



**HAL**  
open science

# From the POD-Galerkin Method to Sparse Manifold Models

Jean-Christophe Loiseau, Steven Brunton, Bernd Noack

► **To cite this version:**

Jean-Christophe Loiseau, Steven Brunton, Bernd Noack. From the POD-Galerkin Method to Sparse Manifold Models. Peter Benner. Handbook of Model Order Reduction, Volume 3, De Gruyter, 2021, 10.1515/9783110499001-009 . hal-02398729

**HAL Id: hal-02398729**

**<https://hal.science/hal-02398729v1>**

Submitted on 15 Jun 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

Jean-Christophe Loiseau, Steven L. Brunton, and Bernd R. Noack

## 9 From the POD-Galerkin method to sparse manifold models

**Abstract:** Reduced-order models are essential for the accurate and efficient prediction, estimation, and control of complex systems. This is especially true in fluid dynamics, where the fully resolved state space may easily contain millions or billions of degrees of freedom. Because these systems typically evolve on a low-dimensional attractor, model reduction is defined by two essential steps: (1) identifying a good state space for the attractor and (2) identifying the dynamics on this attractor. The leading method for model reduction in fluids is Galerkin projection of the Navier–Stokes equations onto a linear subspace of modes obtained via proper orthogonal decomposition (POD). However, there are serious challenges in this approach, including truncation errors, stability issues, difficulty handling transients, and mode deformation with changing boundaries and operating conditions. Many of these challenges result from the choice of a linear POD subspace in which to represent the dynamics. In this chapter, we describe an alternative approach, feature-based manifold modeling (FeMM), in which the low-dimensional attractor and nonlinear dynamics are characterized from typical experimental data: time-resolved sensor data and optional nontime-resolved particle image velocimetry (PIV) snapshots. FeMM consists of three steps: First, the sensor signals are lifted to a dynamic feature space. Second, we identify a sparse human-interpretable nonlinear dynamical system for the feature state based on the sparse identification of nonlinear dynamics (SINDy). Third, if PIV snapshots are available, a local linear mapping from the feature state to the velocity field is performed to reconstruct the full state of the system. We demonstrate this approach, and compare with POD-Galerkin modeling, on the incompressible two-dimensional flow around a circular cylinder. Best practices and perspectives for future research are also included, along with open-source code for this example.

**Keywords:** Reduced-order models, manifold, SINDy, dynamical systems, nonlinear dynamics


**MSC 2010:** 70K50, 70K70, 76D25, 76E30

---

**Jean-Christophe Loiseau**, Laboratoire DynFluid, Arts et Métiers ParisTech, 75013 Paris, France

**Steven L. Brunton**, Department of Mechanical Engineering, University of Washington, Seattle, WA 98195, USA

**Bernd R. Noack**, Center for Turbulence Control, Harbin Institute of Technology, Shenzhen Campus, Room 311 of C block HIT Campus, University Town, Xili, Shenzhen, 518055 People's Republic of China; and Institut für Strömungsmechanik und Technische Akustik (ISTA), Technische Universität Berlin, Müller-Breslau-Straße 8, 10623 Berlin, Germany

Open Access. © 2021 Jean-Christophe Loiseau et al., published by De Gruyter.  This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

