





Robust principal component analysis for modal decomposition of corrupt fluid flows

Isabel Scherl ^{1,*} Benjamin Strom,¹ Jessica K. Shang,² Owen Williams ³
Brian L. Polagye ¹ and Steven L. Brunton ¹

¹Department of Mechanical Engineering, University of Washington, Seattle, Washington 98195, USA

²Department of Mechanical Engineering, University of Rochester, Rochester, New York 14627, USA

³Department of Aeronautics and Astronautics, University of Washington, Seattle, Washington 98195, USA



(Received 19 December 2019; accepted 1 April 2020; published 28 May 2020)

Modal analysis techniques are used to identify patterns and develop reduced-order models in a variety of fluid applications. However, experimentally acquired flow fields may be corrupted with incorrect and missing entries, which may degrade modal decomposition. Here we use robust principal component analysis (RPCA) to improve the quality of flow-field data by leveraging global coherent structures to identify and replace spurious data points. RPCA is a robust variant of principal component analysis, also known as proper orthogonal decomposition in fluids, that decomposes a data matrix into the sum of a low-rank matrix containing coherent structures and a sparse matrix of outliers and corrupt entries. We apply RPCA filtering to a range of fluid simulations and experiments of varying complexities and assess the accuracy of low-rank structure recovery. First, we analyze direct numerical simulations of flow past a circular cylinder at Reynolds number 100 with artificial outliers, alongside similar particle image velocimetry (PIV) measurements at Reynolds number 413. Next, we apply RPCA filtering to a turbulent channel flow simulation from the Johns Hopkins Turbulence database, demonstrating that dominant coherent structures are preserved in the low-rank matrix. Finally, we investigate PIV measurements behind a two-bladed cross-flow turbine that exhibits both broadband and coherent phenomena. In all cases, we find that RPCA filtering extracts dominant coherent structures and identifies and fills in incorrect or missing measurements. The performance is particularly striking when flow fields are analyzed using dynamic mode decomposition, which is sensitive to noise and outliers.

DOI: [10.1103/PhysRevFluids.5.054401](https://doi.org/10.1103/PhysRevFluids.5.054401)

I. INTRODUCTION

The ability to understand, model, and control fluid flows is foundational to advancing technologies in transportation, energy, health, and defense. The challenges these fields pose are not easily solved by first-principles analysis without intense simplification. Instead, we rely on data from simulations and experiments [1–3]. The fidelity of both approaches have improved dramatically, generating vast and increasing volumes of data [4]. However, despite this apparently

*Corresponding author: ischerl@uw.edu

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

high-dimensional data, fluid dynamics are often characterized by the evolution of a few dominant coherent structures that are energetically or dynamically important [5–9]. Thus, even with increasing ambient measurement dimension, the *intrinsic* dimension of the flow may remain relatively low. Modal decomposition techniques are designed to extract these meaningful patterns from high-dimensional fluids data [3,10], resulting in a compact representation that can be used for accurate and efficient reduced-order models and control [11,12].

The majority of modal decompositions are linear [3,10], although emerging techniques in machine learning are providing improved nonlinear pattern extraction [1]. Linear regression and least-squares optimization are particularly widely used, as in the ubiquitous proper orthogonal decomposition (POD) [also known as principal component analysis (PCA)] [5–7,9,13] and the emerging dynamic mode decomposition (DMD) [8,14–16]. POD provides a principled approach to decomposing high-dimensional fluid flow data into a hierarchy of orthogonal modes that are ordered in terms of their ability to capture the energy in the flow; because these modes are orthogonal, it is possible to obtain reduced-order models by Galerkin projection of the Navier-Stokes equations onto a truncated POD basis [17–19]. DMD is a related technique to decompose a flow into spatiotemporal coherent structures that are each constrained to have coherent and linear dynamics in time. The least-squares regression underlying these approaches is highly susceptible to outliers and corrupted data [20,21]. Outliers and corrupt entries differ significantly from the distribution of other measurements [22] and are common in experimental measurement techniques such as particle image velocimetry (PIV). Thus, modal decomposition techniques such as POD/PCA and DMD are *fragile* with respect to outliers. Further, even though POD is robust to Gaussian white noise [23], DMD is sensitive to noisy data [24–26]. Even techniques that are robust to corrupt velocity fields, such as finite-time Lyapunov exponents and Lagrangian coherent structures [27–32], are unable to process velocity fields with missing regions, which is common in experimentally acquired data, so that interpolation must be used to fill in missing vectors. In this work, we explore the use of the robust principal component analysis (RPCA) [21] to process corrupt flow fields, leveraging global correlations in the data. We emphasize the impact of this approach on modal analysis, including POD/PCA and DMD.

A. Experimental challenges

Experimental techniques to measure fluid flows have evolved rapidly over the past century, with the ultimate goal of acquiring full flow fields with high spatial and temporal resolution. Laser-based imaging techniques have evolved from point measurements [33,34] to two-dimensional (2D) and 3D field measurements [35–40]. PIV has since become a cornerstone of experimental fluid mechanics, providing nonintrusive velocity field measurements across a range of applications. Improvements in PIV hardware, including more powerful lasers, higher-resolution and frame-rate cameras, advanced image processing technology, and the development of tomographic PIV are providing unprecedented views into real flows. Despite the undeniable success of PIV, there are several well-known challenges to acquiring clean and accurate data. Multiple factors in the PIV data acquisition and processing pipeline can contribute to velocity vector outliers that degrade the resulting velocity fields. These include inadequate illumination and irregularities in the light field, background speckle, seeding density and nonpassivity of the particles, sharp gradients in flow properties, optical issues, such as alignment and aberration, limited resolution and shot noise in the image recording, and out of plane motion of the particles when measuring in 2D [37,41]. Because of a fundamental trade-off between the quantity and quality of PIV data, in both space and time, researchers continue to push the resolution limits of current systems. Thus, flow fields acquired with PIV often have missing and/or corrupt measurements. This has motivated processing techniques to improve PIV data quality of PIV data [42–44], including predictor-corrector schemes [45], spatial filtering to remove frequencies not possible for the measurement resolution [46], and POD-based background removal [47]. The identification of spurious vectors has been studied extensively [48,49], and the normalized median filter is a robust and well-used method [50]. However, missing

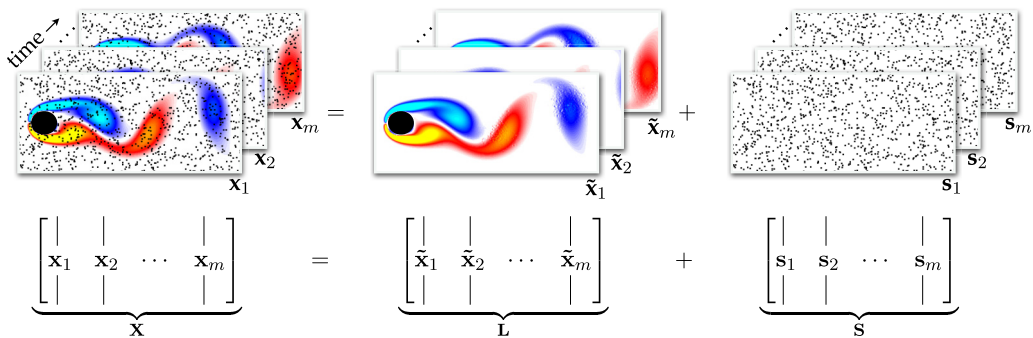


FIG. 1. Schematic of RPCA filtering applied to corrupted flow-field data. Corrupted snapshots are arranged as column vectors in the matrix \mathbf{X} , which is decomposed into the sum of a low-rank matrix \mathbf{L} and a sparse matrix of outliers \mathbf{S} .

vectors often cluster in regions of high shear, presenting a challenge for standard vector validation and interpolation methods that rely on local flow information [50]. In this work, we leverage robust statistics and global spatiotemporal coherent structures across the entire data set to fill in missing measurements and improve modal decomposition of fluid flow fields.

B. Contributions of this work

We investigate the use of RPCA [21], a robust variant of POD/PCA [20,51]. RPCA uses a sparsity-promoting optimization to decompose a data matrix into the sum of a low-rank matrix containing coherent structures and a sparse matrix containing outliers. RPCA was originally popularized in the Netflix matrix completion algorithm for its recommender system [52] and has since been widely used for image and video processing [53], electrical capacitance tomography [54], and voice separation [55]. Here, we use RPCA filtering to process flow-field data from several simulations and experiments. Figure 1 demonstrates the ability of RPCA to uncover and isolate the dominant low-rank coherent structures from sparse outliers in flow data from an idealized example. In addition to directly analyzing and processing flow-field data, we also perform PCA and DMD modal analyses on the data before and after RPCA filtering to assess its performance.

Here we consider a range of simulated and experimentally acquired flow fields of varying complexity to isolate and analyze various aspects of the algorithm applied to data from fluid mechanics. First, we investigate high-fidelity flow fields from direct numerical simulations of a laminar flow past a cylinder and a turbulent channel flow, where it is possible to artificially add corrupt velocity field vectors to compare the RPCA filtered fields with a known ground truth. Next, we apply the method to two experimentally acquired data sets, including a companion laminar flow past a cylinder and measurements of a cross-flow turbine wake. Although there is not a ground-truth model for these flows, it is possible to estimate the effect of RPCA filtering in reducing outliers and corruption by analyzing the DMD spectrum, which has well-stereotyped behavior for such periodic wake flows [24]. In all cases, we show that the RPCA filtered fields yield DMD spectra that are more consistent with a periodic wake in the absence of noise.

This work is organized as follows: First, we present the standard POD/PCA and DMD modal analysis techniques in Sec. II, followed by the RPCA method in Sec. III. Section IV describes the four flow fields used in this analysis. Results of RPCA filtering on these flow fields and its impact on downstream modal analysis are presented in Sec. V.

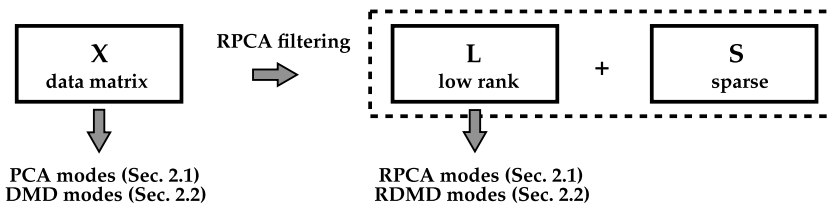


FIG. 2. Flowchart showing how we apply RPCA filtering to a data matrix and analyze the results. Depending on the data set in question, the data matrix may be artificially corrupted prior to RPCA filtering (Sec. IV). The results of principal component analysis and dynamic mode decomposition performed on the data matrix (\mathbf{X}) are referred to as PCA and DMD modes, respectively, whereas those same operations performed on the low-rank matrix (\mathbf{L}) are referred to as RPCA and RDMD modes.

