Modal Analysis of Fluid Flows: An Overview

Kunihiko Taira
Florida State University, Tallahassee, Florida 32310
Steven L. Brunton
University of Washington, Seattle, Washington 98195
Scott T. M. Dawson and Clarence W. Rowley
Princeton University, Princeton, New Jersey 08544
Tim Colonius, Beverley J. McKeon, and Oliver T. Schmidt
California Institute of Technology, Pasadena, California 91125
Stanislav Gordeyev
University of Notre Dame, Notre Dame, Indiana 46556
Vassilios Theofilis
University of Liverpool, Brownlow Hill, England L69 3GH, United Kingdom
and
Lawrence S. Ukeiley
University of Florida, Gainesville, Florida 32611
DOI: 10.2514/1.J056060

I. Introduction

SIMPLE aerodynamic configurations under even modest conditions can exhibit complex flows with a wide range of temporal and spatial features. It has become common practice in the analysis of these flows to look for and extract physically important features, or modes, as a first step in the analysis. This step typically starts with a modal decomposition of an experimental or numerical dataset of the flowfield, or of an operator relevant to the system. We describe herein some of the dominant techniques for accomplishing these modal decompositions and analyses that have seen a surge of activity in recent decades [1–8]. For a nonexpert, keeping track of recent developments can be daunting, and the intent of this document is to provide an introduction to modal analysis that is accessible to the

Kunihiko Taira is an Associate Professor of Mechanical Engineering at the Florida State University. His research focuses on computational fluid dynamics, flow control, and network science. He received his B.S. degree from the University of Tennessee, and he received his M.S. and Ph.D. degrees from the California Institute of Technology. He is a recipient of the 2013 U.S. Air Force Office of Scientific Research and 2016 Office of Naval Research Young Investigator Awards. He is an Associate Fellow of the AIAA.

Steven L. Brunton is an Assistant Professor of Mechanical Engineering and a Data Science Fellow with the eScience Institute at the University of Washington. He received his B.S. degree in Mathematics with a minor in Control and Dynamical Systems from the California Institute of Technology in 2006, and he received his Ph.D. degree in Mechanical and Aerospace Engineering from Princeton University in 2012. He is a Member of the AIAA.

Scott T. M. Dawson is currently a Postdoctoral Scholar within the Graduate Aerospace Laboratories at the California Institute of Technology. Before this, he completed his Ph.D. degree in Mechanical and Aerospace Engineering at Princeton University. His research interests include modeling and control of unsteady aerodynamic systems and wall-bounded turbulence. He is a Member of the AIAA.
larger fluid dynamics community. In particular, we present a brief overview of several of the well-established techniques and clearly lay the framework of these methods using familiar linear algebra. The modal analysis techniques covered in this paper include the proper orthogonal decomposition (POD), balanced proper orthogonal decomposition (balanced POD), dynamic mode decomposition

Clarence W. Rowley is a Professor in Mechanical and Aerospace Engineering. His research interests lie at the intersection of dynamical systems, control theory, and fluid mechanics. A recipient of a National Science Foundation CAREER Award and a U.S. Air Force Office of Scientific Research Young Investigator Award, he has made contributions to various model reduction methods used in fluid mechanics, including proper orthogonal decomposition (POD) for compressible flows, balanced POD, and the connection between dynamic mode decomposition and the Koopman operator. He is an Associate Fellow of the AIAA.

Tim Colonius is the Frank and Ora Lee Marble Professor of Mechanical Engineering at the California Institute of Technology. He is an author of over 250 publications on topics including aeroacoustics, flow instability, turbulence, interfacial phenomena, and numerical methods. He is a Fellow of the American Physical Society and the Acoustical Society of America. He is an Associate Fellow of the AIAA.

Beverley J. McKeon is the Theodore von Kármán Professor of Aeronautics in the Graduate Aerospace Laboratories at the California Institute of Technology. Her research interests include experimental and theoretical contributions in wall turbulence, as well as interdisciplinary approaches to modeling and flow manipulation. She is a Fellow of the American Physical Society and an Associate Fellow of the AIAA.

Oliver T. Schmidt is a Senior Postdoctoral Scholar in the Department of Mechanical and Civil Engineering at the California Institute of Technology. He received his Ph.D. degree in Aeronautical Engineering from the University of Stuttgart in 2014. His research focuses on hydrodynamic stability, physics-based modeling, and computational fluid mechanics. He is a Member of the AIAA.

Stanislav Gordeyev is an Associate Professor in the Department of Aerospace and Mechanical Engineering at the University of Notre Dame. His main research interests include experimental investigation of optical distortions caused by compressible turbulent flows, aerooptical and reduced-order modeling, flow control, and mitigation techniques. He is an Associate Fellow of the AIAA.

Vassilios Theofilis is a Chair in Aerospace Engineering at the University of Liverpool, UK and has been an Alexander von Humboldt Research Fellow in Germany and a Ramón y Cajal Senior Research Fellow in Spain. He received his B.Sc. from the University of Patras, Greece, and both his M.Sc. and Ph.D. from the University of Manchester, UK. His work focuses on instability of flows from the incompressible to the hypersonic regime. He is an Associate Fellow of the AIAA.

Lawrence S. Ukeiley joined the Mechanical and Aerospace Engineering Department at the University of Florida in 2006, where he is an Associate Professor. He has been working with modal decomposition techniques of turbulent flows for more than 20 years. He holds a B.S. degree from Alfred University, along with M.S. and Ph.D. degrees from Clarkson University, all in Mechanical Engineering. He is an Associate Fellow of the AIAA.
In the study of fluid mechanics, there can be distinct physical features that are shared across a variety of flows and even over a wide range of parameters such as the Reynolds number and Mach number [9,10]. Examples of common flow features and phenomena include von Kármán shedding [11–17], Kelvin–Helmholtz instability [18–20], and vortex pairing/merging [21–23]. The fact that these features are often easily recognized through simple visual inspections of the flow, even under the presence of perturbations or variations, provides us with the expectation that the features can be extracted through some mathematical procedure [24]. We can further anticipate that these dominant features provide a means to describe in a low-dimensional form that appears as a complex high-dimensional flow. Moreover, as computational techniques and experimental measurements are advancing their ability in providing large-scale high-fidelity data, the compression of a vast amount of flowfield data to a low-dimensional form is ever more important in studying complex fluid flows and developing models for understanding and modeling their dynamical behaviour.

To briefly illustrate these ideas, let us provide a preview of modal decomposition. In Fig. 1, we present a modal decomposition analysis of two-dimensional laminar separated flow over a flat-plate wing [25,26]. By inspecting the flowfield, we clearly observe the formation of a von Kármán vortex street in the wake as the dominant unsteady feature. A modal decomposition method discussed later (proper orthogonal decomposition [1,27,28]; see Sec. III) can extract the important oscillatory modes of this flow. Moreover, two of these most dominant modes and the mean represent (reconstruct) the flowfield very effectively, as shown in the bottom figure. Additional modes can be included to reconstruct the original flow more accurately, but their contributions are much smaller in comparison to the two unsteady modes shown in this example. What is also encouraging is that the modes seen here share a striking resemblance to the dominant modes for three-dimensional turbulent flow at a much higher Reynolds number of 23,000 with a different airfoil and angle of attack (see Sec. III.B.1).

We refer to modal decomposition as a mathematical technique to extract energetically and dynamically important features of fluid flows. The spatial features of the flow are called (spatial) modes, and they are accompanied by characteristic values, representing either the energy content levels or growth rates and frequencies. These modes can be determined from the flowfield data or from the governing equations. We will refer to modal decomposition techniques that take flowfield data as input to the analysis as data-based techniques. This paper will also present modal analysis methods that require a more theoretical framework or discrete operators from the Navier–Stokes equations, and we will refer to them as operator-based techniques.

The origin of this document lies with an AIAA Discussion Group, titled “Modal Decomposition of Aerodynamic Flows,” formed under the auspices of the Fluid Dynamics Technical Committee. One of the initial charters for this group was to organize an invited session where experts in the areas of modal decomposition methods would provide an introductory crash course on the methods. The intended audience for these talks was the nonspecialist, e.g., a new graduate student or early-career researcher who, in one afternoon, could acquire a compact yet intensive introduction to the modal analysis methods. This session (121-FC-5) appeared at the 2016 AIAA Aviation Conference (13–17 June 2016 in Washington, D.C.) and provided the foundation for the present overview paper.

In this overview paper, we present key modal decomposition and analysis techniques that can be used to study a range of fluid flows. We start by rephrasing the basics of eigenvalue and singular value decompositions as well as pseudospectral analysis in Sec. II, which serve as the backbone for all decomposition and analysis techniques discussed here. We then present data-based modal decomposition techniques: proper orthogonal decomposition in Sec. III, balanced POD in Sec. IV, and dynamic mode decomposition in Sec. V. These sections are then followed by discussions on operator-based modal analysis techniques. The Koopman analysis is briefly discussed in Sec. VI as a generalization of the DMD analysis to encapsulate nonlinear dynamics using a linear (but infinite-dimensional) operator-based framework. The global linear stability analysis and resolvent analysis are presented in Secs. VII and VIII, respectively. Table 1 provides a brief summary of the techniques to facilitate comparison of the methods before engaging in the details of each method.

For each of the methods presented, we provide subsections on overview, description, illustrative examples, and future outlook. We offer in the Appendix an example of how the flowfield data can be arranged into vector and matrix forms in preparation for performing the (data-based) modal decomposition techniques presented here.

*Video recordings of this session have been made available by AIAA on https://www.youtube.com/user/AIAATV/playlists [retrieved 10 October 2017].
Table I Summary of the modal decomposition/analysis techniques for fluid flows presented in the present paper

<table>
<thead>
<tr>
<th>Techniques</th>
<th>Sections</th>
<th>Inputs</th>
<th>General descriptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data-based</td>
<td>POD V</td>
<td>Data (L or NL flow; C and E)</td>
<td>Determines the optimal set of modes to represent data based on $L_2$ norm (energy)</td>
</tr>
<tr>
<td></td>
<td>Balanced POD IV</td>
<td>Data (L forward and L adjoint flow; C)</td>
<td>Gives balancing and adjoint modes based on input-output relation (balanced truncation)</td>
</tr>
<tr>
<td></td>
<td>DMD V</td>
<td>Data (L or NL flow; C and E)</td>
<td>Captures dynamic modes with associated growth rates and frequencies; linear approximation to nonlinear dynamics</td>
</tr>
<tr>
<td>Operator-based</td>
<td>Koopman analysis VI</td>
<td>Theoretical (also see DMD)</td>
<td>Transforms nonlinear dynamics into linear representation but with an infinite-dimensional operator; Koopman modes are approximated by DMD modes</td>
</tr>
<tr>
<td>Global linear stability analysis VII</td>
<td>L NS operators and base flow (C)</td>
<td>Finds linear stability modes about a base flow (i.e., steady state); assumes small perturbations about base flow</td>
<td></td>
</tr>
<tr>
<td>Resolvent analysis VIII</td>
<td>L NS operators and base flow (C)</td>
<td>Provides forcing and response modes based on input-output analysis with respect to a base flow (including time-averaged mean flow); can be applied to turbulent flow</td>
<td></td>
</tr>
</tbody>
</table>

$L$, (linear), NL (nonlinear), C (computational), E (experimental), and NS (Navier–Stokes).

II. Eigenvalue and Singular Value Decompositions

The decomposition methods presented in this paper are founded on the eigenvalue and singular value decompositions of matrices or operators. In this section, we briefly present some important fundamental properties of the eigenvalue and singular value decomposition techniques. We also briefly discuss the concepts of pseudospectra and nonnormality.

Eigenvalue decomposition is performed on a square matrix, whereas singular value decomposition can be applied on a rectangular matrix. Analyses based on the eigenvalue decomposition are usually employed when the range and domain of the matrix or operator are the same [29]. That is, the operator of interest can take a vector and map it into the same space. Hence, eigenvalue decomposition can help examine the iterative effects of the operator (e.g., $A^n$ and $\exp(At) = I + At + \frac{1}{2} A^2 t^2 + \cdots$).

The singular value decomposition, on the other hand, is performed on a rectangular matrix, which means that the domain and range spaces are not necessarily the same. As a consequence, singular value decomposition is not associated with analyzing iterative operators. That is, rectangular matrices cannot serve as propagators. However, singular value decomposition can be applied on rectangular data matrices compiled from dynamical processes (see Sec. II.C and the Appendix for details).

The theories and numerical algorithms for eigenvalue and singular value decompositions are not provided here but are discussed extensively in textbooks by Horn and Johnson [30], Golub and Loan [31], Trefethen and Bau [29], and Saad [32]. Numerical programs and libraries to perform eigenvalue and singular value decompositions are listed in Sec. II.D.

A. Eigenvalue Decomposition

The eigenvalues and eigenvectors of a matrix (linear operator) capture the directions in which vectors can grow or shrink. For a given matrix $A \in \mathbb{C}^{n \times n}$, a vector $v \in \mathbb{C}^n$ and a scalar $\lambda \in \mathbb{C}$ are called an eigenvector and an eigenvalue, respectively, of $A$ if they satisfy

$$Av = \lambda v$$

Note that the eigenvectors are unique only up to a complex scalar. That is, if $v$ is an eigenvector, $\alpha v$ is also an eigenvector (where $\alpha \in \mathbb{C}$). The eigenvectors obtained from computer programs are commonly normalized such that they have unit magnitude. The set of all eigenvalues of $A$ is called a spectrum of $A$.

Although the preceding expression in Eq. (1) appears simple, the concept of an eigenvector has great significance in describing the effect of premultiplying $A$ on a vector. The aforementioned expression states that, if an operator $A$ is applied to its eigenvector (eigendirection), the operation can be captured solely by the multiplication of a scalar $\lambda$, which is the eigenvalue associated with that direction. The magnitude of the eigenvalue tells us whether the operator $A$ will increase or decrease the size of the original vector in that particular direction. If multiplication by $A$ is performed in an iterative manner, the resulting vector from the compound operations can be predominantly described by the eigenvector having the eigenvalue with the largest magnitude, as shown by the illustration in Fig. 2.

