

## ON DYNAMIC MODE DECOMPOSITION: THEORY AND APPLICATIONS

JONATHAN H. TU, CLARENCE W. ROWLEY, DIRK M. LUCHTENBURG,

Dept. of Mechanical and Aerospace Engineering  
Princeton University  
Princeton, NJ 08544, USA

STEVEN L. BRUNTON, AND J. NATHAN KUTZ

Dept. of Applied Mathematics  
University of Washington  
Seattle, WA 98195, USA

**ABSTRACT.** Originally introduced in the fluid mechanics community, dynamic mode decomposition (DMD) has emerged as a powerful tool for analyzing the dynamics of nonlinear systems. However, existing DMD theory deals primarily with sequential time series for which the measurement dimension is much larger than the number of measurements taken. We present a theoretical framework in which we define **DMD as the eigendecomposition of an approximating linear operator**. This generalizes DMD to a larger class of datasets, including nonsequential time series. We demonstrate the utility of this approach by presenting novel sampling strategies that increase computational efficiency and mitigate the effects of noise, respectively. We also introduce the concept of linear consistency, which helps explain the potential pitfalls of applying DMD to rank-deficient datasets, illustrating with examples. Such computations are not considered in the existing literature, but can be understood using our more general framework. In addition, we show that our theory strengthens the connections between DMD and Koopman operator theory. It also establishes connections between DMD and other techniques, including the eigensystem realization algorithm (ERA), a system identification method, and linear inverse modeling (LIM), a method from climate science. We show that under certain conditions, DMD is equivalent to LIM.

**1. Introduction.** Fluid flows often exhibit low-dimensional behavior, despite the fact that they are governed by infinite-dimensional partial differential equations (the Navier–Stokes equations). For instance, the main features of the laminar flow past a two-dimensional cylinder can be described using as few as three ordinary differential equations [1]. To identify these low-order dynamics, such flows are often analyzed using *modal decomposition* techniques, including proper orthogonal decomposition (POD), balanced proper orthogonal decomposition (BPOD), and dynamic mode decomposition (DMD). Such methods describe the fluid state (typically the velocity or vorticity field) as a superposition of empirically computed basis vectors, or “modes.” In practice, the number of modes necessary to capture the gross behavior

---

2000 *Mathematics Subject Classification.* Primary: 37M10, 65P99; Secondary: 47B33.

*Key words and phrases.* Dynamic mode decomposition, Koopman operator, spectral analysis, time series analysis, reduced-order models.

of a flow is often many orders of magnitude smaller than the state dimension of the system (e.g.,  $\mathcal{O}(10)$  compared to  $\mathcal{O}(10^6)$ ).

Since it was first introduced in [2], DMD has quickly gained popularity in the fluids community, primarily because it provides information about the *dynamics* of a flow, and is applicable even when those dynamics are *nonlinear* [3]. A typical application involves collecting a time series of experimental or simulated velocity fields, and from them computing DMD modes and eigenvalues. The modes are spatial fields that often identify coherent structures in the flow. The corresponding eigenvalues define growth/decay rates and oscillation frequencies for each mode. Taken together, the DMD modes and eigenvalues describe the dynamics observed in the time series in terms of oscillatory components. In contrast, POD modes optimally reconstruct a dataset, with the modes ranked in terms of energy content [4]. BPOD modes identify spatial structures that are important for capturing linear input-output dynamics [5], and can also be interpreted as an optimal decomposition of two (dual) datasets [6].

At first glance, it may seem dubious that a nonlinear system could be described by superposition of modes whose dynamics are governed by eigenvalues. After all, one needs a *linear* operator in order to talk about eigenvalues. However, it was shown in [3] that **DMD is closely related to a spectral analysis of the Koopman operator.** The Koopman operator is a linear but infinite-dimensional operator whose modes and eigenvalues capture the evolution of observables describing any (even nonlinear) dynamical system. The use of its spectral decomposition for data-based modal decomposition and model reduction was first proposed in [7]. DMD analysis can be considered to be a numerical approximation to Koopman spectral analysis, and it is in this sense that DMD is applicable to nonlinear systems. In fact, the terms “DMD mode” and “Koopman mode” are often used interchangeably in the fluids literature.

Much of the recent work involving DMD has focused on its application to different flow configurations. For instance, DMD has been applied in the study of the wake behind a flexible membrane [8], the flow around high-speed trains [9], instabilities in annular liquid sheets [10], shockwave-turbulent boundary layer interactions [11], detonation waves [12], cavity flows [8, 13], and various jets [3, 8, 14, 15, 16, 17]. There have also been a number of efforts regarding the numerics of the DMD algorithm, including the development of memory-efficient algorithms [18, 19], an error analysis of DMD growth rates [20], and a method for selecting a sparse basis of DMD modes [21]. Variants of the DMD algorithm have also been proposed, including optimized DMD [22] and optimal mode decomposition [23, 24]. Theoretical work on DMD has centered mainly on exploring connections with other methods, such as Koopman spectral analysis [3, 25, 26], POD [8], and Fourier analysis [22]. Theorems regarding the existence and uniqueness of DMD modes and eigenvalues can be found in [22]. For a review of the DMD literature, we refer the reader to [25].

Many of the papers cited above mention the idea that DMD is able to characterize nonlinear dynamics through an analysis of some approximating linear system. In this work, we build on this notion. We present DMD as an analysis of *pairs* of  $n$ -dimensional data vectors  $(x_k, y_k)$ , in contrast to the sequential time series that are typically considered. From these data we construct a particular linear operator  $A$  and define DMD as the eigendecomposition of that operator (see Definition 1). We show that DMD modes satisfying this definition can be computed

using a modification of the algorithm proposed in [8]. Both algorithms generate the same eigenvalues, with the modes differing by a projection (see Theorem 3).

There is of course no guarantee that analyzing this particular approximating operator is meaningful for data generated by nonlinear dynamics. To this end, we show that our definition strengthens the connections between DMD and Koopman operator theory, extending those connections to include more general sampling strategies. This is important, as it allows us to maintain the interpretation of DMD as an approximation to Koopman spectral analysis. We can then be confident that DMD is useful for characterizing nonlinear dynamics. Furthermore, we show that the connections between DMD and Koopman spectral analysis hold not only when the vectors  $x_k$  are linearly independent, as assumed in [3], but under a more general condition we refer to as linear consistency (see Definition 2). When the data are not linearly consistent, the Koopman analogy can break down and DMD analysis may produce either meaningful or misleading results. We show an example of each and explain the results using our approximating-operator definition of DMD. (For a more detailed investigation of how well DMD eigenvalues approximate Koopman eigenvalues, we refer the reader to [26].)

The generality of our framework has important practical implications as well. To this end, we present examples demonstrating the benefits of applying DMD to non-sequential time series. For instance, we show that nonuniform temporal sampling can provide increased computational efficiency, with little effect on accuracy of the dominant DMD modes and eigenvalues. We also show that noise in experimental datasets can be dealt with by concatenating data from multiple runs of an experiment. The resulting DMD computation produces a spectrum with sharper, more isolated peaks, allowing us to identify higher-frequency modes that are obscured in a traditional DMD computation.

Finally, our framework highlights the connections between DMD and other well-known methods, specifically the eigensystem realization algorithm (ERA) and linear inverse modeling (LIM). The ERA is a control-theoretic method for system identification of linear systems [27, 28, 29]. We show that when computed from the same data, DMD eigenvalues reduce to poles of an ERA model. This connection motivates the use of ERA-inspired strategies for dealing with certain limitations of DMD. LIM is a modeling procedure developed in the climate science community [30, 31]. We show that under certain conditions, DMD is equivalent to LIM. Thus it stands to reason that practitioners of DMD could benefit from an awareness of related work in the climate science literature.

The remainder of this work is organized as follows: in Section 2, we propose and discuss a new definition of DMD. We provide several different algorithms for computing **DMD modes and eigenvalues** that satisfy this new definition and show that these are closely related to the modes and eigenvalues computed using the currently accepted SVD-based DMD algorithm [8]. A number of examples are presented in Section 3. These explore the application of DMD to rank-deficient datasets and nonsequential time series. Section 4 describes the connections between DMD and Koopman operator theory, the ERA, and LIM, respectively. We summarize our results in Section 5.

**2. Theory.** Since its first appearance in 2008 [2], DMD has been defined by an algorithm (the specifics of which are given in Algorithm 1 below). Here, we present a more general, non-algorithmic definition of DMD. Our definition emphasizes data

that are collected as a set of *pairs*  $\{(x_k, y_k)\}_{k=1}^m$ , rather than as a sequential time series  $\{z_k\}_{k=0}^m$ . We show that our DMD definition and algorithm are closely related to the currently accepted algorithmic definition. In fact, the two approaches produce the same DMD eigenvalues; it is only the DMD modes that differ.

**2.1. Standard definition.** Originally, the DMD algorithm was formulated in terms of a companion matrix [2, 3], which highlights its connections to the Arnoldi algorithm and to Koopman operator theory. The SVD-based algorithm presented in [8] is more numerically stable, and is now generally accepted as the defining DMD algorithm; we describe this algorithm below.

Consider a sequential set of data vectors  $\{z_0, \dots, z_m\}$ , where each  $z_k \in \mathbb{R}^n$ . We assume that the data are generated by linear dynamics

$$z_{k+1} = Az_k, \quad (1)$$

for some (unknown) matrix  $A$ . (Alternatively, the vectors  $z_k$  can be sampled from a continuous evolution  $z(t)$ , in which case  $z_k = z(k\Delta t)$  and a fixed sampling rate  $\Delta t$  is assumed.) When DMD is applied to data generated by nonlinear dynamics, it is assumed that there exists an operator  $A$  that approximates those dynamics. The DMD modes and eigenvalues are intended to approximate the eigenvectors and eigenvalues of  $A$ . The algorithm proceeds as follows:

**Algorithm 1 (Standard DMD).**

1. Arrange the data  $\{z_0, \dots, z_m\}$  into matrices

$$X \triangleq [z_0 \quad \dots \quad z_{m-1}], \quad Y \triangleq [z_1 \quad \dots \quad z_m]. \quad (2)$$

2. Compute the (reduced) SVD of  $X$  (see [32]), writing

$$X = U\Sigma V^*, \quad (3)$$

where  $U$  is  $n \times r$ ,  $\Sigma$  is diagonal and  $r \times r$ ,  $V$  is  $m \times r$ , and  $r$  is the rank of  $X$ .

3. Define the matrix

$$\tilde{A} \triangleq U^* Y V \Sigma^{-1}. \quad (4)$$

4. Compute eigenvalues and eigenvectors of  $\tilde{A}$ , writing

$$\tilde{A}w = \lambda w. \quad (5)$$

5. The DMD mode corresponding to the DMD eigenvalue  $\lambda$  is then given by

$$\hat{\varphi} \triangleq Uw. \quad (6)$$

6. If desired, the DMD modes can be scaled in a number of ways, as described in Appendix A.

In this paper, we will refer to the modes produced by Algorithm 1 as *projected DMD modes*, for reasons that will become apparent in Section 2.3 (in particular, see Theorem 3).

**2.2. New definition.** The standard definition of DMD assumes a sequential set of data vectors  $\{z_0, \dots, z_m\}$  in which the order of vectors  $z_k$  is critical. Furthermore, the vectors should (at least approximately) satisfy the relation (1). Here, we relax these restrictions on the data, and consider data *pairs*  $\{(x_1, y_1), \dots, (x_m, y_m)\}$ . We then define DMD in terms of the  $n \times m$  data matrices

$$X \triangleq [x_1 \quad \dots \quad x_m], \quad Y \triangleq [y_1 \quad \dots \quad y_m]. \quad (7)$$

Note that the formulation (2) is a special case of (7), with  $x_k = z_{k-1}$  and  $y_k = z_k$ . In order to relate this method to the standard DMD procedure, we may assume that

$$y_k = \hat{A}x_k$$

for some (unknown) matrix  $\hat{A}$ . However, the procedure here is applicable more generally. We define the DMD modes of this dataset as follows:

**Definition 1** (Exact DMD). For a dataset given by (7), define the operator

$$A \triangleq YX^+, \quad (8)$$

where  $X^+$  is the pseudoinverse of  $X$ . The *dynamic mode decomposition* of the pair  $(X, Y)$  is given by the eigendecomposition of  $A$ . That is, the DMD modes and eigenvalues are the eigenvectors and eigenvalues of  $A$ .

