Information Theoretic Discovery of Lagged Dynamic Mode Decomposition Models

1 Introduction

Equation free model development has enjoyed a years long stretch of continued progress by way of the evolution of dynamic mode decomposition (DMD) methods. Beginning in the fluid dynamics community, DMD has moved out into almost every area of data driven science, and it has merged itself with every major trend in the data sciences as well, in particular machine learning based methods.

Naturally multi-scale and delay structured models reminiscent of statistical AR models have also been explored; see [1, 2].

In some sense then, this work is the natural generaliztion of something like the Aikake Information Criteria used to develop parsimonius ARIMA models in statistics.

2 Entropic Regression DMD

The results in this work are a marriage between the higher-order DMD of [1], and the entropic regression technique for network detection and model construction found in [3, 4]. We now briefly explain both results, and then we show how to bring them together in order to build accurate, yet minimal time lag models.

2.1 Higher Order DMD

Originally presented in [1], though see also the HAVOK method in [5] and Hankel DMD method of [6], we generalize the results of [1] so as to keep each lagged model separate, causal, and formulated over an arbitrary of not necessarily unit incremement lags. We suppose that we have the maximum lag of d time steps. We likewise suppose that we have the choice of say N_l lags as $l_c = \{l_1, l_2, \cdots, l_k, \cdots, l_{N_l}\}$, with $1 = l_1 < l_j < l_{j+1} < l_{N_L} \leq d$. Given times series $\{\mathbf{y}_j\}_{j=0}^{N_T}$, we then define the matrices

$$\mathbf{Y}_{+,d} = (\mathbf{y}_d \cdots \mathbf{y}_{N_T}), \ \mathbf{Y}_- = (\mathbf{y}_0 \cdots \mathbf{y}_{N_T-1}), \ \mathbf{y}_j \in \mathbb{R}^s,$$

and the $s \times (N_T - d + 1)$ shifted mask matrices $\mathbf{M}_l$ where

$$(\mathbf{M}_l)_{mn} = \begin{cases} 0 & m-1 < d-l,\ m-1 > N_T - l \\ 1 & n = m,\ d-l \leq m-1 \leq N_T - l \\ 0 & n \neq m,\ d-l \leq m-1 \leq N_T - l \end{cases}$$

An arbitrary lagged DMD model can then be written as

$$\mathbf{Y}_{+,d} = \sum_{k=1}^{N_L} \mathbf{K}_{l_k} \mathbf{Y}_- \mathbf{M}_{l_k}$$

where we note that each matrix $\mathbf{K}_{l_k}$ is $s \times s$. Using our standard optimization arguments, we can find each matrix $\mathbf{K}_{l_j}$ via the critical-point equation

$$\sum_{k=1}^{N_L} \mathbf{K}_{l_k} \mathbf{Y}_- \mathbf{M}_{l_k} \mathbf{M}_{l_j}^T \mathbf{Y}_-^T = \mathbf{Y}_{+,d} \mathbf{M}_{l_j}^T \mathbf{Y}_-^T.$$

Combining these terms across lags leads to the system

$$\mathbf{K}(l_c) \mathbf{Y}_-(l_c) \mathbf{Y}_-(l_c)^T = \mathbf{Y}_{+,d} \mathbf{Y}_-(l_c)^T.$$

where

$$\mathbf{K}(l_c) = \left(\mathbf{K}_{l_{N_L}} \mathbf{K}_{l_{N_L - 1}} \cdots \mathbf{K}_1 \right),$$

and

$$\mathbf{Y}_-(l_c) = \begin{pmatrix} \mathbf{Y}_- \mathbf{M}_{l_{N_L}} \\ \mathbf{Y}_- \mathbf{M}_{l_{N_L}-1} \\ \vdots \\ \mathbf{Y}_- \mathbf{M}_1 \end{pmatrix}.$$

Note, in [1], models with distinct matrices $\mathbf{K}_{l_j}$ for each lag were eschewed for something much closer to what is now called Hankel DMD. While effective, and in several respects simpler, such an approach does not as readily allow for more thoughtful model selection as we explain in the next section.

That said, a price is paid for working with separate lag matrices when one wants to look at the corresponding one-step Koopman operator, say $\mathcal{K}^\delta$ which would be approximated by

$$\mathcal{K}^\delta \approx \mathbf{K}_a \equiv \begin{pmatrix} 0 & I & 0 & \cdots & 0 & 0 \\ 0 & 0 & I & \ddots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & I & 0 \\ \tilde{\mathbf{K}}_{l_{N_L}} & \tilde{\mathbf{K}}_{l_{N_L}-1} & \cdots & \cdots & \tilde{\mathbf{K}}_2 & \tilde{\mathbf{K}}_1 \end{pmatrix} \tag{1}$$

where each $s \times s$ matrix $\tilde{\mathbf{K}}_j$ is given by

$$\tilde{\mathbf{K}}_j = \begin{cases} \mathbf{K}_{l_k}, & j = l_k \in l_c \\ \mathbf{0}, & j \neq l_k \in l_c \end{cases}$$

Finding the eigenvalues of $\mathbf{K}_a$ is equivalent to finding the roots of the polynomial $p_a(z)$ where

$$p_a(z) = \det\left(\sum_{k=0}^{N_L - 1} \mathbf{K}_{l_{N_L - k}} \; z^k - z^{N_L} \mathbf{I} \right). \tag{2}$$

While not useful in a numerical context, this formula will prove useful when charaterizing the eigenvalues of $\mathbf{K}_a$ relative to the choice of lags l_c.

