The Challenge of Small Data:
Dynamic Mode Decomposition, Redux

Amirhossein Karimi and Tryphon T. Georgiou

Abstract— We revisit the setting and the assumptions that underlie the methodology of Dynamic Mode Decomposition (DMD) in order to highlight caveats as well as potential measures of when the applicability is warranted.

I. INTRODUCTION

Whereas the topic of "big data" dominates current headlines in research publications and popular news analyses alike, the perennial challenge of obtaining reliable models with only limited observation records persists in a wide range of time series applications. Indeed, one often hears the admission from practitioners that the problem is not "big data" but "small data." A case in point is that of time series of flow fields where a exceedingly high-dimension state is observed, or partially observed, albeit over a relatively short time window. It is precisely for these types of applications that Dynamic Mode Decomposition (DMD) and related frameworks were conceived to address [1], [2], [3].

DMD, as introduced by Schmid [4], is a formalism to identify dominant modes in a high-dimensional time series $x_t \in \mathbb{R}^N$, $t \in \{1, 2, \ldots, L\}$, where the dimensionality N of the time series is much larger than the number L of available observations. In its original formulation, DMD takes x_t as a convenient state of an underlying process and thereby dispenses of higher order dynamics that may be hidden in differences between the time series data. The more general situation of higher order dynamics can be treated similarly [5]. The main issue that we discuss in this paper is the pertinence of the assumption in seeking such a state model, and whether a reliable estimate of state dynamics should be expected to reflect the structure of the data. We propose a certain geometric concept, the so-called gap metric, as a tool to provide guidance in selecting suitable dimension for the sought DMD dynamics.

II. THE BASIC DMD RATIONALE

Consider the basic linear dynamical model,

$$x_{t+1} = Ax_t + v_t, \text{ for } 1 \le t \le L-1, \tag{1}$$

where $A \in \mathbb{R}^{N \times N}$, while $v_t \in \mathbb{R}^N$ signifies deviation from linear deterministic dynamics (via the input term v_t that may represent stochastic excitation or contribution of nonlinear terms). The standard formulation of DMD is based on the assumption that the time series under consideration, herein x_t, is dominated by the linear transition mechanism and that,

moreover, the dimension of x_t is much larger than the size of the observation window $t \in \{1, 2, \ldots, L\}$.

The underlying premise of the DMD methodology is that the state vector x_t concentrates along the directions that correspond to the dominant eigendirections of A and, thereby, DMD aims (and has a viable chance) to identify the dynamics that are manifested by restricting the recurrence relation in (1) onto the range of a data matrix

$$X_{1:n-1} := [x_1, x_2, \ldots, x_{n-1}],$$

for n possibly $n \le L$. Thereby, the dynamics are sought in a matrix $A \in \mathbb{R}^{N \times N}$ to satisfy

$$X_{2:n} \simeq AX_{1:n-1}. \tag{2}$$

One readily observes that the operator A, restricted onto the orthogonal complement of $\text{range}(X_{1:n-1})$, namely,

$$A|_{\text{range}(X_{1:n-1})^\perp},$$

is undefined, i.e., it cannot be determined from the data. DMD sets out to determine the action of A precisely on the range of $X_{1:n-1}$. To this end, complete the columns of $X_{1:n-1}$ into a basis $\mathcal{B} := \{x_1, \ldots, x_{n-1}, y_n, \ldots, y_N\}$ for $\mathbb{R}^N$. We tacitly assume that $\{x_1, \ldots, x_{n-1}\}$ are linearly independent. The matrix $Y_{n:N} = [y_n, \ldots, y_N]$ formed out of the added (column) vectors is such that

$$T = [X_{1:n-1}, Y_{n:N}]$$

is an invertible matrix. Selection of $Y_{n:N}$ can be accomplished by taking the singular value decomposition

$$X_{1:n-1} = U\Sigma V^T,$$

of $X_{1:n-1}$, where $U \in O(N)$, $V \in O(n-1)$, and

$$\Sigma = \begin{bmatrix} \sigma_1(X_{1:n-1}) & 0 & 0 & \cdots \\ 0 & \sigma_2(X_{1:n-1}) & 0 & \cdots \\ 0 & 0 & \ddots & \\ \vdots & \vdots & & \end{bmatrix}$$

is the $N \times (n-1)$ matrix with the (non-increasing sequence of) singular values of $X_{1:n-1}$ on the main diagonal, and where $O(k)$ denotes the group of $k \times k$ orthogonal matrices. Then, if after partitioning

$$U = [U_{1:n-1}, U_{n:N}],$$

the selection $Y_{n:N} = U_{n:N}$ presents a convenient option.

