
Patching Non-Euclidean Domains for Neural PDE Solvers

Aleksa Milovanović
Faculty of Sciences
University of Novi Sad
aleksa.milovanovic@dmi.uns.ac.rs

Abstract

Localising a neural field by splitting its domain into patches is standard in flat domains and underlies both finite basis solvers and patch based tokenisation, but a curved surface carries no grid, so the patch family must itself be constructed. We show that overlapping decompositions of the ambient space restrict to a partition of unity on any orientable surface at no cost in generality, and that patches defined by geodesic distance can be used even though the PDE residual differentiates the windows twice while geodesic distance admits no closed form. On a smooth torus patching brings no benefit. On three meshes, with up to ten seeds each, the gain of geodesic patches over a tuned global network grows with how far ambient proximity departs from intrinsic proximity, from a factor of 1.7 to 4.6 on a hand scan. Geodesic patches improve on ambient boxes on every mesh tested, by factors of 1.4 to 2.1 in the mean with $p < 0.001$ under a rank test, and at low distortion the whole gain comes from going intrinsic, since ambient boxes alone do not beat a tuned global network. On a family of progressively squashed tori the advantage of geodesic patches over ambient boxes grows from parity to a factor of 5.5.

1 Introduction

Splitting a domain into overlapping patches and giving each its own small network is a standard way to localise a neural field. In flat domains it underlies finite basis physics informed neural networks [Moseley et al., 2023], where per patch input and output normalisation turn a globally hard target into many locally easy ones, and it shares its roots with the partition of unity method [Melenk and Babuška, 1996], extended PINNs [Jagtap and Karniadakis, 2020] and multilevel decompositions [Dolean et al., 2024]. The same move appears wherever continuous physical data are cut into tokens.

On a curved surface the move is not available, because the object it starts from does not exist. A regular grid, the natural patch family in a box, has no counterpart on a surface of unknown topology, and the intrinsic candidates are not differentiable in the way a PDE residual requires. This matters in practice because neural solvers on surfaces are now formulated in a way that makes patching otherwise straightforward. PINNsur [Jain et al., 2026] writes the surface operators of a physics informed network [Raissi et al., 2019] as projections of Euclidean ones along the unit normal $\mathbf{n}$,

$$\nabla_{\Omega} u = \nabla u - (\nabla u \cdot \mathbf{n}) \mathbf{n}, \quad \Delta_{\Omega} u = \nabla \cdot (\nabla_{\Omega} u) - \mathbf{n}^{\top} J(\nabla_{\Omega} u) \mathbf{n}, \quad (1)$$

where J is the Jacobian, so the solution network consumes ambient coordinates and any orientable topology is admissible. Classical surface finite elements [Dziuk and Elliott, 2013] need a mesh throughout, and learned models on surfaces either diffuse features intrinsically [Sharp et al., 2022] or feed intrinsic coordinates to a neural field [Koestler et al., 2022], but none of these split a neural PDE solver into patches.

We ask what a patch should be on such a surface, and what the choice costs. Our contributions are the following. Ambient decompositions restrict to a valid partition of unity on the surface, which makes patching free in generality (Section 2). Geodesic patches can be made admissible by expanding the operator with the product rule so that the windows enter the loss as data and are never differentiated (Section 3). Finally, on meshes the benefit of patching tracks a measurable property of the geometry, the discrepancy between ambient and intrinsic proximity at the patch scale, which we read as a computable proxy for the reach of the surface relative to the patch radius (Section 4).

2 Ambient patches restrict to the surface

Place a regular grid on the bounding box of Ω and give the cell centred at c_j the separable window $w_j(\mathbf{x}) = \prod_d \rho((x_d - c_{j,d})/h_d)$ with $\rho(s) = (1 - s^2)_+^3$ and $h = \frac{\gamma}{2}$ (cell) for an overlap factor $\gamma > 1$. The bump is C^2 , as the residual differentiates the windows twice. Cells holding no surface samples are discarded, so the patch count follows surface area rather than bounding box volume.

