

# Supplementary Material for “Know the Distortion Before You Draw: A Closed-Form Calculus for Attribute-Preserving Visualizations of Tabular Data”

Martin Radvansky

Martin Radvansky, Jr.

Submitted to *The Visual Computer*.

Both authors: Department of Computer Science, Faculty of Electrical Engineering and Computer Science, VŠB – Technical University of Ostrava, 17. listopadu 2172/15, 708 00 Ostrava-Poruba, Czech Republic.

Corresponding author: Martin Radvansky, [martin.radvansky@vsb.cz](mailto:martin.radvansky@vsb.cz).

## S1 Full proofs

This section contains the complete proofs of the results stated in Sections 3 and 4 of the main text, where each is accompanied by a two- to four-sentence sketch. The notation is that of the main text:  $\Phi = (\phi_1, \dots, \phi_p)$  is a  $p$ -tuple of linearly independent functions in  $L^2([T_{start}, T_{end}])$ ,  $V = \text{span}\{\phi_1, \dots, \phi_p\}$ ,  $G$  is the Gram matrix  $G_{ij} = \int \phi_i \phi_j dt$ ,  $\mathbf{w} \in \mathbb{R}^p$  is the weight vector,  $D_w = \text{diag}(w_{\pi(1)}, \dots, w_{\pi(p)})$ ,  $P$  is the permutation matrix with  $P_{j,\pi(j)} = 1$ , and

$$M = P^\top D_w G D_w P$$

is the design (metric) matrix of the transformation  $A = A_{\Phi, \mathbf{w}, \pi}$ . We write  $\lambda_{\min}$ ,  $\lambda_{\max}$  for the extreme eigenvalues of  $M$  and  $\kappa(M) = \lambda_{\max}/\lambda_{\min}$ .

### S1.1 The classification theorem

**Step 1: the quadratic form of the operator.** Linearity and the bilinearity of the inner product in  $L^2$  give, for any  $\mathbf{x} \in \mathbb{R}^p$ ,

$$\|A_{\Phi, \mathbf{w}, \pi} \mathbf{x}\|_{L^2}^2 = \sum_{j=1}^p \sum_{k=1}^p w_{\pi(j)} x_{\pi(j)} G_{jk} w_{\pi(k)} x_{\pi(k)} = \mathbf{x}^\top (P^\top D_w G D_w P) \mathbf{x} = \mathbf{x}^\top M \mathbf{x}.$$

Applied to  $\mathbf{x} - \mathbf{x}'$ , and using linearity of  $A$ , this is the identity

$$\|A\mathbf{x} - A\mathbf{x}'\|_{L^2}^2 = (\mathbf{x} - \mathbf{x}')^\top M (\mathbf{x} - \mathbf{x}') \quad (1)$$

quoted in the main text.

**Step 2: isometry**  $\Leftrightarrow M = I$ . If  $M = I$ , then  $\|A\mathbf{x} - A\mathbf{x}'\| = \|\mathbf{x} - \mathbf{x}'\|$  for all  $\mathbf{x}, \mathbf{x}'$  directly from Step 1. Conversely, if  $A$  preserves norms, then by the polarization identity it also preserves inner products, so  $\mathbf{x}^\top M \mathbf{x}' = \mathbf{x}^\top \mathbf{x}'$  for all  $\mathbf{x}, \mathbf{x}'$ , and hence  $M = I$ .

**Step 3: claim (i).** For  $w_j = 1$  we have  $D_w = I$  and  $M = P^\top GP$ . Since  $P$  is orthogonal,  $M = I \Leftrightarrow G = I$ , which is exactly the orthonormality of the system  $\{\phi_j\}$ . A linearly independent orthonormal system of  $p$  functions generating  $V$  is an orthonormal basis of  $V$ ; the image of  $A$  is always  $V$  (linear independence). This proves claim (i). The permutation  $\pi$  changes nothing: it corresponds to a relabelling of the basis functions, i.e. to another element of the same orbit (Step 4).

**Step 4: claim (ii), the  $O(p)$  orbit.** Let  $\Phi_0 = (\psi_1, \dots, \psi_p)$  be a fixed orthonormal basis of  $V$  and  $Q \in O(p)$ . The system  $Q\Phi_0$  with functions  $\tilde{\psi}_j = \sum_k Q_{jk}\psi_k$  has Gram matrix  $\tilde{G} = QQ^\top = I$ , hence is an orthonormal basis of  $V$ , and by (i) defines an isometry with image  $V$ . Conversely, if  $\Phi$  is an orthonormal basis of  $V$ , then the change-of-basis matrix  $Q$  between  $\Phi$  and  $\Phi_0$  (i.e.  $\phi_j = \sum_k Q_{jk}\psi_k$ ) satisfies  $I = G_\Phi = QQ^\top$ , so  $Q \in O(p)$  and  $\Phi = Q\Phi_0$ . The set of all isometric generalized Andrews transformations with image  $V$  is therefore exactly  $\{Q\Phi_0 : Q \in O(p)\}$ , which proves (ii).  $\square$

**Remark on weights.** Admitting general weights,  $M = P^\top D_w G D_w P$ , and for an orthonormal  $\Phi$  ( $G = I$ ) we obtain  $M = P^\top D_w^2 P$ . The isometry condition  $M = I$  is then equivalent to  $D_w^2 = I$ , i.e.  $w_j = \pm 1$ : the weights compatible with isometry are at most sign changes, which themselves belong to  $O(p)$ . Any non-trivial weighting ( $|w_j| \neq 1$  for some  $j$ ) necessarily destroys the isometry.

**Remark on discretization.** In the implementation the  $L^2$  inner product is replaced by the rectangle quadrature  $\langle f, g \rangle = \Delta t \sum_m f(t_m)g(t_m)$ , so the discrete curve space is  $\mathbb{R}^T$  with the inner product  $\Delta t \langle \cdot, \cdot \rangle_2$ ; all claims hold unchanged, the global constant  $\sqrt{\Delta t}$  between the Euclidean and  $L^2$  norms cancelling in every metric used. The numerical validity of both claims on the real computational chain is verified by the invariance audit reported in the main text: the deviations of the distance matrices along the orbit are of order  $10^{-15}$ , i.e. at machine precision, against a regression threshold of  $10^{-10}$ .

## S1.2 The distortion theorems

**Proof of the bi-Lipschitz bounds.** By (1),  $\|Az - Az'\|_{L^2}^2 = \mathbf{d}^\top M \mathbf{d}$  for  $\mathbf{d} = \mathbf{z} - \mathbf{z}'$ . The Rayleigh quotient of the symmetric positive semidefinite matrix  $M$  satisfies  $\lambda_{\min} \leq \mathbf{d}^\top M \mathbf{d} / \|\mathbf{d}\|^2 \leq \lambda_{\max}$  for all  $\mathbf{d} \neq 0$ , with equality exactly on the eigenvectors associated with  $\lambda_{\min}$  and  $\lambda_{\max}$ ; taking square roots yields the two-sided bound and the tightness of both constants. The ratio of the upper and lower constants is  $\sqrt{\kappa(M)}$ . For an orthonormal basis,  $M = P^\top D_w^2 P$  (Step 3 above) with eigenvalues  $w_j^2$ , hence  $\sqrt{\kappa(M)} = \max_j |w_j| / \min_j |w_j|$ . Under Definition 1 of the main text the tuple  $\Phi$  is linearly independent, so  $G$  is positive definite, and so is  $M$  whenever all weights are nonzero; if some weight vanishes,  $M$  is singular and  $\kappa = \infty$ .  $\square$

**Proof of the no-free-lunch theorem.** The operator  $A$  is an isometry up to a constant  $c$  if and only if  $\mathbf{d}^\top M \mathbf{d} = c^2 \|\mathbf{d}\|^2$  for all  $\mathbf{d}$ , which by the polarization identity is equivalent to  $M = c^2 I$ . Since  $P$  is orthogonal,  $M = P^\top D_w G D_w P = c^2 I$  is equivalent to  $D_w G D_w = c^2 I$ . From this necessarily  $w_j \neq 0$  for all  $j$  (otherwise  $M$  is singular) and  $G = c^2 D_w^{-2}$ : the Gram matrix is diagonal (the basis is orthogonal) with  $\|\phi_j\|_{L^2}^2 = c^2 / w_{\pi(j)}^2$ . Set  $\psi_j = w_{\pi(j)} \phi_j / c$ ; the system  $\{\psi_j\}$  is orthonormal and

$$Az = \sum_j w_{\pi(j)} z_{\pi(j)} \phi_j = c \sum_j z_{\pi(j)} \psi_j,$$

so the image is, up to the global factor  $c$ , identical to the unweighted transformation in the basis  $\{\psi_j\}$ . By contraposition: if the weighting changes the visualization by more than a global scale, then  $M \neq c^2 I$  and the distortion is  $\sqrt{\kappa(M)} > 1$ .  $\square$

**Proof of the upper bound on nDPE.** Let  $\mathbf{x}$  be the vector of pairwise distances in  $\mathbb{R}^p$  and  $\mathbf{y}$  the vector of the same pairs in curve space; by the bi-Lipschitz bounds,  $y_i = s_i x_i$  for some  $s_i \in [a, b]$  with  $a = \sqrt{\lambda_{\min}}$  and  $b = \sqrt{\lambda_{\max}}$ . After normalizing both distance matrices by their own Frobenius norms, nDPE equals  $\|\mathbf{x}/\|\mathbf{x}\| - \mathbf{y}/\|\mathbf{y}\|\| = \sqrt{2 - 2\cos\theta}$ , where  $\theta$  is the angle between  $\mathbf{x}$  and  $\mathbf{y}$ .

We minimize  $\cos\theta$  over all admissible scaling vectors  $\mathbf{s} \in [a, b]^N$ . Put  $u_i = x_i^2 / \sum_j x_j^2$ , so that  $\mathbf{u}$  is a probability vector supported on the pairs, and observe that

$$\cos\theta = \frac{\sum_i s_i x_i^2}{\sqrt{\sum_i x_i^2} \sqrt{\sum_i s_i^2 x_i^2}} = \frac{\mathbb{E}_{\mathbf{u}}[s]}{\sqrt{\mathbb{E}_{\mathbf{u}}[s^2]}},$$

i.e.  $\cos\theta$  depends on  $(\mathbf{s}, \mathbf{x})$  only through the distribution of  $s$  under  $\mathbf{u}$ . Fix the mean  $m = \mathbb{E}_{\mathbf{u}}[s]$ . Since  $s \mapsto s^2$  is convex on  $[a, b]$ , the second moment  $\mathbb{E}_{\mathbf{u}}[s^2]$  is maximized, and hence  $\cos\theta$  minimized, by the two-point distribution supported on the endpoints  $\{a, b\}$  with that mean. Writing  $q$  for the mass at  $a$  we obtain  $m = qa + (1-q)b$  and  $\mathbb{E}[s^2] = qa^2 + (1-q)b^2$ ; minimizing  $m^2/\mathbb{E}[s^2]$  over  $q \in [0, 1]$  gives  $q = b/(a+b)$ , hence  $m = 2ab/(a+b)$ ,  $\mathbb{E}[s^2] = ab$  and

$$\min_{\mathbf{s} \in [a, b]^N} \cos\theta = \frac{2\sqrt{ab}}{a+b}$$

(the Wielandt relation 9, an equivalent of the Kantorovich inequality 7; the quantity is Gustafson's first antieigenvalue of  $M^{1/2}$ , 8). The optimal mass ratio  $u_1/u_2 = q/(1-q) = b/a$  corresponds to the configuration  $x_2^2/x_1^2 = a/b$ . Substitution yields the bound  $\text{nDPE} \leq \sqrt{2 - 4\sqrt{ab}/(a+b)}$  of the main text.

The minimum is taken over the cube  $[a, b]^N$ , not over point configurations: each  $s_i$  is the Rayleigh quotient  $\mathbf{d}_i^\top M \mathbf{d}_i / \|\mathbf{d}_i\|^2$  of a difference vector, and the difference vectors of  $n$  points in  $\mathbb{R}^p$  are heavily constrained (they span at most  $p$  dimensions and satisfy the triangle inequality). The bound is therefore a supremum over admissible scaling vectors rather than a value attained by any point set. For realizable configurations it is conservative: equality would require concentrating all distance mass on the two extremal eigendirections with no cross pairs, whose Rayleigh quotients are mixed.  $\square$

**Proof of the distortion constant of parallel coordinates.** For uniform hat functions with knot spacing  $h$  we have  $\langle h_k, h_k \rangle = \frac{2h}{3}$  (interior knots),  $\frac{h}{3}$  (boundary knots),  $\langle h_k, h_{k\pm 1} \rangle = \frac{h}{6}$ , and all other products vanish. Normalizing each function to unit  $L^2$  norm,  $e_k = h_k / \|h_k\|$ , the Gram matrix  $G$  has unit diagonal and is tridiagonal with

$$G_{k,k+1} = \frac{h/6}{\sqrt{\frac{2h}{3} \cdot \frac{2h}{3}}} = \frac{1}{4} \quad (\text{interior pairs}), \quad G_{1,2} = G_{p-1,p} = \frac{h/6}{\sqrt{\frac{h}{3} \cdot \frac{2h}{3}}} = \frac{1}{2\sqrt{2}}.$$

Write  $G = I + C$ , where  $C \geq 0$  is tridiagonal with zero diagonal and the couplings above;  $C$  is irreducible. Let  $p \geq 3$  and put

$$v = (1, \sqrt{2}, \sqrt{2}, \dots, \sqrt{2}, 1)^\top \in \mathbb{R}^p.$$

The first row of  $Cv$  is  $\frac{1}{2\sqrt{2}}\sqrt{2} = \frac{1}{2} = \frac{1}{2}v_1$ ; the second row is  $\frac{1}{2\sqrt{2}} \cdot 1 + \frac{1}{4}\sqrt{2} = \frac{\sqrt{2}}{4} + \frac{\sqrt{2}}{4} = \frac{1}{2}v_2$ ; every interior row is  $\frac{1}{4}(\sqrt{2} + \sqrt{2}) = \frac{1}{2}v_k$ ; the remaining two rows follow by the symmetry  $v_k = v_{p+1-k}$ . The displayed second row uses  $C_{23} = \frac{1}{4}$  and therefore presumes that knot 3 is interior, i.e.  $p \geq 4$ ; for  $p = 3$  the middle row carries a boundary coupling on both sides and reads  $\frac{1}{2\sqrt{2}}(1 + 1) = \frac{1}{\sqrt{2}} = \frac{1}{2}v_2$ , so the identity holds there as well. Hence  $Cv = \frac{1}{2}v$  *exactly*, for every  $p \geq 3$ . Since  $v > 0$  componentwise and  $C$  is a nonnegative irreducible matrix, the Perron–Frobenius theorem

identifies  $\frac{1}{2}$  as the spectral radius of  $C$ : no eigenvalue of  $C$  exceeds  $\frac{1}{2}$  in modulus. Conjugating by  $D = \text{diag}((-1)^k)$  maps  $C \mapsto -C$ , so the spectrum of  $C$  is symmetric about zero and  $-\frac{1}{2}$  is attained as well (eigenvector  $Dv$ ). Therefore

$$\lambda_{\min}(G) = \frac{1}{2}, \quad \lambda_{\max}(G) = \frac{3}{2}, \quad \kappa(G) = 3$$

exactly, for every  $p \geq 3$ ; for  $p = 2$  the single coupling equals  $\frac{1}{2}$  and  $\kappa = 3$  by direct computation. The distortion factor is  $\sqrt{3}$ , independently of  $p$  and of the data.

The constant is a property of the unit-norm convention rather than of the hat basis as such: in the unnormalized (interpolatory) convention the Gram matrix is  $\frac{h}{6}B$  with  $B = \text{tridiag}(1, (2, 4, \dots, 4, 2), 1)$ . Its interior band is  $\frac{h}{6}(4 + 2\cos\vartheta) \in (\frac{h}{3}, h)$ , and the boundary rows detach *two* modes below the band, one localized at each end. The geometric vector  $u_k = r^{k-1}$  with  $r = -\frac{1}{2}$  satisfies the first row,  $2u_1 + u_2 = \frac{3}{2}u_1$ , and every interior row,  $u_{k-1} + 4u_k + u_{k+1} = r^{k-1}(r^{-1} + 4 + r) = \frac{3}{2}u_k$ , exactly; its only residual is in the opposite boundary row, where  $u_{p-1} + 2u_p = r^{p-2}(1 + 2r) = 0$  instead of  $\frac{3}{2}u_p$ , i.e. a residual of  $-\frac{3}{2}r^{p-1}$  in the last coordinate. By the reflection symmetry  $k \mapsto p+1-k$  the mirror vector  $\tilde{u}_k = r^{p-k}$  is an equally good candidate.

