# Covariance Matrices and Covariance Operators Theory and Applications

Hà Quang Minh

Functional Analytic Learning Unit
RIKEN Center for Advanced Intelligence Project (AIP), Tokyo

February 2019

# Main Research Directions

1. Vector-valued Reproducing Kernel Hilbert Spaces (RKHS) and Applications
2. Geometrical methods in Machine Learning and Applications

# Geometrical methods in Machine Learning

- Exploit the geometrical structures of data
- Current theoretical focus: Infinite-dimensional generalizations of the geometrical structures of the set of Symmetric Positive Definite (SPD) matrices
- Current computational focus: Geometry of RKHS covariance operators
- Current practical application focus: Image representation by covariance matrices and covariance operators

# Covariance Matrices and Covariance Operators

## Motivations

- Covariance matrices: many applications in computer vision, brain imaging, radar signal processing etc
  - Powerful approach for data representation by encoding input correlations
  - Rich mathematical theories and computational algorithms
  - Very good practical performances
- Covariance operators (infinite-dimensional setting):
  - Nonlinear generalization of covariance matrices
  - Can be much more powerful as a form of data representation
  - Can achieve substantial gains in practical performances

## Symmetric Positive Definite (SPD) matrices

$\mathrm{Sym}^{++}(n)$ = set of $n \times n$ SPD matrices

- Have been studied extensively mathematically
- Numerous practical applications
  - Brain imaging (Arsigny et al 2005, Dryden et al 2009, Qiu et al 2015)
  - Computer vision: object detection (Tuzel et al 2008, Tosato et al 2013), image retrieval (Cherian et al 2013), visual recognition (Jayasumana et al 2015), many more
  - Radar signal processing: Barbaresco (2013), Formont et al 2013
  - Machine learning: kernel learning (Kulis et al 2009)

# Example: Covariance matrix representation of images

- Tuzel, Porikli, Meer (ECCV 2006, CVPR 2006): covariance matrices as region descriptors for images (covariance descriptors)
- 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 $\mathbf{X}$

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

where

- $m$ = number of pixels
- $n$ = number of features at each pixel

# Example: Covariance matrix representation of images

$\mathbf{X} = [x_1, \ldots, x_m]$ = data matrix of size $n \times m$, with $m$ observations

- Empirical mean vector

$$\mu_{\mathbf{X}} = \frac{1}{m}\sum_{i=1}^{m} x_i = \frac{1}{m}\mathbf{X}\mathbf{1}_m, \quad \mathbf{1}_m = (1, \ldots, 1)^T \in \mathbb{R}^m$$

- Empirical covariance matrix

$$C_{\mathbf{X}} = \frac{1}{m}\sum_{i=1}^{m}(x_i - \mu_{\mathbf{X}})(x_i - \mu_{\mathbf{X}})^T = \frac{1}{m}\mathbf{X}J_m\mathbf{X}^T$$

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

# Example: Covariance matrix representation of images

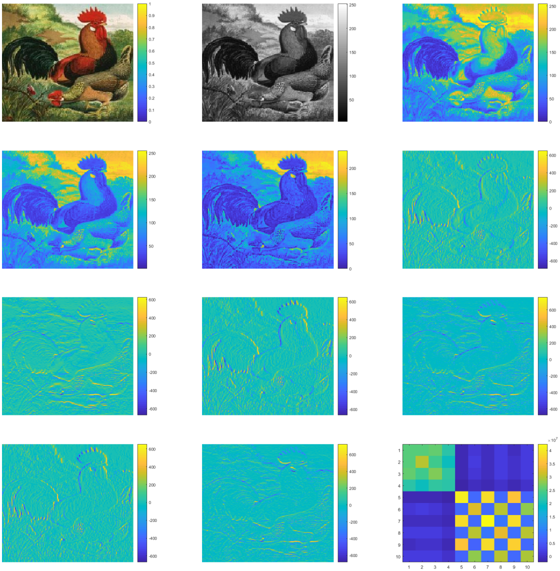
Image $F$ $\Rightarrow$ Data matrix $\mathbf{X}$ $\Rightarrow$ Covariance matrix $C_{\mathbf{X}}$

