Journal of Mathematical Imaging and Vision (2026) 68:68
https://doi.org/10.1007/s10851-026-01337-2

Explainable Binary Classification of Separable Shape Ensembles

Zachary Grey[1] · **Nicholas Fisher**[2] · **Andrew Glaws**[3]

Received: 11 December 2025 / Accepted: 11 July 2026

Abstract

Scientists, engineers, biologists, and technology specialists universally leverage image segmentation to extract shape ensembles containing many thousands of curves representing patterns in observations and measurements. These large curve ensembles facilitate statistical inferences about important changes when comparing and contrasting images. We introduce novel pattern recognition formalisms combined with two-sample hypothesis testing over large ensembles of segmented curves. Our formalism involves accurately approximating eigenspaces of composite integral operators to motivate discrete, dual representations of curves collocated at quadrature nodes. Approximations are projected onto underlying matrix manifolds, and the resulting separable shape tensors constitute Euclidean invariant decompositions of curves into generalized (linear) scale variations and complementary (nonlinear) undulations. Hypothesis testing over the separable feature space can be conceived of as an explainable binary classification mapping from pairs of distributions into a binary conclusion—reject or fail to reject—offering principled decisions with logical implications. With thousands of curves segmented from pairs of images, we demonstrate how data-driven features of separable shape tensors inform explainable binary classification utilizing a product maximum mean discrepancy, absent labeled data, building interpretable feature spaces in seconds without high performance computation, and detecting discrepancies below cursory visual inspections.

Keywords Separable shape tensors · Reproducing kernel Hilbert space · Maximum mean discrepancy

Mathematics Subject Classification 53Z50 · 51F25 · 65D18 · 65D30 · 46E22

1 Introduction

Precise measurement and quantification of shape are ubiquitous in imaging science with applications spanning engineering design [1, 2], materials science [3–6], medical imaging [7, 8], functional data analysis [9, 10], biology [11–13], and ecology [14, 20]. Many of the analyses in these applications benefit from separable characterizations of shape translations (*where is it?*), scale variations (*how big is it?*), rotations/reflections (*how is it oriented?*), and everything else that remains (*how nonlinear is it?*). In several applications, a choice of invariance against one or multiple of types of similarity transforms [21]—among other transformations [22]—leads to natural definitions of *shape*.

In some cases, scientists or engineers may want to answer a simple, binary question: *Is this the same as that*? This naturally leads to the follow-up question: *If they are different, why?* We refer to solutions of this general class of problems as *explainable binary classifications*—hypothesis testing that not only renders a binary decision but also explains *why* the discrepancy exists.

Target applications span diverse imaging modalities, including electron microscopy, X-ray computed tomography, satellite and optical imagery, cell microscopy/histopathology,

Zachary Grey and Nicholas Fisher have contributed equally to this work.

✉ Zachary Grey
zachary.grey@nist.gov

Nicholas Fisher
nicholfi@pdx.edu

Andrew Glaws
andrew.glaws@nlr.gov

[1] Applied & Computational Mathematics Div., National Institute of Standards & Technology, 325 S. Broadway, Boulder, CO 80305, USA

[2] Dept. of Mathematics and Statistics, Portland State University, 1825 SW Broadway, Portland, OR 97201, USA

[3] Computational Science Center, National Laboratory of the Rockies, 15013 Denver West Parkway, Golden, CO 80401, USA

Published online: 06 September 2026

solar UV imaging, Lagrangian averaged vorticity deviations of turbulent flows, and AI-generated imagery. For example, electron backscatter diffraction (EBSD) [23] provides high-resolution segmentation of material microstructures [24, 25]. Processing these via denoising [3] and segmentation [6, 26] yields large ensembles of planar curves. Figure 1 illustrates a segmented EBSD image of deformed ice generated using MTEX [5, 6, 27], alongside an individual 'grain boundary' extracted from the ensemble.

Beyond EBSD, our interpretations are extensible to *any imaging combined with subsequent segmentation informing a large ensemble of planar curves*. Additional examples of alternative image modalities are shown in Fig. 2.

1.1 Problem Statement

> *Given two images: are the curves in one image, up to Euclidean motions, the same or different from that of another and, if they are different, are differences due to linear deformations (anisotropic scale variations) or nonlinear deformations (undulations)?*

In this work, we present an approach to answer the proposed question as a binary response augmented by *explanation through separability* and formal extensions to existing work [33, 34]. Unlike standard artificial intelligence (AI) models—which rely on subjective hand-labeled training and often utilize opaque 'learned features' within an ambiguous 'latent space'—our framework prioritizes quantifiable explanations. This avoids the common pitfalls of AI-generated responses, which, while often articulate, are prone to fabrication or hallucination.

To formally answer the proposed question in an unsupervised manner, we define a binary mapping $H : (\mathcal{E}, \widehat{\mathcal{E}}) \mapsto \{0, 1\}$, where $\mathcal{E}, \widehat{\mathcal{E}}$ are ensembles of feature-encoding random tensors. In our case, 0 rejects the null hypothesis of equality and 1 is a failure to reject (FTR). Separability of the ensemble representations facilitates a decomposition as $H(\mathcal{E}, \widehat{\mathcal{E}}) := H(\mathcal{G}_r, \widehat{\mathcal{G}}_r) \wedge H(\mathcal{P}, \widehat{\mathcal{P}})$. Thus, separate marginal outcomes can be combined to logically explain the aggregate. Stochastic features constituting ensembles make the associated decisions and explanations prone to error, but the compositional framework remains logical by definition.

Here H is a distribution-level binary classifier over pairs of ensembles, not an instance-level label predictor. An instance-level predictor can nonetheless be *built* from H via *context-driven classification*: Fixing one argument of $H(\cdot, \cdot)$ to a curated 'trusted' ensemble implicitly assigns it a label, so sweeping the other argument over discovered ensembles labels each as same kind as context or not. Fixing the trusted ensemble to curves extracted from healthy histopathology tissue, for instance, yields a healthy vs. not healthy decision on each query image—a seemingly binary oversimplification that is nonetheless exactly the diagnostic question that matters in screening. Determining what a discovered ensemble globally represents would require enumerating pairwise comparisons against a full database of trusted references which may be intractable. However, in several cases, we simply need to know whether a discovered ensemble matches a specific trusted context, and H delivers exactly that decision with quantifiable statistical guarantees.

Additionally, the proposed paradigm is focused toward the comparison of ensemble features *between* paired sets of images. However, it is just as reasonable to: i) consider partitioning ensemble features within a single image or ii) clustering features over our novel metric space. Both the 'within' statistics and clustering concepts are interesting and should be studied further. However, in this work, we dedicate our focus toward the problem at hand: comparing pairs of ensembles informed by two (or more) images.

'Difference' here is stochastic and invariant to Euclidean motions—the set of all distance-preserving rotations /reflections/translations (isometries). Randomness arises from aleatoric uncertainty in the imaged objects (Fig. 1) and epistemic uncertainty in segmentation hyperparameters, so decision-making must be stochastic. A naive deterministic approach—e.g., comparing pointwise pixel differences—rejects almost surely absent a perfect registration and therefore carries no information about distributional difference.

A complementary naive approach is to summarize each shape by a handful of intuitive scalars—principal lengths, area, perimeter, sphericity, and so on [25, 28–30]. These are useful for a stochastic problem but amount to a 'lossy' projection of shape: Distinct geometries can share identical hand-picked summaries, so meaningful difference can collapse in unexpected ways. Despite this, commercial materials quality standards [31] are still written around such features, and we expect the shape *representations* developed below to inform improvements to those standards while extending well beyond materials science.

Quantifying these random differences is based on (maximum mean) *discrepancy* [32] over separately approximated distributions of representative shape features. Procedurally:

i) *Decompose* shapes into separate distinguishing features ← separable shape tensors
ii) *Discretize* and 'learn' separate feature distributions from images ← manifold learning
iii) *Test* for statistically significant discrepancies between images over separate feature distributions ← maximum mean discrepancy

In practice, this corresponds to *statistically comparing thousands or more shapes between two ensembles informed by two images* [32].

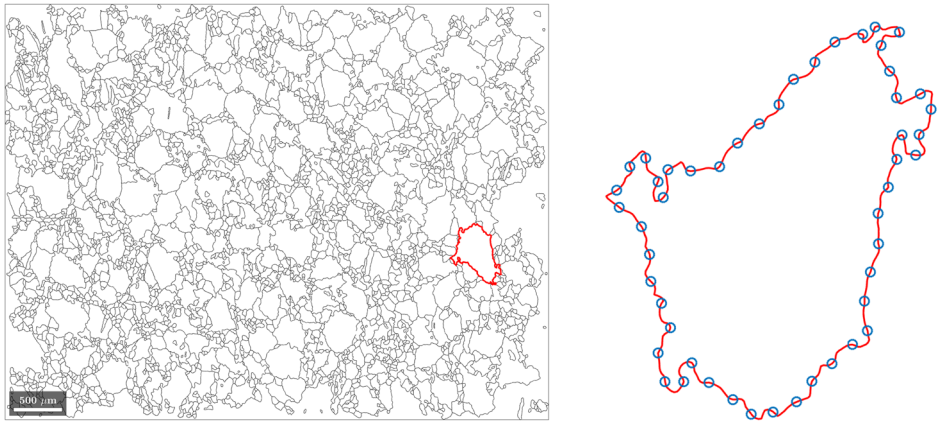

Fig. 1 An example ensemble of thousands of grain boundaries from an EBSD image of ice [5, 6] (left) and an example segmented grain boundary (right) with arc-length reparametrization landmarks (blue circles) generated by an interpolating curve (red). Data are available online [5]. The micron bar in the lower-left corner of the EBSD grain boundaries image reads 500 micrometers

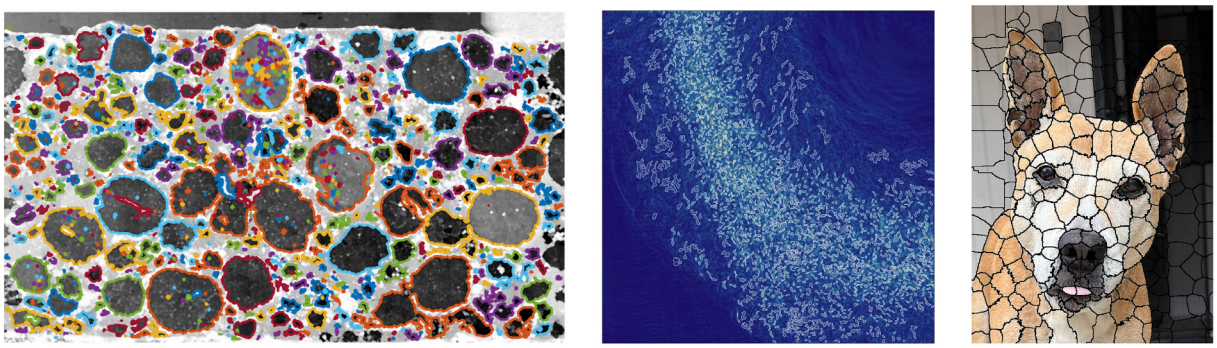

Fig. 2 (left) A scanning electron microscopy of a lithium-ion battery cross section, (middle) Lagrangian coherent structures in the lower-left quadrant of a cyclone modeled by large eddy simulation, (right) a 'superpixel' segmentation of canine named Penni

Shape is only one important piece of the quite literal puzzles in Figs. 1-2. However, we ignore the role of textures intrinsic to the tiling—i.e., the spatial arrangement of the shapes. Instead, we treat the ensemble of shapes as independent realizations from some generating distribution which, at present, *ignores their translations*. Related efforts are focused toward unifying both the topology of the spatial tiling and features of shapes in image.

1.2 Contributions & Outline

The manuscript follows the *decompose* $\mapsto$ *discretize* $\mapsto$ *test* arc introduced above: Sect. 2 together with Sects. 3.1–3.2 develops the SST decomposition; Sects. 3.3–3.4 treat registration and manifold learning; Sect. 3.5 and Sect. 4 carry out the tests. Each contribution below is labeled **[prior art]** if inherited from [34] or **[new]** otherwise and points at the section that develops the concept.

- **[prior art]** [34] Shape polar standardization (3) over $Gr(d,n) \times S^d_{++}$ (Sect. 1.4; supplementary derivations in Appendix A).
- **[prior art]** [34, 39–41, 71] Separable tangent PCA algorithms for intrinsic means and principal geodesic directions on $Gr(d,n)$ and S^d_{++} (reviewed in Sect. 3.4).
- **[new, Section 2]** Interpretation of the composite linear operator as a Hilbert-Schmidt integral operator on $\mathcal{C}_d \cap \mathcal{H}_d$, with PRRTI invariance and eigenfunctions given as scaled evaluation functionals of an $\mathbb{R}^d$-valued RKHS (Definitions 1, 2; Lemmas 1, 2, 3; Theorem 1).
- **[new, Sections 3.1–3.2]** Square root quadrature decomposition (SRQD) as a Nyström-weighted polar decomposition of landmark configurations (Fig. 4).
- **[new, Section 3.3]** Cyclic Procrustes registration against a random asymmetric archetype (Theorem 2).
- **[new, Section 3.5]** Formulation of two-sample testing as the binary mapping $H : (\mathcal{E}, \widehat{\mathcal{E}}) \mapsto \{0, 1\}$ with De Morgan aggregation of marginal MMD tests (Table 1).
- **[new, Section 4]** Decision landscape diagnostic, an EBSD grain boundary study illustrating all four logical conjunctions of Table 1 and a lithium-ion battery SEM study addressing the 'goldilocks problem' of smoothing segmented boundaries without obscuring distinguishing features.

Our primary theoretical contribution is the third bullet: relating the SST configuration space over finite-dimensional matrix manifolds to a functional analysis over an infinite-dimensional RKHS. This generalizes classical configuration-space work of Kendall et al. [21]—which has no demon-

strated connection to an infinite-dimensional space of curves —and permits flexible kernel choice [33] to control regularity. Relative to varifold and Sobolev metric approaches [15–19] that return a single scalar distance, and to AI-based analysis that requires hand-labeled data mapped to opaque latent features, the separable RKHS construction isolates undulation from anisotropic scale and supplies logical explanations via Truth Table 1.

A primary practical contribution is computational efficiency: Registration reduces to finite SVDs collocated at quadrature nodes and cyclic permutations against a single asymmetric archetype (Theorem 2), rather than optimization over diffeomorphisms as in elastic and Sobolev metric frameworks [12, 17, 18]. Thousands of curves can therefore be decomposed and tested in seconds on a conventional machine, making tractable the parameter sweeps of Sect. 4—including the sub-visual smoothing-induced discrepancies in Fig. 18 and the 'goldilocks' segmentation analysis of Fig. 15. To our knowledge, comparisons over such scales and sweeps have not previously been demonstrated for curve ensembles.

1.3 Notation

In general, bold fonts distinguish vector-valued objects from capitalized fonts which represent matrix-valued objects. We reuse common conventions for the following spaces of matrices:

- $GL(d,\mathbb{R})$ is the space of d-by-d full rank matrices with real entries while $GL_+(d,\mathbb{R})$ is the subgroup of $GL(d,\mathbb{R})$ with positive determinant.
- $St(d,n)$ is the Stiefel, the space of $n \times d$ matrices with orthonormal columns while $\mathbb{R}^{n\times d}_*$ is the corresponding non-compact Stiefel of full rank (d) matrices.
- $Gr(d,n)$ is the Grassmannian, the space of d-dimensional subspaces of $\mathbb{R}^n$.
- $O(d)$ is the space of d-by-d orthogonal matrices while $SO(d)$ is the subgroup of $O(d)$ with determinant equal to one.
- S^d_{++} is the cone of d-by-d symmetric positive definite matrices.

Quotient topologies are written $\mathcal{A}/\mathcal{B}$ ('dividing out' the action of $\mathcal{B}$), and square brackets $[\cdot]$ denote the induced equivalence classes (e.g. $[\boldsymbol{c}] \in \mathcal{A}/\mathcal{B}$). Elsewhere, we define new notation explicitly and use ':=' to indicate identification by definition. Angle brackets $\langle\cdot,\cdot\rangle$ denote bilinear inner products; the scalar dot product is written equivalently as $\boldsymbol{a}^\top\boldsymbol{b} = \boldsymbol{a}\cdot\boldsymbol{b}$; vectors are always tall. One conflict arises with a redundant use of α. The bold notation $\boldsymbol{\alpha}$ is a vector of constant coefficients while α defines the significance level of the hypothesis tests.

1.4 Review

Related work [10, 21, 34, 35] defines finite representations of curve data as *configuration spaces*. Following the development of SST [34] for aerodynamic shape representations [36], we let $X = (\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n)^\top$ be n unique *landmarks* drawn from a curve $\boldsymbol{c} : \mathcal{I} \subset \mathbb{R} \to \mathbb{R}^d$ via

$$\boldsymbol{x}_i = \boldsymbol{c}(s_i), \quad i = 1,\ldots,n, \tag{1}$$

with $X \in \mathbb{R}^{n\times d}_*$ full rank to exclude degenerate points/lines [34, 37]. Appendix A offers a more extensive review.

The central object is the thin SVD of the centered landmarks,

$$(X - B(X))^\top \overset{SVD}{=} U\Sigma V^\top, \tag{2}$$

with $B(X) = (1/n)\mathbf{1}_{n,n}X$. The polar standardization [34] yields $X = \widetilde{X}P$ with $\widetilde{X} := VU^\top$ and $P := U\Sigma U^\top \in S^d_{++}$. Identifying $S^d_{++} \cong GL_+(d,\mathbb{R})/SO(d)$ via [39] and $Gr(d,n) \cong \mathbb{R}^{n\times d}_*/GL(d,\mathbb{R})$ via [37], the equivalence class $[\widetilde{X}] \in Gr(d,n)$ captures nonlinear *undulation* while $P \in S^d_{++}$ captures generalized anisotropic *scale*, independent of rotations and reflections. The pair $([\widetilde{X}], P) \in Gr(d,n) \times S^d_{++}$ parametrizes the SST

$$X(\boldsymbol{t},\boldsymbol{\ell}) = \widetilde{X}(\boldsymbol{t})P(\boldsymbol{\ell}), \tag{3}$$

with $\boldsymbol{t} \in \mathbb{R}^r$ $(1 < r < d(n-d))$ and $\boldsymbol{\ell} \in \mathbb{R}^3$ learned by separate tangent PCA [34, 40, 41] over the aggregate ensemble.

A key distinction from classical configuration space analyses [10, 12, 21] is the generalized notion of scale. Classical treatments model scale as a scalar dilation, whereas an SST takes the full cone S^d_{++} of anisotropic deformations, so the product $Gr(d,n) \times S^d_{++}$ separates undulation from anisotropic scale—a separation essential to the hypothesis tests of Sect. 3.5.

An important caveat of this finite representation is that shapes are not precluded from generating self-intersections—i.e., discrete shapes are generally identified with immersions. Moreover, the existing analysis is predicated on hand-picked discretizations of interpolating curves, called *fixed reparametrizations* [34], and lacks a formal interpretation of the entries of $\widetilde{X}$—other than achieving desirable aerodynamic parametrizations and subsequent meshing. This begs the question of how to relate the finite representation of (3) to an (infinite dimensional) analysis of continuous curves to offer improved discretizations and continuous extensions as immersions.

1.5 Preshapes as Boundary Curves

As before, preshapes are defined as open or closed curves $\boldsymbol{c} : \mathcal{I} \to \mathbb{R}^d$ assumed to be immersions and/or embeddings [11, 12, 22, 33, 42, 43]—e.g., for $d = 2$, $\boldsymbol{c}$ in our application is a planar immersion over $\mathcal{I}$ as the unit circle or equivalent compact, connected 1-manifold denoted $\text{Imm}(\mathcal{I}, \mathbb{R})$. Applications are benefited by representations of curves over regularized spaces which control noisy measurement oscillations, e.g., those with increasing Lipschitz constant of curves $\mathcal{C}_d(\lambda_r) \subseteq \mathcal{C}_d(\lambda_{r+1})$ such that $0 < \lambda_r \leq \lambda_{r+1} < \infty$ where

$$\mathcal{C}_d(\lambda_r) := \{\boldsymbol{c} \in \text{Imm}(\mathcal{I}, \mathbb{R}^d) : \|\boldsymbol{c}(u) - \boldsymbol{c}(s)\| \leq \lambda_r |u - s|\} \quad (4)$$

for all $u, s \in \mathcal{I}$, to explicitly control nonlinear *undulations*. Notice, for boundedness it is sufficient to assume the individual components of curve $\boldsymbol{c} = (c_1, c_2, \ldots, c_d)$ are Lipschitz, such that $|c_i(u) - c_i(s)| \leq \lambda_{i,r}|u - s|$ with $\lambda_r \leq \|(\lambda_{1,r}, \lambda_{2,r}, \ldots, \lambda_{d,r})^\top\| < \infty$.

We will generally refer to these curves as *integrable curves* for some $d\mu$, such that $\int_{\mathcal{I}} \boldsymbol{c}\, d\mu < \infty$ and $\int_{\mathcal{I}} \|\dot{\boldsymbol{c}}\|\, d\mu < \infty$ where $\dot{\boldsymbol{c}}$ represents the d component-wise derivative over $\mathcal{I} \subset \mathbb{R}$, with respect to some curve parameter. In other words, by Rademacher's theorem [44], we argue $\boldsymbol{c} \in \mathcal{C}_d(\lambda_r)$ with derivative $\dot{\boldsymbol{c}}$ almost everywhere admits integrable speed, $\|\dot{\boldsymbol{c}}\|$, over bounded $\mathcal{I}$.

1.5.1 Arc-length reparametrization

Conflating the assumed differentiability almost everywhere over a bounded domain $\mathcal{I}$ to weakly differentiable (absolutely) continuous curves is a convenience for reviewing arc-length reparametrizations as arc-length measures. The core developments in this work only require integrable curves over a bounded domain with the arc-length measure, which is sufficient to assume they are at least Lipschitz continuous per (4) but does not necessitate that curves be more regular—similar to arguments in [12, 45].

More precisely, the arc-length function,

$$\xi^{-1}(t) = \int_0^t \|\dot{\boldsymbol{c}}(s)\| d\mu(s), \quad (5)$$

can be expressed according to uniform measure $d\mu(s) = \rho(s)ds$,

$$\rho(s) = \begin{cases} 1, & s \in \mathcal{I} \\ 0, & s \notin \mathcal{I}, \end{cases} \quad (6)$$

where, arbitrarily assuming unit length curves, $\mathcal{I} := [0, 1]$. Via variable substitution, $u = \gamma(s)$ given any nonnegative $\gamma : \mathcal{I} \to \mathcal{I}$ such that $\gamma(0) = 0$ and $\dot{\gamma} > 0$ implies $ds = \dot{\gamma}^{-1}(s)du$. Thus, by assigning $t' := \gamma(t)$,

$$\begin{aligned}(\xi^{-1} \circ \gamma)(t) &= \int_0^{\gamma(t)} \|(\boldsymbol{c} \dot{\circ} \gamma)(s)\| ds \\ &= \int_0^{t'} \|\dot{\boldsymbol{c}}(u)\| |\dot{\gamma}(s)| \dot{\gamma}^{-1}(s) du \\ &= \int_0^{t'} \|\dot{\boldsymbol{c}}(u)\| du \\ &= \xi^{-1}(t'). \end{aligned} \quad (7)$$

Therefore, ξ^{-1} is *invariant to reparametrization*—the range of ξ^{-1} is unmodified by the composition with γ—by arbitrarily renaming t' as t.

