PADERBORN UNIVERSITY

PADERBORN UNIVERSITY

EDMD FOR THE KOOPMAN OPERATOR AND GENERATOR (2)

Feliks Nüske

Magdeburg, 30 July, 2020

EDMD for Reversible Systems

gEDMD

Spectral Mapping Theorem

Proposition

Let κ_i be an eigenvalue of the generator $\mathcal{L}$. Then $e^{\kappa_i \tau}$ is an eigenvalue of the Koopman operator $\mathcal{K}^\tau$ for all lag times τ, with the same eigenfunctions.

- Since $\|\mathcal{K}^\tau\| = 1$, all eigenvalues of $\mathcal{L}$ must lie in the non-positive complex half-plane.
- For a reversible SDE, the spectrum of $\mathcal{L}$ is part of the non-positive real-half line.
- If, in addition, uniform ellipticity holds, we have a simple eigenvalue $\kappa_1 = 0$ and $\kappa_i < 0$ for all $i > 1$.

Pazy, *Semigroups of linear operators and applications to partial differential equations* (2012)

Spectral Decomposition

Consider the reversible generator $\mathcal{L}$. If the spectrum is fully discrete, we can expand the Koopman operator as

$$\begin{aligned}\mathcal{K}^\tau f &= \sum_{i=1}^{\infty} e^{-\kappa_i \tau} \langle f, \psi_i \rangle_\mu \psi_i \\ &= \langle f, 1 \rangle_\mu + \sum_{i=2}^{\infty} e^{-\kappa_i \tau} \langle f, \psi_i \rangle_\mu \psi_i.\end{aligned}$$

Spectral decomposition is related to convergence to equilibrium. If the system is metastable, there must be a cluster of eigenvalues of $\mathcal{L}$ close to zero.

Variational Principle for the Rayleigh Trace

Theorem

Let $\mathcal{K}$ be a self-adjoint operator on Hilbert space H with leading eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_M > R$, where R is an upper bound on the essential spectrum. Then

$$\sum_{k=1}^{M} \lambda_k = \sup_{\mathbb{V}:\dim(\mathbb{V})=M} \sum_{k=1}^{M} \langle \psi_k, \mathcal{K}\psi_k \rangle \, .$$

*The functions ψ_k form an orthonormal basis of the M-dimensional subspace $\mathbb{V}$, and the sum on the right-hand side is called the **Rayleigh trace**. The maximum is attained by choosing $\mathbb{V}$ spanned by the first M exact eigenfunctions.*

Variational Approximation of the Koopman Operator

Proposition

Let the reversible Koopman operator possess leading eigenvalues $1 = \lambda_1 > \lambda_2 \geq \ldots \geq \lambda_M > R$. *Let* $\mathbb{W}$ *be a space of* $N \geq M$ *linearly independent trial functions* ψ_i, $i = 1, \ldots, N$. *The coefficient vectors* v_m *of* $M \leq N$ *mutually orthonormal functions* $\psi_{v_m} \in \mathbb{W}$ *which maximize the Rayleigh trace restricted to* $\mathbb{W}$, *is given by the first M eigenvectors of the generalized eigenvalue problem*

$$Av_k = \hat{\lambda}_k G v_k, \qquad A_{ij} = \left\langle \psi_i, \mathcal{K}^\tau \psi_j \right\rangle_\mu, \qquad G_{ij} = \left\langle \psi_i, \psi_j \right\rangle_\mu.$$

Noé and Nüske, *Multiscale Model. Simul.* (2013)

Comments

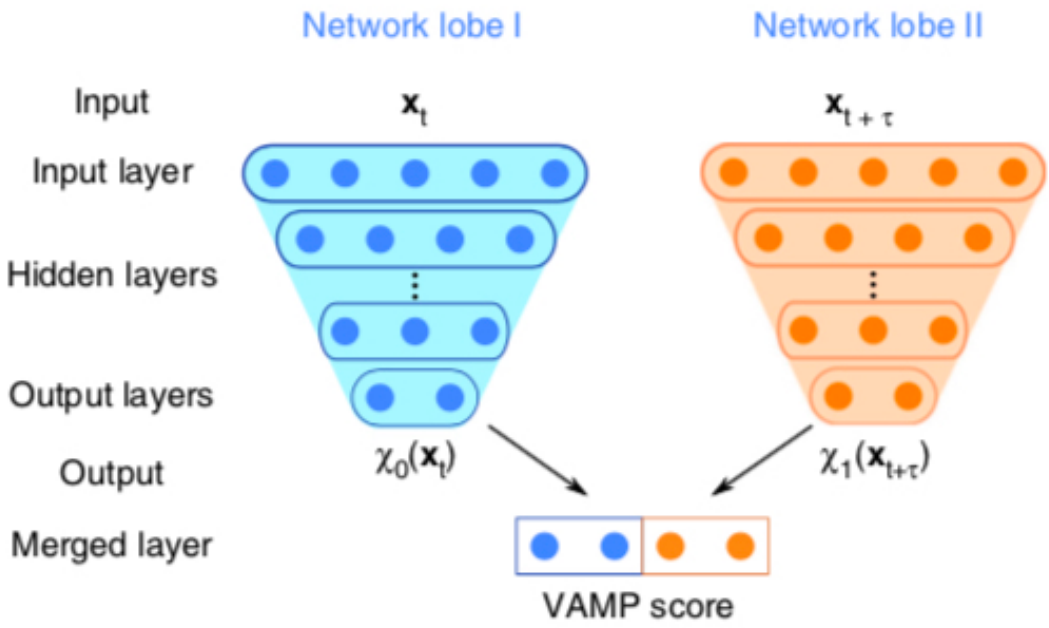
- This result can be generalized to non-reversible systems with compact Koopman operator, applying the variational principle to its SVD.
- We can define a "subspace score"

$$J(\mathbb{W}) = \max_{V \in \mathbb{R}^{N \times m}} \frac{V^T A(\mathbb{W}) V}{V^T G(\mathbb{W}) V},$$

  which can be compared and maximized across subspaces.
- Deep Learning and other ML techniques can be applied based on this score. This seems a very promising direction.
- Open problems: hyper-parameter tuning, interpretation, application to systems with more complex spectrum, ..

Example: VAMPnets

Model the trial space as a neural network and train the network to optimize the "subspace score" above (called VAMP score here):



Mardt et al, *Nature Communications* (2018)

References

Noé and Nüske, A Variational Approach to Modeling Slow Processes in Stochastic Dynamical Systems, *Multiscale Model. Simul.* (2013),

Klus, Nüske, et al, Data-Driven Model Reduction and Transfer Operator Approximation, *J. Nonlinear Sci.* (2018),

Mardt et al, VAMPnets for deep learning of molecular kinetics, *Nature Communications* (2018),

Wu and Noé, Variational Approach for Learning Markov Processes from Time Series Data, *J. Nonlinear Sci.* (2020)

gEDMD

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 $\mathbb{W} = \operatorname{span}\{\psi_i\}_{i=1}^n$:

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

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)

System Identification

Assume that $B \in \mathbb{R}^{N \times d}$ is the matrix such that $x =: g(x) = B^T \cdot \psi(x)$. Then the drift b can be identified via

$$b = (\mathcal{L}g) \approx (LB)^T \cdot \psi.$$

In addition to the drift term, we need to identify the diffusion term. Note that for $h_{ij}(x) = x_i \cdot x_j$, it holds that

$$(\mathcal{L}h_{ij}) = b_i \cdot x_j + b_j \cdot x_i + a_{ij}.$$

Since we already obtained a representation of b in the previous step, we can subtract the first two terms to obtain a_{ij}

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