From Covariance Matrices to Covariance Operators

Data Representation from Finite to Infinite-Dimensional Settings

Hà Quang Minh

Pattern Analysis and Computer Vision (PAVIS)
Istituto Italiano di Tecnologia, ITALY

ISTITUTO ITALIANO
DI TECNOLOGIA

November 30, 2017

Acknowledgement

Collaborators

- Marco San Biagio
- Vittorio Murino

ISTITUTO ITALIANO
DI TECNOLOGIA

From finite to infinite dimensions

Part I - Outline

Finite-dimensional setting

Covariance Matrices and Applications

1. Data Representation by Covariance Matrices
2. Geometry of SPD matrices
3. Machine Learning Methods on Covariance Matrices and Applications in Computer Vision

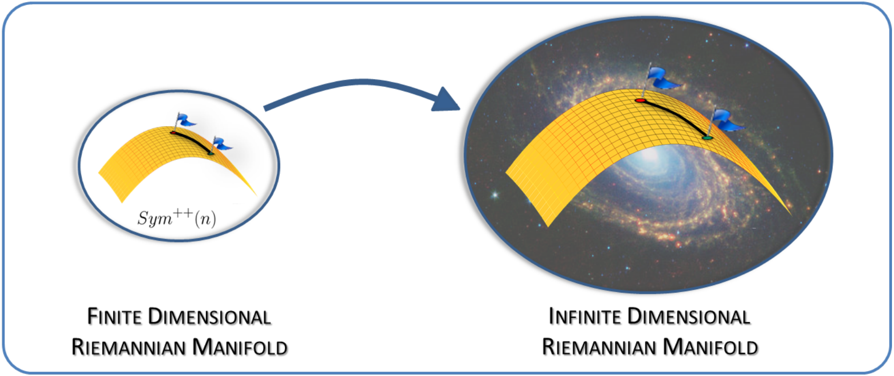
Part II - Outline

Infinite-dimensional setting

Covariance Operators and Applications

1. Data Representation by Covariance Operators
2. Geometry of Covariance Operators
3. Machine Learning Methods on Covariance Operators and Applications in Computer Vision

From finite to infinite-dimensional settings



Part II - Outline

Infinite-dimensional setting

Covariance Operators and Applications

1. Data Representation by Covariance Operators
2. Geometry of Covariance Operators
3. Machine Learning Methods on Covariance Operators and Applications in Computer Vision

Covariance operator representation - Motivation

- Covariance matrices encode linear correlations of input features
- Nonlinearization
  1. Map original input features into a high (generally infinite) dimensional feature space (via kernels)
  2. Covariance operators: covariance matrices of infinite-dimensional features
  3. Encode nonlinear correlations of input features
  4. Provide a richer, more expressive representation of the data

Covariance operator representation

- S.K. Zhou and R. Chellappa. From sample similarity to ensemble similarity: Probabilistic distance measures in reproducing kernel Hilbert space, PAMI 2006
- M. Harandi, M. Salzmann, and F. Porikli. Bregman divergences for infinite-dimensional covariance matrices, CVPR 2014
- H.Q.Minh, M. San Biagio, V. Murino. Log-Hilbert-Schmidt metric between positive definite operators on Hilbert spaces, NIPS 2014
- H.Q.Minh, M. San Biagio, L. Bazzani, V. Murino. Approximate Log-Hilbert-Schmidt distances between covariance operators for image classification, CVPR 2016

Positive Definite Kernels

- $\mathcal{X}$ any nonempty set
- $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is a (real-valued) positive definite kernel if it is symmetric and

$$\sum_{i,j=1}^{N} a_i a_j K(x_i, x_j) \geq 0$$

  for any finite set of points $\{x_i\}_{i=1}^N \in \mathcal{X}$ and real numbers $\{a_i\}_{i=1}^N \in \mathbb{R}$.
- $[K(x_i, x_j)]_{i,j=1}^N$ is symmetric positive semi-definite

Reproducing Kernel Hilbert Spaces

- K a positive definite kernel on $\mathcal{X} \times \mathcal{X}$. For each $x \in \mathcal{X}$, there is a function $K_x : \mathcal{X} \to \mathbb{R}$, with $K_x(t) = K(x, t)$.

$$\mathcal{H}_K = \overline{\{\sum_{i=1}^{N} a_i K_{x_i} : N \in \mathbb{N}\}}$$

  with inner product

$$\langle \sum_i a_i K_{x_i}, \sum_j b_j K_{y_j} \rangle_{\mathcal{H}_K} = \sum_{i,j} a_i b_j K(x_i, y_j)$$

- $\mathcal{H}_K$ = RKHS associated with K (unique).

Reproducing Kernel Hilbert Spaces

- **Reproducing property**: for each $f \in \mathcal{H}_K$, for every $x \in \mathcal{X}$

$$f(x) = \langle f, K_x \rangle_{\mathcal{H}_K}$$

- Abstract theory due to Aronszajn (1950)
- Numerous applications in machine learning (kernel methods)

Examples: RKHS

- The Gaussian kernel $K(x,y) = \exp(-\frac{|x-y|^2}{\sigma^2})$ on $\mathbb{R}^n$ induces the space

$$\mathcal{H}_K = \{||f||^2_{\mathcal{H}_K} = \frac{1}{(2\pi)^n(\sigma\sqrt{\pi})^n}\int_{\mathbb{R}^n} e^{\frac{\sigma^2|\xi|^2}{4}}|\widehat{f}(\xi)|^2 d\xi < \infty\}.$$

- The Laplacian kernel $K(x,y) = \exp\left(-a|x-y|\right)$, $a > 0$, on $\mathbb{R}^n$ induces the space

$$\mathcal{H}_K = \{||f||^2_{\mathcal{H}_K} = \frac{1}{(2\pi)^n}\frac{1}{aC(n)}\int_{\mathbb{R}^n}(a^2+|\xi|^2)^{\frac{n+1}{2}}|\hat{f}(\xi)|^2 d\xi < \infty\}.$$

with $C(n) = 2^n\pi^{\frac{n-1}{2}}\Gamma(\frac{n+1}{2})$

Feature map and feature space

- Geometric viewpoint from machine learning
- Positive definite kernel K on $\mathcal{X} \times \mathcal{X}$ induces feature map $\Phi : \mathcal{X} \to \mathcal{H}_K$

$$\Phi(x) = K_x \in \mathcal{H}_K, \quad \mathcal{H}_K = \text{feature space}$$
$$\langle \Phi(x), \Phi(y) \rangle_{\mathcal{H}_K} = \langle K_x, K_y \rangle_{\mathcal{H}_K} = K(x, y)$$

- Kernelization: Transform linear algorithm depending on

$$\langle x, y \rangle_{\mathbb{R}^n}$$

into nonlinear algorithms depending on

$$K(x, y)$$

Covariance matrices

ρ = Borel probability distribution on $\mathcal{X} \subset \mathbb{R}^n$, with

$$\int_{\mathcal{X}} ||x||^2 d\rho(x) < \infty$$

- Mean vector

$$\mu = \int_{\mathcal{X}} x d\rho(x)$$

- Covariance matrix

$$C = \int_{\mathcal{X}} (x - \mu)(x - \mu)^T d\rho(x)$$

RKHS covariance operators

ρ = Borel probability distribution on $\mathcal{X}$, with

$$\int_{\mathcal{X}} ||\Phi(x)||^2_{\mathcal{H}_K} d\rho(x) = \int_{\mathcal{X}} K(x,x) d\rho(x) < \infty$$

- RKHS mean vector

$$\mu_\Phi = \mathbb{E}_\rho[\Phi(x)] = \int_{\mathcal{X}} \Phi(x) d\rho(x) \in \mathcal{H}_K$$

- For any $f \in \mathcal{H}_K$

$$\langle \mu_\Phi, f \rangle_{\mathcal{H}_K} = \int_{\mathcal{X}} \langle f, \Phi(x) \rangle_{\mathcal{H}_K} d\rho(x) = \int_{\mathcal{X}} f(x) d\rho(x) = \mathbb{E}_\rho[f]$$