Notably, we may reinterpret changes in the scale of velocity, $|\dot{\gamma}(s)|$, (speed) as an integral measure, $d\mu(s) := |\dot{\gamma}(s)|ds$. Thus, the scale of the speed we traverse the curve can instead be identified with an integral (pushforward) measure along the curve, i.e., for $\tilde{\boldsymbol{c}} = \boldsymbol{c} \circ \gamma$ we have

$$\begin{aligned}\tilde{\xi}^{-1}(t) &= \int_0^t \|\dot{\tilde{\boldsymbol{c}}}(s)\| ds \\ &= \int_0^t \|\dot{\boldsymbol{c}}(s)\| |\dot{\gamma}(s)| ds \\ &= \int_0^t \|\dot{\boldsymbol{c}}(s)\| d\mu(s). \end{aligned} \quad (8)$$

Here, γ is any diffeomorphism constituting a *reparametrization*. As a simple identification, we take $d\mu(s) := \rho(s)ds / \|\dot{\boldsymbol{c}}(s)\|$ for $\rho(s)$ uniform so that $\xi^{-1}(t) = t$. Thus $d\mu(s)$ becomes a natural choice of *fixed* integral measure which utilizes a *uniformly weighted* (weak) arc-length reparametrization as our integral measure. Moreover, by composition with differentiation and the norm, the *arc-length measure is invariant to Euclidean actions* on $\boldsymbol{c}$. Other applications may be interested in an alternative weighting (density) ρ, which integrates to one, to collect points near hand-picked landmarks of interest [2, 34].

Alternative treatments may also identify any diffeomorphism γ, up to scale, with the cumulative distribution function and corresponding nonnegative probability density, $\rho \propto \dot{\gamma}(s)$. Moreover, this identification could be extended to the Fourier dual (characteristic function) of γ. These identifications are reminiscent of the *varifolds* [15, 16, 43, 46] and applications of Sobolev metrics [17] and Fisher–Rao metrics in functional data analysis [9, 10] as a promising flexible treatment of reparametrizations.

Sobolev-type metrics are standard for controlling regularity and avoiding vanishing distances over (manifold) topologies of preshapes and shapes [22, 33, 47], but $\mathcal{C}_d(\lambda_r)$ with the 2-norm and weak differentiability suffices here because the subsequent analysis projects onto *finite* matrix manifolds on which distances do not vanish [22]. Although

the space is naturally parametrized by the Lipschitz bounding constant, we do not pursue that formalism in this work and refer simply to $\mathcal{C}_d$ with an unknown finite bound; we nonetheless observe empirically in Sect. 3.4 that the learned submanifold dimensionality r respects the nesting $\mathcal{C}_d(\lambda_r) \subseteq \mathcal{C}_d(\lambda_{r+1})$.

2 Curve FIE

To formalize the reviewed landmark standardization of SSTs, we study a second kind homogeneous Fredholm integral equation (FIE). The SVD (2) of SSTs is equivalently written as the symmetric counterpart as the (thin) scaled eigendecomposition,

$$\frac{1}{n}(X - B(X))(X - B(X))^\top \overset{SVD}{=} V\Sigma_d^2 V^\top. \tag{9}$$

Note, the 'thin SVD' and assumed linear independence of component functions implies Σ_d is d-by-d, $\Sigma_d^2 = \text{diag}(\sigma_1^2, \ldots, \sigma_d^2)$ with $\sigma_1^2 \geq \sigma_2^2 \geq \cdots \geq \sigma_d^2 > 0$. A column partition $V = (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_d)$ defines V as an element of the Stiefel manifold, $V^\top V = \mathbb{I}_d$, such that $V_{ij} = (v_i)_j$ is the i-th (row) entry of the j-th (column) sequence—i.e., $i = 1, \ldots, n$ and $j = 1, \ldots, d$. Equivalently, recalling in (1) that rows of the matrix are curve evaluations, the weighted eigendecomposition can be expressed as

$$\sum_{i=1}^{n} w_i(\boldsymbol{\tau}(s_p) \cdot \boldsymbol{\tau}(s_i))(v_i)_j = \sigma_j^2 (v_p)_j. \tag{10}$$

In (10), $w_i = 1/n$ are constant (over i) weights in the sum, and $\boldsymbol{\tau} : \mathcal{I} \to \mathbb{R}^d$ represents a fixed reparametrization of a 'translated' parametric curve generating data in the rows of $X - B(X)$, i.e., $X - B(X) \approx (\boldsymbol{\tau}(s_1), \ldots, \boldsymbol{\tau}(s_n))^\top$.

This alludes to an interpretation that specific landmark standardizations [34, 35] act as simple Riemann sums for integrable curves. Generally, any constant-weight quadrature rule over a fixed reparametrization according to arbitrary integral measure $d\mu(s)$ can be described as a weighted version of (2) with an appropriate sequence of nodes (s_i) mapping to $(\boldsymbol{x}_i)$. Thus, (10) is interpreted as a quadrature (numerical integration) for any integrable curve.

With this continuous extension in mind, for any integrable curve $\boldsymbol{c} \in \mathcal{C}_d$ over integral measure $d\mu$, define

$$\mathcal{T}[\boldsymbol{c}](s) := \boldsymbol{c}(s) - \int_{\mathcal{I}} \boldsymbol{c}(u) d\mu(u) \tag{11}$$

for all $s \in \mathcal{I}$, and notice $\mathcal{T}[\boldsymbol{c} + \boldsymbol{b}](\cdot) = \mathcal{T}[\boldsymbol{c}](\cdot)$ for arbitrary $\boldsymbol{b} \in \mathbb{R}^d$ which does not depend on the integral parameter. We refer to $\boldsymbol{\tau} := \mathcal{T}[\boldsymbol{c}]$ as the *translated and/or centered curve*.

Definition 1 *(Composite Linear Operator (CLO))* For $\boldsymbol{c} \in \mathcal{C}_d$, the *composite linear operator* $k_{\mathcal{T}}[\boldsymbol{c}] : \mathcal{I} \times \mathcal{I} \to \mathbb{R}$ is defined by

$$k_{\mathcal{T}}[\boldsymbol{c}](s, u) := \mathcal{T}[\boldsymbol{c}](s) \cdot \mathcal{T}[\boldsymbol{c}](u), \tag{12}$$

or equivalently,

$$k_{\mathcal{T}}[\boldsymbol{c}](s, u) = \|\mathcal{T}[\boldsymbol{c}](s)\| \|\mathcal{T}[\boldsymbol{c}](u)\| \cos(\theta), \tag{13}$$

where θ is the angle between pairs of translated landmarks $\boldsymbol{\tau}(s)$ and $\boldsymbol{\tau}(u)$.

Finally, the continuous analog of the weighted eigendecomposition (10) motivates an approximation of the form,

$$\int_{\mathcal{I}} k_{\mathcal{T}}[\boldsymbol{c}](s, u) v_j(u) d\mu(u) = \sigma_j^2 v_j(s), \tag{14}$$

alternatively denoted $\langle k_{\mathcal{T}}[\boldsymbol{c}](s_i, \cdot), v_j \rangle = \sigma_j^2 v_j(s_i)$ in the context of the canonical $L^2(\mathcal{I}, \mu)$ inner product.

Lemma 1 *Let $\boldsymbol{c} \in \mathcal{C}_d$. Then $k_{\mathcal{T}}[\boldsymbol{c}]$ is symmetric, i.e., $k_{\mathcal{T}}[\boldsymbol{c}](s, u) = k_{\mathcal{T}}[\boldsymbol{c}](u, s)$ for all $s, u \in \mathcal{I}$.*

Proof By definition of the scalar dot product, $k_{\mathcal{T}}[\boldsymbol{c}](s, u) = \mathcal{T}[\boldsymbol{c}](s) \cdot \mathcal{T}[\boldsymbol{c}](u) = \mathcal{T}[\boldsymbol{c}](u) \cdot \mathcal{T}[\boldsymbol{c}](s) = k_{\mathcal{T}}[\boldsymbol{c}](u, s)$. □

Lemma 2 *Let $v \in L^2(\mathcal{I}, \mu)$ and $\boldsymbol{c} \in \mathcal{C}_d$. Define the operator $\mathcal{K}_{\mathcal{T}}[\boldsymbol{c}]$ by*

$$(\mathcal{K}_{\mathcal{T}} v)[\boldsymbol{c}](\cdot) := \int_{\mathcal{I}} (\mathcal{T}[\boldsymbol{c}] \cdot \mathcal{T}[\boldsymbol{c}])(\cdot, u) v(u) \, d\mu(u). \tag{15}$$

Then $\mathcal{K}_{\mathcal{T}}[\boldsymbol{c}]$ is a Hilbert–Schmidt (HS) integral operator.

Proof Assuming an integrable curve implies boundedness of $\mathcal{T}[\boldsymbol{c}]$ in (13) and thus

$$\int_{\mathcal{I} \times \mathcal{I}} |k_{\mathcal{T}}[\boldsymbol{c}](s, u)|^2 d\mu(s) d\mu(u) < \infty. \tag{16}$$

The remainder is simply an application of Theorem 2.1 of [48]. In short, Theorem 2.1 in [48] states $\mathcal{K}_{\mathcal{T}}[\boldsymbol{c}]$ is a Hilbert–Schmidt operator and every Hilbert–Schmidt operator on $L^2(\mathcal{I}, \mu)$ is of the form (15) for some unique $k_{\mathcal{T}}[\boldsymbol{c}] : (s, u) \mapsto k_{\mathcal{T}}[\boldsymbol{c}](s, u)$ in $L^2(\mathcal{I} \times \mathcal{I}, \mu \times \mu)$. □

Consequently, by Mercer's theorem [48], we have general $L^2(\mathcal{I}, \mu)$ eigensolutions of the HS integral operator constituting solutions of (14). Thus, with Lipschitz continuity sufficient for weak arc-length integrability, v_j is the j-th eigenfunction satisfying the homogeneous FIE of the second kind [48] for this specific choice of curve-dependent operator.

Notice that composition in $\mathcal{T}$ offers translation invariance of the CLO. In fact, the CLO is invariant to a larger class of transformations on the curve:

Definition 2 *(PRRTI condition)* We say an operator on $\mathcal{C}_d$ is *permutation, rotation, reflection, and translation invariant* (PRRTI) if for any $\boldsymbol{c} \in \mathcal{C}_d$, $R \in O(d)$, and $\boldsymbol{b} \in \mathbb{R}^d$ independent of the curve parameter, the operator is unchanged under $\boldsymbol{c} \mapsto R\boldsymbol{c} + \boldsymbol{b}$.

Lemma 3 *Let $\boldsymbol{c} \in \mathcal{C}_d$. Then the CLO $k_{\mathcal{T}}[\boldsymbol{c}]$ is PRRTI.*

Proof Given $\mathcal{T}[\boldsymbol{c} + \boldsymbol{b}] = \mathcal{T}[\boldsymbol{c}]$ for arbitrary $\boldsymbol{b} \in \mathbb{R}^d$ independent of the curve parameter. Additionally, for arbitrary $R \in O(d)$ independent of the curve parameter,

$$\begin{aligned} k_{\mathcal{T}}[R\boldsymbol{c} + \boldsymbol{b}] &= (R\mathcal{T}[\boldsymbol{c} + \boldsymbol{b}]) \cdot (R\mathcal{T}[\boldsymbol{c} + \boldsymbol{b}]) \\ &= (R\mathcal{T}[\boldsymbol{c}]) \cdot (R\mathcal{T}[\boldsymbol{c}]) \\ &= \mathcal{T}[\boldsymbol{c}] \cdot \left(R^\top R\mathcal{T}[\boldsymbol{c}]\right) \\ &= \mathcal{T}[\boldsymbol{c}] \cdot \mathcal{T}[\boldsymbol{c}] \\ &= k_{\mathcal{T}}[\boldsymbol{c}] \end{aligned} \tag{17}$$

□

Thus, the CLO and the associated integral operation with Euclidean invariance of the arc-length measure fall into the class of PRRTI operators (Definition 2), also referred to as TRI kernels [33]. Lemma 3 implies the angle between pairs of landmarks, θ, is preserved under *Euclidean transforms* according to (13)—i.e., angles between landmarks are unchanged by distance-preserving transformations (isometries). Thus, we can interpret the FIE with the HS integral operator (15) as a separation of general linear (scale) deformations and Euclidean invariant angles between discrete pairs of landmarks (undulations).

2.1 Eigensolutions

Given that our assumed space of preshapes $\mathcal{C}_d$ is not a Hilbert space, we propose a restriction $\boldsymbol{c} \in \mathcal{C}_d \cap \mathcal{H}_d$ where $\mathcal{H}_d$ is a unique $\mathbb{R}^d$-valued Hilbert space of curves defined by a *bounded* reproducing kernel. We elaborate on a generalized dual formulation of (14) and the geometric nature of the sought separability described in [34] to establish a framework for continuous extensions with flexible regularity implied by a chosen kernel.

The preliminaries and developments aggregated in [49, 50] are a particularly helpful motivation. The restriction $\boldsymbol{c} \in \mathcal{C}_d \cap \mathcal{H}_d$ may trivially simplify, $\mathcal{C}_d \cap \mathcal{H}_d = \mathcal{H}_d$, based on the choice of $\mathcal{H}_d$ implied by an appropriate (matrix-valued) reproducing kernel [51]. In other words, redefining $\mathcal{C}_d := \mathcal{H}_d$ with bounded kernel is sufficient to satisfy Lipschitz continuity [49, 50] while strengthening regularity. Ultimately, given ambiguity about the 'true' regularity of curves generating boundary landmarks as data, proposing $\mathcal{C}_d \cap \mathcal{H}_d$ alludes to nuances regarding a flexible choice of bounded kernel uniquely defining the restriction. In this case, $\mathcal{H}_d$ can be designed to consider the most appropriate restriction for a variety of applications.

Utilizing the RKHS definition of [33], the following result establishes a novel interpretation of the eigensolutions to (21) and has a profound implication on both the selection of landmarks—that may otherwise be hand-picked [21]—as well as the overall complexity of computing SSTs:

Definition 3 *(Evaluation functionals of RKHS, [33])* Assume $(\mathcal{H}_d, \langle \cdot, \cdot \rangle_{\mathcal{H}_d})$ is a Hilbert space of $\mathbb{R}^d$-valued functions defined over $\mathcal{I}$, $\mathcal{H}_d$ is a Reproducing Kernel Hilbert Space (RKHS) if evaluation functionals,

$$v : \mathcal{H}_d \to \mathbb{R} \,:\, \boldsymbol{\tau} \mapsto \boldsymbol{\alpha} \cdot \boldsymbol{\tau}(s),$$

are linear and continuous over $\mathcal{H}_d$ and in $\boldsymbol{\alpha} \in \mathbb{R}^d$ for all $s \in \mathcal{I}$, i.e., $v \in \mathcal{H}_d^*$ is a dual representation of the curve.

Theorem 1 *Let $\boldsymbol{c} \in \mathcal{C}_d \cap \mathcal{H}_d$. Then the eigenfunctions of the CLO $k_{\mathcal{T}}[\boldsymbol{c}]$ are PRRTI orthonormal evaluation functionals of $\mathcal{C}_d \cap \mathcal{H}_d$.*

Proof Substituting the definition of the CLO into the generalized FIE,

$$\begin{aligned} \sigma_j^2 v_j(s) &= \langle k_{\mathcal{T}}[\boldsymbol{c}](s, \cdot), v_j \rangle_{\mathcal{H}_d^*} \\ &= \langle \boldsymbol{\tau}(s) \cdot \boldsymbol{\tau}(\cdot), v_j \rangle_{\mathcal{H}_d^*} \\ &= \langle \sum_{j'=1,\ldots,d} \tau_{j'}(s)\tau_{j'}(\cdot), v_j \rangle_{\mathcal{H}_d^*} \\ &= \sum_{j'=1,\ldots,d} \tau_{j'}(s) \langle \tau_{j'}, v_j \rangle_{\mathcal{H}_d^*} \end{aligned} \tag{18}$$

Thus, solving for the eigenfunctions as the (scaled) evaluation functionals implies

$$v_j(s) = \frac{1}{\sigma_j^2} \boldsymbol{\alpha}_j \cdot \boldsymbol{\tau}(s) \tag{19}$$

where $\boldsymbol{\alpha}_j^\top := (\langle \tau_1, v_j \rangle_{\mathcal{H}_d^*}, \ldots, \langle \tau_d, v_j \rangle_{\mathcal{H}_d^*})$. In other words, the j-th eigenfunctions are a linear combination of the component functions of the curve for non-trivial $\boldsymbol{\alpha}_j \neq 0$.

Substituting the solution (19) for $v_j(s)$ back into the definition of $\boldsymbol{\alpha}_j^\top = (\alpha_{j1}, \alpha_{j2}, \ldots, \alpha_{jd})$ reveals an algebraic dual of *eigensolutions* as paired eigenvalues and eigenfunctions:

$$\begin{aligned} \boldsymbol{\alpha}_j &= \frac{1}{\sigma_j^2} \begin{pmatrix} \langle \tau_1, \boldsymbol{\alpha}_j \cdot \boldsymbol{\tau} \rangle_{\mathcal{H}_d^*} \\ \vdots \\ \langle \tau_d, \boldsymbol{\alpha}_j \cdot \boldsymbol{\tau} \rangle_{\mathcal{H}_d^*} \end{pmatrix} \\ &= \frac{1}{\sigma_j^2} \begin{pmatrix} \langle \tau_1, \sum_{j'=1,\ldots,d} \alpha_{jj'} \tau_{j'} \rangle_{\mathcal{H}_d^*} \\ \vdots \\ \langle \tau_d, \sum_{j'=1,\ldots,d} \alpha_{jj'} \tau_{j'} \rangle_{\mathcal{H}_d^*} \end{pmatrix} \end{aligned}$$

$$= \frac{1}{\sigma_j^2}\begin{pmatrix} \sum_{j'=1,\dots,d}\langle\tau_1,\tau_{j'}\rangle_{\mathcal{H}_d^*}\boldsymbol{\alpha}_{jj'} \\ \vdots \\ \sum_{j'=1,\dots,d}\langle\tau_d,\tau_{j'}\rangle_{\mathcal{H}_d^*}\boldsymbol{\alpha}_{jj'} \end{pmatrix}$$
$$= \frac{1}{\sigma_j^2}\begin{pmatrix} \langle\tau_1,\tau_1\rangle_{\mathcal{H}_d^*} & \dots & \langle\tau_1,\tau_d\rangle_{\mathcal{H}_d^*} \\ \vdots & \ddots & \vdots \\ \langle\tau_d,\tau_1\rangle_{\mathcal{H}_d^*} & \dots & \langle\tau_d,\tau_d\rangle_{\mathcal{H}_d^*} \end{pmatrix}\boldsymbol{\alpha}_j$$
$$:= \frac{1}{\sigma_j^2}\langle\boldsymbol{\tau}\otimes\boldsymbol{\tau}\rangle_{\mathcal{H}_d^*}\boldsymbol{\alpha}_j. \tag{20}$$

We have defined $\langle\boldsymbol{\tau}\otimes\boldsymbol{\tau}\rangle_{\mathcal{H}_d^*}$ as the *component-wise integration* of the outer product of centered component functions. Rearranging this expression implies the eigenproblem,

$$\left(\langle\boldsymbol{\tau}\otimes\boldsymbol{\tau}\rangle_{\mathcal{H}_d^*} - \sigma_j^2\mathbb{I}_d\right)\boldsymbol{\alpha}_j = \mathbf{0}, \quad j = 1,\dots,d. \tag{21}$$

Here, $\langle\boldsymbol{\tau}\otimes\boldsymbol{\tau}\rangle_{\mathcal{H}_d^*} := \langle\mathcal{T}[\boldsymbol{c}]\otimes\mathcal{T}[\boldsymbol{c}]\rangle_{\mathcal{H}_d^*} \in S_{++}^d$ is symmetric positive definite (full rank) by symmetry of the inner product $\langle\cdot,\cdot\rangle_{\mathcal{H}_d^*}$ and the assumed linearly independent component functions, τ_j for $j = 1,\dots,d$.

Next, note that $\langle\boldsymbol{\tau}\otimes\boldsymbol{\tau}\rangle_{\mathcal{H}_d^*}$ admits strictly positive real eigendecomposition $\langle\boldsymbol{\tau}\otimes\boldsymbol{\tau}\rangle_{\mathcal{H}_d^*} = A\Sigma_d^2A^\top = (A\Sigma_d)(\Sigma_dA^\top)$ where $A \in O(d)$. That is, columns of $A = (\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\dots,\boldsymbol{\alpha}_d)$ satisfy $\boldsymbol{\alpha}_j\cdot\boldsymbol{\alpha}_{j'} = \delta_{jj'}$ where $\delta_{jj'}$ is Kronecker's delta function. Thus, $\Sigma_dA^\top\boldsymbol{\alpha}_j = \sigma_j\boldsymbol{e}_j$, with $\boldsymbol{e}_j$ the j-th column of the identity, implies

$$\begin{aligned}\langle v_j, v_{j'}\rangle_{\mathcal{H}_d^*} &= \frac{1}{\sigma_j^2\sigma_{j'}^2}\boldsymbol{\alpha}_j^\top\langle\boldsymbol{\tau}\otimes\boldsymbol{\tau}\rangle_{\mathcal{H}_d^*}\boldsymbol{\alpha}_{j'} \\ &= \frac{1}{\sigma_j^2\sigma_{j'}^2}(\Sigma_dA^\top\boldsymbol{\alpha}_j)\cdot(\Sigma_dA^\top\boldsymbol{\alpha}_{j'}) \\ &= \frac{1}{\sigma_j^2\sigma_{j'}^2}(\sigma_j\boldsymbol{e}_j)\cdot(\sigma_{j'}\boldsymbol{e}_{j'}) \\ &= \frac{1}{\sigma_j\sigma_{j'}}\delta_{jj'}. \end{aligned}\tag{22}$$

Normalizing the eigenfunctions to be orthonormal such that $\|\widetilde{v}_j\|_{\mathcal{H}_d} = 1$ results in

$$\widetilde{v}_j(s) := v_j(s)\Big/\sqrt{\langle v_j,v_j\rangle_{\mathcal{H}_d^*}} = \frac{1}{\sigma_j}\boldsymbol{\alpha}_j\cdot\boldsymbol{\tau}(s). \tag{23}$$

□

An important interpretation is established by virtue of Theorem. 1: *Solutions of* (21) *define dual (orthonormal) evaluation functionals* (23) *of (centered)* $\mathbb{R}^d$*-valued curves*. Thus, following the presentation of [33], we can naturally construct a (unique) $\mathbb{R}^d$-valued RKHS as an infinite-dimensional extension constituting a *space of undulations*. This establishes a direct correspondence for exploring improved shape metrics and alternative deformations, such as curl-free and divergence-free deformations, to further improve finite—yet computationally efficient—SST manifold learning detailed in [34]. Thus, we may equivalently refer to undulations as scale-invariant *PRRTI features* of an RKHS.

