Kernel Principal Geodesic Analysis

Suyash P. Awate[1,2], Yen-Yun Yu[2], Ross T. Whitaker[2]

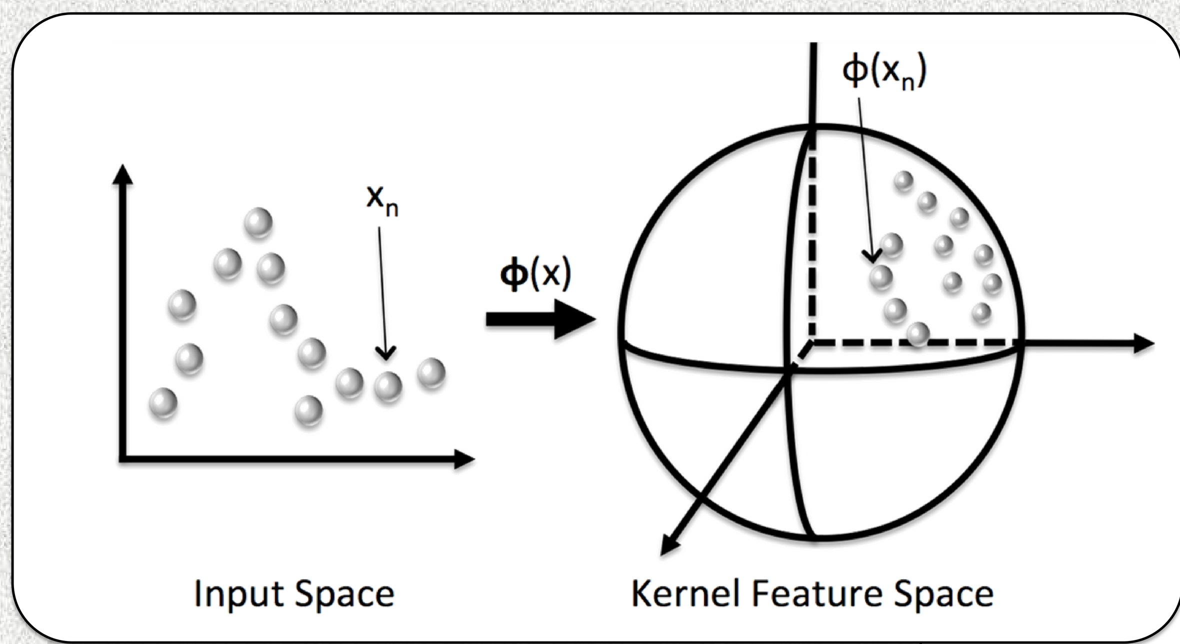
1) Computer Science and Engineering Department, Indian Institute of Technology (IIT) Bombay

2) Scientific Computing and Imaging (SCI) Institute, School of Computing, University of Utah

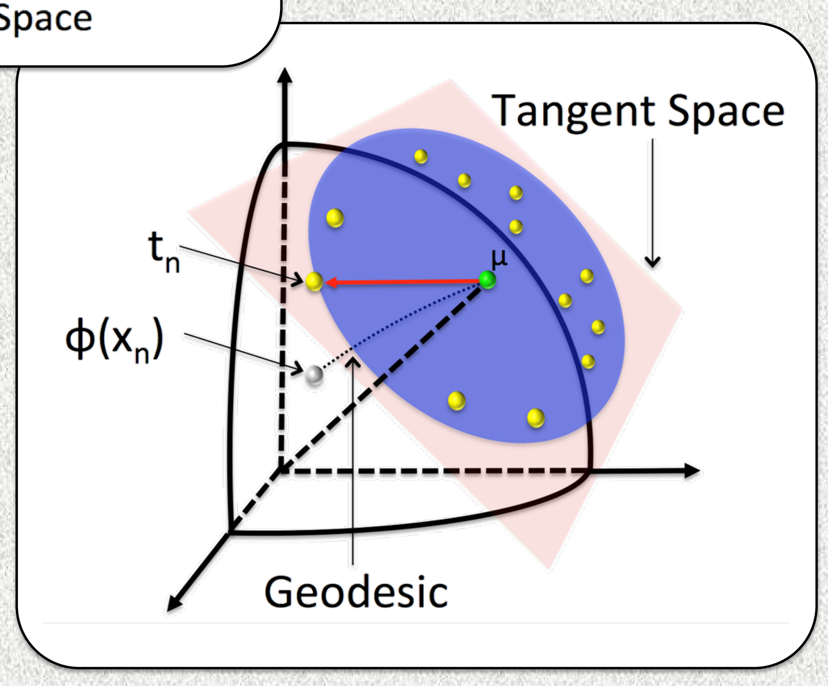
www.sci.utah.edu

Abstract

- **Motivation:** Many popular kernels that map input points to a *hypersphere* in kernel feature space (KFS) (Courty et al., ECML-PKDD, 2012), where Euclidean statistical analysis can make kernel PCA (kPCA) (Scholkopf et al., Neural. Comput.,1998) inefficient.



