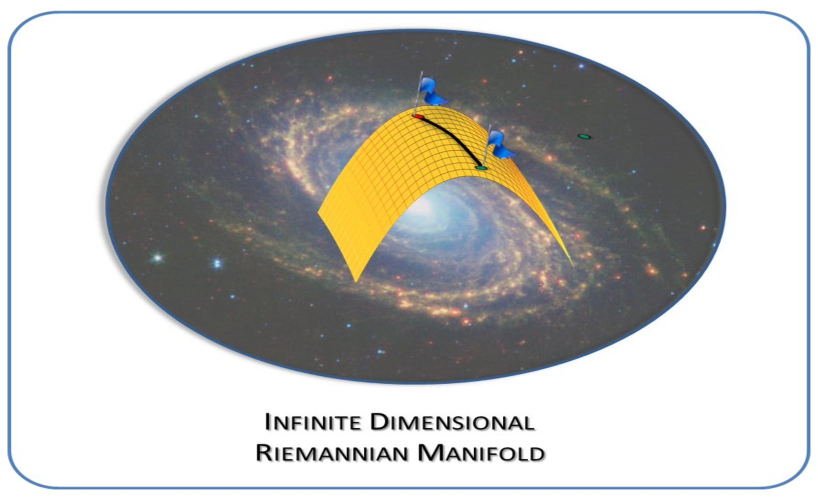
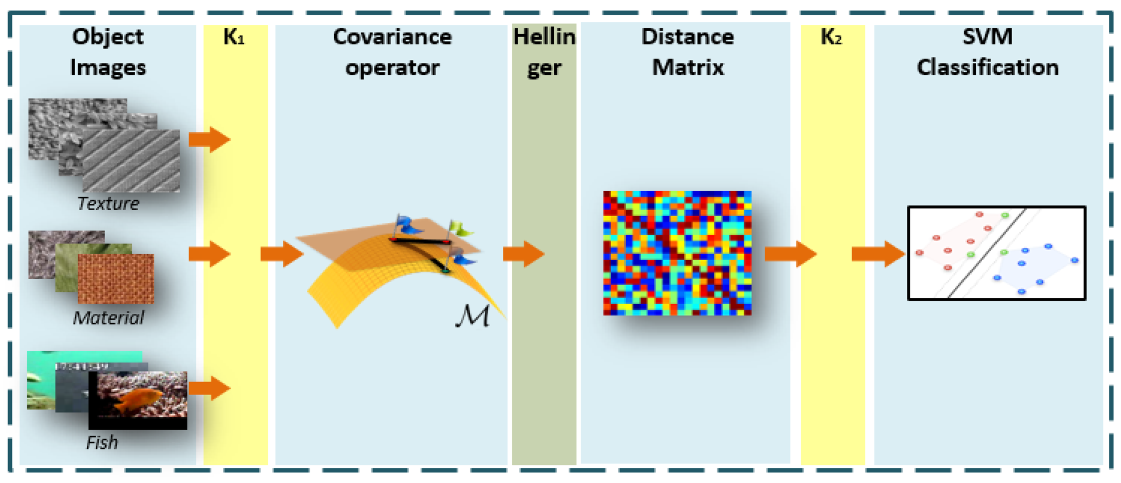
関数解析的学習ユニット　ハクァンミン
Functional Analytic Learning Unit
Ha Quang Minh

Main Research Directions

1) Mathematical structures coming from/related to **Infinite-dimensional Information Geometry** and **Optimal Transport**, in particular in the setting of **reproducing kernel Hilbert spaces** (RKHS), as well as their unification.
2) Theory of RKHS, in particular vector-valued RKHS and operator-valued kernels
3) Novel algorithms and applications in computer vision, functional data analysis, signal processing, brain imaging, and brain computer interfaces.
Current Team members: PI and 1 Postdoctoral researcher

Geometry of covariance operators and Gaussian measures





Two-layer kernel machine with the Hellinger distance in RKHS (for image classification)

Infinite-dimensional distances and divergences

Main result 1: Alpha-Beta Log-Det divergences between positive definite Hilbert-Schmidt operators. A highly general formulation unifying many distances and divergences on the *infinite-dimensional manifold of positive definite Hilbert-Schmidt operators*. Special cases include Rényi, Kullback-Leibler, and Stein divergences and affine-invariant Riemannian distance, between infinite-dimensional covariance operators, Gaussian measures and Gaussian processes.