2.2 Determining Causality through Information Theory and Entropic Regression

Given two time series, say $\{X_j\}_{j=1}^{N_T}$ and $\{Y_j\}_{j=1}^{N_T}$, it is a foundational question to determine if one time series *causes* the other. Said another way, can we find quantitative methods which determine how one time series might drive or ultimately explain the behavior of another?

Motivated by the now celebrated *Granger causality* test, *cf.* [7], in linear time series, [8] introduced the notion of *transfer entropy* to determine the causal relationship between two nonlienar time series. The transfer entropy from X_j to Y_j, say $T_{X \to Y}(j)$ is defined in [8] to be

$$T_{X \to Y}(j) = I(Y_{j+1}, Y_j) - I(Y_{j+1}, Y_j | X_j),$$

where $I(,)$ is the mutual information between two random variables and $I(Y_{j+1}, Y_j | X_j)$ measures the conditional mutual information with conditioning over X_j. Note, if Y_{j+1} is independent of X_j, then $I(Y_{j+1}, Y_j | X_j) = I(Y_{j+1}, Y_j)$ so that $T_{X \to Y}(j) = 0$.

This initial concept of transfer entropy has given rise to a host of modifications and improvements, see in particular [9] and [3], which has ultimately lead to sophisticated software libraries being developed which can determine networks of interactions between time series that accurately account for confounding variables and non-Markovian influences of past states.

2.3 Entropic-Regression-Dynamic-Mode Decomposition

Thus, we can see how the ER approach can be used to enhance the HODMD. We again suppose that the time dynamics is modeled with at most $d-$lags as

$$\mathbf{y}_{j+d} = \sum_{l=1}^{d} \mathbf{K}_l \mathbf{y}_{j+d-l},$$

where $\mathbf{K}_l$ is $s \times s$ matrix describing the appropriate lag effect. To choose the appropriate lags in HODMD, we adapt the method of *entropic-regression* from [4]. There are three stages to the method. In the first we initialize our choice of lag and corresponding HODMD matrix. We then further build these lists through finding those laged models which provide the most information relative to the prior choice of laged model. Finally, we prune by testing whether each chosen lag is necessary relative to the other choices we have made during the build phase. Note, we always have $1 \in l_c$ since this makes all subsequent lags improvements on the basic DMD approach. This process is formalized and detailed as follows.

3 Results

To study the efficacy of our method, we examine its use over numerically generated data from various chaotic dynamical systems. In each case, for a given time series $\{\mathbf{y}_j\}_{j=0}^{N_T}$ we choose a maximum lag d and then use a derived model to reconstruct the original time series after the d^{th} step, i.e. $\{\mathbf{y}_j\}_{j=d}^{N_T}$. In each case, our models are from our proposed ERDMD method and the HODMD method.

3.1 Lorenz-63

We now examine how well our method performs on the standard Lorenz-63 system, given by the equations

$$\begin{aligned}
\dot{y}_1 &= \sigma(y_2 - y_1) \\
\dot{y}_2 &= y_1(\rho - y_3) - y_2 \\
\dot{y}_3 &= y_1 y_2 - \beta y_3
\end{aligned}$$

with $\sigma = 10$, $\rho = 28$, and $\beta = 8/3$. These parameter choices ensure that trajectories are pulled onto a strange attractor and exhibit chaotic dynamics. We test our method on numerically generated data. Time stepping is done with standard Runge–Kutta 4.For our numerics, the time step is $dt = .01$, and we run the simulation from $0 \leq t \leq 22$.

We look at data for $20 \leq t \leq 22$, with $d = 150$, corresponding to 1.5 units of non-dimensional time. We then compare our ER-DMD model to a model using all possible lags for $21.5 \leq t \leq 22$. The ERDMD algorithm converges to $l_c = \{1, 149\}$. The results are seen in Figure 1. As can be seen, all three curves are indistinguishable to the eye, showing that a far more minimal lag model is able to reconstruct dynamics with as much practical accuracy as the full HODMD method.