- Each image is represented by a covariance matrix
- Example of image features

$$\mathbf{f}(x,y) = \left[I(x,y), R(x,y), G(x,y), B(x,y), \left|\frac{\partial R}{\partial x}\right|, \left|\frac{\partial R}{\partial y}\right|, \left|\frac{\partial G}{\partial x}\right|, \left|\frac{\partial G}{\partial y}\right|, \left|\frac{\partial B}{\partial x}\right|, \left|\frac{\partial B}{\partial y}\right|\right]$$

at pixel location $(x,y)$

# Example



# Covariance matrix representation - Properties

- Encode linear correlations (second order statistics) between image features
- Flexible, allowing the fusion of multiple and different features
  - Handcrafted features, e.g. colors and SIFT
  - Convolutional features
- Compact
- Robust to noise

# Covariance matrix representation - generalization

- Covariance representation for video: e.g. Guo et al (AVSS 2010), Sanin et al (WACV 2013)
  - Employ features that capture temporal information, e.g. optical flow
- Covariance representation for 3D point clouds and 3D shapes: e.g. Fehr et al (ICRA 2012, ICRA 2014), Tabias et al (CVPR 2014), Hariri et al (Pattern Recognition Letters 2016)
  - Employ geometric features e.g. curvature, surface normal vectors

# Statistical interpretation

Representing an image by a covariance matrix

is essentially equivalent to

Representing an image by a Gaussian probability density $\rho$ in $\mathbb{R}^n$ with mean zero

Features extracted are random observations of a $n$-dimensional random vector with probability density $\rho$

# Geometry of SPD Matrices

$A, B \in \mathrm{Sym}^{++}(n)$ = set of $n \times n$ SPD matrices

- Euclidean distance $d_E(A,B) = ||A - B||_F$
- Riemannian manifold viewpoint
  - Affine-invariant Riemannian distance (e.g. Pennec et al 2006, Bhatia 2007)

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

  - Log-Euclidean distance (Arsigny et al 2007)

$$d_{\mathrm{logE}}(A,B) = ||\log(A) - \log(B)||_F$$

- Optimal transport viewpoint Bures-Wasserstein-Fréchet distance (Dowson and Landau 1982, Olkin and Pukelsheim 1982, Givens and Shortt 1984, Gelbrich 1990)

$$d_{\mathrm{BW}}(A,B) = \left(\mathrm{tr}[A + B - 2(A^{1/2}BA^{1/2})]\right)^{1/2}$$

Affine-Invariant Metric

- Close connection with Fisher-Rao metric in information geometry (e.g. Amari 1985)
- For two multivariate Gaussian probability densities $\rho_1 \sim \mathcal{N}(\mu, C_1)$, $\rho_2 \sim \mathcal{N}(\mu, C_2)$

$$d_{\mathrm{aiE}}(C_1, C_2) = 2(\text{Fisher-Rao distance between } \rho_1 \text{ and } \rho_2\ )$$

## Bures-Wasserstein Distance

- $\mu_X \sim \mathcal{N}(m_1, A)$ and $\mu_Y \sim \mathcal{N}(m_2, B)$ = Gaussian probability distributions on $\mathbb{R}^n$
- $\mathcal{L}^2$-Wasserstein distance between $\mu_X$ and $\mu_Y$

$$
\begin{aligned}
d^2_{\mathrm{W}}(\mu_X, \mu_Y) &= \inf_{\mu \in \Gamma(\mu_X, \mu_Y)} \int_{\mathbb{R}^n \times \mathbb{R}^n} ||x - y||^2 d\mu(x, y) \\
&= ||m_1 - m_2||^2 + \mathrm{tr}[A + B - 2(A^{1/2} B A^{1/2})^{1/2}]
\end{aligned}
$$