- **Proposed Method:** *Kernel principal geodesic analysis* (kPGA), an adaptation of kPCA, for statistical analysis on hyperspheres in KFS, including algorithms for estimating:
  1) Sample weighted Karcher mean
  2) Eigenanalysis of sample-weighted Karcher covariance



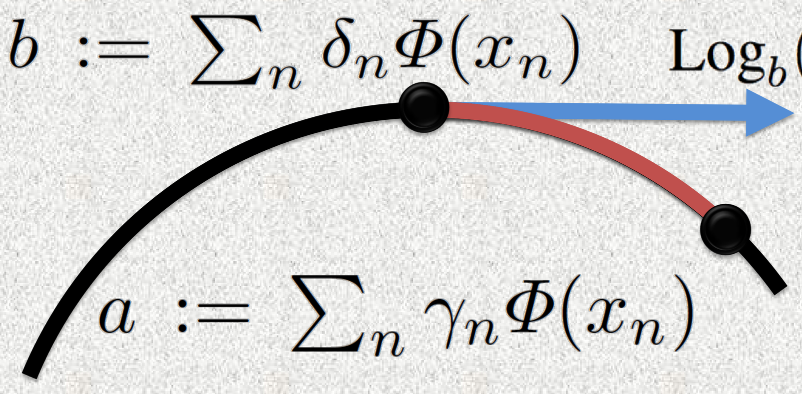
Geometry of Hilbert Sphere in KFS

- **Logarithmic Map in KFS:** Linear combination of points

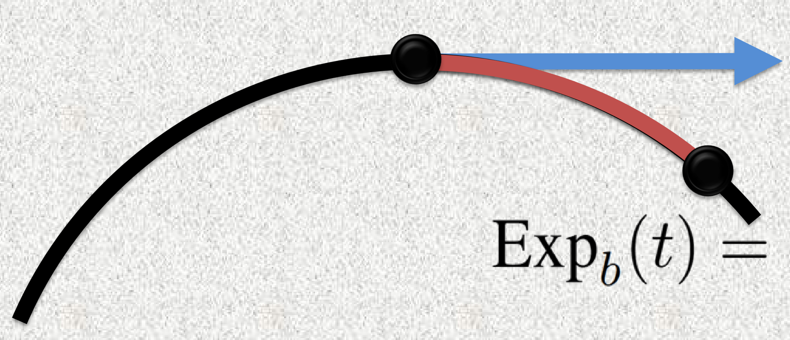
$$b := \sum_n \delta_n \Phi(x_n) \qquad \mathrm{Log}_b(a) = \frac{a - \langle a, b\rangle_{\mathcal{F}} b}{\|a - \langle a, b\rangle_{\mathcal{F}} b\|_{\mathcal{F}}} \arccos(\langle a, b\rangle_{\mathcal{F}})$$

$$= \sum_n \zeta_n \Phi(x_n), \text{ where } \forall n : \zeta_n \in \mathbb{R}$$

$$a := \sum_n \gamma_n \Phi(x_n)$$



- **Exponential Map in KFS:** Linear combination of points



$$t := \sum_n \beta_n \Phi(x_n)$$

$$\mathrm{Exp}_b(t) = \cos(\|t\|_{\mathcal{F}}) b + \sin(\|t\|_{\mathcal{F}}) \frac{t}{\|t\|_{\mathcal{F}}}$$

$$= \sum_n \omega_n \Phi(x_n), \text{ where } \forall n : \omega_n \in \mathbb{R}$$

Sample Karcher Mean

- Linear combination of points; lies on the hypersphere in KFS
- **Gradient-descent algorithm:**
  1) Input: A set of points on hypersphere in KFS:

$$y_m := \sum_n w_{mn} \Phi(x_n) \qquad \{y_m\}_{m=1}^M$$

  2) Initialize

$$\mu^0 = \frac{\sum_m p_m y_m}{\|\sum_m p_m y_m\|_{\mathcal{F}}} = \sum_n \xi_n \Phi(x_n)$$