The two candidates are therefore *near-degenerate*, and the relevant spectral gap is the one between them, of order  $2^{-p}$ , not the  $O(1)$  gap to the interior band. A Kato–Temple argument against the band gap is consequently not available here. What the residual does give, by the min–max principle (equivalently Bauer–Fike), is  $|\lambda(B) - \frac{3}{2}| = O(2^{-p})$  for both detached eigenvalues. Their splitting follows from two-mode degenerate perturbation theory: with  $\tilde{u}^\top(B - \frac{3}{2}I)u = -\frac{3}{2}r^{p-1}$  and  $\|u\|^2 \rightarrow (1 - r^2)^{-1} = \frac{4}{3}$ , the two eigenvalues are  $\frac{3}{2} \mp \frac{9}{8}|r|^{p-1}$ , that is

$$\lambda_{\min}(B) = \frac{3}{2} - \frac{9}{4} \cdot 2^{-p} + o(2^{-p}), \quad \text{equivalently} \quad \lambda_{\min} = \frac{h}{4} - \frac{3h}{8} \cdot 2^{-p} + o(2^{-p}).$$

Numerically  $(\frac{3}{2} - \lambda_{\min}(B)) \cdot 2^p = 2.1712, 2.2415, 2.2493, 2.250000$  for  $p = 8, 12, 16, 32$ , converging to  $\frac{9}{4}$  as predicted. The splitting is symmetric only asymptotically: the second-smallest eigenvalue exceeds  $\frac{3}{2}$  by a deviation that differs from the one below by 7.7 % at  $p = 8$ , by 0.07 % at  $p = 16$ , and by less than machine precision from  $p = 32$  on. Since  $\lambda_{\max} \rightarrow h$ , we obtain  $\kappa \rightarrow 4$  and a distortion tending to 2; numerically  $\kappa = 3.9116, 3.9734, 3.9934, 3.9996$  for  $p = 8, 16, 32, 128$ . A caveat belongs here rather than in a footnote. The two-mode step is a perturbation argument, not a controlled expansion: it identifies the splitting correctly and the numerics confirm the constant  $9/4$  to four digits at  $p = 32$ , but it is not a proof of the rate. The limit  $\kappa \rightarrow 4$  itself does not depend on it and is proved above. A rigorous derivation of the rate is available along the same lines, since  $B$  commutes with the reversal  $k \mapsto p+1-k$  and the eigenvectors therefore split by parity, which turns the two-mode step into an exact block decomposition; we state the rate as numerically confirmed and leave that derivation aside.

The asymmetry between the two conventions is exactly the statement that Löwdin-type renormalization of the axes is not cosmetic, and it has a concrete geometric meaning: since  $\|h_k\|^2 = \frac{2h}{3}$  at interior knots and  $\frac{h}{3}$  at the two boundary knots, the unit-norm convention rescales the first and last axis by  $\sqrt{2}$  relative to the interior ones. The constant  $\sqrt{3}$  therefore describes a parallel-coordinates display with reweighted outer axes, whereas the interpolatory polyline that is actually drawn, and that serves as the **pcp** baseline in the benchmark, has  $\kappa \rightarrow 4$ . On the discrete  $T$ -point quadrature grid used by the implementation the unit-norm value evaluates to  $\kappa \in [2.988, 2.992]$  at  $T = 500$ , the deviation from 3 being rectangle-rule error, while the interpolatory value across the dimensionalities of the nine datasets is  $\kappa \in [3.67, 3.83]$ , i.e. a measured distortion of 1.92–1.96.  $\square$

**Proof of the decodability corollary.** For  $\lambda_{\min} > 0$ ,  $A$  is injective by the lower bi-Lipschitz bound, so  $A^{-1} : V \rightarrow \mathbb{R}^p$  exists, and for  $g = Az$ ,  $g' = Az'$  that bound gives directly  $\|z - z'\| \leq \|g - g'\|_{L^2} / \sqrt{\lambda_{\min}}$ .  $\square$

The restriction of the statement to perturbations lying in the image subspace  $V$  is milder than it looks. The orthogonal projection onto  $V$  is 1-Lipschitz, so the bound survives for the  $P_V$ -projected part of an arbitrary perturbation, and only the component orthogonal to  $V$  escapes it; that component carries no attribute information at all, since it is annihilated by  $A^{-1}P_V$ . The bound nevertheless governs the idealized decoding problem rather than perceptual reading accuracy, because a human reader makes pointwise (sup-norm) rather than  $L^2$  errors.

## S2 Per-dataset results of experiment E1

The main text aggregates experiment E1 into mean ranks, which is the appropriate summary for a comparison across datasets of very different scales: averaged over the bases, ICS ranges from 0.06 on Wine Quality to 5.67 on Iris, a factor of 91, so an average of raw values would be dominated by the easiest dataset. The aggregation does, however, hide both the individual datasets and the fold-to-fold dispersion that the evaluation protocol reports. Tables 1–3 give the underlying values in full: all nine datasets, all seven bases, mean over the five folds with the half-width of the 95 % confidence interval.

Two features of the table are worth reading directly rather than through the ranks. First, the intervals are wide relative to the differences between bases: ordering the seven bases by ICS within each dataset, the 95 % intervals of adjacent bases overlap in 43 of the 54 adjacent pairs. The ranking is therefore a statement about the *consistency of the sign* of the difference across datasets, which is exactly what the Friedman and Holm procedures test, and not a claim of a large effect on any single dataset. Second, the VCI column carries the least ambiguous single effect in the experiment: the visual complexity of the two wavelet bases exceeds that of every smooth basis on every one of the nine datasets without exception, by factors ranging from 4.5 (Breast Cancer) to 55 (Iris).

### S2.1 Aggregate ranks, Friedman statistics and adjusted $p$ -values

Table 4 gives the mean ranks of the seven bases per metric and Table 5 the corresponding Friedman tests with the Nemenyi critical difference. At  $k = 7$  bases and  $n = 9$  datasets that critical difference is 3.00 on a rank scale of seven, which no pair of the four isometric bases attains. The pairwise comparisons behave accordingly: the contrast of Ortho-B-spline against Fourier does not survive the Holm correction even inside its own metric family (ICS  $p = 0.052$ , COI  $p = 0.077$ ), let alone the further multiplicity of eleven metric families. What does survive is what the theory predicts should survive, namely the comparisons against the non-orthonormal controls and the VCI comparisons against the two wavelet bases ( $p = 9.7 \cdot 10^{-4}$  and  $2.4 \cdot 10^{-4}$ , still 0.011 and 0.003 after a conservative correction across all eleven families). Consistently, the Friedman test over the 1-NN column fails to reject the null ( $p = 0.48$ ; Iman–Davenport  $F_F(6, 48) = 0.91$ ,  $p = 0.50$ ). The classification theorem forces this column to be constant only on the orbit of the four orthonormal bases; for the three non-orthonormal controls nothing is enforced, and the test reports that they do not separate either.

### S2.2 Implementation of the ranking procedure

Two implementation details matter for reproducibility. First, before ranking, all metric values are rounded to ten decimal places: the nDPE values of orthonormal bases are exact zeros perturbed by floating-point noise of order  $10^{-16}$ , and without the rounding this consistent machine noise would be ranked as a systematic effect. Second, tied values share the average rank, which is why the four orthonormal bases share the nDPE rank 2.5. For each metric the mean ranks of the seven bases over the nine datasets are compared by the normal-approximation statistic  $z = (R_i - R_j)/\sqrt{k(k+1)/(6n)}$  over all  $\binom{7}{2} = 21$  pairs, and the resulting  $p$ -values are adjusted by the Holm step-down correction within each metric; the critical difference diagrams use the

Nemenyi test, which is the more conservative of the two procedures of Demšar [4]. Because the Friedman  $\chi_F^2$  statistic is known to be conservative at  $n = 9$ , every Friedman table also reports the Iman–Davenport  $F_F$  correction [10]; it changes no conclusion.

Table 1: Per-dataset values of the visual-quality metrics of experiment E1 (all nine datasets, all seven bases). Each cell is the mean over the five folds  $\pm$  the half-width of the 95 % confidence interval (Student’s  $t$ ,  $\text{df} = 4$ ); per-fold standard deviations are in the `e1_basis_impact.csv` output of the repository. VCI is tabulated in units of  $10^{-3}$  because it spans three orders of magnitude across bases. The mean ranks aggregated from these values are given in the rank table of Supplement S2. Part 1 of 3.

| Dataset       | Basis            | ICS                | VCI ( $\cdot 10^{-3}$ ) | COI               |
|---------------|------------------|--------------------|-------------------------|-------------------|
| Iris          | Fourier          | $4.085 \pm 0.627$  | $0.27 \pm 0.02$         | $0.283 \pm 0.041$ |
|               | Legendre         | $4.763 \pm 1.391$  | $0.15 \pm 0.02$         | $0.322 \pm 0.054$ |
|               | Chebyshev        | $5.457 \pm 1.394$  | $0.14 \pm 0.02$         | $0.288 \pm 0.042$ |
|               | Haar             | $5.214 \pm 1.634$  | $18.04 \pm 0.70$        | $0.361 \pm 0.049$ |
|               | Daubechies (db4) | $4.361 \pm 1.732$  | $15.06 \pm 1.70$        | $0.378 \pm 0.034$ |
|               | B-spline         | $5.513 \pm 2.997$  | $0.06 \pm 0.01$         | $0.332 \pm 0.120$ |
|               | Ortho-B-spline   | $10.313 \pm 4.174$ | $0.16 \pm 0.02$         | $0.318 \pm 0.035$ |
| Wine          | Fourier          | $0.886 \pm 0.222$  | $2.49 \pm 0.21$         | $0.497 \pm 0.034$ |
|               | Legendre         | $0.908 \pm 0.226$  | $4.02 \pm 0.38$         | $0.482 \pm 0.034$ |
|               | Chebyshev        | $0.860 \pm 0.217$  | $3.27 \pm 0.32$         | $0.492 \pm 0.032$ |
|               | Haar             | $1.004 \pm 0.248$  | $57.46 \pm 1.74$        | $0.468 \pm 0.017$ |
|               | Daubechies (db4) | $0.964 \pm 0.307$  | $62.67 \pm 1.97$        | $0.473 \pm 0.025$ |
|               | B-spline         | $1.108 \pm 0.393$  | $1.17 \pm 0.05$         | $0.482 \pm 0.062$ |
|               | Ortho-B-spline   | $1.099 \pm 0.204$  | $2.31 \pm 0.08$         | $0.454 \pm 0.043$ |
| Breast Cancer | Fourier          | $0.248 \pm 0.055$  | $12.27 \pm 0.42$        | $0.641 \pm 0.022$ |
|               | Legendre         | $0.149 \pm 0.050$  | $31.82 \pm 1.17$        | $0.694 \pm 0.033$ |
|               | Chebyshev        | $0.138 \pm 0.048$  | $27.09 \pm 0.81$        | $0.695 \pm 0.027$ |
|               | Haar             | $0.451 \pm 0.035$  | $143.91 \pm 3.06$       | $0.528 \pm 0.011$ |
|               | Daubechies (db4) | $0.451 \pm 0.049$  | $150.67 \pm 3.43$       | $0.552 \pm 0.017$ |
|               | B-spline         | $0.695 \pm 0.094$  | $4.20 \pm 0.19$         | $0.460 \pm 0.019$ |
|               | Ortho-B-spline   | $0.567 \pm 0.069$  | $9.01 \pm 0.31$         | $0.505 \pm 0.015$ |

Fig. 1 shows the same comparison as a critical difference diagram.

### S3 Parameter audits and full audit tables

All results in the main text use a single pipeline configuration ( $T = 500$ , cubic B-splines, the db4 wavelet, Silverman bandwidth,  $\text{rcond} = 10^{-12}$ ). This section reports the audits that establish which conclusions depend on that choice: the invariance audit E7 and the clustering audit E5, the ablation grid E11 (the full product  $T \in \{32, \dots, 2048\}$  by seven bases, plus one-factor-at-a-time axes for spline degree, wavelet order and the Löwdin orthonormalization threshold, 594 combinations under 5-fold cross-validation), the COI bandwidth sweep E12, the COI estimator audit E16, the ICS localization audit E17, the clutter grid audit E19 and the stability audit E20 of the located minimum. It also carries the grid measurement of the parallel-coordinates constant of the main text.

#### S3.1 Full tables of the structural columns (E7, E5)

The main text reports the outcome of the two audits of the structural columns in one sentence each; the tables behind those sentences are given here. Table 6 lists, for each of the eleven bases, the largest deviation observed when the basis is rotated by three random  $Q \in O(p)$  on

Table 2: Per-dataset values of experiment E1, part 2 of 3 (continued from Table 1; same format).

| Dataset              | Basis            | ICS               | VCI ( $\cdot 10^{-3}$ ) | COI               |
|----------------------|------------------|-------------------|-------------------------|-------------------|
| Heart Disease        | Fourier          | $0.131 \pm 0.038$ | $2.44 \pm 0.02$         | $0.739 \pm 0.020$ |
|                      | Legendre         | $0.095 \pm 0.033$ | $4.15 \pm 0.14$         | $0.764 \pm 0.016$ |
|                      | Chebyshev        | $0.091 \pm 0.034$ | $3.26 \pm 0.12$         | $0.766 \pm 0.016$ |
|                      | Haar             | $0.142 \pm 0.040$ | $61.55 \pm 0.19$        | $0.714 \pm 0.008$ |
|                      | Daubechies (db4) | $0.140 \pm 0.031$ | $65.14 \pm 0.30$        | $0.708 \pm 0.031$ |
|                      | B-spline         | $0.194 \pm 0.025$ | $1.22 \pm 0.08$         | $0.674 \pm 0.025$ |
|                      | Ortho-B-spline   | $0.142 \pm 0.023$ | $2.40 \pm 0.05$         | $0.662 \pm 0.020$ |
| Wine Quality (White) | Fourier          | $0.062 \pm 0.010$ | $1.78 \pm 0.02$         | $0.727 \pm 0.040$ |
|                      | Legendre         | $0.063 \pm 0.007$ | $2.59 \pm 0.04$         | $0.729 \pm 0.034$ |
|                      | Chebyshev        | $0.064 \pm 0.008$ | $2.16 \pm 0.02$         | $0.728 \pm 0.035$ |
|                      | Haar             | $0.064 \pm 0.013$ | $49.75 \pm 0.39$        | $0.721 \pm 0.022$ |
|                      | Daubechies (db4) | $0.061 \pm 0.014$ | $56.30 \pm 0.44$        | $0.724 \pm 0.038$ |
|                      | B-spline         | $0.065 \pm 0.015$ | $0.85 \pm 0.01$         | $0.721 \pm 0.038$ |
|                      | Ortho-B-spline   | $0.058 \pm 0.013$ | $1.72 \pm 0.01$         | $0.722 \pm 0.029$ |
| Dry Bean             | Fourier          | $2.379 \pm 0.114$ | $3.73 \pm 0.02$         | $0.389 \pm 0.004$ |
|                      | Legendre         | $1.944 \pm 0.106$ | $7.37 \pm 0.05$         | $0.406 \pm 0.004$ |
|                      | Chebyshev        | $1.904 \pm 0.101$ | $5.95 \pm 0.03$         | $0.414 \pm 0.004$ |
|                      | Haar             | $3.717 \pm 0.134$ | $79.05 \pm 0.34$        | $0.292 \pm 0.001$ |
|                      | Daubechies (db4) | $3.706 \pm 0.150$ | $77.22 \pm 0.22$        | $0.301 \pm 0.002$ |
|                      | B-spline         | $6.114 \pm 0.228$ | $0.93 \pm 0.01$         | $0.277 \pm 0.004$ |
|                      | Ortho-B-spline   | $5.667 \pm 0.173$ | $1.86 \pm 0.02$         | $0.303 \pm 0.002$ |

Table 3: Per-dataset values of experiment E1, part 3 of 3 (continued from Table 1; same format).

