Research

Author for correspondence:
Steven L. Brunton
e-mail: sbrunton@uw.edu

THE ROYAL SOCIETY
PUBLISHING

Physics-informed dynamic mode decomposition

Peter J. Baddoo[1], Benjamin Herrmann[2], Beverley J. McKeon[3], J. Nathan Kutz[4] and Steven L. Brunton[5]

[1]Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[2]Department of Mechanical Engineering, University of Chile, Beauchef 851, Santiago, Chile
[3]Graduate Aerospace Laboratories, California Institute of Technology, Pasadena, CA 91125, USA
[4]Department of Applied Mathematics, and [5]Department of Mechanical Engineering, University of Washington, Seattle, WA 98195, USA

PJB, 0000-0002-8671-6952; BJM, 0000-0003-4220-1583; JNK, 0000-0002-6004-2275; SLB, 0000-0002-6565-5118

In this work, we demonstrate how physical principles—such as symmetries, invariances and conservation laws—can be integrated into the *dynamic mode decomposition* (DMD). DMD is a widely used data analysis technique that extracts low-rank modal structures and dynamics from high-dimensional measurements. However, DMD can produce models that are sensitive to noise, fail to generalize outside the training data and violate basic physical laws. Our physics-informed DMD (piDMD) optimization, which may be formulated as a Procrustes problem, restricts the family of admissible models to a matrix manifold that respects the physical structure of the system. We focus on five fundamental physical principles—conservation, self-adjointness, localization, causality and shift-equivariance—and derive several closed-form solutions and efficient algorithms for the corresponding piDMD optimizations. With fewer degrees of freedom, piDMD models are less prone to overfitting, require less training data, and are often less computationally expensive to build than standard

DMD models. We demonstrate piDMD on a range of problems, including energy-preserving fluid flow, the Schrödinger equation, solute advection-diffusion and three-dimensional transitional channel flow. In each case, piDMD outperforms standard DMD algorithms in metrics such as spectral identification, state prediction and estimation of optimal forcings and responses.

1. Introduction

Integrating partial knowledge of physical principles into data-driven techniques is a primary goal of the scientific machine learning (ML) community [1]. Physical principles—such as conservation laws, symmetries and invariances—can be incorporated into ML algorithms in the form of inductive biases, thereby ensuring that the learned models are constrained to the correct physics. Recent successful examples of ML algorithms that have been modified to respect physical principles include neural networks [2–10], kernel methods [11,12], deep generative models [13], data assimilation [14,15] and sparse regression [16–20]. These examples demonstrate that incorporating partially known physical principles into ML architectures can increase the accuracy, robustness and generalizability of the resulting models, while simultaneously decreasing the required training data. In this work, we integrate knowledge of physical principles into one of the most widely used methods in data-driven dynamical systems research: the dynamic mode decomposition (DMD) [21–25].

DMD is a data diagnostic technique that extracts coherent spatial–temporal patterns from high-dimensional time-series data [21,25]. Although DMD originated in the fluid dynamics community [21], the algorithm has since been applied to a wealth of dynamical systems including in epidemiology [26], robotics [27,28], neuroscience [29], quantum control [30], power grids [31] and plasma physics [32,33]. Despite its widespread successes, DMD is highly sensitive to noise [34–36], fails to capture travelling wave physics, and can produce overfit models that do not generalize. We demonstrate that integrating physics into the learning framework can help address these challenges.

Suppose that we are studying a dynamical system defined by $x_{j+1} = F(x_j)$ for an unknown function F. Given a collection of measurements $\{x_1, \ldots, x_m\}$, DMD identifies the best low-rank linear approximation of F. In other words, DMD seeks a low-rank matrix A such that $x_{j+1} \approx A x_j$ for $j = 1, \ldots, m$. Arranging the data measurements into matrices X and Y (see (2.1) for details) allows us to phrase the above formally as

$$\arg\min_{\text{rank}\,(A) \leq r} ||Y - AX||_F. \tag{1.1}$$

After approximately[1] solving (1.1), the DMD process computes the dominant spectral properties of the learned linear operator [21–25].

The rank-r constraint in (1.1) is motivated by the assumed modal structure of the system but does not account for other important physical properties. For example, one limitation of DMD is that the solution of (1.1) lies within the span of Y, so the learned model can fail to generalize outside the training regime. We address this limitation and others by embedding partial physics knowledge into the learning process.

In this work, we incorporate physical principles into the optimization in (1.1) by constraining the solution matrix A to lie on a matrix manifold:

$$\arg\min_{A \in \mathcal{M}} ||Y - AX||_F. \tag{1.2}$$

---

[1]Although the problem (1.1) admits an exact solution [37,38], the vast majority of DMD implementations solve (1.1) *approximately* by projecting the problem into the *r*-dimensional subspace spanned by the first *r* principal components of *X*. For that reason, we focus on comparing our new method to the commonly used exact DMD [39] and optimized DMD algorithms [40] as opposed to the method of [38].

The matrix manifold $\mathcal{M}$ is dictated by the known physics of the system at hand. For example, we may select $\mathcal{M}$ such that its members satisfy certain symmetries that are known to be obeyed by the system. We call (1.2) *physics-informed* DMD (piDMD) as the optimization integrates underlying knowledge of the system physics into the learning framework.[2] Again, the spectral features of the solution of (1.2) can be computed to give insight into the modes that dominate the system and their dynamics. Similarly to DMD, these modes can usually be computed without forming A explicitly. The low-rank DMD regression (1.1) is a special case of piDMD where $\mathcal{M}$ is the manifold of matrices with rank bounded by r. However, piDMD models are not always low rank (see, for example, §4d) but nevertheless have few degrees of freedom due to the physics-informed constraints. Constraining the solution matrix improves the generalizability of the learned model, reduces sensitivity to noise and reduces the demand for large training sets, as evidenced by a series of applications in §4. Similarly to standard DMD, our goal is to decompose the data into its most important modes; as such, our results focus on comparing the spectrum of the learned operator A with that of the true underlying operator.

The optimization problem (1.2) is well known in the statistical literature as a *Procrustes problem* [41–53], which comprises of finding the optimal transformation between two matrices subject to certain constraints on the class of admissible transformations. According to Greek mythology, Procrustes was a bandit who would stretch or amputate the limbs of his victims to force them to fit onto his bed. Herein, X plays the role of Procrustes' victim, Y is the bed and A is the 'treatment' (stretching or amputation).[3] Procrustes problems (1.2) seek to learn the treatment A that best relates the data measurements X and Y. Our recasting of the DMD regression as a Procrustes problem is a new connection, and is the basis of piDMD. This perspective enables us to leverage the substantial extant work on Procrustes problems into new application areas. The literature contains many exact solutions for Procrustes problems, including notable cases of orthogonal matrices [49], and symmetric matrices [45].

In a similar vein to the present work, [54] deployed techniques from group theory and representation theory to derive a data-driven method to study the Koopman operator for equivariant dynamical systems. By contrast, piDMD does not work with the Koopman operator directly but permits a broader class of physical principles, including those that cannot be expressed as equivariances.

A significant challenge in physics-informed ML is to develop implementations that incorporate known physics but also scale to higher dimensions [1]. One contribution of this work is to develop several exact solutions, in terms of standard linear algebra objects, of physically relevant Procrustes problems. Where possible, we use matrix factorizations and low-rank representations to alleviate the computational burden of implementation; it is rarely necessary to form A explicitly when computing its spectral properties or other diagnostics. For complicated matrix manifolds, an exact solution of the Procrustes problem (1.2) may be intractable and we must resort to algorithmic approaches. Fortunately, optimization on matrix manifolds is a mature field [55], and many algorithms are implemented in the open source software package 'Manopt' [56].

The remainder of the paper is arranged as follows. In §2, we provide background information on DMD and Procrustes problems. Then, in §3, we describe the broad framework of piDMD. We consider a range of applications in §4, with a focus on shift-equivariant, conservative, self-adjoint, local and causal systems. Section 5 concludes with a discussion of the limitations of piDMD and suggests several future research directions. Further information is available in electronic supplementary material.

2. Mathematical background

In this section, we provide further details on DMD and Procrustes problems. Throughout the article, we assume that we have access to m snapshots pairs of n features each: $\{(x_j, y_j),\ j =$

---

[2] An open source implementation of piDMD is available in MATLAB at www.github.com/baddoo/piDMD.

[3] This terminology was first introduced by Hurley & Cattell [41].

$1, \ldots, m\}$. For example, x_j may be a discretized fluid flow field at time t_j and y_j may be the flow field at the next time step t_{j+1}. It is convenient to arrange the data into $n \times m$ snapshot matrices of the form

$$X = \begin{bmatrix} | & & | & & | \\ x_1 & \cdots & x_m \\ | & & | & & | \end{bmatrix} = \begin{bmatrix} — & \tilde{x}_1 & — \\ — & \vdots & — \\ — & \tilde{x}_n & — \end{bmatrix} \quad \text{and} \quad Y = \begin{bmatrix} | & & | & & | \\ y_1 & \cdots & y_m \\ | & & | & & | \end{bmatrix} = \begin{bmatrix} — & \tilde{y}_1 & — \\ — & \vdots & — \\ — & \tilde{y}_n & — \end{bmatrix} \quad (2.1)$$

so that the ith rows of X and Y are the measurements of the ith features and the jth columns are the jth temporal snapshots. Henceforth, we use $\tilde{\cdot}$ to represent a row vector.

(a) Dynamic mode decomposition

DMD was initially proposed as a dimensionality reduction technique that extracts dominant spatio-temporal coherent structures from high-dimensional time-series data [21]. In particular, DMD identifies the leading-order spatial eigenmodes of the matrix A in (1.1), along with a linear model for how the amplitudes of these coherent structures evolve in time. DMD has been applied to a range of systems, as summarized in the monograph [24] and the recent review by Schmid [25].

To address the challenges associated with DMD, researchers have derived many variations on the original algorithm, including sparsity promoting DMD [57], DMD with control [58], noise-robust variants [34–36,40], recursive DMD [59], online DMD [60,61], low-rank DMD [38] and versions for under-resolved data in space or time [62–64]. At present, the most widely used variant is 'exact DMD' [23], which phrases the DMD solution in terms of the Moore–Penrose pseudoinverse. Owing to its simplicity and widespread use, most of the comparisons in this paper are made between exact DMD and piDMD.

It is assumed that each x_j and y_j are connected by an unknown dynamical system of the form $y_j = F(x_j)$; for discrete-time dynamics $y_j = x_{j+1}$ and for continuous-time dynamics $y_j = \dot{x}_j$. DMD aims to learn the dominant behaviour of F by finding the best linear approximation for F given the data and then performing diagnostics on that approximation. Thus, DMD seeks the linear operator A that best maps the snapshots in the set $\{x_j\}$ to those in the set $\{y_j\}$:

$$y_j \approx A x_j \quad \text{for } j = 1, \ldots, m. \quad (2.2)$$

Expressed in terms of the snapshot matrices in (2.1), the linear system in (2.2) becomes

$$Y \approx AX, \quad (2.3)$$

and the optimization problem for A is given by (1.1). The minimum-norm solution for A is given by

$$A = YX^{\dagger} = YV\Sigma^{\dagger}U^*, \quad (2.4)$$

where $\dagger$ indicates the Moore–Penrose pseudoinverse [65] and $X = U\Sigma V^*$ is the singular value decomposition.

