PADERBORN UNIVERSITY

PADERBORN UNIVERSITY

KERNEL-BASED APPROXIMATION OF THE KOOPMAN GENERATOR

Feliks Nüske

Magdeburg, 31 July, 2020

gEDMD

Elements of RKHS Theory

Generator EDMD and Kernels

Approximation of the Generator

Recall that for SDEs, the generator is the differential operator:

$$\mathcal{L}\psi(x) = b(x) \cdot \nabla\psi(x) + \frac{1}{2}a(x) : \nabla^2\psi(x).$$

Galerkin approximation to $\mathcal{L}$ on a trial space $\mathbb{W} = \mathrm{span}\{\psi_i\}_{i=1}^n$:

$$\mathbf{L} = \mathbf{G}^{-1}\mathbf{A}, \qquad \mathbf{G}_{ij} = \langle \psi_i, \psi_j \rangle_\mu, \qquad \mathbf{A}_{ij} = \langle \psi_i, \mathcal{L}\psi_j \rangle_\mu.$$

By ergodicity, data-driven approximations are (gEDMD):

$$\mathbf{G} = \frac{1}{M}\mathbf{X}\mathbf{X}^T, \quad \mathbf{A} = \frac{1}{M}\mathbf{X}\mathbf{dX}^T, \quad \mathbf{X}_{im} = \psi_i(X_{t_m}), \quad \mathbf{dX}_{im} = \mathcal{L}\psi_i(X_{t_m}).$$

Klus and **Nüske** et al, *Physica D* (2020) gEDMD

Elements of RKHS Theory

Generator EDMD and Kernels

Reproducing Kernel Hilbert Space

Let $\mathbb{H} \subset C(\mathbb{X})$ be a Hilbert space of continuous functions on a domain $\mathbb{X}$. Assume the point evaluation functional $\delta_x f = f(x)$ is continuous for some $x \in \mathbb{X}$. Then there is a function $k_x(\cdot) \in \mathbb{H}$ such that

$$f(x) = \delta_x f = \langle f,\ k_x(\cdot)\rangle_{\mathbb{H}}$$

for all $f \in \mathbb{H}$. If the above is true for all $x \in \mathbb{X}$, we can define a two-parameter function

$$k(x, y) := k_x(y)$$

which is called the *reproducing kernel* for $\mathbb{H}$.

Definition RKHS

Definition

Let $\mathbb{X}$ be a suitable domain and $\mathbb{H}$ a space of continuous functions $f: \mathbb{X} \to \mathbb{R}$. Then $\mathbb{H}$ is called a *reproducing kernel Hilbert space* (RKHS) with inner product $\langle \cdot, \cdot \rangle_{\mathbb{H}}$ if a function $k: \mathbb{X} \times \mathbb{X} \to \mathbb{R}$ exists such that

1. $\mathbb{H} = \overline{\mathrm{span}\{k(x, \cdot),\ x \in \mathbb{X}\}}$,
2. $\langle f, k(x, \cdot) \rangle_{\mathbb{H}} = f(x)$ for all $f \in \mathbb{H}$.

Basic Properties

1. Symmetry:

$$\begin{aligned} k(x,y) &= \delta_y k_x(\cdot) = \langle k(x,\cdot),\, k(y,\cdot)\rangle_{\mathbb{H}} \\ &= \langle k(y,\cdot),\, k(x,\cdot)\rangle_{\mathbb{H}} = k(y,x) \end{aligned}$$

2. Positive semi-definiteness: let $x_j \in \mathbb{X}$, $c_j \in \mathbb{R}$, $j = 1, \ldots, N$:

$$\sum_{j,k=1}^{N} c_j c_k k(x_j, x_k) = \left\langle \sum_{j=1}^{N} c_j k(x_j,\cdot),\, \sum_{j=1}^{N} c_j k(x_j,\cdot) \right\rangle_{\mathbb{H}} \geq 0.$$

The *Gramian* matrix is always positive semi-definite.

A two-argument function with these properties is called *positive semi-definite*.

Equivalence Result

Theorem (Moore-Aronszajn)

Let $k(\cdot, \cdot)$ be a positive semi-definite function. Then there is a unique RKHS $\mathbb{H}$ with k as its reproducing kernel.

Idea of the Proof.