Constructing the full Koopman one-step approximation $\mathbf{K}_a$ as in Equation (1) allows us to find the affiliated Koopman spectrum as seen in the left

Algorithm 1 ERDMD Method

procedure INITIALIZE
  Set $l_c = \{1\}$ and $l_r = \{2, \cdots, d\}$.
  Find $\mathbf{K}_1 = \arg\min_{\mathbf{K}} \left|\left|\mathbf{Y}_1^+ - \mathbf{K}\mathbf{Y}_-(l_c)\right|\right|_F$. Set $\mathbf{K}(l_c) = (\mathbf{K}_1)$.
end procedure
procedure BUILD
  while $l_r \neq \{\varnothing\}$ **do**
    Given $l_c = \{1\ l_1 \cdots l_j\}$, $\mathbf{K}(\mathbf{l}_c)$, and $l_r = \{2, \cdots, d\} \setminus l_c$
    for $l_{j+1} \in l_r$ **do**
      Define $l_t = l_c \cup \{l_{j+1}\}$ and find

$$\mathbf{K}(l_t) = \arg\min_{\tilde{\mathbf{K}}_\mathbf{1}, \tilde{\mathbf{K}}_{\mathbf{l_1}}, \cdots \tilde{\mathbf{K}}_{\mathbf{l_{j+1}}}} \left|\left|\mathbf{Y}_{l_{j+1}}^+ - \tilde{\mathbf{K}}(l_t)\mathbf{Y}_-(l_t)\right|\right|_F.$$

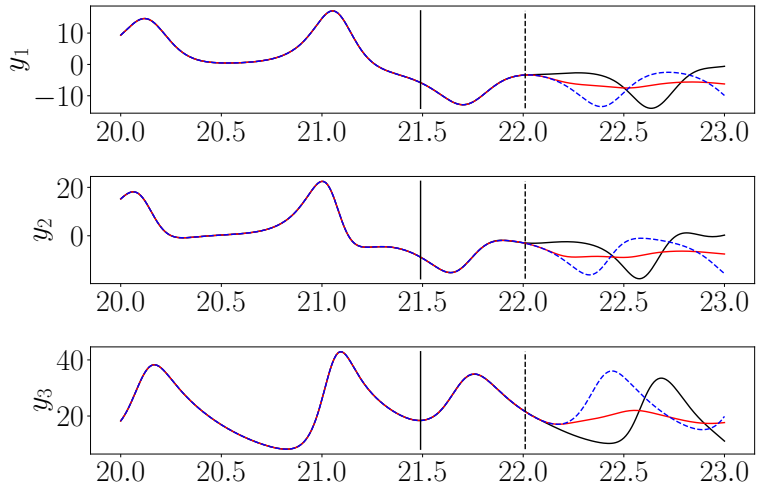
    end for
    Choose l_{j+1} and the corresponding l_t and $\mathbf{K}(l_t)$ to maximize

$$I\left(\mathbf{Y}_d^+, \mathbf{K}(l_t)\mathbf{Y}_-(l_t) \middle| \mathbf{K}(l_c)\mathbf{Y}_-(l_t)\right)$$

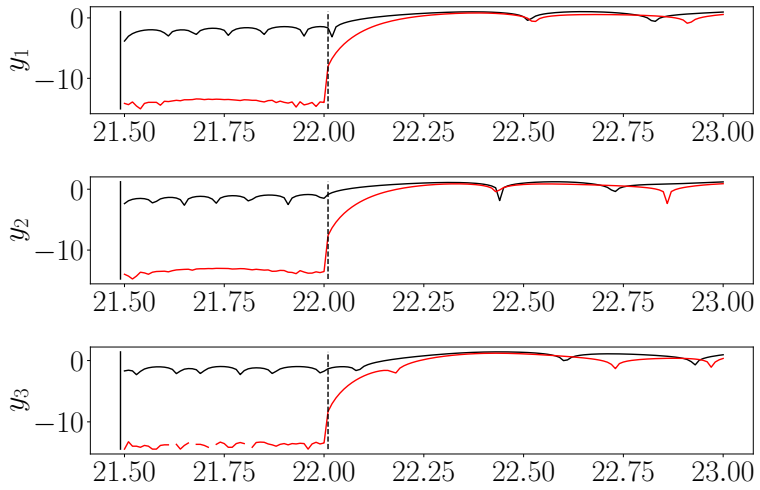
    If choice is statistically significant, update l_c and $\mathbf{K}(l_c)$.
  end while
end procedure
procedure PRUNE
  Given $l_c = \{1, l_1, \cdots, l_{N_L}\}$, set $S \equiv$ True
  while S **do**
    for $l_j \in l_c$ **do**
      Define $l_t = \{1, l_1, \cdots, l_{N_L}\} \setminus \{l_j\}$
      Compute

$$I\left(\mathbf{Y}_{l_{N_L}}^+, \mathbf{K}(l_t)\mathbf{Y}_-(l_t) \middle| \mathbf{K}(l_c)\mathbf{Y}_-(l_t)\right).$$

    end for
    Choose l_t corresponding to minimum information.
    if minimum is statistically insignificant **then**
      Prune corresponding l_j from l_c and $\mathbf{K}(l_c)$.
    else
      $S \equiv$ False
    end if
  end while
end procedure



(a) Trajectories



(b) Error Comparison (semi-log scale)

Figure 1: Direct comparison of ERDMD and all lags HODMD model against the true trajectory for the Lorenz-63 system (a), and error across dimensions for the ERDMD and all lags HODMD method (b). The black line indicates the ERDMD result while the red indicates the all lags HODMD result. The solid vertial bar indicates the maximum lag choice of $d = 150$, while the dashed line indicates the end of the reconstruction interval and the beginning of the forecasting regime. The ERDMD algorithm converges to $l_c = \{1, 149\}$.