In many applications, the state dimension n is very large, and forming or storing A explicitly becomes impractical. Instead, we use a rank-r approximation for A, denoted by $\hat{A}$, where $r \ll n$. To form $\hat{A}$, we construct the optimal rank-r approximation for X using the truncated singular value decomposition [66]: $X \approx U_r \Sigma_r V_r^*$. We then project A onto the leading r principal components of X as

$$\hat{A} = U_r^* A U_r = U_r^* Y V_r \Sigma_r^{-1}. \quad (2.5)$$

It is now computationally viable to compute the eigendecomposition of $\hat{A}$ as

$$\hat{A}\hat{\Psi} = \hat{\Psi}\Lambda. \quad (2.6)$$

The eigenvectors of A can be approximated from the reduced eigenvectors by [23]

$$\Psi = YV\Sigma^{-1}\hat{\Psi}. \quad (2.7)$$

This eigendecomposition is connected to the Koopman operator of the system, and allows reconstructions and predictions [22,24,67,68]. For example, for a discrete-time system (i.e. $y_k = x_{k+1}$) with evenly spaced samples in time, if the eigenvectors form a linearly independent set then

$$x_j = \boldsymbol{\Psi}\boldsymbol{\Lambda}^{j-1}\boldsymbol{b}, \tag{2.8}$$

where the vector $\boldsymbol{b}$ contains the weights of the modes in the initial condition: $\boldsymbol{b} = \boldsymbol{\Psi}^{\dagger}\boldsymbol{x}_1$. From the above, it is clear that the eigenvalues $\boldsymbol{\Lambda}$ govern the temporal behaviour of the system and the eigenvectors $\boldsymbol{\Psi}$ are the spatial modes.

(b) Procrustes problems

Procrustes problems (1.2) comprise finding the optimal transformation between two matrices subject to certain constraints on the class of admissible transformations [44,45,49]. The minimization is usually phrased in the Frobenius norm. Procrustes analysis finds relevance in a wealth of fields including sensory analysis [69], data visualization [70], neural networks [71], climate science [72] and solid mechanics [45]. A summary is available in the monograph [44].

The earliest and most common formulation is the 'orthogonal Procrustes problem' [44,49]. Suppose that we have two sets of measurements X and Y that we know are related by an unitary (orthogonal) transformation (i.e. a rotation or reflection). The goal is to learn the best unitary transformation that relates the data measurements. Thus, A is constrained to be a unitary matrix and the minimization problem is

$$\arg\min_{A^*A=I}||Y - AX||_F. \tag{2.9}$$

The solution to (2.9) was derived by Schönemann [49] as

$$A = \boldsymbol{U}_{YX}\boldsymbol{V}_{YX}^*, \tag{2.10}$$

where $\boldsymbol{U}_{YX}\boldsymbol{\Sigma}_{YX}\boldsymbol{V}_{YX}^* = YX^*$ is a full singular value decomposition. Alternatively, $A = \boldsymbol{U}_P$, where $\boldsymbol{U}_P\boldsymbol{H}_P = YX^*$ is a polar decomposition. The solution is unique if and only if YX^* is full-rank. As we show in §4b, a unitary matrix corresponds to an energy-preserving operator.

There are many solutions for Procrustes problems available in the literature for different matrix constraints [45]. When exact solutions are not possible, algorithmic solutions can be effective [56].

3. Physics-informed dynamic mode decomposition

Incorporating physics into ML algorithms involves supplementing an existing technique with additional biases. Usually, ML practitioners take one of three approaches [1]. First, observational biases can be embedded through data augmentation techniques; however, augmented data are not always available, and incorporating additional training samples can become computationally expensive. Second, physical constraints can be included by suitably penalizing the loss function [8]. Third, inductive biases can be incorporated directly into the ML architecture in the form of mathematical constraints [18]. As noted by Karniadakis *et al.* [1], this third approach is arguably the most principled method since it produces models that strictly satisfy the physical constraints. piDMD falls into the final category. piDMD incorporates physical principles by constraining the matrix manifold of the DMD regression problem (1.2).[4] We exchange the low-rank representation of A typically sought in model order reduction in favour of alternative or additional matrix structures more relevant to the physical problem at hand. In figure 1, we illustrate the matrix structures used in this paper, along with their corresponding physical principles and references for the optimal solutions.

Analysing a system with the piDMD framework requires four steps: modelling, interpretation, optimization and diagnostics:

---

[4]Although the focus of this paper is on the underlying regression problem, we are still primarily concerned with decomposing the data into the form (2.8); hence, piDMD is a modal decomposition algorithm.
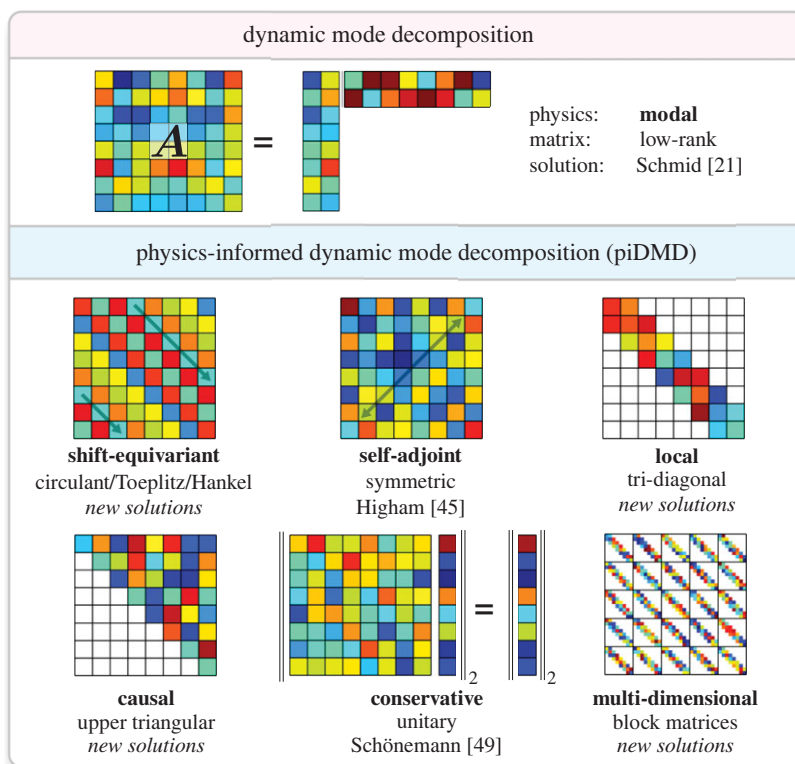
Figure 1. Visual illustrations of the types of matrices, their corresponding physical principles and references to the solutions of the corresponding optimization problem (1.2).

(i) Modelling: When studying a system from a data-driven perspective, it is common to have some partial knowledge of the underlying physics of the system. This first step in piDMD asks the user to summarize the known or suspected physical properties of the system at hand.

(ii) Interpretation: Having determined the physical principle we would like to enforce, we must translate these laws into the matrix manifold to which the linear model should be constrained. This step involves knowledge of the data collection procedure such as the spatial grid or the energy inner product.

(iii) Optimization: Equipped with a target matrix manifold, we may now solve the relevant Procrustes problem (1.2). By suitably modifying the optimization routine, we can guarantee that the resulting model satisfies the physical principle identified in step (i).

(iv) Diagnostics: The final step of piDMD involves extracting physical information from the learned model A. For example, one may wish to analyse the spectrum or modes, compute the resolvent modes [73], perform predictions, or investigate other diagnostics. Similarly to the original DMD formulation, the learned modes can be used to decompose the data into the form (2.8).

These steps are expanded on in electronic supplementary material, §A. As an example, consider the travelling wave solution explored in the figure 2. We begin with the physical principle that the system is shift equivariant (modelling), and this lead us to seek a circulant matrix model (interpretation). We find a solution to the corresponding Procrustes problem (optimization) and investigated the predictive behaviour of the model (diagnostics).
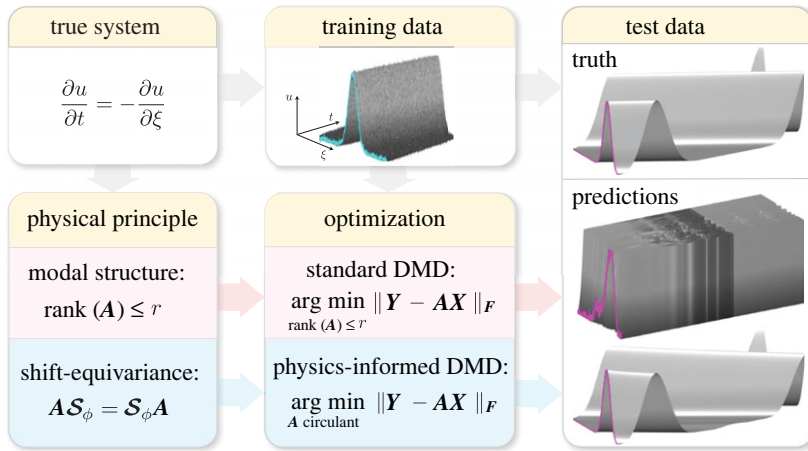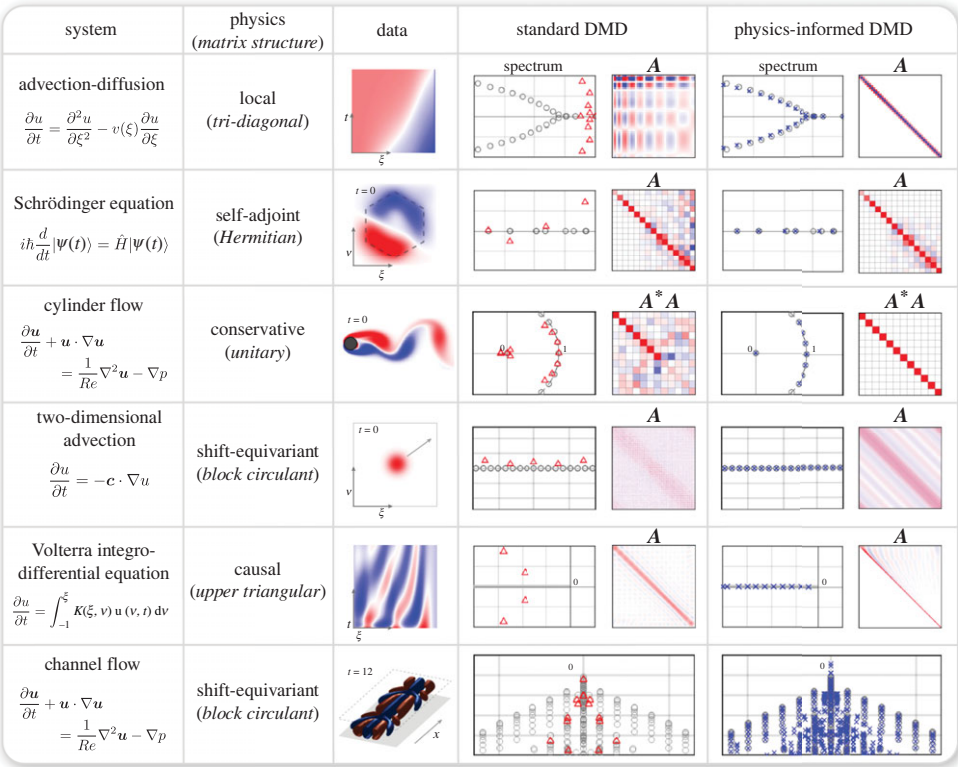
Figure 2. Comparing standard dynamic mode decomposition (specifically, optimized DMD [40]) and physics-informed dynamic mode decomposition applied to the advection equation with incomplete data. The data are contaminated with 2% additive Gaussian noise. Having trained DMD models, we then perform predictions for different initial conditions. Standard DMD fails whereas piDMD produces a faithful prediction. In the figure, $\mathcal{S}_\phi$ represents the shift operator. The performance of DMD improves if the data covers a full cycle of the pulse.

