Autonomous Flows: Exact Finite Interpolation and Uniform Approximation Obstructions

Qiaomu Liu* Qi Lü† Enrique Zuazua‡

Abstract

In dimension at least two, a class of locally Lipschitz vector fields realizes every finite correspondence between distinct inputs and distinct targets at any prescribed positive time, provided that it is linear, its members generate global flows, and it approximates every smooth vector field of bounded support uniformly on bounded sets. The input and target sets may overlap. The proof constructs a smooth autonomous reference flow and finitely many localized correction fields, then uses uniform stability and Brouwer degree to obtain exact endpoints within the approximating class. The argument uses only uniform approximation of the vector fields; it does not require control of their derivatives. Applied to shallow ReLU vector fields, the result gives exact finite interpolation without time-dependent coefficients or additional state variables, together with bounds on width and normalized coefficient strength. In contrast, no continuous autonomous semiflow can exchange and compress two disjoint balls. Its time maps therefore fail to approximate all continuous maps uniformly on compact sets. The results distinguish exact interpolation at finitely many points from uniform control of neighborhoods.

Keywords. neural ordinary differential equations; universal approximation; exact finite-data interpolation; simultaneous controllability; autonomous flows; shallow ReLU networks; uniform neighborhood-routing obstruction.

2020 Mathematics Subject Classification. 93B05, 34A12, 34H05, 41A30, 68T07.

1 Introduction

Finite-data interpolation asks whether a prescribed class of maps can carry each point of one finite configuration to a prescribed point of another. Neural ODEs provide one motivating setting [10, 14]. A control-theoretic treatment of classification, approximation, and transport for neural ODEs [25] provides an important starting point for the line of work to which the present paper contributes. For an autonomous flow, the terminal map has the form $F = \Phi_V^T$: one time-independent vector field must move all samples during one common time interval. This is substantially more rigid than arbitrary finite-point matching by diffeomorphisms, since a diffeomorphism carrying one configuration to another need not be the time-T map of an autonomous flow [7].

The main result of this paper is that this rigidity does not prevent exact finite interpolation in dimension $N \geq 2$. If a linear class of locally Lipschitz autonomous vector fields generates global flows and can approximate smooth compactly supported fields arbitrarily well on bounded sets, then every correspondence between finite configurations with pairwise distinct source points and pairwise distinct target points is realized exactly at every prescribed positive time; see Theorem 2.4. This includes points that remain fixed, assignments in which the target of one point is another input point, and finite cycles such as $x_1 \mapsto x_2 \mapsto x_3 \mapsto x_1$. Thus the source and target configurations need not be disjoint. Pairwise distinct targets are necessary because a flow map is injective.

*School of Mathematics, Sichuan University, Chengdu 610064, China. (Email: `liuqm@stu.scu.edu.cn`). Qiaomu Liu is supported by the NSF of China under grant 12025105.

†School of Mathematics, Sichuan University, Chengdu 610064, China. (Email: `lu@scu.edu.cn`). Qi Lü is supported by the NSF of China under grant 12025105.

‡Corresponding author. Chair for Dynamics, Control, Machine Learning and Numerics (Alexander von Humboldt Professorship), Department of Mathematics, Friedrich–Alexander–Universität Erlangen–Nürnberg, Cauerstrasse 11, 91058 Erlangen, Germany; Chair of Computational Mathematics, Fundación Deusto, Avenida de las Universidades 24, 48007 Bilbao, Spain; Departamento de Matemáticas, Universidad Autónoma de Madrid, Calle Francisco Tomás y Valiente 7, 28049 Madrid, Spain. (Email: `enrique.zuazua@fau.de`).

Proof strategy. The difficulty is to make all endpoint errors vanish while keeping the vector field in the prescribed class. We proceed in four steps.

1. **Construct a smooth reference flow.** Complete the finite correspondence into cycles and fixed points. Realize these by planar rotations and stationary points, then transport the model to the prescribed configuration by a smooth change of coordinates. After time rescaling, this reference flow reaches all targets at time one.
2. **Construct independent endpoint corrections.** Each moving sample has a short trajectory segment that no other marked sample visits during the unit time interval. Perturbations near these segments, together with bumps near stationary samples, vary every endpoint coordinate independently to first order at the smooth reference field.
3. **Approximate the correction fields first.** Choose correction fields from the prescribed class close enough to the ideal ones that their endpoint response matrix, evaluated at the smooth reference field, remains invertible. With these fields fixed, choose a small coefficient ball and then approximate the reference field accurately enough to control the endpoint error throughout that ball.
4. **Remove the endpoint error by degree.** On the boundary of the coefficient ball, the endpoint error differs from the invertible linear response by less than the size of that response. Brouwer degree therefore gives coefficients for which every endpoint error vanishes. Linearity keeps the corrected field in the prescribed class.

The key point is that uniform approximation suffices: we do not approximate derivatives or require differentiability of the endpoint map at a ReLU field. The variational calculation is performed at the smooth reference field, and the final step uses continuity and degree. For the quantitative bounds, we estimate perturbed trajectories in the standard rotation coordinates. This avoids an exponential loss involving the potentially large Lipschitz constant of the transported field. Section 3 gives the construction and estimates.

A ReLU application. Finite sums of ReLU ridge fields are globally Lipschitz and can approximate smooth vector fields arbitrarily well on bounded sets. Hence one time-independent shallow ReLU field exactly interpolates every admissible finite dataset in $\mathbb{R}^N$ for $N \geq 2$, at any prescribed positive time, without temporal switching or state augmentation. Theorem 2.10 gives quantitative bounds on width and time–strength; Proposition 5.2 gives a complementary bounded-distortion lower bound. In dimension one, order preservation rules out universal exact interpolation.

Rigidity at neighborhood scale. The flexibility at the point scale has a sharp limitation at the neighborhood scale. For a radius fixed before the field is chosen, a universal routing requirement would ask that each source ball be sent into a smaller ball around its assigned target. Theorem 2.7 shows that no continuous autonomous semiflow can exchange and compress two disjoint balls in this way: the semigroup identity and Brouwer's fixed-point theorem force a contradiction. The obstruction is structural and does not exclude data-dependent neighborhoods obtained after a particular interpolating flow has been selected. The distinction between pointwise and neighborhood control is a central motivation of recent work [6]: simultaneous control of neighborhoods or cells, rather than pointwise interpolation alone, is the stronger property relevant to universal approximation and generalization estimates. Corollary 2.9 consequently rules out uniform density of autonomous time maps in the continuous maps on compact sets. Thus exact interpolation of finitely many points does not supply the neighborhood control required for those stronger consequences; indeed, autonomy obstructs every universal fixed-radius version of it. This concerns the uniform topology and the full class $C(\mathcal{K}; \mathbb{R}^N)$, not weaker topologies or almost-everywhere approximation. Corollary 2.9 does not characterize which invertible targets are approximable.

Relation to previous work. Finite-ensemble controllability and approximation on mapping spaces have been studied through Lie-algebraic, time-dependent, and switched mechanisms [1, 2, 3, 16, 17]. These mechanisms do not face the same autonomous semigroup constraint. Earlier autonomous neural-ODE interpolation results impose additional geometric or architectural conditions [5]. Recent work [6] formulates the remaining gap that motivates the present study: it develops pointwise interpolation and simultaneous neighborhood control in its semi-autonomous setting, while the corresponding general exact-interpolation question for a single-layer autonomous system remains open there. By contrast, the criterion proved here applies to every linear, globally well-posed, class with this approximation property; shallow ReLU fields are its principal application. Approximation of diffeomorphisms is related but distinct: geometric finite-point relocation supplies diffeomorphisms moving finite configurations in dimension at least two [8, 21], whereas autonomous embeddability remains a separate issue [7]. Further comparisons are given in Section 6.

The shallow autonomous ReLU model used quantitatively is

$$\dot{z}(t) = V(z(t)), \qquad V(x) = \sum_{j=1}^{p} w_j (a_j^\top x + b_j)_+, \qquad (s)_+ := \max\{s, 0\}.$$

Positive homogeneity and coefficient-controlled approximation enter only in the quantitative estimates. Section 2 states the main results; Section 3 proves the endpoint-correction principle; Section 4 derives the ReLU bounds; and Section 5 proves the obstructions. The appendices contain the geometric, approximation, bridge, and closure estimates.

Figure 1 summarizes the dimension- and scale-dependent mechanisms behind these results.

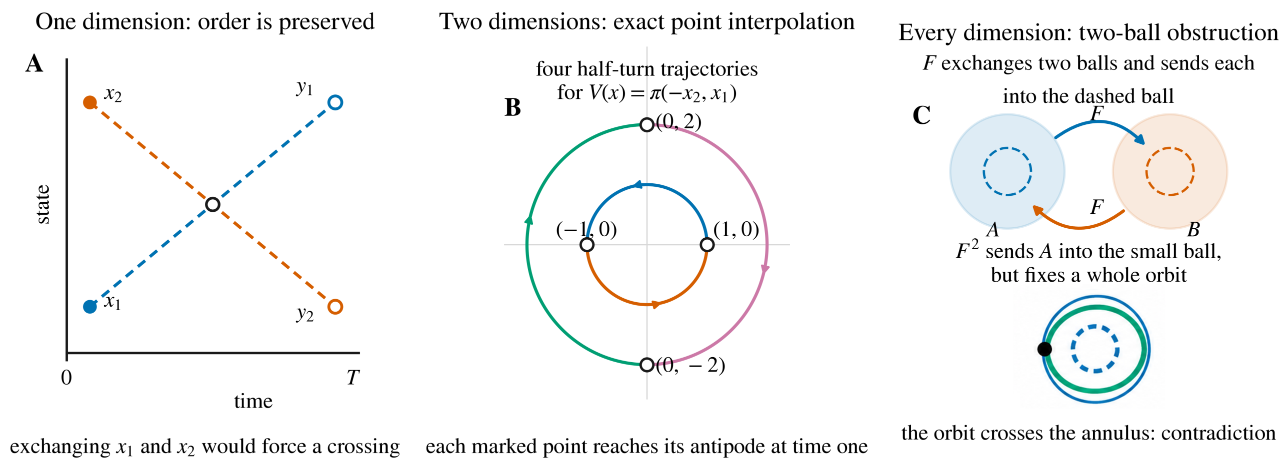

Figure 1: Exact interpolation and the obstruction to routing balls. (A) A one-dimensional flow preserves order, so it cannot exchange two ordered inputs. (B) The field $V(x) = \pi(-x_2, x_1)$ rotates the inputs $(\pm 1, 0)$ and $(0, \pm 2)$ to their antipodes at time one; the four marked trajectories lie on two circles. (C) If $F = \Psi^T$ exchanged and compressed two disjoint closed balls, Brouwer's theorem would give a fixed point of F^2 in the first ball. Its semiflow trajectory would cross the surrounding annulus, where F^2 cannot both fix the crossing point and send it into the smaller target ball. Panel B shows the explicit example; panels A and C illustrate the arguments.

2 Framework and principal results

Several architecture classes organize the comparison. An *autonomous* model uses one field $V(x)$, so its terminal map repeats under iteration. A *semi-autonomous* model has restricted explicit time dependence, for example affine time dependence in thresholds; equivalently, some such schedules can be encoded autonomously after adjoining a clock state. A *fully nonautonomous* model uses $V(t, x)$ and may select distinct temporal stages, although this freedom alone does not imply simultaneous cell controllability. An *augmented-state autonomous* model evolves in a larger state space and can evade restrictions tied to the original-state topology; augmented neural ODEs were introduced to alleviate original-state representational restrictions [13]. Controlled residual dynamics and continuous normalizing flows supply discrete-depth and invertible-flow comparisons, respectively [9, 15]. The distinctions relevant here are exact finite interpolation, fixed-radius neighborhood routing, invertibility, and the availability of temporal scheduling.

In the autonomous shallow network, the trainable "controls" are fixed coefficients of one time-independent vector field; once chosen, they act throughout the entire interval. Nonautonomous control architectures can instead select different operations at different temporal stages. This distinction explains both sides of the paper: endpoint sensitivities permit simultaneous correction of finitely many terminal constraints, while autonomy forces the semigroup identity $\Phi^{2T} = \Phi^T \circ \Phi^T$ on neighborhood behavior.

2.1 Framework: model and expressivity notions

2.1.1 Autonomous ReLU flows and finite data

Let $\mathbb{N}_0 := \{0, 1, 2, \ldots\}$, and set $\log_+(r) := \max\{\log r, 0\}$ for $r > 0$, with $\log_+(0) := 0$. For $d \geq 1$, $x \in \mathbb{R}^d$, and $r > 0$, write

$$\mathrm{B}_r^d(x) := \{z \in \mathbb{R}^d : |z - x| < r\}, \qquad \mathrm{B}_r^d := \mathrm{B}_r^d(0).$$

When $d = N$, we suppress the superscript and write $\mathrm{B}_r(x)$ and B_r. Thus B_r is reserved for a ball, whereas the italic symbol B below denotes the linear part of an affine normalization. For $N \geq 1$ and $p \in \mathbb{N}_0$, define

$$\mathscr{N}_{N,p} := \left\{ x \longmapsto \sum_{j=1}^{p} w_j (a_j^\top x + b_j)_+ : w_j, a_j \in \mathbb{R}^N,\ b_j \in \mathbb{R} \right\}, \qquad \mathscr{N}_N := \bigcup_{p \in \mathbb{N}_0} \mathscr{N}_{N,p}.$$

For a map u on a set A, let $\|u\|_{C(A)} := \sup_{x \in A} |u(x)|$, and let $\mathrm{Lip}_A(u)$ denote its least Lipschitz constant on A. For an integer $s \geq 0$, $\|u\|_{C^s(A)}$ is the maximum of the suprema of all derivatives of order at most s, when they exist, while $\|D^q u(x)\|_{\mathrm{op}}$ denotes the operator norm of the q-linear derivative. Finally, $\mathrm{Diff}_c^\infty(\Omega)$ denotes the group of smooth diffeomorphisms of Ω that agree with Id outside a compact subset of Ω.

The empty sum is the zero field. Each $V \in \mathscr{N}_N$ is globally Lipschitz and has at most linear growth. It therefore generates a unique global flow Φ_V^t, $t \in \mathbb{R}$.

For a parameter list $\theta = ((w_j, a_j, b_j))_{j=1}^p$, write

$$V_\theta(x) := \sum_{j=1}^{p} w_j (a_j^\top x + b_j)_+, \qquad p(\theta) := p.$$

Thus $V_\theta \in \mathscr{N}_N$; different parameter lists may represent the same field.

Let $M \in \mathbb{N}$, set $m := NM$, and let $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^M \subset \mathbb{R}^N \times \mathbb{R}^N$. We identify $(\mathbb{R}^N)^M$ with $\mathbb{R}^m$ and equip it with the Euclidean norm. The dataset is *admissible* if the inputs x_i are pairwise distinct and the targets y_i are pairwise distinct. Fixed constraints $x_i = y_i$ and cross-relations $y_i = x_j$ are allowed. For pairwise distinct inputs, pairwise distinct targets are necessary because every flow map is injective; repeated identical constraints may be discarded. Set $Y := (y_1, \ldots, y_M) \in \mathbb{R}^m$.

Remark 2.1 (Overlapping input and target configurations). Admissibility does not require the union of inputs and targets to contain $2M$ points. Stationary constraints $x_i = y_i$, cross-relations $y_i = x_j$, finite cycles, and partial overlap of the input and target sets are all permitted. The synchronized reference construction forms the directed correspondence on the K distinct vertices, completes its paths into disjoint cycles when necessary, and assigns stationary vertices to fixed components. Thus repetitions in the union create shared vertices, not conflicting endpoint conditions. The construction is carried out in Sections 3.1.1–3.1.4, and the endpoint directions for moving and stationary samples are constructed in Section 3.2; the regularity and derivative estimates are deferred to Appendix A.

Remark 2.2. The restriction $N \geq 2$ in the exact interpolation theorem is structural, rather than a technical artifact of the proof. In one dimension, every flow map preserves order: if $x < x'$, then $\Phi_V^t(x) < \Phi_V^t(x')$ for every $t \in \mathbb{R}$. Indeed, two distinct trajectories cannot meet by uniqueness of solutions. Hence the admissible assignment $x_1 = 0,\ x_2 = 1,\ y_1 = 1,\ y_2 = 0$ cannot be realized by any one-dimensional flow. Thus universal exact interpolation already fails for $N = 1$. In dimensions $N \geq 2$, trajectories have transverse room to route around one another; this geometric freedom is what makes arbitrary finite assignments possible.

2.1.2 Point interpolation and uniform neighborhood routing

Fix $T > 0$. For any locally Lipschitz vector field V generating a global flow, define the terminal, or endpoint, map—the list of all sample locations after time T—by $\mathcal{E}(V) := (\Phi_V^T(x_1), \ldots, \Phi_V^T(x_M)) \in \mathbb{R}^m$. For an admissible dataset, let

$$p_{\min}(\mathcal{D}, T) := \inf\left\{ p(\theta) : \Phi_{V_\theta}^T(x_i) = y_i,\ i = 1, \ldots, M \right\},$$

with the convention that the infimum is $+\infty$ if the constraint set is empty.

A *continuous autonomous semiflow*—a continuous forward-time evolution whose successive time intervals compose— is a family $(\Psi^t)_{t \geq 0}$ of self-maps such that $(t, x) \mapsto \Psi^t(x)$ is continuous on $[0, \infty) \times \mathbb{R}^N$, $\Psi^0 = \mathrm{Id}$, and $\Psi^{t+s} = \Psi^t \circ \Psi^s$ for $s, t \geq 0$.

Definition 2.3 (Universal point-to-neighborhood property). Let $\mathscr{F}$ be a class of autonomous vector fields with global flows.

(i) The class $\mathscr{F}$ has property (P_0) if every admissible finite dataset is interpolated exactly at time T by a field in $\mathscr{F}$.

(ii) For $\delta > 0$, the class $\mathscr{F}$ has property (P_δ) if, for every finite family of centers $x_1, \ldots, x_M$ for which the balls $\overline{\mathrm{B}}_\delta(x_i)$ are pairwise disjoint, every family of pairwise distinct target centers $y_1, \ldots, y_M$, and every $0 < \varepsilon < \delta$, there exists $V \in \mathscr{F}$ such that

$$\Phi_V^T(\overline{\mathrm{B}}_\delta(x_i)) \subset \mathrm{B}_\varepsilon(y_i), \qquad i = 1, \ldots, M.$$

The input radius in (P_δ) is fixed before the vector field is chosen. This order of quantifiers distinguishes universal fixed-radius neighborhood routing from the data-dependent neighborhoods supplied by continuity after a particular interpolating field has been fixed.

We use the notion of *simultaneous cell controllability* (SCC) introduced in [6]: a control architecture has SCC if it can send every finite family of pairwise disjoint compact convex cells into arbitrarily small balls about prescribed targets. SCC includes the two-ball routing configurations used below. For an architecture closed under positive time rescaling, it also implies (P_δ) at each prescribed terminal time.

2.2 Principal structural results

The paper has two principal structural statements. Theorem 2.4 gives the positive approximation-to-interpolation principle, and Theorem 2.7 gives the complementary autonomous semigroup obstruction. Their shallow-ReLU and approximation-theoretic consequences are stated alongside them; the quantitative ReLU refinement is separated below from the two structural statements.

2.2.1 From approximation to exact interpolation

Theorem 2.4 (Approximation on bounded sets implies exact finite interpolation)**.** *Let $N \geq 2$, and let $\mathscr{F} \subset C^{0,1}_{\mathrm{loc}}(\mathbb{R}^N; \mathbb{R}^N)$ be a linear space of vector fields such that every $V \in \mathscr{F}$ generates a global flow. Assume that members of $\mathscr{F}$ can approximate every compactly supported smooth vector field arbitrarily well on every bounded set: for every $u \in C_c^\infty(\mathbb{R}^N; \mathbb{R}^N)$, $R > 0$, and $\varepsilon > 0$, there exists $V \in \mathscr{F}$ such that*

$$\|V - u\|_{C(\overline{\mathrm{B}}_R)} < \varepsilon.$$

Then, for every $T > 0$ and every admissible finite correspondence $\{(x_i, y_i)\}_{i=1}^M$, there exists $V \in \mathscr{F}$ such that $\Phi_V^T(x_i) = y_i$ for all i.

The proof is given in Section 3; the regularity and quantitative estimates needed by the ReLU application are recorded in the appendices.

Shallow-ReLU consequence.