Proposition 1. *Let $\gamma > 1$ and let every point of Ω lie in a retained cell. Then $\sum_k w_k > 0$ on Ω , and the normalised weights $\hat{w}_j = w_j / \sum_k w_k$ are C^2 , nonnegative and sum to one on a neighbourhood of Ω . All three properties are preserved under restriction to Ω .*

The proof is immediate, and the content of the statement is its cost, which is nil. No chart, parametrisation or assumption on topology is needed, so patching is exactly as general as the solver it extends. The ansatz is $u(\mathbf{x}) = s \sum_j \hat{w}_j(\mathbf{x}) N_j((\mathbf{x} - c_j)/h)$ with local networks N_j and an output scale s .

3 Geodesic patches and the product rule

Ambient and intrinsic proximity provably stay within a factor of $\pi/2$ of each other where the patch is smaller than the reach of the surface [Federer, 1959, Boissonnat et al., 2019], the scale below which balls centred on the surface meet it in a single well behaved sheet [Niyogi et al., 2008]. Reach is limited both by curvature and by distinct parts of the surface coming close. On a hand scan a single box can hold two fingers (Figure 1), so one small network becomes responsible for two unrelated parts of the problem. The intrinsic remedy defines windows by distance measured along the surface, $w_j = \rho(d_g(\mathbf{x}, c_j)/h_j)$, with centres from geodesic farthest point sampling and each patch given twice the covering radius of its own geodesic Voronoi cell, so that neighbours overlap. Distances come from the heat method [Crane et al., 2017], whose systems factor once, so an extra centre costs two back substitutions.

The obstacle is that the residual needs two derivatives of d_g , which admits no closed form. We avoid it by never differentiating the windows. Expanding the operator by the product rule,

$$\Delta_\Omega \left(\sum_j \hat{w}_j N_j \right) = \sum_j \left[\hat{w}_j \Delta_\Omega N_j + 2 \nabla_\Omega \hat{w}_j \cdot \nabla_\Omega N_j + N_j \Delta_\Omega \hat{w}_j \right], \quad (2)$$

the windows appear only as values. We obtain $\hat{w}_j$ by barycentric interpolation, its tangential gradient as the per face gradient of the piecewise linear interpolant, and its Laplace–Beltrami value from $-M^{-1}S\hat{w}_j$ with the cotangent stiffness matrix S and the lumped mass matrix M , once, and supply all three to the loss as data. Automatic differentiation then touches only the local networks. The scheme is exact for the discrete windows and cheaper per step than differentiating the blended ansatz, at the cost of fixing the collocation set in advance. The discrete windows satisfy $\sum_j \hat{w}_j = 1$, $\sum_j \nabla_\Omega \hat{w}_j = 0$ and $\sum_j \Delta_\Omega \hat{w}_j = 0$ to about 10^{-7} relative to their largest entries, the round off of single precision.

4 Experiments

Setup. We solve the screened Poisson equation $\Delta_\Omega u - u = f$ on four surfaces, a smooth torus with analytic normals, the SPOT mesh, a torus squashed until its sheets nearly touch, and a hand scan of 51k vertices. On meshes the geometry enters the operator through pointwise values only, using $\Delta_\Omega u = \Delta u - \mathbf{n}^\top (\nabla^2 u) \mathbf{n} - 2H(\nabla u \cdot \mathbf{n})$ with $\mathbf{n}$ from vertex normals and twice the mean curvature $2H$ from $\Delta_\Omega \mathbf{x} = -2H\mathbf{n}$. On the hand the target is manufactured from eigenpairs of the