While each of the four steps can be challenging, the optimization step is often the most difficult. Moreover, any solutions are peculiar to the matrix manifold under consideration: the symmetric Procrustes problem is largely unconnected to, say, the tridiagonal Procrustes problem. Thus, in this article, we present many new solutions for (1.2) with different physics-informed matrix manifold constraints. When exact solutions are not possible (e.g. if the manifold constraint is quite complicated), there are sophisticated algorithmic solutions available [56].

Standard DMD exploits the low-rank structure of A to efficiently perform diagnostics on the learned model. Some of the manifolds we consider (such as tridiagonal and upper triangular) do not have an obvious or useful low-rank approximation. Instead, these matrices often have an alternative structure that can be exploited to perform fast diagnostic tests. For example, tridiagonal matrices rarely have a meaningful low-rank approximation, but nevertheless admit a fast eigendecomposition and singular value decomposition [65].

Some of the matrix manifolds we consider (such as symmetric, triangular, tridiagonal or circulant) can be phrased as linear equality constraints and can, in principle, be solved with linear-equality constrained least squares [65]. However, the number of equality constraints needed is $\mathcal{O}(n^2)$ so the resulting least-squares matrix will have $\mathcal{O}(n^2)$ rows, which is intractably large in most applications. Thus, we avoid phrasing the piDMD constraints in terms of linear equality constraints and instead exploit properties of the matrix manifold to find solutions that can be efficiently implemented (see §4a and §4c, for example).

DMD uses the available degrees of freedom in the model to optimize the reconstruction of the data in a low-dimensional subspace. This can result in overfitting since the algorithm does not respect global properties of the system; as such, the modes usually fail to characterize unseen test data. In piDMD, we exchange some of the available degrees of freedom in the model to enforce a global physical structure. This regularization produces models that are suboptimal in the reconstruction norm but are more generalizable to unseen datasets. The generalizability of piDMD is observed in each of our test cases as the algorithm is able to more accurately learn the true eigenvalues of the underlying system. We emphasize that we are not claiming that the low-rank constraint is always suboptimal. In fact, the low-rank constraint is often best and should be used as a first attempt when analysing data. However, if prior physical knowledge of the system is available then a physics-informed manifold may be more appropriate.

system	physics (*matrix structure*)	data	standard DMD	physics-informed DMD
advection-diffusion $\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial \xi^2} - v(\xi)\frac{\partial u}{\partial \xi}$	local (*tri-diagonal*)		spectrum, A	spectrum, A
Schrödinger equation $i\hbar\frac{d}{dt}|\psi(t)\rangle = \hat{H}|\psi(t)\rangle$	self-adjoint (*Hermitian*)		A	A
cylinder flow $\frac{\partial \boldsymbol{u}}{\partial t} + \boldsymbol{u}\cdot\nabla\boldsymbol{u} = \frac{1}{Re}\nabla^2\boldsymbol{u} - \nabla p$	conservative (*unitary*)		A^*A	A^*A
two-dimensional advection $\frac{\partial u}{\partial t} = -\boldsymbol{c}\cdot\nabla u$	shift-equivariant (*block circulant*)		A	A
Volterra integro-differential equation $\frac{\partial u}{\partial t} = \int_{-1}^{\xi} K(\xi, v)\, u(v, t)\, dv$	causal (*upper triangular*)		A	A
channel flow $\frac{\partial \boldsymbol{u}}{\partial t} + \boldsymbol{u}\cdot\nabla\boldsymbol{u} = \frac{1}{Re}\nabla^2\boldsymbol{u} - \nabla p$	shift-equivariant (*block circulant*)			

Figure 3. A comparison of the models learned by exact DMD and piDMD for a range of applications. piDMD identifies the spectrum of the true operator with higher accuracy than exact DMD. The structure of the model matrices is also illustrated; piDMD models are generally more coherent than those learned by exact DMD. Details of problem sizes are given in the corresponding sections in the electronic supplementary material. In the spectrum subplots, we plot the true eigenvalues as circles, the DMD eigenvalues as triangles and the piDMD eigenvalues as crosses.

4. Applying physics-informed dynamic mode decomposition to enforce canonical physical structures

We now present a range of applications of piDMD. Each of the following sections investigates a detailed application of piDMD for a specific physical principle. A summary is illustrated in figure 3, and the degrees of freedom in each model, and associated computational cost, is discussed in electronic supplementary material, F. The physical principles, corresponding matrix structures, and optimal solutions and extensions are summarized below

- §4a: Shift equivariant: circulant (and symmetric, skew-symmetric or unitary, electronic supplementary material, B.1), low rank (electronic supplementary material, B.2), non-equally spaced samples (electronic supplementary material, B.3), total least squares (electronic supplementary material, B.4), Toeplitz & Hankel (electronic supplementary material, B.5)
- §4b: Conservative: unitary (§2b)
- §4c: Self-adjoint: symmetric (§4c), skew-symmetric (electronic supplementary material, C.1), symmetric in a subspace (§4c(i))
- §4d: Local: tridiagonal (§4d), variable diagonals (electronic supplementary material, D.1), periodic (electronic supplementary material, D.2), symmetric tridiagonal (electronic

supplementary material, D.3), total least squares (electronic supplementary material, D.4), regularized locality (electronic supplementary material, D.5).

§4e: Causal: upper triangular (electronic supplementary material, E)

(a) Shift-equivariant systems

Shift equivariance (also called 'translation equivariance') is ubiquitous in the physical sciences. Spatially homogeneous systems—such as constant coefficient PDEs or convolution operators—are shift equivariant, and thus appear identical with respect to any (stationary) observer. In view of Noether's theorem [74], the conserved quantity related to shift equivariance is (generalized) linear momentum; as such, incorporating shift equivariance into a DMD model means that the model preserves linear momentum. While it may be unusual for practical engineering problems to have a truly homogeneous direction, many problems of basic physical interest—such as wall-bounded turbulent flow—frequently possess a spatially homogeneous dimension. In the sequel, we derive a piDMD formulation for shift-equivariant operators.

We define $\mathcal{S}_\phi$ as the ϕ-shift operator i.e. $\mathcal{S}_\phi v(\xi) = v(\xi + \phi)$ for all test functions v. We say that a space-continuous linear operator $\mathcal{A}$ is shift equivariant if $\mathcal{A}$ commutes with the ϕ-shift operator for all shifts:

$$\mathcal{S}_\phi \mathcal{A} = \mathcal{A} \mathcal{S}_\phi. \tag{4.1}$$

An application of the shift operator shows that if $\mathcal{A}$ is shift equivariant then the quantity $e^{-\lambda \xi} \mathcal{A} e^{\lambda \xi}$ is constant for all ξ. Thus, neglecting boundary conditions for now, $\{e^{\lambda \xi}\}$ are eigenfunctions of $\mathcal{A}$. If the domain is normalized to $[-1, 1]$ and has periodic boundaries, then $\lambda = l\pi i$ for integer l. In this case, the corresponding eigenfunctions form an orthogonal basis and the operator is diagonalized by the known eigenfunctions.

In the following, we assume that we are studying a shift-equivariant system on a periodic domain $\xi \in [-1, 1]$. We move from the continuous formulation to a discretized space and assume that we are studying a function $u(\xi, t)$ and have access to evenly spaced samples of u at $\boldsymbol{\xi} = [-1, -1 + \Delta\xi, \ldots, 1 - \Delta\xi]$. We discuss the case of non-equally spaced samples in electronic supplementary material, B.3. If we define the state variable as $\boldsymbol{x}(t) = [u(\xi_1, t), u(\xi_2, t), \ldots, u(\xi_n, t)]$ then the discrete-space linear operator A that generates $\boldsymbol{x}(t)$ is diagonalized by its eigenvectors

$$A = \mathcal{F} \operatorname{diag}(\hat{a}) \mathcal{F}^{-1}, \tag{4.2}$$

where $\mathcal{F}_{j,k} = e^{2\pi i (j-1)(k-1)/n} / \sqrt{n}$ and $\mathcal{F}^{-1} = \mathcal{F}^*$ and $\{\hat{a}_j\}$ are the unknown eigenvalues. Equation (4.2) is equivalent to stating that A is circulant

$$A = \begin{bmatrix} a_0 & a_{n-1} & \ldots & a_1 \\ a_1 & a_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & a_{n-1} \\ a_{n-1} & \ldots & a_1 & a_0 \end{bmatrix}, \tag{4.3}$$

i.e. $A_{j,k} = a_{(j-k) \bmod n}$. If the boundary conditions are instead Dirichlet or Neumann then A is not circulant but Toeplitz; we solve the corresponding Procrustes problem in electronic supplementary material, B.5.

Substituting (4.2) into (1.2) and noting that the Frobenius norm is invariant to unitary transformations allows (1.2) to be transformed to

$$\arg \min_{\hat{a}} ||\operatorname{diag}(\hat{a})\boldsymbol{\mathcal{X}} - \boldsymbol{\mathcal{Y}}||_F, \tag{4.4}$$

where $\boldsymbol{\mathcal{X}} = \mathcal{F}^* X$ and $\boldsymbol{\mathcal{Y}} = \mathcal{F}^* Y$ are the spatial discrete Fourier transforms of X and Y. As such, $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{Y}}$ can be efficiently formed in $O(mn \log(n))$ operations using fast Fourier transform (FFT) methods [75].

The rows of the cost function (4.4) now decouple to produce n minimization problems:

$$\arg\min_{\hat{a}_j} ||\hat{a}_j \tilde{\boldsymbol{\mathcal{X}}}_j - \tilde{\boldsymbol{\mathcal{Y}}}_j||_F \quad \text{for } 1 \leq j \leq n, \tag{4.5}$$

where $\tilde{\boldsymbol{\mathcal{X}}}_j$ and $\tilde{\boldsymbol{\mathcal{Y}}}_j$ are the jth rows of $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{Y}}$, respectively. An alternative view of this step is that spatial wavenumbers decouple in periodic shift-equivariant systems, so we may analyse each wavenumber individually. The coefficient $\hat{a}_j$ represents the temporal growth of the jth Fourier component. The optimal solution for each eigenvalue follows as:

$$\hat{a}_j = \tilde{\boldsymbol{\mathcal{Y}}}_j \tilde{\boldsymbol{\mathcal{X}}}_j^{\dagger}. \tag{4.6}$$

In principle, standard DMD methods [21,23] could be encouraged to respect shift equivariance by augmenting the data matrices with their shifted counterparts. However, one would strictly need to include n shifts, thus producing nm samples which is usually too large in applications. By contrast, the above solution is extremely efficient via the use of the FFT and the decoupling of wavenumbers.