RKHS covariance operators

- RKHS covariance operator $C_\Phi : \mathcal{H}_K \to \mathcal{H}_K$

$$C_\Phi = \mathbb{E}_\rho[(\Phi(x) - \mu) \otimes (\Phi(x) - \mu)]$$
$$= \int_{\mathcal{X}} \Phi(x) \otimes \Phi(x) d\rho(x) - \mu \otimes \mu$$

- For all $f, g \in \mathcal{H}_K$

$$\langle f, C_\Phi g\rangle_{\mathcal{H}_K} = \int_{\mathcal{X}} \langle f, \Phi(x)\rangle_{\mathcal{H}_K} \langle g, \Phi(x)\rangle_{\mathcal{H}_K} d\rho(x) - \langle \mu, f\rangle_{\mathcal{H}_K} \langle \mu, g\rangle_{\mathcal{H}_K}$$
$$= \mathbb{E}_\rho(fg) - \mathbb{E}_\rho(f)\mathbb{E}_\rho(g)$$

Empirical mean and covariance

- $\mathbf{X} = [x_1, \ldots, x_m]$ = data matrix randomly sampled from $\mathcal{X}$ according to ρ, with m observations
- Informally, Φ gives an infinite feature matrix in the feature space $\mathcal{H}_K$, of size $\dim(\mathcal{H}_K) \times m$

$$\Phi(\mathbf{X}) = [\Phi(x_1), \ldots, \Phi(x_m)]$$

- Formally, $\Phi(\mathbf{X}) : \mathbb{R}^m \to \mathcal{H}_K$ is the bounded linear operator

$$\Phi(\mathbf{X})w = \sum_{i=1}^{m} w_i \Phi(x_i), \quad w \in \mathbb{R}^m$$

Empirical mean and covariance

- Theoretical RKHS mean

$$\mu_{\Phi} = \int_{\mathcal{X}} \Phi(x) d\rho(x) \in \mathcal{H}_K$$

- Empirical RKHS mean

$$\mu_{\Phi(\mathbf{X})} = \frac{1}{m}\sum_{i=1}^{m} \Phi(x_i) = \frac{1}{m}\Phi(\mathbf{X})\mathbf{1}_m \in \mathcal{H}_K$$

Empirical mean and covariance

- Theoretical covariance operator $C_\Phi : \mathcal{H}_K \to \mathcal{H}_K$

$$C_\Phi = \int_{\mathcal{X}} \Phi(x) \otimes \Phi(x) d\rho(x) - \mu \otimes \mu$$

- Empirical covariance operator $C_{\Phi(\mathbf{x})} : \mathcal{H}_K \to \mathcal{H}_K$

$$\begin{aligned} C_{\Phi(\mathbf{X})} &= \frac{1}{m} \sum_{i=1}^{m} \Phi(x_i) \otimes \Phi(x_i) - \mu_{\Phi(\mathbf{X})} \otimes \mu_{\Phi(\mathbf{X})} \\ &= \frac{1}{m} \Phi(\mathbf{X}) J_m \Phi(\mathbf{X})^* \end{aligned}$$

$J_m = I_m - \frac{1}{m}\mathbf{1}_m \mathbf{1}_m^T$ = centering matrix

RKHS covariance operators

- C_{Φ} is a self-adjoint, positive, trace class operator, with a countable set of eigenvalues $\{\lambda_k\}_{k=1}^{\infty}$, $\lambda_k \geq 0$ and

$$\sum_{k=1}^{\infty} \lambda_k < \infty$$

- $C_{\Phi(\mathbf{x})}$ is a self-adjoint, positive, finite rank operator

Covariance operator representation of images

- Given an image F (or a patch in F), at each pixel, extract a feature vector (e.g. intensity, colors, filter responses etc)
- Each image corresponds to a data matrix $\mathbf{X}$

$$\mathbf{X} = [x_1, \ldots, x_m] = n \times m \quad \text{matrix}$$

  where m = number of pixels, n = number of features at each pixel
- Define a kernel K, with corresponding feature map Φ and feature matrix

$$\Phi(\mathbf{X}) = [\Phi(x_1), \ldots, \Phi(x_m)]$$

Covariance operator representation of images

- Each image is represented by covariance operator

$$C_{\Phi(\mathbf{X})} = \frac{1}{m}\Phi(\mathbf{X})J_m\Phi(\mathbf{X})^*$$

- This representation is implicit, since Φ is generally implicit
- Computations are carried out via Gram matrices
- Can be approximated by explicit, low-dimensional approximation of Φ

Part II - Outline

Infinite-dimensional setting

Covariance Operators and Applications

1. Data Representation by Covariance Operators
2. Geometry of Covariance Operators
3. Machine Learning Methods on Covariance Operators and Applications in Computer Vision

Geometry of Covariance Operators

1. Hilbert-Schmidt metric: generalizing the Euclidean metric
2. Manifold of positive definite operators
   - Infinite-dimensional generalization of $\mathrm{Sym}^{++}(n)$
   - Infinite-dimensional affine-invariant Riemannian metric: generalizing finite-dimensional affine-invariant Riemannian metric
   - Log-Hilbert-Schmidt metric: generalizing Log-Euclidean metric
3. Convex cone of positive definite operators
   - Log-Determinant divergences: generalize the finite-dimensional Log-Determinant divergences

Geometry of Covariance Operators

1. Hilbert-Schmidt metric: generalizing the Euclidean metric
2. Manifold of positive definite operators
   - Infinite-dimensional generalization of $\mathrm{Sym}^{++}(n)$
   - Infinite-dimensional affine-invariant Riemannian metric: generalizing finite-dimensional affine-invariant Riemannian metric
   - Log-Hilbert-Schmidt metric: generalizing Log-Euclidean metric
3. Convex cone of positive definite operators
   - Log-Determinant divergences: generalize the finite-dimensional Log-Determinant divergences

Hilbert-Schmidt metric

Generalizing Frobenius inner product and distance on $\mathrm{Sym}^{++}(n)$

$$\langle A, B\rangle_F = \mathrm{tr}(A^T B), \quad d_E(A, B) = ||A - B||_F$$

- $\mathcal{H}$ = infinite-dimensional separable real Hilbert space, with a countable orthonormal basis $\{e_k\}_{k=1}^{\infty}$
- $A : \mathcal{H} \to \mathcal{H}$ = bounded linear operator

$$||A|| = \sup_{x\in\mathcal{H}, x\neq 0} \frac{||Ax||}{||x||} < \infty$$

- Adjoint operator $A^* : \mathcal{H} \to \mathcal{H}$ (transpose if $\mathcal{H} = \mathbb{R}^n$)

$$\langle Ax, y\rangle = \langle x, A^* y\rangle \quad \forall x, y \in \mathcal{H}$$

- Self-adjoint: $A^* = A$

Hilbert-Schmidt metric

Generalizing the Frobenius inner product $\langle A, B \rangle_F = \mathrm{tr}(A^T B)$

- Hilbert-Schmidt inner product

$$\langle A, B \rangle_{\mathrm{HS}} = \mathrm{tr}(A^* B) = \sum_{k=1}^{\infty} \langle e_k, A^* B e_k \rangle = \sum_{k=1}^{\infty} \langle A e_k, B e_k \rangle$$

between Hilbert-Schmidt operators on the Hilbert space $\mathcal{H}$

$$\mathrm{HS}(\mathcal{H}) = \{A : ||A||_{\mathrm{HS}}^2 = \mathrm{tr}(A^* A) = \sum_{k=1}^{\infty} ||A e_k||^2 < \infty\}$$

for any orthonormal basis $\{e_k\}_{k \in \mathbb{N}}$

Hilbert-Schmidt metric

Generalizing the Frobenius norm $||A||_F$

- Hilbert-Schmidt norm

$$||A||^2_{\mathrm{HS}} = \mathrm{tr}(A^*A) = \sum_{k=1}^{\infty} ||Ae_k||^2 < \infty\}$$

- A is Hilbert-Schmidt $\Rightarrow$ A is compact, with a countable set of eigenvalues $\{\lambda_k\}_{k=1}^{\infty}$
- If A is self-adjoint

$$||A||^2_{\mathrm{HS}} = \sum_{k=1}^{\infty} \lambda_k^2$$

Hilbert-Schmidt distance between RKHS covariance operators

- Two random data matrices $\mathbf{X} = [x_1, \ldots, x_m]$ and $\mathbf{Y} = [y_1, \ldots, y_m]$ sampled according to two Borel probability distributions ρ_X and ρ_Y on $\mathcal{X}$ give rise to two empirical covariance operators