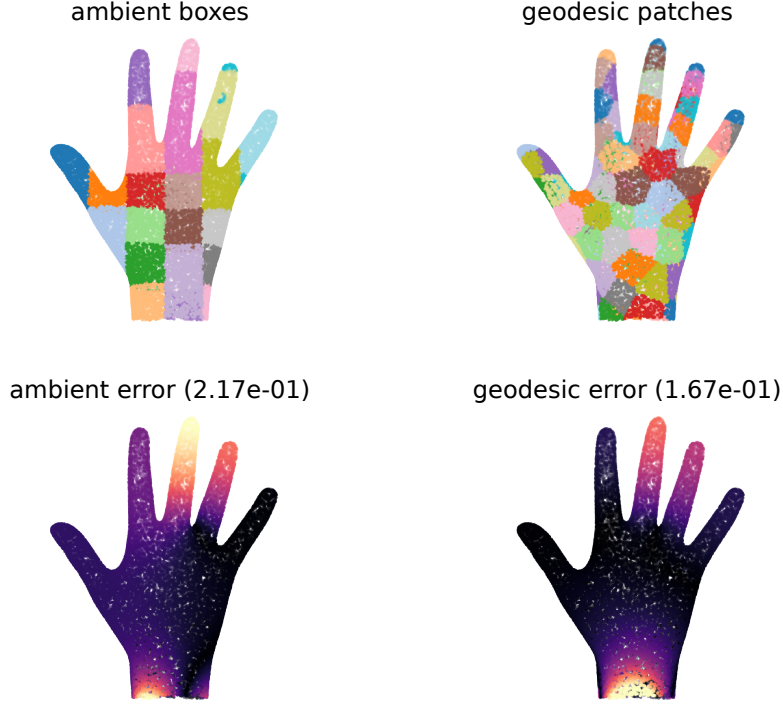


Figure 1: Patches on a hand scan, coloured by dominant window. Ambient boxes at the top left cut straight across the fingers, giving one network responsibility for parts of the surface that are far apart along it. Geodesic patches at the top right follow the surface. Below, the resulting absolute error on a shared scale, for a seed of each method whose error is close to its ten seed mean, with the relative ℓ_2 error in the panel title. Geodesic patches err most at the open wrist, where boundary values are inherited from the reference rather than imposed, and ambient boxes at the fingertips.

discrete operator following the protocol of Jain et al. [2026], on SPOT and the thin torus the reference is a finite element solve for a smooth source, and on the smooth torus the solution is analytic. All networks are SIRENs [Sitzmann et al., 2020]. Every comparison holds the schedule, the samples, the source and the target fixed. The global network is used at its best width, which carries more weights than the patched models, and the weight counts are listed with Table 1. On the meshes the learning rate is swept for *every* method independently and the best value per method is used, since we found a fixed rate to misrepresent the global baseline severely, in line with known training pathologies [Krishnapriyan et al., 2021, Wang et al., 2021]. Each mesh comparison is repeated over ten seeds, eight for the global network on two meshes, which vary the network initialisation and, for geodesic patches, the centre placement. Seed variation is appreciable, so we report means with ranges and assess differences with an exact one sided rank test.

An axis for the geometry. To ask when patching pays we need to say what makes a surface hard. Curvature magnitude alone does not order our surfaces. We instead measure how far ambient proximity departs from intrinsic proximity, by sampling source vertices, taking every vertex inside a Euclidean ball of the typical patch radius, and recording the 95th percentile of the ratio of geodesic to Euclidean distance. A locally flat surface gives approximately one. The statistic does not compete with curvature but combines it with the closeness of distinct sheets, so we treat it as a computable proxy for reach at the patch scale rather than as a quantity in its own right.

Results. Table 1 and Figure 2 give the finding. On the smooth torus patching brings nothing, a properly tuned global network matching the decomposed one. As distortion increases the gain of geodesic patches over the global network grows monotonically, through $1.7\times$ on SPOT and $1.9\times$ on the thin torus to $4.6\times$ on the hand scan, where the global network fails at every width, learning rate

surface	distortion	global	ambient boxes	geodesic patches
smooth torus	1.35	4.83×10^{-4}	4.65×10^{-4}	
SPOT	1.30	1.57×10^{-2}	1.94×10^{-2}	9.07×10^{-3}
thin torus	1.62	1.15×10^{-2}	1.25×10^{-2}	6.08×10^{-3}
hand scan	2.59	7.21×10^{-1}	2.25×10^{-1}	1.58×10^{-1}

Table 1: Relative ℓ_2 error, each method at its own best learning rate, averaged over ten seeds on the meshes and eight for the global network on SPOT and the thin torus. The smooth torus carries an analytic normal field and no mesh, so it is measured once with ambient boxes only, and both of its models have 17k weights. On the meshes the global network has 46k weights, ambient boxes 33k, 13k and 25k and geodesic patches 23k, 13k and 32k on SPOT, the thin torus and the hand. Ranges are shown in Figure 2.