Consider the linear span $\tilde{\mathbb{H}}$ of all finite combinations $\sum_{j=1}^{N} c_j k(x_j, \cdot)$, with semi inner product $\sum_{j,k} c_j d_k k(x_j, x_k)$. It can be shown that this is actually a genuine inner product, and $\tilde{\mathbb{H}}$ is a pre-Hilbert space. Complete it and identify elements $\tilde{f}$ of the completion $\mathbb{H}$ with a function by $f(x) = \left\langle \tilde{f},\ k(x, \cdot) \right\rangle_{\mathbb{H}}$. □

Kernel Trick

In summary, a RKHS is a Hilbert space of functions where inner products can be evaluated by kernel evaluations (*the "kernel trick"*).

Example

- The most popular kernel is the Gaussian $k(x, y) = \exp(-\frac{1}{2\sigma^2}\|x - y\|^2)$.
- Many other kernels with rich properties have been developed for a broad range of applications.

Derivative Reproducing Property

The Reproducing Property can be extended to derivatives if the kernel is smooth:

Theorem

Let $k(\cdot, \cdot) \in C^{2k}(\mathbb{X} \times \mathbb{X})$ *be a positive definite function on an open set* $\mathbb{X}$*. Then all functions in* $\mathbb{H}$ *are* C^k *and we have for all* $|\alpha| \leq k$*:*

$$D^\alpha f(x) = \langle f,\, D^\alpha k(x, \cdot)\rangle_{\mathbb{H}} \,,$$

where the derivative acts on the first argument of k*.*

References

Berlinet and Thomas-Agnan, *Reproducing Kernel Hilbert Spaces in Probability and Statistics* (2004)

Wendland, *Scattered Data Approximation* (2005)

Contents

gEDMD

Elements of RKHS Theory

Generator EDMD and Kernels

Rank-one Operators

In order to approximate the generator on an RKHS, we introduce rank-one operators of the type:

$$\mathcal{T}_x^\alpha : \mathbb{H} \mapsto \mathbb{H}, \; x \in \mathbb{X},$$
$$\mathcal{T}_x^\alpha f := \langle f, \, D^\alpha k(x, \cdot) \rangle_{\mathbb{H}} \, k(x, \cdot).$$

We can verify for all $f, g \in \mathbb{H}$:

$$\begin{aligned} \langle \mathcal{T}_x^\alpha f, \, g \rangle_{\mathbb{H}} &= \langle f, \, D^\alpha k(x, \cdot) \rangle_{\mathbb{H}} \, \langle k(x, \cdot), \, g \rangle_{\mathbb{H}} \\ &= D^\alpha f(x) \, \langle k(x, \cdot), \, g \rangle_{\mathbb{H}} = D^\alpha f(x) g(x). \end{aligned}$$

kgEDMD

Approximating Differential Operators

Consider a formal operator

$$\mathcal{T}_{\mathbb{H}}^{\alpha} f = \int_{\mathbb{X}} w(x) \left\langle f,\; D^{\alpha} k(x,\cdot) \right\rangle_{\mathbb{H}} k(x,\cdot)\, \mathrm{d}\mu(x),$$

where w is a weight function. This is a bounded linear operator on $\mathbb{H}$ if:

$$\int_{\mathbb{X}} |w(x)| \|D^{\alpha} k(x,\cdot)\|_{\mathbb{H}} \|k(x,\cdot)\|_{\mathbb{H}} \,\mathrm{d}\mu(x) < \infty.$$

Approximating Differential Operators

We can then verify that

$$
\begin{aligned}
\langle \mathcal{T}_{\mathbb{H}}^{\alpha} f,\, g \rangle_{\mathbb{H}} &= \left\langle \int_{\mathbb{X}} w(x) \, \langle f,\, D^{\alpha} k(x,\cdot) \rangle_{\mathbb{H}} \, k(x,\cdot) \, \mathrm{d}\mu(x),\, g \right\rangle_{\mathbb{H}} \\
&= \int_{\mathbb{X}} w(x) \, \langle f,\, D^{\alpha} k(x,\cdot) \rangle_{\mathbb{H}} \, \langle k(x,\cdot),\, g \rangle_{\mathbb{H}} \, \mathrm{d}\mu(x) \\
&= \int_{\mathbb{X}} w(x) D^{\alpha} f(x) g(x) \, \mathrm{d}\mu(x).
\end{aligned}
$$

The last term is an inner product in L^2_μ between a function g and the action of a differential operator $w(\cdot)D^{\alpha} f(\cdot)$ on another function f.