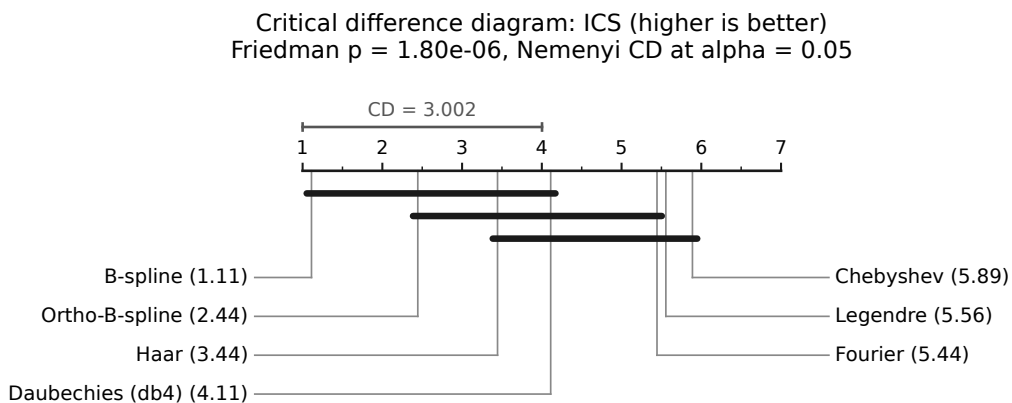
| Dataset        | Basis            | ICS               | VCI ( $\cdot 10^{-3}$ ) | COI               |
|----------------|------------------|-------------------|-------------------------|-------------------|
| Rice           | Fourier          | $0.542 \pm 0.027$ | $0.75 \pm 0.01$         | $0.590 \pm 0.012$ |
|                | Legendre         | $0.545 \pm 0.023$ | $0.84 \pm 0.01$         | $0.576 \pm 0.012$ |
|                | Chebyshev        | $0.487 \pm 0.021$ | $0.87 \pm 0.01$         | $0.604 \pm 0.010$ |
|                | Haar             | $0.759 \pm 0.041$ | $26.80 \pm 0.15$        | $0.506 \pm 0.020$ |
|                | Daubechies (db4) | $0.621 \pm 0.030$ | $33.39 \pm 0.33$        | $0.526 \pm 0.016$ |
|                | B-spline         | $1.491 \pm 0.081$ | $0.20 \pm 0.01$         | $0.267 \pm 0.018$ |
|                | Ortho-B-spline   | $1.052 \pm 0.052$ | $0.45 \pm 0.01$         | $0.389 \pm 0.014$ |
| Early Diabetes | Fourier          | $0.121 \pm 0.021$ | $3.59 \pm 0.05$         | $0.681 \pm 0.039$ |
|                | Legendre         | $0.109 \pm 0.004$ | $6.15 \pm 0.21$         | $0.682 \pm 0.041$ |
|                | Chebyshev        | $0.116 \pm 0.004$ | $5.06 \pm 0.06$         | $0.676 \pm 0.038$ |
|                | Haar             | $0.132 \pm 0.027$ | $76.21 \pm 1.30$        | $0.678 \pm 0.050$ |
|                | Daubechies (db4) | $0.147 \pm 0.024$ | $74.94 \pm 0.82$        | $0.637 \pm 0.046$ |
|                | B-spline         | $0.232 \pm 0.054$ | $1.79 \pm 0.05$         | $0.598 \pm 0.069$ |
|                | Ortho-B-spline   | $0.188 \pm 0.039$ | $3.49 \pm 0.08$         | $0.580 \pm 0.057$ |
| Seeds          | Fourier          | $2.153 \pm 0.824$ | $0.76 \pm 0.06$         | $0.381 \pm 0.068$ |
|                | Legendre         | $1.687 \pm 0.717$ | $0.86 \pm 0.05$         | $0.418 \pm 0.078$ |
|                | Chebyshev        | $1.348 \pm 0.608$ | $0.88 \pm 0.04$         | $0.473 \pm 0.074$ |
|                | Haar             | $2.602 \pm 0.883$ | $26.80 \pm 1.72$        | $0.332 \pm 0.060$ |
|                | Daubechies (db4) | $2.881 \pm 1.105$ | $33.70 \pm 1.31$        | $0.331 \pm 0.054$ |
|                | B-spline         | $3.684 \pm 1.126$ | $0.21 \pm 0.01$         | $0.242 \pm 0.055$ |
|                | Ortho-B-spline   | $2.969 \pm 1.082$ | $0.53 \pm 0.01$         | $0.325 \pm 0.058$ |

Table 4: Mean ranks of functional bases across datasets (lower is better; ties share the average rank), isometric bases first. Visual\_Rank aggregates the visual-quality metrics ICS, VCI and COI only; nDPE, NPS and 1-NN accuracy verify isometry (orthonormal bases tie by construction) and are reported but not aggregated. Non-isometric bases (Isometric = no) are negative controls: their favourable visual ranks are obtained precisely by deforming distances.

| basis         | ICS   | nDPE  | NPS   | VCI   | COI   | 1-NN  | Visual_Rank | Isometric |
|---------------|-------|-------|-------|-------|-------|-------|-------------|-----------|
| ortho_bspline | 2.444 | 2.500 | 2.167 | 2.222 | 2.111 | 3.889 | 2.259       | yes       |
| haar          | 3.444 | 2.500 | 2.500 | 6.333 | 3.333 | 4.167 | 4.370       | yes       |
| fourier       | 5.444 | 2.500 | 2.833 | 3.222 | 5.000 | 3.889 | 4.556       | yes       |
| daubechies    | 4.111 | 2.500 | 2.500 | 6.667 | 3.778 | 4.167 | 4.852       | yes       |
| bspline       | 1.111 | 6.667 | 6.889 | 1.000 | 2.111 | 5.167 | 1.407       | no        |
| chebyshev     | 5.889 | 6.333 | 6.111 | 4.000 | 5.889 | 3.389 | 5.259       | no        |
| legendre      | 5.556 | 5.000 | 5.000 | 4.556 | 5.778 | 3.333 | 5.296       | no        |

Table 5: Friedman test over bases per metric with the Nemenyi critical difference ( $\alpha = 0.05$ ).  $\chi_F^2$  is the Friedman statistic,  $F_F$  its Iman–Davenport correction with  $(k - 1, (k - 1)(n - 1))$  degrees of freedom, which is less conservative at this sample size; the two agree on every metric. For orthonormal bases the isometry columns (nDPE, NPS, 1-NN accuracy) are tied by construction (Corollary B); their Friedman tests are reported for completeness only.

| metric  | k | n | $\chi_F^2$ | $p$      | $F_F$  | $p_F$    | CD <sub>0.05</sub> |
|---------|---|---|------------|----------|--------|----------|--------------------|
| ICS     | 7 | 9 | 36.95      | 1.8e-06  | 17.34  | 1.52e-10 | 3.0024             |
| nDPE    | 7 | 9 | 52.96      | 1.2e-09  | 406.00 | 2.3e-39  | 3.0024             |
| NPS     | 7 | 9 | 51.57      | 2.28e-09 | 169.55 | 1.46e-30 | 3.0024             |
| VCI     | 7 | 9 | 49.43      | 6.12e-09 | 86.50  | 5.04e-24 | 3.0024             |
| COI     | 7 | 9 | 29.62      | 4.64e-05 | 9.72   | 5.38e-07 | 3.0024             |
| kNN_acc | 7 | 9 | 5.52       | 0.479    | 0.91   | 0.496    | 3.0024             |



**Fig. 1** Critical difference diagram of functional bases for Integrated Class Separability (ICS, higher is better; Friedman with Nemenyi post-hoc,  $\alpha = 0.05$ , nine datasets). Bases connected by a bar are not significantly different at this  $n$ ; the ranking of the negative-control B-spline must be read jointly with its distance-preservation failure

all nine datasets. For orthonormal bases the predicted deviation is zero and the acceptance threshold is  $10^{-10}$ ; the largest distance-matrix deviation is  $5.3 \times 10^{-15}$  and the ARI deviation is exactly zero, while every non-orthonormal negative control, which is what makes the audit able to fail, reaches at least 0.16. The residual NPS and 1-NN deviations of the orthonormal rows are neighbour-order flips at tied distances, not distance changes.

Table 6: Maximum deviations of curve-space quantities under random orthogonal rotations of the basis (3 seeds, 9 datasets; experiment E7). Orthonormal bases must show distance-matrix deviations at machine precision (Corollary B); non-orthonormal bases serve as negative controls. NPS, kNN and ARI deviations are bounded by neighbour tie-breaking, not exact zeros.

| Basis            | Orthonormal | Dist. dev. | nDPE dev. | NPS dev. | kNN dev. | ARI dev. |
|------------------|-------------|------------|-----------|----------|----------|----------|
| Daubechies (db4) | yes         | 4.62e-15   | 1.17e-15  | 8.85e-03 | 6.41e-03 | 0.00e+00 |
| Fourier          | yes         | 8.20e-16   | 1.43e-16  | 1.50e-02 | 6.41e-03 | 0.00e+00 |
| Haar             | yes         | 5.32e-15   | 1.66e-15  | 1.38e-02 | 6.41e-03 | 0.00e+00 |
| Ortho-B-spline   | yes         | 1.10e-15   | 1.66e-16  | 1.00e-02 | 6.41e-03 | 0.00e+00 |
| Ortho-Chebyshev  | yes         | 1.05e-15   | 3.18e-16  | 1.27e-02 | 6.41e-03 | 0.00e+00 |
| Ortho-Hat        | yes         | 7.62e-16   | 1.31e-16  | 9.62e-03 | 6.41e-03 | 0.00e+00 |
| Ortho-Legendre   | yes         | 9.50e-16   | 1.59e-16  | 1.38e-02 | 0.00e+00 | 0.00e+00 |
| B-spline         | no          | 4.29e-01   | 1.29e-01  | 1.58e-01 | 8.89e-02 | 2.45e-01 |
| Chebyshev        | no          | 4.21e-01   | 1.31e-01  | 1.76e-01 | 6.67e-02 | 2.32e-01 |
| Hat (PCP)        | no          | 2.41e-01   | 3.89e-02  | 5.33e-02 | 6.35e-02 | 1.34e-01 |
| Legendre         | no          | 1.59e-01   | 6.67e-02  | 6.92e-02 | 1.17e-02 | 9.09e-02 |

Table 7 states the same fact in the applied direction. K-Means in curve space returns bitwise the same Adjusted Rand Index as K-Means in  $\mathbb{R}^p$  for every orthonormal basis and every dataset, which is a mathematical necessity rather than a result; the informative column is the plain B-spline control, whose ARI change prices the departure from the isometric set.

Table 7: K-Means clustering quality (Adjusted Rand Index). For orthonormal bases identical ARI is mathematically necessary (Corollary B) and serves as implementation verification; the plain B-spline basis is a negative control.

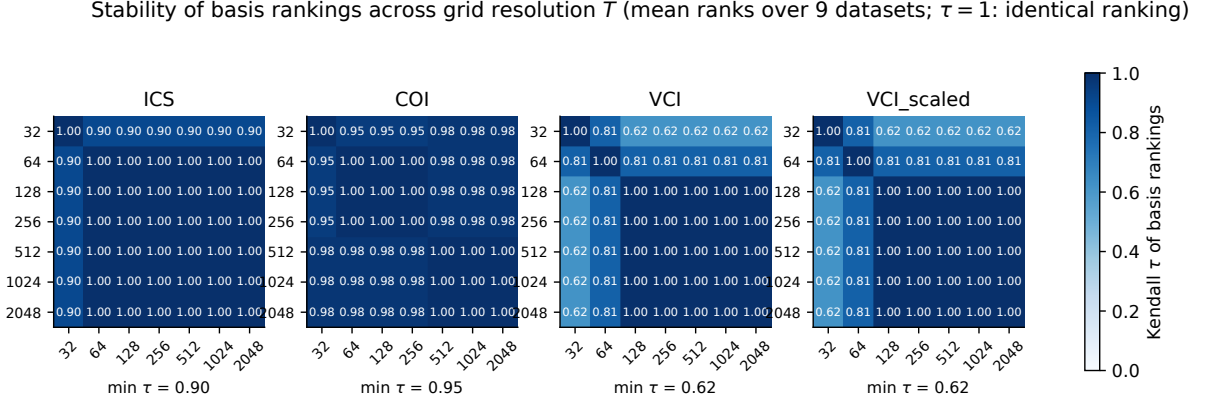
| Dataset              | ARI (original space) | ARI (Ortho-B-spline) | ARI (B-spline) |
|----------------------|----------------------|----------------------|----------------|
| Iris                 | 0.620                | 0.620                | 0.487          |
| Wine                 | 0.897                | 0.897                | 0.713          |
| Breast Cancer        | 0.654                | 0.654                | 0.603          |
| Heart Disease        | 0.447                | 0.447                | 0.429          |
| Wine Quality (White) | 0.034                | 0.034                | 0.023          |
| Dry Bean             | 0.669                | 0.669                | 0.639          |
| Rice                 | 0.682                | 0.682                | 0.674          |
| Early Diabetes       | 0.177                | 0.177                | 0.229          |
| Seeds                | 0.773                | 0.773                | 0.677          |

### S3.2 Stability of the basis ranking (E11, E12)

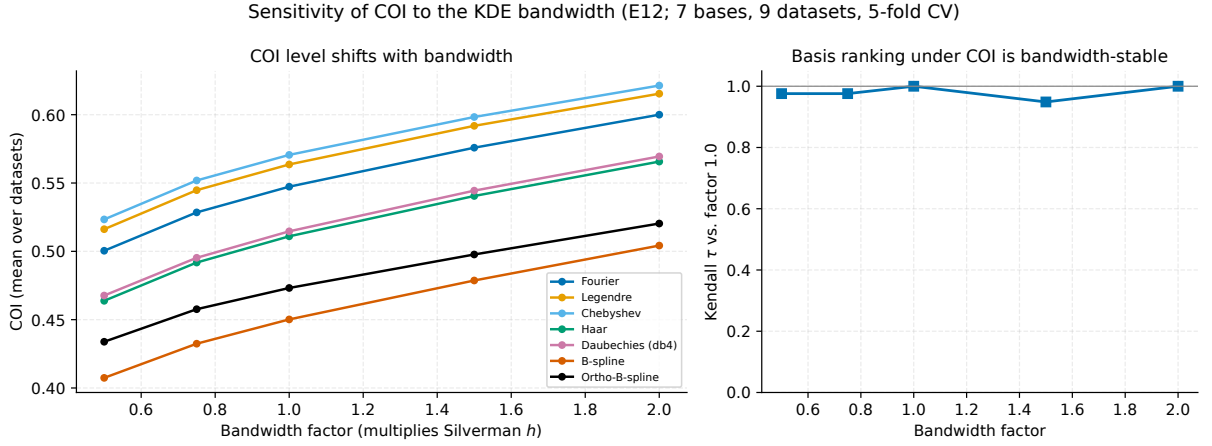
Kendall’s  $\tau$ , throughout in the tie-corrected variant  $\tau_b$  because the mean-rank vectors contain ties, between the basis rankings at different resolutions equals 1.00 for ICS and VCI and  $\geq 0.97$  for COI over  $T \in [128, 2048]$ : the ordering reported in the main text does not depend on the choice of  $T$ . The ranking changes only at extremely coarse grids  $T \leq 64$  ( $\tau$  down to 0.62), where the sampling ceases to resolve the roughness of the fine bases. The ablation grid runs over powers of two, so the study’s own resolution  $T = 500$  is not one of its points;  $T = 512$  is its surrogate,

and the two differ by less than the spacing between neighbouring grid points of the ablation. The COI-based ranking is likewise stable against the KDE bandwidth ( $\tau \geq 0.94$  between mean-rank vectors for factors 0.5–2.0): the absolute COI level grows monotonically with the bandwidth, but the cross-dataset basis ranking does not. Figs. 2 and 3 show the two sweeps.

The remaining axes are second-order. The threshold  $\text{rcond} \in [10^{-14}, 10^{-8}]$  does not change the metrics even in the last significant digits, the Gram matrices being well conditioned; the choice of wavelet within the family (db2–db8, sym4, coif2) changes ICS only within 1.45–1.55; and the spline degree preserves isometry exactly for the orthonormalized variants (nDPE = 0 for all degrees, in agreement with the classification theorem), whereas for the unnormalized ones the deformation grows with the degree (nDPE 0.07  $\rightarrow$  0.17 for degree 1  $\rightarrow$  5).



