\bar{i}}$, hence $V_{\bar{j}}^{\bar{\alpha}} \in \gamma_{\bar{i}}$ and the step is thus concluded.

Step V. *Conclusion.*

We are now ready to conclude the construction. In fact, for every $0 \le i \le K+1$ and every $0 \le \alpha \le N+K$ we define the point $S_i^\alpha = (s_i, y_\alpha) \in \mathcal{R}$, where

$$y_\alpha = b^+ - \frac{\alpha}{N+K}(b^+ - b^-).$$

Then, we let $v(S_i^\alpha) = \boldsymbol{S}_i^\alpha$, and we extend v linearly in every segment $S_i^\alpha S_{i+1}^\alpha$ and in every segment $S_i^\alpha S_i^{\alpha+1}$. Notice that v is now defined on the one-dimensional grid in $\mathcal{R}$ made by all the points whose first coordinate is one of the s_i, or whose second coordinate is one of the y_α; notice also that v coincides with φ on $\partial\mathcal{R}$. By Steps III and IV we obtain that v is injective on the grid and that its image is done by the union of the curves γ_i and of the curves γ^α. In particular, for each rectangle $\mathcal{R}_i^\alpha = [s_i, s_{i+1}] \times [y_\alpha, y_{\alpha+1}]$, the function v is injective on $\partial\mathcal{R}_i^\alpha$, and linear on each of its four sides: therefore, we can apply Corollary 2.14 to each rectangle.

More precisely, let us fix any $0 \le i \le K$ and $0 \le \alpha \le N+K-1$, and let us apply Corollary 2.14 to the rectangle $\mathcal{R}_i^\alpha$, finding a finitely piecewise affine and injective function $v_i^\alpha : \mathcal{R}_i^\alpha \to \mathbb{R}^2$, which extends the function v already defined on $\partial\mathcal{R}_i^\alpha$ and for which the estimates (2.10) are valid. In particular, for every $0 \le t \le 1$ let us define $\gamma(i,\alpha)_t : [y_\alpha, y_{\alpha+1}] \to \mathbb{R}^2$ and $\gamma(i,\alpha)^t : [s_i, s_{i+1}] \to \mathbb{R}^2$ the curves defined by

$$\gamma(i,\alpha)_t(s) = v_i^\alpha((1-t)s_i + ts_{i+1}, s), \quad \gamma(i,\alpha)^t(s) = v_i^\alpha(s, (1-t)y_\alpha + ty_{\alpha+1});$$

hence, (2.10) reads as

$$\begin{aligned} \mathscr{H}^1(\gamma(i,\alpha)_t) &\le (1-t)\mathscr{H}^1(\gamma(i,\alpha)_0) + t\mathscr{H}^1(\gamma(i,\alpha)_1), \\ \mathscr{H}^1(\gamma(i,\alpha)^t) &\le (1-t)\mathscr{H}^1(\gamma(i,\alpha)^0) + t\mathscr{H}^1(\gamma(i,\alpha)^1). \end{aligned} \tag{2.23}$$

Let us finally define $v : \mathcal{R} \to \mathbb{R}^2$ as the function which coincides with v_i^α on each rectangle $\mathcal{R}_i^\alpha$: by construction, we know that v is an injective, finitely piecewise affine function which coincides with φ on $\partial\mathcal{R}$. To conclude, we need then to prove the validity of (2.9).

To do so, let us take any $b^- \le y \le b^+$; then, there exist an index $0 \le \alpha \le N+K-1$ and a number $0 \le t \le 1$ such that $y = (1-t)y_\alpha + ty_{\alpha+1}$. Observe now that the curve $[a^-, a^+] \ni s \mapsto v(s,y)$ is simply the union of the curves $[s_i, s_{i+1}] \ni s \mapsto \gamma(i,\alpha)^t(s)$, with i ranging from 0 to K. Therefore, by (2.23), (2.18) and (2.17) we get

$$\begin{aligned} \int_{a^-}^{a^+} |D_1 v(s,y)|\,ds &= \sum_{i=0}^K \mathscr{H}^1(\gamma(i,\alpha)^t) \le \sum_{i=0}^K (1-t)\mathscr{H}^1(\gamma(i,\alpha)^0) + t\mathscr{H}^1(\gamma(i,\alpha)^1) \\ &= (1-t)\mathscr{H}^1(\gamma^\alpha) + t\mathscr{H}^1(\gamma^{\alpha+1}) \\ &\le (1-t)d_{\mathcal{P}}(\boldsymbol{A}^\alpha, \boldsymbol{B}^\alpha) + td_{\mathcal{P}}(\boldsymbol{A}^{\alpha+1}, \boldsymbol{B}^{\alpha+1}) + N\delta \end{aligned}$$

$$\leq d_{\mathcal{P}}((1-t)\boldsymbol{A}^{\alpha}+t\boldsymbol{A}^{\alpha+1},(1-t)\boldsymbol{B}^{\alpha}+t\boldsymbol{B}^{\alpha+1})+(N+2)\delta$$
$$= d_{\mathcal{P}}(\varphi(a^-,y),\varphi(a^+,y))+(N+2)\delta.$$

It is then enough to integrate in the variable $y \in [b^-, b^+]$ to get

$$\int_{\mathcal{R}} |D_1 v| = \int_{y=b^-}^{b^+} \int_{s=a^-}^{a^+} |D_1 v(s,y)|\, ds\, dy$$
$$\leq \int_{y=b^-}^{b^+} d_{\mathcal{P}}(\varphi(a^-,y),\varphi(a^+,y))\, dy + (b^+ - b^-)(N+2)\delta.$$

The very same argument, done for the "vertical" curves $y \mapsto v(s,y)$ for any $a^- \leq s \leq a^+$, and using then (2.22) and (2.16) in place of (2.18) and (2.17), yields that

$$\int_{\mathcal{R}} |D_2 v| \leq \int_{s=a^-}^{a^+} d_{\mathcal{P}}(\varphi(s,b^-),\varphi(s,b^+))\, dy + (a^+ - a^-)(N+2)\delta.$$

Then, adding the last two estimates and recalling Definition 1.1 one obtains (2.9), thus concluding the proof. □

2.4. The proof of Proposition 2.2

In this last subsection we can give the proof of Proposition 2.2; thanks to the results of the previous subsections, this is now rather simple.

PROOF (OF PROPOSITION 2.2). First of all we notice that, given any piecewise linear and injective map $g : \partial\mathcal{R} \to \mathbb{R}^2$ and any $\sigma \ll 1$, there exists a finitely piecewise affine bijection $\Phi : \mathbb{R}^2 \to \mathbb{R}^2$, bi-Lipschitz with constant at most $1+\sigma$, such that $\Phi \circ g$, which is of course still piecewise linear and injective, is also not aligned. As a consequence, we can assume without loss of generality that the map φ is not aligned.