If $A$ has $n$ linearly independent eigenvectors $v_j$ with corresponding eigenvalues $\lambda_j$ ($j = 1, \ldots, n$), then we have

$$AV = VA$$

where $V = [v_1, v_2, \ldots, v_n] \in \mathbb{C}^{n \times n}$ and $A = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n) \in \mathbb{C}^{n \times n}$. Postmultiplying $V^{-1}$ to the preceding equation, we have

$$A = VAV^{-1}$$

This is called the eigenvalue decomposition. For the eigenvalue decomposition to hold, $A$ needs to have a full set of $n$ linearly independent eigenvectors.\(^\dagger\)

For linear dynamical systems, we often encounter systems for some state variable $x(t) \in \mathbb{C}^n$ described by

$$\dot{x}(t) = Ax(t)$$

with the solution of

$$x(t) = \exp(At)x(0) = V \exp(\Lambda t)V^{-1}x(0)$$

where $x(0)$ denotes the initial condition. Here, the eigenvalues characterize the long-term behavior of linear dynamical systems [6,34] for $x(t)$, as illustrated in Fig. 3. The real and imaginary parts of $\lambda_j$ represent the growth (decay) rate and the frequency at which the state variable evolves in the direction of the eigenvector $v_j$. For a linear system to be stable, all eigenvalues need to be on the left-hand side of the complex plane, i.e., $\text{Re}(\lambda_j) \leq 0$ for all $j$.

For intermediate dynamics, the pseudospectra [33,35,36] can provide insights. The concept of pseudospectra is associated with nonnormality of operators and the sensitivity of the eigenvalues to perturbations. We briefly discuss the pseudospectra in Sec. II.E.

For some problems, there can be a mass matrix $B \in \mathbb{C}^{n \times n}$ that appears on the left-hand side of Eq. (4):

$$B\dot{x} = Ax$$

\(^\dagger\)In such a case, $A$ is called diagonalizable or nondefective. If $A$ is defective, we have $A = VJV^{-1}$, with $J$ being the canonical Jordan form [31,33].
has eigenvalues of \( \lambda = \{1.2, 0.5\} \). Figure 2 Collection of random points (vectors \( x \)) stretched in the direction of the dominant eigenvector \( v_1 \) with iterative operations \( A^k \) for matrix \( A \), which has eigenvalues of \( \lambda_1 = 1.2 \) and \( \lambda_2 = 0.5 \).

Fig. 3 Dynamic response of a linear system characterized by the eigenvalues (stable: \( \text{Re}(\lambda) < 0 \), and unstable: \( \text{Re}(\lambda) > 0 \)). Location of example eigenvalues \( \lambda \) are shown by the symbols with corresponding sample solutions \( \exp(\lambda t) \) in inset plots.

In such a case, we are led to a generalized eigenvalue problem of the form

\[
A v = \lambda B v
\]

(7)

If \( B \) is invertible, we can rewrite the preceding equation as

\[
B^{-1} A v = \lambda v
\]

(8)

and treat the generalized eigenvalue problem as a standard eigenvalue problem. However, it may not be desirable to consider this reformulation if \( B \) is not invertible or if the inversion of \( B \) results in ill conditioning (worsening of scaling) of the problem. Note that generalized eigenvalue problems can also be solved with many numerical libraries, which are similar to the standard eigenvalue problems. See the work of Trefethen and Embree [33] and Golub and Loan [31] for additional details on the generalized eigenvalue problems.

B. Singular Value Decomposition

The singular value decomposition (SVD) is one of the most important matrix factorizations, generalizing the eigendecomposition to rectangular matrices. The SVD has many uses and interpretations, especially for dimensionality reduction, where it is possible to use the SVD to obtain optimal low-rank matrix approximations [37]. The singular value decomposition also reveals how a rectangular matrix or operator stretches and rotates a vector. As an illustrative example, consider a set of vectors \( v_j \in \mathbb{R}^n \) of unit length that describe a sphere. We can premultiply these unit vectors \( v_j \) with a rectangular matrix \( A \in \mathbb{R}^{m \times n} \) as shown in Fig. 4. The semiaxes of the resulting ellipse (ellipsoid) are represented by the unit vectors \( u_j \) and magnitudes \( \sigma_j \). Hence, we can view the singular values to capture the amount of stretching imposed by matrix \( A \) in the directions of the axes of the ellipse.

Generalizing this concept for complex \( A \in \mathbb{C}^{m \times n} \), \( v_j \in \mathbb{C}^n \), and \( u_j \in \mathbb{C}^m \), we have

\[
A v_j = \sigma_j u_j
\]

(9)

In matrix form, the aforementioned relationship can be expressed as

\[
A V = U \Sigma
\]

(10)

where \( U = [u_1, u_2, \ldots, u_m] \in \mathbb{C}^{m \times n} \) and \( V = [v_1, v_2, \ldots, v_n] \in \mathbb{C}^{n \times n} \) are unitary matrices and \( \Sigma \in \mathbb{R}^{m \times n} \) is a diagonal matrix with \( \sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_p \geq 0 \) along its diagonal, where \( p = \min(m,n) \).

Now, multiplying \( V^{-1} = V^* \) from the right side of the preceding equation, we arrive at

\[
A = U \Sigma V^*
\]

(11)

which is referred to as the singular value decomposition. In the preceding equation, \( ^* \) denotes the conjugate transpose. The column vectors \( u_j \) and \( v_j \) of \( U \) and \( V \) are called the left and right singular vectors, respectively. Both of the singular vectors can be determined

---

\( ^* \)Unitary matrices \( U \) and \( V \) satisfy \( U^* = U^{-1} \) and \( V^* = V^{-1} \), with \( ^* \) denoting the conjugate transpose.
up to a complex scalar of magnitude one (i.e., \( e^{i\theta} \), where \( \theta \in [0, 2\pi) \)).

Given a rectangular matrix \( A \), we can decompose the matrix with the SVD in the following graphical manner:

\[
A = \begin{bmatrix}
U & \Sigma & V^* \end{bmatrix}
\]

where we have taken \( m > n \) in this example. Sometimes, the components in \( U \) enclosed by the broken lines are omitted from the decomposition, as they are multiplied by zeros in \( \Sigma \). The decomposition that disregards the submatrices in the broken-line boxes are called the reduced SVD (economy-sized SVD), as opposed to the full SVD.

In a manner similar to the eigenvalue decomposition, we can interpret the SVD as a means to represent the effect of matrix operation merely through the multiplication by scalars (singular values) given the appropriate directions. Because the SVD is applied to a rectangular matrix, we need two sets of basis vectors to span the domain and range of the matrix. Hence, we have the right singular vectors \( V \) that span the domain of \( A \) and the left singular vectors \( U \) that span the range of \( A \), as illustrated in Fig. 4. This is different from the eigenvalue decomposition of a square matrix: in which case, the domain and the range are (generally) the same. Although the eigenvalue decomposition requires the square matrix to be diagonalizable, the SVD (on the other hand) can be performed on any rectangular matrix.

C. Relationship Between Eigenvalue and Singular Value Decompositions

The eigenvalue and singular value decompositions are closely related. In fact, the left and right singular vectors of \( A \in \mathbb{C}^{m \times n} \) are also the orthonormal eigenvectors of \( AA^* \) and \( A^*A \), respectively. Furthermore, the nonzero singular values of \( A \) are the square roots of the nonzero eigenvalues of \( AA^* \) and \( A^*A \). Therefore, instead of the SVD, the eigenvalue decomposition can be performed on \( AA^* \) or \( A^*A \) to solve for the singular vectors and singular values of \( A \). For these reasons, the smaller of the square matrices of \( AA^* \) and \( A^*A \) are often chosen to perform the decomposition in a computationally inexpensive manner as compared to the full SVD. This property is taken advantage of in some of the decomposition methods discussed in the following because flowfield data usually yield a rectangular data matrix that can be very high-dimensional in one direction (e.g., the snapshot POD method [28] in Sec. III).

D. Numerical Libraries for Eigenvalue and Singular Value Decompositions

Eigenvalue and singular value decompositions can be performed with codes that are readily available. We list a few standard numerical libraries to execute eigenvalue and singular value decompositions.

**MATLAB**: In MATLAB®, the command \texttt{eig} finds the eigenvalues and eigenvectors for standard eigenvalue problems as well as generalized eigenvalue problems. The command \texttt{svd} outputs the singular values and the left and right singular vectors. It can also perform the economy-sized SVD. For small- to moderate-sized problems, MATLAB can offer a user-friendly environment to perform modal decompositions. We provide in Table 2 some common examples of \texttt{eig} and \texttt{svd} in use for canonical decompositions.‡‡

**LAPACK**: LAPACK (linear algebra package) offers standard numerical library routines for a variety of basic linear algebra problems, including eigenvalue and singular value decompositions. The routines are written in Fortran 90. See the users’ guide [38].††

**ScALAPACK**: ScALAPACK (scalable LAPACK) comprises high-performance linear algebra routines for parallel distributed memory machines. ScALAPACK solves dense and banded eigenvalue and singular value problems. See the users’ guide [39].‡‡

**ARPACK**: ARPACK (Arnoldi package) is a numerical library, written in FORTRAN 77, that is specialized to handle large-scale eigenvalue problems as well as generalized eigenvalue problems. It can also perform singular value decompositions. The library is available for both serial and parallel computations. See the users’ guide [40].‡‡

E. Pseudospectra

Before we transition our discussion to the coverage of modal analysis techniques, let us consider the pseudospectral analysis [33, 35], which reveals the sensitivity of the eigenvalue spectra with respect to perturbations to the operator. This is also an important concept in studying transient and input–output dynamics, complementing the stability analysis based on eigenvalues. Concepts from the pseudospectral analysis appear later in the resolvent analysis (Sec. VIII).

For a linear system described by Eq. (4) to exhibit stable dynamics, we require all eigenvalues of its operator \( A \) to satisfy \( \text{Re}(\lambda) < 0 \), as illustrated in Fig. 3. Although this criterion guarantees the solution \( x(t) \) to be stable for large \( t \), it does not provide insights into the transient behavior of \( x(t) \). To illustrate this point, let us consider an example of \( A = VAV^{-1} \) with stable eigenvalues of

\[
\lambda_1 = -0.1, \quad \lambda_2 = -0.2
\]

‡‡For additional details, see the documentation available on http://www.mathworks.com [retrieved 10 February 2017].

††Library available online at http://www.netlib.org/lapack [retrieved 10 February 2017].

‡‡Library available online at http://www.netlib.org/scalapack [retrieved 10 February 2017].

‡‡Library available online at http://www.caam.rice.edu/software/ARPACK [retrieved 10 February 2017].
and eigenvectors of
\[ v_1 = \begin{bmatrix} \cos\left(\frac{\pi}{4} - \delta\right) & \sin\left(\frac{\pi}{4} - \delta\right) \end{bmatrix}^T, \]
\[ v_2 = \begin{bmatrix} \cos\left(\frac{\pi}{4} + \delta\right) & \sin\left(\frac{\pi}{4} + \delta\right) \end{bmatrix}^T, \]

where \( \delta \) is a free parameter to choose. Observe that, as \( \delta \) becomes small, the eigenvectors become nearly linearly dependent, which makes the matrix \( A \) ill conditioned.

Providing an initial condition of \( x(t_0) = [1, 0.1]^T \), we can solve Eq. (5) for different values of \( \delta \), as shown in Figs. 5a and 5b. Although all solutions decay to zero due to the stable eigenvalues, the transient growths of \( x_1(t) \) and \( x_2(t) \) become noticeable as \( \delta \to 0 \). The large transient for small \( \delta \) is caused by the eigenvectors becoming nearly parallel, which necessitate large coefficients to represent the solution [i.e., \( x(t) = x_1(t) \psi_1 + x_2(t) \psi_2 \) where \( |x_1| \) and \( |x_2| \to 1 \) during the transient]. As such, the solution grows significantly during the transient before the decay from the negative eigenvalues starts to take over the solution behavior as \( t \) becomes large. Thus, we observe that the transient behavior of the solution is not controlled by the eigenvalues of \( A \). Nonnormal operators (i.e., operators for which \( AA^* \neq A^*A \)) have nonorthogonal eigenvectors and can exhibit this type of transient behavior. Thus, it is important that care is taken when we examine transient dynamics caused by nonnormal operators. In fluid mechanics, the dynamics of shear-dominant flows often exhibit nonnormality.

To further assess the influence of \( A \) on the transient dynamics, let us examine here how the eigenvalues are influenced by perturbations on \( A \). That is, we consider
\[ \Lambda_\epsilon(A) = \{ z \in \mathbb{C} : z \in \Lambda(A + \Delta A), \quad \| \Delta A \| \leq \epsilon \} \]  

This subset of perturbed eigenvalues is known as the \( \epsilon \) pseudospectrum of \( A \). It is also commonly known with the following equivalent definition:
\[ \Lambda_\epsilon(A) = \{ z \in \mathbb{C} : \| zI - A \|^{-1} \geq \epsilon^{-1} \} \]

Note that, as \( \epsilon \to 0 \), we recover the eigenvalues (0 pseudospectrum); and as \( \epsilon \to \infty \), the subset \( \Lambda_\epsilon(A) \) occupies the entire complex domain. To numerically determine the pseudospectra, we can use the following definition based on the minimum singular value of \((zI - A)\):
\[ \Lambda_\epsilon(A) = \{ z \in \mathbb{C} : \sigma_{\text{min}}(zI - A) \leq \epsilon \}
\]

which is equivalent to \( \Lambda_\epsilon(A) \), described by Eqs. (15) and (16). If \( A \) is normal, the pseudospectrum \( \Lambda_\epsilon(A) \) is the set of points away from \( \Lambda_\epsilon(A) \) (eigenvalues) by only less than or equal to \( \epsilon \) on the complex plane. However, as \( A \) becomes nonnormal, the distance between
$\Lambda_0(A)$ and $\Lambda_+(A)$ may become much larger. As will be discussed later, the resolvent analysis in Sec. VIII considers the pseudospectra along the imaginary axis [6] (i.e., $z = \omega i$, where $\omega \in \mathbb{R}$).