**Fig. 2** Stability of basis rankings across grid resolution  $T$  (Kendall  $\tau$  between mean-rank vectors over nine datasets; E11). Rankings are perfectly stable for ICS and VCI and stable to  $\tau \geq 0.97$  for COI over  $T \in [128, 2048]$ ; only very coarse grids ( $T \leq 64$ ) reorder bases under the roughness-based metrics

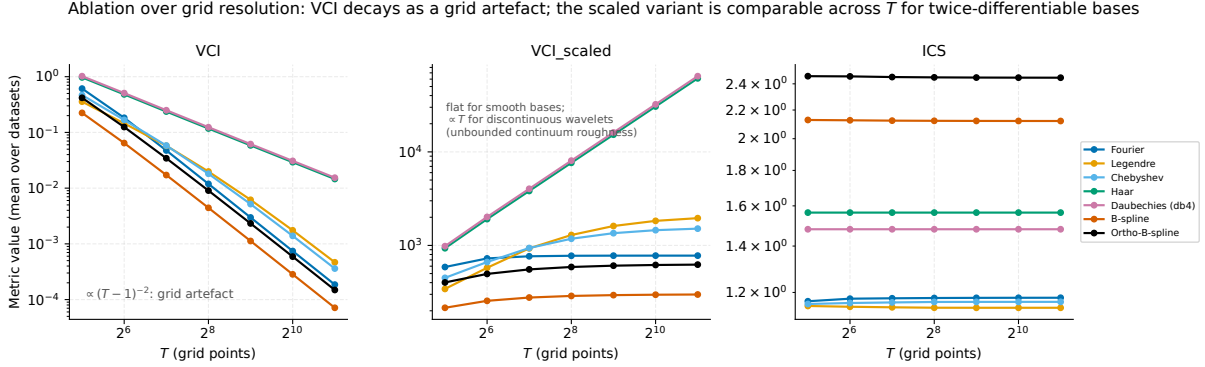


**Fig. 3** Sensitivity of COI to the KDE bandwidth factor (E12). The COI level shifts monotonically with the bandwidth, but the cross-dataset basis ranking is stable (Kendall  $\tau \geq 0.94$  between mean-rank vectors against factor 1.0)

### S3.3 VCI as a grid artefact (E11)

VCI decays  $\propto (T-1)^{-2}$  for all bases, and the scaled variant  $\widetilde{\text{VCI}}$  behaves exactly as predicted: flat for twice-differentiable bases and growing  $\propto T$  for the discontinuous wavelets, whose continuum roughness is unbounded. At fixed  $T$  the basis ranking is identical in both variants, so the

conclusions of the main text remain valid; across different  $T$ , however,  $\widetilde{\text{VCI}}$  must be used. Fig. 4 plots both variants against  $T$ .



**Fig. 4** Ablation over grid resolution (E11, means over datasets). VCI decays  $\propto (T - 1)^{-2}$ , a grid artefact; the scaled variant  $\widetilde{\text{VCI}}$  is flat for twice-differentiable bases and grows  $\propto T$  for discontinuous wavelets. ICS is essentially independent of  $T$

### S3.4 COI estimator audit (E16)

COI averages over the *admissible* classes, those represented by at least two instances in the evaluated sample, since a single instance admits no bandwidth estimate. The distinction is not academic: under 5-fold cross-validation the rarest Wine Quality class (quality 9, five instances in total) contributes exactly one test-fold instance and is dropped in all five folds, so COI is averaged there over 15 class pairs rather than the nominal 21. On the remaining eight datasets the effective class count equals the nominal one in every fold. The exclusion is from the density estimation and from the pair average; the values of the dropped instances still delimit the range over which the inner integral is discretized, and on Wine Quality even that residual influence vanishes, since the dropped instance lies inside the value range of the remaining classes in all five folds and recomputing COI with those rows removed reproduces the published values exactly. The binned KDE used in production was audited against an exact Gaussian KDE on a 2048-point grid at identical bandwidths: the two agree to  $4.7 \cdot 10^{-3}$  at worst pointwise and to  $5 \cdot 10^{-4}$  in the aggregate on all nine datasets.

### S3.5 Full tables of the ICS localization audit (E17)

Table 8 gives the per-dataset detail behind the summary in the main text. For each dataset it reports the range of plain ICS over the seven orthonormal bases, the common value of  $\text{ICS}_{\text{ratio}}$ , the largest relative deviation from that common value among the orthonormal bases (predicted value zero, acceptance threshold  $10^{-10}$ ) and the smallest deviation among the four non-orthonormal controls. The pattern is uniform across datasets: plain ICS varies by factors of 1.14 to 4.38 between isometric bases, and the entire variation disappears when the ratio is taken outside the time average.

### S3.6 Full tables of the clutter grid audit (E19)

Table 9 gives the clutter measures of the three curve-type techniques at five grid resolutions, and Fig. 5 the same audit as a picture. Raw VCI falls by a factor of 231 to 261 over  $T \in [128, 2048]$ , whereas the total variation of the derivative per unit curve energy moves by at most 13.3% and the crossing count per degree of freedom is stable. Under the grid-invariant measure the values settle at 51.6 for parallel coordinates, 99.1 for Ortho-B-spline and 124 for Fourier, and

Table 8: Audit E17 per dataset. Columns 2–4 give the range of plain ICS over the seven orthonormal bases and their ratio; column 5 the common value of  $\text{ICS}_{\text{ratio}}$ , which equals  $\text{tr}(S_B^{(Z)})/\text{tr}(S_W^{(Z)})$  in  $\mathbb{R}^p$ ; column 6 the largest relative deviation from it among the orthonormal bases (predicted value zero, acceptance threshold  $10^{-10}$ ); and column 7 the smallest deviation among the four non-orthonormal controls.

| Dataset        | ICS min | ICS max | max/min | $\text{ICS}_{\text{ratio}}$ | max dev. (ONB)       | min dev. (control)  |
|----------------|---------|---------|---------|-----------------------------|----------------------|---------------------|
| Iris           | 3.276   | 8.103   | 2.47    | 2.603                       | $7.6 \cdot 10^{-14}$ | $3.1 \cdot 10^{-3}$ |
| Seeds          | 1.209   | 2.841   | 2.35    | 2.142                       | $2.8 \cdot 10^{-14}$ | $1.1 \cdot 10^{-2}$ |
| Wine           | 0.727   | 0.991   | 1.36    | 0.780                       | $1.0 \cdot 10^{-14}$ | $5.1 \cdot 10^{-3}$ |
| Early Diabetes | 0.095   | 0.177   | 1.85    | 0.134                       | $2.7 \cdot 10^{-15}$ | $2.4 \cdot 10^{-2}$ |
| Heart Disease  | 0.081   | 0.142   | 1.76    | 0.130                       | $3.8 \cdot 10^{-15}$ | $6.3 \cdot 10^{-2}$ |
| Breast Cancer  | 0.125   | 0.546   | 4.38    | 0.398                       | $3.5 \cdot 10^{-15}$ | $1.7 \cdot 10^{-1}$ |
| Rice           | 0.540   | 1.116   | 2.07    | 0.730                       | $1.4 \cdot 10^{-15}$ | $7.1 \cdot 10^{-3}$ |
| Dry Bean       | 1.917   | 5.892   | 3.07    | 2.890                       | $1.7 \cdot 10^{-15}$ | $1.7 \cdot 10^{-2}$ |
| Wine Quality   | 0.050   | 0.057   | 1.14    | 0.057                       | $1.3 \cdot 10^{-15}$ | $7.6 \cdot 10^{-3}$ |

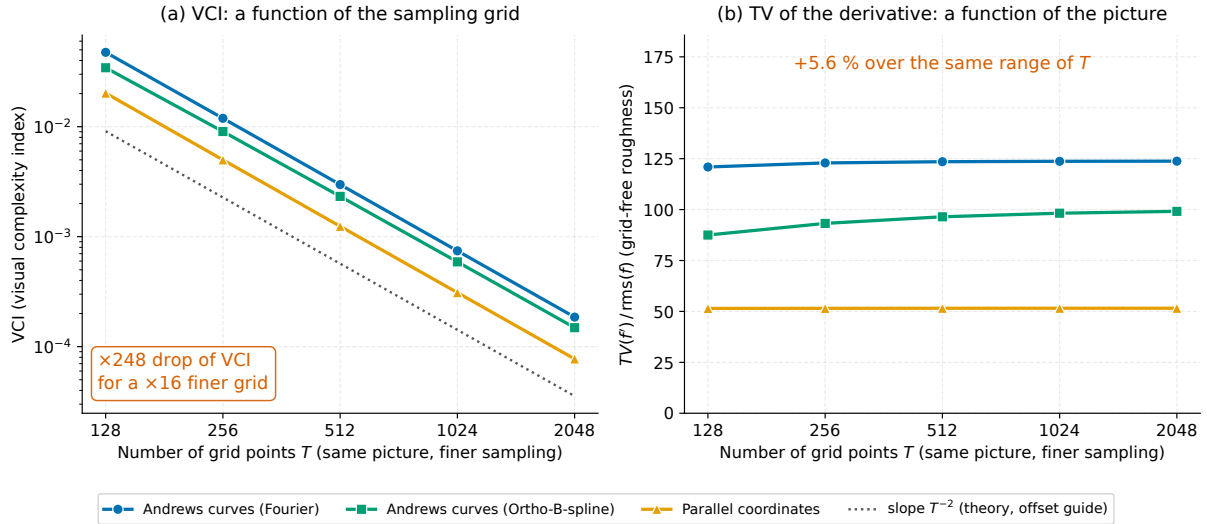
the crossing count per independent degree of freedom at 0.36, 0.52 and 0.67 at  $T = 2048$ ; the ordering of the three techniques is the same under every measure and at every resolution.

Table 9: Audit E19: clutter measures against grid resolution, means over the nine datasets. Raw VCI drifts by roughly two and a half orders of magnitude over the tested range; the total variation of the derivative per unit curve energy (TV) and the crossing count per degree of freedom are stable and give the same ordering at every resolution.

| Technique                | $T$  | VCI     | $\widetilde{\text{VCI}}$ | TV    | Cross./dim |
|--------------------------|------|---------|--------------------------|-------|------------|
| Andrews (Fourier)        | 128  | 0.04748 | 766                      | 120.9 | 0.66       |
| Andrews (Fourier)        | 256  | 0.01192 | 775                      | 122.9 | 0.66       |
| Andrews (Fourier)        | 512  | 0.00298 | 778                      | 123.5 | 0.66       |
| Andrews (Fourier)        | 1024 | 0.00074 | 778                      | 123.7 | 0.67       |
| Andrews (Fourier)        | 2048 | 0.00019 | 778                      | 123.8 | 0.67       |
| Andrews (Ortho-B-spline) | 128  | 0.03436 | 554                      | 87.5  | 0.51       |
| Andrews (Ortho-B-spline) | 256  | 0.00904 | 588                      | 93.2  | 0.51       |
| Andrews (Ortho-B-spline) | 512  | 0.00233 | 607                      | 96.5  | 0.51       |
| Andrews (Ortho-B-spline) | 1024 | 0.00059 | 618                      | 98.2  | 0.52       |
| Andrews (Ortho-B-spline) | 2048 | 0.00015 | 623                      | 99.1  | 0.52       |
| Parallel coordinates     | 128  | 0.02022 | 326                      | 51.5  | 0.32       |
| Parallel coordinates     | 256  | 0.00500 | 325                      | 51.5  | 0.31       |
| Parallel coordinates     | 512  | 0.00124 | 325                      | 51.6  | 0.32       |
| Parallel coordinates     | 1024 | 0.00031 | 324                      | 51.6  | 0.33       |
| Parallel coordinates     | 2048 | 0.00008 | 324                      | 51.6  | 0.36       |

### S3.7 Stability of the located minimum (E20)

The COI profile is used in the main text only comparatively, but an analyst reading a profile also reads the *position* of its minimum, and that quantity is the weaker link. Audit E20 perturbs it in three ways on the case-study datasets: the bandwidth factor over 0.5–2.0, the grid resolution over  $T \in [128, 2048]$  and 200 nonparametric bootstrap resamples. Bandwidth and resolution are harmless; the bootstrap is not. On Seeds the located position ranges over  $t \in [-3.09, +1.08]$ , and the attribute named at the minimum is the modal one in only 48 % of resamples (Early Diabetes 64 %, Heart Disease 70 %). The diagnosis is not that the profile locates the wrong region but that it locates a *region* rather than a point: nearly all bootstrap minima fall inside the near-minimal



**Fig. 5** A clutter column that reports the sampling grid instead of the picture (E19). (a) VCI of three unchanged displays as the number of grid points  $T$  is refined: the value falls by roughly a factor of four per doubling of  $T$ , along the reference slope  $T^{-2}$  that follows from VCI being a second difference of a fixed continuous curve. Nothing about the drawings changes along the horizontal axis, so a statement of the form “technique X is less cluttered by a factor  $c$ ” is in part a statement about the grid. (b) The total variation of the derivative per unit curve energy, measured on the same curves and the same grids, is flat, because it has a continuum limit for smooth curves and for polylines alike. Comparisons across display types of different smoothness are admissible in panel B and not in panel A. The ordering of the three techniques is the same in both, which is why we retain the conclusion about polylines while discarding the measure that first suggested it

band of the full-sample profile, the times within 5% of the profile’s depth, which on Seeds is simply wide. Point estimates of a discriminating time should therefore be reported together with that band.

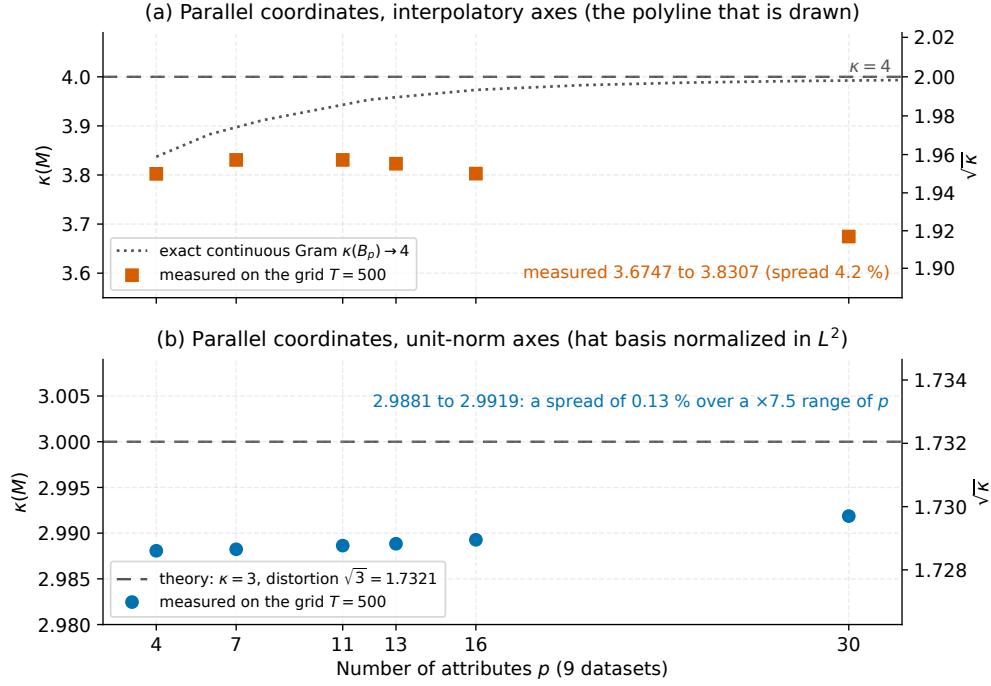
### S3.8 Per-cell fold counts of the weighting audit (E8)

The distortion ranges quoted for discriminative weighting in the main text are taken over the folds with a finite  $\kappa$ , since a single singular fold would send every aggregate to infinity, and that makes some cells thinner than others. The singular folds are concentrated on three datasets: mutual information is singular in all five folds on Early Diabetes, so that dataset contributes nothing to the mutual-information range, in four of five on Heart Disease and in two of five on Breast Cancer, while the single ANOVA case is one fold of Breast Cancer. Every other cell rests on all five folds.