Let then $\eta \ll 1$ be a small constant, depending on $\mathcal{R}$, on φ and on ε and to be specified later. We apply Lemma 2.11 to φ, thus getting the ordinates $\{y_i\}_{i=0}^{M}$, the piecewise linear and injective extension $\tilde{\varphi} : \bigcup_{i=0}^{M-1} \partial\mathcal{R}_i \to \mathbb{R}^2$ of φ, the constants δ_i and the bijections Φ_i as in the claim. In particular, $\tilde{\varphi}$ satisfies (2.4).

For every $1 \leq i < M-1$, the polygon $\mathcal{P}(\varphi_i)$ is a δ_i-tube, up to apply the finitely piecewise affine bijection Φ_i, of bi-Lipschitz constant at most $1+\eta$; hence, we can apply Lemma 2.12 to get a finitely piecewise affine and injective function $v_i : \mathcal{R}_i \to \mathbb{R}^2$ such that $v_i = \varphi_i$ on $\partial\mathcal{R}_i$ and

$$\frac{1}{(1+\eta)^2}\int_{\mathcal{R}_i} \|Dv_i\| \leq \Psi(\varphi_i) + (a^+ - a^- + y_{i+1} - y_i)(N_i+2)\delta_i.$$

Let us now call $\mathcal{R}_{\text{int}} = \bigcup_{i=1}^{M-2} \mathcal{R}_i$, and let $v : \mathcal{R}_{\text{int}} \to \mathbb{R}^2$ be the function coinciding with v_i on every $\mathcal{R}_i$; by construction, it is finitely piecewise affine and injective, and it coincides with $\tilde{\varphi}$ on $\partial\mathcal{R}_{\rm int}$. Since, by Lemma 2.11, for every i the number N_i is smaller than the total number T of vertices of $\mathcal{P}(\varphi)$, while $\delta_i \ll \eta/M$, adding the above inequality for all $1 \leq i < M-1$ we get

$$\frac{1}{(1+\eta)^2}\int_{\mathcal{R}_{\rm int}} \|Dv\| \leq \sum_{i=1}^{M-2} \Psi(\varphi_i) + (a^+ - a^- + b^+ - b^-)T\eta. \tag{2.24}$$

To conclude the definition of v, we need to give it on the two strips $\mathcal{R}_0$ and $\mathcal{R}_{M-1}$: we will concentrate on $\mathcal{R}_0$, the case of $\mathcal{R}_{M-1}$ will then be identical.

Consider the rectangle $\mathcal{R}_0$, and the function φ_0: by construction, φ_0 is piecewise linear and injective, and moreover φ_0 is linear on the two (very short) vertical sides of $\mathcal{R}_0$. As before, we can assume without loss of generality that φ_0 is also not aligned. Then, we can apply once again Lemma 2.11, this time to φ_0 and $\mathcal{R}_0$; more precisely, we apply the "rotated" version of Lemma 2.11, where the horizontal side $[a^-, a^+]$, instead of the vertical one, is subdivided. Thus, we get some abscissae $a^- = x_0 < x_1 < \cdots < x_{P-1} < x_P = a^+$, and a piecewise linear and injective function $\hat{\varphi} : \bigcup_{j=0}^{P-1} \partial\mathcal{R}_0^j \to \mathbb{R}^2$, being $\mathcal{R}_0^j = [x_j, x_{j+1}] \times [y_0, y_1]$, with $\hat{\varphi} = \tilde{\varphi}$ on $\partial\mathcal{R}_0$ and

$$\sum_{j=0}^{P-1} \Psi(\varphi_0^j) < \Psi(\varphi_0) + \eta(a^+ - a^-), \tag{2.25}$$

where φ_0^j is the restriction of $\hat{\varphi}$ to $\partial\mathcal{R}_0^j$. Moreover, for every $1 \leq j \leq P-2$ the polygon $\mathcal{P}(\varphi_0^j)$ is a (rotated) δ_0^j-tube for some $\delta_0^j \ll \eta/P$, up to a $(1+\eta)$-biLipschitz finitely piecewise affine bijection. Thus, we can apply Lemma 2.12 to each rectangle $\mathcal{R}_0^j$ with $1 \leq j \leq P-2$ to find an extension v_0^j of $\hat{\varphi}$ inside the rectangle; exactly as before, let us call v the function which coincides with v_0^j on each rectangle $\mathcal{R}_0^j$, hence in place of (2.24) we find

$$\frac{1}{(1+\eta)^2}\int_{\mathcal{R}_{0,\,\rm int}} \|Dv\| \leq \sum_{j=1}^{P-2} \Psi(\varphi_0^j) + (a^+ - a^- + y_1 - y_0)T\eta, \tag{2.26}$$

where $\mathcal{R}_{0,\,\rm int} = \bigcup_{j=1}^{P-2} \mathcal{R}_0^j$ (notice that the total number of vertices of $\mathcal{P}(\varphi_0)$ is not greater than T by construction). Notice that, up to now, the function v has been defined on $\mathcal{R}_{\rm int} \cup \mathcal{R}_{0,\,\rm int}$, it is by construction injective and finitely piecewise affine, and in addition it coincides with φ on $\partial\mathcal{R} \cap \partial(\mathcal{R}_{\rm int} \cup \mathcal{R}_{0,\,\rm int})$, as well as with $\tilde{\varphi}$ on $\partial\mathcal{R}_{\rm int}$ and with $\hat{\varphi}$ on $\partial\mathcal{R}_{0,\,\rm int}$.

To conclude the definition of v on the whole $\mathcal{R}_0$ we have then only to take care of the two rectangles $\mathcal{R}_0^0$ and $\mathcal{R}_0^{P-1}$; notice that both the rectangles have both the lengths smaller than η. Let us first consider $\mathcal{R}_0^0$: by construction, the function $\hat{\varphi}$ is linear on its left side, as well as on both its horizontal sides; in addition, by Lemma 2.11 $\mathcal{P}(\varphi_0^0)$ is an upper δ-tube (actually, since we have applied the "rotated" version of Lemma 2.11, it would be consistent to speak about a "right δ-tube", instead of an upper one). Then, we are allowed to apply (the rotated version of) Lemma 2.15, finding an extension v of φ_0^0 on the whole rectangle $\mathcal{R}_0^0$ which satisfies

$$\int_{\mathcal{R}_0^0} \|Dv\| \leq 3\eta \mathscr{H}^1(\partial \mathcal{P}(\varphi_0^0)) \leq \eta.$$