Let us return to the example given by Eqs. (13) and (14) and compute the pseudospectra for decreasing $\delta$ of 0.01, 0.001, and 0.0001, as shown in Figs. 5c–5e, respectively. Here, the contours of the $\epsilon$ pseudospectra are drawn for the same values of $\epsilon$. With decreasing $\delta$, the matrix $A$ becomes increasingly nonnormal and susceptible to perturbations. The influence of nonnormality on the spectra is clearly visible with the expanding $\epsilon$ pseudospectra. It should be noticed that some of the pseudospectra contours penetrate into the right-hand side of the complex plane, suggesting that perturbations of such magnitude may thrust the system to become unstable even with stable eigenvalues. This nonnormal feature can play a role in destabilizing the dynamics with perturbations or nonlinearity.

The transient dynamics of $\dot{x} = Ax$ can be related to how the $\epsilon$ pseudospectrum of $A$ expands from the eigenvalues as parameter $\epsilon$ is varied. The pseudospectrum of $A$ can provide a lower bound on the amount of transient amplification by $\|\exp(At)\|$. If $\Lambda_+(A)$ extends a distance $\eta$ into the right half-plane for a given $\epsilon$, it can be shown through Laplace transform that $\|\exp(At)\|$ must be as large as $\eta/\epsilon$ for some $t > 0$. If we let a constant $\kappa$ for $A$ be defined as the supremum of this ratio over all $\epsilon$, the lower bound for the solution can then be shown to take the form of [41]

$$\sup_{\epsilon \geq 0} \|\exp(At)\| \geq \kappa$$

This constant $\kappa$ is referred to as the Kreiss constant, which provides an estimate of how the solution [Eq. (5)] behaves during the transient. This estimate is not obtained from the eigenanalysis but from the pseudospectral analysis. The same concept applies to time-discretized linear dynamics [42]. Readers can find applications of the pseudospectral analysis to fluid mechanics in the works of Trefethen et al. [35], Trefethen and Embree [33], and Schmid [36].

### III. Proper Orthogonal Decomposition

The proper orthogonal decomposition is a modal decomposition technique that extracts modes based on optimizing the mean square of the field variable being examined. It was introduced to the fluid dynamics/turbulence community by Lumley [27] as a mathematical technique to extract coherent structures from turbulent flowfields. The POD technique, also known as the Karhunen–Loève procedure [43,44], provides an objective algorithm to decompose a set of data into a minimal number of basis functions or modes to capture as much energy as possible. The method itself is known under a variety of names in different fields: POD, principal component analysis (PCA), Hotelling analysis, empirical component analysis, quasi-harmonic modes, empirical eigenfunction decomposition, and others. Closely related to this technique is factor analysis, which is used in psychology and economics. Roots of the POD can be traced back to the middle of the 19th century to the matrix diagonalization technique, which is ultimately related to the SVD (Sec. II). Excellent reviews on the POD can be found in [4,45], and chapter 3 of [46].

In applications of POD to a fluid flow, we start with a vector field $q(\xi,t)$ (e.g., velocity) with its temporal mean $\bar{q}(\xi)$ subtracted and assume that the unsteady component of the vector field can be decomposed in the following manner:

$$q(\xi,t) - \bar{q}(\xi) = \sum_{j} \phi_j(\xi,t)$$

where $\phi_j(\xi,t)$ and $a_j$ represent the modes and expansion coefficients, respectively. Here, $\xi$ denotes the spatial vector. This expression represents the flowfield in terms of a generalized Fourier series for some set of basis functions $\phi_j(\xi,t)$. In the framework of the POD, we seek the optimal set of basis functions for given flowfield data. In early applications of the POD, this typically led to modes that were functions of space and time/frequency [47–51], as will also be discussed in the following.

Modern applications of modal decompositions have further sought to split space and time, hence only needing spatial modes. In that context, the preceding equation can be written as

$$q(\xi,t) - \bar{q}(\xi) = \sum_{j} a_j(t)\phi_j(\xi)$$

where the expansion coefficients $a_j$ are now time dependent. Note that Eq. (20) explicitly employs a separation of variables, which may not be appropriate for all problems. The application of the two forms listed previously should depend on the properties of the flow and the information one wishes to extract, as discussed by Holmes et al. [52]. In what follows, we will discuss the properties of the POD by assuming that the desire is to extract a spatially dependent set of modes.

The POD is one of the most widely used techniques in analyzing fluid flows. There are a large number of variations of the POD technique, with applications including fundamental analysis of fluids flows, reduced-order modeling, data compression/reconstruction, flow control, and aerodynamic design optimization. Because the POD serves as the basis and motivation for the development of other modal decomposition techniques, we provide a somewhat detailed overview of the POD in the following.

#### A. Description

**1. Algorithm**

The inputs are snapshots of any scalar (e.g., pressure, temperature) or vector (e.g., velocity, vorticity) field $q(\xi,t)$ over one-, two-, or three-dimensional discrete spatial points $\xi$ at discrete times $t_j$.

The outputs are a set of orthogonal modes $\phi_j(\xi)$ with their corresponding temporal coefficients $a_j(t)$ and energy levels $\lambda_j$ arranged in the order of their relative amount of energy. The fluctuations in the original field are expressed as a linear combination of the modes and their corresponding temporal coefficients:

$$q(\xi,t) - \bar{q}(\xi) = \sum_{j} a_j(t)\phi_j(\xi)$$

We discuss three main approaches to perform the POD of the flowfield data: namely, the spatial (classical) POD method, the snapshot POD method, and the SVD. In the following, we briefly describe these three methods and discuss how they are related to each other.

**Spatial (Classical) POD Method.** With the POD, we determine the set of basis functions that optimally represents the given flowfield data. First, given the flowfield $q(\xi,t)$, we prepare snapshots of the flowfield stacked in terms of a collection of column vectors $x(t)$. That is, we consider a collection of finite-dimensional data vectors that represents the flowfield:

$$x(t) = q(\xi,t) - \bar{q}(\xi) \in \mathbb{R}^n, \quad t = t_1, t_2, \ldots, t_m$$

Here, $x(t)$ is taken to be the fluctuating component of the data vector with its time-averaged value $\bar{q}(\xi)$ removed. Although the data vector can be written as $x(\xi,t)$, we simply write $x(t)$ to emphasize that it is being considered as a snapshot at time $t$. An example of forming the data vector $x(t)$ for a given flowfield is provided in the Appendix.

The objective of the POD analysis is to find the optimal basis vectors that can best represent the given data. In other words, we seek the vectors $\phi_j(\xi)$ in Eq. (20) that can represent $q(\xi)$ in an optimal manner and with the least number of modes. The solution to this problem [37] can be determined by finding the eigenvectors $\phi_j$ and the eigenvalues $\lambda_j$ from

$$R\phi_j = \lambda_j\phi_j, \quad \phi_j \in \mathbb{R}^n, \quad \lambda_1 \geq \ldots \geq \lambda_n \geq 0$$

Because we reserve the symbol $x$ to denote the data vector, we use $\xi$ to represent the spatial coordinates in this paper.
where \( R \) is the covariance matrix of vector \( x(t) \).

\[
R = \sum_{i=1}^{m} x(t_i)x^T(t_i) = XX^T \in \mathbb{R}^{n \times n} \tag{24}
\]

where the matrix \( X \) represents the \( m \) snapshot data being stacked into a matrix form of

\[
X = [x(t_1), x(t_2), \ldots, x(t_m)] \in \mathbb{R}^{n \times m} \tag{25}
\]

The size of the covariance matrix \( n \) is based on the spatial degrees of freedom of the data. For fluid flow data, \( n \) is generally large and is equal to the number of grid points times the number of variables to be considered in the data, as illustrated in Eq. (A2) of the Appendix. See the Appendix for an example of preparing the data matrix from the velocity field data.

The eigenvectors found from Eq. (37) are called the POD modes. It should be noted that the POD modes are orthonormal. That means that the inner product\(^{11}\) between the modes satisfies

\[
\langle \phi_j, \phi_k \rangle \equiv \int \phi_j \cdot \phi_k \, dV = \delta_{jk}, \quad j, k = 1, \ldots, n \tag{26}
\]

Consequently, the eigenvalues \( \lambda_k \) convey how well each eigenvector \( \phi_k \) captures the original data in the \( L_2 \) sense (scaled by \( m \)). When the velocity vector is used for \( x(t) \), the eigenvalues correspond to the kinetic energy captured by the respective POD modes. If the eigenvalues are arranged from the largest to the smallest, in decreasing order, the POD modes are arranged in the order of importance in terms of capturing the kinetic energy of the flowfield.

We can use the eigenvalues to determine the number of modes needed to represent the fluctuations in the flowfield data. Generally, we retain only \( r \) number of modes to express the flow such that

\[
\sum_{j=1}^{r} \lambda_j \approx 1 \tag{27}
\]

With the determination of the important POD modes, we can represent the flowfield only in terms of the finite or truncated series

\[
g(\xi, t) - \bar{g}(\xi) \approx \sum_{j=1}^{r} a_j(t) \phi_j(\xi) \tag{28}
\]

in an optimal manner, effectively reducing the high-dimensional \( (n) \) flowfield to be represented only with \( r \) modes. The temporal coefficients are determined accordingly

\[
a_j(t) = \langle g(\xi, t) - \bar{g}(\xi), \phi_j(\xi) \rangle = \langle x(t), \phi_j \rangle \tag{29}
\]

Method of Snapshots. When the spatial size of the data \( n \) is very large, the size of the correlation matrix \( R = XX^T \) becomes very large \((n \times n)\), making the use of the classical spatial POD method for finding the eigenvectors practically impossible. Sirovich\(^{28}\) pointed out that the temporal correlation matrix will yield the same dominant spatial modes while giving rise to a much smaller and computationally more tractable eigenvalue problem. This alternative approach, called the method of snapshots, takes a collection of snapshots \( x(t_i) \) at discrete time levels \( t_i, i = 1, 2, \ldots, m \), with \( m \ll n \), and it solves an eigenvalue problem of a smaller size \((m \times m)\) to find the POD modes. The number of snapshots \( m \) should be chosen such that important fluctuations in the flowfield are well resolved in time.

The method of snapshots relies on solving an eigenvalue problem of a much smaller size:

\[
X^T X \psi_j = \lambda_j \psi_j, \quad \psi_j \in \mathbb{R}^m, \quad m \ll n \tag{30}
\]

where \( X^T X \) is of size \( m \times m \) instead of the original eigenvalue problem of size \( n \times n \) [Eq. (37)]. Although we are analyzing the smaller eigenvalue problem, the same nonzero eigenvalues are shared by \( X^T X \) and \( XX^T \) and the eigenvectors of these matrices can be related to each other (see Sec. II.C). With the eigenvectors \( \psi_j \) of the aforementioned smaller eigenvalue problem determined, we can recover the POD modes through

\[
\phi_j = X \psi_j \frac{1}{\sqrt{\lambda_j}} \in \mathbb{R}^n, \quad j = 1, 2, \ldots, m \tag{31}
\]

which can equivalently be written in matrix form as

\[
\Phi = X \Psi \Sigma^{-1/2} \tag{32}
\]

where \( \Phi \in \mathbb{R}^{n \times m} \), \( \Psi \in \mathbb{R}^{m \times m} \), and \( \Sigma \in \mathbb{R}^{n \times n} \), with \( m < n \). The matrices \( \Phi \) and \( \Psi \) contain the left and right singular vectors\(^{12}\) of \( X \), and matrix \( \Sigma \) holds the singular values \((\sigma_1, \sigma_2, \ldots, \sigma_m)\) along its diagonal. These singular vectors held in \( \Phi \) and \( \Psi \) are identical to the eigenvectors of \( XX^T \) and \( X^T X \), respectively. Moreover, the singular values and the eigenvalues are related by \( \sigma_j = \lambda_j \). This means that the SVD can be directly performed on \( X \) to determine the POD modes \( \Phi \).

Note that the SVD finds \( \Phi \) from Eq. (33) with a full size of \( n \times n \), as compared to \( \Phi \) from the snapshot approach with a size of \( n \times m \), holding only the leading \( m \) modes.

The terms POD and SVD are often used interchangeably in the literature. However, the SVD is a decomposition technique for rectangular matrices and POD can be seen as a decomposition formalism for which the SVD can be one of the approaches to determine its solution. Although the method of snapshot is preferred for handling large datasets, the SVD-based technique to determine the POD modes is known to be robust against roundoff errors\(^{3}\).

2. Notes

Optimality. The POD modes are computed in the optimal manner in the \( L_2 \) sense\(^{1} \). If the velocity or vorticity field is used to determine the POD modes, the modes are optimal to capture the kinetic energy or enstrophy, respectively, of the flowfield. Moreover, POD decomposition is optimal, not only in terms of minimizing the mean-square error between the signal and its truncated representation but also minimizing the number of modes required to describe the signal for a given error\(^{53}\).

\(^{1}\)Precisely speaking, the covariance matrix is defined as \( R \equiv XX^T / m \) or \( XX^T / (m - 1) \). For clarity of presentation, we drop the factor \( 1/m \) and note that it is lumped into the eigenvalue \( \lambda_k \).

\(^{11}\)For the sake of discussion, we consider here that the flowfield data to be placed on a uniform grid such that scaling due to the size of the cell volume does not need to be taken into account. In general, the cell volume for each data point needs to be included in the formulation to represent this inner product (volume integral). Consequently, the covariance matrix [Eq. (24)] should be written as \( R \equiv XX^T W \), where \( W \) holds the spatial weights. The matrix \( X^T X \) that later appears in Eq. (30) for the method of snapshot would similarly be replaced by \( X^T WX \).

