Aalto University

Hencl, Stanislav; Koski, Aleksis

Extension of planar Hölder homeomorphisms

RESEARCH ARTICLE

Extension of planar Hölder homeomorphisms

Stanislav Hencl[1] | **Aleksis Koski[2]**

[1]Department of Mathematical Analysis,
Charles University, Prague, Czech
Republic

[2]Department of Mathematics and Systems
Analysis, Aalto University, Espoo,
Finland

Correspondence
Aleksis Koski, Department of
Mathematics and Systems Analysis, P.O.
Box 11100, FI-00076 Aalto, Espoo, Finland.
Email: aleksis.koski@aalto.fi

Abstract

Let $\alpha \in (0,1)$. We show that any α-Hölder homeomorphism from the unit circle in the plane to the plane can be extended to an α-Hölder homeomorphism from the whole unit disc.

MSC 2020
26B35, 46E35 (primary), 30C35, 30L05 (secondary)

1 | INTRODUCTION

The well-known Kirszbraun extension theorem tells us that we can extend a Lipschitz mapping from any subset of $\mathbb{R}^n$ into $\mathbb{R}^n$ into a Lipschitz mapping of the whole space. However, things get more difficult once we ask our mapping to be injective as well. Let us denote by $\mathbb{D}$ the unit disc in $\mathbb{R}^2$ and by $\partial\mathbb{D}$ the unit circle. The classical Schoenflies theorem [13, Theorem 10.4] tells us that any continuous injective mapping $\varphi : \partial\mathbb{D} \to \mathbb{R}^2$ can be extended to a homeomorphism from $\mathbb{R}^2$ onto $\mathbb{R}^2$.

The assumption about injectivity is crucial in many applications as it corresponds, for example, to the "noninterpenetration of the matter" in models of nonlinear elasticity, see, for example, Ball [2]. Similarly, it appears naturally in geometric function theory once we want our extension to be quasi-conformal as done by Ahlfors [1] or Tukia and Väisälä [17]. Requiring the extension to be both injective and Lipschitz complicates things even further. Such results were first obtained by Tukia in [15] and [16].

Recently, these problems became important in the solution of the so-called Ball–Evans approximation problem, that is, given a Sobolev $W^{1,p}$ homeomorphism the question is whether there

exists a sequence of diffeomorphisms (or piecewise-linear homeomorphisms) that approximate it in the Sobolev norm. This problem was solved in the planar case for $1 < p < \infty$ by Kovalev, Iwaniec, and Onninen in [8] and the remaining case $p = 1$ by Hencl and Pratelli in [6], see also Mora-Corral [14] and Bellido and Mora-Corall [3] for some initial results in this direction and Daneri and Pratelli [4] for some results about also approximating the inverse mapping. In many of these results, the main crucial step is to establish some extension theorem from a given homeomorphism $\varphi : \partial \mathbb{D} \to \mathbb{R}^2$ to the homeomorphism $h : \mathbb{D} \to \mathbb{R}^2$ (or from $\varphi : \mathbb{R}^2 \to \partial \mathbb{D}$ to $h : \mathbb{R}^2 \to \mathbb{D}$ in [8]) with some control of $\int |Dh|^p$ or of the Lipschitz constant, see, for example, Hencl and Pratelli [6, Theorem 2.1] or Daneri and Pratelli [5]. The extension of Sobolev homeomorphisms from the boundary attracted a lot of independent attention recently, and we recommend, for example, [7, 9] or [10] and references given there for further reading.

The planar extension of Lipschitz homeomorphisms from the circle has been recently improved by Kovalev in [11] and [12], where he was able to obtain a sharp dependence of both the Lipschitz- and bilipschitz constants of the extension in terms of the constants for the original boundary map.

The main aim of the present paper is to study the extension of α-Hölder homeomorphisms from the unit circle $\partial \mathbb{D}$. A mapping φ is called α-Hölder for $\alpha \in (0, 1)$ if there is $C > 0$ with

$$|\varphi(x) - \varphi(y)| \leqslant C|x - y|^\alpha \text{ for every } x, y.$$

Our main result is the following extension theorem from the unit circle $\partial \mathbb{D}$ in the plane.

Theorem 1.1. *Let $\alpha \in (0, 1)$ and let $\varphi : \partial \mathbb{D} \to \mathbb{R}^2$ be a homeomorphism that is α-Hölder continuous. Then, there is a homeomorphism $H : \overline{\mathbb{D}} \to \mathbb{R}^2$ that is α-Hölder and satisfies $H = \varphi$ on $\partial \mathbb{D}$.*

Let us note that it is not possible to use previously known extension procedures. One of the main reasons is that the Lipschitz-property is local and any locally Lipschitz H is globally Lipschitz. However, a locally Hölder continuous map (on small scales) might not be globally Hölder continuous on bigger scales, and thus, we require new ideas.

In our proof, we cut $\mathbb{D}$ into some dyadic parts and we define our h on those parts separately. We define images of boundaries of those dyadic parts by somehow "copying" the behavior of φ on the boundary nearby. Then, we extend the mapping inside those images by using shortest curves inside with the help of some ideas from [6]. Since everything is dyadic in the domain, we can sum the local estimates using a basic geometric series estimate, and we obtain the desired global Hölder estimate at the end.

2 | EXAMPLES AND TECHNICAL LEMMA

In our extension results, we assume that we map a circle to a Jordan curve in the plane. The following simple example shows that the circle cannot be replaced by some non-Lipschitz domain even for a Lipschitz mapping φ. We use the following simple domains with an outer and inner cusp (see Figure 1):

$$\Omega_1 := (-1, 1)^2 \cup \{[x, y] : \ x \in [1, 2), \ |y| < (2 - x)^2\} \text{ and}$$

$$\Omega_2 := (-1, 1)^2 \setminus \{[x, y] : \ x \in [0, 1], |y| < x^2\}.$$
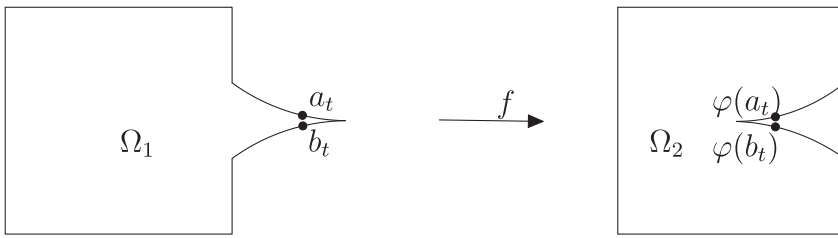
FIGURE 1   Definition of Ω_1, Ω_2 and φ between $\partial\Omega_1$ and $\partial\Omega_2$.

Example 2.1.   There is a Lipschitz homeomorphism $\varphi : \partial\Omega_1 \to \partial\Omega_2$ that does not admit a Lipschitz homeomorphic extension $h : \overline{\Omega_1} \to \overline{\Omega_2}$.

Proof.   We simply define φ as the identity map on $\partial\Omega_1 \cap \partial\Omega_2$, and a simple reflection across the line $x = 1$ from the inner cusp to the outer cusp, that is, we set

$$\varphi(x, y) = \begin{cases} [x, y] & \text{for } [x, y] \in \partial\Omega_1 \cap \partial(-1, 1)^2, \\ [2 - x, y] & \text{for } [x, y] \in \partial\Omega_1 \setminus \partial(-1, 1)^2. \end{cases}$$

This mapping is clearly a homeomorphism from $\partial\Omega_1$ to $\partial\Omega_2$ and it is not difficult to find out that it is Lipschitz.

Given small $t > 0$, we consider two points (see Figure 1)

$$a_t = \left[2 - t, t^2\right] \text{ and } b_t = \left[2 - t, -t^2\right],$$

which are mapped to

$$\varphi(a_t) = \left[t, t^2\right] \text{ and } \varphi(b_t) = \left[t, -t^2\right].$$

The segment $L_t := \{[2 - t, y] : |y| \leqslant t^2\}$ of length $2t^2$ has to be mapped by the extension to some continuum that connects $\varphi(a_t)$ and $\varphi(b_t)$ inside Ω_2. However, the inner distance of $\varphi(a_t)$ and $\varphi(b_t)$ inside Ω_2 is equal to $2\sqrt{t^2 + t^4}$, and hence, the Lipschitz constant of the extension needs to be at least

$$\frac{2\sqrt{t^2 + t^4}}{2t^2} = \frac{\sqrt{1 + t^2}}{t} \text{ and this tends to } \infty \text{ as } t \to 0 + . \qquad \square$$

We next give a simple characterization of curves that admit a Hölder parametrization, classifying all the Jordan curves for which our main theorem is applicable.

Definition 2.2.   Let $p \in [1, \infty)$. A Jordan curve $\Gamma \subset \mathbb{R}^2$ has finite p-content if there is a constant C such that for every finite subdivision $x_1, x_2, \ldots, x_n \in \Gamma$, it holds that

$$\sum_{k=1}^{n-1} |x_{k+1} - x_k|^p \leqslant C.$$

The smallest such C is denoted as $|\Gamma|_p$.

Theorem 2.3. *Let $\alpha \in (0, 1)$. A Jordan curve Γ admits an α-Hölder parametrization $\varphi : [0, 1] \to \Gamma$ if and only if it has finite $\frac{1}{\alpha}$-content.*

Proof. The "only if" part of this statement is trivial, so we will only prove here that a curve with finite $\frac{1}{\alpha}$-content admits such a parametrization.

Let $p = \alpha^{-1}$. Note that any subcurve $\gamma \subset \Gamma$ also has finite p-content, and it is simple to verify that

$$|\Gamma|_p \geqslant |\gamma|_p + |\Gamma \setminus \gamma|_p. \tag{2.1}$$

Hence, given an initial parametrization $\varphi_0 : [0, 1] \to \Gamma$, the map $F(t) = |\varphi_0([0, t])|_p$ defines an increasing surjection from [0,1] to $[0, |\Gamma|_p]$. By setting

$$\varphi(t) = \varphi_0\big(F^{-1}(t)\big),$$

we obtain a parametrization $\varphi : [0, |\Gamma|_p] \to \Gamma$ that satisfies

$$|\varphi([0, t])|_p = t. \tag{2.2}$$

Moreover, by the definition of the p-content $|\varphi([t_1, t_2])|_p$ and combining (2.1) and (2.2), we have

$$|\varphi(t_2) - \varphi(t_1)| \leqslant |\varphi([t_1, t_2])|_p^\alpha \leqslant \big(|\varphi([0, t_2])|_p - |\varphi([0, t_1])|_p\big)^\alpha = (t_2 - t_1)^\alpha.$$

Hence, φ is α-Hölder continuous. The map $t \to \varphi(t|\Gamma|_p)$ also gives such a parametrization on [0,1] instead with Hölder-constant at most $|\Gamma|_p^\alpha$. $\qquad \square$

3 | PROOF OF THE HÖLDER EXTENSION RESULT

Proof of Theorem 1.1. We are hence given an α-Hölder parametrization $\varphi : \partial\mathbb{D} \to \partial\mathbb{Y}$ of the boundary of a Jordan domain $\mathbb{Y}$, and we want to extend it to a Hölder-continuous homeomorphism $H : \mathbb{D} \to \mathbb{Y}$.

