MINIMUM WIDTH FOR UNIVERSAL APPROXIMATION IN $W^{1,p}$ SPACES

Anonymous authors
Paper under double-blind review

ABSTRACT

The minimum width for universal approximation has been extensively studied for the L^p norm, the L^∞ norm, and more recently the $W^{1,\infty}$ norm. A notable observation is that the minimum width for the L^∞ norm is strictly larger than that for the L^p norm for some input/output dimensions; however, no other non-trivial separation between norms is known. In this work, we study the minimum width for the $W^{1,p}$ norm with finite p and ask whether it is separated from the L^p norm and from the $W^{1,\infty}$ norm. We first show that for networks with leaky-RELU-like activation functions, if the output dimension d_{out} is larger than $3/2$ times the input dimension d_{in} and $p \leq d_{\text{out}} - d_{\text{in}}$, then the minimum width for the $W^{1,p}$ norm is d_{out}. This value coincides with that for the L^p norm and if $d_{\text{out}} \leq 2d_{\text{in}}$, it is strictly smaller than that for the $W^{1,\infty}$ norm. We then show that the $W^{1,p}$ norm is separated from the L^p norm through a finer analysis of the case $d_{\text{out}} = 2d_{\text{in}} - 1$ with $d_{\text{in}} \geq 3$, where we characterize the minimum width for all p. Here, the minimum width coincides with that for the L^p norm if $p \leq d_{\text{in}} - 1$, increases by one when p exceeds $d_{\text{in}} - 1$, and coincides with that for the $W^{1,\infty}$ norm if $p > d_{\text{in}}$. Furthermore, if d_{in} is odd, it increases by one once more when p exceeds d_{in}. Hence, the $W^{1,p}$ norm is separated from both the L^p norm and the $W^{1,\infty}$ norm. In addition, unlike that for the L^p norm, the minimum width for the $W^{1,p}$ norm depends on p and even on the parity of d_{in}. We also extend our results to general input and output dimensions, and a general class of activation functions.

1 INTRODUCTION

Neural networks have shown remarkable performance in various practical applications such as computer vision (He et al., 2016), natural language processing (Vaswani, 2017; Brown et al., 2020), and science (Jumper et al., 2021). In some modern applications, neural networks are required to fit not only the function values but also the gradients with respect to the input. For example, physics-informed neural networks (PINNs) solve partial differential equations by fitting the derivatives of networks to the equations (Raissi et al., 2019), and Sobolev training improves the performance of networks by fitting the gradients of networks to those of a target function (Czarnecki et al., 2017). These applications are theoretically supported by the classical universal approximation theorem, which states that two-layer networks can approximate any continuous function if an activation function is non-polynomial (Cybenko, 1989; Hornik et al., 1989; Leshno et al., 1993; Pinkus, 1999). Likewise, it is known that two-layer networks can simultaneously approximate function values and gradients (Hornik et al., 1990; Li, 1996), i.e., the minimum depth for universal approximation is two.

With the emergence of deep learning, the expressive power of deep networks has also been studied. In particular, a line of research has investigated the *minimum width* required for universal approximation as a dual problem to the classical universal approximation problem (Lu et al., 2017; Hanin & Sellke, 2017; Kidger & Lyons, 2020; Park et al., 2021). For the L^p norm with $1 \leq p < \infty$, the minimum width is now well understood and is known to be $\max\{d_{\text{in}}, d_{\text{out}}, 2\}$ for a wide range of activation functions (Park et al., 2021; Cai, 2023; Kim et al., 2024; Shin et al., 2025), where d_{in} and d_{out} denote the input and output dimensions. On the other hand, for the L^∞ norm (i.e., the uniform norm), the exact minimum width is unknown in general: the best known lower bound is $\max\{d_{\text{in}} + 1, d_{\text{out}} + \mathbf{1}_{d_{\text{in}} < d_{\text{out}} \leq 2d_{\text{in}}}\}$ (Hanin & Sellke, 2017; Johnson, 2019; Cai, 2023; Kim et al., 2024), and the best known upper bound is $\min\{d_{\text{in}} + d_{\text{out}}, \max\{2d_{\text{in}} + 1, d_{\text{out}}\}\}$ up to an additive constant depending on the activation function (Hanin & Sellke, 2017; Kidger & Lyons, 2020; Hwang,

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

Table 1: A summary of known bounds and our results on the minimum width $w_{\min}$ for universal approximation from $[0,1]^{d_{\text{in}}}$ to $\mathbb{R}^{d_{\text{out}}}$ with leaky-ReLU-like activation functions (Definition 2.1).

Norm	Reference	Dimensions	Range of p	Min. width $w_{\min}$
L^p	[Cai23, KMP24, SKHP25]†	arbitrary	$[1,\infty)$	$= \max\{d_{\text{in}}, d_{\text{out}}, 2\}$
	[Joh19, Cai23, KMP24]†	arbitrary	∞	$\geq \max\{d_{\text{in}}+1, d_{\text{out}}+\mathbf{1}_{d_{\text{in}}<d_{\text{out}}\leq 2d_{\text{in}}}\}$
	[Hwa25a, Hwa25b]	arbitrary	∞	$\leq \min\{d_{\text{in}}+d_{\text{out}}, \max\{2d_{\text{in}}+1, d_{\text{out}}\}\}$
	[KMP24, Hwa25a]	$(1,2),(2,2),(2,3),(2k,4k-1)$	∞	$= 3,\ 4,\ 4,\ 4k$
$W^{1,p}$	[Hwa25b]	arbitrary	∞	$\leq \min\{d_{\text{in}}+d_{\text{out}}, \max\{2d_{\text{in}}+1, d_{\text{out}}\}\}$
		$d_{\text{out}} = 2d_{\text{in}}-1,\ d_{\text{in}} \geq 3$‡	∞	$= 2d_{\text{in}} + \mathbf{1}_{d_{\text{in}}=\text{odd}}$
	Ours (Thm 3.1)	$d_{\text{in}} \geq 2,\ d_{\text{out}} > 3d_{\text{in}}/2$	$[1, d_{\text{out}}-d_{\text{in}}]$	$= d_{\text{out}}$
	Ours (Thm 3.2)	$d_{\text{out}} = 2d_{\text{in}}-1,\ d_{\text{in}} \geq 3$	$[1, d_{\text{in}}-1]$	$= 2d_{\text{in}}-1$
			$(d_{\text{in}}-1, d_{\text{in}}]$	$= 2d_{\text{in}}$
			(d_{in}, ∞)	$= 2d_{\text{in}} + \mathbf{1}_{d_{\text{in}}=\text{odd}}$
	Ours (Cor 3.3)	$d_{\text{in}} \geq 3$, arbitrary d_{out}	$[1, d_{\text{in}}-1]$	$\leq \min\{d_{\text{in}}+d_{\text{out}}, \max\{d_{\text{out}}, 2d_{\text{in}}-1\}\}$
			$(d_{\text{in}}-1, d_{\text{in}}]$	$\leq \min\{d_{\text{in}}+d_{\text{out}}, \max\{d_{\text{out}}, 2d_{\text{in}}\}\}$
		$d_{\text{in}} \geq 3,\ d_{\text{in}} \leq d_{\text{out}} \leq 2d_{\text{in}}-1$	$(p^\star, \infty)$	$\geq d_{\text{out}}+1$
			(j, ∞)§	$\geq d_{\text{out}}+2$

† also holds for broader classes of activation functions. ‡ [Hwa25b] also identifies $w_{\min}$ for infinitely many other $(d_{\text{in}}, d_{\text{out}})$. ⋆ $p^\star := \max\{2, \lceil (d_{\text{out}}-1)/2 \rceil\}$. § for some odd j with $p^\star < j \leq d_{\text{in}}$.

2025a). Interestingly, these results imply that unlike the minimum depth, the minimum width depends not only on the input and output dimensions but also on the norm that measures the approximation error: the minimum width for the L^∞ norm is strictly larger than that for the L^p norm whenever $d_{\text{in}} \geq \max\{d_{\text{out}}, 2\}$, i.e., the minimum width increases as $p \to \infty$ and the L^∞ norm and the L^p norm are separated.

The minimum width for approximating function values and gradients simultaneously has been studied very recently, and to the best of our knowledge, only in a single work (Hwang, 2025b). They consider the $W^{1,\infty}$ norm, i.e., the L^∞ norm of both the function value and the gradient, and identify the exact minimum width for infinitely many $(d_{\text{in}}, d_{\text{out}})$. Their key idea is to reduce the problem to a topological one: whether a target function can be approximated by a projection of an embedding of $\mathbb{R}^{d_{\text{in}}}$ into $\mathbb{R}^\ell$ for the width ℓ. However, this result does not determine whether the minimum width for the L^∞ norm and that for the $W^{1,\infty}$ norm are different. Hence, other than the separation between the L^p norm and the L^∞ norm, no non-trivial separation is known.

The separation between the L^p norm and the L^∞ norm arises because an L^p approximator is allowed to be inaccurate on a small part of the domain, which a narrow network can exploit, whereas an L^∞ approximator is not. Similar to the L^∞ norm, the $W^{1,\infty}$ norm does not allow an approximator to be inaccurate anywhere, and the minimum width is accordingly large. A natural question is then what happens for the $W^{1,p}$ norm with finite p, which measures an integral of the errors in function values and gradients as in the training losses of Sobolev training and PINNs (Czarnecki et al., 2017; Mishra & Molinaro, 2023), and hence allows inaccuracy on a small part of the domain but still controls the gradient. In particular, it is unknown whether the minimum width for the $W^{1,p}$ norm is larger than that for the L^p norm, and whether it increases as p grows.

1.1 Summary of results

We study the minimum width $w_{\min}$ of networks that can approximate any function from $[0,1]^{d_{\text{in}}}$ to $\mathbb{R}^{d_{\text{out}}}$ in the $W^{1,p}$ norm with a sufficiently large depth. Specifically, we consider leaky-ReLU-like activation functions (Definition 2.1). Our first result is as follows.

Theorem 3.1 (Informal)**.** If $d_{\text{in}} \geq 2$, $d_{\text{out}} > 3d_{\text{in}}/2$ and $1 \leq p \leq d_{\text{out}} - d_{\text{in}}$, then $w_{\min} = d_{\text{out}}$.

Since $d_{\text{out}} > d_{\text{in}}$, $w_{\min} = d_{\text{out}}$ in Theorem 3.1 coincides with the minimum width $\max\{d_{\text{in}}, d_{\text{out}}, 2\}$ for the L^p norm (see Table 1). Namely, if $p \leq d_{\text{out}} - d_{\text{in}}$, approximating gradients in $W^{1,p}$ does not require any additional width compared to approximating function values in L^p. We note that the minimum width for the $W^{1,p}$ norm is trivially at least that for the L^p norm and at most that for the $W^{1,\infty}$ norm, and is non-decreasing in p. On the other hand, if $d_{\text{out}} \leq 2d_{\text{in}}$, the minimum width for the $W^{1,\infty}$ norm is at least $d_{\text{out}} + 1$, since the lower bound $d_{\text{out}} + 1$ for the L^∞ norm (Kim et al., 2024) also holds for the $W^{1,\infty}$ norm. Hence, Theorem 3.1 shows that the $W^{1,p}$ norm and the $W^{1,\infty}$ norm are separated by the minimum width if $3d_{\text{in}}/2 < d_{\text{out}} \leq 2d_{\text{in}}$ and $p \leq d_{\text{out}} - d_{\text{in}}$.

This separation arises for a reason similar to that underlying the separation between the L^p norm and the L^∞ norm. Namely, as an L^p approximator is allowed to be inaccurate on a small part of the domain, a $W^{1,p}$ approximator is allowed to have a large derivative on a small part of the domain, whereas a $W^{1,\infty}$ approximator is not. In the proof of Theorem 3.1, we approximate a target function by an embedding by moving its double points, i.e., pairs of distinct inputs with the same output, apart within a small neighborhood. Such a perturbation requires a large derivative in the neighborhood. However, its $W^{1,p}$ cost vanishes as the neighborhood shrinks if $p \leq d_{\text{out}} - d_{\text{in}}$, where $d_{\text{out}} - d_{\text{in}}$ is the codimension of the set of double points, and the condition $d_{\text{out}} > 3d_{\text{in}}/2$ ensures that this set is well-behaved (e.g., a generic map has no triple points).

We next characterize the minimum width for all finite p in the case $d_{\text{out}} = 2d_{\text{in}} - 1$. In this case, Theorem 3.1 gives $w_{\min} = 2d_{\text{in}} - 1$ for $p \leq d_{\text{in}} - 1$, and what remains is how the minimum width changes for $p > d_{\text{in}} - 1$. As we describe below, our characterization reveals the separation between the $W^{1,p}$ norm and the L^p norm, and how the minimum width changes with p.

Theorem 3.2 (Informal)**.** If $d_{\text{out}} = 2d_{\text{in}} - 1$ and $d_{\text{in}} \geq 3$, then

$$w_{\min} = \begin{cases} 2d_{\text{in}} - 1 & \text{if } 1 \leq p \leq d_{\text{in}} - 1, \\ 2d_{\text{in}} & \text{if } d_{\text{in}} - 1 < p \leq d_{\text{in}}, \\ 2d_{\text{in}} & \text{if } p > d_{\text{in}} \text{ and } d_{\text{in}} \text{ is even}, \\ 2d_{\text{in}} + 1 & \text{if } p > d_{\text{in}} \text{ and } d_{\text{in}} \text{ is odd}. \end{cases}$$

Theorem 3.2 implies the following. If $p \leq d_{\text{in}} - 1$, the minimum width coincides with that for the L^p norm, which agrees with Theorem 3.1 since $d_{\text{out}} - d_{\text{in}} = d_{\text{in}} - 1$. If $p > d_{\text{in}} - 1$, the minimum width is strictly larger than that for the L^p norm, i.e., the L^p norm and the $W^{1,p}$ norm are separated by the minimum width. Furthermore, if $p > d_{\text{in}}$, the minimum width coincides with that for the $W^{1,\infty}$ norm (Hwang, 2025b). In particular, if d_{in} is odd and $d_{\text{in}} - 1 < p \leq d_{\text{in}}$, the minimum width lies strictly between those for the L^p norm and the $W^{1,\infty}$ norm. Hence, unlike that for the L^p norm, the minimum width for the $W^{1,p}$ norm increases with p, once when p exceeds $d_{\text{in}} - 1$ and once more when p exceeds d_{in} if d_{in} is odd, and depends on the parity of d_{in} (see Table 1).

We briefly explain the sources of these phenomena (see Section 4 for details). Recall that a network of width ℓ is an approximation of a projection of an embedding into $\mathbb{R}^\ell$.

- **Separation from L^p.** If $p > d_{\text{in}} - 1$, then $W^{1,p}$ convergence is uniform on suitably chosen $(d_{\text{in}} - 1)$-dimensional spheres. Hence, a $W^{1,p}$ approximator must preserve how a target function links two such spheres, which an embedding into $\mathbb{R}^{2d_{\text{in}}-1}$ cannot do.
- **Transitions at finite p.** The thresholds $d_{\text{in}} - 1$ and d_{in} are the codimensions of the sets of double points of generic maps into $\mathbb{R}^{2d_{\text{in}}-1}$ and $\mathbb{R}^{2d_{\text{in}}}$, respectively. As in the proof of Theorem 3.1, double points can be removed at a vanishing $W^{1,p}$ cost if and only if p is at most this codimension.
- **Parity of d_{in}.** If $p > d_{\text{in}}$, then a network of width $2d_{\text{in}}$ must separate the two inputs of every double point using its additional coordinate. This is possible if d_{in} is even but not if d_{in} is odd, since such a separation fails only if the double points form a closed curve along which the two inputs of each pair are exchanged, and such a curve can exist only if d_{in} is odd.

Finally, we derive upper and lower bounds on the minimum width for general d_{in} and d_{out} from Theorems 3.1 and 3.2 by adding or removing output coordinates (Corollary 3.3). In particular, the separation between the $W^{1,p}$ norm and the L^p norm is not specific to the case $d_{\text{out}} = 2d_{\text{in}} - 1$. Namely, for any $d_{\text{in}} \geq 3$ and $d_{\text{in}} \leq d_{\text{out}} \leq 2d_{\text{in}} - 1$, the minimum width for the $W^{1,p}$ norm is strictly larger than that for the L^p norm if $p > \max\{2, \lceil (d_{\text{out}} - 1)/2 \rceil\}$. We also extend our bounds to general input and output dimensions (Corollary 3.3), and non-polynomial $C^{1,1}$ activation functions (Theorem 3.4).

2 PROBLEM SETUP

2.1 NOTATION

We use d_{in} and d_{out} to denote the input and output dimensions, and $I := [0,1]^{d_{\text{in}}}$ to denote the input domain. For $1 \le p < \infty$ and a measurable $f : I \to \mathbb{R}^{d_{\text{out}}}$, we define

$$\|f\|_{L^p(I)} := \Big(\int_I \|f(x)\|^p \, dx \Big)^{1/p}, \qquad \|f\|_{L^\infty(I)} := \operatorname*{ess\,sup}_{x \in I} \|f(x)\|,$$

where $\|\cdot\|$ denotes the Euclidean norm. For f with a weak derivative Df, we also define

$$\|f\|_{W^{1,p}(I)} := \Big(\|f\|_{L^p(I)}^p + \|Df\|_{L^p(I)}^p \Big)^{1/p}, \qquad \|f\|_{W^{1,\infty}(I)} := \max\big\{ \|f\|_{L^\infty(I)}, \|Df\|_{L^\infty(I)} \big\},$$

where the norm of $Df(x)$ is the Frobenius norm. We denote the corresponding function spaces by $L^p(I; \mathbb{R}^{d_{\text{out}}})$ and $W^{1,p}(I; \mathbb{R}^{d_{\text{out}}})$ for $1 \le p \le \infty$, where the weak derivative is taken on the interior of I. For the L^∞ norm and the $W^{1,\infty}$ norm, we consider the target classes $C(I; \mathbb{R}^{d_{\text{out}}})$ and $C^1(I; \mathbb{R}^{d_{\text{out}}})$, respectively, i.e., the closures of smooth functions under these norms, as in prior work.

A map is smooth on the closed cube I if it admits a smooth extension to an open neighborhood of I. An *embedding* of I into $\mathbb{R}^\ell$ is an injective immersion, i.e., an injective C^1 map with an injective derivative at every point. Since I is compact, an embedding is a homeomorphism onto its image.

2.2 NETWORKS AND MINIMUM WIDTH

Let $\Sigma \subset C^1(\mathbb{R})$ be a set of activation functions. A network of depth L from $\mathbb{R}^{d_{\text{in}}}$ to $\mathbb{R}^{d_{\text{out}}}$ is a function of the form

$$N = A_L \circ S_{L-1} \circ A_{L-1} \circ \cdots \circ S_1 \circ A_1, \tag{2.1}$$

where $A_i : \mathbb{R}^{n_{i-1}} \to \mathbb{R}^{n_i}$ is an affine map with $n_0 = d_{\text{in}}$ and $n_L = d_{\text{out}}$, and $S_i : \mathbb{R}^{n_i} \to \mathbb{R}^{n_i}$ applies activation functions coordinate-wise, i.e., $(S_i(z))_j = \sigma_{i,j}(z_j)$ for some $\sigma_{i,j} \in \Sigma$. We note that activation functions may vary across neurons and layers. We call $\max_{1 \le i < L} n_i$ the width of N and denote by $\mathcal{N}^{\Sigma}_{d_{\text{in}}, d_{\text{out}}, w}$ the set of all networks of the form given in Eq. (2.1) with a width of at most w and arbitrary depth L.

For a normed space $\mathcal{X}$ of functions from I to $\mathbb{R}^{d_{\text{out}}}$, we define the *minimum width* for universal approximation in $\mathcal{X}$ as

$$w^{\mathcal{X}}_{\min}(d_{\text{in}}, d_{\text{out}}; \Sigma) := \min\big\{ w \in \mathbb{N} : \mathcal{N}^{\Sigma}_{d_{\text{in}}, d_{\text{out}}, w} \text{ is dense in } \mathcal{X} \big\}.$$

We use $w^p_{\min}$, $w^\infty_{\min}$, $w^{1,p}_{\min}$, and $w^{1,\infty}_{\min}$ for the cases $\mathcal{X} = L^p(I; \mathbb{R}^{d_{\text{out}}})$, $C(I; \mathbb{R}^{d_{\text{out}}})$ with the L^∞ norm, $W^{1,p}(I; \mathbb{R}^{d_{\text{out}}})$, and $C^1(I; \mathbb{R}^{d_{\text{out}}})$ with the $W^{1,\infty}$ norm, respectively. Throughout this paper, $1 \le p < \infty$ unless stated otherwise, and $w_{\min}$ in Section 1 denotes $w^{1,p}_{\min}(d_{\text{in}}, d_{\text{out}}; \Sigma)$.