\(^{12}\)The matrices \( \Phi \) and \( \Psi \) are orthonormal, i.e., \( \Phi^T \Phi = \Phi \Phi^T = I \) and \( \Psi^T \Psi = \Psi \Psi^T = I \).
The optimality (the fastest-convergent property) of the POD reduces the amount of information required to represent statistically dependent data to a minimum. This crucial feature explains the wide usage of the POD in a process of analyzing data. For this reason, the POD is used extensively in the fields of detection, estimation, pattern recognition, and image processing.

**Reduced-Order Modeling.** The orthogonality of POD modes \([\phi_i, \phi_j] = \delta_{ij}\) is an attractive property for constructing reduced-order models [1,54]. Galerkin projection can be used to reduce high-dimensional discretizations of partial differential equations into reduced-order ordinary differential equation models for the temporal coefficients \(a_j(t)\). POD modes have been used to construct Galerkin projection-based reduced-order models for incompressible [48,55] and compressible [58] flows.

**Traveling Structures.** With real-valued POD modes, traveling structures cannot be represented as a single mode. In general, traveling structures are represented by a pair of stationary POD modes, which are similar but appear shifted in the advection direction. See, for instance, modes 1 and 2 in Figs. 1 and 6. One way to understand the emergence of POD mode pairs for traveling structures is to consider the following traveling sine wave example:

\[
\sin(\xi - ct) = \cos(ct) \sin(\xi) - \sin(ct) \cos(\xi)
\]  

(34)

Here, we have a pair of spatial modes, \(\sin(\xi)\) and \(\cos(\xi)\), which are shifted by a phase of \(\pi/2\). Note that we can combine the pair of modes into a single mode if a complex representation of POD modes is considered [59]. There are variants of the POD analysis specialized for traveling structures [60,61].

**Constraints.** With linear superposition of the POD modes in representing the flowfield, each and every POD mode also satisfies linear constraints, such as the incompressibility constraint and the no-slip boundary condition. This statement asserts that the given data also satisfy these constraints.

**Homogeneous Directions.** For homogeneous, periodic, or stationary (translationally invariant) directions, POD modes reduce to Fourier modes [62].

**Spectral POD.** It is also possible to consider the use of the POD in the frequency domain. Spectral POD provides time-harmonic modes at discrete frequencies from a set of realizations of the temporal Fourier transform of the flowfield. This application of the POD provides an orthogonal basis of modes at discrete frequencies that are optimally ranked in terms of energy because the POD reduces to a harmonic analysis over directions that are stationary or periodic [62,63]. Hence, the method is effective at extracting coherent structures from statistically stationary flows and has been successfully applied in the early applications of the POD, as well as other flows including turbulent jets [64--68]. This frequency-based approach to the POD overcomes some of the weaknesses, and the discussions in the strengths and weaknesses do not directly apply here [69].

Spectral POD can be estimated from a time series of snapshots in the form of Eq. (25) using Welch’s method [70]. First, the data are segmented into a number of \(n_b\) (of potentially overlapping) blocks or realizations, consisting of \(n_{FFT}\) snapshots each. Under the ergodicity hypothesis, each block can be regarded as statistically independent realization of the flow. We proceed by calculating the temporal Fourier transform

\[
\hat{X}^{(l)} = [\hat{x}(\omega_1)^{(l)} \hat{x}(\omega_2)^{(l)} \cdots \hat{x}(\omega_{n_{FFT}})^{(l)}] \in \mathbb{R}^{n_{FFT}}
\]  

(35)

of each block, where superscript \(l\) denotes the \(l\)th block. All realizations of the Fourier transform at a specific frequency \(\omega_k\) are now collected into a new data matrix:

\[
\tilde{X}_{\omega_k} = [\hat{x}(\omega_k)^{(1)} \hat{x}(\omega_k)^{(2)} \cdots \hat{x}(\omega_k)^{(n_b)}]
\]  

(36)

The product \(\tilde{X}_{\omega_k} \tilde{X}^T_{\omega_k}\) forms the cross-spectral density matrix, and its eigenvalue decomposition

\[
\tilde{X}_{\omega_k} \tilde{X}^T_{\omega_k} \phi_{\omega_k,j} = \lambda_{\omega_k,j} \phi_{\omega_k,j}, \quad \phi_{\omega_k,j} \in \mathbb{R}^n,
\]  

(37)

yields the spectral POD modes \(\phi_{\omega_k,j}\) and corresponding modal energies \(\lambda_{\omega_k,j}\), respectively. As an example, the most energetic spectral POD mode of a turbulent jet at one frequency is shown in Fig. 16d. The spectral POD as described here dates back to the early work of Lumley [62] and is not related to the method under the same name proposed by Sieber et al. [71], whose approach blended between the POD and discrete Fourier transform modes by filtering the temporal correlation matrix.

3. **Strengths and Weaknesses**

**Strengths:**

1) The POD gives an orthogonal set of basis vectors with the minimal dimension. This property is useful in constructing a reduced-order model of the flowfield.

2) POD modes are simple to compute using either the (classical) spatial or snapshot methods. The method of snapshots is especially attractive for high-dimensional spatial datasets.

3) Incoherent noise in the data generally appears as high-order POD modes, provided that the noise level is lower than the signal level. The POD analysis can be used to practically remove the incoherent noise from the dataset by simply removing high-order modes from the expansion.

Fig. 6 POD analysis of turbulent flow over a NACA0012 airfoil at \(Re = 23,000\) and \(\alpha = 9\) deg. Shown are the instantaneous and time-averaged streamwise velocity fields and the associated four most dominant POD modes [73,74]. Reprinted with permission from Springer.
4) The POD (PCA) analysis is very widely used in a broad spectrum of studies. It is used for pattern recognition, image processing, and data compression.

Weaknesses:
1) As the POD is based on second-order correlation, higher-order correlations are ignored.
2) The temporal coefficients of spatial POD modes generally contain a mix of frequencies. The spectral POD discussed previously addresses this issue.
3) The POD arranges modes in the order of energy contents and not in the order of the dynamical importance. This point is addressed by balanced POD and DMD analyses.
4) It is not always clear how many POD modes should be kept, and there are many different truncation criteria.

B. Illustrative Examples
1. Turbulent Separated Flow over an Airfoil

We present an example of applying the POD analysis on the velocity field obtained from three-dimensional large-eddy simulation (LES) of turbulent separated flow over a NACA 0012 airfoil [72]. The flow is incompressible with spanwise periodicity at $Re = 23,000$ and $\alpha = 9^\circ$ deg. Visualized in Fig. 6 (left) are the instantaneous and time-averaged streamwise velocities on a spanwise slice. We can observe that there are large-scale vortical structures in the wake from von Kármán shedding, yielding spatial and temporal fluctuations about the mean flow. Also present are the finer-scale turbulent structures in the flow. Performing the POD on the flowfield data, we can find the dominant modes [73,74]. Here, the first four dominant POD modes with the percentage of kinetic energy held by the modes are shown in Fig. 6 (middle and right). The four modes shown together capture approximately 19% of the unsteady fluctuations over the examined domain. Modes 1 and 2 (first pair) represent the most dominant fluctuations in the flowfield, possessing equal levels of kinetic energy, and amounting to oscillatory (periodic) modes. Modes 3 and 4 (second pair) represent the subharmonic spatial structures of modes 1 and 2 in this example. Compared to the laminar case shown in Fig. 1, the number of modes required to reconstruct this turbulent flow is increased due to the emergence of multiple spatial scales and higher-dimensionality of the turbulent flow. The dominant features of the shown POD modes share similarities with the laminar flow example shown in Fig. 1, despite the large difference in the Reynolds numbers.

2. Compressible Open-Cavity Flows

As the second sample application of the POD, let us look at an analysis of the velocity field from open-cavity flow experiments [75]. In this study, the snapshot POD was applied to two-component particle image velocimetry data acquired with several different freestream Mach numbers from 0.2 through 0.73. Here, the dominant mode (mode 1) contains between 15 and 20% of the energy, with 50% represented by the first seven modes. This study reveals the similarity of the modes among four of the freestream Mach numbers investigated, even though there were differences in the mean flow patterns. The first five POD modes associated with the vertical velocity component are presented in Fig. 7. In this figure, we can observe a representation of the vortical structures in the cavity shear layer with similar wavelengths, regardless of the freestream Mach number. Further quantitative analysis of the similarity of the modes was verified by checking the orthogonality between the modes for the various applications. The similarity among the modes implies that the underlying turbulence had the same structure, at least over the range of freestream Mach numbers investigated, regardless of the mean flow differences.

C. Outlook

The POD has been the bedrock of modal decomposition techniques to extract coherent structures for unsteady fluid flows.

![Fig. 7 POD modes of the vertical fluctuating velocity for flows over a rectangular cavity at Mach numbers of $M_\infty = 0.19$ and 0.29 [75]. Reprinted with permission from AIP Publishing.](http://arc.aiaa.org/DOI: 10.2514/1.J056060)
To address some of the shortcomings of the standard POD analysis, many variations have emerged: namely, the balanced POD [76] (see Sec. IV), the split POD [27], the sequential POD [78,79], and the joint POD [81], among others. A number of overlapping studies have emerged to bridge the gap between the POD and other decomposition methods, revisiting some of the early POD discussions by Lumley [27] and George [63]. Recently, theoretical connections have been made between the spectral POD and several other methods, including resolvent analysis (Sec. VIII) [69,82,83] and other data-based methods, including the spatial POD described earlier and dynamic mode decomposition (Sec. V) [69].

One of the most attractive properties of POD modes is orthogonality. This feature allows us to develop models that are low in order and sparse. Taking advantage of such a property, there have been efforts to construct reduced-order models based on Galerkin projection to capture the essential flow physics [1,48,55–57] and to implement model-based closed-loop flow control [54,84]. More recently, there are emerging approaches that take advantage of the POD modes to model and control fluid flows, leveraging cluster-based analysis [85] and networked-oscillator representation [22] of complex unsteady flow systems.

IV. Balanced Proper Orthogonal Decomposition

Balanced proper orthogonal decomposition is a modal decomposition technique that can extract two sets of modes for specified inputs and outputs. Here, the inputs are typically external disturbances or actuation used for flow control. The outputs are typically the available sensor measurements or the quantities we want to capture with a model (for instance, they could be amplitudes of the POD modes).

This method is an approximation of a technique called balanced truncation [86], which is a standard method used in control theory that balances the properties of controllability and observability. The most controllable states correspond to those that are most easily excited by the inputs, and the most observable states correspond to those that excite large future outputs. In a reduced-order model, we wish to retain both the most controllable modes and the most observable modes; but, the difficulty is that, for some systems (particularly systems that are nonnormal, which arise in many shear flows), states that have very small controllability might have very large observability, and vice versa.Balancing involves determining a coordinate system in which the most controllable directions in state space are also the most observable directions. We then truncate the states that are the least controllable/observable.

Balanced POD is closely related to the POD: both procedures produce a set of modes that describe the coherent structures in a given fluid flow, and the computations required are similar (both involve the SVD). However, there are some important differences. The POD provides a single set of modes that is orthogonal and ranked by energy content. In contrast, the balanced POD provides two sets of modes (balancing modes and adjoint modes), which form a biorthogonal set and are ranked by controllability/observability (which we can think of as the importance to the input–output dynamics). With both the POD and balanced POD, a quantity $q(ξ, t)$ is expanded as

$$q(ξ, t) = \sum_{j=1}^{n} a_j(ξ) φ_j(ξ)$$

(38)

where $φ_j$ are the POD modes or direct balancing modes, and $a_j(t)$ are scalar temporal coefficients. For the POD, the modes $φ_j$ are orthonormal (which means $⟨φ_j, φ_k⟩ = δ_{jk}$), so the coefficients $a_j$ are computed by $a_j(t) = (q, φ_j)$. With balanced POD, biorthogonality of the balancing modes $φ_j$ and adjoint modes $ψ_j$ means that these satisfy $⟨φ_j, ψ_k⟩ = δ_{jk}$, and the coefficients $a_j$ are then determined by $a_j(t) = (q, ψ_j)$.

The dataset used for the balanced POD is also quite specific: it consists of the linear responses of the system to impulse inputs (one time series for each input), as well as impulse responses of an adjoint system (one adjoint response for each output). It is these adjoint simulations that enable the balanced POD to determine the observability (or sensitivity) of different states, which makes the procedure so effective for nonnormal systems. However, because adjoint information is required, it is usually not possible to apply the balanced POD to experimental data. It has been shown that a system identification method called the eigen-system realization algorithm (ERA) [87] produces reduced-order models that are equivalent to balanced POD-based models, without the need for adjoint responses, and can therefore be used on experimental data [88]. For the full details of the balanced POD, see the work of Rowley [76] or the second edition of the book by Holmes et al. ([1] Chap. 5). A description of a related method was also given in the work of Wilcox and Peraire [89].

A. Description

1. Algorithm

As the balanced POD analysis is founded on linear state-space systems, the inputs to the balanced POD should be obtained from linear dynamics.

The inputs are two sets of snapshots from a linearized forward simulation and a companion adjoint simulation.

The outputs are sets of balancing modes and adjoint modes ranked in the order of the Hankel singular values. These modes comprise a coordinate transform that balances the controllability and observability of the system.

The balanced POD is based on the concept of balanced truncation that provides a balancing measure between controllability and observability in the transformed coordinate. This approach seeks for the balancing transform $Φ$ and its inverse transform $Ψ$ that can diagonalize and equate the (empirical) controllability and observability Gramians, $W_c$ and $W_ω$, respectively, such that $ΨW_cΨ^T = W_ωΦ^TΦ = Σ$ is a diagonal matrix. Although the controllability Gramian can be determined from the forward simulation, the observability Gramian requires results from the adjoint simulation.