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

- Closed-form expressions for $||C_{\Phi(\mathbf{X})} - C_{\Phi(\mathbf{Y})}||_{\mathrm{HS}}$ in terms of the $m \times m$ Gram matrices

$$K[\mathbf{X}] = \Phi(\mathbf{X})^*\Phi(\mathbf{X}), (K[\mathbf{X}])_{ij} = K(x_i, x_j),$$
$$K[\mathbf{Y}] = \Phi(\mathbf{Y})^*\Phi(\mathbf{Y}), (K[\mathbf{Y}])_{ij} = K(y_i, y_j),$$
$$K[\mathbf{X}, \mathbf{Y}] = \Phi(\mathbf{X})^*\Phi(\mathbf{Y}), (K[\mathbf{X}, \mathbf{Y}])_{ij} = K(x_i, y_j)$$
$$K[\mathbf{Y}, \mathbf{X}] = \Phi(\mathbf{Y})^*\Phi(\mathbf{X}), (K[\mathbf{Y}, \mathbf{x}])_{ij} = K(y_i, x_j)$$

Hilbert-Schmidt distance between RKHS covariance operators

Proposition

The Hilbert-Schmidt distance between $C_{\Phi(\mathbf{X})}$ and $C_{\Phi(\mathbf{Y})}$

$$||C_{\Phi(\mathbf{X})} - C_{\Phi(\mathbf{Y})}||^2_{\mathrm{HS}} = \frac{1}{m^2}\langle J_m K[\mathbf{X}], K[\mathbf{X}]J_m\rangle_F - \frac{2}{m^2}\langle J_m K[\mathbf{X},\mathbf{Y}], K[\mathbf{X},\mathbf{Y}]J_m\rangle_F + \frac{1}{m^2}\langle J_m K[\mathbf{Y}], K[\mathbf{Y}]J_m\rangle_F.$$

The Hilbert-Schmidt inner product between $C_{\Phi(\mathbf{X})}$ and $C_{\Phi(\mathbf{Y})}$

$$\langle C_{\Phi(\mathbf{X})}, C_{\Phi(\mathbf{Y})}\rangle_{\mathrm{HS}} = \frac{1}{m^2}\langle J_m K[\mathbf{X},\mathbf{Y}], K[\mathbf{X},\mathbf{Y}]J_m\rangle_F.$$

Hilbert-Schmidt distance - Computational complexity

- Requires the calculation of $m \times m$ Gram matrices and their pointwise multiplication. The computational complexity required is $O(m^2)$.
- For set of N data matrices $\{X_j\}_{j=1}^N$, then for computing all the pairwise Hilbert-Schmidt distances/inner products between the corresponding covariance operators $\{C_{\Phi(\mathbf{x}_j)}\}_{j=1}^N$, the computational complexity required is $O(N^2 m^2)$.

Geometry of Covariance Operators

1. Hilbert-Schmidt metric: generalizing the Euclidean metric
2. Manifold of positive definite operators
    - Infinite-dimensional generalization of $\mathrm{Sym}^{++}(n)$
    - Infinite-dimensional affine-invariant Riemannian metric: generalizing finite-dimensional affine-invariant Riemannian metric
    - Log-Hilbert-Schmidt metric: generalizing Log-Euclidean metric
3. Convex cone of positive definite operators
    - Log-Determinant divergences: generalize the finite-dimensional Log-Determinant divergences

Infinite-dimensional generalization of $\mathrm{Sym}^{++}(n)$

- The infinite-dimensional setting is significantly different from the finite-dimensional setting
- The functions $\log$, $\det$, different norms $|| \ ||$ are defined for specific classes of operators
- The regularization $(A + \gamma I)$ is necessary both theoretically and empirically

Infinite-dimensional generalization of $\mathrm{Sym}^{++}(n)$

- Motivation : Generalizing the Log-Euclidean distance

$$d_{\mathrm{logE}}(A, B) = ||\log(A) - \log(B)||_F, \quad A, B \in \mathrm{Sym}^{++}(n)$$

  to the setting where A, B are self-adjoint, positive, bounded operators on a separable Hilbert space $\mathcal{H}$

- Two issues to consider
  1. Generalization of the principal matrix log function
  2. Generalization of the Frobenius inner product and norm (the Hilbert-Schmidt norm is not sufficient)

Infinite-dimensional generalization of $\mathrm{Sym}^{++}(n)$

Generalizing the principal matrix $\log(A)$ function

- $A \in \mathrm{Sym}^{++}(n)$, with eigenvalues $\{\lambda_k\}_{k=1}^n$ and orthonormal eigenvectors $\{\mathbf{u}_k\}_{k=1}^n$

$$A = \sum_{k=1}^{n} \lambda_k \mathbf{u}_k \mathbf{u}_k^T, \quad \log(A) = \sum_{k=1}^{n} \log(\lambda_k) \mathbf{u}_k \mathbf{u}_k^T$$

- $A : \mathcal{H} \to \mathcal{H}$ self-adjoint, positive, compact operator, with eigenvalues $\{\lambda_k\}_{k=1}^\infty$, $\lambda_k > 0$, $\lim_{k\to\infty} \lambda_k = 0$, and orthonormal eigenvectors $\{\mathbf{u}_k\}_{k=1}^\infty$

$$A = \sum_{k=1}^{\infty} \lambda_k (\mathbf{u}_k \otimes \mathbf{u}_k), \quad (\mathbf{u}_k \otimes \mathbf{u}_k) w = \langle \mathbf{u}_k, w \rangle \mathbf{u}_k$$

$$\log(A) = \sum_{k=1}^{\infty} \log(\lambda_k)(\mathbf{u}_k \otimes \mathbf{u}_k), \quad \lim_{k\to\infty} \log(\lambda_k) = -\infty$$

Infinite-dimensional generalization of $\mathrm{Sym}^{++}(n)$

First problem: unboundedness of $\log(A)$ since $\lim_{k\to\infty}\lambda_k = 0$

$$\log(A) = \sum_{k=1}^{\infty} \log(\lambda_k)(\mathbf{u}_k \otimes \mathbf{u}_k), \quad \lim_{k\to\infty} \log(\lambda_k) = -\infty$$

Resolution: positive definite operators

- Strict positivity, i.e. $\lambda_k > 0 \;\; \forall k \in \mathbb{N}$ is not sufficient. We need A to be positive definite

$$\langle Ax, x\rangle \geq M_A ||x||^2, \quad M_A > 0$$

- The eigenvalues of A are bounded from below by $M_A > 0$

Infinite-dimensional generalization of $\mathrm{Sym}^{++}(n)$

- Finite-dimensional setting: $\{\lambda_k\}_{k=1}^{n}$, $\lambda_k > 0$

$$A \text{ is positive definite} \Longleftrightarrow A \text{ is strictly positive}$$

- Infinite-dimensional setting: $\{\lambda_k\}_{k=1}^{\infty}$, $\lambda_k > 0$, $\lim_{k\to\infty} \lambda_k = 0$

$$A \text{ is positive definite} \Longleftrightarrow \begin{cases} A \text{ is strictly positive and} \\ A \text{ is invertible} \end{cases}$$

Infinite-dimensional generalization of $\mathrm{Sym}^{++}(n)$

- First problem: unboundedness of $\log(A)$
- Resolution: positive definite operators
- Regularization: examples of positive definite operators are regularized operators of the form

$$\{(A + \gamma I) > 0 \mid \gamma \in \mathbb{R}, \gamma > 0\}$$

where A is self-adjoint, compact, positive. Then

$$\log(A + \gamma I)$$

is well-defined and bounded.