# Geometry of SPD Matrices

## Convex cone viewpoint

- Alpha Log-Determinant divergences (Chebbi and Moakher, 2012)

$$d^{\alpha}_{\mathrm{logdet}}(A,B) = \frac{4}{1-\alpha^2}\log\frac{\det(\frac{1-\alpha}{2}A + \frac{1+\alpha}{2}B)}{\det(A)^{\frac{1-\alpha}{2}}\det(B)^{\frac{1+\alpha}{2}}}, \quad -1 < \alpha < 1$$

- Limiting cases

$$d^{1}_{\mathrm{logdet}}(A,B) = \lim_{\alpha\to 1} d^{\alpha}_{\mathrm{logdet}}(A,B) = \mathrm{tr}(B^{-1}A - I) - \log\det(B^{-1}A)$$

$$d^{-1}_{\mathrm{logdet}}(A,B) = \lim_{\alpha\to -1} d^{\alpha}_{\mathrm{logdet}}(A,B) = \mathrm{tr}(A^{-1}B - I) - \log\det(A^{-1}B)$$

- Are generally not metrics

# Alpha Log-Determinant divergences

- $\alpha = 0$: Symmetric Stein divergence (also called $S$-divergence)

$$d^0_{\mathrm{logdet}}(A, B) = 4\left[\log\left(\frac{A+B}{2}\right) - \frac{1}{2}\log\det(AB)\right] = 4d^2_{\mathrm{stein}}(A, B)$$

- Sra (NIPS 2012):

$$d_{\mathrm{stein}}(A, B) = \sqrt{\log\left(\frac{A+B}{2}\right) - \frac{1}{2}\log\det(AB)}$$

is a metric (satisfying positivity, symmetry, and triangle inequality)

# Statistical Interpretation

## Alpha Log-Determinant Divergences

- Close connection with Kullback-Leibler and Rényi divergences
- For two multivariate Gaussian probability densities $\rho_1 \sim \mathcal{N}(\mu, C_1)$, $\rho_2 \sim \mathcal{N}(\mu, C_2)$

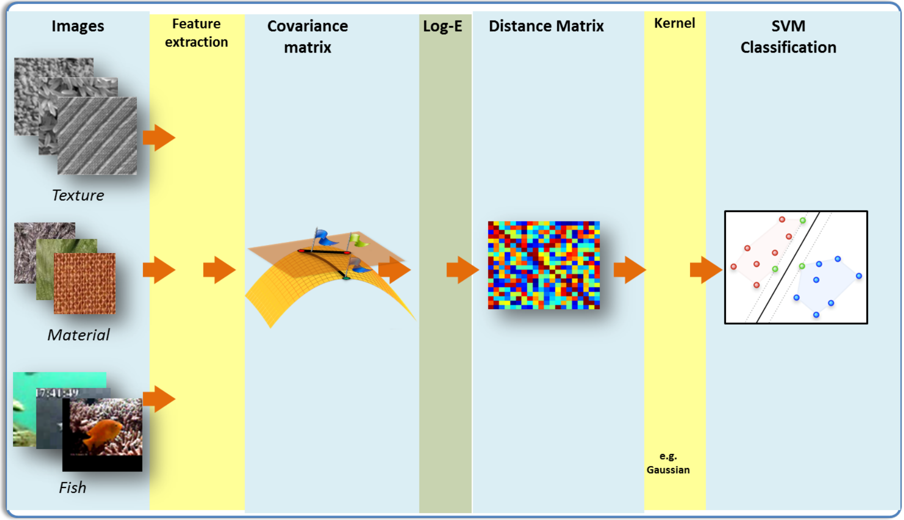
$$d^{\alpha}_{\mathrm{logdet}}(C_1, C_2) = \textit{constant}(\text{a Rényi divergence between } \rho_1 \text{ and } \rho_2)$$

$$d^{1}_{\mathrm{logdet}}(C_1, C_2) = 2(\text{Kullback-Leibler divergence between } \rho_1 \text{ and } \rho_2)$$

