Vector-valued Reproducing Kernel Hilbert Spaces

with applications to Function Extension and Image Colorization

Hà Quang Minh
minh.haquang@iit.it

Istituto Italiano di Tecnologia
Outline of the Talk

- Brief Review of Scalar-valued RKHS
- Vector-valued RKHS
- Function Extension: 2 algorithms
- Application: Image Colorization
Colorization: Function extension

Figure 1: About 0.96% of color is given
Positive Definite Kernels

- $X$ any nonempty set
- $K : X \times X \to \mathbb{R}$ 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 $\{x_i\}_{i=1}^{N} \in X$ and real numbers $\{a_i\}_{i=1}^{N} \in \mathbb{R}$.

- Complex-valued kernels are often encountered in complex analysis.
Abstract theory due to Aronszajn (1950).

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

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

with inner product

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

$\mathcal{H}_K = \text{RKHS associated with } K \text{ (unique).}$
Reproducing Property: for each $f \in \mathcal{H}_K$, for every $x \in X$

$$f(x) = \langle f, K_x \rangle_K$$

Assumption

$$\kappa = \sup_{x \in X} \sqrt{K(x, x)} < \infty$$

Then

$$\|f\|_{\infty} \leq \kappa \|f\|_K$$
Examples: RKHS

For $s > n/2$, the Sobolev space $H^s(\mathbb{R}^n)$, with

$$\| f \|_{H^s(\mathbb{R}^n)}^2 = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \left| (1 + |\xi|^2)^{s/2} \hat{f}(\xi) \right|^2 d\xi < \infty,$$

is an RKHS, with kernel

$$K(x, y) = \frac{1}{(2\pi)^n} \frac{1}{(1 + |\xi|^2)^s} (x - y)$$
Examples: RKHS

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

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

The Laplacian kernel $K(x, y) = \exp(-a|x-y|)$, $a > 0$, 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 $C(n) = 2^n \pi^{\frac{n-1}{2}} \Gamma\left(\frac{n+1}{2}\right)$
Examples: RKHS

- The Laplacian kernel has less smoothing effect than the Gaussian kernel (may be useful if we do not want very smooth functions)

- Generalization of the Gaussian kernel:
  \[ K(x, y) = \exp\left(-\frac{|x-y|^p}{\sigma^2}\right), \text{ where } 0 \leq p \leq 2 \] (Schoenberg 1938).
Outline of the Talk

- Brief Review of Scalar-valued RKHS
- **Vector-valued** RKHS
- Function Extension: 2 algorithms
- Application: Image Colorization
Vector-valued RKHS

- Laurent Schwartz (1964): very general framework for RKHS of functions with values in locally convex topological spaces


- Here we will focus on RKHS of functions with values in a Hilbert space.
Operator-valued kernels

- $D$ a nonempty set, $\mathcal{W}$ a real Hilbert space with the inner product $\langle \cdot, \cdot \rangle_\mathcal{W}$, $\mathcal{L}(W)$ the Banach space of bounded linear operators on $\mathcal{W}$.

- A function $K : D \times D \to \mathcal{L}(\mathcal{W})$ is said to be an **operator-valued positive definite kernel** if for each pair $(x, y) \in D \times D$, $K(x, y) \in \mathcal{L}(\mathcal{W})$ is a self-adjoint operator and

$$
\sum_{i,j=1}^{N} \langle w_i, K(x_i, x_j)w_j \rangle_\mathcal{W} \geq 0
$$

for every finite set of points $\{x_i\}_{i=1}^{N}$ in $D$ and $\{w_i\}_{i=1}^{N}$ in $\mathcal{W}$, where $N \in \mathbb{N}$.
Vector-valued RKHS

- $\mathcal{W}^D = \text{vector space of all functions } f : D \rightarrow \mathcal{W}$.