Let $\mathcal{H}$ be a separable Hilbert space, $\dim(\mathcal{H}) = \infty$, $A : \mathcal{H} \to \mathcal{H}$ a positive, self-adjoint compact operator with eigenvalues $\{\lambda_k\}_{k=1}^{\infty}$

Extended (unitized) trace class operators $\mathrm{Tr}_X(\mathcal{H}) = \{A + \gamma I \ : \ A \in \mathrm{Tr}(\mathcal{H}), \gamma \in \mathbb{R}\}$

Extended trace: $\mathrm{tr}_X(A + \gamma I) = \mathrm{tr}(A) + \gamma$

Fredholm determinant $\det(A + I) = \prod_{k=1}^{\infty}(\lambda_k + 1)$, $A \in \mathrm{Tr}(\mathcal{H}), \gamma = 1$

Extended Fredholm determinant. $\det_X(A + \gamma I) = \gamma \det[(A/\gamma) + I]$, $\gamma \neq 0$

Extended Hilbert-Schmidt operators $\mathrm{HS}_X(\mathcal{H}) = \{A + \gamma I \ : \ A \in \mathrm{HS}(\mathcal{H}), \gamma \in \mathbb{R}\}$

Extended Hilbert-Schmidt norm $||A + \gamma I||^2_{\mathrm{HS}_X} = ||A||^2_{\mathrm{HS}} + \gamma^2$

Hilbert-Carleman determinant $\det_2(I + A) = \det[(I + A)\exp(-A)]$, $A \in \mathrm{HS}(\mathcal{H}), \gamma = 1$

Extended Hilbert-Carleman determinant $\det_{2X}(A + \gamma I) = \det_X[(A + \gamma I)\exp(-A/\gamma)]$, $\gamma \neq 0$,

Definition 1 (**Alpha-Beta Log-Det divergences between positive definite Hilbert-Schmidt operators**). *Let $\alpha > 0$, $\beta > 0$, $r \in \mathbb{R}$, $r \neq 0$ be fixed. For $(A + \gamma I) > 0, (B + \mu I) > 0$, $A, B \in \mathrm{HS}(\mathcal{H})$,*

$$D_r^{(\alpha,\beta)}[(A + \gamma I), (B + \mu I)] = \frac{1}{\alpha\beta} \log \left[\left(\frac{\gamma}{\mu}\right)^{r(\delta - \frac{\alpha}{\alpha+\beta})} \det_X \left(\frac{\alpha(\Lambda + \frac{\gamma}{\mu}I)^{r(1-\delta)} + \beta(\Lambda + \frac{\gamma}{\mu}I)^{-r\delta}}{\alpha + \beta} \right) \right]$$

where $\delta = \frac{\alpha\gamma^r}{\alpha\gamma^r + \beta\mu^r}$, $Z + \frac{\gamma}{\mu}I = (A + \gamma I)(B + \mu I)^{-1}$.

$$D_r^{(\alpha,0)}[(A + \gamma I), (B + \mu I)] = \frac{1}{\alpha^2}[(\frac{\mu}{\gamma})^r - 1](1 + r\log\frac{\mu}{\gamma}) - \frac{1}{\alpha^2}(\frac{\mu}{\gamma})^r \log\det_{2X}([(A + \gamma I)^{-1}(B + \mu I)]^r)$$

$$D_r^{(0,\beta)}[(A + \gamma I), (B + \mu I)] = \frac{1}{\beta^2}[(\frac{\gamma}{\mu})^r - 1](1 + r\log\frac{\gamma}{\mu}) - \frac{1}{\beta^2}(\frac{\gamma}{\mu})^r \log\det_{2X}([(B + \mu I)^{-1}(A + \gamma I)]^r)$$