# Kernel methods with Log-Euclidean metric

- S. Jayasumana, R. Hartley, M. Salzmann, H. Li, and M. Harandi. Kernel methods on the Riemannian manifold of symmetric positive definite matrices. CVPR 2013.
- S. Jayasumana, R. Hartley, M. Salzmann, H. Li, and M. Harandi. Kernel methods on Riemannian manifolds with Gaussian RBF kernels, PAMI 2015.
- P. Li, Q. Wang, W. Zuo, and L. Zhang. Log-Euclidean kernels for sparse representation and dictionary learning, ICCV 2013
- D. Tosato, M. Spera, M. Cristani, and V. Murino. Characterizing humans on Riemannian manifolds, PAMI 2013

# Kernel methods with Log-Euclidean metric for image classification



# Material classification

Example: KTH-TIPS2b data set

Cork Wool Lettuce leaf Aluminium foil Corduroy Linen Cotton Brown bread White bread Wood Cracker

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

# Object recognition

Example: ETH-80 data set

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

# Numerical results

Better results with covariance operators (later)!

| Method | KTH-TIPS2b | ETH-80 |
| --- | --- | --- |
| *E* | 55.3% ($\pm$7.6%) | 64.4% ($\pm$0.9%) |
| *Stein* | 73.1% ($\pm$8.0%) | 67.5% ($\pm$0.4%) |
| *Log-E* | 74.1 % ($\pm$7.4%) | 71.1% ($\pm$1.0%) |

# Comparison of metrics

Results from Cherian et al (PAMI 2013) using Nearest Neighbor

| Method | Texture | Activity |
|---|---|---|
| *Affine-invariant* | 85.5% | 99.5% |
| *Stein* | 85.5% | 99.5% |
| *Log-E* | 82.0% | 96.5% |

Texture: images from Brodatz and CURET datasets
Activity: videos from Weizmann, KTH, and UT Tower datasets

# 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

# From covariance matrices

$\mathbf{X} = [x_1, \ldots, x_m]$ = data matrix with $m$ observations, sampled according to some probability distribution $\rho$ on the input space $\mathcal{X} = \mathbb{R}^n$

- Empirical mean vector

$$\mu_{\mathbf{X}} = \frac{1}{m}\sum_{i=1}^{m} x_i = \frac{1}{m}\mathbf{X}\mathbf{1}_m, \quad \mathbf{1}_m = (1, \ldots, 1)^T \in \mathbb{R}^m$$

- Empirical covariance matrix

$$C_{\mathbf{X}} = \frac{1}{m}\sum_{i=1}^{m}(x_i - \mu_{\mathbf{X}})(x_i - \mu_{\mathbf{X}})^T = \frac{1}{m}\mathbf{X}J_m\mathbf{X}^T$$

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

# To RKHS covariance operators

- $\mathbf{X} = [x_1, \ldots, x_m]$ = data matrix randomly sampled according to $\rho$ on the input space $\mathcal{X}$, with $m$ observations
- Positive definite kernel $K$, RKHS $\mathcal{H}_K$, feature map $\Phi : \mathcal{X} \to \mathcal{H}_K$
- Informally, $\Phi$ 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