Now, let us introduce the forward and adjoint linear systems and present the snapshot-based balanced POD technique [76] that is analogous to the snapshot-based POD method. Following standard state-space notation from control theory, the forward and adjoint simulations solve

$$\begin{cases} \dot{x} = Ax + Bu \\ y = Cx \end{cases}$$

(39)

and

$$\begin{cases} \dot{z} = A^Tz + C^Tv \\ w = B^tz \end{cases}$$

(40)

For the forward dynamics [Eq. (39)], $x(t) ∈ \mathbb{R}^n$, $u(t) ∈ \mathbb{R}^p$, and $y(t) ∈ \mathbb{R}^q$ are the state, input, and output vectors; and $A ∈ \mathbb{R}^{nxn}$, $B ∈ \mathbb{R}^{nxp}$, and $C ∈ \mathbb{R}^{qxn}$ are the state, input, and output matrices, respectively. For the adjoint system [Eq. (40)], $z(t) ∈ \mathbb{R}^n$, $v(t) ∈ \mathbb{R}^p$, and $w(t) ∈ \mathbb{R}^q$ are the adjoint state, input, and output vectors.

Based on the solution of linear and adjoint simulations, we can construct the data matrices $X$ and $Z$ (also see the Appendix). For simplicity, let us consider in the following a single-input single-output system with $p = q = 1$, for which we can construct the data matrices as

$$X = [x(t_1) \ x(t_2) \ … \ x(t_m)] ∈ \mathbb{R}^{nxm}$$

(41)

and

$$Z = [z(t_1) \ z(t_2) \ … \ z(t_m)] ∈ \mathbb{R}^{nxm}$$

(42)
For a multi-input multioutput system, the solutions from the linear and adjoint simulations can be stacked in an analogous manner to construct $X$ and $Z$, as discussed by Rowley [76]. Consequently, the empirical controllability and observability Gramians are given by $W_c \approx XX^T$ and $W_o \approx ZZ^T$, respectively. The balancing transform $\Phi$ and its inverse $\Psi$ can simply be found as

$$\Phi = XV\Sigma^{-1/2} \quad \text{and} \quad \Psi = ZU\Sigma^{-1/2}$$

(43)

where $U$, $\Sigma$, and $V$ are determined from the SVD of the matrix product $Z^TX$:

$$Z^TX = U\Sigma V^T$$

(44)

The columns $\phi_j$ of $\Phi$ and $\psi_j$ of $\Psi$ correspond to the balancing and adjoint modes ranked in the order of $\Sigma = \text{diag}(\sigma_1, \sigma_2, \ldots, \sigma_m)$, which are referred to as the Hankel singular values.

2. Notes

**Biorthogonality.** The property of biorthogonality $(\phi_j, \psi_k) = \delta_{jk}$ can be used as a projection to derive a reduced-order model known as the Petrov–Galerkin model [90].

**Eigenvalue Realization Algorithm.** If we are only interested in deriving a reduced-order model based on balanced truncation without the need to access the balancing and adjoint modes, we can use the ERA and remove the requirement to perform the adjoint simulation. For additional details on the derivation of the models and the ERA, the readers can refer to the work of Juang and Pappa [87] and Ma et al. [88].

3. Strengths and Weaknesses

**Strengths.**

1) The balanced POD is particularly attractive for capturing the dynamics of nonnormal systems with large transient growth [90]. Because the POD ranks the modes based on energy content, nonnormal characteristics of the flow may not be captured. In contrast, with the balanced POD models capture small energy perturbations that are highly observable (typically through the adjoint modes). For this reason, balanced POD models typically perform better than POD models for nonnormal systems.

2) The balanced POD provides an input–output model suitable for feedback control [90].

**Weaknesses.**

1) Snapshots from adjoint simulations are needed, which makes the balanced POD analysis difficult or impossible to perform with experimental measurements. However, the ERA can be used instead if only the balanced-POD-based model is sought for without access to the balancing and adjoint modes [87,88].

2) Both forward and adjoint simulations should be based on linear dynamics (although various extensions to nonlinear systems have been introduced [91,92]).

B. Illustrative Example

**Control of Wake Behind a Flat-Plate Wing.** The balanced POD has been applied to analyze and control the unsteady wake behind a flat-plate wing at $Re = 100$. In the work of Ahuja and Rowley [93], they performed linearized forward and adjoint simulations of flow over a flat-plate wing to determine balancing and adjoint modes, as shown in Fig. 8 (left). The balancing modes resemble those of the traditional POD modes, but the adjoint modes highlight regions of the flow that can trigger large perturbations downstream. These modes are used to develop models and closed-loop controllers to stabilize the naturally unstable fluid flows, as depicted in Fig. 8 (right). Although their work necessitated snapshots from the forward and adjoint simulations, the requirement for adjoint simulations was later removed by the use of the ERA in the work by Ma et al. [88]. Although the balancing and adjoint modes were not revealed from the ERA, the resulting reduced-order model was shown to be identical to the balanced-POD-based model. This was numerically demonstrated using the same flat-plate wing problem along with a successful implementation of observer-based feedback control.

C. Outlook

Although the balanced POD has been successfully applied to model and control a number of fluid systems [90–93,96], further research could validate and extend its use to a wider range of fluids applications. Work toward this goal currently includes application to the control of nonlinear systems [97], direct application to unstable systems [98], and use with harmonically forced data [99]. Further work could also seek algorithmic variants that improve efficiency. Examples of work in this direction include analytic treatment of impulse response tails [100] and the use of randomized methods [101,102]. Although, at present, the need for adjoint simulations makes the balanced POD unsuitable for experimental data, future work could possibly remove this restriction by making use of the connections that the balanced POD shares with methods such as the ERA and DMD.

V. Dynamic Mode Decomposition

Dynamic mode decomposition provides a means to decompose time-resolved data into modes, with each mode having a single characteristic frequency of oscillation and growth/decay rate. DMD is based on the eigendecomposition of a best-fit linear operator that approximates the dynamics present in the data. This technique was first introduced to the fluids community in an American Physical

---

Fig. 8 Use of balanced POD analysis for feedback stabilization of the unstable wake behind a flat plate ($Re = 100, \alpha = 35$ deg) [93]. First and third balancing and adjoint modes are shown on the left. Baseline lift history is shown with a dashed line, and controlled cases with different initiation times of feedback control using the balanced-POD-based reduced-order model are plotted in solid lines on the right. Reprinted with permission from Cambridge University Press.
The matrix $\mu$ and find the eigenvalues and eigenvectors of $X$ given by $A$. The DMD eigenvalues and modes are then defined as the eigenvalues and eigenvectors of $X$. In this case, it is not efficient to compute $A$ explicitly, so we normally use an algorithm such as the pseudoinverse of $X$.

In many ways, DMD may be viewed as combining favorable aspects of both the POD and the discrete Fourier transform. It has been widely applied beyond fluid dynamics: in finance [109], video processing [110–112], epidemiology [113], robotics [114], and neuroscience [115].

2. Notes

Computationally Efficient Algorithms. A fast method to perform DMD in real time on large datasets was recently proposed by Hemati et al. [116]. A library of tools for computing variants of DMD is available [14]. A parallelized implementation of DMD (as well as other system identification/modal decomposition techniques) was described by Belson et al. [117].

Sparsity. It is often desirable to represent a dataset sparsely, i.e., in terms of a small number of DMD modes. Because DMD modes are not orthogonal and have no objective ranking (as POD modes have), this is not an easy task. A number of variants of DMD has been proposed to provide such a sparse representation: for instance, the optimized DMD [108] and the optimal mode decomposition [118] are two such variants.

Connections with Other Methods. In many situations, DMD is equivalent to a discrete Fourier transform [105,108]. In addition, DMD shares algorithmic similarities with a number of other techniques, such as the eigensystem realization algorithm [107] and linear inverse modeling [104,107] (a technique used in climate science). If DMD is applied to linearized flow about its steady state, the extracted DMD modes capture the global modes (also see Sec. VII).

Connections with Other Methods. In many situations, DMD is equivalent to a discrete Fourier transform [105,108]. In addition, DMD shares algorithmic similarities with a number of other techniques, such as the eigensystem realization algorithm [107] and linear inverse modeling [104,107] (a technique used in climate science). If DMD is applied to linearized flow about its steady state, the extracted DMD modes capture the global modes (also see Sec. VII).

A. Description

1. Algorithm

The inputs are a set of snapshot pairs from fluids experiments or simulations, where the two snapshots in each pair are separated by a constant interval of time. Often, this will come from a time series of data.

The outputs are DMD eigenvalues and modes. The modes are spatial structures that oscillate and/or grow/decay at rates given by the corresponding eigenvalues. These come from the eigendecomposition of a best-fit linear operator that approximates the dynamics present in the data.

We begin by collecting snapshots of data and arranging them as columns of matrices $X$ and $X$, such that

$$X = [x(1), x(2), \ldots, x(m)] \in \mathbb{R}^{n \times m}$$

In DMD, we approximate the relationship between the snapshots in a linear manner such that

$$X = AX$$

The matrix $A$ may be defined by $A = XX^+$, where $X^+$ denotes the pseudoinverse of $X$. The DMD eigenvalues and modes are then defined as the eigenvalues and eigenvectors of $A$ [107]. It is common that the number of snapshots is smaller than the number of components of each snapshot ($m \ll n$). In this case, it is not efficient to compute $A$ explicitly, so we normally use an algorithm such as the following, which is similar to algorithm 2 in the work of Tu et al. [107]:

1. Perform the reduced SVD (Sec. II.B) of $X$, letting $X = U \Sigma V^T$.

2. Truncate the SVD by considering only the first $r$ columns of $U$ and $V$, as well as the first $r$ rows and columns of $\Sigma$ (with the singular values ordered by size), to obtain $U_r$, $\Sigma_r$, and $V_r$. Note that this is optional.

3. Let $\hat{A} = U_r^T A U_r = U_r^T X V_r \Sigma_r^{-1} \in \mathbb{R}^{r \times r}$.

4. Find the eigenvalues $\lambda_j$ and eigenvectors $\tilde{v}_j$ of $\hat{A}$, with $\hat{A} \tilde{v}_j = \lambda_j \tilde{v}_j$.

5. Every nonzero $\lambda_j$ is a DMD eigenvalue, with the corresponding DMD mode given by

$$v_j = \mu_j^{-1} X \Sigma_r^{-1} \tilde{v}_j$$

It is common to compute the projected DMD modes, which are simply $P v_j = U_j \tilde{v}_j$, where $P = U \Sigma^T$ is the orthogonal projection onto the first $r$ POD modes of the data in $X$.

Note that matrix $A$ in Eq. (46) is related to operator $exp(\Delta t)$ in Eq. (5), with $\Delta t = t_{i+1} - t_i$. Hence, the eigenvalues are related by

$$\lambda_j = \frac{1}{\Delta t} \log(\mu_j)$$

Using this relationship, the growth/decay rates and frequencies of the DMD modes can be inferred by examining the real and imaginary components of $\lambda_j$.

In addition to the DMD algorithms presented, there are a number of variants. Some of them are discussed below in the Notes. For further details, refer to the work of Rowley and Dawson [2], Tu et al. [107], and Kutz et al. [4].
3. Strengths and Weaknesses

Strengths.
1) DMD does not require any a priori assumptions or knowledge of the underlying dynamics. DMD is an entirely data-driven analysis.
2) DMD can be applied to many types of data, or even concatenations of disparate data sources.
3) Under certain conditions, DMD gives a finite-dimensional approximation to the Koopman operator, which is an infinite-dimensional linear operator that can be used to describe nonlinear dynamics (see Sec. VI).
4) DMD modes can isolate specific dynamic structures (associated with a particular frequency).
5) DMD has proven to be quite customizable, in the sense that a number of proposed modifications address the weaknesses outlined in the following.

Weaknesses.
1) It can be difficult (or at least subjective) to determine which modes are the most physically relevant (i.e., there is no single correct way to rank eigenvalue importance, unlike other methods such as POD).
2) DMD typically requires time-resolved data to identify dynamics, although extensions exist [107,120].
3) If DMD is used for system identification (without any modifications), the resulting model will be linear.
4) DMD can be unreliable for nonlinear systems. In particular, for a nonlinear system, we must be careful to choose a sufficiently rich set of measurements (in each snapshot). Without care, the connection with the Koopman operator and the underlying dynamical system may be lost. Furthermore, for nonlinear systems with complex (e.g., chaotic) dynamics, there are further complications that could limit the applicability of DMD and related algorithms.
5) The outputs of DMD can be sensitive to noisy data, which were shown empirically [127] and analytically [128]. The effect of process noise (that is, a disturbance that affects the dynamics of the system) has also been investigated [129]. However, there are algorithms that are robust to sensor noise [128,130].
6) DMD should generally be used only for autonomous systems (i.e., the governing equations should have no time dependence or external inputs, unless these are explicitly accounted for [124]).
7) DMD modes are not orthogonal. This has a number of drawbacks: for instance, if the modes are used as a basis/coordinate system for a reduced-order model, the model will have additional terms due to the spatial inner product between different modes being nonzero. Note that a recent variant, recursive DMD [131], considers orthogonalized DMD modes.
8) DMD relies fundamentally on the separation of variables, as does POD, and hence does not readily extend to traveling wave problems.
9) DMD does not typically work well for systems with highly intermittent dynamics. However, multiresolution [132] and time-delay [133] variants show promise for overcoming this weakness.

B. Illustrative Example

1. Jet in Crossflow

We show an example from Rowley et al. [105], where DMD is applied to three-dimensional jet-in-crossflow direct numerical simulation (DNS) data. A typical snapshot of the complex flow is visualized in Fig. 9 (top) using the $\lambda_2$ criterion. The results of applying DMD to a sequence of 251 snapshots are shown in Fig. 9 (bottom). The bottom-left plot shows the frequency and amplitude of each DMD mode, with two modes having large amplitudes visualized in the bottom plots. Note that these modes capture very different flow structures, with each having a different characteristic frequency of oscillation identified by DMD.