### S3.9 The parallel-coordinates constant on the quadrature grid (E15)

The distortion of parallel coordinates is a constant of the axis convention rather than a measurement, and the main text states the two exact continuum values. Fig. 6 shows what the same quantity does on the quadrature grid actually used for drawing, over  $p = 4$  to 30: the interpolatory polyline approaches  $\kappa = 4$  from below and the unit-norm convention stays constant to within 0.13% of  $\kappa = 3$ . Neither panel depends on which dataset is drawn. The measured values of panel (a) lie below the exact ones and, on this fixed grid, fall as  $p$  grows, which is quadrature error rather than a departure from the continuum result: refining the grid at fixed  $p$  recovers the exact value to two decimal places, and Table 10 in Supplement S4 gives the convergence together

with the rate.



Both conventions are measured on the same fixed quadrature grid  $T = 500$ ; the offset below the continuous limit is the quadrature error and grows with  $p$ .

**Fig. 6** The distortion of parallel coordinates is a constant, not a measurement. Each panel plots the condition number  $\kappa(M)$  of the hat basis against the number of attributes  $p$ , with the corresponding distortion  $\sqrt{\kappa(M)}$  on the right-hand axis. (a) The interpolatory polyline that is actually drawn: the exact eigenvalues of the continuum Gram matrix approach  $\kappa = 4$  from below (dotted), and the values measured on the quadrature grid stay in a narrow band beneath it. (b) The unit-norm axis convention:  $\kappa = 3$  exactly in the continuum, and constant to within 0.13 % over  $p = 4$  to 30 on the grid. Neither panel depends on which dataset is drawn, which is the point: the price of a parallel-coordinates plot is fixed by the axis convention alone, and the two conventions must not be quoted interchangeably

## S4 Data, metrics and evaluation protocol in full

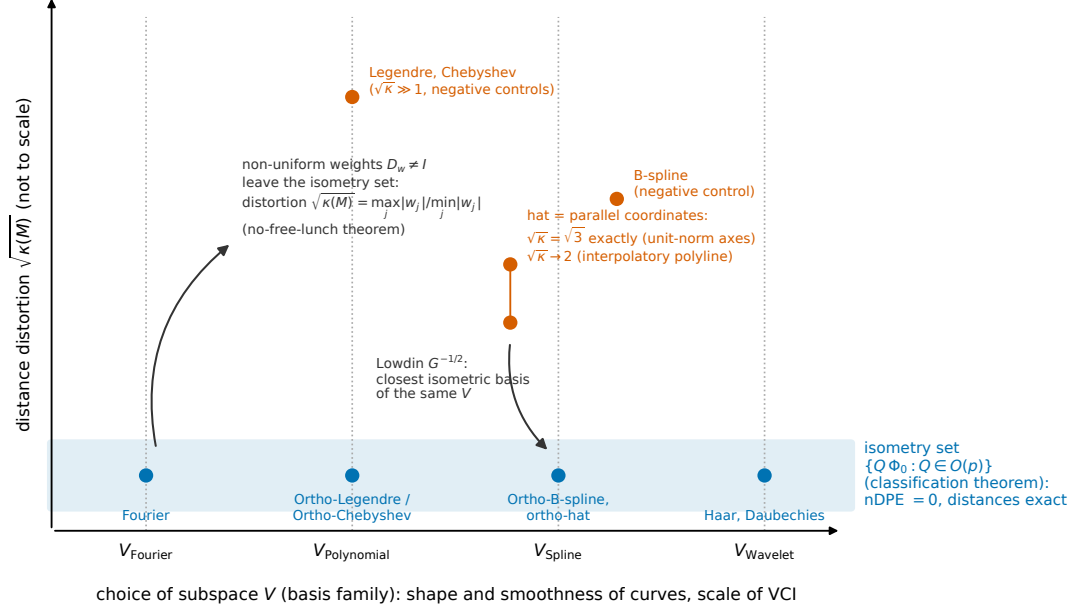
This section collects the material summarized in Section 5 of the main text: the geography of the design space, the discretization details, the nine datasets, the full definition and known limitations of each metric, the estimator audits, the per-experiment protocol table, the fold algorithm, and the reproducibility statement.

### S4.1 The design space

Fig. 7 draws the geography of the design space.

### S4.2 Discretization

In the implementation the  $L^2$  inner product is approximated by the rectangle rule on a grid of  $T = 500$  points, so the Euclidean distance of sampled curves is exactly  $1/\sqrt{\Delta t}$  times the  $L^2$  distance; this global constant cancels in all metrics used, and the discrete and continuous formulations of the classification theorem therefore correspond. Three consequences are worth recording.



**Fig. 7** The design space of generalized Andrews transforms (schematic; vertical axis not to scale). Choosing the subspace  $V$  (horizontal) fixes the shape and smoothness of the curves; within each  $V$ , the orthonormal bases form the orbit  $\{Q \Phi_0 : Q \in O(p)\}$ , the isometry set of the classification theorem of the main text, on which distances are exact. Parallel coordinates (hat basis) sit off the set in a band fixed by the axis convention alone:  $\sqrt{\kappa} = \sqrt{3}$  exactly for unit-norm axes and  $\sqrt{\kappa} \rightarrow 2$  for the interpolatory polyline that is usually drawn, and they are sent to their closest isometric representative (ortho-hat) by Löwdin orthonormalization; non-orthonormal polynomial and B-spline bases serve as negative controls. Non-uniform weighting leaves the isometry set with a distortion computable in advance from the no-free-lunch theorem

First, the rectangle rule is chosen over the trapezoidal rule precisely because it is exactly proportional to the Euclidean inner product of the sampled curves; without that agreement, nDPE would measure quadrature error rather than isometry. For the same reason Löwdin orthonormalization is applied to the Fourier basis as well, whose Gram matrix on the sampled grid departs from  $I$  by an  $O(1/T)$  artefact that would otherwise be misread as a disadvantage of the classical transform. Note also that **legendre** in the experiments denotes the sampled, uncorrected basis: its nonzero nDPE is to a substantial extent this discretization artefact, not a property of Legendre polynomials.

Second, the two axis conventions of the hat basis differ by more than a scale. Since  $\|h_k\|^2 = \frac{2h}{3}$  at interior knots but only  $\frac{h}{3}$  at the two boundary knots, normalizing to unit norm rescales the first and last axis by  $\sqrt{2}$  relative to the interior ones, so the exact value  $\kappa = 3$  describes a parallel-coordinates plot with reweighted outer axes rather than the textbook display.

Third, the limit statement  $\kappa \rightarrow 4$  for the interpolatory polyline is a continuum result, valid when  $T/p \gg 1$ . On a fixed grid the measured  $\kappa$  of the interpolatory basis in fact *decreases* over the range of  $p$  studied here (3.831 at  $p = 7$  against 3.675 at  $p = 30$ ), while the exact continuum values *increase* (3.898 against 3.993). The two directions are not in conflict, because the rectangle rule resolves each of the  $p - 1$  spline intervals with  $T/(p - 1)$  points: at fixed  $T$  the quadrature error grows with  $p$  and swamps the continuum trend. Experiment E15c settles this by refining the grid at fixed  $p$  instead (Table 10). Doubling  $T$  divides the relative error by 1.98 on average, so the error is  $O(1/T)$ ; at  $T = 32\,000$  the measured values are 3.896 and 3.987, that is within 0.03% and 0.14% of the exact ones, against 1.7% and 8.0% on the drawing grid

$T = 500$ . Across the whole sweep the relative error scaled by  $T/p$  stays in  $[1.13, 1.45]$ , which identifies the error as  $O(p/T)$  and the gap as quadrature error alone; the unit-norm convention, whose exact  $\kappa$  does not depend on  $p$ , never departs from 3 by more than 0.82% anywhere in the sweep. On the drawing grid itself the measured values are  $\kappa \in [2.988, 2.992]$  for the unit-norm basis and  $\kappa \in [3.67, 3.83]$  for the interpolatory polyline across the nine dataset dimensionalities, that is a measured distortion of 1.92 to 1.96 for the polyline.

Table 10: Convergence of the measured  $\kappa$  of the interpolatory hat basis to its exact continuum value as the quadrature grid is refined at fixed  $p$  (E15c). For each dimensionality the table gives  $\kappa$  measured on a grid of  $T$  points and its signed relative deviation from the exact value in percent. The exact value is  $\kappa(B)$  with  $B = \text{tridiag}(1, (2, 4, \dots, 4, 2), 1)$ , the continuum Gram matrix being  $\frac{h}{6}B$ . Each doubling of  $T$  halves the deviation, so the gap between the measured and the exact value is rectangle-rule error and nothing else; on a fixed grid it grows with  $p$ , because each spline interval is then covered by fewer grid points. The grid used for drawing throughout the study is  $T = 500$ .

| Grid $T$ | $p = 7$  |          | $p = 13$ |          | $p = 30$ |          |
|----------|----------|----------|----------|----------|----------|----------|
|          | $\kappa$ | Dev. [%] | $\kappa$ | Dev. [%] | $\kappa$ | Dev. [%] |
| 250      | 3.7664   | -3.36    | 3.6956   | -6.67    | 3.4191   | -14.36   |
| 500      | 3.8307   | -1.71    | 3.8230   | -3.46    | 3.6747   | -7.96    |
| 1 000    | 3.8639   | -0.86    | 3.8901   | -1.76    | 3.8268   | -4.15    |
| 2 000    | 3.8806   | -0.43    | 3.9247   | -0.89    | 3.9072   | -2.14    |
| 4 000    | 3.8890   | -0.22    | 3.9422   | -0.45    | 3.9495   | -1.08    |
| 8 000    | 3.8933   | -0.11    | 3.9510   | -0.22    | 3.9709   | -0.54    |
| 16 000   | 3.8954   | -0.05    | 3.9555   | -0.11    | 3.9817   | -0.27    |
| 32 000   | 3.8965   | -0.03    | 3.9577   | -0.06    | 3.9871   | -0.14    |
| Exact    | 3.8975   | –        | 3.9599   | –        | 3.9925   | –        |

### S4.3 Datasets

The study uses nine publicly available datasets, listed with their instance, attribute and class counts in Table 1 of the main text.

1. **Iris:** the classical botanical dataset, used to verify basic separability.
2. **Wine:** chemical analyses of 178 wines of three cultivars (13 attributes in the UCI distribution); tests the bases at medium dimensionality. This is not the same as Wine Quality: it is a distinct dataset by different authors [6].
3. **Breast Cancer Wisconsin:** a diagnostic dataset with 30 attributes [15, 17]; the high dimensionality stresses display readability.
4. **Heart Disease:** clinical data from the Cleveland Clinic [5]; for this study reduced to binary classification, absence of disease (grade 0) versus presence (grades 1–4 pooled); all 303 instances are retained.
5. **Wine Quality (White):** 4898 instances of white wine, 11 physico-chemical attributes [3]. The target is a sensory quality score on the scale 3–9, i.e. 7 classes. No discretization is performed; the original grades are kept so that plot labels match the data.
6. **Dry Bean:** a large agronomic dataset of 13 611 bean seeds with 7 morphological classes [13].
7. **Rice (Cammeo and Osmancik):** morphological features of two rice varieties [2].

8. **Early Stage Diabetes Risk:** a medical dataset (520 samples, 16 mostly binary categorical attributes, 2 classes) [11]. It tests the ability of the bases to handle binary categorical input, which differs markedly in character from the continuous morphological attributes of the other datasets.
9. **Seeds:** 210 instances of three wheat varieties (Kama, Rosa and Canadian) [1].

Detailed download links and archived versions of the data are available in the project documentation in `src/data/sources.md`.

#### S4.4 The six metrics in full

The formulas are given in Section 5 of the main text; what follows is the reading of each metric, its intended use and its known limitations.

**Integrated Class Separability (ICS).** ICS quantifies the global ability of the visualization to “unfold” multivariate class clusters into readable bundles of curves, extending Fisher’s linear discriminant into the function space. The scatters  $S_B(t)$  and  $S_W(t)$  are sums of squares without degrees-of-freedom normalization, so the ratio is not an  $F$  statistic, and the regularizer  $\epsilon = 10^{-9}$  breaks its exact scale invariance; at the energy normalization used the effect is negligible, which reinforces the rule that ICS is interpreted only relatively. **Meaning:** ICS states how large the average separation of the class means is relative to their internal variability along the whole interval. **Utility:** High ICS indicates that classes are separated along the value axis at many times; whether a human observer actually reads them apart is a perceptual question that this paper does not test. **Limitation:** The metric is not bounded above; under near-perfect separation ( $S_W \rightarrow 0$ ) it grows without limit. It is therefore interpreted strictly relatively (by ranking bases), never by absolute value. Because the scatters are unnormalized sums of squares, both scale linearly with  $n$  and their ratio does not. Experiment E24 checks this directly on Dry Bean, the largest dataset of the study: over five subsample sizes and five draws each, the sample varies by a factor of 10 (from 13611 down to 1361 instances) while ICS stays between 5.41 and 5.96, a factor of 1.10 with no monotone trend. Had ICS scaled with the sample size, the two factors would have matched. What ICS does and does not certify about the *design* is the subject of audit E17, reported in the main text.

**Normalized Distance Preservation Error (nDPE).** **Reference:**  $D_X$  is always computed over the *unweighted* standardized space  $Z$  (test fold); thanks to this the metric also captures the deformation caused by weighting, which measurably breaks the isometry. Where the evaluated sample exceeds 2000 rows the matrices are computed on a deterministic subsample of that size (seed 42); since the sample here is the test fold, this triggers for Dry Bean alone. **Meaning:** nDPE indicates how faithfully “closeness” in the plot represents “similarity” in the data. A value of 0 denotes perfect isometry.

**Neighbourhood Preservation Score (NPS).** **Meaning:** NPS states what fraction of a point’s local neighbourhood survives the transformation. We use  $k = 5$ ; neighbourhoods are searched within the evaluated sample (the test fold, not the full dataset) on the same deterministic query subsample as nDPE, and the reference is again the unweighted standardized space  $Z$ . **External validation (audit E18):** Five of the six metrics are introduced in this paper, and NPS is checked against instruments the community already trusts. Audit E18 recomputes, over all 63 (dataset, basis) cells of E1 and under the same fold protocol, the trustworthiness and continuity measures of Kaski et al. [12] and Venna and Kaski [16] at the same neighbourhood size, and compares them with NPS. The agreement is monotone and strong (Spearman  $\rho = 0.94$  against trustworthiness and 0.93 against continuity over all cells), and it is not merely

the isometric-versus-non-isometric contrast: restricted to the non-orthonormal cells alone, where there is real variation to explain,  $\rho$  is still 0.64 and 0.58. The two are nevertheless not interchangeable on an absolute scale, and in a direction worth stating: NPS is the harsher instrument, falling to 0.71–0.76 on the bases where trustworthiness stays above 0.98. We therefore read NPS comparatively, as everywhere else in this paper.

**Visual Complexity Index (VCI).** To support non-differentiable bases (Haar) VCI is defined as the mean discrete roughness (second difference) normalized by the root-mean-square value of the curve,  $\text{rms}(f_i) = \sqrt{T^{-1} \sum_m f_{i,m}^2}$ . **Meaning:** VCI measures “visual noise”: the more the curves wiggle, the higher the VCI. **Scaled variant:** The RMS normalization makes the metric independent of curve scale, but the value depends on the density of the sampling grid, so comparisons are valid only at identical  $T$  (in this study  $T = 500$  for all bases). For comparisons across resolutions we introduce

$$\widetilde{\text{VCI}} = (T - 1)^2 \cdot \text{VCI}. \quad (2)$$

For twice-differentiable bases the scaled variant (2) converges, up to an interval-length constant, to  $\int |f''| dt / \text{rms}(f)$  and is asymptotically independent of  $T$ ; for discontinuous wavelets (Haar, Daubechies) no local roughness functional converges, since jumps have unbounded continuum roughness, and  $\widetilde{\text{VCI}}$  grows  $\propto T$ . At fixed  $T$  both variants give identical basis rankings. The grid behaviour of both is audited in the main text.