Since C^1 functions are dense in $W^{1,p}(I; \mathbb{R}^{d_{\text{out}}})$ and $\|f\|_{L^p(I)} \le \|f\|_{W^{1,p}(I)} \le 2^{1/p - 1/q} \|f\|_{W^{1,q}(I)} \le 2^{1/p} \|f\|_{W^{1,\infty}(I)}$ for $p \le q$ by Hölder's inequality, we have

$$w^p_{\min} \le w^{1,p}_{\min} \le w^{1,q}_{\min} \le w^{1,\infty}_{\min} \qquad \text{for } 1 \le p \le q < \infty. \tag{2.2}$$

Likewise, $w^\infty_{\min} \le w^{1,\infty}_{\min}$. Hence, lower bounds on $w^p_{\min}$ and $w^\infty_{\min}$ also hold for $w^{1,p}_{\min}$ and $w^{1,\infty}_{\min}$, and upper bounds on $w^{1,\infty}_{\min}$ also hold for $w^{1,p}_{\min}$ and $w^\infty_{\min}$ (see Table 1).

2.3 ACTIVATION FUNCTIONS

We consider the following class of activation functions.

Definition 2.1. We say that $\Sigma \subset C^1(\mathbb{R})$ is *leaky*-RELU-*like* if every $\sigma \in \Sigma$ is non-decreasing and for every $\beta > 0$ there exists $\sigma_\beta \in \Sigma$ such that

$$\lim_{t \to \infty} \sigma'_\beta(t) = 1, \qquad \lim_{t \to -\infty} \sigma'_\beta(t) = \beta, \qquad \lim_{\beta \to 1} \sup_{t \in \mathbb{R}} |\sigma'_\beta(t) - 1| = 0. \tag{2.3}$$

Namely, Σ contains smooth non-decreasing variants of leaky-ReLU with all slopes $\beta > 0$ in the negative part. Examples include the leaky-softplus variants $\sigma_\beta(t) = \beta t + (1-\beta)\log(1+e^t)$ (Chan et al., 2023) and $\sigma_\beta(t) = \log(1+e^t) - \log(1+e^{-\beta t})$ (Dey et al., 2022), normalized bendable linear units $\sigma_\beta(t) = \frac{1+\beta}{2}t + \frac{1-\beta}{2}(\sqrt{1+t^2}-1)$ (Godfrey, 2019), and leaky SmeLU $\sigma_\beta(t) = \beta(t+h) + \frac{1-\beta}{4h}\left((t+h)_+^2 - (t-h)_+^2\right)$ (Shamir et al., 2020). The asymptotic conditions (2.3) slightly strengthen those of Hwang (2025b, Definition 3.3) to address a technical gap, and monotonicity is an additional requirement used to transfer geometric results to networks (Proposition 2.2). We note that our results are not restricted to leaky-ReLU-like activation functions. Specifically, for any set of non-polynomial $C^{1,1}$ increasing activation functions, our lower bounds hold as they are, and our upper bounds hold with a width increase of one (Theorem 3.4).

2.4 Reduction to a geometric problem

Following Hwang (2025b), we reduce the minimum width problem to a geometric one. For $\ell \geq \max\{d_{\text{in}}, d_{\text{out}}\}$, let $\iota_{d_{\text{in}},\ell}(x) := (x, 0_{\ell-d_{\text{in}}})$ be the canonical inclusion of $\mathbb{R}^{d_{\text{in}}}$ into $\mathbb{R}^\ell$ and $\pi_{\ell,d_{\text{out}}}(z_1,\dots,z_\ell) := (z_1,\dots,z_{d_{\text{out}}})$ be the projection onto the first d_{out} coordinates. We define the class of projected diffeomorphisms and the *geometric width* $\Omega_{1,p}(d_{\text{in}}, d_{\text{out}})$ as

$$\mathcal{A}^\ell_{d_{\text{in}},d_{\text{out}}} := \left\{\pi_{\ell,d_{\text{out}}} \circ \Phi \circ \iota_{d_{\text{in}},\ell}|_I : \Phi \in \text{Diff}^\infty(\mathbb{R}^\ell)\right\}, \tag{2.4}$$

$$\Omega_{1,p}(d_{\text{in}}, d_{\text{out}}) := \min\left\{\ell \geq \max\{d_{\text{in}}, d_{\text{out}}\} : \mathcal{A}^\ell_{d_{\text{in}},d_{\text{out}}} \text{ is dense in } W^{1,p}(I;\mathbb{R}^{d_{\text{out}}})\right\}, \tag{2.5}$$

where $\text{Diff}^\infty(\mathbb{R}^\ell)$ denotes the set of smooth diffeomorphisms. Since every smooth embedding of I into $\mathbb{R}^\ell$ with $\ell > d_{\text{in}}$ extends to a diffeomorphism of $\mathbb{R}^\ell$, $\mathcal{A}^\ell_{d_{\text{in}},d_{\text{out}}}$ coincides with the class of projected embeddings:

$$\mathcal{A}^\ell_{d_{\text{in}},d_{\text{out}}} = \left\{\pi_{\ell,d_{\text{out}}} \circ e : e \in \text{Emb}^\infty(I,\mathbb{R}^\ell)\right\}. \tag{2.6}$$

Here, $\text{Emb}^\infty(I,\mathbb{R}^\ell)$ is the set of smooth embeddings. Thus, $\Omega_{1,p}(d_{\text{in}}, d_{\text{out}})$ is the smallest ℓ such that any function in $W^{1,p}(I;\mathbb{R}^{d_{\text{out}}})$ can be approximated by a projection of an embedding of I into $\mathbb{R}^\ell$.

The following proposition shows that the minimum width for the $W^{1,p}$ norm coincides with the geometric width.

Proposition 2.2. *Let Σ be leaky-ReLU-like, $1 \leq p < \infty$, and $\ell \geq \max\{d_{\text{in}}, d_{\text{out}}\}$. Then, the closures of $\mathcal{N}^\Sigma_{d_{\text{in}},d_{\text{out}},\ell}$ and $\mathcal{A}^\ell_{d_{\text{in}},d_{\text{out}}}$ in $W^{1,p}(I;\mathbb{R}^{d_{\text{out}}})$ coincide. In particular, we have $w^{1,p}_{\min}(d_{\text{in}}, d_{\text{out}}; \Sigma) = \Omega_{1,p}(d_{\text{in}}, d_{\text{out}})$.*

Proposition 2.2 is the $W^{1,p}$ analog of Hwang (2025b, Theorem 4.6), which is stated for the $W^{1,\infty}$ norm. The proof of Proposition 2.2 is deferred to Appendix B.1. By Proposition 2.2, it suffices to analyze $\Omega_{1,p}(d_{\text{in}}, d_{\text{out}})$, and all results in Section 3 are stated for both $w^{1,p}_{\min}$ and $\Omega_{1,p}$.

3 Main results

In this section, we present our main results on the minimum width for the $W^{1,p}$ norm. By Proposition 2.2, all results are stated for both $w^{1,p}_{\min}(d_{\text{in}}, d_{\text{out}}; \Sigma)$ with leaky-ReLU-like Σ and the geometric width $\Omega_{1,p}(d_{\text{in}}, d_{\text{out}})$. Proof sketches are given in Section 4, and full proofs are deferred to the appendix.

3.1 The case of $d_{\text{out}} > 3d_{\text{in}}/2$ and $1 \leq p \leq d_{\text{out}} - d_{\text{in}}$

Our first result identifies the minimum width when the output dimension is larger than $3d_{\text{in}}/2$ and p is at most the difference between the output and input dimensions. The proof is given in Appendix C.

Theorem 3.1. *Let Σ be leaky-ReLU-like, $d_{\text{in}} \geq 2$, $d_{\text{out}} > 3d_{\text{in}}/2$, and $1 \leq p \leq d_{\text{out}} - d_{\text{in}}$. Then,*

$$w^{1,p}_{\min}(d_{\text{in}}, d_{\text{out}}; \Sigma) = \Omega_{1,p}(d_{\text{in}}, d_{\text{out}}) = d_{\text{out}}.$$

Since $w^p_{\min}(d_{\text{in}}, d_{\text{out}}; \Sigma) \geq \max\{d_{\text{in}}, d_{\text{out}}\} = d_{\text{out}}$ (Cai, 2023) and $w^{1,p}_{\min} \geq w^p_{\min}$ by Eq. (2.2), the lower bound in Theorem 3.1 is trivial, and the content of Theorem 3.1 is the upper bound. To establish

this upper bound, we use a triangulation to organize deformations that remove the singularities of a generic smooth target with arbitrarily small $W^{1,p}$ error when $p \le d_{\text{out}} - d_{\text{in}}$.

Theorem 3.1 has the following implications.

- Since $w^p_{\min} = d_{\text{out}}$ for $d_{\text{out}} \ge \max\{d_{\text{in}}, 2\}$ (Shin et al., 2025), $w^{1,p}_{\min} = w^p_{\min}$ if $d_{\text{out}} > 3d_{\text{in}}/2$ and $p \le d_{\text{out}} - d_{\text{in}}$. Namely, approximating gradients in $W^{1,p}$ does not require any additional width compared to approximating function values in L^p in this regime.
- If $d_{\text{in}} < d_{\text{out}} \le 2d_{\text{in}}$, then $w^{1,\infty}_{\min} \ge w^{\infty}_{\min} \ge d_{\text{out}} + 1$ by the L^∞ lower bound of Kim et al. (2024) and Eq. (2.2). Hence, if $3d_{\text{in}}/2 < d_{\text{out}} \le 2d_{\text{in}}$ and $p \le d_{\text{out}} - d_{\text{in}}$, then $w^{1,p}_{\min} < w^{\infty}_{\min}, w^{1,\infty}_{\min}$. Namely, approximating both the values and gradients in $W^{1,p}$ requires a smaller width than approximating the values in L^∞.

3.2 The case of $d_{\text{out}} = 2d_{\text{in}} - 1$ and $1 \le p < \infty$

Our second result characterizes the minimum width for all p when $d_{\text{out}} = 2d_{\text{in}} - 1$ and $d_{\text{in}} \ge 3$.

Theorem 3.2. *Let Σ be leaky-*ReLU*-like, $d_{\text{in}} \ge 3$, and $1 \le p < \infty$. Then,*

$$w^{1,p}_{\min}(d_{\text{in}}, 2d_{\text{in}} - 1; \Sigma) = \Omega_{1,p}(d_{\text{in}}, 2d_{\text{in}} - 1) = \begin{cases} 2d_{\text{in}} - 1 & \text{if } 1 \le p \le d_{\text{in}} - 1, \\ 2d_{\text{in}} & \text{if } d_{\text{in}} - 1 < p \le d_{\text{in}}, \\ 2d_{\text{in}} & \text{if } p > d_{\text{in}} \text{ and } d_{\text{in}} \text{ is even}, \\ 2d_{\text{in}} + 1 & \text{if } p > d_{\text{in}} \text{ and } d_{\text{in}} \text{ is odd}. \end{cases} \tag{3.1}$$

The first case of Eq. (3.1) is the special case of Theorem 3.1 with $d_{\text{out}} = 2d_{\text{in}} - 1$, since $2d_{\text{in}} - 1 > 3d_{\text{in}}/2$ for $d_{\text{in}} \ge 3$ and $d_{\text{out}} - d_{\text{in}} = d_{\text{in}} - 1$. The remaining cases consist of the upper bound $2d_{\text{in}}$ for $p \le d_{\text{in}}$, the lower bound $2d_{\text{in}}$ for $p > d_{\text{in}} - 1$, and the upper and lower bounds $2d_{\text{in}} + \mathbf{1}_{d_{\text{in}}=\text{odd}}$ for $p > d_{\text{in}}$. Here, the upper bound for $p > d_{\text{in}}$ follows from $w^{1,\infty}_{\min}(d_{\text{in}}, 2d_{\text{in}} - 1; \Sigma) = 2d_{\text{in}} + \mathbf{1}_{d_{\text{in}}=\text{odd}}$ (Hwang, 2025b) and Eq. (2.2), so the content of Theorem 3.2 for $p > d_{\text{in}}$ is the lower bound. The proof of Theorem 3.2 is given in Appendix A. The general proof idea follows from a phase transition in the approximation behavior in the $W^{1,p}$ norm: for small p, one can untangle self-intersections at a very small $W^{1,p}$ approximation cost, whereas for large p, Sobolev embeddings enforce uniform convergence and thus expose some fundamental topological limitations.

Theorem 3.2 has the following implications.

- If $p > d_{\text{in}} - 1$, then $w^{1,p}_{\min} \ge 2d_{\text{in}} > 2d_{\text{in}} - 1 = w^p_{\min}$ (Shin et al., 2025), i.e., the $W^{1,p}$ norm and the L^p norm are separated by the minimum width. Combined with the first case of Eq. (3.1), this shows that the threshold $d_{\text{out}} - d_{\text{in}}$ in Theorem 3.1 is tight when $d_{\text{out}} = 2d_{\text{in}} - 1$.
- If $p > d_{\text{in}}$, then $w^{1,p}_{\min} = w^{1,\infty}_{\min}$ (Hwang, 2025b). On the other hand, if $d_{\text{in}} - 1 < p \le d_{\text{in}}$ and d_{in} is odd, then $w^{1,p}_{\min} = 2d_{\text{in}}$ lies strictly between $w^p_{\min} = 2d_{\text{in}} - 1$ and $w^{1,\infty}_{\min} = 2d_{\text{in}} + 1$.
- Unlike the minimum width for the L^p norm that does not depend on p, the minimum width for the $W^{1,p}$ norm is a non-decreasing step function of p that increases at $p = d_{\text{in}} - 1$, and once more at $p = d_{\text{in}}$ if d_{in} is odd. In particular, it depends on the parity of d_{in} for $p > d_{\text{in}}$, which is inherited from the parity dependence of $w^{1,\infty}_{\min}$ (Hwang, 2025b), whereas it does not depend on the parity for $p \le d_{\text{in}}$. Here, the thresholds $d_{\text{in}} - 1$ and d_{in} are the codimensions of the sets of double points of generic maps from I to $\mathbb{R}^{2d_{\text{in}}-1}$ and to $\mathbb{R}^{2d_{\text{in}}}$, respectively (see Section 4).

3.3 Extension to general input and output dimensions

We can extend Theorem 3.2 to general input and output dimensions by adding output coordinates for the upper bound and removing them for the lower bound.

Corollary 3.3. *Let Σ be leaky-*ReLU*-like, $d_{\text{in}} \ge 3$, and $1 \le p < \infty$. Then, for any $d_{\text{out}} \ge 1$,*

$$w^{1,p}_{\min}(d_{\text{in}}, d_{\text{out}}; \Sigma) \le \begin{cases} \min\{d_{\text{in}} + d_{\text{out}}, \max\{d_{\text{out}}, 2d_{\text{in}} - 1\}\} & \text{if } p \le d_{\text{in}} - 1, \\ \min\{d_{\text{in}} + d_{\text{out}}, \max\{d_{\text{out}}, 2d_{\text{in}}\}\} & \text{if } d_{\text{in}} - 1 < p \le d_{\text{in}}, \\ \min\{d_{\text{in}} + d_{\text{out}}, \max\{d_{\text{out}}, 2d_{\text{in}} + 1\}\} & \text{if } p > d_{\text{in}}. \end{cases} \tag{3.2}$$

Furthermore, for any $d_{\text{in}} \le d_{\text{out}} \le 2d_{\text{in}} - 1$ *and* $j^* := \max\{3, \lceil (d_{\text{out}} + 1)/2 \rceil\}$,

$$w^{1,p}_{\min}(d_{\text{in}}, d_{\text{out}}; \Sigma) = \Omega_{1,p}(d_{\text{in}}, d_{\text{out}}) \ge \begin{cases} d_{\text{out}} + 1 & \text{if } p > j^* - 1, \\ d_{\text{out}} + 2 & \text{if } p > j \text{ for some odd } j \in [j^*, d_{\text{in}}]. \end{cases} \tag{3.3}$$

The upper bound in Eq. (3.2) follows from Theorem 3.2 by approximating $(f, 0) \in W^{1,p}(I; \mathbb{R}^{2d_{\text{in}}-1})$ and discarding the last output coordinates if $d_{\text{out}} \le 2d_{\text{in}} - 1$, and from the $W^{1,\infty}$ upper bound $\min\{d_{\text{in}} + d_{\text{out}}, \max\{2d_{\text{in}} + 1, d_{\text{out}}\}\}$ (Hwang, 2025b) and Eq. (2.2) otherwise. For the specific case $d_{\text{out}} = 2d_{\text{in}}$ and $p \le d_{\text{in}}$, using the $W^{1,\infty}$ result naively only gives us $2d_{\text{in}} + 1$. Getting $2d_{\text{in}}$ would require an extra density argument.

The lower bound in Eq. (3.3) follows from Theorem 3.2 by passing to lower-dimensional domains. However, one needs to be careful in doing so, as restricting a $W^{1,p}$-convergent sequence of networks to a prescribed boundary face can fail to preserve convergence. Instead, we use Fubini's theorem to select convergent interior slices and adjoin the required output coordinates directly to the underlying geometric embedding. The complete proof of Corollary 3.3 is given in Appendix B.

Since $w^p_{\min} = d_{\text{out}}$ for $d_{\text{out}} \ge \max\{d_{\text{in}}, 2\}$ (Shin et al., 2025), Eq. (3.3) shows that the separation between the $W^{1,p}$ norm and the L^p norm is not specific to the case $d_{\text{out}} = 2d_{\text{in}} - 1$: for any $d_{\text{in}} \ge 3$ and $d_{\text{in}} \le d_{\text{out}} \le 2d_{\text{in}} - 1$, $w^{1,p}_{\min}(d_{\text{in}}, d_{\text{out}}; \Sigma) > w^p_{\min}(d_{\text{in}}, d_{\text{out}}; \Sigma)$ if $p > \max\{2, \lceil (d_{\text{out}} - 1)/2 \rceil\}$.

3.4 Other activation functions

Our bounds extend to a fixed non-polynomial $C^{1,1}$ activation function at the cost of at most one additional hidden neuron per layer.

Theorem 3.4. *Let* $d_{\text{in}}, d_{\text{out}} \in \mathbb{N}$, $1 \le p < \infty$, *and let* $\Sigma \subset C^1(\mathbb{R})$ *contain a non-polynomial activation function* $\sigma \in C^{1,1}(\mathbb{R})$. *Then,*

$$w^{1,p}_{\min}(d_{\text{in}}, d_{\text{out}}; \Sigma) \le \Omega_{1,p}(d_{\text{in}}, d_{\text{out}}) + 1.$$

In addition, if every activation in Σ *is non-decreasing, then* $\Omega_{1,p}(d_{\text{in}}, d_{\text{out}}) \le w^{1,p}_{\min}(d_{\text{in}}, d_{\text{out}}; \Sigma)$.

Proof. To see this, put $\ell = \Omega_{1,p}(d_{\text{in}}, d_{\text{out}})$ and fix $f \in W^{1,p}(I; \mathbb{R}^{d_{\text{out}}})$ and $\varepsilon > 0$. By definition, there is $g = \pi_{\ell, d_{\text{out}}} \circ \Phi \circ \iota_{d_{\text{in}}, \ell}|_I \in \mathcal{A}^{\ell}_{d_{\text{in}}, d_{\text{out}}}$ with $\|f - g\|_{W^{1,p}(I)} < \varepsilon/2$. By Hwang (2025b, Theorem 4.1), the smooth diffeomorphism Φ can be approximated arbitrarily closely in C^1 on a compact neighborhood of $\iota_{d_{\text{in}}, \ell}(I)$ by a σ-network F of width $\ell + 1$. The affine inclusion and projection can be absorbed into the first and last layers, so $N := \pi_{\ell, d_{\text{out}}} \circ F \circ \iota_{d_{\text{in}}, \ell}$ has the same width. The chain rule gives $\|N - g\|_{W^{1,p}(I)} < \varepsilon/2$ for a sufficiently accurate approximation, and hence $\|N - f\|_{W^{1,p}(I)} < \varepsilon$. The lower bound follows from the proof of Propositions B.1 and 2.2, whose argument for this direction requires only C^1 regularity and monotonicity. □

Consequently, every geometric upper bound in Theorems 3.1 and 3.2 and Corollary 3.3 gives a network upper bound with a width increase of one. For example, if $d_{\text{in}} \ge 2$, $d_{\text{out}} > 3d_{\text{in}}/2$, and $1 \le p \le d_{\text{out}} - d_{\text{in}}$, then $w^{1,p}_{\min}(d_{\text{in}}, d_{\text{out}}; \Sigma) \le d_{\text{out}} + 1$.