Approximating the Generator

For the generator of an SDE, we can make the definitions:

$$\mathcal{L}_{\mathbb{H}} f = \int_{\mathbb{X}} \left[\sum_i b_i(x) \left\langle D^{e_i} k(x,\cdot),\ f \right\rangle_{\mathbb{H}} + \frac{1}{2} \sum_{i,j} a_{ij}(x) \left\langle D^{e_i + e_j} k(x,\cdot),\ f \right\rangle_{\mathbb{H}} \right] k(x,\cdot)\, \mathrm{d}\mu(x),$$

$$\mathcal{C}_{00} f = \int_{\mathbb{X}} \langle k(x,\cdot),\ f \rangle_{\mathbb{H}}\, k(x,\cdot)\, \mathrm{d}\mu(x).$$

Galerkin Approximation

Theorem

Assume that $\mathbb{H} \subset \mathcal{D}(\mathcal{L})$*, and that*

$$\int |a_{ij}(x)| \| D^{e_i + e_j} k(x, \cdot) \|_{\mathbb{H}} \| k(x, \cdot) \|_{\mathbb{H}} \, \mathrm{d}\mu(x) < \infty,$$

$$\int |b_i(x)| \| D^{e_i} k(x, \cdot) \|_{\mathbb{H}} \| k(x, \cdot) \|_{\mathbb{H}} \, \mathrm{d}\mu(x) < \infty,$$

$$\int \| k(x, \cdot) \|_{\mathbb{H}}^2 \, \mathrm{d}\mu(x) < \infty.$$

Then, for all $f, g \in \mathbb{H}$*,*

$$\langle \mathcal{L} f, g \rangle_\mu = \langle \mathcal{L}_{\mathbb{H}} f, g \rangle_{\mathbb{H}}, \qquad \langle f, g \rangle_\mu = \langle \mathcal{C}_{00} f, g \rangle_{\mathbb{H}}.$$

Klus, Nüske, and Hamzi, *Entropy* (2020)

Numerical Algorithm

Data-driven approximations of the RKHS operators:

$$\mathrm{d}k(x_m) = \frac{1}{2}\sum_{i,j} a_{ij}(x_m) D^{e_i+e_j} k(x_m, \cdot) + \sum_i b_i(x_m) D^{e_i} k(x_m, \cdot),$$

$$\widehat{\mathcal{C}}_{00} = \frac{1}{M}\sum_{m=1}^{M} \langle k(x_m, \cdot), \cdot\rangle_{\mathbb{H}}\, k(x_m, \cdot),$$

$$\widehat{\mathcal{L}}_{\mathbb{H}} = \frac{1}{M}\sum_{m=1}^{M} \langle \mathrm{d}k(x_m), \cdot\rangle_{\mathbb{H}}\, k(x_m, \cdot).$$

Convergence: $\|\mathcal{L}_{\mathbb{H}} - \widehat{\mathcal{L}}_{\mathbb{H}}\|_{\mathrm{HS}} \to 0$, $\|\mathcal{C}_{00} - \widehat{\mathcal{C}}_{00}\|_{\mathrm{HS}} \to 0$ with $m \to \infty$.

Klus, Nüske, and Hamzi, *Entropy* (2020)

Comments

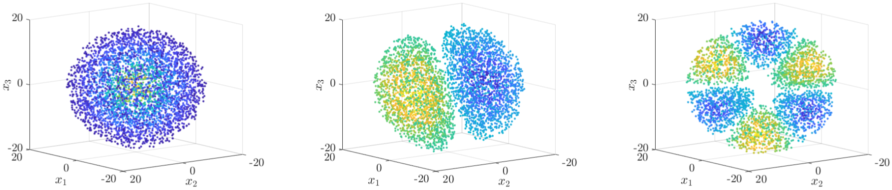
- Finite-dimensional matrix representations of these operators can be obtained. These matrices require evaluations of drift / diffusion, and of derivatives of the kernel function, at the data sites.
- The same formalism can be applied to derive kernel approximations for other differential operators, e.g. Schrödinger operators.

Klus, Nüske, and Hamzi, *Entropy* (2020)

Example

Schrödinger operator for the hydrogen atom in $\mathbb{R}^3$:

$$\mathcal{T} = -\frac{\hbar^2}{2m}\Delta - \frac{e^2}{4\pi\epsilon_0 \|x\|}$$



Discussion

- \+ Large approximation spaces can be used implicitly to approximate the Koopman generator.
- \+ Approximation quality of RKHS has been studied for a long time.
- +- Requires solution of matrix eigenvalue problems of the same size as the data.
- \- These problems may be poorly conditioned.
- \- Hyper-parameters of the kernel function need to be tuned.