side of Figure 2. We see that because of the strong separation in lags as seen in l_c that the affiliated characteristic polynomial $p_a(z)$ is given explicitly by

$$p_a(z) = \det\left(\mathbf{K}_{149} + (\mathbf{K}_1 - z\mathbf{I})\, z^{149}\right).$$

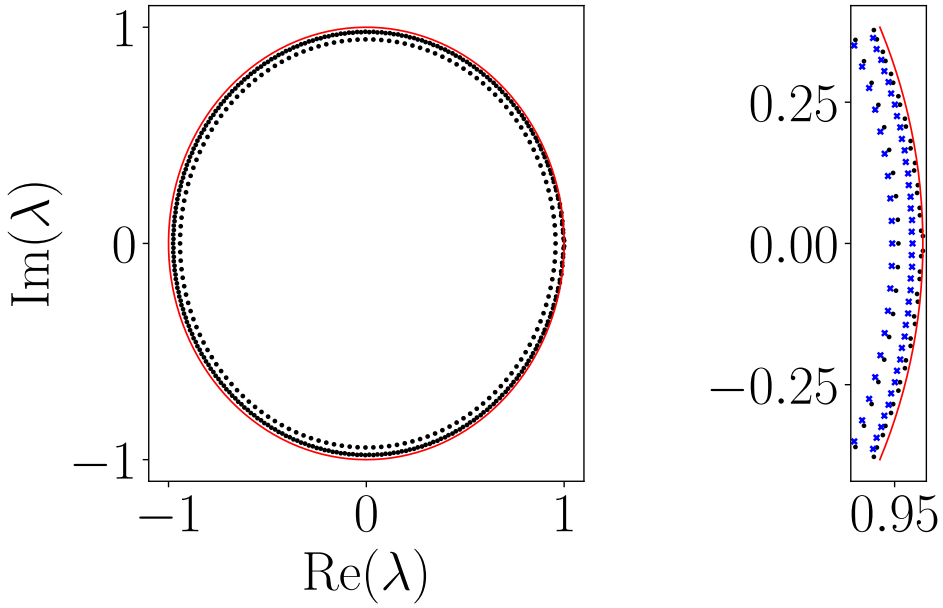


Figure 2: Spectrum of corresponding ERDMD Koopman operator for Lorenz-63 with $d = 150$ on the left side, with a detail comparison to the reduced approximation (blue crosses) on the right side near (1,0). The ERDMD algorithm converges to $l_c = \{1, 149\}$. The solid/red line is the unit circle, provided for reference.

If we look at the interior of the unit disc so that $|z| < 1$, then our innermost eigenvalues to leading order are found from the roots of $\tilde{p}_a(z)$ where

$$\tilde{p}_a(z) = \det\left(\mathbf{K}_{149} + \mathbf{K}_1 z^{149}\right)$$

or the leading roots are found by finding the generalized eigenvalues $\tilde{z} = z^{149}$ of the two matrix problem $\mathbf{K}_{149} + \mathbf{K}_1\tilde{z}$. Looking at this spectrum in Figure 2 (b), we see that most of the features in the spectrum seen in the full spectrum on the left are present in the right. Thus, the damping in the dynamics comes almost entirely from the disparity in lag values. Looking at Figure 2 (c), which shows a detailed comparison of the two spectra for complex phase ϕ such that $|\phi| < \frac{\pi}{8}$, we see that the full spectrum gets closer to the unit circle and even has two modes which just cross the unit circle. Otherwise, the reduced model well describes the damping modes, though we also see that the lag matrices $\mathbf{K}_1$ and $\mathbf{K}_{149}$ balance for more delicate dynamics along the unit circle.

By way of contrast though, if we set the maximum lag $d = 100$, and build reconstructions for $21 \leq t \leq 22$, we find that the ERDMD results differ markedly in terms of the determined lags though ultimately not in terms of accuracy; see Figure 3. In this case, the ERDMD algorithm converges onto

the lag choices

$$l_c = \{1, 15, 26, 35, 45, 48, 68, 73, 97, 99\}\,.$$

By choosing a lag horizon which does not fully capture the approximate period of oscillation in the dynamics, we need significantly more information to accurately reconstruct the data, though still nowhere as much as the full HODMD model.

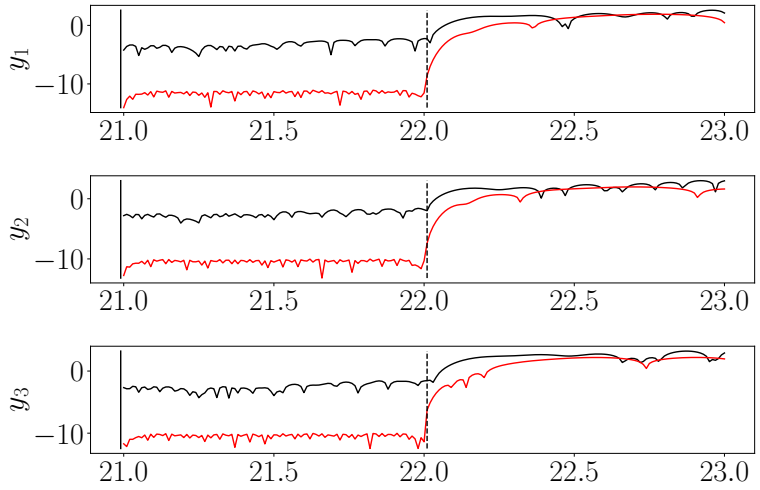


Figure 3: Comparison of ERDMD and all lags HODMD model against true trajectory for Lorenz-63 system. The ERDMD reconstruction is in solid black in the figure. The vertial bar indicates the maximum lag choice of $d = 100$. The ERDMD algorithm converges to $l_c = \{1, 15, 26, 35, 45, 48, 68, 73, 97, 99\}$.