II. MODAL ANALYSIS

Extracting coherent structures from high-dimensional data has been a central challenge in fluid mechanics for decades. Here we review two leading modal decomposition techniques for data from fluid mechanics, the POD, also known as PCA (Sec. II A), and the DMD (Sec. II B). Both methods apply equally well to data from simulations or experiments. We use these two modal decomposition techniques to assess the effectiveness of RPCA filtering, in processing and correcting corrupt flow fields. Both techniques are based on the singular value decomposition (SVD) [56–59] and there are several detailed discussions of these decompositions [3,8,10,20]. The RPCA algorithm is explained in Sec. III.

In this work, we follow the flowchart shown in Fig. 2. RPCA filtering is applied to a data matrix (\mathbf{X}) which is bisected into the low-rank structure (\mathbf{L}) and sparse (\mathbf{S}) subspaces. From there, RPCA modes and DMD modes are calculated from the low-rank data. We also calculate the POD/PCA and DMD modes on the data matrix.

A. Proper orthogonal decomposition

Proper orthogonal decomposition—referred to as PCA throughout results—is a widely used method to identify spatially correlated coherent structures from data, decomposing the flow field into a linear combination of orthogonal modes that are arranged hierarchically by energy content. There are several variants of POD [3,6,7,9,13,60], and we will present a variant of the *snapshot POD* of Sirovich that relies on the numerically stable SVD [20,60]. First, flow-field data (e.g., a velocity or vorticity field) is measured or computed on a discrete spatial grid, and m snapshots of these flow fields are collected at various times t_1, t_2, \dots, t_m . The flow-field data at time t_k may be reshaped into a column vector $\mathbf{x}_k = \mathbf{x}(t_k) \in \mathbb{R}^n$, where n denotes the number of flow variables times the number of spatial grid locations. Next, a data matrix is formed by arranging the column vectors \mathbf{x}_k in a matrix \mathbf{X} :

$$\mathbf{X} = \begin{bmatrix} | & | & \cdots & | \\ \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_m \\ | & | & \cdots & | \end{bmatrix}. \quad (1)$$

Finally, POD modes are obtained by computing the singular value decomposition of $\mathbf{X} \in \mathbb{R}^{n \times m}$:

$$\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T, \quad (2)$$

where superscript T defines the matrix transpose, $\mathbf{U} \in \mathbb{R}^{n \times n}$, $\mathbf{\Sigma} \in \mathbb{R}^{n \times m}$, and $\mathbf{V} \in \mathbb{R}^{m \times m}$. The columns of \mathbf{U} are *POD modes* with the same dimension as a flow field \mathbf{x} . POD modes are orthonormal so that $\mathbf{U}^T\mathbf{U} = \mathbf{I}$; similarly, $\mathbf{V}^T\mathbf{V} = \mathbf{I}$. Moreover, the columns of \mathbf{U} (respectively, rows of \mathbf{V}^T) are arranged in order of their importance in describing the data. The importance of each

mode (i.e., column of \mathbf{U}) is given by the corresponding entry of the non-negative, diagonal matrix of singular values $\mathbf{\Sigma} \in \mathbb{R}^{n \times m}$.

The matrix \mathbf{X} will exhibit *low-rank structure*, so that it is well approximated by the first $r \ll m < n$ columns of \mathbf{U} and \mathbf{V} :

$$\mathbf{X} \approx \mathbf{U}_r \mathbf{\Sigma}_r \mathbf{V}_r^T, \quad (3)$$

where \mathbf{U}_r and \mathbf{V}_r denote the first r columns of each matrix and $\mathbf{\Sigma}_r$ denotes the first $r \times r$ sub-block of $\mathbf{\Sigma}$. In fact, the Eckart-Young theorem states that this is the *optimal* rank- r approximation of the matrix \mathbf{X} in a least-squares sense. More details about the SVD can be found in Ref. [20].

After truncating all but the first r dominant modes, a flow-field snapshot \mathbf{x} may be approximated by a linear combination of these modes:

$$\mathbf{x} \approx \sum_{k=1}^r \mathbf{u}_k \alpha_k,$$

where α_k is the *POD mode coefficient*.

Because the POD modal basis is orthogonal, it is possible to obtain a reduced-order nonlinear dynamical system for the evolution of the coefficients $\alpha_k(t)$ in time via Galerkin projection of the Navier-Stokes equations onto the POD basis. In this way, the POD basis may be thought of a data-driven generalization of the Fourier basis that is tailored to a particular flow field. POD is also closely related to PCA [51], the Karhunen-Loève decomposition [61], empirical orthogonal functions [62], or the Hotelling transform [63,64].

B. Dynamic mode decomposition

DMD is a modal decomposition technique that simultaneously identifies spatially coherent modes that are constrained to have the same linear behavior in time, given by oscillations at a fixed frequency with growth or decay [8,15]. Thus, the dynamic mode decomposition provides a dimensionality reduction into a set of spatial modes along with a linear model for how these modes evolve in time. This is in contrast to POD, which results in orthogonal modes arranged in terms of energy content and without consideration of dynamics. However, in many formulations, DMD is closely related to POD, and may be thought of as a linear combination of POD modes that results in linear evolution in time. DMD also has deep connections to nonlinear dynamical systems via Koopman operator theory [8,14,16,65,66].

In the original formulation of DMD [15], the snapshots in the data matrix in Eq. (1) are spaced evenly in time, so that $t_k = k\Delta t$ with Δt sufficiently small to resolve the highest frequencies in the dynamics. Generalizations have since been formulated to allow for nonsequential time series [16,67] and for data that is under-resolved in space [68,69] or time [70]; however, for simplicity, we will present the standard *exact DMD* formulation of Tu *et al.* [16] with evenly spaced and sequential snapshots. DMD seeks to identify the leading eigenvalues and eigenvectors of the best-fit linear operator \mathbf{A} that evolves snapshots forward in time:

$$\mathbf{x}_{k+1} \approx \mathbf{A}\mathbf{x}_k. \quad (4)$$

The eigenvectors ϕ of \mathbf{A} have the dimensions of a flow field and correspond to spatiotemporal coherent structures whose dynamics in time evolve according to the associated eigenvalue γ .

In practice, this operator is identified by first splitting the data in Eq. (1) into two matrices:

$$\mathbf{X} = \begin{bmatrix} | & | & \cdots & | \\ \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_{m-1} \\ | & | & & | \end{bmatrix} \quad \mathbf{X}' = \begin{bmatrix} | & | & \cdots & | \\ \mathbf{x}_2 & \mathbf{x}_3 & \cdots & \mathbf{x}_m \\ | & | & & | \end{bmatrix}, \quad (5a)$$

and then solving for the best-fit operator that satisfies

$$\mathbf{X}' \approx \mathbf{A}\mathbf{X} \quad (6)$$

via the following *least-squares* optimization problem:

$$\mathbf{A} = \underset{\mathbf{A}}{\operatorname{argmin}} \|\mathbf{X}' - \mathbf{A}\mathbf{X}\|_F = \mathbf{X}'\mathbf{X}^\dagger \approx \mathbf{X}'\mathbf{V}_r\boldsymbol{\Sigma}_r^{-1}\mathbf{U}_r^T. \quad (7)$$

Here we are minimizing the Frobenius norm $\|\cdot\|_F$ via the pseudoinverse $\mathbf{X}^\dagger = \mathbf{V}\boldsymbol{\Sigma}^{-1}\mathbf{U}^T \approx \mathbf{V}_r\boldsymbol{\Sigma}_r^{-1}\mathbf{U}_r^T$.

In practice, the matrix \mathbf{A} is far too large to analyze directly, and instead, we project \mathbf{A} onto an r -dimensional POD subspace, given by the columns of \mathbf{U}_r :

$$\tilde{\mathbf{A}} = \mathbf{U}_r^T \mathbf{A} \mathbf{U}_r = \mathbf{U}_r^T \mathbf{X}' \mathbf{V}_r \boldsymbol{\Sigma}_r^{-1}. \quad (8)$$

The eigenvalues of $\tilde{\mathbf{A}}$ are the same as the eigenvalues of \mathbf{A} , which are known as the *DMD eigenvalues*. They are computed via the eigendecomposition of the $r \times r$ matrix $\tilde{\mathbf{A}}$:

$$\tilde{\mathbf{A}}\mathbf{W} = \mathbf{W}\boldsymbol{\Gamma}. \quad (9)$$

Finally, the corresponding *DMD modes* are reconstructed using the full-dimensional data along with the reduced eigenvectors in \mathbf{W} :

$$\boldsymbol{\Phi} = \mathbf{X}'\mathbf{V}_r\boldsymbol{\Sigma}_r^{-1}\mathbf{W}. \quad (10)$$

This formula for the eigenvectors is from the exact DMD algorithm [8,16]; the original formulation of Schmid [15] computes modes as $\boldsymbol{\Phi} = \mathbf{U}_r\mathbf{W}$.

With the DMD modes $\boldsymbol{\Phi}$ and eigenvalues $\boldsymbol{\Gamma}$ it is possible to reconstruct the state at time $k\Delta t$,

$$\mathbf{x}_k = \sum_{j=1}^r \boldsymbol{\phi}_j \gamma^{k-1} b_j = \boldsymbol{\Phi}\boldsymbol{\Gamma}^{k-1}\mathbf{b}, \quad (11)$$

where the vector \mathbf{b} of mode amplitudes is generally computed as

$$\mathbf{b} = \boldsymbol{\Phi}^\dagger \mathbf{x}_1. \quad (12)$$

More principled approaches to select the few dominant modes have been considered based on sparsity-promoting optimization [71].

The spectral expansion above may also be written in continuous time by introducing the continuous eigenvalues $\omega = \log(\gamma)/\Delta t$:

$$\mathbf{x}(t) = \sum_{j=1}^r \boldsymbol{\phi}_j e^{\omega_j t} b_j = \boldsymbol{\Phi} \exp(\boldsymbol{\Omega}t)\mathbf{b}, \quad (13)$$

where $\boldsymbol{\Omega}$ is a diagonal matrix containing the continuous-time eigenvalues ω_j .

DMD is known to be extremely sensitive to noisy data [24–26], and the eigenvalues specifically suffer from a bias that is not reduced with increasing data. There are several modifications to make DMD more robust to noise, including averaging forward-time and backward-time operators [25], total least squares [26], and variable projection [67]. For periodic wake data, as explored in three of the examples in this paper, the discrete-time eigenvalues should occur in complex conjugate pairs $\gamma, \bar{\gamma}$ exactly on the unit circle in the complex plane for clean data. Similarly, the continuous-time eigenvalues should be in complex conjugate pairs $\pm i\omega$ on the imaginary axis, indicating pure oscillations with no growth or decay [24,72]. In Ref. [24], Bagheri characterized the perturbative effect of noise on these eigenvalues, deriving an asymptotic expression for how high frequency eigenvalues become increasingly affected by noise. If the true continuous-time eigenvalue should be $\pm i\omega$ in the absence of noise, Bagheri showed that in the presence of perturbative white noise with magnitude $\epsilon \ll 1$ the observed eigenvalue pair is

$$\pm i\omega - \epsilon C\omega^2 + \mathcal{O}(\epsilon^2), \quad (14)$$

where C is a sensitivity constant. Thus, low noise levels cause a spurious real-valued damping $-\epsilon C \omega^2$ that is quadratic in the frequency. We will make extensive use of this property to assess the quality of our RPCA filtered fields by computing the ratio of the best-fit factor ϵC before and after applying RPCA filtering. ϵC should decrease as a consequence of reduced noise and corruption.