Step 1. Suitable decomposition of the domain and target. To construct the extension, we will employ a suitable decomposition of the target and domain sides. We will split the unit disk $\mathbb{D}$ into dyadic regions $U_{n,k}$, each with diameter at most 2^{-n}, and also split the domain side into appropriate regions $V_{n,k}$ depending on the behavior of the boundary map φ. The aim is to define an extension $h : U_{n,k} \to V_{n,k}$ in each region and to show that it is Hölder-continuous with uniform constant C in each region.

At a later stage of the proof, we will be able to use the dyadic structure of the sets $U_{n,k}$ to conclude that the extension h will, in fact, be Hölder-continuous in the whole unit disk. The initial extension h will not be a homeomorphism as it will map some open sets onto curves and map some of the interior of $\partial\mathbb{D}$ onto the boundary $\partial\mathbb{Y}$, but it will define a monotone map (preimage of each point is connected), and we will take care in the construction in order for there to be an injectification process at the end where h may be modified in an arbitrarily small way to give a homeomorphic extension $H : \mathbb{D} \to \mathbb{Y}$ of φ. The fact that this modification is as small as we

wish (on each dyadic level as well) is going to let us argue that H inherits the Hölder-continuity estimates from h.

We start by constructing the dyadic sets $U_{n,k}$. We first split the unit circle $\partial \mathbb{D}$ into dyadic arcs $I_{n,k}$, $n = 1, 2, \dots$ and $k = 1, \dots, 2^n$. The amount of initial arcs or their position does not matter. For the purposes of simplifying the notation here, we may suppose that the arcs $I_{n,k}$ are instead intervals on the real line by flattening out the circle locally. The disk $\mathbb{D}$ may be locally interpreted as the upper half space for this purpose.

We now let $T_{n,k}$ be the isosceles triangle in the upper half space with base $I_{n,k}$ and two equal angles of size $\pi/4 + 100^{-n}$ against the base. The apex vertex of $T_{n,k}$ is denoted as $p_{n,k}$ and the endpoints of $I_{n,k}$ are denoted by $a_{n,k}$ and $b_{n,k}$. If we suppose that $I_{n+1,k'}$ and $I_{n+1,k'+1}$ are the two dyadic children of $I_{n,k}$, then we let $U_{n,k} = T_{n,k} \setminus (I_{n+1,k'} \cup I_{n+1,k'+1})$ denote the region between each successive dyadic parent triangle and its children (see Figure 8 for the idea how $U = U_{n,k}$ looks like).

We now wish to explain how the sets $V_{n,k}$ will be defined. Given the boundary map φ, we let $A_{n,k} = \varphi(a_{n,k})$ and $B_{n,k} = \varphi(b_{n,k})$. Moreover, let $C_{n,k}$ be the image of the middle point of $I_{n,k}$ under φ. We now may need to make a certain adjustment to this construction.

Let us call a point $\omega \in \partial \mathbb{Y}$ a *good point* if there exists a line segment inside $\mathbb{Y}$ with endpoint ω. We wish to assume that the dyadic image points given by $A_{n,k}$ are good points for all n, k. If this is not the case, we argue as follows. For every point $\omega \in \partial \mathbb{Y}$, it is possible to choose a good point $\omega' \in \partial \mathbb{Y}$ arbitrarily close to ω by considering a point $P \in \mathbb{Y}$ very close to ω, considering the largest disk of center P that lies entirely within $\mathbb{Y}$, and picking ω' as an intersection point of $\partial \mathbb{Y}$ with the boundary of this disk. This way of defining ω' guarantees that it is a good point. As P approaches ω, it can be shown that ω' approaches ω as well.

By appropriately using this observation inductively, we may replace each point $A_{n,k}$ with a good point $A'_{n,k}$ so that the preimages of $A'_{n,k}$ under φ form another decomposition of the unit circle into intervals $I'_{n,k}$ that have length uniformly comparable to the length of $I_{n,k}$ for all n, k. If this is the case, it is not difficult to construct a self-map of the unit disk that is bilipschitz and maps each $I_{n,k}$ to $I'_{n,k}$. Since composition by bilipschitz-maps does not affect Hölder-continuity estimates, we may as well suppose that the points $A_{n,k}$ were good points to begin with and not deal with additional notation.

Let $\Gamma_{n,k}$ denote the unique shortest curve from $A_{n,k}$ to $B_{n,k}$ inside $\overline{\mathbb{Y}}$. Moreover, let $\Gamma^+_{n,k}$ and $\Gamma^-_{n,k}$ denote the corresponding curves for the dyadic children in this case. The curves $\Gamma_{n,k}, \Gamma^+_{n,k}$, and $\Gamma^-_{n,k}$ may overlap each other and the boundary $\partial \mathbb{Y}$ but not cross each other. The region $V_{n,k}$ will be defined as a set bounded by these three curves, but let us first fix a parametrization for each of them.

Step 2. Fixing a parametrization. Let us now define a parametrization of the curve $\Gamma_{n,k}$ that will be used later to define the values of the extension h of φ in part of the boundary of $U_{n,k}$. The main goal here is to define this parametrization as a Hölder-continuous map from the interval $I_{n,k}$ to $\Gamma_{n,k}$.

For purposes of easier presentation, let us drop subscripts for a moment and denote $I = I_{n,k} = [a, b]$. Let $\Gamma \in \overline{\mathbb{Y}}$ be the shortest curve between $\varphi(a)$ and $\varphi(b)$. We wish to define a parametrization $\tau : I \to \Gamma$ by "projecting" the parametrization given by φ on the boundary. Given a point $z \in I$, we let $\tau(z)$ be the point on Γ that is closest to $\varphi(z)$ with respect to the internal distance in $\mathbb{Y}$ (see Figure 2). Such a point is determined uniquely since the connected components of Γ that do not touch $\varphi(I)$ must be locally concave toward $\varphi(I)$ due to Γ being a shortest curve (see, e.g., [6, Step 5 of the proof of Theorem 2.1] for the properties of the shortest curves). This defines the
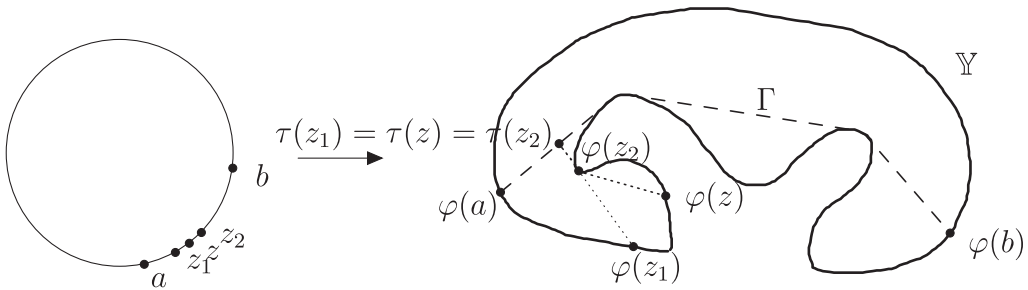
FIGURE 2 Definition of the curve Γ and its parametrization τ. The whole part between z_1 and z_2 is mapped to the same point $\tau(z)$.

parametrization τ as a monotone map from I to Γ, with loss of injectivity being possible since the preimage of a single point on Γ may be an interval in I (e.g., points between z_1 and z_2 in Figure 2 are all mapped to the same points $\tau(z)$). We will now show that τ inherits the Hölder-continuity of φ. The lack of injectivity of τ will not pose a problem later as we can eventually injectify this map with an arbitrarily small change to the Hölder-constant.

Let $x, y \in I$ and let β_{xy} denote the part of $\varphi(I)$ between $\varphi(x)$ and $\varphi(y)$. It will be enough to show that the inequality

$$|\tau(x) - \tau(y)| \leqslant C \operatorname{diam}(\beta_{xy}) \tag{3.1}$$

holds, since the diameter of β_{xy} satisfies $\operatorname{diam}(\beta_{xy}) \leqslant C|x - y|^\alpha$ due to the Hölder-continuity of φ on all subintervals.

We may suppose that the shortest curves γ_x and γ_y that connect $\varphi(x)$ to $\tau(x)$ and $\varphi(y)$ to $\tau(y)$, respectively, within $\overline{\mathbb{Y}}$ are, in fact, straight line segments. This is due to the following fact. Let I_x denote the preimage of $\tau(x)$ under τ, which is either a single point or a closed interval containing x. Since the closest point on Γ with respect to the internal distance from $\varphi(x')$ is $\tau(x)$ for all $x' \in I_x$, by geometry, we see that if x' is one of the endpoints of I_x, such a shortest curve must be a straight line segment between $\varphi(x')$ and $\tau(x)$ (e.g., in Figure 2, our I_z is between z_1 and z_2 and shortest line between $\varphi(z_1)$ and $\tau(z_1)$ or $\varphi(z_2)$ and $\tau(z_2)$ is a segment). Let x' be the endpoint of I_x that is closest to y. Since $\beta_{x'y} \subset \beta_{xy}$, it is enough to prove (3.1) for x' instead of x because $\tau(x') = \tau(x)$. Hence, we suppose that γ_x and γ_y are straight line segments.