The electronic supplementary material includes alternative solutions for cases where the system is shift equivariant and symmetric, skew-symmetric or unitary (electronic supplementary material, B.1), low rank (electronic supplementary material, B.2), the samples are not equally spaced (electronic supplementary material, B.3), the total least-squares case (electronic supplementary material, B.4), and when the system is Toeplitz or Hankel (electronic supplementary material, B.5).

(i) Example: plane channel flow

We now use this shift-equivariant piDMD to compute the modes of the linearized Navier–Stokes equations from a single simulation. We consider the incompressible flow inside a plane channel of size $2\pi \times 2 \times 2\pi$ along the spatial x, y and z coordinates that indicate the streamwise, wall-normal and spanwise directions, respectively. The configuration considers a Reynolds number of $Re = 2000$ based on the channel half-height and the centreline velocity, and periodic boundary conditions in the open faces of the channel; hence the flow is homogeneous in the x- and z-directions. We use the spectral code *Channelflow* [76] to perform direct numerical simulations (DNS) with the same numerical configurations as that presented in [73].

The dataset investigated is generated from a DNS of the response of the laminar flow with parabolic velocity profile to a localized perturbation in the wall-normal velocity component of the form

$$v(x, y, z, 0) = \left(1 - \frac{r^2}{c_r^2}\right)(\cos(\pi y) + 1)e^{(-r^2/c_r^2 - y^2/c_y^2)}, \tag{4.7}$$

where $r^2 = (x - \pi)^2 + (z - \pi)^2$, the parameters are set to $c_r = 0.7$ and $c_y = 0.6$, and the amplitude of the perturbation was scaled to have an energy-norm of 10^{-5} to ensure that the effect of nonlinearity is negligible. This initial condition, which was first studied in [77] and also in [73], is a model of a disturbance that could be generated in experiments using a spanwise and streamwise periodic array of axisymmetric jets injecting fluid perpendicular to the wall. The dataset obtained from this simulation is then comprised of a sequence of 300 snapshots of the three-dimensional velocity-perturbation field recorded every 0.5 time units. The data is corrupted with 2% Gaussian noise.

Our aim is to learn the spectrum of the underlying linear operator that *best* describes the evolution of the DNS snapshots, and to thus obtain a reliable predictive model for the dynamics. When directly applying exact DMD to this dataset, the algorithm attempts to find global modes for the three-dimensional velocity field and their spectrum. This is challenging because modes associated with different streamwise and spanwise wavenumbers may be mixed, leading to spurious eigenvalues, as shown in figure 4. However, we know that the flow is homogeneous in the x and z coordinates, and is therefore shift-equivariant in these directions. Hence, we
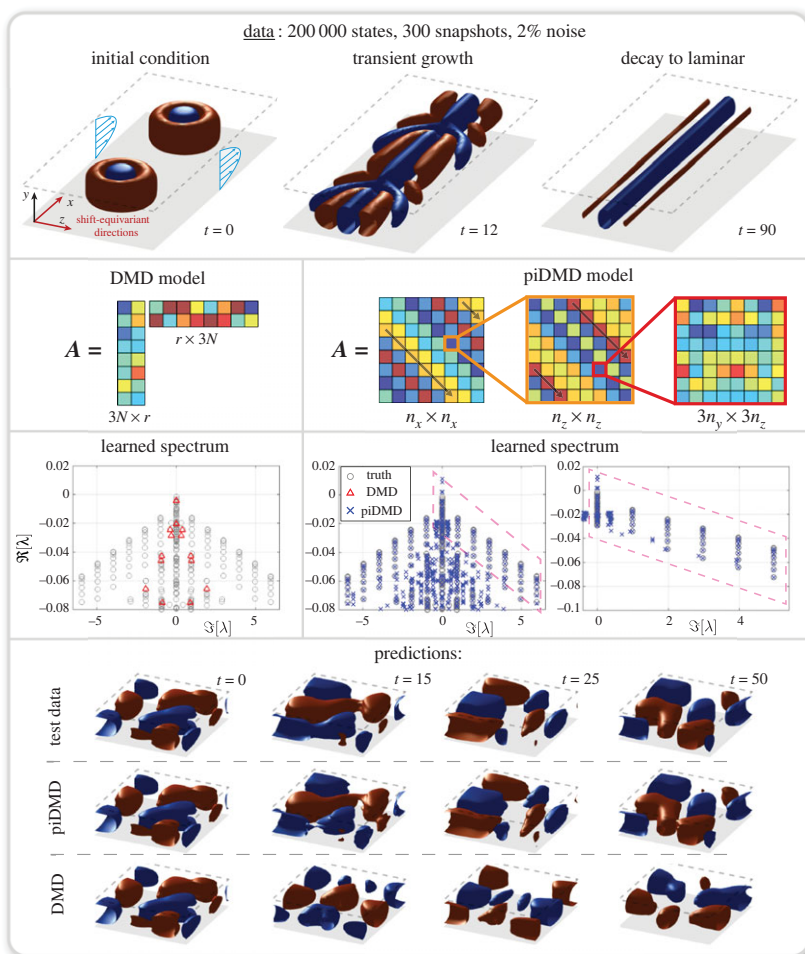
Figure 4. Learning the spectrum of the linearized Navier–Stokes equations from velocity measurements of the response to a localized disturbance in a channel flow. piDMD embeds the shift-equivariant structure of the Navier–Stokes equations into the model learning framework, thus learning the spectrum of the linearized operator with improved accuracy over exact DMD. The training data are corrupted with 2% noise. The predictions of the learned DMD and piDMD modes are also compared. The isosurfaces are surfaces of constant vertical velocity.

can leverage piDMD to incorporate this property, forcing the resulting matrix to respect, by construction, the structure with three nested levels shown in figure 4. In practice, this amounts to reshaping every snapshot, taking the FFT in the dimensions corresponding to x and z, and performing exact DMD on the y-dependent Fourier amplitudes for every streamwise and spanwise wavenumber tuple.

The spectra learned from both approaches are compared in figure 4, showing that piDMD results in a far more accurate and complete eigenvalue spectrum. Importantly, piDMD accurately identifies the most dynamically significant (least damped) eigenvalues, as illustrated in the close-up plot. piDMD identifies three unstable eigenvalues, but the corresponding modes are of low amplitude. By contrast, exact DMD identifies very few relevant eigenvalues despite the model's comparatively high rank of 200. As a result, the predictive ability of exact DMD is poor while piDMD produces very accurate predictions of unseen test data (bottom row).

(b) Conservative systems

Conservation laws are foundational to all areas of science. Identifying quantities that remain constant in time—such as mass, momentum, energy, electric charge and probability—allow us to derive universal governing equations, apply inductive reasoning, and understand basic physical mechanisms. In this section, we demonstrate how conservation laws can be incorporated into the DMD framework.

Suppose that we are studying a system that we know conserves energy. In applications of DMD, it is implicitly assumed that measurements of the state have been suitably weighted so that the square of the 2-norm corresponds to the energy of the state [73]: $E(x) = ||x||_2^2$. In these variables, the original optimization problems (1.1), (1.2) equate to finding the model A that minimizes the energy of the errors between the true and predicted states (y_k and Ax_k, respectively). Thus, if A represents a discrete-time linear dynamical system ($y_k = x_{k+1} = Ax_k$) then A is an energy preserving operator if and only if

$$E(Ax) = ||Ax||_2^2 = ||x||_2^2 = E(x) \quad \text{for all } x \in \mathbb{R}^n. \tag{4.8}$$

In words, (4.8) states that A does not change the energy of the system but merely redistributes energy between the states. In matrix terminology, (4.8) holds if and only if A is *unitary*. Therefore, for conservative systems, the optimization problem (1.2) is the orthogonal Procrustes problem described in §2b, and the solution is given by (2.10). The eigenvalues of A lie on the unit circle, and the eigenvectors are orthogonal. Thus, the eigenvectors oscillate in time, with no growth or decay. Since the solution of the orthogonal Procrustes problem (2.10) requires the full SVD of an $n \times n$ matrix, it can be more computationally efficient to first project onto the leading POD modes and build a model therein. In this case, the model is only energy preserving within the subspace spanned by the leading POD modes.

Noise is a substantial problem for most DMD methods, and a common remedy is to phrase the DMD optimization (1.1) as a total least-squares problem [35]. Perhaps surprisingly, the solution to the orthogonal Procrustes problem (2.10) is also the solution to the total least-squares problem when the solution is constrained to be orthogonal [47]. Thus, the solution (2.10) is statistically optimal even when there is noise in both X and Y, as long as the noise has the same Gaussian distribution in X and Y. The solutions obtained by energy-preserving piDMD and total least-squares DMD [78] are often similar. Further, there are other optimizations closely related to the orthogonal Procrustes problem including the case where A lies on the Steifel manifold (i.e. when A is rectangular with orthonormal columns [52]), there is a weighting matrix [51], missing data [46] or high amounts of noise [50].

(i) Example: $Re = 100$ flow past a cylinder

We now apply this energy-preserving piDMD to study the flow past a cylinder, which is a benchmark problem for modal decompositions [24]. The Reynolds number is $Re = DU/\nu = 100$, where D is the diameter of the cylinder, U is the free-stream velocity and ν is the kinematic viscosity. The data consist of 151 samples of vorticity measurements, corresponding to five periods of vortex shedding, taken on a grid of 199×449 points. We note that the 2-norm of the measurements is approximately constant in time. Note that the cylinder flow problem is only approximately conservative due to the balance of the applied pressure and viscous loss; in reality, the (numerical) energy of the state vector exhibits small oscillations. The data are contaminated with 20% Gaussian noise, as illustrated in the top left panel of figure 5. We truncate the data to the first 15 POD modes and learn a standard DMD model, and an energy-preserving piDMD model. piDMD learns a more accurate representation of the leading-order dynamics than standard DMD. For the high-order modes, the eigenvalues learned by DMD exhibits (well-known) spurious damping [34] whereas the eigenvalues learned by piDMD are purely oscillatory and remain on the imaginary axis. The trajectories of the POD coefficients are also more accurate for piDMD,
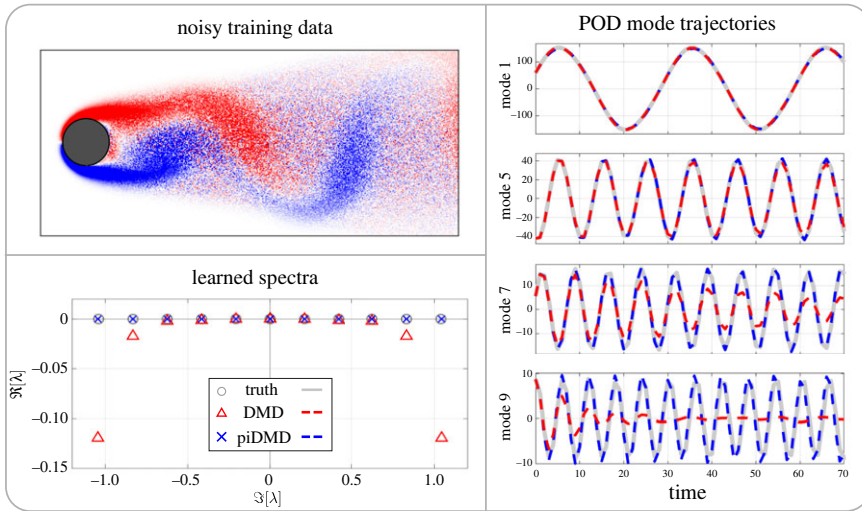
Figure 5. Learning the spectrum of the flow past a cylinder with noisy measurements with DMD and energy-preserving piDMD.

and do not exhibit the spurious energy loss caused by noise. Note that total least-squares DMD exhibits comparable performance to piDMD in this problem.