III. ROBUST EXTRACTION OF FLUID COHERENT STRUCTURES

Techniques based on least-squares regression, such as POD/PCA and DMD, are highly susceptible to outliers and corrupted data, making them *fragile* with respect to some experimental measurement errors. Outliers and corruption are defined as data points that differ significantly from the statistical distribution of the majority of the data set [22], so that they cannot be considered as the original data plus a small-to-moderate amount of white noise. To mitigate this sensitivity, Candès *et al.* [21] have developed RPCA that seeks to decompose a data matrix \mathbf{X} into a structured low-rank matrix \mathbf{L} that is characterized by dominant coherent structures and a sparse matrix \mathbf{S} containing outliers and corrupt data:

$$\mathbf{X} = \mathbf{L} + \mathbf{S}. \quad (15)$$

The principal components of \mathbf{L} are *robust* to outliers and corrupt data, which are isolated in \mathbf{S} . This decomposition, also referred to as a filter, has profound implications for many modern problems of interest, including video surveillance (where the background objects appear in \mathbf{L} and foreground objects appear in \mathbf{S}), facial recognition (eigenfaces are in \mathbf{L} and shadows, occlusions, etc., are in \mathbf{S}), natural language processing and latent semantic indexing, and ranking problems.¹ Standard PCA/POD is effective at removing white noise that is smaller than the relevant singular values in the data [23]; however, it is not able to remove outliers. Instead, RPCA is used to correct outliers that differ significantly from the distribution of the other observations.

Mathematically, the goal is to find \mathbf{L} and \mathbf{S} that satisfy the following:

$$\min_{\mathbf{L}, \mathbf{S}} \text{rank}(\mathbf{L}) + \|\mathbf{S}\|_0 \quad \text{subject to } \mathbf{L} + \mathbf{S} = \mathbf{X}. \quad (16)$$

$\|\mathbf{S}\|_0$ counts the number of nonzero elements in \mathbf{S} , quantifying how *sparse* it is. $\text{rank}(\mathbf{L})$ is the number of nonzero singular values in \mathbf{L} , quantifying how many linearly independent rows and columns describe the data. However, neither the $\text{rank}(\mathbf{L})$ nor the $\|\mathbf{S}\|_0$ terms are convex, making this optimization intractable. Similarly to the compressed sensing problem, it is possible to solve for \mathbf{L} and \mathbf{S} with *high probability* using a convex relaxation of (16):

$$\min_{\mathbf{L}, \mathbf{S}} \|\mathbf{L}\|_* + \lambda_0 \|\mathbf{S}\|_1 \quad \text{subject to } \mathbf{L} + \mathbf{S} = \mathbf{X}, \quad (17)$$

where $\|\cdot\|_*$ is the nuclear norm, given by the sum of singular values which is a proxy for the rank of the matrix, and $\|\cdot\|_1$ is the 1-norm of the matrix viewed as a vector, given by the sum of the magnitudes of each entry in the matrix, which is a proxy for the $\|\cdot\|_0$ norm of a matrix; the hyperparameter λ_0 is given by $\lambda_0 = \lambda / \sqrt{\max(n, m)}$. The solution to (17) converges to the solution of (16) with high probability if $\lambda = 1$, where n and m are the dimensions of \mathbf{X} , given that \mathbf{L} is not sparse and \mathbf{S} is not low rank. In the examples below, these assumptions may only be partially valid, so the optimal value of λ may vary slightly. The convex problem in (17) is known as *principal component pursuit* and may be solved using the augmented Lagrange multiplier (ALM) algorithm.

¹The ranking problem may be thought of in terms of the Netflix prize for matrix completion. In the Netflix prize, a large matrix of preferences is constructed, with rows corresponding to users and columns corresponding to movies. This matrix is sparse, as most users only rate a handful of movies. The Netflix prize seeks to accurately fill in the missing entries of the matrix, revealing the likely user rating for movies the user has not seen.

Specifically, an augmented Lagrangian may be constructed:

$$\mathcal{L}(\mathbf{L}, \mathbf{S}, \mathbf{Y}) = \|\mathbf{L}\|_* + \lambda_0 \|\mathbf{S}\|_1 + \langle \mathbf{Y}, \mathbf{X} - \mathbf{L} - \mathbf{S} \rangle + \frac{\nu}{2} \|\mathbf{X} - \mathbf{L} - \mathbf{S}\|_F^2, \quad (18)$$

where \mathbf{Y} is the matrix of Lagrange multipliers and ν is a hyperparameter. We then solve for \mathbf{L}_k and \mathbf{S}_k to minimize \mathcal{L} , update the Lagrange multipliers

$$\mathbf{Y}_{k+1} = \mathbf{Y}_k + \nu(\mathbf{X} - \mathbf{L}_k - \mathbf{S}_k),$$

and iterate until convergence. In this work, an inexact ALM implementation from Ref. [73] is used. The alternating directions method (ADM) [74,75] provides another simple procedure.

After the low-rank matrix \mathbf{L} is obtained, it is possible to compute robust POD/PCA modes (Sec. II A) as in Eq. (2):

$$\mathbf{L} = \mathbf{U}\Sigma\mathbf{V}^T. \quad (19)$$

Henceforth, we refer to the modes in \mathbf{U} from \mathbf{L} as RPCA modes. We note that in many flow applications it is important to subtract the mean flow before computing POD, which allows the POD eigenvalues to be interpreted as the variance of fluctuations and the expansion to respect boundary conditions by construction. However, before computing RPCA, it may be difficult to obtain an accurate mean flow estimate. Instead, we advocate computing the RPCA, then subtracting the mean of \mathbf{L} from itself, and finally computing POD on the mean-subtracted low-rank matrix; this final POD step will remove small amounts of white noise.

Similarly, it is also possible to compute robust DMD (RDMD) modes and eigenvalues. We note that the RDMD in this work should not be confused with the recursive DMD of Noack *et al.* [76], which uses the same acronym. We also note that this decomposition is similar in spirit to the coherent vortex simulation approach [77], which separates turbulent flows into coherent and random parts based on a wavelet decomposition. However, RPCA does not perform this decomposition using a universal basis, such as wavelets, that rely on scale separation but rather based on statistical correlations in the data.

IV. MODEL FLOWS

We demonstrate RPCA filtering on several example data sets of varying complexity, drawn from direct numerical simulations (DNS) and PIV data from experiments. Figure 3 provides an overview of the four example flow fields.

A. Cylinder flow

Flow past a cylinder is a canonical example in fluid mechanics. We consider data from DNS at a diameter-based Reynolds number of 100 and from PIV measurements at Reynolds number 413 [78].

The DNS data are generated by simulating the two-dimensional incompressible Navier-Stokes equations using the immersed boundary projection method [79,80]. The computational domain comprises four nested grids: The finest grid covers a domain of 9×4 and the largest grid covers a domain of 72×32 , where lengths are nondimensionalized by the cylinder diameter. Each grid contains 449×199 points with a resolution of 50 points per cylinder diameter. The time step is $\Delta t = 0.02$ and data are sampled at intervals of $10\Delta t$ (30 times the vortex shedding frequency) with $m = 150$ snapshots saved, covering 5 vortex shedding cycles. The DNS provides a benchmark, where the uncorrupted flow field is known, to quantitatively assess performance of RPCA filtering on data with artificial salt-and-pepper corruption. Corrupted sample points are chosen uniformly in space and time at a given rate, and both the u and v velocity components at each selected location are randomly assigned a value of ± 10 times the standard deviation of the streamwise velocity data. In addition, we consider a second case where corrupted sample points are chosen with a bias toward regions of high vorticity or shear magnitude, which is more physically realistic for PIV data. In this

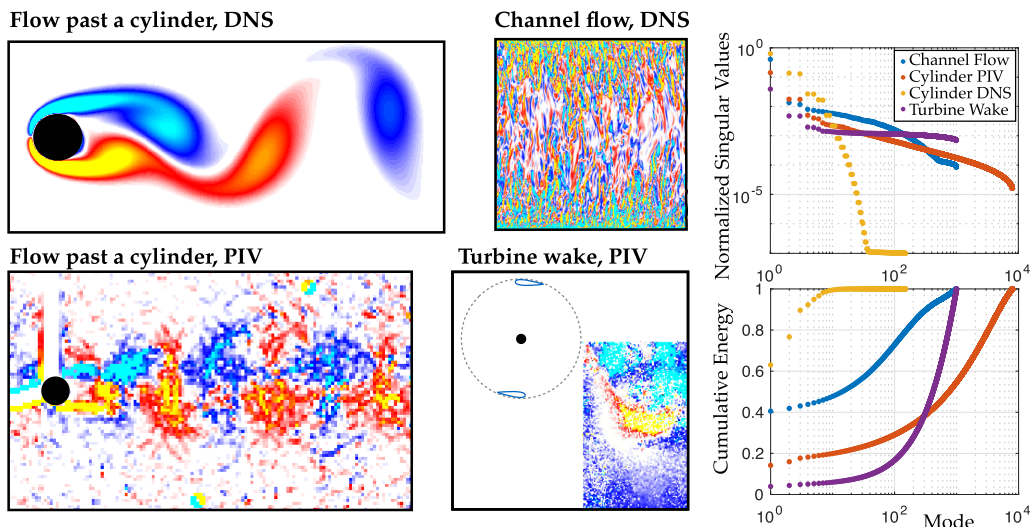


FIG. 3. Left: Example flow-field data. Right: Singular value spectrum for each data set. Mean flow travels from left to right in all cases.

second case, we select measurements for corruption based on a probability density given by $\alpha + |\omega|$, where α is a small positive constant and $|\omega|$ is the absolute value of the vorticity; when the rate of corruption is sufficiently high, these corrupted fields begin to resemble the uniformly corrupted cases, but with more corruption in vortex cores. Because vorticity is calculated from velocity fields using a finite-difference derivative, there is a higher rate of corruption in the vorticity fields than in the velocity fields.

The PIV data has frame size of 135×80 grid points with a resolution of 8 points per cylinder diameter. Data are sampled at a rate of 20 Hz (125 times the shedding frequency) with $m = 8000$ snapshots saved, which corresponds to 64 vortex shedding cycles.

B. Turbulent channel flow

For a more complex and multiscale flow, we consider DNS data from a forced, fully developed turbulent channel flow data with a friction velocity Reynolds number of $Re_\tau = 1000$, from the Johns Hopkins Turbulence Database [81]. This example provides a test case to see how turbulent kinetic energy at various scales is filtered depending on the level of added noise. The addition of noise is similar to the cylinder DNS where randomly selected sample points of the streamwise and cross-stream velocity fields are assigned a value of ± 10 standard deviations of the streamwise velocity data. Due to the size of the full data set, we only consider two-dimensional fields on the midplane, with a 512×512 grid of three component velocity measurements spanning the channel width. Data are sampled at a rate of 966 times the mean flow-through time with $m = 1000$ snapshots.