Limiting case: Infinite-dimensional affine-invariant Riemannian distance (Larotonda, 2007)

$$\lim_{\alpha \to 0} D_{2\alpha}^{(\alpha,\alpha)}[(A + \gamma I), (B + \mu I)] = \frac{1}{2}||\log[(A + \gamma I)^{-1/2}(B + \mu I)(A + \gamma I)^{-1/2}]||^2_{\mathrm{HS}_X}$$

Finite-dimensional case. $A, B \in \mathrm{Sym}^{++}(n)$ ($n \times n$ symmetric, positive definite matrices), $\gamma = \mu = 0$. Alpha-Beta Log-Det divergences (Cichocki et al 2015) $D^{(\alpha,\beta)}(A, B) = \frac{1}{\alpha\beta}\log\det\left[\frac{\alpha(AB^{-1})^{\beta} + \beta(AB^{-1})^{-\alpha}}{\alpha+\beta}\right]$, $\alpha > 0, \beta > 0$, Affine-invariant Riemannian (Fisher-Rao) distance $d_{\mathrm{aiE}}(A, B) = ||\log(A^{-1/2}BA^{-1/2})||_F$, Alpha Log-Det (Rényi) divergences $D^{(\alpha,1-\alpha)}(A, B) = \frac{1}{\alpha(1-\alpha)}\log\frac{\det(\alpha A + (1-\alpha)B)}{\det(A)^{\alpha}\det(B)^{1-\alpha}}$, $0 < \alpha < 1$, Kullback-Leibler divergence $D^{(0,1)}(A, B) = \mathrm{tr}(B^{-1}A - I) - \log\det(B^{-1}A)$, Stein divergence $D^{(1/2,1/2)}(A, B) = 4[\log\det(\frac{A+B}{2}) - \frac{1}{2}\log\det(AB)]$

****

Main result 2: A unified formulation for the infinite-dimensional Bures-Wasserstein metric between covariance operators and Gaussian measures (from Optimal Transport) and the Log-Euclidean/Log-Hilbert-Schmidt metrics

Definition 2 (**Alpha Procrustes distance between positive definite Hilbert-Schmidt operators**). *Let $\gamma > 0$, $\alpha \in \mathbb{R}$, $\alpha \neq 0$ be fixed. For $(A + \gamma I) > 0, (B + \mu I) > 0$, $A, B \in \mathrm{HS}(\mathcal{H})$,*

$$d^{\alpha}_{\mathrm{proHS}}[(A + \gamma I), (B + \gamma I)] = \min_{(I+U) \in \mathbb{U}(\mathcal{H}) \cap \mathrm{HS}_X(\mathcal{H})} \left\| \frac{(A + \gamma I)^{\alpha} - (B + \gamma I)^{\alpha}(I + U)}{\alpha} \right\|_{\mathrm{HS}_X}$$

$$= \frac{1}{|\alpha|}(\mathrm{tr}[(A + \gamma I)^{2\alpha} + (B + \gamma I)^{2\alpha} - 2[(A + \gamma I)^{\alpha}(B + \gamma I)^{2\alpha}(A + \gamma I)^{\alpha}]^{1/2}])^{1/2}$$

Special case: Bures-Wasserstein distance. For $A, B \in \mathrm{Sym}^+(\mathcal{H}) \cap \mathrm{Tr}(\mathcal{H})$,

$$\lim_{\gamma \to 0} d^{1/2}_{\mathrm{proHS}}[(A + \gamma I), (B + \gamma I)] = 2(\mathrm{tr}[A + B - 2(A^{1/2}BA^{1/2})^{1/2}])^{1/2}$$

Limiting case: Log-Hilbert-Schmidt distance (H.Q.M. et al 2014).

$$\lim_{\alpha \to 0} d^{\alpha}_{\mathrm{proHS}}[(A + \gamma I), (B + \gamma I)] = ||\log(A + \gamma I) - \log(B + \gamma I)||_{\mathrm{HS}_X}$$