<https://doi.org/10.1515/9783110499001-009>

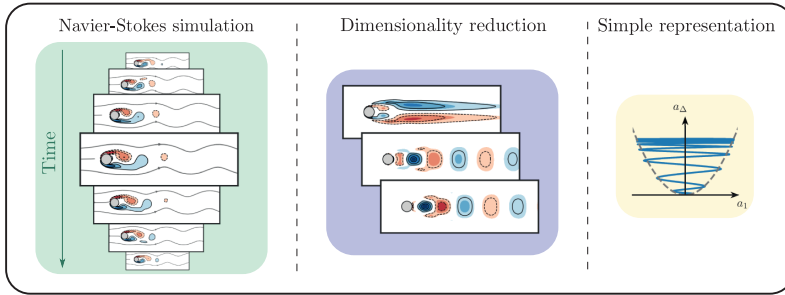
## 9.1 Introduction

Understanding, modeling, and controlling complex fluid flows is a central focus in many scientific, technological, and industrial applications, including energy (e. g., wind, tidal, and combustion), transportation (e. g., planes, trains, and automobiles), security (e. g., airborne contamination), and medicine (e. g., artificial hearts and artificial respiration). Improved models of engineering flows have the potential to dramatically improve performance in these systems through optimization and control, resulting in practical gains such as drag reduction, lift increase, and mixing enhancement [38, 21, 98, 85, 25]. Although the Navier–Stokes equations provide a detailed mathematical model, this representation may be difficult to use for engineering design, optimization, and control. Instead, they are commonly discretized into a high-dimensional, nonlinear dynamical system with many degrees of freedom and multiscale interactions. These equations are nonetheless expensive to simulate, making them unwieldy for iterative optimization or in-time control. They may also obscure the underlying physics, which often evolves on a low-dimensional attractor [49, 77]. The various fidelities of model description were described by [115]: *white-box* describes an accurate evolution equation based on first principles (e. g., Navier–Stokes discretization), *gray-box* describes a low-dimensional model approximating the full state (e. g., proper orthogonal decomposition [POD]-Galerkin models), and *black-box* describes input–output models that lack a connection to the full state space (e. g., neural networks).

In the following, we outline related reduced-order models as our point of departure in Section 9.1.1 and foreshadow proposed innovations of this study in Section 9.1.2.

### 9.1.1 Related reduced-order models as point of departure

Reduced-order models provide low-dimensional descriptions of the underlying fluid behavior in a compact and computationally efficient representation. This is illustrated in Figure 9.1, where, starting from full-state velocity snapshots obtained from direct numerical simulation, one extracts the leading coherent structures in order to obtain a low-dimensional representation of the system’s dynamics. There are many techniques for reduced-order modeling, ranging from physical reductions to purely data-driven methods, and nearly everything in between. POD [100, 14, 49] provides a low-rank modal decomposition of fluid flow field data, extracting the most energetic modes. It is then possible to Galerkin project the Navier–Stokes equations onto these modes, resulting in an approximate, low-dimensional model in terms of mode coefficients [78, 28]. POD-Galerkin models are widely used, as they are interpretable, gray-box models, and it is straightforward to reconstruct the high-dimensional flow field from the low-dimensional model via POD modes. The first pioneering example of [4] featured wall



**Figure 9.1:** Illustration of reduced-order modeling. Starting from a direct numerical simulation of the Navier–Stokes equations (left), the dominant spatio-temporal coherent structures are extracted from a set of velocity snapshots (center). The temporal evolution of these structures then provides a simplified representation of the system’s dynamics (right) amenable to modeling.

turbulence, almost three decades ago. Subsequent POD models have been developed for the transitional boundary layer [83], the mixing layer [111, 114], the cylinder wake [33, 42], and the Ahmed body wake [80], to name only a few.

POD-Galerkin modeling is challenging for changing domains [18], changing boundary conditions [45], and slow deformation of the modal basis [5]. Standard Galerkin projection can also be expected to suffer from stability issues [82, 90, 29], although including energy-preserving constraints may improve the long-time stability and performance of nonlinear models [7, 31]. POD-Galerkin models tend to be valid for a narrow range of operating conditions, near those of the data set used to generate the POD modes. Transients also pose a challenge to POD modeling. Refs. [77] and [106] demonstrate the ability of a low-dimensional model to reproduce nonlinear transients of the von Kármán vortex shedding past a two-dimensional cylinder, provided the projection basis includes a *shift mode* quantifying the distortion between the linearly unstable base flow and marginally stable mean flow. These techniques have been extended to include the effect of wall actuation [45, 81].

In addition to the physics-informed Galerkin projection, data-driven modeling approaches are prevalent in fluid dynamics [21, 85]. For example, dynamic mode decomposition (DMD) [50, 86, 55], the eigensystem realization algorithm (ERA) [51], Koopman analysis [72, 73, 109, 116], cluster-based reduced-order models [53], NARMAX models [15, 95, 120, 44], and network analysis [76] have all been used to identify dynamical systems models from fluids data, without relying on prior knowledge of the underlying Navier–Stokes equations. DMD models are readily obtained directly from data, and they provide interpretability in terms of flow structures, but the resulting models are linear, and the connection to nonlinear systems is tenuous unless DMD is enriched with nonlinear functions of the data [116, 55]. Neural networks have long been used for flow modeling and control [74, 122, 56, 54], and recently deep neural networks have been used for Reynolds-averaged turbulence modeling [59]. However, many machine learning methods may be prone to overfitting, have limited

interpretability, and make it difficult to incorporate known physical constraints. Parsimony has thus become an overarching goal when using machine learning to model nonlinear dynamics. In the seminal work of [16] and [91] governing dynamics and conservation laws are discovered using genetic programming along with a Pareto analysis to balance model accuracy and complexity, preventing overfitting.