Similarity transformation with T bring A into the form

$$\begin{bmatrix} S & S_{12} \\ S_{21} & S_{22} \end{bmatrix},$$

since

$$A\left[X_{1:n-1}, Y_{n:N}\right] = \left[X_{1:n-1}, Y_{n:N}\right] \begin{bmatrix} S & S_{12} \\ S_{21} & S_{22} \end{bmatrix}.$$

Thus,

$$AX_{1:n-1} = X_{1:n-1}S + Y_{n:N}S_{21}.$$

Assuming that A leaves $\mathrm{range}(X_{1:n-1})$ invariant, the intertwining relation

$$AX_{1:n-1} = X_{1:n-1}S,$$

holds and S represents the restriction of A onto the $\mathrm{range}(X_{1:n-1})$. Thus, assuming that (2) holds with equality,

$$X_{2:n} = X_{1:n-1}S, \tag{3}$$

captures the action of A on the range of $X_{1:n-1}$ and can be used to determine S. Finally, because the columns of $X_{2:n-1}$ are shared with a shift between $X_{1:n-1}$ and $X_{2:n}$, S has the companion structure

$$S = \begin{bmatrix} 0 & 0 & 0 & \dots & 0 & -s_{n-1} \\ 1 & 0 & 0 & \dots & 0 & -s_{n-2} \\ 0 & 1 & 0 & \dots & 0 & -s_{n-3} \\ \vdots & & \ddots & & & \vdots \\ 0 & 0 & 0 & \dots & 1 & -s_1 \end{bmatrix},$$

where the last column can be easily identified by solving (3).

Since in general the linear transformation A does not leave $\mathrm{range}(X_{1:n-1})$ entirely invariant, and thereby (2) does not hold with equality, suitable approximation is carried out to obtain S. For instance, the vector $s = (s_{n-1}, \dots, s_1)^T$ can be obtained as

$$\mathrm{argmin}\{\|x_n - X_{1:n-1}s\| \mid s \in \mathbb{R}^{n-1}\}, \tag{4}$$

with $\|\cdot\|$ denoting (typically, and herein) the Euclidean norm, and to this end several alternative numerical schemes have been proposed (such as Arnoldi and SVD based) [4], [1]. This is the typical scenario for DMD applications.

Regularizations

An alternative approach is to regularize the problem by penalizing perturbation from the recorded values in data matrix $X_{1:n-1}$ as well, e.g., by solving instead (the nonlinear problem)

$$\mathrm{argmin}\{\|x_n - \hat{X}_{1:n-1}s\| + \epsilon\|\hat{X}_{1:n-1} - X_{1:n-1}\|\},$$

over $s \in \mathbb{R}^{n-1}$ and $\hat{X}_{1:n-1} \in \mathbb{R}^{N \times (n-1)}$, for a choice of regularizing parameter $\epsilon > 0$. This option is especially reasonable in case (2) fails to hold with equality due to stochastic noise or the (small) effect of nonlinear dynamics, or in cases where prior information dictates specific structural features, e.g., see [6], [7].

Higher order dynamics

We note that in cases when higher order dynamics are at play and $\mathbb{R}^N$ is insufficient as a choice of state-space, an option is to account for lagged values of x_t and thereby select as a candidate state vector, e.g., for the case of one lag,

$$\xi_t = [x_t^T, \ x_{t-1}^T]^T.$$

Very little changes in the basic setting [5]. In this case, one seeks an A matrix of twice the size to now satisfy $\Xi_{3:n} \simeq A\Xi_{2:n-1}$, cf. (2). Thence, a matrix S as before, with companion structure, such that

$$\Xi_{3:n} = \Xi_{2:n-1}S,$$

with $\Xi_{k:\ell} := [\xi_k, \xi_{k+1}, \dots, \xi_\ell]$, assuming $k < \ell$, cf. (3). Thus, without loss of generality we will only discuss the basic setting without further expanding neither on higher order dynamics nor on the relevance of various choices for regularization.