Given many $\boldsymbol{c}$, as approximations or interpolations of segmented points with distinct cardinality, we can rapidly compute eigensolutions of (21) to define the corresponding orthonormal eigenfunctions (23) as continuous duals. Numerically, the interpretation of Theorem. 1 motivates discrete approximations benefited by higher order quadratures of $\langle\boldsymbol{\tau}\otimes\boldsymbol{\tau}\rangle_{\mathcal{H}_d^*}$ compared to that of a Riemann sum in the original presentation of SST [34]—i.e., highly accurate approximations without closed form solutions. This offers an improved criteria for the selection of various preshape discretizations for configuration spaces. Specifically, we can maintain row-wise evaluations stored in X as *collocations at quadrature nodes which only depend on a choice of fixed* $d\mu$ *for all curves*—i.e., equivariant to reparametrization and independent of any given $\boldsymbol{c}$.

In other cases, it may be possible to motivate closed form solutions to (21)—e.g., $\boldsymbol{\tau}$ given as a specific spline/polynomial. However, we assume a high-order numerical quadrature rule will be required to compute eigenfunctions out of convenience and flexibility when experimenting with a variety of approximations and interpolations of data. Therefore, we desire numerical methods to compute quadrature weights and nodes typically with respect to parametric forms of $d\mu$ utilizing ρ as some nonnegative scalar-valued function integrating to one. One possibility is employing a recursive implementation by Lanczos-Stieltjes [52–54] to determine nodes and weights for arbitrary $\rho : \mathcal{I} \to \mathbb{R}_{\geq 0}$ integrating to one. Lanczos–Stieltjes methods potentially enable the exploration of more sophisticated reparametrizations but, in this setting, we continue to assume ρ is uniform.

2.1.1 Smooth Approximations

Alternatively, provided additional regularity of the curves is reasonable or useful for an application, given the d-tuple of orthonormal components $\widetilde{v}_j$, we can approximate dual representations arbitrarily well utilizing interpolation with orthogonal polynomials [55–57],

$$\mathcal{O}(s\,;\boldsymbol{\theta},\boldsymbol{\ell}) = \widetilde{\mathcal{V}}(s\,;\boldsymbol{\theta})P(\boldsymbol{\ell}). \tag{24}$$

In this case, the infinite-dimensional analog of (3) is a dual representation over coefficients $\boldsymbol{\theta}$ defining the ∞-by-d quasi-matrix $\widetilde{\mathcal{V}} = (\widetilde{v}_1,\dots,\widetilde{v}_d)$ with 'columns' represented by expansions over orthogonal Legendre polynomials—assuming uniform speed reparametrization and uniform ρ.

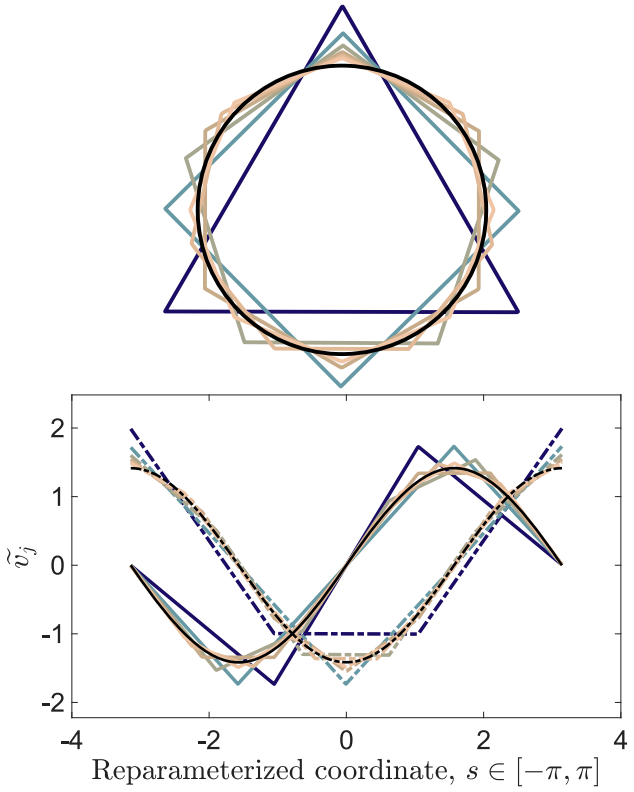

Fig. 3 (top) Piecewise linear polygonal shapes with identical generalized scales and increasing number of faces from triangle to octagon. (bottom) Corresponding orthonormal eigenfunctions. Solid lines correspond to $j = 1$ while dashed lines are $j = 2$ and colors match the corresponding shape above. The eigenfunctions are distinct, and thus, each polygon is distinct but appears to naturally converge to the eigenfunctions of the circle (black curve/functions) with an increasing number of faces

The open-source software package *chebfun* is useful for constructing such approximations [58].

When more regular curve representations are hypothesized, convenient, and beneficial—e.g., the reconstruction of reduced dimensional shapes depicted in Fig. 9—we reference Weierstrass approximation theorem to argue that polynomial representations are still dense in $\mathcal{C}_d$, a subset of absolutely continuous curves, over compact domains [57]. Therefore, with finite measurement precision in any imaging modality, we can get arbitrarily close (under measurement precision) with quasi-matrices [34, 55, 56].

In short, the interpretation of Theorem. 1 motivates discretizations by n curve evaluations at quadrature nodes instead of an arbitrary sequence. These discretizations are finite representations of: i) dual evaluation functionals, i.e., scale-invariant PRRTI features, in a unique matrix-valued RKHS [33], and/or ii) arbitrarily good quasi-matrix approximations built from orthogonal polynomials [55, 56].

2.2 Geometric Interpretation

Let's explore some simple examples computing (21) by hand and demonstrate the nature of these PRRTI features. Parametrizing a circle with $d = 2$, $\boldsymbol{c}(s) := (\cos(2\pi s), \sin(2\pi s))^\top$ for $\mathcal{I} := [0, 1]$. In this case, the standardization achieved by computing $\widetilde{v}_j$ should simply scale the original component functions to have norm one. We obtain eigenvalues $\sigma_1^2 = \sigma_2^2 = 1/2$ indicating that component functions covary equally and normalized eigenfunctions $\widetilde{v}_1(s) = \pm\sqrt{2}\cos(2\pi s)$ and $\widetilde{v}_2(s) = \pm\sqrt{2}\sin(2\pi s)$ according to $\boldsymbol{\alpha}_j = \pm\boldsymbol{e}_j$. Clearly, $\|\widetilde{v}_j\|_{\mathcal{H}_d} = 1$ for $j = 1, 2$ and $\langle v_i, v_j \rangle_{\mathcal{H}_d^*} = \delta_{ij}$.

Next, consider a simple linear scaling of the circle into an ellipse such that

$$\boldsymbol{c}(s) := \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix} \begin{pmatrix} \cos(2\pi s) \\ \sin(2\pi s) \end{pmatrix}. \tag{25}$$

Transforming to orthonormal eigenfunctions gives the same result as the circle, $\widetilde{v}_1(s) = \pm\sqrt{2}\cos(2\pi s)$ and $\widetilde{v}_2(s) = \pm\sqrt{2}\sin(2\pi s)$. However, the component functions covary differently and the eigenvalues are subsequently scaled as $\sigma_1^2 = a^2/2$ and $\sigma_2^2 = b^2/2$. With identical eigenfunctions, we conclude that a circle *equally undulates* as an ellipse and their distinction is due, entirely, to linear scale variations.

Lastly, if we now arbitrarily rotate the ellipse,

$$\boldsymbol{c}(s) := \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \begin{pmatrix} a\cos(2\pi s) \\ b\sin(2\pi s) \end{pmatrix}, \tag{26}$$

we obtain phase shifted eigenfunctions, $\widetilde{v}_1(s) = \pm\sqrt{2}\cos(2\pi s + \theta)$ and $\widetilde{v}_2(s) = \pm\sqrt{2}\sin(2\pi s + \theta)$, but (up to shifted reparametrization) our conclusion persists—all three of these curves are equally undulating with the only distinction being linear scale variations according to $\sigma_1^2 = a^2/2$ and $\sigma_2^2 = b^2/2$. In contrast, examples of piecewise linear polygons with increasing face counts admitting distinct undulations but identical anisotropic scales are shown in Fig. 3.

Notice that area, perimeter, and unsigned scalar curvature of the ellipse are distinct from the circle. Hence, *functionals of hand-picked features alone may indicate distinctions between each of these equally undulating curves*, but their distinction, up to phase shift, is due entirely to scale parameters a and b—which may or may not be well-defined by a subset of hand-picked features. In our case, a Hilbert space of bounded shift invariant kernels (convolutions or mollifiers) will eliminate the possibility of distinguishing between phase shifted eigenfunctions.

These examples are intended to motivate that the eigenproblem (21) is a principal components analysis (PCA) or eigendecomposition of the second central moments (varia-

tion) of the continuous curve, $\langle \mathcal{T}[\boldsymbol{c}] \otimes \mathcal{T}[\boldsymbol{c}] \rangle_{\mathcal{H}_d^*} = A\Sigma^2 A^\top$. However, the eigenfunctions of the corresponding integral kernel offer a definition of undulation as dual evaluation functionals of the curve which are 'factored' by linear scale variations informed by simple PCA.

Based on these conclusions, orthonormal (standardized) parametric forms satisfying $\|\boldsymbol{v}\|_{\mathcal{H}_d} = 1$ represented by dual evaluation functionals with (bounded) *periodic shift invariant kernel* equivalently constitute *expansions of undulation*. We may refer to such an expansion as a *Mercer series of undulation* over the construction of a dual space of PRRTI features to approximate the tuple of shape eigenfunctions, $\widetilde{\boldsymbol{v}} := (\widetilde{v}_1, \ldots, \widetilde{v}_d)^\top$, with any desired regularity.

3 PRRTI Numerics

Having formalized the SST decomposition and its RKHS interpretation in Sect. 2, we now turn to the numerical machinery that instantiates it and progressively moves from the *decompose* stage into *discretize* and on to *test*. The developments below assume an upstream image segmentation step as preprocessing has produced sets of ordered landmarks or integrable curves; we do not assume nor prescribe segmentation types, and related registration/alignment literature [10, 59–61] is complementary.

3.1 Nyström's Method for Curves

With explicit knowledge of the curve's construction, we may be able to compute eigenfunctions exactly. However, *absent explicit definitions of curves* and for convenience in exploring a variety of sufficiently smooth (regular) interpolations and approximations, we elaborate on a spectral method as a re-weighting of the SVD (2) for approximating eigenfunctions.

The development draws heavily from [62] and Section 12.1.5 of [48]. Essentially, we want the SVD of a symmetric positive definite matrix $K \in \mathbb{R}^{n\times n}$, e.g., $K_{ii'} = k_{\mathcal{T}}[\boldsymbol{c}](s_i, s_{i'})$ and Lemma 1, akin to the eigendecomposition (10) as the symmetric counterpart of (2),

$$K \overset{SVD}{=} V\Sigma_d^2 V^\top. \tag{27}$$

In [48, 62], it is noted that the columns of $V\Sigma_d^{-1}$ form a more stable basis for computation than those of the matrix K. This is especially important when the matrix K is fixed by some given data. The main contribution of [62] was to note that a *weighted* SVD could be computed by viewing the above SVD as a discretization of the HS integral eigenvalue problem. We can reinterpret this work to our ends. That is, rather than a more computationally stable basis for the native space of the kernel, we seek a more accurate approximation of the eigenfunctions than those originally proposed for SST. The result is simultaneous approximation and collocation (at quadrature nodes) of eigenfunctions as PRRTI features of curve.

As a matrix factorization of K, utilizing the weighted SVD,

$$K \overset{SVD}{=} W^{-1/2} V \Sigma_d^2 V^\top W^{-1/2}. \tag{28}$$

Thus, $V\Sigma_d^2 V^\top$ is the SVD of $W^{1/2} K W^{1/2}$ instead of K. Here, W is the diagonal matrix of positive weights that arise from the discretization of $(\mathcal{K}_{\mathcal{T}} v)[\boldsymbol{c}]$ using the quadrature rule, e.g.,

$$\int_{\mathcal{I}} \boldsymbol{\tau}(s) d\mu(s) \approx \sum_{i=1}^{n} w_i \boldsymbol{\tau}(s_i), \tag{29}$$

such that weights and nodes are determined by the fixed, weighted arc-length integral measure, $d\mu(s)$. To work out the appropriate weights for the specific case of uniformly weighted arc-length measures with variable speed, we make a few key observations. Since the translated curve $\boldsymbol{\tau}(s)$ is closed, the components of its parametrization are periodic. Thus, when given a sufficiently regular curve and sampling a uniform partition of $\mathcal{I}$ over arc-length, the necessary derivatives can be approximated to spectral accuracy using Fourier differentiation [63].

Given periodic integrand, integration over the period, and sampling on a uniform partition, the trapezoid rule will exhibit spectral accuracy assuming sufficient regularity of the curves. Thus, taking $W := W_\Delta W_\mu$ where the positive diagonal matrices W_Δ and W_μ contain the weights of the trapezoid rule and the contribution of the (pushforward) measure μ, respectively. In particular, $W_\Delta = \Delta u \mathbb{I}_n$ where Δu represents the arc-length gauge of the resulting discretization of the curve. For example, if we utilize a uniform partition with n-subintervals of $\mathcal{I} := [-\pi, \pi]$ to discretize then $\Delta u = 2\pi/n$. As a convention, we assume n subintervals for numerical integration and omit the duplicate $n+1$ point which closes the polygonal representation—i.e., $\boldsymbol{x}_1 = \boldsymbol{x}_{n+1}$ is the closure condition for our planar curves.

For example, computing weights of the pushforward measure can be accomplished using periodic spectral differentiation over a uniform grid,

$$W_\mu = \operatorname{diag}\left(((D_n T_n) \odot (D_n T_n))\, \mathbf{1}_{d,1} \right)^{-1/2}, \tag{30}$$

where $\odot$ is the Hadamard product and

$$T_n := \begin{pmatrix} \boldsymbol{\tau}^\top(s_1) \\ \vdots \\ \boldsymbol{\tau}^\top(s_n) \end{pmatrix} = \begin{pmatrix} \tau_1(s_1) & \ldots & \tau_d(s_1) \\ \tau_1(s_2) & \ldots & \tau_d(s_2) \\ \vdots & \vdots & \vdots \\ \tau_1(s_n) & \ldots & \tau_d(s_n) \end{pmatrix} \in \mathbb{R}_*^{n\times d}.$$

Additionally, D_n is the $n \times n$ Fourier differentiation matrix with entries, for all $i, i' = 1, \ldots, n$,

$$(D_n)_{ii'} = \begin{cases} 0, & i = i' \\ \frac{1}{2}(-1)^{i+i'} \cot\left(\frac{s_i - s_{i'}}{2}\right), & i \neq i', \end{cases} \tag{31}$$

where n is even, $s_i = i\Delta u$, and $\mathbf{1}_{d,1}$ is the column vector of all ones with length d [63]. Making this choice results in the standard eigenvalue problem,

$$\sum_{i=1}^{n} w_i k(\boldsymbol{\tau}(s_{i'}), \boldsymbol{\tau}(s_i)) v(s_i) = \widetilde{\sigma}^2 v(s_{i'}) \tag{32}$$

with linear kernel $k(\boldsymbol{\tau}(s_{i'}), \boldsymbol{\tau}(s_i)) = \boldsymbol{\tau}(s_{i'}) \cdot \boldsymbol{\tau}(s_i)$, generalizing the weighted extension posed in (10) and written in matrix–vector format as $KW\boldsymbol{v} = \widetilde{\sigma}^2\boldsymbol{v}$ where $\widetilde{\sigma}^2$ denotes the eigenvalues of the discrete problem. Moreover, this problem is not symmetric. Thus, we write

$$W^{1/2} K W^{1/2} \widetilde{\boldsymbol{v}} = \widetilde{\sigma}^2 \widetilde{\boldsymbol{v}} \tag{33}$$

where the symmetric matrix $W^{1/2}KW^{1/2}$ has the eigenvector $\widetilde{\boldsymbol{v}} = W^{1/2}\boldsymbol{v}$, a *weighted* version of $\boldsymbol{v}$.

3.2 Functional Approximation

Extending the discussion of Nyström method for curves, we develop numerical approximations of (19)-(21) by simplifying to a $d \times d$ eigenproblem as opposed to the larger $n \times n$ problem in (33).

From (12.15) in [48] and following the approach of Sect. 2.1 to obtain numerical approximations of (20) and (21), we have the SVD basis from the Nyström method written as

$$\boldsymbol{v}^\top(s) = \mathbf{k}^\top(s) W V \Sigma_d^{-2} \tag{34}$$

with

$$\begin{aligned} \mathbf{k}^\top(s) &= (K(s, s_1), \cdots, K(s, s_n)) \\ &= \boldsymbol{\tau}^\top(s) \begin{pmatrix} \tau_1(s_1) & \tau_1(s_2) & \cdots & \tau_1(s_n) \\ \vdots & \vdots & \cdots & \vdots \\ \tau_d(s_1) & \tau_d(s_2) & \cdots & \tau_d(s_n) \end{pmatrix} \\ &= \boldsymbol{\tau}^\top(s) T_n^\top. \end{aligned}$$

The expression $\boldsymbol{v}^\top(s) := (v_1(s), \ldots, v_d(s))$ represents the d-tuple of evaluation functionals (19), as a row vector, approximated with a quadrature rule and collocated at s. Now, as in Sect. 2.1, we will rewrite this expression without explicit dependence on the eigenvectors, $V \in \mathbb{R}_*^{n \times d}$, in the right-hand side. Importantly, without a closed form expression for the curve, $\boldsymbol{\tau}(s)$, (34) only returns eigenfunctions evaluated (i.e., collocated) at the quadrature nodes.

Note that K of (27) satisfies $\operatorname{rank}(K) = d$ by assumption and, hence,

$$\begin{aligned} \boldsymbol{v}^\top(s) &= \mathbf{k}^\top(s) W V \Sigma_d^{-2} \\ &= \boldsymbol{\tau}^\top(s) T_n^\top W (\mathbf{v}_1/\sigma_1^2, \mathbf{v}_2/\sigma_2^2, \cdots, \mathbf{v}_d/\sigma_d^2) \end{aligned} \tag{35}$$

where $\mathbf{v}_j$, $j = 1, \ldots, d$, are the column vectors of V collocated at quadrature nodes. Moreover, with a quadrature of $\boldsymbol{\alpha}_j$ as defined in the proof of Theorem. 1, $\boldsymbol{\alpha}_j := T_n^\top W \mathbf{v}_j$, and substituting into (35) we obtain

$$\boldsymbol{v}^\top(s) = \boldsymbol{\tau}^\top(s)(\boldsymbol{\alpha}_1/\sigma_1^2, \boldsymbol{\alpha}_2/\sigma_2^2, \cdots, \boldsymbol{\alpha}_d/\sigma_d^2). \tag{36}$$

Numerically, (36) is (19) in the context of the Nyström method and evaluates eigenfunctions given the translated curve, $\boldsymbol{\tau}$, and approximations of $\boldsymbol{\alpha}_j$'s.

Letting $A := (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_d) \in \mathbb{R}^{d \times d}$ and collocating (36) at quadrature nodes result in

$$V = \begin{pmatrix} \boldsymbol{v}^\top(s_1) \\ \vdots \\ \boldsymbol{v}^\top(s_n) \end{pmatrix} = T_n A \Sigma_d^{-2},$$

again, with $\Sigma_d^{-2} := \operatorname{diag}(1/\sigma_1^2, 1/\sigma_2^2, \ldots, 1/\sigma_d^2)$. Recalling $\boldsymbol{\alpha}_j := T_n^\top W \mathbf{v}_j$ implies $T_n^\top W V = A$ and substituting our collocation $V = T_n A \Sigma_d^{-2}$, in place of (20), we have $T_n^\top W (T_n A \Sigma_d^{-2}) = A$. Finally, rearranging reveals the anticipated identification as a weighted eigendecomposition,

$$T_n^\top W T_n \overset{SVD}{=} A \Sigma_d^2 A^\top, \tag{37}$$

specifically such that $A \in O(d)$. In short, orthonormal $\boldsymbol{\alpha}_j$, $1 \leq j \leq d$ as the columns of orthogonal A, are obtained simultaneously using an appropriate scheme to compute the *weighted* decomposition (37), instead of separately via (21), and (36) may be subsequently employed to evaluate the eigenfunctions at any point s, if needed.

Moreover, to obtain $\widetilde{V}$ with orthonormal columns for numerical implementations [34], we can equivalently compute the thin weighted SVD, or *square root quadrature decomposition*, (SRQD)

$$T_n^\top \sqrt{W} \overset{SVD}{=} A \Sigma_d \widetilde{V}^\top, \tag{38}$$

for all curves in the ensemble utilizing rapid rank-d decomposition schemes. A visual add is given in Fig. 4.