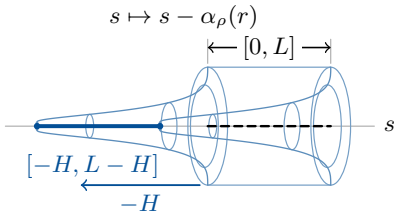
4 Proof Sketches

In this section, we give brief sketches of the proofs. We begin with input and output dimensions $(d_{\text{in}}, 2d_{\text{in}} - 1)$, where the singular set is one-dimensional. This case provides a useful introduction to the general construction and, in particular, makes the lower-bound obstruction easier to see.

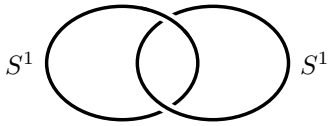
4.1 The case of $d_{\text{out}} = 2d_{\text{in}} - 1$ and $1 \le p < \infty$

We note that the transfer result above allows us to work entirely with the geometric width $\Omega_{1,p}(d_{\text{in}}, 2d_{\text{in}} - 1)$, so the problem is purely a geometric. This problem is, in short:

How many additional target dimensions are needed
for embeddings to be dense in the Sobolev space?



(a) Radial deformation.



(b) Two circles linked once in $\mathbb{R}^3$.

Figure 1: Geometric constructions for the upper and lower bounds. The radial deformation shifts each transverse sphere by $-\alpha_\rho(r)$, moving the center arc to $[-H, L - H]$ while fixing the outer boundary $r = R_0$. The linked circles illustrate the lower-bound obstruction.

Thus, we start from a smooth target function $f : [0,1]^{d_{\text{in}}} \to \mathbb{R}^{2d_{\text{in}}-1}$ and ask when it can be approximated by the projection of an embedding. There is an immediate problem: f need not be one-to-one. To deal with this, we call $x \neq y$ a *double-point pair* when $f(x) = f(y)$. The crucial points regarding the output dimension are the following: For a generic map into $\mathbb{R}^{2d_{\text{in}}-1}$, the source points participating in such pairs form one-dimensional curves; after adding one ambient coordinate, a generic map into $\mathbb{R}^{2d_{\text{in}}}$ has only finitely many double point pairs; and in $\mathbb{R}^{2d_{\text{in}}+1}$, a generic map has no double points (Whitney, 1944, Section 3 and Section 22).

Therefore, the two thresholds in the main theorem are determined by the number of source directions available to make the $W^{1,p}$ cost of removing self-intersecting pairs arbitrarily small. In this regard, we introduce the following lemma.

Lemma 4.1. *Let $q \geq 2$ and fix $H > 0$. There are smooth functions $a_\rho : B^q_R \to [0, H]$, supported in a fixed ball such that $a_\rho(0) = H$, and for every $1 \leq p \leq q$, $\|a_\rho\|_{W^{1,p}(B^q_R)} \to 0$.*

Proof Sketch. Fix $0 < R_0 < R$. Choose a smooth cutoff $\beta : \mathbb{R} \to [0,1]$ equal to one on $(-\infty, 0]$ and zero on $[1, \infty)$, and put

$$a_\rho(z) = H\beta\left(\frac{\log(|z|/\rho)}{\log(R_0/\rho)}\right) \quad (\rho < |z| < R_0),$$

extended by H for $|z| \leq \rho$ and by zero for $|z| \geq R_0$. Then $a_\rho \to 0$ in $W^{1,p}$. □

We now explain the basic idea of using Lemma 4.1 to remove self-intersections. For example, consider two d_{in}-planes ($d_{\text{in}} \geq 3$) parameterized by

$$F_1(s, z) = (s, z, 0), \qquad F_2(t, w) = (t, 0, w), \qquad z, w \in \mathbb{R}^{d_{\text{in}}-1}.$$

We use their bounded pieces with $0 \leq s, t \leq L$ and $|z|, |w| < R$, which intersect along $z = w = 0$. Take the radial cutoff $a_\rho(z) = \alpha_\rho(|z|)$ with $\alpha_\rho(0) = H > L$. Choose a smooth cutoff $\chi = 1$ on $[0, L]$, supported in a longer interval, with $H\|\chi'\|_\infty < 1$, and set

$$T_\rho(s, z) = \big(s - \chi(s)\alpha_\rho(|z|), z\big), \qquad F_{1,\rho} = F_1 \circ T_\rho.$$

The derivative condition makes T_ρ invertible. If $z = 0$, the new first coordinate is $s - H < 0$, outside the second piece; if $z \neq 0$, the unchanged middle coordinate prevents an intersection. Thus, the pieces become disjoint.

For $1 \leq p \leq d_{\text{in}} - 1$, Lemma 4.1 makes the Sobolev cost of this radial detour arbitrarily small. Applying such detours along the selected arcs allows us to avoid self-intersections, which provides one direction of the following lemma.

Lemma 4.2 (First Threshold Phenomenon)**.** *For $d_{\text{in}} \geq 3$, the set of smooth embeddings $[0,1]^{d_{\text{in}}} \to \mathbb{R}^{2d_{\text{in}}-1}$ is dense in $W^{1,p}$ if and only if $p \leq d_{\text{in}} - 1$.*

For the lower bound when $p > d_{\text{in}} - 1$, choose a target map that sends two $(d_{\text{in}} - 1)$-spheres to linked images. After selecting suitable nearby radii, a sufficiently small $W^{1,p}$ error implies uniform closeness on both spheres by the Sobolev embedding:

$$W^{1,p}(S^{d_{\text{in}}-1}) \hookrightarrow C^0(S^{d_{\text{in}}-1}) \qquad (p > d_{\text{in}} - 1).$$

Their linking is, therefore, preserved. However, if the approximating map were an embedding, the first sphere would bound the image of its source ball, which is disjoint from the second sphere, forcing the linking to be zero. Hence, width $2d_{\text{in}} - 1$ is impossible above $p = d_{\text{in}} - 1$.

In $\mathbb{R}^{2d_{\text{in}}}$, generic self-intersections are isolated points. A similar detour uses a radial cutoff in all d_{in} source variables, giving a vanishing $W^{1,p}$ cost for $p \leq d_{\text{in}}$. Applying it to a generic perturbation of $(f, 0)$ yields $\Omega_{1,p}(d_{\text{in}}, 2d_{\text{in}} - 1) = 2d_{\text{in}}$ for $d_{\text{in}} - 1 < p \leq d_{\text{in}}$, which gives the following lemma:

Lemma 4.3 (Second Threshold Phenomenon)**.** *For $d_{\text{in}} \geq 2$ and $p \leq d_{\text{in}}$, the set of smooth embeddings $[0,1]^{d_{\text{in}}} \to \mathbb{R}^{2d_{\text{in}}}$ is dense in $W^{1,p}$.*

Thus, for $d_{\text{in}} - 1 < p \leq d_{\text{in}}$, we have $\Omega_{1,p}(d_{\text{in}}, 2d_{\text{in}} - 1) = 2d_{\text{in}}$.

For $p > d_{\text{in}}$, these local changes can no longer have vanishing Sobolev cost. At width $2d_{\text{in}}$, the extra coordinate must instead separate every double point pair without substantially changing the first $2d_{\text{in}} - 1$ coordinates. This is described by the following lemma.

Lemma 4.4. *The following statements are true:*

- *For even d_{in}, every smooth $f : [0,1]^{d_{\text{in}}} \to \mathbb{R}^{2d_{\text{in}}-1}$ can be approximated in C^1 by a map g for which there is a scalar h such that (g, h) is an embedding into $\mathbb{R}^{2d_{\text{in}}}$.*
- *For odd d_{in}, there is a smooth target f and a C^0-neighborhood of f such that no map g in that neighborhood admits such an h.*

For $p > d_{\text{in}}$, Morrey's embedding $W^{1,p}(I) \hookrightarrow C^0(I)$ reduces the lower-bound argument to the uniform obstruction stated in the lemma. The number of additional coordinates needed to overcome this obstruction depends on the parity of d_{in}, reflecting the topological distinction discussed in the uniform approximation setting by Hwang (2025a, Sections 4.4 and 5).

Combining the lemmas gives the four regimes in the main theorem:

- width $2d_{\text{in}} - 1$ while the one-dimensional double-point set can be removed at vanishing cost;
- width $2d_{\text{in}}$ once only isolated double points can be removed this way;
- above $p = d_{\text{in}}$, width $2d_{\text{in}}$ or $2d_{\text{in}} + 1$ according to whether an additional scalar coordinate can globally separate every double-point pair.

4.2 The case of $d_{\text{out}} > 3d_{\text{in}}/2$ and $1 \leq p \leq d_{\text{out}} - d_{\text{in}}$

In the case $(d_{\text{in}}, 2d_{\text{in}} - 1)$, the condition $p \leq d_{\text{in}} - 1$ arises from localizing a deformation in the $d_{\text{in}} - 1$ source directions that are transverse to a double-point curve. More generally, for $3d_{\text{in}}/2 < d_{\text{out}} \leq 2d_{\text{in}}$, the generic singular set has dimension $2d_{\text{in}} - d_{\text{out}}$, leaving $d_{\text{out}} - d_{\text{in}}$ transverse source directions. The same Sobolev suppressing effect therefore gives $p \leq d_{\text{out}} - d_{\text{in}}$.

The remaining difficulty is that the previous construction uses finitely many disjoint arcs, whereas a higher-dimensional singular set need not admit such a decomposition. To overcome this, we use a smooth triangulation to organize the construction into finitely many detours around lower-dimensional cells. We carry out these detours inductively in order of increasing dimension, moving the embedded input cube away from each cell while keeping the previously treated regions fixed. A final deformation near the boundary of the selected singular subset makes the cube disjoint from it, yielding an embedding after composition with f. Localizing each step in the normal directions makes the total $W^{1,p}$ cost arbitrarily small. For $d_{\text{out}} > 2d_{\text{in}}$, generic maps are already embeddings.

5 Conclusion

In this paper, we established exact minimum widths for $W^{1,p}$ universal approximation with finite p in several dimension regimes for leaky-ReLU-like activation functions, and extended our bounds to non-polynomial $C^{1,1}$ activation functions with a width increase of at most one. The results show that the minimum width can depend on both p and the parity of the input dimension, and can differ from the widths required in L^p and $W^{1,\infty}$. While our exact characterizations cover certain ranges of input and output dimensions, we expect that our combination of Sobolev estimates and topological arguments can help close the remaining gaps between the upper and lower bounds. Additionally, extending this analysis to $W^{k,p}$ for $k \geq 2$, under suitable smoothness assumptions on the activations, would clarify how higher derivative requirements affect the minimum width.

AI USE STATEMENT

We used ChatGPT and Claude models in writing this paper. Specifically, we used them to polish the writing and to check the text for accuracy and consistency. We also used them to write proofs. Namely, the core ideas of our proofs, such as tracking singularities via the generic-singularity theorem of Whitney (1944) and the linking number argument for the lower bound in Theorem 3.2, were developed by the authors, and the models turned detailed sketches of the authors into formal arguments and checked their geometric and analytic details. In addition, we used them for ideation in two places. For Theorem 3.2 with odd d_{in} and $d_{\text{in}} - 1 < p \leq d_{\text{in}}$, our initial approach did not resolve this case, and the lemma closing this gap (Section 4) was obtained with the assistance of the models. For the proof of Theorem 3.1 (Appendix C), we used the models to find relevant literature on geometry, and some ideas in the arguments on the spine came from the models. We note that all arguments in this paper were verified by the authors, and the authors take full responsibility for its content.

REPRODUCIBILITY STATEMENT

This paper is theoretical and contains no experiments. All definitions and assumptions are stated in Section 2. All results in Section 3 are stated with their full hypotheses, Section 4 gives proof sketches, and complete proofs are provided in the appendix.

REFERENCES

R. A. Adams and J. J. F. Fournier. *Sobolev Spaces*. Academic Press, second edition, 2003.

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. In *Annual Conference on Neural Information Processing Systems (NeurIPS)*, 2020.

Yongqiang Cai. Achieve the minimum width of neural networks for universal approximation. In *International Conference on Learning Representations (ICLR)*, 2023.

Robin Chan, Sarina Penquitt, and Hanno Gottschalk. Lu-net: Invertible neural networks based on matrix factorization. In *2023 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–10. IEEE, 2023.

George Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematics of control, signals and systems*, 2(4):303–314, 1989.

Wojciech M Czarnecki, Simon Osindero, Max Jaderberg, Grzegorz Swirszcz, and Razvan Pascanu. Sobolev training for neural networks. In *Annual Conference on Neural Information Processing Systems (NeurIPS)*, 2017.

Debojyoti Dey, Bhaskar Mukhoty, and Purushottam Kar. AGGLIO: Global optimization for locally convex functions. In *5th Joint International Conference on Data Science & Management of Data (9th ACM IKDD CODS and 27th COMAD)*, pp. 37–45. ACM, 2022. doi: 10.1145/3493700.3493708.

Luke B. Godfrey. An evaluation of parametric activation functions for deep learning. In *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, pp. 3006–3011. IEEE, 2019. doi: 10.1109/SMC.2019.8913972.

M. Golubitsky and V. Guillemin. *Stable Mappings and Their Singularities*. Springer, 1973.

Boris Hanin and Mark Sellke. Approximating continuous functions by ReLU nets of minimal width. *arXiv preprint arXiv:1710.11278*, 2017.

A. Hatcher. *Algebraic Topology*. Cambridge University Press, 2002.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

M. W. Hirsch. *Differential Topology*. Springer, 1976.

Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5):359–366, 1989.

Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks. *Neural Networks*, 3(5): 551–560, 1990.

J. F. P. Hudson and E. C. Zeeman. On regular neighbourhoods. *Proceedings of the London Mathematical Society*, s3-14(4):719–745, 1964. URL https://doi.org/10.1112/plms/s3-14.4.719.

Geonho Hwang. Minimum width for deep, narrow mlp: A diffeomorphism approach. In *Annual Conference on Neural Information Processing Systems (NeurIPS)*, 2025a.

Geonho Hwang. Optimal minimum width for the universal approximation of continuously differentiable functions by deep narrow MLPs. In *Annual Conference on Neural Information Processing Systems (NeurIPS)*, 2025b.

Jesse Johnson. Deep, skinny neural networks are not universal approximators. In *International Conference on Learning Representations (ICLR)*, 2019.

J. Jost. *Riemannian Geometry and Geometric Analysis*. Springer, sixth edition, 2011. URL https://doi.org/10.1007/978-3-642-21298-7.

John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, et al. Highly accurate protein structure prediction with alphafold. *Nature*, 596(7873):583–589, 2021.

Patrick Kidger and Terry Lyons. Universal approximation with deep narrow networks. In *Conference on Learning Theory (COLT)*, 2020.

Namjun Kim, Chanho Min, and Sejun Park. Minimum width for universal approximation using ReLU networks on compact domain. In *International Conference on Learning Representations (ICLR)*, 2024.

J. M. Lee. *Introduction to Riemannian Manifolds*, volume 176 of *Graduate Texts in Mathematics*. Springer, 2018. URL https://doi.org/10.1007/978-3-319-91755-9.

A. Lerario, C. Meroni, and D. Zuddas. On smooth functions with two critical values. *Revista Matemática Complutense*, 37:907–920, 2024. URL https://doi.org/10.1007/s13163-023-00484-z.

Moshe Leshno, Vladimir Ya Lin, Allan Pinkus, and Shimon Schocken. Multilayer feedforward networks with a nonpolynomial activation function can approximate any function. *Neural networks*, 6(6):861–867, 1993.

Xin Li. Simultaneous approximations of multivariate functions and their derivatives by neural networks with one hidden layer. *Neurocomputing*, 12(4):327–343, 1996.

Zhou Lu, Hongming Pu, Feicheng Wang, Zhiqiang Hu, and Liwei Wang. The expressive power of neural networks: A view from the width. In *Annual Conference on Neural Information Processing Systems (NeurIPS)*, 2017.

Siddhartha Mishra and Roberto Molinaro. Estimates on the generalization error of physics-informed neural networks for approximating PDEs. *IMA Journal of Numerical Analysis*, 43(1):1–43, 2023. doi: 10.1093/imanum/drab093.

R. S. Palais. Extending diffeomorphisms. *Proceedings of the American Mathematical Society*, 11: 274–277, 1960.

Sejun Park, Chulhee Yun, Jaeho Lee, and Jinwoo Shin. Minimum width for universal approximation. In *International Conference on Learning Representations (ICLR)*, 2021.

Allan Pinkus. Approximation theory of the mlp model in neural networks. *Acta numerica*, 8:143–195, 1999.

Maziar Raissi, Paris Perdikaris, and George E Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, 2019.

C. P. Rourke and B. J. Sanderson. *Introduction to Piecewise-Linear Topology*. Springer, 1972.

Gil I. Shamir, Dong Lin, and Lorenzo Coviello. Smooth activations and reproducibility in deep networks. *arXiv preprint arXiv:2010.09931*, 2020.

Jonghyun Shin, Namjun Kim, Geonho Hwang, and Sejun Park. Minimum width for universal approximation using squashable activation functions. In *International Conference on Machine Learning (ICML)*, 2025.

A Vaswani. Attention is all you need. In *Annual Conference on Neural Information Processing Systems (NeurIPS)*, 2017.

C. T. C. Wall. *Differential Topology*, volume 156 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, 2016. URL https://doi.org/10.1017/CBO9781316597835. English text available at https://www.math.auckland.ac.nz/~hekmati/Books/Wall.pdf.

J. H. C. Whitehead. On C^1-complexes. *Annals of Mathematics*, 41(4):809–824, 1940.

H. Whitney. The singularities of a smooth n-manifold in $(2n-1)$-space. *Annals of Mathematics*, 45(2): 247–293, 1944. URL https://webhomes.maths.ed.ac.uk/~v1ranick/papers/singular.pdf.

A Proof of the Sobolev Phase Diagram

Write $I = [0,1]^{d_{\text{in}}}$. Throughout this section, all maps called smooth on a compact set extend smoothly to an open neighborhood of that set; an embedding is injective on I and has an injective differential at every point of I, including its boundary ∂I. Boundary values are not prescribed. We use the Euclidean norm for pointwise values and the Frobenius norm for derivatives.

One should view these results through three relevant ambient dimensions, each carrying a distinct geometric meaning: in dimension $2d_{\text{in}} - 1$, the visible approximant must itself be an embedding; in dimension $2d_{\text{in}}$, one additional scalar coordinate may be used as a height function to distinguish double points; and in dimension $2d_{\text{in}} + 1$, general position provides an embedding directly without this scalar-lifting restriction. The key analytical question is therefore when self-intersections can be removed at arbitrarily small Sobolev cost.

Theorem A.1 (Geometric Sobolev phase diagram)**.** *For $d_{\text{in}} \geq 3$ and $1 \leq p < \infty$,*

$$\Omega_{1,p}(d_{\text{in}}, 2d_{\text{in}} - 1) = \begin{cases} 2d_{\text{in}} - 1, & 1 \leq p \leq d_{\text{in}} - 1, \\ 2d_{\text{in}}, & d_{\text{in}} - 1 < p \leq d_{\text{in}}, \\ 2d_{\text{in}}, & p > d_{\text{in}} \text{ and } d_{\text{in}} \text{ is even}, \\ 2d_{\text{in}} + 1, & p > d_{\text{in}} \text{ and } d_{\text{in}} \text{ is odd}. \end{cases} \tag{A.1}$$

Whitney's singularity theory supplies the generic collision geometry for smooth maps. The proof below establishes both the Sobolev cost of resolving these collisions and the topological obstructions to improving the resulting widths, organized around four regimes:

1. For $1 \leq p \leq d_{\text{in}} - 1$, we move the source cube away from selected double curves and rank-defect points using a radial capacity cutoff in $d_{\text{in}} - 1$ transverse variables (Section A.3).
2. For $p > d_{\text{in}} - 1$, we place linked spherical slices on concentric annuli to rule out approximation by embeddings into $\mathbb{R}^{2d_{\text{in}}-1}$ (Section A.4).
3. For $d_{\text{in}} - 1 < p \leq d_{\text{in}}$, we remove the isolated double points of a generic immersion into $\mathbb{R}^{2d_{\text{in}}}$ by guiding one sheet through the free boundary using a cutoff in all d_{in} source variables (Section A.5).
4. For $p > d_{\text{in}}$, the even-dimensional upper bound follows from Hwang's stronger C^1 result, while in odd dimension we prove a uniformly C^0-stable twisted double-cover obstruction (Sections A.6 and A.7).