Infinite-dimensional generalization of $\mathrm{Sym}^{++}(n)$

- Generalizing the Log-Euclidean distance

$$d_{\mathrm{logE}}(A, B) = ||\log(A) - \log(B)||_F, \quad A, B \in \mathrm{Sym}^{++}(n)$$

  to the setting where A, B are self-adjoint, positive, bounded operators on a separable Hilbert space $\mathcal{H}$

- Two issues to consider
  1. Generalization of the principal matrix $\log$ function
  2. Generalization of the Frobenius inner product and norm (the Hilbert-Schmidt norm is not sufficient)

Infinite-dimensional generalization of $\mathrm{Sym}^{++}(n)$

Second problem: The identity operator I is not Hilbert-Schmidt:

$$||I||_{\mathrm{HS}} = \mathrm{tr}(I) = \infty$$

- For $\gamma \neq 1$

$$||\log(A + \gamma I)||^2_{\mathrm{HS}} = \sum_{k=1}^{\infty}[\log(\lambda_k + \gamma)]^2 = \infty$$

- For $\gamma \neq \nu$

$$d(\gamma I, \nu I) = ||\log(\gamma/\nu) I||_{\mathrm{HS}} = |\log(\gamma/\nu)| \; ||I||_{\mathrm{HS}} = \infty$$

Infinite-dimensional generalization of $\mathrm{Sym}^{++}(n)$

- Second problem: The identity I is not Hilbert-Schmidt
- Resolution: Extended (unitized) Hilbert-Schmidt algebra (Larotonda, *Differential Geometry and Its Applications*, 2007)

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

- Extended (unitized) Hilbert-Schmidt inner product

$$\langle A + \gamma I, B + \nu I\rangle_{\mathrm{eHS}} = \langle A, B\rangle_{\mathrm{HS}} + \gamma\nu$$

i.e. the scalar operators γI are orthogonal to the Hilbert-Schmidt operators

$$||A + \gamma I||^2_{\mathrm{eHS}} = ||A||^2_{\mathrm{HS}} + \gamma^2, \quad ||I||_{\mathrm{eHS}} = 1$$

Infinite-dimensional generalization of $\mathrm{Sym}^{++}(n)$

- Hilbert manifold of positive definite (unitized) Hilbert-Schmidt operators (Larotonda 2007)

$$\Sigma(\mathcal{H}) = \{A + \gamma I > 0 : A^* = A, A \in \mathrm{HS}(\mathcal{H}), \gamma \in \mathbb{R}\}$$

- Infinite-dimensional manifold
- For $(A + \gamma I) \in \Sigma(\mathcal{H})$

$$||\log(A + \gamma I)||^2_{\mathrm{eHS}} = \left\|\log\left(\frac{A}{\gamma} + I\right)\right\|^2_{\mathrm{HS}} + (\log\gamma)^2 < \infty$$

Infinite-dimensional generalization of $\mathrm{Sym}^{++}(n)$

Manifold $\mathrm{Sym}^{++}(n)$ (SPD matrices) generalizes to

Manifold $\Sigma(\mathcal{H})$ (positive definite operators)

Two major differences with the finite-dimensional setting

1. $\log(A)$ is unbounded for a compact operator A
2. The identity operator I is not Hilbert-Schmidt

Resolutions

1. The regularization form $(A + \gamma I)$, which is positive definite is necessary both theoretically and empirically
2. Extended (unitized) Hilbert-Schmidt inner product and norm

Infinite-dimensional generalization of $\mathrm{Sym}^{++}(n)$

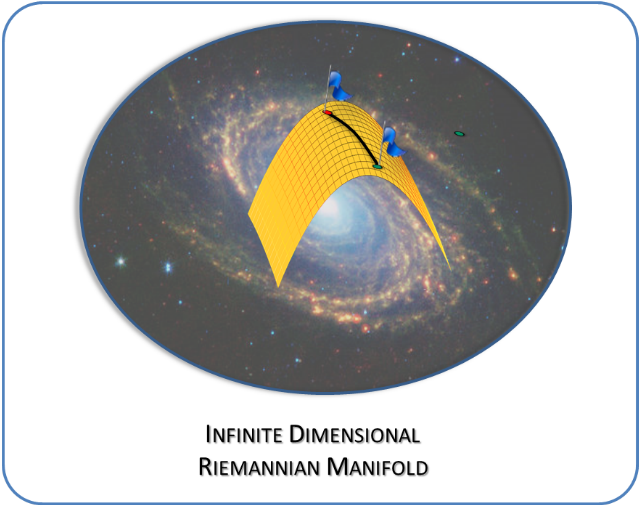
Manifold $\mathrm{Sym}^{++}(n)$ (SPD matrices) generalizes to

Manifold $\Sigma(\mathcal{H})$ (positive definite operators)

The following are both well-defined on $\Sigma(\mathcal{H})$

- Generalization of Affine-invariant Riemannian metric
- Generalization of Log-Euclidean metric

Infinite-dimensional generalization of $\mathrm{Sym}^{++}(n)$



Geometry of Covariance Operators

1. Hilbert-Schmidt metric: generalizing the Euclidean metric
2. Manifold of positive definite operators
   - Infinite-dimensional generalization of $\mathrm{Sym}^{++}(n)$
   - Infinite-dimensional affine-invariant Riemannian metric: generalizing finite-dimensional affine-invariant Riemannian metric
   - Log-Hilbert-Schmidt metric: generalizing Log-Euclidean metric
3. Convex cone of positive definite operators
   - Log-Determinant divergences: generalize the finite-dimensional Log-Determinant divergences

Affine-invariant Riemannian metric

- Affine-invariant Riemannian metric: Larotonda (2005), Larotonda (2007), Andruchow and Varela (2007), Lawson and Lim (2013)
- Larotonda, Nonpositive curvature: A geometrical approach to Hilbert-Schmidt operators, *Differential Geometry and Its Applications*, 2007

In the setting of RKHS covariance operators

- H.Q.M. Affine-invariant Riemannian distance between infinite-dimensional covariance operators, *Geometric Science of Information*, 2015

Affine-invariant Riemannian metric

Tangent space

- Finite-dimensional

$$T_P(\mathrm{Sym}^{++}(n)) \cong \mathrm{Sym}(n), \quad \forall P \in \mathrm{Sym}^{++}(n)$$

- Infinite-dimensional: (Larotonda 2007)

$$T_P(\Sigma(\mathcal{H})) \cong \mathcal{H}_{\mathbb{R}} \quad \forall P \in \Sigma(\mathcal{H})$$
$$\mathcal{H}_{\mathbb{R}} = \{A + \gamma I \ : \ A^* = A, A \in \mathrm{HS}(\mathcal{H}), \gamma \in \mathbb{R}\}$$

Riemannian metric

- Finite-dimensional: For $P \in \mathrm{Sym}^{++}(n)$

$$\langle A, B \rangle_P = \langle P^{-1/2} A P^{-1/2}, P^{-1/2} B P^{-1/2} \rangle_F$$

where $A, B \in \mathrm{Sym}(n)$

- Infinite-dimensional: (Larotonda 2007) For $P \in \Sigma(\mathcal{H})$

$$\begin{aligned}&\langle (A+\gamma I), (B+\nu I) \rangle_P \\ &= \langle P^{-1/2}(A+\gamma I)P^{-1/2}, P^{-1/2}(B+\nu I)P^{-1/2} \rangle_{\mathrm{eHS}}\end{aligned}$$

where $(A+\gamma I), (B+\nu I) \in \mathcal{H}_{\mathbb{R}}$

Affine invariance

- Finite-dimensional

$$\langle CAC^T, CBC^T \rangle_{CPC^T} = \langle A, B \rangle_P$$

- Infinite-dimensional: For any invertible $C + \delta I$, $C \in \mathrm{HS}(\mathcal{H})$

$$\langle (C+\delta I)(A+\gamma I)(C+\delta I)^*, (C+\delta I)(B+\nu I)(C+\delta I)^* \rangle_{(C+\delta I)P(C+\delta I)^*}$$
$$= \langle (A+\gamma I), (B+\nu I) \rangle_P$$