**Remark 1.** The operator  $A$  in (8) is the least-squares/minimum-norm solution to the potentially over- or under-constrained problem  $AX = Y$ . That is, if there is an exact solution to  $AX = Y$  (which is always the case if the vectors  $x_k$  are linearly independent), then the choice (8) minimizes  $\|A\|_F$ , where  $\|A\|_F = \text{Tr}(AA^*)^{1/2}$  denotes the Frobenius norm. If there is no  $A$  that exactly satisfies  $AX = Y$ , then the choice (8) minimizes  $\|AX - Y\|_F$ .

When  $n$  is large, as is often the case with fluid flow data, it may be inefficient to compute the eigendecomposition of the  $n \times n$  matrix  $A$ . In some cases, even storing  $A$  in memory can be prohibitive. Using the following algorithm, the DMD modes and eigenvalues can be computed without an explicit representation or direct manipulations of  $A$ .

**Algorithm 2** (Exact DMD).

1. Arrange the data pairs  $\{(x_1, y_1), \dots, (x_m, y_m)\}$  into matrices  $X$  and  $Y$ , as in (7).
2. Compute the (reduced) SVD of  $X$ , writing  $X = U\Sigma V^*$ .
3. Define the matrix  $\tilde{A} \triangleq U^*YV\Sigma^{-1}$ .
4. Compute eigenvalues and eigenvectors of  $\tilde{A}$ , writing  $\tilde{A}w = \lambda w$ . Each nonzero eigenvalue  $\lambda$  is a DMD eigenvalue.
5. The DMD mode corresponding to  $\lambda$  is then given by

$$\varphi \triangleq \frac{1}{\lambda} YV\Sigma^{-1}w. \quad (9)$$

6. If desired, the DMD modes can be scaled in a number of ways, as described in Appendix A.

**Remark 2.** When  $n \gg m$ , the above algorithm can be modified to reduce computational costs. For instance, the SVD of  $X$  can be computed efficiently using the method of snapshots [33]. This involves computing the correlation matrix  $X^*X$ . The product  $U^*Y$  required to form  $\tilde{A}$  can be cast in terms of a product  $X^*Y$ , using (3) to substitute for  $U$ . If  $X$  and  $Y$  happen to share columns, as is the case for sequential time series, then  $X^*Y$  will share elements with  $X^*X$ , reducing the number of new computations required. (See Section 3.1 for more on sequential versus nonsequential time series.)

Algorithm 2 is nearly identical to Algorithm 1 (originally presented in [8]). In fact, the only difference is that the DMD modes are given by (9), whereas in Algorithm 1, they are given by (6). This modification is subtle, but important, as we discuss in Section 2.3.

**Remark 3.** Though the original presentations of DMD [3, 8] assume  $X$  and  $Y$  of the form given by (2), Algorithm 1 does not make use of this structure. That is, the algorithm can be carried out for the general case of  $X$  and  $Y$  given by (7). (This point has been noted previously, in [24].)

**Theorem 1.** *Each pair  $(\varphi, \lambda)$  generated by Algorithm 2 is an eigenvector/eigenvalue pair of  $A$ . Furthermore, the algorithm identifies all of the nonzero eigenvalues of  $A$ .*

*Proof.* From the SVD  $X = U\Sigma V^*$ , we may write the pseudoinverse of  $X$  as

$$X^+ = V\Sigma^{-1}U^*,$$

so from (8), we find

$$A = YV\Sigma^{-1}U^* = BU^*, \quad (10)$$

where

$$B \triangleq YV\Sigma^{-1}. \quad (11)$$

In addition, we can rewrite (4) as

$$\tilde{A} = U^*YV\Sigma^{-1} = U^*B. \quad (12)$$

Now, suppose that  $\tilde{A}w = \lambda w$ , with  $\lambda \neq 0$ , and let  $\varphi = \frac{1}{\lambda}Bw$ , as in (9). Then

$$A\varphi = \frac{1}{\lambda}BU^*Bw = B\frac{1}{\lambda}\tilde{A}w = Bw = \lambda\varphi.$$

In addition,  $\varphi \neq 0$ , since if  $Bw = 0$ , then  $U^*Bw = \tilde{A}w = 0$ , so  $\lambda = 0$ . Hence,  $\varphi$  is an eigenvector of  $A$  with eigenvalue  $\lambda$ .

To show that Algorithm 2 identifies all of the nonzero eigenvalues of  $A$ , suppose  $A\varphi = \lambda\varphi$ , for  $\lambda \neq 0$ , and let  $w = U^*\varphi$ . Then

$$\tilde{A}w = U^*BU^*\varphi = U^*A\varphi = \lambda U^*\varphi = \lambda w.$$

Furthermore,  $w \neq 0$ , since if  $U^*\varphi = 0$ , then  $BU^*\varphi = A\varphi = 0$ , and  $\lambda = 0$ . Thus,  $w$  is an eigenvector of  $\tilde{A}$  with eigenvalue  $\lambda$ , and is identified by Algorithm 2.  $\square$

**Remark 4.** Algorithm 2 may also be used to find certain eigenvectors with  $\lambda = 0$  (that is, in the nullspace of  $A$ ). In particular, if  $\tilde{A}w = 0$  and  $\varphi = YV\Sigma^{-1}w \neq 0$ , then  $\varphi$  is an eigenvector with  $\lambda = 0$  (and is in the image of  $Y$ ); if  $\tilde{A}w = 0$  and  $YV\Sigma^{-1}w = 0$ , then  $\varphi = Uw$  is an eigenvector with  $\lambda = 0$  (and is in the image of  $X$ ). However, DMD modes corresponding to zero eigenvalues are usually not of interest, since they do not play a role in the dynamics.

Next, we characterize the conditions under which the operator  $A$  defined by (8) satisfies  $Y = AX$ . We emphasize that this does not require that  $Y$  is generated from  $X$  through linear dynamics defined by  $A$ ; we place no restrictions on the data pairs  $(x_k, y_k)$ . To this end, we introduce the following definition:

**Definition 2** (Linear consistency). Two  $n \times m$  matrices  $X$  and  $Y$  are *linearly consistent* if, whenever  $Xc = 0$ , then  $Yc = 0$  as well.

Thus  $X$  and  $Y$  are linearly consistent if and only if the nullspace of  $Y$  contains the nullspace of  $X$ . If the vectors  $x_k$  (columns of  $X$ ) are linearly independent, the nullspace of  $X$  is  $\{0\}$  and linear consistency is satisfied trivially. However, linear consistency does not imply that the columns of  $X$  are linearly independent. (Linear consistency will play an important role in establishing the connection between DMD modes and Koopman modes, as discussed in Section 4.1.)

The notion of linear consistency makes intuitive sense if we think of the vectors  $x_k$  as inputs and the vectors  $y_k$  as outputs. Definition 2 follows from the idea that two identical inputs should not have different outputs, generalizing the idea to arbitrary linearly dependent sets of inputs. It turns out that for linearly consistent data, the approximating operator  $A$  relates the datasets exactly ( $AX = Y$ ), even if the data are generated by nonlinear dynamics.

**Theorem 2.** *Define  $A = YX^+$ , as in Definition 1. Then  $Y = AX$  if and only if  $X$  and  $Y$  are linearly consistent.*

*Proof.* First, suppose  $X$  and  $Y$  are not linearly consistent. Then there exists  $v$  in the nullspace of  $X$  (denoted  $\mathcal{N}(X)$ ) such that  $Yv \neq 0$ . But then  $AXv = 0 \neq Yv$ , so  $AX \neq Y$  for any  $A$ .

Conversely, suppose  $X$  and  $Y$  are linearly consistent: that is,  $\mathcal{N}(X) \subset \mathcal{N}(Y)$ . Then

$$Y - AX = Y - YX^+X = Y(I - X^+X).$$

Now,  $X^+X$  is the orthogonal projection onto the range of  $X^*$  (denoted  $\mathcal{R}(X^*)$ ), so  $I - X^+X$  is the orthogonal projection onto  $\mathcal{R}(X^*)^\perp = \mathcal{N}(X)$ . Thus, since  $\mathcal{N}(X) \subset \mathcal{N}(Y)$ , it follows that  $Y(I - X^+X) = 0$ , so  $Y = AX$ .  $\square$

We note that given Definition 1, is natural to define adjoint DMD modes as the eigenvectors of  $A^*$  (or, equivalently, the left eigenvectors of  $A$ ). Computing such modes can be done easily with slight modifications to Algorithm 2. Let  $z$  be an adjoint eigenvector of  $\tilde{A}$ , so that  $z^* \tilde{A} = \lambda z^*$ . Then one can easily verify that  $\psi \triangleq Uz$  is an adjoint eigenvector of  $A$ :  $\psi^* A = \lambda \psi^*$ . It is interesting to note that while the DMD modes corresponding to nonzero eigenvalues lie in the image of  $Y$ , the adjoint DMD modes lie in the image of  $X$ .

**2.3. Comparing definitions.** Algorithm 1 (originally presented in [8]) has come to dominate among DMD practitioners due to its numerical stability. Effectively, it has become the working definition of DMD. As mentioned above, it differs from Algorithm 2 only in that the (projected) DMD modes are given by

$$\hat{\varphi} \triangleq Uw,$$

where  $w$  is an eigenvector of  $\tilde{A}$ , while the exact modes DMD modes are given by (9) as

$$\varphi \triangleq \frac{1}{\lambda} YV\Sigma^{-1}w.$$

Since  $U$  contains left singular vectors of  $X$ , we see that the original modes defined by (6) lie in the image of  $X$ , while those defined by (9) lie in the image of  $Y$ . As a result, the modes  $\hat{\varphi}$  are not eigenvectors of the approximating linear operator  $A$ . Are they related in any way to the eigenvectors of  $A$ ? The following theorem establishes that they are, and motivates the terminology *projected* DMD modes.

**Theorem 3.** *Let  $\tilde{A}w = \lambda w$ , with  $\lambda \neq 0$ , and let  $\mathbb{P}_X$  denote the orthogonal projection onto the image of  $X$ . Then  $\hat{\varphi}$  given by (6) is an eigenvector of  $\mathbb{P}_X A$  with eigenvalue  $\lambda$ . Furthermore, if  $\varphi$  is given by (9), then  $\hat{\varphi} = \mathbb{P}_X \varphi$ .*

*Proof.* From the SVD  $X = U\Sigma V^*$ , the orthogonal projection onto the image of  $X$  is given by  $\mathbb{P}_X = UU^*$ . Using (10) and (12), and recalling that  $U^*U$  is the identity



matrix, we have

$$\begin{aligned}\mathbb{P}_X A \hat{\varphi} &= (UU^*)(BU^*)(Uw) = U(U^*B)w \\ &= U\tilde{A}w = \lambda Uw = \lambda \hat{\varphi}.\end{aligned}$$

Then  $\hat{\varphi}$  is an eigenvector of  $\mathbb{P}_X A$  with eigenvalue  $\lambda$ . Moreover, if  $\varphi$  is given by (9), then

$$U^* \varphi = \frac{1}{\lambda} U^* B w = \frac{1}{\lambda} \tilde{A} w = w,$$

so  $\mathbb{P}_X \varphi = UU^* \varphi = Uw = \hat{\varphi}$ .  $\square$

Thus, the modes  $\hat{\varphi}$  determined by Algorithm 1 are simply the projection of the modes determined by Algorithm 2 onto the range of  $X$ . Also, note that  $U^* \varphi = U^* \hat{\varphi} = w$ .

**Remark 5.** If the vectors  $\{y_k\}$  lie in the span of the vectors  $\{x_k\}$ , then  $\mathbb{P}_X A = A$ , and the projected and exact DMD modes are identical. For instance, this is the case for a sequential time series (as in (2)) when the last vector  $z_m$  is a linear combination of  $\{z_0, \dots, z_{m-1}\}$ .

In addition to providing an improved algorithm, [8] also explores the connection between (projected) DMD and POD. For data generated by linear dynamics  $z_{k+1} = Az_k$ , [8] derives the relation  $\tilde{A} = U^* A U$  and notes that we can interpret  $\tilde{A}$  as the correlation between the matrix of POD modes  $U$  and the matrix of time-shifted POD modes  $AU$ . Of course,  $\tilde{A}$  also determines the DMD eigenvalues. Exact DMD is based on the definition of  $A$  as the least-squares/minimum-norm solution to  $AX = Y$ , but does not require that equation to be satisfied exactly. Even so, we can combine (10) and (12) to find that  $\tilde{A} = U^* A U$ . Thus exact DMD preserves the interpretation of  $\tilde{A}$  in terms of POD modes, extending it from sequential time series to generic data matrices, and without making any assumptions about the dynamics relating  $X$  and  $Y$ .