Throughout this section, every coordinate chart, tubular strip, and displacement height is fixed before its radial cutoff concentration parameter tends to zero; implicit constants may depend on those fixed choices, but never on the cutoff parameter. This strict order of choices is what makes the critical-exponent estimates meaningful.

A.1 Standard Reductions and Whitney Genericity

We record only the standard facts needed later. Since the cube I satisfies the segment condition, restrictions of $C^\infty(\mathbb{R}^{d_{\text{in}}}; \mathbb{R}^{d_{\text{out}}})$ maps are dense in $W^{1,p}(I; \mathbb{R}^{d_{\text{out}}})$ for every $1 \leq p < \infty$ (Adams & Fournier, 2003, Theorem 3.22). We will also use the following elementary composition fact without further comment: if $\psi_j \to \psi$ strongly in $W^{1,p}$, the images of ψ_j and ψ lie in a fixed compact set, and F is smooth on a neighborhood of that set, then $F \circ \psi_j \to F \circ \psi$ strongly in $W^{1,p}$. Indeed, the chain rule gives

$$D(F \circ \psi_j) - D(F \circ \psi) = DF(\psi_j)(D\psi_j - D\psi) + \big(DF(\psi_j) - DF(\psi)\big)D\psi.$$

The first term converges to zero in L^p because DF is bounded on the fixed compact set. For the second, $DF(\psi_j) - DF(\psi)$ is uniformly bounded and converges to zero in measure. A standard truncation argument then gives convergence after multiplication by the fixed $D\psi \in L^p$: on $\{|D\psi| \leq M\}$, bounded convergence in measure on a finite-measure set implies convergence in L^p, while the contribution of $\{|D\psi| > M\}$ is uniformly small for large M. The zeroth-order term is handled in the same way.

For the lower bounds, the only analytic input is the supercritical Sobolev–Morrey embedding. On the cube and on the sphere, respectively,

$$\|v\|_{C^0(I)} \le C_{d_{\mathrm{in}},p}\|v\|_{W^{1,p}(I)}, \qquad p > d_{\mathrm{in}}, \tag{A.2}$$

$$\|v\|_{C^0(S^{d_{\mathrm{in}}-1})} \le C_{d_{\mathrm{in}},p}\|v\|_{W^{1,p}(S^{d_{\mathrm{in}}-1})}, \qquad p > d_{\mathrm{in}} - 1, \tag{A.3}$$

by Adams & Fournier (2003, Theorem 4.12), applied directly on I and through a finite smooth atlas on $S^{d_{\mathrm{in}}-1}$.

We also freely identify projected embeddings with the geometric class from Section 2. By Hwang (2025b, Lemma 4.2), which invokes the ambient extension theorem of Palais (1960, Theorem C), every smooth embedding $e : I \hookrightarrow \mathbb{R}^\ell$ with $\ell \ge d_{\mathrm{in}}$ is of the form $e = \Phi \circ \iota_{d_{\mathrm{in}},\ell}|_I$ for some $\Phi \in \mathrm{Diff}^\infty(\mathbb{R}^\ell)$. Hence

$$\mathcal{A}^\ell_{d_{\mathrm{in}},d_{\mathrm{out}}} = \{\pi_{\ell,d_{\mathrm{out}}} \circ e : e \in \mathrm{Emb}^\infty(I,\mathbb{R}^\ell)\}. \tag{A.4}$$

It remains to record the generic collision structures used in the constructions. In target dimensions $2d_{\mathrm{in}}$ and $2d_{\mathrm{in}}+1$, standard jet and multijet transversality give the following.

Lemma A.2 (General position)**.** *Every neighborhood-smooth cube map admits arbitrarily small $C^1(I)$ perturbations with the following properties:*

1. *in target $\mathbb{R}^{2d_{\mathrm{in}}}$, the perturbed map is an immersion with finitely many self-transverse unordered double pairs, no triple values, and no double pairs involving ∂I;*
2. *in target $\mathbb{R}^{2d_{\mathrm{in}}+1}$, the perturbed map is a smooth embedding.*

Proof. For maps $\mathbb{R}^{d_{\mathrm{in}}} \to \mathbb{R}^{d_{\mathrm{out}}}$, the corank-one first-jet stratum has codimension $d_{\mathrm{out}} - d_{\mathrm{in}} + 1 > d_{\mathrm{in}}$ when $d_{\mathrm{out}} \in \{2d_{\mathrm{in}}, 2d_{\mathrm{in}}+1\}$, so generic maps are immersions. The expected dimensions of the double- and triple-point loci are $2d_{\mathrm{in}} - d_{\mathrm{out}}$ and $3d_{\mathrm{in}} - 2d_{\mathrm{out}}$. Thus multijet transversality leaves only isolated self-transverse double pairs when $d_{\mathrm{out}} = 2d_{\mathrm{in}}$, and no double points when $d_{\mathrm{out}} = 2d_{\mathrm{in}}+1$; triple points are absent in both cases. A double pair with one preimage on a proper boundary face has expected dimension at most $(d_{\mathrm{in}}-1) + d_{\mathrm{in}} - 2d_{\mathrm{in}} = -1$, so it is also generically absent. These are standard consequences of Wall (2016, Theorem 4.7.7(i)–(ii)) and relative multijet transversality (Golubitsky & Guillemin, 1973, Chapter II, Section 4). Finiteness in the $2d_{\mathrm{in}}$ case follows because an immersion is locally injective near the diagonal and the remaining zero-dimensional double locus is compact. □

For maps into $\mathbb{R}^{2d_{\mathrm{in}}-1}$ we use Whitney's classical semi-regular form. Write $\mathcal{D}_f := \{x \in S^{d_{\mathrm{in}}} : \exists\, y \ne x,\ f(y) = f(x)\}$ for the source double locus.

Lemma A.3 (Whitney-generic maps)**.** *Let $d_{\mathrm{in}} \ge 3$. Every smooth map $f_0 : S^{d_{\mathrm{in}}} \to \mathbb{R}^{2d_{\mathrm{in}}-1}$ admits arbitrarily small C^1 perturbations f such that:*

1. *the critical locus $\Sigma_f = \{x : \mathrm{rank}\, Df_x < d_{\mathrm{in}}\}$ consists of finitely many isolated cross-caps, locally equivalent to*
$$(t, x') \longmapsto (x', t^2, tx'), \qquad x' \in \mathbb{R}^{d_{\mathrm{in}}-1}; \tag{A.5}$$
2. *every regular double pair is self-transverse and there are no triple values;*
3. *the completed source double locus $\widehat{\mathcal{D}}_f = \overline{\mathcal{D}_f} = \mathcal{D}_f \cup \Sigma_f$ is a compact smooth one-manifold, hence a finite union of circles, and $T_c\widehat{\mathcal{D}}_f = \ker Df_c$ at each cross-cap;*
4. *the partner relation extends to a smooth involution on $\widehat{\mathcal{D}}_f$ fixing precisely the cross-caps. Each orbit of circle components is either (a) two circles exchanged by the involution, (b) one circle with a fixed-point-free involution, or (c) one circle with a reflection having two fixed points.*

Proof. Whitney's complete semi-regularity theorem gives the cross-cap normal form, self-transversality of regular double points, and absence of triple values after an arbitrarily small perturbation (Whitney, 1944, Section 3, Lemmas 1–2). In the local model, the double points $(t, 0)$ and

$(-t, 0)$ approach the cross-cap along the kernel direction, so adjoining the cross-caps closes the regular double locus to a compact smooth one-manifold and extends the partner relation by reflection. Whitney's global description of the resulting involution on each circle gives the three orbit types above (Whitney, 1944, Section 22, pp. 290–291). □

A.2 The Critical Sobolev Cutoff

Every geometric detour below must displace the core of a tube by a definite macroscopic distance $H > 0$. While a uniformly small perturbation cannot achieve this, a perturbation concentrated in sufficiently many transverse directions can have arbitrarily small $W^{1,p}$ norm. An ordinary power-law rescaled bump suffices in the subcritical range $p < d$, whereas the following logarithmic capacity profile makes the Sobolev cost vanish all the way through the critical exponent $p = d$.

Lemma A.4. *Let $d \geq 2$, $1 \leq p \leq d$, $0 < R_0 < R$, and $H > 0$. There exist smooth radial functions $a_j : B^d_R \to [0, H]$, equal to H on a neighborhood of the origin and zero for $|z| \geq R_0$, such that $a_j \to 0$ strongly in $W^{1,p}(B^d_R)$.*

Proof. Choose inner radii $\rho_j = R_0 e^{-j}$ so that $L_j = \log(R_0/\rho_j) = j$, and fix a smooth cutoff $\beta : \mathbb{R} \to [0, 1]$ equal to one on a neighborhood of $(-\infty, 0]$ and zero on a neighborhood of $[1, \infty)$. Set

$$a_j(z) = \begin{cases} H, & |z| \leq \rho_j, \\ H\beta\left(\dfrac{\log(|z|/\rho_j)}{L_j}\right), & \rho_j < |z| < R_0, \\ 0, & |z| \geq R_0. \end{cases}$$

Because β is locally constant near 0 and 1, each a_j is smooth across both joining spheres $|z| = \rho_j$ and $|z| = R_0$ as well as at the origin. Since $a_j(z) \to 0$ pointwise for every $z \neq 0$ and $|a_j| \leq H$, dominated convergence gives $\|a_j\|_{L^p(B^d_R)} \to 0$. On the transition annulus $\rho_j < |z| < R_0$, the chain rule gives $|Da_j(z)| \leq CH/(L_j|z|)$, so integrating in polar coordinates yields

$$\int_{B^d_R} |Da_j|^p \, dz \leq \frac{CH^p}{L_j^p} \int_{\rho_j}^{R_0} r^{d-p-1} \, dr. \tag{A.6}$$

For $p < d$, the radial integral is uniformly bounded above by $R_0^{d-p}/(d-p)$, giving a derivative energy of $O(L_j^{-p}) \to 0$. At the critical exponent $p = d$, the radial integral evaluates to $\log(R_0/\rho_j) = L_j$, giving $O(L_j^{1-d}) \to 0$ because $d \geq 2$. □

Notice that the peak height $a_j(0) = H$ remains fixed while $\|a_j\|_{W^{1,p}(B^d_R)} \to 0$. Thus Lemma A.4 controls only the integrated Sobolev error, not the maximum pointwise displacement or maximum derivative. In particular, the failure of this cutoff estimate for $p > d$ does not by itself constitute a proof of a width lower bound; establishing the lower bounds above the critical thresholds requires the independent topological obstructions constructed in Sections A.4 and A.7.

Furthermore, all geometric coordinate charts used with Lemma A.4 below are fixed and smooth on open neighborhoods of their compact supports before the cutoff index j is chosen. If a sequence of local cylinder deformations satisfies $T_j \to \mathrm{id}$ strongly in $W^{1,p}$ with images remaining in a fixed compact set, and Θ is a fixed smooth chart, then decomposing the differential as

$$D(\Theta \circ T_j) - D\Theta = D\Theta(T_j)(DT_j - I) + \big[D\Theta(T_j) - D\Theta\big]$$

and the composition observation from Section A.1, together with the fixed change-of-variables formula, shows that conjugating T_j by Θ preserves strong $W^{1,p}$ convergence to the identity.

A.3 Removing Double Curves

Fix a Whitney-generic map $f : S^{d_{\mathrm{in}}} \to \mathbb{R}^{2d_{\mathrm{in}}-1}$ from Lemma A.3. We leave f unchanged and instead move the embedded source cube away from a finite collection of bad arcs containing all critical points and meeting every double pair.

Write the completed double circles as $C_1, \dots, C_N$. On each C_i, choose a small closed regular arc B_i so that f is injective on $\bigcup_i B_i$ and the sets $f(B_i)$ are pairwise disjoint. This is possible because every fiber has at most two points and each circle contains infinitely many regular points. Then $S_0 := \bigcup_{i=1}^{N} \big(C_i \setminus \operatorname{int}_{C_i} B_i\big)$ is a finite disjoint union of compact arcs containing all cross-caps and meeting every double pair. Consequently, f is an injective immersion on $S^{d_{\text{in}}} \setminus S_0$.

Choose a small closed ball $V \Subset S^{d_{\text{in}}} \setminus (\iota_I(I) \cup S_0)$. Since $d_{\text{in}} \geq 3$, relative transversality allows us to extend the endpoints of the finitely many arcs in S_0 by pairwise disjoint smooth tails ending in V, while keeping the original arcs fixed (Golubitsky & Guillemin, 1973, Chapter II, Section 4). Parametrize each resulting arc by $\gamma : (-A, L + A) \to S^{d_{\text{in}}}$ so that every point of the arc meeting $\iota_I(I)$ lies in the core $\gamma([0, L])$ and both tails lie in V. Let S be the union of these core arcs, so $S_0 \subset S$.

Fix $H > L$ and $A > H$, and choose $\chi \in C_c^\infty((-A, L + A); [0, 1])$ with $\chi \equiv 1$ near $[0, L]$ and $H\|\chi'\|_\infty < 1/2$. In a tubular coordinate system (s, z) around each extended arc, take the cutoff a_j from Lemma A.4 in transverse dimension $d_{\text{in}} - 1$ and define

$$T_j(s, z) = \big(s - \chi(s)a_j(z), z\big). \tag{A.7}$$

For fixed z, the derivative in the s-direction is at least $1/2$, so T_j is a diffeomorphism of the cylinder, equal to the identity near its boundary. On the core, $T_j(s, 0) = (s - H, 0)$ moves $[0, L]$ into the exterior tail, while

$$DT_j - I = \begin{pmatrix} -\chi'(s)a_j(z) & -\chi(s)Da_j(z) \\ 0 & 0 \end{pmatrix}.$$

Thus $T_j \to \mathrm{id}$ strongly in $W^{1,p}$ for every $1 \leq p \leq d_{\text{in}} - 1$.

Choose disjoint tubular embeddings around the finitely many extended arcs and conjugate the maps T_j into $S^{d_{\text{in}}}$, extending by the identity outside the tubes. The tubular neighborhoods are fixed before j varies. We obtain $\Psi_j \in \mathrm{Diff}^\infty(S^{d_{\text{in}}})$ such that

$$\Psi_j(\iota_I(I)) \cap S = \varnothing, \qquad \Psi_j \circ \iota_I \longrightarrow \iota_I \quad \text{strongly in } W^{1,p}(I; \mathbb{R}^{d_{\text{in}}+1}). \tag{A.8}$$

The convergence follows from the fixed-chart composition observation in Section A.1.

Lemma A.5. *If $d_{\text{in}} \geq 3$ and $1 \leq p \leq d_{\text{in}} - 1$, then smooth embeddings $I \hookrightarrow \mathbb{R}^{2d_{\text{in}}-1}$ are dense in $W^{1,p}(I; \mathbb{R}^{2d_{\text{in}}-1})$.*

Proof. Given $u \in W^{1,p}(I; \mathbb{R}^{2d_{\text{in}}-1})$ and $\varepsilon > 0$, first approximate u by a neighborhood-smooth map, extend it to $S^{d_{\text{in}}}$ by a chart cutoff, and then take a Whitney-generic perturbation f with total error below $2\varepsilon/3$. Since $\Psi_j(\iota_I(I))$ avoids $S \supset S_0$ and f is an injective immersion off S_0, the maps $e_j = f \circ \Psi_j \circ \iota_I$ are smooth embeddings. Equation (A.8) and the fixed-smooth-map composition observation imply $e_j \to f|_I$ strongly in $W^{1,p}$. Choosing j large enough gives $\|e_j - u\|_{W^{1,p}} < \varepsilon$. □

A.4 The Linked-Sphere Obstruction

To prove that the threshold $p = d_{\text{in}} - 1$ is sharp, we need a target witness that *every* nearby embedding must fail to approximate. We place a linked sphere pair on two concentric annular plateaus in the cube: a small bulk $W^{1,p}$ error yields a low-energy spherical slice in each annulus, and Morrey's inequality on those spheres preserves their linking number, whereas a global embedding of the cube would provide a disjoint ball filling and force the linking number to vanish.

Lemma A.6 (Linked-sphere obstruction)**.** *If $d_{\text{in}} \geq 3$ and $p > d_{\text{in}} - 1$, there exist a smooth map $f : I \to \mathbb{R}^{2d_{\text{in}}-1}$ and $c > 0$ such that every smooth embedding $g : I \hookrightarrow \mathbb{R}^{2d_{\text{in}}-1}$ satisfies $\|g - f\|_{W^{1,p}(I)} \geq c$.*

Proof. For continuous maps $a, b : S^{d_{\text{in}}-1} \to \mathbb{R}^{2d_{\text{in}}-1}$ with disjoint images, use the mod-2 linking invariant

$$\mathrm{lk}_2(a, b) = \deg_2\left[(u, v) \longmapsto \frac{a(u) - b(v)}{|a(u) - b(v)|}\right] \in \mathbb{F}_2. \tag{A.9}$$

Here $\deg_2$ is mapping degree modulo two between $S^{d_{\text{in}}-1} \times S^{d_{\text{in}}-1}$ and $S^{2d_{\text{in}}-2}$. We use its standard homotopy and intersection properties (Hatcher, 2002, Sections 2.2 and 3.3): linking is unchanged

under homotopies through disjoint pairs, and vanishes if a extends continuously over a d_{in}-ball $D^{d_{\text{in}}}$ disjoint from $b(S^{d_{\text{in}}-1})$. For the latter assertion, the normalized difference map extends over $D^{d_{\text{in}}} \times S^{d_{\text{in}}-1}$, so its boundary degree is zero.

In $\mathbb{R}^{2d_{\text{in}}-1} = \mathbb{R}^{d_{\text{in}}}_x \times \mathbb{R}^{d_{\text{in}}-1}_y$, take the model spheres

$$L_1 = \{(x,0) : |x| = 1\}, \qquad L_2 = \{(x,y) : x_2 = \cdots = x_{d_{\text{in}}} = 0,\ (x_1-1)^2 + |y|^2 = 1/4\}.$$

They are disjoint, and the flat disk $\{(x,0) : |x| \le 1\}$ bounded by L_1 meets L_2 once transversely at $x = (1/2, 0, \dots, 0)$, $y = 0$; thus their mod-2 linking number is one. Parametrize them for $\theta = (\theta_1, \dots, \theta_{d_{\text{in}}}) \in S^{d_{\text{in}}-1}$ by

$$h_1(\theta) = (\theta, 0), \qquad h_2(\theta) = \big((1 + \theta_1/2, 0, \dots, 0),\ (\theta_2, \dots, \theta_{d_{\text{in}}})/2\big).$$

Writing $d_0 = \operatorname{dist}(L_1, L_2) > 0$, any continuous perturbations of both parametrizations by less than $d_0/4$ in C^0 stay linked, as the straight-line homotopy remains a disjoint pair.