Affine-invariant Riemannian metric

Geodesics

- Geodesically complete Riemannian manifold
- Unique geodesic joining $A, B \in \Sigma(\mathcal{H})$

$$\gamma_{AB}(t) = A^{1/2}(A^{-1/2}BA^{-1/2})^t A^{1/2}$$

$$\gamma_{AB}(0) = A, \quad \gamma_{AB}(1) = B$$

- Exponential map $\mathrm{Exp}_P : T_P(\Sigma(\mathcal{H})) \to \Sigma(\mathcal{H})$

$$\mathrm{Exp}_P(V) = P^{1/2}\exp(P^{-1/2}VP^{-1/2})P^{1/2}$$

is defined for all $V \in T_P(\Sigma(\mathcal{H}))$

Affine-invariant Riemannian metric

Riemannian distance

- Finite-dimensional

$$d_{\mathrm{aiE}}(A, B) = ||\log(A^{-1/2}BA^{-1/2})||_F$$

- Infinite-dimensional: Riemannian distance between $(A + \gamma I)$, $(B + \nu I) \in \Sigma(\mathcal{H})$

$$d_{\mathrm{aiHS}}[(A + \gamma I), (B + \nu I)]$$
$$= ||\log[(A + \gamma I)^{-1/2}(B + \nu I)(A + \gamma I)^{-1/2}]||_{\mathrm{eHS}}$$

Affine-invariant Riemannian distance

Proposition (H.Q.M. - GSI 2015)

$$d^2_{\mathrm{aiHS}}[(A+\gamma I),(B+\nu I)] \qquad \dim(\mathcal{H}) = \infty$$

$$= \mathrm{tr}\left\{\log\left[\left(\frac{A}{\gamma}+I\right)^{-1}\left(\frac{B}{\nu}+I\right)\right]\right\}^2 + \left(\log\frac{\gamma}{\nu}\right)^2$$

$$d^2_{\mathrm{aiHS}}[(A+\gamma I),(B+\nu I)] = d^2_{\mathrm{aiE}}[(A+\gamma I),(B+\nu I)] \qquad \dim(\mathcal{H}) < \infty$$

$$= \mathrm{tr}\left\{\log\left[\left(\frac{A}{\gamma}+I\right)^{-1}\left(\frac{B}{\nu}+I\right)\right]\right\}^2 + \left(\log\frac{\gamma}{\nu}\right)^2 \dim(\mathcal{H})$$

$$-2\left(\log\frac{\gamma}{\nu}\right)\mathrm{tr}\left\{\log\left[\left(\frac{A}{\gamma}+I\right)^{-1}\left(\frac{B}{\nu}+I\right)\right]\right\}$$

Affine-invariant Riemannian distance between RKHS covariance operators

For $\gamma > 0, \nu > 0$,

$$d_{\text{aiHS}}[(C_{\Phi(\mathbf{X})} + \gamma I_{\mathcal{H}_K}), (C_{\Phi(\mathbf{Y})} + \nu I_{\mathcal{H}_K})]$$
$$= d_{\text{aiHS}}\left[\left(\frac{1}{m}\Phi(\mathbf{X})J_m\Phi(\mathbf{X})^* + \gamma I_{\mathcal{H}_K}\right), \left(\frac{1}{m}\Phi(\mathbf{X})J_m\Phi(\mathbf{X})^* + \nu I_{\mathcal{H}_K}\right)\right]$$

has a closed form expression via the $m \times m$ Gram matrices

$$K[\mathbf{X}] = \Phi(\mathbf{X})^*\Phi(\mathbf{X}), (K[\mathbf{X}])_{ij} = K(x_i, x_j),$$
$$K[\mathbf{Y}] = \Phi(\mathbf{Y})^*\Phi(\mathbf{Y}), (K[\mathbf{Y}])_{ij} = K(y_i, y_j),$$
$$K[\mathbf{X}, \mathbf{Y}] = \Phi(\mathbf{X})^*\Phi(\mathbf{Y}), (K[\mathbf{X}, \mathbf{Y}])_{ij} = K(x_i, y_j)$$
$$K[\mathbf{Y}, \mathbf{X}] = \Phi(\mathbf{Y})^*\Phi(\mathbf{X}), (K[\mathbf{Y}, \mathbf{x}])_{ij} = K(y_i, x_j)$$

Affine-invariant Riemannian distance between RKHS covariance operators

Theorem (H.Q.M. - GSI 2015)

For $\dim(\mathcal{H}_K) = \infty$, $\gamma > 0, \nu > 0$,

$$d^2_{\text{aiHS}}[(C_{\Phi(\mathbf{X})} + \gamma I_{\mathcal{H}_K}), (C_{\Phi(\mathbf{Y})} + \nu I_{\mathcal{H}_K})]$$

$$= \mathrm{tr}\left\{ \log \left[\begin{pmatrix} C_{11} & C_{12} & C_{13} \\ C_{21} & C_{22} & C_{23} \\ C_{11} & C_{12} & C_{13} \end{pmatrix} + I_{3m} \right] \right\}^2 + \left(\log \frac{\gamma}{\nu} \right)^2$$

Affine-invariant Riemannian distance between RKHS covariance operators

$$C_{11} = \frac{1}{\nu m} J_m K[\mathbf{Y}] J_m, \quad C_{21} = \frac{1}{\sqrt{\gamma\nu} m} J_m K[\mathbf{X}, \mathbf{Y}] J_m$$

$$C_{12} = -\frac{1}{\sqrt{\gamma\nu} m} J_m K[\mathbf{Y}, \mathbf{X}] J_m \left(I_m + \frac{1}{\gamma m} J_m K[\mathbf{X}] J_m \right)^{-1}$$

$$C_{13} = -\frac{1}{\gamma\nu m^2} J_m K[\mathbf{Y}, \mathbf{X}] J_m \left(I_m + \frac{1}{\gamma m} J_m K[\mathbf{X}] J_m \right)^{-1} J_m K[\mathbf{X}, \mathbf{Y}] J_m$$

$$C_{22} = -\frac{1}{\gamma m} J_m K[\mathbf{X}] J_m \left(I_m + \frac{1}{\gamma m} J_m K[\mathbf{X}] J_m \right)^{-1}$$

$$C_{23} = -\frac{1}{\sqrt{\gamma^3 \nu} m^2} J_m K[\mathbf{X}] J_m \left(I_m + \frac{1}{\gamma m} J_m K[\mathbf{X}] J_m \right)^{-1} J_m K[\mathbf{X}, \mathbf{Y}] J_m$$

Affine-invariant Riemannian distance between RKHS covariance operators

Theorem (H.Q.M. - GSI 2015)

For $\dim(\mathcal{H}_K) < \infty, \gamma > 0, \nu > 0$,

$$d^2_{\mathrm{aiHS}}[(C_{\Phi(\mathbf{X})} + \gamma I_{\mathcal{H}_K}), (C_{\Phi(\mathbf{Y})} + \nu I_{\mathcal{H}_K})]$$

$$= \mathrm{tr}\left\{\log\left[\begin{pmatrix} C_{11} & C_{12} & C_{13} \\ C_{21} & C_{22} & C_{23} \\ C_{11} & C_{12} & C_{13} \end{pmatrix} + I_{3m}\right]\right\}^2 + \left(\log\frac{\gamma}{\nu}\right)^2 \dim(\mathcal{H}_K)$$

$$- 2\left(\log\frac{\gamma}{\nu}\right)\mathrm{tr}\left\{\log\left[\begin{pmatrix} C_{11} & C_{12} & C_{13} \\ C_{21} & C_{22} & C_{23} \\ C_{11} & C_{12} & C_{13} \end{pmatrix} + I_{3m}\right]\right\}$$

Affine-invariant Riemannian distance between RKHS covariance operators

Special case

For linear kernel $K(x, y) = \langle x, y \rangle$, $x, y \in \mathbb{R}^n$

$$d_{\mathrm{aiHS}}[(C_{\Phi(\mathbf{X})} + \gamma I_{\mathcal{H}_K}), (C_{\Phi(\mathbf{Y})} + \nu I_{\mathcal{H}_K})] = d_{\mathrm{aiE}}[(C_{\mathbf{X}} + \gamma I_n), (C_{\mathbf{Y}} + \nu I_n)]$$

This can be used to verify the correctness of an implementation

Affine-invariant Riemannian distance between RKHS covariance operators

- For $m \in \mathbb{N}$ fixed, $\gamma \neq \nu$,

$$\lim_{\dim(\mathcal{H}_K) \to \infty} d_{\mathrm{aiHS}}[(C_{\Phi(\mathbf{X})} + \gamma I_{\mathcal{H}_K}), (C_{\Phi(\mathbf{Y})} + \nu I_{\mathcal{H}_K})] = \infty$$

- In general, the infinite-dimensional formulation cannot be approximated by the finite-dimensional counterpart.