Let Γ' denote the part of Γ between $\tau(x)$ and $\tau(y)$. We claim that it is enough to prove (3.1) when the interior of Γ' does not touch the curve $\varphi(I)$. Indeed, if the interior does touch $\varphi(I)$, then it must be at a point on β_{xy} as the part of $\varphi(I)$ not equal to β_{xy} cannot cross γ_x, γ_y or Γ and hence has no access to Γ'. Supposing that (3.1) has been proven when the interior of Γ' does not intersect β_{xy}, we explain now how the general case would follow. For this, note that if $u, v \in \Gamma' \cap \beta_{xy}$, then trivially $|u - v| \leqslant \operatorname{diam}(\beta_{xy})$. Hence, by picking u, v on Γ' as close to the endpoints $\tau(x)$ and $\tau(y)$ as possible (with respect to parametrization), we may split Γ' into three parts: One part from $\tau(x)$ to u, one part from u to v, and one part from v to $\tau(y)$. For clarification, it is possible for any of these parts to be a singleton as well. If $\tau(x)$ and u are not equal, by definition of u, there cannot be a point on β_{xy} intersecting the part of Γ' between $\tau(x)$ and u. But now by assumption, we can apply (3.1) in this part and obtain that $|\tau(x) - u| \leqslant C \operatorname{diam}(\beta_{xy})$. Similarly, $|v - \tau(y)| \leqslant C \operatorname{diam}(\beta_{xy})$, and combining these estimates with the one for $|u - v|$ gives (3.1) while only increasing the constant by a factor 3.
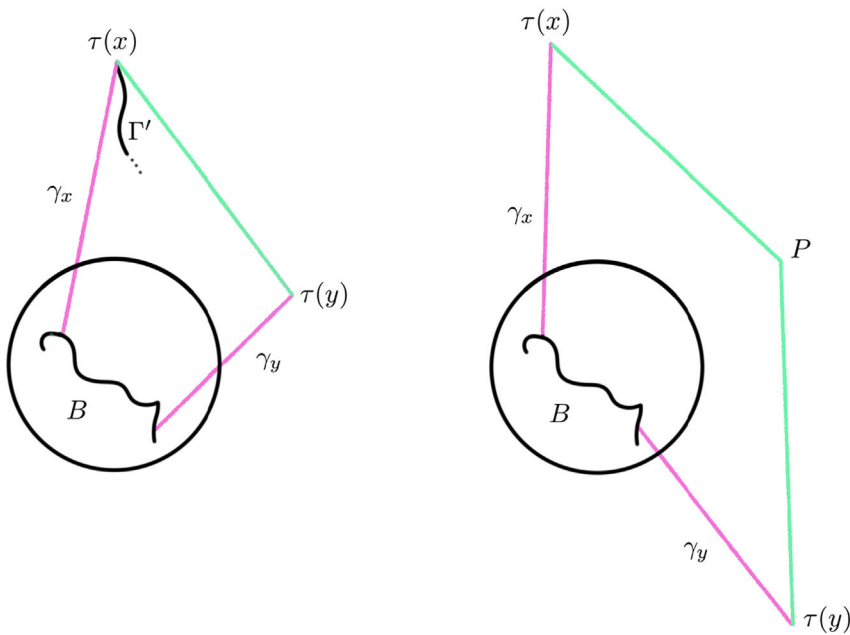
FIGURE 3   Two situations that both lead to a contradiction.

Hence, we may suppose that the interior of Γ' does not intersect β_{xy}. Since Γ' is a shortest curve and does not intersect γ_x or γ_y beside the mutual endpoints, this means that Γ' is locally concave toward the interior of the region Ω bounded by the four curves $\beta_{xy}, \gamma_x, \Gamma'$, and γ_y (see, e.g., [6, Step 5 of the proof of Theorem 2.1]). For the sake of contradiction in proving (3.1), let us suppose that $|\tau(x) - \tau(y)| > 100 \operatorname{diam}(\beta_{xy})$. Now if both $|\tau(x) - \varphi(x)|$ and $|\tau(y) - \varphi(y)|$ are smaller than $40 \operatorname{diam}(\beta_{xy})$, then

$$|\tau(x) - \tau(y)| \leqslant |\tau(x) - \varphi(x)| + |\varphi(x) - \varphi(y)| + |\tau(y) - \varphi(y)| \leqslant 81 \operatorname{diam}(\beta_{xy}),$$

a contradiction. Hence, we may suppose without loss of generality that $|\tau(x) - \varphi(x)| > 40 \operatorname{diam}(\beta_{xy})$.

Let us cover β_{xy} by a ball B of radius $2 \operatorname{diam}(\beta_{xy})$ (exact positioning does not matter here). Hence, the line segment γ_x starts inside of B and ends outside of B. We now note that γ_y must also do the same, as otherwise γ_y would have no intersection point with ∂B and Γ' would have to intersect ∂B somewhere. This would give a contradiction with the definition of $\tau(x)$: To connect $\varphi(x)$ with Γ, it would be shorter to go from x to a point in $\gamma_x \cap \partial B$ and then along ∂B to the point in $\Gamma' \cap \partial B$ and the total length would be shorter than $40 \operatorname{diam}(\beta_{xy})$, giving us a contradiction with the definition of $\tau(x)$ and $|\tau(x) - \varphi(x)| > 40 \operatorname{diam}(\beta_{xy})$.

Consider now a line ℓ picked so that it passes through $\tau(x)$ but the sets γ_x, γ_y, and B all lie on one (closed) side of ℓ (e.g., in Figure 3, our ℓ could be between $\tau(x)$ and $\tau(y)$ on the first picture and between $\tau(x)$ and P on the second picture). We call this side the "good side" just to have a name for it. It is not difficult to see that Γ' also lies inside this good side of ℓ since Γ' is concave toward β_{xy} by the property of minimal curves. Of course, it could happen that Γ' contains both $\tau(x)$ and $\tau(y)$ and Γ' is a straight segment and then it lies inside γ.
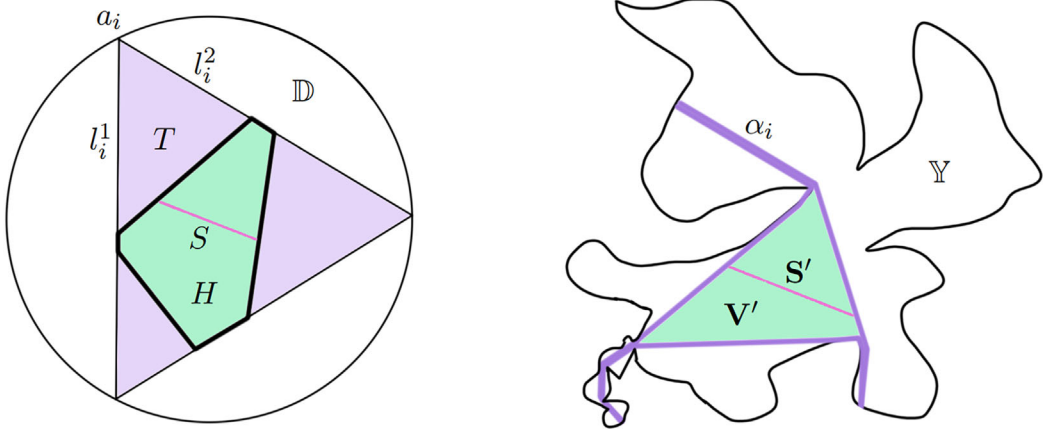
FIGURE 4   The domain and target for the initial case.

Thus, any line that passes through $\tau(x)$ (or $\tau(y)$) and contains the sets γ_x, γ_y, and B on one side of it automatically also contains Γ' on that side, we call this property $(*)$. Now the claim (3.1) will be fairly easy to prove. If the line segment L_{xy} from $\tau(x)$ to $\tau(y)$ does not intersect B, then due to property $(*)$, the curve Γ' must lie within a region bounded by $L_{xy}, \gamma_x, \gamma_y$, and the part of ∂B between γ_x and γ_y. Hence, L_{xy} must form an angle greater or equal to $\pi/2$ with respect to both γ_x and γ_y due to the fact that $\tau(x)$ and $\tau(y)$ are the closest points on Γ' from points on γ_x and γ_y, as otherwise we could find closer points on Γ' (e.g., in the situation of Figure 3 on the left close to $\tau(x)$, we could slightly cut the corner and we would go from γ_x to Γ' in a shorter way). This automatically shows (3.1) as then $|\tau(x) - \tau(y)|$ must be less than the diameter of B. If the line segment from $\tau(x)$ to $\tau(y)$ does intersect B, then we can find another point P so that the segments $\tau(x)P$ and $P\tau(y)$ form acute angles with γ_x and γ_y, respectively, and by property $(*)$, they must contain Γ' on the side with acute angles, see Figure 3. This is a contradiction as then there must be points on Γ' besides $\tau(x)$ that are closer to points on γ_x than $\tau(x)$ is (in the situation of Figure 3 on the right close to $\tau(x)$, we could again slightly cut the corner). This proves the claim and hence the Hölder-continuity of τ.

Step 3. The initial case. We detail here the first step of the construction of the extension h by giving an initial set $T \subset \mathbb{D}$ in which we define the map h on. The rest of the construction will be a repetition of this idea with respect to each dyadic interval.

Let us start the dyadic decomposition of $\partial \mathbb{D}$ by fixing three initial arcs $I_{0,1}, I_{0,2}$, and $I_{0,3}$ of length $2\pi/3$ whose union is the whole circle. Hence, we have an equilateral triangle T in $\mathbb{D}$ with vertices $a_i, i = 1, 2, 3$ given by the endpoints of these three arcs.

On the target side, there are corresponding curves $\Gamma_{0,i}, i = 1, 2, 3$ given by the shortest curves in $\overline{\mathbb{Y}}$ between the image points $\varphi(a_i), i = 1, 2, 3$. These curves may partially overlap, as the two curves $\Gamma_{0,i}$ that share a mutual endpoint $\varphi(a_j)$ may have a common part given by a curve α_j (see, e.g., α_i in Figure 4). In the case where there is no overlap, we simply consider α_j as the singleton $\varphi(a_j)$. However, each curve $\Gamma_{0,i}$ has some part that does not overlap with any of the others, and these three curves together define a Jordan curve that bounds a domain $\mathbf{V}'$. The particular thing to note about $\mathbf{V}'$ is that due to the fact that its boundary consists of three nonintersecting shortest curves, these three curves must be concave toward the interior of $\mathbf{V}'$ (again by [6, Step 5 of the proof of Theorem 2.1]), see Figure 4 and also Figure 6 for possible shapes of $\mathbf{V}'$. We let V denote the union of $\mathbf{V}'$ and the common parts $\alpha_j, j = 1, 2, 3$. Using the projected parametrization τ defined
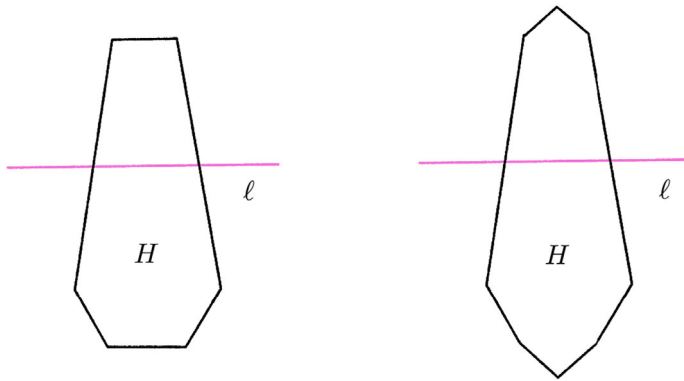
FIGURE 5   Adjusting the hexagon.

in previous step, we may define a boundary map $\tau' : \partial T \to \partial V$ that is monotone on each of the sides of T and is Hölder-continuous with estimates in terms of the original boundary map φ.

We now aim to extend the map τ' to the interior of the triangle T to give a map $h_0 : T \to V$. The preimages of the curves $\alpha_i \subset V, i = 1, 2, 3$ are fairly easy to choose. For each i, we look at the two preimages l_i^1 and l_i^2 of α_i under τ', which are line segments starting from a_i on the two sides of ∂T, see Figure 4. Now for each point $P \in \alpha_i$, we look at the preimages $p_1 \in l_i^1$ and $p_2 \in l_i^2$ under τ' and connect p_1 and p_2 via a line segment L. Then, for each $z \in L$, we simply define that $h_0(z) = P$. This effectively defines the map h_0 on the triangle bounded by l_i^1 and l_i^2. The map is Hölder-continuous as it is simply a linear interpolation of the two Hölder-continuous maps (given by τ') on l_i^1 and l_i^2.