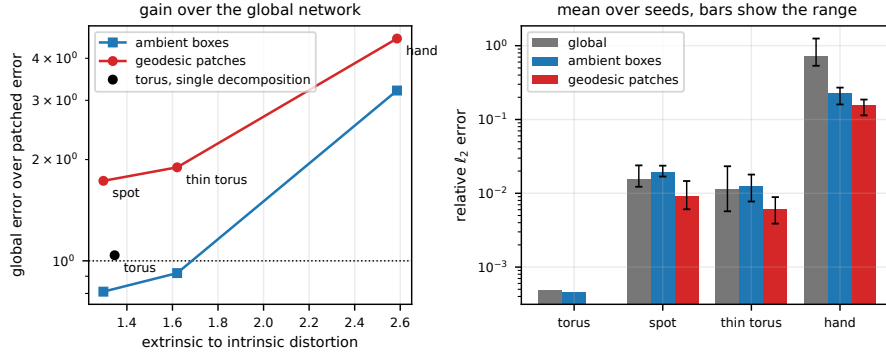


Figure 2: Left, the error of the global network divided by that of each patched model. Geodesic patches gain on every mesh, from $1.7\times$ on SPOT to $4.6\times$ on the hand, while ambient boxes are worse than the global network on SPOT and the thin torus and gain only on the hand. The smooth torus, measured with ambient boxes only, is drawn in black. Right, mean errors with the range over seeds.

and seed we tried. Ambient boxes behave differently. They are worse than the tuned global network on SPOT and the thin torus and help only on the hand, so at low distortion the whole gain comes from going intrinsic. Curvature does not order the surfaces the same way, since the thin torus carries the largest curvature but a smaller gain than the hand. The smooth torus does not fit the ordering. Its patches are wider than its tube, so at their scale it reads as distorted as SPOT, yet patching brings nothing there. Distortion thus orders the three meshes but does not separate the smooth torus from SPOT. Geodesic patches improve on ambient boxes on all three meshes, by 2.13 , 2.06 and 1.43 in the mean, with exact one sided rank test values of 5×10^{-6} , 2×10^{-5} and 2×10^{-4} , all below 10^{-3} after a Bonferroni correction for the three comparisons. On the hand the geodesic model carries a quarter more weights than the ambient one, so that margin is the least clean of the three.

A controlled family. Four surfaces give one ordering over four points, so we also vary a single geometry continuously. Figure 3 squashes one torus while keeping the topology, the source and the patch count fixed. On the round torus the two patch families tie. As the sheets approach, the error of ambient boxes grows thirtyfold and overtakes that of the global network, while the error of geodesic patches grows sixfold, so their advantage over ambient boxes rises to $5.5\times$.

5 Discussion

The construction is general and its setup is cheap, and the measurement indicates when to use it rather than giving a blanket recommendation. Several limitations are worth stating. We study one operator, screened Poisson, with smooth targets, and test neither anisotropy nor boundary layers nor time dependence. Geodesic patches require a mesh, so they do not reach the point cloud setting that the underlying solver supports, and the collocation set must be fixed in advance because the window data are tied to it. Patching is not free in time either. Every local network is evaluated at every point,

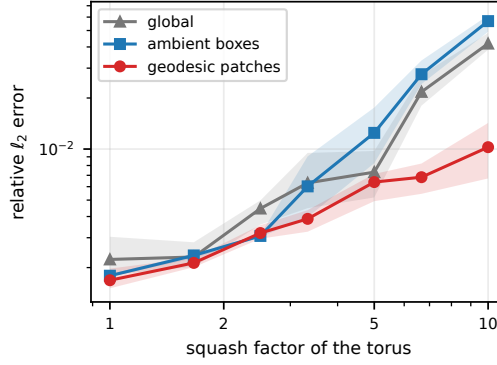


Figure 3: A torus squashed by a factor of up to ten, with one source and the same patch count for both families, each method at the learning rate selected on the meshes, three seeds per point and bands showing the range.

so a patched run takes between two and a half and nine and a half times longer than the global one, although the geodesic setup itself costs under a minute (Appendix D). The distortion statistic is a proxy whose value depends on the patch scale at which it is computed, and a surface can be hard for reasons it does not capture.