**Curve Overlap Index (COI).** The average runs over the *admissible* classes  $\mathcal{C}_{\text{eff}}$ , those represented by at least two instances in the evaluated sample,  $\mathcal{C}_{\text{eff}} = |\mathcal{C}_{\text{eff}}|$ . A single instance admits no bandwidth estimate, so a class below that threshold is dropped rather than assigned a degenerate density. The distinction is not academic: under 5-fold cross-validation the rarest Wine Quality class (quality 9, five instances in total) contributes exactly one test-fold instance and is therefore dropped in *all five* folds, so COI is averaged over 15 class pairs there rather than the nominal 21 (audit E16). On the remaining eight datasets  $\mathcal{C}_{\text{eff}} = \mathcal{C}$  in every fold, so this affects one dataset of nine. The exclusion is from the density estimation and from the pair average; the values of the dropped instances still delimit the range over which the inner integral is discretized. On Wine Quality even that residual influence vanishes: the dropped instance lies inside the value range of the remaining classes in all five folds, and recomputing COI with those rows removed reproduces the published values exactly. We use a Gaussian kernel with bandwidth given by the robust variant of Silverman’s rule:

$$h_c(t) = 0.9 \cdot \min\left(\hat{\sigma}_c(t), \frac{IQR_c(t)}{1.34}\right) \cdot n_c^{-1/5} \quad (3)$$

The inner integral in the definition of COI is computed by the trapezoidal rule on a grid of 256 points covering the value range extended by  $3h$  on each side, with  $h$  given by (3). The KDE is implemented by the binned method (binning to a common grid followed by Gaussian filtering), which is an order of magnitude faster than the direct sum; audit E16 quantifies the approximation by recomputing the profile with an exact Gaussian KDE on a 2048-point grid at identical bandwidths, and the two agree to  $4.7 \cdot 10^{-3}$  at worst pointwise and to  $5 \cdot 10^{-4}$  in the aggregate on all nine datasets. One further branch is exercised only in degenerate slices: where every admissible class has zero bandwidth (constant values), no density exists, and the pairwise overlap is set to 1 for classes sharing the same constant and 0 otherwise. **Meaning:** COI states how much the “shadows” (densities) of the individual classes mix. A value of 1 means identical distributions. The lower endpoint is nominal rather than attainable: a strictly positive Gaussian kernel makes the two densities overlap everywhere, so COI has a positive floor that depends on the bandwidth and the sample size, and the metric, like ICS, is interpreted comparatively. **Implementation:** In the tabulated results the outer sum is evaluated on a grid subset with

step 5 (i.e. 100 slices out of  $T = 500$ ), because KDE at all  $T$  points would be the computational bottleneck; COI heatmaps use the full grid.

**1-NN classifier accuracy in curve space.** The sixth metric is the accuracy of a 1-NN classifier trained on the curves of the training fold and evaluated on the test fold. **Reason for inclusion:** ICS is a Fisher ratio of between- to within-class variance, and weighting by the ANOVA F-score is also a Fisher ratio, so measuring the benefit of weighting by ICS would be a circular argument (“maximizing the F-score increases the F-score”). The 1-NN accuracy is independent of ICS, is evaluated on held-out data (thus also capturing overfitting of the weights to the training fold), and is therefore the appropriate criterion whenever the benefit of weighting is at issue; in this paper it belongs to the verification group.

## S4.5 Evaluation protocol and preprocessing

The comparative experiments of this paper (E1, E8, E10, E10b, E11, E12, E21 and E22) were produced by stratified 5-fold cross-validation. Standardization (z-score parameters), discriminative weights and any supervised attribute ordering are estimated exclusively from the training fold; metrics are computed on the test fold. Several parts of the study deliberately use a different protocol, and Table 11 states which, so that no result of this paper has to be read under a blanket guarantee: the exploratory gallery, the clustering check E5 and the audits E17 and E19 are computed on the full dataset, the invariance audit E7 on a single stratified split, the hat-basis audit E15 is analytic and uses no data at all, and the timing benchmark runs on the full Dry Bean data. For E5 and E7 the choice is deliberate rather than an oversight: both are implementation verifications of an exact algebraic identity rather than estimates of out-of-sample quality, and a fold split would only add noise to a quantity whose predicted value is zero.

Without the train–test separation, the weights would see the same labels against which the quality of the visualization is subsequently judged (label leakage). Results are aggregated as mean, standard deviation and 95 % confidence interval over folds (Student’s  $t$ -distribution,  $df = k - 1$ ). The number of folds is limited by the size of the smallest class: in Wine Quality the rarest class (quality 9) has five samples, so each fold contains at least one. One sample is nevertheless below the admissibility threshold of the COI estimator, which needs at least two per class, so that class is excluded from COI on that dataset. Its effect on ICS runs the other way and is worth stating, because it is not neutral: a class represented by a single instance in a fold contributes nothing to the within-class scatter  $S_W(t)$ , while it still contributes its full squared deviation from the grand mean, weighted by  $n_c = 1$ , to the between-class scatter  $S_B(t)$ . A singleton class can therefore only raise the measured ICS of every design alike. Since ICS is read comparatively across designs on the same dataset, the ranking is unaffected, but the absolute Wine Quality value is inflated and is not comparable with the other datasets. We keep the class rather than pooling the extreme grades, because the alternative would silently redefine the dataset, but every class-conditional statement about Wine Quality is to be read as being about its six well-populated classes. The “variance” ordering is computed from the variances of the raw training data (after standardization all columns have variance 1 and the ordering would be degenerate). All experiments of this paper keep the original attribute order, and non-uniform weights enter only the distortion audit E8; the two levers are not exercised here as design choices, only priced. During data download, missing values are imputed by the column mean and categorical attributes are integer-encoded (LabelEncoder); both happen before the CV split (this concerns the Heart Disease dataset only). We declare the mean imputation explicitly as a deliberate, label-free preprocessing exception: the imputation constant is computed on the full dataset and therefore uses test-fold feature values (a mild, unsupervised form of train–test contamination), affecting only the handful of missing entries in Heart Disease; no labels enter this step.

Table 11: Evaluation protocol per experiment. Not every experiment uses cross-validation: the gallery, the clustering check and three of the audits deliberately do not, for the reasons given in the text. The labels index the computational pipeline of the accompanying software package, which is broader than this study, so they are not contiguous here. The last column names the place where each experiment is reported.

| Exp. | Role                                          | Protocol                | Reported in                                        |
|------|-----------------------------------------------|-------------------------|----------------------------------------------------|
| –    | Gallery of visual profiles                    | full data, no split     | main text, controlled comparison;<br>Supplement S6 |
| E1   | Seven bases $\times$ nine datasets            | stratified 5-fold CV    | main text, controlled comparison;<br>Supplement S2 |
| E5   | K-Means in curve space vs. $\mathbb{R}^p$     | full data, no split     | main text, benchmark audit                         |
| E7   | Invariance on the $O(p)$ orbit                | single stratified split | main text, benchmark audit                         |
| E8   | nDPE prediction and bound                     | stratified 5-fold CV    | main text, controlled comparison                   |
| E10  | PCP / RadViz / Star baselines                 | stratified 5-fold CV    | main text, controlled comparison                   |
| E10b | Anchor-order sensitivity                      | stratified 5-fold CV    | Supplement S5                                      |
| E11  | Ablation: $T$ , degree, wavelet, <b>rcond</b> | stratified 5-fold CV    | main text, benchmark audit; Supplement S3          |
| E12  | Ablation: COI bandwidth                       | stratified 5-fold CV    | main text, benchmark audit; Supplement S3          |
| E15  | Hat-basis axis conventions                    | analytic, data-free     | Supplements S1, S3 and S4                          |
| E16  | COI estimator audit                           | stratified 5-fold CV    | this section                                       |
| E17  | ICS localization audit                        | full data, no split     | main text, benchmark audit; Supplement S3          |
| E18  | NPS against trustworthiness/continuity        | stratified 5-fold CV    | this section                                       |
| –    | Timing benchmark                              | full Dry Bean, no split | Supplement S5                                      |
| E19  | Clutter measures against the grid             | full data, no split     | main text, benchmark audit; Supplement S3          |
| E20  | Stability of the located window               | full data, bootstrap    | Supplement S3                                      |
| E21  | Dimensionality-reduction reference            | stratified 5-fold CV    | main text, controlled comparison;<br>Supplement S5 |
| E22  | Nine in-class techniques, one protocol        | stratified 5-fold CV    | main text, controlled comparison;<br>Supplement S5 |
| E23  | Gallery of orbit representatives              | full data, no split     | main text, controlled comparison                   |
| E24  | ICS against the sample size                   | full data, subsampled   | this section                                       |

Algorithm 1 describes the evaluation of one cross-validation fold exactly as implemented in the accompanying code (module `src/evaluation.py`). Its call to `GETBASIS` carries the Löwdin correction: the correction is applied to every basis whose name bears the “ortho\_” prefix and, for the reason given above, to the Fourier basis as well. The total complexity of the transformation including orthonormalization is  $O(npT + p^2T + p^3)$ : the  $O(p^2T)$  term for the Gram matrix at  $T = 500$  dominates  $O(p^3)$  for all  $p$  in this study. Fig. 8 places the fold evaluation in the pipeline as a whole.

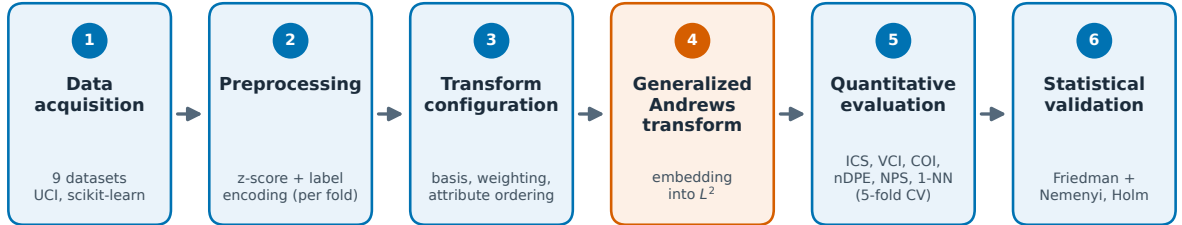
---

**Algorithm 1** Evaluation of a single cross-validation fold (after `evaluate_fold`)

---

- 1: **Input:** training fold  $(X_{tr}, y_{tr})$ , test fold  $(X_{te}, y_{te})$
  - 2: **Input:** basis  $B$ , weighting method  $W$ , ordering  $\pi$
  - 3:  $(\mu, \sigma) \leftarrow \text{FITZSCORE}(X_{tr})$  *training fold only*
  - 4:  $Z_{tr} \leftarrow \frac{X_{tr} - \mu}{\sigma}$
  - 5:  $Z_{te} \leftarrow \frac{X_{te} - \mu}{\sigma}$
  - 6:  $\mathbf{w} \leftarrow \text{CALCULATEWEIGHTS}(Z_{tr}, y_{tr}, W)$  *training fold only*
  - 7:  $\mathbf{w} \leftarrow \mathbf{w} \sqrt{p / \|\mathbf{w}\|^2}$  *energy normalization*
  - 8:  $\pi \leftarrow \text{RESOLVEORDERING}(\pi, Z_{tr}, y_{tr}, \sigma_{raw}^2)$  *training fold only*
  - 9:  $\Phi \leftarrow \text{GETBASIS}(B, p, T)$  *Löwdin-corrected*
  - 10:  $F_{tr} \leftarrow (Z_{tr}[\cdot, \pi] \odot \mathbf{w}_\pi) \Phi$
  - 11:  $F_{te} \leftarrow (Z_{te}[\cdot, \pi] \odot \mathbf{w}_\pi) \Phi$
  - 12: **Output:** ICS, VCI, COI on  $F_{te}$
  - 13: **Output:** nDPE, NPS of  $F_{te}$  against unweighted  $Z_{te}$
  - 14: **Output:** `kNN_acc`( $F_{tr} \rightarrow F_{te}$ )
- 

**Experimental pipeline**



**Fig. 8** Process-flow diagram of the experimental pipeline for the evaluation of generalized Andrews curves, from data acquisition to statistical validation

### S4.6 Ethics and reproducibility

All nine datasets are public, de-identified archives from the UCI Machine Learning Repository or scikit-learn; no new human data were collected for this study, so no ethics approval was required.

To ensure reproducibility, all source code is released as a public repository at <https://github.com/MartinRaSt/andrews-distortion-calculus> and will be archived with a citable DOI on acceptance. Every number, table and figure in this paper is regenerated from the raw data by two commands, with no result transcribed by hand, and every experiment is checkpointed per row so an interrupted run resumes rather than recomputing. All experiments were run with the fixed seed `RANDOM_SEED = 42`, which covers the fold splits, the subsampling inside the metrics and the weighting estimators. The environment is pinned to exact versions: Python 3.12 with NumPy 2.2.6, SciPy 1.17.1, scikit-learn 1.8.0 [14], pandas 2.3.3, PyWavelets 1.9.0 and Matplotlib

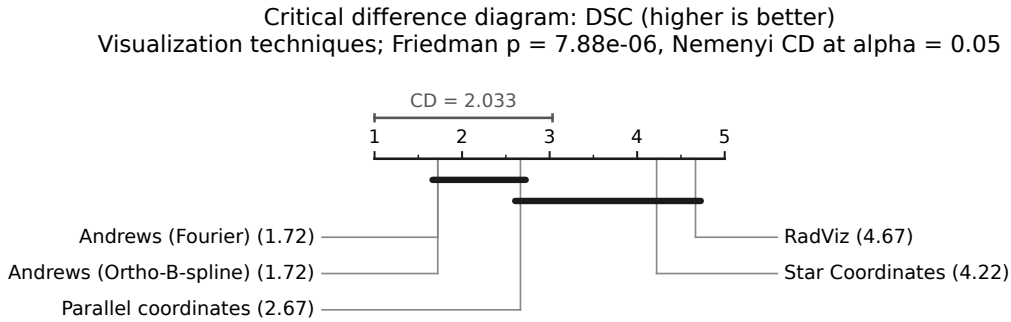
3.10.8; the full snapshot ships with the package. Regeneration is verified to be byte-identical across runs by the package’s own test suite.

The computational-cost measurements reported in Supplement S5 were performed on an AMD Ryzen 9 9950X CPU (16 cores), 128 GB RAM, Windows 10, over the full Dry Bean dataset. Times are averages of 100 repetitions (basis generation) and 30 repetitions (matrix transformation), with a warm-up run excluded from the measurement (`time.perf_counter`).

## S5 Sensitivity of the technique comparison and out-of-class details

### S5.1 The conservative Nemenyi gate at $k = 5$

The main text carries the nine-technique inference with bootstrap intervals, an omnibus permutation test and four pre-specified contrasts, because at  $k = 9$  the studentized-range procedures cannot separate any pair by construction. Fig. 9 adds the conservative gate that can still be read at  $k = 5$ : restricted to the five classical techniques the Nemenyi critical difference is 2.03 rather than 4.00, and at that resolution both Andrews variants are significantly better than both planar projections, with parallel coordinates in between and separable from neither group. This is the group separation the main text reports, obtained by the most conservative procedure available at this sample size.



**Fig. 9** Critical difference diagram for Distance Consistency (higher is better; Friedman with Nemenyi post-hoc,  $\alpha = 0.05$ , nine datasets), restricted to the five classical techniques so that the critical difference is 2.03 rather than 4.00. It is to be read as a conservative gate: both Andrews variants are significantly better than both planar projections, and parallel coordinates fall in between and are separable from neither group at this  $n$ . The same diagram drawn over all nine techniques would connect every pair, which is why the nine-technique evidence is presented in the main text as bootstrap intervals instead

### S5.2 Anchor-order sensitivity of RadViz and Star Coordinates (E10b)

RadViz and Star Coordinates depend on the order in which the per-attribute anchors are placed on the circle, and the main text fixes that order to the dataset column order. Audit E10b measures how much this choice matters. Across ten seeded anchor permutations the Distance Consistency of the two techniques spans 0.20 (RadViz) and 0.17 (Star) per dataset on average, at most 0.31, and their 1-NN accuracy 0.18 and 0.17 on average, at most 0.39. The spread is therefore not negligible, and we do not claim that the reported configuration is optimal. It does not, however, change the ordering of the techniques: even the best of the ten permutations exceeds the weakest curve-type technique in Distance Consistency on only one of the nine datasets (Iris).

### S5.3 Control for the input normalization

The curve techniques receive z-scored input while RadViz and Star Coordinates use the min-max normalization of the literature, fitted on the training fold, and nDPE and NPS are measured for all techniques against the same z-scored reference space. Part of the projections' reported loss could in principle precede any projection at all, so we report a control: nDPE and NPS recomputed for every technique against the reference space matched to its *own* input normalization. Running RadViz on z-scored input is not an available control, because min-max normalization is invariant under per-attribute affine maps, so the min-max of the z-scores is exactly the min-max of the raw data and the representation would be unchanged; only the reference space can be varied. The correction is small and changes nothing: RadViz moves from nDPE 0.451 to 0.442 and NPS 0.275 to 0.289, Star from 0.396 to 0.380 and 0.287 to 0.300, with the curve techniques unaffected by construction. The gap to the curve techniques is an order of magnitude larger than the normalization effect, so the ranking of the main text stands.