Corollary 2.5 (Exact finite interpolation by shallow autonomous ReLU flows)**.** *For every $N \geq 2$, every admissible finite correspondence in $\mathbb{R}^N$, and every prescribed time $T > 0$, there exists one time-independent field*

$$V(x) = \sum_{j=1}^{p} w_j (a_j^\top x + b_j)_+$$

whose time-T flow satisfies $\Phi_V^T(x_i) = y_i$ for all i. This includes stationary constraints, cross-relations, finite cycles, and arbitrary overlap between the input and target configurations.

This follows from Theorem 2.4, as detailed in Section 4.1.

Remark 2.6 (Reusable criterion beyond ReLU). Theorem 2.4 is not specific to ReLU: any linear dictionary class with the same approximation-on-bounded-sets property and global well-posedness assumptions inherits exact finite interpolation. In particular, globally Lipschitz nonpolynomial ridge activations, such as the logistic sigmoid or tanh, provide natural examples, since their finite ridge expansions are globally Lipschitz and classical universal approximation theorems provide the required approximation on bounded sets [11, 18, 23]. Thus the qualitative exact-interpolation result extends to these activations. The quantitative width and normalized-strength estimates below, however, remain ReLU-specific because they use coefficient-controlled approximation and positive homogeneity.

2.2.2 Autonomous semigroup obstruction

Theorem 2.7 (Semigroup obstruction to two-cell routing)**.** *The following statement contains no neural-network hypothesis. Let $N \geq 1$, $T > 0$, and $\delta > 0$. Suppose $|a - b| > 2\delta$ and $0 < \varepsilon < \delta$. No continuous autonomous semiflow $(\Psi^t)_{t\geq 0}$ on $\mathbb{R}^N$ can satisfy*

$$\Psi^T(\overline{\mathrm{B}}_\delta(a)) \subset \mathrm{B}_\varepsilon(b), \qquad \Psi^T(\overline{\mathrm{B}}_\delta(b)) \subset \mathrm{B}_\varepsilon(a).$$

Corollary 2.8 (Sharp point-to-neighborhood threshold)**.** *Let $N \geq 2$, $T > 0$, and $\delta \geq 0$. Then*

$$\mathscr{N}_N \text{ has property } (\mathsf{P}_\delta) \quad \Longleftrightarrow \quad \delta = 0.$$

In particular, no original-state autonomous architecture whose admissible evolutions are continuous semiflows can satisfy SCC.

Corollary 2.9 (Failure of uniform universal approximation by autonomous time maps)**.** *Let $N \geq 1$ and $T > 0$. There exist a compact ball $\mathcal{K} \subset \mathbb{R}^N$, a map $f \in C_c^\infty(\mathbb{R}^N; \mathbb{R}^N)$, and a constant $c > 0$ such that*

$$\left\|\Psi^T - f\right\|_{C(\mathcal{K})} \geq c$$

for every continuous autonomous semiflow $(\Psi^t)_{t\geq 0}$ on $\mathbb{R}^N$. Consequently, time-T maps of continuous autonomous semiflows are not dense in $C(\mathcal{K}; \mathbb{R}^N)$ in the uniform topology.

Proof. Fix $\delta > 0$ and choose $a, b \in \mathbb{R}^N$ with $|a - b| > 4\delta$. There are disjoint open neighborhoods of $\overline{\mathrm{B}}_\delta(a)$ and $\overline{\mathrm{B}}_\delta(b)$ and smooth compactly supported cutoff functions η_a, η_b that equal one on the respective closed balls and whose supports lie in those neighborhoods. Set $f := b\eta_a + a\eta_b$, and let $\mathcal{K}$ be a closed ball containing both supports. Then $f \equiv b$ on $\overline{\mathrm{B}}_\delta(a)$ and $f \equiv a$ on $\overline{\mathrm{B}}_\delta(b)$.

If $\left\|\Psi^T - f\right\|_{C(\mathcal{K})} < \delta$, choose ε strictly between this norm and δ. The uniform estimate on the two closed source balls then gives

$$\Psi^T(\overline{\mathrm{B}}_\delta(a)) \subset \mathrm{B}_\varepsilon(b), \qquad \Psi^T(\overline{\mathrm{B}}_\delta(b)) \subset \mathrm{B}_\varepsilon(a),$$

contradicting Theorem 2.7. Hence the asserted bound holds with $c = \delta$. ☐

Combining Remark 2.2, Corollary 2.5, and Theorem 2.7 gives the sharp point-to-neighborhood picture. Universal exact point interpolation fails in dimension one by order preservation, while for the shallow autonomous ReLU class it holds in every dimension $N \geq 2$; at every fixed positive radius, however, universal neighborhood routing fails for continuous autonomous semiflows. Equivalently, for $N \geq 2$,

$$\mathscr{N}_N \text{ has property } (\mathsf{P}_\delta) \quad \Longleftrightarrow \quad \delta = 0.$$

Thus the positive and negative results meet at the point scale: exact finite interpolation is compatible with autonomy, whereas universal fixed-radius routing is not.

2.3 Quantitative refinement for shallow ReLU flows

For the quantitative statement, set

$$\mathcal{V} := \{x_1, \dots, x_M, y_1, \dots, y_M\}_{\text{distinct}}, \qquad K := |\mathcal{V}|.$$

For $K \geq 2$, define

$$\operatorname{sep}(\mathcal{V}) := \min_{v \neq v'} |v - v'|, \qquad \operatorname{diam}(\mathcal{V}) := \max_{v,v'} |v - v'|,$$

and the affine aspect ratio

$$\mathfrak{a}(\mathcal{V}) := \inf_{L_{\text{geom}} \in \mathrm{GL}(N)} \frac{\operatorname{diam}(L_{\text{geom}}\mathcal{V})}{\operatorname{sep}(L_{\text{geom}}\mathcal{V})}.$$

For $K = 1$, set $\mathfrak{a}(\mathcal{V}) = 1$.

For an affine normalization $S(x) = B(x - x_*)$, $B \in \mathrm{GL}(N)$, and a parameter list $\theta = ((w_j, a_j, b_j))_{j=1}^p$, define

$$\mathrm{Str}_S(\theta) := \sum_{j=1}^p |Bw_j| \sqrt{\left|B^{-\top} a_j\right|^2 + \left|a_j^\top x_* + b_j\right|^2}, \qquad \|\theta\|_{S,2}^2 := \sum_{j=1}^p \left(|Bw_j|^2 + \left|B^{-\top} a_j\right|^2 + \left|a_j^\top x_* + b_j\right|^2\right).$$

The main bound is stated in terms of these two representation-dependent quantities. Their functional interpretation, the positive-homogeneity balancing identity, and the approximation exponents used in the proof are introduced later, at the point where they enter the argument.

Theorem 2.10 (Quantitative shallow-ReLU realization)**.** *For every $N \geq 2$, there exists $C_N \geq 2$, depending only on N, with the following property. Let $T > 0$, and let $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^M$ be admissible. There exist $B \in \mathrm{GL}(N)$, $x_* \in \mathbb{R}^N$, the affine map $S(x) = B(x - x_*)$, and a parameter list θ such that*

$$\Phi_{V_\theta}^T(x_i) = y_i, \quad i = 1, \dots, M, \tag{2.1}$$

$$\max\{p(\theta), T\,\mathrm{Str}_S(\theta)\} \leq \mathfrak{U}_N(\mathcal{D}) := \big(C_N K \mathfrak{a}(\mathcal{V})\big)^{C_N K}.$$

If $K \geq 2$, the affine map can be chosen so that

$$S(\mathcal{V}) \subset \overline{\mathrm{B}}_{1/2}, \qquad \mathrm{sep}(S(\mathcal{V})) \geq \frac{1}{4\mathfrak{a}(\mathcal{V})}. \tag{2.2}$$

If $K = 1$, one may take $S = \mathrm{Id}$ and the empty parameter list. In all cases, the units may be balanced without changing the represented field, so that

$$\|\theta\|_{S,2} \leq \sqrt{\frac{2\mathfrak{U}_N(\mathcal{D})}{T}}. \tag{2.3}$$

In particular,

$$p_{\min}(\mathcal{D}, T) \leq \mathfrak{U}_N(\mathcal{D}). \tag{2.4}$$

The proof combines the finite-dimensional endpoint-correction construction of Section 3, the quantitative bridge of Proposition 4.3, and the coefficient-controlled ReLU approximation and width–strength accounting of Section 4; the bridge estimates are established in Appendix C.

3 Finite-dimensional endpoint correction for autonomous flows

We follow the four steps described in the introduction. Section 3.1 constructs a smooth autonomous reference field, and Section 3.2 constructs correction fields with identity first-order endpoint response. Section 3.3 shows that approximate correction fields retain an invertible response at the smooth reference field, controls the error after approximation of the reference field, and removes the residual by Brouwer degree. Section 3.4 combines these ingredients within the prescribed class. Quantitative bounds on the geometric construction are deferred to Appendix A and used in the ReLU refinement.

For the fixed-order regularity estimates used later, set

$$s_N := \left\lfloor \frac{N+3}{2} \right\rfloor + 2. \tag{3.1}$$

The order s_N is used to index the derivative bounds for the transported reference field and endpoint directions. It is not an assumption of Theorem 2.4: the qualitative exactification argument itself uses only smoothness, while the explicit C^{s_N} bounds are recorded in Appendix A for the quantitative analysis of Section 4.

3.1 Smooth autonomous reference realizations

3.1.1 Directed correspondence graphs and cycle completion

Associate with the dataset the directed graph whose distinct vertices form

$$\mathcal{V} := \{x_1, \dots, x_M, y_1, \dots, y_M\}_{\text{distinct}},$$

and whose prescribed edges are $x_i \to y_i$.

Lemma 3.1 (Cycle completion of an admissible correspondence)**.** *Every connected component of the directed data graph is a directed path, a directed cycle, or a self-loop. By adding one unconstrained edge from the terminal vertex to the initial vertex of each directed path, the graph becomes a disjoint union of directed cycles and self-loops. No prescribed edge is changed and no additional constrained sample is introduced.*

Proof. Pairwise distinct inputs imply that every vertex has out-degree at most one, while pairwise distinct targets imply that every vertex has in-degree at most one. A finite directed graph with these two properties has only path, cycle, and self-loop components. A nonclosed path has a unique initial vertex with no incoming edge and a unique terminal vertex with no outgoing edge. Adding the single terminal-to-initial edge closes that component without changing the outgoing edge of any vertex that already carries a prescribed assignment. Repeating this independently for every path gives the claimed disjoint cycle decomposition. ☐

Figure 2 illustrates the three component types and the auxiliary edge used to close a path.

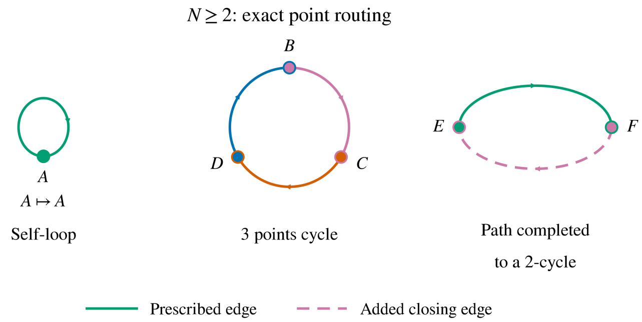

Figure 2: The three component types used in the reference construction. Left: a stationary constraint is represented by a self-loop. Center: a prescribed three-point cycle. Right: a directed path is completed to a two-cycle by one auxiliary closing edge (dashed). The auxiliary edge changes no original endpoint constraint. The displayed curves are schematic reference routes at the data vertices, not the standard-model trajectories of $z_{\nu,k}$ under f_0 introduced below.

3.1.2 Standard cyclic and stationary models

Index the completed components by positive integers ν, put $c_\nu := 4\nu e_1$, and let a nontrivial component contain $n_\nu \geq 2$ vertices. Place its standard representatives at

$$z_{\nu,k} = c_\nu + \cos(2\pi k/n_\nu)e_1 + \sin(2\pi k/n_\nu)e_2, \qquad k = 0, \dots, n_\nu - 1.$$

Let $\mathsf{R}_{\text{plane}} e_1 = e_2$, $\mathsf{R}_{\text{plane}} e_2 = -e_1$, and $\mathsf{R}_{\text{plane}} e_j = 0$ for $j \geq 3$. Fix $\chi \in C_c^\infty((1/2, 3/2))$, with $0 \leq \chi \leq 1$ and $\chi \equiv 1$ on $[3/4, 5/4]$, and define

$$f_{0,\nu}(z) := \omega_\nu \chi(|z - c_\nu|)\mathsf{R}_{\text{plane}}(z - c_\nu), \qquad \omega_\nu := \frac{2\pi}{n_\nu}.$$

Place each self-loop at the corresponding center c_ν, and let f_0 be the sum of the component fields. Their supports are disjoint. On every standard cycle, $\chi = 1$, so the flow is the planar rotation of angular speed $2\pi/n_\nu$; hence

$$\Phi^1_{f_0}(z_{\nu,k}) = z_{\nu,k+1\,(\mathrm{mod}\,n_\nu)}.$$

A standard self-loop lies at a center where $f_0 = 0$, and is therefore fixed. Thus the completed correspondence is realized exactly by one smooth autonomous standard field. For each fixed derivative order s, the norm $\|f_0\|_{C^s}$ is bounded by a constant depending only on N, s, and the fixed cutoff template, independently of the number of components. Indeed, the component supports are disjoint, their angular speeds satisfy $\omega_\nu \leq \pi$, and translation of the fixed cutoff does not change its derivative bounds.

3.1.3 Finite-point relocation by localized point pushes

The standard vertices must now be transported to the actual data configuration. The following quantitative finite-transitivity statement is also the geometric input for the later upper bound.

Lemma 3.2 (Quantitative finite-point relocation)**.** *Fix $N \geq 2$, $s \geq 2$, $K \geq 1$, and $R \geq 1$. Let $\mathbf{z} = (z_1, \dots, z_K)$ and $\mathbf{w} = (w_1, \dots, w_K)$ be ordered configurations of distinct points in B_R, and put*

$$d_{\text{cfg}} := \begin{cases} 1, & K = 1, \\ \min\{1, \operatorname{sep}(\{z_j\}), \operatorname{sep}(\{w_j\})\}, & K \geq 2. \end{cases}$$

If an open set Ω_{tr} contains $\overline{\mathrm{B}}_{R+4K+4}$, there is $H \in \mathrm{Diff}_c^\infty(\Omega_{\text{tr}})$ such that $H(z_j) = w_j$ and

$$1 + \max_{1 \leq q_{\text{der}} \leq s} \left(\|D^{q_{\text{der}}} H\|_\infty + \left\|D^{q_{\text{der}}} H^{-1}\right\|_\infty\right) \leq \left(\frac{C_{\text{tr},N,s}(1+K)(1+R)}{d_{\text{cfg}}}\right)^{C_{\text{tr},N,s} K}, \tag{3.2}$$

where $C_{\text{tr},N,s} \geq 2$ depends only on N and s.

Proof sketch. Choose mutually separated parking points outside the ball containing the source and target configurations. Move the source points to the parking points one at a time, and then move the parking points to the targets one at a time; all points not currently moving are treated as obstacles. For each individual move from *a* to b, one can choose a distant intermediate point ξ so that both segments $[a,\xi]$ and $[\xi,b]$ stay a positive distance from every obstacle. This is the only step where $N \geq 2$ is essential: the set of forbidden directions is a finite union of small spherical caps and cannot exhaust the available directions.

Along each clear segment, apply a compactly supported smooth cylinder push that sends its initial endpoint to its terminal endpoint and is the identity near the cylinder boundary. Because the support cylinder avoids all obstacles, the push fixes every point not currently moving. Composing at most $4K$ such pushes gives a compactly supported diffeomorphism satisfying $H(z_j) = w_j$. The explicit cylinder map, its inverse, the support verification, and the product–chain estimates leading to (3.2) are given in Appendix A. ☐

3.1.4 Smooth autonomous reference realization

Proposition 3.3 (Smooth autonomous reference realization)**.** *Let $N \geq 2$ and let $\mathcal{D}$ be admissible. There exists a compactly supported smooth autonomous field f whose unit-time flow satisfies*

$$\Phi_f^1(x_i) = y_i, \qquad i = 1, \ldots, M.$$

The marked trajectories remain pairwise distinct at every common time. In the normalized setting

$$\mathcal{V} \subset \overline{\mathrm{B}}_{1/2}, \qquad \mathrm{sep}(\mathcal{V}) = \rho, \qquad 0 < \rho \leq 1,$$

the same construction admits support and fixed-order derivative bounds depending only on N, K, ρ; these bounds are recorded in Appendix A.

Proof. Choose an ordering $(\mathsf{v}_1, \ldots, \mathsf{v}_K)$ of the actual vertices and the corresponding ordering $(\mathsf{v}_1^0, \ldots, \mathsf{v}_K^0)$ of the standard vertices. Lemma 3.2 gives a compactly supported smooth diffeomorphism H satisfying

$$H(\mathsf{v}_j^0) = \mathsf{v}_j \quad (j = 1, \ldots, K), \qquad H(x_i^0) = x_i, \quad H(y_i^0) = y_i \quad (i = 1, \ldots, M). \tag{3.3}$$

Define the pushforward field $f := H_* f_0$, namely

$$(H_* u)(x) := DH(H^{-1}x)u(H^{-1}x).$$

Uniqueness of solutions gives the exact conjugacy

$$\Phi_f^t = H \circ \Phi_{f_0}^t \circ H^{-1}. \tag{3.4}$$

Since every prescribed standard edge is traversed in unit time, (3.3) and (3.4) yield $\Phi_f^1(x_i) = y_i$. Standard marked trajectories are either disjoint component circles or distinct synchronized positions on the same circle; conjugation by a diffeomorphism preserves this collision-free property. ☐

3.2 Independent endpoint corrections

For the unit-time reference field, write $\gamma_i(t) := \Phi_f^t(x_i)$, and let $U_i(t,s)$ denote the fundamental matrix

$$\partial_t U_i(t,s) = Df(\gamma_i(t))U_i(t,s), \qquad U_i(s,s) = I_N.$$

For a perturbation field h, define the endpoint response

$$\mathcal{L}_f(h)_i := \int_0^1 U_i(1,t)h(\gamma_i(t))\,\mathrm{d}t, \qquad \mathcal{L}_f(h) := (\mathcal{L}_f(h)_1, \ldots, \mathcal{L}_f(h)_M).$$

The usual variational equation gives $D\mathcal{E}(f)[h] = \mathcal{L}_f(h)$.

3.2.1 Private trajectory tubes

For a nonstationary datum, let γ_i^0 be its standard circular trajectory and let $\nu(i)$ denote its completed component. On $I_{\mathrm{tube}} := (1/4, 3/4)$, introduce longitudinal–transverse coordinates

$$\Theta_i^0(t,z) := \gamma_i^0(t) + z_1\big(\gamma_i^0(t) - c_{\nu(i)}\big) + \sum_{a=2}^{N-1} z_a e_{a+1}, \qquad (t,z) \in I_{\mathrm{tube}} \times \mathbb{R}^{N-1}.$$

Because only finitely many marked trajectories are involved, one may choose a common radius $\varepsilon_{\mathrm{tube}} > 0$ sufficiently small that the restriction

$$\Theta_i^0 : I_{\mathrm{tube}} \times \mathrm{B}_{2\varepsilon_{\mathrm{tube}}}^{N-1} \longrightarrow O_i^0$$

is a diffeomorphism onto an open set O_i^0, and the middle part of this tube meets no constrained standard trajectory except its own centerline:

$$O_i^0 \cap \gamma_j^0([0,1]) = \varnothing \quad (j \neq i), \qquad O_i^0 \cap \gamma_i^0([0,1]) = \gamma_i^0(I_{\mathrm{tube}}). \tag{3.5}$$

This is the geometric localization that permits one endpoint to be varied without affecting the others. The injectivity, quantitative separation, and inverse-regularity estimates for Θ_i^0 are proved in Appendix A.3.