The resulting right singular vectors $\widetilde{V}$ are the weighted eigenvectors, $\widetilde{V} = W^{1/2}V$, of (33) in the Nyström method. Thus, *fast rank-d SRQDs can efficiently map thousands of preshapes to discrete PRRTI features*. Equivalently, with A

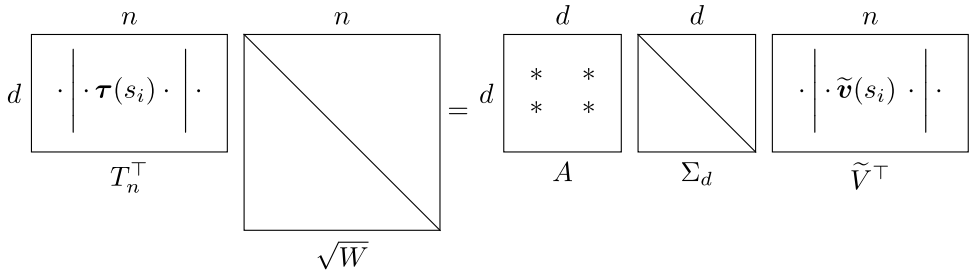

Fig. 4 SRQD of a translated (centered) curve where W is a diagonal of quadrature weights w_i, $\boldsymbol{\tau}(s_i)$ are centered landmarks evaluated at quadrature nodes, and $\widetilde{\boldsymbol{v}}(s_i)$ are the d-tuples of evaluation functionals at quadrature nodes

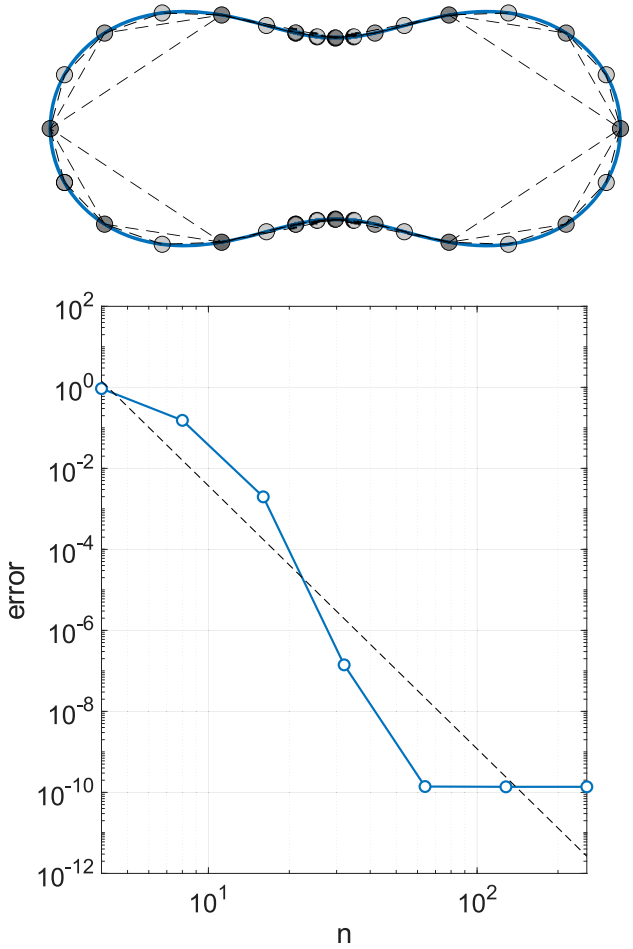

Fig. 5 (top) An increasing number of landmarks collocated at quadrature nodes over a Cassini oval. Dashed lines are a visual cue connecting nodes in a particular order corresponding to 8, 16, and 32 landmarks each. Replicated nodes, $\{\boldsymbol{\tau}(s_i)\}_{n=8} \subset \{\boldsymbol{\tau}(s_i)\}_{n=16} \subset \{\boldsymbol{\tau}(s_i)\}_{n=32}$, over the increasing total number of nodes have darker shading. (bottom) Convergence plot over the same Cassini oval with increasing number of quadrature nodes, n. Error is the maximum 2-norm difference in component functions over the curve parameter taken between approximated PRRTI features and a reference solution with 8192 nodes. A black dashed curve is shown with corresponding rate of ~ 6.5 to emphasize the spectral nature of convergence

orthogonal, this fixed-order numerical integration constitutes $\{T_n\} \mapsto \{(\widetilde{X}, P)\}$ as a *weighted polar decomposition*,

$$\begin{aligned}\sqrt{W}T_n &= \widetilde{V}\Sigma_d A^\top \\ &= \widetilde{V}(A^\top A)\Sigma_d A^\top \\ &= \widetilde{X}P,\end{aligned} \tag{39}$$

where $\widetilde{X} := \widetilde{V}A^\top$ are rotations and reflections of the eigenfunctions into a *unique view* and $P := A\Sigma_d A^\top$ are the corresponding decomposed (separated) generalized scale variations.

Thus, the SRQD acts as a mapping into separated pieces of generalized (anisotropic) scale variations $P \in S^d_{++}$, and representative orthonormal (preshape) undulations $\widetilde{X}$ of the Stiefel manifold, i.e., $\widetilde{X}^\top\widetilde{X} = \mathbb{I}_d$. Consequently, Theorem. 1 motivates rapid and accurate weighted decompositions collocated at n quadrature nodes, *which depend exclusively on the chosen integral measure*, instead of increasing sequences $(s_i)_{i=1}^n$ from alternative reparametrizations or optimization schemes to align pairs of curves. We saturate accuracy in the approximations over all curves in the ensemble by taking n large, e.g., $n = 500$, while only increasing the computational expense linearly over increasing n given fixed d.

To validate our interpretation, we offer an example in Fig. 5. Given $\boldsymbol{\tau}$ as a Cassini oval, we utilize (37) to inform approximations of (36) which are reevaluated at 10, 000 consistent points over the curve parameter to study convergence rates. Figure 5 emphasizes the spectral nature of convergence. In practice, we utilize interpolations of data in composition with high-order arc-length reparametrization approximations to achieve unit speed and uniformly distributed landmarks—i.e., trivializing $W_\mu \propto \mathbb{I}_n$ (Fig. 7).

3.3 Cyclic Procrustes

In other applications, like aerodynamic design [34], ensembles may be supplied with some fixed reparametrization establishing a convention for the initial starting landmark and orientation of curves. However, when an ensemble is the result of an image segmentation, the data lack a consistent starting landmark for registration/alignment. In other words, the first row of the discretization, T_n or X, is ambiguous

for closed segmented curves from an image. Additionally, as demonstrated in Sect. 2.2, undulation is unique *up to phase shift*.

When we lack extrinsic alignment conventions to phase shift collocations, we propose solutions to the following (discrete) alignment problem for registering preshape data *a priori*:

Theorem 2 (Separable Cyclic Procrustes) *Given translated discrete shapes* $X, Y \in \mathbb{R}_*^{n\times d}$ *(full rank),* $\mathcal{N} = \{1, \ldots, n\} \subset \mathbb{N}$*, p-cycles as powers of the permutation matrix*

$$C(p) = \begin{bmatrix} \mathbf{0} & \mathbb{I}_{n-1} \\ 1 & \mathbf{0}^\top \end{bmatrix}^p$$

for all $p \in \mathcal{N}$*, and* $\mathcal{L}(p) := Y^\top C(p) X$ *full rank,* (p_*, R_*) *is an optimal solution of*

$$\underset{p\in\mathcal{N},\, R\in O(d)}{\text{minimize}} \ \|C(p)X - YR\|_F$$

if, and only if,

$$p_* = \underset{p\in\mathcal{N}}{\text{argmax}} \, \|\mathcal{L}(p)\|_1$$

where $\mathcal{L}(p_*) \overset{SVD}{=} U_*\Omega_*V_*^\top$ *such that* $R_* = U_*V_*^\top$.

Proof Maximization of our derived parametric problem statement is equivalent to maximization of the form of Umeyama's Lagrangian—defined in the proof of the lemma in [64] and denoted here as $\mathcal{U}$. Rewritten absent the sign convention for general $R \in O(d)$ and composition over p, Umeyama's Lagrangian is

$$\begin{aligned} \mathcal{U}(p) &= \|C(p)X\|_F^2 + \|YR\|_F^2 - 2\|\mathcal{L}(p)\|_1 \\ &= \|X\|_F^2 + \|Y\|_F^2 - 2\|\mathcal{L}(p)\|_1. \end{aligned} \tag{40}$$

Next, consider an equivalent problem stated with an extraneous constraint identifying parameter separability similar to [65],

$$\begin{aligned} &\underset{p\in\mathcal{N}}{\text{maximize}} \quad \|\mathcal{L}(p)\|_1 \\ &\text{such that} \quad R = \pi_{\mathcal{U}}(\mathcal{L}(p)). \end{aligned}$$

However, distinct from [65], the bijection,

$$\pi_{\mathcal{U}}(\cdot) := (\cdot)(V_p\Omega_p^{-1}V_p^\top),$$

is defined by the SVD, $\mathcal{L}(p) \overset{SVD}{=} U_p\Omega_pV_p^\top$, such that $\Omega_p = \text{diag}(\omega_1(p), \omega_2(p), \ldots, \omega_d(p))$ with $\omega_1(p) \geq \omega_2(p) \geq \cdots \geq \omega_d(p) > 0$ and $U_p, V_p \in O(d)$. Utilizing the definition of $\pi_{\mathcal{U}}$, notice

$$\begin{aligned} \langle \pi_{\mathcal{U}}(\mathcal{L}(p)), \mathcal{L}(p)\rangle_F &= \langle \mathcal{L}(p)(V_p\Omega_p^{-1}V_p^\top), \mathcal{L}(p)\rangle_F \qquad (\star) \\ &= \langle U_p\Omega_pV_p^\top(V_p\Omega_p^{-1}V_p^\top), \mathcal{L}(p)\rangle_F \\ &= \langle U_pV_p^\top, U_p\Omega_pV_p^\top\rangle_F \\ &= \text{tr}(V_p\Omega_pV_p^\top) \\ &= \|\mathcal{L}(p)\|_1 \end{aligned}$$

and

$$\pi_{\mathcal{U}}(\mathcal{L}(p)) = U_pV_p^\top. \qquad (\star\star)$$

Next, we argue the (extraneous) parametric constraint, $R = \pi_{\mathcal{U}}(\mathcal{L}(p))$ for any p, is best. By definition, $\pi_{\mathcal{U}}(\mathcal{L}(p))$ is best if and only if $\langle \pi_{\mathcal{U}}(\mathcal{L}(p)), \mathcal{L}(p)\rangle_F > \langle R, \mathcal{L}(p)\rangle_F$ for all $R \neq \pi_{\mathcal{U}}(\mathcal{L}(p)) \in O(d)$ and $p \in \mathcal{N}$. Equivalently, it is sufficient to show

$$\langle R - \pi_{\mathcal{U}}(\mathcal{L}(p)), \mathcal{L}(p)\rangle_F < 0$$

for all $R \neq \pi_{\mathcal{U}}(\mathcal{L}(p)) \in O(d)$. Given $(\star)$,

$$\begin{aligned} &\langle R - \pi_{\mathcal{U}}(\mathcal{L}(p)), \mathcal{L}(p)\rangle_F = \langle R, \mathcal{L}(p)\rangle_F - \\ &\quad \langle \pi_{\mathcal{U}}(\mathcal{L}(p)), \mathcal{L}(p)\rangle_F \\ &= \langle R, \mathcal{L}(p)\rangle_F - \|\mathcal{L}(p)\|_1. \end{aligned}$$

Then, by the original argument of orthogonal Procrustes solution [66–68],

$$\begin{aligned} \langle R, \mathcal{L}(p)\rangle_F &= \text{tr}(R^\top U_p\Omega_pV_p^\top) \\ &= \text{tr}(V_p^\top R^\top U_p\Omega_p) \\ &= \langle Q_p, \Omega_p\rangle_F \end{aligned}$$

for some $Q_p = U_p^\top RV_p \in O(d)$ which satisfies $\langle Q_p, \Omega_p\rangle_F \leq \langle \mathbb{I}_d, \Omega_p\rangle_F = \|\mathcal{L}(p)\|_1$ with equality when $Q_p = \mathbb{I}_d$. However, $Q_p = \mathbb{I}_d$ if and only if $R = U_pV_p^\top$ and thus, by $(\star\star)$, $R = \pi_{\mathcal{U}}(\mathcal{L}(p))$. Hence, $\langle R, \mathcal{L}(p)\rangle_F - \|\mathcal{L}(p)\|_1 < 0$ for all $R \neq \pi_{\mathcal{U}}(\mathcal{L}(p)) \in O(d)$. □

Theorem 2 aligns each segmented curve to a fixed 'archetype' Y collocated at the same quadrature nodes and implicitly defines a consistent starting landmark (a 'best' discrete phase shift) over the rows of X. Uniqueness requires the archetype to be asymmetric[1]; in practice, a random draw from the pooled ensemble suffices, and we normalize X and Y to unit length (dilating Y to match the scale of X) for numerical stability. Figure 6 shows registration to within machine precision under an arbitrary rotation and cyclic permutation, and related continuous treatments [59] accelerate alignment

[1] A 'featureless' circle produces a constant objective.

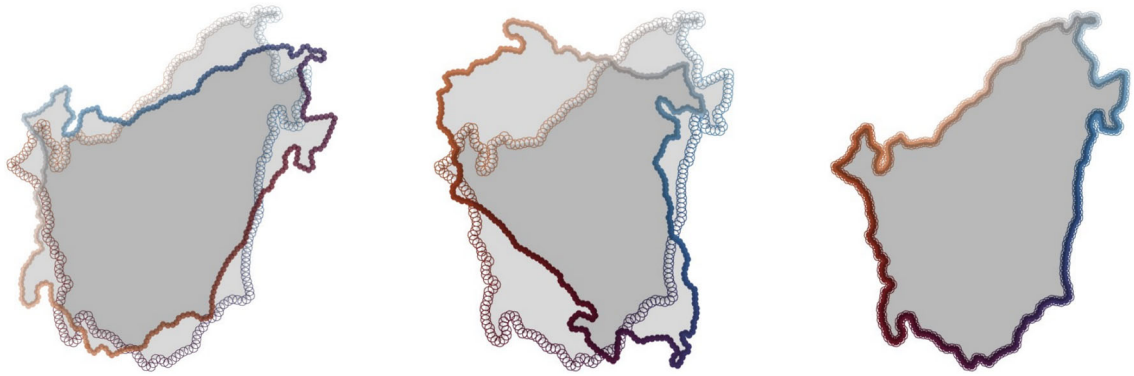

Fig. 6 (left) An arbitrary cyclic permutation and rotation of a uniformly discrete grain shape with colors indicating row index, $n = 500$. (center) An orthogonal Procrustes match. (right) A cyclic Procrustes match

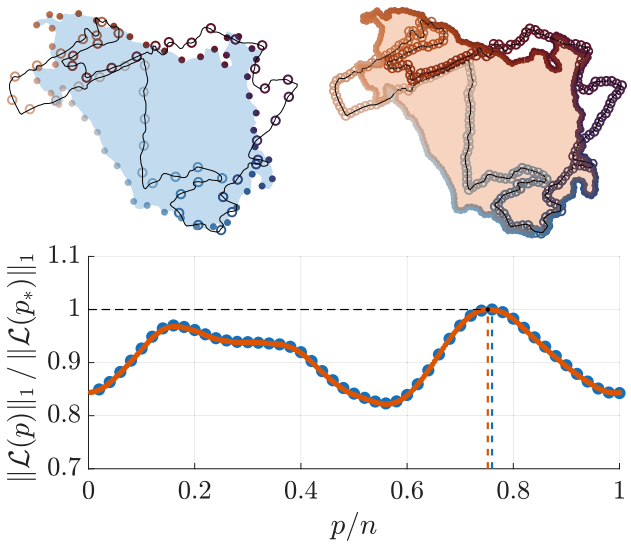

Fig. 7 Cyclic Procrustes matching at low ($n = 50$) and high ($n = 250$) levels of uniform arc-length reparametrization. The fixed archetype is shown with a black curve. The fill color of the matched shape corresponds to the levels of refinement ($n = 50$ blue and $n = 250$ orange) in the discrete objective function evaluations. The landmark colors correspond to the registered indices of the shape against the archetype

via fast Fourier transforms. We note that cyclic Procrustes itself induces a modified Procrustean metric [10] over discrete shapes and could serve as a standalone discrepancy measure on shape ensembles, independent of the SST and MMD machinery developed in the remainder of this work.

A natural alternative is to align against a 'central shape' such as the Fréchet mean. However, the Fréchet mean is coupled with alignment—alignment needs an archetype, the mean needs aligned data—and the iterative scheme (align $\mapsto$ mean $\mapsto$ repeat) may or may not converge for the discrete cyclic permutation. Theorem 2 instead guarantees an optimal alignment to any template, so the key requirement is asymmetry, not Fréchet optimality. Figures 11–12 confirm empirically that the random archetype registration is stable across n and, somewhat remarkably, that coordinate normalization alone stabilizes decisions even under random cyclic permutations (Fig. 8).

3.4 Ensemble Manifold Learning

With an ensemble of accurate, discrete, cyclically aligned, preshape PRRTI features $\{\widetilde{X}\}$ and their corresponding generalized scales $\{P\}$ now in hand, we pivot from constructing these matrices to learning with thousands of them informed by thousands of curves spanning the pair of images. Supplemental algorithmic details appear in Appendix A.

Naturally, to retain the sought properties of Lemma 3 (PRRTI features) and separation of generalized scale $P \in S^d_{++} \cong GL_+(d,\mathbb{R})/SO(d)$ in the representative *preshape* Stiefel discretizations $\widetilde{X}$, we must project approximated eigenfunctions onto the equivalence classes $[\widetilde{X}] \in \mathbb{R}^{n\times d}_*/GL_+(d,\mathbb{R})$ constituting the *shape* of undulations.

As described in [37], this quotient space is identified with the Grassmannian, $Gr(d,n) \cong \mathbb{R}^{n\times d}_*/GL(d,\mathbb{R})$. Note, $[\widetilde{V}] = [\widetilde{X}]$ but $\widetilde{X}$ serves as the *representative* preshape paired with scale variations P in the polar decomposition. With $\{(\widetilde{X}, P)\}$ as SRQD transformed data from a the aggregate ensemble informed by both segmented images, we leverage the *tangent PCA*[2] (tPCA) submanifold learning procedure [34, 40], separately, over underlying matrix manifolds $\{[\widetilde{X}]\} \subset Gr(d,n)$ and $\{P\} \subset S^d_{++}$. This procedure, with thousands of shapes, executes in seconds on a conventional laptop.

For example, tangent PCA proceeds with an ensemble of discrete matrix-valued PRRTI features $\{[\widetilde{X}]\}$ first approximating the Fréchet mean over $Gr(n,d)$, denoted $[\widetilde{V}_0]$. With the Fréchet mean establishing a local origin for a coordinate frame, we apply PCA to the image of the inverse exponential

[2] Note a correction of the misnomer. This analysis is more precisely described as *tangent PCA* as opposed to previous descriptions as *Principal Geodesic Analysis* (PGA).

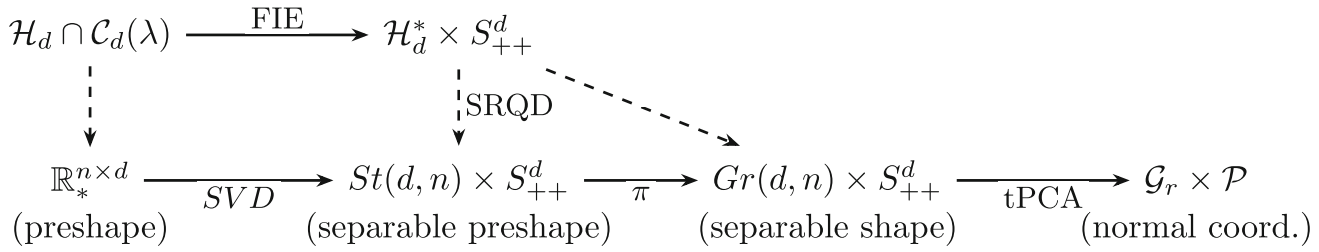

Fig. 8 Various named spaces and essential transformations to map from translated curves to normal coordinates over a section of the product submanifold. Here, π represents a normal projection of the first factor onto the equivalence classes of the Grassmannian. Dashed arrows indicate mappings into discrete representatives

at $[\widetilde{V}_0]$ to determine normal coordinates, $\boldsymbol{t} \in \mathbb{R}^r$, over an r-dimensional subspace, $1 \leq r \leq d(n-d)$. The span of this subspace parametrizes undulations over the matrix manifold as a local section $\mathcal{G}_r \subseteq Gr(d,n)$, i.e.,

$$\mathcal{G}_r := \{[X] \in Gr(d,n) \,:\, [X] = \mathrm{Exp}_{[\widetilde{V}_0]}(E_r \boldsymbol{t})\}, \tag{41}$$

where E_r is the 'learned' matrix[3] with columns constituting an ordered orthonormal basis of $\mathbb{R}^r \cong T_{[\widetilde{V}_0]}\mathcal{G}_r$. When selecting representative preshapes, *classic orthogonal Procrustes* is very useful for searching equivalence classes $[\widetilde{X}]$ to align a preshape $\widetilde{X}$ as desired [34].

This same learning procedure is repeated, separately, with $\{P\}$ to determine the intrinsic mean $P_0 \in \mathcal{P} \subseteq S_{++}^d$ and tangent basis spanning $T_{P_0}\mathcal{P} \subseteq T_{P_0}S_{++}^d$ facilitated by the algorithms in [39]. In our materials application, $\mathcal{P}$ is taken to be all of S_{++}^d but other applications may benefit from coordinates along subspaces of $T_{P_0}S_{++}^d$. Finally, consistent with [34], we utilize $\mathcal{G}_r \times \mathcal{P}$ as a *product submanifold* of separable (pre-)shape tensors parametrized with smooth right inverse $T_n(\boldsymbol{t}, \boldsymbol{\ell}) = \widetilde{X}(\boldsymbol{t})P(\boldsymbol{\ell})$ over normal coordinates $(\boldsymbol{t}, \boldsymbol{\ell}) \in T_{[\widetilde{V}_0]}\mathcal{G}_r \times T_{P_0}\mathcal{P}$.

The result of this learning is a mapping into local normal coordinates, $\{([\widetilde{X}_q], P_q)\} \mapsto \{(\boldsymbol{t}_q, \boldsymbol{\ell}_q)\} \subset \mathbb{R}^{r+3}$, over an r-submanifold, $\mathcal{G}_r \subset Gr(2,n)$. In composition, given an ensemble of curves $\{\boldsymbol{c}_q\}$ for $q = 1, 2, \ldots, N$ and fixed quadrature nodes $\{s_i\}$ for $i = 1, 2, \ldots, n$, the developed interpretation informs a procedure,

$$\{\mathcal{T}[\boldsymbol{c}_q](\{s_i\})\} \overset{SRQD}{\mapsto} \{\widetilde{X}_q P_q\} \tag{42}$$

$$\mapsto \{([\widetilde{X}_q], P_q)\} \tag{43}$$

$$\mapsto \{(\boldsymbol{t}_q, \boldsymbol{\ell}_q)\} \subset \mathbb{R}^{r+3}. \tag{44}$$

All algorithmic details mapping N SRQD's into normal coordinates, progressing from (42) to (44), are described in [34].

[3] Note $\mathrm{Exp}_{[\widetilde{V}_0]}(\cdot)$ is composed with an appropriate reshaping of the matrix–vector multiply [34].

