About “Theoretical Foundations for Higher Order Dynamic Mode Decompositions” by J. Rosenfeld

Corbinian Schlosser

January 19, 2021

Setting

Given are:

1. A set $X \subset \mathbb{R}^n$ (constraint set for our dynamical system),
2. dynamics $f : \mathbb{R}^n \to \mathbb{R}^n$ for the differential equation $\dot{x} = f(x)$ (for the first part we may assume that f is known, later we want to reconstruct the dynamical behaviour from data),
3. a kernel k (for example the Gauß kernel $k(x, y) = e^{-x^T y}$),
4. data $x_0 = x(t_0), \ldots, x_N = x(t_N) \in \mathbb{R}^n$

The used concepts and advantages

1. Koopman operator

The used concepts and advantages

1. Koopman operator
2. Its generator

The used concepts and advantages

1. Koopman operator
2. Its generator
3. **R**eproducing **K**ernel **H**ilbert **spaces** (RKHS)

The used concepts and advantages

1. Koopman operator
2. Its generator
3. **R**eproducing **K**ernel **H**ilbert **spaces** (RKHS)
   $\rightsquigarrow$ continuous point evaluation + Hilbert spaces

The used concepts and advantages

1. Koopman operator
2. Its generator
3. **R**eproducing **K**ernel **H**ilbert **spaces** (RKHS)
   $\rightsquigarrow$ continuous point evaluation + Hilbert spaces

Advantages

1. Koopman operators fit well with data

The used concepts and advantages

1. Koopman operator
2. Its generator
3. **R**eproducing **K**ernel **H**ilbert **spaces** (RKHS)
   $\rightsquigarrow$ continuous point evaluation + Hilbert spaces

Advantages

1. Koopman operators fit well with data
2. Working with the Koopman operator directly typically requires equidistance sampling times, i.e. knowledge/data of the form $x(t_0 + k \cdot \Delta t)$ for $0 \leq k \leq N$. Working with the generator allows non-uniform sampling data as well.

The used concepts and advantages

1. Koopman operator
2. Its generator
3. **R**eproducing **K**ernel **H**ilbert **spaces** (RKHS)
   $\rightsquigarrow$ continuous point evaluation + Hilbert spaces

Advantages

1. Koopman operators fit well with data
2. Working with the Koopman operator directly typically requires equidistance sampling times, i.e. knowledge/data of the form $x(t_0 + k \cdot \Delta t)$ for $0 \leq k \leq N$. Working with the generator allows non-uniform sampling data as well.
3. Doesn't require the system to be forward complete

The used concepts and advantages

1. Koopman operator
2. Its generator
3. **R**eproducing **K**ernel **H**ilbert **spaces** (RKHS)
   $\rightsquigarrow$ continuous point evaluation + Hilbert spaces

Advantages

1. Koopman operators fit well with data
2. Working with the Koopman operator directly typically requires equidistance sampling times, i.e. knowledge/data of the form $x(t_0 + k \cdot \Delta t)$ for $0 \leq k \leq N$. Working with the generator allows non-uniform sampling data as well.
3. Doesn't require the system to be forward complete
4. Hilbert spaces are nice, but typically $\mathrm{L}^2(\mu)$ does not enjoy continuous point evaluation $\mathrm{L}^2 \ni g \mapsto g(x)$
   $\rightsquigarrow$ (well defined) explicit calculations

The Koopman operator

Let X be some (topological,compact,...) space with dynamics $\varphi_t : X \to X$ for $t \in \mathbb{R}_+$ on X. Let L be some function space $(\mathcal{C}(X), \mathrm{L}^2(X, \mu)$,...).

The Koopman operator

Let X be some (topological,compact,...) space with dynamics $\varphi_t : X \to X$ for $t \in \mathbb{R}_+$ on X. Let L be some function space $(\mathcal{C}(X), \mathrm{L}^2(X,\mu)$,...). Then the Koopman operator is a *linear* lift/representation of the dynamical system by

$$\begin{array}{ccc} X & \xrightarrow{(\varphi_t)_{t\in\mathbb{R}_+}} & X \\ \downarrow & \circlearrowleft & \downarrow \\ L & \xrightarrow{(T_t)_{t\in\mathbb{R}}} & L \end{array}$$

with $T_t g := g \circ \varphi_t$.