3.2.2 Independent endpoint directions

Choose smooth cutoffs $\vartheta \in C_c^\infty(I_{\mathrm{tube}})$ and $\zeta_i \in C_c^\infty(\mathrm{B}^{N-1}_{2\varepsilon_{\mathrm{tube}}})$ such that $\int_0^1 \vartheta(t)\,\mathrm{d}t = 1$ and $\zeta_i(0) = 1$. Let $U_i^0(t,s)$ be the fundamental matrix of f_0 along γ_i^0. For $a = 1, \ldots, N$, define on O_i^0

$$k_{i,a}^0(\Theta_i^0(t,z)) := \vartheta(t)\zeta_i(z)U_i^0(1,t)^{-1}e_a,$$

and extend it by zero. By (3.5), this field vanishes along all constrained trajectories except the ith, while on its own centerline the fundamental matrix is exactly cancelled. Therefore

$$\int_0^1 U_j^0(1,t)k_{i,a}^0(\gamma_j^0(t))\,\mathrm{d}t = \delta_{ij}e_a. \tag{3.6}$$

For a self-loop datum, take a compactly supported bump equal to e_a near the stationary point, with support in a neighborhood where the standard field vanishes and which is disjoint from every other marked unit-time path; the same identity follows because the fundamental matrix there is the identity.

Transport and precondition these directions by setting

$$h_{i,a} := H_*\left(\sum_{b=1}^{N}[DH(y_i^0)^{-1}]_{ba}k_{i,b}^0\right). \tag{3.7}$$

Differentiating the flow conjugacy gives

$$U_i(t,s) = DH(\gamma_i^0(t))U_i^0(t,s)DH(\gamma_i^0(s))^{-1}. \tag{3.8}$$

Equations (3.6)–(3.8) imply $\mathcal{L}_f(h_{i,a})_j = \delta_{ij}e_a$. Relabeling the $m = NM$ fields gives the central endpoint-frame identity

$$\begin{bmatrix}\mathcal{L}_f(h_1) & \cdots & \mathcal{L}_f(h_m)\end{bmatrix} = I_m. \tag{3.9}$$

Proposition 3.4 (Independent endpoint corrections)**.** *The reference field from Proposition 3.3 admits compactly supported smooth directions $h_1, \ldots, h_m$ satisfying* (3.9)*. In the normalized setting, their support and fixed-order C^{s_N} bounds are recorded in Appendix A.4.*

Proof. The endpoint identity has just been proved. Compact support follows from the private-tube construction and the compact support of $H - \mathrm{Id}$. Smooth zero extension and the correction norms are verified in Appendix A.4. ◻

Figure 3 summarizes the private-tube localization and the resulting independent endpoint responses.

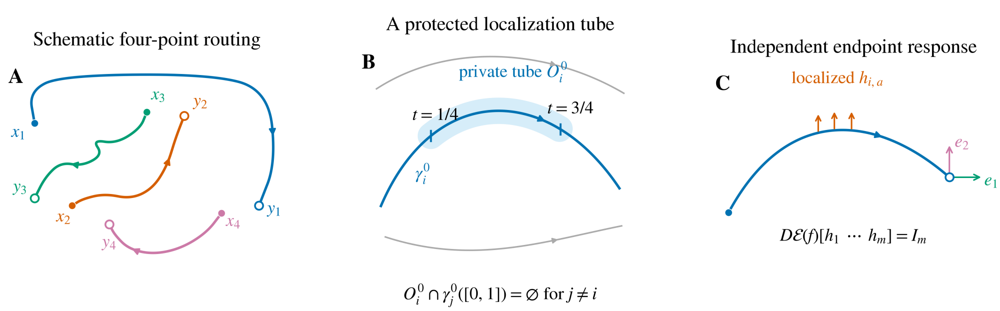

Figure 3: Localized corrections for the smooth reference flow. (A) Marked reference paths join the prescribed inputs and targets. (B) In standard coordinates, a tube around the middle part of one moving path meets no other marked path during the unit time interval. (C) After transport and preconditioning, the correction fields give independent first-order endpoint changes and the identity response matrix in (3.9); two response directions are shown schematically. Stationary samples use local bumps instead of tubes.

3.3 Stability and Brouwer-degree correction

3.3.1 Uniform endpoint stability

Lemma 3.5 (Uniform endpoint stability on a parameter ball)**.** *Let $f, g_1, \dots, g_m$ be locally Lipschitz and put $G_\alpha := \sum_{\ell=1}^m \alpha_\ell g_\ell$. Assume that, for $|\alpha| \le r$, all trajectories of $f + G_\alpha$ issued from $x_1, \dots, x_M$ exist on $[0,1]$ and remain in $\overline{\mathrm{B}}_{R_0}$. Fix any $R > R_0$. Then there are $\eta_0 > 0$ and $C > 0$ such that, whenever V is locally Lipschitz, every field $V + G_\alpha$, $|\alpha| \le r$, has a global flow, and*

$$\|V - f\|_{C(\overline{\mathrm{B}}_R)} \le \eta \le \eta_0,$$

one has

$$\sup_{|\alpha| \le r} |\mathcal{E}(V + G_\alpha) - \mathcal{E}(f + G_\alpha)| \le C\eta.$$

Proof. For each initial point and parameter $|\alpha| \le r$, stop the trajectory of $V + G_\alpha$ at its first exit from B_R. On the stopped interval, subtract the two integral equations. Since only finitely many locally Lipschitz fields g_ℓ occur and $|\alpha| \le r$, the family $f + G_\alpha$ has a common Lipschitz constant on $\overline{\mathrm{B}}_R$. Gronwall's inequality therefore gives

$$\sup_{0 \le t \le 1} |z_{V,\alpha}(t) - z_{f,\alpha}(t)| \le C_0 \eta$$

uniformly in α and in the finitely many initial points, as long as the stopped trajectory is defined. Choose η_0 so that $C_0 \eta_0 < R - R_0$. Then a first exit from B_R is impossible, so the estimate holds on the whole interval $[0,1]$. Evaluating at time one and stacking the M endpoint blocks proves the claim. ☐

3.3.2 Brouwer-degree correction

Lemma 3.6 (Degree correction)**.** *Let $F : \overline{\mathrm{B}}_r^m \to \mathbb{R}^m$ be continuous, let $J \in \mathbb{R}^{m \times m}$ be invertible, and set $\sigma := \min_{|v|=1} |Jv|$. If*

$$\sup_{|\alpha| = r} |F(\alpha) - J\alpha| < \sigma r,$$

then some $\alpha_* \in \mathrm{B}_r^m$ satisfies $F(\alpha_*) = 0$.

Proof. For $0 \le \lambda \le 1$, set

$$\mathscr{H}_\lambda(\alpha) := J\alpha + \lambda\big(F(\alpha) - J\alpha\big).$$

On $|\alpha| = r$,

$$|\mathscr{H}_\lambda(\alpha)| \ge |J\alpha| - |F(\alpha) - J\alpha| > 0.$$

Thus the homotopy does not meet the origin on the boundary, and

$$\deg(F, \mathrm{B}_r^m, 0) = \deg(J, \mathrm{B}_r^m, 0) = \operatorname{sgn}(\det J) \ne 0.$$

The existence property of Brouwer degree gives a zero of F in B_r^m. ☐

3.3.3 The endpoint-correction principle

Lemma 3.7 (Local endpoint expansion)**.** *Suppose that*

$$\mathcal{E}(f) = Y, \qquad [\mathcal{L}_f(h_1) \ \cdots \ \mathcal{L}_f(h_m)] = I_m,$$

with $f, h_1, \dots, h_m \in C_c^\infty(\mathbb{R}^N; \mathbb{R}^N)$. Then there exist $R > 0$ and $\varepsilon_0 > 0$ such that, whenever $g_1, \dots, g_m$ are locally Lipschitz and

$$\max_\ell \|g_\ell - h_\ell\|_{C(\overline{\mathrm{B}}_R)} < \varepsilon_0,$$

the matrix

$$J := [\mathcal{L}_f(g_1) \ \cdots \ \mathcal{L}_f(g_m)]$$

is invertible. Writing $G_\alpha := \sum_{\ell=1}^m \alpha_\ell g_\ell$, there exist $r_0, C > 0$ such that, for $|\alpha| \le r_0$,

$$\mathcal{E}(f + G_\alpha) = Y + J\alpha + R(\alpha), \qquad |R(\alpha)| \le C|\alpha|^2.$$

Proof. Choose R so that all reference trajectories lie in B_{R-2}. The map $h \mapsto \mathcal{L}_f(h)$ is continuous in the uniform norm on $\overline{\mathrm{B}}_R$, hence J remains invertible if g_ℓ are sufficiently close to h_ℓ.

Fix such g_ℓ, and set

$$A_g := \left(\sum_{\ell=1}^m \|g_\ell\|_{C(\overline{\mathrm{B}}_R)}^2 \right)^{1/2}, \qquad L_g := \left(\sum_{\ell=1}^m \mathrm{Lip}_{\overline{\mathrm{B}}_R}(g_\ell)^2 \right)^{1/2}.$$

Both constants are finite. Let z_i^α be the trajectory of $f + G_\alpha$ issued from x_i, stopped at its first exit from B_{R-1}, and write $e_i^\alpha := z_i^\alpha - \gamma_i$. On the stopped interval, Cauchy–Schwarz gives

$$|G_\alpha(z)| \le |\alpha| A_g, \qquad |G_\alpha(z) - G_\alpha(z')| \le |\alpha| L_g |z - z'|.$$

Thus, with $L_f := \mathrm{Lip}_{\overline{\mathrm{B}}_R}(f)$, the integral equations and Gronwall's inequality give

$$\|e_i^\alpha\|_{C([0,t])} \le A_g e^{L_f} |\alpha|, \qquad 0 \le t \le 1.$$

After decreasing r_0, the right-hand side is < 1, so the stopped trajectory cannot reach $\partial \mathrm{B}_{R-1}$. Consequently every z_i^α exists on $[0,1]$, remains in B_{R-1}, and satisfies the displayed estimate.

Let ζ_i^α solve

$$\dot\zeta_i^\alpha = Df(\gamma_i)\zeta_i^\alpha + G_\alpha(\gamma_i), \qquad \zeta_i^\alpha(0) = 0.$$

Variation of constants first gives $\|\zeta_i^\alpha\|_{C([0,1])} \le C_1|\alpha|$. For $w_i^\alpha := e_i^\alpha - \zeta_i^\alpha$, subtraction of the two equations yields

$$\dot w_i^\alpha = Df(\gamma_i) w_i^\alpha + \big[f(\gamma_i + e_i^\alpha) - f(\gamma_i) - Df(\gamma_i)e_i^\alpha\big] + \big[G_\alpha(\gamma_i + e_i^\alpha) - G_\alpha(\gamma_i)\big].$$

Since f is C^2 on $\overline{\mathrm{B}}_R$, Taylor's formula and the preceding bounds imply, uniformly in i and t,

$$|f(\gamma_i + e_i^\alpha) - f(\gamma_i) - Df(\gamma_i)e_i^\alpha| \le \tfrac12 \left\|D^2 f\right\|_{C(\overline{\mathrm{B}}_R)} |e_i^\alpha|^2 = O(|\alpha|^2),$$

and

$$|G_\alpha(\gamma_i + e_i^\alpha) - G_\alpha(\gamma_i)| \le |\alpha| L_g |e_i^\alpha| = O(|\alpha|^2).$$

A second application of Gronwall's inequality therefore gives

$$\|z_i^\alpha - \gamma_i - \zeta_i^\alpha\|_{C([0,1])} = \|w_i^\alpha\|_{C([0,1])} \le C_2|\alpha|^2.$$

By variation of constants,

$$\zeta_i^\alpha(1) = \sum_{\ell=1}^m \alpha_\ell \, \mathcal{L}_f(g_\ell)_i.$$

Stacking the endpoint blocks and using $\mathcal{E}(f) = Y$ yields

$$\mathcal{E}(f + G_\alpha) = Y + J\alpha + O(|\alpha|^2),$$

which proves the claim. $\square$

Proposition 3.8 (Robust endpoint correction)**.** *Under the assumptions of Lemma 3.7, there exist $R > 0$ and $\varepsilon_0 > 0$ such that, for every family $g_1, \dots, g_m$ satisfying*

$$\max_\ell \|g_\ell - h_\ell\|_{C(\overline{\mathrm{B}}_R)} < \varepsilon_0,$$

there are $r, \eta_0 > 0$ with the following property: if

$$\|V_{\mathrm{base}} - f\|_{C(\overline{\mathrm{B}}_R)} < \eta_0$$

and every

$$V_{\mathrm{base}} + \sum_{\ell=1}^m \alpha_\ell g_\ell, \qquad |\alpha| \le r,$$

has a global flow, then some $\alpha_* \in \mathrm{B}_r^m$ satisfies

$$\mathcal{E}\left(V_{\mathrm{base}} + \sum_{\ell=1}^m \alpha_{*,\ell} g_\ell\right) = Y.$$

The robustness statement uses two continuity facts. First, the endpoint-response operator $g \mapsto \mathcal{L}_f(g)$ is continuous with respect to uniform perturbations on $\overline{\mathrm{B}}_R$, so sufficiently accurate approximations of the directions h_ℓ preserve invertibility of the endpoint-response matrix. Second, once $g_1, \dots, g_m$ are fixed, Lemma 3.5 gives uniform continuous dependence of $\mathcal{E}(V_{\mathrm{base}} + G_\alpha)$ on the base field for $|\alpha| \le r$. Consequently, the admissible radii and tolerances may depend on the chosen approximate frame, in particular through its local Lipschitz bounds, but no derivative approximation of the base field is required.

Proof. Fix $g_1, \dots, g_m$ and let

$$J = [\mathcal{L}_f(g_1) \ \cdots \ \mathcal{L}_f(g_m)], \qquad \sigma := \min_{|v|=1} |Jv| > 0.$$

By Lemma 3.7, after choosing $r > 0$ small enough,

$$|\mathcal{E}(f + G_\alpha) - Y - J\alpha| \le \frac{\sigma}{4}|\alpha|, \qquad |\alpha| \le r.$$

Lemma 3.5 then gives $\eta_0 > 0$ such that

$$\sup_{|\alpha|\le r} |\mathcal{E}(V_{\text{base}} + G_\alpha) - \mathcal{E}(f + G_\alpha)| \le \frac{\sigma r}{4}$$

whenever $\|V_{\text{base}} - f\|_{C(\overline{\mathrm{B}}_R)} < \eta_0$.

Set

$$F(\alpha) := \mathcal{E}(V_{\text{base}} + G_\alpha) - Y.$$

For the fixed correction fields and base field, F is continuous on the closed parameter ball: the vector fields depend continuously on the coefficients, have a common local Lipschitz bound there, and their relevant trajectories remain in the common trapping ball. For $|\alpha| = r$,

$$|F(\alpha) - J\alpha| \le \frac{\sigma r}{2} < \sigma r.$$

Lemma 3.6 therefore yields $\alpha_* \in \mathrm{B}_r^m$ with $F(\alpha_*) = 0$, proving the result. $\square$

Figure 4 illustrates the residual-enclosure argument underlying the degree step.

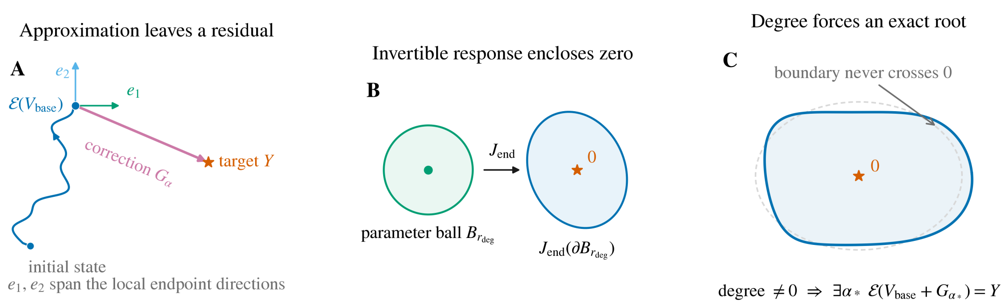

Figure 4: The degree argument in endpoint space, shown in two dimensions. Let $F(\alpha)$ be the stacked endpoint error after adding the correction with coefficients α. The linear response J maps the parameter sphere to a boundary separated from zero. The estimate $\|F(\alpha) - J\alpha\| < \sigma r$ on $|\alpha| = r$, where σ is the smallest singular value of J, keeps the straight homotopy between F and J away from zero on that boundary. Since J is invertible, the degree is nonzero and F vanishes for a coefficient vector inside the ball.

3.4 Proof of the approximation-to-interpolation theorem

Proof of Theorem 2.4. **Step 1: reference field.** It is enough to consider $T = 1$, since linearity of $\mathscr{F}$ implies that $V \in \mathscr{F}$ entails $T^{-1}V \in \mathscr{F}$, and $\Phi^T_{T^{-1}V} = \Phi^1_V$. Proposition 3.3 provides a compactly supported smooth field f with $\mathcal{E}(f) = Y$.

Step 2: correction fields. Proposition 3.4 provides compactly supported smooth fields $h_1, \dots, h_m$ with identity first-order endpoint response. Proposition 3.8 supplies a correction tolerance ε_0 on a ball $\overline{\mathrm{B}}_R$.

Step 3: approximation in the prescribed class. First use the approximation-on-bounded-sets hypothesis to choose $g_1, \dots, g_m \in \mathscr{F}$ within this tolerance. For these fixed approximate correction fields, Proposition 3.8 supplies $r > 0$ and $\eta_0 > 0$. Use the same approximation property once more to choose $V_{\text{base}} \in \mathscr{F}$ with

$$\|V_{\text{base}} - f\|_{C(\overline{\mathrm{B}}_R)} < \eta_0.$$

Step 4: exact correction. Since $\mathscr{F}$ is linear, every field

$$V_{\text{base}} + \sum_{\ell=1}^{m} \alpha_\ell g_\ell, \qquad |\alpha| \le r,$$

belongs to $\mathscr{F}$, and by assumption it has a global flow. Proposition 3.8 therefore yields a parameter α_* realizing all endpoints exactly. Rescaling time gives the conclusion for arbitrary $T > 0$. $\square$

Remark 3.9 (Qualitative exact interpolation beyond ReLU). Theorem 2.4 isolates the qualitative core of the positive result. It uses only linear closure, local Lipschitz regularity, global well-posedness of the candidate flows, and approximation of compactly supported smooth fields on bounded sets. In particular, neither piecewise linearity nor positive homogeneity of ReLU is required for exact finite interpolation itself; see also Remark 2.6 for the corresponding extension to other activation functions.

This separates two levels of the positive theory. Qualitative exact interpolation follows from approximation on bounded sets together with a finite-dimensional endpoint correction and Brouwer degree. By contrast, the quantitative width and normalized-strength estimates in Theorem 2.10 use specifically quantitative shallow-network approximation estimates and the positive homogeneity of ReLU.

4 Shallow ReLU realizations: qualitative and quantitative consequences

4.1 Qualitative consequence of the abstract theorem

Proof of Corollary 2.5. The class $\mathscr{N}_N$ is linear, every one of its fields is globally Lipschitz and hence globally well posed, and classical universal approximation for continuous nonpolynomial ridge activations gives the required coordinatewise approximation on every compact ball [18, 23]. In particular, each compactly supported smooth vector field can be approximated there by a finite vector-valued ReLU ridge expansion. Theorem 2.4 therefore applies. Positive time rescaling preserves $\mathscr{N}_N$ by scalar closure. □

4.2 Quantitative preliminaries

Recall the fixed regularity order s_N from (3.1). The ReLU approximation rate uses the additional dimension-dependent exponent

$$\kappa_N := \frac{2N}{N+3}. \tag{4.1}$$

Unlike s_N, which indexes the fixed-order regularity of the geometric construction, κ_N enters only in the quantitative approximation argument below.

4.2.1 Parameter conventions and normalized strength

For a parameter list $\theta = ((w_j, a_j, b_j))_{j=1}^{p}$, write

$$\mathrm{Str}(\theta) := \sum_{j=1}^{p} |w_j| \sqrt{|a_j|^2 + b_j^2},$$

and, when a shallow field is equipped with this particular list, also write $p(V_\theta) := p(\theta)$ and $\mathrm{Str}(V_\theta) := \mathrm{Str}(\theta)$. These are representation-dependent quantities. The affine-normalized versions $\mathrm{Str}_S(\theta)$ and $\|\theta\|_{S,2}$ were introduced in the quantitative refinement of Section 2.2 solely to state the coordinate-invariant quantitative theorem.

Let $S(x) = B(x - x_*)$ and let $\widetilde{V}(z) := BV(B^{-1}z + x_*)$ be the conjugated field. Directly from the shallow representation,

$$\max\{\mathrm{Lip}(\widetilde{V}), |\widetilde{V}(0)|\} \le \mathrm{Str}_S(\theta). \tag{4.2}$$

Thus $T\,\mathrm{Str}_S(\theta)$ is a genuine dynamical resource: it controls the logarithmic bi-Lipschitz distortion and the displacement generated by the normalized flow over time T, not merely the number of parameters.