This leaves a convex hexagonal region $H \subset T$ on the interior of which the map h_0 has not been defined yet. The boundary values of h_0 on ∂H are already determined, and three of the sides S_1, S_3, and S_5 of H are mapped to single points under h_0, while the three other sides are mapped to concave curves (the boundary of $\mathbf{V}'$). To define h_0, we will employ a direct version of the shortest curve extension method. For clarity, let us denote by $\varphi_0 = h_0|_{\partial H}$ the already defined boundary values of h_0.

Let us start by fixing a "horizontal" direction given by a line ℓ, and we consider the "vertical" direction to be the one orthogonal to ℓ. There are a few additional requirements we are able to impose. First, we may assume that any two points in H can be connected by a curve consisting of at most five line segments that are either horizontal or vertical. This is due to the fact that H is convex and the practical idea here is that if H is very thin, then we should choose the vertical direction to be parallel to the "direction of thinness" of H.

Moreover, we may suppose that the angle between the horizontal direction ℓ and all of the sides of H is controlled from below by some absolute constant $c > 0$. This is not immediately possible if H is very thin and it has two sides that are close to being parallel to the horizontal direction we need to impose for the previous requirement, see Figure 5. But in this case, we make a small adjustment to these sides to replace them with two line segments that have a large angle toward the horizontal direction. Such an adjustment can be realized as a bilipschitz transformation of H with absolute control on the constant, so it does not affect future estimates, although it does transform the hexagon H into an octagon. But we will later only use the fact that H has a fixed finite amount of sides, so this does not matter.

With these assumptions, H may be written as the union of line segments parallel to ℓ, the first and last of which consist of simply a single vertex of H. We call these segments horizontal
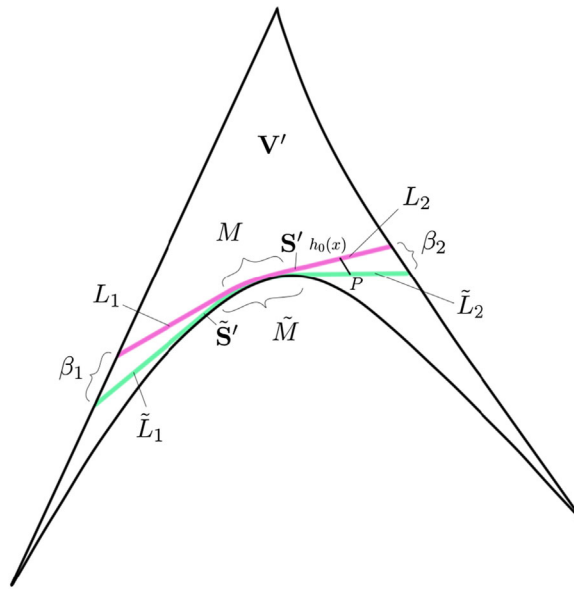
FIGURE 6　Shortest curves $\mathbf{S}'$ and $\tilde{\mathbf{S}}'$ and related objects within the region $\mathbf{V}'$.

segments on H. For each such segment S having endpoints s_1 and s_2, we define the map h_0 on S by defining h_0 to map S onto the shortest curve $\mathbf{S}'$ in $\mathbf{V}'$ between the points $\varphi_0(s_1)$ and $\varphi_0(s_2)$ on $\partial \mathbf{V}'$, parametrized at constant speed, see Figure 6. This defines a monotone map $h_0 : H \to \mathbf{V}'$, but it can be injectified with an arbitrarily small modification. Hence, it remains to verify the Hölder-continuity of such a map h_0. It is enough to verify the required estimates separately for the horizontal and the vertical directions, since we assumed that any two points can be connected by at most five horizontal/vertical segments (the lengths of such segments may also be chosen to be less than the distance of those two points).

First of all, if two points x_1, x_2 belong to a common horizontal segment S, then we may compute due to the choice of constant speed parametrization that

$$|h_0(x_1) - h_0(x_2)| \leqslant |x_1 - x_2| \frac{|\mathbf{S}'|}{|S|}.$$

Note now that due to the fact that $\mathbf{V}'$ had the special structure of being bounded by three concave curves, any shortest curve $\mathbf{S}'$ within $\mathbf{V}'$ must satisfy $|\mathbf{S}'| \leqslant C \operatorname{diam}(\mathbf{S}')$ for some absolute constant C. Moreover, due to the Hölder-estimates for the boundary values φ_0, we obtain that $\operatorname{diam}(\mathbf{S}') \leqslant C|s_1 - s_2|^\alpha$. Combining these gives

$$|h_0(x_1) - h_0(x_2)| \leqslant C|x_1 - x_2| \frac{|s_1 - s_2|^\alpha}{|s_1 - s_2|} \leqslant C|x_1 - x_2|^\alpha.$$

Let now x and $\tilde{x}$ belong to different horizontal segments S and $\tilde{S}$ but be on the same vertical line. Due to the fact that H only has six sides, we may assume that the left and right endpoints of S and $\tilde{S}$ belong to the same sides of H, respectively. We may also assume that

$$|\tilde{S}| \leqslant |S|, \quad |x - \tilde{x}| \leqslant |S|, \tag{3.2}$$

the second part being because if this was not the case, we could connect x to $\tilde{x}$ with two horizontal segments along S and $\tilde{S}$ and one boundary curve between s_1 and $\tilde{S}$, each of whose length is less than a constant times $|x - \tilde{x}|$. In this case, we could conclude the Hölder-continuity estimate simply from the previous estimate for the horizontal direction and the Hölder-continuity of the boundary map.

Now due to the concavity of the boundary curves in $\mathbf{V}'$, the shortest curves $\mathbf{S}' = h_0(S)$ and $\tilde{\mathbf{S}}' = h_0(\tilde{S})$ on the target side in $\mathbf{V}'$ must behave in a specific way. We aim to use this to put the possible shortest curves $\mathbf{S}'$ into a finite amount of different categories in order to be able to assume that $\mathbf{S}'$ and $\tilde{\mathbf{S}}'$ belong to the same category. We do this as follows.

Due to the fact that $\mathbf{V}'$ is bounded by three concave curves, a shortest curve $\mathbf{S}'$ in $\mathbf{V}'$ may only intersect at most one of these concave curves at points different from the endpoints of $\mathbf{S}'$. Thus, such a curve $\mathbf{S}'$ consists of up to two line segments starting from either endpoint of $\mathbf{S}'$ and possibly a middle section that goes along one of the boundary curves of $\mathbf{V}'$. We can categorize each such curve by whether it is just a simple segment, a segment combined with a concave part, and so on. We also put curves in a different category depending on which of the three boundary curves of $\mathbf{V}'$ they intersect with. Finally, we may also do the following: If one of the boundary curves of $\mathbf{V}'$ is "too curved," meaning the angle of its tangent changes by over $\pi/2$ (in fact, at most one of the three curves may be like this as the total angular change of all three is bounded by π), then we split it into two halves on which the angle stays below $\pi/2$. We then also categorize the shortest curves S' into different categories if they start or end on different halves of such a "too curved" boundary part.

As we move the horizontal segment S vertically in the domain, the category of S' can only change a finite amount of times. Note that at the point where a category changes from one to another, there is always one curve $\mathbf{S}'$ that lies at the edge of these two categories and where one or more of its parts (segments or convex part) may be just singletons. We will say that such a curve $\mathbf{S}'$ belongs to both categories to avoid technicalities. In any case, the whole point of putting these curves into categories is that since there are only a finite number of categories and a finite number of ways such categories can appear in order, we may assume that $\mathbf{S}'$ and $\tilde{\mathbf{S}}'$ belong to the same category. Note that the bounds for the finite numbers that appear here are absolute.

Let β_1 denote the part of $\partial\mathbf{V}'$ between $\varphi_0(s_1)$ and $\varphi_0(\tilde{s}_1)$ that does not contain $\varphi_0(s_2)$, and define β_2 analogously, see Figure 6. Note that due to the fact that $\mathbf{S}'$ and $\tilde{\mathbf{S}}'$ belong to the same category and the fact that we made sure not to include the case where, say, s_1 and $\tilde{s}_1$ belong to a "too curved" boundary part, we must actually have the length estimate

$$|\beta_1| \leqslant 2|\varphi_0(s_1) - \varphi_0(\tilde{s}_1)| \tag{3.3}$$

and the same for β_2 as these are convex curves where the angular change is less than $\pi/2$. There are now a few cases to consider depending on which category $\mathbf{S}'$ and $\tilde{\mathbf{S}}'$ belong to, but they are handled very similarly, so we only consider the most complicated one.

Let us thus consider the case where $\mathbf{S}'$ and $\tilde{\mathbf{S}}'$ both consist of a line segment, a middle part, and another line segment. We denote this by $\mathbf{S}' = L_1 \cup M \cup L_2$ and $\tilde{\mathbf{S}}' = \tilde{L}_1 \cup \tilde{M} \cup \tilde{L}_2$. By geometry, one of the inclusions $M \subset \tilde{M}$ or $\tilde{M} \subset M$ must hold. Since we assumed that $|\tilde{S}| \leqslant |S|$ in (3.2), we should treat these cases separately, but they are similar enough that we will only consider the case $M \subset \tilde{M}$. The situation is now as in Figure 6.

Given the point $h_0(x) \in \mathbf{S}'$, we let P denote the point on $\tilde{\mathbf{S}}'$ that is obtained from $h_0(x)$ as follows. If $h_0(x) \in L_1$, we traverse from $h_0(x)$ in the direction of the vector $\varphi(\tilde{s}_1) - \varphi(s_1)$ until we hit a point on $\mathbf{S}'$ which is defined as P. If $h_0(x) \in M$, then $P = h_0(x)$, and if $h_0(x) \in L_2$, we traverse in the

direction of the vector $\varphi(\tilde{s}_2) - \varphi(s_2)$ to hit a point $P \in \mathbf{S}'$. We may suppose that we are in the case $h_0(x) \in L_1$, as the case $h_0(x) \in L_2$ is purely analogous and the case $h_0(x) \in M$ is considered very similarly. Due to the concavity of the curve $\tilde{M}$, it is geometrically clear that we have the internal distance estimate $d_{\mathbf{V}'}(h_0(x), P) \leqslant |\beta_1|$ and therefore also $|h_0(x) - P| \leqslant |\beta_1|$. It remains to estimate $|P - h_0(\tilde{x})|$.