We can likewise find the spectrum of $\mathbf{K}_a$ as seen in Figure 4. In this case, the larger spread of lags makes ready identification of positions in the spectrum to lag structure more difficult, though we see that The more uniform spread in lag values makes any estimates of the spectrum using reduced models less useful beyond identifying the most strongly damped modes.

Rossler Equation

To see how the ERDMD method works on problems with multiple scales, we now look at modeling dynamics coming from the Rossler system, given

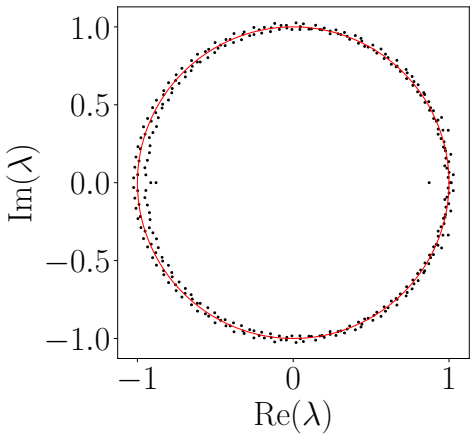


Figure 4: Spectrum of corresponding ERDMD Koopman operator for Lorenz-63 for $d = 100$. The ERDMD algorithm converges to $l_c = \{1, 15, 26, 35, 45, 48, 68, 73, 97, 99\}$. The solid/red line is the unit circle, provided for comparison.

by the equations

$$\begin{aligned}\dot{y}_1 &= -y_2 - y_3\\ \dot{y}_2 &= y_1 + a y_2\\ \dot{y}_3 &= b + y_3(y_1 - c)\end{aligned}$$

where $a = .1$, $b = .1$, $c = 14$. To see the role the multiple time scales play in this proble, letting $\epsilon = .1$, then we see $a = b = \epsilon$ and $c = 1.4/\epsilon$. Letting $\tau = t/\epsilon$ and setting $y_3 = \epsilon^2 \tilde{y}_3(t, \tau)$, to leading order, in y_1 and y_2 we find

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = e^{\epsilon t/2} \begin{pmatrix} \cos(t) & -\sin(t) \\ \sin(t) & \cos(t) \end{pmatrix} \begin{pmatrix} y_{1,0} \\ y_{2,0} \end{pmatrix} + \mathcal{O}(\epsilon^2),$$

so that we have $\mathcal{O}(1)$ planar oscillations complimented by slow growth away from the origin. Likewise, in $\tilde{y}_3$ we find

$$\partial_\tau \tilde{y}_3 + \epsilon \partial_t \tilde{y}_3 = 1 - 1.4\tilde{y}_3 + \epsilon y_1 \tilde{y}_3.$$

This then motivates the expansion

$$\tilde{y}_3(t, \tau) = c(t) e^{-1.4\tau} + \frac{1}{1.4}\left(1 - e^{-1.4\tau}\right) + \epsilon \tilde{y}_{3,1}(t, \tau) + \cdots,$$

which, to remove secularities then gets us the equation for $c(t)$ of the form

$$\frac{dc}{dt} = y_1(t)\left(c - \frac{1}{1.4}\right).$$

So we have

$$c(t) = \frac{1}{1.4} + \left(c(0) - \frac{1}{1.4} \right) e^{\int_0^t y_1(s) ds},$$

so that $\tilde{y}_3$ stays small until the slow growth in y_1 due to the $e^{\epsilon t/2}$ term pushes the dynamics out of the plane.

We build reconstructions for $25 \leq t \leq 40$. To get accurate results, we chose $d = 1000$, for which choice the method converged to the set l_c with

$$l_c = \{1, 3, 170, 436, 553, 665, 988\}.$$

As can be seen in Figure 5, our accuracy relies on capturing a full departure from the fast through the slow manifold. This also explains the need for such a large choice of d relative to what was used for the Lorenz-63 system.

We likewise can find the affiliated spectrum as seen in Figure 6.