To further clarify the connections between exact and projected DMD, consider the following alternative algorithm for computing DMD eigenvalues and exact DMD modes:

**Algorithm 3** (Exact DMD, alternative method).

1. Arrange the data pairs  $\{(x_1, y_1), \dots, (x_m, y_m)\}$  into matrices  $X$  and  $Y$ , as in (7).
2. Compute the (reduced) SVD of  $X$ , writing  $X = U\Sigma V^*$ .
3. Compute an orthonormal basis for the column space of  $\begin{bmatrix} X & Y \end{bmatrix}$ , stacking the basis vectors in a matrix  $Q$  such that  $Q^*Q = I$ . For example,  $Q$  can be computed by singular value decomposition of  $\begin{bmatrix} X & Y \end{bmatrix}$ , or by a  $QR$  decomposition.
4. Define the matrix

$$\tilde{A}_Q \triangleq Q^* A Q, \tag{13}$$

where  $A = YV\Sigma^{-1}U^*$ , as in (10).

5. Compute eigenvalues and eigenvectors of  $\tilde{A}_Q$ , writing  $\tilde{A}_Q w = \lambda w$ . Each nonzero eigenvalue  $\lambda$  is a DMD eigenvalue.
6. The DMD mode corresponding to  $\lambda$  is then given by

$$\varphi \triangleq Qw.$$

7. If desired, the DMD modes can be scaled in a number of ways, as described in Appendix A.



To see that  $\varphi$  computed by Algorithm 3 is an exact DMD mode, we first observe that because the columns of  $Q$  span the column space of  $Y$ , we can write  $Y = QQ^*Y$ , and thus  $A = QQ^*A$ . Then we find that

$$A\varphi = QQ^*A\varphi = QQ^*AQv = Q\tilde{A}_Qv = \lambda Qv = \lambda\varphi.$$

We emphasize that because the above algorithm requires both an SVD of  $X$  and a QR decomposition of the augmented matrix  $[X \ Y]$ , it is more costly than Algorithm 2, which should typically be used in practice. However, Algorithm 3 proves useful in showing how exact DMD can be interpreted as a simple extension of the standard algorithm for projected DMD (see Algorithm 1). Recall that in projected DMD, one constructs the matrix  $\tilde{A} = U^*AU$ , where  $U$  arises from the SVD of  $X$ . One then finds eigenvectors  $w$  of  $\tilde{A}$ , and the projected DMD modes have the form  $\hat{\varphi} = Uw$ . Algorithm 3 is precisely analogous, with  $U$  replaced by  $Q$ : one constructs  $\tilde{A}_Q = Q^*AQ$ , finds eigenvectors  $v$  of  $\tilde{A}_Q$ , and the exact DMD modes have the form  $\varphi = Qv$ . In the case of a sequential time series, where  $X$  and  $Y$  are given by (2), projected DMD projects  $A$  onto the space spanned by the first  $m$  vectors  $\{z_0, \dots, z_{m-1}\}$  (columns of  $X$ ), while exact DMD projects  $A$  onto the space spanned by all  $m+1$  vectors  $\{z_0, \dots, z_m\}$  (columns of  $X$  and  $Y$ ). In this sense, exact DMD is perhaps more natural, as it uses all of the data, rather than leaving out the last vector.

The case of a sequential time series is so common that it deserves special attention. For this case, we provide a variant of Algorithm 3 that is computationally advantageous.

**Algorithm 4** (Exact DMD, sequential time series).

1. Arrange the data  $\{z_0, \dots, z_m\}$  into matrices  $X$  and  $Y$ , as in (2).
2. Compute the (reduced) SVD of  $X$ , writing  $X = U\Sigma V^*$ .
3. Compute a vector  $q$  such that the columns of

$$Q = [U \quad q] \tag{14}$$

form an orthonormal basis for  $\{z_0, \dots, z_m\}$ . For instance,  $q$  may be computed by one step of the Gram-Schmidt procedure applied to  $z_m$ :

$$p = z_m - UU^*z_m, \quad q = \frac{p}{\|p\|}. \tag{15}$$

(If  $p = 0$ , then take  $Q = U$ ; in this case, exact DMD is identical to projected DMD, as discussed in Remark 5.)

4. Define the matrix  $\tilde{A} \triangleq U^*YV\Sigma^{-1}$ , as in (4).
5. Compute eigenvalues and eigenvectors of  $\tilde{A}$ , writing  $\tilde{A}w = \lambda w$ . Each nonzero eigenvalue  $\lambda$  is a DMD eigenvalue.
6. The DMD mode corresponding to  $\lambda$  is then given by

$$\varphi \triangleq Uw + \frac{1}{\lambda}qq^*Bw, \tag{16}$$

where  $B = YV\Sigma^{-1}$ .

7. If desired, the DMD modes can be scaled in a number of ways, as described in Appendix A.

Here, (16) is obtained from (9), noting that  $B = QQ^*B$ , so

$$Bw = QQ^*Bw = UU^*Bw + qq^*Bw = \lambda Uw + qq^*Bw.$$

From Algorithm 4, we see that an exact DMD mode  $\varphi$  can be viewed as a projected DMD mode  $\hat{\varphi}$  (calculated from (6)) plus a correction (the last term in (16)) that lies in the nullspace of  $A$ .

**3. Applications.** In this section we discuss the practical implications of a DMD framework based on Definition 1. Specifically, we extend DMD to nonsequential datasets using the fact that Definition 1 places no constraints on the structure of the data matrices  $X$  and  $Y$ . This allows for novel temporal sampling strategies that we show to increase computational efficiency and mitigate the effects of noise, respectively. We also present canonical examples that demonstrate the potential benefits and pitfalls of applying DMD to rank-deficient data. Such computations are not discussed in the existing DMD literature, but can be understood in our linear algebra-based framework.

**3.1. Nonsequential sampling.** In Section 2.2, there were no assumptions made on the data contained in  $X$  and  $Y$ . However, DMD is typically applied to data that come from a dynamical system, for instance one whose evolution is given by

$$z \mapsto f(z),$$

with  $z \in \mathbb{R}^n$ . Often, the data consist of direct measurements of the state  $z$ . More generally, one could measure some function of the state  $h(z)$  (see Section 4.1).

Consider a set of vectors  $\{z_1, \dots, z_m\}$ . These vectors need not comprise a sequential time series; they do not even need to be sampled from the same dynamical trajectory. There is no constraint on how the vectors  $z_k$  are sampled from the state space. We pair each vector  $z_k$  with its image under the dynamics  $f(z_k)$ . This yields a set of data pairs  $\{(z_1, f(z_1)), \dots, (z_m, f(z_m))\}$ . Arranging these vectors as in (7) yields matrices

$$X \triangleq [z_1 \ \cdots \ z_m], \quad Y = [f(z_1) \ \cdots \ f(z_m)], \quad (17)$$

with which we can perform either projected or exact DMD. The case of sequential time series, for which  $z_{k+1} = f(z_k)$ , is simply a special case.

**Remark 6.** We observe that only the pairwise correspondence of the columns of  $X$  and  $Y$  is important, and not the overall ordering. That is, permuting the order of the columns of  $X$  and  $Y$  has no effect on the matrix  $A = YX^+$ , or on the subsequent computation of DMD modes and eigenvalues, so long as the same permutation is applied to the columns of both  $X$  and  $Y$ . This is true even for data taken from a sequential time series (i.e.,  $X$  and  $Y$  as given in (2)).

Recall from Definition 1 and Remark 1 that DMD can be interpreted as an analysis of the best-fit linear operator relating  $X$  and  $Y$ . This operator relates the columns of  $X$  to those of  $Y$  in a pairwise fashion. For  $X$  and  $Y$  as in (17), the columns of  $X$  are mapped to those of  $Y$  by the (generally) nonlinear map  $f$ . Regardless, DMD provides eigenvalues and eigenvectors of the linear map  $A = YX^+$ . Then we can interpret DMD as providing an analysis of the best-fit linear approximation to  $f$ . If  $X$  and  $Y$  are linearly consistent, we can also interpret DMD using the formalism of Koopman operator theory (see Section 4.1).

The use of (17) in place of (2) certainly offers greater flexibility in the sampling strategies that can be employed for DMD (see Sections 3.2.3 and 3.2.4). However, it is important to note that for sequential time series, there exist memory-efficient variants of Algorithm 1 [18, 19]. These improved algorithms take advantage of the overlap in the columns of  $X$  and  $Y$  (when defined as in (2)) to avoid redundant

computations; the same strategies can be applied to Algorithms 2, 3, and 4. This is not possible for the more general definitions of  $X$  and  $Y$  given by (7) and (17).

**3.2. Examples.** The following section presents examples demonstrating the utility of a DMD theory based on Definition 1. The first two examples consider DMD computations involving rank-deficient datasets, which are not treated in the existing DMD literature. We show that in some cases, DMD can still provide meaningful information about the underlying dynamical system, but in others, the results can be misleading. The second two examples use the generalized approach described in Section 3.1 to perform DMD analysis using nonsequential datasets. First, we use nonuniform sampling to dramatically increase the efficiency of DMD computations. Then, we concatenate time series taken from multiple runs of an experiment, reducing the effects of noise.

**3.2.1. Stochastic dynamics.** Consider a system with stochastic dynamics

$$z_{k+1} = \lambda z_k + n_k, \quad (18)$$

where each  $z_k \in \mathbb{R}$ . We choose a decay rate  $\lambda = 0.5$  and let  $n_k$  be white noise with variance  $\sigma^2 = 10$ . (This system was first used as a test of DMD in [24].) Figure 1 (left) shows a typical trajectory for an initial condition  $z_0 = 0$ . If we apply DMD to this trajectory, we estimate a decay rate  $\tilde{\lambda} = 0.55$ . This is despite the fact that the nominal (noiseless) trajectory is simply given by  $z_k = 0$ ; a global, linear analysis of the trajectory shown in Figure 1 (left) would identify a stationary process ( $\tilde{\lambda} = 0$ ).

Because the existing DMD literature focuses on high-dimensional systems, existing DMD theory deals primarily with time series whose elements are linearly independent. As such, it cannot be applied to explain the ability of DMD to accurately estimate the dynamics underlying this noisy data (a rank-one time series). Recalling Definition 1, we can interpret DMD in terms of a linear operator that relates the columns of a data matrix  $X$  to those of  $Y$ , in column-wise pairs. Figure 1 (right) shows the time series from Figure 1 (left) plotted in this pairwise fashion. We see that though the data are noisy, there is clear evidence of a linear relationship between  $z_k$  and  $z_{k+1}$ . For rank-deficient data, DMD approximates the

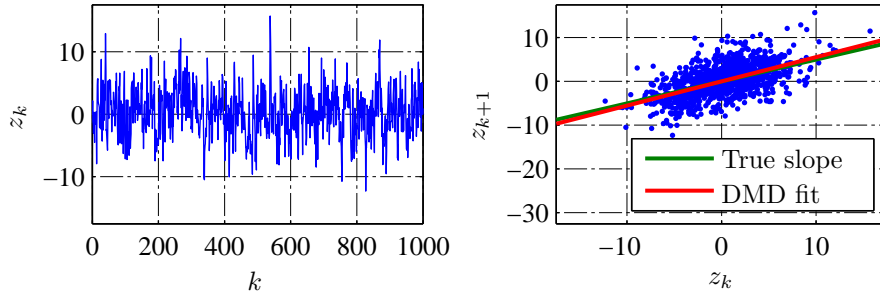


FIGURE 1. (Left) Typical trajectory of the noisy one-dimensional system governed by (18). (Right) Scatter plot showing the correlation of  $z_{k+1}$  and  $z_k$ . DMD is able to identify the relationship between future and past values of  $z$  even though the dataset is rank-deficient.

dynamics relating  $X$  and  $Y$  through a least-squares fit, and so it is no surprise that we can accurately estimate  $\lambda$  from this time series.

**3.2.2. Standing waves.** Because each DMD mode has a corresponding DMD eigenvalue (and thus a corresponding growth rate and frequency), DMD is often used to analyze oscillatory behavior, whether the underlying dynamics are linear or nonlinear. Consider data describing a standing wave:

$$z_k = \cos(k\theta)q, \quad k = 0, \dots, m, \quad (19)$$

where  $q$  is a fixed vector in  $\mathbb{R}^n$ . For instance, such data can arise from the linear system

$$\begin{aligned} u_{k+1} &= (\cos \theta)u_k - (\sin \theta)v_k \\ v_{k+1} &= (\sin \theta)u_k + (\cos \theta)v_k \end{aligned} \quad (u_0, v_0) = (q, 0), \quad (20)$$

where  $(u_k, v_k) \in \mathbb{R}^{2n}$ . If we measure only the state  $u_k$ , then we observe the standing wave (19). Such behavior can also arise in nonlinear systems, for instance by measuring only one component of a multi-dimensional limit cycle.

