Spectral Kernels for Probabilistic Analysis and Clustering of Shapes

Loic Le Folgoc, Aditya V. Nori, and Antonio Criminisi

Microsoft Research Cambridge, UK.

Abstract. We propose a framework for probabilistic shape clustering based on kernel-space embeddings derived from spectral signatures. Our root motivation is to investigate practical yet principled clustering schemes that rely on geometrical invariants of shapes rather than explicit registration. To that end we revisit the use of the Laplacian spectrum and introduce a parametric family of reproducing kernels for shapes, extending WESD [12] and shape DNA [20] like metrics. Parameters provide control over the relative importance of local and global shape features, can be adjusted to emphasize a scale of interest or set to uninformative values. As a result of kernelization, shapes are embedded in an infinite-dimensional inner product space. We leverage this structure to formulate shape clustering via a Bayesian mixture of kernel-space Principal Component Analysers. We derive simple variational Bayes inference schemes in Hilbert space, addressing technicalities stemming from the infinite dimensionality. The proposed approach is validated on tasks of unsupervised clustering of sub-cortical structures, as well as classification of cardiac left ventricles w.r.t. pathological groups.

1 Introduction

This paper introduces a family of spectral kernels for the purpose of probabilistic analysis and clustering of shapes. Statistical shape analysis spans a range of applications in computer vision, medical imaging and computational anatomy: object recognition, segmentation, detection and modelling of pathologies, etc. Many approaches have been developed, including landmark based representations and active shape models [4,2,15], medial representations [11] and Principal Geodesic Analysis [7], deformable registration and diffeomorphometry [5,26,9]. In many applications, the relevant information is invariant to the pose of the object and is encoded instead within its intrinsic geometry. It may then be advantageous to circumvent the challenges of explicit registration, relying on representations that respect these invariants. Spectral shape descriptors [20,19], built from the spectrum of the Laplace(-Beltrami) operator over the object surface or volume, have achieved popularity for object retrieval [3], analysis and classification of anatomical structures [16,8,24], structural or functional inter-subject mapping [17,14,21]. In [13,12], Konukoglu et al. introduce a $\underline{W}$eighted $\underline{S}$pectral $\underline{D}$istance (WESD) on shapes with two appealing properties: a gentle dependency on the finite spectrum truncation, and a parameter p to emphasize finer or coarser scales. For the purpose of shape analysis and clustering, it would be useful to define not merely a metric, but also an inner product structure. This is a prerequisite for many traditional statistical analysis methods such as Principal Component Analysis (PCA).

Our first contribution is to derive a parametric family of spectral (reproducing) kernels that effectively provide an inner product structure while preserving the multiscale aspect of WESD. As a result, shapes are embedded in an infinite dimensional Hilbert space. Our second contribution is a probabilistic Gaussian mixture model in kernel space, presented within a variational Bayes framework. The kernel space mixture of PCA is of interest in its own right and widely applicable. To the authors' knowledge it has not been proposed previously [6]. The two contributions are coupled to yield a straightforward shape clustering algorithm – a mixture of PCA in spectral kernel space. This approach is validated on tasks of unsupervised and supervised clustering on 69 images from the "Leaf Shapes" database, 240 3D sub-cortical brain structures from the LPBA40 dataset and 45 3D left ventricles from the Sunnybrook cardiac dataset.

2   Background: Laplace operator, heat-trace & WESD

In [13,12], WESD is derived by analyzing the sensitivity of the *heat-trace* to the Laplacian spectrum. Let $\Omega \subset \mathbb{R}^d$ an object (a closed bounded domain) with sufficiently regular boundary $\partial\Omega$, and define its Laplace operator Δ_Ω as

$$[\Delta_\Omega f](\mathbf{x}) \triangleq \sum_{i=1}^d \frac{\partial^2}{\partial x_i^2} f(\mathbf{x}), \quad \forall \mathbf{x} \in \Omega \tag{1}$$

for any sufficiently smooth real-valued function f. The Laplacian spectrum is the infinite set $0 \le \lambda_1 \le \cdots \le \lambda_n \le \cdots, \lambda_n \to +\infty$, of eigenvalues for the Dirichlet problem:

$$\begin{cases} \Delta_\Omega f + \lambda f = 0 & \text{in } \Omega \\ f = 0 & \text{on } \partial\Omega. \end{cases} \tag{2}$$

Denoting by ϕ_n the associated $L^2(\Omega)$-orthonormal eigenfunctions, let $K_\Omega(t, \mathbf{x}, \mathbf{y}) \triangleq \sum_{n \ge 1} \exp\{-\lambda_n t\}\phi_n(\mathbf{x})\phi_n(\mathbf{y})$ the heat kernel. The heat kernel is the fundamental solution to the heat equation $\partial_t K_\Omega(t, \mathbf{x}, \mathbf{y}) = \Delta_\Omega K_\Omega(t, \mathbf{x}, \mathbf{y})$ over Ω, with $K_\Omega(0, \mathbf{x}, \mathbf{y}) \equiv \delta_\mathbf{x}(\mathbf{y})$ and Dirichlet boundary conditions. It is at the basis of a variety of point matching techniques in the computer vision and computer graphics literature, as $K_\Omega(t, \mathbf{x}, \mathbf{x})$ encodes information about the local neighbourhood of $\mathbf{x}$ [10]. Similarly the trace of the heat kernel a.k.a. the *heat-trace*, $Z_\Omega(t) \triangleq \int_\Omega K_\Omega(t, \mathbf{x}, \mathbf{x})d\mathbf{x} = \sum_{n=1}^{+\infty} e^{-\lambda_n t}$, summarizes information about local and global invariants of the shape [12]. It provides a convenient link between the intrinsic geometry of the object and the spectrum. For two shapes Ω_λ and Ω_ξ with respective spectrum $\boldsymbol{\lambda}$ and $\boldsymbol{\xi}$, let $\Delta_{\lambda,\xi}^n$ quantify the influence of the change in the nth eigenmode on the heat-trace:

$$\Delta_{\lambda,\xi}^n \triangleq \int_0^{+\infty} |e^{-\lambda_n t} - e^{-\xi_n t}| \, dt = \frac{|\lambda_n - \xi_n|}{\lambda_n \xi_n}. \tag{3}$$

The pseudo-metric WESD is obtained by summing the contributions of all modes:

$$\rho_p(\Omega_\lambda, \Omega_\xi) \triangleq \left[\sum_{n=1}^{+\infty} (\Delta_{\lambda,\xi}^n)^p\right]^{1/p} = \left[\sum_{n=1}^{+\infty} \left(\frac{|\lambda_n - \xi_n|}{\lambda_n \xi_n}\right)^p\right]^{1/p}. \tag{4}$$

Konukoglu et al. show that WESD is well-defined for $p > d/2$. The key element is due to Weyl [25,10], who proved that the eigenvalues λ behave asymptotically as $\lambda_n \sim \Lambda_n$ when $n \to +\infty$,

$$\Lambda_n \triangleq 4\pi^2 \left(\frac{n}{B_d V_\Omega} \right)^{2/d} , \tag{5}$$

where B_d is the volume of the unit ball in $\mathbb{R}^d$ and V_Ω the volume of Ω. Furthermore from Eq. (2), WESD is made invariant to isotropic rescaling by multiplying the spectrum by $V_\Omega^{2/d}$. Although we refer to objects, shapes and their spectrum interchangeably throughout the article, WESD is a pseudo-metric as objects with different shapes may share the same spectrum. Last but not least, it is multi-scale: its sensitivity to finer scales decreases as p becomes higher. While control over the scale is appealing, the interleaving with the parameter p is somewhat inconvenient. Indeed the metric only defines an inner product structure for $p = 2$. Because this structure is critical for linear statistical analysis, WESD is instead typically used in conjunction with non-linear embedding schemes. We further comment that the choice of measure w.r.t. the time t in the integral of Eq. (3) is arbitrary. This observation turns out to be key for our analysis: the time t is crucial in modulating the sensitivity of the heat-trace $Z_\Omega(t)$ to coarser or finer scales.