Recently, [22] introduced the sparse identification of nonlinear dynamics (SINDy), which identifies parsimonious nonlinear models from data. SINDy follows the principle of Ockham's razor, resting on the assumption that there are only a few important terms that govern the dynamics of a system, so that the equations are sparse in the space of possible functions. Sparse regression is then used to efficiently determine the fewest terms in the dynamics required to accurately represent the data, preventing overfitting. Because SINDy is based on linear algebra (i. e., the nonlinear dynamics are represented as a linear combination of candidate nonlinear functions), the method is readily extended to incorporate known physical constraints [61]. In general, it is possible to obtain nonlinear models using genetic programming or SINDy on POD or DMD mode coefficients, which make these methods *gray box*, having a transformation from the model back to the high-dimensional, interpretable state space. However, models developed on POD/DMD mode coefficients may still suffer from fundamental challenges of traditional POD-Galerkin models, such as capturing changing boundary conditions, moving geometry, and varying operating condition.

### 9.1.2 Contribution of this work

In this work, we introduce a new gray-box modeling procedure that yields interpretable nonlinear models from measurement data. The method is applied to the well-investigated two-dimensional transient flow past a circular cylinder with slow change of the base flow and varying coherent structures [105]. In particular, we develop sparse interpretable nonlinear models only from the temporal amplitudes  $a_1(t)$  and  $a_2(t)$  of the leading vortex shedding POD modes, hereafter denoted as our features. Second, a sparse dynamical model is identified in this feature space. For the following step, full-state measurement data are assumed to be available. Combining the nonlinear correlations existing between the various POD modes with techniques from Grassmann manifold interpolation enables us to obtain highly accurate estimates of the flow field both in the vicinity of the linearly unstable base flow and the marginally stable flow. This mapping provides significantly more accurate flow reconstruction, as compared to a POD-Galerkin model of the same order. To summarize, the resulting gray-box modeling procedure has the following beneficial features: (i) it captures nonlinear physics, (ii) it is based on a simple, noninvasive computational algorithm, (iii) the resulting model is interpretable in terms of nonlinear interaction physics and generalized modes (optional with full-state data), and (iv) modeling

feature vectors is more robust to mode deformation, moving geometry, and varying operating condition.

The chapter is organized as follows: Section 9.2 provides an overview of the flow configuration considered in this work, namely, the incompressible, two-dimensional flow past a circular cylinder at  $Re = 100$ . Based on velocity snapshots obtained from direct numerical simulations, two different reduced-order modeling strategies are presented in Sections 9.3 and 9.4. First, Section 9.3 introduces the canonical POD-Galerkin reduced-order model and discusses its main limitations. Then, Section 9.4 presents a highly accurate low-order model identified using recent advances in machine learning. Finally, Sections 9.5 and 9.6 summarize our key findings, highlight some connections with previous works, and provide the reader with good practices and possible future directions to extend this work.

This contribution closely relates to three chapters of Volumes 1 and 2 of this handbook. Starting point is the POD-Galerkin method [12, Chapter 2]. A transient cylinder wake illustrates the benefits from manifold interpolation [124, Chapter 7]: A two-dimensional manifold is more accurate than a POD expansion with 50 modes. The resulting dynamical system on this manifold is significantly simplified by SINDy [12, Chapter 7].