C. Cross-flow turbine wake

Finally, we consider PIV wake data from a cross-flow turbine experiment conducted at the University of Washington. Cross-flow turbines can be used to extract power from wind and water currents for renewable energy generation. This flow exhibits both coherent and broadband phenomena and provides a challenging test-case RPCA filtering. The frame consists of 158×98 grid points, at a resolution of 99 points per rotor diameter. Data are sampled at a rate of 32 times the blade-pass frequency with $m = 1000$ snapshots. Vectors were calculated using a multigrid, multipass algorithm with adaptive image deformation [82]. Resulting vector fields were then

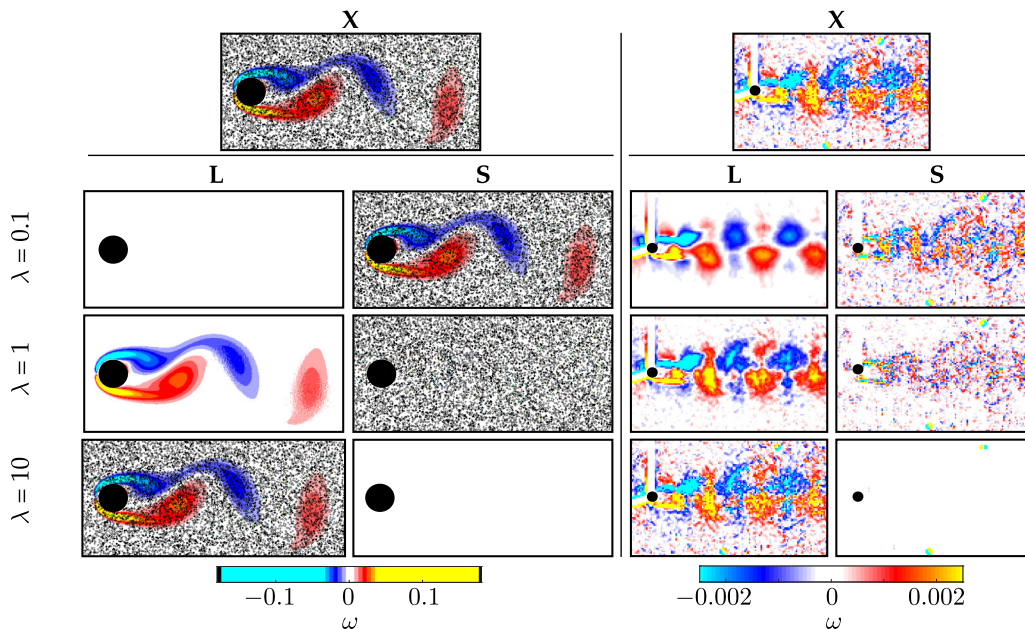


FIG. 4. RPCA filtering removes noise and outliers in the flow past a cylinder (black circle), from DNS (left) with 10% of velocity field measurements corrupted with salt-and-pepper noise, and PIV measurements (right). All frames show resultant vorticity fields. As the parameter λ is decreased, RPCA filtering is more aggressive, eventually incorrectly identifying coherent flow structures as *outliers*.

validated using a normalized median filter with potential replacement by secondary correlation peaks. The cross-correlation and validation steps result in missing data, particularly in regions of high vorticity and shear. To apply RPCA filtering, these missing values are randomly assigned a value of ± 10 standard deviations of the streamwise flow data, in contrast to the experimental cylinder wake where missing measurements were previously interpolated.

V. RESULTS

We now explore the ability of RPCA filtering to isolate and remove noise and corruption from the example flow fields. We will begin with the simulated and experimental flow past a cylinder, followed by data from the Johns Hopkins turbulent channel flow simulation, and ending with the experimental wake of a cross-flow turbine.

A. Cylinder flow

Figure 4 shows the results of RPCA filtering for flow past a cylinder, providing a side-by-side comparison of PIV and corrupted DNS data. Although the Reynolds numbers differ by a factor of four, the flow fields are qualitatively similar, characterized by periodic, laminar vortex shedding. For $\lambda = 1$, the data are correctly segmented with the coherent flow in **L** and the sparse corruption in **S**. When λ is too small, RPCA filtering is overly aggressive, incorrectly including relevant flow structures in **S**, and when λ is too large, the corruption is not filtered.

For the experimental data in Fig. 4 (right), the optimal value of λ is less clear. For $\lambda = 0.1$, the low-rank field **L** is visually smoother than the field at $\lambda = 1$, but the sparse matrix **S** contains a significant portion of the wake structures, indicating overfiltering. This filtering becomes more pronounced in the movies, where it is clear that much of the high-frequency “noise” in the bypass flow is actually free-stream turbulence, which is consistent with the turbulence intensity of the

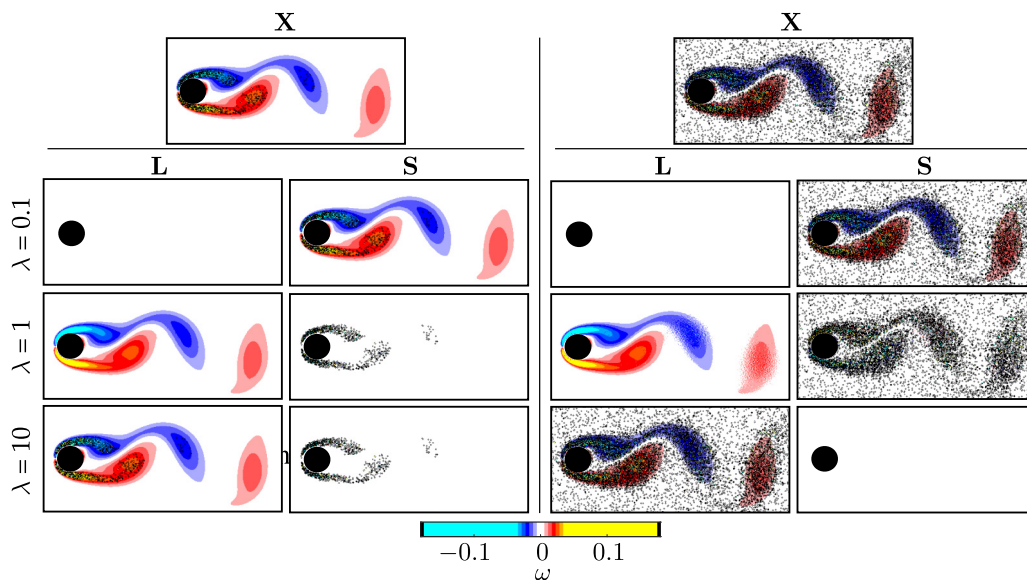


FIG. 5. RPCA filtering removes vorticity-biased corruption from simulated flow past a cylinder at Reynolds number 100. Unlike results shown in Fig. 4, corrupt entries are concentrated in regions of high vorticity instead of being uniformly distributed. In the flow on the left, $\eta = 1\%$ of the velocity field measurements are corrupted and on the right $\eta = 10\%$ of the velocity field measurements are corrupted. All frames show resultant vorticity fields.

experiments. Further, as subsequently discussed, when we compute the RPCA modes, it is clear that the $\lambda = 0.1$ case is heavily filtering out all but the first three modes. Thus, it appears that the theoretically optimal value $\lambda = 1$ has the best performance, although there may be a trade-off between filtering ambient free-stream turbulence and coherent structures of interest in experiments.

Figure 5 shows the results of RPCA filtering on the simulated data for the second case of vorticity-biased corruption. Again, in all cases, the theoretically optimal value of $\lambda = 1$ yields the best segmentation of the corruption into the matrix \mathbf{S} . When the rate of corruption is increased from 1% (left) to 10% (right), the free-stream flow begins to become corrupted, resembling the uniform corruption case in Fig. 4. The mean error and relative nuclear norm of the low-rank matrix (\mathbf{L}) compared to the true, uncorrupted data (\mathbf{X}) are shown in Fig. 6 for varying percentages of corrupt entries. Statistically, results are similar for vorticity-biased and randomly-distributed corruption. In both cases, RPCA filtering is remarkably robust to corruption, even for corruption in excess of

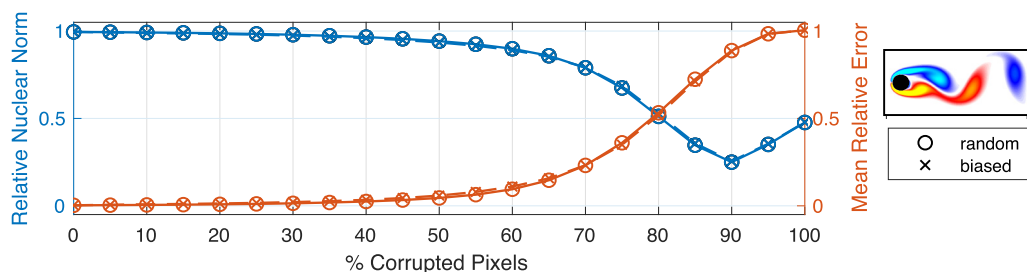


FIG. 6. Error ($\|\mathbf{X}_{\text{uncorrupted}} - \mathbf{L}\|_F / \|\mathbf{X}_{\text{uncorrupted}}\|_F$) and relative nuclear norm $[\|\mathbf{L}\|_* / \|\mathbf{X}\|_*] = \text{sum}(\sigma_L) / \text{sum}(\sigma_{X, \text{uncorrupted}})$ of the low-rank matrix \mathbf{L} compared with the uncorrupted data \mathbf{X} for varying percentages of corruption.

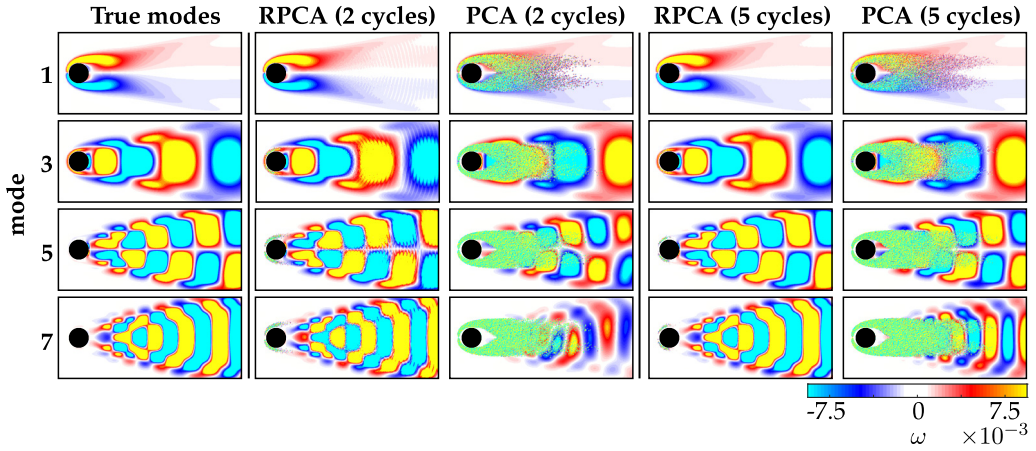


FIG. 7. Odd PCA vorticity modes of the cylinder simulations from Fig. 5 with 1% of velocity measurements corrupted with a bias toward regions of high vorticity.