Let X be some (topological,compact,...) space with dynamics $\varphi_t : X \to X$ for $t \in \mathbb{R}_+$ on X. Let L be some function space $(\mathcal{C}(X), \mathrm{L}^2(X,\mu)$,...). Then the Koopman operator is a *linear* lift/representation of the dynamical system by

$$\begin{array}{ccc} X & \xrightarrow{(\varphi_t)_{t\in\mathbb{R}_+}} & X \\ \psi\downarrow & \circlearrowleft & \downarrow\psi \\ L & \xrightarrow{(T_t)_{t\in\mathbb{R}}} & L \end{array}$$

with $T_t g := g \circ \varphi_t$.

Some properties of the Koopman operator

Let $L = \mathcal{C}(X)$ then

1. T_t is linear, bounded ($\|T_t\| = 1$), an algebrahomomorphism ($T_t(g \cdot h) = T_t g \cdot T_t h$)

Some properties of the Koopman operator

Let $L = \mathcal{C}(X)$ then

1. T_t is linear, bounded ($\| T_t \| = 1$), an algebrahomomorphism ($T_t(g \cdot h) = T_t g \cdot T_t h$)
2. T_t is strongly continuous, i.e. $T_t g \to T_{t_0} g$ as $t \to t_0$ for all $g \in \mathcal{C}(X)$

Some properties of the Koopman operator

Let $L = \mathcal{C}(X)$ then

1. T_t is linear, bounded ($\|T_t\| = 1$), an algebrahomomorphism ($T_t(g \cdot h) = T_t g \cdot T_t h$)
2. T_t is strongly continuous, i.e. $T_t g \to T_{t_0} g$ as $t \to t_0$ for all $g \in \mathcal{C}(X)$
3. T_t is a semigroup, i.e. $T_{t+s} = T_t T_s$

Some properties of the Koopman operator

Let $L = \mathcal{C}(X)$ then

1. T_t is linear, bounded ($\|T_t\| = 1$), an algebrahomomorphism ($T_t(g \cdot h) = T_t g \cdot T_t h$)
2. T_t is strongly continuous, i.e. $T_t g \to T_{t_0} g$ as $t \to t_0$ for all $g \in \mathcal{C}(X)$
3. T_t is a semigroup, i.e. $T_{t+s} = T_t T_s$
   Hence "$T_t = e^{tA}$" for the generator $Ag = \nabla g \cdot f$

Linear methods and some occuring problems

Remark (Venga, Venga)

With a linear description come linear methods! As for example spectral theory :) which allows a qualitative analysis of the semigroup and its longtime behaviour.

Remark (Venga, Venga)

With a linear description come linear methods! As for example spectral theory :) which allows a qualitative analysis of the semigroup and its longtime behaviour.

Remark (Just a different perspective does not make the problem easier!)

We have the following difficulties on $L = \mathcal{C}(X)$*:*

Linear methods and some occuring problems

Remark (Venga, Venga)

With a linear description come linear methods! As for example spectral theory :) which allows a qualitative analysis of the semigroup and its longtime behaviour.

Remark (Just a different perspective does not make the problem easier!)

We have the following difficulties on $L = \mathcal{C}(X)$:

1. *The function space is typically infinite dimensional*
2. *The Koopman operator is not compact*
3. *The spectrum is given by $\sigma(T_t) = \overline{B_1(0)} \subset \mathbb{C}$*
4. *In general T_t is not a spectral operator, there does not exists a spectral representation of the form $T_t g = \int_{\sigma(T_t)} z \, dE(z) g$ for some spectral measure E.*

Applications

Koopman theory is widely applied, theoretically (ergodic theory) and for applications (Milan, Didier, Victor, Igor Mezic,...) and even *spectral results* exist (Milan).

Applications

Koopman theory is widely applied, theoretically (ergodic theory) and for applications (Milan, Didier, Victor, Igor Mezic,...) and even *spectral results* exist (Milan). **D**ynamic **M**ode **D**ecomposition (DMD) is a typical example: given snapshots $x_0 = x(t_0), x_1 = x(t_0 + \Delta t), \ldots, x_N = x(t_0 + N\Delta t)$. Find the **matrix** A such that

$$\left\| A \begin{pmatrix} x_0 & x_1 & \cdots & x_{N-1} \end{pmatrix} - \begin{pmatrix} x_{k+1} & x_{k+2} & \cdots & x_N \end{pmatrix} \right\|$$

is minimal.

