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 , we generalize the results of [1] so as to keep each lagged model separate, causal, and formulated over an arbitrary of not necessarily unit increment 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 $\mathbf{l}_c = (l_1, l_2, \cdots, l_k, \cdots, l_{N_l})$, 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

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

where

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

and

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

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.

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.* [5], in linear time series, [6] 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 [6] 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 [7] 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]. This begins with finding the one lag matrix $\mathbf{K}_1$. We then define the choice of lag matrices, say $\mathbf{K}_c$ as $\mathbf{K}_1$ and the lag choices via the vector $l_c = (1)$. We then:

-

We then look for a lag $l_* > 1$ and corresponding matrices $\mathbf{K}_1$ and $\mathbf{K}_{l_*}$ such that

$1, l_*, \mathbf{K}_1, \mathbf{K}_{l_*} = \arg \max_{2 \le l \le d} I\left(\mathbf{y}(t_{j+l}), \mathbf{K}_1 \mathbf{y}(t_{j+l-1}) + \mathbf{K}_l \mathbf{s}(t_j) | \mathbf{K}_p \mathbf{y}(t_{j+l-1})\right).$

If the corresponding information is statistically significant (assessed through a straightforward hypothesis test), then we include the lag and corresponding lag matrix in our HODMD. This process is repeated until there are no remaining significant lags. In turn then, we test whether each included lag is still signficant if it is ignored so that we prune our initial list of chosen lags. As in [3], this process creates a set of lags which removes intermediate or confounding stastical relationships, producing the most meaningful list of essentially causitive lags within our HODMD framework. We label this method the Entropic-Regression DMD (ERDMD). The details of the algorithm are explained in the Appendix.

Again, collecting the EWT scales into the vector $\mathbf{s}(t)$, our HODMD model is trying to determine those lags to use in the $d-$lag model

$$\mathbf{s}\left(t_{j+d}\right) = \sum_{l=1}^{d} \mathbf{K}_l \mathbf{s}\left(t_{j+d-l}\right),$$

where d is a maximum value chosen by the user. We suppose that our sampled time series is given by $\{\mathbf{s}_j\}_{j=0}^{N_T}$ where $\mathbf{s}_j = \mathbf{s}(t_j)$. To begin the ER-DMD algorithm, we first find

$$\mathbf{K}_1 = \arg\min_{\mathbf{K}} \left\| \mathbf{S}_1^+ - \mathbf{K}\mathbf{S}_{1,1}^- \right\|_F$$

where for lag l we define the matrices $\mathbf{S}_{j,l}^-$ and $\mathbf{S}_j^+$ to be

$$\mathbf{S}_{j,k}^- = \left(\mathbf{s}_{j-k} \ \mathbf{s}_{j-k+1} \ \cdots \mathbf{s}_{N_T-k}\right), \ \mathbf{S}_j^+ = \left(\mathbf{s}_j \ \mathbf{s}_{j+1} \ \cdots \mathbf{s}_{N_T}\right),$$

and where $\|\cdot\|_F$ is the Frobenius norm. Taking a greedy-algorithm approach, we then fix $\mathbf{K}_1$ in all subsequent computations.

At the next step, we then find any given next choice for a lag matrix $\mathbf{K}_l$ for $2 \leq l \leq d$ via

$$\mathbf{K}_l = \arg\min_{\mathbf{K}} \left\| \mathbf{S}_l^+ - \mathbf{K}_1 \mathbf{S}_{l,1}^- - \mathbf{K}\mathbf{S}_{l,l}^- \right\|_F$$

Defining I_l such that

$$I_l = I\left(\mathbf{S}_l^+, \mathbf{K}_1 \mathbf{S}_{l,1}^- + \mathbf{K}_l \mathbf{S}_{l,l}^- \,\middle|\, \mathbf{K}_1 \mathbf{S}_{l,1}^-\right),$$

we find $I_{l_*} = \max_{2 \leq l \leq d} I_l$ and then test if $I_{l_*} > 0$ using a hypothesis test based off of building null distributions using random shuffles of the time series.

Proceeding in this way, where we note that we fix each chosen lag matrix $\mathbf{K}_{l_j}$ before proceeding to find the next one, we can build a list of proposed lags, say $\ell_{pr} = (1, l_1, \cdots, l_c)$, and corresponding lag matrices $\mathbf{K}_{l_j}$, which we order such that $1 < l_j < l_{j+1} \leq d$. That said, building off the theorems and algorithms in [3] and [4], we also need to potentially prune our proposed lag list. To that end, we define $I_{\ell_{pr} \backslash l_k}$ where

$$I_{\ell_{pr} \backslash l_j} = I\left(\mathbf{S}_d^+, \sum_j \mathbf{K}_{l_j} \mathbf{S}_{d,l_j}^- \,\middle|\, \sum_{j \neq k} \mathbf{K}_{l_j} \mathbf{S}_{d,l_j}^-\right).$$

We then find $\tilde{I}_{j_*} = \min_j I_{\ell_{pr} \backslash l_j}$. If $\tilde{I}_j$ is not greater than zero in a statistical sense, then l_{j_*} and its corresponding matrix $\mathbf{K}_{l_{j_*}}$ are pruned. This process is then repeated until there are no lags remaining or all lags are statistically significant.

References

[1] Soledad Le Clainche and José M. Vega. Higher order dynamic mode decomposition. *SIAM Journal on Applied Dynamical Systems*, 16(2):882–925, 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] C. W. J. Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37(3):424–438, 1969.

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

[7] 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.