(c) Self-adjoint systems

Self-adjoint systems are another important class of linear systems that arise frequently in solid mechanics, heat and mass transfer, fluid mechanics and quantum mechanics. When studying a system known to be self-adjoint, we can use piDMD to constrain our model A to lie in the manifold of symmetric (Hermitian) matrices, such that $A = A^*$. Symmetric matrices have real eigenvalues and, by the spectral theorem, are diagonalizable. This restriction places a significant constraint on the learned model and can substantially reduce the sensitivity of DMD to noise.

The minimum-norm solution of the symmetric Procrustes problem [79] is

$$A = U_X L U_X^*, \tag{4.9}$$

where the entries of L are

$$L_{i,j} = \overline{L_{j,i}} = \begin{cases} \dfrac{\sigma_i \overline{C_{j,i}} + \sigma_j C_{i,j}}{\sigma_i^2 + \sigma_j^2} & \text{if } \sigma_i^2 + \sigma_j^2 \neq 0, \\ 0 & \text{otherwise,} \end{cases} \tag{4.10}$$

where $C = U_X^* Y V_X$ and $X = U_X \Sigma_X V_X^*$ is the SVD of X, and σ_i is the ith singular value. The solution of the skew-symmetric case is similar (electronic supplementary material, C.1), and algorithmic solutions are available when A is instead constrained to be positive definite [48]. For large-scale systems, a low-rank approximation to A can be obtained by projecting the system onto the first r principal components of X. This low-rank projection preserves the self-adjointness of the model.

Note that the symmetric Procrustes problem is connected to the orthogonal Procrustes problem (§4b) via the fact that every unitary matrix U can be expressed as the exponential of a Hermitian matrix H as $U = \exp(iH)$.

(i) Learning energy states of the Schrödinger equation

Possibly the most famous self-adjoint operator is the quantum Hamiltonian, $\hat{H}$ [80]. In quantum mechanics, physical quantities of interest, such as position, momentum, energy and spin, are
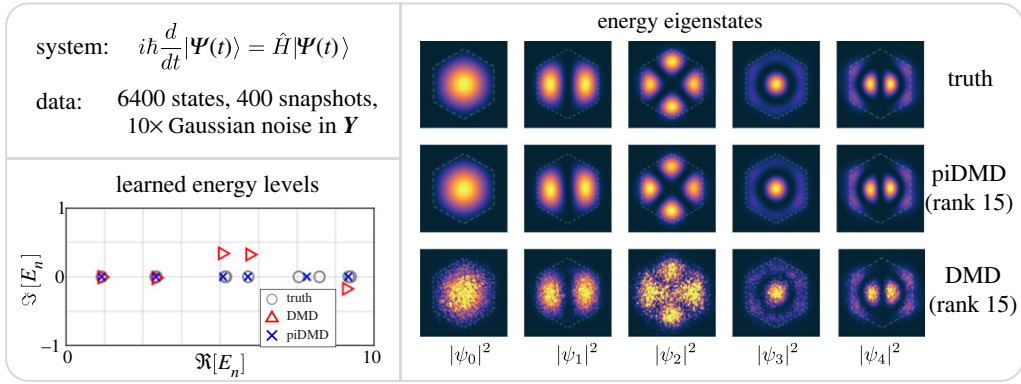
Figure 6. Learning the eigenstates of a quantum Hamiltonian of a hexagonal well with the self-adjoint piDMD.

represented by self-adjoint linear operators called 'observables', of which one example is the Hamiltonian. The Hamiltonian describes the evolution of the probabilistic wave function via the time-dependent Schrödinger equation

$$i\hbar \frac{\mathrm{d}}{\mathrm{d}t}|\Psi(t)\rangle = \hat{H}|\Psi(t)\rangle. \tag{4.11}$$

Solutions of (4.11) take the form of an eigenmode expansion

$$\Psi(\boldsymbol{\xi}, t) = \sum_{j=1}^{\infty} \alpha_j \, \psi_j(\boldsymbol{\xi}) \, \mathrm{e}^{-\mathrm{i}E_j t/\hbar}, \tag{4.12}$$

where E_j is an energy level (eigenvalue) of $\hat{H}$ with corresponding eigenstate ψ_j, the coefficients α_j are determined by the initial distribution of the wave function, and $\boldsymbol{\xi}$ is the spatial coordinate.

We consider a wave function evolving according to the two-dimensional Schrödinger's equation subject to an unknown Hamiltonian function. The unknown potential function is taken to be a finite well in a hexagonal shape. We collect measurements of the wave function and its velocity and train an exact DMD model and a piDMD model. In practice, empirical measurements of the wave function can be obtained through optical tomographic methods [81,82] or sequential measurements of complementary variables [83]. The data matrices X and Y are formed from measurements of $|\Psi(t)\rangle$ and $i\hbar(\mathrm{d}/\mathrm{d}t)|\Psi(t)\rangle$, respectively, which are themselves formed by superposing eigenfunctions obtained by a finite difference method. Thus, the matrices X and Y are connected by an unknown self-adjoint Hamiltonian $\hat{H}$ and we may apply our self-adjoint piDMD optimization. The measurements of Y are contaminated with a large amount (200%) of Gaussian noise but the measurements of X are clean.

Constraining the learned Hamiltonian to be self-adjoint reduces the sensitivity of the learning process to noise, as evidenced by figure 6. From the bottom left panel, we see that the energy levels of the piDMD model are physically consistent insofar as they are all real and positive. By contrast, the energy levels learned by standard DMD exhibit a spurious imaginary component, which will produce unrealistic growth/decay of the eigenstates. Both methods miss one eigenstate ($E_n \sim 7.5$) since this particular mode is not well represented in the data. Additionally, the eigenstates learned by piDMD are much less noisy than those learned by standard DMD. This can be explained by noting that the eigenstates learned by piDMD lie in the span of X, which is clean, whereas the eigenstates learned by standard DMD lie in the span of Y, which is noisy. In the electronic supplementary material, lemma C.1 proves that the symmetric piDMD model (4.9) is less sensitive to noise than the exact DMD model (2.4). In summary, a physics-aware approach enables a more accurate identification of quantum energy levels and the associated eigenstates.

(d) Spatially local systems

A hallmark of most physical systems is *spatial locality*. Points close to one another will usually have a stronger coupling than points that are far from one another. For example, basic physical processes such as advection and diffusion are spatially local. In most DMD applications, the data are collected from a spatial grid, but the grid rarely plays a role in the DMD analysis beyond forming the energy norm [73]. In fact, the output of the exact DMD algorithm remains invariant to unitary transformations of the rows and columns of the data matrices, so that randomly shuffling the rows (i.e. the spatial locations) will result in identical DMD models with the corresponding shuffling of the rows of the modes [62,84]. Knowledge of the underlying grid from which the data were collected enables us to bias the learning process to prioritize relationships that we expect to have a strong local coupling.

Herein we consider the one-dimensional case; the analysis generalizes straightforwardly to higher dimensions. We assume that the entries of x are samples of a function u taken at n grid points $\{\xi_i\}$. The grid points can be arranged sequentially so that $\xi_i < \xi_{i+1}$. By the spatial locality principle, we expect states that are spatially close to one another to have a stronger coupling than states that are spatially far from one another. For example, we may expect entries close to the diagonal of A to be larger than entries far from the diagonal. Then the entries of A would satisfy

$$|A_{i,j}| \geq |A_{i,k}| \quad \text{for } |j - i| < |k - i|. \tag{4.13}$$

Equation (4.13) is merely a heuristic, and we would expect it to hold on average rather than for each (i, j, k). Additionally, (4.13) is a difficult and expensive condition to implement in practice. Instead, we consider an alternative version of spatial locality where we only permit coupling between states that are sufficiently close

$$A_{i,k} = 0 \quad \text{for } d < |k - i|. \tag{4.14}$$

Equation (4.14) describes a d-diagonal matrix. If $d = 0$ then states can only affect themselves and A is diagonal (comparable to (4.4)). A more interesting case is if $d = 1$ and we only allow coupling between adjacent states. Then A is tridiagonal: the entries of A are zero except the leading, upper and lower diagonals

$$A = \begin{bmatrix} \beta_1 & \gamma_1 & & & \\ \alpha_2 & \beta_2 & \gamma_2 & & \\ & \alpha_3 & \ddots & \ddots & \\ & & \ddots & \ddots & \gamma_{n-1} \\ & & & \alpha_n & \beta_n \end{bmatrix}. \tag{4.15}$$

We now solve the optimization problem (1.2) when $\mathcal{M}$ is the manifold of tridiagonal matrices. Electronic supplementary material, D includes many solutions for more general diagonal-like structures including longer-range and variable coupling (electronic supplementary material, D.1), spatially periodic local systems (electronic supplementary material, D.2), self-adjoint (electronic supplementary material, D.3), total-least squares (electronic supplementary material, D.4) and weaker local structures (electronic supplementary material, D.5). Since the rows of A are decoupled, we can row-wise expand the Frobenius norm in (1.2) to obtain n smaller minimization problems

$$\arg\min_{\alpha_i, \beta_i, \gamma_i} ||\alpha_i \tilde{x}_{i-1} + \beta_i \tilde{x}_i + \gamma_i \tilde{x}_{i+1} - \tilde{y}_i||_2 \quad \text{for } 1 \leq i \leq n, \tag{4.16}$$

with the convention $\alpha_1 = \gamma_n = 0$ and $\tilde{x}_0 = \tilde{x}_{n+1} = \mathbf{0}$. Recall that $\tilde{x}_i$ and $\tilde{y}_i$ are the ith rows of X and Y, respectively (2.1). Each minimization problem (4.16) has the minimum-norm solution

$$\begin{bmatrix} \alpha_i & \beta_i & \gamma_i \end{bmatrix} = \tilde{y}_i \begin{bmatrix} \tilde{x}_{i-1} \\ \tilde{x}_i \\ \tilde{x}_{i+1} \end{bmatrix}^\dagger \quad \text{for } 1 \leq i \leq n. \tag{4.17}$$

Each of the n minimizations costs $\mathcal{O}(m)$ so the total cost is $\mathcal{O}(mn)$.
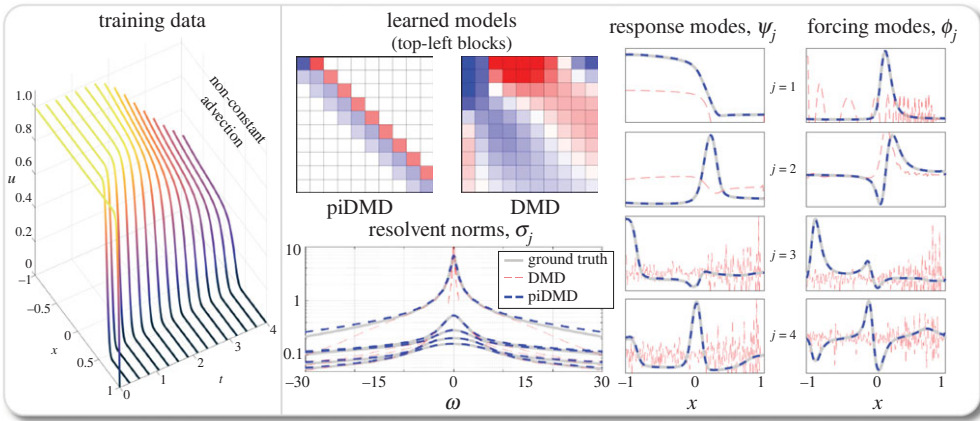
Figure 7. Data-driven resolvent analysis of the advection-diffusion equation for DMD and piDMD. The piDMD model is constrained to be tridiagonal and therefore respects the spatial locality of the system. Incorporating this physical structure results a significant improvement in the identification of the resolvent norm, response modes and forcing modes.