Positive homogeneity of ReLU also allows each unit to be balanced independently. After zero units are discarded,

$$\inf_{\lambda_1,\dots,\lambda_p>0} \left\|((\lambda_j^{-1} w_j, \lambda_j a_j, \lambda_j b_j))_{j=1}^{p}\right\|_{S,2}^{2} = 2\,\mathrm{Str}_S(\theta). \tag{4.3}$$

Consequently, $2\,\mathrm{Str}_S(\theta)$ is the least squared normalized Euclidean parameter norm among unitwise positive-homogeneity rescalings representing the same vector field. This separates the functional size of the realized vector field from an arbitrary choice of unit scaling.

Remark 4.1. The representation-dependent quantity $\mathrm{Str}_S(\theta)$ therefore has two roles: it controls the conjugated vector field through (4.2), and after balancing it controls the normalized Euclidean parameter size through (4.3). These facts are needed only for the quantitative ReLU realization and for the bounded-strength obstruction; they play no role in the qualitative density-to-exactness theorem.

The coefficient-controlled ridge approximation below uses the exponent κ_N from (4.1), while the fixed order s_N from (3.1) indexes the geometric derivative bounds in Appendix A and the quantitative endpoint-frame bridge.

4.2.2 Coefficient-controlled ReLU approximation

Lemma 4.2 (Coefficient-controlled approximation of smooth vector fields)**.** *Let $s > (N+3)/2$ be an integer, let $R \geq 1$, and let $u \in C_c^s(\mathbb{R}^N;\mathbb{R}^N)$. For every $0 < \varepsilon_{\mathrm{app}} \leq 1$, set*

$$\mathcal{B}_{R,s}(u) := 2C_{\mathrm{emb},N,s}(1+R)^s \left\|u\right\|_{C^s},$$

$$n_{\mathrm{app}}(R,s;u,\varepsilon_{\mathrm{app}}) := \begin{cases} 0, & u = 0, \\ \left\lceil \left(\dfrac{A_N\sqrt{N}\,\mathcal{B}_{R,s}(u)}{\varepsilon_{\mathrm{app}}} \right)^{\kappa_N} \right\rceil, & u \neq 0. \end{cases} \tag{4.4}$$

and

$$\mathcal{P}_{R,s}(u,\varepsilon_{\mathrm{app}}) := N n_{\mathrm{app}}(R,s;u,\varepsilon_{\mathrm{app}}), \qquad \mathcal{S}_{R,s}(u) := N\sqrt{1+R^{-2}}\,\mathcal{B}_{R,s}(u). \tag{4.5}$$

Then there is a shallow ReLU field V_{app} such that

$$\|V_{\mathrm{app}} - u\|_{C(\overline{\mathrm{B}}_R)} \leq \varepsilon_{\mathrm{app}}, \qquad p(V_{\mathrm{app}}) \leq \mathcal{P}_{R,s}(u,\varepsilon_{\mathrm{app}}), \qquad \mathrm{Str}(V_{\mathrm{app}}) \leq \mathcal{S}_{R,s}(u).$$

This ReLU-specific approximation input is proved in Appendix B.

4.3 A quantitative endpoint-correction bridge

The following proposition quantifies the compact region and tolerances left implicit in Proposition 3.8; its proof is in Appendix C.

For a family $\mathbf{g} = (g_1,\ldots,g_m)$ of globally Lipschitz vector fields, define the correction-family budget

$$\mathfrak{B}(\mathbf{g}) := \left(\sum_{\ell=1}^m \big(|g_\ell(0)|^2 + \mathrm{Lip}(g_\ell)^2\big) \right)^{1/2}.$$

Proposition 4.3 (Quantitative endpoint-correction bridge)**.** *For every $N \geq 2$ there is $C_N \geq 2$, depending only on N, with the following property. Let $\mathcal{D}$ be an admissible unit-time dataset with*

$$\mathcal{V} \subset \overline{\mathrm{B}}_{1/2}, \qquad K := |\mathcal{V}| \geq 2, \qquad \rho := \mathrm{sep}(\mathcal{V}) \in (0,1], \qquad m := NM.$$

The reference field and endpoint correction directions in Propositions 3.3 and 3.4 can be chosen with $R_{\mathrm{tr}} = 8K+6$ so that

$$\mathcal{E}(f) = Y, \qquad [\mathcal{L}_f(h_1) \ \cdots \ \mathcal{L}_f(h_m)] = I_m, \qquad \operatorname{supp} f \cup \bigcup_{\ell=1}^m \operatorname{supp} h_\ell \subset \mathrm{B}_{R_{\mathrm{tr}}}.$$

There also exist an active approximation radius $R_* \geq R_{\mathrm{tr}}$ and a frame tolerance $0 < \varepsilon_* \leq 1$, depending only on the normalized dataset, such that, with

$$\Gamma_N(K,\rho) := \left(C_N \frac{K}{\rho} \right)^{C_N K},$$

one has

$$\max\left\{ R_*,\ \varepsilon_*^{-1},\ \|f\|_{C^{s_N}},\ \max_{1\leq\ell\leq m} \|h_\ell\|_{C^{s_N}} \right\} \leq \Gamma_N(K,\rho). \tag{4.6}$$

For every scalar budget $\mathfrak{b} \geq 1$, there are

$$0 < r_* = r_*(\mathcal{D},\mathfrak{b}) \leq \frac{1}{8}, \qquad 0 < \eta_* = \eta_*(\mathcal{D},\mathfrak{b}) \leq 1,$$

satisfying

$$\max\{r_*^{-1}, \eta_*^{-1}\} \leq \Gamma_N(K,\rho)(1+\mathfrak{b})^2, \tag{4.7}$$

with the following property. If globally Lipschitz fields $g_1,\ldots,g_m$ and a locally Lipschitz field V_{base} satisfy

$$\max_{1\leq\ell\leq m} \|g_\ell - h_\ell\|_{C(\overline{\mathrm{B}}_{R_{\mathrm{tr}}})} \leq \varepsilon_*, \qquad \mathfrak{B}(\mathbf{g}) \leq \mathfrak{b}, \qquad \|V_{\mathrm{base}} - f\|_{C(\overline{\mathrm{B}}_{R_*})} \leq \eta_*, \tag{4.8}$$

and every field $V_{\mathrm{base}} + \sum_{\ell=1}^m \alpha_\ell g_\ell$, $|\alpha| \leq r_*$, has a global flow, then there exists $\alpha_* \in \mathrm{B}_{r_*}^m$ such that

$$\mathcal{E}\Big(V_{\mathrm{base}} + \sum_{\ell=1}^m \alpha_{*,\ell} g_\ell\Big) = Y. \tag{4.9}$$

Moreover, every sample trajectory entering the endpoint homotopy for $|\alpha| \leq r_*$, both for $f + \sum_\ell \alpha_\ell g_\ell$ and for $V_{\mathrm{base}} + \sum_\ell \alpha_\ell g_\ell$, remains in B_{R_*} for $0 \leq t \leq 1$.

Proof. See Appendix C. ∎

4.4 Controlled shallow-ReLU realization in normalized coordinates

The affine aspect ratio defined in Section 2.2 removes overall scale and anisotropic coordinate distortion before the configuration is routed. Assume first that $K \geq 2$. By the definition of the infimum $\mathfrak{a}(\mathcal{V})$, choose a 2-near minimizer $L_{\mathrm{aff}} \in \mathrm{GL}(N)$ such that

$$\frac{\mathrm{diam}(L_{\mathrm{aff}}\mathcal{V})}{\mathrm{sep}(L_{\mathrm{aff}}\mathcal{V})} \leq 2\mathfrak{a}(\mathcal{V}).$$

Fix $x_* \in \mathcal{V}$ and put

$$B := \frac{L_{\mathrm{aff}}}{2\,\mathrm{diam}(L_{\mathrm{aff}}\mathcal{V})}, \qquad S(x) := B(x - x_*).$$

Then $S(\mathcal{V}) \subset \overline{\mathrm{B}}_{1/2}$, and with $\rho := \mathrm{sep}(S\mathcal{V})$,

$$\rho = \frac{\mathrm{sep}(L_{\mathrm{aff}}\mathcal{V})}{2\,\mathrm{diam}(L_{\mathrm{aff}}\mathcal{V})} \geq \frac{1}{4\mathfrak{a}(\mathcal{V})}. \tag{4.10}$$

Work in these normalized coordinates and at unit time. Take the reference field f, endpoint directions $h_1, \dots, h_m$, radius R_*, and frame tolerance ε_* supplied by Proposition 4.3.

Define the aggregate ReLU strength envelope of the ideal endpoint directions by

$$\mathcal{S}_h := \left(\sum_{\ell=1}^{m} \mathcal{S}_{R_{\mathrm{tr}},s_N}(h_\ell)^2\right)^{1/2}, \qquad \mathfrak{b}_h := \max\{1, \sqrt{2}\,\mathcal{S}_h\}.$$

Use the budget-dependent part of Proposition 4.3 with $\mathfrak{b} = \mathfrak{b}_h$, obtaining r_* and η_*. Now set

$$\mathcal{P}_h := \sum_{\ell=1}^{m} \mathcal{P}_{R_{\mathrm{tr}},s_N}(h_\ell, \varepsilon_*), \qquad \mathcal{P}_f := \mathcal{P}_{R_*,s_N}(f, \eta_*), \qquad \mathcal{S}_f := \mathcal{S}_{R_*,s_N}(f).$$

Lemma 4.2 gives shallow ReLU fields $g_1, \dots, g_m$ such that

$$\max_\ell \|g_\ell - h_\ell\|_{C(\overline{\mathrm{B}}_{R_{\mathrm{tr}}})} \leq \varepsilon_*, \qquad \mathrm{Str}(g_\ell) \leq \mathcal{S}_{R_{\mathrm{tr}},s_N}(h_\ell),$$

and a shallow ReLU field V_{base} such that

$$\|V_{\mathrm{base}} - f\|_{C(\overline{\mathrm{B}}_{R_*})} \leq \eta_*, \qquad p(V_{\mathrm{base}}) \leq \mathcal{P}_f, \qquad \mathrm{Str}(V_{\mathrm{base}}) \leq \mathcal{S}_f.$$

For each shallow ReLU field g, equation (4.2) with the identity normalization gives $|g(0)|, \mathrm{Lip}(g) \leq \mathrm{Str}(g)$. Hence

$$\mathfrak{B}(\mathbf{g})^2 \leq 2\sum_{\ell=1}^{m} \mathrm{Str}(g_\ell)^2 \leq 2\mathcal{S}_h^2 \leq \mathfrak{b}_h^2.$$

Every shallow ReLU field is globally Lipschitz, so all flow hypotheses in Proposition 4.3 are automatic. The bridge therefore gives $\alpha_* \in \mathrm{B}^m_{r_*}$ for which, with $G_\alpha := \sum_{\ell=1}^{m} \alpha_\ell g_\ell$,

$$\mathcal{E}(V_{\mathrm{base}} + G_{\alpha_*}) = Y. \tag{4.11}$$

Thus $\widehat{V} := V_{\mathrm{base}} + G_{\alpha_*}$ is one shallow autonomous ReLU field realizing all normalized endpoints exactly.

4.5 Width–strength accounting and proof of the quantitative theorem

Concatenate the parameter lists of V_{base} and the scaled fields $\alpha_{*,\ell} g_\ell$, and call the resulting list $\widehat{\theta}$. Then

$$\widehat{p} := p(\widehat{\theta}) \leq \mathcal{P}_f + \mathcal{P}_h, \qquad \mathrm{Str}(\widehat{\theta}) \leq \mathcal{S}_f + r_*\mathcal{S}_h. \tag{4.12}$$

Appendix D closes these quantities under one coarse envelope. Hence, after enlarging a dimension-only constant C_N,

$$\widehat{p},\ \mathrm{Str}(\widehat{\theta}) \leq \left(C_N \frac{K}{\rho}\right)^{C_N K}. \tag{4.13}$$

Proof of Theorem 2.10. For $K = 1$, admissibility forces the unique sample to be stationary and the empty parameter list suffices. Suppose $K \geq 2$ and use the normalized realization above. Write $\widehat{V}(z) = \sum_{j=1}^{\widehat{p}} \widehat{w}_j (\widehat{a}_j^\top z + \widehat{b}_j)_+$, and define

$$w_j := \frac{1}{T} B^{-1} \widehat{w}_j, \qquad a_j := B^\top \widehat{a}_j, \qquad b_j := \widehat{b}_j - a_j^\top x_*. \tag{4.14}$$

Then $V_\theta(x) = T^{-1}B^{-1}\widehat{V}(Sx)$, and uniqueness of solutions gives $S \circ \Phi^t_{V_\theta} = \Phi^{t/T}_{\widehat{V}} \circ S$. At $t = T$, equation (4.11) yields the original endpoint constraints. The affine pullback of each ridge unit is again one ridge unit, so the width is unchanged, while $T\,\mathrm{Str}_S(\theta) = \mathrm{Str}(\widehat{\theta})$. Combining (4.13) with (4.10) gives

$$p(\theta),\ T\,\mathrm{Str}_S(\theta) \le \big(C_N K \mathfrak{a}(\mathcal{V})\big)^{C_N K} = \mathfrak{U}_N(\mathcal{D}).$$

Finally, the positive-homogeneity balancing identity (4.3) gives (2.3). This proves (2.1)–(2.4). □

4.6 Interpretation, coordinate bounds, and examples

For a directly computable estimate in the original coordinates, define

$$\Gamma_{\mathrm{coord}}(\mathcal{V}) := \begin{cases} 1, & K = 1, \\ \dfrac{1 + \max_{z\in\mathcal{V}} |z|}{\min\{1, \mathrm{sep}(\mathcal{V})\}}, & K \ge 2. \end{cases}$$

Corollary 4.4 (Coordinate width–strength and individual-parameter bound)**.** *Let $N \ge 2$ and $T > 0$. The constant C_N in Theorem 2.10 may be chosen so that every admissible dataset admits a parameter list θ satisfying $\Phi^T_{V_\theta}(x_i) = y_i$ for $i = 1, \dots, M$ and*

$$p(\theta),\ T\,\mathrm{Str}(\theta) \le \big(C_N M \Gamma_{\mathrm{coord}}(\mathcal{V})\big)^{C_N M}. \tag{4.15}$$

Writing $W_\theta = [w_1 \ \cdots \ w_p]$, taking the rows of A_θ to be $a_j^\top$, and writing $b_\theta = (b_j)_{j=1}^p$, the units may moreover be balanced so that

$$\Big(\|W_\theta\|_{\mathrm{F}}^2 + \|A_\theta\|_{\mathrm{F}}^2 + \|b_\theta\|_2^2\Big)^{1/2} \le \sqrt{\frac{2}{T}}\,\big(C_N M \Gamma_{\mathrm{coord}}(\mathcal{V})\big)^{C_N M/2}. \tag{4.16}$$

In particular, the same right-hand side bounds every individual scalar parameter and also the operator norms of the two weight matrices.

The proof of Corollary 4.4 is given in Appendix D.

An explicit width-four example.
Take $N = 2$, $T = 1$, and the four inputs $x_1 = (1, 0)$, $x_2 = (-1, 0)$, $x_3 = (0, 2)$, and $x_4 = (0, -2)$, with targets $y_i = -x_i$. Define four ReLU units with zero biases by

j	1	2	3	4
a_j	$(1,0)$	$(-1,0)$	$(0,1)$	$(0,-1)$
w_j	$(0,\pi)$	$(0,-\pi)$	$(-\pi,0)$	$(\pi,0)$.

Since $s = s_+ - (-s)_+$, their sum is exactly

$$V(x) = \sum_{j=1}^{4} w_j (a_j^\top x)_+ = \pi(-x_2, x_1).$$

Hence $\Phi^t_V(x) = R_{\pi t}x$, where $R_{\pi t}$ is planar rotation through angle πt, and $\Phi^1_V(x_i) = -x_i = y_i$ for all four points. The field has width $p = 4$ and raw strength $\mathrm{Str}(\theta) = \sum_j |w_j||a_j| = 4\pi$. Figure 1 samples these four exact trajectories from the analytic formula. This small instance is far cheaper than the deliberately coarse worst-case envelope in Theorem 2.10.

Interpretation of the quantitative envelope. The coarse exponential dependence on K comes from protected finite-point relocation and derivative propagation through H; the remaining losses come from meeting the bridge tolerances. Optimizing the scaling is open.

Remark 4.5 (Time–strength scaling and the dimension-only constant)*.* Let $\widetilde{V}(z) := BV_\theta(B^{-1}z + x_*)$ and $\widetilde{F} := \Phi^T_{\widetilde{V}}$. By (4.2), $\mathrm{Lip}(\widetilde{V}), |\widetilde{V}(0)| \le \mathrm{Str}_S(\theta)$. Grönwall's inequality, applied to the forward and inverse flows, gives

$$e^{-T\,\mathrm{Str}_S(\theta)}|z - z'| \le |\widetilde{F}(z) - \widetilde{F}(z')| \le e^{T\,\mathrm{Str}_S(\theta)}|z - z'|,$$

so $T\,\mathrm{Str}_S(\theta)$ controls the logarithmic bi-Lipschitz distortion. The same argument with the growth bound $|\widetilde{V}(z)| \le \mathrm{Str}_S(\theta)(1 + |z|)$ yields

$$|\widetilde{F}(z) - z| \le (1 + |z|)\big(e^{T\,\mathrm{Str}_S(\theta)} - 1\big).$$

Hence any exact interpolant satisfies, for every nonstationary datum,

$$T\,\mathrm{Str}_S(\theta) \geq \log\left(1 + \frac{|Sy_i - Sx_i|}{1 + |Sx_i|}\right).$$

Under the normalization (2.2), $|Sx_i| \leq 1/2$ and $|Sy_i - Sx_i| \geq [4\mathfrak{a}(\mathcal{V})]^{-1}$, so

$$T\,\mathrm{Str}_S(\theta) \geq \log\left(1 + \frac{1}{6\mathfrak{a}(\mathcal{V})}\right).$$

Thus the inverse-T scaling in Theorem 2.10 reflects a genuine time–strength tradeoff rather than an artifact of the construction.

Here C_N depends only on N and absorbs fixed cutoff, approximation, and finite-order calculus constants; all data dependence remains explicit through K and $\mathfrak{a}(\mathcal{V})$. For the normalized unit-time list $\widehat{\theta}$, the pullback satisfies $T\,\mathrm{Str}_S(\theta) = \mathrm{Str}(\widehat{\theta})$.

Remark 4.6 (Two elementary exact width values). If all samples are stationary, then $p_{\min}(\mathcal{D}, T) = 0$. If $M = 1$ and $x_1 \neq y_1$, the one-unit field $V(x) = T^{-1}(y_1 - x_1)(0^\top x + 1)_+$ realizes the datum exactly, and hence $p_{\min}(\mathcal{D}, T) = 1$. Indeed, $p = 0$ gives only the zero vector field and therefore cannot move x_1 to the distinct point y_1.

5 Structural and metric obstructions

This section collects two complementary limitations of autonomous flows. In one dimension, universal exact point interpolation already fails by order preservation; see Remark 2.2. In dimensions $N \geq 2$, exact point interpolation is possible, but autonomy still imposes a structural obstruction to universally quantified fixed-radius neighborhood routing, as expressed by Theorem 2.7. We first prove this semigroup obstruction and its sharp point-to-neighborhood consequence, and then record the distinct metric obstruction caused by bounded time–strength.

5.1 Semigroup obstruction to neighborhood routing

The pointwise exact-interpolation theory does not extend to the universally quantified fixed-radius property. The following proof uses only continuity, autonomy, and the semigroup identity; it is independent of network width or approximation accuracy.

Proof of Theorem 2.7. Let $C_a := \overline{\mathrm{B}}_\delta(a)$, $C_b := \overline{\mathrm{B}}_\delta(b)$, and suppose, to the contrary, that $\Psi^T(C_a) \subset \mathrm{B}_\varepsilon(b)$ and $\Psi^T(C_b) \subset \mathrm{B}_\varepsilon(a)$. Because $\varepsilon < \delta$, $\mathrm{B}_\varepsilon(b) \subset C_b$. The semigroup identity gives

$$\Psi^{2T}(C_a) = \Psi^T(\Psi^T(C_a)) \subset \Psi^T(C_b) \subset \mathrm{B}_\varepsilon(a). \tag{5.1}$$

Choose ρ with $\varepsilon < \rho < \delta$ and set $K_a := \overline{\mathrm{B}}_\rho(a)$. Since $K_a \subset C_a$, equation (5.1) implies

$$\Psi^{2T}(K_a) \subset \mathrm{B}_\varepsilon(a) \subset K_a.$$

Thus $\Psi^{2T}|_{K_a} : K_a \to K_a$ is a continuous self-map of a nonempty compact convex set. Brouwer's fixed-point theorem yields $z \in K_a$ such that

$$\Psi^{2T}(z) = z. \tag{5.2}$$

The strict inclusion above implies $z \in \mathrm{B}_\varepsilon(a)$, while $\Psi^T(z) \in \mathrm{B}_\varepsilon(b)$. Since $|a - b| > 2\delta > 2\varepsilon$, these two balls are disjoint, so z lies on a nonconstant $2T$-periodic orbit.