Denote by $\mathrm{part}_\gamma(a, b)$ the part of a curve γ between two points $a, b \in \gamma$. Note that since $\tilde{\mathbf{S}}'$ is a shortest curve, we must have the estimate

$$|\mathrm{part}_{\tilde{\mathbf{S}}'}(\varphi_0(\tilde{s}_1), P)| \leqslant |\beta_1| + |\mathrm{part}_{\mathbf{S}'}(\varphi_0(s_1), h_0(x))| + d_{\mathbf{V}'}(h_0(x), P).$$

Similarly, for $\mathbf{S}'$, we get

$$|\mathrm{part}_{\mathbf{S}'}(\varphi_0(s_1), h_0(x))| \leqslant |\beta_1| + |\mathrm{part}_{\tilde{\mathbf{S}}'}(\varphi_0(\tilde{s}_1), P)| + d_{\mathbf{V}'}(h_0(x), P).$$

Since $d_{\mathbf{V}'}(h_0(x), P) \leqslant |\beta_1|$, we find the comparison estimate

$$\big| |\mathrm{part}_{\mathbf{S}'}(\varphi_0(s_1), h_0(x))| - |\mathrm{part}_{\tilde{\mathbf{S}}'}(\varphi_0(\tilde{s}_1), P)| \big| \leqslant 2|\beta_1|. \tag{3.4}$$

A similar argument shows that

$$\big| |\mathbf{S}'| - |\tilde{\mathbf{S}}'| \big| \leqslant |\beta_1| + |\beta_2|. \tag{3.5}$$

Note next that since x and $\tilde{x}$ are on the same vertical line, since the horizontal segments S and $\tilde{S}$ have endpoints on the same sides of H, and since the angle between the sides of H and the horizontal direction is bounded absolutely from below, we have that

$$\big| |x - s_1| - |\tilde{x} - \tilde{s}_1| \big| \leqslant C_1 |s_1 - \tilde{s}_1| \leqslant C_2 |x - \tilde{x}|. \tag{3.6}$$

By triangle inequality, we also have the direct estimate

$$\big| |S| - |\tilde{S}| \big| \leqslant |s_1 - \tilde{s}_1| + |s_2 - \tilde{s}_2| \leqslant C|x - \tilde{x}|. \tag{3.7}$$

Due to the constant speed parametrization of the curves $\mathbf{S}'$ and $\tilde{\mathbf{S}}'$, we may now combine (3.4)–(3.7) to calculate that

$$\begin{aligned}
|P - h_0(\tilde{x})| &\leqslant |\mathrm{part}_{\tilde{\mathbf{S}}'}(P, h_0(\tilde{x}))| \\
&= \big| |\mathrm{part}_{\tilde{\mathbf{S}}'}(\varphi_0(\tilde{s}_1), h_0(\tilde{x}))| - |\mathrm{part}_{\tilde{\mathbf{S}}'}(\varphi_0(\tilde{s}_1), P)| \big| \\
&\leqslant \big| |\mathrm{part}_{\tilde{\mathbf{S}}'}(\varphi_0(\tilde{s}_1), h_0(\tilde{x}))| - |\mathrm{part}_{\mathbf{S}'}(\varphi_0(s_1), h_0(x))| \big| + 2|\beta_1| \\
&= \left| \frac{|\tilde{s}_1 - \tilde{x}|}{|\tilde{S}|} |\tilde{\mathbf{S}}'| - \frac{|s_1 - x|}{|S|} |\mathbf{S}'| \right| + 2|\beta_1| \\
&\leqslant \left| \frac{|\tilde{s}_1 - \tilde{x}|}{|\tilde{S}|} - \frac{|s_1 - x|}{|S|} \right| |\mathbf{S}'| + 3|\beta_1| + |\beta_2| \\
&\leqslant \left| \frac{|\tilde{s}_1 - \tilde{x}| - |s_1 - x|}{|S|} \right| |\mathbf{S}'| + C\frac{|\mathbf{S}'|}{|S|} |x - \tilde{x}| + 3|\beta_1| + |\beta_2| \\
&\leqslant C_2 \frac{|\mathbf{S}'|}{|S|} |x - \tilde{x}| + 3|\beta_1| + |\beta_2|.
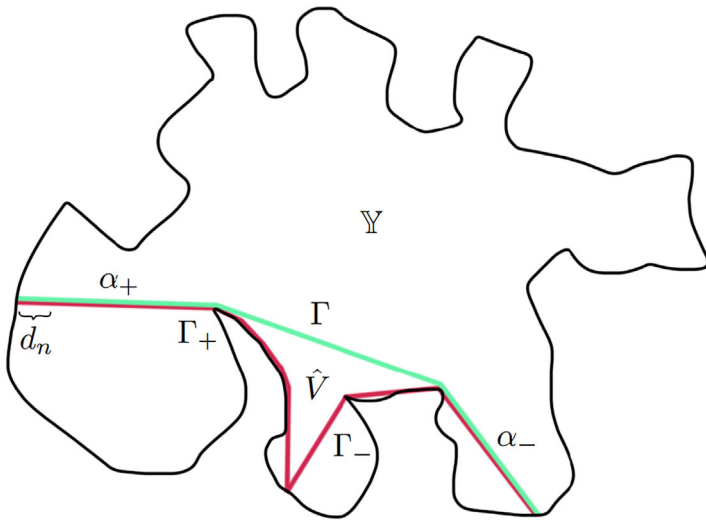\end{aligned}$$

FIGURE 7 The configuration on the target side.

Note now that since the curve $\mathbf{S'}$ is a shortest curve in the domain $\mathbf{V'}$, by the fact that $\mathbf{S'}$ may not turn more than π degrees, we find the estimate $|\mathbf{S'}| \leqslant C \operatorname{diam}(\mathbf{S'})$. The Hölder-continuity of the boundary values φ_0 gives that $\operatorname{diam}(\mathbf{S'}) \leqslant C|S|^\alpha$. Combining these and remembering that we assumed $|x - \tilde{x}| \leqslant |S|$ in (3.2), for the first term above, we have the estimate

$$\frac{|\mathbf{S'}|}{|S|}|x - \tilde{x}| \leqslant C|S|^{1-\alpha}|x - \tilde{x}| \leqslant C|x - \tilde{x}|^\alpha.$$

Combining (3.3), (3.6), and the Hölder-continuity of φ_0 gives that

$$|\beta_i| \leqslant C|x - \tilde{x}|^\alpha, \ i = 1, 2.$$

This finally yields that

$$|h_0(x) - h_0(\tilde{x})| \leqslant |h_0(x) - P| + |P - h_0(\tilde{x})| \leqslant C|x - \tilde{x}|^\alpha.$$

Hence, the extension h_0 as defined is Hölder-continuous with the same exponent and comparable constants to the original boundary map φ.

Step 4. The general construction. Let us keep the indices k, n fixed for ease of presentation and consider the set $V = V_{n,k}$ defined as the union of $\Gamma = \Gamma_{n,k}$, its two children Γ_+ and Γ_-, and the region bounded by them. Note that the curves Γ_+ and Γ_- may overlap with Γ and also with each other. In fact, due to the definition of, say, Γ and Γ_+ being shortest curves starting at the point A, their intersection is either the single point $\varphi(a)$ or one connected curve α_+ starting from $\varphi(a)$. Similarly, α_- is defined as the common part of Γ and Γ_-, and $\alpha_\pm$ as the common part of Γ_+ and Γ_-. The Jordan domain $\hat{V}$ is defined by removing from V the common parts α_+, α_-, and $\alpha_\pm$, see Figure 7. The boundary $\partial\hat{V}$ consists of three curves that are concave toward its interior due to the assumption that they are nonintersecting shortest curves.

Let us now recall the set $U = U_{n,k}$ given on the domain side as the region between the curve $\gamma = \gamma_{n,k}$ (consisting of two sides of the isosceles triangle on top of $I = I_{n,k}$) and its two dyadic children
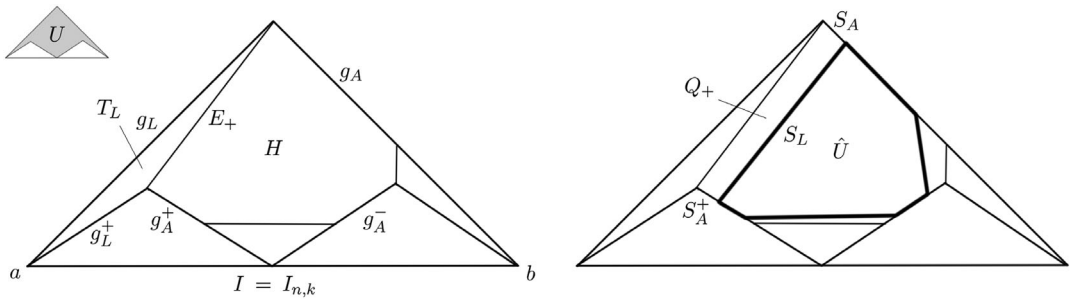
FIGURE 8 Different configurations on the domain side.

γ_+ and γ_-, see Figure 8 for the shape of U. In the second step, we have defined a parametrization $\tau : I \to \Gamma$ that is Hölder-continuous with the correct estimates. The map τ was defined for each dyadic interval and we will sometimes distinguish this map by writing the dependence as $\tau = \tau_\Gamma$, as opposed to τ_{Γ_+} or τ_{Γ_-} that are also maps defined on subintervals of I but not necessarily equal to τ_Γ. We would like to lift the parametrization τ from the interval I to define a parametrization $\tau' : \gamma \to \Gamma$ and repeat this for the children γ_+ and γ_- to define a "boundary map" $\tau' : \partial U \to V$, but unfortunately, this lifting process cannot be done directly as there may be issues with the fact that the curve $\gamma \subset \partial U$ has very small angles with respect to γ_+ and γ_- at the two endpoints of I, causing possible issues with Hölder-estimates if the parametrization τ_Γ has some notable difference with τ_{Γ_+} and τ_{Γ_-} near these points.

Instead, we aim to split each curve γ into three parts g_A, g_L, and g_R. The two parts g_L and g_R will be line segments starting from the two endpoints a, b of I, and they will be mapped to Γ so that their images are very close to the endpoints $\varphi(a)$ and $\varphi(b)$ of Γ. In fact, at the start of the construction (after fixing the $\Gamma_{n,k}$:s), we fix a sequence (d_n) of numbers decreasing very quickly to zero, and on each dyadic level n, the curves g_L, g_R will be mapped to parts of Γ that have length d_n. The sequence d_n may be chosen to tend to zero arbitrarily fast without impacting Hölder-continuity estimates, and we will use this fact later to be able to say that the extension map h is practically a constant map in the part of U near to the small angles at the endpoints a, b of I, and hence, we do not need to worry about Hölder-estimates there.

We opt to call the part $g_A \subset \gamma$ the *active part* of γ, and its image on Γ will be most of Γ. The parts g_A, g_L, and g_R are defined as follows, see Figure 8. If the curve γ is not on the first dyadic level, it has a parent curve γ_P and shares one endpoint with γ_P, say the left endpoint a, for example. In this case, g_L will be the left line segment of γ, and g_R will be the part of the right line segment of γ obtained as a line segment whose vertical projection to the x-axis is the same as the right line segment of the rightmost child curve γ_-. If γ is on the first dyadic level, we do the process of defining g_L and g_R via the dyadic children as outlined in the previous sentence for both sides of γ. The active part g_A is what is left of γ.