(i) Example: data-driven resolvent analysis

We consider a system of the form

$$u_t = \mathcal{A}u + f(\xi, t), \tag{4.18}$$

and aim to design a control strategy for the forcing f via input–output analysis [73,85–87]. Fourier transforming in time and rearranging produces $\hat{u} = R_\omega \hat{f}$ where $R_\omega = (i\omega - \mathcal{A})^{-1}$ is the resolvent of $\mathcal{A}$ and the hats $\hat{\cdot}$ denote the Fourier transform in time. We wish to understand the form of the forcing $\hat{f}$ that produces the largest response in $\hat{u}$. Formally, the optimal forcing (ϕ_1), response (ψ_1) and gain (σ_1) for each frequency component satisfy

$$\phi_1 = \arg\max_{||\phi_1||_2 = 1} ||R_\omega \phi_1||_2, \quad \sigma_1 = ||R_\omega \phi_1||_2, \quad \psi_1 = \sigma_1^{-1} R_\omega \phi_1. \tag{4.19}$$

We may also investigate the higher-order singular triplets (σ_j, ϕ_j, ψ_j) that satisfy equivalent conditions with the additional requirement of orthogonality of $\{\phi_j\}$ and $\{\psi_j\}$. Equivalently, we wish to determine the Hilbert–Schmidt decomposition of R_ω [88]:

$$R_\omega = \sum_{j=1}^{\infty} \sigma_j \psi_j(\xi) \langle \phi_j, \cdot \rangle, \tag{4.20}$$

which is analogous to the SVD of the discretization of R_ω.

Recently, the authors have proposed 'data-driven resolvent analysis' as an efficient method for learning the singular triplets of a system purely from data measurements [73]. The method is analogous to DMD, except instead of computing the eigenmodes via eigendecomposition, we compute the resolvent modes via a singular value decomposition. In the following example, we demonstrate that piDMD can be integrated into data-driven resolvent analysis to provide physics-informed diagnostic information about the optimal forcings and responses of the system.

We consider the problem of designing a strategy to control the concentration of a solute that is governed by unknown dynamics. The solute concentration u is governed by the advection-diffusion operator $\mathcal{A}u = u_{\xi\xi} - v(\xi)u_\xi$ where v represents a non-constant advection term (here taken to be a random function) and boundary conditions $u_\xi(-1) = u(1) = 0$. A typical simulation is illustrated in the left panel of figure 7. The advection-diffusion operator is a canonical example of a non-normal system where the typical eigenvalue analysis fails to provide meaningful information [89,90]. Our goal is to determine the forcings $\hat{f}$ that produce the largest increase in solute in the frequency domain.

We solve the advection-diffusion equation in Chebfun [91] and collect 2000 snapshots of u evaluated at 100 evenly spaced grid points. The snapshots matrices X and Y are formed from measurements of u and u_t, respectively. We apply data-driven resolvent analysis with DMD and piDMD and compute the learned operators, resolvent norm and forcing and response modes. For both algorithms, we approximate u_t from the measurements of u via finite differences. The piDMD model is constrained to be tridiagonal, and the DMD model has rank 40. From the visualizations of the learned models in the centre of figure 7, we see that the standard DMD identifies spurious relationships between the states, whereas piDMD, by construction, identifies a stronger coupling between adjacent states. The model learned by piDMD is reminiscent of the finite difference matrix; indeed, local piDMD may be viewed as a data-driven discovery of a finite difference stencil.

The benefit of piDMD is evidenced in its accurate identification of the resolvent modes (for $\omega = 1$), as seen in the right panel of figure 7. The true resolvent modes are computed in Chebfun (see section 12 of [92]). The modes learned by standard DMD are highly inaccurate for all j, and are completely incoherent for $j > 2$. By contrast, piDMD accurately determines the mode shapes, and associated gains for a range of frequencies. The physics-informed approach even determines the correct boundary conditions of the response and forcing modes, and resolves the complicated structures associated with the non-normality of the system. In particular, enforcing spatial locality enables piDMD to identify modes that are poorly represented by the data, which allows for diagnostic analysis that generalizes outside of the training regime. Equipped with these data-driven modes, one could now design a low-rank control strategy for the solute concentration.

Analogously to finite difference methods, the accuracy of local piDMD involves a trade-off between the grid sizes in space and time. Future work should focus on clarifying these issues.

(e) Causal systems

Causality is the process by which the behaviour of one state (the cause) influences another state (the effect). Many systems possess a spatially causal structure where each state depends only on 'upstream' states and are unaffected by 'downstream' states. Specifically, we may express the jth state as a function of the $n - j + 1$ upstream states only: $y_j = f(x_j, x_{j+1}, \ldots x_n)$. For example, y_n depends only on x_n whereas y_1 depends only all the elements of x. Accordingly, if we know that the system at hand possesses a spatially causal structure, we may seek a piDMD model that is upper triangular. Such structures are typical of a system with causal features, for example, time-delay coordinates [93], feed-forward/strict-feedback systems [94,95] and other control systems [96].

As in §4d, we may decouple the rows of A in (1.2) and expand the cost function row-wise to obtain n smaller minimization problems

$$\arg\min_{A_{j,j:n}} ||A_{j,j:n} X_{j:n,:} - \tilde{y}_j||_2 \quad \text{for } 1 \leq j \leq n. \tag{4.21}$$

where we have employed the MATLAB colon notation $j : n$ to indicate entries from j to n and $:$ to indicate entries from 1 to m. Here, $\tilde{y}_j$ represents the jth row of Y. The (minimum norm) solution of (4.21) is obtained via the pseudoinverse of each sub-block of X as

$$A_{j,j:n} = \tilde{y}_j X_{j:n,:}^{\dagger} \quad \text{for } 1 \leq j \leq n. \tag{4.22}$$

Naively evaluating (4.22) for each j would require building n pseudoinverses for a total cost of $\mathcal{O}(n^2 m \min(m,n))$. Fortunately, consecutive blocks of X (e.g. $X_{j-1:n,:}$ and $X_{j:n,:}$) are related by a rank-1 update, so we can solve every problem in (4.22) for a total of $\mathcal{O}(nm \min(n,m))$ operations (see electronic supplementary material, E.1 for further details). However, this approach is very numerically unstable when applied to realistic data with high-condition number. Thus, in electronic supplementary material, E.2, we derive a more stable alternative.

We illustrate this causal structure on a simple example defined by the Volterra-type integro-differential equation

$$\frac{\partial u}{\partial t}(\xi, t) = \int_{-1}^{\xi} K(\xi, \nu) u(\nu, t)\, d\nu \quad -1 \leq \xi \leq 1. \tag{4.23}$$

Similar systems have been used to model transmission lines in neural networks during bursting activity [97] and the spread of disease in epidemics [98]. Taking $K(\xi, \nu) = \sqrt{1 - \xi^2}\sqrt{1 - \nu^2}$ and $u(\xi, 0) = \exp(-\xi^2)$, we numerically simulate (4.23) and construct X and Y from measurements of u only (velocity measurements are not used). The results of piDMD are illustrated in the penultimate row of figure 3, and indicate that piDMD is able to learn the leading eigenvalues while standard DMD fails. Note that the true spectrum here is continuous.

5. Conclusion and outlook

This work presents piDMD, a physics-aware modal decomposition technique that extracts coherent structures from high-dimensional time-series data. Rephrasing the DMD regression (1.1) as a Procrustes problem (1.2) shows that partially known physics can be incorporated into the DMD framework by enforcing a matrix manifold constraint determined by known physics. We have applied piDMD to five of the most fundamental physical principles: conservation laws, self-adjointness, shift-equivariance, locality and causality. In several of these cases, we have presented new 'exact' solutions of the corresponding optimization problems. The solutions usually require comparable computational effort to the exact DMD method, but sometimes piDMD models are far cheaper to compute. The examples presented demonstrate that piDMD can exhibit superior performance compared to classical DMD methods that do not account for the physics of the system.

The framework developed herein offers exciting new opportunities and challenges for data-driven dynamical systems, numerical linear algebra, convex optimization and statistics. Fortunately, many research advances developed for classical DMD can also be leveraged to enhance piDMD. We conclude the paper with a brief discussion of the limitations of piDMD, and suggest promising future research directions.

Each of the Procrustes steps of the piDMD framework outlined in §3 pose distinct challenges. In some scenarios where the physics is poorly understood, determining suitable physical laws to impose on the model (step i, 'modelling') can be challenging. For problems with intricate geometries and multiple dimensions, interpreting the physical principle as a matrix manifold (step ii, 'interpretation') can be the roadblock, as the manifold can become exceedingly complicated. We have already commented on the challenges and opportunities posed by step iii (optimization), and suggested algorithmic solutions [56]. Finally, the typically large state dimension and number of samples can obfuscate step iv when the matrix manifold does not easily admit the desired diagnostics. For example, if A has a complicated banded structure, as in local piDMD, it may not be amenable to a fast eigendecomposition or SVD.

The solutions presented herein exhibit various levels of sensitivity to noise. As is typical of DMD methods, most solutions are unbiased with respect to Y, but can be quite sensitive to noise in X. Some of the solutions we have presented are relatively insensitive (§4b) or can be reformulated in the total least-squares sense (electronic supplementary material, B.4, D.4). A full characterization of the sensitivity of piDMD should be performed on a case-by-case basis for different manifold constraints, and will be the subject of future investigations.

The Procrustes problem (1.2) is not the only means of incorporating physical principles into the DMD process. Closely related to Procrustes problems are 'nearest matrix' problems [99] where, given a matrix $\hat{A}$, we seek the closest matrix on some matrix manifold $\mathcal{M}$

$$\arg\min_{A \in \mathcal{M}} ||A - \hat{A}||. \tag{5.1}$$

Equation (5.1) can be viewed as a special case of the Procrustes problem (1.2) where $Y = \hat{A}$ and $X = I$. As such, nearest matrix problems are generally easier to solve than Procrustes problems.

The nearest matrix problem (5.1) can be used to project any DMD model onto the nearest physically consistent model, or within optimization routines, such as proximal gradient methods or constrained gradient descent, to constrain the solution to the feasible set.