### S5.4 A second dimensionality: the same techniques on Wine

Fig. 10 repeats the four-technique comparison of the main text on a dataset of higher dimension.

### S5.5 Out-of-class reference block: protocol and per-dataset values (E21)

Experiment E21 runs PCA, t-SNE and UMAP through the same leakage-free 5-fold protocol as the in-class comparison and scores them with the same technique-neutral metrics, so that the reference values of the main text are produced by an identical pipeline rather than quoted from the literature. Three protocol points matter for reading them. First, PCA and UMAP are fitted on the training fold and applied to the held-out fold, which is the same train-then-apply discipline used for the standardizer and the weights everywhere else in this paper. Second, t-SNE has no out-of-sample extension, so it is fitted on the union of the training and test folds; the resulting transductive advantage is one of the reasons its numbers are not comparable with the in-class ranking. Third, for the same reason we also report a joint-fit UMAP variant as an internal audit of how much of the difference is attributable to the fitting regime rather than to the method.

None of these values enters a rank, a Friedman test or a critical difference diagram anywhere in this paper, for the reason given in the main text: the techniques do not satisfy the data contract that defines the comparison class, so a ranking across the two classes would credit the curve techniques for a structural column and penalize the projections for a problem they do not pose.

Table 12 gives the per-dataset values and the hyperparameters actually used.

### S5.6 Audit of the distortion predictions (E8)

Fig. 11 shows the audit summarized in the main text: the nDPE predicted from the metric matrix  $M$  against the nDPE measured on the curves, for all 324 combinations of dataset, basis and weighting, together with the closed-form upper bound. The lower-left cluster of panel (a) consists of the exact zeros of the unweighted orthonormal bases, and panel (b) shows that no measured value reaches the bound.

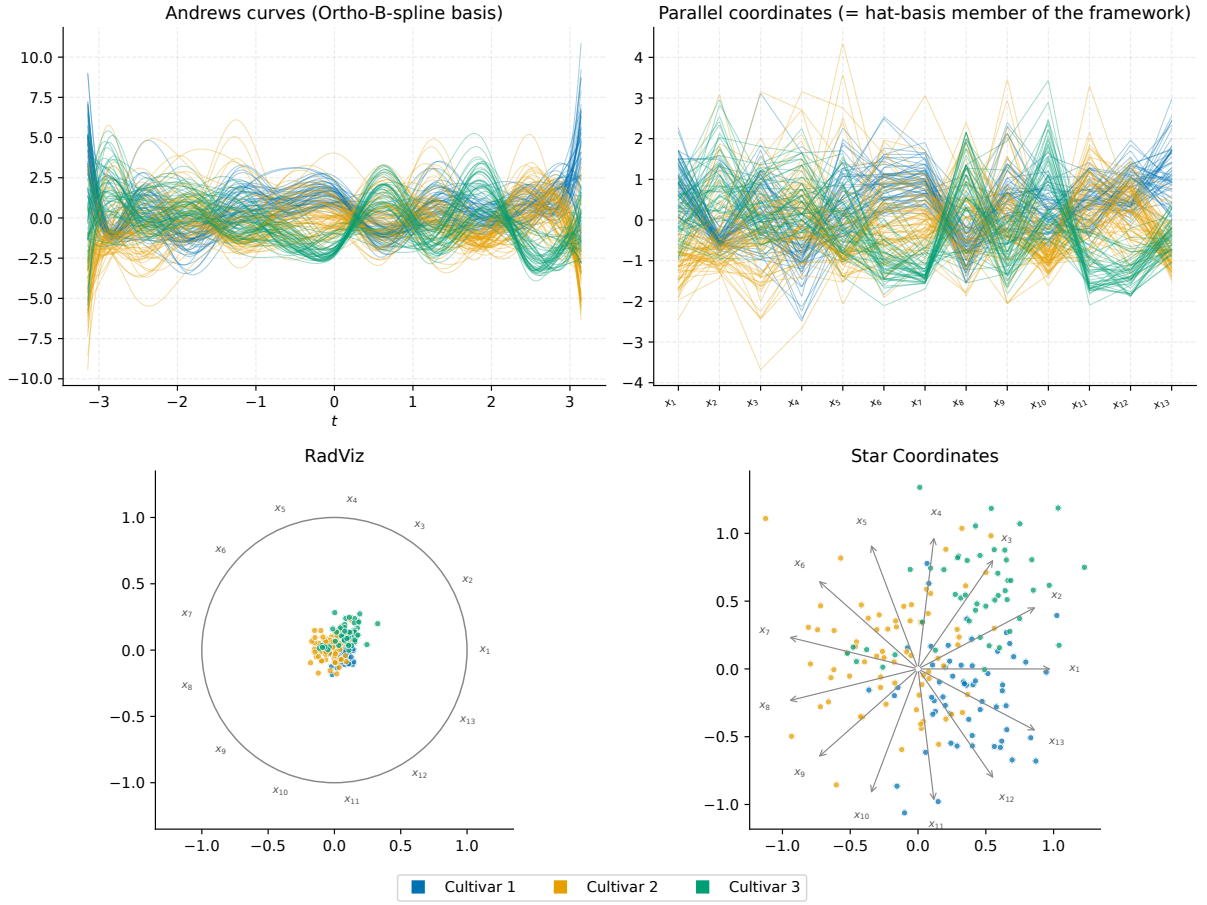
### S5.7 Computational cost of the bases

Table 13 gives the measured transformation cost per basis on the largest dataset. All bases are inexpensive at this scale; the orthonormalized variants pay a one-off Gram-matrix construction that is amortized over all curves of a dataset.

Table 12: Per-dataset values behind the aggregated reference block of the main text (experiment E21). Each cell is the mean over the five cross-validation folds; the spread over folds is in the `*_std` and `*_ci95` columns of `results/data/e21_dr_reference.csv`, from which this table is generated. ICS is omitted here for the same reason as in the main text. The last row of each block is the transductive audit described in the main text. Hyperparameters actually used, recorded per run: per-row values in the `params` column of the CSV.

| Dataset              | Technique               | nDPE  | NPS   | 1-NN  | DSC   |
|----------------------|-------------------------|-------|-------|-------|-------|
| Breast Cancer        | PCA                     | 0.226 | 0.375 | 0.912 | 0.933 |
|                      | t-SNE                   | 0.419 | 0.482 | 0.954 | 0.935 |
|                      | UMAP                    | 0.420 | 0.386 | 0.935 | 0.931 |
|                      | UMAP (joint fit, audit) | 0.411 | 0.455 | 0.946 | 0.907 |
| Dry Bean             | PCA                     | 0.152 | 0.156 | 0.814 | 0.848 |
|                      | t-SNE                   | 0.359 | 0.469 | 0.900 | 0.888 |
|                      | UMAP                    | 0.317 | 0.236 | 0.878 | 0.898 |
|                      | UMAP (joint fit, audit) | 0.320 | 0.314 | 0.879 | 0.891 |
| Early Diabetes       | PCA                     | 0.352 | 0.328 | 0.933 | 0.833 |
|                      | t-SNE                   | 0.373 | 0.482 | 0.967 | 0.844 |
|                      | UMAP                    | 0.523 | 0.386 | 0.854 | 0.677 |
|                      | UMAP (joint fit, audit) | 0.508 | 0.418 | 0.935 | 0.677 |
| Heart Disease        | PCA                     | 0.354 | 0.307 | 0.710 | 0.822 |
|                      | t-SNE                   | 0.390 | 0.488 | 0.775 | 0.759 |
|                      | UMAP                    | 0.439 | 0.385 | 0.753 | 0.835 |
|                      | UMAP (joint fit, audit) | 0.440 | 0.482 | 0.762 | 0.805 |
| Iris                 | PCA                     | 0.059 | 0.876 | 0.853 | 0.867 |
|                      | t-SNE                   | 0.333 | 0.841 | 0.953 | 0.833 |
|                      | UMAP                    | 0.358 | 0.776 | 0.893 | 0.907 |
|                      | UMAP (joint fit, audit) | 0.378 | 0.816 | 0.940 | 0.867 |
| Rice                 | PCA                     | 0.134 | 0.357 | 0.891 | 0.913 |
|                      | t-SNE                   | 0.251 | 0.629 | 0.883 | 0.918 |
|                      | UMAP                    | 0.276 | 0.493 | 0.887 | 0.912 |
|                      | UMAP (joint fit, audit) | 0.269 | 0.570 | 0.882 | 0.916 |
| Seeds                | PCA                     | 0.117 | 0.727 | 0.890 | 0.924 |
|                      | t-SNE                   | 0.231 | 0.719 | 0.924 | 0.924 |
|                      | UMAP                    | 0.268 | 0.626 | 0.905 | 0.924 |
|                      | UMAP (joint fit, audit) | 0.269 | 0.665 | 0.924 | 0.924 |
| Wine                 | PCA                     | 0.287 | 0.609 | 0.961 | 0.961 |
|                      | t-SNE                   | 0.390 | 0.705 | 0.944 | 0.933 |
|                      | UMAP                    | 0.401 | 0.575 | 0.916 | 0.933 |
|                      | UMAP (joint fit, audit) | 0.467 | 0.644 | 0.944 | 0.927 |
| Wine Quality (White) | PCA                     | 0.334 | 0.099 | 0.563 | 0.200 |
|                      | t-SNE                   | 0.374 | 0.440 | 0.633 | 0.203 |
|                      | UMAP                    | 0.412 | 0.289 | 0.477 | 0.183 |
|                      | UMAP (joint fit, audit) | 0.472 | 0.366 | 0.557 | 0.174 |

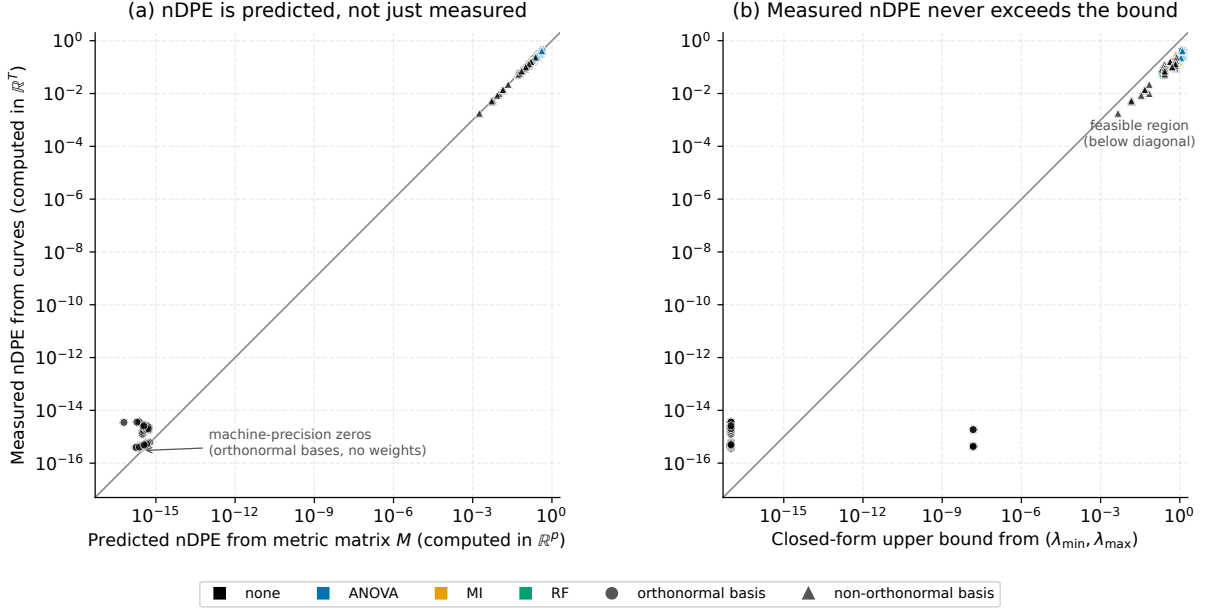
One dataset, four techniques of the same design space (Wine,  $p = 13$ , z-score for curves, min--max for 2D projections)



**Fig. 10** The four in-class techniques on Wine ( $p = 13$ , three cultivars), which differs from Seeds in dimension rather than in the number of classes. The two attribute-preserving displays keep their resolution, whereas RadViz contracts almost the whole sample into a small central cloud: with thirteen anchors spread around the circle the barycentric average of the normalized attributes concentrates near the centre, and the classes overlap there. This is the aggregation cost quantified for the whole comparison class by the nDPE and NPS columns of the main-text ranking, shown here on one dataset

Table 13: Transformation cost per basis on the largest dataset (Dry Bean): basis generation (mean of 100 runs) and matrix transform of the whole dataset (mean of 30 runs), in milliseconds; warm-up run excluded.

| Basis            | Basis generation (ms) | Transform (ms) | Total (ms) |
|------------------|-----------------------|----------------|------------|
| Fourier          | 0.23                  | 11.51          | 11.75      |
| Chebyshev        | 0.41                  | 12.37          | 12.78      |
| B-spline         | 0.99                  | 11.88          | 12.87      |
| Daubechies (db4) | 0.50                  | 13.28          | 13.77      |
| Haar             | 0.49                  | 13.29          | 13.78      |
| Ortho-B-spline   | 1.15                  | 12.73          | 13.88      |
| Legendre         | 0.89                  | 13.65          | 14.54      |



**Fig. 11** Distortion calculus (E8). (a) nDPE predicted from the metric matrix  $M$  in  $\mathbb{R}^p$  agrees with the value measured on curves in  $\mathbb{R}^T$  to machine precision for all 324 combinations; the lower-left cluster consists of the exact zeros of unweighted orthonormal bases. (b) Measured nDPE never exceeds the closed-form upper bound derived from  $(\lambda_{\min}, \lambda_{\max})$ , reaching at most 48 % of it. Colour encodes the weighting method, marker the orthonormality of the basis

### S5.8 Full means of the in-class comparison

The in-class table of the main text reports mean ranks, which is the appropriate summary across datasets of very different scales. Table 14 gives the underlying means with their standard deviations for all nine techniques and every metric.

### S5.9 Does the geometric price buy anything? (E25)

The distortion calculus prices a weighting scheme without looking at the data, and the audit of the distortion predictions in the main text reports prices spanning four orders of magnitude. That number is a statement about the  $L^2$  geometry alone. Experiment E25 pairs it with the other side of the trade by joining the aggregates of E3, which measures visualization quality under four weighting schemes, with those of E8, which measures the distortion of the same combinations. For every (dataset, basis, scheme) cell with finite  $\kappa$  we take the difference against the unweighted design of the same basis on the same dataset and correlate it with  $\log \sqrt{\kappa(M)}$ .

Two of the six metrics move with the price and two do not, and the split is instructive. Class overlap improves, and improves more where more was paid (Spearman  $\rho = +0.47$ ,  $p < 0.001$  over 52 cells). Class separability improves substantially on average (+1.139 in ICS) but without any relation to the price ( $\rho = +0.04$ ,  $p = 0.779$ ), which is what the ICS localization audit of the main text leads one to expect of a metric that rewards localization. Discriminability does not move at all ( $\rho = +0.00$ ,  $p = 0.990$ ). The two verification metrics degrade with the price, which is not a finding but the theorem restated: nDPE and NPS are algebraically tied to  $\kappa$ .

The practical reading is narrower than “weighting is expensive” and more useful. The price is real, it is computable in advance, and it is paid for a reduction of class overlap; it is not paid for accuracy, and a reader who assumes otherwise is assuming something these data do not support.