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

# RKHS covariance operators

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

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

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

# RKHS covariance operators

- Theoretical mean

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

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

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

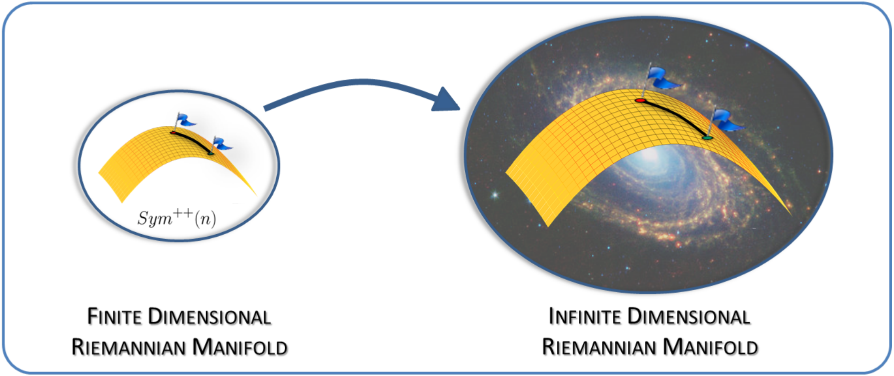
# Geometry of Covariance Operators

- H.Q. Minh et al. Log-Hilbert-Schmidt metric between positive definite operators on Hilbert spaces, *NIPS 2014*
  - Infinite-dimensional generalization of the Log-Euclidean Riemannian metric on the manifold of SPD matrices
  - Closed form formulas in the case of RKHS covariance operators
- H.Q. Minh. Affine-invariant Riemannian distance between infinite-dimensional covariance operators, *Geometric Science of Information 2015*
- H.Q.Minh, M. San Biagio, L. Bazzani, V. Murino. Approximate Log-Hilbert-Schmidt Distances between Covariance Operators for Image Classification, *CVPR 2016*

# Geometry of Covariance Operators

- H.Q. Minh. Infinite-dimensional Log-Determinant divergences between positive definite trace class operators, *Linear Algebra and its Applications 2017*
  - Infinite-dimensional generalization of the Alpha Log-Determinant divergences on the convex cone of SPD matrices
  - Closed form formulas in the case of RKHS covariance operators
- H.Q. Minh. Infinite-Dimensional Log-Determinant Divergences II: Alpha-Beta divergences, under review *Information Geometry* `https://arxiv.org/abs/1610.08087`
- H.Q. Minh. Log-Determinant divergences between positive definite Hilbert-Schmidt operators, *Geometric Science of Information 2017*
- H.Q. Minh. Infinite-Dimensional Log-Determinant Divergences III: Log-Euclidean and Log-Hilbert-Schmidt divergences, *Information Geometry and Its Applications 2018*

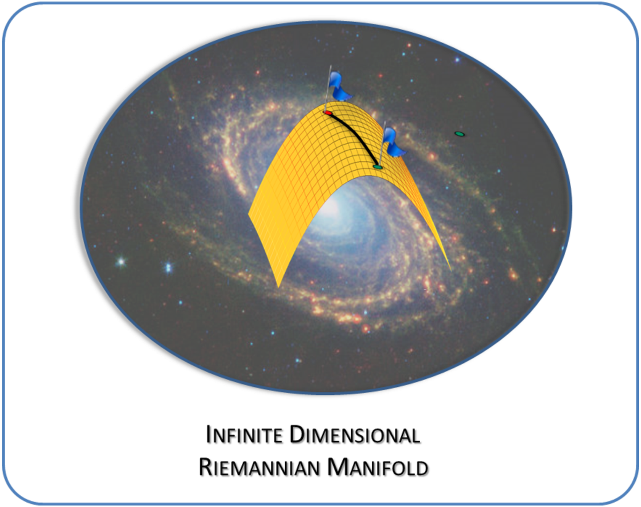
# From finite to infinite-dimensional settings



# Infinite-dimensional generalizations

- Substantially different from the finite-dimensional formulations
- Problems: $A$ = strictly positive, self-adjoint compact operator (e.g. covariance operator)
  1. Eigenvalues $\lambda_k \to 0$ as $k \to \infty$
  2. $\frac{1}{\lambda_k} \to \infty$ and $\log(\lambda_k) \to -\infty$
  3. $A^{-1}$ is unbounded
  4. $\log(A)$ is unbounded
  5. $\det(A)$ is always zero

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



# Geometry of positive definite operators

- Larotonda (*Differential Geometry and Its Applications* 2007): generalization of the manifold $\mathrm{Sym}^{++}(n)$ of SPD matrices to the infinite-dimensional Hilbert manifold