Notice that the last inequality is true as soon as the perimeter of the polygon $\mathcal{P}(\varphi_0^0)$ is less than $1/3$; but in fact, this is true up to take η small enough, because the continuity of φ immediately implies that $\mathcal{P}(\varphi_0^0)$ can be taken arbitrarily small, up to take $\eta \ll 1$. Since the very same can be done on the rectangle $\mathcal{R}_0^{P-1}$, we end up with a piecewise linear and injective function v on the whole bottom rectangle $\mathcal{R}_0$, which by (2.26) and (2.25) satisfies

$$\begin{aligned}\int_{\mathcal{R}_0} \|Dv\| &\leq (1+\eta)^2 \Big(\sum_{j=1}^{P-2} \Psi(\varphi_0^j) + (a^+ - a^- + y_1 - y_0) T\eta \Big) + 2\eta \\ &\leq (1+\eta)^2 (\Psi(\varphi_0) + (a^+ - a^- + b^+ - b^-)(T+1)\eta + 2\eta).\end{aligned} \tag{2.27}$$

Of course, the very same construction done for the bottom rectangle $\mathcal{R}_0$ can be done also for the top rectangle $\mathcal{R}_{M-1}$, so we find a last extension of v on the top rectangle such that

$$\int_{\mathcal{R}_{M-1}} \|Dv\| \leq (1+\eta)^2 (\Psi(\varphi_{M-1}) + (a^+ - a^- + b^+ - b^-)(T+1)\eta + 2\eta). \tag{2.28}$$

Altogether, our final function $v : \mathcal{R} \to \mathbb{R}^2$ is fintely piecewise affine and injective by construction, it coincides with φ on $\partial \mathcal{R}$, and putting together (2.24), (2.27), (2.28) and (2.4) we obtain

$$\begin{aligned}\int_{\mathcal{R}} \|Dv\| &= \int_{\mathcal{R}_{\text{int}}} \|Dv\| + \int_{\mathcal{R}_0} \|Dv\| + \int_{\mathcal{R}_{M-1}} \|Dv\| \\ &\leq (1+\eta)^2 \Big(\sum_{i=0}^{M-1} \Psi(\varphi_i) + (a^+ - a^- + b^+ - b^-)(3T+2)\eta + 4\eta \Big) \\ &\leq (1+\eta)^2 (\Psi(\varphi) + (a^+ - a^- + b^+ - b^-)(3T+3)\eta + 4\eta) \leq \Psi(\varphi) + \varepsilon,\end{aligned}$$

where the last inequality is clearly true up to choose η small enough, depending only on $\mathcal{R}$ and on φ. The proof is then concluded. □

3. The proof of Theorem A: the general case

This section is devoted to show the general case of Theorem A. Since we have already proved the result in the special case of piecewise linear boundary data, the idea is simply to reduce ourselves to that case, decomposing the rectangle as a countable but locally finite union of rectangles, on each of which the piecewise linear case can be applied.

PROOF (OF THEOERM A). We start by observing, once again, that it is enough to find a piecewise affine homeomorphism $v \in \mathrm{Ext}(\varphi)$ such that (1.2) holds true, because then (1.1) follows directly from Lemma 2.1.

For every $a^- < t < a^+$ and for every $b^- < s < b^+$, we call

$$\begin{aligned} A_s &= (a^-, s), & B_s &= (a^+, s), & C_t &= (t, b^-), & D_t &= (t, b^+), \\ \boldsymbol{A}_s &= \varphi(A_s), & \boldsymbol{B}_s &= \varphi(B_s), & \boldsymbol{C}_t &= \varphi(C_t), & \boldsymbol{D}_t &= \varphi(D_t). \end{aligned}$$

Let us now fix an arbitrary $a^- < t < a^+$, and let us call $\mathcal{R}_1$ and $\mathcal{R}_2$ the two rectangles in which $\mathcal{R}$ is divided by the segment C_tD_t, being $\mathcal{R}_1$ the left one. We want to extend φ to C_tD_t in such a way that, calling φ_1 and φ_2 the restrictions of this extension to $\partial\mathcal{R}_1$ and to $\partial\mathcal{R}_2$, one has

$$\Psi(\varphi_1) + \Psi(\varphi_2) < \Psi(\varphi) + \frac{\varepsilon}{4}. \tag{3.1}$$

To do so, we will argue in a way very similar to what already done in Step III of the proof of Lemma 2.11. More precisely, let us call $\delta = \delta(\mathcal{R}, \varepsilon)$ a small constant, to be specified later, and let γ be a curve, contained in the interior of $\mathcal{P}(\varphi)$, which connects $\boldsymbol{C}_t$ with $\boldsymbol{D}_t$, and such that $\mathscr{H}^1(\gamma) \leq d_{\mathcal{P}}(\boldsymbol{C}_t, \boldsymbol{D}_t) + \delta$: this curve will eventually be the image of C_tD_t under the extension of φ. Notice that, for every two points $\boldsymbol{P}$ and $\boldsymbol{Q}$ in γ, the length of γ between $\boldsymbol{P}$ and $\boldsymbol{Q}$ is smaller than $d_{\mathcal{P}}(\boldsymbol{P}, \boldsymbol{Q}) + \delta$. As an immediate consequence, if we call $\mathcal{P}_1$ and $\mathcal{P}_2$ the two sets in which $\mathcal{P}$ is divided by γ, being $\mathcal{P}_1$ the one containing $\boldsymbol{C}_\tau$ for every $\tau < t$, we get that

$$d_{\mathcal{P}_i}(\boldsymbol{C}_\tau, \boldsymbol{D}_\tau) \leq d_{\mathcal{P}}(\boldsymbol{C}_\tau, \boldsymbol{D}_\tau) + \delta \tag{3.2}$$

for every $a^- < \tau < b^+$, where $i = 1$ if $\tau \leq t$ and $i = 2$ if $\tau \geq t$.

Notice now that each point $\boldsymbol{A}_s$ belongs to $\mathcal{P}_1$, while $\boldsymbol{B}_s$ belongs to $\mathcal{P}_2$; therefore, the geodesic in $\mathcal{P}$ between each $\boldsymbol{A}_s$ and the corresponding $\boldsymbol{B}_s$ must intersect the curve γ at least once, and we call $\boldsymbol{E}_s$ the last point of this intersection (that is, the one closest to $\boldsymbol{B}_s$ on the curve). Notice that, by construction,

$$d_{\mathcal{P}_1}(\boldsymbol{A}_s, \boldsymbol{E}_s) + d_{\mathcal{P}_2}(\boldsymbol{E}_s, \boldsymbol{B}_s) \leq d_{\mathcal{P}}(\boldsymbol{A}_s, \boldsymbol{B}_s) + \delta.$$

Moreover, by uniqueness of the geodesics we readily obtain that, whenever $\sigma > s$, the point $\boldsymbol{E}_\sigma$ is "above" $\boldsymbol{E}_s$, that is, $\boldsymbol{E}_\sigma$ belongs to the part of γ connecting $\boldsymbol{E}_s$ and $\boldsymbol{D}_t$. The function $s \mapsto \boldsymbol{E}_s$ is then an increasing map from (b^-, b^+) to γ: this function is in general neither continuous, now injective, nor surjective, but exactly as in Step III of the proof of Lemma 2.11 we can easily modify it to get a continuous bijection $g : [b^-, b^+] \to \gamma$ so that

$$\int_{s=b^-}^{b^+} d_{\mathcal{P}_1}(\boldsymbol{A}_s, g(s)) + d_{\mathcal{P}_2}(g(s), \boldsymbol{B}_s)\, ds \leq \int_{s=b^-}^{b^+} d_{\mathcal{P}}(\boldsymbol{A}_s, \boldsymbol{B}_s)\, ds + 2\delta(b^+ - b^-).$$

Putting together this estimate and (3.2), and extending of course φ on C_tD_t as $\varphi(t,s) = g(s)$, we immediately get the validity of (3.1), up to choose δ small enough.