Suppose we compute DMD modes and eigenvalues from data satisfying (19). By construction, the columns of the data matrix  $X$  will be spanned by the single vector  $q$ . As such, the SVD of  $X$  will generate a matrix  $U$  with a single column, and the matrix  $\tilde{A}$  will be  $1 \times 1$ . Then there will be precisely one DMD eigenvalue  $\lambda$ . Since  $z$  is real-valued, then so is  $\lambda$ , meaning it captures only exponential growth/decay, and no oscillations. This is despite the fact that the original data are known to oscillate with a fixed frequency.

What has gone wrong? It turns out that, in this example, the data are not linearly consistent (see Definition 2). To see this, let  $x$  and  $y$  be vectors with components  $x_k = \cos(k\theta)$  and  $y_k = \cos((k+1)\theta)$  (for  $k = 0, \dots, m-1$ ), so the data matrices become

$$X = qx^T, \quad Y = qy^T.$$

Then  $X$  and  $Y$  are not linearly consistent unless  $\theta$  is a multiple of  $\pi$ .

For instance, the vector  $a = (-\cos \theta, 1, 0, \dots, 0)$  is in  $\mathcal{N}(X)$ , since  $x^T a = 0$ . However,  $y^T a = -\cos^2 \theta + \cos 2\theta = \sin^2 \theta$ , so  $a \notin \mathcal{N}(Y)$  unless  $\theta = j\pi$ . Also, note that if  $\theta = \pi$ , then the columns of  $X$  simply alternate sign, and in this case DMD yields the (correct) eigenvalue  $-1$ . As such, even though the data in this example arise from the linear system (20), there is no  $A$  such that  $Y = AX$  (by Theorem 2), and DMD fails to capture the correct dynamics. This example underscores the importance of linear consistency.

We note that in practice, we observe the same deficiency when the data do not satisfy (19) exactly, so long as the dynamics are dominated by such behavior (a standing wave). Thus the presence of random noise, which may increase the rank of the dataset, does not alleviate the problem. This is not surprising, as the addition of random noise should not enlarge the subspace in which the oscillation occurs. However, if we append the measurement with a time-shifted value, i.e., performing DMD on a sequence of vectors  $[z_k^T \ z_{k+1}^T]^T$ , then the data matrices  $X$  and  $Y$  become linearly consistent, and we are able to identify the correct oscillation frequency. (See Section 4.2 for an alternative motivation for this approach.)

**3.2.3. Nonuniform sampling.** Systems with a wide range of time scales can be challenging to analyze. If data are collected too slowly, dynamics on the fastest time scales will not be captured. On the other hand, uniform sampling at a high frequency can yield an overabundance of data, which can prove challenging to deal

with numerically. Such a situation can be handled using the following sampling strategy:

$$X \triangleq \begin{bmatrix} | & | & & | \\ z_0 & z_P & \cdots & z_{(m-1)P} \\ | & | & & | \end{bmatrix}, \quad Y \triangleq \begin{bmatrix} | & | & & | \\ z_1 & z_{P+1} & \cdots & z_{(m-1)P+1} \\ | & | & & | \end{bmatrix}, \quad (21)$$

where we again assume dynamics of the form  $z_{k+1} = f(z_k)$ . The columns of  $X$  and  $Y$  are separated by a single iteration of  $f$ , capturing its fastest dynamics. However, the tuning parameter  $P$  allows for a separation of time scales between the flow map iteration and the rate of data collection.

We demonstrate this strategy using a flow control example. Consider the flow past a two-dimensional cylinder, which is governed by the incompressible Navier–Stokes equations. We simulate the dynamics using the fast immersed boundary projection method detailed in [34, 35]. The (non-dimensionalized) equations of motion are

$$\begin{aligned} \frac{\partial \vec{u}}{\partial t} + (\vec{u} \cdot \nabla) \vec{u} &= -\nabla p + \frac{1}{\text{Re}} \nabla^2 \vec{u} + \int_{\partial \mathcal{B}} \vec{f}(\vec{x}) \delta(\vec{x} - \vec{\xi}) d\vec{\xi} \\ \nabla \cdot \vec{u} &= 0, \end{aligned}$$

where  $\vec{u}$  is the velocity,  $p$  is the pressure, and  $\vec{x}$  is the spatial coordinate. The Reynolds number  $\text{Re} \triangleq U_\infty D / \nu$  is a nondimensional parameter defined by the freestream velocity  $U_\infty$ , the cylinder diameter  $D$ , and the kinematic viscosity  $\nu$ .  $\partial \mathcal{B}$  is the union of the boundaries of any bodies in the flow.  $\vec{f}$  is a boundary force that can be thought of a Lagrange multiplier used to enforce the no-slip boundary condition.  $\delta$  is the Dirac delta function.

We consider a cylinder with diameter  $D = 1$  placed at the origin, in a flow with freestream velocity  $U_\infty = 1$  and Reynolds number  $\text{Re} = 100$ . Convergence tests show that our simulations are accurate for an inner-most domain with  $(x, y) \in [15, 15] \times [-5, 5]$  and a  $1500 \times 500$  mesh of uniformly-spaced points. At this Reynolds number, the flow is globally unstable. As a step towards computing a reduced-order model using BPOD [5], we restrict the linearized dynamics to their stable subspace. The system is actuated using a disk of vertical velocity downstream of the cylinder and sensed using localized measurements of vertical velocity placed along the flow centerline. This setup is based on flow control benchmark proposed in [36] and is illustrated in Figure 2 (left).

The impulse response of this system is shown in Figure 2 (right). We see that from  $t = 200$  to  $t = 500$ , the dynamics exhibit both a fast and slow oscillation. Suppose we want to identify the underlying frequencies and corresponding modes using DMD. For instance, this would allow us to implement the more efficient and more accurate analytic tail variant of BPOD [18]. In order to capture the fast frequency, we must sample the system every 50 timesteps, with each timestep corresponding to  $\Delta t = 0.02$ . (This is in order to satisfy the Nyquist–Shannon sampling criterion.) As such, we let  $z_k = z(50k\Delta t)$ .

Figure 3 and Table 1 compare the DMD eigenvalues computed using **uniform sampling and nonuniform sampling**. (Referring back to (21), the former corresponds to  $P = 1$  and the latter to  $P = 10$ .) We see that the dominant eigenvalues agree, with less than 10% error in all cases; we use the DMD eigenvalues computed with uniform sampling as truth values. However, the larger errors occur for modes with norms on the order of  $10^{-5}$ , two orders of magnitude smaller than those of the dominant two DMD modes. As such, these modes have negligible contribution to

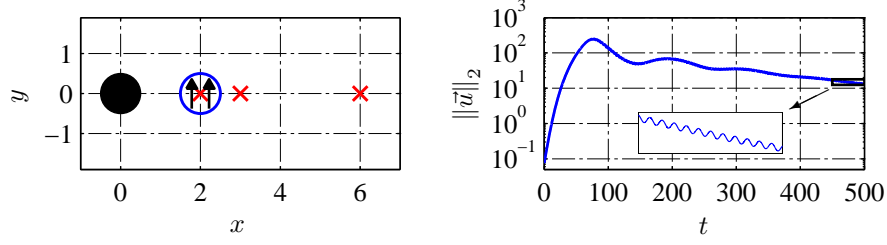


FIGURE 2. (Left) Schematic showing the placement of sensors ( $\times$ ) and actuators ( $\circ$ ) used to control the flow past a two-dimensional cylinder. (Right) Kinetic energy of the corresponding impulse response (restricted to the stable subspace). After an initial period of non-normal growth, oscillations with both short and long time scales are observed.

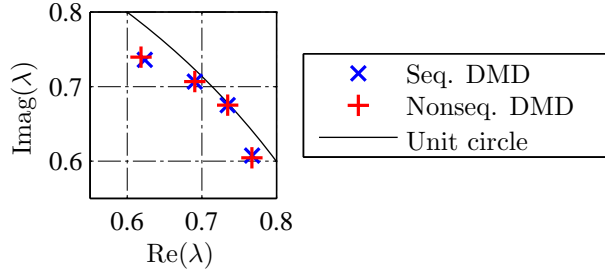


FIGURE 3. DMD estimates of the eigenvalues governing the decay of the impulse response shown in Figure 2 (right). The slowest decaying eigenvalues are captured well with both uniform sampling (sequential DMD) and nonuniform sampling (nonsequential DMD).

the evolution of the impulse response, and the error in the corresponding eigenvalues is not significant. The dominant DMD modes show similar agreement, as seen in Figure 4. This agreement is achieved despite using 90% less data in the nonuniform sampling case, which corresponds to an 85.8% reduction in computation time.

**3.2.4. Combining multiple trajectories.** DMD is often applied to experimental data, which are typically noisy. While filtering or phase averaging can be done to eliminate noise prior to DMD analysis, this is not always desirable, as it may remove features of the true dynamics. In POD analysis, the effects of noise can be averaged out by combining multiple trajectories in a single POD computation. We can take the same approach in DMD analysis using (17).

Consider multiple dynamic trajectories, indexed by  $j$ :  $\{z_k^j\}_{k=0}^{m_j}$ . These could be multiple runs of an experiment, or particular slices of a single, long trajectory. (The latter might be useful in trying to isolate the dynamics of a recurring dynamic event.) Suppose there are a total of  $J$  trajectories. DMD can be applied to the

TABLE 1. Comparison of DMD eigenvalues\*

Frequency			Growth rate		
Seq. DMD	Nonseq. DMD	Error	Seq. DMD	Nonseq. DMD	Error
0.118	0.118	0.00%	0.998	0.998	0.00%
0.127	0.127	0.01%	0.988	0.988	0.00%
0.107	0.106	3.40%	0.979	0.977	2.10%
0.138	0.139	7.50%	0.964	0.964	0.05%

\* Row order corresponds to decreasing mode norm.

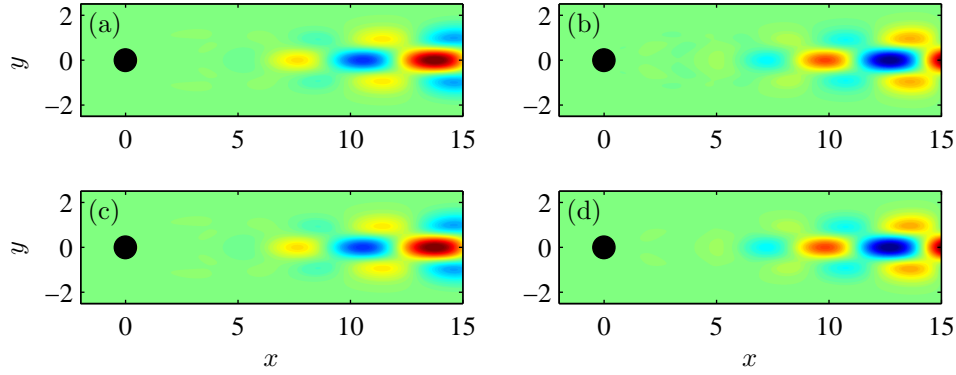


FIGURE 4. Comparison of dominant DMD modes computed from the impulse response shown in Figure 2 (right), illustrated using contours of vorticity. For each of the dominant frequencies, modes computed using nonuniform sampling (nonsequential DMD; bottom row) match those computed using uniform sampling (sequential DMD; top row). (For brevity, only the real part of each mode is shown; similar agreement is observed in the imaginary parts.) (a)  $f = 0.118$ , uniform sampling; (b)  $f = 0.127$ , uniform sampling; (c)  $f = 0.118$ , nonuniform sampling; (d)  $f = 0.127$ , nonuniform sampling.

entire ensemble of trajectories by defining

$$\begin{aligned}
 X &\triangleq \begin{bmatrix} | & & | & & | & & | & & | \\ z_0^0 & \dots & z_{m_0-1}^0 & z_0^1 & \dots & z_{m_1-1}^1 & \dots & z_0^J & \dots & z_{m_J-1}^J \\ | & & | & & | & & | & & | \end{bmatrix}, \\
 Y &\triangleq \begin{bmatrix} | & & | & & | & & | & & | \\ z_1^0 & \dots & z_{m_0}^0 & z_1^1 & \dots & z_{m_1}^1 & \dots & z_1^J & \dots & z_{m_J}^J \\ | & & | & & | & & | & & | \end{bmatrix}.
 \end{aligned} \tag{22}$$