  3) Iteratively update until convergence

$$\mu^{i+1} = \mathrm{Exp}_{\mu^i}\left(\frac{\tau^i}{M} \sum_m p_m \mathrm{Log}_{\mu^i}(y_m)\right)$$

Sample Karcher Covariance

- **Covariance in Tangent Space at Mean:**

$$C := \frac{1}{M} \sum_m p_m z_m \otimes z_m, \text{ where } z_m = \mathrm{Log}_\mu(y_m) = \sum_{n'} \beta_{n'm} \Phi(x_{n'})$$

$$C = \sum_{n'} \sum_{n''} E_{n'n''} \Phi(x_{n'}) \otimes \Phi(x_{n''}),$$

$$\text{where } E_{n'n''} = \frac{1}{M} \sum_m p_m \beta_{n'm} \beta_{n''m}$$

- **Eigenfunctions of Covariance:**
  Linear combinations of points; lies in the tangent space in KFS

$$v = \frac{1}{\lambda} \sum_{n'} \sum_{n''} E_{n'n''} \Phi(x_{n'}) \otimes \Phi(x_{n''}) v = \sum_{n'} \alpha_{n'} \Phi(x_{n'}),$$

$$\text{where } \alpha_{n'} = \sum_{n''} \frac{E_{n'n''}}{\lambda} \langle \Phi(x_{n''}), v \rangle_{\mathcal{F}}$$

PGA on Hilbert Sphere in KFS

- **Generalized Eigenanalysis Problem:**

$$KEK\alpha = \lambda K\alpha$$

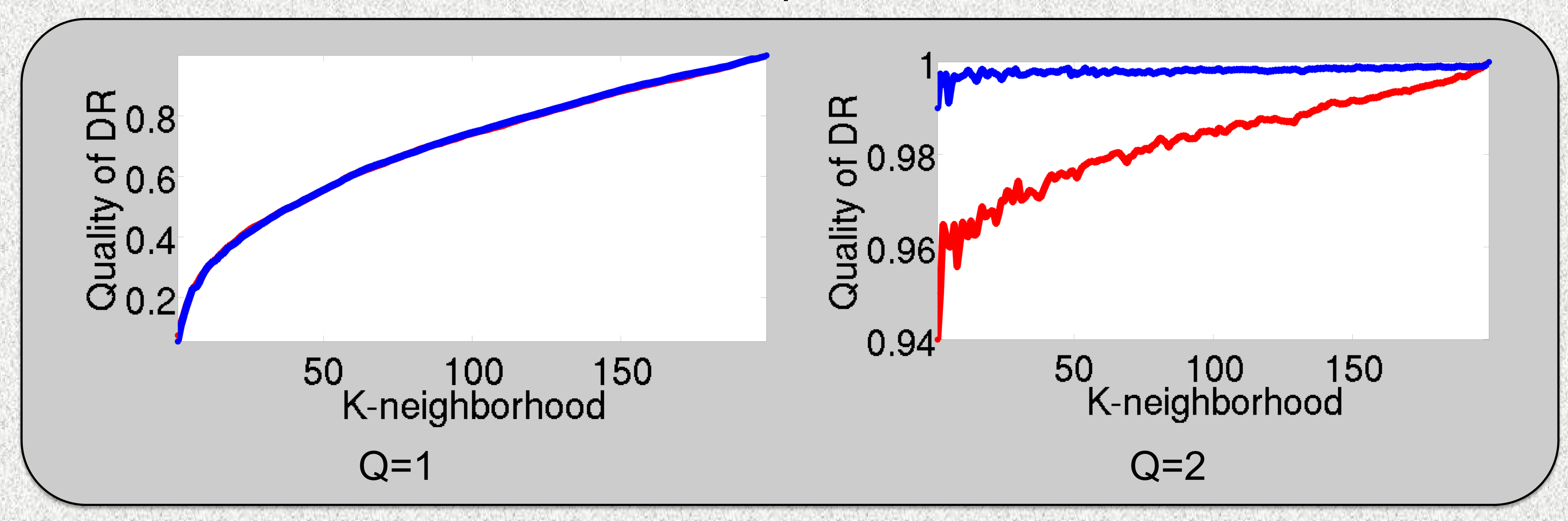
where K is the Gram matrix, E is as defined before.