2. Canonical Separated Flow with Control

Let us present a second example [134] of DMD (and POD) analysis performed on three-dimensional separated turbulent flow over a finite-thickness plate with an elliptical leading edge at $Re = 100,000$. The plate is aligned with the freestream, and separation is induced by imposing a steady blowing/suction boundary condition above the plate. The flowfield is obtained from LES, and the wake dynamics in the example is modified by a synthetic jet actuator placed on the top surface of the wing. First, we consider the dominant and secondary POD modes, shown in Fig. 10 (left). We observe that some interactions between the actuation input (at the 0.6 chord location) and the wake are captured by these...
POD modes

Dominant mode

Secondary mode

DMD modes

Fundamental mode

Secondary mode

Fig. 10 Comparison of POD and DMD modes for separated three-dimensional turbulent flow over a flat-plate wing with synthetic-jet actuation on the top surface [134]. The fundamental and secondary DMD modes correspond to the actuation frequencies of 4.40 (blue) and its superharmonic (red). Reprinted with permission from J. Tu.

energetically dominant modes. Also shown on the right are the DMD modes corresponding to the actuation frequency and its superharmonic. Although the fundamental DMD mode is similar to the dominant POD mode, the secondary mode is able to clearly identify the synchronization of the actuator input with the downstream wake. In contrast, POD modes comprise spatial structures having a distribution of temporal frequencies that makes pinpointing the POD mode to a specific frequency difficult.

C. Outlook

Although DMD has quickly become a widely used method for analyzing fluid flow data, there remain challenges and applications that have yet to be fully addressed. Connections between DMD and the Koopman operator indicate its potential to model (and control) nonlinear systems. However, choosing suitable observables to give an accurate finite-dimensional approximation to the Koopman operator generally remains an open question. Algorithmic improvements that have been and should continue to be made will allow DMD to remain a practical tool to analyze increasingly large fluid flow datasets.

VI. Koopman Analysis

Koopman analysis provides an important alternative perspective to classical dynamical systems theory [135] for the description of complex systems. The Koopman operator was introduced in the early 1930s [136] to show how the dynamics of Hamiltonian systems could be described by an infinite-dimensional linear operator on the space of observable functions of the state of the system. Recently, this theory has been at the center of efforts for the data-driven characterization of complex systems [137]. There is particular interest in obtaining finite-rank approximations to the linear Koopman operator. Practically, dynamic mode decomposition (Sec. V) is the most implemented numerical framework for Koopman mode decomposition in fluids [103–105]. There have been a number of excellent in-depth reviews on Koopman analysis recently [106,138].

Koopman analysis not only provides a set of modes, of which DMD modes are a subset, but also a set of eigenvalues that determine the modal dynamics and a set of Koopman eigenfunctions that serve as intrinsic observable functions. In many contexts, Koopman analysis provides an equation-free method [139] to extract coherent structures and dynamics from data measurements of a complex system; these coherent structures are related to POD modes in fluids [1].

A. Description

1. Algorithm

   The inputs are a nonlinear dynamical system:
   \[ x_{i+1} = f(x_i) \]  
   \[(50)\]

   The state \( x \) typically lives in a vector space, such as \( \mathbb{R}^n \) or \( \mathbb{C}^n \), although the theory is defined more generally on curved manifolds. Although the aforementioned dynamical system is written in discrete time, the theory also holds for continuous dynamical systems.

   The output is the infinite-dimensional linear operator \( U_t \) that describes the evolution of scalar observables \( g(x) \) on state space:
   \[ U_t g = g \circ f \]
   \[(51)\]

   Here, \( \circ \) denotes function composition. This operator advances all functions in the Hilbert space of measurement functions, and it holds for all states \( x \). An implication is that measurements are advanced in time:
   \[ U_t g(x_i) = g \circ f(x_i) = g(x_{i+1}) \]  
   \[(52)\]

   An alternative description, related to DMD, is as follows:
   The inputs are data snapshots of observables on the state space of a dynamical system.

   The outputs are the modal decomposition and linear dynamical system describing modal evolution (see Sec. V; DMD).

2. Notes

   **Approximation by DMD.** A major applied goal of the Koopman analysis is the identification of eigenfunctions from data. These eigenfunctions provide an intrinsic coordinate system, along which the dynamics appear linear. The Koopman mode decomposition is usually approximated via the dynamic mode decomposition, described in Sec. V. DMD is essentially a linear regression of data onto dynamics, and augmenting the measurement data with nonlinear functions of the state may enable a more accurate approximation of the Koopman operator [125] and resulting eigenfunctions. Another promising algorithm computes Koopman eigenfunctions based on the diffusion operator [140].

   **Hamiltonian Systems.** For Hamiltonian systems, the Koopman operator is unitary, meaning that the inner product of any two observable functions remains the same before and after the operator. Unitarity is a familiar concept, as the discrete Fourier transform and the POD basis both provide unitary coordinate transformations. In the original paper of Koopman [136], connections were drawn between the Koopman eigenvalue spectrum and both conserved quantities and integrability.

3. Strengths and Weaknesses

   **Strengths:**

   1) The Koopman analysis provides an alternative operator-theoretic perspective to dynamical systems and allows nonlinear systems to be represented and analyzed using linear techniques.

   2) For example, note that \( \sin(x) \ast x^2 = \sin(x^2) \).
2) The Koopman operator is amenable to standard spectral decomposition in terms of eigenvalues and eigenvectors.

3) Powerful techniques in control theory apply to linear systems and may improve nonlinear control performance via Koopman linear systems [141].

**Weaknesses:**

1) By introducing the Koopman operator, we trade finite-dimensional nonlinear dynamics for infinite-dimensional linear dynamics. Dealing with infinite-dimensional operators and obtaining low-order approximations is challenging, although there are a larger number of computational techniques to analyze linear operators than there are for differential geometry on manifolds.

2) It may be difficult to find Koopman eigenfunctions, which define an intrinsic observable measurement coordinate system. Without Koopman eigenfunctions, it may be difficult or impossible to obtain a finite-dimensional subspace of the Hilbert space of measurement functions that remains closed under the Koopman operator. However, if discovered, a Koopman invariant subspace defines a measurement system where measurements are propagated by a finite-dimensional linear dynamical system.

3) The integration of control theory and Koopman analysis remains to be fully developed.

**B. Illustrative Example**

**Simple Nonlinear Dynamical System.** As an example, consider the following simple nonlinear dynamical system in two variables with a single fixed point at the origin [107,142]:

\[
\begin{align*}
\dot{x}_1 &= \mu x_1, \\
\dot{x}_2 &= \lambda (x_2 - x_1^2)
\end{align*}
\]

(53)

For \( \lambda \ll \mu < 0 \), the fixed point at the origin is stable and there is a slow manifold given by \( x_2 = x_1^2 \); trajectories quickly attract onto this manifold before converging to the origin (Fig. 11a). Introducing a nonlinear change of coordinates, given by \((y_1, y_2, y_3) = (x_1, x_2, x_1^2)\), the nonlinear dynamics become linear:

\[
\begin{bmatrix}
\dot{y}_1 \\
\dot{y}_2 \\
\dot{y}_3
\end{bmatrix} =
\begin{bmatrix}
\mu & 0 & 0 \\
0 & \lambda & -\lambda \\
0 & 0 & 2\mu
\end{bmatrix}
\begin{bmatrix}
y_1 \\
y_2 \\
y_3
\end{bmatrix}
\]

(54)

The nonlinear system in Eq. (53) and the linear Koopman system in Eq. (54) are shown in Fig. 11. The system in Eq. (54) is defined on a manifold before converging to the origin (Fig. 11a). Introducing a nonlinear change of coordinates, given by \((y_1, y_2, y_3) = (x_1, x_2, x_1^2)\), the nonlinear dynamics become linear:

The preceding example may appear exceedingly simple, but there are a few direct applications of the Koopman analysis for more complex systems because it is challenging to discover the coordinate transformations that linearize the problem. In fluid dynamics, Bagheri [143] analyzed the Koopman mode decomposition of the flow past a cylinder, relating the decomposition to a POD–Galerkin with a shift mode [56]. Related work investigated the Liouville equation [144], which resulted in cluster-based reduced-order models in fluid systems [85]. In power electronics, Koopman mode decomposition has been used to model and predict instabilities [145,146]. However, DMD is used in most fluid applications to approximate the Koopman mode decomposition, although the quality of this approximation depends on the measurements used for DMD [4]. True Koopman linearization of complex systems in fluid dynamics relies on the choice of good observable functions that provide linear embeddings for the nonlinear dynamics, as in the preceding example. The extended DMD [125] provides one approach to augment DMD with nonlinear measurements of the state, potentially improving the approximation of Koopman eigenfunctions.

**C. Outlook**

The promise of Koopman analysis hinges on the discovery of good observable functions that provide a coordinate transformation in which measurements behave linearly. Identifying these Koopman eigenfunctions from data is both a major goal and central challenge moving forward, making it a focus of research efforts. Dynamic mode decomposition, which is a workhorse of Koopman analysis, implicitly uses linear observable functions. In fluid flows, these often take the form of direct velocity field measurements from numerical simulations or particle image velocimetry in experiments. In other words, the observable function is an identity map on the fluid flow state. This set of linear observables is often too limited to describe the rich dynamics observed in fluid systems. Recently, DMD has been extended to include a richer set of nonlinear observable functions, providing the ability to effectively analyze nonlinear systems [125]. To avoid overfitting, sparse regression is emerging as a principled technique to select active terms in the dynamical system [147], and similar techniques may be used to improve the identification of observable functions. Recent techniques also construct Koopman eigenfunctions using a regularized advection–diffusion operator [140] or approximate the Koopman operator in delay coordinates [133,148]. Discovering Koopman eigenfunctions is a central computational issue to obtain closed systems that may be used for nonlinear control and estimation, and the ultimate success of this
analysis will depend on our ability to accurately and efficiently approximate these eigenfunctions from data. It is likely that developments in machine learning, such as deep learning, will continue to advance these capabilities.

VII. Global Linear Stability Analysis

Global linear stability analysis solves the linearized Navier–Stokes equations pertinent to an underlying base flow with multiple inhomogeneous spatial directions. The introduction of a modal formulation for the linear perturbations converts the linearized equations of motion (initial-value problem) to an eigenvalue problem with respect to the base flow, which needs to be an exact steady (or unsteady) solution of the equations of fluid motion. The term global is used to distinguish the analysis from the classic local linear stability theory [6,34,149], where the base flow is independent of two coordinate directions. The eigenvalue problem delivers information on the spectrum of the linearized Navier–Stokes operator at asymptotically large times. Much like in classic local linear stability theory, the eigenvalue determines whether the corresponding eigenvector grows (exponentially) in time or space and delivers information on the frequency of the linear perturbation (Sec. II and Fig. 3). The eigenvector describes the spatial structure of the linear perturbation and may be used to reconstruct the full flowfield at conditions consistent with the linear approximation.

Global modes are solutions to the two- or three-dimensional partial derivative eigenvalue problems to which the linearized Navier–Stokes equations may be recast. They describe disturbances developing upon base flows that vary inhomogeneously in two (biglobal) or three (triglobal) spatial directions [150]. Thus, global modes are the counterparts, in base flows that vary in multiple spatial dimensions, of well-known linear instabilities developing in one-dimensional base flows such as the inviscid Kelvin–Helmholtz instabilities, viscous Tollmien–Schlichting, and crossflow modes. When the two- or three-dimensional base flow contains portions of flow that could be analyzed with local theory, the spatial distribution of the global eigenfunction is seen to contain structures related to the eigenfunction delivered by local analysis, e.g., Tollmien–Schlichting waves in the boundary layer on a flat plate [151] and on the downstream wall of an open cavity [152]. The applications of global stability analysis to study linear stability of two- and three-dimensional base flows have been reviewed recently [5] and are continuously expanding.

A. Description

1. Algorithm

The input is the steady or time-periodic laminar base flow, inhomogeneous in two or all three spatial directions, at a given Reynolds number and Mach number. The base flow can be a steady state or an unstable steady state. For the outputs, the global linear modal stability theory delivers global modes (spatial distribution pattern) and the associated growth rates and frequencies of small-amplitude perturbations about the base state.

Unlike the aforementioned POD, balanced POD, or DMD techniques, global stability analysis is not based on the snapshots of the flowfield. Global modes are found numerically by discretizing the Navier–Stokes equations, which result from the decomposition of any flow quantity $q$ into a base flow $q_0$ and small-amplitude perturbations $q'$ (i.e., $||q'||/||q_0|| \ll 1$):

$$q(\xi, \eta, \tau) = q_0(\xi, \eta) + q'(\xi, \eta)$$  \hspace{1cm} (55)

We can then examine the growth or decay of the perturbations with respect to the base state $q_0$. Substitution of the flow variable in the form of Eq. (55) into the Navier–Stokes equations and neglecting the quadratic terms of small perturbation $q'$ yields the linearized Navier–Stokes equations.

Now, as an example, let us assume the base flow $q_0 = q_0(\xi)$ to be steady and homogeneous (e.g., periodic in space) in the $\zeta$ direction and perform a biglobal stability analysis. In this case, the perturbation can assume the form of

$$q(\xi, \eta, \zeta, t) = \hat{q}(\xi, \eta) e^{i(\beta \zeta - \omega t)} + \text{complex conjugate}$$  \hspace{1cm} (56)

where $\beta$ is a real wave number in the $\zeta$ direction. The real and imaginary components of the complex number $\omega = \omega_r + i \omega_i$, respectively, correspond to the frequency and the growth/decay rate of the amplitude function of the global mode $\hat{q}(\xi, \eta)$. There are two main approaches to find the global stability modes and the associated eigenvalues. Namely, they are the matrix-based approach and the time-stepping approach, which are described in the following.

Various techniques have also been summarized in the review by Theofilis [5], where the relative merits of matrix-based and time-stepping techniques were discussed.