50% of the measurements. This laminar vortex shedding example is an ideal application for RPCA filtering, as the true flow field is low rank and the corruption is sparse; it is unlikely that this will hold as well for data exhibiting broadband turbulence.

1. PCA analysis for cylinder wake flows

We now investigate the impact of RPCA filtering on modal decompositions. Figures 7 and 8 show four leading PCA and RPCA modes for 1% and 10% vorticity-biased corruption, respectively. The first mode corresponds to the mean flow, and the remaining modes come in energetic pairs where the corresponding coefficients α_{2j} and α_{2j+1} oscillate sinusoidally at the same frequency but $\pi/2$ out of phase, sweeping out a circle in the phase plane. Thus, we only show one mode, \mathbf{u}_{2j+1} from each of the first three energetic mode pairs. In all cases, the RPCA modes show dramatic improvement, while significant artifacts remain in the PCA modes. We also investigate the effect of increasing the amount of data, and there is a clear improvement in RPCA modes from two to five vortex shedding cycles; in contrast, the PCA modes do not improve appreciably with more data.

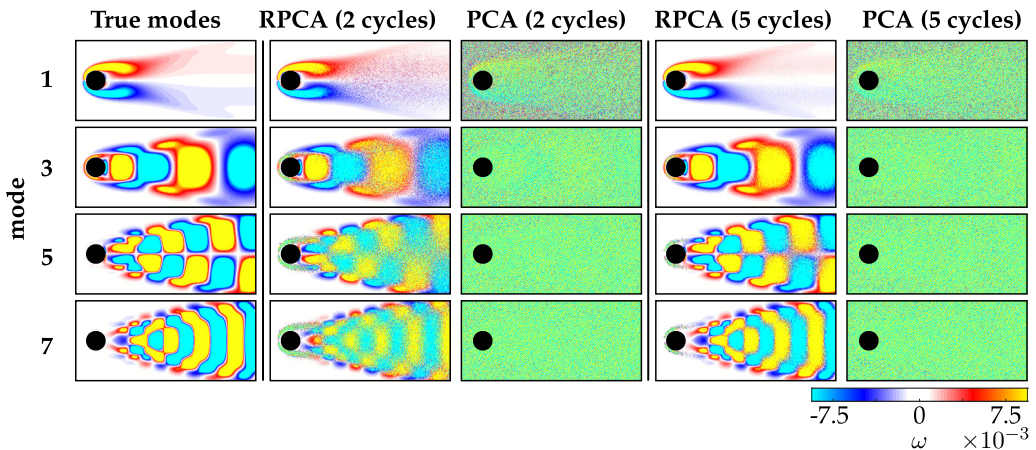


FIG. 8. Odd PCA vorticity modes of the cylinder simulations from Fig. 5 with 10% of velocity measurements corrupted with a bias toward regions of high vorticity.

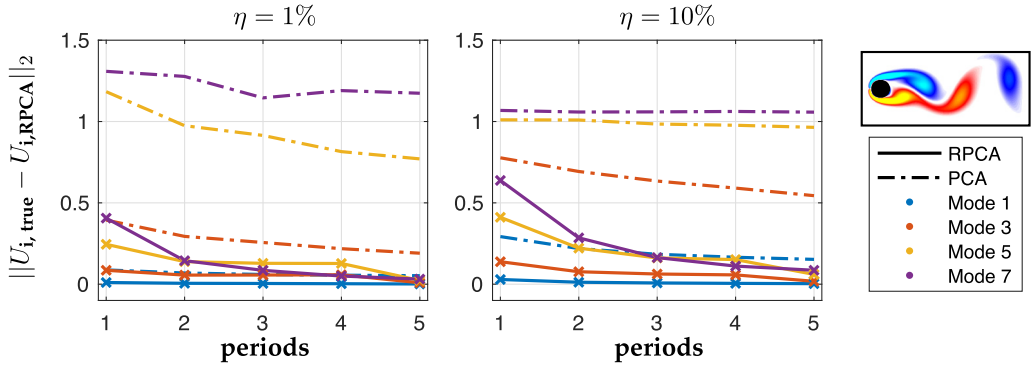


FIG. 9. L_2 error between the true PCA modes of the clean cylinder simulation data \mathbf{X} and the RPCA and PCA modes for corrupted data with 1% (left) and 10% (right) vorticity-biased corruption.

To quantify the improvement observed above, we compute the L_2 error between the PCA and RPCA modes of corrupted data and the PCA modes for the clean data (i.e., DNS results) as a function of the number of shedding periods. As shown in Fig. 9, the RPCA mode velocity fields quickly converge to a small error as the amount of data is increased for both the 1% and 10% corruption cases, while the PCA modes converge much more slowly and still have considerable error after five shedding periods are included in the analysis.

The modes for the PIV data for the cylinder flow are shown in Fig. 10. This figure highlights the effect of λ , the sparsity hyperparameter, which was previously discussed with respect to Fig. 4. In this case, we do not have a clean ground-truth data set to compare against. Although the flow field in Fig. 4 appears to have less corruption for $\lambda = 0.1$, here we see that all RPCA modes after the first three modes are heavily filtered, as seen in the rapid drop off in the singular values after the third mode. The corresponding modes are highly corrupt, further supporting that $\lambda = 0.1$ is not a good choice. In contrast, the RPCA modes for the theoretically optimal $\lambda = 1$ case appear to have slightly less free-stream corruption than the PCA modes. Also, as expected, for a large enough value of λ , the RPCA filtering has little effect on the modes.

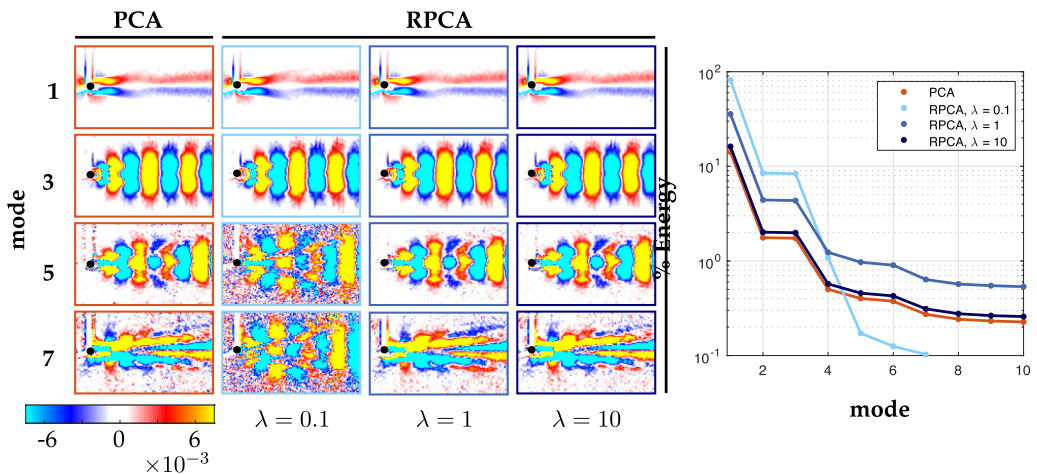


FIG. 10. Odd PCA vorticity modes of the experimental cylinder data for PCA and RPCA at $\lambda = 0.1, 1,$ and 10 , along with their singular values.

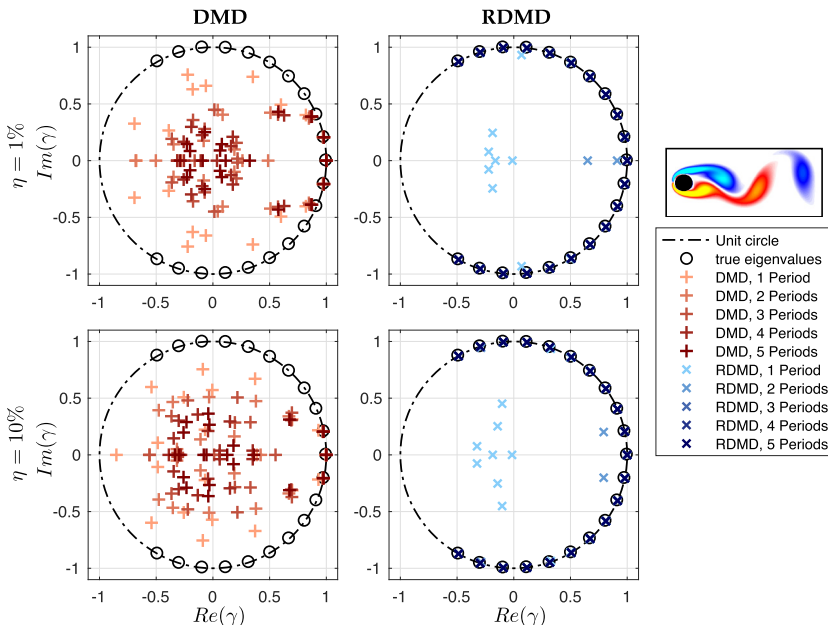


FIG. 11. Discrete-time DMD eigenvalues for the simulated cylinder data for small and large amounts of corruption and for increasing amounts of training data. In all cases, the RPCA-filtered DMD results dramatically outperform the standard DMD results.

2. DMD analysis for cylinder wake flows

DMD is known to be quite sensitive to noisy data, making this a challenging test case for RPCA filtering. Figure 11 shows the discrete-time eigenvalues for the cylinder DNS data with vorticity-biased corruption. For the cylinder wake, the uncorrupted or true DMD eigenvalues may be computed from the noiseless data, and they are equally spaced on the unit circle in the complex plane. In all cases, the RPCA-filtered DMD (RDMD) data result in dramatically better agreement with the uncorrupted or true DMD eigenvalues compared with the corrupted DMD eigenvalues. Even with only a single period of data and $\eta = 10\%$ corruption, the RDMD values capture the first six low-frequency mode pairs; in contrast, even with five periods of data and as little as $\eta = 1\%$ corruption, corrupted DMD only captures the first two low-frequency mode pairs, and with considerably more spurious damping. To see this more clearly, we plot the eigenvalues in continuous-time in Fig. 12, where the x axis is the imaginary eigenvalue component and the y axis is the real eigenvalue component, which is a standard way to plot DMD eigenvalues [15]. Here, the best-fit parabolas for the RDMD eigenvalues and the first seven corrupted DMD eigenvalues are shown in dashed lines. The curvature of these parabolas is directly related to the noise amplitude, as in Eq. (14) from Ref. [24]. The same continuous-time eigenvalue plot is shown for the PIV cylinder wake data in Fig. 13. In both cases, we see that the parabolic eigenvalue fit for the RDMD eigenvalues has a smaller curvature than for the corrupted DMD eigenvalues, indicating a quantitative and significant reduction in noise.