3 Shape Spectral Kernels

We now introduce a parametric family of spectral kernels $K_{\alpha,\beta}(\boldsymbol{\lambda}, \cdot)$. These kernels can be interpreted as inner products in some infinite dimensional Hilbert space $\mathcal{H}_{\alpha,\beta}$ and will constitute the basis for the mixture of kernel PCA model of section 4. The parameters α, β control the influence of coarser and finer scales on the metric. Let us introduce the form of the kernels without further delay, postponing details related to its derivation. For $\alpha, \beta \leq 0$, let

$$K_{\alpha,\beta}(\boldsymbol{\lambda}, \boldsymbol{\xi}) \triangleq \sum_{n=1}^{+\infty} \frac{1}{(\beta + \lambda_n)^\alpha} \frac{1}{(\beta + \xi_n)^\alpha} . \tag{6}$$

The kernel definition is subject to matters of convergence of the series, discussed in the electronic appendix. The series is shown to converge for $\alpha > \frac{d}{4}$, while variants of the kernel can be defined for $\alpha > \frac{d}{4} - \frac{1}{2}$. Because it is defined from a convergent series, $K_{\alpha,\beta}(\boldsymbol{\lambda}, \boldsymbol{\xi})$ can be approximated arbitrarily well from a finite term truncation. It shares this property with WESD, unlike shape 'DNA' [20]. Furthermore, invariance to rescaling of the shape can be obtained as for WESD by normalizing the spectrum by $V_\Omega^{2/d}$.

Effect of the parameters α, β. For the sake of relating the kernel to WESD and discussing its behaviour under various pairs of (α, β), let us introduce the corresponding metric via the polarization identity $\rho_{\alpha,\beta}(\Omega_\lambda, \Omega_\xi) \triangleq [K_{\alpha,\beta}(\boldsymbol{\lambda}, \boldsymbol{\lambda}) - 2 \cdot K_{\alpha,\beta}(\boldsymbol{\lambda}, \boldsymbol{\xi}) + K_{\alpha,\beta}(\boldsymbol{\xi}, \boldsymbol{\xi})]^{1/2}$. This leads to:

$$\rho_{\alpha,\beta}(\Omega_\lambda, \Omega_\xi) = \left[\sum_{n=1}^{+\infty} \left(\frac{(\beta + \lambda_n)^\alpha - (\beta + \xi_n)^\alpha}{(\beta + \lambda_n)^\alpha \cdot (\beta + \xi_n)^\alpha} \right)^2 \right]^{1/2} , \tag{7}$$
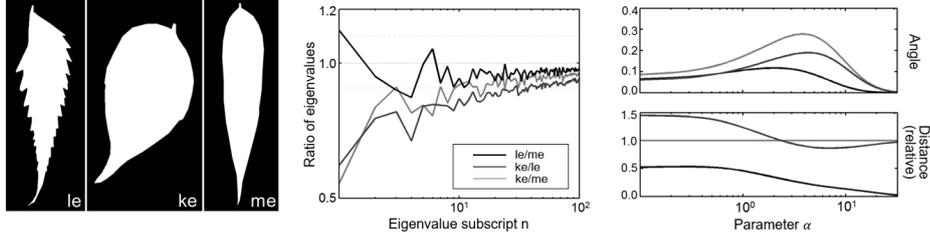
Fig. 1. A multiscale family of spectrum based metrics. (Left) Example shapes "ke", "le", "me" from the "Leaf Shapes" dataset[2]. "ke" is rounder, "me" is more elongated, "le" is in-between but its boundary is jagged. (Right) Effect of α on the distance and angle between each pair. Distances are normalized w.r.t. the ke/me pair. Larger (resp. smaller) values of α emphasize coarser (finer) scale features. For instance, the dissimilarity of the pair le/ke relatively to that of the pair le/me shrinks (from $6\times$ to under $3\times$) at finer scales as the focus shifts from, say, purely the aspect ratio to also account for *e.g.* the smooth/jagged boundary. (Middle) Ratio of Laplacian eigenvalues for the three pairs. Small eigenvalues relate to global features (*e.g.* rounder, elongated), thus small eigenvalues of "ke" are even further away from "le", "me".

and may be compared to $\rho_p(\Omega_\lambda, \Omega_\xi)$ of Eq. (4). Firstly, for $\alpha = 1$, $\beta = 0$ and $p = 2$, WESD and the spectral kernel metric coincide: $\rho_2 = \rho_{1,0}$. In that sense $\rho_{\alpha,0}$ extends the "Euclidean" WESD while providing additional parameters to control the weight of coarser and finer scales. Recall that the larger eigenvalues λ_n relate to finer local details, while the first eigenvalues $\lambda_1, \lambda_2 \ldots$ relate to global invariants (*e.g.* volume, boundary surface). When α increases to large values, global shape features gain overwhelming importance and, in the limit of $\alpha \to +\infty$, only λ_1 matters. Inversely, in the limit of $\alpha \to 0$, the nth term of the series (before squaring) behaves as $\rho_{0,0}^n(\Omega_\lambda, \Omega_\xi) \sim |\log \frac{\lambda_n}{\xi_n}|$, which is sensitive only to the *relative* ratio of eigenvalues, as opposed to the actual values λ_n, ξ_n. In that sense it gives equal weight to all eigenvalues. Most eigenvalues relate to finer scale features, hence smaller α values emphasize these details.

For $\lambda_n, \xi_n \ll \beta$, the nth series term behaves as $|\lambda_n - \xi_n|^2$, which does not penalize the eigenvalue magnitude whatsoever. This is reminiscent of the shape DNA [20,24] metric $\rho_{\text{DNA}}(\Omega_\lambda, \Omega_\xi) \triangleq \sum_{n=1}^{N_{\max}} |\lambda_n - \xi_n|$. β acts similarly to $N_{\max}$ in selecting a range of relevant eigenvalues for the metric: as it grows, larger eigenvalues are given more importance. Finally, for $\alpha, \beta \to +\infty$ such that $\frac{\alpha}{\beta} = \text{const.} = t_{\alpha,\beta}$, the nth series term (unsquared) behaves as $\rho_{\alpha,\beta}^n(\Omega_\lambda, \Omega_\xi) \sim \exp\{-\lambda_n t_{\alpha,\beta}\} - \exp\{-\xi_n t_{\alpha,\beta}\}$, that is the nth term of $Z_\lambda(t_{\alpha,\beta}) - Z_\xi(t_{\alpha,\beta})$. Hence for large "informative" values of α and β, the ratio α/β selects a heat diffusion time-scale of interest and β the spread around that value. Alternatively, somewhat neutral choices of parameters with a balance between coarser and finer scales are obtained for β small ($\beta = 0$ or $\beta \simeq \Lambda_1$) and α small (*e.g.* $\alpha \leq d/2$). The discussion will be clearer from a closer look at the link between the choice of α, β and the corresponding choice of time integration.