Recap & concluding thoughts

The goal of DMD is to identify dominant modes that capture the relation between successive vectors of the time series. These are the roots of the polynomial

$$\mathbf{s}(\lambda) = \lambda^{n-1} + s_1\lambda^{n-2} + \dots + s_{n-1}.$$

An underlying premise of the framework is that the time series does not depart significantly from being quasi-stationary. This can only hold if the observed dynamics result from a "tug-of-war" mechanism that provides excitation and saturation at the same time (*a la* fluctuation-dissipation). Such a dynamical mechanism can be based in either or both, a stochastic excitation or nonlinear contributions, as in (1), where v_t may represent either. This understanding suggests that the effectiveness of DMD and relevance of the underlying dynamical structure may be quantified by the geometric relation between subspaces spanned by successive collections of time series samples x_t. From a more practical perspective, the effectiveness of DMD, by necessity, rests on how close the subspaces spanned by $X_{2:n}$ and $X_{1:n-1}$ are.

To this end, below, we explore the use of geometric concepts that quantify how well the above expectations are reflected in the data. Specifically, we introduce the analogue of partial autocorrelation coefficients that can serve to identify the size of the state-space that can usefully be exploited to identify dominant dynamics.

III. INNOVATION PARAMETERS (IP'S)

The effectiveness of DMD in modeling the underlying dynamics rests on the relation between the subspaces spanned by $\{x_\ell, x_{\ell+1}, \dots, x_m\}$, over a progression of intervals $[\ell, m]$ of indices and over varying window sizes.

Consider first intervals $[1, k-1]$ and $[2, k]$. We seek to quantify the new information that is contained in the last vector x_k as compared to the previous ones. Specifically,

we consider how introducing these new data point x_k impacts the distance (angle) between the subspaces spanned by $X_{1:k-1}$ and $X_{2:k}$. Evidently, the angle between these subspaces relates to the discrepancy in (2) from holding with identity.

We will similarly consider relations between subspaces corresponding to adjacent windows $[\ell, \ell + k - 1]$ and $[\ell + 1, \ell + k]$, and how angles between such subspaces change with the indices ℓ and k.

Angles and the gap metric

The distance between subspaces $\mathcal{X}_1, \mathcal{X}_2 \subseteq \mathcal{X}$, of a Hilbert space $\mathcal{X}$, is naturally quantified by the angle operator

$$R_{12} := \Pi_{\mathcal{X}_1}|_{\mathcal{X}_2^\perp},$$

where $\Pi_{\mathcal{X}_1}$ denotes orthogonal projection onto $\mathcal{X}_1$ and $|_{\mathcal{X}_2^\perp}$ the restriction onto the orthogonal complement of $\mathcal{X}_2$. Herein, we will be concerned with finite dimensional Euclidean spaces. In this case, provided the subspaces have equal dimension,

$$\|\Pi_{\mathcal{X}_1}|_{\mathcal{X}_2^\perp}\| = \|\Pi_{\mathcal{X}_2}|_{\mathcal{X}_1^\perp}\|.$$

This common value is equal to $\|\Pi_{\mathcal{X}_1} - \Pi_{\mathcal{X}_2}\|$ and defines a *bona fide* metric between subspaces [8], [9]. This is referred to as the *gap metric*

$$d(\mathcal{X}_1, \mathcal{X}_2) := \|\Pi_{\mathcal{X}_1} - \Pi_{\mathcal{X}_2}\|.$$

Thence,

$$\theta(\mathcal{X}_1, \mathcal{X}_2) := \arcsin(d(\mathcal{X}_1, \mathcal{X}_2))$$

represents an angular distance between the two subspaces. In case their dimensions do not match, the gap is the maximal norm of the two angle operators, and equals $d(\mathcal{X}_1, \mathcal{X}_2) = 1$, giving $\theta(\mathcal{X}_1, \mathcal{X}_2) = \frac{\pi}{2}$.

We remark that the gap metric between the graphs (infinite dimensional subspaces) of dynamical systems is a natural metric to quantify uncertainty in the context of feedback theory, and as such has been a chapter in modern robust control [10], [11], [12]. Herein we are only concerned with the geometry of finite dimensional subspaces spanned by the vectorial entries of a time series.