B. Turbulent channel flow

The trade-off between filtering corruption and small-scale structures is also apparent in the turbulent channel flow DNS. Unlike the cylinder wake, this flow field contains broadband turbulent phenomena across multiple spatial and temporal scales. Figure 14 shows RPCA filtering for various levels of corruption, sweeping across the tuning parameter λ . The corresponding turbulent kinetic

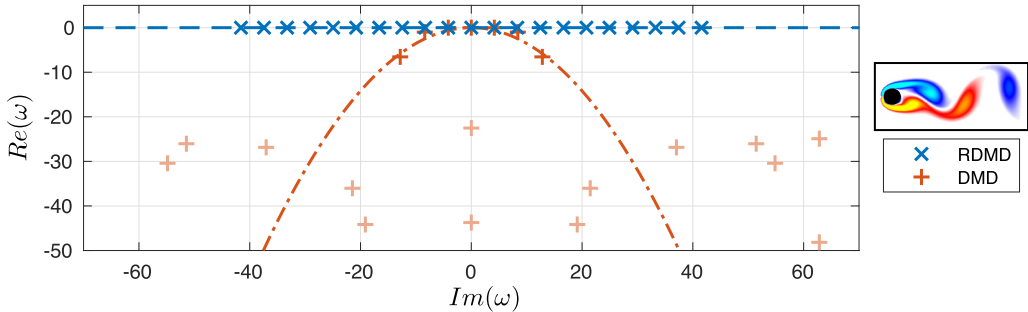


FIG. 12. Continuous-time DMD eigenvalues for the simulated cylinder data along with parabolic eigenvalue fits to estimate the error as in Ref. [24]. Here we use five vortex shedding periods with $\eta = 1\%$ corrupt values. The RDMD parabolic coefficient is approximately 2×10^4 times smaller than the DMD coefficient.

energy (TKE) is shown in Fig. 15, providing a summary of the various scales that are filtered. The value of λ that preserves the true TKE spectrum varies with the degree of velocity field corruption. In the uncorrupted case ($\eta = 0$), we can clearly see the effect of filtering on the turbulent coherent structures, indicating that some fine-scale structures are filtered for $\lambda = 1$. As the degree of corruption increases to $\eta = 2\%$, we see that the curves for $\lambda \leq 2$ remain relatively unchanged, although the $\lambda > 2$ curves begin to exhibit spurious high-frequency spatial structures (i.e., corruption is present in \mathbf{L}). As the rate of corruption increases to $\eta = 10\%$, spurious high-frequency energy also appears for $\lambda = 2$. In this case, it is clear that the optimal filtering value λ changes with the level of corruption. For relatively limited corruption, a larger value of λ may be used but must be decreased toward the theoretically optimal value of $\lambda = 1$ for higher levels of corruption. Finally, we note that, unlike the cylinder wake cases, it is not surprising that $\lambda = 1$ is suboptimal because the channel flow is not fundamentally low rank which deviates from an underlying assumption of the RPCA algorithm.

C. Cross-flow turbine wake

As a final example, we consider the use of RPCA filtering to identify outliers and fill in missing PIV data collected in the wake of a cross-flow turbine, as shown in Fig. 16. There are several stages in the PIV processing pipeline where RPCA filtering could be applied, including after initial cross-correlation, after conventional normalized median filter vector validation, and after

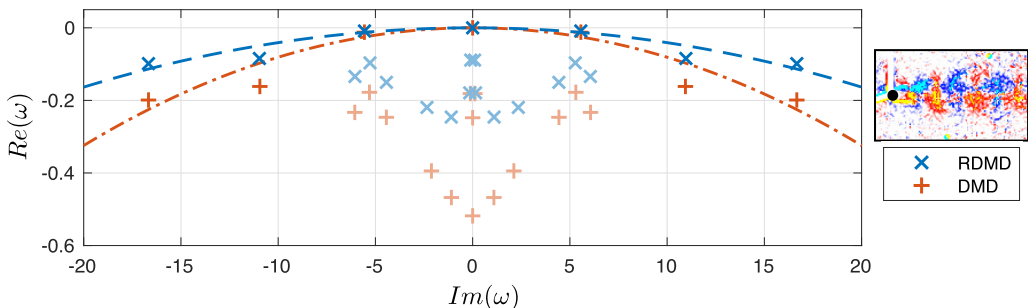


FIG. 13. Continuous-time DMD eigenvalues for the PIV cylinder data along with parabolic eigenvalue fits to estimate the error as in Ref. [24]. The RDMD parabolic coefficient is approximately two times smaller than the DMD coefficient.

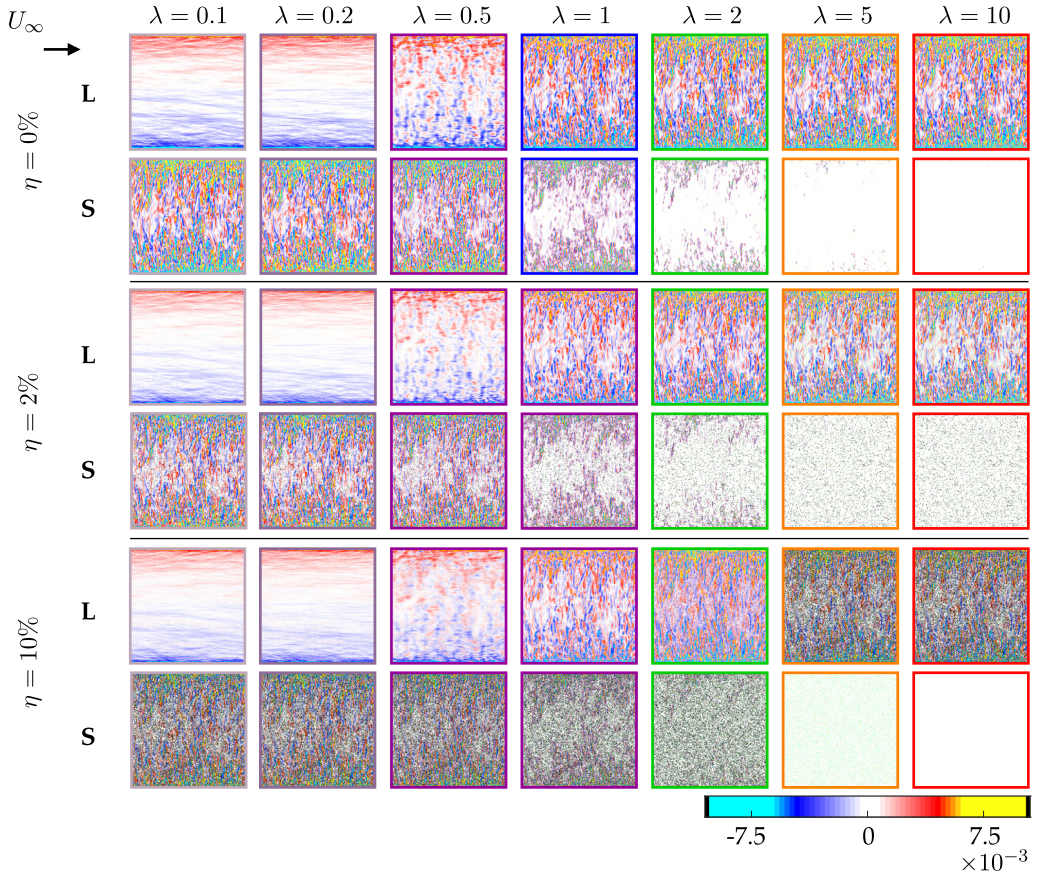


FIG. 14. RPCA filtering for turbulent channel flow vorticity fields with various levels of added noise and tuning parameter λ ; η represents the percentage of corrupted measurements in the velocity fields. The border colors match the color of the curve at the corresponding η in Fig. 15.

linear interpolation. For the cases shown here, we use $\lambda = 1.6$, which results in a velocity in the bypass flow, or lower third of the frame, that visually matches the frequency content of the unfiltered data. There are enough missing velocity vectors (23% and 20%, respectively) to degrade the effectiveness of both median filtering and interpolation. In contrast, RPCA filtering produces flow fields that capture dominant coherent structures for either cross-correlated or median-filtered fields. Finally, by investigating the standard deviation of all flow fields collected at a given turbine angular position (i.e., phase), it is clear that the RPCA filtering can be used to remove artifacts introduced by linear interpolation. This is consistent with the intuition that vector validation and interpolation should fail in these regions where there is high density of missing data that are spatially clustered.

The continuous-time DMD eigenvalues for the cross-flow turbine wake are shown in Fig. 17. In this plot, the parabolic fits for DMD eigenvalues computed after interpolation and RDMD eigenvalues computed after vector validation are displayed as dotted lines. For this case, the coefficient of the parabolic fit for the RDMD-based eigenvalues is six times smaller than the parabolic fit coefficient for the DMD-based eigenvalues. This demonstrates a significant quantitative improvement of the DMD spectrum using RPCA filtering to process the data.

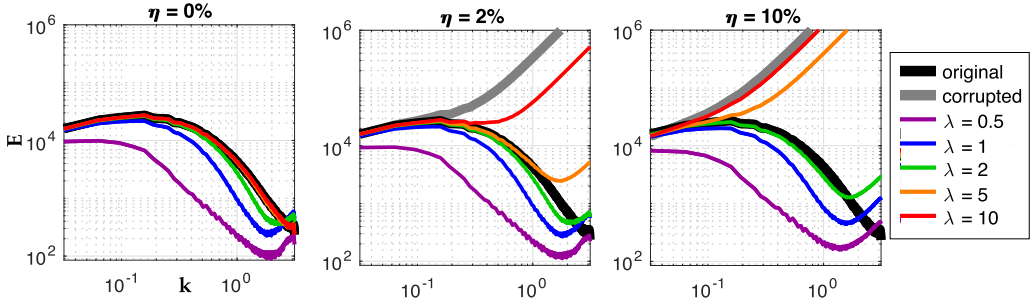


FIG. 15. TKE spectra for various levels of corruption and RPCA filtering. The TKE profiles provide a summary of the filtering that occurs at various scales. As corruption increases, the filtering remove more high-frequency information.

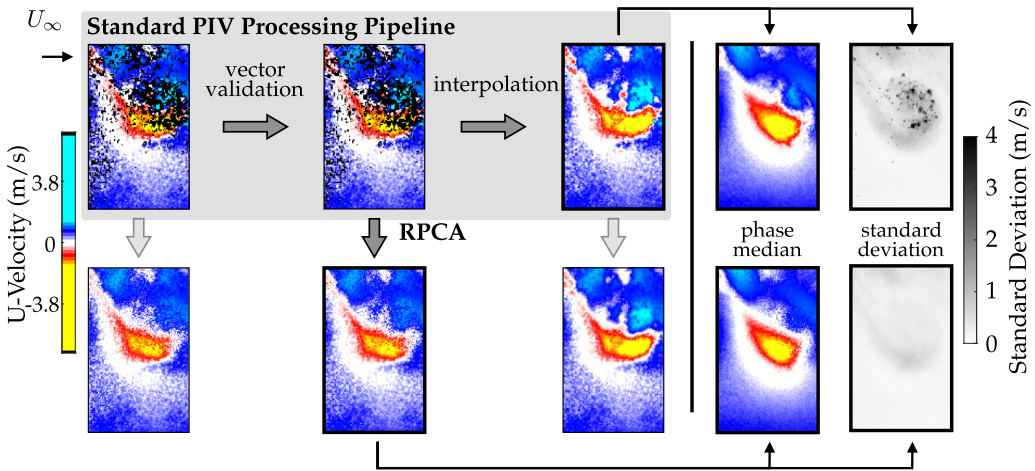


FIG. 16. RPCA filtering of cross-flow turbine wake PIV data. The standard PIV processing pipeline (top row) includes several steps where RPCA filtering can be applied (bottom row). In the cross-correlated streamwise velocity field (top left), 23% of the velocity vectors are missing. Vector validation reduces the missing vectors to 20%. Finally, linear interpolation is used to fill in these missing vectors. In all cases, RPCA filtering captures the relevant phase-averaged coherent structures with fewer outliers and missing data, which appear as dark spots in the standard deviation plot.