For every $s \geq 0$, the semigroup identity and (5.2) give

$$\Psi^{2T}(\Psi^s(z)) = \Psi^s(\Psi^{2T}(z)) = \Psi^s(z). \tag{5.3}$$

Let $\gamma(s) := \Psi^s(z)$, $0 \leq s \leq T$. Its initial point lies in $\mathrm{B}_\varepsilon(a)$, whereas

$$|\gamma(T) - a| \geq |a - b| - |\gamma(T) - b| > 2\delta - \varepsilon > \delta.$$

By continuity, there exists $s_* \in (0, T)$ such that $|\gamma(s_*) - a| = (\varepsilon + \delta)/2$. Set $q_{\mathrm{orb}} := \gamma(s_*)$. Then $q_{\mathrm{orb}} \in C_a \setminus \overline{\mathrm{B}}_\varepsilon(a)$. Equation (5.1) gives $\Psi^{2T}(q_{\mathrm{orb}}) \in \mathrm{B}_\varepsilon(a)$, while (5.3) gives $\Psi^{2T}(q_{\mathrm{orb}}) = q_{\mathrm{orb}}$, a contradiction. ☐

Figure 5 summarizes the closed-ball return construction and the periodic-orbit contradiction.

Proof of Corollary 2.8. The implication at $\delta = 0$ is Corollary 2.5. Let $\delta > 0$ and suppose that (P_δ) holds. Choose $a, b \in \mathbb{R}^N$ with $|a - b| > 2\delta$ and $0 < \varepsilon < \delta$. Applying (P_δ) with the two target centers exchanged gives

$$\Phi^T(\overline{\mathrm{B}}_\delta(a)) \subset \mathrm{B}_\varepsilon(b), \qquad \Phi^T(\overline{\mathrm{B}}_\delta(b)) \subset \mathrm{B}_\varepsilon(a),$$

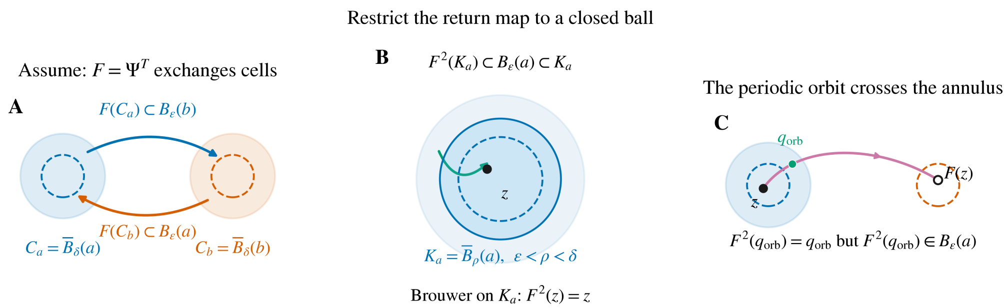

Figure 5: Why an autonomous semiflow cannot exchange and compress two balls. Write $F = \Psi^T$. (A) Assume F sends each closed source ball of radius δ into the open ball of radius $\varepsilon < \delta$ around the other center. (B) For $\varepsilon < \rho < \delta$, the second iterate maps the closed ball K_a of radius ρ into itself, so Brouwer's theorem gives $F^2(z) = z$. (C) Since F^2 commutes with Ψ^s, every point on the semiflow path from z to $F(z)$ is fixed by F^2. This path crosses the annulus at q with $\varepsilon < |q - a| < \delta$, but the assumed return inclusion requires $F^2(q)$ to lie inside the ε-ball, a contradiction.

contradicting Theorem 2.7. Applying SCC to the compact convex cells $\overline{\mathrm{B}}_\delta(a)$ and $\overline{\mathrm{B}}_\delta(b)$ would imply, a fortiori, the open-ball routing configuration forbidden by Theorem 2.7. Hence SCC fails as well. □

Remark 5.1 (Structural meaning of the two-cell obstruction). The obstruction is structural rather than a capacity bound: autonomy forces $\Psi^{2T} = \Psi^T \circ \Psi^T$, producing the strict return and periodic-orbit contradiction above. It is independent of width and parameter magnitude and concerns a radius fixed before the dynamics are chosen; data-dependent or almost-everywhere routing is not excluded.

5.2 Metric obstruction under bounded strength

A distinct limitation follows from quantitative invertibility: bounded time–strength prevents arbitrary collapse of within-cell spread. The following deterministic push-forward estimate is not a statistical generalization or learning-rate statement.

Within-cell squared-error floor for piecewise-constant targets. Let $q \in \mathbb{N}$, and let $A_1, \ldots, A_q \subset \mathbb{R}^N$ be pairwise disjoint Borel sets. Put $\mathcal{A} := (A_1, \ldots, A_q)$, let μ be a probability measure supported on their union, and put $\pi_k := \mu(A_k) > 0$ and $\mu_k := \pi_k^{-1}\mu|_{A_k}$. Assume that every μ_k has finite second moment. For target vectors $r_1, \ldots, r_q \in \mathbb{R}^N$, define $f_*(x) := r_k$ for $x \in A_k$, and define f_* arbitrarily outside $\bigcup_{k=1}^q A_k$. Given an affine map $S(x) = B(x - x_*)$, the normalized squared population risk of a predictor $F : \mathbb{R}^N \to \mathbb{R}^N$ is

$$\mathcal{R}_S(F) := \int_{\mathbb{R}^N} |B(F(x) - f_*(x))|^2 \, \mathrm{d}\mu(x). \tag{5.4}$$

Let $\nu_k := S_\# \mu_k$. The corresponding normalized within-cell spread is

$$\mathcal{W}_S(\mu, \mathcal{A}) := \sum_{k=1}^q \pi_k \operatorname{Var}(\nu_k), \qquad \overline{z}_\nu := \int z \, \mathrm{d}\nu(z), \qquad \operatorname{Var}(\nu) := \int |z - \overline{z}_\nu|^2 \, \mathrm{d}\nu(z). \tag{5.5}$$

The identity

$$\operatorname{Var}(\nu) = \frac{1}{2} \iint |z - z'|^2 \, \mathrm{d}\nu(z) \, \mathrm{d}\nu(z') \tag{5.6}$$

will be used below.

Even outside the two-cell exchange configuration, quantitative invertibility imposes the following compression cost.

Proposition 5.2 (Bounded-distortion obstruction to neighborhood compression). *Let $V_\theta \in \mathscr{N}_N$, let $F_\theta := \Phi_{V_\theta}^T$, and let $S(x) = B(x - x_*)$, with $B \in \mathrm{GL}(N)$, be an affine change of coordinates. For the piecewise-constant regression problem in* (5.4)–(5.5)*, the conjugated terminal map $\widetilde{F} = S \circ F_\theta \circ S^{-1}$ obeys the co-Lipschitz estimate*

$$|\widetilde{F}(z) - \widetilde{F}(z')| \geq e^{-T \operatorname{Str}_S(\theta)} |z - z'|.$$

Consequently, bounded time–strength prevents arbitrary contraction of pairwise distances and within-region variance. As an application,

$$\mathcal{R}_S(F_\theta) \geq e^{-2T\,\mathrm{Str}_S(\theta)}\mathcal{W}_S(\mu, \mathcal{A}). \tag{5.7}$$

Consequently, for every budget $Q \geq 0$,

$$\inf_{\theta:\, T\,\mathrm{Str}_S(\theta)\leq Q} \mathcal{R}_S(F_\theta) \geq e^{-2Q}\mathcal{W}_S(\mu, \mathcal{A}). \tag{5.8}$$

Equivalently, every parameter list with $\mathcal{R}_S(F_\theta) > 0$ satisfies the compression-cost bound

$$T\,\mathrm{Str}_S(\theta) \geq \frac{1}{2}\log_+\left(\frac{\mathcal{W}_S(\mu, \mathcal{A})}{\mathcal{R}_S(F_\theta)}\right). \tag{5.9}$$

Corollary 5.3 (Within-cell squared-error floor for piecewise-constant targets)**.** *Let $N \geq 2$ and $T > 0$. Assume, in addition, that the targets $r_1, \ldots, r_q$ are pairwise distinct, and choose one point $x_k \in A_k$ from each cell. Apply Theorem 2.10 to $\mathcal{D} = \{(x_k, r_k)\}_{k=1}^q$, and let S be the affine normalization supplied by that theorem. The resulting exact interpolant satisfies*

$$\mathcal{R}_S(F_\theta) \geq e^{-2\mathfrak{U}_N(\mathcal{D})}\mathcal{W}_S(\mu, \mathcal{A}).$$

Proof of Proposition 5.2. Conjugate the field by S, setting $\widetilde{V}(z) := BV_\theta(B^{-1}z + x_*)$ and $\widetilde{F} := \Phi^T_{\widetilde{V}} = S \circ F_\theta \circ S^{-1}$. By (4.2), $\mathrm{Lip}(\widetilde{V}) \leq \mathrm{Str}_S(\theta)$. Because $\widetilde{V}$ is globally Lipschitz, its flow is defined for both positive and negative times. Grönwall's inequality for the forward flow gives the usual upper Lipschitz estimate. Applying the same estimate to the inverse flow $\Phi^{-T}_{\widetilde{V}}$ and rearranging gives

$$\left|\widetilde{F}(z) - \widetilde{F}(z')\right| \geq e^{-T\,\mathrm{Str}_S(\theta)}\left|z - z'\right|. \tag{5.10}$$

Fix a cell A_k, and let Z, Z' be independent random variables with law $\nu_k = S_\#\mu_k$, and set $Z_T := \widetilde{F}(Z)$, $Z'_T := \widetilde{F}(Z')$. The variance decomposition and (5.10) imply

$$\int \left|\widetilde{F}(z) - S(r_k)\right|^2 \mathrm{d}\nu_k(z) \geq \mathrm{Var}(Z_T) = \frac{1}{2}\mathbb{E}\left|Z_T - Z'_T\right|^2 \geq \frac{1}{2}e^{-2T\,\mathrm{Str}_S(\theta)}\mathbb{E}\left|Z - Z'\right|^2 = e^{-2T\,\mathrm{Str}_S(\theta)}\,\mathrm{Var}(\nu_k),$$

where the last identity is (5.6). Multiply by π_k and sum over k to obtain (5.7). The budget estimate (5.8) is immediate. Rearranging (5.7), and using the trivial lower bound $T\,\mathrm{Str}_S(\theta) \geq 0$, gives (5.9). ☐

Proof of Corollary 5.3. The selected dataset is admissible because the cells are disjoint and the targets are pairwise distinct. Theorem 2.10 provides an exact interpolant satisfying $T\,\mathrm{Str}_S(\theta) \leq \mathfrak{U}_N(\mathcal{D})$. Insert this upper bound into (5.7). ☐

Remark 5.4 (Two-level obstruction and relative compression cost)*.* Theorem 2.7 is strength-independent and structural, whereas Proposition 5.2 is metric and applies beyond the two-cell pattern. If an exact interpolant satisfies $T\,\mathrm{Str}_S(\theta_Q) \leq Q$, then

$$\mathcal{R}_S(F_{\theta_Q}) \geq e^{-2Q}\mathcal{W}_S(\mu, \mathcal{A}).$$

Hence, when $\mathcal{W}_S(\mu, \mathcal{A}) > 0$, the requirement $\mathcal{R}_S(F_{\theta_Q}) \leq \varepsilon_{\mathrm{rel}}^2\mathcal{W}_S(\mu, \mathcal{A})$ forces

$$T\,\mathrm{Str}_S(\theta_Q) \geq \log(1/\varepsilon_{\mathrm{rel}}).$$

6 Discussion, scope, and open problems

The results above concern finite-sample exact interpolation and its autonomous neighborhood-scale limitations. They do not establish statistical generalization, density estimation, or distributional approximation. Repeated labels require a noninjective readout or target regions, and Proposition 5.2 is a deterministic bounded-distortion estimate. This section records the principal distinctions and open directions.

6.1 Comparisons and scope

Measure transport versus labeled routing. Autonomous transport of one probability measure to another is a different problem from routing prescribed labels attached to finitely many particles. A pushforward specifies a terminal distribution but does not prescribe a terminal point for every initial particle. Accordingly, the two-cell obstruction does not transfer verbatim to measure transport. Relevant exact and approximate transport results are [12, 20, 4]; their relation to autonomous neural or vector-field classes under common-time and regularity constraints remains a separate controllability question.

Approximation inputs versus endpoint exactness. Classical density results for nonpolynomial ridge networks provide accurate vector-field approximants [11, 18, 23]; the quantitative ReLU argument used here additionally draws on coefficient-controlled approximation and smoothness-to-variation estimates [19, 26, 27]. Exact satisfaction of the endpoint system, however, comes from the finite-dimensional correction and degree argument, not from density alone. Preserving finitely many equilibria during vector-field approximation [22] is likewise distinct from prescribing arbitrary nonstationary values of one common autonomous time map.

6.2 Open problems and research directions

Autonomous time maps and weaker neighborhood notions. Corollary 2.9 rules out uniform universal approximation of the full class of continuous maps, but it leaves open a characterization of maps in the uniform closure of autonomous time maps. The two-ball obstruction gives one condition stable under small uniform errors. A useful next step is to characterize this closure within specified classes of smooth invertible targets and to compare it with closure in weaker topologies. At the neighborhood scale, it remains to determine which data-dependent radii, target regions, or readouts permit useful alternatives to universal fixed-radius routing.

Model classes, closure, and stability. Theorem 2.4 uses linear closure so that the base approximant and finitely many endpoint corrections remain in the class. It is natural to ask whether weaker algebraic or local closure conditions suffice, and whether analogous exactification principles hold for fixed-width, sparse, or otherwise nonlinear parameterized dictionaries. Separately, the theorem produces an interpolating field for each admissible dataset but does not select one continuously: understanding continuous or Lipschitz selection, and the deterioration of endpoint-frame conditioning near collisions or loss of separation, is open.

Readouts and measure-level controllability. Repeated class labels cannot be represented directly by an injective flow map, but a noninjective readout, target regions, or an augmented state may remove this immediate obstruction [24]. Which interpolation, margin, or neighborhood guarantees survive after such a readout? At measure level, for which pairs (μ_0, μ_1) can a prescribed autonomous neural or vector-field class satisfy $F_{\#}\mu_0 = \mu_1$ at one common terminal time, exactly or approximately, and which injectivity, regularity, support-geometry, and common-time restrictions remain after labeled endpoint constraints are removed? The answer may differ for atomic, absolutely continuous, and singular measures.

Complexity and construction. The quantitative envelope is deliberately coarse. Improving its dependence on sample number and affine geometry, proving matching lower bounds for width and path strength, and finding algorithmically implementable and numerically stable exact-interpolation procedures remain open; the endpoint coordinates developed here may provide useful variables for such constructions.

6.3 Conclusion

In dimension at least two, uniform approximation of smooth compactly supported vector fields within a linear class whose members generate global flows is sufficient for exact finite interpolation. The proof combines a smooth reference construction with localized endpoint corrections and a degree argument. It applies in particular to shallow ReLU fields and gives quantitative bounds on width and normalized coefficient strength. The two-ball obstruction shows that this interpolation property does not extend to universal control of fixed neighborhoods. Determining the optimal interpolation cost and the maps that remain approximable under autonomy are natural next questions.

Acknowledgements

The authors thank Antonio Álvarez-López for fruitful suggestions. Enrique Zuazua was supported by the European Research Council (ERC) under the European Union's Horizon Europe research and innovation programme (grant agreement No. 101096251, CoDeFeL); the Air Force Office of Scientific Research under award number FA8655-22-1-7012; the Alexander von Humboldt Professorship program; the European Union's Horizon Europe MSCA project ModConFlex (HORIZON-MSCA-2021-DN-01, project 101073558); SURE-AI: The Norwegian Centre for Sustainable, Risk-Averse, and Ethical AI, grant 357482, Research Council of Norway; Grant PID2023-146872OB-I00-DyCMaMod of MICIU (Spain); and the COST Actions CA24122, *Multiscale Stochastics, Patterns, and Analysis of Combinatorial Environments*, and CA24136, *Interactions between Control Theory and Machine Learning.*

A Geometric construction and endpoint-correction estimates

This appendix records the geometric and fixed-order estimates for the reference realization and endpoint correction directions constructed in Section 3.

Constant and norm conventions. An unsubscripted C denotes a constant depending only on N and, where stated, on a fixed derivative order; its value may change between displayed lines. For a derivative of order q, $\|D^q H\|_\infty$ denotes the supremum of its q-linear operator norm (and $\|DH\|_\infty$ the usual operator norm). Fixed dimension-dependent equivalence constants between this convention and coordinatewise C^s norms are absorbed into named bounds. Named constants are retained when they enter the quantitative ledger in Appendix D.

A.1 Localized cylinder pushes and derivative bounds

Lemma A.1 (Quantitative axial push)**.** *For every $\ell > 0$, $0 < r \le 1$, and fixed integer $s \ge 1$, there exists $h_{\ell,r} \in \mathrm{Diff}^\infty(\mathbb{R})$ such that*

$$h_{\ell,r}(0) = \ell, \qquad h_{\ell,r}(\sigma) = \sigma \quad \text{for } \sigma \notin [-r, \ell + r],$$

and, with constants depending only on s,

$$\inf h'_{\ell,r} \ge c_s \frac{r}{\ell + r}, \qquad 1 + \max_{1 \le q \le s} \left(\|h_{\ell,r}^{(q)}\|_\infty + \|(h_{\ell,r}^{-1})^{(q)}\|_\infty \right) \le \left(\frac{C_s(1+\ell)}{r} \right)^{C_s}.$$

Proof. Fix a smooth step function $\chi \in C^\infty(\mathbb{R}; [0,1])$ that vanishes on $(-\infty, 0]$, equals one on $[1, \infty)$, and is flat at 0 and 1. Put

$$\delta := \frac{r^2}{100(\ell + r)}$$

and define

$$\psi_-(\sigma) := \chi\left(\frac{\sigma + r}{\delta}\right) \chi\left(\frac{-\sigma}{\delta}\right), \qquad \psi_+(\sigma) := \chi\left(\frac{\sigma}{\delta}\right) \chi\left(\frac{\ell + r - \sigma}{\delta}\right).$$

Then $\psi_-\psi_+ \equiv 0$, $\operatorname{supp} \psi_- \subset [-r, 0]$, $\operatorname{supp} \psi_+ \subset [0, \ell + r]$, and, writing $I_\pm := \int_{\mathbb{R}} \psi_\pm$,

$$r - 2\delta \le I_- \le r, \qquad \ell + r - 2\delta \le I_+ \le \ell + r.$$

Set

$$\lambda(\sigma) := 1 + \frac{\ell}{I_-}\psi_-(\sigma) - \frac{\ell}{I_+}\psi_+(\sigma), \qquad h_{\ell,r}(\sigma) := -r + \int_{-r}^{\sigma} \lambda(\tau)\, \mathrm{d}\tau$$

on $[-r, \ell + r]$, and extend by the identity outside this interval. The normalized masses give $h_{\ell,r}(0) = \ell$ and $h_{\ell,r}(\ell + r) = \ell + r$. Moreover, $\lambda \ge 1$ on $(-r, 0)$, while on $(0, \ell + r)$, since $\delta \le r/100$,

$$\lambda \ge 1 - \frac{\ell}{I_+} \ge \frac{r - 2\delta}{\ell + r - 2\delta} \ge \frac{49}{50} \frac{r}{\ell + r}.$$

Thus $h_{\ell,r}$ is an increasing smooth diffeomorphism.