As presented, piDMD strictly requires that the model A lies on the chosen matrix manifold $\mathcal{M}$. In many applications, such as when the data are very noisy or the physical laws and/or constraints are only approximately understood, it may be more appropriate to merely 'encourage' A towards $\mathcal{M}$. In such cases, the piDMD regression (1.2) becomes

$$\arg \min_{A} ||Y - AX||_F + \lambda R(A), \tag{5.2}$$

where the first term is the reconstruction loss, R is a physics-informed regularizer and λ is a user-defined constant that tunes the relative importance of the reconstruction and regularization. For example, $R(A)$ could represent the distance between A and $\mathcal{M}$, which can be computed by solving the nearest matrix problem (5.1). An example of such a problem is solved in electronic supplementary material, D.5.

DMD models may be trained in discrete or continuous time. In discrete time with a constant sampling rate, the snapshots in Y are $y_k = x_{k+1}$, whereas in the continuous time, we have $y_k = \dot{x}_k$. If the continuous-time operator is A_C then the discrete time operator is $A_D = \exp(\Delta t A_C)$. In special cases, the matrices A_C and A_D may lie on the same manifold, such as the upper-triangular and circulant manifolds; however, in general, they do not. For example, if A_C is tridiagonal, then A_D is not generally tridiagonal. Accordingly, if only discrete-time measurements are available but the matrix manifold is imposed on the continuous-time operator, then the optimization problem to solve is

$$\arg \min_{A \in \mathcal{M}} ||Y - \exp(\Delta t A)X||_F, \tag{5.3}$$

which is a much more difficult optimization problem.

Although this paper has focused on linear systems, one could apply the underlying principles to neural networks by constraining the weight matrix of each layer to a specific manifold. The chosen matrix manifolds dictate the neural network architecture by enforcing specific relationships between the neurons. For example, a tridiagonal weight matrix would mean that only certain neurons are connected, while a circulant weight matrix is representative of a convolutional layer. In addition to enforcing physical structures, these constraints have numerical advantages: Jing *et al.* [100] demonstrated that constraining a network's weight matrices to be unitary can address vanishing/exploding gradients. These ideas will be investigated in future work.

References

1. Karniadakis GE, Kevrekidis IG, Lu L, Perdikaris P, Wang S, Yang L. 2021 Physics-informed machine learning. *Nat. Rev. Phys.* **3**, 422–440. (doi:10.1038/s42254-021-00314-5)

2. Behler J, Parrinello M. 2007 Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett.* **98**, 146401. (doi:10.1103/PhysRevLett.98.146401)

3. Greydanus S, Dzamba M, Yosinski J. 2019 Hamiltonian neural networks. In *Adv. Neural Inf. Process. Syst.*, vol. 32. Red Hook, NY: Curran Associates.

4. Reichstein M, Camps-Valls G, Stevens B, Jung M, Denzler J, Carvalhais N, Prabhat . 2019 Deep learning and process understanding for data-driven Earth system science. *Nature* **566**, 195–204. (doi:10.1038/s41586-019-0912-1)

5. Raissi M, Perdikaris P, Karniadakis GE. 2019 Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comput. Phys.* **378**, 686–707. (doi:10.1016/j.jcp.2018.10.045)

6. Cranmer M, Greydanus S, Hoyer S, Battaglia P, Spergel D, Ho S. 2020 Lagrangian neural networks. (http://arxiv.org/abs/2003.04630)

7. Raissi M, Yazdani A, Karniadakis GE. 2020 Hidden fluid mechanics: learning velocity and pressure fields from flow visualizations. *Science* **367**, 1026–1030. (doi:10.1126/science.aaw4741)

8. Wang R, Maddix D, Faloutsos C, Wang Y, Yu R. 2020 Bridging physics-based and data-driven modeling for learning dynamical systems. *Proc. Mach. Learn. Res.* **144**, 385–398.

9. Lu L, Jin P, Pang G, Zhang Z, Karniadakis GE. 2021 Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators. *Nat. Mach. Intell.* **3**, 218–229. (doi:10.1038/s42256-021-00302-5)

10. Li Z, Kovachki N, Azizzadenesheli K, Liu B, Bhattacharya K, Stuart A, Anandkumar A. 2021 Fourier neural operator for parametric partial differential equations. In *9th Int. Conf. Learn. Represent.*, *Vienna, Austria*, 3–7 May.

11. Baddoo PJ, Herrmann B, McKeon BJ, Brunton SL. 2021 Kernel learning for robust dynamic mode decomposition: linear and nonlinear disambiguation optimization (LANDO). (http://arxiv.org/abs/2106.01510)

12. Klus S, Gelß P, Nüske F, Noé F. 2021 Symmetric and antisymmetric kernels for machine learning problems in quantum physics and chemistry. *Mach. Learn.: Sci. Technol.* **2**, 045016.

13. Shah V, Joshi A, Ghosal S, Pokuri B, Sarkar S, Ganapathysubramanian B, Hegde C. 2019 Encoding invariances in deep generative models. (http://arxiv.org/abs/1906.01626)

14. Darakananda D, Eldredge J, de Castro da Silva A, Colonius T, Williams D. 2018 EnKF-based dynamic estimation of separated flows with a low-order vortex model. In *AIAA Aerosp. Sci. Meet. 2018*, no. 210059 (Reston, Virginia), American Institute of Aeronautics and Astronautics.

15. Le Provost M, Baptista R, Marzouk Y, Eldredge JD. 2022 A low-rank ensemble Kalman filter for elliptic observations. *Proc. R. Soc. A* **478**, 20220182. (doi:10.1098/rspa.2022.0182)

16. Brunton SL, Proctor JL, Kutz JN. 2016 Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proc. Natl Acad. Sci. USA* **113**, 3932–3937. (doi:10.1073/pnas.1517384113)

17. Rudy SH, Brunton SL, Proctor JL, Kutz JN. 2017 Data-driven discovery of partial differential equations. *Sci. Adv.* **3**, e1602614. (doi:10.1126/sciadv.1602614)

18. Loiseau JC, Brunton SL. 2018 Constrained sparse Galerkin regression. *J. Fluid Mech.* **838**, 42–67. (doi:10.1017/jfm.2017.823)

19. Guan Y, Brunton SL, Novosselov I. 2021 Sparse nonlinear models of chaotic electroconvection. *R. Soc. Open Sci.* **8**, 202367. (doi:10.1098/rsos.202367)

20. Zanna L, Bolton T. 2020 Data-driven equation discovery of ocean mesoscale closures. *Geophys. Res. Lett.* **47**, e2020GL088376. (doi:10.1029/2020GL088376)

21. Schmid PJ. 2010 Dynamic mode decomposition of numerical and experimental data. *J. Fluid Mech.* **656**, 5–28. (doi:10.1017/S0022112010001217)

22. Rowley CW, Mezić I, Bagheri S, Schlatter P, Henningson DS. 2009 Spectral analysis of nonlinear flows. *J. Fluid Mech.* **641**, 115–127. (doi:10.1017/S0022112009992059)

23. Tu JH, Rowley CW, Luchtenburg DM, Brunton SL, Kutz JN. 2014 On dynamic mode decomposition: theory and applications. *J. Comput. Dyn.* **1**, 391–421.

24. Kutz JN, Brunton SL, Brunton BW, Proctor JL. 2016 *Dynamic mode decomposition: data-driven modeling of complex systems*. Philadelphia, PA: SIAM.

25. Schmid PJ. 2022 Dynamic mode decomposition and its variants. *Annu. Rev. Fluid Mech.* **54**, 225–254. (doi:10.1146/annurev-fluid-030121-015835)

26. Proctor JL, Eckhoff PA. 2015 Discovering dynamic patterns from infectious disease data using dynamic mode decomposition. *Int. Health* **7**, 139–145. (doi:10.1093/inthealth/ihv009)

27. Berger E, Sastuba M, Vogt D, Jung B, Amor HB. 2015 Estimation of perturbations in robotic behavior using dynamic mode decomposition. *J. Adv. Rob.* **29**, 331–343. (doi:10.1080/01691864.2014.981292)

28. Abraham I, Murphey TD. 2019 Active learning of dynamics for data-driven control using Koopman operators. *IEEE Trans. Rob.* **35**, 1071–1083. (doi:10.1109/TRO.2019.2923880)

29. Brunton BW, Johnson LA, Ojemann JG, Kutz JN. 2016 Extracting spatial–temporal coherent patterns in large-scale neural recordings using dynamic mode decomposition. *J. Neurosci. Methods* **258**, 1–15. (doi:10.1016/j.jneumeth.2015.10.010)

30. Goldschmidt A, Kaiser E, DuBois JL, Brunton SL, Kutz JN. 2021 Bilinear dynamic mode decomposition for quantum control. *New J. Phys.* **23**, 033035. (doi:10.1088/1367-2630/abe972)

31. Susuki Y, Mezić I, Hikihara T. 2011 Coherent swing instability of power grids. *J. Nonlinear Sci.* **21**, 403–439. (doi:10.1007/s00332-010-9087-5)

32. Taylor R, Kutz JN, Morgan K, Nelson BA. 2018 Dynamic mode decomposition for plasma diagnostics and validation. *Rev. Sci. Instrum.* **89**, 053501. (doi:10.1063/1.5027419)

33. Kaptanoglu AA, Morgan KD, Hansen CJ, Brunton SL. 2020 Characterizing magnetized plasmas with dynamic mode decomposition. *Phys. Plasmas* **27**, 032108. (doi:10.1063/1.5138932)

34. Bagheri S. 2014 Effects of weak noise on oscillating flows: linking quality factor, Floquet modes, and Koopman spectrum. *Phys. Fluids* **26**, 094104. (doi:10.1063/1.4895898)

35. Dawson ST, Hemati MS, Williams MO, Rowley CW. 2016 Characterizing and correcting for the effect of sensor noise in the dynamic mode decomposition. *Exp. Fluids* **57**, 42. (doi:10.1007/s00348-016-2127-7)

36. Hemati MS, Rowley CW, Deem EA, Cattafesta LN. 2017 De-biasing the dynamic mode decomposition for applied Koopman spectral analysis. *TCFD* **31**, 349–368.

37. Sondermann D. 1986 Best approximate solutions to matrix equations under rank restrictions. *Stat. Hefte* **27**, 57–66. (doi:10.1007/BF02932555)

38. Héas P, Herzet C. 2022 Low-Rank dynamic mode decomposition: an exact and tractable solution. *J. Nonlinear Sci.* **32**, 1–24.

39. Tu JH, Rowley CW, Luchtenburg DM, Brunton SL, Kutz JN. 2014 On dynamic mode decomposition: theory and applications. *J. Comput. Dyn.* **1**, 391–421.

40. Askham T, Kutz JN. 2018 Variable projection methods for an optimized dynamic mode decomposition. *SIAM J. Appl. Dyn. Syst.* **17**, 380–416. (doi:10.1137/M1124176)

41. Hurley JR, Cattell RB. 1962 The Procrustes program: producing direct rotation to test a hypothesized factor structure. *Behav. Sci.* **7**, 258–262. (doi:10.1002/bs.3830070216)

42. Söderkvist I. 1993 Perturbation analysis of the orthogonal Procrustes problem. *BIT* **33**, 687–694.

43. Trendafilov NT, Lippert RA. 2002 The multimode Procrustes problem. *Linear Algebra Appl.* **349**, 245–264. (doi:10.1016/S0024-3795(02)00253-7)

44. Gower JC, Dijksterhuis GB. 2007 *Procrustes problems*. Oxford, NJ: Oxford University Press.

45. Higham NJ. 1988 The symmetric Procrustes problem. *BIT* **28**, 133–143. (doi:10.1007/BF01934701)

46. ten Berge JM, Kiers HA, Commandeur JJ. 1993 Orthogonal Procrustes rotation for matrices with missing values. *Br. J. Math. Stat. Psychol.* **46**, 119–134.