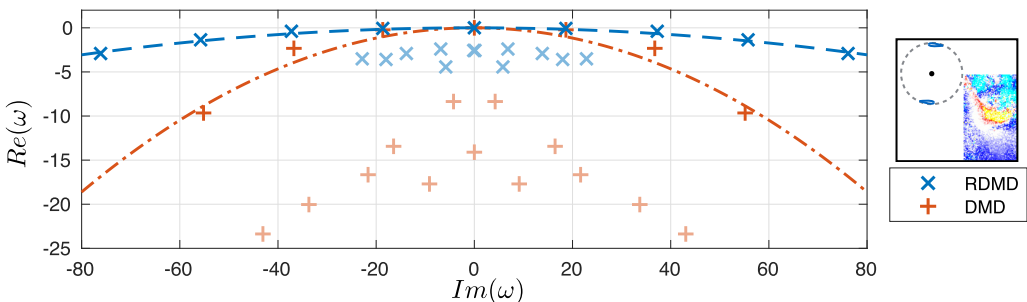


FIG. 17. Continuous-time DMD eigenvalues for the turbine wake PIV data, along with parabolic eigenvalue fits to estimate the error as in Ref. [24]. The RDMD parabolic coefficient is approximately six times smaller than the DMD coefficient.

VI. CONCLUSIONS AND DISCUSSION

In this work, we have demonstrated the ability of RPCA filtering to effectively recover dominant coherent structures from corrupt flow fields with missing measurements. Unlike standard POD/PCA, which is based on least squares and is susceptible to outliers and corruption, RPCA utilizes sparse optimization to separate a data matrix into a low-rank matrix containing correlated structures and a sparse matrix containing the spurious entries.

We apply RPCA filtering to several types of fluid flow data (DNS and PIV), ranging from laminar vortex shedding behind a circular cylinder, to fully turbulent channel flow DNS, and concluding with an experimental flow past a cross-flow turbine. These flows exhibit a variety of phenomena and a range of measurement quality. The DNS examples provide us with a baseline, where it is possible to add corruption to quantitatively assess the performance of RPCA. For flow past a cylinder in DNS, RPCA filtering is extremely effective at separating the true flow field from considerable corruption, with robust recovery even in flow fields with excess of 50% of the measurements corrupted. In the experimental counterpart, RPCA is still able to remove large outliers and corruption, although there is a trade-off between filtering the background turbulence and coherent structures in the wake. The fully turbulent channel flow DNS provides an opportunity to more fully explore this trade-off in a controlled setting, where we can incrementally increase the corruption ratio and observe the filtering effects on various spatial frequencies. As expected, an increasingly aggressive filtering leads to degradation at higher wave numbers, although dominant coherent structures are robustly preserved. Finally, the wake behind a cross-flow turbine provides a practical real-world flow that directly benefits from improved PIV processing. In all three wake flows we also assess the performance of RPCA filtering to yield more accurate modal decompositions. Although we do not have ground-truth measurements and modal decompositions, except in the case of direct numerical simulations, we know that continuous-time DMD eigenvalues should be arranged on the imaginary axis in the complex plane for clean data, and deviations from this may be quantified using the derivation from Bagheri [24]. In all three cases, we see considerable reduction in spurious damping, indicating the denoising effectiveness of RPCA. Based on these results, we believe that RPCA can be a valuable algorithm in the arsenal of PIV processing and filtering techniques, particularly when the processing pipeline culminates in modal analysis.

There are a number of future directions motivated by this work. First, RPCA depends on the hyperparameter λ , and a better understanding of how to objectively choose λ for different conditions is important. Because RPCA is based on sparse, nonconvex optimization, it is also likely that improved optimization techniques may improve speed and robustness. Although this work considered three-dimensional flows, the data comprised two-dimensional cross sections, and the current analysis could be extended to flow volumes. In principle, the RPCA method should generalize, although there may be computational scaling challenges. Recent results have extended RPCA from linear subspaces to manifolds [83], so it may be possible to robustly characterize fluids data that is well described by a low-dimensional manifold [84], rather than a low-dimensional POD subspace. Nonstationary flows may be more challenging for this method, as the bulk distribution will drift. Similarly shocks may be erroneously flagged as outliers; however, this may provide an opportunity to identify shocks in the data. Investigating these flows is an important avenue of future work. It would also be useful to extend this work to PIV measurements of other turbine configurations [85]. Finally, the quality of the RPCA filtered flow fields for additional downstream analyses should be assessed for example, in dynamical systems modeling via Galerkin projection [19] or regression [18] onto the filtered modes and in control [86].

The code and videos for this work are available [87,88].

ACKNOWLEDGMENTS

We gratefully acknowledge funding from the Army Research Office (Grant No. ARO W911NF-19-1-0045) and Naval Facilities Engineering Command. We also thank Jared Callahan, Nathan Kutz, Kazuki Maeda, and Joshua Proctor for valuable discussions.

-
- [1] S. L. Brunton, B. R. Noack, and P. Koumoutsakos, Machine learning for fluid mechanics, *Annu. Rev. Fluid Mech.* **52**, 477 (2020).
 - [2] K. Duraisamy, G. Iaccarino, and H. Xiao, Turbulence modeling in the age of data, *Annu. Rev. Fluid Mech.* **51**, 357 (2019).
 - [3] K. Taira, S. L. Brunton, S. T. M. Dawson, C. W. Rowley, T. Colonius, B. J. McKeon, O. T. Schmidt, S. Gordeyev, V. Theofilis, and L. S. Ukeiley, Modal analysis of fluid flows: An overview, *AIAA J.* **55**, 4013 (2017).
 - [4] A. Pollard, L. Castillo, L. Danaila, and M. Glauser, *Whither Turbulence and Big Data in the 21st Century?* (Springer, Berlin, 2016).
 - [5] N. Aubry, P. Holmes, J. L. Lumley, and E. Stone, The dynamics of coherent structures in the wall region of a turbulent boundary layer, *J. Fluid Mech.* **192**, 115 (1988).
 - [6] G. Berkooz, P. Holmes, and J. L. Lumley, The proper orthogonal decomposition in the analysis of turbulent flows, *Annu. Rev. Fluid Mech.* **25**, 539 (1993).
 - [7] P. J. Holmes, J. L. Lumley, G. Berkooz, and C. W. Rowley, *Turbulence, Coherent Structures, Dynamical Systems and Symmetry*, Cambridge Monographs in Mechanics, 2nd ed. (Cambridge University Press, Cambridge, UK, 2012).
 - [8] J. N. Kutz, S. L. Brunton, B. W. Brunton, and J. L. Proctor, *Dynamic Mode Decomposition: Data-Driven Modeling of Complex Systems* (SIAM, Philadelphia, PA, 2016).
 - [9] J. L. Lumley, *Stochastic Tools in Turbulence* (Academic Press, New York, 1970).
 - [10] K. Taira, M. S. Hemati, S. L. Brunton, Y. Sun, K. Duraisamy, S. Bagheri, S. Dawson, and C.-A. Yeh, Modal analysis of fluid flows: Applications and outlook, [arXiv:1903.05750](https://arxiv.org/abs/1903.05750).
 - [11] S. L. Brunton and B. R. Noack, Closed-loop turbulence control: Progress and challenges, *Appl. Mech. Rev.* **67**, 050801 (2015).
 - [12] C. W. Rowley and S. T. M. Dawson, Model reduction for flow analysis and control, *Annu. Rev. Fluid Mech.* **49**, 387 (2017).
 - [13] A. Towne, O. T. Schmidt, and T. Colonius, Spectral proper orthogonal decomposition and its relationship to dynamic mode decomposition and resolvent analysis, *J. Fluid Mech.* **847**, 821 (2018).
 - [14] C. W. Rowley, I. Mezić, S. Bagheri, P. Schlatter, and D. S. Henningson, Spectral analysis of nonlinear flows, *J. Fluid Mech.* **641**, 115 (2009).
 - [15] P. J. Schmid, Dynamic mode decomposition of numerical and experimental data, *J. Fluid Mech.* **656**, 5 (2010).
 - [16] J. H. Tu, C. W. Rowley, D. M. Luchtenburg, S. L. Brunton, and J. N. Kutz, On dynamic mode decomposition: Theory and applications, *J. Comput. Dyn.* **1**, 391 (2014).
 - [17] K. Carlberg, M. Barone, and H. Antil, Galerkin v. least-squares petrov–galerkin projection in nonlinear model reduction, *J. Comput. Phys.* **330**, 693 (2017).
 - [18] J.-C. Loiseau and S. L. Brunton, Constrained sparse Galerkin regression, *J. Fluid Mech.* **838**, 42 (2018).
 - [19] B. R. Noack, K. Afanasiev, M. Morzynski, G. Tadmor, and F. Thiele, A hierarchy of low-dimensional models for the transient and post-transient cylinder wake, *J. Fluid Mech.* **497**, 335 (2003).
 - [20] S. L. Brunton and J. N. Kutz, *Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control* (Cambridge University Press, Cambridge, UK, 2019).
 - [21] E. J. Candès, X. Li, Y. Ma, and J. Wright, Robust principal component analysis? *J. ACM* **58**, 11 (2011).
 - [22] F. E. Grubbs, Procedures for detecting outlying observations in samples, *Technometrics* **11**, 1 (1969).
 - [23] M. Gavish and D. L. Donoho, The optimal hard threshold for singular values is $4/\sqrt{3}$, *IEEE Trans. Inf. Theory* **60**, 5040 (2014).