We demonstrate this approach using experimental data from a bluff-body wake experiment. A finite-thickness flat plate with an elliptical leading edge is placed in a uniform oncoming flow. Figure 5 shows a schematic of the experimental setup. We capture snapshots of the velocity field in the wake behind the body using a time-resolved particle image velocimetry (PIV) system. (For details on the PIV



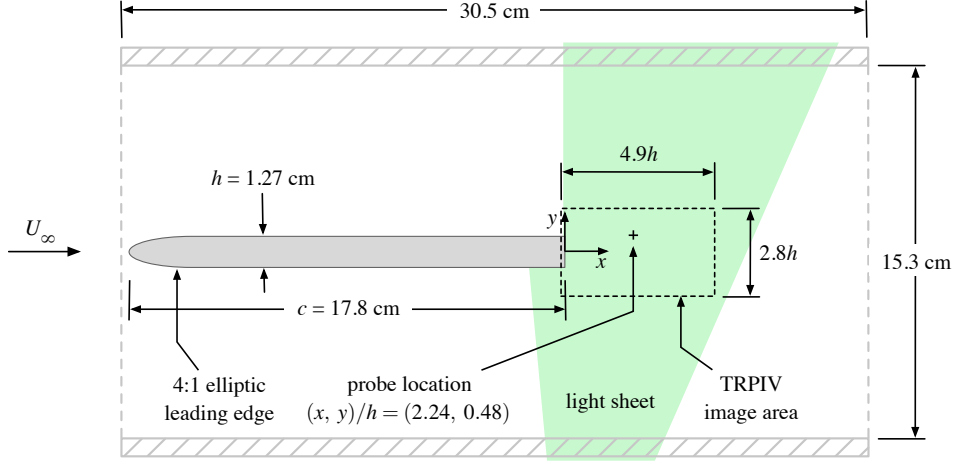


FIGURE 5. Schematic of setup for bluff-body wake experiment.

data acquisition, see [37].) Multiple experimental runs are conducted, with approximately 1,400 velocity fields captured in each run. This corresponds to the maximum amount of data that can be collected per run.

Due to the high Reynolds number ( $Re = 50,000$ ), the flow is turbulent. As such, though we observe a standard von Kármán vortex street (see Figure 6), the familiar vortical structures are contaminated by turbulent fluctuations. Figure 7 (left) shows a DMD spectrum computed using PIV data from a single experimental run.\*\* The spectrum is characterized by a harmonic set of peaks, with the dominant peak corresponding to the wake shedding frequency. The corresponding modes are shown in Figure 8 (a–c). We see that the first pair of modes (see Figure 8 (a)) exhibits approximate top-bottom symmetry, with respect to the centerline of the body. The second pair of modes (see Figure 8 (b)) shows something close to top-bottom antisymmetry, though variations in the vorticity contours make this antisymmetry inexact. The third pair of modes (see Figure 8 (c)) also shows approximate top-bottom symmetry, with structures that are roughly spatial harmonics of those seen in the first mode pair.

These modal features are to be expected, based on two-dimensional computations of a similar flow configuration [38]. However, when computed from noise-free simulation data, the symmetry/antisymmetry of the modes is more exact. Figures 7 (right) and 8 (d–g) show that when five experimental runs are used, the experimental DMD results improve, more closely matching computational results. In the DMD spectrum (see Figure 7 (right)), we again observe harmonic peaks, with a fundamental frequency corresponding to the shedding frequency. The peaks are more isolated than those in Figure 7 (left); in fact, we observe a fourth frequency peak, which is not observed in the single-run computation. The modal structures,

\*\*Instead of simply plotting the mode norms against their corresponding frequencies, as is generally done, we first scale the mode norms by  $\lambda^m$ . This reduces the height of spectral peaks corresponding to modes with large norm but quickly decaying eigenvalues. For dynamics known to lie on an attractor, such peaks can be misleading; they do not contribute to the long-time evolution of the system.

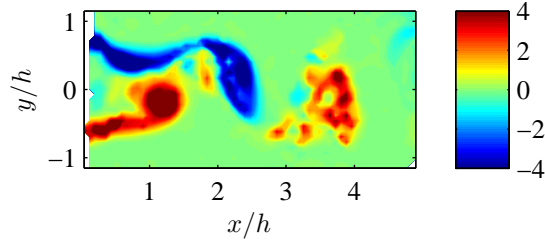


FIGURE 6. Typical vorticity field from the bluff-body wake experiment depicted in Figure 5. A clear von Kármán vortex street is observed, though the flow field is contaminated by turbulent fluctuations.

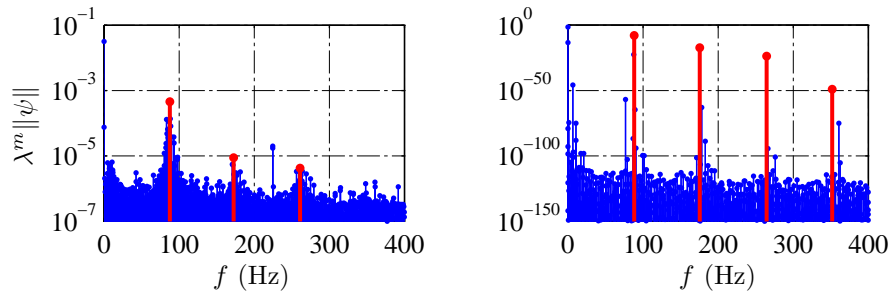


FIGURE 7. Comparison of DMD spectra computed using a single experimental run (left) and five experimental runs (right). When multiple runs are used, the spectral peaks are more isolated and occur at almost exactly harmonic frequencies. Furthermore, a fourth harmonic peak is identified; this peak is obscured in the single-run DMD computation. (Spectral peaks corresponding to modes depicted in Figure 8 are shown in red.)

shown in Figure 8 (d–g), display more obvious symmetry and antisymmetry, respectively. The structures are also smoother and more elliptical.

**4. Connections to other methods.** In this section we discuss how DMD relates to other methods. First, we show that our definition of DMD preserves, and even strengthens, the connections between DMD and Koopman operator theory. Without these connections, the use of DMD to analyze nonlinear dynamics appears dubious, since there seems to be an underlying assumption of (approximately) linear dynamics (see Section 3.1), as in (1). One might well question whether such an approximation would characterize a nonlinear system in a meaningful way. However, so long as DMD can be interpreted as an approximation to Koopman spectral analysis, there is a firm theoretical foundation for applying DMD in analyzing nonlinear dynamics.

Second, we explore the links between DMD and the eigensystem realization algorithm (ERA). The close relationship between the two methods provides motivation for the use of strategies from the ERA in DMD computations where rank is a problem. Finally, we show that under certain assumptions, DMD is equivalent to

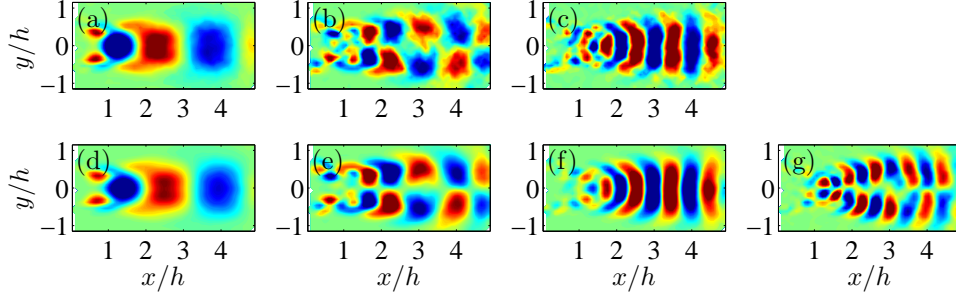


FIGURE 8. Representative DMD modes, illustrated using contours of vorticity. (For brevity, only the real part of each mode is shown.) The modes computed using multiple runs (bottom row) have more exact symmetry/antisymmetry and smoother contours. Furthermore, with multiple runs four dominant mode pairs are identified; the fourth spectral peak is obscured in the single-run computation (see Figure 7). (a)  $f = 87.75$  Hz, single run; (b)  $f = 172.6$  Hz, single run; (c)  $f = 261.2$  Hz, single run; (d)  $f = 88.39$  Hz, five runs; (e)  $f = 175.6$  Hz, five runs; (f)  $f = 264.8$  Hz, five runs; (g)  $f = 351.8$  Hz, five runs.

linear inverse modeling (LIM), a method developed in the climate science community decades ago. The link between the two methods suggests that practitioners of DMD may benefit from an awareness of the LIM literature, and vice versa.

**4.1. Koopman spectral analysis.** We briefly introduce the concepts of Koopman operator theory below and discuss how they relate to the theory outlined in Section 2. The connections between Koopman operator theory and projected DMD were first explored in [3], but only in the context of sequential time series. Here, we extend those results to more general datasets, doing so using exact DMD (defined in Section 2.2).

Consider a discrete-time dynamical system

$$z \mapsto f(z), \quad (23)$$

where  $z$  is an element of a finite-dimensional manifold  $M$ . The Koopman operator  $\mathcal{K}$  acts on scalar functions  $g : M \rightarrow \mathbb{R}$  or  $\mathbb{C}$ , mapping  $g$  to a new function  $\mathcal{K}g$  given by

$$\mathcal{K}g(z) \triangleq g(f(z)). \quad (24)$$

Thus the Koopman operator is simply a composition or pull-back operator. We observe that  $\mathcal{K}$  acts linearly on functions  $g$ , even though the dynamics defined by  $f$  may be nonlinear.

As such, suppose  $\mathcal{K}$  has eigenvalues  $\lambda_j$  and eigenfunctions  $\theta_j$ , which satisfy

$$\mathcal{K}\theta_j(z) = \lambda_j\theta_j(z), \quad j = 0, 1, \dots \quad (25)$$

An *observable* is simply a function on the state space  $M$ . Consider a vector-valued observable  $h : M \rightarrow \mathbb{R}^n$  or  $\mathbb{C}^n$ , and expand  $h$  in terms of these eigenfunctions as

$$h(z) = \sum_{j=0}^{\infty} \theta_j(z) \tilde{\varphi}_j, \quad (26a)$$

where  $\tilde{\varphi}_j \in \mathbb{R}^n$  or  $\mathbb{C}^n$ . (We assume each component of  $h$  lies in the span of the eigenfunctions.) We refer to the vectors  $\tilde{\varphi}_j$  as *Koopman modes* (after [3]).

Applying the Koopman operator, we find that

$$h(f(z)) = \sum_{j=0}^{\infty} \lambda_j \theta_j(z) \tilde{\varphi}_j. \quad (26b)$$

For a sequential time series, we can repeatedly apply the Koopman operator (see (24) and (25)) to find

$$h(f^k(z)) = \sum_{j=0}^{\infty} \lambda_j^k \theta_j(z) \tilde{\varphi}_j. \quad (27)$$

A set of modes  $\{\tilde{\varphi}_j\}$  and eigenvalues  $\{\lambda_j\}$  must satisfy (26) in order to be considered Koopman modes and eigenvalues.

Now consider a set of arbitrary initial states  $\{z_0, z_1, \dots, z_{m-1}\}$  (not necessarily a trajectory of the dynamical system (23)), and let

$$x_k = h(z_k), \quad y_k = h(f(z_k)).$$

As before, construct data matrices  $X$  and  $Y$ , whose columns are  $x_k$  and  $y_k$ . This is similar to (17), except that here we measure an observable, rather than the state itself. As long as the data matrices  $X$  and  $Y$  are linearly consistent (see Definition 2), we can determine a relationship between DMD modes and Koopman modes, as follows.