## 9.2 Benchmark configuration and dynamics

The flow configuration considered is the canonical two-dimensional incompressible viscous flow past a circular cylinder at  $Re = 100$ , based on the free-stream velocity  $U_\infty$ , the cylinder diameter  $D$ , and the kinematic viscosity  $\nu$ . This Reynolds number is well above the critical Reynolds number ( $Re_c = 48$ ) for the onset of the two-dimensional vortex shedding [118, 104, 94] and below the critical Reynolds number ( $Re_c = 188$ ) for the onset of three-dimensional instabilities [119, 8, 117]. Its dynamics are governed by the incompressible Navier-Stokes equations

$$\begin{aligned} \frac{\partial \mathbf{u}}{\partial t} + \nabla \cdot (\mathbf{u} \otimes \mathbf{u}) &= -\nabla p + \frac{1}{Re} \nabla^2 \mathbf{u}, \\ \nabla \cdot \mathbf{u} &= 0, \end{aligned} \quad (9.1)$$

where  $\mathbf{u} = (u, v)^T$  and  $p$  are the velocity and pressure fields, respectively. The center of the cylinder has been chosen as the origin of the reference frame  $\mathbf{x} = (x, y)$ , where  $x$  denotes the streamwise coordinate and  $y$  denotes the spanwise coordinate. This study considers the same computational domain as in [77, 61, 63], extending from  $x = -5$  to  $x = 15$  in the streamwise direction and from  $y = -5$  to  $y = 5$  in the spanwise direction. A uniform velocity profile is prescribed at the inflow, a classical stress-free boundary condition is used at the outflow, and free-slip boundary conditions are used on

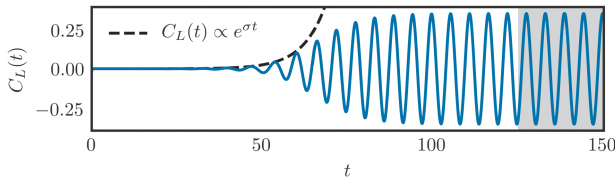
the lateral boundaries of the computational domain. The open-source spectral element solver NEK5000 [41] is used to solve the equations with a third-order accurate temporal integration. For the sake of reproducibility, all of the files required to rerun the simulations presented in this work are freely available at the following address: <https://www.github.com/loiseaujc> along with an illustrative Jupyter Notebook.

### 9.2.1 Direct numerical simulation

Figure 9.2 depicts the evolution of the lift coefficient  $C_L$  as a function of time. This direct numerical simulation (DNS) has been initialized with

$$\mathbf{u}(\mathbf{x}, 0) = \mathbf{u}_b + \epsilon \Re(\hat{\mathbf{u}})(\mathbf{x}),$$

where  $\mathbf{u}_b$  is the linearly unstable base flow and  $\Re(\hat{\mathbf{u}})(\mathbf{x})$  is the real part of the linearly unstable eigenmode normalized such that its amplitude is equal to unity (see Section 9.2.2 for more details). The parameter  $\epsilon$ , fixing the initial amplitude of the perturbation, was set such that the initial energy of the perturbation is of the order  $10^{-6}$ .



**Figure 9.2:** Time series of the instantaneous lift coefficient  $C_L(t)$ , from the linearly unstable base flow to the marginally stable mean flow, obtained by direct numerical simulation. The black dashed line depicts the exponential growth predicted by linear stability analysis while the gray shaded region highlights the window over which flow snapshots have been collected for the POD analysis presented in Section 9.3.

Three different phases are clearly visible in the time evolution of  $C_L(t)$ , namely, a period of exponential growth for  $0 \leq t \leq 60$ , the onset of nonlinear saturation for  $60 \leq t \leq 100$ , and finally the constant amplitude quasi-harmonic oscillatory regime for  $t \geq 100$  characteristic of the von Kármán vortex street. The nonlinear saturation mechanism is briefly described hereafter. The nonlinear interaction of the instability mode with itself produces Reynolds stresses that distort the underlying base flow which, in turn, modifies the shape of the instability mode. This distortion also induces a frequency shift, the flow oscillating at a frequency almost 30% larger in its final saturated state compared to that predicted by linear stability analysis of the base flow. This process continues until an equilibrium is achieved, balancing the influence of the perturbation's Reynolds stresses onto the instantaneous mean flow and the feedback this mean flow has onto the instantaneous growth rate of the perturbation. When

this equilibrium is reached, the flow is in a marginally stable state [9] and the amplitude of the perturbation no longer grows. For a complete description of this stabilizing nonlinear feedback mechanism, interested readers are referred to the self-consistent model presented in [68] or the weakly nonlinear analyses conducted by [96] and [27].