Computations in kernel space. Before outlining derivations from the heat trace, let us note an attractive computational property of the family of spectral kernels. For the sake of "linear" statistical analysis (*e.g.*, k-means, PCA), weighted sums of kernel representers $K_{\alpha,\beta}(\boldsymbol{\lambda}, \cdot)$ of shapes $\boldsymbol{\lambda}$ are typically involved. Usually such sums cannot be explicitly computed, but their projections on a given element $K_{\alpha,\beta}(\boldsymbol{\xi}, \cdot)$ can, since $(K_{\alpha,\beta}(\boldsymbol{\lambda}, \cdot)|K_{\alpha,\beta}(\boldsymbol{\xi}, \cdot))_{\mathcal{H}_{\alpha,\beta}} = K_{\alpha,\beta}(\boldsymbol{\lambda}, \boldsymbol{\xi})$. This is known as the kernel "trick". As a drawback many kernel extensions of widespread linear statistical schemes have a time complexity tied to the square or cube of the number of observations in the dataset, instead of the *dimensionality* of the data.

In the present case it turns out that computations can be done explicitly. Let $\Phi_{\alpha,\beta}: \boldsymbol{\lambda} \mapsto (\cdots (\lambda_n + \beta)^{-\alpha} \cdots)$ map shapes to the space I^2 of l_2 sequences. By construction $K_{\alpha,\beta}(\boldsymbol{\lambda}, \boldsymbol{\xi}) = \langle \Phi_{\alpha,\beta}(\boldsymbol{\lambda})|\Phi_{\alpha,\beta}(\boldsymbol{\xi})\rangle_{I^2}$. Addition, scalar multiplication and inner product on elements $K_{\alpha,\beta}(\boldsymbol{\lambda}, \cdot)$ of $\mathcal{H}_{\alpha,\beta}$ are equivalent to explicit addition, scalar multiplication and inner product on elements $\Phi_{\alpha,\beta}(\boldsymbol{\lambda})$ of I^2. For instance, given two shapes $\boldsymbol{\lambda}$ and $\boldsymbol{\xi}$, their mean is given by $\Phi_{\alpha,\beta}(\boldsymbol{\chi})$, defined by $(\chi_n + \beta)^{-\alpha} \triangleq \frac{1}{2}(\lambda_n + \beta)^{-\alpha} + \frac{1}{2}(\xi_n + \beta)^{-\alpha}$. Any finite N-term truncation of the infinite kernel series is equivalent to an N-term truncation of the l^2 sequence, so that the mean of $|\mathcal{D}|$ points $\{K_{\alpha,\beta}(\boldsymbol{\lambda}_i, \cdot)\}_{1 \le i \le |\mathcal{D}|}$ can for instance be stored as an N-tuple. Moreover, the eigenmodes of the Gram matrix $[K_{\alpha,\beta}(\boldsymbol{\lambda}_i, \boldsymbol{\lambda}_j)]_{1 \le i,j \le |\mathcal{D}|}$ can be obtained from the eigenmodes of the $N \times N$ covariance matrix $\sum_i \Phi_{\alpha,\beta}(\boldsymbol{\lambda}_i)\Phi_{\alpha,\beta}(\boldsymbol{\lambda}_i)^{\intercal}$ of the truncated $\Phi_{\alpha,\beta}(\boldsymbol{\lambda}_i)$ tuples. Hence the computational complexity depends on the truncated dimensionality rather than the dataset size, whenever advantageous. Truncation error bounds are given in the electronic appendix.

Derivation of the spectral kernels from the heat trace. Similarly to WESD, the proposed kernels are derived by quantifying the influence of the change in the nth eigenmode on the heat-trace. However we consider a variety of measures for the integration w.r.t. time. Specifically let $p_{\alpha,\beta}(t) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} \exp\{-\beta t\} t^{\alpha-1}$ the probability density function of the gamma distribution with positive shape parameter α and rate parameter β (formally including improper cases, $\alpha = 0$ or $\beta = 0$). We extend Eq. (3) by integrating w.r.t. $p_{\alpha,\beta}$:

$$\Delta_{\alpha,\beta}^n(\boldsymbol{\lambda}, \boldsymbol{\xi}) \triangleq \int_0^{+\infty} |e^{-\lambda_n t} - e^{-\xi_n t}| \cdot p_{\alpha,\beta}(t) dt, \tag{8}$$

$$= \beta^{\alpha} \cdot \left| \frac{(\beta + \lambda_n)^{\alpha} - (\beta + \xi_n)^{\alpha}}{(\beta + \lambda_n)^{\alpha} \cdot (\beta + \xi_n)^{\alpha}} \right|, \tag{9}$$

$$= \beta^{\alpha} \cdot \left| \frac{1}{(\beta + \lambda_n)^{\alpha}} - \frac{1}{(\beta + \xi_n)^{\alpha}} \right|. \tag{10}$$

We obtain a (pseudo-)metric on shapes and retrieve Eq. (7) by aggregating the contributions of all modes: $\rho_{\alpha,\beta}(\Omega_{\lambda}, \Omega_{\xi})^2 \triangleq \sum_{n=1}^{+\infty} (\Delta_{\alpha,\beta}^n(\boldsymbol{\lambda}, \boldsymbol{\xi}))^2$. Moreover $K_{\alpha,\beta}(\boldsymbol{\lambda}, \boldsymbol{\xi})$ defined as in Eq. (6) is positive definite as a kernel over spectra, and defines up to renormalization an inner product consistent with the metric $\rho_{\alpha,\beta}$, *i.e.* $[\rho_{\alpha,\beta}(\Omega_{\lambda}, \Omega_{\xi})/\beta^{\alpha}]^2 = K_{\alpha,\beta}(\boldsymbol{\lambda}, \boldsymbol{\lambda}) - 2 \cdot K_{\alpha,\beta}(\boldsymbol{\lambda}, \boldsymbol{\xi}) + K_{\alpha,\beta}(\boldsymbol{\xi}, \boldsymbol{\xi})$.

4   Probabilistic clustering via mixtures of kernel PCA (mkPCA)

We now introduce a probabilistic mixture model of Principal Component Analysers in kernel space, tackling the inference in a variational Bayesian framework. Let $K : (x, y) \in \mathcal{X} \times \mathcal{X} \mapsto K(x, y) = K_x(y) \in \mathbb{R}$ a reproducing kernel with $(L_K f)(x) \triangleq \int_{\mathcal{X}} K(x, y) f(y) dy$ a compact operator over $L_2(\mathcal{X})$. Denote by $\mathcal{H}$ the associated reproducing kernel Hilbert space. For $f, g \in \mathcal{H}$, $\langle f | g \rangle_{L_2} \triangleq \int_{\mathcal{X}} fg$ or simply $\langle f | g \rangle$ is the L_2 inner-product, whereas $\mathcal{H}$ is endowed with the inner product $(f | g)_{\mathcal{H}} = \langle f | L_K^{-1} g \rangle_{L_2}$ or simply $(f | g)$. $\| f \|_{\mathcal{H}} \triangleq (f | f)^{1/2}$ or $\| f \|$ stands for the norm of $\mathcal{H}$. Finally we use the bra-ket notation, $|f) \triangleq L_K^{-1/2} f$ and $(f| \triangleq |f)^{\mathsf{T}}$. The main technical hurdles stem from the infinite dimensionality of the data to model: probability density functions are not properly defined, and the full normalization constants cannot be computed. In solving this obstacle, the decomposition of the covariance into isotropic noise and low rank structure is key.