47. Arun KS. 1992 A unitarily constrained total least squares problem in signal processing. *SIAM J. Matrix Anal. Appl.* **13**, 729–745. (doi:10.1137/0613046)

48. Gillis N, Sharma P. 2018 A semi-analytical approach for the positive semidefinite Procrustes problem. *Linear Algebra Appl.* **540**, 112–137. (doi:10.1016/j.laa.2017.11.023)

49. Schönemann PH. 1966 A generalized solution of the orthogonal Procrustes problem. *Psychometrika* **31**, 1–10.

50. Pumir T, Singer A, Boumal N. 2021 The generalized orthogonal Procrustes problem in the high noise regime. *Inf. Inference A J. IMA* **10**, 921–954. (doi:10.1093/imaiai/iaaa035)

51. Viklands T. 2006 *Algorithms for the weighted orthogonal Procrustes problem and other least squares problems*. PhD thesis, Umeå University.

52. Eldén L, Park H. 1999 A Procrustes problem on the Stiefel manifold. *Numer. Math.* **82**, 599–619.

53. Andersson LE, Elfving T. 1997 A constrained Procrustes problem. *SIAM J. Matrix Anal. Appl.* **18**, 124–139. (doi:10.1137/S0895479894277545)

54. Salova A, Emenheiser J, Rupe A, Crutchfield JP, D'Souza RM. 2019 Koopman operator and its approximations for systems with symmetries. *Chaos* **29**, 093128. (doi:10.1063/1.5099091)

55. Absil PA, Mahony R, Sepulchre R. 2009 *Optimization algorithms on matrix manifolds*. Princeton, NJ: Princeton University Press.
56. Boumal N, Mishra B, Absil PA, Sepulchre R. 2014 Manopt, a Matlab toolbox for optimization on manifolds. *J. Mach. Learn. Res.* **15**, 1455–1459.
57. Jovanović MR, Schmid PJ, Nichols JW. 2014 Sparsity-promoting dynamic mode decomposition. *Phys. Fluids* **26**, 024103.
58. Proctor JL, Brunton SL, Kutz JN. 2016 Dynamic mode decomposition with control. *SIAM J. Appl. Dyn. Syst.* **15**, 142–161. (doi:10.1137/15M1013857)
59. Noack BR, Stankiewicz W, Morzynski M, Schmid PJ. 2016 Recursive dynamic mode decomposition of a transient cylinder wake. *J. Fluid Mech.* **809**, 843–872. (doi:10.1017/jfm.2016.678)
60. Hemati MS, Williams MO, Rowley CW. 2014 Dynamic mode decomposition for large and streaming datasets. *Phys. Fluids (1994-present)* **26**, 111701. (doi:10.1063/1.4901016)
61. Zhang H, Rowley CW, Deem EA, Cattafesta LN. 2019 Online dynamic mode decomposition for time-varying systems. *SIAM J. Appl. Dyn. Syst.* **18**, 1586–1609. (doi:10.1137/18M1192329)
62. Brunton SL, Proctor JL, Tu JH, Kutz JN. 2015 Compressed sensing and dynamic mode decomposition. *J. Comput. Dyn.* **2**, 165–191.
63. Gueniat F, Mathelin L, Pastur L. 2015 A dynamic mode decomposition approach for large and arbitrarily sampled systems. *Phys. Fluids* **27**, 025113.
64. Tu JH, Rowley CW, Kutz JN, Shang JK. 2014 Spectral analysis of fluid flows using sub-Nyquist-rate PIV data. *Exp. Fluids* **55**, 1–13. (doi:10.1007/s00348-014-1805-6)
65. Golub GH, Van Loan CF. 2013 *Matrix computations*, vol. 3. Baltimore, MD: JHU Press.
66. Eckart C, Young G. 1936 The approximation of one matrix by another of lower rank. *Psychometrika* **1**, 211–218. (doi:10.1007/BF02288367)
67. Mezić I. 2013 Analysis of fluid flows via spectral properties of the Koopman operator. *Annu. Rev. Fluid Mech.* **45**, 357–378.
68. Brunton SL, Budišić M, Kaiser E, Kutz JN. 2021 Modern Koopman theory for dynamical systems. (http://arxiv.org/abs/2102.12086)
69. Dijksterhuis G. 1994 Procrustes analysis in studying sensory-instrumental relations. *Food Qual. Prefer.* **5**, 115–120. (doi:10.1016/0950-3293(94)90016-7)
70. Cox MAA, Cox TF. 2008 Multidimensional scaling. In *Handbook of Data Visualization*, pp. 315–347. Berlin: Springer.
71. Zhang X, Zou J, He K, Sun J. 2016 Accelerating very deep convolutional networks for classification and detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **38**, 1943–1955. (doi:10.1109/TPAMI.2015.2502579)
72. Richman MB. 1986 Rotation of principal components. *J. Climatol.* **6**, 293–335. (doi:10.1002/joc.3370060305)
73. Herrmann B, Baddoo PJ, Semaan R, Brunton SL, McKeon BJ. 2021 Data-driven resolvent analysis. *J. Fluid Mech.* **918**, A10. (doi:10.1017/jfm.2021.337)
74. Noether E. 1918 Invariante variationsprobleme. *Nachrichten von der Gesellschaft der Wissenschaften zu Güttingen, Mathematisch-Physikalische Klasse* **1918**, 235–257.
75. Dutt A, Rokhlin V. 1993 Fast Fourier transforms for nonequispaced data. *SIAM J. Sci. Comput.* **14**, 1368–1393. (doi:10.1137/0914081)
76. Gibson JF. 2014 Channelflow: a spectral Navier-Stokes simulator in C++. tech. rep., U. New Hampshire. Channelflow.org.
77. Ilak M, Rowley CW. 2008 Modeling of transitional channel flow using balanced proper orthogonal decomposition. *Phys. Fluids* **20**, 034103. (doi:10.1063/1.2840197)
78. Dawson ST, Hemati MS, Williams MO, Rowley CW. 2016 Characterizing and correcting for the effect of sensor noise in the dynamic mode decomposition. *Exp. Fluids* **57**, 1–19. (doi:10.1007/s00348-016-2127-7)
79. Higham NJ. 2008 *Functions of matrices: theory and computation*. Society for Industrial and Applied Mathematics.
80. Griffiths DJ. 1995 *Introduction to quantum mechanics*. Cambridge, UK: Cambridge University Press.
81. Smithey DT, Beck M, Raymer MG, Faridani A. 1993 Measurement of the Wigner distribution and the density matrix of a light mode using optical homodyne tomography: application to squeezed states and the vacuum. *Phys. Rev. Lett.* **70**, 1244–1247. (doi:10.1103/PhysRevLett.70.1244)

82. Vogel K, Risken H. 1989 Determination of quasiprobability distributions in terms of probability distributions for the rotated quadrature phase. *Phys. Rev. A* **40**, 2847–2849. (doi:10.1103/PhysRevA.40.2847)

83. Lundeen JS, Sutherland B, Patel A, Stewart C, Bamber C. 2011 Direct measurement of the quantum wavefunction. *Nature* **474**, 188–191. (doi:10.1038/nature10120)

84. Brunton SL, Kutz JN. 2019 *Data-driven science and engineering: machine learning, dynamical systems, and control*. Cambridge, UK: Cambridge University Press.

85. McKeon BJ, Sharma AS. 2010 A critical-layer framework for turbulent pipe flow. *J. Fluid Mech.* **658**, 336–382. (doi:10.1017/S002211201000176X)

86. Jovanović MR. 2021 From bypass transition to flow control and data-driven turbulence modeling: an input–output viewpoint. *Annu. Rev. Fluid Mech.* **53**, 311–345.

87. Trefethen LN, Trefethen AE, Reddy SC, Driscoll TA. 1993 Hydrodynamic stability without eigenvalues. *Science* **261**, 578–584. (doi:10.1126/science.261.5121.578)

88. Kato T. 1966 *Perturbation theory for linear operators*. Classics in Mathematics. Berlin, Heidelberg: Springer.

89. Reddy SC, Trefethen LN. 1994 Pseudospectra of the convection-diffusion operator. *SIAM J. Appl. Math.* **54**, 1634–1649. (doi:10.1137/S0036139993246982)

90. Trefethen LN, Embree M. 2005 *Spectra and pseudospectra: the behavior of nonnormal matrices and operators*. Princeton, NJ: Princeton University Press.

91. Driscoll TA, Hale N, Trefethen LN. 2014 *Chebfun guide*. Oxford, UK: Pafnuty Publications.

92. Aurentz JL, Trefethen LN. 2017 Block operators and spectral discretizations. *SIAM Rev.* **59**, 423–446. (doi:10.1137/16M1065975)

93. Brunton SL, Brunton BW, Proctor JL, Kaiser E, Nathan Kutz J. 2017 Chaos as an intermittently forced linear system. *Nat. Commun.* **8**, 19. (doi:10.1038/s41467-017-00030-8)

94. Krishnamurthy P, Khorrami F. 2007 Generalized state scaling and applications to feedback, feedforward, and nontriangular nonlinear systems. *IEEE Trans. Automat. Contr.* **52**, 102–108. (doi:10.1109/TAC.2006.886534)

95. Teel AR. 1996 A nonlinear small gain theorem for the analysis of control systems with saturation. *IEEE Trans. Automat. Contr.* **41**, 1256–1270. (doi:10.1109/9.536496)

96. Annaswamy AM, Baillieul J. 1994 Adaptive control of nonlinear systems with a triangular structure. *IEEE Trans. Automat. Contr.* **39**, 1411–1428.

97. Jackiewicz Z, Rahman M, Welfert BD. 2008 Numerical solution of a Fredholm integro-differential equation modelling $\dot{\theta}$-neural networks. *Appl. Math. Comput.* **195**, 523–536.

98. Medlock J, Kot M. 2003 Spreading disease: integro-differential equations old and new. *Math. Biosci.* **184**, 201–222. (doi:10.1016/S0025-5564(03)00041-5)

99. Higham NJ. 1989 Matrix nearness problems and applications. In *Applications of Matrix Theory* (eds MJC Gover, S Barnett), pp. 1–27. Oxford, UK: Oxford University Press.

100. Jing L, Shen Y, Dubcek T, Peurifoy J, Skirlo S, LeCun Y, Tegmark M, Soljačić M. 2017 Tunable efficient unitary neural networks (EUNN) and their application to RNNs. In *34th Int. Conf. Mach. Learn. ICML 2017*, *Sydney, Australia*, vol. 4, pp. 2753–2761.

101. Baddoo PJ, Herrmann B, McKeo BJ, Nathan Kutz J, Brunton SL. 2023 Physics-informed dynamic mode decomposition. Figshare. (doi:10.6084/m9.figshare.c.6423942)