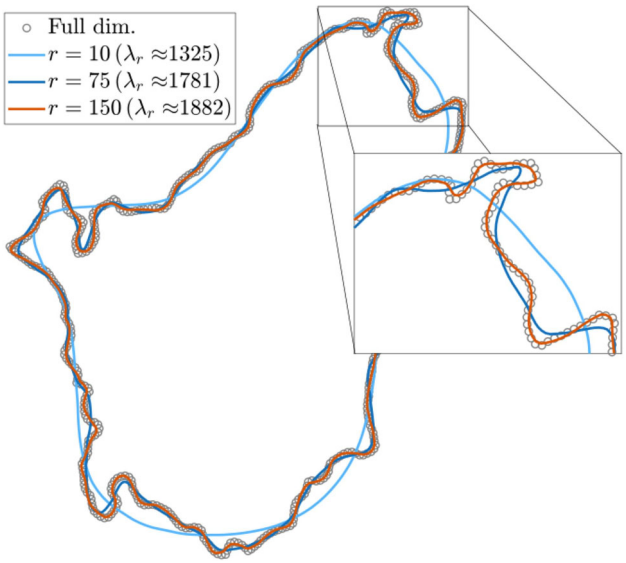

Fig. 9 Empirical effect of the sought shape regularization (4) over changing submanifold dimensionality, r, with $n = 500$ landmarks from a reparametrized spline. Lipschitz constants of the curves are approximated using *chebfun* [58], an open-source toolbox being utilized to compute quasi-matrix interpolation (24) of the reduced dimension analogs

Changing dimensionality r by accumulating terms in the ordered basis expansion over normal coordinates, $\boldsymbol{t} \in \mathbb{R}^r \cong T_{[\widetilde{V}_0]}\mathcal{G}_r$, can empirically achieve the sought regularization of (4) in the representation of the shapes. This observed mechanism can promote a biasing of undulating shape features away from potentially noisy variations in the segmented boundaries. An example of the empirical effects of this regularization is shown in Fig. 9. In later sections, we demonstrate how this choice of undulation dimensionality can be used to make subsequent inferences more robust against noisy measurements and image processing.

3.5 Maximum Mean Discrepancy

Having formalized the approximation of curve dual evaluation functionals with (shape) discretizations parametrized

Table 1 A logical conjunction (AND gate) for the two separable hypothesis tests (left of the double vertical lines) using pMMD

Case	$(\rho_{\boldsymbol{t}} = \widehat{\rho}_{\boldsymbol{t}})$	$(\rho_{\boldsymbol{\ell}} = \widehat{\rho}_{\boldsymbol{\ell}})$	pMMD$[\mathcal{H} \oplus \mathcal{Q}, \cdot, \cdot]$
$(1 \wedge 1)$	FTR	FTR	$\Longrightarrow$ FTR
$(1 \wedge 0)$	FTR	Reject	$\Longrightarrow$ Reject
$(0 \wedge 1)$	Reject	FTR	$\Longrightarrow$ Reject
$(0 \wedge 0)$	Reject	Reject	$\Longrightarrow$ Reject

Logical 'true' corresponds to *failure to reject* (FTR) the null hypothesis; logical 'false' corresponds to rejection

over a learned product submanifold, we explore an explainable binary classification of the learned undulation and scale coordinates $(\boldsymbol{t}, \boldsymbol{\ell}) \in T_{[\widetilde{V}_0]}\mathcal{G}_r \times T_{P_0}\mathcal{P}$ defined at respective intrinsic means $[\widetilde{V}_0] \in \mathcal{G}_r \subseteq Gr(n, d)$ and $P_0 \in \mathcal{P} \subseteq S^d_{++}$. Related work [69] explores similar concepts, albeit, modulo curve dilation over elastic metrics and Kendall shape spaces while utilizing a more general measure theory coined DISCO analysis [70].

We take the Borel σ-algebra over $T_{[\widetilde{V}_0]}\mathcal{G}_r \times T_{P_0}\mathcal{P}$ in a normal coordinate neighborhood—i.e., Hopf–Rinow theorem implies geodesic completeness[4] is equivalent to a metric space which is sufficient to generate the σ-algebra [73]. Thus, consider $(\boldsymbol{t}, \boldsymbol{\ell}) \sim \boldsymbol{\rho}$ as some joint distribution given by the smallest σ-finite (product) measure of random shape (normal) coordinates generated by the metric (feature) space.

The goal of maximum mean discrepancy (MMD) [32, 74] is to determine if finite observations of independent and identically distributed (i.i.d.) random variables defined on a topological space, with respective probability measures, coincide or not. That is, given observations $\{(\boldsymbol{t}_1, \boldsymbol{\ell}_1), \ldots, (\boldsymbol{t}_N, \boldsymbol{\ell}_N)\}$ and $\{(\widehat{\boldsymbol{t}}_1, \widehat{\boldsymbol{\ell}}_1), \ldots, (\widehat{\boldsymbol{t}}_{\widehat{N}}, \widehat{\boldsymbol{\ell}}_{\widehat{N}})\}$ assumed i.i.d. from $\boldsymbol{\rho}$ and $\widehat{\boldsymbol{\rho}}$, respectively, can we test whether $\boldsymbol{\rho} \neq \widehat{\boldsymbol{\rho}}$? Moreover, with separable parameters and in a complementary fashion, can we test whether $(\rho_{\boldsymbol{t}} = \widehat{\rho}_{\boldsymbol{t}})$ *and* $(\rho_{\boldsymbol{\ell}} = \widehat{\rho}_{\boldsymbol{\ell}})$ given shape features informed by an ensemble extracted from a pair of images?

This formal problem statement is the quantitative analog of the proposed explainable binary classification: *test if distributions of random shape coordinates partitioned between two images are discrepant, modulo Euclidean actions, and whether those discrepancies result from distinct distributions over undulation,* $\{\boldsymbol{t}\}$ *from the first image versus* $\{\widehat{\boldsymbol{t}}\}$ *from the second, or generalized scale,* $\{\boldsymbol{\ell}\}$ *from the first image versus* $\{\widehat{\boldsymbol{\ell}}\}$ *from the second.*

Instead of testing for equivalence, we look for statistically significant differences. This complementary separable formalism is simply De Morgan's theorem, $(\rho_{\boldsymbol{t}} = \widehat{\rho}_{\boldsymbol{t}}) \wedge (\rho_{\boldsymbol{\ell}} = \widehat{\rho}_{\boldsymbol{\ell}}) \leftrightarrow \neg((\rho_{\boldsymbol{t}} \neq \widehat{\rho}_{\boldsymbol{t}}) \vee (\rho_{\boldsymbol{\ell}} \neq \widehat{\rho}_{\boldsymbol{\ell}}))$. Truth Table 1 details this formalism where *logical truth* corresponds to *failure to reject* (FTR) the null hypothesis of equivalence, $(\rho_{(\cdot)} = \widehat{\rho}_{(\cdot)})$, when comparing against approximated thresholds as discussed in [32].

Concretely, setting $\mathcal{E} := \boldsymbol{\rho}$ and $\widehat{\mathcal{E}} := \widehat{\boldsymbol{\rho}}$ as the arguments of H in Sect. 1, the MMD hypothesis test is an unsupervised, distribution-level binary classifier whose separable decomposition populates the four rows of Truth Table 1. Again, such a framework extends naturally to supervised or context-driven classification over labeled ensembles.

Because the separable test is a logical conjunction of two marginal MMD tests, a union bound gives joint Type I error at most 2α under the intersection null. We therefore adopt the Bonferroni correction and run each marginal at $\alpha/2$ so that the joint level is bounded by α. For Sect. 4 at $\alpha = 0.02$ (hence $\alpha/2 = 0.01$), all rejections satisfy $p \ll 0.01$ and all FTRs satisfy $p \gg 0.02$. The De Morgan identity underlying Truth Table 1 is a logical formalism for aggregating results while Bonferroni is statistical correction on the aggregated results.

Here, as in [32], we utilize MMD defined over a unit ball in an appropriate RKHS, $\mathcal{F}$,

$$\mathrm{MMD}[\mathcal{F}, \boldsymbol{\rho}, \widehat{\boldsymbol{\rho}}] := \sup_{\psi \in \mathcal{F}} |\mathbb{E}_{\boldsymbol{\rho}}[\psi] - \mathbb{E}_{\widehat{\boldsymbol{\rho}}}[\psi]|. \tag{45}$$

We note that $\mathcal{F}$ is, presently, distinct from the $\mathbb{R}^d$-valued RKHSs motivated by Theorem. 1, but the measure theory is extensible to spaces of curves and alternative metric spaces [69, 70]. For simplicity and enabling the empirical regularization depicted in Fig. 9, we take $\mathcal{F}$ as an RKHS representing features of 'learned' parameter distributions over reduced dimensional normal coordinates informed by the tangent PCA of undulations and generalized scales. Future work is aimed at reconciling the two RKHSs—one as a functional space of curve equivalences and the other representing a distribution of learned features.

In (45), $\mathbb{E}_{(\cdot)}$ represents expectation (integration) with respect to the identified probability measures, $\boldsymbol{\rho}$ versus $\widehat{\boldsymbol{\rho}}$, which are informed by samples from distinct measurements or observations—i.e., distributions of shape features from distinct images in our case. Solutions to (45) are motivated empirically utilizing a choice of symmetric positive definite kernel, $f \in \mathcal{F}$, and Lemma 6 in [32] such that we can express the squared population MMD over $\mathcal{F}$ as

$$\mathrm{MMD}^2[\mathcal{F}, \boldsymbol{\rho}, \widehat{\boldsymbol{\rho}}] = \boldsymbol{E}_{\boldsymbol{\rho}}[f] + \boldsymbol{E}_{\widehat{\boldsymbol{\rho}}}[f] - 2\boldsymbol{E}_{\boldsymbol{\rho}\widehat{\boldsymbol{\rho}}}[f]. \tag{46}$$

This rewrite utilizes a simplification via the 'tower property' such that $\boldsymbol{E}_{\boldsymbol{\rho}}[f] := \mathbb{E}_{\boldsymbol{\rho}}[\,\mathbb{E}_{\boldsymbol{\rho}}[f|\boldsymbol{\theta}]\,]$ is a double expectation over $\boldsymbol{\rho}$, similarly for $\boldsymbol{E}_{\widehat{\boldsymbol{\rho}}}$, and $\boldsymbol{E}_{\boldsymbol{\rho}\widehat{\boldsymbol{\rho}}} := \mathbb{E}_{\boldsymbol{\rho}}[\,\mathbb{E}_{\widehat{\boldsymbol{\rho}}}[f|\boldsymbol{\theta}]\,]$ is a double expectation over both.

Out of convenience, we simply extend this framework with the *product metric* for any p-norm of pairs of separate MMD

[4] We utilize $\mathcal{G}_r$ with the usual tangential metric [41, 71] and S^d_{++} with the affine-invariant metric [39, 72] for geodesic completeness.

metrics,

$$\begin{aligned}&\mathrm{pMMD}[\mathcal{H}\oplus\mathcal{Q},(\rho_{\boldsymbol{t}},\rho_{\boldsymbol{\ell}}),(\widehat{\rho}_{\boldsymbol{t}},\widehat{\rho}_{\boldsymbol{\ell}})]\\&:=\left\|\left(\mathrm{MMD}[\mathcal{H},\rho_{\boldsymbol{t}},\widehat{\rho}_{\boldsymbol{t}}]\mathrm{MMD}[\mathcal{Q},\rho_{\boldsymbol{\ell}},\widehat{\rho}_{\boldsymbol{\ell}}]\right)\right\|_p.\end{aligned}\tag{47}$$

This extension trivially satisfies the logical implications depicted in Truth Table 1 by composition with the corresponding norm of thresholds. Moreover, the use of the product metric is a somewhat natural choice by virtue of the product submanifold construction for the space of separable shape tensors. However, definition of pMMD is merely a formalism to demonstrate the existence of a logical conjunction over the separate feature tests comparing the images. In practice, we simply compute the hypothesis tests separately and draw the aggregated conclusion as a consequence of this formal construction.

Separability does not assume nor imply statistical independence of $(\boldsymbol{t},\boldsymbol{\ell})$. Distinct normal coordinates describe distinct shape features. The separable decomposition in Fig. 9 visualizes this distinction. Varying $\boldsymbol{t}$ alone modulates boundary roughness (curve regularity) while varying $\boldsymbol{\ell}$ alone modulates anisotropic size. However, physical processes—e.g. recrystallization correlations driving both grain growth and boundary roughness—can couple them. A straightforward (un-pursued) diagnostic is to test whether the joint factors as a product of marginals via $\mathrm{MMD}[\mathcal{F},\boldsymbol{\rho},\rho_{\boldsymbol{t}}\rho_{\boldsymbol{\ell}}]$ [32] (the Hilbert–Schmidt information criterion). Whether or not the joint factors, the product metric (47) remains valid and the aggregate conclusion of Truth Table 1 still holds. Decomposition simply determines whether direct comparison of the joint distributions would add explanatory content beyond the marginal tests.

The test statistic is the biased empirical estimator of [32], computed separately for undulation and scale coordinates, with $\mathcal{F}$ realized as the RKHS with the (characteristic) Gaussian kernel under the standard median pairwise-distance heuristic.

Two permutation operations enter the workflow. Cyclic Procrustes permutation (Theorem 2) acts on the rows of each landmark matrix to register every shape against the random asymmetric archetype, once and prior to testing. Sample permutations act on the ensemble labels. The $N+\widehat{N}$ pre-registered coordinate vectors for either ensemble are pooled and randomly reassigned to two groups of sizes $N,\widehat{N}$; then, MMD is recomputed. This ensemble label resampling is called *bootstrapping*. The permutation p-value is rejected at significance α when it is below $\alpha/2$ per the Bonferroni adjustment.

In summary, given ensembles of segmented curve features as data from a pair of images, separate hypothesis tests over one image $\{\boldsymbol{t}\}\sim\rho_{\boldsymbol{t}}$ versus the other $\{\widehat{\boldsymbol{t}}\}\sim\widehat{\rho}_{\boldsymbol{t}}$, and likewise $\{\boldsymbol{\ell}\}\sim\rho_{\boldsymbol{\ell}}$ versus $\{\widehat{\boldsymbol{\ell}}\}\sim\widehat{\rho}_{\boldsymbol{\ell}}$, offer logical tests to *explain* differences aggregated with the product maximum mean discrepancy (pMMD). Numerical examples are depicted and described in Figs. 16–19 emphasizing each row of the logical conjunction shown in Truth Table 1. Algorithm 1 collects the full workflow.

Algorithm 1 Pseudo-code for explainable binary classification of separable shape ensembles

Require: Curve ensembles $\{\boldsymbol{c}_q\}_{q=1}^{N}$ and $\{\widehat{\boldsymbol{c}}_q\}_{q=1}^{\widehat{N}}$; quadrature rule $\{(w_i,s_i)\}$ of cardinality n; undulation rank r; total number of bootstrap samples B, significance level α.

Ensure: Marginal and joint decisions $(H_{\boldsymbol{t}},H_{\boldsymbol{\ell}},H)$ keyed to Truth Table 1.

1: **fix reparametrization**: collocate at n quadrature nodes $\{s_i\}$ to form landmarks, e.g.,
$$\mathcal{T}[\boldsymbol{c}_q](\{s_i\}).$$
(Sections 1.5, 3.1)

2: **cyclic Procrustes**: sample random asymmetric archetype and register every landmark matrix against it *once* (Theorem 2).

3: **SRQD**: weighted polar decomposition of each registered matrix, e.g.,
$$(\widetilde{X}_q,P_q)\in Gr(d,n)\times S^d_{++}.$$
(Section 3.2, Figure 4)

4: **tPCA**: approximate Frechét means $[\widetilde{V}_0]\in\mathcal{G}_r$ and $P_0\in\mathcal{P}$, then project each $(\widetilde{X}_q,P_q)$ into an r-dimensional orthogonal basis of normal coordinates, e.g.,
$$(\boldsymbol{t}_q,\boldsymbol{\ell}_q)\in T_{[\widetilde{V}_0]}\mathcal{G}_r\times T_{P_0}\mathcal{P}.$$
(Section 3.4)

5: **pMMD**: bootstrap separate MMD values B times over $\{\boldsymbol{t}\}$ and $\{\boldsymbol{\ell}\}$.

6: **test**: approximate p-values and compare to Bonferroni at level $\alpha/2$.

7: **return**
$$(H_{\boldsymbol{t}},H_{\boldsymbol{\ell}})\in\{0,1\}^2$$

4 Numerical Experiments

We begin with a summary investigating material micrographs with EBSD followed by a study involving a more general grayscale imaging modality. Descriptions and results of EBSD numerical experiments are illustrated and explained in Figs. 16–19. The combined numerical experiments enable explainable binary classifications of ice data [5, 6] and TRIP780 steel data measured at the National Institute of Standards and Technology (NIST) Material Measurement Laboratory (MML). In aggregate, all examples span Truth Table 1.

The preprocessing of EBSD data to extract sorted vertices of segmented boundaries is accomplished with the open-source toolbox MTEX [27], version 5.10.2. In conjunction with [32], for all examples, a Gaussian kernel is utilized to approximate MMD and the corresponding scale parameter is inferred using the median Euclidean distance heuristic over a subset of normal coordinates. Coordinates are not

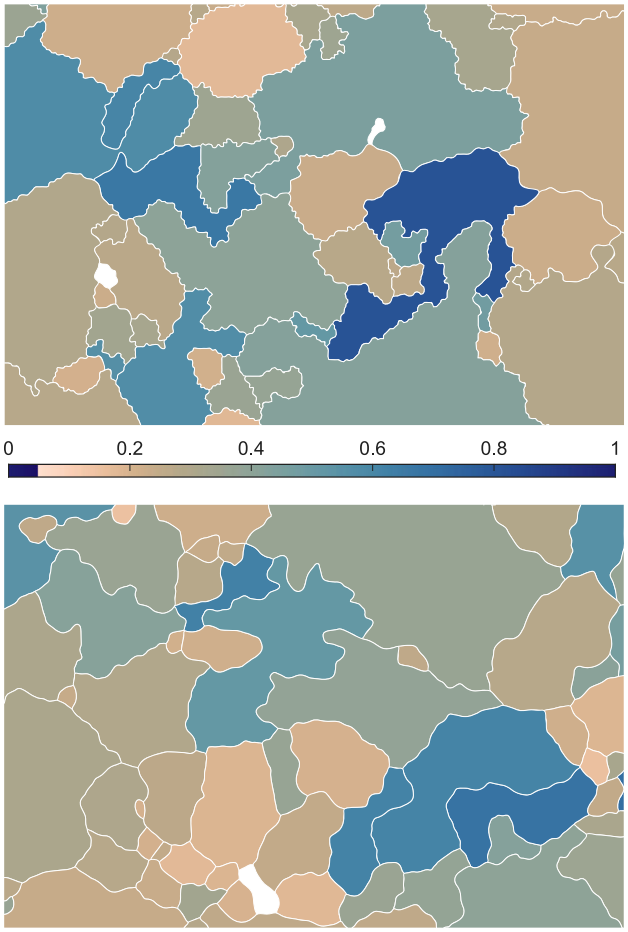

Fig. 10 (Example $0 \wedge 1$) Magnified inspection of small scale grain structures subjected to different image preprocessing. Colors of the grains correspond to the normalized distances of scale-invariant PRRTI features from the Fréchet mean over the learned Grassmannian submanifold. The top image has no smoothing applied to the input orientation map before segmentation. The bottom image utilizes increased smoothing of the boundaries according to MTEX algorithms (parameter equal to five)

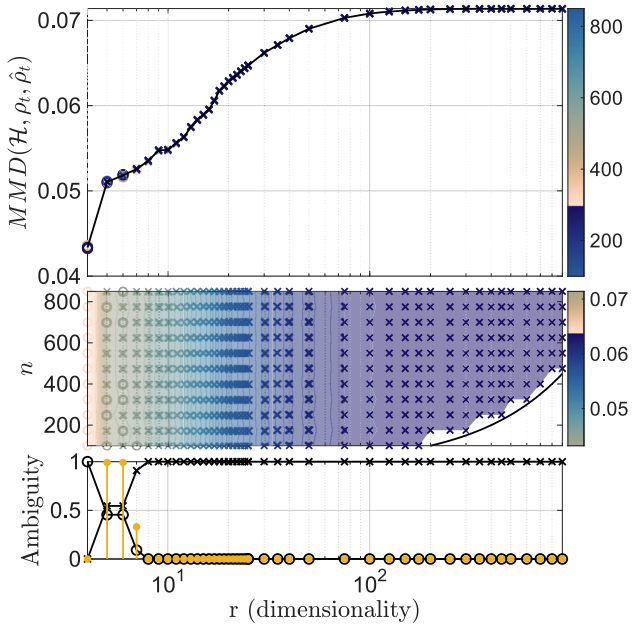

Fig. 11 (Example $0 \wedge 1$) Changes in the undulation MMD value over shape dimensionality, r, and number of quadrature nodes, n, without coordinate normalization but including cyclic Procrustes registrations. (top) Colors indicate the number of quadrature nodes while crosses correspond to rejecting the null hypothesis and circles correspond to failure to reject based on a significance level $\alpha = 0.02$. The solid black line is the conditional average of MMD over quadrature nodes. (middle) Colors indicate MMD value over both r (horizontal axis) and n (vertical axis). The black line in the bottom right corner of the contour plot represents the upper bound of the Grassmannian intrinsic dimensionality. (bottom) Ambiguity in decision-making emphasizing the proportion of failure to reject (circles) and reject (crosses) over collocation levels. The yellow stem plot represents the ambiguity as the scaled product of these two proportions

normalized—i.e., scaled by approximate components of variation inferred from a subset of coordinates—for the results in Figs. 16–19, but the effect of this (often default) coordinate normalization is studied in Fig. 12. All decisions in the following experiments are based on a significance level $\alpha = 0.02$.

All of the depicted numerical experiments utilize an undulation dimensionality of $r = 150$, $n = 500$ quadrature nodes—see Fig. 9 for a representative approximation with this dimensionality—and three (full) dimensions of scale. The filled colors of the shapes in Figs. 16–19 represent a normalized shape distance over the product submanifold measured from the approximate mean shape. Note, EBSD experiments and subsequent conclusions are predicated on the version of the MTEX software (5.10.2) used to segment and smooth grain boundaries.

These numerical experiments span the four rows of Truth Table 1 and emphasize four scenarios where the method delivers benefits to practitioners investigating EBSD micrographs:

i) $(1 \wedge 1)$, Fig. 16: no over-sensitivity to statistical rejection—no useful discrepancy in undulation or scale.
ii) $(1 \wedge 0)$, Fig. 17: conclusions consistent with human observation—visible grain size differences yield a scale-only rejection.
iii) $(0 \wedge 1)$, Fig. 18: sub-visual discrepancy detection—boundary smoothing, a preprocessing step ubiquitous in EBSD (and in image segmentation pipelines more broadly), induces undulation-only differences invisible to cursory human inspection that are nonetheless flagged as a statistically significant rejection. This sensitivity is among the method's most compelling demonstrations.
iv) $(0 \wedge 0)$, Fig. 19: detection of a faulty measurement—an out-of-focus instrument is flagged by simultaneous rejection on both factors.