Kuramoto–Sivashinsky Equation

To look at a more intricate and higher dimensional example, we now study the Kuramoto–Sivashinsky (KS) equation given by

$$u_t + u_{xx} + u_{xxxx} + uu_x = 0, \ u(x + 2L, t) = u(x, t).$$

See [10] for an extensive bibliography with regards to details and relavent proofs of facts used in this paper. Introducing the rescalings

$$\tilde{t} = \frac{t}{T}, \ \tilde{x} = \frac{\pi}{L} x, \ u = A\tilde{u},$$

and taking the balances

$$A = \frac{L}{\pi T}, \ T = \left(\frac{L}{\pi} \right)^2,$$

we get the equivalent KS equation (dropping tildes for ease of reading)

$$u_t + u_{xx} + \nu u_{xxxx} + uu_x = 0, \ \nu = \left(\frac{\pi}{L} \right)^2.$$

Looking at the linearized dispersion relationship $\omega(k) = k^2 - \nu k^4$, we see that the ν parameter acts as a viscous damping term. Thus, as the system size L is increased, the effective viscosity is decreased, thereby allowing for more complex dynamics to emerge. As is now well known, for L sufficiently large, a fractional-dimensional-strange attractor forms which both produces intricate spatio-temporal dynamics while also allowing for a far simpler representation of said dynamics. It is has been shown in many different works (see for example [11]) that $L = 11$ generates a strange attractor with dimension between eight and nine, and that this is about the smallest value of L

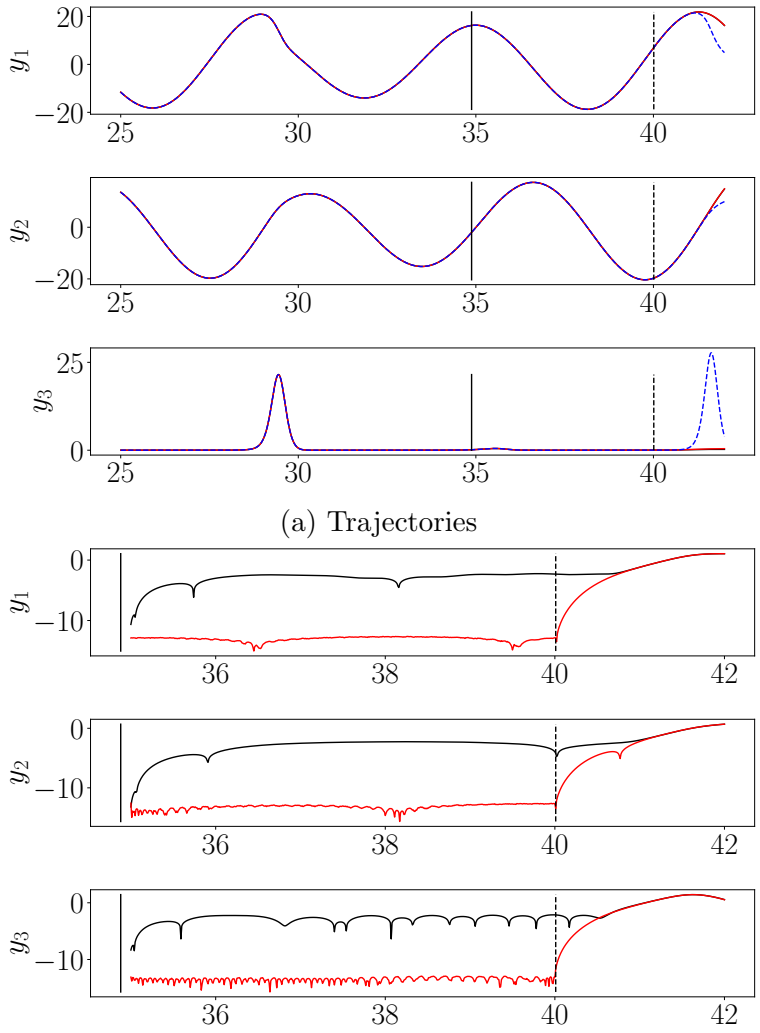


Figure 5: Comparison of ERDMD and all lags HODMD model against true trajectory for the Rossler system. The ERDMD reconstruction is in solid black in the figure. The vertial bar indicates the maximum lag choice of $d = 1000$. The ERDMD algorithm converges to $l_c = \{1, 3, 170, 436, 553, 665, 988\}$.

which is guaranteed to generate chaotic dynamics. We therefore set $L = 11$ throughout the remainder of this section.

To study ERDMD on the KS equation, we use KS data numerically generated by a pseudo-spectral in space and fourth-order exponential-differencing Runge-Kutta in time method [12] of lines approach. For the pseudo-spectral method, $K = 128$ total modes are used giving an effective spatial mesh width

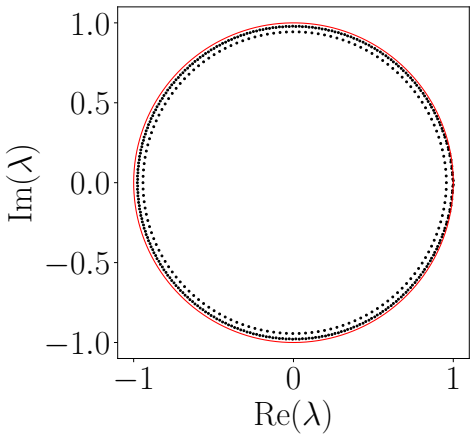


Figure 6: Spectrum of corresponding ERDMD Koopman operator for the Rossler system. The ERDMD algorithm converges to $l_c = \{1, 3, 170, 436, 553, 665, 988\}$. The solid/red line is the unit circle, provided for comparison.

of $2L/K = .172$, while the time step for the Runge-Kutta scheme is set to $\delta t = .25$. After a burn in time of $t_b = 10$, we generated a simulation of length $t_f = (L/\pi)^4 \approx 160$ to allow for nonlinear effects to fully manifest. This trajectory was then separated via a POD into space and time modes; see [13]. Taking $N_s = 12$ modes captured 98.6% of the total energy.