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

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

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

  for any orthonormal basis $\{e_k\}_{k=1}^{\infty}$

- $A$ self-adjoint $||A||^2_{\mathrm{HS}} = \sum_{k=1}^{\infty} \lambda_k^2$
- Generalization of the affine-invariant Riemannian metric

Generalizing Log-Euclidean distance $d_{\mathrm{logE}}(A, B) = ||\log(A) - \log(B)||$

- Log-Hilbert-Schmidt distance

$$d_{\mathrm{logHS}}[(A + \gamma I), (B + \nu I)] = ||\log(A + \gamma I) - \log(B + \nu I)||_{\mathrm{eHS}}$$

- Extended Hilbert-Schmidt norm

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

- Extended Hilbert-Schmidt inner product

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

# Log-Hilbert-Schmidt distance

Why $\log(A+\gamma I)$? Why extended Hilbert-Schmidt norm?

- $A \in \mathrm{Sym}^{++}(n)$, 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 $\{\mathbf{u}_k\}_{k=1}^\infty$

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

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

# Log-Hilbert-Schmidt distance

Why $\log(A+\gamma I)$? Why extended Hilbert-Schmidt norm?

- $\log(A)$ is unbounded
- $\log(A+\gamma I)$ is bounded
- Hilbert-Schmidt norm

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

- The extended Hilbert-Schmidt norm

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

$$= \sum_{j=1}^{\infty}[\log(\frac{\lambda_k}{\gamma}+1)]^2 + (\log\gamma)^2 < \infty$$