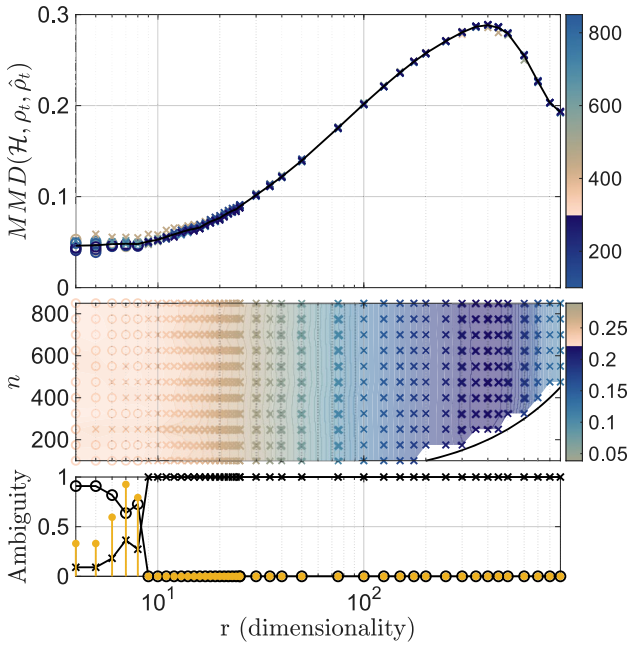

Fig. 12 (Example $0 \wedge 1$) Changes in the undulation MMD value over shape dimensionality, r, and number of quadrature nodes, n, with coordinate normalization. Shape data are not registered with cyclic Procrustes—i.e., data represent random row-wise permutations. Description is consistent with Fig. 11. Note the distinct MMD scales compared to Fig. 11

The $(0 \wedge 1)$ case of Fig. 18 was designed by taking case $(1 \wedge 1)$ and processing the same boundaries with and without smoothing; Fig. 10 magnifies the difference. The null is rejected at $\alpha = 0.02$ despite the discrepancy being largely invisible to cursory inspection, which makes the method useful for *detecting the presence of any preprocessing or synthetic generation of data* that hand-picked features would miss.

4.1 Decision Landscapes

The explainable binary classification in all cases is predicated on the selected shape dimensionality, r, number of quadrature nodes, n, and kernel heuristics like scale parameter selection and coordinate normalization. However, given the computational efficiency of the method, we can easily motivate parameter studies to understand the *decision landscape* over either set of separate coordinates. That is, decision landscapes are the primary diagnostic for selecting the quadrature size n and the undulation rank r. We focus on determining r and n empirically in the case $(0 \wedge 1)$ for decisions over Grassmannian undulations, $\text{MMD}(\mathcal{H}, \rho_t, \widehat{\rho_t})$.

Spectral convergence of Nyström quadrature on smooth closed curves means a modest n suffices; Figs. 11 and 12 confirm decisions are stable across $n \in [100, 500]$, so n should simply be large enough to approximate PRRTI features of the worst-case boundary in the ensemble. The rank r is more consequential but should *not* be chosen to maximize variance coverage as in a traditional PCA. Smaller r acts as a *dimension reduction* that regularizes shapes against noisy undulations (Fig. 9), which is often desirable [75]. We instead study decision variability directly via the *ambiguity* (48)—a scaled Bernoulli variation in the rejection proportion conditioned on r—and choose r with limited ambiguity so that decisions are stable against the chosen regularization level.

As an example decision landscape, we reuse PIL184 data from the designed case $(0 \wedge 1)$ of Fig. 18 with different image processing but now include duplicate grain shapes in the comparison. In other words, the first set of undulation coordinates $\{\boldsymbol{t}\}$ are computed without smoothing while the second set of undulation coordinates $\{\widehat{\boldsymbol{t}}\}$ are computed using a nominal level of smoothing—MTEX smoothing parameter equal to five. Duplicate grain shapes are included—i.e., comparing two identical sample sets of PIL184 boundaries with $N = 933$—up to smoothing thus emphasizing discrepancies attributed entirely to boundary smoothness while eliminating variability due to partitioning the imaging data as in Fig. 18.

Figures 11 and 12 depict *decision landscapes* over changing shape dimensionality and quadrature nodes. Figure 11 utilizes cyclic Procrustes registrations against a random archetype from the full ensemble but does not include (normal) coordinate normalization. In contrast, Fig. 12 depicts results obtained by scaling normal coordinates but randomly permuting rows of the discrete grain shapes. Coordinate normalization is achieved by scaling all coordinates with the approximated component-wise variation in the noisy samples, $\{\boldsymbol{t}\}$.

Given the changing decision landscape, it is helpful to introduce *ambiguity* as scaled Bernoulli variance,

$$4\mathbb{P}_n(H = 1|r)(1 - \mathbb{P}_n(H = 1|r)), \tag{48}$$

where $H \in \{0, 1\}$ is a Bernoulli trial representing the binary decision contrasting distributions of undulation coordinates, $\boldsymbol{t}, \widehat{\boldsymbol{t}} \in T_{[\widetilde{V}_0]}\mathcal{G}_r$, and $\mathbb{P}_n(\cdot|r)$ is the conditional probability of failing to reject the null hypothesis at a particular shape dimensionality. The probability is estimated as the sum of binary FTR conditions over the total number of quadrature discretizations at the corresponding level of r (dimensionality). By definition, an ambiguity of one corresponds to an entirely random guess which should not be trusted in practice.

Examining Fig. 11, empirical regularization over reduced shape dimensionality—as described in Fig. 9—results in no statistically significant differences between the distributions of undulation when $r \leq 7$. In other words, without sufficient dimensionality representing undulations, empirical regularization results in no significant discrepancy. Moreover, there is very limited variability in MMD values over different

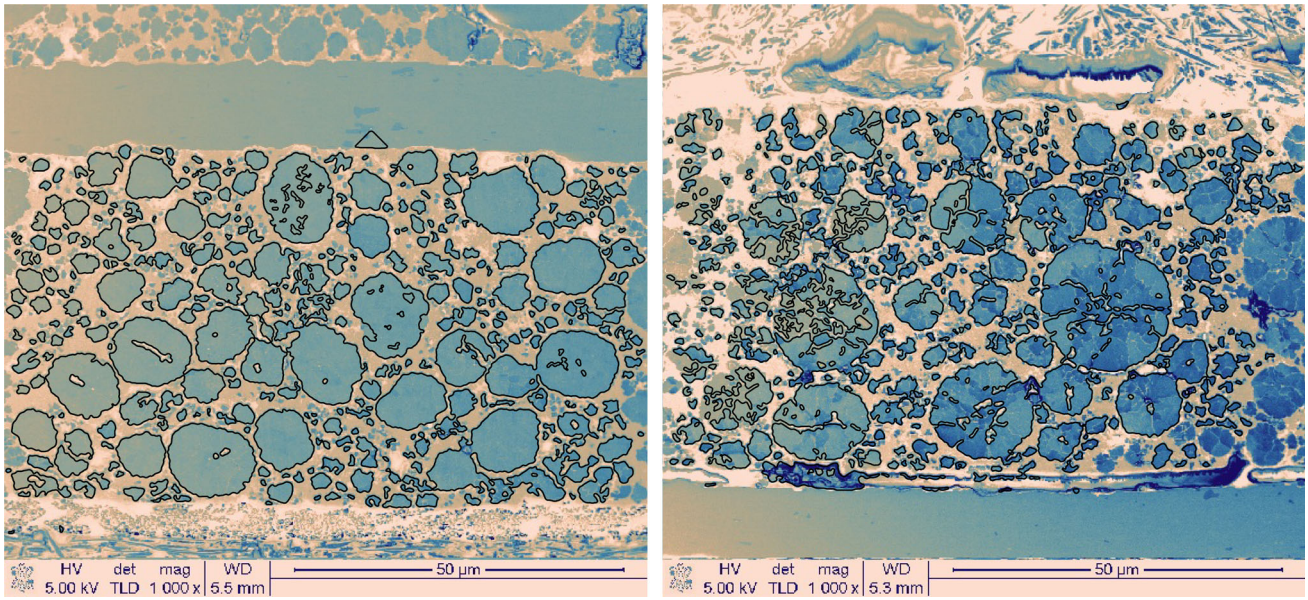

Fig. 13 (top) Cross section of a 'new' condition lithium-ion battery with segmented black curves over scalar-valued image intensity. (bottom) Cross section of a 'used' condition lithium-ion battery with segmented curves (black boundaries) over scalar-valued image intensity. The new condition battery exhibits less fracturing in the NMC particles compared to the battery which has been subjected to accelerated cycling in the lab. A smoothing length scale $\sigma = 5$ is used to extract the depicted curves

quadrature collocations. This is a testament to the stability of the cyclic Procrustes random archetype registrations over changing n. At $r = 8$ and beyond, the binary decision consistently rejects the null hypothesis indicating statistically significant differences in undulation. Ambiguity is nonzero over $5 \leq r \leq 7$ and peaks at $r = 5, 6$ with maximum ambiguity of approximately one.

Examining Fig. 12, coordinate normalization also appears to modulate variability in approximated MMD values despite randomly permuted quadrature nodes. The only notable (visible) conditional variation in MMD value occurs over small shape dimensionality, $r \leq 20$. We reject the null hypothesis consistently beyond $r = 8$ while $r = 7$ corresponds to maximum ambiguity of approximately one. Surprisingly, MMD values in Fig. 12 do not exhibit the anticipated monotonicity over r as in Fig. 11. This may suggest that increased dimensionality can promote cancelations in discrepancy given otherwise random curve alignments unlike the cyclic Procrustes alignments of Fig. 11 exhibiting cumulative (monotonic) discrepancy over dimensionality.

The global trends between the two decision landscapes are distinct but the ability to circumvent row-wise registrations by simply normalizing coordinates could offer massive computational advantages—i.e., *optimal alignments may be less important than previously hypothesized if utilizing coordinate normalization*. In practice, decision landscapes should be utilized to understand sensitivities to decision-making in any replicated or extended work.

4.2 Segmentation Efficacy

Across imaging applications, feature segmentation is a challenging step in analysis workflows. Despite modern advancements, segmentation methods remain highly dependent on multiple factors, including image quality and hyperparameters—e.g., smoothing parameters, thresholding values, training with hand-labeled data, etc. Variability in segmentation algorithms can result in different ensembles of shapes extracted from a given image, which can complicate subsequent assessments of patterns in the image. Here, we briefly examine the utility of explainable binary classification for introducing novel, quantifiable metrics to compare and contrast the efficacy of a given image segmentation.

The use-case for this study is motivated by the need to characterize degradation in lithium-ion batteries to better understand battery aging and performance decay. Improved degradation models can lead to improved battery design, charging, and replacement strategies. Scanning electron microscope (SEM) imaging is often used to enabling high-resolution visualizations of batteries throughout aging. In particular, fracturing and formation of fissures in the nickel-manganese-cobalt (NMC) particles visible in SEM are notable markers of this decay. Robust analysis of the degradation process relies on meaningful characterization of the damage to these NMC particles, making this an appropriate test case of studying the interplay between segmentation approaches and explainable binary classification.

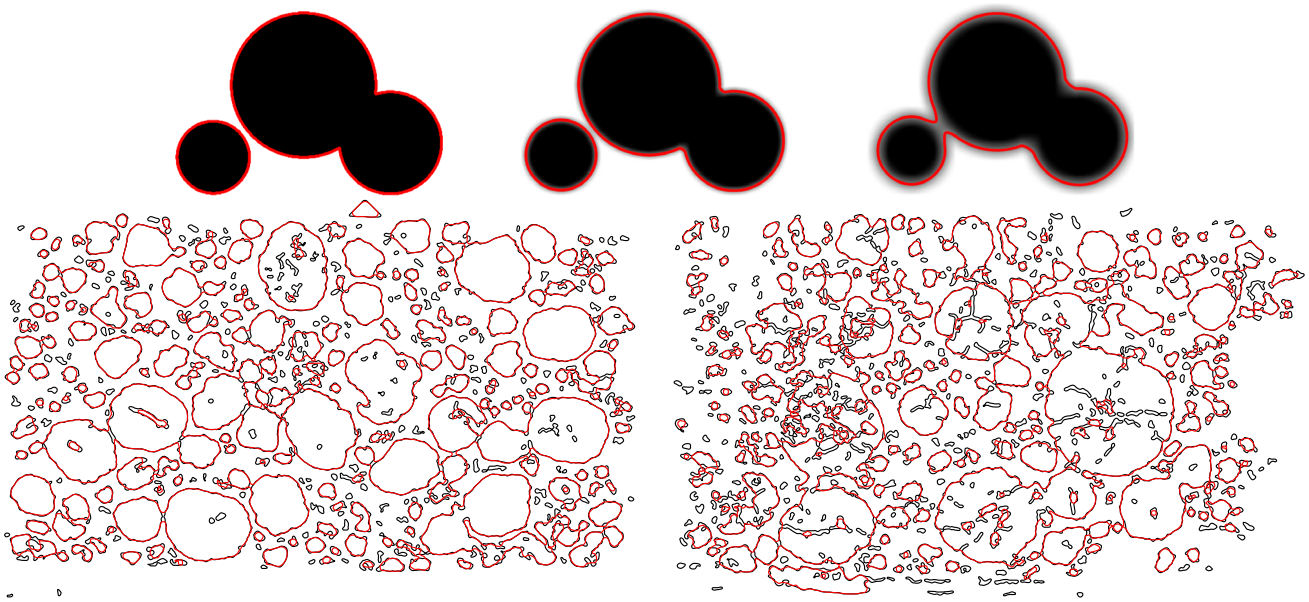

Fig. 14 (top) Effect of increased Gaussian length scale convolutions applied to a simple binary mask. From left to right, the scale is increased as $\sigma = 1, 15, 30$ over a 400-by400 pixel grid—i.e., the relative size of these simple circular objects versus the image resolution is much greater than the NMC particles in an SEM. At larger scales, increased smoothing begins to blur and merge boundaries reducing the ensemble size. (bottom) For reference, ensembles from the same pair of images in Fig. 13 are segmented at $\sigma = 5$ (black boundaries) and $\sigma = 15$ (red boundaries). Clearly, increased smoothing begins to obscure the important distinguishing fractures and fissures of the 'used' condition battery

To quantify the nature of the fractures and fissures which form during lithium-ion battery degradation, we run simple morphological segmentation over SEM images for three 'new' condition battery and three 'used' condition battery. An example pair of new and used condition image segmentations are shown in Fig. 13. Each image is 1024-by-1024 and down-selected to a subregion of interest containing the NMC particles.

The output of these segmentations is a binary mask which is then smoothed with a Gaussian convolution corresponding to a parametric length scale, σ, to remove pixel 'stepping.' The length scale is sized according to the number of pixels and we use the 0.5 level-set of the smoothed mask to extract corresponding boundary curves. Figure 14 depicts the effect of smoothing the binary mask to reduce noisy pixel stepping in the segmentation.

Notice the segmentation of any given image is imperfect—some particle boundaries are inappropriately merged, others are omitted entirely or partially, and artifacts are present. However, in aggregate, the ensemble of extracted curves appear to represent distinct patterns which admit a seemingly obvious dependency on the fractured state of the NMC particles. The goldilocks paradigm of interest: *How much smoothing is required to denoise and achieve consistent decision-making prior to excessively blurring the mask?* Too little smoothing and high-frequency undulations may obscure relevant decision making. Too much smoothing and we obfuscate the segmentation results.

We apply explainable binary classification to the separable shape ensembles extracted from simple, smoothed morphological segmentations over three pairs of new and used battery SEM images—i.e., six SEM images in total. Shape dimensionality, r, is varied from 5 to 100 in increments of 5, (a separate view of the data emphasizes converged MMD values at approximately $r = 50$) quadrature levels n are taken sparsely at 100, 500, 1000 given limited variability in MMD values over n, and we sweep over the convolution length scale, $\sigma = 0, \ldots, 30$, in increments of one. The undefined value $\sigma = 0$ corresponds to simple black-white boundary extraction without smoothing applied to the binary mask.

Approximated MMD results depicted and aggregated as conditional averages at corresponding smoothing scales are visualized in Fig. 15. We notice small amounts of ambiguity initially with steadily increasing discrepancy. Over the range $4 \leq \sigma \leq 11$, all decisions result in a consistent rejection of the null hypothesis. Finally, at $\sigma = 13$, the decision landscape inverts presumably due to the increasingly merged and smoothed shape objects. As anticipated by the simple example in Fig. 14, ensemble sample sizes are shown to steadily decrease with increased smoothing length scales. MMD also appears to increase again at the increasingly smaller ensemble sizes—i.e., the few remaining shapes are distinct but the distributions of normal coordinates are no longer sig-

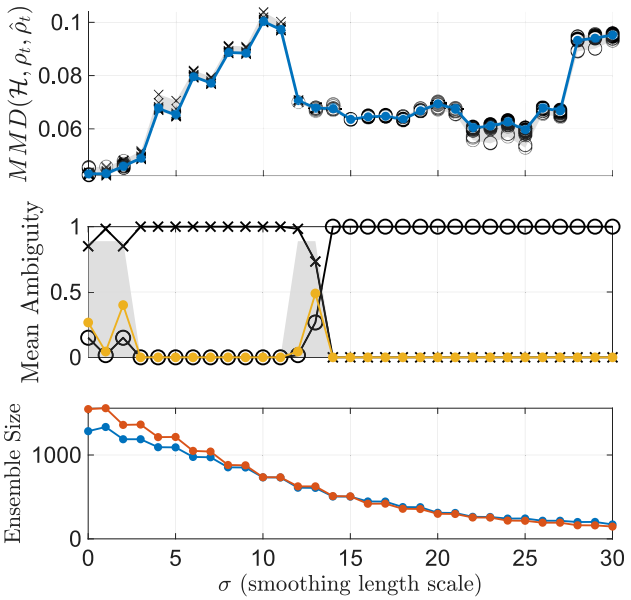

Fig. 15 (top) Conditionally averaged undulation MMD over smoothing scales applied to the segmentation binary mask. Raw data are depicted with crosses or circles depending on the decision to reject or failure to reject the null hypothesis, respectively. (middle) Mean ambiguity over dimensionality at the corresponding convolution length scale. The gray shaded region is a min-max envelope over all combinations of r and n. (bottom) Changing ensemble sizes from three segmented images of 'new' condition batteries (blue) versus three segmented images of 'used' condition batteries (orange) over smoothing length scales

nificantly discrepant. Our intuition beyond $\sigma = 13$ is that the ensemble has digressed into a sparse collection of seemingly random undulations with increasing discrepancy, albeit insignificant, due to the reduced ensemble size.

In practice, this analysis would suggest that achieving the largest number of segmented shapes in an ensemble with consistent decision-making occurs near $\sigma = 4$ or 5. This study could be used to set parameter ranges for more effective segmentation when detecting failure modalities in subsequent battery experiments. It appears that $r \approx 50$, $n \approx 500$, and $\sigma \approx 5$ are suitable candidates to detect the fractured nature for this level of battery degradation while maximizing the number of segmented objects.

4.3 Computational Burdens

We report timings in Table 2 using subsets of curves segmented from the battery data. The SRQD column subsumes the arc-length reparametrization and cyclic Procrustes alignment it invokes internally. The pMMD two-sample test is evaluated on an even split ($N/2$ versus $N/2$) with $B = 1000$ bootstrap permutations. Two of the three stages are reported in two configurations. The *sequential* single-threaded reference exposes clean scaling laws on the algorithm itself. The *parallel* configuration follows the default Python implementation, namely multiprocessing for SRQD's per-curve work and Numba's parallel JIT for the pMMD permutation loop. Measurements are taken using a Python implementation at $n = 500$ landmarks and $r = 50$ reduced dimensions, run on a 16-core / 32-thread x86-64 desktop CPU (up to 4.3 GHz boost) with 64 GB of system memory under Windows 11 and Python 3.13.

Table 2 Wall-clock timings (seconds) for the principal pipeline stages as the ensemble size N increases, reported side by side for sequential (seq., single-threaded) and parallel (par.) configurations of a Python implementation

N (total)	SRQD (seq./par.)	tPCA (seq./par.)	pMMD (seq./par.)
100	**0.43**/2.29	0.03/–	0.07/ <**0.01**
500	**1.93**/2.43	0.15/–	0.93/**0.03**
1000	3.74/**2.52**	0.36/–	4.92/**0.07**
2000	7.41/**3.01**	0.66/–	31.91/**0.36**
η_{seq}	0.96/ 0.08	1.10/ –	2.09 / 1.61

The SRQD column subsumes the first three steps of Algorithm 1, and the tPCA column does not parallelize. The final row reports the empirical power-law exponent η_{seq} fit to $time(N) \approx C\,N^{\eta}$ via ordinary least squares. Values near unity indicate linear scaling in ensemble size

In the sequential configuration, the aggregated per-curve stages labeled 'SRQD' (planar reparametrization, cyclic Procrustes, and SRQD) scale linearly ($\eta \approx 1$) and the pMMD bootstrap scales quadratically ($\eta \approx 2$) because each of the $B = 1000$ permutations touches the pooled $N \times N$ Gaussian kernel matrix. The per-curve stages and pMMD permutations are embarrassingly parallel. The sequential SRQD slope of 0.96 reflects a small-N transient at $N = 100$, and refitting on $N \geq 250$ (not shown) recovers $\eta = 1.00$. The parallel SRQD column carries an approximately three-second multiprocess pool initialization cost. This cost dominates the small-N entries and is responsible for the column remaining nearly flat through $N = 1000$. Among the per-curve stages aggregated into the SRQD column, cyclic Procrustes contributes a low constant factor courtesy of its FFT-based implementation, and is not separately accelerated.

Numba's parallel JIT drops the $N = 2000$ pMMD entry from 31.9 s to 0.36 s and removes it from the critical path. The tPCA step contains an inherently sequential outer iteration because Karcher's algorithm to approximate the Fréchet mean is iterative—details in Appendix A and [34]. However, across both configurations tPCA is among the cheapest columns at every N. The dominant column at large N is pMMD in the sequential configuration and SRQD in the parallel configuration. Full software packages in MATLAB and Python are forthcoming.

5 Conclusion

We established a formal connection between separable shape tensors and the dual evaluation functionals of [33], developed a spectral (Nyström) approximation of SSTs that is accurate for sufficiently regular curves, and combined the resulting features with cyclic Procrustes alignment and product-MMD hypothesis testing. Four EBSD examples spanning Truth Table 1 and a lithium-ion battery SEM segmentation study illustrate the framework end to end.

Future work will expound on the identified relationships between finite matrix manifold representations of shape and the formal interpretations of infinite-dimensional analogs detailed in [33]. In short, our efforts will be focused toward: i) formalizing bounds on the approximation error of reduced dimension SSTs and ii) infinite-dimensional extensions of pMMD explainable hypothesis testing over the $\mathbb{R}^d$-valued RKHSs of [33]. We also hope to improve explanations to understand which patterns of shape, in more detail, contribute to approximated discrepancies.

Appendix A SST Review

This appendix collects supplementary derivations and algorithmic context underlying the separable shape tensor representation (3) of Sect. 1.4 and the tangent PCA review of Sect. 3.4. More details about these developments can be found in related work [34].