Suppose we compute DMD modes from  $X$  and  $Y$ , giving us eigenvectors and eigenvalues that satisfy

$$A\varphi_j = \lambda_j \varphi_j,$$

where  $A = YX^+$ . If the matrix  $A$  has a full set of eigenvectors, then we can expand each column  $x_k$  of  $X$  as

$$h(z_k) = x_k = \sum_{j=0}^{n-1} c_{jk} \varphi_j, \quad (28a)$$

for some constants  $c_{jk}$ . For linearly consistent data, we have  $Ax_k = y_k$ , by Theorem 2, so

$$\begin{aligned} h(f(z_k)) &= y_k = Ax_k \\ &= \sum_{j=0}^{n-1} c_{jk} A\varphi_j \\ &= \sum_{j=0}^{n-1} \lambda_j c_{jk} \varphi_j. \end{aligned} \quad (28b)$$

Comparing to (26), we see that in the case of linearly consistent data matrices (and diagonalizable  $A$ ), the DMD modes correspond to Koopman modes, and the DMD eigenvalues to Koopman eigenvalues, where the constants  $c_{jk}$  are given by  $c_{jk} = \theta_j(z_k)$ . (We note, however, that we typically do not know the states  $z_k$ , nor do we know the Koopman eigenfunctions  $\theta_j$ .) If the data arise from a sequential time series, we can scale the modes to subsume the constants  $c_{jk}$ , following [22]. (For more details see Appendix A.)

The relationship between Koopman modes and DMD modes is similar to that established in [3] (and discussed further in [22]), but the present result differs in some

important respects. First, the result in [3] uses projected DMD modes (see Section 2.1) and requires a sequential time series; here, we use exact DMD modes (see Section 2.2) and do not require a sequential time series. Furthermore, [3] assumes the vectors  $x_k$  are linearly independent; here, we impose the weaker requirement of linear consistency.

We note that when  $n$  is large, it may be impractical to compute all of the DMD modes. For instance, when  $n \gg m$ ,  $A$  will have a large nullspace, and one might use Algorithm 2 to compute only those DMD modes with nonzero eigenvalues. Suppose the rank of  $A$  is  $r$  and we write

$$x_k = \sum_{j=0}^{r-1} c_{jk} \varphi_j + \sum_{j=r}^{n-1} c_{jk} \varphi_j, \quad (29a)$$

where the first sum contains only DMD modes with nonzero eigenvalues. We still have

$$y_k = \sum_{j=0}^{r-1} \lambda_j c_{jk} \varphi_j, \quad (29b)$$

since all modes in the second sum have zero eigenvalues. As such, those modes don't contribute to the dynamics, and we can neglect those vectors as an error term that gets projected out by the dynamics. Contrast this with the DMD reconstruction of a sequential time series using projected DMD, where the residual instead appears at the end of the time series (see 3.13–3.14 in [3]).

Though the Koopman analogy provides a firm mathematical foundation for applying DMD to data generated by nonlinear systems, it is limited by the fact that it relies on (28). The derivation of this equation requires making a number of assumptions, namely that the data are linearly consistent and that the matrix  $A$  has a full set of eigenvectors (e.g., this holds when the eigenvalues of  $A$  are distinct). When these assumptions do not hold, there is no guarantee that DMD modes will closely approximate Koopman modes. For instance, in some systems DMD modes and eigenvalues closely approximate those of the Koopman operator near an attractor, but not far from it [26]. DMD may also perform poorly when applied to dynamics whose Koopman spectral decomposition contains Jordan blocks [26]. In contrast, an understanding of DMD built on Definition 1 holds even when these conditions break down.

**4.2. The eigensystem realization algorithm (ERA).** The ERA is a control-theoretic method for system identification and model reduction [27, 28, 29]. Applicable to linear systems, the ERA takes input-output data and from them computes a minimal realization of the underlying dynamics. In this section, we show that while DMD and the ERA were developed in different contexts, they are closely related: the low-order linear operators central to each method are related by a similarity transform. This connection suggests that strategies used in ERA computations could be leveraged for gain in DMD computations. Specifically, it provides a motivation for appending the data matrices  $X$  and  $Y$  with time-shifted data to overcome rank problems (as suggested in Section 3.2.2).

Consider the linear, time-invariant system

$$\begin{aligned} x_{k+1} &= Ax_k + Bu_k \\ y_k &= Cx_k + Du_k, \end{aligned} \quad (30)$$

where  $x \in \mathbb{R}^n$ ,  $u \in \mathbb{R}^p$ , and  $y \in \mathbb{R}^q$ . (The matrix  $A$  defined here is not necessarily related to the one defined in Section 2.) We refer to  $x$  as the state of the system,  $u$  as the input, and  $y$  as the output.

The goal of the ERA is to identify the dynamics of the system (30) from a time history of  $y$ . Specifically, in the ERA we sample outputs from the impulse response of the system. We collect the sampled outputs in two sets

$$\mathcal{H} \triangleq \{CB, CA^P B, \dots, CA^{(m-1)P} B\}$$

$$\mathcal{H}' \triangleq \{CAB, CA^{P+1} B, \dots, CA^{(m-1)P+1} B\},$$

where we sample the impulse response every  $P$  steps. The elements of  $\mathcal{H}$  and  $\mathcal{H}'$  are commonly referred to as Markov parameters of the system (30).

We then form the Hankel matrix by stacking the elements of  $\mathcal{H}$  as

$$H \triangleq \begin{bmatrix} CB & CA^P B & \dots & CA^{m_c P} B \\ CA^P B & CA^{2P} B & \dots & CA^{(m_c+1)P} B \\ \vdots & \vdots & \ddots & \vdots \\ CA^{m_o P} B & CA^{(m_o+1)P} B & \dots & CA^{(m_c+m_o)P} B \end{bmatrix}, \quad (31)$$

where  $m_c$  and  $m_o$  can be chosen arbitrarily, subject to  $m_c + m_o = m - 1$ . The time-shifted Hankel matrix is built from the elements of  $\mathcal{H}'$  in the same way:

$$H' \triangleq \begin{bmatrix} CAB & CA^{P+1} B & \dots & CA^{m_c P+1} B \\ CA^{P+1} B & CA^{2P+1} B & \dots & CA^{(m_c+1)P+1} B \\ \vdots & \vdots & \ddots & \vdots \\ CA^{m_o P+1} B & CA^{(m_o+1)P+1} B & \dots & CA^{(m_c+m_o)P+1} B \end{bmatrix}. \quad (32)$$

Next, we compute the (reduced) SVD of  $H$ , giving us

$$H = U \Sigma V^*.$$

Let  $U_r$  consist of the first  $r$  columns of  $U$ . Similarly, let  $\Sigma_r$  be the upper left  $r \times r$  submatrix of  $\Sigma$  and let  $V_r$  contain the first  $r$  columns of  $V$ . Then the  $r$ -dimensional ERA approximation of (30) is given by the reduced-order system

$$\begin{aligned} \xi_{k+1} &= A_r \xi_k + B_r u_k \\ \eta_k &= C_r \xi_k + D_r u_k, \end{aligned} \quad (33)$$

where

$$\begin{aligned} A_r &= \Sigma_r^{-1/2} U_r^* H' V_r \Sigma_r^{-1/2} \\ B_r &= \text{the first } p \text{ columns of } \Sigma_r^{1/2} V_r^* \\ C_r &= \text{the first } q \text{ rows of } U_r \Sigma_r^{1/2} \\ D_r &= D. \end{aligned} \quad (34)$$

Now, suppose we use  $H$  and  $H'$  to compute DMD modes and eigenvalues as in Algorithm 2, with  $X = H$  and  $Y = H'$ . Recall from (5) that the DMD eigenvalues and modes are determined by the eigendecomposition of the operator

$$\tilde{A} = U^* H' V \Sigma^{-1},$$

with  $U$ ,  $\Sigma$ , and  $V$  defined as above. Comparing with (34), we see that if one keeps all of the modes from ERA (i.e., choosing  $r$  equal to the rank of  $H$ ), then  $A_r$  and  $\tilde{A}$

are related by a similarity transform, with

$$A_r = \Sigma^{-1/2} \tilde{A} \Sigma^{1/2},$$

so  $A_r$  and  $\tilde{A}$  have the same eigenvalues. Furthermore, if  $v$  is an eigenvector of  $A_r$ , with

$$A_r v = \lambda v,$$

then  $w = \Sigma^{1/2} v$  is an eigenvector of  $\tilde{A}$ , since

$$\tilde{A} w = \tilde{A} \Sigma^{1/2} v = \Sigma^{1/2} A_r v = \lambda \Sigma^{1/2} v = \lambda w.$$

Then  $w$  can be used to construct either the exact or projected DMD modes (see (9) and (6)).

We see that algorithmically, the ERA and DMD are closely related: given two matrices  $H$  and  $H'$ , applying the ERA produces a reduced-order operator  $A_r$  that can be used to compute DMD modes and eigenvalues. However, the ERA was originally developed to characterize linear input-output systems, for which (33) approximates (30). (In practice, it is often applied to experimental data, for which the underlying dynamics may be only approximately linear.) On the other hand, DMD is designed to analyze data generated by any dynamical system; the system can be nonlinear and may have inputs (e.g.,  $x_{k+1} = f(x_k, u_k)$ ) or not (e.g.,  $x_{k+1} = f(x_k)$ ).

We note that in the ERA, we collect two sets of data  $\mathcal{H}$  and  $\mathcal{H}'$ , then arrange that data in matrices  $H$  and  $H'$ . In doing so we are free to choose the values of  $m_c$  and  $m_o$ , which determine the shapes of  $H$  and  $H'$  in (31) and (32). Interpreting DMD using the Koopman formalism of Section 4.1, each column of  $H$  corresponds to a particular value of an observable, where the observable function is a vector of outputs at  $m_o + 1$  different timesteps. Each column of the matrix  $H'$  then contains the value of this observable at the next timestep. A more typical application of DMD would use  $m_o = 0$  and  $m_c = m - 1$ . Allowing for  $m_o > 0$  is equivalent to appending the data matrices with rows of time-shifted Markov parameters. Doing so can enlarge the rank of  $H$  and  $H'$ , increasing the accuracy of the reduced-order system (33) for the ERA. For DMD, it can overcome the rank limitations that prevent the correct characterization of standing waves (as suggested in Section 3.2.2).

**4.3. Linear inverse modeling (LIM).** In this section, we investigate the connections between DMD and LIM. To set up this discussion, we briefly introduce and define a number of terms used in the climate science literature. This is followed by a more in-depth description of LIM. Finally, we show that under certain conditions, LIM and DMD are equivalent.

**4.3.1. Nomenclature.** *Empirical orthogonal functions* (EOFs) were first introduced in 1956 by Lorenz [39]. Unique to the climate science literature, EOFs simply arise from the application of principal component analysis (PCA) [40, 41, 42] to meteorological data [30]. As a result, EOF analysis is equivalent to PCA, and thus also to POD and SVD. (We note that in PCA and EOF analysis, the data mean is always subtracted, so that the results can be interpreted in terms of variances; this is often done for POD as well.)

In practice, EOFs are often used as a particular choice of *principal interaction patterns* (PIPs), a concept introduced in 1988 by Hasselmann [43]. The following discussion uses notation similar to that found in [44], which provides a nice review of PIP concepts. Consider a dynamical system with a high-dimensional state



$x(t) \in \mathbb{R}^n$ . In some cases, such a system may be approximately driven by a lower-dimensional system with state  $z(t) \in \mathbb{R}^r$ , where  $r < n$ . To be precise, we say that  $x$  and  $z$  are related as follows:

$$\begin{aligned} z_{k+1} &= F(z_k; \alpha) + \text{noise} \\ x_k &= Pz_k + \text{noise}, \end{aligned}$$

where  $\alpha$  is a vector of parameters. From this we see that given a knowledge of  $z$  and its dynamics,  $x$  is completely specified by the static map  $P$ , aside from the effects of noise. Though in general  $P$  cannot be inverted, given a measurement of  $x$ , we can approximate  $z$  using a least-squares fit:

$$z_k = (P^T P)^{-1} P^T x_k.$$

In climate science, the general approach of modeling the dynamics of a high-dimensional variable  $x$  through a lower-dimensional variable  $z$  is referred to as *inverse modeling*. The inverse model described above requires definitions of  $F$ ,  $P$ , and  $\alpha$ . Generally,  $F$  is chosen based on physical intuition. Once that choice is made,  $P$  and  $\alpha$  are fitted simultaneously. The PIPs are the columns of  $P$  for the choice of  $P$  (and  $\alpha$ ) that minimizes the error

$$\epsilon(P, \alpha) \triangleq E \left( \left\| x_{k+1} - x_k - P \left( F(z_k; \alpha) - z_k \right) \right\| \right),$$

where  $E$  is the expected value operator [44]. In general, the choice of  $P$  is not unique.