We next install this fixed pair on annuli in the source cube. Choose disjoint closed balls $B_1, B_2 \Subset \operatorname{int} I$ with centers c_i and radii $0 < r_0 < r_1 < r_2 < r_3 < \min_i \operatorname{rad}(B_i)$. Let $\eta \in C^\infty_c((r_0, r_3))$ equal one on $[r_1, r_2]$, and set

$$f(c_i + r\theta) = \eta(r) h_i(\theta) \text{ in } B_i, \qquad f = 0 \text{ outside } B_1 \cup B_2. \tag{A.10}$$

This map is smooth on I, including at the ball centers and boundaries. For any smooth $g : I \to \mathbb{R}^{2d_{\text{in}}-1}$, put $u = g - f$. The tangential gradient identity $\nabla_{S^{d_{\text{in}}-1}}[u(c_i + r\theta)] = rDu(c_i + r\theta)|_{T_\theta S^{d_{\text{in}}-1}}$ and polar coordinates give, since $r \in [r_1, r_2]$ is bounded above and away from zero,

$$\int_{r_1}^{r_2} \|u(c_i + r\,\cdot)\|^p_{W^{1,p}(S^{d_{\text{in}}-1})}\, dr \le C \|u\|^p_{W^{1,p}(B_i)}. \tag{A.11}$$

Select a radius $\widehat{r}_i \in [r_1, r_2]$ where the integrand does not exceed its average. Since $f = h_i$ on these spherical slices,

$$\|g(c_i + \widehat{r}_i\,\cdot) - h_i\|_{W^{1,p}(S^{d_{\text{in}}-1})} \le C \|g - f\|_{W^{1,p}(I)}. \tag{A.12}$$

By the spherical Sobolev–Morrey inequality (A.3), a sufficiently small bulk Sobolev error makes both uniform errors on the selected spheres less than $d_0/4$, forcing the two image spheres to have linking number one. If g were an embedding, however, the image of the source ball $\overline{B}(c_1, \widehat{r}_1)$ would fill the first sphere disjointly from the second sphere $g(\partial B(c_2, \widehat{r}_2))$, forcing their linking number to be zero. Choosing a fixed error threshold $c > 0$ from (A.12) proves the asserted positive distance. □

A.5 Removing Isolated Double Points

Adding one ambient coordinate changes the collision geometry: Lemma A.2 leaves only finitely many isolated, self-transverse double pairs in $\mathbb{R}^{2d_{\text{in}}}$. We remove them one at a time. A purely local separation would merely move the intersection, so the modified sheet is guided along the partner sheet to the free boundary of the cube and then out of the image.

Lemma A.7 (Adapted escape strip)**.** *Let $d_{\text{in}} \ge 2$ and let $F : I \to \mathbb{R}^{2d_{\text{in}}}$ be a smooth immersion with finitely many self-transverse double pairs, no triple values, and no double pairs involving ∂I. For a double pair $a \ne b$ with $F(a) = F(b)$, there exist $H, R > 0$, a coordinate embedding ψ near $\overline{B}^{d_{\text{in}}}_R$ with $\psi(0) = a$ and $\psi(\overline{B}^{d_{\text{in}}}_R) \subset \operatorname{int} I$, and a smooth embedding Θ near $[0, H] \times \overline{B}^{d_{\text{in}}}_R$ such that*

$$\Theta(0, z) = F(\psi(z)), \tag{A.13}$$

$$\Theta\big((0, H] \times (B^{d_{\text{in}}}_R \setminus \{0\})\big) \cap F(I) = \varnothing, \tag{A.14}$$

$$\Theta(H, 0) \notin F(I). \tag{A.15}$$

After shrinking R, the disk $\psi(B^{d_{\text{in}}}_R)$ contains no other double-point preimage and its only equal-image pair with the rest of I is (a, b).

Proof. By the standing convention at the beginning of the appendix, F extends smoothly to a neighborhood of I; after shrinking that neighborhood, the extension remains an immersion. Choose a smooth embedded path γ in I from b to a point y in the relative interior of a boundary face, avoiding a

small closed disk about a and all other double-point preimages. Such a path exists for $d_{\text{in}} \geq 2$. Since the only partner of b is a and no other point of γ is a double-point preimage, $F \circ \gamma$ is embedded.

There is no double pair involving y, so y is the unique point of I mapping to $F(y)$. Choose a source neighborhood U_y on which the extension of F is an embedding. Since $F(y) \notin F(I \setminus U_y)$ and $I \setminus U_y$ is compact, these two compact sets are positively separated near $F(y)$. We may therefore extend γ transversely through the boundary, inside U_y, by a short exterior segment whose image is disjoint from $F(I)$ except at its initial point; extend slightly farther so that its terminal image lies outside $F(I)$. Reparametrize the resulting compact arc so that $\gamma(0) = b$ and its terminal point is $\gamma(H)$.

Take a sufficiently thin source tubular strip $\Gamma(s, u)$, $u \in B_\eta^{d_{\text{in}}-1}$, around the extended path and put $E = F \circ \Gamma$, using the fixed smooth extension of F outside I. Because the center arc has embedded image and the extension is immersive, compactness and local injectivity imply, after shrinking η, that E is a smooth embedding of the strip. Near $q = F(a) = F(b)$, choose a coordinate disk $\psi : B_R^{d_{\text{in}}} \to \operatorname{int} I$ about a and write $A(z) = F(\psi(z))$, with $A(0) = q$. Self-transversality of the pair (a, b) says that $DA(0)(\mathbb{R}^{d_{\text{in}}})$ is complementary to the tangent space of the E-sheet at q.

The strip base is contractible, so its normal quotient bundle is trivial. Choose a smooth complementary frame $N(s, u)$ along E, with $N(0, 0) = N_0 := DA(0)$ (Hirsch, 1976, Chapter 4), and choose a smooth cutoff λ on the strip base that equals one near $(0, 0)$. Define

$$\Xi(s, u, z) = E(s, u) + N(s, u)z + \lambda(s, u)\big(A(z) - q - N_0 z\big).$$

Writing $R_A(z) = A(z) - q - N_0 z$, Taylor's theorem gives $|R_A(z)| = O(|z|^2)$ and $|DR_A(z)| = O(|z|)$ as $z \to 0$. Hence $\|R_A\|_{C^1(\overline{B}_R)} \to 0$ as $R \downarrow 0$. The uncorrected tubular map $(s, u, z) \mapsto E(s, u) + N(s, u)z$ is an embedding for a sufficiently small fixed normal radius; since embeddings of a compact source are open in the C^1 topology, shrinking R makes Ξ an embedding as well. Moreover, $\Xi(s, u, 0) = E(s, u)$ and $\Xi(0, 0, z) = A(z)$, the second identity following from $\lambda(0, 0) = 1$ and $N(0, 0) = N_0$.

It remains to choose the tube thin enough to control its intersections with $F(I)$. Near q, the absence of triple values and self-transversality imply that the image of F consists only of the two local sheets parametrized by E and A. Away from this crossing, the compact center subarc inside $F(I)$ is disjoint from the image under F of the complement of a slightly larger source strip, and therefore has positive separation from that complement. Inside the larger strip, the partner sheet is exactly E, i.e. the set $z = 0$ in the coordinates above. Along the short exterior continuation past y, the axis and, after shrinking the normal radius, its terminal tube are disjoint from $F(I)$. By compactness of the center arc, one normal radius works simultaneously on all these regions. Consequently the image of Ξ meets $F(I)$ only in the partner sheet $\{z = 0\}$ and, at $s = 0$, the initial a-sheet.

Finally restrict to the central slice $u = 0$ and set $\Theta(s, z) = \Xi(s, 0, z)$. The identities above give (A.13); the intersection description gives (A.14); and the exterior endpoint gives (A.15). Shrinking R once more excludes every other double-point preimage from $\psi(B_R^{d_{\text{in}}})$, so the only equal-image pair of this disk with the rest of I is (a, b). □

Lemma A.8 (Removal of one double point)**.** *Under Lemma A.7, for $1 \leq p \leq d_{\text{in}}$ and $\varepsilon > 0$, there is a smooth immersion $P : I \to \mathbb{R}^{2d_{\text{in}}}$ with $\|P - F\|_{W^{1,p}(I)} < \varepsilon$, equal to F near ∂I, whose double pairs are exactly the other double pairs of F and whose local maps near the preimages of every remaining double pair are unchanged.*

Proof. Fix the strip Θ and chart ψ, choose $0 < R_0 < R$, and take the cutoff $a_j : B_R^{d_{\text{in}}} \to [0, H]$ from Lemma A.4 in dimension $d = d_{\text{in}}$. Define

$$P_j(x) = \begin{cases} \Theta(a_j(z), z), & x = \psi(z),\ |z| < R, \\ F(x), & x \notin \psi(B_R^{d_{\text{in}}}). \end{cases} \tag{A.16}$$

Because $a_j = 0$ near $|z| = R$, the two formulas glue smoothly. The graph $z \mapsto (a_j(z), z)$ is an embedding, so the modified disk remains an immersed embedded sheet. If $z \neq 0$ and $a_j(z) > 0$, (A.14) excludes a collision with the unmodified image; if $z = 0$, then $a_j(0) = H$ and (A.15) does the same. When $a_j(z) = 0$, the point is unchanged, and the final assertion of Lemma A.7 rules out any new partner outside the disk. Finally, injectivity of Θ excludes collisions between two modified points. Thus the chosen pair is removed. Since the modification is supported in a disk disjoint from

the preimages of every other double pair, $P_j = F$ on neighborhoods of those preimages, so all remaining pairs and their local maps are unchanged.

All changes occur in the fixed chart $\psi(B^{d_{\text{in}}}_{R_0})$. Bounded first derivatives of Θ and Lipschitz continuity of $D\Theta$ on the fixed compact tube give

$$|\Theta(a_j(z), z) - \Theta(0, z)| \le C|a_j(z)|, \qquad \big|D_z[\Theta(a_j(z), z)] - D_z\Theta(0, z)\big| \le C\big(|a_j(z)| + |Da_j(z)|\big).$$

After the fixed coordinate change $x = \psi(z)$,

$$\|P_j - F\|^p_{W^{1,p}(I)} \le C \int_{B^{d_{\text{in}}}_R} \big(|a_j|^p + |Da_j|^p\big)\, dz \longrightarrow 0. \tag{A.17}$$

Choose j large enough. □

Lemma A.9 (Critical-dimensional embedding density)**.** *If $d_{\text{in}} \ge 2$ and $1 \le p \le d_{\text{in}}$, smooth embeddings $I \hookrightarrow \mathbb{R}^{2d_{\text{in}}}$ are dense in $W^{1,p}(I; \mathbb{R}^{2d_{\text{in}}})$.*

Proof. Given a target $u \in W^{1,p}(I; \mathbb{R}^{2d_{\text{in}}})$ and $\varepsilon > 0$, choose a neighborhood-smooth map u_0 with $\|u - u_0\|_{W^{1,p}} < \varepsilon/3$. By Lemma A.2, choose a C^1-small perturbation F of u_0 with $\|F - u_0\|_{W^{1,p}} < \varepsilon/3$ that is an immersion with finitely many transverse interior double pairs and no triple or boundary double values. If there are no double pairs, F is already an embedding. Otherwise let their number be N and iterate Lemma A.8 N times, choosing the error at each stage smaller than $\varepsilon/(3N)$. Each step leaves the local maps near the remaining double pairs unchanged, so the hypotheses needed at the next stage are preserved. The final map is an injective immersion of the compact cube, hence a smooth embedding, and the triangle inequality gives total $W^{1,p}$ error less than ε. □

Remark A.10. The escape strip uses the free boundary of I. The construction does not assert a corresponding density statement for closed source manifolds or for approximation with prescribed boundary values.

A.6 The Supercritical Even-Dimensional Case

For $p > d_{\text{in}}$ and even d_{in}, no new finite-p construction is needed. Hwang proves the stronger C^1 geometric result $\Omega(d_{\text{in}}, 2d_{\text{in}} - 1) = 2d_{\text{in}}$ for even d_{in} (Hwang, 2025b, Lemma 5.3), where Ω is defined using C^1-density of the same class of projected embeddings. Since $C^1(I)$ convergence implies $W^{1,p}(I)$ convergence on the compact cube, this immediately gives

$$\Omega_{1,p}(d_{\text{in}}, 2d_{\text{in}} - 1) \le 2d_{\text{in}} \qquad (p > d_{\text{in}},\ d_{\text{in}} \text{ even}). \tag{A.18}$$

The matching lower bound already follows from the linked-sphere obstruction, which excludes width $2d_{\text{in}} - 1$ for every $p > d_{\text{in}} - 1$.

A.7 A Uniformly Stable Odd-Dimensional Obstruction

When $d_{\text{in}} \ge 3$ is odd, Whitney's fixed-point-free orbit type (b) can occur, producing a connected two-to-one double cover of an image circle. An exact scalar lift h is impossible for such a loop because the odd height difference $h(x) - h(y)$ would have to be a continuous nowhere-vanishing function on a connected circle that changes sign under the antipodal exchange. To obtain a Sobolev lower bound for $p > d_{\text{in}}$, we must prove more: this obstruction persists for *all sufficiently uniformly (C^0) close maps*, without requiring any derivative control.

To do this, we begin with an explicit local model that realizes this connected double cover. Let $d_{\text{in}} \ge 3$ be odd, put $q = d_{\text{in}} - 1$ (which is even), and choose the block-diagonal orthogonal complex structure J on $\mathbb{R}^q$ with 2×2 blocks $\left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$. For $t \in \mathbb{R}$, set

$$B_t = e^{tJ/2}, \qquad A_t z = \big(\cos(t/2) B_t z,\ \sin(t/2) B_t z\big) \in \mathbb{R}^q \oplus \mathbb{R}^q. \tag{A.19}$$

Because both trigonometric scalar factors and B_t change sign at $t + 2\pi$, the family A_t is smooth and 2π-periodic on $S^1 = \mathbb{R}/2\pi\mathbb{Z}$. Each A_t is a linear isometry onto a q-plane, and the images $\operatorname{im} A_t$ and $\operatorname{im} A_{t+\pi}$ at antipodal angles are orthogonal complementary q-planes in $\mathbb{R}^{2q}$, since their inner product carries the scalar factor $\cos(t/2)(-\sin(t/2)) + \sin(t/2)\cos(t/2) = 0$.

Embed a source solid torus $S^1 \times B^q_\rho \Subset \operatorname{int} I$ and choose a target tubular embedding $\Xi : S^1 \times B^{2q}_{2\rho} \to \mathbb{R}^{2q+1} = \mathbb{R}^{2d_{\text{in}}-1}$ around a standard circle. On a smaller coaxial source tube, define

$$f(t, z) = \Xi(2t, A_t z). \tag{A.20}$$

A radial cutoff extends this smoothly over I without changing it near the core $\{z = 0\}$. The model map is an immersion on the tube: its base derivative has speed $2 \neq 0$ and its normal z-derivative A_t has full rank q. Equality of two model values forces either the same point ($t' = t$, $z' = z$) or antipodal core points $t' = t + \pi$ and $z = z' = 0$, by complementarity of the two q-planes. These double pairs are self-transverse, and the ordered double locus in this pair neighborhood is the isolated circle

$$C = \big\{\big((t, 0), (t + \pi, 0)\big) : t \in S^1\big\}, \tag{A.21}$$

with partner swap $\tau : t \mapsto t + \pi$, forming a connected double cover of C/τ. Coincidences involving the global extension outside the tube do not enter a sufficiently small isolating neighborhood of C in the deleted product.

Uniform C^0 perturbations need not preserve the smooth double locus, so we encode the obstruction in the deleted-product quotient; this yields a stability statement requiring no derivative control on the perturbing map.

Lemma A.11. *For a smooth $f : I \to \mathbb{R}^{2d_{\text{in}}-1}$, $d_{\text{in}} \geq 2$, write $\widetilde{I} = (\operatorname{int} I)^2 \setminus \{(x, x)\}$, $\tau(x, y) = (y, x)$, and $d_f(x, y) = f(x) - f(y)$. Suppose $d_f^{-1}(0)$ has an isolated compact circle C along which d_f is transverse to zero, with $\tau(C) = C$ and connected cover $C \to C/\tau$. Then some $\delta > 0$ has the following property: no smooth g with $\|g - f\|_\infty < \delta$ admits a continuous height h for which (g, h) is injective.*

Proof. Choose a small τ-invariant closed tubular neighborhood $U \Subset \widetilde{I}$ of C, which is a compact smooth $2d_{\text{in}}$-manifold with boundary on which τ acts freely. Its boundary misses the zero set $d_f^{-1}(0)$; put $\alpha = \min_{\partial U} |d_f| > 0$ and choose $0 < \delta < \alpha/4$. For $\|g - f\|_\infty < \delta$, the linear homotopy from d_f to d_g has norm at least $\alpha - 2\delta > 0$ on ∂U. If (g, h) were injective, $H(x, y) = h(x) - h(y)$ would be continuous and odd on U, and because U misses the diagonal, (d_g, H) has no common zero on U; hence $\mu := \min_U \big(|d_g|^2 + |H|^2\big)^{1/2} > 0$. A perturbation of d_g smaller than $\mu/2$ cannot vanish where $H = 0$. This auxiliary tolerance $\mu/2$ may depend on the proposed height h; the uniform neighborhood size δ depends only on f and U.

On the $2d_{\text{in}}$-dimensional quotient manifold $\bar{U} = U/\tau$, odd maps $U \to \mathbb{R}^{2d_{\text{in}}-1}$ correspond to sections of the rank-$(2d_{\text{in}} - 1)$ vector bundle

$$\xi = (U \times \mathbb{R}^{2d_{\text{in}}-1})/\!\sim \longrightarrow \bar{U}, \qquad \big((x, y), v\big) \sim \big((y, x), -v\big). \tag{A.22}$$

View the homotopy between d_f and d_g as a section of the pullback of ξ to $\bar{U} \times [0, 1]$ and make it constant near the time endpoints. Relative transversality for sections (Hirsch, 1976, Chapter 3; Golubitsky & Guillemin, 1973, Chapter II, Section 4) gives a smooth perturbation, fixed near time zero and the lateral boundary $\partial\bar{U} \times [0, 1]$, for which the full section and its terminal restriction at $s = 1$ are transverse to the zero section. Choose the lifted perturbation smaller than $\min\{(\alpha - 2\delta)/2, \mu/2\}$. Then there are no lateral zeros, and H stays nonzero on the lifted terminal zero set. Because the bundle has rank $2d_{\text{in}} - 1$, the transverse zero locus $W \subset \bar{U} \times [0, 1]$ has dimension $\dim(\bar{U} \times [0, 1]) - (2d_{\text{in}} - 1) = (2d_{\text{in}} + 1) - (2d_{\text{in}} - 1) = 2$ and is therefore a compact smooth surface with boundary $\partial W = (C/\tau) \sqcup Z_1$, where $Z_1 \subset \bar{U} \times \{1\}$ is the terminal zero one-manifold. Pulling back the double covering $U \times [0, 1] \to \bar{U} \times [0, 1]$ along W gives a two-sheeted cover $\widetilde{W} \to W$. The perturbed terminal section need not be a difference map; it is used only to construct this cobordism, with separation by H preserved independently.

Let $a \in H^1(W; \mathbb{F}_2)$ be the cohomology class classifying the double cover $\widetilde{W} \to W$, whose Kronecker evaluation $\langle a, [\gamma]\rangle$ on a homology loop class $[\gamma] \in H_1(W; \mathbb{F}_2)$ is one precisely when a circuit around γ exchanges the sheets (Hatcher, 2002, Sections 1.3 and 3.1). Because $C \to C/\tau$ is connected, $\langle a, [C/\tau]\rangle = 1$. On every circle component Z of Z_1, the unique lift on which the odd nonzero function satisfies $H > 0$ is a continuous section, trivializing the cover over Z and giving $\langle a, [Z]\rangle = 0$. With mod-2 coefficients, $\partial[W, \partial W] = [\partial W] \in H_1(\partial W; \mathbb{F}_2)$ holds for any compact

surface, even without orientability (Hatcher, 2002, Section 3.3). If $\jmath : \partial W \hookrightarrow W$ is the boundary inclusion, exactness of the pair sequence gives $\jmath_*[\partial W] = 0$ in $H_1(W; \mathbb{F}_2)$, hence

$$0 = \langle a, \jmath_*[\partial W] \rangle = \langle a, [C/\tau] \rangle + \sum_{Z \subset Z_1} \langle a, [Z] \rangle = 1,$$

a contradiction in $\mathbb{F}_2$, including when Z_1 is empty. □

Applying Lemma A.11 to the explicit model (A.20) and its connected double circle (A.21), we obtain a smooth $f : I \to \mathbb{R}^{2d_{\text{in}}-1}$ and $\delta > 0$ such that no smooth g with $\|g - f\|_{C^0(I)} < \delta$ admits an injective continuous scalar lift.

For $p > d_{\text{in}}$, the cube Sobolev–Morrey inequality (A.2) gives

$$\|g - f\|_{C^0(I)} \le C_{d_{\text{in}},p} \|g - f\|_{W^{1,p}(I)}. \tag{A.23}$$

Every $g \in \mathcal{A}^{2d_{\text{in}}}_{d_{\text{in}},2d_{\text{in}}-1}$ has a smooth scalar lift, given by the last ambient coordinate of its defining embedding. The obstruction therefore places its witness f at positive $W^{1,p}$ distance, which is at least $\delta/C_{d_{\text{in}},p}$, from that class.