RKHS covariance operators

Let $\mathcal{X}$ be a separable topological space, K a continuous positive definite kernel on $\mathcal{X} \times \mathcal{X}$, then the corresponding RKHS $\mathcal{H}_K$ is a separable Hilbert space. Consider the feature map $\Phi : \mathcal{X} \to \mathcal{H}_K$, so that $K(x, y) = \langle\Phi(x), \Phi(y)\rangle_{\mathcal{H}_K}$ $\forall (x, y) \in \mathcal{X} \times \mathcal{X}$. Let $\mathbf{X} = [x_1, \ldots, x_m]$ be a data matrix randomly sampled from $\mathcal{X}$ according to some probability distribution. This defines the bounded operator $\Phi(\mathbf{X}) : \mathbb{R}^m \to \mathcal{H}_K$ by $\Phi(\mathbf{X})\mathbf{b} = \sum_{j=1}^m b_j\Phi(x_j)$, $\mathbf{b} \in \mathbb{R}^m$, or equivalently the (infinite) data matrix $\Phi(\mathbf{X}) = [\Phi(x_1), \ldots, \Phi(x_m)]$ in $\mathcal{H}_K$. Empirical covariance operator

$$C_{\Phi(\mathbf{X})} = \frac{1}{m}\Phi(\mathbf{X})J_m\Phi(\mathbf{X})^* : \mathcal{H}_K \to \mathcal{H}_K,$$

$J_m = I_m - \frac{1}{m}\mathbf{1}_m\mathbf{1}_m^T$ being the centering matrix. For two covariance operators $C_{\Phi(\mathbf{X})}$ and $C_{\Phi(\mathbf{Y})}$, the distances and divergences

$$d[(C_{\Phi(\mathbf{X})} + \gamma I_{\mathcal{H}_K}), (C_{\Phi(\mathbf{Y})} + \mu I_{\mathcal{H}_K})],$$

all have closed forms expressed in terms of the corresponding Gram matrices.

References

[1] H.Q.Minh. Alpha-Beta Log-Determinant divergences between positive definite trace class operators. **Information Geometry 2019**.

[2] H.Q.Minh. Infinite-dimensional Log-Determinant divergences between positive definite Hilbert-Schmidt operators. **Positivity 2019**.

[3] H.Q.Minh. A unified formulation for the Bures-Wasserstein and Log-Euclidean/Log-Hilbert-Schmidt distances between positive definite operators. **Geometric Science of Information 2019**.

[4] J.C. Guella. On the universality of operator-valued positive definite kernels, *in preparation*.

Operator-valued kernels

Main result: Generalization of results on the universality of positive definite kernels from the scalar to operator-valued setting.

Theorem 1. *Let $F : [0, \infty) \to \mathcal{L}(\mathcal{H})$ be an ultraweakly continuous function for which the kernel $K : \mathbb{R}^m \times \mathbb{R}^m \to \mathcal{L}(\mathcal{H})$ given by $K(x, y) := F(\|x - y\|)$ is positive definite for every $m \in \mathbb{N}$ and $F(0) \in \mathcal{L}(\mathcal{H})$ is a trace class operator. Then $\mathcal{H}_{K_F} \subset C(\mathbb{R}^m, \mathcal{H})$ and the following are equivalent*

(i) The kernel K is strictly positive definite.

(ii) The kernel K is universal.

(iii) For every $v \in \mathcal{H} \setminus \{0\}$ the function $t \in [0, \infty) \to \langle F(t)v, v\rangle_{\mathcal{H}}$ is non-constant.

Moreover, $\mathcal{H}_{K_F} \subset C_0(\mathbb{R}^m, \mathcal{H})$ if and only if for every $v \in \mathcal{H} \setminus \{0\}$ the function $t \in [0, \infty) \to \langle F(t)v, v\rangle_{\mathcal{H}} \in C_0([0, \infty))$. Under this additional hypothesis, the following are equivalent

(i) The kernel K is strictly positive definite.

(ii) The kernel K is C_0−universal.

(iii) For every $v \in \mathcal{H} \setminus \{0\}$ the function $t \in [0, \infty) \to \langle F(t)v, v\rangle_{\mathcal{H}}$ is nonzero.