Affine-invariant Riemannian distance - Computational complexity

- Requires the multiplications and inversions of $m \times m$ matrices, and the eigenvalue computation for a $3m \times 3m$ matrix. Total computational complexity required is $O(m^3)$.
- For a set of N data matrices $\{\mathbf{X}_j\}_{j=1}^N$, in order to compute all the pairwise distances between the corresponding regularized covariance operators, the total computational complexity is $O(N^2 m^3)$.

Geometry of Covariance Operators

1. Hilbert-Schmidt metric: generalizing the Euclidean metric
2. Manifold of positive definite operators
    - Infinite-dimensional generalization of $\mathrm{Sym}^{++}(n)$
    - Infinite-dimensional affine-invariant Riemannian metric: generalizing finite-dimensional affine-invariant Riemannian metric
    - Log-Hilbert-Schmidt metric: generalizing Log-Euclidean metric
3. Convex cone of positive definite operators
    - Log-Determinant divergences: generalize the finite-dimensional Log-Determinant divergences

Log Hilbert-Schmidt metric

- H.Q.Minh, M. San Biagio, V. Murino. Log-Hilbert-Schmidt metric between positive definite operators on Hilbert spaces, NIPS 2014
- H.Q.Minh, M. San Biagio, L. Bazzani, V. Murino. Approximate Log-Hilbert-Schmidt distances between covariance operators for image classification, CVPR 2016

Log-Hilbert-Schmidt metric

- Generalization from $\mathrm{Sym}^{++}(n)$ to $\Sigma(\mathcal{H})$

$$\odot : \Sigma(\mathcal{H}) \times \Sigma(\mathcal{H}) \to \Sigma(\mathcal{H})$$
$$(A + \gamma I) \odot (B + \nu I) = \exp[\log(A + \gamma I) + \log(B + \nu I)]$$
$$\circledast : \mathbb{R} \times \Sigma(\mathcal{H}) \to \Sigma(\mathcal{H})$$
$$\lambda \circledast (A + \gamma I) = \exp[\lambda \log(A + \gamma I)] = (A + \gamma I)^{\lambda}, \quad \lambda \in \mathbb{R}$$

- $(\Sigma(\mathcal{H}), \odot, \circledast)$ is a vector space
  - $\odot$ acting as vector addition
  - $\circledast$ acting as scalar multiplication

Log-Hilbert-Schmidt metric

- $(\Sigma(\mathcal{H}), \odot, \circledast)$ is a vector space
- Log-Hilbert-Schmidt inner product

$$\langle A + \gamma I, B + \nu I \rangle_{\text{logHS}} = \langle \log(A + \gamma I), \log(B + \nu I) \rangle_{\text{eHS}}$$
$$||A + \gamma I||_{\text{logHS}} = ||\log(A + \gamma I)||_{\text{eHS}}$$

- $(\Sigma(\mathcal{H}), \odot, \circledast, \langle\ ,\ \rangle_{\text{logHS}})$ is a Hilbert space
- Log-Hilbert-Schmidt distance is the Hilbert distance

$$\begin{aligned} d_{\text{logHS}}(A + \gamma I, B + \nu I) &= ||\log(A + \gamma I) - \log(B + \nu I)||_{\text{eHS}} \\ &= ||(A + \gamma I) \odot (B + \nu I)^{-1}||_{\text{logHS}} \end{aligned}$$

Log-Hilbert-Schmidt distance between RKHS covariance operators

The distance

$$d_{\mathrm{logHS}}[(C_{\Phi(\mathbf{X})} + \gamma I_{\mathcal{H}_K}), (C_{\Phi(\mathbf{Y})} + \nu I_{\mathcal{H}_K})]$$
$$= d_{\mathrm{logHS}}\left[\left(\frac{1}{m}\Phi(\mathbf{X})J_m\Phi(\mathbf{X})^* + \gamma I_{\mathcal{H}_K}\right), \left(\frac{1}{m}\Phi(\mathbf{Y})J_m\Phi(\mathbf{Y})^* + \nu I_{\mathcal{H}_K}\right)\right]$$

has a closed form in terms of $m \times m$ Gram matrices

$$K[\mathbf{X}] = \Phi(\mathbf{X})^*\Phi(\mathbf{X}), (K[\mathbf{X}])_{ij} = K(x_i, x_j),$$
$$K[\mathbf{Y}] = \Phi(\mathbf{Y})^*\Phi(\mathbf{Y}), (K[\mathbf{Y}])_{ij} = K(y_i, y_j),$$
$$K[\mathbf{X}, \mathbf{Y}] = \Phi(\mathbf{X})^*\Phi(\mathbf{Y}), (K[\mathbf{X}, \mathbf{Y}])_{ij} = K(x_i, y_j)$$
$$K[\mathbf{Y}, \mathbf{X}] = \Phi(\mathbf{Y})^*\Phi(\mathbf{X}), (K[\mathbf{Y}, \mathbf{x}])_{ij} = K(y_i, x_j)$$

Log-Hilbert-Schmidt distance between RKHS covariance operators

$$\frac{1}{\gamma m} J_m K[\mathbf{X}] J_m = U_A \Sigma_A U_A^T, \quad \frac{1}{\nu m} J_m K[\mathbf{Y}] J_m = U_B \Sigma_B U_B^T,$$

$$A^* B = \frac{1}{\sqrt{\gamma \nu} m} J_m K[\mathbf{X}, \mathbf{Y}] J_m$$

$$C_{AB} = 1_{N_A}^T \log(I_{N_A} + \Sigma_A) \Sigma_A^{-1} (U_A^T A^* B U_B \circ U_A^T A^* B U_B) \Sigma_B^{-1} \log(I_{N_B} + \Sigma_B) 1_{N_B}$$

Log-Hilbert-Schmidt distance between RKHS covariance operators

Theorem (H.Q.M. et al - NIPS2014)

Assume that $\dim(\mathcal{H}_K) = \infty$. *Let* $\gamma > 0$, $\nu > 0$. *The Log-Hilbert-Schmidt distance between* $(C_{\Phi(\mathbf{X})} + \gamma I_{\mathcal{H}_K})$ *and* $(C_{\Phi(\mathbf{Y})} + \nu I_{\mathcal{H}_K})$ *is*

$$\begin{aligned} d^2_{\mathrm{logHS}}[(C_{\Phi(\mathbf{X})} + \gamma I_{\mathcal{H}_K}), (C_{\Phi(\mathbf{Y})} + \nu I_{\mathcal{H}_K})] &= \mathrm{tr}[\log(I_{N_A} + \Sigma_A)]^2 + \mathrm{tr}[\log(I_{N_B} + \Sigma_B)]^2 \\ &\quad - 2C_{AB} + (\log\gamma - \log\nu)^2 \end{aligned}$$

The Log-Hilbert-Schmidt inner product between $(C_{\Phi(\mathbf{X})} + \gamma I_{\mathcal{H}_K})$ *and* $(C_{\Phi(\mathbf{Y})} + \nu I_{\mathcal{H}_K})$ *is*

$$\langle (C_{\Phi(\mathbf{X})} + \gamma I_{\mathcal{H}_K}), (C_{\Phi(\mathbf{Y})} + \nu I_{\mathcal{H}_K}) \rangle_{\mathrm{logHS}} = C_{AB} + (\log\gamma)(\log\nu)$$

Log-Hilbert-Schmidt distance between RKHS covariance operators

Theorem (H.Q.M. et al - NIPS2014)

Assume that $\dim(\mathcal{H}_K) = \infty$. *Let* $\gamma > 0$.
The Log-Hilbert-Schmidt norm of the operator $(C_{\Phi(\mathbf{x})} + \gamma I_{\mathcal{H}_K})$ *is*

$$||(C_{\Phi(\mathbf{x})} + \gamma I_{\mathcal{H}_K})||^2_{\mathrm{logHS}} = \mathrm{tr}[\log(I_{N_A} + \Sigma_A)]^2 + (\log\gamma)^2$$

Log-Hilbert-Schmidt distance between RKHS covariance operators

Theorem (H.Q.M. et al - NIPS2014)