Innovation parameters and PARCOR's

In order to assess the consistency of successive measurements of the time series we consider gaps between subspaces spanned by successive segments, e.g., $\mathrm{range}(X_{1:k})$ and $\mathrm{range}(X_{2:k+1})$ for different values of k. We refer to these as *innovation parameters (IP)*

$$\boxed{r_k := d(\mathrm{range}(X_{1:k}), \mathrm{range}(X_{2:k+1}))}$$

In geometric terms, r_k is the sine of the angular distance

$$\theta_k := \arcsin(r_k)$$

between $\Pi_{\mathrm{span}(x_2,\ldots,x_k)^\perp} x_1$ and $\Pi_{\mathrm{span}(x_2,\ldots,x_k)^\perp} x_{k+1}$, i.e., between the projections of x_1, x_{k+1} onto the orthogonal complement of the span of the intermediate vectors $\{x_1, \ldots, x_k\}$. Similarly, we define

$$\boxed{r_{\ell,k} := d(\mathrm{range}(X_{\ell:\ell+k-1}), \mathrm{range}(X_{\ell+1:\ell+k}))}$$

to capture the same dependence between successive subspaces from a different starting point ℓ.

The innovation parameters relate to the partial correlation coefficients (PARCOR) in time-series analysis [13]. Specifically, if $\mathbf{X}_k$, for $k \in \mathbb{Z}$, denotes a stationary time series, the PARCOR's are the cosines of the angles between

$$\mathbf{X}_\ell - \mathbb{E}\{\mathbf{X}_\ell | \mathbf{X}_{\ell+1}, \ldots, \mathbf{X}_{\ell+k-1}\} \text{ and}$$
$$\mathbf{X}_{\ell+k} - \mathbb{E}\{\mathbf{X}_{\ell+k} | \mathbf{X}_{\ell+1}, \ldots, \mathbf{X}_{\ell+k-1}\},$$

where in the conditioning, for $k = 1$, we define the set $\{\mathbf{X}_{\ell+1}, \ldots, \mathbf{X}_{\ell+k-1}\}$ as empty. Thus, these also coincide with the cosines of the angles between the spans of the random variables $\{\mathbf{X}_\ell, \ldots, \mathbf{X}_{\ell+k-1}\}$ and $\{\mathbf{X}_{\ell+1}, \ldots, \mathbf{X}_{\ell+k}\}$.

Besides one set of parameters corresponding to sines and the other to cosines, the main difference between IP's and PARCORs is that the latter are typically defined for stationary stochastic processes, in that the kernel

$$\mathcal{K}(i,j) := \langle x_i, x_j \rangle$$

has a Toeplitz structure [13], unlike the case of IP's which do not have necessarily a Toeplitz structure, as the geometric relations in the data sequence $x_1, x_2, \ldots$ are not shift-invariant in general, which often necessitates exploring the double indexing in $r_{\ell,k}$.

Recursive computation of innovation parameters

Efficient code for computing the innovation parameters for large data sets and size of vectors can be devised based on a recursive scheme that orthonormalizes successive vectors in the data base.

Specifically, consider a basis for the span of $X_{1:k-1}$ to consist of x_1 and the orthonormal columns of a matrix $U_{2:k-1}$. Likewise, the span of $X_{2:k}$ consist of x_k and the orthonormal columns of a matrix $U_{2:k-1}$. Define the orthogonal projection onto the orthogonal complement of the range of $U_{2:k-1}$

$$\Pi_{\mathrm{range}(U_{2:k-1})^\perp} = I - U_{2:k-1}U_{2:k-1}^T.$$

Then the angle between the span of $X_{1:k-1}$ and that of $X_{2:k}$ coincides with the angle between

$$(I - \Pi_{\mathrm{range}(U_{2:k-1})^\perp})x_1 \text{ and } (I - \Pi_{\mathrm{range}(U_{2:k-1})^\perp})x_k.$$

The computation of the innovation parameters can be carried our recursively as follows:

Alternatively, the same computation can be carried out in Matlab utilizing the "econ" feature that optimizes computations for large data sets. E.g., in order to compute