Let $A := (1 + \ell)/r \ge 1$. Since $\delta^{-1} \le 100A^2$, fixed-cutoff scaling gives $\|\psi_\pm^{(q)}\|_\infty \le C_q A^{2q}$. Also $\ell/I_- \le 2A$ and $\ell/I_+ \le 2$. Hence the derivatives of $h_{\ell,r}$ through every fixed order are bounded by a fixed power of A. The lower bound on $h'_{\ell,r}$ gives $\|(h_{\ell,r}^{-1})'\|_\infty \le CA$; repeated differentiation of $h_{\ell,r} \circ h_{\ell,r}^{-1} = \mathrm{Id}$ and the one-dimensional Faà di Bruno formula give the same fixed-power bound for higher inverse derivatives. Enlarging C_s proves the claim. $\square$

Lemma A.2 (Finite-order composition and pushforward bounds)**.** *Fix $N \ge 1$ and an integer $s \ge 1$. There exists $C_{\mathrm{comp},N,s} \ge 1$, depending only on N and s, with the following properties.*

(i) *If $P_1, \dots, P_n$ are C^s diffeomorphisms and, for some $D \ge 1$,*

$$1 + \max_{1 \le j \le n} \max_{1 \le q \le s} \left(\|D^q P_j\|_\infty + \left\|D^q P_j^{-1}\right\|_\infty \right) \le D,$$

then, for $H = P_n \circ \cdots \circ P_1$,

$$1 + \max_{1 \le q \le s} \left(\|D^q H\|_\infty + \left\|D^q H^{-1}\right\|_\infty \right) \le \left(C_{\mathrm{comp},N,s} D\right)^{C_{\mathrm{comp},N,s}\, n}.$$

(ii) *If H is a C^{s+1} diffeomorphism, $u \in C_c^s(\mathbb{R}^N; \mathbb{R}^N)$, and*

$$\mathfrak{H}_{s+1} := 1 + \max_{1 \le q \le s+1} \left(\|D^q H\|_\infty + \left\|D^q H^{-1}\right\|_\infty \right),$$

then

$$\|H_* u\|_{C^s} \le C_{\mathrm{comp},N,s}\big(1+\|u\|_{C^s}\big)\mathfrak{H}_{s+1}^{C_{\mathrm{comp},N,s}}.$$

Proof. For part (i), let Π_q denote the set of partitions of $\{1,\dots,q\}$. The set-partition form of the multivariate Faà di Bruno formula gives

$$D^q(F\circ G)(x)[v_1,\dots,v_q] = \sum_{\pi\in\Pi_q} D^{|\pi|}F(G(x))\Big[D^{|B|}G(x)[v_i]_{i\in B}\Big]_{B\in\pi}.$$

The point needed below is that every term satisfies $\sum_{B\in\pi}|B| = q$.

Set

$$H_k := P_k\circ\cdots\circ P_1, \qquad k=1,\dots,n.$$

We claim that, for a sufficiently large constant $C_0 = C_0(N,s)\ge 2$,

$$\|D^q H_k\|_\infty \le (C_0 D)^{C_0 q k}, \qquad 1\le q\le s, \quad 1\le k\le n.$$

The case $k=1$ follows from the hypothesis. Suppose the claim holds for H_k. Since $H_{k+1} = P_{k+1}\circ H_k$, Faà di Bruno yields

$$\begin{aligned}
\|D^q H_{k+1}\|_\infty &\le C_{N,s}\sum_{\pi\in\Pi_q}\Big\|D^{|\pi|}P_{k+1}\Big\|_\infty \prod_{B\in\pi}\Big\|D^{|B|}H_k\Big\|_\infty \\
&\le C_{N,s}|\Pi_q| D \max_{\pi\in\Pi_q}\prod_{B\in\pi}(C_0 D)^{C_0|B|k} \\
&= C_{N,s}|\Pi_q| D (C_0 D)^{C_0 q k}.
\end{aligned}$$

Here the last identity is precisely where $\sum_{B\in\pi}|B| = q$ is used. Since $q\le s$, the numbers $|\Pi_q|$ are bounded in terms of s alone; after enlarging C_0,

$$\|D^q H_{k+1}\|_\infty \le (C_0 D)^{C_0 q(k+1)}.$$

Thus the exponent grows only linearly with the number of composed factors, rather than through an iterated C^s composition bound.

Taking $k=n$ and enlarging the constant gives

$$1+\max_{1\le q\le s}\|D^q H\|_\infty \le \big(C_{\mathrm{comp},N,s}D\big)^{C_{\mathrm{comp},N,s}n}.$$

Moreover, $H^{-1} = P_1^{-1}\circ\cdots\circ P_n^{-1}$, and the inverse factors satisfy the same derivative hypothesis. Applying the same argument to this reverse-order composition gives the corresponding estimate for H^{-1} and proves part (i).

For part (ii), put

$$K := H^{-1}, \qquad \Lambda := \mathfrak{H}_{s+1}, \qquad A := DH\circ K, \qquad B := u\circ K.$$

Then

$$H_* u = AB.$$

For $0\le r\le s$, Faà di Bruno and $\Lambda\ge 1$ give

$$\|D^r B\|_\infty \le C_{N,s}\,\|u\|_{C^s}\,\Lambda^r.$$

Indeed, for $r\ge 1$,

$$D^r B(x)[v_1,\dots,v_r] = \sum_{\pi\in\Pi_r} D^{|\pi|}u(K(x))\Big[D^{|B'|}K(x)[v_i]_{i\in B'}\Big]_{B'\in\pi},$$

and every term contains at most r factors involving derivatives of K.

Similarly,

$$\|D^r A\|_\infty \le C_{N,s}\Lambda^{r+1}, \qquad 0\le r\le s.$$

Indeed, the corresponding Faà di Bruno terms contain $D^{|\pi|+1}H$, together with $|\pi|$ derivatives of K. In particular, derivatives of H are needed only through order $r+1\le s+1$, while derivatives of H^{-1} are needed only through order $r\le s$.

The finite-order Leibniz rule now yields, for $0\le q\le s$,

$$\begin{aligned}
\|D^q(H_* u)\|_\infty &\le C_{N,s}\sum_{r=0}^q\binom{q}{r}\|D^r A\|_\infty\,\big\|D^{q-r}B\big\|_\infty \\
&\le C_{N,s}\,\|u\|_{C^s}\sum_{r=0}^q\binom{q}{r}\Lambda^{r+1}\Lambda^{q-r} \\
&\le C_{N,s}\,\|u\|_{C^s}\,\Lambda^{q+1}.
\end{aligned}$$

The useful bookkeeping here is $(r+1)+(q-r)=q+1$: the total power of Λ is independent of how the q derivatives are distributed between the two factors. Hence

$$\|H_*u\|_{C^s} \le C_{N,s}\,\|u\|_{C^s}\,\Lambda^{s+1}.$$

Since $\Lambda \ge 1$, enlarging $C_{\mathrm{comp},N,s}$ absorbs this stronger estimate into the form stated in the lemma.

Finally, $\operatorname{supp}(H_*u) \subset H(\operatorname{supp} u)$, so H_*u is compactly supported. This completes the proof. □

Proof of Lemma 3.2. We first treat the case $K \ge 2$.

Choose parking points $p_j = (R+2+3j)e_1$, $j=1,\dots,K$. First move the z_j's to the p_j's one at a time, and then move the p_j's to the w_j's. The points that are not currently moving are regarded as obstacles. During the first phase they are a mixture of original and parking points, and during the second phase they are a mixture of parking and target points. The parking points are mutually separated by three and lie at distance at least five from $\overline{\mathrm{B}}_R$. Consequently, in every move, both the initial and terminal positions of the moving point are at distance at least d_{cfg} from every obstacle, and the obstacles are mutually d_{cfg}-separated.

A clear two-segment route. Consider one move from a to b, with obstacles $o_1,\dots,o_{K-1}$. All current points lie in $\overline{\mathrm{B}}_{R+3K+2}$. Put $R_{\mathrm{sh}} := R+4K+1$ and $\mathcal{A}_{\mathrm{sh}} := \{\xi : R_{\mathrm{sh}}-1 < |\xi| < R_{\mathrm{sh}}\}$. Writing ω_N for the unit-ball volume, set $c_N^{\mathrm{sh}} := N\omega_N/2^{N-1}$. Since $R_{\mathrm{sh}} \ge 2$,

$$|\mathcal{A}_{\mathrm{sh}}| = \omega_N\big(R_{\mathrm{sh}}^N - (R_{\mathrm{sh}}-1)^N\big) \ge N\omega_N(R_{\mathrm{sh}}-1)^{N-1} \ge c_N^{\mathrm{sh}} R_{\mathrm{sh}}^{N-1}.$$

Fix a dimension-only constant $C_N^{\mathrm{cap}} \ge 1$, large enough to absorb the spherical-cap and radial-integration constants below, including the fixed factor 8^{N-1}. Define

$$C_N^{\mathrm{mid}} := 2 + \frac{2C_N^{\mathrm{cap}}}{c_N^{\mathrm{sh}}}, \qquad \delta_{\mathrm{tr}} := \frac{d_{\mathrm{cfg}}}{C_N^{\mathrm{mid}}(1+K)(1+R_{\mathrm{sh}})}.$$

In particular, $4\delta_{\mathrm{tr}} \le d_{\mathrm{cfg}}/2$.

Fix $a_{\mathrm{end}} \in \{a,b\}$ and one obstacle o_j. If $[a_{\mathrm{end}},\xi]$, with $\xi \in \mathcal{A}_{\mathrm{sh}}$, intersects $\mathrm{B}_{4\delta_{\mathrm{tr}}}(o_j)$, choose an intersection point q. Since $4\delta_{\mathrm{tr}} < |o_j - a_{\mathrm{end}}|$, the intersection does not occur at a_{end}, and

$$(q-a_{\mathrm{end}})^{\top}(o_j - a_{\mathrm{end}}) \ge |o_j - a_{\mathrm{end}}|\big(|o_j-a_{\mathrm{end}}| - 4\delta_{\mathrm{tr}}\big) > 0.$$

Thus the angle θ_{ang} between $\xi - a_{\mathrm{end}}$ and $o_j - a_{\mathrm{end}}$ satisfies $0 \le \theta_{\mathrm{ang}} < \pi/2$ and $\sin\theta_{\mathrm{ang}} \le 4\delta_{\mathrm{tr}}/|o_j - a_{\mathrm{end}}| \le 4\delta_{\mathrm{tr}}/d_{\mathrm{cfg}} \le 1/2$. Hence $\theta_{\mathrm{ang}} \le 2\sin\theta_{\mathrm{ang}} \le 8\delta_{\mathrm{tr}}/d_{\mathrm{cfg}}$, so the direction $(\xi - a_{\mathrm{end}})/|\xi - a_{\mathrm{end}}|$ belongs to a spherical cap of angular radius

$$\theta_0 := \frac{8\delta_{\mathrm{tr}}}{d_{\mathrm{cfg}}} \le 1.$$

For any $e \in \mathbb{S}^{N-1}$, spherical coordinates give

$$\mathcal{H}^{N-1}\big(\{\omega \in \mathbb{S}^{N-1} : \angle(\omega,e) \le \theta_0\}\big) = |\mathbb{S}^{N-2}| \int_0^{\theta_0} (\sin\theta)^{N-2}\,\mathrm{d}\theta \le \frac{|\mathbb{S}^{N-2}|}{N-1}\theta_0^{N-1}.$$

Moreover, $|a_{\mathrm{end}}| \le R+3K+2 < R_{\mathrm{sh}}$, so every $\xi \in \mathcal{A}_{\mathrm{sh}}$ satisfies $|\xi - a_{\mathrm{end}}| \le 2R_{\mathrm{sh}}$. Enlarging the bad set to the corresponding cone and integrating radially about a_{end} therefore gives

$$\begin{aligned} |\{\xi \in \mathcal{A}_{\mathrm{sh}} : [a_{\mathrm{end}},\xi] \cap \mathrm{B}_{4\delta_{\mathrm{tr}}}(o_j) \ne \varnothing\}| &\le \frac{|\mathbb{S}^{N-2}|}{N-1}\theta_0^{N-1}\int_0^{2R_{\mathrm{sh}}} r^{N-1}\,\mathrm{d}r \\ &\le C_N^{\mathrm{cap}} R_{\mathrm{sh}}^N \left(\frac{\delta_{\mathrm{tr}}}{d_{\mathrm{cfg}}}\right)^{N-1}. \end{aligned}$$

After a union bound over the two endpoints and $K-1$ obstacles, the ratio of bad volume to $|\mathcal{A}_{\mathrm{sh}}|$ is, since $0 < \delta_{\mathrm{tr}}/d_{\mathrm{cfg}} = [C_N^{\mathrm{mid}}(1+K)(1+R_{\mathrm{sh}})]^{-1} < 1$ and $N \ge 2$, at most

$$\frac{2C_N^{\mathrm{cap}}}{c_N^{\mathrm{sh}}} K R_{\mathrm{sh}} \left(\frac{\delta_{\mathrm{tr}}}{d_{\mathrm{cfg}}}\right)^{N-1} \le \frac{2C_N^{\mathrm{cap}}}{c_N^{\mathrm{sh}} C_N^{\mathrm{mid}}} \frac{K}{1+K}\frac{R_{\mathrm{sh}}}{1+R_{\mathrm{sh}}} < 1.$$

Thus some $\xi_{\mathrm{sh}} \in \mathcal{A}_{\mathrm{sh}}$ makes both $[a,\xi_{\mathrm{sh}}]$ and $[\xi_{\mathrm{sh}},b]$ stay at distance at least $4\delta_{\mathrm{tr}}$ from every obstacle. This angular argument is the only place where $N \ge 2$ is essential.

One straight-cylinder push. We now use the quantitative axial construction above.

Now consider an oriented segment of length $\ell_{\mathrm{seg}} > 0$ and a radius $0 < r_{\mathrm{cyl}} \le 1$. After an orthogonal change of coordinates, write points as $(\sigma,\xi) \in \mathbb{R}\times\mathbb{R}^{N-1}$, with endpoints $(0,0)$ and $(\ell_{\mathrm{seg}},0)$, and take $h_{\mathrm{ax}} := h_{\ell_{\mathrm{seg}},r_{\mathrm{cyl}}}$ from Lemma A.1.

Choose a fixed $\zeta \in C_c^\infty(\mathrm{B}_1^{N-1})$, $0 \le \zeta \le 1$, with $\zeta \equiv 1$ on $\mathrm{B}_{1/3}^{N-1}$, and put $\zeta_r(\xi) := \zeta(\xi/r_{\mathrm{cyl}})$. Define

$$\mathscr{P}_{\mathrm{cyl}}(\sigma,\xi) := \big(\sigma + \zeta_r(\xi)(h_{\mathrm{ax}}(\sigma) - \sigma), \xi\big).$$

Its axial derivative is $(1-\zeta_r)+\zeta_r h'_{\mathrm{ax}}>0$, so $\mathscr{P}_{\mathrm{cyl}}$ is a diffeomorphism, is the identity near the cylinder boundary, and sends $(0,0)$ to $(\ell_{\mathrm{seg}},0)$. For clarity, the inverse axial coordinate $\sigma=\sigma(\tau,\xi)$ is determined by $\tau=\sigma+\zeta_r(\xi)(h_{\mathrm{ax}}(\sigma)-\sigma)$. Its axial derivative $d=1-\zeta_r(\xi)+\zeta_r(\xi)h'_{\mathrm{ax}}(\sigma)$ is a convex combination of 1 and h'_{ax}, hence is bounded below by a positive constant times $r_{\mathrm{cyl}}/(\ell_{\mathrm{seg}}+r_{\mathrm{cyl}})$. Implicit differentiation gives $\partial_\tau\sigma=1/d$ and $\partial_{\xi_j}\sigma=-\partial_{\xi_j}\zeta_r(\xi)(h_{\mathrm{ax}}(\sigma)-\sigma)/d$. The displacement and cutoff bounds, followed by repeated implicit differentiation, bound every fixed-order inverse derivative by a fixed power of $(1+\ell_{\mathrm{seg}})/r_{\mathrm{cyl}}$. Thus there is a constant $C_{\mathrm{seg,N,s}}\geq 1$, depending only on N,s, such that

$$1+\max_{1\leq q_{\mathrm{der}}\leq s}\left(\|D^{q_{\mathrm{der}}}\mathscr{P}_{\mathrm{cyl}}\|_\infty+\left\|D^{q_{\mathrm{der}}}\mathscr{P}_{\mathrm{cyl}}^{-1}\right\|_\infty\right)\leq\left(\frac{C_{\mathrm{seg,N,s}}(1+\ell_{\mathrm{seg}})}{r_{\mathrm{cyl}}}\right)^{C_{\mathrm{seg,N,s}}}. \tag{A.1}$$

In the routed construction take $r_{\mathrm{cyl}}=\delta_{\mathrm{tr}}/2$. The support then lies within distance $<\delta_{\mathrm{tr}}$ of the segment and hence fixes every current obstacle.

Composition for $K\geq 2$. There are at most $2K$ point moves and hence at most $4K$ segment pushes. Every routed segment has $\ell_{\mathrm{seg}}\leq 2R_{\mathrm{sh}}$, while $r_{\mathrm{cyl}}^{-1}=2C_N^{\mathrm{mid}}(1+K)(1+R_{\mathrm{sh}})/d_{\mathrm{cfg}}$ and $1+R_{\mathrm{sh}}\leq 4(1+K)(1+R)$. Consequently,

$$\frac{1+\ell_{\mathrm{seg}}}{r_{\mathrm{cyl}}}\leq\left(\frac{C_N^{\mathrm{len}}(1+K)(1+R)}{d_{\mathrm{cfg}}}\right)^3,\qquad C_N^{\mathrm{len}}:=8\max\{1,C_N^{\mathrm{mid}}\}. \tag{A.2}$$

Apply Lemma A.2(i) to the at most $4K$ segment pushes and to the reverse-order inverse composition. Choose

$$C_{\mathrm{tr},N,s}:=12C_{\mathrm{comp},N,s}\big(1+C_{\mathrm{seg},N,s}+C_N^{\mathrm{len}}\big).$$

Combining Lemma A.2(i) with (A.1) and (A.2) gives (3.2). Each push fixes all current obstacles and all already parked or placed points. Moreover, every routed segment lies in $\mathrm{B}_{R_{\mathrm{sh}}}$, and every support cylinder lies in $\mathrm{B}_{R_{\mathrm{sh}}+1}=\mathrm{B}_{R+4K+2}\Subset\Omega_{\mathrm{tr}}$. This proves the lemma for $K\geq 2$.

It remains to treat $K=1$. If $z_1=w_1$, take $H=\mathrm{Id}$. Otherwise apply the straight-cylinder construction above directly to the single segment $[z_1,w_1]$, with $\ell_{\mathrm{seg}}=|w_1-z_1|\leq 2R$ and $r_{\mathrm{cyl}}:=1/4$. No routing is needed. The support cylinder lies within distance $\sqrt{2}\,r_{\mathrm{cyl}}<1$ of $[z_1,w_1]$, and hence in $\mathrm{B}_{R+1}\Subset\Omega_{\mathrm{tr}}$. By (A.1),

$$1+\max_{1\leq q_{\mathrm{der}}\leq s}\left(\|D^{q_{\mathrm{der}}}H\|_\infty+\left\|D^{q_{\mathrm{der}}}H^{-1}\right\|_\infty\right)\leq\big(8C_{\mathrm{seg},N,s}(1+R)\big)^{C_{\mathrm{seg},N,s}}.$$

Since here $d_{\mathrm{cfg}}=1$, $1+K=2$, and the above choice of $C_{\mathrm{tr},N,s}$ satisfies $C_{\mathrm{tr},N,s}\geq 12C_{\mathrm{seg},N,s}$, this is bounded by the right-hand side of (3.2). Thus (3.2) also holds for $K=1$, completing the proof. ☐

A.2 Quantitative output for the transported reference field

Assume the normalized configuration satisfies

$$\mathcal{V}\subset\overline{\mathrm{B}}_{1/2},\qquad \mathrm{sep}(\mathcal{V})=\rho,\qquad 0<\rho\leq 1.$$

Put $R_{\mathrm{vert}}:=4K+1$, $d_{\mathrm{tr}}:=\min\{1,4/K,\rho\}$, and $R_{\mathrm{tr}}:=8K+6$. The standard vertices lie in $\mathrm{B}_{R_{\mathrm{vert}}}$ and have separation at least $4/K$. Applying Lemma 3.2 with $R=R_{\mathrm{vert}}$ and derivative order s_N+3 gives the diffeomorphism appearing in (3.3), supported in $\mathrm{B}_{R_{\mathrm{tr}}}$, together with

$$\mathfrak{H}:=1+\max_{1\leq j\leq s_N+3}\left(\left\|D^jH\right\|_\infty+\left\|D^jH^{-1}\right\|_\infty\right),$$
$$\mathfrak{H}\leq\left[C_{\mathrm{tr},N,s_N+3}(1+K)(1+R_{\mathrm{vert}})d_{\mathrm{tr}}^{-1}\right]^{C_{\mathrm{tr},N,s_N+3}K}.$$

Since $d_{\mathrm{tr}}^{-1}\leq K/\rho$, this envelope is bounded by $(C_NK/\rho)^{C_NK}$ after enlarging a dimension-only constant.