PCA as a latent linear variable model. Let $\phi_k \in \mathcal{H}$, $k = 1 \cdots r$, a finite set of basis functions. Assume $f = \sum_{1 \le k \le r} \phi_k w_k + \sigma \epsilon$ is generated by a linear combination of the ϕ_k's plus some scaled white noise $\epsilon \sim \mathcal{N}(0, L_K)$. Further assume that $\boldsymbol{w} \sim \mathcal{N}(0, \mathrm{I})$ where $\boldsymbol{w} \triangleq (w_1 \cdots w_r)$. Then $f \sim \mathcal{N}(0, \sigma^2 L_K + \Phi \Phi^{\mathsf{T}})$ follows a Gaussian process, where $\Phi \triangleq (\cdots \phi_k \cdots)$. It is a natural extension of the probabilistic PCA model of [23], conveniently decomposing the variance as the sum of an isotropic part $\sigma^2 L_K$ (w.r.t. $\| \cdot \|$) and of a low-rank part $\Phi \Phi^{\mathsf{T}}$. The first term accounts for noise in the data, while the latter one captures the latent structure. Furthermore $|f) = \sum_{1 \le k \le r} |\phi_k) w_k + \sigma |\epsilon)$ also follows a Gaussian distribution $\mathcal{N}(0, \sigma^2 \mathrm{I} + |\Phi)(\Phi|)$. We use this latter form from now on as it most closely resembles finite dimensional linear algebra. By analogy in the finite case, the following linear algebra updates hold: $(\sigma^2 \mathrm{I} + \Phi \Phi^{\mathsf{T}})^{-1} = \sigma^{-2}[\mathrm{I} - \Phi(\sigma^2 \mathrm{I}_r + \Phi^{\mathsf{T}} \Phi)^{-1} \Phi^{\mathsf{T}}]$ and $|\sigma^2 \mathrm{I} + \Phi \Phi^{\mathsf{T}}| = |\sigma^2 \mathrm{I}| \cdot |\mathrm{I}_r + \sigma^{-2} \Phi^{\mathsf{T}} \Phi|$. The former Woodbury matrix identity and latter matrix determinant lemma express, using only readily computable terms, the resulting change for (resp.) the inverse and determinant under a low rank matrix perturbation. In particular the determinant conveniently factorizes into a constant term $|\sigma^2 \mathrm{I}|$, which is improper in the infinite dimensional case, and a well-defined term that depends on Φ. This let us compute normalization constants and other key quantities required for inference and model comparison, up to a constant.

Probabilistic mixture of kernel PCA. Let $\mathcal{D} = \{x_i\}_{i=1}^{|\mathcal{D}|}$ a set of observations and K_{x_i} the kernel embedding of the ith point. Assume K_{x_i} is drawn from a mixture of M Gaussian components, depending on the state of a categorical variable $c_i \in \{1 \cdots M\}$. Let z_{im} the binary gate variable that is set to 1 if $c_i = m$ and 0 otherwise. Let $|K_{x_i})_{\mathcal{H}} | z_{im} = 1 \sim \mathcal{N}(\mu_m, C_m)$ i.i.d. according to a Gaussian process, where $C_m \triangleq \sigma^2 \mathrm{I} + |\Phi_m)(\Phi_m|$ and σ^2 is a fixed parameter. $\Phi_m \triangleq (\cdots \phi_{km} \cdots)$ concatenates an arbitrary number of (random) basis functions $\phi_{km} \in \mathcal{H}$. Denoting $\boldsymbol{z}_i \triangleq \{z_{im}\}_{1 \le m \le M}$, the $\boldsymbol{z}_i$'s are i.i.d. following a categorical distribution, *i.e.* $p(z_i | \boldsymbol{\pi}) = \prod_{1 \le m \le M} \pi_m^{z_{im}}$, where $\boldsymbol{\pi} \triangleq \{\pi_m\}_{1 \le m \le M}$. A conjugate Dirichlet prior is taken over the mixture proportions, $p(\boldsymbol{\pi} | \kappa_0) \propto \prod_{1 \le m \le M} \pi_m^{\kappa_0 - 1}$. (μ_m, Φ_m) is endowed (formally) with the improper prior $\mathcal{N}(\mu_m | \mathrm{m}_0, \eta_0^{-1} C_m) \cdot |C_m|^{-\gamma_0/2} \exp -\frac{1}{2} \mathrm{tr}(s_0 C_m^{-1})$. As a constraint for the op-

timal Φ_m to have finite rank, $\gamma_0^{-1}s_0 \leq \sigma^2$. The model and its fixed hyperparameters $\{\eta_0, \gamma_0, s_0, \kappa_0, \sigma^2\}$ will be shortened as $\mathcal{M}$.

In what follows, bolded quantities stand for the concatenation of their normal font counterpart across all values of the missing indices. For instance $\boldsymbol{z}_m \triangleq \{z_{im}\}_{1 \leq i \leq |\mathcal{D}|}$ and $\boldsymbol{z}_i \triangleq \{z_{im}\}_{1 \leq m \leq M}$. Variational Bayesian updates can be derived in closed form for the family $q(\boldsymbol{z}) \triangleq q_{\boldsymbol{z}}(\boldsymbol{z})q_{\boldsymbol{\theta}}(\boldsymbol{\theta})$ of variational posteriors where the mixture assignment variables $\boldsymbol{z}_i$'s and the model hyperparameters $\boldsymbol{\theta} = \{\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Phi}\}$ for all mixtures are factorized, provided that $q_{\boldsymbol{\theta}}(\boldsymbol{\theta}) \triangleq q_{\boldsymbol{\pi}}(\boldsymbol{\pi}|\boldsymbol{\mu}, \boldsymbol{\Phi})\delta_{\hat{\boldsymbol{\mu}}}(\boldsymbol{\mu}|\boldsymbol{\Phi})\delta_{\hat{\boldsymbol{\Phi}}}(\boldsymbol{\Phi})$ is further restrained to a Dirac (point mass) distribution over the Φ_m's and μ_m's. They closely follow their finite dimensional counterpart as in *e.g.* [23,1]. For all model variables but $\boldsymbol{\Phi}, \boldsymbol{\mu}$, model conjugacies can be exploited and variational posteriors have the same functional form as the corresponding priors. Specifically, $q_{\boldsymbol{z}}(\boldsymbol{z}) = \prod_{i=1}^{|\mathcal{D}|} q_{\boldsymbol{z}_i}(\boldsymbol{z}_i)$ and $q_{\boldsymbol{\pi}}(\boldsymbol{\pi}|\boldsymbol{\mu}, \boldsymbol{\Phi}) = q_{\boldsymbol{\pi}}(\boldsymbol{\pi})$, with $q_{\boldsymbol{z}_i}(\boldsymbol{z}_i) = \prod_{m=1}^{M} \rho_{im}^{z_{im}}$ categorical and $q_{\boldsymbol{\pi}}(\boldsymbol{\pi}) = \mathcal{D}(\boldsymbol{\pi}|\boldsymbol{\kappa})$ Dirichlet, denoting $\boldsymbol{\kappa} \triangleq \{\kappa_m\}_{m=1}^{M}$. In addition $\hat{\mu}_m$ is the mode of $p(\mu_m|\Phi_m, \mathcal{D}, \mathcal{M}) = \mathcal{N}(\mu_m \,|\, |m_m), \eta_m^{-1}C_m)$, $\hat{\Phi}_m$ maximizes the posterior $p(\Phi_m \,|\langle \boldsymbol{z}_m \rangle_{q_{\boldsymbol{z}}}, \mathcal{D}, \mathcal{M}) \propto |C_m|^{-\gamma_m/2} \exp -\frac{1}{2}\text{tr}\{S_m C_m^{-1}\}$. The updates for all hyperparameters involved are as follows:

$$\kappa_m = |\mathcal{D}|\bar{\pi}_m + \kappa_0 \tag{11}$$

$$\eta_m = |\mathcal{D}|\bar{\pi}_m + \eta_0 \tag{12}$$

$$\gamma_m = |\mathcal{D}|\bar{\pi}_m + \gamma_0 \tag{13}$$

$$m_m = \frac{|\mathcal{D}|\bar{\pi}_m}{\eta_m} \cdot \bar{\mu}_m + \frac{\eta_0}{\eta_m} \cdot m_0 \tag{14}$$

$$S_m = \eta_m \bar{\Sigma}_m^{+0} + s_0 I \tag{15}$$

$$|\Phi_m) = \sqrt{\left(\gamma_m^{-1}S_m - \sigma^2 I\right)_+} \tag{16}$$

$$\rho_{im} \propto \tilde{\pi}_m \cdot \frac{1}{|I + \sigma^{-2}(\Phi_m|\Phi_m)_{\mathcal{H}}|^{1/2}} \exp\left\{-\frac{1}{2}\chi_{im}^2\right\} \tag{17}$$

where $\sum_{m=1}^{M} \rho_{im} = 1$ for all i, $(\cdot)_+$ stands for the projection on the set of positive semidefinite matrices and:

$$\bar{\pi}_m \triangleq \frac{1}{|\mathcal{D}|} \sum_{i=1}^{|\mathcal{D}|} \rho_{im} \tag{18}$$

$$\bar{\mu}_m \triangleq \frac{1}{|\mathcal{D}|\bar{\pi}_m} \sum_{i=1}^{|\mathcal{D}|} \rho_{im} K_{x_i} \tag{19}$$

$$\delta_{im} \triangleq K_{x_i} - m_m, \quad \delta_{0m} \triangleq m_0 - m_m \tag{20}$$

$$\bar{\Sigma}_m^{+0} \triangleq \frac{1}{\eta_0 + |\mathcal{D}|\bar{\pi}_m} \left(\eta_0 \,|\delta_{0m})(\delta_{0m}| + \sum_{i=1}^{|\mathcal{D}|} \rho_{im} \,|\delta_{im})\,(\delta_{im}| \right) \tag{21}$$

$$\chi_{im}^2 \triangleq \frac{1}{\sigma^2} \left(\|\delta_{im}\|^2 - (\delta_{im}|\Phi_m)\left(\sigma^2 I + (\Phi_m|\Phi_m)\right)^{-1}(\Phi_m|\delta_{im}) \right) \tag{22}$$

$$\tilde{\pi}_m \triangleq \exp \langle \log \pi_m \rangle_{q_{\pi_m}} = \exp\left\{ \psi(\kappa_m) - \psi(\textstyle\sum_{m'=1}^{M} \kappa_{m'}) \right\} \tag{23}$$

The derivation of Eq. (16) closely mirrors the one found in [23]. Fortunately, computations of infinite dimensional quantities $\bar{\Sigma}_m^{+0}$ and S_m need not be conducted explicitly, exploiting generic relationships between the eigenmodes of Gram and scatter matrices. Indeed more explicitly we have $|\Phi_m) = |U_m)\Delta_m$, up to right multiplication by an arbitrary rotation, with the following notations. $\Delta_m \triangleq \text{diag}[(\gamma_m^{-1}d_{km} - \sigma^2)^{1/2}]$ is the $r_m \times r_m$ diagonal matrix such that d_{km} is the kth biggest eigenvalue of S_m among those r_m eigenvalues strictly bigger than $\gamma_m\sigma^2$. The kth column of $|U_m)$ is the corresponding eigenvector $|U_m)_k$ of S_m. Remark that S_m can be rewritten as the sum of $s_0 I$ and $|\mathcal{D}|+1$ rank one terms of the form $|g_{im})(g_{im}|$. The non-zero eigenvalues of S_m are the same as those of the $(|\mathcal{D}|+1) \times (|\mathcal{D}|+1)$ matrix $(G|G) = [\cdots (g_{im}|g_{jm}) \cdots]_{ij}$. The $|U_m)_k = |G) e_k$ are obtained from the eigenvectors e_k of the latter matrix. Moreover, computations in Eqs. (17) and (22) simplify by noting that $(\Phi_m|\Phi_m) = \Delta_m^2$ is diagonal.

High-level overview. The scheme proceeds analogously to the finite dimensional case. Update Eqs. (11)-(17) can be understood as follows, disregarding terms stemming from priors for simplicity. Each data point i is softly assigned to the mth mixture component with a probability ρ_{im} that depends on the mixing proportion π_m and the point likelihood under the mth component model. Then for each component, the mixing proportion is set to the average of the responsibilities ρ_{im}. The mean m_m is updated to the empirical mean, with data points weighted by ρ_{im}. Similarly the covariance $C_m = \sigma^2 I + |\Phi_m)(\Phi_m|$ is (implicitly) set to the empirical covariance, shrinking all eigenvalues by σ^2 (down to a minimum of 0), after which $|\mathcal{D}|$ non-zero directions remain at most. The algorithm iterates until convergence.

Lower bound. The lower bound on the evidence can be computed (appendix) up to a constant and is given formally by $\sum_i \log\left(\sum_m \rho_{im}^u\right) - \text{KL}[q_{\boldsymbol{\theta}}\|p(\boldsymbol{\theta})]$, where ρ_{im}^u is the unnormalized soft-assignment of point i to cluster m (right hand side of Eq. (17)).

Initialization, choice of σ^2. The mkPCA algorithm is initialized via k-means clustering. In the proposed model σ^2 controls the level of noise in the data. The more noise, the more data variability it accounts for, and the lesser variability attributed to latent structure. Bayesian inference of σ^2 is difficult in infinite dimension, hence not pursued here. If applicable, cross-validation is sound. Alternatively the following k-means based heuristic can be used. Let d_i the distance of the ith data point to the centroid of its assigned k-means cluster. $\sum_{1 \le i}^{|D|} d_i^2 / |D|$ gives an estimate of the average variability in the dataset due to both noise and structure. We set σ^2 to a small fraction of that value. The a posteriori analysis of the learned clusters (*e.g.* number of eigenmodes, magnitude of eigenvalues) also helps verify the soundness of the setting.

5   Experiments & Results

The proposed framework is implemented in MATLAB, based on the publicly available implementation of WESD[3]. Unless otherwise specified, the scale invariant kernel is

---

[3] http://www.nmr.mgh.harvard.edu/ enderk/software.html
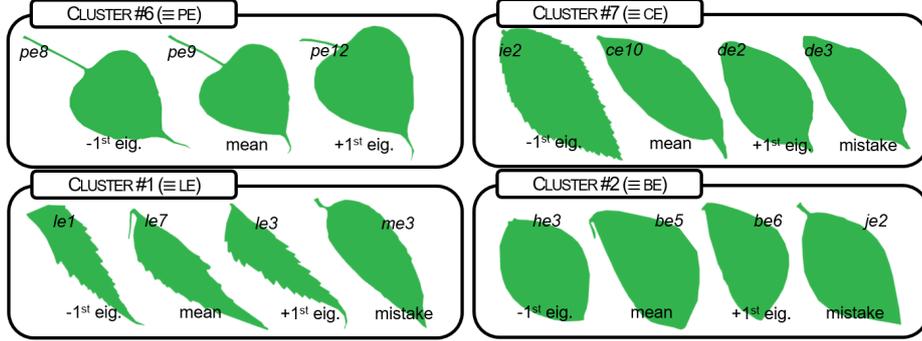
Fig. 2. Inferred leaf clusters. 4 clusters out of 7 are shown, with representative examples for the mean and ± 2st.d. along the first eigenmode. An example mistake is shown if available. For each leaf, the three character code is its database index. The cluster assignment of the leaf is deemed wrong whenever the letter prefix differs from that of the cluster (chosen by majority voting).

used and truncated at $N = 200$ eigenvalues; and σ^2 is set to $0.05 \times$ the k-means average square error. We experiment with values of $\alpha \geq d/2$, and $\beta = 0$ or $\equiv \lambda_1$. Other hyper-parameters are set to uninformative values: $\kappa_0 = \gamma_0 = 10^{-6}$, $\eta_0 = 10^{-15}$, $s_0 = \gamma_0 \sigma^2 / 10$ and m_0 to the data mean. The best (based on the lower bound) of 10 mkPCA runs (100 iterations) with k-mean initialization is typically selected.