2278

Algorithm 1: Recursive computation of IP's

Data: Given $X_{1:n} \in \mathbb{R}^{N \times n}$
Initialization: $k = 1$, $u_1 = x_1/\|x_1\|$, $u_2 = x_2/\|x_2\|$,
$u_{\text{first}} = u_1 - \langle u_1, u_2 \rangle u_2$, $u_{\text{last}} = u_2$, $U = [u_2]$;
while $k < n - 1$ **do**
    $u_{\text{last}} = x_{k+2}$;
    $u_{\text{last}} = u_{\text{last}} - UU'u_{\text{last}}$;
    $u_{\text{last}} = u_{\text{last}}/\|u_{\text{last}}\|$;
    $r_k = \sin(\text{acos}(\langle u_{\text{first}}, u_{\text{last}} \rangle))$;
    $U = [U \quad u_{\text{last}}]$;
    $u_{\text{first}} = u_{\text{first}} - \langle u_{\text{first}}, u_{\text{last}} \rangle u_{\text{last}}$;
    $u_{\text{first}} = u_{\text{first}}/\|u_{\text{first}}\|$;
    $k = k + 1$;
end

r_n set $Y_1 = X_{1:n-1}$ and $Y_2 = X_{2:n}$, and compute U_i for $i \in \{1, 2\}$ with the command $[U_i, \Sigma_i, V_i] = \text{svd}(Y_i,' \text{econ}')$. Since,

$$\Pi_{\text{range}(Y_i)} = U_i U_i',$$

with U_i an isometry, the gap between the two subspaces is

$$\|U_1 U_1'(I - U_2 U_2')\|^2 = \|U_1' - \underbrace{(U_1' U_2)}_{M} U_2'\|^2$$
$$= \|(U_1' - M U_2')(U_1 - U_2 M')\|$$
$$= \|I - MM' - MM' + MM'\|$$
$$= \|I - MM'\|$$

Therefore, the gap between $\text{range}(Y_1)$ and $\text{range}(Y_2)$ is

$$\sqrt{1 - \sigma_{\min}(M)^2}$$

with $M = U_1' U_2$ and $\sigma_{\min}(M)$ denotes the smallest singular value of M.

We proceed to motivate and explain the use and relevance of the IP's in selecting a suitable size n for the dynamics sought via DMD on a case study. An additional technical result will be presented along with the example, which highlights the fact that under- or over-estimating the value for n leads to significant errors in identifying the correct dynamics. The example we consider is that of an almost periodic series.

IV. A CASE STUDY

We consider time series data that represent a persistent vorticity of a periodically fluctuating fluid flow field in the wake behind a circular cylinder. This dataset can be generated by publicly accessible code in [14]. The two-dimensional Navier–Stokes equations are numerically solved at Reynolds number 100, to obtain these data. At this Reynolds number the flow undergoes a laminar vortex shedding which can be thought of as a stable limit cycle. The data are collected after simulations converge to steady-state vortex shedding. The reader is referred to [1] for more details on how these data set is extracted. At each of 151 snapshots, the values of vorticity are stacked up in a column of a
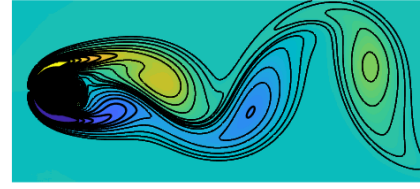
(a) $t = 1$
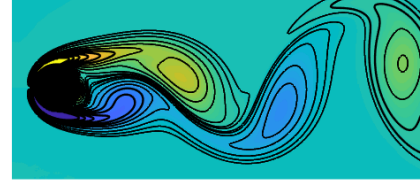
(b) $t = 5$
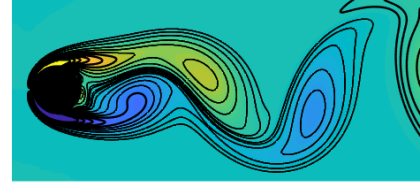
(c) $t = 9$

Fig. 1: Vorticity field around a cylinder wake.

data matrix X which is of size 89351×151. The images of the vorticity field at successive timestamps $t \in \{1, 5, 9\}$ are depicted in Fig. 1. The color-coded velocity fluctuations reveal the mechanism of vortex shedding.