- For each $x \in D$ and $w \in \mathcal{W}$, we form a function $K_xw = K(., x)w \in \mathcal{W}^D$ defined by
  
  $$(K_xw)(y) = K(y, x)w \quad \text{for all } y \in D.$$ 

Consider the set

$\mathcal{H}_0 = \text{span}\{K_xw \mid x \in D, \ w \in \mathcal{W}\} \subset \mathcal{W}^D$. For

$f = \sum_{i=1}^N K_x w_i$, $g = \sum_{i=1}^N K_y z_i \in \mathcal{H}_0$, we define

$$\langle f, g \rangle_{\mathcal{H}_K} = \sum_{i,j=1}^N \langle w_i, K(x_i, y_j)z_j \rangle_{\mathcal{W}}.$$
Taking the closure of $\mathcal{H}_0$ gives the Hilbert space $\mathcal{H}_K$.

The **reproducing property** is

$$\langle f(x), w \rangle_{\mathcal{W}} = \langle f, K_x w \rangle_{\mathcal{H}_K} \quad \text{for all} \quad f \in \mathcal{H}_K.$$

For each $x \in D$ and $f \in \mathcal{H}_K$:

$$\|f(x)\|_{\mathcal{W}} \leq \sqrt{\|K(x, x)\|} \|f\|_{\mathcal{H}_K}. $$
Simple example: let $k(x, y)$ be a real-valued positive definite kernel and $B$ a positive definite matrix. Then

$$K(x, y) = Bk(x, y)$$

is a matrix-valued kernel, which induces a vector-valued RKHS
Outline of the Talk

- Brief Review of Scalar-valued RKHS
- Vector-valued RKHS
- Function Extension: 2 algorithms
- Application: Image Colorization
Function Extension

- $D \subset \Omega$ are closed sets in a complete separable metric space
- $f : D \to \mathcal{W}$,
- Goal: extend $f : D \to \mathcal{W}$ to $F : \Omega \to \mathcal{W}$, such that $F$ is close to $f$ on the smaller set $D$, and reasonably well-behaved on the larger set $\Omega$. 
Extension Operator

- Assume we have a kernel $K : \Omega \times \Omega \to \mathcal{W}$.
- Assume that $K(x, x)$ is compact for each $x$, and that
  \[ \sup_{x \in \Omega} \|K(x, x)\| < \infty. \]
- For $f : D \to \mathcal{W}$, define $L_K : L_2^{\mu}(D; \mathcal{W}) \to \mathcal{H}_K(\Omega)$, with
  \[ L_K f(x) = \int_D K(x, y) f(y) d\mu(y), \]
  for every $x \in \Omega$. This defines an extension operator.

The adjoint operator $L_K^* : \mathcal{H}_K(\Omega) \to L_2^{\mu}(D; \mathcal{W})$ is the restriction operator:

\[ L_K^* F = F|_D \]
Function Extension

Find the extension function $F : \Omega \rightarrow \mathcal{W}$ by solving the minimization problem

$$\inf_{F \in \mathcal{H}_K(\Omega)} \|f - L_K^* F\|_{L^2_\mu(D;\mathcal{W})}^2 + \gamma \|F\|_{\mathcal{H}_K(\Omega)}^2,$$

This problem has a unique solution

$$F_\gamma = (L_K L_K^* + \gamma I)^{-1} L_K f$$
Function Extension: Spectral Algorithm

- Scalar version: Coifman-Lafon (2005)
- Considered as an operator $L^2_\mu(D; \mathcal{W}) \to L^2_\mu(D; \mathcal{W})$, $L_K$ is compact, positive, with orthonormal spectrum $(\lambda_k, \phi_k)_{k=1}^{\infty}$.
- Eigenfunction extension: for $\lambda_k > 0$, we extend $\phi_k : D \to \mathcal{W}$ to $\Phi_k : \Omega \to \mathcal{W}$ by

$$
\Phi_k(x) = \frac{1}{\lambda_k} \int_D K(x, y) \phi_k(y) d\mu(y), \quad \text{for } x \in \Omega.
$$

- To be numerically reliable, one may want to consider only $\lambda_k > \delta$, for some given $\delta > 0$. 
Function Extension: Spectral Algorithm

- Compute the eigenvalues and eigenfunctions $\{(\lambda_k, \phi_k)\}$ of $L_K : L^2_\mu(D; \mathcal{W}) \to L^2_\mu(D; \mathcal{W})$.