We may connect the active part g_A with the active parts g_A^+ and g_A^- of γ_+ and γ_- by taking the convex hull of their union, which defines a hexagonal region $H \subset U$ that is uniformly bilipschitz-equivalent to a ball of radius 2^{-n}. We now start defining the boundary map $\tau' : \partial U \to \Gamma \cup \Gamma_+ \cup \Gamma_- \subset V$, and we will also define τ' on the whole of ∂H.

Let us again suppose that we are in the situation where g_L is the whole left segment of γ and g_R is a part of the right segment of γ. We first set aside two subcurves Γ_L and Γ_R of the curve Γ that begin at each endpoint and have length d_n and d_{n+1}. Note that since both of the endpoints of Γ were assumed to be good points in the beginning, there are no technicalities here and, choosing d_n small
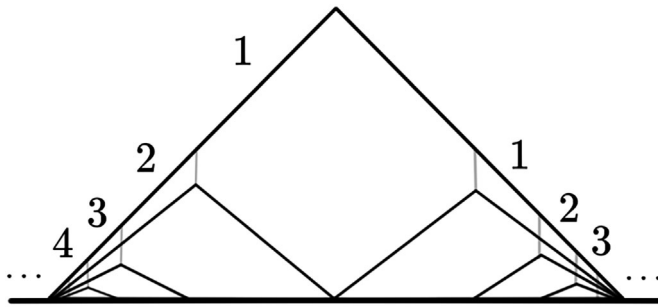
FIGURE 9   Dividing the segments g_L and g_R into dyadic parts on which to define τ'.

enough, these parts will always be either line segments or concave curves only touching one side of $\partial \mathbb{Y}$ near their endpoint. Let Γ_m denote the remaining middle part of Γ. We may now make a linear change of variables to transform the parametrization $\tau : \tau^{-1}(\Gamma_m) \to \Gamma_m$ into a parametrization $\tau' : g_A \to \Gamma_m$ from the active part of γ to Γ_m (in the case where we are on the first dyadic level, a minor modification to this is sufficient). The Hölder-estimates are inherited since $|g_A| \approx |I|$.

We next turn to define τ' on the curves g_L and g_R that will be mapped to Γ_L and Γ_R, respectively. We would like to impose the following rule that affects parametrizations across all dyadic levels: If a point $Q \in g_L$ and a point on $Q_+ \in g_L^+$ (with g_L^+ denoting the part g_L for the child curve γ_+) have the same projection to the x-axis, then they should be mapped to points on $\tau'(Q) \in \Gamma$ and $\tau'(Q_+) \in \Gamma_+$ that are "equally far away from $\varphi(a)$" meaning that

$$| \operatorname{part}_\Gamma(\varphi(a), \tau'(Q))| = | \operatorname{part}_{\Gamma_+}(\varphi(a), \tau'(Q_+))|. \tag{3.8}$$

We also impose an analogous condition for pairs $Q \in g_R$, $Q_- \in g_R^-$ having the same projection to the x-axis.

Note that Γ and Γ_+ may be the same curve near their mutual endpoint $\varphi(a)$, in which case this just says that the images of the points Q, Q_+ under τ' are equal. In fact, we make the following standing assumption here. If Γ and Γ_+ do overlap, meaning that α_+ is not equal to a singleton, then we choose d_n so small that $d_n \leqslant |\alpha_+|$ (here α_+ depends on n but also k, so we take a minimum over k). We also assume $d_n \leqslant |\alpha_-|$ and $d_n \leqslant |\alpha_\pm|$ as long as the curves are not singletons. This forces us to always map g_R and g_L under τ' to subsets of α_+ or α_- if these are not singletons.

The consequence of such a rule is that to define τ' on g_L and g_R, it is enough to define τ' on the part of g_L that does not project to any point that is also a projection from the child curve g_L^+. The definition of the map τ' on any other point of g_L will be obtained inductively from successive dyadic generations, and since we were in the case where g_R projects to the same set as g_R^-, we need not define τ' on any point on g_R yet. On the part of g_L where τ' needs to be defined, we may simply set it to map this part (which is a segment) with constant speed to the part of Γ that contains the points of length d_{n+1} to d_n from the endpoint $\varphi(a)$. In practice, this means that on g_L, the map τ' will consist of maps defined on a sequence of segments (see Figure 9) tending to the endpoint a in dyadic fashion having length comparable to 2^{-N}, $N = n, n+1, ...$, and each such segment is mapped with constant speed to a part of Γ with length $d_N - d_{N+1}$.

Hence, τ' has been defined on all parts of γ (and the curves γ for successive dyadic generations). We next define τ' on ∂H, where we only need to define τ' on the three sides of the hexagon ∂H not already contained in γ, γ_+, or γ_-. Let us call these segments E_+, E_-, and $E_\pm$, and the

definition of τ' on each such segment is similar enough that we may only do so on E_+ which is the segment between the nonmutual endpoints $a_L \in g_L$ and $a_L^+ \in g_L^+$, see Figure 8. To define τ' on E_+, we will define an image curve F_+ within V so that $\tau' : E_+ \to F_+$ is simply a constant speed parametrization. The goal is simply that the length of F_+ is at most $2d_n$. But since the internal distance (meaning path distance within the closed set V here) of the points $\tau'(a_L)$ and $\tau'(a_L^+)$ within V has to be at most $d_n + d_{n+1}$ since

$$| \operatorname{part}_\Gamma(\varphi(a), \tau'(a_L))| = d_n \quad \text{and} \quad | \operatorname{part}_{\Gamma_+}(\varphi(a), \tau'(a_L^+))| = d_{n+1},$$

this choice of F_+ is always possible. More precisely, we define F_+ to be the shortest curve in $\overline{V}$ between its endpoints $\tau'(a_L)$ and $\tau'(a_L^+)$.

Note that if Γ and Γ_+ overlap (so that α_+ is not a singleton), then F_+ may simply be chosen as the curve $\operatorname{part}_{\alpha_+}(\tau'(a_L), \tau'(a_L^+))$. Here, $\tau'(a_L), \tau'(a_L^+) \in \alpha_+$ by our earlier assumption that if α_+ is not a singleton, then $d_n < |\alpha_+|$.

Thus, now τ' has been defined as a constant speed map from E_+ to F_+ and since d_n was chosen small enough, we can assume to have the same Hölder-estimates for τ' on E_+ as the original boundary map.

It is now simple to also verify the Hölder-continuity of τ' on the whole of $\partial U \cup \partial H$. On each of the six sides of ∂H, the map τ' was already verified to have the correct Hölder-estimates, and since the hexagon ∂H has no small angles, this holds on the whole boundary as well. All the other parts of ∂U are mapped to curves where the small enough choice of the sequence (d_n) will guarantee Hölder-continuity.

Next, we finally define the extension h by extending the map τ' to the whole of U. Let us start with the region $H \subset U$. We must first account for the possible overlapped parts $\alpha_+, \alpha_-,$ and $\alpha_\pm$ in the target. Let us take the whole preimage B of these parts under τ'. Let us for now make a standing assumption that none of these parts are singletons, and we will cover the other case later. We explain here how this affects things around the preimage of α_+ and the cases of α_- and $\alpha_\pm$ are handled very similarly. Since g_L and g_L^+ are part of this preimage B, there are also some points on the boundary of H that are mapped to α_+: the curve E_+ in particular belongs to B but also some part of g_A and g_A^+ belongs to B. We may, however, assume that g_A is not fully contained in the preimage B — this may be achieved simply by choosing d_n small enough so that $|\Gamma| - |\alpha_+| > d_n$. Similarly, we may assume that g_A^+ and g_A^- are not contained in B. Hence, there are some subsets (line segments) of $g_A, g_A^+,$ and g_A^- that are not in B, and these three line segments themselves bound yet another hexagon which we denote by $\hat{U}$, see Figure 8. Now $\hat{U}$ is still convex but may be fairly thin, for example.

On three of the sides of $\hat{U}$ (the sides on the active parts), the map τ' was already defined. For each of the other three sides, the map τ' has the same value on both endpoints (this value is the endpoint of $\alpha_+, \alpha_-,$ or $\alpha_\pm$ not on $\partial \mathbb{Y}$), so we extend τ' as a constant map on these three sides. Now τ' has been defined on the whole boundary of the convex hexagon $\hat{U}$, and the image of $\partial \hat{U}$ under τ' is the boundary of the region $\hat{V}$ bounded by three concave curves. Hence, at this part of the construction, we may repeat the idea of the initial case from Step 3 where exactly the same situation was addressed to extend a boundary map from a convex hexagonal region to a region bounded by three concave curves. The Hölder-estimates obtained there show that h is Hölder-continuous in $\hat{U}$ with the correct estimates.

Now the set $H \setminus \hat{U}$ divides into up to three convex quadrilateral regions. For simplicity, we consider here the case of such a region Q_+ that has one side given by E_+ and shares points with
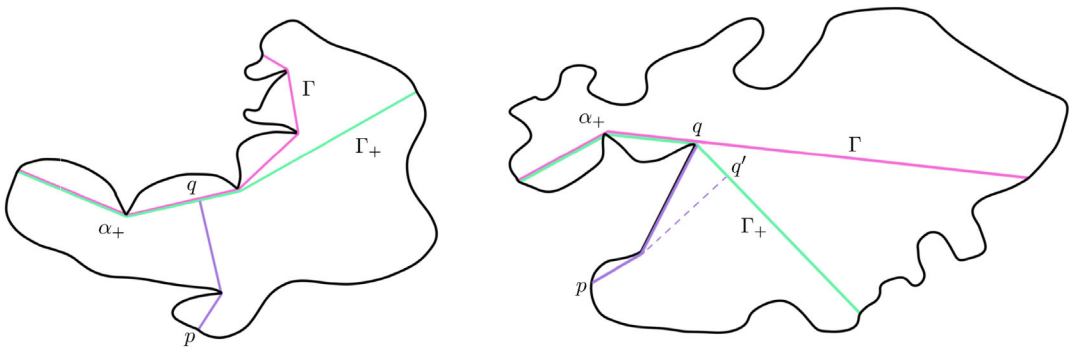
FIGURE 10 The shortest paths from p to $q \in \alpha_+$. If q is an endpoint of α_+, the situation may change.

the preimage of α_+. In fact, the whole boundary of Q_+ is mapped to α_+ under τ'. Let us label the sides of Q_+ as S_L, S_A, S_A^+, and E_+. Here, $S_A \subset g_A$ and $S_A^+ \subset g_A^+$. We wish to first compare how the map τ' behaves on the sets S_A and S_A^+, both mapped to α_+ injectively. On the respective endpoints of S_A and S_A^+ that also belong to S_L, τ' attains the same value. The other endpoints of S_A and S_A^+ (which are a_L and a_L^+) are mapped to the curve α_+ and are of curve length d_n and d_{n+1} away from the endpoint $\varphi(a)$, respectively. We distinguish a subsegment $Z \subset S_A^+$ for which $\tau'(g_L^+) \cup \tau'(Z) = \tau'(g_L)$, meaning that the endpoint of Z not on E_+ is also mapped to curve length d_n away from $\varphi(a)$. We then need a helpful remark.