Matrix-Based Approach:

The matrix-based approach determines the global modes by solving a generalized eigenvalue problem. The substitution of the assumed form of Eq. (56) into the linearized Navier–Stokes equations results in

$$A(q_0, \beta)\hat{q} = \omega B\hat{q}$$  \hspace{1cm} (57)

which is written in terms of a generalized eigenvalue problem where $\omega$ is the eigenvalue and $\hat{q}$ is the global stability mode (eigenvector). The matrix $A$ is dependent on the base state $q_0$, as well as the wave number $\beta$ and flow parameters, such as the Reynolds number and Mach number. In general, $B$ is invertible for compressible flow but is not invertible for incompressible flow due to the incompressibility constraint. For detailed discussions on the linearized Navier–Stokes equations, see the work of Schmid and Henningson [6] and Theofilis [5].

The matrices $A$ and $B$ are approximately of size $n_{\text{grid}} \times n_{\text{var}}$, where $n_{\text{grid}}$ and $n_{\text{var}}$ are the number of grid points and number of variables, respectively. If the operators are of modest size, the matrices may be formed and stored on memory. For larger-sized operators, the eigenvalues can be determined without storing the operators but by only relying on matrix–vector operations (matrix-free approach) [5,153–155]. We can also distribute the matrix entries over several processors on a cluster and use appropriate libraries (e.g., ScalAPACK) for the linear algebra operations [156].

Time-Stepping Approach:

In the time-stepping approach [5,157–159], a code is employed to compute the linear operators of the Navier–Stokes equations (Fréchet derivative [160]), without the need to store the matrix. Because, typically, the numerical discretization of the operators does not expand the solution on bases that automatically satisfy the boundary conditions for the perturbations, spurious modes can be introduced. We need to extract the physically meaningful modes from the numerically obtained part of the eigenspectrum.

2. Notes

Computational Effort: With either the matrix-based approach or the time-stepping approach, the computational effort required to find eigenvalues and eigenfunctions can be very high. Hence, the use of sparse yet high-order spatial discretization methods can be beneficial [161,162]. If the flow of interest has symmetry, it can be exploited to reduce the computational cost of solving two- and three-dimensional eigenvalue problems [159,163,164].

Choice of Base Flow: For stable flows, the steady-state solution to the Navier–Stokes equations can be found by integrating the solution in time until all unsteadiness is eliminated. If the unstable steady state is needed, techniques such as the selective frequency damping method [165] or a Newton–Krylov-type iterative solver [166,167] can be used. We note that time-periodic flow can also be used as the base flow [168], as we will see in the first illustrative example in the following [169].
Occasionally, researchers compute global modes for base flows that are not themselves solutions of the governing equations (e.g., the mean of a turbulent flow). In these cases, the resulting modes are not associated with a question of stability of the base flow but may be useful in modeling and identifying the frequency content of large-scale coherent structures [154,170]. Additional details with a cautionary note were provided by Sipp and Lebedev [171].

Global Nonmodal/Transient-Growth Analysis. Global modes may represent either stable or unstable disturbances with respect to the base flow. When unstable modes are identified, the base flow itself is said to be asymptotically unstable, meaning that at least one perturbation will grow exponentially in time and the equilibrium is not expected to be observed in nature. As discussed in Sec. I.E, systems that are nonnormal can exhibit significant growth of linear combinations of modal perturbations, known as transient growth, which is a phenomenon that may occur even when the global modes are all stable [33,36,172]. Nonnormality of the governing linear operator implies that global transient-growth analysis may deliver qualitatively different short-time behavior of linear combinations of perturbations when modal analysis predicts only asymptotically stable perturbations. Global transient-growth analysis has unraveled initial optimal perturbations that are qualitatively different from their corresponding global eigenmodes on the cylinder [173], low-pressure turbine blades [174,175], and several stalled NACA airfoils [169].

3. Strengths and Weaknesses

Strengths:
1) A spectrum of (discrete and continuous) eigenmodes can be determined, especially with the matrix-based approach.
2) Modal analysis, based on the solution of the eigenvalue problem (see Sec. II), determines whether unstable modes exist. Exponential growth implies that flow dynamics will be asymptotically dominated by the characteristics of the most unstable eigenmode (or the least stable mode if stable).
3) If only stable modes are found, global nonmodal/transient-growth analysis, using a linear combination of the leading members of the eigenvalue spectrum or a solution of a related SVD problem (see Sec. II), can determine the level of energy growth of small-amplitude perturbations over a short time horizon.

Weaknesses:
1) Global linear stability analysis is inherently linear.
2) The base flow provides the spatially variable coefficients of the underlying partial derivative eigenvalue/initial-value problem. Any algorithm for the solution of the latter problems can be influenced by the quality (accuracy) of the underlying base flow.
3) The analysis requires a base flow that is an exact solution of the equations of motion. Although counterexamples exist in both classic [176] and global linear stability theories [177], an analysis of mean turbulent flow requires validation of the turbulence closures employed and typically aims at the prediction of the frequencies and spatial structure of coherent turbulent structures.

B. Illustrative Examples

1. Large-Scale Separation Cells on Stalled Airfoils

He et al. [169] employed a suite of matrix-forming and time-stepping techniques to reexamine large-scale separation patterns on the suction side of wings in near-stall flight conditions. The origins of these structures were explained through systematic application of primary and secondary linear global eigenvalue and transient-growth stability analyses to massively separated spanwise homogeneous laminar flows over spanwise-periodic wings of different thicknesses and cambers. At low chord Reynolds numbers, the dominant flow structure arising from either primary modal or nonmodal mechanisms is associated with two- or three-dimensional Kelvin–Helmholtz eigenmodes. As the wing aspect ratio is shortened, the stationary three-dimensional global eigenmode discovered by Theofilis et al. [178] is amplified more than the Kelvin–Helmholtz mode and stall cells may arise through this primary modal linear amplification mechanism [179]. As the Reynolds number increases, linear amplification of the two-dimensional Kelvin–Helmholtz global eigenmode leads to a time-periodic wake that, in turn, is linearly unstable with respect to two distinct classes of three-dimensional secondary (Floquet) eigenmodes, which peak at short- and long-spanwise wavelengths, respectively. The short-wavelength Floquet eigenmode is the stronger of the two at moderate and high Reynolds numbers. The three-dimensional spatial structure of the short-wavelength mode can be seen in Fig. 12a, whereas the wall-streamline pattern resulting from its linear superposition upon the underlying time-periodic base flow gives rise to stall-cell-like patterns on the wing surface, as can be seen in Fig. 12b.

2. Finite-Span Open-Cavity Flow

The second example leverages the time-stepper-based approach to gain insights into the critical conditions and spatial characteristics of distinct classes of eigenmodes of incompressible flow over a rectangular three-dimensional lateral-wall-bounded open cavity. The analysis has been performed in a triglobal setting [159]. The leading traveling shear-layer mode responsible for transition in this geometry, as well as the next in significant stationary and traveling centrifugal modes, respectively, are shown in Fig. 13 at the slightly subcritical Reynolds number of Re = 1050. A detailed temporal biglobal analysis of the related spanwise homogeneous open cavity has also been performed [155,161], and the effect of the presence of lateral walls on the linear instability mechanisms in this class of flows can now be quantified on the basis of global stability theory.

C. Outlook

With the availability of enhanced computational resources, the use of temporal and spatial biglobal stability analyses is becoming prevalent to examine a range of fluid flows. An area in which the theory is expected to become particularly useful is laminar–turbulent transition prediction in hypersonic flow, in support of increasing efforts to understand and control flow phenomena critical to the design of next-generation vehicles. Unlike typical incompressible or
supersonic conditions at which turbulent flow prevails in flight, hypersonic flow is mostly laminar and the underlying base flows are exact solutions of the equations of motion, which can be computed either using a continuum assumption [180] or by employing direct simulation Monte Carlo methods [181–184]. As an example, a linear global instability analysis has recently addressed hypersonic flow over the HIFIRE-5 (Hypersonic International Flight Research Experimentation) elliptic cone [164,185]. The spatial biglobal eigenvalue problem has been solved, and four classes of (leading) centerline, attachment-line, crossflow, and second Mack modes have been recovered as two-dimensional global mode amplitude functions, underlying the potential of the theory to provide predictions of laminar–turbulent transition at hypersonic speeds.

In addition to taking advantage of the symmetry present in flows of interest, we can incorporate the parabolized assumption [186] into a global stability analysis. For instance, three-dimensional inhomogeneous base flows that depend strongly on two directions and weakly on the third spatial direction are an extension of the nonlocal stability analysis, which is based on the parabolized stability equations. Examples of successful application include the system of trailing vortices behind an aircraft wing [187–189], the wake of an isolated roughness element in supersonic flow [190], streaks in a boundary layer [191], corner flows [192], and models of duct intakes [193]. A computationally expensive triglobal stability analysis is emerging to provide insights into the influence of three-dimensionality on the stability of flows [157–159]. It is anticipated that there will be growing use of these approaches to analyze increasingly complex flows.

As a companion to the global stability analysis, the adjoint global stability analysis [194] can provide efficient responses to questions on flow receptivity, sensitivity, and control. Global modes are the right eigenvectors of the eigenvalue problem, whereas the left eigenvectors are the adjoint global modes. The adjoint linearized Navier–Stokes equations, which are at the heart of the linear instability analysis and linear control methodologies [194], can be solved with some added effort to the global eigenvalue problem. At its simplest implementation, a single solution of the global eigenvalue problem delivers the eigenvalues, as well as linear perturbations and their adjoint variables as the right and left eigenvectors, respectively.

VIII. Resolvent Analysis

The resolvent is defined in the context of fluid mechanics as the linear operator relating an input forcing to a linear system to the output. For the complete incompressible Navier–Stokes equations, the output is the (divergence-free) velocity field and the input to the resolvent (linear dynamics) is provided by the nonlinear term, as shown in Fig. 14. The resolvent analysis is therefore a complement to eigenvalue and global stability analyses, which are linear analyses with perturbations assumed to be sufficiently small such that the nonlinear advective forcing term (second order in the perturbation) can be neglected. More generally, inputs to the resolvent analysis can include actuation and other kinds of forcing terms added to the governing equations, and outputs can be any observable.

The resolvent analysis relies on the pseudospectrum of the linearized Navier–Stokes operator rather than the spectrum itself (see Sec. II.E). For stable base flows, the damped modes of the linear system can be superposed to construct the response to stationary (real-frequency) inputs. Even for these stable flows, when the operator is nonnormal (most fluid systems linearized about spatially nonuniform flows give rise to nonnormal operators), stationary inputs can be amplified. For these systems, there are typically a limited number of highly amplified inputs, implying that the resolvent can be effectively approximated at low rank. An SVD of the resolvent (in the discrete setting) can be used to identify these inputs and their corresponding outputs, as well as to provide a low-rank approximation of the input–output dynamics of the full system. High-gain inputs are useful for flow control efforts by showing where actuation will produce the largest effect.

![Fig. 14 Schematic of a divergence-free projection of the resolvent formulation of the incompressible Navier–Stokes equations, showing the nonlinear term f providing input forcing to the linear dynamics (described by the resolvent operator), with velocity output, u [i.e., q’ in Eq. (58)]](image-url)
A. Description

1. Algorithm

The inputs are a real-valued frequency $\omega$ and a linear time-invariant input–output system (operator) including its boundary conditions; for example, the linearized Navier–Stokes equations with an appropriate assumed base or mean velocity profile $\bar{q}$. These must typically have been discretized in space using a finite difference, finite element, spectral, or other numerical method.

The outputs are a ranked set of modes (total number equal to the number of degrees of freedom of the discretized system) that represent the spatial pattern of inputs, outputs, and the positive gain (amplification) from input to output. High-gain modes represent stationary (statistically steady) inputs that are most amplified at the specified frequency.

Consider $q = \bar{q} + q'$ and the spatially discretized governing equations for the perturbation variable expressed in the compact form

$$M \frac{dq'}{dt} + Aq' = f(q') \quad (58)$$

where the vector $q'$ represents the fluctuating discrete flow variables; the (typically sparse) matrices $M$ and $A = A(\bar{q})$ represent the linearization about a time-invariant flow $\bar{q}$; and $f$ denotes the remaining nonlinear terms, as well as any additional forcing terms added to the equations. Such terms can include, for example, stochastic forcing (synthetic turbulence) added to inflow regions of spatially developing flows. For statistically steady (stationary) flows, we may work in the frequency domain. By expressing

$$q' = \hat{q} e^{i\omega t} \quad \text{and} \quad f = \hat{f} e^{i\omega t} \quad (59)$$

for temporal frequency $\omega \in \mathbb{R}$, Eq. (58) can be written as

$$(i\omega M + A)\hat{q} = \hat{f} \quad (60)$$

Proceeding formally, we write $\hat{q} = R\hat{f}$, where $R = (i\omega M + A)^{-1}$ is the resolvent operator. For a particular frequency $\omega$, the maximum amplification $G$ (defined in a chosen norm $\|q\|_Q$, where $Q$ is a positive definite matrix) of any forcing may be found:

$$G = \max_j \frac{\hat{q}^* \hat{Q} \hat{q}}{\hat{f}^* \hat{Q} \hat{f}} = \max_j \frac{\int \hat{Q} \hat{R} \hat{Q} \hat{f}}{\int \hat{Q} \hat{f}} \quad (61)$$

The solution to this maximization problem can be found through the SVD of $R$. The largest singular value represents the (squared) gain, and the left and right singular vectors represent the optimal response and forcing, respectively.

In the context of the incompressible Navier–Stokes equations, the resolvent $R$ is the linear operator that relates velocity perturbation $u'$ and pressure perturbation $p'$ to the nonlinear term $f$:

$$q' = \begin{bmatrix} u' \\ p' \end{bmatrix} = Rf \quad (62)$$

where $L_A$ is the linear Navier–Stokes (advection and diffusion) operator evaluated at $\bar{k}$, $V$ and $V^T$ are the gradient and divergence operators, and $I$ is the identity operator. As such, the resolvent operator can be obtained by a relatively minor extension of code for the more common Orr–Sommerfeld–Squire analysis. Of course, the system may be formulated in terms of either velocity–vorticity or primitive variables. For an analogous derivation of the resolvent for compressible flow, see the work of Jeun et al. [196].