Assume that $\dim(\mathcal{H}_K) < \infty$. *Let* $\gamma > 0$, $\nu > 0$. *The Log-Hilbert-Schmidt distance between* $(C_{\Phi(\mathbf{X})} + \gamma I_{\mathcal{H}_K})$ *and* $(C_{\Phi(\mathbf{Y})} + \nu I_{\mathcal{H}_K})$ *is*

$$
\begin{aligned}
& d^2_{\mathrm{logHS}}[(C_{\Phi(\mathbf{X})} + \gamma I_{\mathcal{H}_K}), (C_{\Phi(\mathbf{Y})} + \nu I_{\mathcal{H}_K})] \\
&= \mathrm{tr}[\log(I_{N_A} + \Sigma_A)]^2 + \mathrm{tr}[\log(I_{N_B} + \Sigma_B)]^2 - 2C_{AB} \\
&+ 2(\log \frac{\gamma}{\nu})(\mathrm{tr}[\log(I_{N_A} + \Sigma_A)] - \mathrm{tr}[\log(I_{N_B} + \Sigma_B)]) \\
&+ (\log \gamma - \log \nu)^2 \dim(\mathcal{H}_K)
\end{aligned}
$$

Log-Hilbert-Schmidt distance between RKHS covariance operators

Theorem (H.Q.M. et al - NIPS2014)

Assume that $\dim(\mathcal{H}_K) < \infty$. *Let* $\gamma > 0$, $\nu > 0$. *The Log-Hilbert-Schmidt inner product between* $(C_{\Phi(\mathbf{X})} + \gamma I_{\mathcal{H}_K})$ *and* $(C_{\Phi(\mathbf{Y})} + \nu I_{\mathcal{H}_K})$ *is*

$$
\begin{aligned}
&\langle (C_{\Phi(\mathbf{X})} + \gamma I_{\mathcal{H}_K}), (C_{\Phi(\mathbf{Y})} + \nu I_{\mathcal{H}_K}) \rangle_{\log\mathrm{HS}} \\
&= C_{AB} + (\log \nu)\mathrm{tr}[\log(I_{N_A} + \Sigma_A)] \\
&+ (\log \gamma)\mathrm{tr}[\log(I_{N_B} + \Sigma_B)] + (\log \gamma \log \nu)\dim(\mathcal{H}_K)
\end{aligned}
$$

The Log-Hilbert-Schmidt norm of $(C_{\Phi(\mathbf{X})} + \gamma I_{\mathcal{H}_K})$ *is*

$$
\begin{aligned}
||(C_{\Phi(\mathbf{X})} + \gamma I_{\mathcal{H}_K})||^2_{\log\mathrm{HS}} &= \mathrm{tr}[\log(I_{N_A} + \Sigma_A)]^2 + 2(\log \gamma)\mathrm{tr}[\log(I_{N_A} + \Sigma_A)] \\
&+ (\log \gamma)^2 \dim(\mathcal{H}_K)
\end{aligned}
$$

Log-Hilbert-Schmidt distance between RKHS covariance operators

Special case

For linear kernel $K(x,y) = \langle x, y\rangle$, $x, y \in \mathbb{R}^n$

$$d_{\mathrm{logHS}}[(C_{\Phi(\mathbf{X})} + \gamma I_{\mathcal{H}_K}), (C_{\Phi(\mathbf{Y})} + \nu I_{\mathcal{H}_K})] = d_{\mathrm{logE}}[(C_{\mathbf{X}} + \gamma I_n), (C_{\mathbf{Y}} + \nu I_n)]$$

$$\langle (C_{\Phi(\mathbf{X})} + \gamma I_{\mathcal{H}_K}), (C_{\Phi(\mathbf{Y})} + \nu I_{\mathcal{H}_K})\rangle_{\mathrm{logHS}} = \langle (C_{\mathbf{X}} + \gamma I_n), (C_{\mathbf{Y}} + \nu I_n)\rangle_{\mathrm{logE}}$$

$$||(C_{\mathbf{X}} + \gamma I_{\mathcal{H}_K})||_{\mathrm{logHS}} = ||(C_{\mathbf{X}} + \gamma I_n)||_{\mathrm{logE}}$$

These can be used to verify the correctness of an implementation

Log-Hilbert-Schmidt distance between RKHS covariance operators

- For $m \in \mathbb{N}$ fixed, $\gamma \neq \nu$,

$$\lim_{\dim(\mathcal{H}_K)\to\infty} d_{\mathrm{logHS}}[(C_{\Phi(\mathbf{X})} + \gamma I_{\mathcal{H}_K}), (C_{\Phi(\mathbf{Y})} + \nu I_{\mathcal{H}_K})] = \infty$$

- In general, the infinite-dimensional formulation cannot be approximated by the finite-dimensional counterpart.

Log-Hilbert-Schmidt distance - Computational complexity

- Requires the SVD and multiplications of Gram matrices of size $m \times m$. The computational complexity required is $O(m^3)$.
- For a set of N data matrices $\{\mathbf{X}_j\}_{j=1}^N$, then the computational complexity required for computing all pairwise distances between the corresponding regularized covariance operators is $O(N^2 m^3)$.

Approximate methods for reducing computational costs

- M. Faraki, M. Harandi, and F. Porikli, Approximate infinite-dimensional region covariance descriptors for image classification, ICASSP 2015
- H.Q. Minh, M. San Biagio, L. Bazzani, V. Murino. Approximate Log-Hilbert-Schmidt distances between covariance operators for image classification, CVPR 2016
- Q. Wang, P. Li, W. Zuo, and L. Zhang. RAID-G: Robust estimation of approximate infinite-dimensional Gaussian with application to material recognition, CVPR 2016

Geometry of Covariance Operators

1. Hilbert-Schmidt metric: generalizing the Euclidean metric
2. Manifold of positive definite operators
   - Infinite-dimensional generalization of $\mathrm{Sym}^{++}(n)$
   - Infinite-dimensional affine-invariant Riemannian metric: generalizing finite-dimensional affine-invariant Riemannian metric
   - Log-Hilbert-Schmidt metric: generalizing Log-Euclidean metric
3. Convex cone of positive definite operators
   - Log-Determinant divergences: generalize the finite-dimensional Log-Determinant divergences

Log-Determinant divergences

- Zhou and Chellappa (PAMI 2006), Harandi et al (CVPR 214): finite-dimensional RKHS
- H.Q.M. Infinite-dimensional Log-Determinant divergences between positive definite trace class operators, *Linear Algebra and its Applications*, 2017
- H.Q.M. Log-Determinant divergences between positive definite Hilbert-Schmidt operators, *Geometric Science of Information*, 2017

Summary - Geometry of Covariance Operators

- Hilbert-Schmidt distance and inner product
- Affine-invariant Riemannian distance
- Log-Hilbert-Schmidt distance and inner product
- Explicit formulas via Gram matrices in the case of RKHS covariance operators

Infinite-dimensional setting

Covariance Operators and Applications

1. Data Representation by Covariance Operators
2. Geometry of Covariance Operators
3. Machine Learning Methods on Covariance Operators and Applications in Computer Vision

Kernels with Log-Hilbert-Schmidt metric

$(\Sigma(\mathcal{H}), \odot, \circledast, \langle\ ,\ \rangle_{\mathrm{logHS}})$ is a Hilbert space

Theorem (H.Q.M. et al - NIPS 2014)