Table 14: In-class comparison extended by four techniques published in 2022–2023 (experiment E22, upper block) against the baselines of experiment E10 (lower block). Identical protocol for every row: stratified 5-fold cross-validation with the same seed and the same splits; scaler, min–max parameters, anchor order and rotation are fitted on the training fold only and the metrics are computed on the held-out fold. Cells are mean  $\pm$  standard deviation across the 9 datasets of the per-dataset cross-validated means. nDPE and NPS are measured against the common z-scored reference space. Crossings are the mean number of intersections per pair of curves on a 200-row subsample; they are undefined for the scatter displays (RadViz3D, RadViz, Star Coordinates), where the cell carries a dash. The anchor order of the optimized RadViz minimizes the effectiveness error and is searched on the training fold. The optimally smooth Andrews representative is the Fourier transform of the training-fold PCA scores, i.e. another element of the same  $O(p)$  isometry orbit – which is why its nDPE and NPS coincide with the classical Andrews row while its curves are measurably smoother.

| Technique                             | ICS               | nDPE              | NPS               |
|---------------------------------------|-------------------|-------------------|-------------------|
| RadViz3D                              | $2.251 \pm 3.746$ | $0.388 \pm 0.082$ | $0.354 \pm 0.157$ |
| RadViz (optimized anchor order)       | $2.525 \pm 3.372$ | $0.391 \pm 0.088$ | $0.335 \pm 0.170$ |
| Angle-uniform parallel coordinates    | $2.783 \pm 3.627$ | $0.159 \pm 0.061$ | $0.736 \pm 0.080$ |
| Andrews (optimally smooth orbit rep.) | $1.487 \pm 1.740$ | $0.000 \pm 0.000$ | $0.998 \pm 0.004$ |
| Andrews (Fourier)                     | $1.179 \pm 1.399$ | $0.000 \pm 0.000$ | $0.998 \pm 0.005$ |
| Andrews (Ortho-B-spline)              | $2.451 \pm 3.463$ | $0.000 \pm 0.000$ | $0.998 \pm 0.004$ |
| Parallel coordinates                  | $2.491 \pm 3.166$ | $0.076 \pm 0.017$ | $0.825 \pm 0.053$ |
| RadViz                                | $1.891 \pm 3.291$ | $0.451 \pm 0.094$ | $0.275 \pm 0.163$ |
| Star Coordinates                      | $1.578 \pm 2.174$ | $0.396 \pm 0.053$ | $0.287 \pm 0.174$ |

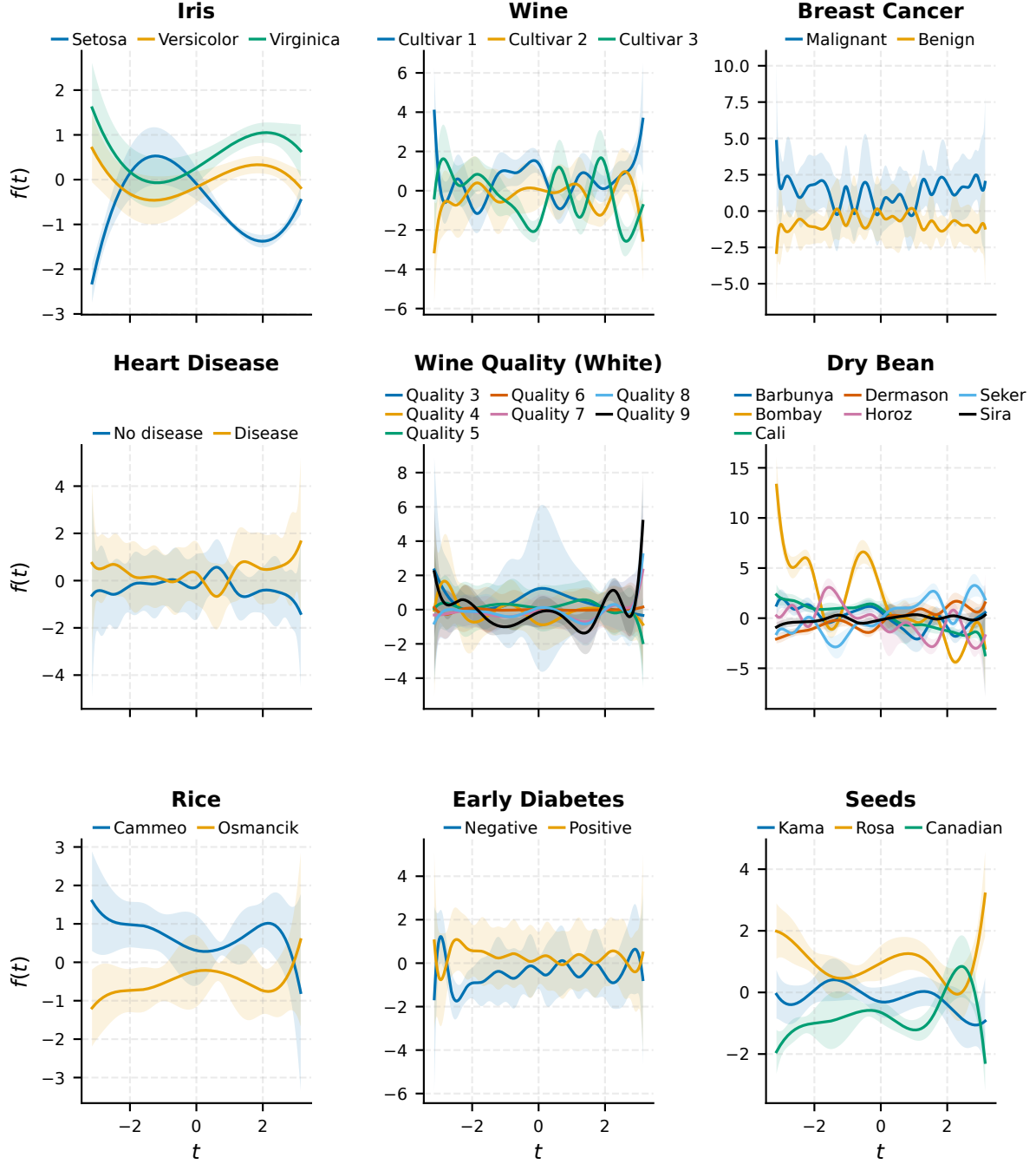
  

| Technique                             | 1-NN              | DSC               | Crossings     |
|---------------------------------------|-------------------|-------------------|---------------|
| RadViz3D                              | $0.784 \pm 0.101$ | $0.737 \pm 0.192$ | –             |
| RadViz (optimized anchor order)       | $0.774 \pm 0.115$ | $0.762 \pm 0.201$ | –             |
| Angle-uniform parallel coordinates    | $0.882 \pm 0.111$ | $0.824 \pm 0.208$ | $3.7 \pm 2.3$ |
| Andrews (optimally smooth orbit rep.) | $0.881 \pm 0.113$ | $0.841 \pm 0.199$ | $3.4 \pm 2.1$ |
| Andrews (Fourier)                     | $0.881 \pm 0.113$ | $0.841 \pm 0.199$ | $7.7 \pm 4.6$ |
| Andrews (Ortho-B-spline)              | $0.881 \pm 0.113$ | $0.841 \pm 0.199$ | $6.0 \pm 3.6$ |
| Parallel coordinates                  | $0.876 \pm 0.110$ | $0.827 \pm 0.201$ | $3.7 \pm 2.3$ |
| RadViz                                | $0.704 \pm 0.131$ | $0.684 \pm 0.178$ | –             |
| Star Coordinates                      | $0.724 \pm 0.128$ | $0.718 \pm 0.185$ | –             |

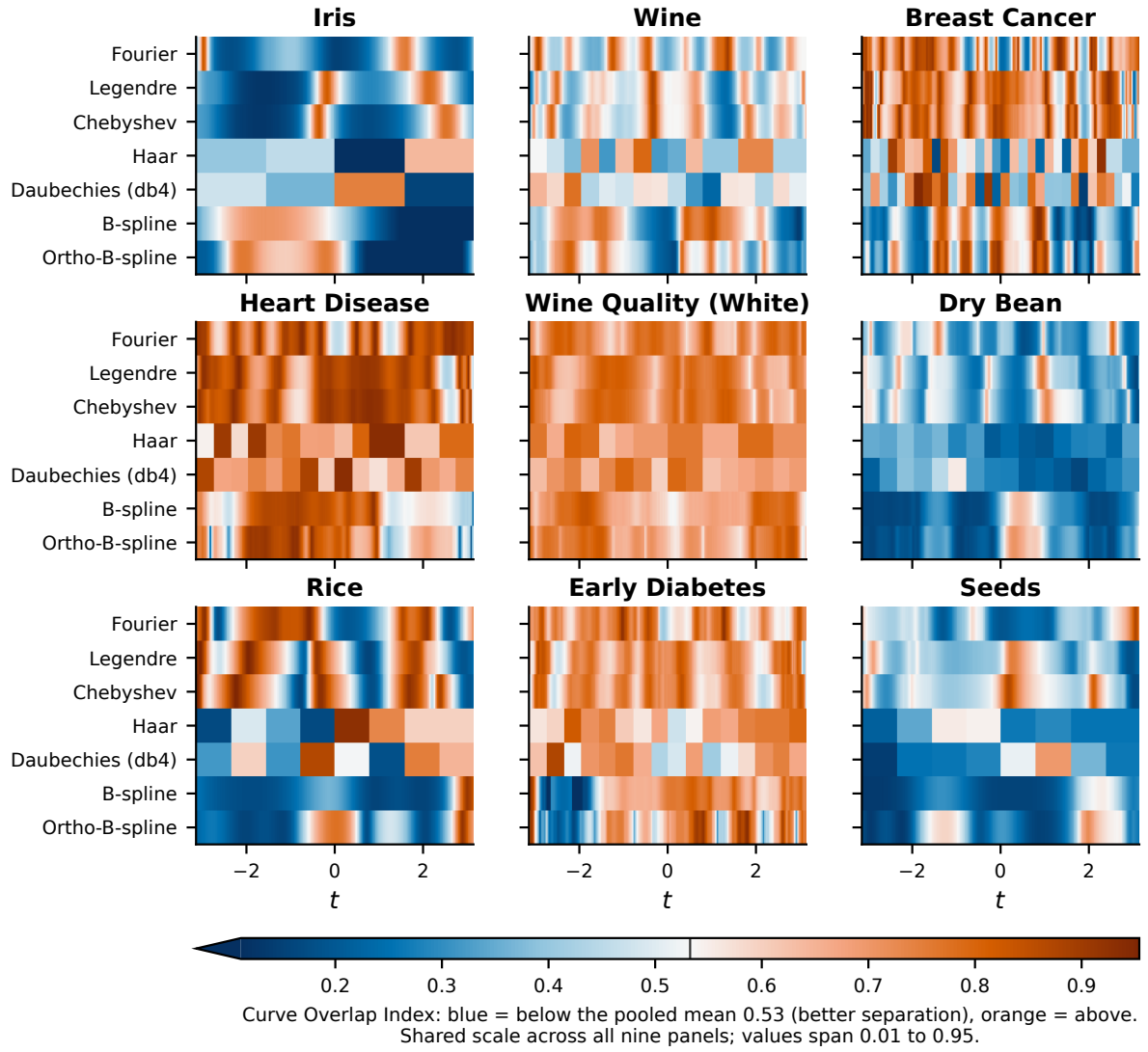
## S6 Gallery of visual profiles and COI heatmaps

The gallery is exploratory. It is rendered on the full data with standardization fitted on all data, whereas every comparative table in the main text comes from the leakage-free 5-fold cross-validation protocol, so figures and tables may differ in detail; the protocol of every experiment is stated individually in the protocol table of the main text, including the audits that deliberately do not use cross-validation. Fig. 12 gives the class summaries and Fig. 13 the COI profiles.

Two readings of the gallery are worth recording, and both are qualitative. The summary curves separate visibly on the datasets where every quantitative metric also separates the classes, and overlap on Heart Disease and Wine Quality, which are the two datasets on which the comparative tables are weakest. The COI heatmaps show the structural difference between global and localized bases directly: a global basis spreads whatever discrimination exists over the whole interval, whereas a localized basis concentrates it into contiguous bands. That is the same localization effect that audit E17 quantifies for ICS, seen here as a picture rather than as a number.



**Fig. 12** Class summaries (mean  $\pm$  std, Ortho-B-spline basis) for all nine datasets. Separated bands indicate class structure directly readable from the curves; overlapping bands mark hard datasets (Heart Disease, Wine Quality)



**Fig. 13** COI profiles (basis by time) for all nine datasets on a shared diverging scale centred on the pooled mean overlap, so the panels stay directly comparable; pooled values span 0.01–0.95 and the colour window is [0.111, 0.954], clipping the lowest 1.4% of cells. Blue marks the time windows where the classes overlap less than the pooled mean, that is, where they separate; orange marks the windows where they overlap more. Localized bases concentrate discrimination into contiguous blue bands; translating these windows back into attribute names is an attribution question and outside the scope of this paper

## References

- [1] Charytanowicz M, Niewczas J, Kulczycki P, et al (2010) Complete gradient clustering algorithm for features analysis of X-ray images. In: Information Technologies in Biomedicine, Advances in Intelligent and Soft Computing, vol 69. Springer, p 15–24, [https://doi.org/10.1007/978-3-642-13105-9\\_2](https://doi.org/10.1007/978-3-642-13105-9_2)
- [2] Cinar I, Koklu M (2019) Classification of rice varieties using artificial intelligence methods. International Journal of Intelligent Systems and Applications in Engineering 7(3):188–194. <https://doi.org/10.18201/ijisae.2019355381>
- [3] Cortez P, Cerdeira A, Almeida F, et al (2009) Modeling wine preferences by data mining

- from physicochemical properties. *Decision Support Systems* 47(4):547–553. <https://doi.org/10.1016/j.dss.2009.05.016>
- [4] Demšar J (2006) Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research* 7:1–30. URL <https://www.jmlr.org/papers/v7/demsar06a.html>
  - [5] Detrano R, Janosi A, Steinbrunn W, et al (1989) International application of a new probability algorithm for the diagnosis of coronary artery disease. *The American Journal of Cardiology* 64(5):304–310. [https://doi.org/10.1016/0002-9149\(89\)90524-9](https://doi.org/10.1016/0002-9149(89)90524-9)
  - [6] Forina M, Armanino C, Castino M, et al (1986) Multivariate data analysis as a discriminating method of the origin of wines. *Vitis* 25(3):189–201. <https://doi.org/10.5073/vitis.1986.25.189-201>
  - [7] Greub W, Rheinboldt W (1959) On a generalization of an inequality of L. V. Kantorovich. *Proceedings of the American Mathematical Society* 10(3):407–415. <https://doi.org/10.1090/S0002-9939-1959-0105028-3>
  - [8] Gustafson KE, Rao DKM (1997) *Numerical Range: The Field of Values of Linear Operators and Matrices*. Universitext, Springer, New York, <https://doi.org/10.1007/978-1-4613-8498-4>
  - [9] Householder AS (1964) *The Theory of Matrices in Numerical Analysis*. Blaisdell, New York
  - [10] Iman RL, Davenport JM (1980) Approximations of the critical region of the Friedman statistic. *Communications in Statistics – Theory and Methods* 9(6):571–595. <https://doi.org/10.1080/03610928008827904>
  - [11] Islam MF, Ferdousi R, Rahman S, et al (2020) Likelihood prediction of diabetes at early stage using data mining techniques. In: *Computer Vision and Machine Intelligence in Medical Image Analysis, Advances in Intelligent Systems and Computing*, vol 992. Springer, p 113–125, [https://doi.org/10.1007/978-981-13-8798-2\\_12](https://doi.org/10.1007/978-981-13-8798-2_12)
  - [12] Kaski S, Nikkilä J, Oja M, et al (2003) Trustworthiness and metrics in visualizing similarity of gene expression. *BMC Bioinformatics* 4(1):48. <https://doi.org/10.1186/1471-2105-4-48>
  - [13] Koklu M, Ozkan IA (2020) Multiclass classification of dry beans using computer vision and machine learning techniques. *Computers and Electronics in Agriculture* 174:105507. <https://doi.org/10.1016/j.compag.2020.105507>
  - [14] Pedregosa F, Varoquaux G, Gramfort A, et al (2011) Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* 12:2825–2830
  - [15] Street WN, Wolberg WH, Mangasarian OL (1993) Nuclear feature extraction for breast tumor diagnosis. In: *Biomedical Image Processing and Biomedical Visualization*, SPIE, pp 861–870, <https://doi.org/10.1117/12.148698>
  - [16] Venna J, Kaski S (2006) Local multidimensional scaling. *Neural Networks* 19(6-7):889–899. <https://doi.org/10.1016/j.neunet.2006.05.014>
  - [17] Wolberg WH, Street WN, Mangasarian OL (1994) Machine learning techniques to diagnose breast cancer from image-processed nuclear features of fine needle aspirates. *Cancer Letters* 77(2–3):163–171. [https://doi.org/10.1016/0304-3835\(94\)90099-X](https://doi.org/10.1016/0304-3835(94)90099-X)