The pushforward identity

$$f(x)=DH(H^{-1}x)\,f_0(H^{-1}x)$$

fits Lemma A.2(ii). Since s_N depends only on N and the envelope $\mathfrak{H}$ controls more than the required s_N+1 derivatives, choose a dimension-only constant $C_{\mathrm{pc},N}\geq C_{\mathrm{comp},N,s_N}+1$ large enough to absorb the fixed product–composition constants below. Then

$$M_f:=C_{\mathrm{pc},N}(1+\|f_0\|_{C^{s_N}})\mathfrak{H}^{C_{\mathrm{pc},N}},\qquad \|f\|_{C^{s_N}}\leq M_f.$$

The compact support of $H-\mathrm{Id}$ also shows that H maps its support ball onto itself and that all reference trajectories remain in $\mathrm{B}_{R_{\mathrm{tr}}}$.

We next record the regularity estimates for the private tubes and endpoint-frame directions from Section 3.2.

A.3 Private-tube geometry

Take $\varepsilon_{\mathrm{tube}} := 1/(100K)$. For $|z| < 2\varepsilon_{\mathrm{tube}}$, one has $3/4 < 1 + z_1 < 5/4$, while the angular interval swept by I_{tube} has length at most $\pi/2$. The longitudinal derivative of Θ_i^0 has size comparable to $\omega_{\nu(i)} \asymp n_{\nu(i)}^{-1}$, and the transverse derivatives form an orthogonal frame up to uniformly bounded factors. Hence Θ_i^0 is injective on $I_{\mathrm{tube}} \times \mathrm{B}_{2\varepsilon_{\mathrm{tube}}}^{N-1}$, its derivative has singular values bounded below by $1/K$, and it is a smooth diffeomorphism onto O_i^0.

For $t \in \overline{I}_{\mathrm{tube}}$, the centerline is at distance at least $2\sin(\pi/(4n_{\nu(i)})) \geq 1/K$ from every different constrained arc on the same cycle; arcs in distinct components are separated by at least two. Since the tube radius is less than $1/(50K)$, this proves the privacy property (3.5). Repeated differentiation of Θ_i^0 and its inverse yields, for every fixed integer $s \geq 1$,

$$1 + \left\|\Theta_i^0\right\|_{C^{s+1}} + \left\|(\Theta_i^0)^{-1}\right\|_{C^{s+1}(O_i^0)} \leq C_{\mathrm{tube},N,s} K^{2s+2}. \tag{A.3}$$

The power of K records the only small singular value, namely the longitudinal one.

A.4 Smooth zero extension and endpoint-correction norms

For completeness, one may take

$$\vartheta(t) := \begin{cases} \dfrac{\exp\left(-\frac{1}{1-100(t-1/2)^2}\right)}{\displaystyle\int_{2/5}^{3/5} \exp\left(-\frac{1}{1-100(s-1/2)^2}\right) \mathrm{d}s}, & |t-1/2| < 1/10, \\ 0, & |t-1/2| \geq 1/10, \end{cases}$$

and

$$\zeta_{\mathrm{fr}}(z) := \begin{cases} \exp\bigl(1 - 1/(1-|z|^2)\bigr), & |z| < 1, \\ 0, & |z| \geq 1, \end{cases} \qquad \zeta_i(z) := \zeta_{\mathrm{fr}}(z/\varepsilon_{\mathrm{tube}}).$$

Their supports are compactly contained in the coordinate cylinder, so the fields $k_{i,a}^0$ defined in Section 3.2 extend smoothly by zero. Derivatives of ζ_i through order s cost at most a fixed multiple of $(100K)^s$; derivatives of $U_i^0(1,t)^{-1}$ are uniformly bounded; and composition with $(\Theta_i^0)^{-1}$ is controlled by (A.3). Thus

$$\max_{i,a} \left\|k_{i,a}^0\right\|_{C^s} \leq C_{\mathrm{fr},N,s} K^{2s^2+4s+2}.$$

For a self-loop, the fixed bump construction has the same conclusion with a dimension-only bound.

For the transported directions in (3.7), the preconditioning matrix $DH(y_i^0)^{-1}$ is bounded by $\mathfrak{H}$. Thus the C^{s_N} norm of the standard field inside the pushforward is bounded by a fixed multiple of $\mathfrak{H}\, C_{\mathrm{fr},N,s_N} K^{2s_N^2+4s_N+2}$. Applying Lemma A.2(ii) and enlarging $C_{\mathrm{pc},N}$ to absorb this additional factor gives

$$M_h := C_{\mathrm{pc},N} C_{\mathrm{fr},N,s_N} K^{2s_N^2+4s_N+2} \mathfrak{H}^{C_{\mathrm{pc},N}}, \qquad \max_{1\leq \ell \leq m} \|h_\ell\|_{C^{s_N}} \leq M_h.$$

No additional endpoint argument is needed, since the identity (3.9) was already established in the body of the paper.

B Coefficient-controlled shallow-ReLU approximation

This appendix contains the only approximation-theoretic input specific to the shallow ReLU dictionary. It proves Lemma 4.2 with simultaneous control of uniform error, width, and coefficient strength. No endpoint or degree argument is used here.

B.1 Coefficient-controlled shallow-ReLU approximation

Let $R > 0$ and $u \in C(\overline{\mathrm{B}}_R)$. Define the normalized product-dictionary ReLU variation

$$\gamma_R(u) := \inf\left\{ \|\lambda\|_{\mathrm{TV}} : u(x) = \int_{\mathbb{S}^{N-1}\times[-1,1]} \left(v^\top \frac{x}{R} + \beta_{\mathrm{rid}}\right)_+ \mathrm{d}\lambda(v, \beta_{\mathrm{rid}}),\ \forall x \in \overline{\mathrm{B}}_R \right\}, \tag{B.1}$$

where λ ranges over finite signed Radon measures. If no representation exists, set $\gamma_R(u) = +\infty$. We shall also use the joint-sphere dictionary $(a^\top z + b)_+$, $(a, b) \in \mathbb{S}^N$, employed by the smoothness embedding below. The product variation in (B.1) is no larger than the corresponding joint-sphere variation, while the latter is at most $\sqrt{2}\,\gamma_R(u)$.

Indeed, decompose the joint sphere into $a \neq 0, |b| \le |a|$, $a \neq 0, b > |a|$, $a \neq 0, b < -|a|$, and $a = 0$. On the first set, $(a^\top z + b)_+ = |a|\big((a/|a|)^\top z + b/|a|\big)_+$. On $b > |a|$ the atom is affine on $\overline{\mathrm{B}}_1$ and, with $v = a/|a|$,

$$(a^\top z + b)_+ = \frac{|a| + b}{2}(v^\top z + 1)_+ + \frac{b - |a|}{2}(-v^\top z + 1)_+.$$

If $a = 0, b = 1$, fix any $v \in \mathbb{S}^{N-1}$ and use $1 = \frac{1}{2}(v^\top z + 1)_+ + \frac{1}{2}(-v^\top z + 1)_+$, whereas $a = 0, b = -1$ and $b < -|a|$ give the zero atom. These transformations do not increase total variation. Conversely, for $|v| = 1$ and $|\beta_{\mathrm{rid}}| \le 1$, $(v^\top z + \beta_{\mathrm{rid}})_+ = \sqrt{1 + \beta_{\mathrm{rid}}^2}\big((v/\sqrt{1+\beta_{\mathrm{rid}}^2})^\top z + \beta_{\mathrm{rid}}/\sqrt{1+\beta_{\mathrm{rid}}^2}\big)_+$, whose cost is at most $\sqrt{2}$. This proves the stated comparison.

The next two lemmas record the precise Siegel and Yang–Zhou specializations used below; the scaling is $z = x/R$, with the preceding joint-sphere/product-dictionary conversion.

Lemma B.1 (Coefficient-controlled uniform approximation)**.** *There is a finite $A_N > 0$, depending only on N, such that for every $R > 0$, $u \in C(\overline{\mathrm{B}}_R)$, strict variation budget $\mathcal{B} > \gamma_R(u)$, and integer $n_{\mathrm{rid}} \ge 1$, there is an n_{rid}-term scalar ReLU network*

$$u_{n_{\mathrm{rid}}}(x) = \sum_{j=1}^{n_{\mathrm{rid}}} c_j \left(v_j^\top \frac{x}{R} + \beta_j^{\mathrm{rid}}\right)_+$$

satisfying

$$\sum_{j=1}^{n_{\mathrm{rid}}} |c_j| \le \mathcal{B}, \quad \mathrm{Lip}(u_{n_{\mathrm{rid}}}) \le \mathcal{B}/R, \quad \sum_{j=1}^{n_{\mathrm{rid}}} |c_j| \sqrt{R^{-2} + (\beta_j^{\mathrm{rid}})^2} \le \sqrt{1 + R^{-2}}\,\mathcal{B}. \tag{B.2}$$

$$\|u - u_{n_{\mathrm{rid}}}\|_{C(\overline{\mathrm{B}}_R)} \le A_N \mathcal{B} n_{\mathrm{rid}}^{-1/\kappa_N}, \tag{B.3}$$

Here κ_N is defined in (4.1)*.*

Proof of Lemma B.1. Theorem 3 of Siegel [26], specialized to $k = 1$, derivative order zero, and ambient dimension N, gives the rate $n_{\mathrm{rid}}^{-1/2-3/(2N)} = n_{\mathrm{rid}}^{-1/\kappa_N}$ on the unit ball and, in the notation of that result, chooses the approximant in $\Sigma^1_{n_{\mathrm{rid}}}$, so that the sum of the absolute coefficients is at most one. Apply it to $u(R\,\cdot)/\mathcal{B}$, whose product-dictionary variation is strictly smaller than one. Multiplication by $\mathcal{B}$ and the change of variables $z = x/R$ give (B.3) and the coefficient bound. Finally, every normalized atom is R^{-1}-Lipschitz in the physical variable, so $\mathrm{Lip}(u_{n_{\mathrm{rid}}}) \le R^{-1}\sum_{j=1}^{n_{\mathrm{rid}}} |c_j| \le \mathcal{B}/R$. The augmented norm of its inner parameter $(v_j/R, \beta_j^{\mathrm{rid}})$ is at most $\sqrt{R^{-2}+1}$, which proves (B.2). Zero coefficients pad the representation to exactly n_{rid} terms, while $\mathcal{B} > \gamma_R(u)$ avoids assuming attainment of the variation infimum. Continuity identifies the source L^∞ error with the uniform error on the closed ball, so the pullback gives the stated norm on $\overline{\mathrm{B}}_R$. □

For an integer $s \ge 1$, define

$$\|u\|_{C^s_R,\mathrm{ext}} := \inf\left\{\|E\|_{C^s(\mathbb{R}^N)} : E(z) = u(Rz) \text{ for } |z| \le 1\right\}. \tag{B.4}$$

Lemma B.2 (Smoothness-to-variation embedding)**.** *If the integer $s > (N+3)/2$, then there exists $C_{\mathrm{emb},N,s} < \infty$, depending only on N, s, such that*

$$\gamma_R(u) \le C_{\mathrm{emb},N,s}\,\|u\|_{C^s_R,\mathrm{ext}} \tag{B.5}$$

whenever the right-hand side is finite.

Proof of Lemma B.2. Theorem 2.1 of Yang–Zhou [27], with $k = 1$, $d = N$, and smoothness index s, applies because $s > (N+3)/2$. In the notation of that theorem it bounds the joint-sphere variation of $z \mapsto u(Rz)$ by a constant depending only on N, s times its global-extension Hölder norm. At the integer index s, their convention writes $s = (s-1)+1$, so this is a $C^{s-1,1}$ norm. It is controlled by the C^s norm in (B.4), by the multivariate mean-value theorem. The joint-to-product conversion established above does not increase total variation. Taking the infimum over extensions proves (B.5). Because both sides are continuous, the almost-everywhere identity in the cited function-space result holds pointwise on $\overline{\mathrm{B}}_1$. □

Proof of Lemma 4.2. If $u = 0$, take the empty network. Suppose $u \neq 0$, and write $u = (u^{(1)}, \ldots, u^{(N)})$. For each coordinate, the global function $z \mapsto u^{(k)}(Rz)$ is an admissible extension in (B.4), with C^s-norm at most $(1+R)^s \|u\|_{C^s}$. Lemma B.2 therefore gives $\gamma_R(u^{(k)}) \le C_{\mathrm{emb},N,s}(1+R)^s\|u\|_{C^s} = \frac{1}{2}\mathcal{B}_{R,s}(u)$. Thus $\mathcal{B}_{R,s}(u)$ is a strict variation budget for every nonzero coordinate. The definition of $n_{\mathrm{app}}(R, s; u, \varepsilon_{\mathrm{app}})$ gives $A_N \mathcal{B}_{R,s}(u) n_{\mathrm{app}}(R, s; u, \varepsilon_{\mathrm{app}})^{-1/\kappa_N} \le \varepsilon_{\mathrm{app}}/\sqrt{N}$. Apply Lemma B.1 to each nonzero scalar coordinate with this common budget and term count; use the zero scalar network for zero coordinates. Concatenating the coordinate networks gives V_{app}. The Euclidean error is at most $\sqrt{N(\varepsilon_{\mathrm{app}}/\sqrt{N})^2} = \varepsilon_{\mathrm{app}}$, and the total number of units is at most $Nn_{\mathrm{app}}(R, s; u, \varepsilon_{\mathrm{app}}) = \mathcal{P}_{R,s}(u, \varepsilon_{\mathrm{app}})$. Finally, (B.2) gives $\mathrm{Str}(V_{\mathrm{app}}) \le N\sqrt{1 + R^{-2}}\,\mathcal{B}_{R,s}(u) = \mathcal{S}_{R,s}(u)$, which proves all three assertions. $\square$

C Quantitative estimates for the endpoint-correction bridge

This appendix proves Proposition 4.3 by quantifying the approximation region and tolerances in terms of the normalized data geometry and $\mathfrak{B}(\mathbf{g})$; the degree step is invoked from Section 3.

For $\tau, L \ge 0$, write

$$\mathfrak{g}_\tau(L) := \int_0^\tau e^{L(\tau - t)}\,\mathrm{d}t = \begin{cases} (e^{L\tau} - 1)/L, & L > 0, \\ \tau, & L = 0. \end{cases}$$

Also set $L_{\mathrm{std}} := \|Df_0\|_{C^0}$. All constants involving only f_0, its fixed cutoff template, and the dimension are treated as standard-model constants.

C.1 Endpoint-correction conditioning

Let $f, h_1, \ldots, h_m$ be the reference objects from Propositions 3.3 and 3.4. Along the marked standard trajectories, the cutoff is constant and the linearized moving dynamics are rotations; the stationary components have identity fundamental matrices. Thus the standard fundamental matrices have operator norm one. The conjugacy formula (3.8) gives

$$M_U := \max_{1 \le i \le M} \int_0^1 \|U_i(1, t)\|_{\mathrm{op}}\,\mathrm{d}t \le \|DH\|_\infty \left\|DH^{-1}\right\|_\infty .$$

Choose

$$\varepsilon_* := \frac{1}{8\sqrt{mM}\max\{1, M_U\}}. \tag{C.1}$$

Suppose that $g_1, \ldots, g_m$ satisfy the first condition in (4.8). Since every reference trajectory lies in $\mathrm{B}_{R_{\mathrm{tr}}}$, the definition of $\mathcal{L}_f$ gives, column by column,

$$\|\mathcal{L}_f(g_\ell) - \mathcal{L}_f(h_\ell)\|_{\mathbb{R}^m} \le \sqrt{M}\,M_U\varepsilon_* \le \frac{1}{8\sqrt{m}}.$$

Hence, for

$$\mathsf{J}_{\mathrm{end}} := [\mathcal{L}_f(g_1)\ \cdots\ \mathcal{L}_f(g_m)],$$

the Frobenius bound and (3.9) imply

$$\|\mathsf{J}_{\mathrm{end}} - I_m\|_{\mathrm{op}} \le \frac{1}{8}.$$

Consequently we may fix

$$\sigma_* := \frac{3}{4}, \qquad |\mathsf{J}_{\mathrm{end}}\alpha| \ge \sigma_*|\alpha| \quad (\alpha \in \mathbb{R}^m).$$

C.2 Correction-family normalization and quadratic endpoint expansion

Write

$$a(z) := (DH(z))^{-1}, \qquad H_1 := \|DH\|_\infty, \qquad H_{-1} := \|a\|_\infty, \qquad H_2 := \left\|D^2H\right\|_\infty .$$

For $\widetilde{g}_\ell(z) := a(z)g_\ell(H(z))$, put $\widetilde{G}_\alpha := \sum_{\ell=1}^m \alpha_\ell\widetilde{g}_\ell$. Since $H - \mathrm{Id}$ is supported in $\mathrm{B}_{R_{\mathrm{tr}}}$, the map H sends that ball onto itself. Moreover,

$$Da(z)[v] = -a(z)D^2H(z)[v]a(z).$$

The Lipschitz estimate is global even though the fields g_ℓ need not be bounded on $\mathbb{R}^N$. Inside $\mathrm{B}_{R_{\mathrm{tr}}}$, the identity $H(\mathrm{B}_{R_{\mathrm{tr}}}) = \mathrm{B}_{R_{\mathrm{tr}}}$ and linear growth bound $g_\ell(H(z))$. Outside that ball, H and a are the identity and $Da = 0$. Thus the term involving Da occurs only where the correction field is bounded. Product and chain rules, interpreted almost everywhere for Lipschitz correction fields, give the global Lipschitz estimate. If $\mathfrak{B}(\mathbf{g}) \le \mathfrak{b}$, then the linear-growth estimate $|g_\ell(x)| \le |g_\ell(0)| + \mathrm{Lip}(g_\ell)|x|$, Cauchy–Schwarz, and the preceding product rule give the aggregate bounds

$$\begin{aligned} A_{\mathfrak{b}} &:= (1+R_{\mathrm{tr}})H_{-1}\mathfrak{b}, \\ L_{\mathfrak{b}} &:= \Big[(1+R_{\mathrm{tr}})H_{-1}^2 H_2 + H_{-1}H_1\Big]\mathfrak{b}, \end{aligned} \tag{C.2}$$

so that

$$\left(\sum_{\ell=1}^m |\widetilde{g}_\ell(0)|^2\right)^{1/2} \le A_{\mathfrak{b}}, \qquad \left(\sum_{\ell=1}^m \mathrm{Lip}(\widetilde{g}_\ell)^2\right)^{1/2} \le L_{\mathfrak{b}}.$$

Choose

$$r_0 := \frac{1}{4(1+A_{\mathfrak{b}}+L_{\mathfrak{b}})}. \tag{C.3}$$

For $|\alpha| \le r_0$, Cauchy–Schwarz gives

$$|\widetilde{G}_\alpha(0)| \le \frac{1}{4}, \qquad \mathrm{Lip}(\widetilde{G}_\alpha) \le \frac{1}{4}. \tag{C.4}$$

The pullback of $f + G_\alpha$, where $G_\alpha := \sum_\ell \alpha_\ell g_\ell$, is exactly $f_0 + \widetilde{G}_\alpha$.