To study the ERDMD method, we choose $d = 200$, which for $dt = .25$ corresponds to a lag time of $t = 60$. With this choice, the ERDMD method finds l_c to be

$$l_c = \{1, 123, 141, 158\}.$$

The results for reconstruction can be seen in Figure 7, we where we look at times $10 \leq t \leq 66$. As can be seen, the ERDMD does quite well, though we note that if we look for longer reconstructions, errors do start to appear more rapidly for the ERDMD method than the full HODMD method.

We see in Figure 8 (a) the corresponding full spectrum affiliated with our method. Again, the strong separation of lags in l_c motivate looking at the roots of the reduced polynomial

$$\tilde{p}_a(z) = \det\left(z^{35}\mathbf{K}_{123} + z^{17}\mathbf{K}_{141} + \mathbf{K}_{158}\right).$$

Note, letting $\tilde{z} = z^{17}$, we can write

$$z^{35} = \tilde{z}^2 e^{\ln \tilde{z}/17} \approx \tilde{z}^2$$

so long as $|z|$ is not too close to zero. We can then find the affiliated approxmiate companion matrix representation for $\tilde{p}_a(z)$ in the form

$$\tilde{\mathbf{K}}_a = \begin{pmatrix} 0 & I \\ -\tilde{\mathbf{K}}_2 & -\tilde{\mathbf{K}}_1 \end{pmatrix}$$

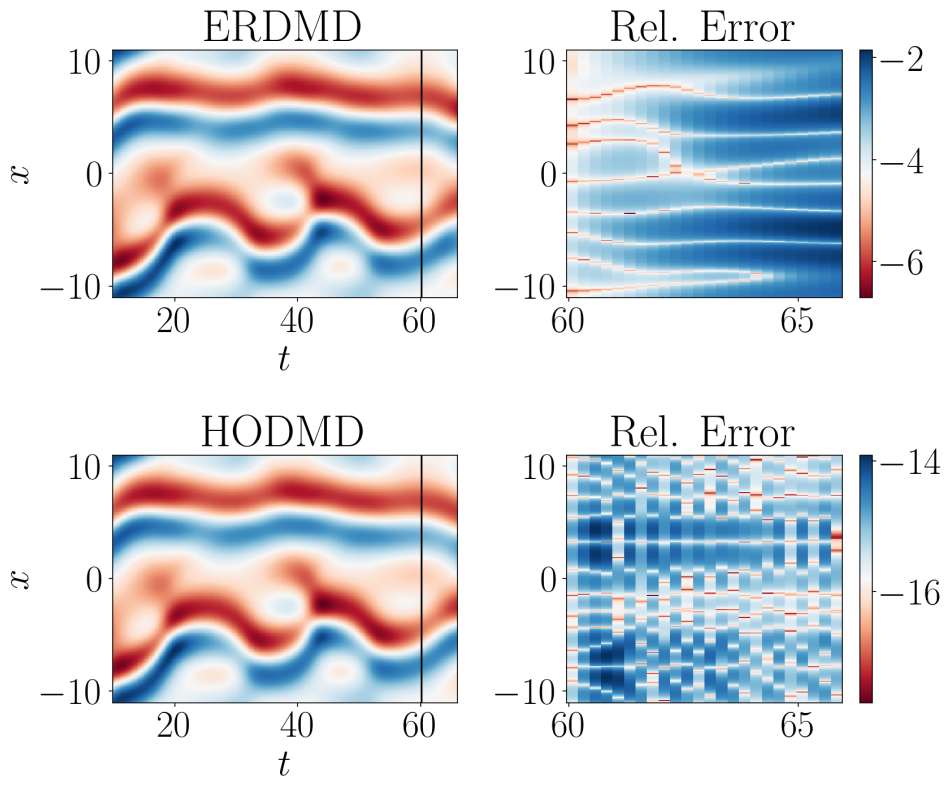


Figure 7: Comparison of ERDMD and all lags HODMD model against true trajectory for the KS equation. The vertial bar indicates the maximum lag choice of $d = 200$. The ERDMD algorithm converges to $l_c = \{1, 123, 141, 158\}$.

where

$$\tilde{\mathbf{K}}_1 = \mathbf{K}_{123}^{-1}\mathbf{K}_{141},\ \tilde{\mathbf{K}}_2 = \mathbf{K}_{123}^{-1}\mathbf{K}_{158}.$$

The reduced spectrum is seen in Figure 8 (b), where we again can see the most damped modes come from the highest lags in our Koopman model. We can also look at the case of $|z| > 1$, in which case our asymptotic arguments would lead us to just look at the largest eigenvalues of $\mathbf{K}_1$. These results are compared in Figure 8 (c), where we see the most unstable modes of the full spectrum correspond to the largest magnitude eigenvalues of $\mathbf{K}_1$. Thus there is a clear separation in the Koopman spectrum across lag values, which is to say time scales in the dynamics.