### 9.2.2 Stability of the steady solution

Given a fixed point  $\mathbf{u}_b$  of the Navier–Stokes equations, the dynamics of an infinitesimal perturbation  $\mathbf{u}'$  evolving in its vicinity are governed by

$$\begin{aligned} \frac{\partial \mathbf{u}'}{\partial t} + \nabla \cdot (\mathbf{u}_b \otimes \mathbf{u}' + \mathbf{u}' \otimes \mathbf{u}_b) &= -\nabla p' + \frac{1}{\text{Re}} \nabla^2 \mathbf{u}', \\ \nabla \cdot \mathbf{u}' &= 0. \end{aligned} \quad (9.2)$$

Introducing the normal mode ansatz  $\mathbf{u}'(\mathbf{x}, t) = \hat{\mathbf{u}}(\mathbf{x})e^{\lambda t}$ , this set of equations can be recast into the following generalized eigenvalue problem:

$$\lambda \begin{bmatrix} \mathcal{I} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \hat{\mathbf{u}} \\ \hat{p} \end{bmatrix} = \begin{bmatrix} -\nabla \cdot (\mathbf{u}_b \otimes \cdot + \cdot \otimes \mathbf{u}_b) + \frac{1}{\text{Re}} \nabla^2 & -\nabla \\ \nabla \cdot & 0 \end{bmatrix} \begin{bmatrix} \hat{\mathbf{u}} \\ \hat{p} \end{bmatrix}. \quad (9.3)$$

The linear stability of the base flow  $\mathbf{u}_b$  is then governed by the real part of the eigenvalue  $\lambda$ . In the rest of this work, the linearly unstable flow  $\mathbf{u}_b$  has been obtained using the selective damping approach [1] while the eigenpairs of the linearized Navier–Stokes operator have been computed using a time stepper Arnoldi algorithm [37, 6, 60, 62]. Interested readers are referred to [30, 108, 97] for exhaustive reviews about hydrodynamic instabilities.

The vorticity field of the linearly unstable base flow  $\mathbf{u}_b$  at  $\text{Re} = 100$  is depicted in Figure 9.3a. To the best of our knowledge, this is the only fixed point of the Navier–Stokes equations known for this flow configuration. Its linear stability has been extensively investigated [43, 96, 68, 27], and it is now well known that the bifurcation occurring at  $\text{Re}_c \approx 48$  is a supercritical Andronov–Poincaré–Hopf bifurcation eventually giving rise to the canonical Bénard–von Kármán vortex street. The vorticity field of the corresponding unstable eigenmode is shown in Figure 9.3b. This complex–conjugate pair of eigenmodes is the only unstable pair before the onset of three-dimensionality.

From a dynamical system point of view, one thus concludes that, although our discretized system is of the order  $10^6$  dimensions, the unstable linear subspace of the fixed point is only two-dimensional, i. e., only two degrees of freedom are required to describe the evolution of the system within this linear subspace. Let us furthermore consider the following *stable and unstable manifold theorem* [46].

**Theorem 1.** *Let  $E$  be an open subset of  $\mathbb{R}^n$  containing the origin, let  $f \in C^1(E)$ , and let  $\phi_t$  be the flow of the nonlinear system*

$$\frac{d\mathbf{a}}{dt} = f(\mathbf{a}).$$



Suppose that  $f(0) = 0$  and that the Jacobian matrix  $\mathbf{L} = Df(0)$  has  $k$  eigenvalues with negative real part and  $n - k$  eigenvalues with positive real part. Then, there exists a  $k$ -dimensional manifold  $W^s$  tangent to the stable subspace  $E^s$  of the linear system

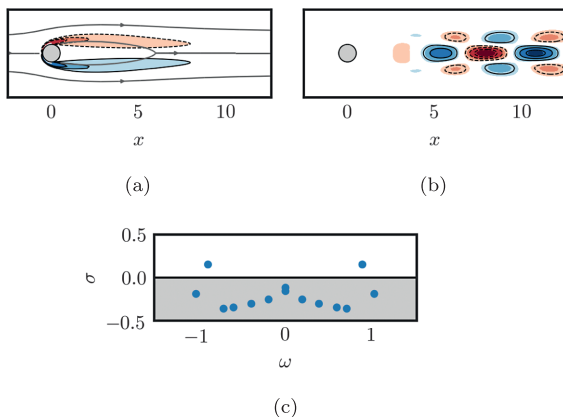
$$\frac{d\mathbf{a}}{dt} = \mathbf{L}\mathbf{a}$$

at  $\mathbf{a}_0 = 0$ . Similarly, there exists an  $(n - k)$ -dimensional unstable manifold  $W^u$  tangent to the unstable subspace  $E^u$ .