For $|\alpha| \le r_0$, let

$$Z_i^\alpha(t) := H^{-1}(\Phi^t_{f+G_\alpha}(x_i)), \qquad \gamma_i^0(t) := \Phi^t_{f_0}(x_i^0).$$

Thus

$$\dot{Z}_i^\alpha = f_0(Z_i^\alpha) + \widetilde{G}_\alpha(Z_i^\alpha), \qquad Z_i^\alpha(0) = x_i^0. \tag{C.5}$$

The linear-growth estimate from (C.4) and Gronwall's inequality show that all these trajectories remain in $\overline{\mathrm{B}}_{R_{\mathrm{par}}}$, where

$$R_{\mathrm{par}} := e^{1/4}R_{\mathrm{vert}} + (\|f_0\|_{C^0} + 1/4)\mathfrak{g}_1(1/4).$$

Define

$$\begin{gathered} E_{\mathrm{disp}} := (A_{\mathfrak{b}} + R_{\mathrm{par}}L_{\mathfrak{b}})\mathfrak{g}_1(L_{\mathrm{std}} + 1/4), \\ E_{\mathrm{rem}} := \mathfrak{g}_1(L_{\mathrm{std}})\left(\frac{1}{2}\left\|D^2 f_0\right\|_{C^0} E_{\mathrm{disp}}^2 + L_{\mathfrak{b}}E_{\mathrm{disp}}\right). \end{gathered}$$

Set $D_i^\alpha := Z_i^\alpha - \gamma_i^0$. Subtracting the integral equations gives

$$\|D_i^\alpha\|_{C([0,1])} \le E_{\mathrm{disp}}|\alpha|.$$

Let z_i^α solve

$$\dot{z}_i^\alpha = Df_0(\gamma_i^0)z_i^\alpha + \widetilde{G}_\alpha(\gamma_i^0), \qquad z_i^\alpha(0) = 0,$$

and set $R_i^\alpha := D_i^\alpha - z_i^\alpha$. Taylor's formula and the Lipschitz bound on $\widetilde{G}_\alpha$ give

$$\|R_i^\alpha\|_{C([0,1])} \le E_{\mathrm{rem}}|\alpha|^2.$$

At $t = 1$, let $\widetilde{\mathsf{J}}_{\mathrm{end}}\alpha$ denote the stack of the vectors $z_i^\alpha(1)$. The fundamental-matrix conjugacy yields

$$\mathsf{J}_{\mathrm{end}} = \mathrm{diag}\big(DH(y_1^0), \dots, DH(y_M^0)\big)\widetilde{\mathsf{J}}_{\mathrm{end}}.$$

Since $\mathcal{Q}(\alpha) := \mathcal{E}(f + G_\alpha)$ is obtained by applying H to the stacked standard endpoints, Taylor's formula gives

$$\|\mathcal{Q}(\alpha) - Y - \mathsf{J}_{\mathrm{end}}\alpha\|_{\mathbb{R}^m} \le E_{\mathrm{quad}}|\alpha|^2, \qquad |\alpha| \le r_0,$$

where

$$E_{\mathrm{quad}} := \sqrt{M}\left(H_1 E_{\mathrm{rem}} + \frac{1}{2}H_2 E_{\mathrm{disp}}^2\right).$$

Set

$$r_* := \begin{cases} r_0/2, & E_{\mathrm{quad}} = 0, \\ \min\left\{r_0/2, \dfrac{\sigma_*}{8E_{\mathrm{quad}}}\right\}, & E_{\mathrm{quad}} > 0. \end{cases} \tag{C.6}$$

Then $r_* \le 1/8$, and

$$\|\mathcal{Q}(\alpha) - Y - \mathsf{J}_{\mathrm{end}}\alpha\|_{\mathbb{R}^m} \le \frac{\sigma_*}{8}|\alpha|, \qquad |\alpha| \le r_*. \tag{C.7}$$

C.3 Active approximation radius and endpoint stability

Define

$$R_{\text{safe}} := R_{\text{par}} + 2, \qquad R_* := 1 + \max\{R_{\text{tr}}, R_{\text{safe}}\}. \tag{C.8}$$

Since $H(\mathrm{B}_{R_{\text{tr}}}) = \mathrm{B}_{R_{\text{tr}}}$ and H is the identity outside that ball, $H(\overline{\mathrm{B}}_{R_{\text{safe}}}) \subset \mathrm{B}_{R_*}$. Every trajectory of $f_0 + \widetilde{G}_\alpha$, $|\alpha| \le r_*$, lies in $\overline{\mathrm{B}}_{R_{\text{par}}}$ and hence has distance at least two from $\partial \mathrm{B}_{R_{\text{safe}}}$.

Fix $0 < \eta \le 1$, and suppose a locally Lipschitz field V_{base} satisfies

$$\|V_{\text{base}} - f\|_{C(\overline{\mathrm{B}}_{R_*})} \le \eta.$$

For $|\alpha| \le r_*$, define

$$\widehat{Z}_i^\alpha(t) := H^{-1}(\Phi^t_{V_{\text{base}}+G_\alpha}(x_i)).$$

As long as $\widehat{Z}_i^\alpha$ remains in $\mathrm{B}_{R_{\text{safe}}}$, its standard-coordinate equation differs from (C.5) by at most $H_{-1}\eta$. Put

$$L_{\text{stab}} := L_{\text{std}} + r_* L_{\mathfrak{b}}, \qquad E_{\text{exit}} := H_{-1}\mathfrak{g}_1(L_{\text{stab}}), \qquad E_{\text{end}} := \sqrt{M}\, H_1 E_{\text{exit}}.$$

Because $r_* \le r_0/2$, definition (C.3) gives $r_* L_{\mathfrak{b}} \le 1/8$, and hence $L_{\text{stab}} \le L_{\text{std}} + 1/8$. Scalar Gronwall, stopped at the first exit from $\mathrm{B}_{R_{\text{safe}}}$, yields

$$\max_{1\le i\le M} \left\| \widehat{Z}_i^\alpha - Z_i^\alpha \right\|_{C([0,1])} \le E_{\text{exit}}\eta, \qquad \|\mathcal{E}(V_{\text{base}} + G_\alpha) - \mathcal{Q}(\alpha)\|_{\mathbb{R}^m} \le E_{\text{end}}\eta,$$

provided $E_{\text{exit}}\eta \le 1$. The first inequality rules out a first exit because the reference path stays at distance two from the boundary; the second follows after applying H at the endpoints and stacking the M blocks.

Choose

$$\eta_* := \frac{1}{2}\min\left\{1, \frac{1}{E_{\text{exit}}}, \frac{\sigma_* r_*}{8E_{\text{end}}}\right\}. \tag{C.9}$$

For every V_{base} satisfying the bridge hypothesis and every $|\alpha| \le r_*$, the stopped-trajectory argument gives

$$\|\mathcal{E}(V_{\text{base}} + G_\alpha) - \mathcal{Q}(\alpha)\|_{\mathbb{R}^m} \le \frac{\sigma_* r_*}{16}. \tag{C.10}$$

The same estimates imply continuous dependence of the endpoint map on α on $\overline{\mathrm{B}}^m_{r_*}$. In particular, all sample trajectories in both homotopies remain in B_{R_*}, which proves the active-radius assertion in Proposition 4.3.

C.4 Degree closure and quantitative dependence

On $|\alpha| = r_*$, equations (C.7) and (C.10) imply

$$\|\mathcal{E}(V_{\text{base}} + G_\alpha) - Y - \mathsf{J}_{\text{end}}\alpha\|_{\mathbb{R}^m} \le \frac{3\sigma_* r_*}{16} < |\mathsf{J}_{\text{end}}\alpha|.$$

Lemma 3.6 therefore gives $\alpha_* \in \mathrm{B}^m_{r_*}$ satisfying (4.9).

For the bridge envelope, Appendix A bounds

$$R_{\text{tr}},\ \mathfrak{H},\ \|f\|_{C^{s_N}},\ \max_\ell \|h_\ell\|_{C^{s_N}},\ M_U$$

by $(C_N K/\rho)^{C_N K}$; hence (C.1) has the same bound for ε_*^{-1}, while (C.8) gives the geometry-only bound for R_*. Moreover, (C.2) gives

$$A_{\mathfrak{b}} + L_{\mathfrak{b}} \le \left(C_N \frac{K}{\rho}\right)^{C_N K} \mathfrak{b}.$$

The definitions of E_{disp}, E_{rem}, and E_{quad} then show, after another enlargement of the same dimension-only constant, that

$$E_{\text{disp}} \le \Gamma_N(K,\rho)(1+\mathfrak{b}), \qquad E_{\text{rem}} + E_{\text{quad}} \le \Gamma_N(K,\rho)(1+\mathfrak{b})^2.$$

Therefore (C.3) and (C.6) give

$$r_*^{-1} \le \Gamma_N(K,\rho)(1+\mathfrak{b})^2.$$

Since $r_* L_{\mathfrak{b}} \le 1/8$, the Gronwall factor is independent of $\mathfrak{b}$, so $E_{\text{exit}} + E_{\text{end}} \le \Gamma_N(K,\rho)$; thus (C.9) gives

$$\eta_*^{-1} \le \Gamma_N(K,\rho)(1+\mathfrak{b})^2.$$

After one final enlargement of C_N, these estimates are precisely (4.6) and (4.7). This completes the proof of Proposition 4.3.

D Closure of quantitative bounds and coordinate pullback

This appendix closes the bridge and ReLU approximation budgets under the envelope of Theorem 2.10.

Closure of the quantitative envelope.

Lemma D.1 (Closure of the coarse envelope)**.** *Fix N, let $K \geq 2$, $0 < \rho \leq 1$, and put $X := K/\rho$. Call a nonnegative quantity q coarsely admissible if $q \leq (c_N X)^{c_N K}$ for some constant $c_N \geq 2$ depending only on N. A finite collection of coarsely admissible quantities remains coarsely admissible, with one common dimension-only constant, after any fixed finite sequence of the following operations:*

(i) multiplication by a dimension-only constant, finite sums, finite products, and fixed positive powers;

(ii) multiplication by fixed powers of K, M, or $m = NM$;

(iii) taking the reciprocal of a positive quantity defined as a dimension-only multiple of a finite minimum, provided the reciprocals of the entries of that minimum are already coarsely admissible;

(iv) applying $q \mapsto \lceil q^{\kappa_N} \rceil$.

The same conclusion holds for a sum or Euclidean aggregate indexed by $1 \leq \ell \leq m$, provided the summands are bounded by one common coarsely admissible quantity.

Proof. Since $M \leq K$, one has $m = NM \leq NK$, while $X = K/\rho \geq K \geq 2$. Consequently every fixed power of K, M, or m is absorbed by $(C_N X)^{C_N K}$ after increasing C_N. A fixed finite sum contributes only a fixed multiplicative factor, and a fixed product or positive power only multiplies the exponent by a fixed amount. These changes are again absorbed by increasing C_N. Moreover,

$$\left(\min_{1\leq j\leq J} q_j\right)^{-1} = \max_{1\leq j\leq J} q_j^{-1} \leq \sum_{j=1}^{J} q_j^{-1}, \qquad \lceil q^{\kappa_N} \rceil \leq 1 + q^{\kappa_N}.$$

Finally, if $0 \leq q_\ell \leq q_{\max}$ for $1 \leq \ell \leq m$, then

$$\sum_{\ell=1}^{m} q_\ell \leq m q_{\max}, \qquad \left(\sum_{\ell=1}^{m} q_\ell^2\right)^{1/2} \leq \sqrt{m}\, q_{\max}.$$

Thus every listed operation preserves the claimed form. Because only finitely many operations occur, the largest of the resulting dimension-only constants works simultaneously for all of them. ☐

Appendix A makes R_{tr}, $\|f\|_{C^{s_N}}$, and $\max_\ell \|h_\ell\|_{C^{s_N}}$ coarsely admissible. By (4.4)–(4.5),

$$\mathcal{S}_h \leq \sqrt{m} \max_\ell \mathcal{S}_{R_{\mathrm{tr}},s_N}(h_\ell)$$

is therefore coarsely admissible, and so is $\mathfrak{b}_h = \max\{1, \sqrt{2}\mathcal{S}_h\}$.

The bridge geometry estimate (4.6) makes R_* and ε_*^{-1} coarsely admissible. Since the budget enters (4.7) only through the fixed square $(1+\mathfrak{b}_h)^2$, Lemma D.1 also makes r_*^{-1} and η_*^{-1} coarsely admissible. Applying the ReLU approximation formulas at $(R_{\mathrm{tr}}, h_\ell, \varepsilon_*)$ and (R_*, f, η_*) then gives the same conclusion for $\mathcal{P}_h$, $\mathcal{P}_f$, and $\mathcal{S}_f$. Since $r_* \leq 1/8$, both final quantities $\mathcal{P}_f + \mathcal{P}_h$ and $\mathcal{S}_f + r_*\mathcal{S}_h$ are coarsely admissible. Thus, after enlarging one dimension-only constant,

$$\max\left\{R_{\mathrm{tr}}, R_*, \|f\|_{C^{s_N}}, \max_\ell \|h_\ell\|_{C^{s_N}}, \varepsilon_*^{-1}, r_*^{-1}, \eta_*^{-1}, \mathcal{P}_f + \mathcal{P}_h, \mathcal{S}_f + r_*\mathcal{S}_h\right\} \leq \left(C_N \frac{K}{\rho}\right)^{C_N K}. \tag{D.1}$$

Equations (4.12) and (D.1) prove (4.13).

D.1 Proof of the coordinate bound

Proof of Corollary 4.4. If all samples are stationary, the empty parameter list proves the claim. Otherwise $K \geq 2$. Put $R_{\mathrm{data}} := 1 + \max_{z\in\mathcal{V}} |z|$ and $S_{\mathrm{coord}}(x) := x/(2R_{\mathrm{data}})$. The normalized vertices lie in $\mathrm{B}_{1/2}$, and their separation is $\mathrm{sep}(S_{\mathrm{coord}}\mathcal{V}) = \mathrm{sep}(\mathcal{V})/(2R_{\mathrm{data}})$, so $\mathrm{sep}(S_{\mathrm{coord}}\mathcal{V})^{-1} \leq 2\Gamma_{\mathrm{coord}}(\mathcal{V})$. Apply the normalized construction summarized in (4.13) to the data $S_{\mathrm{coord}}\mathcal{D}$, and pull its parameter list back by S_{coord} and by time T, exactly as in (4.14). If $\widehat{\theta}$ is the normalized list, this scalar pullback satisfies

$$T\,\mathrm{Str}(\theta) \leq 2R_{\mathrm{data}}\,\mathrm{Str}(\widehat{\theta}). \tag{D.2}$$

Indeed, each pulled-back ridge contributes

$$|\widehat{w}_j| \sqrt{|\widehat{a}_j|^2 + (2R_{\mathrm{data}})^2 \widehat{b}_j^2} \le 2R_{\mathrm{data}} \, |\widehat{w}_j| \sqrt{|\widehat{a}_j|^2 + \widehat{b}_j^2}$$

after multiplication by T.

Let $C_{0,N}$ denote a dimension-only constant for which the normalized estimate (4.13) holds. Since $K \le 2M$ and $\mathrm{sep}(S_{\mathrm{coord}}\mathcal{V})^{-1} \le 2\Gamma_{\mathrm{coord}}(\mathcal{V})$, that estimate gives

$$p(\theta), \ \mathrm{Str}(\widehat{\theta}) \le \big(4C_{0,N} M \Gamma_{\mathrm{coord}}(\mathcal{V})\big)^{2C_{0,N}M}.$$

Choose the final constant C_N so that $C_N \ge 4C_{0,N}$ and $C_N \ge 2C_{0,N} + 1$. Since $M\Gamma_{\mathrm{coord}}(\mathcal{V}) \ge 1$, this implies

$$\big(4C_{0,N} M \Gamma_{\mathrm{coord}}(\mathcal{V})\big)^{2C_{0,N}M} \le \big(C_N M \Gamma_{\mathrm{coord}}(\mathcal{V})\big)^{C_N M - 1}.$$

Finally, $2R_{\mathrm{data}} \le 2\Gamma_{\mathrm{coord}}(\mathcal{V}) \le C_N M \Gamma_{\mathrm{coord}}(\mathcal{V})$. Combining this with (D.2) proves (4.15). Balancing now gives $\|W_\theta\|_{\mathrm{F}}^2 + \|A_\theta\|_{\mathrm{F}}^2 + \|b_\theta\|_2^2 = 2\,\mathrm{Str}(\theta)$, which proves (4.16). □

References

[1] A. Agrachev and C. Letrouit, *Generic controllability of equivariant systems and applications to particle systems and neural networks*, Ann. Inst. H. Poincaré C Anal. Non Linéaire, 43 (2026), pp. 639–668.

[2] A. Agrachev and A. Sarychev, *Control in the spaces of ensembles of points*, SIAM J. Control Optim., 58 (2020), pp. 1579–1596.

[3] ——, *Control on the manifolds of mappings with a view to the deep learning*, J. Dyn. Control Syst., 28 (2022), pp. 989–1008.

[4] A. Álvarez-López, B. Geshkovski, and D. Ruiz-Balet, *Constructive approximate transport maps with normalizing flows*, Appl. Math. Optim., 92 (2025). Article 33.

[5] A. Álvarez-López, A. Hadj Slimane, and E. Zuazua, *Interplay between depth and width for interpolation in neural ODEs*, Neural Netw., 180 (2024). Article 106640.

[6] A. Álvarez-López, L. Liverani, and E. Zuazua, *Constructive interpolation and generalization rates for neural ODEs: a control perspective*, 2026. arXiv preprint arXiv:2606.00469v2.

[7] J. Arango and A. Gómez, *Diffeomorphisms as time one maps*, Aequationes Math., 64 (2002), pp. 304–314.

[8] A. Banyaga, *The Structure of Classical Diffeomorphism Groups*, vol. 400 of Mathematics and Its Applications, Kluwer Academic Publishers, Dordrecht, 1997.

[9] J. Behrmann, W. Grathwohl, R. T. Q. Chen, D. Duvenaud, and J.-H. Jacobsen, *Invertible residual networks*, in Proceedings of the 36th International Conference on Machine Learning, vol. 97 of Proceedings of Machine Learning Research, 2019, pp. 573–582.

[10] R. T. Q. Chen, Y. Rubanova, J. Bettencourt, and D. Duvenaud, *Neural ordinary differential equations*, in Advances in Neural Information Processing Systems 31, 2018, pp. 6571–6583.

[11] G. Cybenko, *Approximation by superpositions of a sigmoidal function*, Math. Control Signals Syst., 2 (1989), pp. 303–314.

[12] N. De Nitti and X. Fernández-Real, *Optimal transport of measures via autonomous vector fields*, 2025. arXiv preprint arXiv:2405.06503v2.

[13] E. Dupont, A. Doucet, and Y. W. Teh, *Augmented neural ODEs*, in Advances in Neural Information Processing Systems 32, 2019, pp. 3140–3150.

[14] W. E, *A proposal on machine learning via dynamical systems*, Commun. Math. Stat., 5 (2017), pp. 1–11.

[15] W. Grathwohl, R. T. Q. Chen, J. Bettencourt, I. Sutskever, and D. Duvenaud, *FFJORD: Free-form continuous dynamics for scalable reversible generative models*, in International Conference on Learning Representations, 2019.

[16] E. Grong and S. Vega-Molino, *Controllability on landmark manifolds for shapes and neural ODEs*, Math. Control Signals Syst., 38 (2026), pp. 275–289.

[17] M. Gugat, *Exact ensemble controllability for neural differential equations via neural interpolation*, 2026. arXiv preprint arXiv:2607.21112v1.

[18] K. Hornik, *Approximation capabilities of multilayer feedforward networks*, Neural Netw., 4 (1991), pp. 251–257.

[19] J. M. Klusowski and A. R. Barron, *Approximation by combinations of ReLU and squared ReLU ridge functions with ℓ^1 and ℓ^0 controls*, IEEE Trans. Inf. Theory, 64 (2018), pp. 7649–7656.

[20] C.-K. Lee, F. Coeurdoux, Y. Chen, S. Tang, P. Potaptchik, Y. Du, M. S. Albergo, and E. Vanden-Eijnden, *Beckmann transport models: From autonomous flows to one-step maps*, 2026. arXiv preprint arXiv:2608.01692v3.

[21] P. W. Michor and C. Vizman, *n-transitivity of certain diffeomorphism groups*, Acta Math. Univ. Comenianae (N.S.), 63 (1994), pp. 221–225.

[22] F. G. Pacifico, D. Fanelli, L. Buffoni, L. Chicchi, D. Febbe, and R. Marino, *Exact fixed-point constraints in neural-ODEs with provable universality*, 2026. arXiv preprint arXiv:2605.10613v1.

[23] A. Pinkus, *Approximation theory of the MLP model in neural networks*, Acta Numer., 8 (1999), pp. 143–195.

[24] D. Ruiz-Balet, E. Affili, and E. Zuazua, *Interpolation and approximation via momentum ResNets and neural ODEs*, Syst. Control Lett., 162 (2022), p. 105182.

[25] D. Ruiz-Balet and E. Zuazua, *Neural ODE control for classification, approximation, and transport*, SIAM Rev., 65 (2023), pp. 735–773.

[26] J. W. Siegel, *Optimal approximation of zonoids and uniform approximation by shallow neural networks*, Constr. Approx., 62 (2025), pp. 441–469.

[27] Y. Yang and D.-X. Zhou, *Optimal rates of approximation by shallow* ReLU^k *neural networks and applications to nonparametric regression*, Constr. Approx., 62 (2025), pp. 329–360.