A.8 Assembling the Phase Diagram

Proof of Theorem A.1. By smooth density and (A.4), it suffices to approximate a neighborhood-smooth target $u : I \to \mathbb{R}^{2d_{\text{in}}-1}$ by projected embeddings.

If $1 \le p \le d_{\text{in}} - 1$, Lemma A.5 gives density already in ambient dimension $2d_{\text{in}} - 1$. If $d_{\text{in}} - 1 < p \le d_{\text{in}}$, Lemma A.6 excludes width $2d_{\text{in}} - 1$, while Lemma A.9 applied to the zero-padded target $(u, 0)$ gives width $2d_{\text{in}}$. This proves the first two regimes.

Now suppose $p > d_{\text{in}}$. The linked-sphere obstruction still excludes width $2d_{\text{in}} - 1$. If d_{in} is even, (A.18) gives the matching upper bound $2d_{\text{in}}$. If d_{in} is odd, the stable double-cover obstruction above provides a smooth target at positive $W^{1,p}$ distance from $\mathcal{A}^{2d_{\text{in}}}_{d_{\text{in}},2d_{\text{in}}-1}$ by Sobolev–Morrey, so width $2d_{\text{in}}$ is impossible. Finally, Lemma A.2(2), applied to $(u, 0, 0) : I \to \mathbb{R}^{2d_{\text{in}}+1}$ and followed by projection onto the first $2d_{\text{in}} - 1$ coordinates, gives width $2d_{\text{in}} + 1$. This proves the remaining two regimes.

At the critical exponents $p = d_{\text{in}} - 1$ and $p = d_{\text{in}}$, the smaller widths are attained directly: Lemma A.4 gives derivative energies $O(L_j^{2-d_{\text{in}}})$ in the $(d_{\text{in}} - 1)$-transverse detour and $O(L_j^{1-d_{\text{in}}})$ in the d_{in}-dimensional graph replacement. Both vanish for $d_{\text{in}} \ge 3$, so no endpoint Morrey argument is used. □

B Implications for the Minimum-Width Problem

Appendix A is geometric. Here we transfer its conclusions to neural networks and prove the dimension-change statement used in Corollary 3.3. The only network-specific input is Hwang's $W^{1,\infty}$ approximation theorem for ambient diffeomorphisms.

Throughout, $I = [0, 1]^{d_{\text{in}}}$, and the value and derivative norms are the Euclidean and Frobenius norms from Section 2.1.

B.1 Transfer Between Projected Embeddings and Networks

We use Hwang (2025b, Theorem 4.1): for a leaky-ReLU-like activation family, every smooth diffeomorphism of $\mathbb{R}^\ell$ can be approximated locally in $W^{1,\infty}$ by width-ℓ networks. Our Definition 2.1 is slightly stronger than Hwang's Definition 3.3. Indeed, for every fixed $\beta > 0$, $\frac{\sigma_\beta(t)}{t} = \frac{\sigma_\beta(0)}{t} + \frac{1}{t}\int_0^t \sigma'_\beta(s)\, ds \to 1$ as $t \to +\infty$, and similarly $\sigma_\beta(t)/t \to \beta$ as $t \to -\infty$, by Cesàro convergence. Thus the subfamily $\{\sigma_\beta : \beta > 0\} \subset \Sigma$ satisfies Hwang's tail-ratio hypotheses, while the uniform near-identity derivative condition is already part of Definition 2.1.

Proposition B.1. *Let Σ satisfy Definition 2.1, let $1 \le p < \infty$, and let $\ell \ge \max\{d_{\text{in}}, d_{\text{out}}\}$. Then*

$$\overline{\mathcal{N}^{\Sigma}_{d_{\text{in}},d_{\text{out}},\ell}}^{W^{1,p}} = \overline{\mathcal{A}^{\ell}_{d_{\text{in}},d_{\text{out}}}}^{W^{1,p}} \quad \text{in } W^{1,p}(I;\mathbb{R}^{d_{\text{out}}}). \tag{B.1}$$

Consequently,

$$w^{1,p}_{\min}(d_{\text{in}}, d_{\text{out}}; \Sigma) = \Omega_{1,p}(d_{\text{in}}, d_{\text{out}}). \tag{B.2}$$

Proof of Propositions 2.2 and B.1. Fix $\ell \ge \max\{d_{\text{in}}, d_{\text{out}}\}$. If $g = \pi_{\ell,d_{\text{out}}} \circ \Phi \circ \iota_{d_{\text{in}},\ell}|_I \in \mathcal{A}^{\ell}_{d_{\text{in}},d_{\text{out}}}$, then Hwang's Theorem 4.1 gives width-ℓ networks $\widetilde{N}_j : \mathbb{R}^{\ell} \to \mathbb{R}^{\ell}$ converging to Φ locally in $W^{1,\infty}$ on a neighborhood of $\iota_{d_{\text{in}},\ell}(I)$. Since $\widetilde{N}_j$ and Φ are C^1, these essential-supremum bounds hold pointwise on that neighborhood and may be restricted to the included cube. Absorbing the inclusion and projection into the first and last affine layers therefore yields $N_j \in \mathcal{N}^{\Sigma}_{d_{\text{in}},d_{\text{out}},\ell}$ with $N_j \to g$ in $C^1(I)$, hence in $W^{1,p}(I)$. Therefore $\mathcal{A}^{\ell}_{d_{\text{in}},d_{\text{out}}} \subset \overline{\mathcal{N}^{\Sigma}_{d_{\text{in}},d_{\text{out}},\ell}}^{W^{1,p}}$.

Conversely, we use the same regularization idea as in the proof of Hwang (2025b, Proposition 4.5), but spell out the affine step because here we need the stronger per-width closure statement. Fix $N \in \mathcal{N}^{\Sigma}_{d_{\text{in}},d_{\text{out}},\ell}$ and pad all hidden layers to dimension ℓ. After absorbing the input inclusion and output projection, N is the restriction of a finite composition of square affine maps and coordinate-wise nondecreasing C^1 maps on $\mathbb{R}^{\ell}$ (with the identity on dummy coordinates). Each square affine map can be perturbed arbitrarily little to an invertible affine map: if M is its linear part, then $\det(M + tI)$ is a nonzero polynomial in t, so it is nonzero for arbitrarily small t. Each nondecreasing $\sigma \in C^1(\mathbb{R})$ can, on any fixed compact interval, be approximated in C^1 by smooth increasing diffeomorphisms $s_{\varepsilon,\delta}(t) = (\eta_{\varepsilon} * \sigma)(t) + \delta t$, where η_{ε} is a standard nonnegative mollifier and $\delta > 0$. Indeed, $s'_{\varepsilon,\delta} \ge \delta$, and hence $s_{\varepsilon,\delta}$ is a smooth strictly increasing bijection $\mathbb{R} \to \mathbb{R}$ with smooth inverse; moreover $s_{\varepsilon,\delta} \to \sigma$ in C^1 on compact intervals as $\varepsilon, \delta \downarrow 0$. Choose slightly enlarged compact sets containing all intermediate images of I. By taking the finitely many perturbations successively small, every perturbed intermediate image remains in the chosen compact set, and finite composition is continuous in the C^1 topology there. Thus the square composition is a global diffeomorphism $\Phi_j : \mathbb{R}^{\ell} \to \mathbb{R}^{\ell}$ and $\pi_{\ell,d_{\text{out}}} \circ \Phi_j \circ \iota_{d_{\text{in}},\ell}|_I \to N$ in $C^1(I)$. Hence $\mathcal{N}^{\Sigma}_{d_{\text{in}},d_{\text{out}},\ell} \subset \overline{\mathcal{A}^{\ell}_{d_{\text{in}},d_{\text{out}}}}^{W^{1,p}}$, which proves (B.1). Finally, the standard L^p lower bound recalled in Section 2.2, together with $w^{1,p}_{\min} \ge w^{p}_{\min}$, excludes universal neural widths below $\max\{d_{\text{in}}, d_{\text{out}}\}$. Taking the least admissible width in (B.1) proves (B.2). □

Combining Proposition B.1 with Theorem A.1 immediately gives the neural phase diagram in Theorem 3.2; no separate corollary is needed here.

B.2 General Dimension Arguments

We next prove the dimension-change inequality used in Corollary 3.3. We would like to highlight that strong $W^{1,p}$ convergence on a d_{in}-dimensional cube does not imply tangential $W^{1,p}$ convergence on a prescribed lower-dimensional face. Instead, a slightly more careful argument with Fubini's theorem is needed to select an interior slice separately for each approximant.

For $r \ge 0$, write $I_r = [0,1]^r$, with I_0 a singleton; in particular, $I_{d_{\text{in}}} = I$. This notation is used only when comparing different input dimensions.

Lemma B.2. *Let $1 \le {d_{\text{in}}}' \le d_{\text{in}}$, $1 \le d_{\text{out}} \le {d_{\text{out}}}'$, and $1 \le p < \infty$. Then*

$$\Omega_{1,p}({d_{\text{in}}}', {d_{\text{out}}}') \le \Omega_{1,p}(d_{\text{in}}, d_{\text{out}}) + ({d_{\text{out}}}' - d_{\text{out}}). \tag{B.3}$$

Proof. Set $\ell = \Omega_{1,p}(d_{\text{in}}, d_{\text{out}})$ and $r = {d_{\text{out}}}' - d_{\text{out}}$. By smooth density in $W^{1,p}$, it suffices to approximate a smooth target $F = (F', F'') : I_{{d_{\text{in}}}'} \to \mathbb{R}^{d_{\text{out}}} \times \mathbb{R}^r$.

Extend the first block trivially in the extra input variables by $\widetilde{F}'(x,t) = F'(x)$ for $x \in I_{{d_{\text{in}}}'}$ and $t \in I_{d_{\text{in}} - {d_{\text{in}}}'}$. By the definition of ℓ and the projected-embedding representation (A.4), there exist smooth embeddings $e_k : I \hookrightarrow \mathbb{R}^{\ell}$ such that, with $u_k(x,t) := \pi_{\ell,d_{\text{out}}} e_k(x,t) - F'(x)$, we have $u_k \longrightarrow 0$ in $W^{1,p}(I)$.

Assume first that $d_{\text{in}}' < d_{\text{in}}$. Fubini's theorem gives

$$\int_{I_{d_{\text{in}}-d_{\text{in}}'}} \|u_k(\cdot, t)\|^p_{W^{1,p}(I_{d_{\text{in}}'})}\, dt \le \|u_k\|^p_{W^{1,p}(I)}, \tag{B.4}$$

because the derivatives tangent to the slice are a subset of the full weak derivative. Hence one may choose an *interior* point $t_k \in (0,1)^{d_{\text{in}}-d_{\text{in}}'}$ such that $\|u_k(\cdot, t_k)\|_{W^{1,p}(I_{d_{\text{in}}'})} \le \|u_k\|_{W^{1,p}(I)}$. Thus $\pi_{\ell,d_{\text{out}}} e_k(\cdot, t_k) \to F'$ in $W^{1,p}(I_{d_{\text{in}}'})$. When $d_{\text{in}}' = d_{\text{in}}$, simply omit the slicing step.

Let $\bar{e}_k(x) = e_k(x, t_k)$ when $d_{\text{in}}' < d_{\text{in}}$, and let $\bar{e}_k = e_k$ when $d_{\text{in}}' = d_{\text{in}}$. Write $\bar{e}_k(x) = (a_k(x), b_k(x)) \in \mathbb{R}^{d_{\text{out}}} \times \mathbb{R}^{\ell - d_{\text{out}}}$ and define $E_k(x) := \big(a_k(x), F''(x), b_k(x)\big) \in \mathbb{R}^{\ell+r}$. Projection of E_k onto its (a_k, b_k) coordinates recovers $\bar{e}_k$. Restriction to a slice preserves injectivity and injectivity of the differential, so $\bar{e}_k$ is an embedding in either case. It follows immediately that E_k is an embedding as well. Its first d_{out}' coordinates satisfy $(a_k, F'') \to (F', F'') = F$ in $W^{1,p}$, with the second block represented exactly. Equation (A.4) therefore identifies this projected embedding with a member of $\mathcal{A}^{\ell+r}_{d_{\text{in}}', d_{\text{out}}'}$. Hence $\Omega_{1,p}(d_{\text{in}}', d_{\text{out}}') \le \ell + r$, which is (B.3). □

B.3 Proof of Corollary 3.3

Put $W := w^{1,p}_{\min}(d_{\text{in}}, d_{\text{out}}; \Sigma) = \Omega_{1,p}(d_{\text{in}}, d_{\text{out}})$, where the equality is Proposition B.1. We first prove the upper bounds. For every smooth target $f : I \to \mathbb{R}^{d_{\text{out}}}$, the graph $x \mapsto (f(x), x)$ is an embedding into $\mathbb{R}^{d_{\text{out}}+d_{\text{in}}}$, so smooth density and (A.4) give

$$\Omega_{1,p}(d_{\text{in}}, d_{\text{out}}) \le d_{\text{in}} + d_{\text{out}}. \tag{B.5}$$

If $d_{\text{out}} \le 2d_{\text{in}} - 1$, pad the target with $2d_{\text{in}} - 1 - d_{\text{out}}$ zero output coordinates, apply Theorem A.1, and discard the added visible coordinates. This gives the bounds $2d_{\text{in}} - 1$, $2d_{\text{in}}$, and $2d_{\text{in}} + 1$ in the three exponent regimes of Corollary 3.3; combining them with (B.5) gives the stated minima.

If $d_{\text{out}} = 2d_{\text{in}}$, Lemma A.9 gives $\Omega_{1,p}(d_{\text{in}}, 2d_{\text{in}}) = 2d_{\text{in}}$ for $p \le d_{\text{in}}$. For $p > d_{\text{in}}$, Lemma A.2(2) applied to $(f, 0)$ gives width $2d_{\text{in}} + 1$. If $d_{\text{out}} \ge 2d_{\text{in}} + 1$, the same jet and multijet transversality argument makes embeddings $I \hookrightarrow \mathbb{R}^{d_{\text{out}}}$ C^1-dense, so $\Omega_{1,p}(d_{\text{in}}, d_{\text{out}}) = d_{\text{out}}$. This proves all upper bounds.

For the lower bounds, assume $d_{\text{in}} \le d_{\text{out}} \le 2d_{\text{in}} - 1$ and set $j^* := \max\{3, \lceil (d_{\text{out}} + 1)/2 \rceil\}$. For any $j \in \{j^*, \dots, d_{\text{in}}\}$, Lemma B.2 with input dimension j and output dimension $2j - 1$ gives

$$W \ge \Omega_{1,p}(j, 2j-1) - (2j - 1 - d_{\text{out}}). \tag{B.6}$$

Taking $j = j^*$, if $p > j^* - 1$, Theorem A.1 gives $\Omega_{1,p}(j^*, 2j^* - 1) \ge 2j^*$, and therefore $W \ge 2j^* - (2j^* - 1 - d_{\text{out}}) = d_{\text{out}} + 1$. If there is an odd $j \in \{j^*, \dots, d_{\text{in}}\}$ with $p > j$, the odd-dimensional case of Theorem A.1 gives $\Omega_{1,p}(j, 2j-1) = 2j+1$, and (B.6) yields $W \ge (2j+1) - (2j-1-d_{\text{out}}) = d_{\text{out}} + 2$. These are exactly the lower bounds in Corollary 3.3.

C Proof of Theorem 3.1

In this section, we prove the upper bound for $3d_{\text{in}}/2 < d_{\text{out}}$. Identify $I := [0,1]^{d_{\text{in}}}$ with a cube in a coordinate chart of $S^{d_{\text{in}}}$, and denote its inclusion by $\iota_I : I \to S^{d_{\text{in}}}$. Our goal in this section is to prove Theorem 3.1.

C.1 Generic maps

We consider the case where $3d_{\text{in}}/2 < d_{\text{out}} < 2d_{\text{in}}$. If $d_{\text{out}} > 2d_{\text{in}}$, transversality makes smooth embeddings $S^{d_{\text{in}}} \to \mathbb{R}^{d_{\text{out}}}$ dense in C^1. Therefore

$$\Omega_{1,p}(d_{\text{in}}, d_{\text{out}}) = d_{\text{out}} \qquad (1 \le p < \infty).$$

If $d_{\text{out}} = 2d_{\text{in}}$, it is obvious from Lemma A.9.

We first consider a sufficiently well-behaved generic smooth map whose singular and self-intersection sets are well understood. The following lemma is a special case of a more general result of Wall (2016, Theorem 4.7.7(iv), pp. 119–121).

Put $q := d_{\text{out}} - d_{\text{in}}$ and $d := 2d_{\text{in}} - d_{\text{out}}$. Thus, we have $d_{\text{in}} = d + q, d_{\text{out}} = d + 2q$ for $1 \le d < q$. For a smooth map $f : M^{d_{\text{in}}} \to \mathbb{R}^{d_{\text{out}}}$, set

$$\Sigma_f := \{x \in M : \operatorname{rank} Df_x < d_{\text{in}}\}, \qquad D_f := \{x \in M : \exists y \neq x,\ f(y) = f(x)\}.$$

Lemma C.1. *Let $M^{d_{\text{in}}}$ be closed and $3d_{\text{in}}/2 < d_{\text{out}} < 2d_{\text{in}}$. Every smooth map $M \to \mathbb{R}^{d_{\text{out}}}$ can be approximated arbitrarily closely in C^1 by a smooth map f satisfying $|f^{-1}(z)| \le 2$ for $z \in \mathbb{R}^{d_{\text{out}}}$, and we have*

$$\left.\begin{array}{c} x \neq y, \\ f(x) = f(y) \end{array}\right\} \implies \begin{cases} \operatorname{rank} Df_x = \operatorname{rank} Df_y = d_{\text{in}}, \\ Df_x(T_xM) + Df_y(T_yM) = \mathbb{R}^{d_{\text{out}}}. \end{cases}$$

Moreover,

$$\widehat{D}_f := \overline{D_f} = D_f \cup \Sigma_f$$

is a compact embedded smooth d-manifold without boundary.

Proof. By Wall (2016, Theorem 4.7.7(iv), pp. 119–121), there is a dense set of maps satisfying the two displayed conditions for which

$$\dim D_f = 2d_{\text{in}} - d_{\text{out}} = d, \qquad \overline{D_f} = D_f \cup \Sigma_f$$

is a smooth submanifold of M. It is compact because M is compact. □

Removing $\widehat{D}_f$ from the domain of f would yield an injective immersion. For technical reasons, however, we only remove a suitably chosen smooth submanifold $S \subset \widehat{D}_f$ such that every component of S has nonempty boundary. The following lemma shows that such an S can be chosen while ensuring that f remains an injective immersion on its complement.

Lemma C.2. *Let f satisfy the conclusions of Lemma C.1. There is a compact embedded smooth d-manifold with boundary $S \subset \widehat{D}_f$ such that every component of S has nonempty boundary, $\Sigma_f \subset S$, and*

$$x \neq y, \quad f(x) = f(y) \quad \implies \quad \{x, y\} \cap S \neq \varnothing.$$

Consequently, $f|_{M \setminus S}$ is an injective immersion.

Proof. If $\widehat{D}_f = \varnothing$, take $S = \varnothing$. Assume henceforth that $\widehat{D}_f \neq \varnothing$, and let $C_1, \dots, C_N$ be its connected components. There are finitely many such components because $\widehat{D}_f$ is a compact manifold. Moreover, $D_f \cap C_i$ is dense in C_i for every i.

We choose points $x_i \in C_i \cap D_f$ inductively. Suppose that $x_1, \dots, x_{i-1}$ have been chosen, and set

$$E_{i-1} := \bigcup_{j<i} f^{-1}\big(\{f(x_j)\}\big).$$

Since every fiber of f contains at most two points, E_{i-1} contains at most $2(i-1)$ points. Since $d \ge 1$, the manifold C_i is positive-dimensional. Therefore the dense set $D_f \cap C_i$ cannot be contained in the finite set E_{i-1}, and we may choose

$$x_i \in (C_i \cap D_f) \setminus E_{i-1}.$$

It follows that $f(x_1), \dots, f(x_N)$ are pairwise distinct.