- Compute the expansion coefficients $a_k$’s of $f$ in the basis $\{\phi_k\}$: $f = \sum_k a_k \phi_k$

- Compute $F_\delta = \sum_{k, \lambda_k \geq \delta} \frac{\lambda_k}{\lambda_k + \gamma} a_k \Phi_k$, for some $\delta > 0$

- Alternatively, to take care of the case $\lambda_k = 0$, compute directly

$$F_\gamma(x) = \sum_{k=1}^\infty \frac{a_k}{\lambda_k + \gamma} \int_D K(x, y) \phi_k(y) d\mu(y)$$
Assume now that $D = \{x_i\}_{i=1}^m$, with $w_i = f(x_i)$.

An algorithm with real kernel-based flavor:

$$F_\gamma = \arg \min_{F \in \mathcal{H}_K(\Omega)} \frac{1}{m} \sum_{i=1}^m ||F(x_i) - w_i||^2_\mathcal{W} + \gamma||F||^2_{\mathcal{H}_K(\Omega)}.$$  

This has a unique solution $F_\gamma = \sum_{i=1}^m K_{x_i} a_i$, with $F_\gamma(x) = \sum_{i=1}^m K(x, x_i) a_i$, where the vectors $a_i \in \mathcal{W}$ satisfy the $m$ linear equations

$$\sum_{j=1}^m K(x_i, x_j) a_j + m \gamma a_i = w_i.$$  

Vector-valued Reproducing Kernel Hilbert Spaces – p. 22/41
Compare two algorithms

- **Spectral**: theoretically more general ($D$ can be either discrete or continuous)

- If $D$ is discrete and $\mu$ is the uniform distribution, then Least square and Spectral are the same analytically.

- Numerically, Least square is easier to implement and should be expected to be more stable (involves solving well-conditioned systems of linear equations, vs finding eigenvalues/eigenfunctions of the Spectral method).

- The basis functions in Least square are exact (based on the given data points)

- Here we will focus on the Least square method for numerical work
Outline of the Talk

- Brief Review of Scalar-valued RKHS
- Vector-valued RKHS
- Function Extension
- Application: Image Colorization
Joint work with Sung Ha Kang (Georgia Tech) and Triet Le (Yale)

\( \Omega \) is the given grayscale image

\( D \subset \Omega \) is the given region with colors (often very small).

The initial function here is \( f : D \to \mathbb{R}^3 \) (red, green, blue)

Goal: extend the colors to all of \( \Omega \).

Nonlocal kernel

- Simplest scenario: all the colors are independent.

\[ K(x, y) = \text{diag}(k_1(x, y), k_2(x, y), k_3(x, y)) \] where each \( k_i \) is a scalar-valued kernel.

- Here we will use scalar-valued kernels of the form

\[ k(x, y) = \exp\left(-\frac{|g_r(x) - g_r(y)|^p}{\sigma_1}\right) \exp\left(-\frac{|x - y|^p}{\sigma_2}\right) \]

where \( g_r(x) \) is the patch of radius \( r \) centered at \( x \), of size \((2r + 1) \times (2r + 1)\), with \( g \) denoting the gray level.

- Extend the color function using least square RKHS
Chromaticity and Brightness Model

For sharper resulting images, we consider the CB model of color.

\[ f(x) = B(x)C(x), \] where \( B(x) \) is the brightness, and \( C(x) = (r(x), g(x), b(x)) \in S^2 \).

**Assumption:** we are given the brightness \( B(x) \) on all of \( \Omega \), but \( C(x) \) only on \( D \).

**Need:** to extend \( C(x) \) to all of \( \Omega \).

**Problem:** the set of \( S^2 \)-valued functions is not a vector space.
**Stereographic Projection**

**Solution** for the $S^2$-valued Chromaticity function:

**Stereographic projection**

Since the colors are all nonnegative and for symmetry, we need a **symmetric** stereographic projection that projects from the first quadrant.