"Leaf Shapes" dataset: unsupervised mining of leaf species. The dataset consists of 69 binary images of leaves of 10 different subtypes (*e.g.* "ce", "je"). Each type contains 3 to 12 examples. We aim at (partially) retrieving these subtypes using *unsupervised* clustering. We use the lower bound (and discard small clusters, $\pi_m < 5\%$). 7 clusters are identified. Fig. 2 displays 4 of the 7 clusters. For quantitative analysis a label is assigned to each cluster by majority voting, and points are given the label of the maximum a posteriori cluster. The retrieval accuracy for $\alpha = 0.5$, $\beta = 100$ is $71\% \pm 1\%$. It is robust w.r.t. α, β ($\geq 68\%$ over 20 runs with chosen values $\alpha \in [0.5, 5]$, $\beta \in [0, 500]$, $67.5\% \pm 3\%$ over 100 runs with random values in that range). Mislabelling is mostly due to the fact that smaller subtypes with 3 to 4 images do not have their own cluster.

LPBA40 dataset: clustering sub-cortical structures. We proceed with unsupervised clustering of 40×6 volumes of left and right caudates, hippocampuses and putamens from the public LPBA40 dataset [22], using 10 clusters maximum. For qualitative assessment, the data is projected on the 2 largest eigenmodes of each cluster (Fig. 3 for the 6 main clusters). As an indicator of class separation in the learned model, we quantify here again the agreement between the true labels and those assigned by MAP, with 16% average misclassification rate for $\alpha = 1$, $\beta = 0$ ($< 23\%$ across a wide parameter range).

Learning left ventricle anomalies. The Sunnybrook cardiac dataset [18] (SCD) consists of 45 left ventricle volumes: a heart failure (HF) group (24 cases), an hypertrophy (HYP) group (12) and a healthy (N) group (9). We evaluate the proposed framework for (maximum a posteriori) classification, modelling each group as a 1 (HYP, N) or 2 (HF)
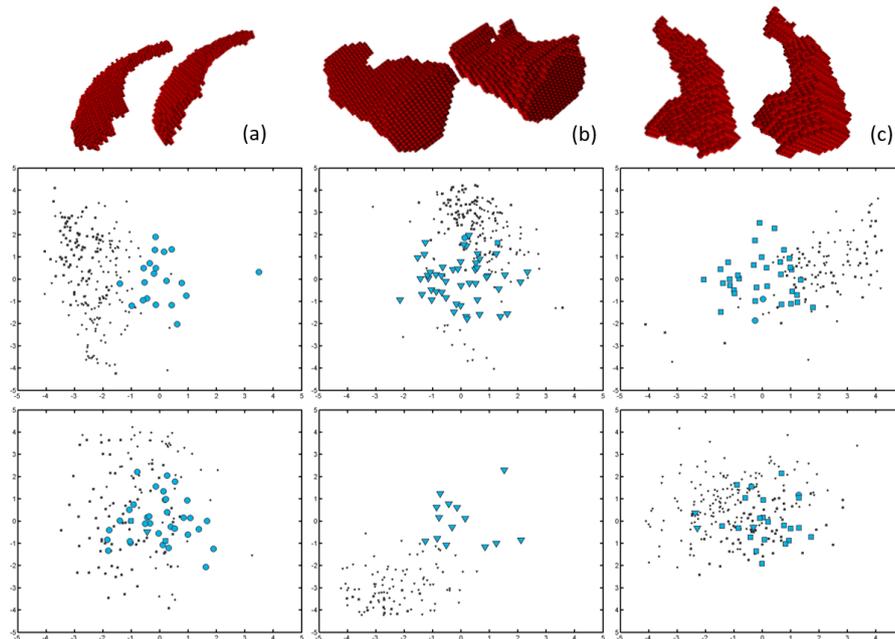
Fig. 3. Unsupervised clustering of subcortical structures. Example (a) caudate nuclei, (b) hippocampi (c) putamina. Below each sub-structure, 2 clusters widely associated with it are shown (projection onto the two first eigenmodes of each cluster, range ± 5 st.d.). The following symbols are used: circle $\equiv$ caudate, triangle $\equiv$ hippocampus, square $\equiv$ putamen. Large blue data points belong to the displayed cluster, small black dots do not (overlap largely due to projection).

component mkPCA. Fig. 4 (Left) reports the test accuracy for various hyperparameters. For each parameter setting, mean and standard deviation are computed over 50 runs, training on a random fold of $2/3$ of the dataset and testing on the complement. The baseline of 53.3% accuracy corresponds to systematically predicting HF. Blood/muscle LV volumes are highly correlated to HF/HYP pathologies and constitute natural predictors that are expected to perform well. The spectrum based mkPCA model classifies better on average ($\sim 75\% \pm 8\%$ across many settings) than a volume based mPCA ($65\% \pm 9\%$) despite the small training set and low-resolution of the raw data. Fig. 5 shows typical (often sensible) classification mistakes. Finally, Fig. 4 evidences strong correlations between eigenmodes of HF/HYP clusters and the cavity/muscle volumes.

6   Conclusion

We proposed a framework for probabilistic clustering of shapes. It couples ideas from the fields of spectral shape analysis and Bayesian modelling to circumvent both the challenges of explicit registration and the recourse to non-linear embeddings. Firstly, a multiscale family of spectral kernels for shapes is introduced. It builds on existing work on the Laplace spectrum (shape DNA, WESD), going beyond by endowing the space of objects with an inner product structure. This is required for many of the most
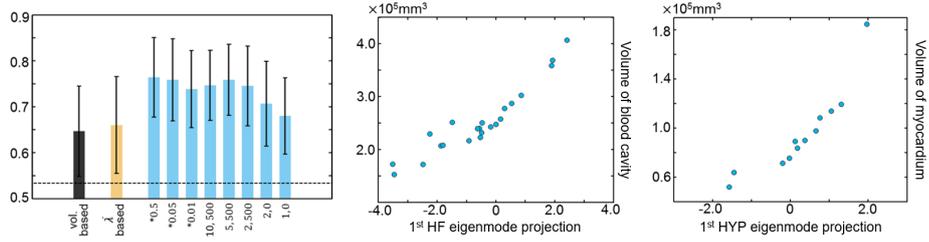
Fig. 4. Sunnybrook Cardiac dataset supervised clustering analysis. (Left) Classification accuracy under various settings (mean and $\pm$ st.d.). Black: volume-based classification baseline. Yellow: normalized spectrum-based classification. Blue: non-normalized spectrum-based, for various α, β, σ^2 triplets. Default when not specified is $\alpha = 10$, $\beta = 10^3$ and σ^2 set to $0.05\times$ the k-means average error. The last five bars correspond to various α, β pairs, the previous 3 to other σ^2 settings. (Middle) Blood volume as a function of the 1st eigenmode projection for HF cases. (Right) Left ventricle myocardium volume as a function of 1st eigen. projection for HYP cases.

widespread statistical analysis algorithms. Secondly a probabilistic mixture of kernel space PCA is designed, working out technicalities stemming from the infinite dimensionality. We experimented with tasks of supervised and unsupervised clustering on 69 images from the "Leaf Shapes" database, 240 3D sub-cortical brain structures from the LPBA40 dataset and 45 3D left ventricles from the Sunnybrook cardiac dataset.