- **If Mercer kernel is symmetric positive-definite, standard eigenanalysis problem:**

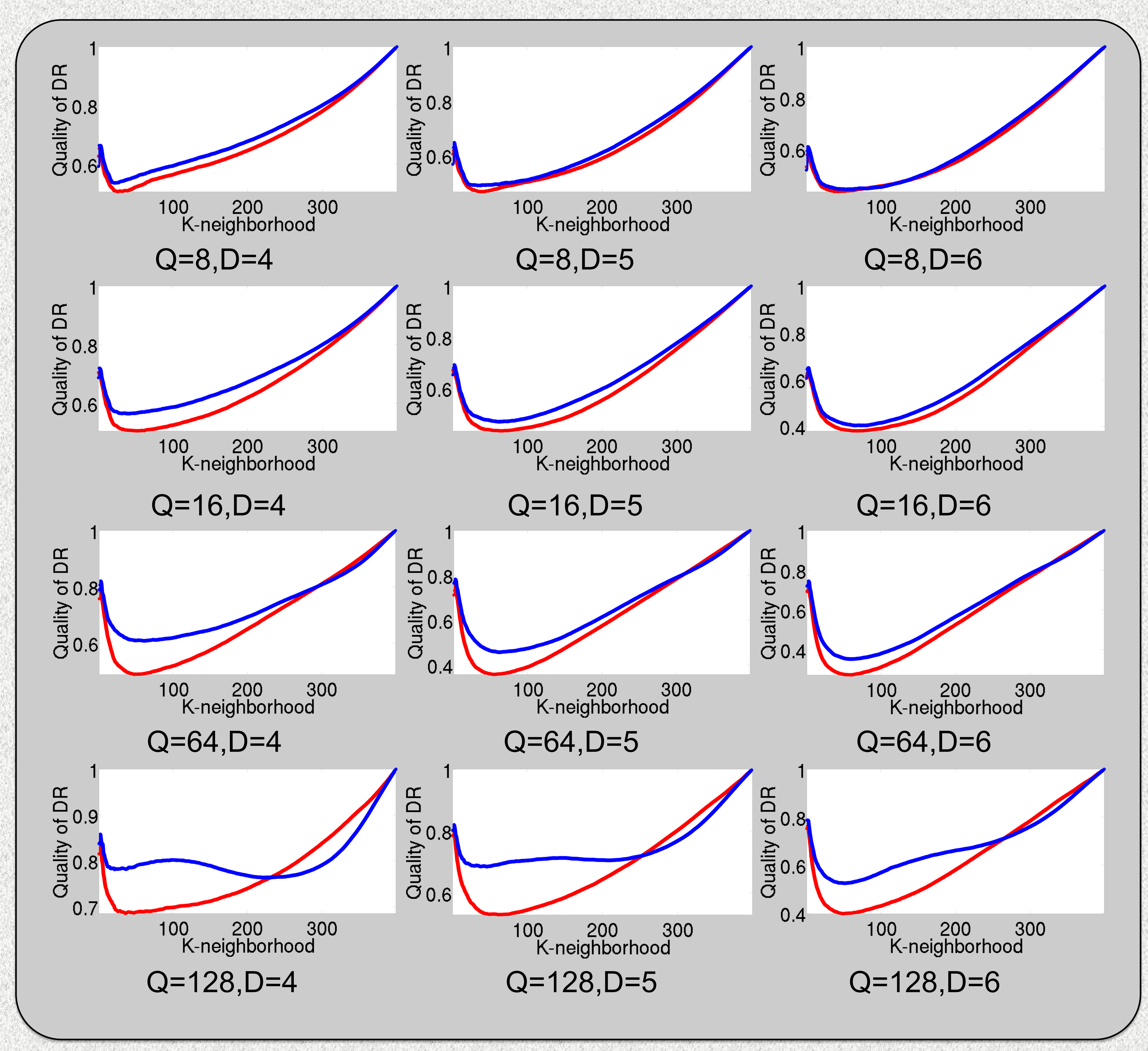
$$EK\alpha = \lambda\alpha$$

Results

- **Nonlinear Dimension Reduction**
  1) Evaluation via co-ranking matrix (Lee and Verleysen, Neurocomputing, 2009)
     a) Data in the high-dimensional space
     b) Data in the lower-dimensional embedding
  2) kPGA compared with kPCA
     a) Simulated data sampled from a low-dimensional von Mises Fisher distribution. Inner-product kernel.



     b) ORL face image database. Polynomial kernel using normalized images.



- **Clustering**
  1) Gaussian Mixture Model fitted via Expectation Maximization
  2) kPGA compared with kPCA and spectral clustering
  3) UCI machine learning repository