The DMD formalism, and specifically (4), is applied to identify the apparent modes of oscillation. The resulting modes are dramatically affected by the choice of n in (4). Important points that are highlighted below by this example are as follows:

i) The time series is very close to being periodic. This can been seen in a variety of ways, including standard spectral or Fourier analysis. However, here, we compute the sequence of innovation parameters that quantify how far the subspaces spanned by sliding windows of data, of varying width, are from each other in the gap metric. Fig. 2 shows $r_k = r_{1,k}$ as a function of k. A dimple that repeats with period 30 indicates periodicity. It turns out that exact periodicity of the r_k's, even when the time series is very close to being periodic is masked by numerical sensitivity that we will comment later on (discussion leading to, and Proposition 1).

ii) Fig. 3 shows the color-coded values of $r_{\ell,k}$ as a function of ℓ vs. k. Specifically, 50 snapshots are drawn as rows. The ℓth row corresponds to the gap $d(\text{range}(X_{\ell:k}), \text{range}(X_{\ell:k+1}))$, where k sweeps from
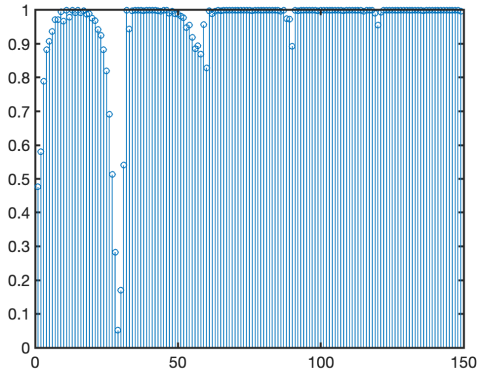
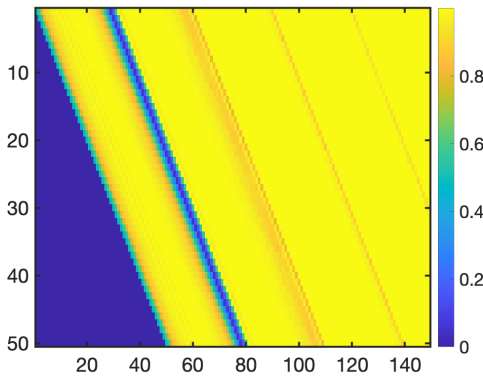Fig. 2: $r_k = d(\mathrm{range}(X_{1:k}), \mathrm{range}(X_{2:k+1}))$ vs. k

Fig. 3: $d(\mathrm{range}(X_{\ell:k}), \mathrm{range}(X_{\ell:k+1}))$ color-coded as function of starting time ℓ and window size k

$\ell + 1$ to the last one. The first row, for instance, corresponds to the values illustrated in Fig. 2. One can observe that at each row the minimum gap occurs at the 30th timestamp. This strongly suggests the use of a time-window of size $n = 30$ to find the DMD modes. Periodicity is evident in Fig. 3; the decreasing dimples with period 30 are repeated with regularity starting from any chosen starting point ℓ (cf. discussion leading to Proposition 1).

iii) The eigenvalues of S (DMD eigenvalues of the sought dynamics) are shown in Fig. 4 for $n \in \{20, 30, 40\}$. It is observed that their distribution is dramatically affected by the chosen size of the subspaces to compare in (2), namely n.

iv) For $n = 30$ the eigenvalues of S shown in Fig. 4 have modulus $\simeq 1$, in agreement with the observed periodic structure of the flow field. Exact periodicity of the time series results in equispaced eigenvalues, and this is (almost) the case here.

At this point we would like to explain the source of the apparent diminishing of periodic dimples in Fig. 2 with period 30. As noted earlier, the gap
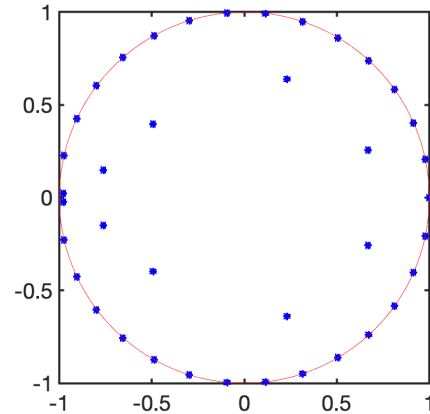
$$r_k = d(\mathrm{range}(X_{1:k-1}), \mathrm{range}(X_{2:k}))$$
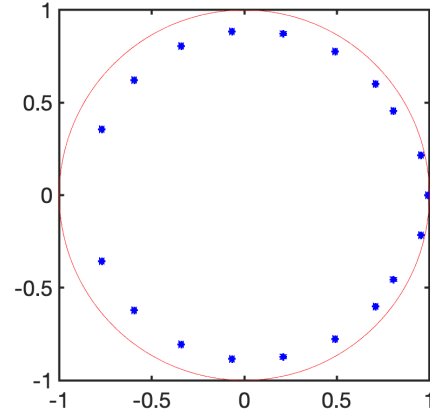
(a) DMD eigenvalues for $n = 20$
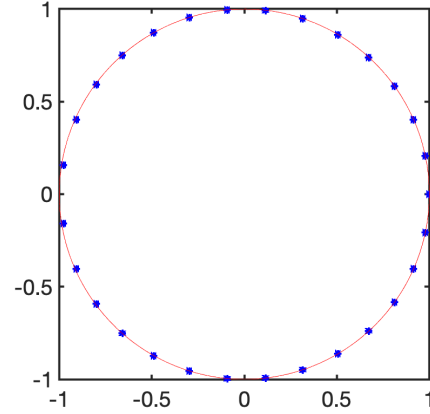
(b) DMD eigenvalues for $n = 30$