References

1. Archambeau, C., Verleysen, M.: Robust Bayesian clustering. Neural Networks 20(1), 129–138 (2007) 7
2. Belongie, S., Malik, J., Puzicha, J.: Shape matching and object recognition using shape contexts. IEEE Trans. Pattern. Anal. Mach. Intell. 24(4), 509–522 (2002) 1
3. Bronstein, A.M., Bronstein, M.M., Guibas, L.J., Ovsjanikov, M.: Shape Google: Geometric words and expressions for invariant shape retrieval. ACM Trans Graph 30(1), 1 (2011) 1
4. Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J.: Active Shape Models – Their training and application. Comput Vis Image Underst 61(1), 38–59 (1995) 1
5. Durrleman, S., Prastawa, M., Charon, N., Korenberg, J.R., Joshi, S., Gerig, G., Trouvé, A.: Morphometry of anatomical shape complexes with dense deformations and sparse parameters. NeuroImage 101, 35–49 (2014) 1
6. Filippone, M., Camastra, F., Masulli, F., Rovetta, S.: A survey of kernel and spectral methods for clustering. Pattern Recognition 41(1), 176–190 (2008) 2
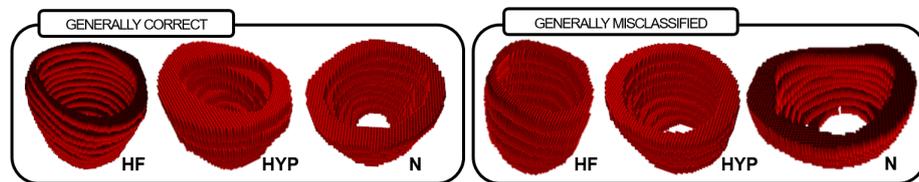
Fig. 5. Example left ventricles with test misclassification rate $> 25\%$ (right) or $< 2\%$ (left).

7. Fletcher, P.T., Lu, C., Pizer, S.M., Joshi, S.: Principal Geodesic Analysis for the study of nonlinear statistics of shape. IEEE T Med Imaging 23(8), 995–1005 (2004) [1]

8. Germanaud, D., Lefèvre, J., Toro, R., Fischer, C., Dubois, J., Hertz-Pannier, L., Mangin, J.F.: Larger is twistier: Spectral analysis of gyrification (spangy) applied to adult brain size polymorphism. NeuroImage 63(3), 1257–1272 (2012) [1]

9. Gori, P., Colliot, O., Marrakchi-Kacem, L., Worbe, Y., Poupon, C., Hartmann, A., Ayache, N., Durrleman, S.: A Bayesian framework for joint morphometry of surface and curve meshes in multi-object complexes. Medical Image Analysis 35, 458–474 (2017) [1]

10. Grebenkov, D.S., Nguyen, B.T.: Geometrical structure of Laplacian eigenfunctions. SIAM Review 55(4), 601–667 (2013) [2, 3]

11. Joshi, S., Pizer, S., Fletcher, P.T., Yushkevich, P., Thall, A., Marron, J.S.: Multiscale deformable model segmentation and statistical shape analysis using medial descriptions. IEEE T Med Imaging 21(5), 538–550 (2002) [1]

12. Konukoglu, E., Glocker, B., Criminisi, A., Pohl, K.M.: WESD–Weighted Spectral Distance for measuring shape dissimilarity. IEEE Trans. Pattern. Anal. Mach. Intell. 35(9), 2284–2297 (2013) [1, 2]

13. Konukoglu, E., Glocker, B., Ye, D.H., Criminisi, A., Pohl, K.M.: Discriminative segmentation-based evaluation through shape dissimilarity. IEEE T Med Imaging 31(12), 2278–2289 (2012) [1, 2]

14. Lombaert, H., Arcaro, M., Ayache, N.: Brain transfer: Spectral analysis of cortical surfaces and functional maps. In: IPMI. pp. 474–487. Springer International Publishing (2015) [1]

15. Myronenko, A., Song, X.: Point set registration: Coherent point drift. IEEE Trans. Pattern. Anal. Mach. Intell. 32(12), 2262–2275 (2010) [1]

16. Niethammer, M., Reuter, M., Wolter, F.E., Bouix, S., Peinecke, N., Koo, M.S., Shenton, M.E.: Global medical shape analysis using the Laplace-Beltrami spectrum. In: MICCAI. pp. 850–857. Springer (2007) [1]

17. Ovsjanikov, M., Ben-Chen, M., Solomon, J., Butscher, A., Guibas, L.: Functional maps: a flexible representation of maps between shapes. ACM Trans Graph 31(4), 30 (2012) [1]

18. Radau, P., Lu, Y., Connelly, K., Paul, G., Dick, A., Wright, G.: Evaluation framework for algorithms segmenting short axis cardiac MRI. The MIDAS Journal 49 (2009) [9]

19. Raviv, D., Bronstein, M.M., Bronstein, A.M., Kimmel, R.: Volumetric heat kernel signatures. In: Proceedings of the ACM Workshop on 3D Object Retrieval. pp. 39–44. ACM (2010) [1]

20. Reuter, M., Wolter, F.E., Peinecke, N.: Laplace–Beltrami spectra as Shape-DNA of surfaces and solids. Computer-Aided Design 38(4), 342–366 (2006) [1, 3, 4]

21. Shakeri, M., Lombaert, H., Datta, A.N., Oser, N., Létourneau-Guillon, L., Lapointe, L.V., Martin, F., Malfait, D., Tucholka, A., Lippé, S., et al.: Statistical shape analysis of subcortical structures using spectral matching. Computerized Medical Imaging and Graphics (2016) [1]

22. Shattuck, D.W., Mirza, M., Adisetiyo, V., Hojatkashani, C., Salamon, G., Narr, K.L., Poldrack, R.A., Bilder, R.M., Toga, A.W.: Construction of a 3D probabilistic atlas of human cortical structures. Neuroimage 39(3), 1064–1080 (2008) [9]

23. Tipping, M.E., Bishop, C.M.: Mixtures of probabilistic Principal Component Analyzers. Neural Comput. 11(2), 443–482 (1999) [6, 7, 8]

24. Wachinger, C., Golland, P., Kremen, W., Fischl, B., Reuter, M., ADNI, et al.: Brainprint: a discriminative characterization of brain morphology. NeuroImage 109, 232–248 (2015) [1, 4]

25. Weyl, H.: Das asymptotische Verteilungsgesetz der Eigenwerte linearer partieller Differentialgleichungen. Mathematische Annalen 71(4), 441–479 (1912) [3]

26. Zhang, M., Fletcher, P.T.: Bayesian Principal Geodesic Analysis for estimating intrinsic diffeomorphic image variability. Medical Image Analysis 25(1), 37–44 (2015) [1]

A Existence of the spectral kernels

The existence of the spectral kernel of Eq. (6) with parameters α, β is conditioned on the convergence of the series of term $(\beta + \lambda_n)^{-\alpha}(\beta + \xi_n)^{-\alpha}$, over all possible spectra $\boldsymbol{\lambda}, \boldsymbol{\xi}$. From the asymptotic behaviour of Eq. (5), the spectral kernel is well-defined if and only if $4\alpha/d > 1$.