Choose pairwise disjoint open neighborhoods $V_i \subset \mathbb{R}^{d_{\text{out}}}$ of $f(x_i)$. Since $x_i \in D_f$, the conclusions of Lemma C.1 imply that $\operatorname{rank} Df_{x_i} = d_{\text{in}}$. Thus f is an embedding on some neighborhood of x_i. After shrinking this neighborhood, choose a smoothly embedded closed coordinate d-ball $B_i \subset C_i$, containing x_i in its relative interior, such that

$$B_i \subset C_i \setminus \Sigma_f, \qquad f|_{B_i} \text{ is injective}, \qquad f(B_i) \subset V_i.$$

Set

$$B := \bigcup_{i=1}^{N} B_i.$$

Then $f|_B$ is injective. Now define

$$S := \widehat{D}_f \setminus \bigcup_{i=1}^{N} \operatorname{int} B_i,$$

where $\operatorname{int} B_i$ denotes the relative interior of B_i in C_i. For each i, the set $C_i \setminus \operatorname{int} B_i$ is connected: if $d = 1$, then C_i is a circle, while if $d \geq 2$, any path crossing B_i can be rerouted along the connected sphere ∂B_i. Consequently, these sets are precisely the components of S, and each has nonempty boundary ∂B_i. Thus S is a compact embedded smooth d-manifold with boundary. Moreover, $\Sigma_f \subset S$, because every B_i is contained in $C_i \setminus \Sigma_f$.

Finally, suppose that $x \neq y$ and $f(x) = f(y)$. Then $x, y \in D_f \subset \widehat{D}_f$. If neither point belonged to S, both would belong to

$$\bigcup_{i=1}^{N} \operatorname{int} B_i \subset B,$$

contradicting the injectivity of $f|_B$. Hence $\{x, y\} \cap S \neq \varnothing$. Since S also contains Σ_f, the restriction $f|_{M \setminus S}$ is an injective immersion. □

C.2 Sobolev suppressing

Throughout this subsection, the unit sphere

$$S^{d_{\text{in}}} := \{x \in \mathbb{R}^{d_{\text{in}}+1} : |x| = 1\}$$

is equipped with the Riemannian metric induced by the Euclidean inner product on $\mathbb{R}^{d_{\text{in}}+1}$. Unless stated otherwise, orthogonal complements and projections, normal-bundle norms and radii, Levi–Civita and normal connections, and volume forms on $S^{d_{\text{in}}}$ and its submanifolds are taken with respect to this metric and its restrictions. Sobolev norms of sphere-valued maps, however, are computed extrinsically using the inclusion $S^{d_{\text{in}}} \subset \mathbb{R}^{d_{\text{in}}+1}$.

Definition C.3 (Bundle metric and compatible connection)**.** Let $\pi : E \to N$ be a smooth real vector bundle, with fiber $E_y := \pi^{-1}(y)$. A *bundle metric* is a smooth map

$$\langle \cdot, \cdot \rangle : E \times_N E \longrightarrow \mathbb{R}, \qquad E \times_N E := \{(v, w) \in E \times E : \pi(v) = \pi(w)\},$$

whose restriction $\langle \cdot, \cdot \rangle_y : E_y \times E_y \to \mathbb{R}$ is an inner product for every $y \in N$. Thus, for $\xi, \zeta \in \Gamma^\infty(E)$, the notation $\langle \xi, \zeta \rangle$ denotes the smooth function

$$y \longmapsto \langle \xi(y), \zeta(y) \rangle_y \quad \text{in } C^\infty(N).$$

Here $\Gamma^\infty(E)$ denotes the space of smooth sections of E, and $\Gamma^\infty(TN)$ the space of smooth vector fields on N. A connection on E is an $\mathbb{R}$-linear map

$$\nabla : \Gamma^\infty(E) \longrightarrow \Gamma^\infty(T^*N \otimes E)$$

satisfying

$$\nabla(f\xi) = df \otimes \xi + f \nabla \xi \qquad \text{for } f \in C^\infty(N),\ \xi \in \Gamma^\infty(E).$$

For $X \in \Gamma^\infty(TN)$ and $\xi \in \Gamma^\infty(E)$, define $\nabla_X \xi := (\nabla \xi)(X) \in \Gamma^\infty(E)$. Equivalently,

$$\Gamma^\infty(TN) \times \Gamma^\infty(E) \longrightarrow \Gamma^\infty(E), \qquad (X, \xi) \longmapsto \nabla_X \xi,$$

is $\mathbb{R}$-bilinear and, for all $f \in C^\infty(N)$, satisfies

$$\nabla_{fX} \xi = f \nabla_X \xi, \qquad \nabla_X (f\xi) = X(f)\xi + f \nabla_X \xi.$$

It is *compatible* with the bundle metric if, for all $X \in \Gamma^\infty(TN)$ and $\xi, \zeta \in \Gamma^\infty(E)$,

$$X\big(\langle \xi, \zeta \rangle\big) = \langle \nabla_X \xi, \zeta \rangle + \langle \xi, \nabla_X \zeta \rangle.$$

For a smooth curve $\gamma : [0, 1] \to N$, define the pullback bundle by

$$\gamma^* E := \{(t, v) \in [0, 1] \times E : \pi(v) = \gamma(t)\}, \qquad p_\gamma : \gamma^* E \longrightarrow [0, 1], \quad p_\gamma(t, v) := t.$$

A smooth section $t \mapsto (t, V(t))$ of $\gamma^* E$ is canonically identified with a smooth map $V : [0,1] \to E$ satisfying $\pi \circ V = \gamma$; thus $V(t) \in E_{\gamma(t)}$ for every $t \in [0,1]$. The connection ∇ determines the covariant derivative along γ,

$$\frac{D^\nabla}{dt} : \Gamma^\infty(\gamma^* E) \longrightarrow \Gamma^\infty(\gamma^* E).$$

For $v \in E_{\gamma(0)}$, parallel transport is the map

$$P_\gamma : E_{\gamma(0)} \longrightarrow E_{\gamma(1)}, \qquad P_\gamma v := V(1),$$

where $V \in \Gamma^\infty(\gamma^* E)$ is the unique section satisfying $V(0) = v$ and $\frac{D^\nabla V}{dt} = 0$. Compatibility implies that P_γ is a linear isometry (Jost, 2011, Lemma 4.2.1). If $N \subset S^{d_{\mathrm{in}}} \subset \mathbb{R}^{d_{\mathrm{in}}+1}$ is a smooth embedded submanifold, possibly with boundary, then

$$\nu_y N := \left\{ z \in T_y S^{d_{\mathrm{in}}} : \langle z, w \rangle_{\mathbb{R}^{d_{\mathrm{in}}+1}} = 0 \text{ for every } w \in T_y N \right\}$$

defines its normal bundle $\nu N \to N$. The metric specified above restricts to a bundle metric on νN. Let $\overline{\nabla}$ be the Levi–Civita connection of this metric on $S^{d_{\mathrm{in}}}$. Fiberwise orthogonal projection then gives the normal connection

$$\nabla^\perp : \Gamma^\infty(TN) \times \Gamma^\infty(\nu N) \longrightarrow \Gamma^\infty(\nu N), \qquad \nabla^\perp_X \xi := \mathrm{pr}_{\nu N}\left(\overline{\nabla}_X \xi\right),$$

where $\overline{\nabla}$ is understood along N and $\mathrm{pr}_{\nu N} : TS^{d_{\mathrm{in}}}|_N \to \nu N$ is the fiberwise orthogonal projection. This connection is compatible with the restricted metric (Lee, 2018, Propositions 2.16 and 8.6).

Using Lemma A.4, we extend a diffeomorphism of a submanifold to the ambient sphere with an arbitrarily small $W^{1,p}$ change on the embedded cube.

Lemma C.4. *Let $N^r \subset S^{d_{\mathrm{in}}}$ be compact, possibly with boundary, and let $h \in \mathsf{Diff}^\infty(N)$ be smoothly isotopic to id_N through diffeomorphisms that fix a common neighborhood of ∂N. Let $\kappa : I \to S^{d_{\mathrm{in}}}$ be a fixed smooth embedding. If $\ell := d_{\mathrm{in}} - r \geq 2$, $1 \leq p \leq \ell$, then for every $\eta > 0$, there is $H \in \mathsf{Diff}^\infty(S^{d_{\mathrm{in}}})$ such that*

$$H|_N = h, \qquad \|H \circ \kappa - \kappa\|_{W^{1,p}(I;\mathbb{R}^{d_{\mathrm{in}}+1})} < \eta.$$

For any chosen isotopy h_t as above, H can be chosen equal to the identity outside any prescribed neighborhood in $S^{d_{\mathrm{in}}}$ of

$$C := \overline{\{y \in N : h_t(y) \neq y \text{ for some } t \in [0,1]\}}.$$

Proof. Fix such an isotopy h_t, $0 \leq t \leq 1$, with $h_0 = \mathsf{id}_N$ and $h_1 = h$. Put

$$C := \overline{\{y \in N : h_t(y) \neq y \text{ for some } t\}} \Subset \operatorname{int} N.$$

Fix an open neighborhood $\mathcal{V}$ of C in $S^{d_{\mathrm{in}}}$, and choose open sets $U, U_1 \subset N$ such that

$$C \Subset U \Subset U_1 \Subset (\operatorname{int} N) \cap \mathcal{V}.$$

Since every h_t fixes $N \setminus C$ and is bijective, $h_t(U) = U$, and h_t is the identity near ∂U. Equip νN with the bundle metric and compatible normal connection defined above. Parallel transport along $u \mapsto h_{ut}(y)$ gives smooth linear isometries

$$P_{y,t} : \nu_y N \longrightarrow \nu_{h_t(y)} N.$$

By the tubular neighborhood theorem and $\overline{U_1} \Subset \operatorname{int} N$, choose $R > 0$ such that, for $W \in \{U, U_1\}$, writing

$$\mathbb{D}_R(\nu N|_W) := \{(y, v) : y \in W,\ v \in \nu_y N,\ |v| < R\},$$

the normal exponential map restricts to a smooth embedding

$$\Theta : \mathbb{D}_R(\nu N|_{U_1}) \longrightarrow S^{d_{\mathrm{in}}}, \qquad \Theta(y, v) := \exp_y^{S^{d_{\mathrm{in}}}}(v).$$

This map is a diffeomorphism onto its open image, which is contained in $\mathcal{V}$. It satisfies $\Theta(y, 0) = y$ and

$$\Theta^{-1}(N) = \{(y, 0) : y \in U_1\}.$$

Restricting Θ to the smaller base set, put

$$\mathcal{T} := \Theta\big(\mathbb{D}_R(\nu N|_U)\big).$$

Thus $\mathcal{T}$ is an open subset of $S^{d_{\text{in}}}$ contained in the prescribed neighborhood $\mathcal{V}$, and $\mathcal{T} \cap N = U$.

Since $h_t(U) = U$ and $P_{y,t}$ is an isometry from $\nu_y N$ to $\nu_{h_t(y)} N$, the formula

$$\widetilde{h}_t : \mathbb{D}_R(\nu N|_U) \longrightarrow \mathbb{D}_R(\nu N|_U), \qquad \widetilde{h}_t(y, v) := \big(h_t(y), P_{y,t} v\big),$$

defines a smooth bundle diffeomorphism for every $t \in [0, 1]$. Its inverse is

$$\widetilde{h}_t^{-1}(y, v) = \big(h_t^{-1}(y), P^{-1}_{h_t^{-1}(y),t} v\big).$$

Each $\widetilde{h}_t$ preserves the normal radius $|v|$.

Fix $0 < R_0 < R$ and let a_j be given by Lemma A.4 with $H = 1$. Choose $0 < \rho_j < R_0$ such that $a_j = 1$ for $|z| \le \rho_j$. Write $a_j(w) = \alpha_j(|w|)$ for $w \in B^\ell_R$, and define the smooth fiberwise radial function

$$A_j : \mathbb{D}_R(\nu N|_U) \longrightarrow [0, 1], \qquad A_j(y, v) := \alpha_j(|v|).$$

This definition is smooth along the zero section because α_j is constant near zero. On the disk bundle, set

$$\widehat{H}_j(y, v) := \widetilde{h}_{A_j(y,v)}(y, v).$$

Because $\widetilde{h}_t$ preserves $|v|$, the function A_j has the same value at a point and at its image under $\widetilde{h}_t$. Consequently $\widehat{H}_j$ is a diffeomorphism, with

$$\widehat{H}_j^{-1}(y, v) = \widetilde{h}^{-1}_{A_j(y,v)}(y, v).$$

Moreover, h_t is the identity off C, and the parallel transport along the resulting constant path is also the identity there. Since $A_j = 0$ when $|v| \ge R_0$, we have

$$\operatorname{supp}\big(\widehat{H}_j - \mathsf{id}\big) \subset \{(y, v) : y \in C,\ v \in \nu_y N,\ |v| \le R_0\} \Subset \mathbb{D}_R(\nu N|_U).$$

Now define the ambient map

$$H_j(x) := \begin{cases} \Theta \circ \widehat{H}_j \circ \Theta^{-1}(x), & x \in \mathcal{T}, \\ x, & x \in S^{d_{\text{in}}} \setminus \mathcal{T}. \end{cases}$$

The two branches agree on a neighborhood of $\partial\mathcal{T}$. Conjugating the displayed inverse of $\widehat{H}_j$ gives a smooth global inverse for H_j. Hence $H_j \in \mathsf{Diff}^\infty(S^{d_{\text{in}}})$ and

$$\operatorname{supp}(H_j - \mathsf{id}) \subset \Theta\big(\{(y, v) : y \in C,\ v \in \nu_y N,\ |v| \le R_0\}\big) \Subset \mathcal{T} \subset \mathcal{V}.$$

In the bundle coordinates supplied by Θ, the ambient map H_j is represented by $\widehat{H}_j$; explicitly,

$$\widehat{H}_j(y, v) = \begin{cases} \widetilde{h}_1(y, v) = \big(h(y), P_{y,1} v\big), & |v| \le \rho_j, \\ \widetilde{h}_{A_j(y,v)}(y, v), & \rho_j < |v| < R_0, \\ \widetilde{h}_0(y, v) = (y, v), & |v| \ge R_0. \end{cases}$$

Thus, it agrees with the full lifted diffeomorphism near the zero section, interpolates through the lifted isotopy on the transition annulus, and is the identity for $|v| \ge R_0$. For $y \in U$,

$$H_j(y) = \Theta\big(\widehat{H}_j(y, 0)\big) = \Theta(h(y), 0) = h(y).$$

For $y \in N \setminus U$, both $H_j(y)$ and $h(y)$ equal y. Thus $H_j|_N = h$. Moreover,

$$\big|P_{y,A_j(y,v)} v\big| = |v|,$$

so $\widehat{H}_j$ preserves the normal radius.

Put $K := \kappa(I)$. Since κ is a fixed smooth embedding of the compact d_{in}-cube I into the d_{in}-manifold $S^{d_{\mathrm{in}}}$, it is a diffeomorphism onto K, and the relevant change-of-variables Jacobians are uniformly bounded. Regarding $H_j - \mathsf{id}_{S^{d_{\mathrm{in}}}}$ as an $\mathbb{R}^{d_{\mathrm{in}}+1}$-valued map, the chain rule gives

$$D(H_j \circ \kappa - \kappa)(x) = D\big(H_j - \mathsf{id}_{S^{d_{\mathrm{in}}}}\big)_{\kappa(x)} D\kappa(x).$$

Consequently, the effect of κ can be absorbed into a constant $C_\kappa > 0$, and

$$\begin{aligned}\|H_j \circ \kappa - \kappa\|^p_{W^{1,p}(I;\mathbb{R}^{d_{\mathrm{in}}+1})} &= \int_I \left(|H_j(\kappa(x)) - \kappa(x)|^p + |D(H_j \circ \kappa - \kappa)(x)|^p\right) dx \\ &\le C_\kappa \int_K \left(|H_j(x) - x|^p + \big|D\big(H_j - \mathsf{id}_{S^{d_{\mathrm{in}}}}\big)_x\big|^p\right) d\,\mathrm{vol}_{S^{d_{\mathrm{in}}}}(x).\end{aligned}$$

Thus, up to the fixed multiplicative constant C_κ, it suffices to estimate $H_j - \mathsf{id}$ on K.

Define the smooth map

$$\mathcal{G} : \mathbb{D}_R(\nu N|_U) \times [0,1] \longrightarrow S^{d_{\mathrm{in}}} \subset \mathbb{R}^{d_{\mathrm{in}}+1}, \qquad \mathcal{G}(u,t) := \Theta\big(\widetilde{h}_t(u)\big).$$

Choose finitely many pairs of open sets $V_\alpha \Subset W_\alpha \Subset U$, $\alpha = 1, \dots, N_{\mathrm{ch}}$, such that the V_α cover C, together with orthonormal bundle trivializations

$$\tau_\alpha : \nu N|_{W_\alpha} \xrightarrow{\cong} W_\alpha \times \mathbb{R}^\ell, \qquad \tau_\alpha(y,v) = \big(y, \tau_{\alpha,y} v\big),$$

where each $\tau_{\alpha,y} : \nu_y N \to \mathbb{R}^\ell$ is a linear isometry. Since $\widehat{H}_j - \mathsf{id}$ is supported over C, it suffices to estimate in these source charts. Fix one such trivialization τ_α. For $y \in V_\alpha$ and $z \in B^\ell_R$, set

$$v := \tau^{-1}_{\alpha,y} z \in \nu_y N, \qquad u := (y,v) = \tau^{-1}_\alpha(y,z) \in \mathbb{D}_R(\nu N|_{V_\alpha}).$$

Thus y is the base point, v is the actual normal vector, z is its coordinate vector in $\mathbb{R}^\ell$, and u is the corresponding point of the disk bundle. Since $\tau_{\alpha,y}$ is an isometry, the previously defined fiberwise function A_j has the coordinate expression

$$(A_j \circ \tau^{-1}_\alpha)(y,z) = A_j(u) = \alpha_j(|v|) = \alpha_j(|z|) = a_j(z).$$

In particular,

$$D(A_j \circ \tau^{-1}_\alpha)_{(y,z)} = \big(0, Da_j(z)\big).$$

The point $\widetilde{h}_t(u)$ need not lie in the chosen bundle chart, but $\mathcal{G}$ is defined globally on the disk bundle. We regard its values as lying in $\mathbb{R}^{d_{\mathrm{in}}+1}$, so all differences and derivatives below are taken in the ambient Euclidean space. For $x = \Theta(u)$, the definitions of H_j, $\widehat{H}_j$, and $\mathcal{G}$ give

$$H_j(x) = \Theta\big(\widehat{H}_j(u)\big) = \Theta\big(\widetilde{h}_{A_j(u)}(u)\big) = \mathcal{G}\big(u, A_j(u)\big), \qquad x = \Theta(u) = \mathcal{G}(u,0).$$

Consequently,

$$\begin{aligned}H_j(x) - x &= \mathcal{G}\big(u, A_j(u)\big) - \mathcal{G}(u,0) \\ &= \int_0^{A_j(u)} \partial_t \mathcal{G}(u,t)\, dt.\end{aligned}$$

The chain rule and the fundamental theorem of calculus also give

$$\begin{aligned}D_u\left[(H_j - \mathsf{id}_{S^{d_{\mathrm{in}}}}) \circ \Theta\right](u) = &\int_0^{A_j(u)} \partial_t D_u \mathcal{G}(u,t)\, dt \\ &+ \partial_t \mathcal{G}\big(u, A_j(u)\big) \otimes DA_j(u).\end{aligned}$$

The support lies in the fixed compact set $\{(y,v) : y \in C,\ v \in \nu_y N,\ |v| \le R_0\}$, and $t \in [0,1]$. Smoothness of $\mathcal{G}$ on the finitely many chart closures and bounded $D\Theta^{-1}$ therefore yield

$$|H_j(x) - x| + \big|D\big(H_j - \mathsf{id}_{S^{d_{\mathrm{in}}}}\big)_x\big| \le C_0\big(A_j(u) + |DA_j(u)|\big).$$

Since $H_j = \mathsf{id}$ outside $\mathcal{T}$, change variables by Θ on its support and sum over the finitely many orthonormal trivializations chosen above. Using the coordinate expression for A_j displayed above, the bounded Jacobians and the finite overlap are absorbed into a constant. Combining this observation with the preceding estimates gives

$$\begin{aligned}\|H_j \circ \kappa - \kappa\|^p_{W^{1,p}(I;\mathbb{R}^{d_{\mathrm{in}}+1})} &\le C_1 \int_U \int_{B^\ell_R} \big(a_j(z)^p + |Da_j(z)|^p\big)\, dz\, d\,\mathrm{vol}_N(y) \\ &\le C_1\, \mathrm{vol}_N(U) \left(\|a_j\|^p_{L^p(B^\ell_R)} + \|Da_j\|^p_{L^p(B^\ell_R)}\right) \longrightarrow 0.\end{aligned}$$

Taking a sufficiently large j and setting $H := H_j$ proves the norm estimate. Since $\mathcal{V}$ was arbitrary and $\mathrm{supp}(H_j - \mathsf{id}) \subset \mathcal{V}$, it also proves the final support assertion. □

C.3 DETOUR AROUND A FINITE COMPLEX

To construct the detours, we perturb auxiliary maps into embeddings that avoid prescribed obstacles while preserving a fixed region.