Remark. Given a point p on $\varphi(I)$ and letting q denote the point on Γ that is the closest point to p with respect to the internal distance in $\overline{\mathbb{Y}}$, we have the following: If q lies on α_+ but is not an endpoint of α_+, then q is also the closest point to p on Γ_+ with respect to the internal distance, see Figure 10 on the left. To see this, consider the minimal boundary component of $\partial\mathbb{Y}$ that contains p and has endpoints on Γ_+. Then, the part of Γ_+ between these two endpoints has to be concave toward p and the closest point q' on Γ_+ to p with respect to the internal distance lies on this part. Now if $q' \neq q$, the three curves given by Γ_+ and the two shortest curves Γ_q from p to q and $\Gamma_{q'}$ from p to q' bound a region Δ whose three boundary curves are all concave toward the interior. Moreover, since q' was the closest point to p on Γ_+, it must hold that the angle between $\Gamma_{q'}$ and Γ_+ toward the interior of Δ is at least $\pi/2$. Since q was not an endpoint of α_+ and $\alpha_+ \subset \Gamma_+$, it must also hold that the angle between Γ_q and Γ_+ is at least $\pi/2$. Now the sum of two interior angles in the region Δ is at least π, which gives a contradiction as this cannot happen when each of the boundary curves of Δ is concave toward the interior.

Note, however, that if q is the other endpoint of α_+, it is possible that $q' \neq q$ if the situation is as in Figure 10 on the right. This is the only exceptional case, and to avoid additional consideration of this case in the future, we will simply choose to make an arbitrarily small modification to the parametrization map $\tau (= \tau_\Gamma)$ in this situation that does not affect the Hölder-estimates for τ (and by proxy τ'). This modification is done in a way where the preimage of all such exceptional points p under φ is not mapped to q by τ_Γ but rather to an arbitrarily small part of the curve Γ starting at the end of α_+.

The conclusion we make from the remark in the previous two paragraphs is that the parametrization map τ that can be defined for both the curve Γ and the curve Γ_+ (we distinguish these maps by τ and τ_+) behaves exactly the same on the preimage of α_+, meaning that

$\tau(x) = \tau_+(x)$ for $x \in \tau^{-1}(\alpha_+)$. Hence, the map τ' that was obtained from the map τ by a linear change of variables also satisfies that $\tau'|_{S_A}$ is simply a linear change of variables of $\tau'|_Z$. This lets us divide Q_+ into a triangle T' with two sides given by Z and E^+ and a quadrilateral Q' with three sides S_A, S_L, and $S_A^+ \setminus Z$.

On the quadrilateral Q', we simply define h as a linear interpolation between the maps $\tau'|_{S_A}$ and $\tau'|_{S_A^+ \setminus Z}$. More precisely, since Q' may be represented as $Q' = \{tS_A + (1 - t)(S_A^+ \setminus Z) : 0 \leqslant t \leqslant 1\}$, we define h on each intermediate segment $tS_A + (1 - t)(S_A^+ \setminus Z)$ as simply the map τ' composed with the natural projection from $tS_A + (1 - t)(S_A^+ \setminus Z)$ to S_A. This gives a Hölder-continuous map $h : Q' \to \alpha_+$ with estimates inherited from τ' since the map is practically just τ' in one direction and constant in the other.

On the triangle T', the map τ' maps the two sides E^+ and Z both onto the same part of α_+ between d_{n+1} and d_n length from $\varphi(a)$. The parametrizations here may be slightly different as the map τ' on E^+ was simply defined via constant speed parametrization, while Z it is inherited from the map τ. Nevertheless, since both parametrizations satisfy appropriate Hölder-estimates, and due to d_n being small enough, we may simply linearly interpolate between these two parametrizations in T' to define an extension map $h : T' \to \tau'(Z) \subset \alpha_+$ with the correct Hölder-estimates.

This defines the map h on Q_+, and hence, it is now defined on the whole hexagon $H \subset U$. The same idea as in the previous paragraph may be applied to define the extension h on the triangle T_L with sides g_L and g_L^+. In fact, if we remove from g_L the part that projects to the x-axis as the same set as g_L^+, this splits the triangle T_L into two triangles. In one triangle, we may use the x-axis projection property (3.8) to define the extension h simply as a constant map on vertical segments, while on the remaining triangle (whose one side is S_L), we may use the same idea as in the previous paragraph. This defines the extension h on T_L and the same idea may be used to define h on the whole of U.

Let us now address the case where α_+ is a singleton. Then, we would modify the above construction as follows. First, we note that the curves Γ and Γ_+ must be concave toward each other near their common point $\varphi(a)$, so we may choose the earlier curve F_+ as simply a line segment between the respective endpoints given by points at length d_n and d_{n+1} on Γ and Γ_+, respectively (if needed, we can decrease d_n beforehand so that this segment F_+ has no additional intersections with other curves). We recall that τ' maps the edge E_+ of the hexagon H to this segment F_+, and in this case, E_+ is also an edge of the hexagon $\hat{U}$ mapped to $\hat{V}$ previously. But in this case, we must instead construct the extension h to map $\hat{U}$ to the set W obtained by removing from $\hat{V}$ the small part W_+ near $\varphi(a)$ that is cut away from it by the segment F_+. If some of the curves α_- or $\alpha_\pm$ are also singletons, this is reflected in the definition of W by also cutting away appropriate parts W_- and/or $W_\pm$.

Hence, the set W consists three concave boundary curves and up to three other boundary parts that are arbitrarily small segments between these curves. We would like to still use the idea from the "initial case" to define an extension $h : \hat{U} \to W$, but the argument is not applicable without some modifications as it assumed that the target boundary did not have the additional small segments such as F_+. Nevertheless, one may verify that as long as these segments are small enough, there is no essential change to the argument or estimates. Hence, this is just a matter of choosing the numbers d_n to be appropriately small depending on the curve Γ and its children. In fact, one may even repeat the arguments of the initial construction in this situation, and notice that the shortest curves in W are still only consisting of at most two segments and a concave part in the middle, so the proof of Hölder-estimates will go through similarly as well.
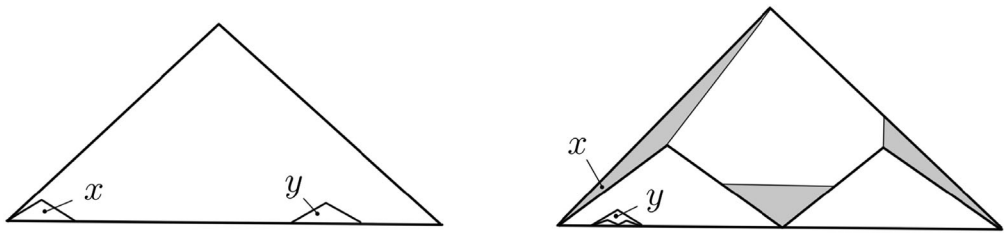
FIGURE 11   The two cases for positioning of x and y. On the right, the triangular regions mapped very close to the boundary are highlighted.

On the triangle T_L, we define the map $h : T_L \to W_+$ as an extension of the boundary values τ'. For such an extension, we may actually use the idea from the "initial case" since W_+ is bounded by two concave curves and a segment. But any reasonable extension method here will also work as long as the number d_n is chosen small enough since any bounds obtained will get uniformly smaller as d_n tends to zero. Some care should also be taken here to respect the way τ' was defined dyadically on the boundary curves g_L and g_L^+ of T_L, recall Figure 9. Each such piece of g_L defines a "dyadic slice" of T_L that projects to the same region on the x-axis as g_L, and we may define the extension h on T_L so that on the mth dyadic slice, we have estimates controlled by d_{n+m-1} instead of d_n. This observation is fairly trivial but necessary to get the correct global bounds as we approach the boundary.

Step 5. Global Hölder continuity. In the previous part of the construction, we were able to define the extension h in each set $U = U_{n,k}$ and verified that it satisfies the appropriate Hölder-continuity estimates locally in each such set. The fact that h maps $U_{n,k}$ as a monotone map to $V_{n,k}$ means that this already defines h as a monotone map of the whole disk $\mathbb{D}$ to the target domain $\mathbb{Y}$. In this part of the proof, we will verify that the map h is also Hölder-continuous in the whole disk $\mathbb{D}$ with the appropriate estimates inherited from its local behavior.

To do this, consider two points $x, y \in \mathbb{D}$. Thus, $x \in U_{n,k}$ and $y \in U_{n',k'}$ for some n, k, n', k', where we may suppose that $n \leqslant n'$. Let N be such that $2^{-N} \approx |x - y|$. We split into two cases depending on how far x and y are apart compared to the level of their dyadic sets, see Figure 11.

Case 1. $N \leqslant n$. In this case, we may traverse from x to y via two sequences of sets $U_x^n, U_x^{n-1}, \dots, U_x^N$ and $U_y^N, U_y^{N+1}, \dots, U_y^{n'}$. Each of these sets is one of the dyadic sets $U_{m,l}$ considered before, with $x \in U_x^n, y \in U_y^{n'}$ and the middle set $U_x^N = U_y^N$ is from the dyadic level N and hence having diameter comparable to 2^{-N}. Moreover, the dyadic level changes by at most one in each step that allows us to use the Hölder-continuity of h in each set to estimate that

$$|h(x) - h(y)| \leqslant \sum_{l=n}^{N} \operatorname{diam}\left(h\left(U_x^l\right)\right) + \sum_{l=n'}^{N} \operatorname{diam}\left(h\left(U_y^l\right)\right)$$

$$\leqslant C \sum_{l=n}^{N} 2^{-l\alpha} + C \sum_{l=n'}^{N} 2^{-l\alpha}$$

$$\leqslant 2C \sum_{l=N}^{\infty} 2^{-l\alpha}$$
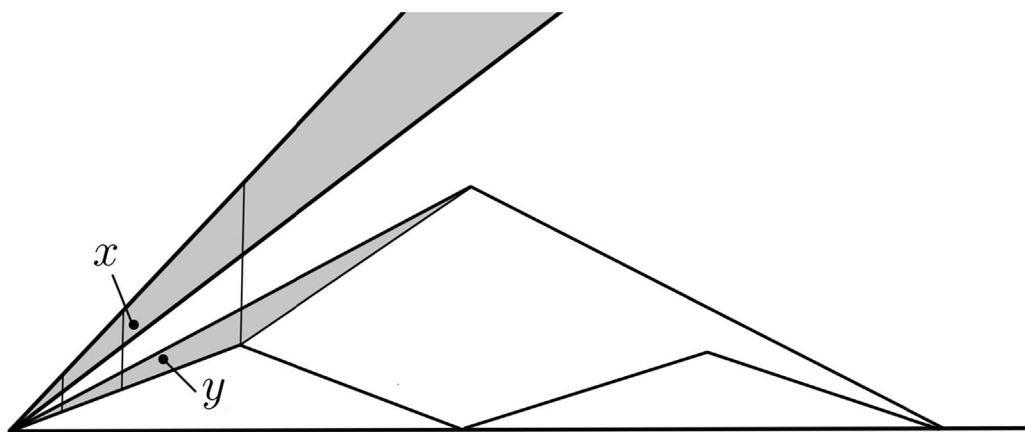
$$\leqslant C' 2^{-N\alpha}.$$

FIGURE 12   Points x and y very close to each other.