Applications

Koopman theory is widely applied, theoretically (ergodic theory) and for applications (Milan, Didier, Victor, Igor Mezic,...) and even *spectral results* exist (Milan). **D**ynamic **M**ode **D**ecomposition (DMD) is a typical example: given snapshots $x_0 = x(t_0), x_1 = x(t_0 + \Delta t), \ldots, x_N = x(t_0 + N\Delta t)$. Find the **matrix** A such that

$$\left\| A \begin{pmatrix} x_0 & x_1 & \cdots & x_{N-1} \end{pmatrix} - \begin{pmatrix} x_{k+1} & x_{k+2} & \cdots & x_N \end{pmatrix} \right\|$$

is minimal. The singular values/vectors of A are called Koopman modes.

Remark

In true Koopman fashion this can be extended to consider $g_j(x_i)$ for some functions $g_1, \ldots, g_m$. This leads to **E***xtended DMD.*

Applications

Koopman theory is widely applied, theoretically (ergodic theory) and for applications (Milan, Didier, Victor, Igor Mezic,...) and even *spectral results* exist (Milan). **D**ynamic **M**ode **D**ecomposition (DMD) is a typical example: given snapshots $x_0 = x(t_0), x_1 = x(t_0 + \Delta t), \ldots, x_N = x(t_0 + N\Delta t)$. Find the **matrix** A such that

$$\left\| A \begin{pmatrix} x_0 & x_1 & \cdots & x_{N-1} \end{pmatrix} - \begin{pmatrix} x_{k+1} & x_{k+2} & \cdots & x_N \end{pmatrix} \right\|$$

is minimal. The singular values/vectors of A are called Koopman modes.

Remark

*In true Koopman fashion this can be extended to consider $g_j(x_i)$ for some functions $g_1, \ldots, g_m$. This leads to **E**xtended DMD. A natural setting to consider EDMD is on a RKHS!*

What is a RKHS?

Two perspectives

What is a RKHS?

Two perspectives

1. Hilbert space: a RKHS is a Hilbert space H of functions on X such that the point evaluation $g \mapsto g(x)$ is continuous for all $x \in X$.

What is a RKHS?

Two perspectives

1. Hilbert space: a RKHS is a Hilbert space H of functions on X such that the point evaluation $g \mapsto g(x)$ is continuous for all $x \in X$.
   By Riesz-Frechet for all $x \in X$ there exists $k_x \in H$ such that
   $$\langle g, k_x \rangle = g(x)$$
   for all $g \in H$.
2. positive semidefinite kernel functions: A function $k : X \times X \to \mathcal{C}$ is called a psd kernel if for all $x_1, \ldots, x_N \in X$ the matrix $(k(x_i, x_j))_{i,j=1}^N$ is psd.

Theorem (Aronszajn's theorem)

Both perspectives coincide (by $k(x, y) = k_x(y)$).

What is a RKHS?

Two perspectives

1. Hilbert space: a RKHS is a Hilbert space H of functions on X such that the point evaluation $g \mapsto g(x)$ is continuous for all $x \in X$.
   By Riesz-Frechet for all $x \in X$ there exists $k_x \in H$ such that
   $$\langle g, k_x \rangle = g(x)$$
   for all $g \in H$.
2. positive semidefinite kernel functions: A function $k : X \times X \to \mathcal{C}$ is called a psd kernel if for all $x_1, \dots, x_N \in X$ the matrix $(k(x_i, x_j))_{i,j=1}^N$ is psd.

Theorem (Aronszajn's theorem)

Both perspectives coincide (by $k(x, y) = k_x(y)$).

Example

$(\mathbb{R}^n, \langle \cdot, \cdot \rangle)$ with $X = \{1, \dots, n\}$, $H^1([0, 1])$ with $X = [0, 1]$, l^2 with $X = \mathbb{N}$.