In the scale-invariant case $\overline{\boldsymbol{\lambda}} \triangleq V_\Omega^{2/d} \boldsymbol{\lambda}$, the domain of existence can be made larger. To see this let us introduce a suitable variant of the kernel definition that does not change the underlying pseudo-metric. Let $\overline{\boldsymbol{\lambda}}$ and $\overline{\boldsymbol{\xi}}$ be volume renormalized spectra as before and $\overline{\zeta}_n \triangleq 4\pi^2(\frac{n}{B_d})^{2/d}$. Define the spectral kernel $\overline{K}_{\alpha,\beta}$ as:

$$\overline{K}_{\alpha,\beta}(\overline{\boldsymbol{\lambda}}, \overline{\boldsymbol{\xi}}) \triangleq \sum_{n=1}^{+\infty} \left(\frac{1}{(\beta + \overline{\lambda}_n)^\alpha} - \frac{1}{(\beta + \overline{\zeta}_n)^\alpha} \right) \left(\frac{1}{(\beta + \overline{\xi}_n)^\alpha} - \frac{1}{(\beta + \overline{\zeta}_n)^\alpha} \right). \quad (24)$$

Intuitively, $\overline{K}_{\alpha,\beta}$ takes the spectrum $\overline{\zeta}_n$ (corresponding to the asymptotic behaviour) as reference and aggregates perturbations w.r.t. this reference. The corresponding metric $\overline{\rho}_{\alpha,\beta}(\overline{\boldsymbol{\lambda}}, \overline{\boldsymbol{\xi}}) = \|\overline{K}_{\alpha,\beta}(\overline{\boldsymbol{\lambda}}, \cdot) - \overline{K}_{\alpha,\beta}(\overline{\boldsymbol{\xi}}, \cdot)\|_{\mathcal{H}_{\alpha,\beta}}$ has exactly the same form as $\rho_{\alpha,\beta}(\Omega_\lambda, \Omega_\xi)$ after replacing the unnormalized spectra by their normalized counterparts, $\overline{\rho}_{\alpha,\beta}(\overline{\boldsymbol{\lambda}}, \overline{\boldsymbol{\xi}}) = [\sum_{n=1}^{+\infty}(\Delta_{\alpha,\beta}^n(\overline{\boldsymbol{\lambda}}, \overline{\boldsymbol{\xi}}))^2]^{1/2}$, since the reference terms cancel out in the difference. For the same reason, the results returned by the proposed kernel clustering algorithm (section 4) under this reparametrization are also unchanged (up to numerical errors), compared to $K_{\alpha,\beta}(\overline{\boldsymbol{\lambda}}, \overline{\boldsymbol{\xi}})$. Moreover $\overline{K}_{\alpha,\beta}$ exists if and only if $\overline{\rho}_{\alpha,\beta}$ exists (same convergence domain). Because both $\overline{\lambda}_n, \overline{\xi}_n \to \overline{\zeta}_n$ when $n \to +\infty$, the series term $\Delta_{\alpha,\beta}^n(\overline{\boldsymbol{\lambda}}, \overline{\boldsymbol{\xi}})$ decreases in $o((\beta + \overline{\zeta}_n)^{-\alpha})$. To find the domain of convergence of the series, the next order in the asymptotic behaviour of the spectrum is required. For sufficiently regular objects the following finer approximation of Eq. (25) is conjectured to hold [10]:

$$\lambda_n = \frac{1}{V_\Omega^{2/d}} \overline{\zeta}_n \pm C(\partial\Omega) \cdot n^{1/d} + o(n^{1/d}), \quad (25)$$

and with that assumption the domain of definition of $\overline{K}_{\alpha,\beta}$ is $\frac{4\alpha+2}{d} > 1$.

B Kernel truncation error bounds

The following upper bound on the error $\mathcal{E}_{\boldsymbol{\lambda},\boldsymbol{\xi}}^{(N)} \triangleq |K_{\alpha,\beta} - K_{\alpha,\beta}^{(N)}|(\boldsymbol{\lambda}, \boldsymbol{\xi})$ introduced by truncation of the kernel at the Nth term in the series holds. It is a shape-independent bound up to volume normalization:

$$\mathcal{E}_{\boldsymbol{\lambda},\boldsymbol{\xi}}^{(N)} \leq K_{\alpha,\boldsymbol{\lambda},\boldsymbol{\xi}} \cdot \left(\frac{d+2}{d}\right)^{2\alpha} (1/N)^{\frac{4\alpha-d}{d}}, \quad (26)$$

where $K_{\alpha,\boldsymbol{\lambda},\boldsymbol{\xi}} \triangleq \frac{d}{4\alpha-d} \cdot 1/(\Lambda_1\Xi_1)^\alpha$ and Λ_1, Ξ_1 are defined as in Eq. (5). This also translates into a bound of the truncation error for the square of the distance $\rho_{\alpha,\beta}(\Omega_\lambda, \Omega_\xi)$ via the polarization identity: $|\rho_{\alpha,\beta}^2 - \rho_{\alpha,\beta}^{2(N)}|(\boldsymbol{\lambda}, \boldsymbol{\xi}) \leq 4\mathcal{E}_{\boldsymbol{\lambda},\boldsymbol{\xi}}^{(N)}$. Alternatively, the following (generally tighter) bound is derived by direct computation, for $N \geq d + 2$:

$$|\rho_{\alpha,\beta}^2 - \rho_{\alpha,\beta}^{2(N)}|(\boldsymbol{\lambda}, \boldsymbol{\xi}) \leq \rho_{\alpha,\boldsymbol{\lambda},\boldsymbol{\xi}}^2 \cdot (1/N)^{\frac{4\alpha-d}{d}}, \quad (27)$$

where $\rho_{\alpha,\boldsymbol{\lambda},\boldsymbol{\xi}}^2 \triangleq \frac{d}{4\alpha-d} \max[(\Lambda_1 \frac{d}{d+2})^{-\alpha} - (c_d\xi_1)^{-\alpha}, (\Xi_1 \frac{d}{d+2})^{-\alpha} - (c_d\lambda_1)^{-\alpha}]^2$, with $c_d \triangleq 1 + a(d)/d$, $a(1) \leq 2.64$, $a(2) \leq 2.27$ and for $d \geq 3$, $a(d) \triangleq 2.2 - 4\log(1 + \frac{d-3}{50})$. The bound depends only on the volume and first eigenvalue of the objects. These truncation error bounds on the squared distance easily translate onto the distance itself. Note that Eq. (26) derives from the lower bound [30,12] on the eigenvalues $\lambda_n \geq \Lambda_n d/(d+2)$, while Eq. (27) also makes use of the upper bound [27,12] $\lambda_n \leq c_d\lambda_1 n^{2/d}$.

References

27. Cheng, Q.M., Yang, H.: Bounds on eigenvalues of dirichlet laplacian. Mathematische Annalen 337(1), 159–175 (2007)
28. Grebenkov, D.S., Nguyen, B.T.: Geometrical structure of Laplacian eigenfunctions. SIAM Review 55(4), 601–667 (2013) 2, 3
29. Konukoglu, E., Glocker, B., Criminisi, A., Pohl, K.M.: WESD–Weighted Spectral Distance for measuring shape dissimilarity. IEEE Trans. Pattern. Anal. Mach. Intell. 35(9), 2284–2297 (2013) 1, 2
30. Li, P., Yau, S.T.: On the schrödinger equation and the eigenvalue problem. Communications in Mathematical Physics 88(3), 309–318 (1983)