Hasselmann also introduced the notion of *principal oscillation patterns* (POPs) in his 1988 paper [43]. Again, we use the notation found in [44]. Consider a system with unknown dynamics. We assume that we can approximate these dynamics with a linear system

$$x_{k+1} = Ax_k + \text{noise}.$$

If we multiply both sides by  $x_k^T$  and take expected values, we can solve for  $A$  as

$$A = E(x_{k+1} x_k^T) E(x_k x_k^T)^{-1}. \quad (35)$$

The eigenvectors of  $A$  are referred to as POPs. That is, POPs are eigenvectors of a particular linear approximation of otherwise unknown dynamics.

Even within the climate science literature, there is some confusion between PIPs and POPs. This is due to the fact that POPs can be considered a special case of PIPs. In general, PIPs are basis vectors spanning a low-dimensional subspace useful for reduced-order modeling. Suppose we model our dynamics with the linear approximation described above, and do not reduce the order of the state. If we then express the model in its eigenvector basis, we are choosing our PIPs to be POPs.

**4.3.2. Linear Markov models/linear inverse modeling (LIM).** In 1989, Penland derived a method for computing a linear, discrete-time system that approximates the trajectory of a stochastic, continuous-time, linear system, which he referred to as a linear Markov model [30]. We describe this method, which came to be known as LIM, using the notation found in [31]. Consider an  $n$ -dimensional Markov process

$$\frac{dx}{dt} = Bx(t) + \xi(t),$$

where  $\xi(t)$  is white noise with covariance

$$Q = E(\xi(t)\xi^T(t)).$$

We assume the mean of the process has been removed. The covariance of  $x$  is given by

$$\Lambda = E(x(t)x^T(t)). \quad (36)$$

One can show that the following must hold:

$$\begin{aligned} B\Lambda + \Lambda B^T + Q &= 0 \\ E(x(t+\tau)x^T(t)) &= \exp(B\tau)\Lambda. \end{aligned}$$

(See [30] for details.)

Defining the Green's function

$$\begin{aligned} G(\tau) &\triangleq \exp(B\tau) \\ &= E(x(t+\tau)x^T(t))\Lambda^{-1}, \end{aligned} \quad (37)$$

we can say that given a state  $x(t)$ , the most probable state time  $\tau$  later is

$$x(t+\tau) = G(\tau)x(t).$$

The operator  $G(\tau)$  is computed from snapshots of the continuous-time system and has the same form as the linear approximation used in POP analysis (see (35)). We note that we arrive at the same model if we apply linear stochastic estimation to snapshots of the state  $x$ , taking  $x(t)$  and  $x(t+\tau)$  to be the unconditional and conditional variables, respectively. (This was done in [37] to identify a model for the evolution of POD coefficients in a fluid flow.)

When this approach is applied to a nonlinear system, it can be shown that  $G(\tau)$  is equivalent to a weighted average of the nonlinear dynamics, evaluated over an ensemble of snapshots [45]. This is in contrast to a typical linearization, which involves evaluating the Jacobian of the dynamics at a fixed point. If the true dynamics are nearly linear, these two approaches will yield nearly the same model. However, if nonlinear effects are significant,  $G(\tau)$  will be closer to the ensemble average, and arguably a better model than a traditional linearization [45].

In [31], this method was applied to compute a linear Markov model in the space of EOF coefficients. This is an example of inverse modeling (equivalently, PIP analysis); a high-dimensional variable is modeled via a projection onto a lower-dimensional EOF subspace. Due to the assumption of linear dynamics, this approach came to be known as *linear inverse modeling*. The combination of PIP and POP concepts in this early work has contributed to the continuing confusion between PIPs and POPs in the climate science literature today.

**4.3.3. Equivalence to projected DMD.** In both exact and projected DMD, the DMD eigenvalues are given by the eigenvalues of the projected linear operator  $\tilde{A}$  (see (4)). The projected DMD modes are computed by lifting the eigenvectors of  $\tilde{A}$  to the original space via the left singular vectors  $U$  (see (6)). In [31], the eigendecomposition of a low-order linear model  $G(\tau)$  is computed and the low-order eigenvectors lifted to the original space via EOFs, in the same way as in projected DMD. The similarity in these two approaches is obvious. Recall that left singular vectors and EOFs are equivalent, so long as they are computed from the same data. Then to prove that LIM-based eigenvector analysis is equivalent to projected DMD, we simply have to show the equivalence of  $G(\tau)$  and  $\tilde{A}$ .

Consider two  $n \times m$  data matrices  $X$  and  $Y$ , with columns  $x_j = x(t_j)$  and  $y_j = x(t_j + \tau)$ , respectively.  $X$  and  $Y$  may or may not share columns. As in (3), we

assume that the EOFs to be used for LIM are computed from  $X$  alone, giving us

$$X = U\Sigma V^*,$$

where the columns of  $U$  are the EOFs. The EOF coefficients of  $X$  and  $Y$  are given by

$$\hat{X} = U^*X, \quad \hat{Y} = U^*Y, \quad (38)$$

whose columns we denote by  $\hat{x}_j$  and  $\hat{y}_j$ , respectively.

In order to show that  $G(\tau)$  and  $\tilde{A}$  are equivalent, we must reduce (37) to (4). Because we are interested in the equivalence of LIM and projected DMD when the former is performed in the space of EOF coefficients, we replace all instances of  $x$  in (36) and (37) with  $\hat{x}$ . Recall that the expected value of  $a$ , for an ensemble  $\{a_j\}_{j=0}^{m-1}$ , is given by

$$E(a_j) \triangleq \frac{1}{m} \sum_{j=0}^{m-1} a_j. \quad (39)$$

Then we can rewrite (36) as

$$\begin{aligned} \Lambda &= \frac{1}{m} \sum_{j=0}^{m-1} \hat{x}_j \hat{x}_j^* \\ &= \frac{1}{m} \hat{X} \hat{X}^* \\ &= \frac{1}{m} U^* X X^* U \\ &= \frac{1}{m} U^* U \Sigma^2 \\ &= \frac{1}{m} \Sigma^2, \end{aligned}$$

using the fact that  $XX^*U = U\Sigma^2$ , by the definition of left singular vectors. This result, along with (3), allows us to rewrite (37) as

$$\begin{aligned} G(\tau) &= \left( \frac{1}{m} \sum_{j=0}^{m-1} \hat{y}_j \hat{x}_j^* \right) (m\Sigma^{-2}) \\ &= \hat{Y} \hat{X}^* \Sigma^{-2} \\ &= U^* Y X^* U \Sigma^{-2} \\ &= U^* Y V \Sigma U^* U \Sigma^{-2} \\ &= U^* Y V \Sigma^{-1}. \end{aligned}$$

(Recall that  $x(t_j + \tau) = y_j$ , and  $x(t_j) = x_j$ .) From (4), we then have  $G(\tau) = \tilde{A}$ , and we see that DMD and LIM are built on the same low-dimensional, approximating linear dynamics.

We emphasize that this equivalence relies on a number of assumptions. First, we assume that we perform LIM in the space of EOF coefficients. Second, we assume that the EOFs are computed from  $X$  alone. This may not be an intuitive choice if  $X$  and  $Y$  are completely distinct, but for a sequential snapshot sequence where  $X$  and  $Y$  differ by a single column, this is not a significant difference.

Given these assumptions, the equivalence of projected DMD and LIM gives us yet another way to interpret DMD analysis. If the data mean is removed, then the low-order map that generates the DMD eigenvalues and eigenvectors is simply the one that yields the statistically most likely state in the future. (This is the case for both exact and projected DMD, as both are built on the same low-order linear map.) In a small sense, the DMD framework is more general, as the interpretation provided by Definition 1 holds even for data that are not mean-subtracted. Then again, in LIM the computation of the EOFs is completely divorced from the modeling procedure, allowing for a computation using both  $X$  and  $Y$ . Nevertheless, the similarities between the two methods suggests that practitioners of DMD would be well-served in studying and learning from the climate science/LIM literature.

**5. Conclusions.** We have presented a new definition in which DMD is defined to be the **eigendecomposition of an approximating linear operator**. Whereas existing DMD theory focuses on full-rank, sequential time series, our theory applies more generally to pairs of data vectors. At the same time, our DMD algorithm is only a slight modification of the commonly used, SVD-based DMD algorithm. It also preserves, and even strengthens, the links between DMD and Koopman operator theory. Thus our framework can be considered to be an extension of existing DMD theory to a more general class of datasets.

For instance, when analyzing data generated by a dynamical system, we require only that the columns of the data matrices  $X$  and  $Y$  be related by the dynamics of interest, in a pairwise fashion. Unlike existing DMD algorithms, we do not require that the data come from uniform sampling of a single time series, nor do we require that the columns of  $X$  and  $Y$  overlap. We demonstrated the utility of this approach using two numerical examples. In the first, we sampled a trajectory nonuniformly, significantly reducing computational costs. In the second, we concatenated multiple datasets in a single DMD computation, effectively averaging out the effects of noise. Our generalized interpretation of DMD also proved useful in explaining the results of DMD computations involving rank-deficient datasets. Such computations may provide either meaningful or misleading information, depending on the dataset, and are not treated in the existing DMD literature.

Finally, we showed that DMD is closely related to both the eigensystem realization algorithm (ERA) and linear inverse modeling (LIM). In fact, under certain conditions DMD and LIM are equivalent. We used the connection between DMD and the ERA to motivate a strategy for dealing with the inability of DMD to correctly characterize standing waves. An interesting future direction would be to explore whether or not lessons learned from past applications of LIM can similarly inform strategies for future applications of DMD.

**Acknowledgments.** The authors gratefully acknowledge support from the Air Force Office of Scientific Research (AFOSR), the National Science Foundation (NSF), and the Federal Aviation Administration (FAA). In addition, the authors thank Peter J. Schmid, Shervin Bagheri, Matthew O. Williams, Kevin K. Chen, and Jay Y. Qi for their valuable insights. Specifically, the authors acknowledge Jay Y. Qi for first observing the limitations of DMD with regard to standing wave behavior, during his study of the Kuramoto–Sivashinsky equation. Experimental data from the bluff-body wake discussed in Section 3.2.4 were collected by John Griffin, Adam Hart, Louis N. Cattafesta III, and Lawrence S. Ukeiley.

**Appendix A. Choices for mode scaling.** DMD practitioners often identify DMD modes of interest using their norms; it is generally assumed that modes with large norms are more dynamically important. Sometimes, as in Figure 7, the norms are weighted by the magnitude of the corresponding DMD eigenvalues to penalize spurious modes with large norms but quickly decaying contributions to the dynamics. However, Definition 1 allows for arbitrary scalings of the DMD modes. Then in order to use mode norms as a criterion for selecting interesting DMD modes, an appropriate scaling must be chosen. (We note that a nice approach to scaling was developed in [21], which considers an sparse representation where only a small number of DMD modes have nonzero magnitude. We do not pursue this approach here, but the ideas in [21] apply equally well to exact DMD modes and to projected DMD modes.)

Below, we discuss a number of possible options for scaling, any of which could be appropriate depending on the application. Our discussion is split into two parts: first, we consider general scalings motivated by linear algebra concepts. Second, we consider scalings that are appropriate for sequential time series. We pay particular attention to the cost of computing the scaling factors, as DMD is primarily applied to high-dimensional datasets for which data manipulations can be costly. For instance, due to computational costs, in practice only a minimal set of modes is computed. This prohibits any scaling that requires first computing all of the DMD modes.

**A.1. Generic data.** Perhaps the most obvious scaling for an eigenvector is one such that each eigenvector has unit norm. This can certainly be applied in computing DMD modes, and such scaling factors can be computed on a mode-by-mode basis. For instance, one may first normalize the eigenvectors of  $\tilde{A}$  ( $w$  in (5)). Then the projected DMD modes  $\hat{\varphi} = \mathbb{P}_X \varphi = Uw$  also have unit norm, since  $\hat{\varphi}^* \hat{\varphi} = w^* U^* U w = w^* w = 1$ . The scalings in (6) and (9) are then natural, recalling from Theorem 3 that  $\hat{\varphi} = \mathbb{P}_X \varphi$ .