Notice that the only requirement for γ was to be a curve, contained in the interior of $\mathcal{P}(\varphi)$, connecting $\boldsymbol{C}_t$ and $\boldsymbol{D}_t$ and with length smaller than $d_{\mathcal{P}}(\boldsymbol{C}_t, \boldsymbol{D}_t) + \delta$; as a consequence, without loss of generality we can assume φ to be locally piecewise linear on C_tD_t: that is, for every $\eta \ll 1$ the function φ is piecewise linear on the segment connecting $(t, a^- + \eta)$ with $(t, a^+ - \eta)$. Notice also that, of course, an identical argument allows to divide the rectangle $\mathcal{R}$ with a horizontal segment, instead of a vertical one.

Repeating the above "cutting argument" countably many times in the obvious way, we can write $\mathcal{R}$ as the countable but locally finite union of essentially disjoint rectangles $\mathcal{R}_i \subset\subset \mathcal{R}$, extending also the function φ to the union of all the boundaries of the rectangles, so that

$$\sum_{i \in \mathbb{N}} \Psi(\varphi_i) < \Psi(\varphi) + \frac{\varepsilon}{2},$$

where φ_i is the restriction to $\partial\mathcal{R}_i$ of the extended function φ.

Notice now that, by construction, each function φ_i is piecewise linear on the boundary of $\mathcal{R}_i$, because each rectangle $\mathcal{R}_i$ is a positive distance apart from $\partial\mathcal{R}$. Hence, we can apply Proposition 2.2 on each rectangle to get a finitely piecewise affine function $v_i : \mathcal{R}_i \to \mathbb{R}^2$ with $v_i = \varphi_i$ on $\partial\mathcal{R}_i$ and such that

$$\int_{\mathcal{R}_i} \|Dv_i\| \le \Psi(\varphi_i) + \frac{\varepsilon}{2^{i+1}}.$$

And finally, putting together all these functions v_i, we obtain a piecewise affine function $v : \mathcal{R} \to \mathbb{R}^2$, coinciding with φ on $\partial\mathcal{R}$, and satisfying (1.2). □

Remark 3.1. From our construction, it is clear which are the optimal functions in (1.1). In fact, a function u realizes the minimum if and only if its restriction to any horizontal and vertical segment in $\mathcal{R}$ is a geodesic in $\mathcal{P}(\varphi)$ between the endpoints. And in turn, this is possible if and only if φ is a convex quadrilateral and its four sides are the image of the four sides of $\mathcal{R}$. In all the other cases, the infimum in (1.1) is not a minimum. However, it is still clear which are the minimizing sequences: more precisely, a sequence $\{u_j\} \in \mathrm{BV}(\mathcal{R}) \cap \mathrm{Ext}(\varphi)$ is a minimizing sequence for (1.1) if and only if the images of almost every horizontal and vertical segment in $\mathcal{R}$ have lengths which converge to the geodesic distance between the endpoints.

4. The proof of Theorem 1.3 and Theorem 1.4

This section is devoted to prove Theorem 1.3 and Theorem 1.4; the first one will be a very simple consequence of the latter. In order to present the proof, we first need to check that the geodesic distances decrease, if we linearize the Jordan curves. Let us be more precise.

DEFINITION 4.1 (ε-linearization of a Jordan curve). Let τ be a Jordan curve with finite length, and let $\varepsilon > 0$ be much smaller than the diameter of $\mathcal{P}(\tau)$. Let $\widehat{\boldsymbol{A}_1\boldsymbol{B}_1}, \widehat{\boldsymbol{A}_2\boldsymbol{B}_2}, \dots, \widehat{\boldsymbol{A}_N\boldsymbol{B}_N}$ be finitely many essentially disjoint arcs contained in τ. Let φ be the closed curve obtained from τ by replacing each arc $\widehat{\boldsymbol{A}_i\boldsymbol{B}_i}$ with the segment $\boldsymbol{A}_i\boldsymbol{B}_i$. We will say that φ is an *ε-linearization of* τ if φ is injective and every arc $\widehat{\boldsymbol{A}_i\boldsymbol{B}_i}$ has length at most ε and intersects φ only on the segment $\boldsymbol{A}_i\boldsymbol{B}_i$ (but not necessarily only at $\boldsymbol{A}_i$ and $\boldsymbol{B}_i$). The ε-linearization is said *complete* if the union of the arcs $\widehat{\boldsymbol{A}_i\boldsymbol{B}_i}$ is the whole τ, hence φ is piecewise linear.

LEMMA 4.2. *Let φ be an ε-linearization of some Jordan curve τ of finite length. Then, for every $i, j \in \{1, 2, \dots, N\}$ one has $d_{\mathcal{P}(\varphi)}(\boldsymbol{A}_i, \boldsymbol{B}_j) \le d_{\mathcal{P}(\tau)}(\boldsymbol{A}_i, \boldsymbol{B}_j)$.*

PROOF. If a point $\boldsymbol{D}$ belongs to $\tau \setminus \mathcal{P}(\varphi)$, then it must be contained in some arc $\widehat{\boldsymbol{A}_i\boldsymbol{B}_i}$. Let then $\widehat{\boldsymbol{PQ}}$ be the shortest arc of τ containing $\boldsymbol{D}$ and such that $\boldsymbol{P}$ and $\boldsymbol{Q}$ belong to the segment $\boldsymbol{A}_i\boldsymbol{B}_i$. The curve $\widehat{\boldsymbol{PQ}} \cup \boldsymbol{PQ}$ is then a Jordan curve, whose internal part we denote by $\mathcal{Z}$. Since by construction $\partial\mathcal{Z}$ intersect φ only in the segment $\boldsymbol{PQ}$, then the whole $\varphi \setminus \boldsymbol{PQ}$ is entirely contained either outside $\mathcal{Z}$ or inside $\mathcal{Z}$, and the second possibility is excluded by the fact that $\mathcal{P}(\tau)$ has diameter much larger than ε.