Lemma C.5. *Let X^d be a compact smooth manifold, possibly with corners, and let $Y^{d_{\mathrm{in}}}$ be a smooth manifold with $2d < d_{\mathrm{in}}$. Let $Z \subset Y$ be a finite union of locally closed smooth submanifolds of dimension less than $d_{\mathrm{in}} - d$. Suppose $f_0 : X \to Y$ is smooth and, for some neighborhood U of a compact set $C \subset X$, the restriction $f_0|_U$ is an embedding and $f_0(U \setminus C) \cap Z = \varnothing$. Then f_0 can be approximated arbitrarily closely in C^∞ by smooth embeddings $f : X \to Y$ agreeing with f_0 near C and satisfying $f(X \setminus C) \cap Z = \varnothing$.*

Proof. Working with smooth extensions near X, apply jet and multijet transversality using perturbations supported away from a smaller neighborhood of C (Golubitsky & Guillemin, 1973, Chapter II, Corollary 4.11 and Theorem 4.13). The dimension assumptions exclude rank loss, intersections with Z outside the fixed neighborhood, and double points; for a pair with one point fixed, perturbing the other still gives transversality to the diagonal. On the fixed neighborhood the required properties hold by assumption. Compactness of X turns the resulting injective immersion into an embedding. □

We first avoid a single simplex at arbitrarily small $W^{1,p}$ cost while fixing a neighborhood of the previously treated complex.

Lemma C.6. *Let $I = [0,1]^{d_{\mathrm{in}}}$ and let $\mathcal{T}$ be a smooth triangulation of $S^{d_{\mathrm{in}}}$. Let $\mathcal{A} \subset \mathcal{T}$ be a finite subcomplex with $\dim \mathcal{A} \le s$. Write $A := |\mathcal{A}|$. Let $\sigma \in \mathcal{T}$ be an s-simplex such that $\sigma \notin \mathcal{A}$ and every proper face of σ belongs to $\mathcal{A}$. Suppose that $\kappa : I \to S^{d_{\mathrm{in}}}$ is a smooth embedding and*

$$\kappa(I) \cap A = \varnothing, \qquad 2(s+1) < d_{\mathrm{in}}, \qquad 1 \le p \le d_{\mathrm{in}} - s - 1.$$

For every $\epsilon > 0$, there is $H \in \mathsf{Diff}^\infty(S^{d_{\mathrm{in}}})$ such that

$$H(\kappa(I)) \cap (A \cup \sigma) = \varnothing, \qquad \|H \circ \kappa - \kappa\|_{W^{1,p}(I;\mathbb{R}^{d_{\mathrm{in}}+1})} < \epsilon,$$

and $H = \mathsf{id}$ on a neighborhood of A.

Proof. Put $K := \kappa(I), O := S^{d_{\mathrm{in}}} \setminus K$. If $K \cap \sigma = \varnothing$, take $H = \mathsf{id}$. Otherwise, extend κ to a slightly larger smooth d_{in}-cell and apply Palais's transitivity theorem (Palais, 1960, Theorem B and Corollary 2). It gives an ambient diffeomorphism of $S^{d_{\mathrm{in}}}$ carrying the standard cube onto K. Consequently,

$$O \cong \mathbb{R}^{d_{\mathrm{in}}}, \qquad \pi_i(S^{d_{\mathrm{in}}}, O) = 0 \quad (1 \le i < d_{\mathrm{in}}).$$

Assume first that $s \ge 1$. Since $K \cap A = \varnothing$ and $\partial\sigma \subset A$, choose closed s-balls satisfying

$$K \cap \sigma \subset \operatorname{int} B_0, \qquad B_0 \Subset \operatorname{int} B_1, \qquad B_1 \Subset \operatorname{int} B_2, \qquad B_2 \Subset \operatorname{int} \sigma.$$

Put $R := B_2 \setminus \operatorname{int} B_1$. Since B_2 is contractible, there exists a smooth unit normal field ν to $\operatorname{int}\sigma$ along B_2. For some $\delta > 0$, the map

$$\chi : B_2 \times [-\delta, \delta] \longrightarrow S^{d_{\mathrm{in}}}, \qquad \chi(u,t) := \exp_u^{S^{d_{\mathrm{in}}}}(t\nu_u),$$

is an embedding such that

$$\chi(u,0) = u, \qquad \chi^{-1}(\sigma) = B_2 \times \{0\}, \qquad \chi\big(B_2 \times [-\delta,\delta]\big) \cap A = \varnothing,$$

and

$$\chi\big((B_2 \setminus \operatorname{int} B_0) \times [-\delta,\delta]\big) \subset O.$$

Put $X := B_2 \times [-1,1]$, $C := B_2 \times \{0\}$, and

$$L := \big((B_2 \setminus \operatorname{int} B_0) \times [-1,1]\big) \cup \big(B_2 \times ([-1,-1/2] \cup [1/2,1])\big).$$

Since $\pi_s(S^{d_{\mathrm{in}}}, O) = 0$, the maps $\chi(\cdot, \pm\delta)|_{B_0}$ can be homotoped into O as maps of pairs. After extending χ slightly, use homotopy extension and relative smooth approximation to obtain $F_0 : X \to S^{d_{\mathrm{in}}}$, smooth on a neighborhood of X, with $F_0 = \chi$ near C and $F_0(L) \Subset O$.

Apply Lemma C.5 to F_0, fixing C and taking $Z = A \cup \sigma$, decomposed into simplex interiors. The dimension conditions follow from $2(s+1) < d_{\text{in}}$ and $(s+1) + s < d_{\text{in}}$. A sufficiently small perturbation gives a smooth embedding $F : X \to S^{d_{\text{in}}}$, equal to χ near C, such that

$$F^{-1}(A) = \varnothing, \qquad F^{-1}(\sigma) = C, \qquad F(L) \Subset O.$$

Consequently,

$$F^{-1}(K) \Subset \operatorname{int} B_0 \times (-1/2, 1/2).$$

Round the corners of $F(B_2 \times [-3/4, 3/4])$, leaving a neighborhood of $F(C) \cup F(B_1 \times [-2/3, 2/3])$ unchanged. The resulting smooth closed $(s+1)$-ball W satisfies

$$W \cap A = \varnothing, \qquad W \cap \sigma = B_2, \qquad K \cap W \Subset \operatorname{int} W.$$

Moreover, $F(B_1 \times [-2/3, 2/3]) \Subset \operatorname{int} W$. Choose closed $(s+1)$-balls D, E inside $F(\operatorname{int} B_1 \times (-1/2, 1/2))$ and $F(\operatorname{int} B_1 \times (1/2, 2/3))$, respectively, with $K \cap W \subset \operatorname{int} D$. Then $D \cap R = \varnothing$, $E \cap (D \cup \sigma) = \varnothing$, and both lie in the connected set $F(\operatorname{int} B_1 \times (-2/3, 2/3)) \subset (\operatorname{int} W) \setminus R$.

If $s = 0$, relative general position and $2 < d_{\text{in}}$ give a smooth embedded arc $W \subset S^{d_{\text{in}}}$ satisfying

$$\sigma \in \operatorname{int} W, \qquad \partial W \subset O, \qquad W \cap A = \varnothing, \qquad W \cap \sigma = \{\sigma\}.$$

Put $R := \varnothing$. Choose disjoint closed intervals $D, E \subset \operatorname{int} W$ such that

$$K \cap W \subset \operatorname{int} D, \qquad E \cap (D \cup \sigma) = \varnothing.$$

In both cases, D, E lie in the same component of $(\operatorname{int} W) \setminus R$. By isotopy extension, choose a smooth isotopy h_t of W, supported in a common compact subset of this component, with $h_0 = \mathrm{id}_W$ and $h_1(D) \subset \operatorname{int} E$. Its support is disjoint from $A \cup \overline{\sigma \setminus W}$. Since $\operatorname{codim} W = d_{\text{in}} - s - 1 \geq 2$ and $p \leq d_{\text{in}} - s - 1$, apply Lemma C.4 with $N = W$, $h = h_1$, and $\eta = \epsilon$. The support assertion gives $H \in \mathsf{Diff}^\infty(S^{d_{\text{in}}})$ with $H|_W = h_1$, $H = \mathrm{id}$ near $A \cup \overline{\sigma \setminus W}$, and

$$\|H \circ \kappa - \kappa\|_{W^{1,p}(I;\mathbb{R}^{d_{\text{in}}+1})} < \epsilon.$$

Since $H(W) = W$ and $K \cap \sigma \subset W$,

$$H(K) \cap W = h_1(K \cap W) \subset E, \qquad H(K) \cap (\sigma \setminus W) = \varnothing.$$

Together with $E \cap \sigma = \varnothing$ and $H^{-1}(A) = A$, this proves $H(K) \cap (A \cup \sigma) = \varnothing$. □

Applying the preceding lemma simplex by simplex moves the embedded cube away from an entire finite complex.

Lemma C.7. *Let $P \subset S^{d_{\text{in}}}$ be the underlying set of a finite subcomplex of a smooth triangulation, with $\dim P \leq r$. Let $\kappa_0 : I \to S^{d_{\text{in}}}$ be a smooth embedding. If*

$$2(r+1) < d_{\text{in}}, \qquad 1 \leq p \leq d_{\text{in}} - r - 1,$$

then, for every $\varepsilon > 0$, there is $G \in \mathsf{Diff}^\infty(S^{d_{\text{in}}})$ such that

$$G(\kappa_0(I)) \cap P = \varnothing, \qquad \|G \circ \kappa_0 - \kappa_0\|_{W^{1,p}(I;\mathbb{R}^{d_{\text{in}}+1})} < \varepsilon.$$

Proof. If $P = \varnothing$, take $G = \mathrm{id}$. Otherwise, order its closed simplices $\sigma_1, \dots, \sigma_N$ by nondecreasing dimension and put

$$P_0 := \varnothing, \qquad P_j := \bigcup_{i=1}^{j} \sigma_i.$$

Every proper face of σ_j lies in P_{j-1}. All norms in this proof are taken in $W^{1,p}(I;\mathbb{R}^{d_{\text{in}}+1})$.

For the initial step, $\dim \sigma_1 = 0$. Apply Lemma C.6 with $A = P_0, \sigma = \sigma_1, \kappa = \kappa_0$. Here $2 \leq 2(r+1) < d_{\text{in}}$ and $p \leq d_{\text{in}} - r - 1 \leq d_{\text{in}} - 1$, so the lemma applies. Choose H_1 so that, for $\kappa_1 := H_1 \circ \kappa_0$,

$$\kappa_1(I) \cap P_1 = \varnothing, \qquad \|\kappa_1 - \kappa_0\|_{W^{1,p}} < 2^{-1}\varepsilon.$$

Suppose that $1 \le j < N$ and $\kappa_j(I) \cap P_j = \varnothing$. Set $s := \dim \sigma_{j+1}$. The ordering and the hypotheses give

$$\dim P_j \le s, \qquad 2(s+1) \le 2(r+1) < d_{\text{in}}, \qquad p \le d_{\text{in}} - r - 1 \le d_{\text{in}} - s - 1.$$

Lemma C.6, applied to $(A, \sigma, \kappa) = (P_j, \sigma_{j+1}, \kappa_j)$, gives H_{j+1}. Choose it so that $\kappa_{j+1} := H_{j+1} \circ \kappa_j$ satisfies

$$\kappa_{j+1}(I) \cap P_{j+1} = \varnothing, \qquad \|\kappa_{j+1} - \kappa_j\|_{W^{1,p}} < 2^{-(j+1)}\varepsilon.$$

This closes the induction.

Set $G := H_N \circ \cdots \circ H_1$. Then $G \circ \kappa_0 = \kappa_N$, and the triangle inequality gives

$$\|G \circ \kappa_0 - \kappa_0\|_{W^{1,p}} \le \sum_{j=1}^{N} \|\kappa_j - \kappa_{j-1}\|_{W^{1,p}} < \varepsilon.$$

Since $P_N = P$, the result follows. □

C.4 The boundary detour

To avoid a manifold with boundary, we first isolate a lower-dimensional spine; its complement supplies collar coordinates for the final displacement.

Lemma C.8. *Let $S \subset S^{d_{\text{in}}}$ be a compact embedded smooth d-manifold whose components all have nonempty boundary. There are a smooth triangulation of $S^{d_{\text{in}}}$, a finite subcomplex*

$$P \subset \operatorname{int} S, \qquad \dim P \le d - 1,$$

and a diffeomorphism

$$\Theta : \partial S \times [0, \infty) \longrightarrow S \setminus P, \qquad \Theta(y, 0) = y.$$

Proof. Choose a smooth triangulation of $S^{d_{\text{in}}}$ compatible with S (Whitehead, 1940, Theorem 6 and its Addendum; Theorems 7 and 10). Applied componentwise to the induced triangulation of S, the classical spine construction gives, after subdivision, a finite subcomplex $P \subset \operatorname{int} S$ of dimension at most $d - 1$ onto which S collapses (Hudson & Zeeman, 1964, Section 2, p. 727). The subdivisions of S extend to the ambient triangulation by Rourke & Sanderson (1972, Lemma 2.17, p. 19). These refinements remain smooth, so P is a subcomplex of a smooth triangulation of $S^{d_{\text{in}}}$.

Thus S is a smooth regular neighborhood of P in its smooth double, in the terminology of Lerario et al. (2024, Section 2). The construction in Lerario et al. (2024, proof of Lemma 10, pp. 913–914) gives a diffeomorphism $F : \partial S \times (0, 1] \to S \setminus P$. After reparametrizing the boundary factor so that $F(y, 1) = y$, set $\Theta(y, t) := F(y, e^{-t})$. □

Proposition C.9. *Let $S \subset S^{d_{\text{in}}}$ be a compact embedded smooth d-manifold such that every component of S has nonempty boundary. If $1 \le d < d_{\text{in}}/2$, for $1 \le p \le d_{\text{in}} - d$, then, for every $\eta > 0$, there is $\Psi \in \mathsf{Diff}^{\infty}(S^{d_{\text{in}}})$ such that*

$$\Psi(I) \cap S = \varnothing, \qquad \|\Psi \circ \iota_I - \iota_I\|_{W^{1,p}(I;\mathbb{R}^{d_{\text{in}}+1})} < \eta.$$

Proof. Let P and Θ be given by Lemma C.8. Since $\dim P \le d - 1$, the assumptions give

$$2((d-1)+1) = 2d < d_{\text{in}}, \qquad p \le d_{\text{in}} - d.$$

Apply Lemma C.7 with $r = d - 1$, $\kappa_0 = \iota_I$, and error $\eta/2$. We obtain $G \in \mathsf{Diff}^{\infty}(S^{d_{\text{in}}})$ such that, for

$$\kappa := G \circ \iota_I, \qquad K := \kappa(I),$$

$$K \cap P = \varnothing, \qquad \|\kappa - \iota_I\|_{W^{1,p}} < \frac{\eta}{2}.$$

Since $K \cap P = \varnothing$, compactness and the diffeomorphism Θ give $L > 0$ such that

$$K \cap S \subset \Theta(\partial S \times [0, L]).$$

Extend Θ, still denoted by Θ, across an exterior collar to $\partial S \times (-a, \infty)$, where the points with negative second coordinate lie outside S. Choose an increasing diffeomorphism $\varphi : (-a, \infty) \to (-a, \infty)$ such that, for some $0 < a_0 < a$ and $R > L$,

$$\operatorname{supp}(\varphi - \mathsf{id}) \Subset (-a_0, R), \qquad \varphi([0, L]) \subset (-a, 0).$$

Put

$$N := \Theta(\partial S \times [-a_0, R]).$$
$$h_t(\Theta(y, r)) := \Theta\big(y, (1-t)r + t\varphi(r)\big)$$

defines an isotopy of the compact d-manifold N, equal to the identity near ∂N, because

$$\frac{\partial}{\partial r}\big((1-t)r + t\varphi(r)\big) = (1-t) + t\varphi'(r) > 0.$$

Put

$$C := \overline{\{x \in N : h_t(x) \neq x \text{ for some } t\}} \Subset \operatorname{int} N.$$

Since $\varphi([0, L]) \subset (-a, 0)$,

$$\Theta(\partial S \times [0, L]) \subset C.$$

Choose an open $V \subset N$ with $C \Subset V \Subset \operatorname{int} N$, and choose a tube T over V so small that

$$\overline{T} \cap P = \varnothing, \qquad T \cap S = \{\Theta(y, r) \in V : r \geq 0\}.$$

The normal codimension of N is $d_{\text{in}} - d \geq 2$. Apply the construction in Lemma C.4 to $h := h_1$, using the isotopy h_t, this tube, and the fixed embedding κ. Choose H such that

$$\|H \circ \kappa - \kappa\|_{W^{1,p}} < \frac{\eta}{2}.$$

Outside T the map H is the identity. Inside T, its construction preserves the normal radius; on the zero section, the choices above send $0 \leq r \leq L$ to $r < 0$ and preserve $r < 0$. Hence

$$H(K) \cap S = \varnothing.$$

With $\Psi := H \circ G$, the triangle inequality gives the required estimate. □

C.5 Proof of Theorem 3.1

Proof. Let $u \in W^{1,p}(I; \mathbb{R}^{d_{\text{out}}})$ and $\varepsilon > 0$. Choose a smooth map f_0 on a neighborhood of I such that

$$\|f_0 - u\|_{W^{1,p}} < \frac{\varepsilon}{3}.$$

Extend f_0 smoothly to $S^{d_{\text{in}}}$. Lemma C.1 gives $f : S^{d_{\text{in}}} \to \mathbb{R}^{d_{\text{out}}}$ such that

$$\|f|_I - f_0\|_{W^{1,p}} < \frac{\varepsilon}{3}.$$

Let $S \subset \widehat{D}_f$ be given by Lemma C.2. Since

$$d < q, \qquad p \leq q = d_{\text{in}} - d,$$

Proposition C.9 gives $\Psi_j \in \mathsf{Diff}^\infty(S^{d_{\text{in}}})$ such that

$$\psi_j := \Psi_j \circ \iota_I, \qquad \psi_j(I) \cap S = \varnothing, \qquad \psi_j \longrightarrow \iota_I \quad \text{in } W^{1,p}(I; \mathbb{R}^{d_{\text{in}}+1}).$$

Set

$$e_j := f \circ \psi_j.$$

By Lemma C.2, $f|_{S^{d_{\text{in}}} \setminus S}$ is an injective immersion. Hence e_j is a smooth embedding. The chain rule and the convergence of ψ_j give

$$e_j \longrightarrow f|_I \quad \text{in } W^{1,p}(I; \mathbb{R}^{d_{\text{out}}}).$$

For all sufficiently large j, the three estimates above imply

$$\|e_j - u\|_{W^{1,p}} < \varepsilon.$$

Since $d_{\text{in}} < d_{\text{out}}$, $\Phi_j \in \mathsf{Diff}^\infty(\mathbb{R}^{d_{\text{out}}})$ with

$$e_j = \Phi_j \circ \iota_{d_{\text{in}}, d_{\text{out}}}|_I.$$

Thus $e_j \in \mathcal{A}^{d_{\text{out}}}_{d_{\text{in}}, d_{\text{out}}}$, which proves the density. The definition of $\Omega_{1,p}$ gives $\Omega_{1,p}(d_{\text{in}}, d_{\text{out}}) = d_{\text{out}}$. □

For $d_{\text{out}} = 2d_{\text{in}}$, we remove isolated double points by modifying the target sheets, using the cutoff above in d_{in} dimensions.