(c) DMD eigenvalues for $n = 40$

Fig. 4: DMD eigenvalues from vorticity field data.

is the sine of the angle between

$$\xi_1 := \Pi_{\mathrm{span}(x_2,\dots,x_{k-1})^\perp} x_1, \text{ and}$$
$$\xi_k := \Pi_{\mathrm{span}(x_2,\dots,x_{k-1})^\perp} x_k.$$

Assuming that the series is k-periodic, the angle between ξ_1 and ξ_k is zero and $x_k \in \mathrm{span}(x_1,\dots,x_{k-1})$. Likewise,

$$x_{k+1} \in \mathrm{span}(x_2,\dots,x_k) = \mathrm{span}(x_1,\dots,x_{k-1}).$$

Denote

$$\xi_{\mathrm{next}} := \Pi_{\mathrm{span}(x_2,\dots,x_{k-1})^\perp} x_{k+1},$$

and observe that the angle to ξ_k, and therefore ξ_1 too, is zero. Then

$$\begin{aligned}
r_{k+1} &= d(\mathrm{range}(X_{1:k}), \mathrm{range}(X_{2:k+1})) \\
&= d(\mathrm{span}(\xi_1, X_{2:k-1}, \xi_k), \mathrm{span}(X_{2:k-1}, \xi_k, \xi_{\mathrm{next}})) \\
&= d(\mathrm{span}(\xi_1, \xi_k), \mathrm{span}(\xi_k, \xi_{\mathrm{next}})) = 0,
\end{aligned}$$

with all three vectors $\xi_1, \xi_k, \xi_{\mathrm{next}}$ co-linear. However, a small perturbation in each has a significant effect. Indeed, for arbitrarily small δ's,

$$\begin{aligned}
&d(\mathrm{span}(\xi_1 + \delta_1, \xi_k + \delta_k), \mathrm{span}(\xi_k + \delta_k, \xi_{\mathrm{next}} + \delta)) \\
&= d(\Pi_{\mathrm{span}(\xi_k + \delta_k)^\perp}(\xi_1 + \delta_1), \Pi_{\mathrm{span}(\xi_k + \delta)^\perp}(\xi_{\mathrm{next}} + \delta))
\end{aligned}$$

can take any value on $[0, 1]$. We recast the claim as follows.

Proposition 1: Consider a vector $\xi \in \mathbb{R}^N$ and perturbations $\xi_i = \xi + \delta_i$, for $i \in \{1, 2\}$, with $\delta_i \perp \xi$. Then

$$d(\mathrm{span}(\xi + \delta_1, \xi), \mathrm{span}(\xi + \delta_2, \xi)) = d(\mathrm{span}(\delta_1), \mathrm{span}(\delta_2)).$$

The proof is elementary. What this statement helps exemplify (and prove) is that in cases where elements that determine the span of interest are almost co-linear, the angles between the subspaces are very sensitive to errors. A more precise mathematical statement can be worked out that involves the conditioning number of the matrix $X_{1:k}$ in our earlier setting.

V. Concluding remarks

In many applications it is often the case that only a limited number of data samples are available for modeling an otherwise exceedingly high dimensional process. The dimensionality of the process, which may represent visual or distributional fields, in conjunction with the limited observation record requires careful analysis. It is precisely this regime of "small data," i.e., "few samples," that has been a challenge in traditional signal analysis since its inception [15], and has led to entropic regularization among other methodologies. DMD represents a more recent development that aims to identify suitable linear dynamics that can explain the data.