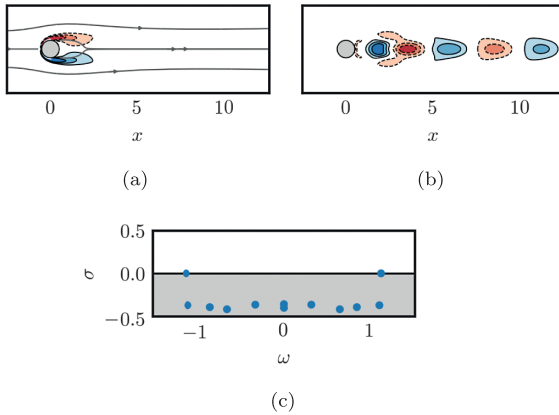
This theorem is of crucial importance for the understanding of the reduced-order model to be discussed in Section 9.4. Indeed, although we will eventually consider the nonlinear evolution of our  $10^6$ -dimensional system, we will see that this evolution can be described by a very simple dynamical system evolving onto a two-dimensional parabolic manifold originating from the aforementioned unstable subspace  $E^u$ .

### 9.2.3 Stability of the mean flow

For the flow configuration considered herein, the linearly unstable base flow  $\mathbf{u}_b(\mathbf{x})$  and the mean flow  $\bar{\mathbf{u}}(\mathbf{x})$  computed from DNS differ quite significantly from one another, notably in the size of the recirculation bubble (see Figures 9.3a and 9.4a). Consequently, predictions of the spatio-temporal characteristics of the fluctuation obtained by linear stability analysis of the base flow might be misleading.



**Figure 9.3:** (a) Vorticity field of the linearly unstable base flow for the two-dimensional cylinder flow at  $Re = 100$ . (b) Real part of the leading unstable mode's vorticity field. In both figures, blue shaded contours (solid lines) highlight regions of positive vorticity, while red shaded ones (dashed lines) highlight those of negative vorticity. In (a), a few streamlines are plotted (light gray) to highlight the extent of the recirculation bubble. (c) Eigenspectrum of the corresponding linearized Navier–Stokes operator.



**Figure 9.4:** (a) Vorticity field of the marginally stable mean flow for the two-dimensional cylinder flow at  $Re = 100$ . (b) Real part of the marginal mode's vorticity field. In both figures, blue shaded contours (solid lines) highlight regions of positive vorticity, while red shaded ones (dashed lines) highlight those of negative vorticity. In (a), a few streamlines are plotted (light gray) to highlight the extent of the recirculation bubble. (c) Eigenspectrum of the corresponding linearized Navier–Stokes operator.

Even though the mean flow  $\bar{\mathbf{u}}(\mathbf{x})$  is not a solution of the stationary Navier–Stokes equations, it has now become quite standard nonetheless to linearize the Navier–Stokes equations in its vicinity as to study its linear stability [9, 110, 11]. The eigenspectrum of the corresponding linearized Navier–Stokes operator is depicted in Figure 9.4c. As shown in [9], the leading eigenvalues have a zero real part, indicating that this mean flow is marginally stable. Moreover, while the frequency predicted by linear stability analysis of the base flow differs by almost 30% from the one recorded in direct numerical simulation, the one predicted by stability analysis of the mean flow almost is a perfect match. This mismatch results from the strong distortion induced by the instability mode as it saturates nonlinearly. Similarly, the eigenmode shown in Figure 9.4b provides a much better representation of the spatial characteristics of the fluctuations observed in DNS. For extensive details and theoretical justifications about mean flow stability analysis, interested readers are referred to [65, 9, 96, 68, 69, 110, 71, 11, 70] and references therein.