Let now $\mathcal{Z}$ and $\mathcal{Z}'$ be two different zones, corresponding to the points $\boldsymbol{P}$ and $\boldsymbol{Q}$, and $\boldsymbol{P}'$ and $\boldsymbol{Q}'$ respectively. Since by construction the open arcs $\widehat{\boldsymbol{PQ}}$ and $\widehat{\boldsymbol{P}'\boldsymbol{Q}'}$ are disjoint, and they are both outside $\mathcal{P}(\varphi)$, then the zones $\mathcal{Z}$ and $\mathcal{Z}'$ are either disjoint or contained one into the other. As a consequence, the union of the zones (which are at most countably many, by construction) can be written as a disjoint union of zones $\mathcal{Z}_\alpha$, $\alpha \in \mathbb{N}$, each one corresponding to the points $\boldsymbol{P}_\alpha$ and $\boldsymbol{Q}_\alpha$, removing those zones which are contained in some bigger one.

Let us now take any point $\boldsymbol{S} \in \mathcal{P}(\tau) \setminus \mathcal{P}(\varphi)$. By construction, there exists some point $\boldsymbol{D} \in \tau \setminus \mathcal{P}(\varphi)$ such that the open segment $\boldsymbol{SD}$ does not intersect neither τ nor φ; as a consequence, $\boldsymbol{S}$ is contained in the closure of the zone corresponding to the point $\boldsymbol{D}$, hence we deduce that

$$\mathcal{P}(\tau) \setminus \mathcal{P}(\varphi) \subseteq \bigcup_{\alpha \in \mathbb{N}} \overline{\mathcal{Z}_\alpha}. \tag{4.1}$$

Let us now take a geodesic $\gamma : [0, 1] \to \mathbb{R}^2$ between $\boldsymbol{A}_i$ and $\boldsymbol{B}_j$ inside $\mathcal{P}(\tau)$. If $\gamma^{-1}(\mathcal{Z}_1)$ is not empty, let s_α and t_α be respectively its infimum and supremum, and let $\gamma_1 : [0, 1] \to \mathbb{R}^2$ be the continuous curve, linear in $[s_\alpha, t_\alpha]$ and coinciding with γ outside of (s_α, t_α). If $\gamma^{-1}(\mathcal{Z}_1) = \emptyset$, we simply set $\gamma_1 = \gamma$. Of course γ_1 is shorter than γ, and by construction its image is contained in $\mathcal{P}(\varphi) \cup (\mathcal{P}(\tau) \setminus \overline{\mathcal{Z}_1})$. In the very same way, starting from γ_1 we build a shorter curve γ_2 contained in $\mathcal{P}(\varphi) \cup (\mathcal{P}(\tau) \setminus (\overline{\mathcal{Z}_1} \cup \overline{\mathcal{Z}_2}))$. Continuing with the obvious recursion, we end up with a curve $\tilde{\gamma}$, shorter than γ, and contained in $\mathcal{P}(\varphi) \cup (\mathcal{P}(\tau) \setminus (\bigcup_{\alpha \in \mathbb{N}} \overline{\mathcal{Z}_\alpha})) \subseteq \mathcal{P}(\varphi)$ by (4.1). Therefore,

$$d_{\mathcal{P}(\varphi)}(\boldsymbol{A}_i, \boldsymbol{B}_j) \le \mathscr{H}^1(\tilde{\gamma}) \le \mathscr{H}^1(\gamma) = d_{\mathcal{P}(\tau)}(\boldsymbol{A}_i, \boldsymbol{B}_j),$$

concluding the proof. □

The following corollary is trivial.

Corollary 4.3. *Let $\mathcal{R} = [a^-, a^+] \times [b^-, b^+]$ be a rectangle, let $\tau : \partial\mathcal{R} \to \mathbb{R}^2$ be a parametrized Jordan curve with finite length, and let $\varphi : \partial\mathcal{R} \to \mathbb{R}^2$ be an ε-linearization of τ such that $\varphi(A_i) = \varphi(\tau^{-1}(\boldsymbol{A}_i)) = \boldsymbol{A}_i$ and $\varphi(B_i) = \varphi(\tau^{-1}(\boldsymbol{B}_i)) = \boldsymbol{B}_i$ for every $1 \le i \le N$. Then, for every $S, T \in \partial\mathcal{R}$ one has*

$$d_{\mathcal{P}(\varphi)}(\varphi(S), \varphi(T)) \le d_{\mathcal{P}(\tau)}(\tau(S), \tau(T)) + 2\varepsilon.$$

We are now ready to prove Theorem 1.4.

Proof (of Theorem 1.4). By the result of [8] we can limit ourselves to look for piecewise affine homeomorphisms. Let us then fix any function $\delta \mapsto \eta(\delta)$ as in Definition 1.2, and any homeomorphism $u \in \mathrm{BV}(\Omega; \mathbb{R}^2)$. We divide the proof in two steps; first, we consider the strict convergence with respect to the Manhattan norm $\|\cdot\|$ in $\mathbb{R}^2$, and then with respect to the standard norm.

Step I. Strict convergence with respect to the Manhattan norm.

Let $j \in \mathbb{N}$ be any number. We can write Ω as a countable, but locally finite, union of rectangles $\mathcal{R}_i$, $i \in \mathbb{N}$, all with diameter smaller than $1/j$, in such a way that the restriction of u to any side of any rectangle has finite total variation. Since u is continuous, up to take the rectangles small enough we can assume that

$$\operatorname{diam}(u(\mathcal{R}_i)) < K_i := \min\left\{ \eta(\operatorname{dist}(\mathcal{R}_i, \mathbb{R}^2 \setminus \Omega)), \frac{1}{j \max\{1 + |x|^3, x \in \mathcal{R}_i\}} \right\} \tag{4.2}$$

for every $i \in \mathbb{N}$. Let us call $\mathcal{G}$ the union of the boundaries of the rectangles $\mathcal{R}_i$, and let $\tau : \mathcal{G} \to \mathbb{R}^2$ be the restriction of u to the grid $\mathcal{G}$. It is possible to define an injective function $\varphi : \mathcal{G} \to \mathbb{R}^2$ such that, for every rectangle $\mathcal{R}_i$, the restriction φ_i of φ to $\partial\mathcal{R}_i$ is a complete ε_i-linearization of the restriction τ_i of τ to $\partial\mathcal{R}_i$, where $\varepsilon_i < K_i - \operatorname{diam}(u(\mathcal{R}_i))$ is an arbitrary number much smaller than the diameter of $u(\mathcal{R}_i)$: a proof of an even more general fact can be found for instance in [5, Proposition 4.17]. In particular, we can choose

$$\varepsilon_i < \frac{1}{j 2^{i+1} \mathscr{H}^1(\partial\mathcal{R}_i)}$$

such that also

$$(4.3) \qquad \forall Q \in u(\mathcal{R}_i),\ \forall P \in u(\Omega), \quad |P - Q| < \varepsilon_i \Rightarrow |u^{-1}(P) - u^{-1}(Q)| < \frac{1}{j}.$$

By Definition 1.1 and Corollary 4.3, we get then

$$\Psi(\varphi_i) < \Psi(\tau_i) + \varepsilon_i \mathscr{H}^1(\partial\mathcal{R}_i) < \Psi(\tau_i) + \frac{1}{j2^{i+1}}.$$