RKHS and Koopman

RKHS are widely used in data sciences and neural networks due to the bounded point evaluation. But their explicit description is hard generally!

Remark

The Koopman operator (on a RKHS) has the form $T_t g = g \circ \varphi_t$*. For which* $g \in H$ *do we have* $T_t g = g \circ \varphi_t \in H$*, i.e. what is the domain of* T_t *on* H*?*

RKHS and Koopman

RKHS are widely used in data sciences and neural networks due to the bounded point evaluation. But their explicit description is hard generally!

Remark

The Koopman operator (on a RKHS) has the form $T_t g = g \circ \varphi_t$. For which $g \in H$ do we have $T_t g = g \circ \varphi_t \in H$, i.e. what is the domain of T_t on H? Result: The adjoint T_t^* is densely defined.

The generator and its adjoint

For the generator $Ag = \nabla g \cdot f$, the question whether $Ag \in H$ for $g \in H$ is also difficult.

The generator and its adjoint

For the generator $Ag = \nabla g \cdot f$, the question whether $Ag \in H$ for $g \in H$ is also difficult. But the adjoint has a good property, which is closely connected to occupation measures!

The generator and its adjoint

For the generator $Ag = \nabla g \cdot f$, the question whether $Ag \in H$ for $g \in H$ is also difficult. But the adjoint has a good property, which is closely connected to occupation measures!

Definition

Let $\gamma : [0, T] \to X$ be continuous and H be a RKHS on X. Then the map

$$g \mapsto \int_0^T g(\gamma(t))\, dt$$

is continuous on H and the corresponding Riesz functional is denoted by Γ_γ, i.e. $\langle g, \Gamma_\gamma \rangle_H = \int_0^T g(\gamma(t))\, dt$.

Lemma

Let $\dot{\gamma} = f(\gamma(t))$*. Then* $\Gamma_\gamma \in D(A^*)$ *and*

$$A^*\Gamma_\gamma = k(\gamma(T), \cdot) - k(\gamma(0), \cdot).$$

Lemma

Let $\dot{\gamma} = f(\gamma(t))$*. Then* $\Gamma_\gamma \in D(A^*)$ *and*

$$A^*\Gamma_\gamma = k(\gamma(T), \cdot) - k(\gamma(0), \cdot).$$

Proof.

For $g \in H$

$$\begin{aligned} \langle \Gamma_\gamma, Ag \rangle &= \int_0^T \nabla g \cdot f(\gamma)(t)\, dt = \int_0^T \frac{\mathrm{d}}{\mathrm{dt}} g(\gamma(t))\, dt \\ &= g(\gamma(T)) - g(\gamma(0)) \\ &\overset{RKHS}{=} \langle g, k(\gamma(T), \cdot)\rangle - \langle g, k(\gamma(0), \cdot)\rangle. \end{aligned}$$

□

How does it help?

How does it help?
This allows to "approximate" the generator (respectively) its adjoint from data.

1. Knowing snapshots $x_0, \ldots, x_N$ (not necessarily uniformly sampled in time!) and the kernel k we know the functions $k(x_i, \cdot)$.
2. The integral map Γ_γ can be approximated by quadrature:

$$\tilde{\Gamma}_\gamma(x) := \sum_{i=1}^{N}(t_i - t_{i-1})k(x, \gamma(t_i)).$$

3. We can solve for A^* in $A^*\tilde{\Gamma}_{\gamma_k} = k(\gamma(T_k), \cdot) - k(\gamma(0), \cdot)$ for $k = 0, \ldots, N$.

Remark

Approximating A^* by $\tilde{A}^*$, i.e. approximating A by $\tilde{A}$ gives approximations of "$T_t = e^{tA}$" by $e^{t\tilde{A}}$ and any eigenvector g of $\tilde{A}$ with eigenvalue λ gives $e^{t\tilde{A}}g = e^{\lambda t}g$.

Remark

Approximating A^* by $\tilde{A}^*$, i.e. approximating A by $\tilde{A}$ gives approximations of "$T_t = e^{tA}$" by $e^{t\tilde{A}}$ and any eigenvector g of $\tilde{A}$ with eigenvalue λ gives $e^{t\tilde{A}}g = e^{\lambda t}g$.