We also report what did not hold. We initially expected patching to repair the failure of a global network on solutions with large dynamic range, which is a limitation reported for the underlying solver. Once the learning rate of the baseline is scaled with width the failure disappears entirely and the two approaches tie, so that reported failure appears to be a property of the training configuration rather than of the method. The same applies to the overlap factor, which looked like a first order hyperparameter at one learning rate and did not behave monotonically at another.

Acknowledgments and Disclosure of Funding

The author gratefully acknowledges the AXIOM HPC facility and support provided by the Scientific Computing Research Group (SCORG) at the Faculty of Sciences, University of Novi Sad.

References

- J.-D. Boissonnat, A. Lieutier, and M. Wintraecken. The reach, metric distortion, geodesic convexity and the variation of tangent spaces. *Journal of Applied and Computational Topology*, 2019.
- K. Crane, C. Weischedel, and M. Wardetzky. The heat method for distance computation. *Communications of the ACM*, 2017.
- V. Dolean, A. Heinlein, S. Mishra, and B. Moseley. Multilevel domain decomposition-based architectures for physics informed neural networks. *Computer Methods in Applied Mechanics and Engineering*, 2024.
- G. Dziuk and C. M. Elliott. Finite element methods for surface PDEs. *Acta Numerica*, 2013.
- H. Federer. Curvature measures. *Transactions of the American Mathematical Society*, 1959.
- A. D. Jagtap and G. E. Karniadakis. Extended physics informed neural networks. *Communications in Computational Physics*, 2020.
- P. Jain, N. Kairanda, P. Y. Chen, and O. Stein. PINNs on curved surfaces. *arXiv:2605.27308*, 2026.
- L. Koestler, D. Grittner, M. Moeller, D. Cremers, and Z. Löhner. Intrinsic neural fields, learning functions on manifolds. In *ECCV*, 2022.

- A. S. Krishnapriyan, A. Gholami, S. Zhe, R. M. Kirby, and M. W. Mahoney. Characterizing possible failure modes in physics informed neural networks. In *NeurIPS*, 2021.
- J. M. Melenk and I. Babuška. The partition of unity finite element method, basic theory and applications. *Computer Methods in Applied Mechanics and Engineering*, 1996.
- B. Moseley, A. Markham, and T. Nissen-Meyer. Finite basis physics informed neural networks. *Advances in Computational Mathematics*, 2023.
- P. Niyogi, S. Smale, and S. Weinberger. Finding the homology of submanifolds with high confidence from random samples. *Discrete and Computational Geometry*, 2008.
- M. Raissi, P. Perdikaris, and G. E. Karniadakis. Physics informed neural networks, a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 2019.
- N. Sharp, S. Attaiki, K. Crane, and M. Ovsjanikov. DiffusionNet, discretization agnostic learning on surfaces. *ACM Transactions on Graphics*, 2022.
- V. Sitzmann, J. Martel, A. Bergman, D. Lindell, and G. Wetzstein. Implicit neural representations with periodic activation functions. In *NeurIPS*, 2020.
- S. Wang, Y. Teng, and P. Perdikaris. Understanding and mitigating gradient flow pathologies in physics informed neural networks. *SIAM Journal on Scientific Computing*, 2021.

Appendix

This appendix records the runs behind every number in the paper, the learning rate sweeps that support the claim that both sides of each comparison were tuned, the definition of the distortion statistic, the cost of each method, the effect of the number of patches and implementation details.