Another approach is to consider the relationship between the (exact) DMD modes and the adjoint DMD modes. Recall that adjoint DMD modes are constructed via  $\psi \triangleq Uz$ , where  $z^* \tilde{A} = \lambda z^*$ . We observe that exact DMD modes are orthogonal to adjoint DMD modes with different eigenvalues, since

$$\begin{aligned} (\lambda_j \psi_j^*) \varphi_k &= \psi_j^* A \varphi_k = \psi_j^* (\lambda_k \varphi_k) \\ \implies (\lambda_j - \lambda_k) \psi_j^* \varphi_k &= 0. \end{aligned}$$

So as long as  $\lambda_j \neq \lambda_k$ , then  $\psi_j^* \varphi_k$  is zero. If we scale the adjoint eigenvectors  $z$  such that  $z^* w = 1$ , leaving  $\|w\| = 1$ , we observe that

$$\begin{aligned} \psi_k^* \varphi_k &= \frac{1}{\lambda} (U z_k)^* (Y V \Sigma^{-1} w_k) \\ &= \frac{1}{\lambda} z_k^* \tilde{A} w_k = 1. \end{aligned}$$

Thus, the exact DMD modes and adjoint DMD modes form a biorthogonal set; there is no additional scaling necessary after computing the DMD modes via (9). (We note that it is also possible to let  $\|z\| = 1$  and scale  $w$  to achieve the above result.)

**A.2. Sequential time series.** Based on (27), projected DMD modes are often scaled such that the sum of the modes equals the first data vector,  $x_0$  [22]. This is also natural for exact DMD modes. However, representing  $x_0$  using eigenvectors of

$A$  generally requires eigenvectors whose eigenvalues are zero; these modes are not identified by Algorithm 2. Fortunately, it turns out that we can compute a scaling based on  $x_0$  without computing those modes.

Let  $\Phi = YV\Sigma^{-1}W\Lambda^{-1}$  be the matrix whose columns are the DMD modes computed by Algorithm 2 (those with nonzero DMD eigenvalues), where  $\tilde{A}W = W\Lambda$ . In order to satisfy (27) (for  $k = 1$ ), we must have

$$\Phi\Lambda d = y_0, \quad (40)$$

where  $d = (d_1, \dots, d_j)$  with  $d_j = \theta_j(z)$  in (27). We wish to solve (40) for  $d$ . Note that, as long as the data matrices  $X$  and  $Y$  are linearly consistent (Definition 2), we have  $y_0 = Ax_0$ , so  $y_0$  is in the range of  $A$ , and therefore in the range of  $\Phi$ , and (40) has a solution. This solution may be found, for instance, by QR decomposition of  $\Phi$ . Then  $d_j$  contains the “amplitude” of each mode  $\varphi_j$ , and if desired, the modes may be renormalized by these amplitudes. (If no exact solution exists, we can instead solve for the least-squares solution.)

For large problems, it can be costly to compute the full set of DMD modes  $\Phi$ . In this case, one may solve for  $d$  using a pseudoinverse of  $\Phi$ , following [19]:

$$\begin{aligned} d &= \Lambda^{-1}(\Phi^*\Phi)^{-1}\Phi^*y_0 \\ &= (W^*\Sigma^{-1}V^*Y^*YV\Sigma^{-1}W)^{-1}W^*\Sigma^{-1}V^*Y^*y_0. \end{aligned}$$

In this approach, one does not need to explicitly calculate  $\Phi$ ; however, one does need the product  $Y^*Y$ , which is expensive to compute for large  $n$ . Consider that when  $n \gg m$ , the most expensive step in the DMD algorithm is computing  $X^*X$ , required to compute the SVD of  $X$  using the method of snapshots [33]. Then for generic  $Y$ , computing scaling factors from  $Y^*Y$  is nearly as costly as the rest of the DMD algorithm, and may not be viable. Furthermore, in this approach, the condition number of  $\Phi$  is squared, so for poorly conditioned matrices, the solution can be much less accurate than alternatives such as QR decomposition. If, however,  $X$  and  $Y$  describe a sequential time-series (see (2)), then  $Y^*Y$  and  $X^*X$  will contain many of the same elements. In this case,  $X^*X$  has already been computed and computing  $Y^*Y$  requires only  $m$  new inner products.

## REFERENCES

- [1] B. R. Noack, K. Afanasiev, M. Morzyński, G. Tadmor, and F. Thiele. A hierarchy of low-dimensional models for the transient and post-transient cylinder wake. *J. Fluid Mech.*, 497:335–363, December 2003.
- [2] P. J. Schmid and J. Sesterhenn. Dynamic mode decomposition of numerical and experimental data. In *61st Annual Meeting of the APS Division of Fluid Dynamics*. American Physical Society, November 2008.
- [3] C. W. Rowley, I. Mezic, S. Bagheri, P. Schlatter, and D. S. Henningson. Spectral analysis of nonlinear flows. *J. Fluid Mech.*, 641:115–127, December 2009.
- [4] Philip J. Holmes, John L. Lumley, Gal Berkooz, and Clarence W. Rowley. *Turbulence, Coherent Structures, Dynamical Systems and Symmetry*. Cambridge University Press, Cambridge, UK, 2nd edition, 2011.
- [5] C. W. Rowley. Model reduction for fluids, using balanced proper orthogonal decomposition. *Int. J. Bifurcat. Chaos*, 15(3):997–1013, March 2005.
- [6] J. R. Singler. Optimality of balanced proper orthogonal decomposition for data reconstruction. *Numerical Functional Analysis and Optimization*, 31(7):852–869, 2010.
- [7] I. Mezić. Spectral properties of dynamical systems, model reduction and decompositions. *Nonlin. Dynam.*, 41(1–3):309–325, August 2005.
- [8] P. J. Schmid. Dynamic mode decomposition of numerical and experimental data. *J. Fluid Mech.*, 656:5–28, August 2010.

- [9] T. W. Muld, G. Efraimsson, and D. S. Henningson. Flow structures around a high-speed train extracted using proper orthogonal decomposition and dynamic mode decomposition. *Comput. Fluids*, 57:87–97, March 2012.
- [10] D. Duke, D. Honnery, and J. Soria. Experimental investigation of nonlinear instabilities in annular liquid sheets. *J. Fluid Mech.*, 691:594–604, January 2012.
- [11] M. Grilli, P. J. Schmid, S. Hickel, and N. A. Adams. Analysis of unsteady behaviour in shockwave turbulent boundary layer interaction. *J. Fluid Mech.*, 700:16–28, June 2012.
- [12] L. Massa, R. Kumar, and P. Ravindran. Dynamic mode decomposition analysis of detonation waves. *Phys. Fluids*, 24(6), June 2012.
- [13] A. Seena and H. J. Sung. Dynamic mode decomposition of turbulent cavity flows for self-sustained oscillations. *Int. J. Heat Fluid Fl.*, 32(6):1098–1110, December 2011.
- [14] P. J. Schmid. Application of the dynamic mode decomposition to experimental data. *Exp. Fluids*, 50(4):1123–1130, April 2011.
- [15] P. J. Schmid, L. Li, M. P. Juniper, and O. Pust. Applications of the dynamic mode decomposition. *Theor. Comp. Fluid Dyn.*, 25(1–4):249–259, June 2011.
- [16] P. J. Schmid, D. Violato, and F. Scarano. Decomposition of time-resolved tomographic PIV. *Exp. Fluids*, 52(6):1567–1579, June 2012.
- [17] O. Semeraro, G. Bellani, and F. Lundell. Analysis of time-resolved PIV measurements of a confined turbulent jet using POD and Koopman modes. *Exp. Fluids*, 53(5):1203–1220, November 2012.
- [18] J. H. Tu and C. W. Rowley. An improved algorithm for balanced POD through an analytic treatment of impulse response tails. *J. Comput. Phys.*, 231(16):5317–5333, June 2012.
- [19] B. A. Belson, J. H. Tu, and C. W. Rowley. A parallelized model reduction library. *ACM T. Math. Software*, 2013 (accepted).
- [20] D. Duke, J. Soria, and D. Honnery. An error analysis of the dynamic mode decomposition. *Exp. Fluids*, 52(2):529–542, February 2012.
- [21] M. R. Jovanović, P. J. Schmid, and J. W. Nichols. Sparsity-promoting dynamic mode decomposition. arXiv:1309.4165v1 [physics.flu-dyn], September 2013.
- [22] K. K. Chen, J. H. Tu, and C. W. Rowley. Variants of dynamic mode decomposition: Boundary condition, Koopman, and Fourier analyses. *J. Nonlinear Sci.*, 22:887–915, 2012.
- [23] P. J. Goulart, A. Wynn, and D. Pearson. Optimal mode decomposition for high dimensional systems. In *Proceedings of the 51st IEEE Conference on Decision and Control*, December 2012.
- [24] A. Wynn, D. Pearson, B. Ganapathisubramani, and P. J. Goulart. Optimal mode decomposition for unsteady flows. *J. Fluid Mech.*, 733:473–503, October 2013.
- [25] I. Mezić. Analysis of fluid flows via spectral properties of the Koopman operator. *Annu. Rev. Fluid Mech.*, 45:357–378, 2013.
- [26] S. Bagheri. Koopman-mode decomposition of the cylinder wake. *J. Fluid Mech.*, 726:596–623, July 2013.
- [27] B. L. Ho and R. E. Kalman. Effective construction of linear state-variable models from input/output data. *Proceedings of the Third Annual Allerton Conference on Circuit and System Theory*, pages 449–459, 1965.
- [28] J. N. Juang and R. S. Pappa. An eigensystem realization-algorithm for modal parameter-identification and model-reduction. *J. Guid. Control Dynam.*, 8(5):620–627, 1985.
- [29] Z. Ma, S. Ahuja, and C. W. Rowley. Reduced-order models for control of fluids using the eigensystem realization algorithm. *Theor. Comp. Fluid Dyn.*, 25(1–4):233–247, June 2011.
- [30] C. Penland. Random forcing and forecasting using Principal Oscillation Pattern analysis. *Mon. Weather Rev.*, 117(10):2165–2185, October 1989.
- [31] C. Penland and T. Magorian. Prediction of Niño 3 sea-surface temperatures using linear inverse modeling. *J. Climate*, 6(6):1067–1076, June 1993.
- [32] L. N. Trefethen and D. Bau III. *Numerical Linear Algebra*. SIAM, Philadelphia, 1997.
- [33] L. Sirovich. Turbulence and the dynamics of coherent structures. 2. Symmetries and transformations. *Q. Appl. Math.*, 45(3):573–582, October 1987.
- [34] K. Taira and T. Colonius. The immersed boundary method: a projection approach. *J. Comput. Phys.*, 225(2):2118–2137, August 2007.
- [35] T. Colonius and K. Taira. A fast immersed boundary method using a nullspace approach and multi-domain far-field boundary conditions. *Comput. Method Appl. M.*, 197(25–28):2131–2146, 2008.



- [36] B. R. Noack, G. Tadmor, and M. Morzyński. Actuation models and dissipative control in empirical Galerkin models of fluid flows. In *Proceedings of the American Control Conference*, pages 5722–5727, 2004.
- [37] J. H. Tu, J. Griffin, A. Hart, C. W. Rowley, L. N. Cattafesta III, and L. S. Ukeiley. Integration of non-time-resolved PIV and time-resolved velocity point sensors for dynamic estimation of velocity fields. *Exp. Fluids*, 54(2):1429, February 2013.
- [38] J. H. Tu, C. W. Rowley, E. Aram, and R. Mittal. Koopman spectral analysis of separated flow over a finite-thickness flat plate with elliptical leading edge. AIAA Paper 2011-38, 49th AIAA Aerospace Sciences Meeting and Exhibit, January 2011.
- [39] E. N. Lorenz. Empirical orthogonal functions and statistical weather prediction. Technical report, Massachusetts Institute of Technology, December 1956.
- [40] K. Pearson. LIII. on lines and planes of closest fit to systems of points in space. *Philos. Mag.*, 2(7–12):559–572, July 1901.
- [41] H. Hotelling. Analysis of a complex of statistical variables into principal components. *J. Educ. Psychol.*, 24:417–441, September 1933.
- [42] H. Hotelling. Analysis of a complex of statistical variables into principal components. *J. Educ. Psychol.*, 24:498–520, October 1933.
- [43] K. Hasselmann. PIPs and POPs: the reduction of complex dynamical-systems using Principal Interaction and Oscillation Patterns. *J. Geophys. Res.-Atmos.*, 93(D9):11015–11021, September 1988.
- [44] H. von Storch, G. Bürger, R. Schnur, and J. S. von Storch. Principal oscillation patterns: a review. *J. Climate*, 8(3):377–400, March 1995.
- [45] M. B. Blumenthal. Predictability of a coupled ocean-atmosphere model. *J. Climate*, 4(8):766–784, August 1991.

*E-mail address:* jhtu@berkeley.edu

*E-mail address:* cwwrowley@princeton.edu

*E-mail address:* dluchten@princeton.edu

*E-mail address:* sbrunton@uw.edu

*E-mail address:* kutz@uw.edu