The following kernels $K : \Sigma(\mathcal{H}) \times \Sigma(\mathcal{H}) \to \mathbb{R}$ *are positive definite*

$$K[(A+\gamma I), (B+\nu I)] = (c + \langle A+\gamma I, B+\nu I\rangle_{\mathrm{logHS}})^d$$

$$c \geq 0, d \in \mathbb{N}$$

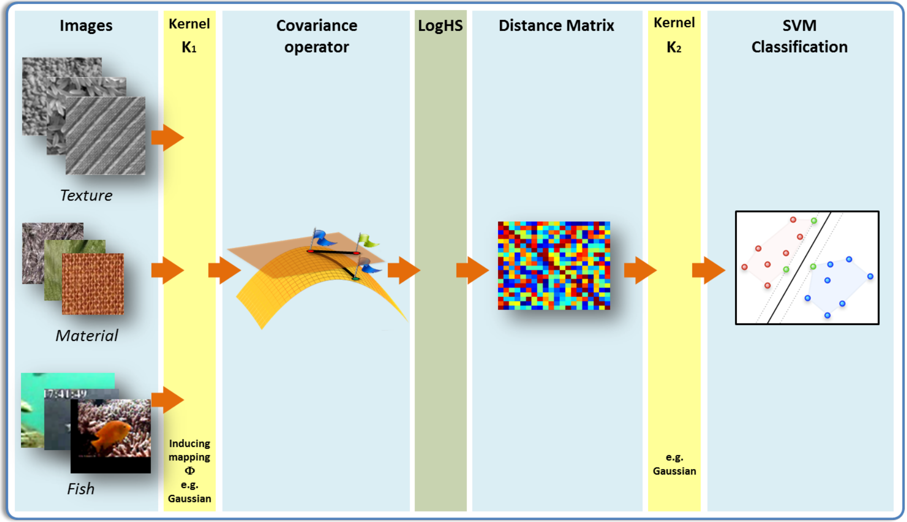
$$K[(A+\gamma I), (B+\nu I)] = \exp(-\frac{1}{\sigma^2}||\log(A+\gamma I) - \log(B+\nu I)||^p_{\mathrm{eHS}})$$

$$0 < p \leq 2, \sigma \neq 0$$

Two-layer kernel machine with Log-Hilbert-Schmidt metric

1. First layer: kernel K_1, inducing covariance operators
2. Second layer: kernel K_2, defined using the Log-Hilbert-Schmidt distance or inner product between the covariance operators

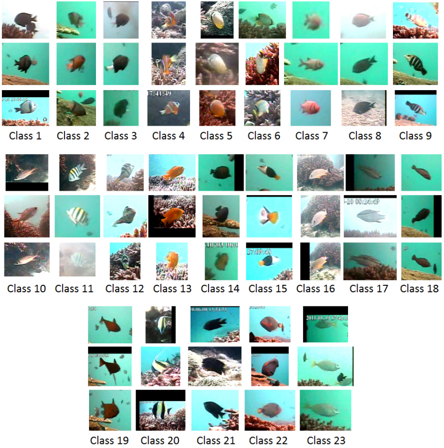
Two-layer kernel machine with Log-Hilbert-Schmidt metric



Experiments in Image Classification

- Multi-class image classification using multi-class SVM with the Gaussian kernel
- Each image is represented by one covariance operator
- Fish recognition
- Material classification

Fish recognition



(Boom et al, *Ecological Informatics*, 2014)

Fish recognition

Using R,G,B channels

Method	Accuracy
E	26.9% ($\pm$3.5%)
Stein	43.9% ($\pm$3.5%)
Log-E	42.7% ($\pm$3.4%)
HS	50.2% ($\pm$2.2%)
Log-HS	56.7% ($\pm$2.9%)

Material classification

Example: KTH-TIPS2b data set (Caputo et al, *ICCV*, 2005)

$$\mathbf{f}(x,y) = \left[R(x,y), G(x,y), B(x,y), \left|G^{0,0}(x,y)\right|, \ldots \left|G^{3,4}(x,y)\right|\right]$$

Material classification

Method	KTH-TIPS2b
E	55.3% (±7.6%)
Stein	73.1% (±8.0%)
Log-E	74.1 % (±7.4%)
HS	79.3% (±8.2%)
Log-HS	81.9% (±3.3%)
Log-HS (CNN)	96.6% (±3.4%)

CNN features = MatConvNet features

Object recognition

Example: ETH-80 data set

$$\mathbf{f}(x, y) = [x, y, I(x, y), |I_x|, |I_y|]$$

Object recognition

Results obtained using approximate Log-HS distance

Method	ETH-80
E	64.4%(±0.9%)
Stein	67.5% (±0.4%)
Log-E	71.1%(±1.0%)
HS	93.1 % (±0.4)
Approx-LogHS	95.0% (±0.5%)

Summary: Kernel methods with covariance operators

- Two-layer kernel machine with Log-Hilbert-Schmidt metric
- Substantial gains in performance compared to finite-dimensional case, but higher computational cost
- Approximate methods for reducing computational costs
  - H.Q. Minh, M. San Biagio, L. Bazzani, V. Murino. Approximate Log-Hilbert-Schmidt distances between covariance operators for image classification, CVPR 2016

Summary: Infinite-dimensional setting

1. Data Representation by Covariance Operators
2. Geometry of Covariance Operators
   - Hilbert-Schmidt distance and inner product
   - Affine-invariant Riemannian distance and inner product
   - Log-Hilbert-Schmidt distance and inner product
3. Machine Learning Methods on Covariance Operators and Applications in Computer Vision
   - Two-layer kernel machine with Log-Hilbert-Schmidt metric
   - Experiments in image classification

Summary: Infinite-dimensional setting

- Significantly different from finite-dimensional case
- The classes of operators, i.e. positive definite operators, extended Hilbert-Schmidt operators, must be defined carefully
- Log-Euclidean metric, affine-invariant Riemannian metric, Bregman divergences can all be generalized to infinite-dimensional setting
- Can obtain substantial gains in performance compared to finite-dimensional case, but with higher computational costs

Summary: Infinite-dimensional setting

- Finite-dimensional approximation methods can be applied under certain conditions, but not in general
- Still undergoing active development, both theoretically and computationally

Conclusion

1. Overview of the finite-dimensional setting
   - Covariance matrix representation
   - Geometry of SPD matrices
   - Kernel methods with covariance matrices
2. Generalization to the infinite-dimensional setting via kernel methods
   - Covariance operator representation
   - Geometry of covariance operators
   - Kernel methods with covariance operators

Some future directions

- Applications of covariance operators
  - Computer vision
  - Image and signal processing
  - Machine learning and statistics
  - Other fields
- Covariance matrices and covariance operators in deep learning
  - C. Ionescu, O. Vantzos, and C. Sminchisescu. Matrix backpropagation for deep networks with structured layers, ICCV 2015
  - Q. Wang, P. Li, W. Zuo, and L. Zhang. RAID-G: Robust estimation of approximate infinite-dimensional Gaussian with application to material recognition, CVPR 2016

Some future directions

- Connection between the geometry of positive definite operators and infinite-dimensional Gaussian probability measures
  - Finite-dimensional setting

$$\text{Affine-invariant Riemannian metric} \Longleftrightarrow \text{Fisher-Rao metric}$$
$$\text{Log-Determinant divergences} \Longleftrightarrow \text{Rény divergences}$$

  - Infinite-dimensional setting: upcoming work

References

- M. Faraki, M. Harandi, and F. Porikli, Approximate infinite-dimensional region covariance descriptors for image classification, ICASSP 2015
- M. Harandi, M. Salzmann, and F. Porikli. Bregman divergences for infinite-dimensional covariance matrices, CVPR 2014
- Larotonda. Nonpositive curvature: A geometrical approach to Hilbert-Schmidt operators, *Differential Geometry and Its Applications*, 2007
- H.Q.Minh. Affine-invariant Riemannian distance between infinite-dimensional covariance operators, *Geometric Science of Information*, 2015
- H.Q.Minh. Infinite-dimensional Log-Determinant divergences between positive definite trace class operators, *Linear Algebra and its Applications*, 2017

References

- H.Q.Minh. Log-Determinant divergences between positive definite Hilbert-Schmidt operators, *Geometric Science of Information*, 2017
- H.Q.Minh, M. San Biagio, V. Murino. Log-Hilbert-Schmidt metric between positive definite operators on Hilbert spaces, NIPS 2014
- H.Q.Minh, M. San Biagio, L. Bazzani, V. Murino. Approximate Log-Hilbert-Schmidt distances between covariance operators for image classification, CVPR 2016
- Q. Wang, P. Li, W. Zuo, and L. Zhang. RAID-G: Robust estimation of approximate infinite-dimensional Gaussian with application to material recognition, CVPR 2016
- S.K. Zhou and R. Chellappa. From sample similarity to ensemble similarity: Probabilistic distance measures in reproducing kernel Hilbert space, PAMI 2006

Thank you for listening!
Questions, comments, suggestions?