A Per seed results

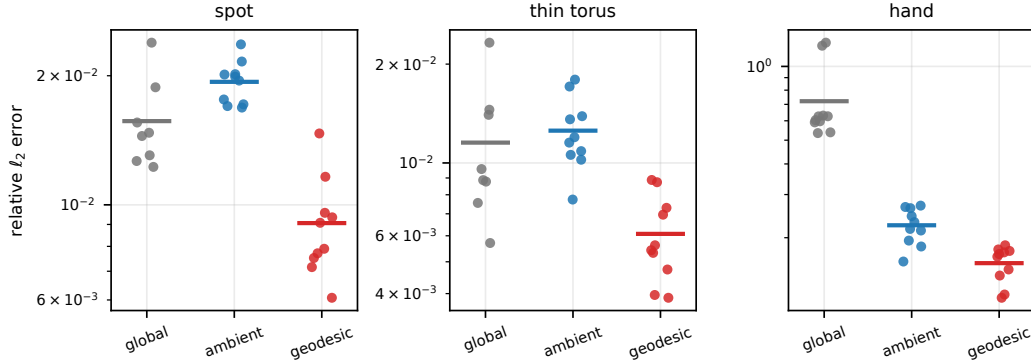


Figure 4: Every run behind Table 1. Dots are individual seeds and the bar is the mean. Seed variation within a method is appreciable, so the comparison rests on rank tests rather than on point estimates.

surface	method	runs	mean	min	max
SPOT	global	8	1.57	1.23	2.39
	ambient	10	1.94	1.69	2.37
	geodesic	10	0.91	0.61	1.47
thin torus	global	8	1.15	0.57	2.32
	ambient	10	1.25	0.77	1.79
	geodesic	10	0.61	0.39	0.89
hand scan	global	10	72.1	53.5	125.0
	ambient	10	22.5	16.0	27.0
	geodesic	10	15.8	11.4	18.6

Table 2: Relative ℓ_2 error in percent over all runs. Seeds vary the network initialisation and, for geodesic patches, the placement of centres.

Exact one sided Mann Whitney tests give p of 5×10^{-6} , 2×10^{-5} and 2×10^{-4} for geodesic patches against ambient boxes on SPOT, the thin torus and the hand, and 3×10^{-4} , 1.5×10^{-3} and 5×10^{-6} for geodesic patches against the global network. With ten runs against ten the smallest attainable value is 5×10^{-6} , reached where the two samples do not overlap. Ambient boxes beat the global network only on the hand, with $p = 5 \times 10^{-6}$, and not on SPOT or the thin torus, with p of 0.99 and 0.84.

B Learning rate sweeps

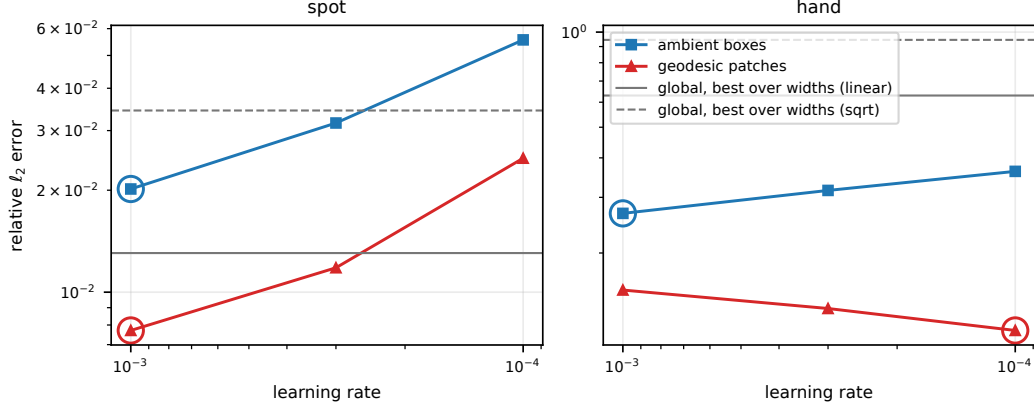


Figure 5: Learning rate is swept for each patch family independently, with the selected value circled, and the global baseline is drawn at the best value over its own sweep of widths and two scaling rules. The two families do not share an optimum, since geodesic patches on the hand prefer a rate ten times smaller than ambient boxes, which is why a single shared rate cannot be used.

Tuning both sides is not a formality. In an earlier version of this study the global baseline was measured at one fixed rate and appeared to fail outright on two problems, by factors of 57 and 121. Both apparent failures disappeared once the peak rate was scaled with width, and on one of them the comparison first inverted and then settled at parity once both sides were tuned. The numbers reported in this paper are the result of repeating every comparison with both sides swept.