References

[1] Soledad Le Clainche and José M. Vega. Higher order dynamic mode decomposition. *SIAM Journal on Applied Dynamical Systems*, 16(2):882–

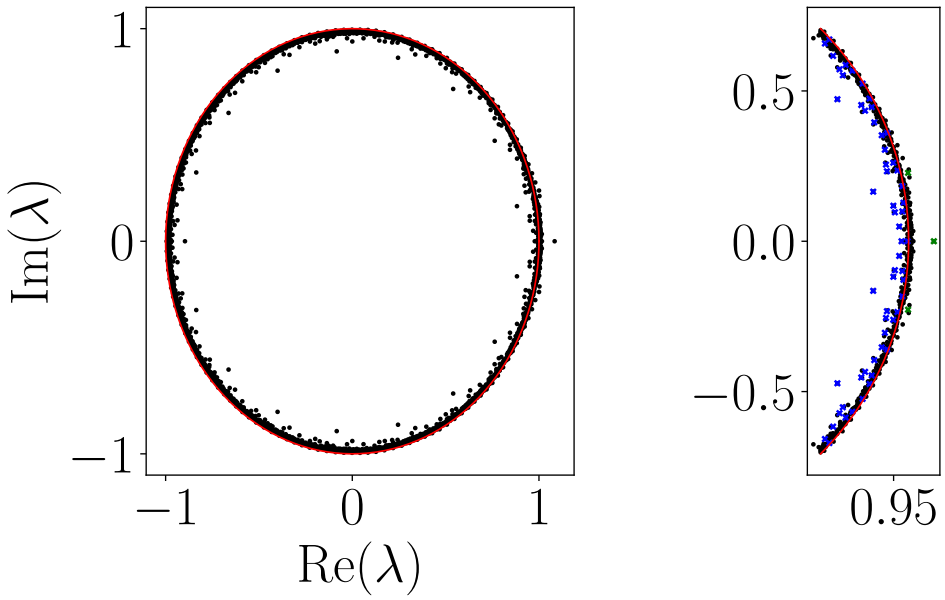


Figure 8: Full spectrum (a) and reduced spectrum (b) of corresponding ERDMD Koopman operator for Kuramoto–Sivashinsky. The ERDMD algorithm converges to $l_c = \{1, 123, 141, 158\}$. The solid/red line is the unit circle, provided for comparison. In (c), we compare the full spectrum (black/dots) to the interior reduced approximation (blue/x) and the outer reduced approximation (green/x) in a region with angular apperature of $\pi/8$.

925, 2017.

[2] Daniel J. Alford-Lago, Christopher W. Curtis, Alexander T. Ihler, and Katherine A. Zawdie. Scale-separated dynamic mode decomposition and ionospheric forecasting. *Radio Science*, 58(8):e2022RS007637, 2023.

[3] Jie Sun, Dane Taylor, and Erik M. Bollt. Causal network inference by optimal causation entropy. *SIAM Journal on Applied Dynamical Systems*, 14(1):73–106, 2015.

[4] A. A. R. AlMomani, Jie Sun, and Erik M. Bollt. How entropic regression beats the outliers problem in nonlinear system identification. *Chaos*, 30:013107, 2020.

[5] S.L. Brunton, B.W. Brunton, J.L. Proctor, , E. Kaiser, and J.N. Kutz. Chaos as an intermittently forced linear system. *Nature Comm.*, 8, 2017.

[6] H. Arbabi and I. Mezic. Ergodic theory, dynamic mode decomposition, and computation of spectral properties of the Koopman operator. *SIAM Appl. Dyn. Sys.*, 16:2096–2126, 2017.

[7] C. W. J. Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37(3):424–438, 1969.

[8] Thomas Schreiber. Measuring information transfer. *Phys. Rev. Lett.*, 85:461–464, Jul 2000.

[9] Luca Faes, Giandomenico Nollo, and Alberto Porta. Information-based detection of nonlinear granger causality in multivariate processes via a nonuniform embedding technique. *Phys. Rev. E*, 83:051112, May 2011.

[10] J.C. Robinson. *Infinite Dimensional Dynamical Systems.* Cambridge University Press, Cambridge, UK, 2001.

[11] N. B. Budanur, P. Cvitanović, R. L. Davidchack, and E. Siminos. Reduction of so(2) symmetry for spatially extended dynamical systems. *Phys. Rev. Lett.*, 114:084102, Feb 2015.

[12] AK Kassam and L.N. Trefethen. Fourth-order time-stepping for stiff PDEs. *SIAM J. Sci. Comp.*, 26:1214–1233, 2005.

[13] G. Berkooz, P. Holmes, J.L. Lumley, and C.W. Rowley. *Turbulence, Coherent Structures, Dynamical Systems, and Symmetry.* Cambridge University Press, Cambridge, UK, 2012.