- Projection point: $\left(-\frac{1}{\sqrt{3}}, -\frac{1}{\sqrt{3}}, -\frac{1}{\sqrt{3}}\right)$
- Projection plane: $X + Y + Z = 0$
Stereographic Projection

Forward projection from $S^2$ onto $X + Y + Z = 0$:

$X = \frac{3x-(x+y+z)}{\sqrt{3}(x+y+z+\sqrt{3})}$, $Y = \frac{3y-(x+y+z)}{\sqrt{3}(x+y+z+\sqrt{3})}$, $Z = \frac{3z-(x+y+z)}{\sqrt{3}(x+y+z+\sqrt{3})}$,

Inverse projection from $X + Y + Z = 0$ onto $S^2$:

$x = \frac{2\sqrt{3}X+1-(X^2+Y^2+Z^2)}{\sqrt{3}(1+X^2+Y^2+Z^2)}$, $y = \frac{2\sqrt{3}Y+1-(X^2+Y^2+Z^2)}{\sqrt{3}(1+X^2+Y^2+Z^2)}$, $z = \frac{2\sqrt{3}Z+1-(X^2+Y^2+Z^2)}{\sqrt{3}(1+X^2+Y^2+Z^2)}$. 
Image Colorization Algorithm

- Given: Brightness $B(x)$ on all of $\Omega$ and Chromaticity on small subset $D \subset \Omega$
- Project $C(x) : D \rightarrow S^2$ to $C(x) : D \rightarrow \mathbb{R}^2$
- Extend $C(x)$ to $\Omega \rightarrow \mathbb{R}^2$ using the least square algorithm in the RKHS induced by the nonlocal kernel above (kernel constructed using $B(x)$)
- Project the results back onto $S^2$ to get the extended Chromaticity function from $\Omega \rightarrow S^2$
- Multiply the resulting Chromaticity with the given Brightness to obtain the final answer.
Colorization Algorithm - Complexity

- Involves solving 2 systems of linear equations, each of size $m \times m$, where $m = |D|$
- Evaluation step involves computing kernel matrix of size $m \times M$, where $M = |\Omega|$
- Main computation time is in computing the kernel
- Explicit and unique solution, no iteration required
Figure 2: $p = 1, r = 1, \sigma_1 = 0.5, \sigma_2 = 1$. About 0.5% of color is given.
Figure 3: $p = 1$, $r = 1$, $\sigma_1 = 0.5$, $\sigma_2 = 1$. About 1% of color is given.
Numerical Examples

Figure 4: $p = 1$, $r = 1$, $\sigma_1 = 0.5$, $\sigma_2 = 1$. About 1% of color is given.
Numerical Examples

Figure 5: $p = 1$, $r = 2$, $\sigma_1 = 0.5$, $\sigma_2 = 2$. About 0.96% of color is given.
Figure 6: The colorization result with $r = 0$, $p = 2$, $\sigma_1 = 0.001$, and $\sigma_2 = 10$. 
Figure 7: Chromaticity and Brightness model via Stereographic Projection vs. RGB channel: $\rho = 1$, $r = 2$, $\sigma_1 = 0.5$, and $\sigma_2 = 10$
Numerical Examples

Figure 8: $p = 2$, $r = 2$, $\sigma_1 = 0.1$, and $\sigma_2 = 10$
Figure 9: The colorization result with $r = 10$, $p = 1.5$, $\sigma_1 = 0.4$, $\sigma_2 = 10$. Less than 2% of color is given.
Conclusion

- Operator-valued positive definite kernels and their induced vector-valued RKHS
- Use of RKHS for the problem of function extension (vector-valued)
- An application in Image Colorization
Acknowledgement

- Institute for Pure and Applied Mathematics (IPAM)
- German Research Foundation (DFG)
- Hausdorff Institute for Mathematics (HIM) in Bonn (Junior Program in Analysis, September-October 2008)