Since the function φ_i is injective and piecewise linear, Proposition 2.2 provides us with some finitely piecewise affine extension $v_i : \mathcal{R}_i \to \mathbb{R}^2$ such that

$$(4.4) \qquad \int_{\mathcal{R}_i} \|Dv_i\| \le \Psi(\varphi_i) + \frac{1}{j2^{i+1}} \le \Psi(\tau_i) + \frac{1}{j2^i}.$$

Let us now call $u_j : \Omega \to \mathbb{R}^2$ the function coinciding with v_i on each $\mathcal{R}_i$. By construction, u_j is a piecewise affine function on Ω and for every $i \in \mathbb{N}$ one has $\|u_j - u\|_{L^\infty(\mathcal{R}_i)} \le \operatorname{diam}(u(\mathcal{R}_i)) + \varepsilon_i < K_i$. Hence, by (4.2) we immediately get on one hand that u_j uniformly coincides with u at $\partial\Omega$, and on the other hand that $\|u - u_j\|_{L^1(\Omega)} < 4\pi/j$. In addition, by (4.4) and Lemma 2.1 we get

$$\int_\Omega \|Du_j\| = \sum_{i\in\mathbb{N}} \int_{\mathcal{R}_i} \|Dv_i\| \le \frac{1}{j} + \sum_{i\in\mathbb{N}} \Psi(\tau_i) \le \frac{1}{j} + \sum_{i\in\mathbb{N}} \|Du\|(\mathcal{R}_i) = \|Du\|(\Omega) + \frac{1}{j}.$$

Repeating this construction for every $j \in \mathbb{N}$, we get a sequence $\{u_j\}$ of piecewise affine functions, each one uniformly coinciding with u at $\partial\Omega$, which is converging to u in $L^1(\Omega)$, and such that $\limsup \|Du_j\|(\Omega) \le \|Du\|(\Omega)$. Then, the sequence $\{u_j\}$ is converging to u in the strict BV sense, with respect to the Manhattan norm $\|\cdot\|$.

Notice that, since on each $\mathcal{R}_i$ one has $\|u_j - u\|_{L^\infty(\mathcal{R}_i)} < K_i < 1/j$, then $\|u_j - u\|_{L^\infty(\Omega)} < 1/j$, so u_j is converging to u also uniformly. Moreover, take any point $P \in u(\Omega)$: there exists some $i \in \mathbb{N}$ such that $P \in u_j(\mathcal{R}_i)$, hence by construction there is some $Q \in u(\mathcal{R}_i)$ such that $|Q - P| < \varepsilon_i$. As a consequence, since the diameter of $\mathcal{R}_i$ is less than $1/j$, by (4.3) we have

$$|u^{-1}(P) - u_j^{-1}(P)| \le |u^{-1}(Q) - u_j^{-1}(P)| + |u^{-1}(Q) - u^{-1}(P)| < \frac{2}{j},$$

hence also the uniform convergence of u_j^{-1} to u^{-1}.

<u>Step II.</u> *Strict convergence with respect to the standard norm.*

We now consider the strict convergence with respect to the standard norm. In fact, notice that the Manhattan norm is equivalent to the standard norm, but the corresponding strict convergences are not equivalent (while so are the corresponding strong convergences, as well as the corresponding weak* convergences).

Let us decompose $Du = \nu|Du|$, the function $\nu : \Omega \to \mathbb{S}^1$ being defined $|Du|$-a.e., and let us fix an integer $j \in \mathbb{N}$. Then, for every $1 \leq \alpha \leq j$ we define the set

$$\Omega_\alpha = \left\{ x \in \Omega : \nu(x) \in \left[(\alpha - 1)\frac{2\pi}{j}, \alpha\frac{2\pi}{j} \right) \right\},$$

with the usual identification of $\mathbb{S}^1$ with $[0, 2\pi)$. Notice that the sets Ω_α are disjoint, and they cover Ω up to $|Du|$-negligible sets. Let then $K_\alpha \subseteq \Omega_\alpha$ be compact sets satisfying

$$|Du|(\Omega_\alpha \backslash K_\alpha) \leq \frac{|Du|(\Omega_\alpha)}{j}. \tag{4.5}$$

Since these are finitely many disjoint compact sets, there exists some $\varepsilon > 0$ such that the distance between any two of these sets, and between any of these sets and $\mathbb{R}^2 \backslash \Omega$, is much larger than ε.

Let us now concentrate ourselves on a given α, and let us consider rectangles with two sides parallel to the direction $(\alpha - 1)\frac{2\pi}{j}$, which we will call "$\alpha$-rotated rectangles"; notice that, up to now, we have always only considered rectangles with two horizontal and two vertical sides, which corresponds to the case $\alpha = 1$. We can clearly cover the compact set K_α with finitely many essentially disjoint α-rotated rectangles $\mathcal{R}_i^\alpha$, $1 \leq i \leq N(\alpha)$, having sides smaller than ε, in such a way that the restriction of u to any side of any rectangle has finite total variation; up to take these rectangles small enough, we can assume also that

$$|Du|\left(\bigcup_{i=1}^{N(\alpha)} \mathcal{R}_i^\alpha \backslash \Omega_\alpha \right) \leq \frac{|Du|(\Omega_\alpha)}{j^2}. \tag{4.6}$$

Up to renumbering, we can find $0 \leq N^-(\alpha) \leq N(\alpha)$ such that

$$|Du|(\mathcal{R}_i^\alpha \backslash \Omega_\alpha) \leq \frac{|Du|(\mathcal{R}_i^\alpha)}{j} \quad \Leftrightarrow \quad i \leq N^-(\alpha). \tag{4.7}$$

As a consequence, by (4.6) we get

$$\begin{aligned}
|Du|(\Omega_\alpha) &\geq j^2 |Du|\left(\bigcup_{i=1}^{N(\alpha)} \mathcal{R}_i^\alpha \backslash \Omega_\alpha \right) \geq j^2 |Du|\left(\bigcup_{i > N^-(\alpha)} \mathcal{R}_i^\alpha \backslash \Omega_\alpha \right) \\
&= j^2 \sum_{i > N^-(\alpha)} |Du|(\mathcal{R}_i^\alpha \backslash \Omega_\alpha) \\
&\geq j \sum_{i > N^-(\alpha)} |Du|(\mathcal{R}_i^\alpha) = j|Du|\left(\bigcup_{i > N^-(\alpha)} \mathcal{R}_i^\alpha \right),
\end{aligned}$$

thus from the fact that all the $\mathcal{R}_i^\alpha$ cover K_α and (4.5) we deduce

$$|Du|\Big(\bigcup_{i\le N^-(\alpha)} \mathcal{R}_i^\alpha\Big) \ge |Du|(K_\alpha) - |Du|\Big(\bigcup_{i>N^-(\alpha)} \mathcal{R}_i^\alpha\Big) \ge \Big(1-\frac{2}{j}\Big)|Du|(\Omega_\alpha).$$