If it is possible to approximate $g = \mathrm{Id}$ by eigenfunctions, it is possible to predict the future of the system by "easier" dynamics.

Higher order ODEs

Setting: The same!

Higher order ODEs

Setting: The same! (That is only snapshots on the state space!)

Higher order ODEs

Setting: The same! (That is only snapshots on the state space!)
Additional knowledge: The system is driven by a higher order ODE (for instance acceleration systems from physics)

Higher order ODEs

Setting: The same! (That is only snapshots on the state space!)
Additional knowledge: The system is driven by a higher order ODE (for instance acceleration systems from physics)
Problem: Snapshots only on the state space - so typically information about the derivative is derived from numerical differentiation. $\rightsquigarrow$ not robust.

What do they do in this setting? Its similar

The procedure is as follows

1. Describe the Koopman operator and its generator
2. For higher order systems use higher order derivatives of $T_t g = g \circ \varphi_t$.

What do they do in this setting? Its similar

The procedure is as follows

1. Describe the Koopman operator and its generator
2. For higher order systems use higher order derivatives of $T_t g = g \circ \varphi_t$.For second order systems $\ddot{\gamma} = f(\gamma)$ consider the operator

$$Bg := \left(\frac{\mathrm{d}}{\mathrm{dt}}\right)^2 \bigg|_{t=0} g \circ \gamma = \dot{\gamma} D^2 g(\gamma(t))\dot{\gamma} + \nabla g(\gamma(t))\ddot{\gamma}(t).$$

What do they do in this setting? Its similar

The procedure is as follows

1. Describe the Koopman operator and its generator
2. For higher order systems use higher order derivatives of $T_t g = g \circ \varphi_t$.For second order systems $\ddot{\gamma} = f(\gamma)$ consider the operator

$$Bg := \left(\frac{\mathrm{d}}{\mathrm{dt}}\right)^2 \Big|_{t=0} g \circ \gamma = \dot{\gamma} D^2 g(\gamma(t))\dot{\gamma} + \nabla g(\gamma(t))\ddot{\gamma}(t).$$

In order to turn B into a well defined operator on a good space Rosenfeld et al. introduce vector values RKHS which consist of a triple of RKHS: A range space Y (for example $H^1(0,1)$), $\tilde{H} \subset \mathcal{C}^2(\mathbb{R}^n)$ such that for all $\theta \in \mathcal{C}^2([0, T], \mathbb{R}^n)$ the map $\tilde{h} \mapsto \tilde{h} \circ \theta$ is a bounded map from $\tilde{H}$ to Y. And a pull-back RKHS is considered.

What do they do in this setting? Its similar

The procedure is as follows

1. Describe the Koopman operator and its generator
2. For higher order systems use higher order derivatives of $T_t g = g \circ \varphi_t$. For second order systems $\ddot{\gamma} = f(\gamma)$ consider the operator

$$Bg := \left(\frac{\mathrm{d}}{\mathrm{dt}}\right)^2 \Big|_{t=0} g \circ \gamma = \dot{\gamma} D^2 g(\gamma(t))\dot{\gamma} + \nabla g(\gamma(t))\ddot{\gamma}(t).$$

3. Define/find (integration) functionals on this RKHS for which we know the action of B (and we can approximate those functionals from data)
4. Here those are given by Cauchy's formula for antiderivatives ($h'' \mapsto h$).

What do they do in this setting? Its similar

The procedure is as follows

1. Describe the Koopman operator and its generator
2. For higher order systems use higher order derivatives of $T_t g = g \circ \varphi_t$. For second order systems $\ddot{\gamma} = f(\gamma)$ consider the operator

$$Bg := \left(\frac{\mathrm{d}}{\mathrm{dt}}\right)^2 \bigg|_{t=0} g \circ \gamma = \dot{\gamma} D^2 g(\gamma(t))\dot{\gamma} + \nabla g(\gamma(t))\ddot{\gamma}(t).$$

3. Define/find (integration) functionals on this RKHS for which we know the action of B (and we can approximate those functionals from data)
4. Here those are given by Cauchy's formula for antiderivatives ($h'' \mapsto h$).
   Since B differentiates twice along γ, the fundamental theorem of calculus and integration by parts gives explicit expressions of the action of B on the second antiderivative operator.