# Log-Hilbert-Schmidt distance between RKHS covariance operators

The distance

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

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

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

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

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

# Example: Log-Hilbert-Schmidt distance between RKHS covariance operators

Closed form expression

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

*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*

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

# Log-Hilbert-Schmidt distance between RKHS covariance operators

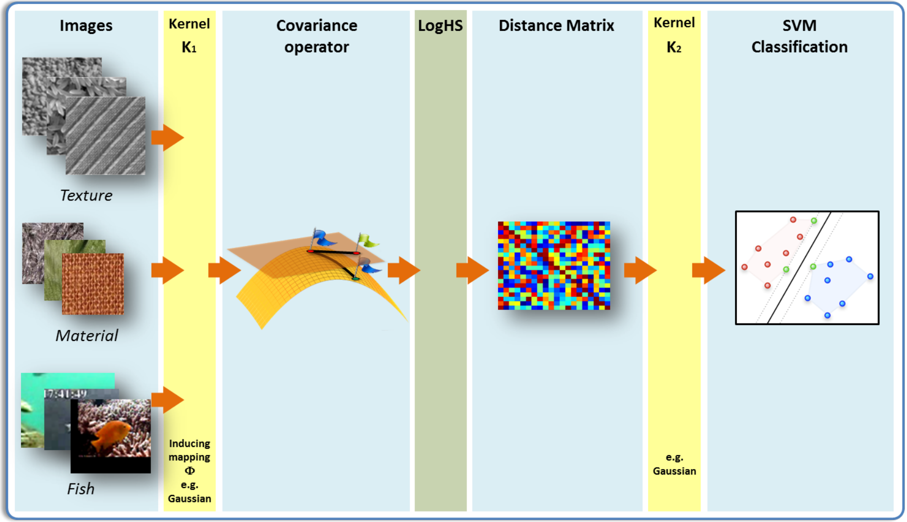
Closed form expression

**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*

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

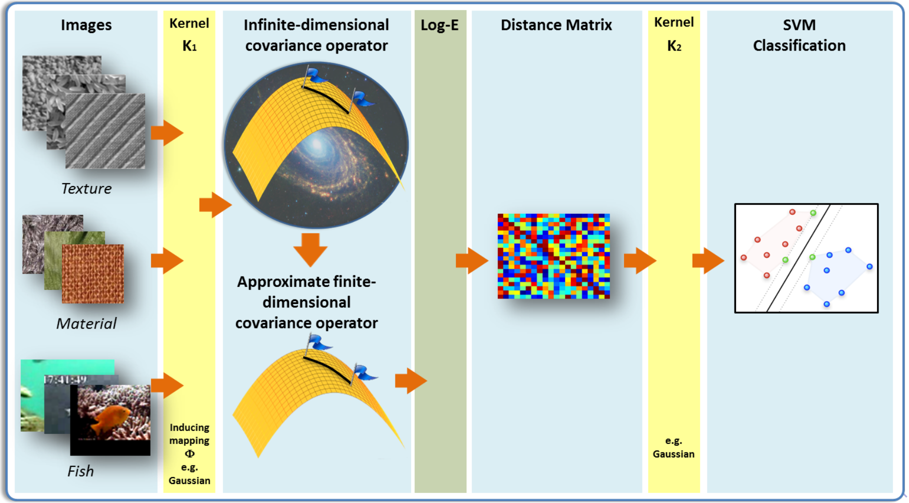
# Example: Two-layer kernel machine for image classification (H.Q.Minh et al - NIPS 2014)



# Approximate methods for reducing computational complexity

- 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

# Two-layer kernel machine with the approximate Log-Hilbert-Schmidt distance (H.Q.M et al CVPR 2016)



# Example: Object recognition

Example: ETH-80 data set

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

# Example: Object recognition

Results obtained using approximate Log-HS distance (H.Q.M et al, CVPR 2016)

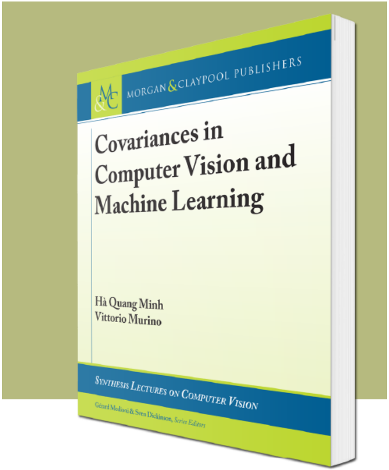
| Method | ETH-80 |
|---|---|
| *Euclidean* | 64.4%($\pm$0.9%) |
| *Stein* | 67.5% ($\pm$0.4%) |
| *Log-Euclidean* | 71.1%($\pm$1.0%) |
| *HS* | 93.1 % ($\pm$0.4) |
| Approx-LogHS | 95.0% ($\pm$0.5%) |

# Further detail

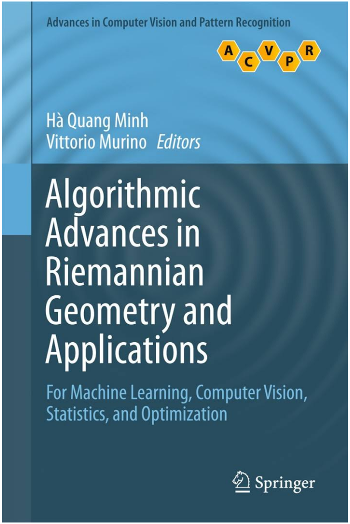
- H.Q. Minh and V. Murino. *Covariances in Computer Vision and Machine Learning*, Morgan & Claypool Publishers, 2017
- H.Q. Minh and V. Murino. From Covariance Matrices to Covariance Operators: Data Representation from Finite to Infinite-Dimensional Settings. In *Algorithmic Advances in Riemannian Geometry and Applications*, Springer, 2017
- H.Q. Minh. International Conference on Computer Vision (ICCV 2017) Tutorial, http://www.covariance2017.eu/

# Exposition

Covariance representation in computer vision
From finite to infinite-dimensional settings



# Exposition


Advances in Computer Vision and Pattern Recognition
A C V P R
Hà Quang Minh
Vittorio Murino Editors
Algorithmic Advances in Riemannian Geometry and Applications
For Machine Learning, Computer Vision, Statistics, and Optimization
Springer

Thank you for listening!

Questions, comments, suggestions?