Let us then call $V \subset\subset \Omega$ the union of all the rectangles $\mathcal{R}_i^\alpha$ with $1 \le i \le N^-(\alpha)$ and $1 \le \alpha \le j$, so that adding over $1 \le \alpha \le j$ the last estimate gives

$$|Du|(V) \ge \Big(1-\frac{2}{j}\Big)|Du|(\Omega). \tag{4.8}$$

Notice now that V is done by finitely many essentially disjoint rectangles, in fact each $\mathcal{R}_i^\alpha$ and each $\mathcal{R}_l^\beta$ with $\beta \ne \alpha$ have strictly positive distance by construction. We can cover $\Omega\setminus V$ with countably many essentially disjoint quadrilaterals $\widetilde{\mathcal{R}}_m$, $m \in \mathbb{N}$, in such a way that each $\widetilde{\mathcal{R}}_m$ can be transformed into a rectangle with a $(1+1/j)$-biLipschitz finitely piecesise affine homeomorphism, and also in such a way that the quadrilaterals $\widetilde{\mathcal{R}}_m$ are locally finitely many in Ω; this means that for every $\delta > 0$ there are only finitely many of these rectangles which have distance from $\mathbb{R}^2\setminus\Omega$ larger than δ; in particular, only finitely many of these rectangles are close to the rectangles $\mathcal{R}_i^\alpha$ in V. The existence of this covering comes through a simple geometrical argument; in particular, the sides of the quadrilaterals $\widetilde{\mathcal{R}}_m$ are generally much smaller than those of the rectangles $\mathcal{R}_i^\alpha$ in V.

Let us now call $\mathcal{G}$ the grid made by the union of all the sides of the quadrilaterals $\widetilde{\mathcal{R}}_m$ and of the rectangles $\mathcal{R}_i^\alpha$ with $1 \le \alpha \le j$, $i \le N^-(\alpha)$. As in Step I, we let $\tau : \mathcal{G} \to \mathbb{R}^2$ be the restriction of u to $\mathcal{G}$, and we find an injective function $\varphi : \mathcal{G} \to \mathbb{R}^2$ which is a complete ε_i^α-linearization (resp., a complete ε_m-linearization) of the restriction of τ to $\partial\mathcal{R}_i^\alpha$ (resp., to $\partial\widetilde{\mathcal{R}}_m$) for every $1 \le \alpha \le j$, $1 \le i \le N^-(\alpha)$ (resp., for every $m \in \mathbb{N}$). We will now consider separately each rectangle or quadrilateral.

Let us start with a quadrilateral $\widetilde{\mathcal{R}}_m$: by construction, up to the $(1+1/j)$-biLipschitz finitely piecewise affine homeomorphism Φ_m this corresponds to a rectangle, call it $\mathcal{R}_m = \Phi_m(\widetilde{\mathcal{R}}_m)$, and $\varphi_m = \varphi \circ \Phi_m^{-1} : \partial\mathcal{R}_m \to \mathbb{R}^2$ is injective and piecewise linear. We can of course assume that $\mathcal{R}_m$ has horizontal and vertical sides, hence we can directly apply Proposition 2.2 to $\mathcal{R}_m$ to find a finitely piecewise affine homeomorphism $v_m : \mathcal{R}_m \to \mathbb{R}^2$, coinciding with φ_m on $\partial\mathcal{R}_m$, and satisfying, also thanks to Lemma 2.1, to Corollary 4.3, and up to have chosen ε_m small enough,

$$\begin{aligned}\int_{\mathcal{R}_m} \|Dv_m\| &\le \Psi(\varphi_m) + \frac{1}{2^{m+2}j} \le \Psi(\tau \circ \Phi^{-1}_{m|_{\partial\mathcal{R}_m}}) + \frac{1}{2^{m+1}j}\\ &\le \|D(u\circ\Phi_m^{-1})\|(\mathcal{R}_m) + \frac{1}{2^{m+1}j}.\end{aligned}$$

Now, keep in mind that the Manhattan norm is not invariant for rotations, hence we cannot say that $\|Du\|(\widetilde{\mathcal{R}}_m)$ is close to $\|D(u\circ\Phi_m^{-1})\|(\mathcal{R}_m)$, even if the bi-

Lipschitz constant of Φ_m is very close to 1. Nevertheless, since $|v| \leq \|v\| \leq \sqrt{2}|v|$ for every $v \in \mathbb{R}^2$, calling $\tilde{v}_m = v_m \circ \Phi_m$ we have

$$\begin{aligned}(4.9)\qquad \int_{\widetilde{\mathcal{R}}_m} |D\tilde{v}_m| &\leq \Big(1+\frac{1}{j}\Big)\int_{\mathcal{R}_m} |Dv_m| \leq \Big(1+\frac{1}{j}\Big)\int_{\mathcal{R}_m} \|Dv_m\| \\ &\leq \Big(1+\frac{1}{j}\Big)\Big(\|D(u\circ\Phi_m^{-1})\|(\mathcal{R}_m) + \frac{1}{2^{m+1}j}\Big) \\ &\leq \Big(1+\frac{1}{j}\Big)\Big(\sqrt{2}|D(u\circ\Phi_m^{-1})|(\mathcal{R}_m) + \frac{1}{2^{m+1}j}\Big) \\ &\leq \Big(1+\frac{1}{j}\Big)\Big(\sqrt{2}\frac{j+1}{j}|Du|(\widetilde{\mathcal{R}}_m) + \frac{1}{2^{m+1}j}\Big) \leq 6|Du|(\widetilde{\mathcal{R}}_m) + \frac{1}{2^m j}.\end{aligned}$$

Let us now consider a rectangle $\mathcal{R}_i^\alpha$, with $1 \leq i \leq N^-(\alpha)$. We denote by $\|\cdot\|_\alpha$ the Manhattan norm rotated of an angle $(\alpha-1)\frac{2\pi}{j}$, so the usual Manhattan norm is $\|\cdot\| = \|\cdot\|_1$. Notice that $\|v\| = |v|$ holds true if $v = (\alpha-1)\frac{2\pi}{j} \in \mathbb{S}^1$, and more in general by definition of Ω_α we have that

$$(4.10)\qquad \|v(x)\|_\alpha \leq \Big(1+\frac{2\pi}{j}\Big)|v(x)| \quad \forall x \in \Omega_\alpha.$$

We can then apply the "α-rotated version" of Proposition 2.2 to the restriction of φ to the α-rotated rectangle $\mathcal{R}_i^\alpha$: keeping in mind Corollary 4.3, and up to have chosen a sufficiently small constant ε_i^α, we find then a finitely piecewise affine function $v_i^\alpha : \mathcal{R}_i^\alpha \to \mathbb{R}^2$, coinciding with φ on the boundary and such that

$$(4.11)\qquad \int_{\mathcal{R}_i^\alpha} |Dv_i^\alpha| \leq \int_{\mathcal{R}_i^\alpha} \|Dv_i^\alpha\|_\alpha \leq \|Du\|_\alpha(\mathcal{R}_i^\alpha) + \frac{1}{j^2 2^i}.$$