Landmark configurations as elements of a non-compact Stiefel manifold. The matrix $X = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)^\top$ built from landmarks (1) is assumed to be an element of the non-compact Stiefel manifold $\mathbb{R}^{n\times d}_*$ [37]. That is, we assume $X \in \mathbb{R}^{n\times d}_*$ to exclude the degenerate cases of points or lines in $\mathbb{R}^d$ as objects representing curves [34]—necessitating that X is full rank. In other words, *we assume the component functions* $\boldsymbol{c} = (c_1, \ldots, c_d)$ *of the curve are linearly independent*.

In practice, we are often only given sequences of landmarks $(\boldsymbol{x}_i)$ and have no knowledge of the 'true' underlying $\boldsymbol{c}$ nor the parameter s beyond assuming the identification (1). However, this characterization of the data implies landmarks are ordered according to some orientation of the unknown curve thus constituting a sequence, $(\boldsymbol{x}_i)$, encoded over the row index of X. Thus, it is reasonable to build a variety of interpolations and approximations—e.g., piecewise spline interpolations over chordal parametrization—which can be shown to converge to the unknown $\boldsymbol{c}$ with an increasing number of given landmarks [34, 38].

Linear deformations and undulation as a quotient. Given the SVD (2), we can elaborate on the decomposition into complementary factors of (discrete) undulation and invertible d-by-d linear deformations—i.e., in a complementary sense, an 'undulation' is the set of all shape representations which cannot be achieved by linear transforms as right group actions on $X - B(X)$, with $B(X) = (1/n)\mathbf{1}_{n,n}X$ the unbiased centering. Ignoring non-deforming translations, for $\boldsymbol{c}(s) = M^\top \widetilde{\boldsymbol{c}}(s)$,

$$\begin{aligned} X^\top &= (\boldsymbol{c}(s_1), \ldots, \boldsymbol{c}(s_n)) \\ &= M^\top (\widetilde{\boldsymbol{c}}(s_1), \ldots, \widetilde{\boldsymbol{c}}(s_n)) \\ &= (\widetilde{X}M)^\top \end{aligned} \tag{A.1}$$

where the right action of $M \in GL(d, \mathbb{R})$ on $\widetilde{X} = (\widetilde{\boldsymbol{c}}(s_1), \ldots, \widetilde{\boldsymbol{c}}(s_n))^\top \in \mathbb{R}^{n\times d}_*$ is akin to linear deformations of the preshape. Given (centered) X, we seek a decomposition $X = \widetilde{X}M$ to describe how X *undulates* as $[\widetilde{X}]$, an equivalence class modulo right group action over $GL(d, \mathbb{R})$, versus how it linearly deforms over subsets of $M \in GL(d, \mathbb{R})$. This motivates the identification of discrete shape *undulation* as $[\widetilde{X}] \in Gr(d, n)$ with $Gr(d, n) \cong \mathbb{R}^{n\times d}_*/GL(d, \mathbb{R})$ [37].

Two landmark standardizations (affine vs polar). Two geometric interpretations of the SVD (2) are called *landmark standardizations* [10, 34, 35] to compute *representative undulations* $\widetilde{X}$ (up to rotations and reflections):

- The affine standardization [35] $X = \widetilde{X}M$ such that $\widetilde{X} := V$ and $M := \Sigma U^\top \in GL_+(d, \mathbb{R})$.
- The polar standardization [34] $X = \widetilde{X}P$ such that $\widetilde{X} := VU^\top$ and $P := U\Sigma U^\top \in S^d_{++}$.

The polar choice is adopted throughout this paper because it separates undulation from scale variations independent of rotations and reflections: following [39], S^d_{++} is identified with $GL_+(d, \mathbb{R})/SO(d)$, so that $[\widetilde{X}]$ encodes nonlinear undulating features over $\mathbb{R}^{n\times d}_*/GL(d, \mathbb{R})$ while P encodes generalized (anisotropic) *scales* over $GL_+(d, \mathbb{R})/SO(d)$. This quotient reading is the geometric justification for the separable shape tensor (3) introduced in the main body.

Tangent PCA algorithmic details. Section 3.4 states the tangent PCA (tPCA) procedure over $\mathcal{G}_r \subseteq Gr(d, n)$ and $\mathcal{P} \subseteq S^d_{++}$ at a high level; we collect here the algorithmic components from [34, 39–41, 71] that inform the *discretize→test* pipeline. Given a sample $\{\boldsymbol{p}_k\}_{k=1}^N$ on a Riemannian manifold (either $Gr(d, n)$ or S^d_{++}), the intrinsic (Fréchet) mean $\boldsymbol{p}_0$ is computed by the Karcher iteration [73]

$$\boldsymbol{p} \leftarrow \mathrm{Exp}_{\boldsymbol{p}}\left(\frac{1}{N}\sum_{k=1}^{N}\mathrm{Log}_{\boldsymbol{p}}(\boldsymbol{p}_k)\right), \quad \text{until}$$

$$\left\| \frac{1}{N} \sum_{k=1}^{N} \mathrm{Log}_{\boldsymbol{p}}(\boldsymbol{p}_k) \right\| < \epsilon, \tag{A.2}$$

and the reduced principal directions in $T_{\boldsymbol{p}_0}M$ are obtained from the thin SVD,

$$\sqrt{N-1}T := \left[\mathrm{vec}(\mathrm{Log}_{\boldsymbol{p}_0}(\boldsymbol{p}_1)), \ \ldots, \ \mathrm{vec}(\mathrm{Log}_{\boldsymbol{p}_0}(\boldsymbol{p}_N)) \right], \tag{A.3}$$

$T \overset{SVD}{=} U\Sigma V^\top$. retaining the first r left singular vectors U_r as an orthonormal basis of $T_{\boldsymbol{p}_0}\mathcal{G}_r$ (or, for S^d_{++}, the full three-dimensional span at P_0); the columns of $U_r^\top \boldsymbol{N} = \Sigma_r V_r^\top$ are the normal coordinates. The map $\mathrm{vec} : \mathbb{R}^{n\times d} \to \mathbb{R}^{nd}$ column-stacks a tangent matrix; since $\mathrm{tr}(A^\top B) = \mathrm{vec}(A)^\top \mathrm{vec}(B)$, the SVD (A.3) preserves the Frobenius inner product induced on the tangent space, and the reshape vec^{-1} recovers matrix-valued directions for composition with $\mathrm{Exp}_{\boldsymbol{p}_0}$. Over S^d_{++}, vec stores only the $d(d+1)/2$ unique symmetric entries.

The intrinsic maps on $Gr(d, n)$ are given with Stiefel representatives [41, 71]: For a direction $\Delta \in T_{[\widetilde{X}]}Gr(d, n)$ with thin SVD $\Delta = \widehat{U}\widehat{\Sigma}\widehat{V}^\top$, the Grassmann exponential is

$$\mathrm{Exp}_{[\widetilde{X}]}(\Delta) = \left[\widetilde{X}\widehat{V}\cos(\widehat{\Sigma})\widehat{V}^\top + \widehat{U}\sin(\widehat{\Sigma})\widehat{V}^\top \right], \tag{A.4}$$

and, for a second representative $\widetilde{Y}$ with projection $(I_n - \widetilde{X}\widetilde{X}^\top)\widetilde{Y}(\widetilde{X}^\top\widetilde{Y})^{-1} = \bar{U}\bar{\Sigma}\bar{V}^\top$, the Grassmann logarithm is

$$\mathrm{Log}_{[\widetilde{X}]}([\widetilde{Y}]) = \bar{U}\ \arctan(\bar{\Sigma})\ \bar{V}^\top. \tag{A.5}$$

Each map requires a single thin SVD, so the complexity is $O(nd^2)$—*linear in n for fixed ambient dimension d*, a key enabler of the ensemble-scale computations reported in Sect. 4. On S^d_{++} with the affine-invariant metric [39, 72], the corresponding maps reduce to matrix exponentials and logarithms, $\mathrm{Exp}_P(S) = P^{1/2}\exp\big(P^{-1/2}SP^{-1/2}\big)P^{1/2}$ and $\mathrm{Log}_P(D) = P^{1/2}\log\big(P^{-1/2}DP^{-1/2}\big)P^{1/2}$.

One caveat of the Grassmann logarithm (A.5) is that $\mathrm{Exp}_{[\widetilde{X}]} \circ \mathrm{Log}_{[\widetilde{X}]}(\widetilde{Y}) = \widehat{Y}$ with $[\widehat{Y}] = [\widetilde{Y}]$ but with a possible rotational mismatch between representatives; a modified logarithm in [71] (Alg. 5.3) removes this mismatch at the cost of a second SVD. In this work, the mismatch is irrelevant for computing normal coordinates but motivates the cyclic Procrustes registration of Sect. 3.3 when a fixed representative is needed across the ensemble. A product manifold geodesic $(\mathrm{Exp}_{1,[\widetilde{V}_0]}(t\Delta), \mathrm{Exp}_{2,P_0}(tS))$ on $\mathcal{G}_r \times \mathcal{P}$ then parametrizes the separable shape tensor (3) componentwise, consistent with the Riemannian-product construction [34].

Appendix B EBSD Examples

This appendix contains high-resolution images for the numerical experiments discussed in the numerical experiments section of the paper. Captions contain relevant descriptions of results spanning Table 1.

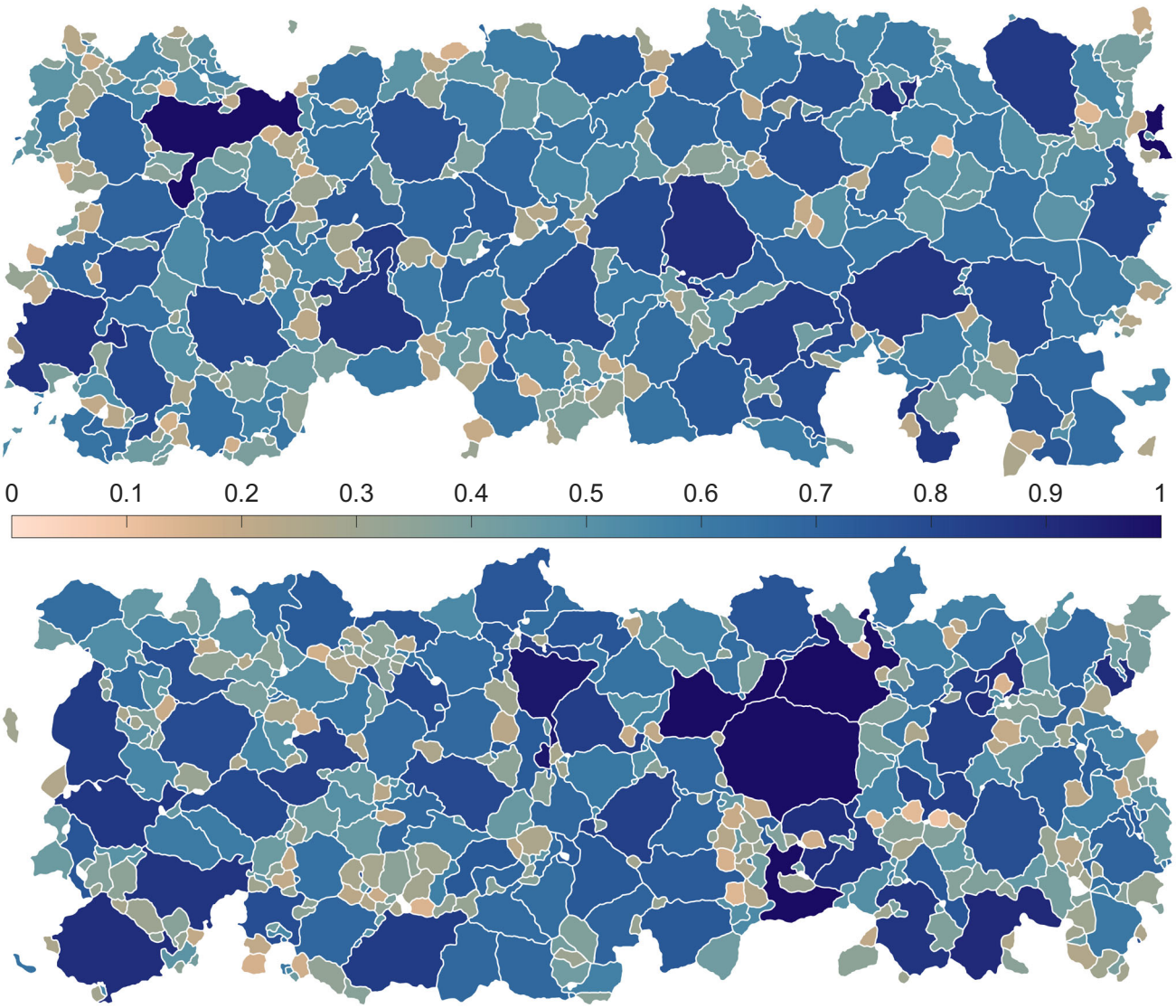

Fig. 16 (Example $1 \wedge 1$) Comparison of locally partitioned subsets of a larger EBSD images from a common ice sample [5] (PIL184) with $N = 933$ total curves—i.e., the set of curves is split approximately in half. No grains are duplicated between the partition in this example. The approximate inferences suggest the same undulations and same scales with significance level of 1%. This constitutes an example of the first row of Truth Table 1, i.e., $1 \wedge 1$

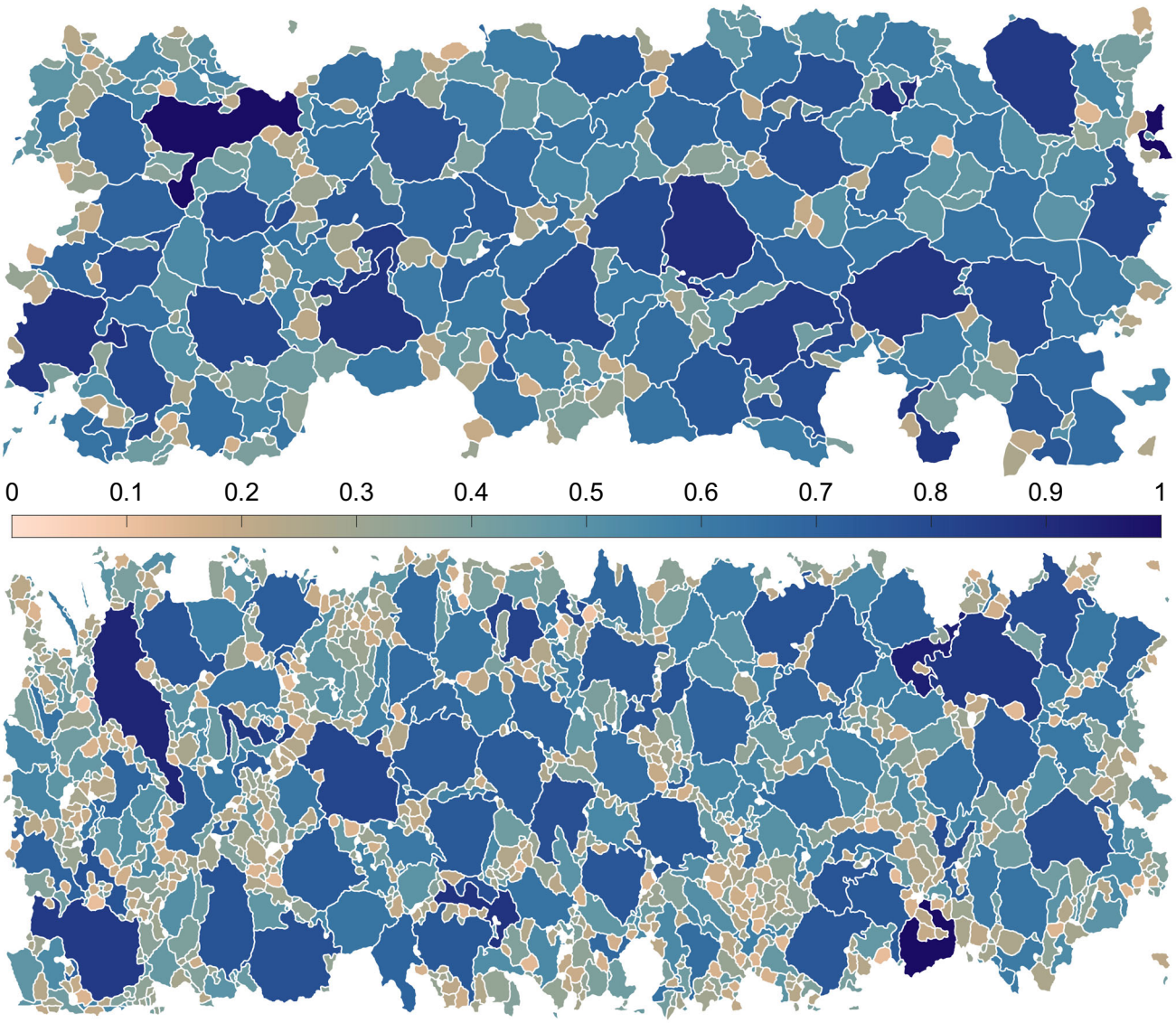

Fig. 17 (Example $1 \wedge 0$) Comparison of two different EBSD images of ice samples [5] (PIL184 and PIL185) with $N = 2566$ curves in total. Both images depict cropped shape selections. (there are additional shapes beyond the selection depicted in the images) The approximate inferences suggest the same undulations but different scales between the two images with significance level of 1%. This constitutes an example of the second row of Truth Table 1, i.e., $1 \wedge 0$

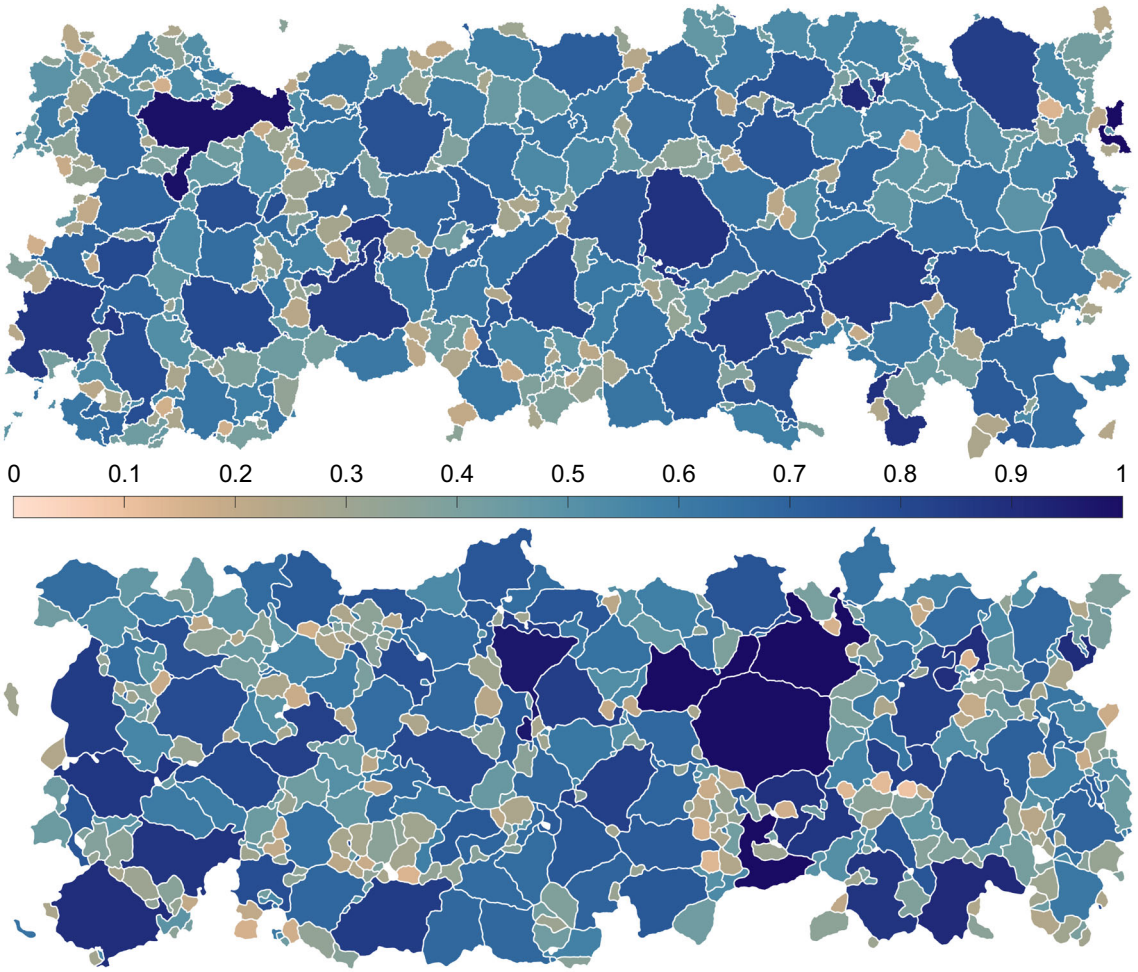

Fig. 18 (Example $0 \wedge 1$) Comparison of locally partitioned subsets of a larger EBSD images from a common ice sample [5] (PIL184) with $N = 933$ total curves. No shapes or boundaries are duplicated between the partition and this partition is consistent with that of Fig. 16. However, using a default MTEX 'smooth' routine [26], the upper partition (top image) is not smoothed while the bottom is smoothed to an extent (MTEX smoothing parameter equal to 5). Magnifying and contrasting the top image with the bottom image of Fig. 17 it is possible to see small differences in the nonlinearities of shapes. The difference is nearly visually indistinguishable. Yet, the approximate inferences suggest differences in undulations and common scales. This constitutes an example of the third row of Truth Table 1, i.e., $0 \wedge 1$, with significance level of 1%. We note, given the partition into distinct grain subsets and subsequently reduced sample size, this decision landscape has a more volatile ambiguity profile than comparing identical grains up to smoothing