- [24] S. Bagheri, Effects of weak noise on oscillating flows: Linking quality factor, Floquet modes, and Koopman spectrum, *Phys. Fluids* **26**, 094104 (2014).
- [25] S. T. M. Dawson, M. S. Hemati, M. O. Williams, and C. W. Rowley, Characterizing and correcting for the effect of sensor noise in the dynamic mode decomposition, *Exp. Fluids* **57**, 42 (2016).
- [26] M. S. Hemati, C. W. Rowley, E. A. Deem, and L. N. Cattafesta, De-biasing the dynamic mode decomposition for applied Koopman spectral analysis, *Theor. Comput. Fluid Dyn.* **31**, 349 (2017).
- [27] M. Farazmand and G. Haller, Computing Lagrangian coherent structures from their variational theory, *Chaos* **22**, 013128 (2012).
- [28] M. A. Green, C. W. Rowley, and A. J. Smits, The unsteady three-dimensional wake produced by a trapezoidal pitching panel, *J. Fluid Mech.* **685**, 117 (2011).
- [29] G. Haller, Lagrangian coherent structures from approximate velocity data, *Phys. Fluids* **14**, 1851 (2002).
- [30] S. G. Raben, S. D. Ross, and P. P. Vlachos, Computation of finite-time Lyapunov exponents from time-resolved particle image velocimetry data, *Exp. Fluids* **55**, 1638 (2014).
- [31] S. C. Shadden, K. Katija, M. Rosenfeld, J. E. Marsden, and J. O. Dabiri, Transport and stirring induced by vortex formation, *J. Fluid Mech.* **593**, 315 (2007).
- [32] S. C. Shadden, F. Lekien, and J. E. Marsden, Definition and properties of Lagrangian coherent structures from finite-time Lyapunov exponents in two-dimensional aperiodic flows, *Physica D* **212**, 271 (2005).
- [33] J. W. Foreman Jr, E. W. George, and R. D. Lewis, Measurement of localized flow velocities in gases with a laser doppler flowmeter, *Appl. Phys. Lett.* **7**, 77 (1965).
- [34] Y. Yeh and H. Z. Cummins, Localized fluid flow measurements with an he–ne laser spectrometer, *Appl. Phys. Lett.* **4**, 176 (1964).
- [35] R. J. Adrian, Particle-imaging techniques for experimental fluid mechanics, *Annu. Rev. Fluid Mech.* **23**, 261 (1991).
- [36] R. J. Adrian, Twenty years of particle image velocimetry, *Exp. Fluids* **39**, 159 (2005).
- [37] R. J. Adrian and J. Westerweel, *Particle Image Velocimetry* (Cambridge University Press, Cambridge, UK, 2011), Vol. 30.
- [38] M. Raffel, C. E. Willert, F. Scarano, C. J. Kähler, S. T. Wereley, and J. Kompenhans, *Particle Image Velocimetry: A Practical Guide* (Springer, Berlin, 2018).
- [39] J. Westerweel, G. E. Elsinga, and R. J. Adrian, Particle image velocimetry for complex and turbulent flows, *Annu. Rev. Fluid Mech.* **45**, 409 (2013).
- [40] C. E. Willert and M. Gharib, Digital particle image velocimetry, *Exp. Fluids* **10**, 181 (1991).
- [41] H. Huang, D. Dabiri, and M. Gharib, On errors of digital particle image velocimetry, *Meas. Sci. Technol.* **8**, 1427 (1997).
- [42] D. Garcia, A fast all-in-one method for automated post-processing of PIV data, *Exp. Fluids* **50**, 1247 (2011).
- [43] D. P. Hart, PIV error correction, *Exp. Fluids* **29**, 13 (2000).
- [44] J. Nogueira, A. Lecuona, and P. A. Rodriguez, Data validation, false vectors correction and derived magnitudes calculation on PIV data, *Meas. Sci. Tech.* **8**, 1493 (1997).
- [45] F. F. J. Schrijer and F. Scarano, Effect of predictor–corrector filtering on the stability and spatial resolution of iterative PIV interrogation, *Exp. Fluids* **45**, 927 (2008).
- [46] S. Discetti, A. Natale, and T. Astarita, Spatial filtering improved tomographic PIV, *Exp. Fluids* **54**, 1505 (2013).
- [47] M. A. Mendez, M. Raiola, A. Masullo, S. Discetti, A. Ianiro, R. Theunissen, and J.-M. Buchlin, Pod-based background removal for particle image velocimetry, *Exp. Therm. Fluid Sci.* **80**, 181 (2017).
- [48] J. Duncan, D. Dabiri, J. Hove, and I, Universal outlier detection for particle image velocimetry (PIV) and particle tracking velocimetry (PTV) data, *Meas. Sci. Technol.* **21**, 057002 (2010).
- [49] J. Westerweel, Efficient detection of spurious vectors in particle image velocimetry data, *Exp. Fluids* **16**, 236 (1994).
- [50] J. Westerweel and F. Scarano, Universal outlier detection for PIV data, *Exp. Fluids* **39**, 1096 (2005).
- [51] K. Pearson, On lines and planes of closest fit to systems of points in space, *Philos. Mag.* **2**, 559 (1901).
- [52] J. Wong, C. Colburn, E. Meeks, and S. Vedaraman, RAD–outlier detection on big data, 2015.

- [53] T. Bouwmans and E. H. Zahzah, Robust PCA via principal component pursuit: A review for a comparative evaluation in video surveillance, *Comp. Vis. Image Under.* **122**, 22 (2014).
- [54] J. Lei, S. Liu, X. Wang, and Q. Liu, An image reconstruction algorithm for electrical capacitance tomography based on robust principle component analysis, *Sensors* **13**, 2076 (2013).
- [55] P.-S. Huang, S. Deeann Chen, P. Smaragdis, and M. Hasegawa-Johnson, Singing-voice separation from monaural recordings using robust principal component analysis, in *Proceedings of the IEEE ICASSP*, (IEEE, Los Alamitos, CA, 2012), p. 57–60.
- [56] P. A. Businger and G. H. Golub, Algorithm 358: Singular value decomposition of a complex matrix [f1, 4, 5], *Commun. ACM* **12**, 564 (1969).
- [57] G. Golub and W. Kahan, Calculating the singular values and pseudo-inverse of a matrix, *J. Soc. Ind. Appl. Math., Ser. B* **2**, 205 (1965).
- [58] G. H. Golub and C. Reinsch, Singular value decomposition and least squares solutions, *Numer. Math.* **14**, 403 (1970).
- [59] G. H. Golub and C. F. Van Loan, An analysis of the total least squares problem, *SIAM J. Numer. Anal.* **17**, 883 (1980).
- [60] L. Sirovich, Turbulence and the dynamics of coherent structures, parts I-III, *Q. Appl. Math.* **45**, 561 (1987).
- [61] K. Karhunen, Über lineare Methoden in der Wahrscheinlichkeitsrechnung, Vol. 37, *Annales Academiae-Scientiarum Fennicae, Ser. A. I* (Academia Scientiarum Fennica, Helsinki, 1947).
- [62] E. N. Lorenz, Empirical orthogonal functions and statistical weather prediction, Technical Report, Massachusetts Institute of Technology (1956).
- [63] H. Hotelling, Analysis of a complex of statistical variables into principal components, *Journal of Educational Psychology* **24**, 417 (1933).
- [64] H. Hotelling, Analysis of a complex of statistical variables into principal components, *Journal of Educational Psychology* **24**, 498 (1933).
- [65] I. Mezić, Spectral properties of dynamical systems, model reduction and decompositions, *Nonlin. Dynam.* **41**, 309 (2005).
- [66] I. Mezic, Analysis of fluid flows via spectral properties of the Koopman operator, *Annu. Rev. Fluid Mech.* **45**, 357 (2013).
- [67] T. Askham and J. N. Kutz, Variable projection methods for an optimized dynamic mode decomposition, *SIAM J. Appl. Dyn. Syst.* **17**, 380 (2018).
- [68] S. L. Brunton, J. L. Proctor, J. H. Tu, and J. N. Kutz, Compressed sensing and dynamic mode decomposition, *J. Comput. Dyn.* **2**, 165 (2015).
- [69] F. Gueniat, L. Mathelin, and L. Pastur, A dynamic mode decomposition approach for large and arbitrarily sampled systems, *Phys. Fluids* **27**, 025113 (2015).
- [70] J. H. Tu, C. W. Rowley, J. N. Kutz, and J. K. Shang, Spectral analysis of fluid flows using sub-Nyquist-rate PIV data, *Exp. Fluids* **55**, 1805 (2014).
- [71] M. R. Jovanović, P. J. Schmid, and J. W. Nichols, Sparsity-promoting dynamic mode decomposition, *Phys. Fluids* **26**, 024103 (2014).
- [72] S. Bagheri, Koopman-mode decomposition of the cylinder wake, *J. Fluid Mech.* **726**, 596 (2013).
- [73] A. Sobral, T. Bouwmans, and El-hadi Zahzah, Lrslibrary: Low-rank and sparse tools for background modeling and subtraction in videos, in *Robust Low-Rank and Sparse Matrix Decomposition: Applications in Image and Video Processing* (CRC Press, Boca Raton, FL, 2015).
- [74] Z. Lin, M. Chen, and Yi Ma, The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices, [arXiv:1009.5055](https://arxiv.org/abs/1009.5055).
- [75] X. Yuan and J. Yang, Sparse and low-rank matrix decomposition via alternating direction methods (unpublished).
- [76] B. R. Noack, W. Stankiewicz, M. Morzynski, and P. J. Schmid, Recursive dynamic mode decomposition of a transient cylinder wake, *J. Fluid Mech.* **809**, 843 (2016).
- [77] M. Farge and K. Schneider, Coherent vortex simulation (CVS), a semi-deterministic turbulence model using wavelets, *Flow, Turbul. Combust.* **66**, 393 (2001).

- [78] J. K. Shang, Flexibility and curvature effects on vortex dynamics and fluid-structure interactions, Ph.D. thesis, Princeton University, 2015.
- [79] T. Colonius and K. Taira, A fast immersed boundary method using a nullspace approach and multi-domain far-field boundary conditions, *Comput. Methods Appl. Mech. Eng.* **197**, 2131 (2008).
- [80] K. Taira and T. Colonius, The immersed boundary method: A projection approach, *J. Comput. Phys.* **225**, 2118 (2007).
- [81] J. Graham, K. Kanov, X. I. A. Yang, M. Lee, N. Malaya, C. C. Lalescu, R. Burns, G. Eyink, A. Szalay, R. D. Moser, and C. Meneveau, A web services accessible database of turbulent channel flow and its use for testing a new integral wall model for LES, *J. Turbul.* **17**, 181 (2016).
- [82] F. Scarano, Iterative image deformation methods in PIV, *Meas. Sci. Tech.* **13**, R1 (2001).
- [83] He Lyu, N. Sha, S. Qin, M. Yan, Y. Xie, and R. Wang, Manifold denoising by nonlinear robust principal component analysis, in *Advances in Neural Information Processing Systems* (2019), pp. 13390–13400.
- [84] J.-C. Loiseau, B. R. Noack, and S. L. Brunton, Sparse reduced-order modeling: Sensor-based dynamics to full-state estimation, *J. Fluid Mech.* **844**, 459 (2018).
- [85] A. Posa, C. M. Parker, M. C. Leftwich, and E. Balaras, Wake structure of a single vertical axis wind turbine, *Int. J. Heat Fluid Flow* **61**, 75 (2016).
- [86] Z. P. Berger, P. R. Shea, M. G. Berry, B. R. Noack, S. Gogineni, and M. N. Glauser, Active flow control for high speed jets with large window piv, *Flow, Turbul. Combust.* **94**, 97 (2015).
- [87] Matlab code: github.com/ischerl/RPCA-PIV.
- [88] Videos: tinyurl.com/RPCA-PIV.