C The distortion statistic

For a source vertex we compute the geodesic distance field by the heat method, take every vertex whose Euclidean distance to the source lies between $0.15h$ and h , and record the 95th percentile of the ratio of geodesic to Euclidean distance. The statistic is the mean of that quantity over forty randomly chosen sources. A locally flat surface returns approximately one. The radius h is the mean covering radius of geodesic patches placed at the patch count of the decomposition each surface uses in Table 1, 48 on the smooth torus, 94 on SPOT, 36 on the thin torus and 72 on the hand. The value depends on h . At the patch counts of the geodesic models it reads 1.31, 1.62 and 2.35 on the three meshes, so their ordering does not change.

surface	distortion	scaled curvature	geodesic gain	ambient gain
smooth torus	1.35	13.0		$1.04\times$
SPOT	1.30	46.5	$1.73\times$	$0.81\times$
thin torus	1.62	203.7	$1.90\times$	$0.92\times$
hand scan	2.59	90.1	$4.58\times$	$3.21\times$

Table 3: Gain is the error of the global network divided by that of the patched model. Distortion orders the three meshes in the same way as the gain of geodesic patches, whereas curvature, here the 95th percentile of $|2H|$ scaled by the bounding box diagonal, does not, since the thin torus carries the largest curvature and only an intermediate gain. The two quantities are not independent, as distortion at the patch scale also grows with curvature.

D Cost

surface	global	ambient	geodesic	geodesic setup
SPOT	12.9	121.8	61.8	0.1
thin torus	13.5	40.9	35.0	0.1
hand scan	13.3	96.6	85.7	0.4

Table 4: Median wall clock minutes per run for 30,000 steps on twelve CPU cores of a shared cluster with two processor generations, so individual times vary. Setup covers data preparation and, for geodesic patches, the heat method factorisation, farthest point sampling and the window data. Patched models are slower because every local network is evaluated at every point, which a sparse evaluation over the few nonzero windows would remove.

E Number of patches

patches	weights	flat, 10^{-3}	flat, $3 \cdot 10^{-4}$	multilevel, 10^{-3}	multilevel, $3 \cdot 10^{-4}$
18	6.4k	6.70	4.75	6.70	4.75
48	16.9k	11.86	4.65	2.01	3.30
130	45.9k	16.15	6.86	2.20	1.47

Table 5: Relative ℓ_2 error times 10^4 for a smooth field on the torus as ambient boxes are refined, at two learning rates. Flat refinement retrains all patches and does not improve with more of them at either rate. Multilevel refinement freezes the coarser level and trains only a correction, whose weights add to those listed, and improves at both rates.

F Implementation

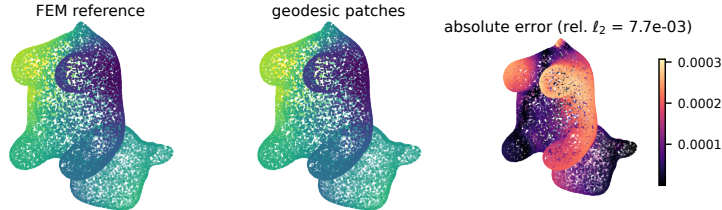


Figure 6: The geodesic solution on SPOT for one run against the finite element reference, with the absolute error on the right.

The implementation is in JAX, and all mesh operations, namely the cotangent Laplacian, area uniform sampling with element provenance, vertex normals, mean curvature, heat geodesics, geodesic farthest point sampling and the finite element reference solves, are written directly on NumPy and SciPy, so the package carries no geometry processing dependency. The projected operator reproduces $\Delta_{\Omega}(xyz) = -12xyz$ on the unit sphere to 1.3×10^{-15} and the closed form parametric expression on the torus to 7.1×10^{-14} . The curvature form matches the same reference to 7.8×10^{-14} when given the exact normal and mean curvature, and to a relative 7×10^{-4} when given the vertex normals and cotangent mean curvature of a torus mesh with 18,432 vertices. The discrete windows of equation (2) satisfy $\sum_j \hat{w}_j = 1$, $\sum_j \nabla_{\Omega} \hat{w}_j = 0$ and $\sum_j \Delta_{\Omega} \hat{w}_j = 0$ at the collocation points to about 10^{-7} relative to their largest entries.

The numbers in this paper rest on two hundred and fifteen independent training runs.