Now, keeping in mind (4.10) and (4.7), we can evaluate

$$\begin{aligned}\|Du\|_\alpha(\mathcal{R}_i^\alpha) &= \|Du\|_\alpha(\mathcal{R}_i^\alpha \cap \Omega_\alpha) + \|Du\|_\alpha(\mathcal{R}_i^\alpha \setminus \Omega_\alpha) \\ &\leq \Big(1+\frac{2\pi}{j}\Big)|Du|(\mathcal{R}_i^\alpha \cap \Omega_\alpha) + \sqrt{2}|Du|(\mathcal{R}_i^\alpha \setminus \Omega_\alpha) \\ &\leq \Big(1+\frac{2\pi+\sqrt{2}}{j}\Big)|Du|(\mathcal{R}_i^\alpha),\end{aligned}$$

hence from (4.11) we obtain

$$(4.12)\qquad \int_{\mathcal{R}_i^\alpha} |Dv_i^\alpha| \leq \Big(1+\frac{2\pi+\sqrt{2}}{j}\Big)|Du|(\mathcal{R}_i^\alpha) + \frac{1}{j^2 2^i}.$$

We can finally define the function $u_j : \Omega \to \mathbb{R}^2$, as the function which coincides with v_i^α on every rectangle $\mathcal{R}_i^\alpha$, and with $\tilde{v}_m$ on every quadrilateral $\widetilde{\mathcal{R}}_m$. It is obvious by construction that u_j is a piecewise affine function on Ω, and arguing exactly as in Step I we also have that, up to have chosen sufficiently small rectangles, u_j uniformly coincides with u on $\partial\Omega$ and the sequence $\{u_j\}$ converges to u in $L^1(\Omega)$. We have then only to check the strict BV convergence of u_j to u. Adding the estimates (4.12) over $1 \leq i \leq \alpha$ we get

$$\int_{\bigcup_{i=1}^{N^-(\alpha)} \mathcal{R}_i^\alpha} |Du_j| \leq \Big(1 + \frac{2\pi + \sqrt{2}}{j}\Big)|Du|\Big(\bigcup_{i=1}^{N^-(\alpha)} \mathcal{R}_i^\alpha\Big) + \frac{1}{j^2},$$

and adding over $1 \leq \alpha \leq j$ we have

$$\int_V |Du_j| \leq \Big(1 + \frac{2\pi + \sqrt{2}}{j}\Big)|Du|(V) + \frac{1}{j} \leq \Big(1 + \frac{2\pi + \sqrt{2}}{j}\Big)|Du|(\Omega) + \frac{1}{j}.$$

Instead, adding (4.9) over all $m \in \mathbb{N}$, and recalling (4.8), we obtain

$$\int_{\Omega\setminus V} |Du_j| \leq 6|Du|(\Omega\setminus V) + \frac{1}{j} \leq \frac{12}{j}|Du|(\Omega) + \frac{1}{j},$$

so altogether we have

$$\int_\Omega |Du_j| \leq \Big(1 + \frac{2\pi + \sqrt{2} + 12}{j}\Big)|Du|(\Omega) + \frac{2}{j},$$

and the strict convergence of u_j to u is then proved. □

Theorem 1.3 now immediately follows.

Proof (of Theorem 1.3). Let $u \in \mathrm{BV}(\Omega) \cap \mathrm{Ext}(\varphi)$ be any homeomorphism. By Theorem 1.4, for every $\varepsilon > 0$ we get a piecewise linear homeomorphisms (or a diffeomorphism) $v \in \mathrm{Ext}(\varphi)$ such that $\int_\Omega |Dv| < |Du|(\Omega) + \varepsilon$ and $\|v - u\|_{L^1(\Omega)} < \varepsilon$. In particular, $v \in W^{1,1}(\Omega)$, so we deduce the thesis. □

Remark 4.4. It is standard to strengthen the claim of Theorem 1.4 as follows. Assume that there exists a piecewise linear Jordan curve $\Gamma \subseteq \partial\Omega$ with positive distance from the set $\partial\Omega\setminus P$, such that u is continuous up to Γ and piecewise linear there. Then, the sequence $\{u_j\}$ can be taken in such a way that every u_j is finitely piecewise affine in a neighborhood of Γ. In particular, if Ω is a polygon and u is piecewise linear on $\partial\Omega$, then each function u_j is finitely piecewise affine. Analogously, in Theorem 1.3, if φ is piecewise linear on some piecewise linear Jordan curve $\Gamma \subseteq \partial\Omega$ with positive distance from $\partial\Omega\setminus\Gamma$, then the infimum of the energy remains the same also if one considers only piecewise affine functions $u \in \mathrm{Ext}(\varphi)$ which are finitely piecewise affine on a neighborhood of Γ. To prove the stronger claim of Theorem 1.4 (from which the stronger claim of Theorem 1.3 trivially follows as above), only a slight modification of our proof is needed. Namely, at the beginning of Step I, instead of decomposing Ω as a countable but locally finite union of rectangles $\mathcal{R}_i$, we have to make a covering with a countable but locally finite union of quadrilaterals $\mathcal{R}_i$, in such a way that those which are not rectangles are only finitely many, they are all uniformly bi-Lipschitz copies of rectangles, they all have exactly one side on Γ, on which u is linear, and these sides cover the whole Γ. The very same modification has to be done with the quadrilaterals $\widetilde{R}_m$ covering $\Omega \setminus V$ in Step II.

References

[1] L. Ambrosio - N. Fusco - D. Pallara, *Functions of Bounded Variation and Free Discontinuity Problems*, Oxford University Press (2000).

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

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

[4] S. Hencl - P. Koskela - J. Onninen, *Homeomorphisms of Bounded Variation*, Arch. Ration. Mech. Anal., 186 (2007), no. 3, 351–360.

[5] S. Hencl - A. Pratelli, *Diffeomorphic Approximation of $W^{1,1}$ Planar Sobolev Homeomorphisms*, J. Eur. Math. Soc., 20 (2018), no. 3, 597–656.

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

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

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

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

[10] A. Pratelli - E. Radici, *Approximation of planar* BV *homeomorphisms by diffeomorphisms*, preprint (2017).

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

Received 20 October 2017,
and in revised form 8 March 2018.

Aldo Pratelli
Department of Mathematics
University of Erlangen
Cauerstrasse 11
90158 Erlangen, Germany
pratelli@math.fau.de

Emanuela Radici
Department of Information Engineering
Computer Science and Mathematics
University of L'Aquila
Via Vetoio 1
67100 Coppito (AQ), Italy
emanuela.radici@univaq.it