Historically, DMD has roots and ramifications that relate to theory of the Koopman operator [16], [17], [18]. Data that originate from periodic and quasi-periodic attractors of nonlinear dynamics can also be dealt with in the same framework [1]. Thus the concept of the gap metric, as a tool to quantify how subspaces spanned by data impact modeling assumptions, is expected to be applicable in this more general

setting. The present work summarizes some of the findings in a developing treatise into the topic of extracting dynamics from high dimension distributional fields [19], specifically, the relevance of the gap metric as a tool to provide guidance in selecting appropriate dimensionality for models for such processes.

References

[1] J. N. Kutz, S. L. Brunton, B. W. Brunton, and J. L. Proctor, *Dynamic mode decomposition: data-driven modeling of complex systems*. SIAM, 2016.

[2] M. O. Williams, I. G. Kevrekidis, and C. W. Rowley, "A data–driven approximation of the Koopman operator: Extending dynamic mode decomposition," *Journal of Nonlinear Science*, vol. 25, no. 6, pp. 1307–1346, 2015.

[3] J. L. Proctor, S. L. Brunton, and J. N. Kutz, "Dynamic mode decomposition with control," *SIAM Journal on Applied Dynamical Systems*, vol. 15, no. 1, pp. 142–161, 2016.

[4] P. J. Schmid, "Dynamic mode decomposition of numerical and experimental data," *Journal of fluid mechanics*, vol. 656, pp. 5–28, 2010.

[5] S. Le Clainche and J. M. Vega, "Higher order dynamic mode decomposition," *SIAM Journal on Applied Dynamical Systems*, vol. 16, no. 2, pp. 882–925, 2017.

[6] M. R. Jovanović, P. J. Schmid, and J. W. Nichols, "Sparsity-promoting dynamic mode decomposition," *Physics of Fluids*, vol. 26, no. 2, p. 024103, 2014.

[7] C. Dicle, H. Mansour, D. Tian, M. Benosman, and A. Vetro, "Robust low rank dynamic mode decomposition for compressed domain crowd and traffic flow analysis," in *2016 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 2016, pp. 1–6.

[8] G. Stewart and J.-G. Sun, "Matrix perturbation theory academic press," *San Diego*, 1990.

[9] T. Kato, *Perturbation theory for linear operators*. Springer Science & Business Media, 2013, vol. 132.

[10] T. T. Georgiou, "On the computation of the gap metric," *Systems & Control Letters*, vol. 11, no. 4, pp. 253–257, 1988.

[11] T. T. Georgiou and M. C. Smith, "Optimal robustness in the gap metric," in *Proceedings of the 28th IEEE Conference on Decision and Control,*. IEEE, 1989, pp. 2331–2336.

[12] K. Zhou and J. C. Doyle, *Essentials of robust control*. Prentice hall Upper Saddle River, NJ, 1998, vol. 104.

[13] P. Stoica and R. L. Moses, *Spectral analysis of signals*. Pearson Prentice Hall, NJ, 2005.

[14] C. R. Rowley. Github repository. [Online]. Available: https://github.com/cwrowley/ibpm

[15] J. P. Burg, D. G. Luenberger, and D. L. Wenger, "Estimation of structured covariance matrices," *Proceedings of the IEEE*, vol. 70, no. 9, pp. 963–974, 1982.

[16] I. Mezic and A. Banaszuk, "Comparison of systems with complex behavior: spectral methods," in *Proceedings of the 39th IEEE Conference on Decision and Control (Cat. No. 00CH37187)*, vol. 2. IEEE, 2000, pp. 1224–1231.

[17] I. Mezić, "Spectral properties of dynamical systems, model reduction and decompositions," *Nonlinear Dynamics*, vol. 41, no. 1, pp. 309–325, 2005.

[18] C. W. Rowley, I. Mezić, S. Bagheri, P. Schlatter, D. Henningson *et al.*, "Spectral analysis of nonlinear flows," *Journal of fluid mechanics*, vol. 641, no. 1, pp. 115–127, 2009.

[19] A. Karimi, "Statistical learning in Wasserstein space," Ph.D. dissertation, University of California, Irvine, 2021.