Two closely related formulations of the resolvent analysis are common in the literature. In the input–output formulation, an output quantity is defined as the state vector premultiplied by a matrix that is chosen to extract an observable of interest [196]. Similarly, the forcing is premultiplied by a (not necessarily) different matrix to restrict the input–output relation to forcings of interest. In the linear frequency response approach [197], the gain is defined as the quotient of the input and output in terms of their energy defined through (not necessarily different) norms.

For a given $\omega$, the SVD procedure described previously in Sec. II returns right singular vectors corresponding to the most dangerous (most amplified) inputs or forcing modes, and left singular vectors corresponding to the associated response modes, with gains given by the corresponding singular values. The example code for a resolvent analysis of turbulent pipe flow in primitive variables is available online for general use [198]. The analysis proceeds, depending on the specific objective. For example, the globally most amplified input can be sought, the forcing and response modes can be compared with observations, or data can be projected onto the resolvent basis [195,199].

2. Notes

Origin. The resolvent is a familiar construction in the study of forced, ordinary differential equations and, more generally, linear operators and their spectra. It arises in control theory and in eigenvalue/eigenvector perturbation analysis. Identified by Schmid and Henningson [6], its use as a tool to study structures in transitional and turbulent flows seems to have been stimulated by Farrell and Ioannou [200], who considered a stochastically forced Navier–Stokes system, and follows contemporaneous work on understanding transient growth in stable base flows with pseudospectra [35].

Choice of Base Flow. Similar to the linear global stability analysis discussed in the previous section, the choice of the base flow $q_0$ is important in resolvent analysis because operator $A$ and forcing input $f$ are established with respect to the base state $q_0$. If the (stable or unstable) steady-state solution to the Navier–Stokes equations is available, it can be used as the base flow [6,35] and the nonlinear term can be assumed to be negligible within the context of linear analysis. In this case, the input forcing must be provided from an external source. We also note that a resolvent analysis has been performed about the time-averaged flow $q$. In such a case, the nonlinear term is not negligible but is considered as an internal source of the forcing mechanism [195], as illustrated in Fig. 14. For this reason, a resolvent analysis has been used to examine how fluctuations including those from nonlinear effects are amplified or attenuated with respect to the time-averaged flow. When applying a resolvent analysis to the temporal mean of inherently unsteady or turbulent flows, it is advisable to examine whether the operator $A = A(q_0)$ has stable eigenvalues to separate the forced response from the unforced response.

Implementation in Different Geometries. The use of a resolvent analysis is increasing in fluid dynamics to study both transitional and fully turbulent flows. Flows that have been analyzed in terms of the most amplified (forced) modes include, among others, canonical flows such as boundary layers, Couette and Poiseuille flows [199,201], jets [196,197,202,203], backward-facing steps [204], and cavity flows [205].

Computational Costs. Resolvent modes can be routinely and efficiently computed for base flows that are homogeneous in two spatial dimensions (one-dimensional base flow). A resolvent analysis for flows that are homogeneous in one spatial dimension (two-dimensional base flow) incurs significantly more computing costs [206]. With current computational resources, fully inhomogeneous...
flows represent an unmet challenge, although flows with narrowband perturbations can be tackled [205].

Wall Boundary Conditions and Control. A resolvent analysis admits any linear control formulation through the boundary conditions. For example, passive, (a compliant surface [207]), open-loop (dynamic roughness [208]), and active (opposition control [209]) control inputs at the wall have been investigated in wall turbulence.

Wave Packets and Coherent Structures. Although formally a tool for linear systems, a resolvent analysis can provide insights into coherent structures in turbulent flow by analyzing the response of the system associated with the linear resolvent operator formed using the turbulent mean flow. This is not necessarily equivalent to linearization because the nonlinear forcing is required to sustain an otherwise stable system. Note that the mean velocity itself can be recovered via an extension of the analysis shown in Fig. 14. Resolvent modes can be identified, and the corresponding inputs can be thought to represent how such coherent structures are forced through nonlinear (triadic) interactions among modes at other frequencies, or through other stochastic inputs to the system (for example, noise added at an inflow in a DNS to excite turbulence). A particular success has been the recovery of packets of hairpin vortices in wall turbulence from resolvent modes [210].

The resolvent-based approach is also particularly useful to analyze convectively unstable configurations that exhibit no intrinsic dynamics. Despite the absence of an unstable linear global mode (see Sec. VIII) in such cases, coherent structures are often observed in such flows. A typical example is a turbulent jet under most operating conditions. Here, continuous forcing through the background turbulence sustains large-scale coherent structures in the jet shear layer over a wide range of frequencies. The resolvent analysis allows identification of these wave packets as optimal responses, and it associates each response mode to its corresponding optimal input via the gain of the input–output relation (see second example in the following).

![Fig. 15 Isocontours at half the maximum value for each component (warm/light colors denote positive value, and cold/dark colors denote negative value) of the real part of the first resolvent response mode pair \((k_\xi, \pm k_\zeta, c = \omega/k_\xi) = (1 h, 0.5 h, 0.5 h U_{cL})\) in channel flow at a friction Reynolds number of \(R_e^+ = 3000\). Here, \(U_{cL}\) is the centerline velocity and \(h^+\) is the channel half-height normalized by the viscous length scale.

3. Strengths and Weaknesses

Strengths:

1) A resolvent analysis can use the mean flow as the base state (instead of the exact solution to the Navier–Stokes equations), even in the case of turbulent flows.

2) When resolvent analysis is applied to turbulent mean flows, large-scale coherent structures can be interpreted as modal solutions that are sustained through forcing by the turbulent background.

3) The resolvent operator can be formulated to describe the linear dynamics associated with the governing Navier–Stokes equations.

4) The resolvent analysis identifies the form of the most amplified inputs, the corresponding output, and the gain.

5) The resolvent operator is often low rank, with an underlying mathematical structure that can be used to understand the coherent flow structure and permit exploitation of state-of-the-art matrix approximation techniques.

Weaknesses. A resolvent analysis identifies the most amplified inputs and outputs, whereas the product of the amplification (singular value) and the weight of the nonlinear forcing determines the observations. Information on the nonlinear forcing term is therefore required to predict the most energetic observed modes and synthesize nonlinear models.

B. Illustrative Examples

1. Turbulent Channel Flow

We show in Fig. 15 an example of the typical form of the first (i.e., most amplified) resolvent response modes in turbulent channel flow. Here, a left- and right-going pair of modes (i.e., \(\pm k_\zeta\)) are considered with wave numbers and wave speeds representative of the very large-scale motions in wall turbulence. For this mode, the amplitudes of the three velocity components have relative magnitudes \((|u|, |v|, |w|) = (1, 0.05, 0.17)\), i.e., the turbulent kinetic energy of the mode is mostly associated with the streamwise velocity fluctuations. The isocontours of velocity correspond to a quasi-streamwise vortex...
Resolvent modes have recently been shown to provide an efficient basis to represent exact coherent solutions of the Navier–Stokes equations [214], suggesting that the analysis may both provide an inexpensive approximate continuation method within families of solutions and aid the search for new exact solutions. Formal connections between the resolvent analysis, Koopman analysis, DMD, and exact nonlinear traveling wave solutions have recently been outlined [216]. This is out of the scope of this paper, but it may be of interest to expert practitioners.

IX. Conclusions

An overview of the modal decomposition/analysis techniques that are widely used to examine a variety of fluid flows was provided. Specifically, the POD, balanced POD, DMD, Koopman analysis, global linear stability analysis, and resolvent analysis were presented. The modal structures extracted by these techniques could shed light on different aspects of the flowfield with some shared similarities. For the data-based methods, the POD analysis captures the most energetic modes and the balanced POD analysis captures the most controllable and observable modes, providing balancing and adjoint modes. DMD (and Koopman analysis) extracts the dynamic modes along with their growth rates and frequencies from the flowfield data. Although POD and DMD methods can use flowfield data from numerical simulation and experimental measurements, the balanced POD requires adjoint flow data, which makes it applicable to numerical studies only. Operator-based analysis techniques of the Koopman analysis, global linear stability analysis, and resolvent analysis were also presented. Koopman analysis rigorously connected DMD to nonlinear dynamical systems. The global linear stability and resolvent analyses can be used to examine growth or decay characteristics of perturbations with respect to a given base or mean flow. These two approaches require discretized operators from the Navier–Stokes equations to perform the modal stability analysis.

The examples shown in this paper focused mostly on fluid flow analysis. However, modes obtained from these techniques can also be used to develop reduced-order models, which can capture the dynamics of the flow with significantly lower computational costs. Such models can be useful in closed-loop flow control [54, 59, 84, 85, 90, 93, 217] as well as aerodynamic design [218, 219].

A large number of references were provided throughout the paper so that readers can seek additional insights as needed. It is hoped that this document can serve as a stepping stone for the readers to become familiar with various modal analysis techniques, analyze a variety of complex flow physics problems, and further advance the developments of these modal analysis techniques.

Appendix: Data Matrix

We present an example of formatting the flowfield data into a matrix form $X$ in preparation to perform the modal decomposition analysis. For simplicity, we assume that the grid is structured with uniform spacing and the data are sampled with a constant time step. Here, we take data from a two-dimensional velocity field $q = (u, v)$ and construct $X$. In general, the data to be examined can be from an unstructured or nonuniform grid, if scaled appropriately, or processed to reside on a uniform grid. The formulation discussed here can be easily extended to scalar variables or three-dimensional vector fields.

Consider the velocity field $q$ made available from the simulation or experiment:

$$q_{ij} = (u_{ij}, v_{ij}) = (u(\xi_i, \eta_j), v(\xi_i, \eta_j))$$

where the spatial coordinates $(\xi_i, \eta_j)$ with $i = 1, \ldots, n_i$ and $j = 1, \ldots, n_j$ are for a colocated setup, as illustrated in Fig. A1. To construct the data matrix $X$, we need to organize the data in the form of a column vector at each time instance. An example of stacking the snapshot data to form a column vector $x(t)$ at time $t = t_k$ can be written as
Fig. A1 Example of discrete velocity data on a uniform collocated grid.

\[
\begin{bmatrix}
  v_{i1} & v_{i2} & \cdots & v_{in} \\
  v_{i2} & v_{i3} & \cdots & v_{in} \\
  \vdots & \vdots & \ddots & \vdots \\
  v_{i(n-1)} & v_{i(n-1)} & \cdots & v_{in}
\end{bmatrix}
\]

\[\mathbf{X} = \{\mathbf{x}(t_1), \mathbf{x}(t_2), \ldots, \mathbf{x}(t_m)\} \in \mathbb{R}^{n \times m} \quad (A2)\]

In general, this data matrix is a tall and skinny matrix for fluid flows modal decompositions through the POD, balanced POD, and DMD.

Note that the preceding stacking and unstacking operations are for a two-dimensional vector, which yield a column vector size of \(n = 2n_x n_y\). The unstacking operation is useful for visualizing the decomposition modes once decomposition is performed on the flowfield data. For scalar data, we only need to consider a particular component. For three-dimensional vector data, we can append the third component to the preceding formulation. Once the column data vectors are available at the desired time levels, we can collect them into a data matrix:

\[\mathbf{X} = [\mathbf{x}(t_1), \mathbf{x}(t_2), \ldots, \mathbf{x}(t_m)] \in \mathbb{R}^{m \times n} \quad (A3)\]

In general, this data matrix is a tall and skinny matrix for fluid flows \((n \gg m)\). With this data matrix at hand, we are ready to perform modal decompositions through the POD, balanced POD, and DMD.

Acknowledgments

This paper was one of the major outcomes from the AIAA Discussion Group (Fluid Dynamics Technical Committee) entitled “Modal Decomposition of Aerodynamics Flows” organized by Kunihiko Taira and Douglas R. Smith [U.S. Air Force Office of Scientific Research (AFOSR) Program Officer, Unsteady Aerodynamics and Turbulent Flows]. D. R. Smith has been instrumental in creating, organizing, and supporting this discussion group at every stage. He, in fact, has been one of the contributors to the present overview paper from conception to final editing. We gratefully acknowledge the contributions made by D. R. Smith to this document and for his continued support of much of this work. The authors also thank the fruitful discussions with the members of the discussion group and greatly acknowledge the generous support from the following agencies: U.S. Air Force Office of Scientific Research, Army Research Office, Defense Advanced Research Projects Agency, U.S. Department of Energy, Deutsche Forschungsgemeinschaft, the European Office of Aerospace Research and Development, the National Science Foundation, and the Office of Naval Research. The authors acknowledge the anonymous referees for providing insightful comments on the paper. K. Taira is grateful to his research group members for providing extensive feedback throughout the preparation of this paper.

References


doi:10.1109/TPWRS.2010.2130369

doi:10.1109/TPWRS.2011.2183625

doi:10.1073/pnas.1517384113

doi:10.1135/i.40626004

doi:10.1016/S0376-0421(03)00030-1

doi:10.1016/j.euromechflu.2007.09.004


doi:10.1017/jfm.2016.778

doi:10.1017/jfm.2016.440

doi:10.1017/s0022112007002777

doi:10.1017/s0022112099007259

doi:10.1063/1.3112738

doi:10.1017/s0022112009006272

doi:10.1007/s00162-010-0183-9

doi:10.1017/s0022112085000277


doi:10.1098/rsta.2006.1843

doi:10.1007/s00162-010-0183-9

doi:10.1017/s0022112085000277


doi:10.1098/rsta.2006.1843

doi:10.1007/s00162-010-0183-9

doi:10.1017/s0022112085000277


doi:10.1098/rsta.2006.1843

doi:10.1007/s00162-010-0183-9

doi:10.1017/s0022112085000277


doi:10.1098/rsta.2006.1843