Since $2^{-N} \approx |x - y|$, this gives the correct Hölder-continuity estimate.

Case 2. $N > n$. We consider a few possibilities. If x and y are in the same set $U_{n,k}$ or y is in one of the two child sets of $U_{n,k}$, then there is nothing to prove as we only need to use the local Hölder estimate for h at most twice to get the required estimate.

Otherwise x must be in one of the triangular regions in $U_{n,k}$ that were mapped to sets of distance at most d_n from the corresponding boundary point under h, see the right part of Figure 11, as otherwise the distance $|x - y|$ would be too large.

If now $U_{n',k'}$ does not share a boundary point with $U_{n,k}$ (as in the right part of Figure 11), then we can argue using the idea from Case 1 to assume that it does. Indeed, Case 1 lets us replace y with any point y' belonging to a set $U_{n',m}$ within the same generation n' as long as $|y - y'| \approx 2^{-N}$. Geometrically, it is simple to see that if $U_{n',k'}$ did not already share a boundary point with $U_{n,k}$, then we may pick the new point y' from such a set instead. Hence, we are assumed to be in a configuration similar to Figure 12.

In this case, we must recall the construction of the map h inside such triangular regions, where it was defined dyadically by splitting the region into dyadic slices depending on their projection to the x-axis as in Figure 9 and the discussion at the end of Step 4. Let us suppose that x lies in the mth dyadic slice, mapped to a small set of diameter less than d_{n+m-1}. We may also suppose that y belongs to the corresponding dyadic slice within $U_{n',k'}$, as otherwise we may again replace y by such a point while this time using the Hölder-estimates inside $U_{n',k'}$.

Now both x and y are mapped within distance d_{n+m-1} of each other. We may now conclude by assuming that the sequence d_n was picked to be small enough so that we get the desired estimate. To be precise, since we assumed that x and y belonged to corresponding dyadic slices that were not neighbors, the distance $|x - y|$ is bounded from below in terms of some function $n + m - 1$ (depending on our initial choice of angles for the curves $\gamma_{n,k}$ on the domain side). Hence, the inequality $d_{n+m-1} \leqslant C|x - y|^{\alpha}$ may be obtained by small enough choice of sequence (d_n).

Step 6. Injectification of the map. In this part of the proof, we explain how the map h may be modified slightly to produce a homeomorphic extension $H : \mathbb{D} \to \mathbb{Y}$ instead. The map h itself has the potential issue of not being injective on each set $U_{n,k}$. For example, in the case where $\Gamma = \Gamma_{n,k}$ and its child curve Γ_+ intersect along a curve α_+, then the preimage of α_+ will be a set where we
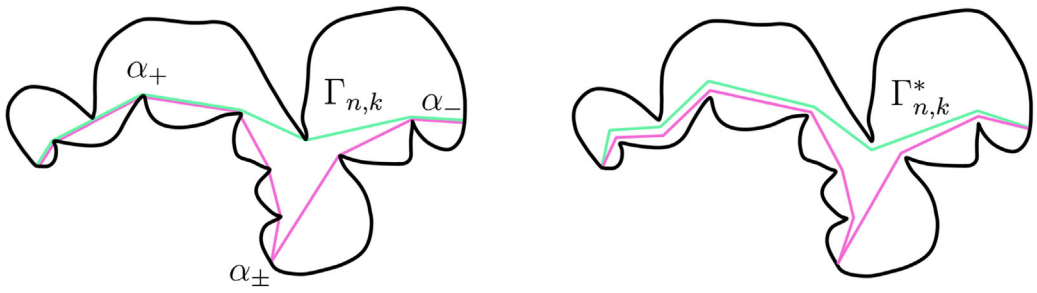
FIGURE 13   Injectification of the map.

have loss of injectivity. The curves $\Gamma_{n,k}$ may also touch $\partial\mathbb{Y}$ outside of their endpoints as they were simply defined as shortest curves inside $\overline{\mathbb{Y}}$.

But since the endpoints of each curve $\Gamma_{n,k}$ were "good points" in $\partial\mathbb{Y}$, it is fairly clear that we may move each curve slightly toward the interior of $\mathbb{Y}$ and even replace each such curve by a piecewise linear approximation instead. See Figure 13 for an illustration. Here, we may choose to make the piecewise linear approximation such that it produces a finite number of pieces in each compact subset of $\mathbb{Y}$, possibly allowing for fine adjustments as we near the boundary at the endpoints of $\Gamma_{n,k}$ (to account for the dyadic way the map h was defined near the preimages of such endpoints, recall again Figure 9). This may also be done in such a way that none of the curves $\Gamma_{n,k}$ intersect within the interior of $\mathbb{Y}$ and hence produce topologically the same configuration as the curves $\gamma_{n,k}$ on the domain side. It is also clear that such modifications may be done in an arbitrarily small fashion on each dyadic level, if necessary so that the adjustment also gets smaller and smaller as we near an endpoint of $\Gamma_{n,k}$.

The extension map h will be modified a follows. First, we adjust the parametrization τ' to map instead to the new image curves $\Gamma_{n,k}^*$ in place of $\Gamma_{n,k}$. At this point, we may also injectify the parametrization τ' in parts where it may have not been injective before (certain intervals could be mapped into single points, recall Figure 2). This can also be done in an arbitrarily small way, again adjusting for increasingly smaller changes as we near the boundary.

On the sets where h was already injective before, it is fairly clear that the aforementioned changes to the curves $\Gamma_{n,k}$ only produce very small perturbations of the construction of h. On the preimages of the possible overlapped curves α_+, α_-, and $\alpha_\pm$ under h, we must redefine h as a map H that now sends the preimages to an appropriate region defined by the two replacements of such curves. For example, if originally α_+ was the intersection of Γ and Γ_+, then after the fact that we have replaced Γ and Γ_+ by two curves Γ^* and Γ_+^* that do not intersect each other, there are parts $\alpha^* \subset \Gamma^*$ and $\alpha_+^* \subset \Gamma_+^*$ that correspond to the part that was α_+ before. Since the adjustment may be done in an arbitrarily small way, the other endpoints of the curves α^* and α_+^* (which are now distinct points) may be connected by an arbitrarily small curve ℓ^*, in fact a line segment if the earlier perturbation is done appropriately. Furthermore, within the region defined by α^*, α_+^*, and ℓ^*, it will be possible to define an isotopy α^t, $t \in [0,1]$, between α^* and α_+^* with the other endpoint of α^t moving on the segment ℓ^*.

Recalling that the map h was defined on all preimages of the curve α_+ as some form of homotopy between different parametrizations of α_+. For example, in the quadrilateral $Q' \subset Q_+$, the map h was simply defined by linear interpolation between the boundary map τ' on two opposite sides of Q'. If we now modify this linear interpolation by mapping to the curves α^t, where t runs over the whole interval $[0,1]$, instead then this produces an injective version of the map h in this

set. As usual, since the modifications were arbitrarily small, the change to the Hölder-constant may also be assumed to be arbitrarily small here. □

JOURNAL INFORMATION
The *Journal of the London Mathematical Society* is wholly owned and managed by the London Mathematical Society, a not-for-profit Charity registered with the UK Charity Commission. All surplus income from its publishing programme is used to support mathematicians and mathematics research in the form of research grants, conference grants, prizes, initiatives for early career researchers and the promotion of mathematics.

REFERENCES
1. L. V. Ahlfors, *Extension of quasiconformal mappings from two to three dimensions*, Proc. Natl. Acad. Sci. U.S.A. **51** (1964), 768–771.
2. J. Ball, *Global invertibility of Sobolev functions and the interpenetration of matter*, Proc. Roy. Soc. Edinburgh Sect. A **88** (1981), no. 3–4, 315–328.
3. J. C. Bellido and C. Mora-Corral, *Approximation of Hölder continuous homeomorphisms by piecewise affine homeomorphisms*, Houston J. Math. **37** (2011), no. 2, 449–500.
4. S. Daneri and A. Pratelli, *Smooth approximation of bi-Lipschitz orientation-preserving homeomorphisms*, Ann. Inst. H. Poincaré Anal. Non Linéaire **31** (2014), 567–589.
5. S. Daneri and A. Pratelli, *A planar bi-Lipschitz extension theorem*, Adv. Calc. Var. **8** (2015), no. 3, 221–266.
6. S. Hencl and A. Pratelli, *Diffeomorphic approximation of $W^{1,1}$ planar Sobolev homeomorphisms*, J. Eur. Math. Soc. **20** (2018), no. 3, 597–656.
7. S. Hencl, A. Koski, and J. Onninen, *Sobolev homeomorphic extensions from two to three dimensions*, arXiv 2112.14767.
8. T. Iwaniec, L. V. Kovalev, and J. Onninen, *Diffeomorphic Approximation of Sobolev Homeomorphisms*, Arch. Rational Mech. Anal. **201** (2011), 1047–1067.
9. P. Koskela, A. Koski, and J. Onninen, *Sobolev homeomorphic extensions onto John domains*, J. Funct. Anal. **279** (2020), 17pp.
10. A. Koski and J. Onninen, *Sobolev homeomorphic extensions*, J. Eur. Math. Soc. **23** (2021), no. 12, 4065–4089.
11. P. Kovalev, *Symmetrization and extension of planar bi-Lipschitz maps*, Ann. Acad. Sci. Fenn. Math. **43** (2018), no. 1, 541–556.
12. P. Kovalev, *Optimal extension of Lipschitz embeddings in the plane*, Bull. London Math. Soc. **51** (2019), no. 4, 622–632.
13. E. E. Moise, *Geometric topology in dimensions 2 and 3*, Graduate Texts in Mathematics, vol. 47, Springer, New York-Heidelberg, 1977.
14. C. Mora-Corral, *Approximation by piecewise affine homeomorphisms of Sobolev homeomorphisms that are smooth outside a point*, Houston J. Math. **35** (2009), 515–539.
15. P. Tukia, *The planar Schönflies theorem for Lipschitz maps*, Ann. Acad. Sci. Fenn. Ser. A I Math. **5** (1980), 49–72.
16. P. Tukia, *Extension of quasisymmetric and Lipschitz embeddings of the real line into the plane*, Ann. Acad. Sci. Fenn. Ser. A I Math. **6** (1981), 89–94.
17. P. Tukia and J. Väisälä, *Quasiconformal extension from dimension n to $n + 1$*, Ann. of Math. (2) **115** (1982), no. 2, 331–348.