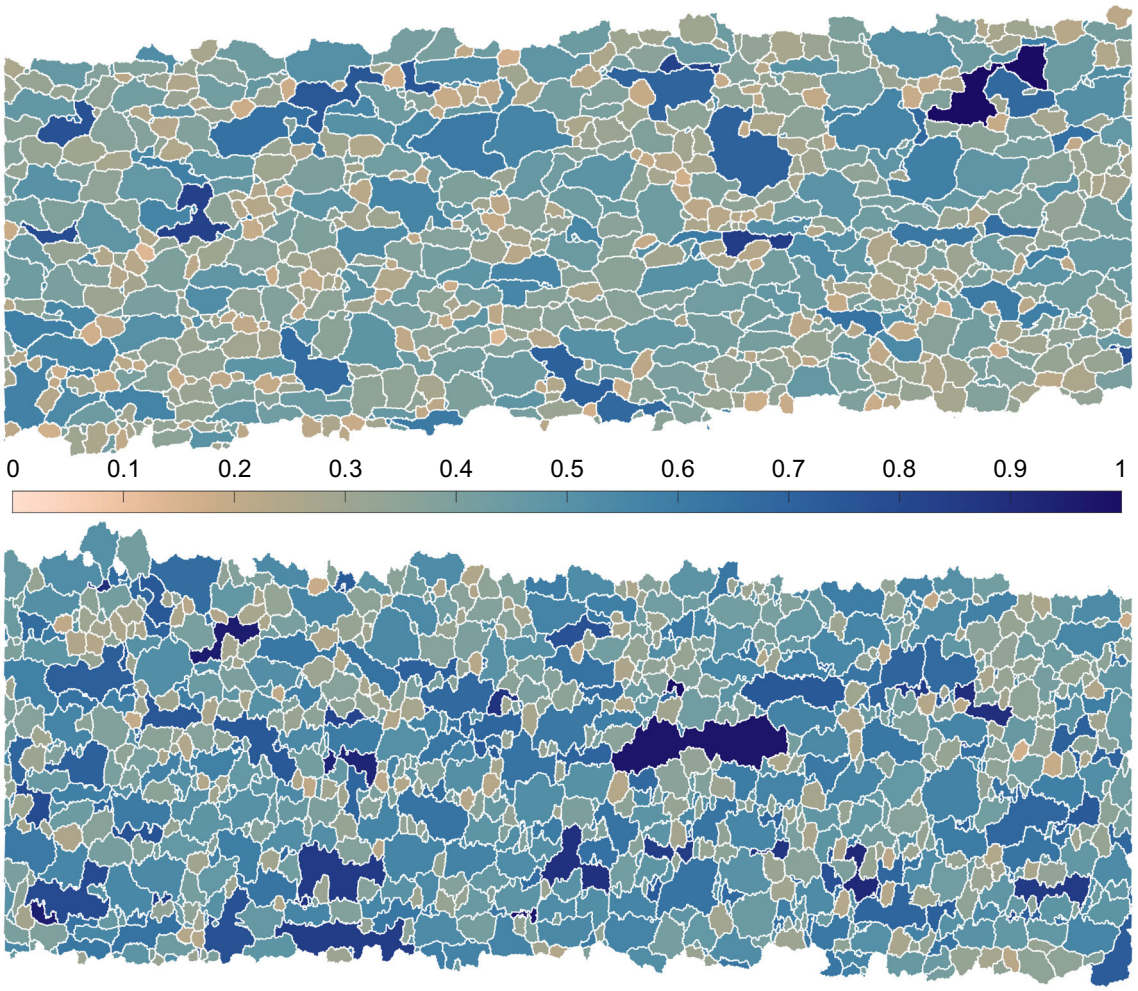

Fig. 19 (Example $0 \wedge 0$) Comparison of two different EBSD images of TRIP780 steel samples with $N = 5605$ curves in total. Both images depict cropped shape selections to place emphasis on small scale features of undulation. The approximate inferences suggest differences in both undulations and scales between the two images with significance level of 1%. This constitutes an example of the last (fourth) row of Truth Table 1, i.e., $0 \wedge 0$. Note, the hypothesis from a NIST materials scientist, Adam Creuziger, is that the bottom sample was collected from a measurement which was 'out-of-focus'

Acknowledgements We greatly appreciate Adam Creuziger and Günay Doğan at NIST for their guidance regarding applications and important data considerations. The views expressed in the article do not necessarily represent the views of the DOE or the US Government. This work is US Government work and not protected by US copyright. The US Government retains and the publisher, by accepting the article for publication, acknowledges that the US Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for US Government purposes. Certain equipment, instruments, software, or materials are identified in this paper in order to specify the experimental procedure adequately. Such identification is not intended to imply recommendation or endorsement of any product or service by NIST, nor is it intended to imply that the materials or equipment identified are necessarily the best available for the purpose.

Author Contributions Z.G. and N.F. made equal contributions to the theoretical work and preparation of the manuscript text. A.G. facilitated data collection, the application of maximum mean discrepancies, and provided revisions the text and visualizations.

Funding This work was partially support by a National Institute of Standards & Technology (NIST) Building the Future (BTF) internal funding initiative and the second author is supported in part by funds from the National Science Foundation (NSF) Research Training Group (RTG) grant (DMS-2136228). This work was authored in part by the National Laboratory of the Rockies for the U.S. Department of Energy (DOE), operated under Contract No. DE-AC36-08GO28308. This work was supported by the U.S. Department of Energy Office of Critical Minerals and Energy Innovation.

Data Availability Open-source ice micrograph data are available at: https://doi.org/10.6084/m9.figshare.13456550. The MTEX open-

source EBSD processing software is available at: https://mtex-toolbox.github.io/.

Declarations

Conflicts of Interest/Competing Interests All authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest in the subject matter or materials discussed in this manuscript.

Competing interests The authors declare no competing interests.

References

1. Grey, Z.J., Constantine, P.G.: Active subspaces of airfoil shape parameterizations. AIAA J. **56**(5), 2003–2017 (2018). https://doi.org/10.2514/1.J056054
2. Doronina, O.A., Grey, Z.J., Glaws, A.: G2aero: a python package for separable shape tensors. J. Open Sour. Softw. **8**(89), 5408 (2023). https://doi.org/10.21105/joss.05408
3. Atindama, E., Lef, P., Doğan, G., Athavale, P.: Restoration of noisy orientation maps from electron backscatter diffraction imaging. Int. Mater. Manuf. Innov. **12**(3), 251–266 (2023). https://doi.org/10.1007/s40192-023-00304-8
4. Bachmann, F., Hielscher, R., Jupp, P.E., Pantleon, W., Schaeben, H., Wegert, E.: Inferential statistics of electron backscatter diffraction data from within individual crystalline grains. J. Appl. Crystallogr. **43**(6), 1338–1355 (2010). https://doi.org/10.1107/S002188981003027X
5. Fan, S., Prior, D.J., Cross, A.J., Goldsby, D.L., Hager, T., Negrini, M., Qi, C.: EBSD data for synthetic polycrystalline pure water ice samples deformed at high homologous temperatures. FigShare (2020). https://doi.org/10.6084/m9.figshare.13456550.v2
6. Fan, S., Prior, D.J., Cross, A.J., Goldsby, D.L., Hager, T.F., Negrini, M., Qi, C.: Using grain boundary irregularity to quantify dynamic recrystallization in ice. Acta Mater. **209**, 116810 (2021). https://doi.org/10.1016/j.actamat.2021.116810
7. Mang, A., Gholami, A., Davatzikos, C., Biros, G.: Claire: a distributed-memory solver for constrained large deformation diffeomorphic image registration. SIAM J. Sci. Comput. **41**(5), C548–C584 (2019). https://doi.org/10.1137/18M1207818
8. Durrleman, S., Prastawa, M., Charon, N., Korenberg, J.R., Joshi, S., Gerig, G., Trouvé, A.: Morphometry of anatomical shape complexes with dense deformations and sparse parameters. Neuroimage **101**, 35–49 (2014). https://doi.org/10.1016/j.neuroimage.2014.06.043
9. Tucker, J.D., Wu, W., Srivastava, A.: Generative models for functional data using phase and amplitude separation. Comput. Stat. Data Anal. **61**, 50–66 (2013). https://doi.org/10.1016/j.csda.2012.12.001
10. Srivastava, A., Klassen, E.P.: Functional and shape data analysis, vol. 1, Springer (2016)
11. Hartman, E., Sukurdeep, Y., Klassen, E., Charon, N., Bauer, M.: Elastic shape analysis of surfaces with second-order sobolev metrics: a comprehensive numerical framework. Int. J. Comput. Vision **131**(5), 1183–1209 (2023). https://doi.org/10.1007/s11263-022-01743-0
12. Srivastava, A., Klassen, E., Joshi, S.H., Jermyn, I.H.: Shape analysis of elastic curves in euclidean spaces. IEEE Trans. Pattern Anal. Mach. Intell. **33**(7), 1415–1428 (2010). https://doi.org/10.1109/TPAMI.2010.184
13. Hagwood, C., Bernal, J., Halter, M., Elliott, J., Brennan, T.: Testing equality of cell populations based on shape and geodesic distance. IEEE Trans. Med. Imaging **32**(12), 2230–2237 (2013). https://doi.org/10.1109/TMI.2013.2278467
14. Lynch, H.J.: Satellite remote sensing for wildlife research in the polar regions. Mar. Technol. Soc. J. **57**(3), 43–50 (2023). https://doi.org/10.4031/MTSJ.57.3.1
15. Kaltenmark, I., Charlier, B., Charon, N.: A general framework for curve and surface comparison and registration with oriented varifolds. Comput. Vis. Pattern Recognit. (CVPR) (2017). https://doi.org/10.1109/CVPR.2017.487
16. Pierson, E., Daoudi, M., Arguillère, S.: 3D shape sequence of human comparison and classification using current and varifolds. In: Proceedings of the IEEE/CVF International Conference on 3D Vision (3DV) (2022), pp. 1–10
17. Bauer, M., Bruveris, M., Harms, P., Møller-Andersen, J.: A numerical framework for Sobolev metrics on the space of curves. SIAM J. Imag. Sci. **10**(1), 47–73 (2017). https://doi.org/10.1137/16M1066282
18. Hartman, E., Pierson, E.: Varigrad: A novel feature vector architecture for geometric deep learning on unregistered data. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), (2023)
19. Hartman, E., Sukurdeep, Y., Charon, N., Klassen, E., Bauer, M.: Supervised deep learning of elastic SRV distances on the shape space of curves. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2021), pp. 4425–4433. ()
20. Gonçalves, B.C., Lynch, H.J.: Fine-scale sea ice segmentation for high-resolution satellite imagery with weakly-supervised cnns. Remote Sens. **13**(18), 3562 (2021). https://doi.org/10.3390/rs13183562
21. Kendall, D.G., Barden, D., Carne, T.K., Le, H.: Shape and shape theory (John Wiley & Sons, 2009). https://doi.org/10.1002/9780470317006
22. Bauer, M., Bruveris, M., Michor, P.W.: Overview of the geometries of shape spaces and diffeomorphism groups. J. Math. Imag. Vis. **50**, 60–97 (2014). https://doi.org/10.1007/s10851-013-0490-z
23. Schwartz, A.J., Kumar, M., Adams, B.L., Field, D.P.: Electron backscatter diffraction in materials science, vol. 2, Springer (2009)
24. Saville, A.I., Vogel, S.C., Creuziger, A., Benzing, J.T., Pilchak, A.L., Nandwana, P., Klemm-Toole, J., Clarke, K.D., Semiatin, S.L., Clarke, A.J.: Texture evolution as a function of scan strategy and build height in electron beam melted ti-6al-4v. Addit. Manuf. **46**, 102118 (2021). https://doi.org/10.1016/j.addma.2021.102118
25. Stoudt, M.R., Williams, M.E., Levine, L.E., Creuziger, A., Young, S.A., Heigel, J.C., Lane, B.M., Phan, T.Q.: Location-specific microstructure characterization within in625 additive manufacturing benchmark test artifacts. Int. Mater. Manuf. Innov. **9**, 54–69 (2020). https://doi.org/10.1007/s40192-020-00172-6
26. Bachmann, F., Hielscher, R., Schaeben, H.: Texture analysis with MTEX-free and open source software toolbox. Solid State Phe-

nom. **160**, 63–68 (2010). https://doi.org/10.4028/www.scientific.net/SSP.160.63
27. Hielscher, R., Schaeben, H.: A novel pole figure inversion method: specification of the MTEX algorithm. J. Appl. Crystallogr. **41**(6), 1024–1037 (2008). https://doi.org/10.1107/S0021889808030112
28. A. International. ASTM E562-19: test method for determining volume fraction by systematic manual point count (2019). https://doi.org/10.1520/E0562-19
29. A. International. ASTM E930-18: test methods for estimating the largest grain observed in a metallographic section (ALA Grain Size) (2018). https://doi.org/10.1520/E0930-18
30. A. International. ASTM E112-24: test methods for determining average grain size (2024). https://doi.org/10.1520/E0112-24
31. Kazakov, A., Kiselev, D., Kazakova, E., Voort, G.V.: ASTM E1268: from improvement to the new standard practice for assessing the degree of banding or orientation of microstructures by automatic image analysis, In: Symposium Commemorating 100 Years of E04 Development of Metallography Standards (ASTM International, 2019), pp. 1–11. https://doi.org/10.1520/STP160720170206
32. Gretton, A., Borgwardt, K.M., Rasch, M.J., Schölkopf, B., Smola, A.: A kernel two-sample test. J. Mach. Learn. Res. **13**(1), 723–773 (2012). https://doi.org/10.5555/2188385.2188410
33. Micheli, M., Glaunès, J.A.: Matrix-valued kernels for shape deformation analysis. Geo. Imag. Comput. **1**(1), 57–139 (2014)
34. Grey, Z.J., Doronina, O.A., Glaws, A.: Separable shape tensors for aerodynamic design. J. Comput. Des. Eng. **10**(1), 468–487 (2023). https://doi.org/10.1093/jcde/qwac140
35. Bryner, D., Klassen, E., Le, H., Srivastava, A.: 2D affine and projective shape analysis. IEEE Trans. Pattern Anal. Mach. Intell. **36**(5), 998–1011 (2014). https://doi.org/10.1109/TPAMI.2013.199
36. Doronina, O.A., Grey, Z.J., Glaws, A.: *Grassmannian shape representations for aerodynamic applications*, In: Presented at the 36th AAAI Conference on Artificial Intelligence (AAAI-22) (2022) https://openreview.net/forum?id=1RRU6ud9YC
37. Absil, P.A., Mahony, R., Sepulchre, R.: *optimization algorithms on matrix manifolds* (Princeton University Press, 2008). https://doi.org/10.1515/9781400830244
38. Floater, M.S.: Arc length estimation and the convergence of polynomial curve interpolation. BIT Numer. Math. **45**, 679–694 (2005). https://doi.org/10.1007/s10543-005-0031-2
39. Fletcher, P.T., Lu, C., Joshi, S.: *Statistics of shape via principal geodesic analysis on Lie groups*, In: 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings., vol. 1 (IEEE, 2003), pp. I–I. https://doi.org/10.1109/CVPR.2003.1211342
40. Fletcher, P.T. Joshi, S.: In computer vision and mathematical methods in medical and biomedical image analysis (Springer, 2004), pp. 87–98. https://doi.org/10.1007/978-3-540-27816-0_8
41. Edelman, A., Arias, T.A., Smith, S.T.: The geometry of algorithms with orthogonality constraints. SIAM J. Matrix Anal. Appl. **20**(2), 303–353 (1998). https://doi.org/10.1137/S0895479895290954
42. Mumford, D.B., Michor, P.W.: Riemannian geometries on spaces of plane curves. J. Eur. Math. Soc. **8**(1), 1–48 (2006). https://doi.org/10.4171/jems/37
43. Younes, L., Michor, P.W., Shah, J.M., Mumford, D.B.: A metric on shape space with explicit geodesics. Rendiconti Lincei **19**(1), 25–57 (2008). https://doi.org/10.4171/rlm/506
44. Rudin, W.: Real and complex analysis, McGraw-Hill Inc (1987)
45. Bruveris, M.: Optimal reparametrizations in the square root velocity framework. SIAM J. Math. Anal. **48**(6), 4335–4354 (2016). https://doi.org/10.1137/15M1014693
46. Charon, N., Trouvé, A.: The varifold representation of nonoriented shapes for diffeomorphic registration. SIAM J. Imag. Sci. **6**(4), 2547–2580 (2013). https://doi.org/10.1137/130918885
47. Michor, P.W., Mumford, D.: Vanishing geodesic distance on spaces of submanifolds and diffeomorphisms. Doc. Math. **10**, 217–245 (2005)
48. Fasshauer, G., McCourt, M.:*Kernel-based approximation methods using Matlab*, vol. 19 (World Scientific Publishing Company, 2015). https://www.worldscientific.com/worldscibooks/10.1142/9335#t=toc
49. Fiedler, C.: Lipschitz and holder continuity in reproducing Kernel Hilbert spaces. arXiv preprint arXiv:2310.18078 (2023). https://arxiv.org/abs/2310.18078
50. Alpay, D., Jorgensen, P.E.: New characterizations of reproducing kernel hilbert spaces and applications to metric geometry. Opuscula Math. **41**(3), 283–300 (2021). https://doi.org/10.7494/OpMath.2021.41.3.283
51. Drake, K.P., Fuselier, E.J., Wright, G.B.: Implicit surface reconstruction with a curl-free radial basis function partition of unity method. SIAM J. Sci. Comput. **44**(5), A3018–A3040 (2022). https://doi.org/10.1137/22M1474485
52. Gautschi, W.: On generating orthogonal polynomials. SIAM J. Sci. Stat. Comput. **3**(3), 289–317 (1982). https://doi.org/10.1137/0903018
53. Glaws, A., Constantine, P.G.: Gauss-christoffel quadrature for inverse regression: applications to computer experiments. Stat. Comput. **29**, 429–447 (2019). https://doi.org/10.1007/s11222-018-9816-4
54. Constantine, P.G., Phipps, E.T.: A lanczos method for approximating composite functions. Appl. Math. Comput. **218**(24), 11751–11762 (2012). https://doi.org/10.1016/j.amc.2012.05.009
55. Trefethen, L.N.: Householder triangularization of a quasimatrix. IMA J. Numer. Anal. **30**(4), 887–897 (2010). https://doi.org/10.1093/imanum/drp018
56. Townsend, A., Trefethen, L.N.: Continuous analogues of matrix factorizations. Proc. Roy. Soc. A Math. Phys. Eng. Sci. **471**(2173), 20140585 (2015). https://doi.org/10.1098/rspa.2014.0585
57. Trefethen, L.N.: Approximation theory and approximation practice,, SIAM (2019). (extended edition)
58. Driscoll, T.A. Hale, N. Trefethen, L.N.: Chebfun Guide (Pafnuty Publications, 2014). http://www.chebfun.org/docs/guide/
59. Dogan, G., Bernal, J., Hagwood, C.R.: *A fast algorithm for elastic shape distances between closed planar curves*, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2015), pp. 4222–4230. https://doi.org/10.1109/CVPR.2015.7299050
60. Al-Aifari, R., Daubechies, I., Lipman, Y.: Continuous procrustes distance between two surfaces. Commun. Pure Appl. Math. **66**(6), 934–964 (2013). https://doi.org/10.1002/cpa.21444
61. Rangarajan, A., Chui, H., Bookstein, F.L.: *The softassign Procrustes matching algorithm*, In: Information Processing in Medical Imaging, ed. by J. Duncan, G. Gindi (Springer Berlin Heidelberg, Berlin, Heidelberg, 1997), pp. 29–42. https://doi.org/10.1007/3-540-63046-5_3
62. De Marchi, S., Santin, S.: A new stable basis for radial basis function interpolation. J. Comput. Appl. Math. **253**, 1–13 (2013)
63. Trefethen, L.N.: Spectral methods in Matlab (SIAM, 2000). https://epubs.siam.org/doi/book/10.1137/1.9780898719598
64. Umeyama, S.: Least-squares estimation of transformation parameters between two point patterns. IEEE Trans. Pattern Anal. Mach. Intell. **13**(4), 376–380 (1991). https://doi.org/10.1109/34.88573
65. Golub, G.H., Pereyra, V.: The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate. SIAM J. Numer. Anal. **10**(2), 413–432 (1973). https://doi.org/10.1137/0710036
66. Gibson, W.: On the least-squares orthogonalization of an oblique transformation. Psychometrika **27**(2), 193–195 (1962). https://doi.org/10.1007/BF02289637

67. Schönemann, P.H.: A generalized solution of the orthogonal procrustes problem. Psychometrika **31**(1), 1–10 (1966). https://doi.org/10.1007/BF02289451
68. Lawrence, J., Bernal, J., Witzgall, C.: A purely algebraic justification of the kabsch-umeyama algorithm. J. Res. Nat. Inst. Stand. Technol. **124**, 1 (2019). https://doi.org/10.6028/jres.124.028
69. Zhang, R., Ogden, R.T., Picard, M., Srivastava, A.: Nonparametric k-sample test on shape spaces with applications to mitochondrial shape analysis. J. R. Stat. Soc.: Ser. C: Appl. Stat. **71**(1), 51–69 (2022). https://doi.org/10.1111/rssc.12521
70. Rizzo, M.L., Székely, G.J.: DISCO analysis: a nonparametric extension of analysis of variance. Ann. Appl. Stat. **4**(2), 1034–1055 (2010). https://doi.org/10.1214/09-AOAS245
71. Bendokat, T., Zimmermann, R., Absil, P.A.: A Grassmann manifold handbook: basic geometry and computational aspects. arXiv preprint arXiv:2011.13699 (2020). https://doi.org/10.48550/arXiv.2011.13699
72. Pennec, X.: In Riemannian geometric statistics in medical image analysis, ed. by X. Pennec, S. Sommer, T. Fletcher (Academic Press, 2020), pp. 75–134. https://doi.org/10.1016/B978-0-12-814725-2.00010-8
73. Pennec, X.: Probabilities and statistics on riemannian manifolds: basic tools for geometric measurements, in *NSIP*, vol. 3 (1999), pp. 194–198. https://hal.inria.fr/inria-00615833
74. Fortet, R., Mourier, E.: Convergence de la répartition empirique vers la répartition théorique. Annales scientifiques de l'École Normale Supérieure **70**, 267–285 (1953)
75. Doronina, O.A., Lee, B., Grey, Z.J., Glaws, A.: Aerodynamic sensitivities over separable shape tensors. AIAA J. **63**, 2707–2720 (2025)

Zach Grey is a staff PhD within the applied mathematics division at the National Institute of Standards & Technology. His research is towards model-based dimension-reduction and manifold-learning methods for imaging, signal-processing, pattern/shape analysis, and optimization challenges. His work bridges mathematical theory of differential geometry with practical data-driven applications. He also has previous private industry experience developing solutions for the optimization and uncertainty quantification of gas turbine engine flowpath components.

Nicholas Fisher received a PhD in Computational and Applied Mathematics from the Colorado School of Mines in 2019. From 2023 to 2026 he was a Postdoctoral Scholar at Portland State University. His research interests center on applications of kernel based methods in data science, approximation theory, and numerical analysis with a particular emphasis on learning dynamical systems from data and the numerical solution of fractional differential equations.

Andrew Glaws is a researcher in applied mathematics in the Computational Science Center at NLR. He joined the lab as a postdoc in January 2019 to work on physics-informed deep learning for energy systems. His research focuses on enhancing scientific research into renewable energy and energy efficient problems using machine learning, artificial intelligence, and other data-driven methods. He has collaborated with domain scientists in a variety of energy-related fields, including wind and solar energy, climate science, buildings energy analysis, bioenergy, and battery technology. Prior to joining NLR, Andrew completed his Ph.D. in computer science at the University of Colorado Boulder, researching the use of parameter reduction methods for computational experiments.