### 9.3 POD-Galerkin projection of the Navier–Stokes equations

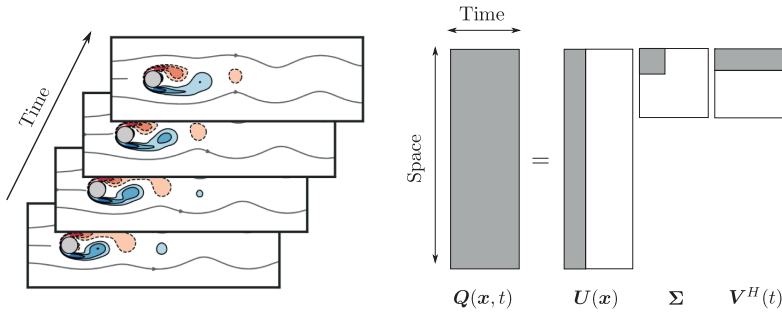
POD [100, 49] provides a low-rank modal decomposition of fluid flow field data, extracting the most energetic modes. It is then possible to project the Navier–Stokes onto the span of these POD modes, resulting in an approximate low-dimensional model

governing the evolution of the mode coefficients. POD-Galerkin models are widely used as they are interpretable gray-box models and it is straightforward to reconstruct the high-dimensional state vector of the original system from the low-dimensional model via the POD modes. The first pioneering example of [4] featured wall turbulence, over three decades ago. Subsequent POD models have been developed for the transitional boundary layer [83], the mixing layer [111, 114], the cylinder wake [33, 77, 42], and the Ahmed body wake [80], to name only a few. In the present section, dimensionality reduction via POD analysis is first presented in Section 9.3.1. Then, Sections 9.3.2 to 9.3.5 discuss the derivation of the reduced-order model from the Navier–Stokes equations and its properties, as well as its accuracy and limitations.

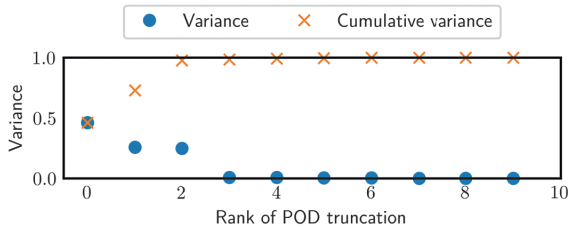
### 9.3.1 Dimensionality reduction – POD analysis

A large number of systems, including but not limited to fluid flows, are governed by high-dimensional nonlinear dynamics. Nonetheless, because most of these nonlinear dynamical systems are dissipative by nature, their dynamics are likely to evolve onto a lower-dimensional attractor characterized by a few dominant coherent structures containing a significant portion of the system’s energy [49]. Given a high-dimensional data set, the aim of *dimensionality reduction* is thus to extract a low-dimensional embedding capturing most of the variability of the original data. One of the most widely used techniques for dimensionality reduction is *POD*. It is also known as principal component analysis (PCA) in statistics and machine learning, as Kosambi–Karhunen–Loève transform in signal processing, or as empirical orthogonal functions in meteorological science, and it is closely related to singular value decomposition (see Figure 9.5). For the sake of conciseness, the mathematical details of POD will not be discussed herein. For more details, interested readers are referred to [100] and [14]. Note additionally that POD is discussed at length in this book series; see for instance Chapters 2 and 12 of Volume 1.

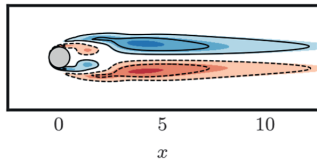
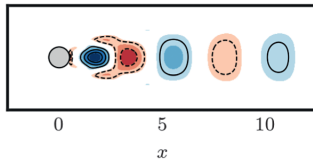
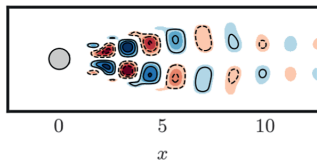
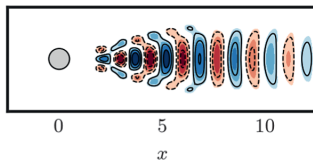
The gray shaded region in Figure 9.2 highlights the window over which snapshots of the base flow-subtracted fluctuation have been collected for the present POD analysis at a sampling rate approximately 25 times higher than the circular frequency of the natural vortex shedding. Figure 9.6a depicts the fraction of the fluctuation’s kinetic energy captured by each of the first 10 POD modes along with its cumulative sum. Note that, because we have considered base flow-subtracted fluctuations rather than mean flow-subtracted ones, the leading POD mode corresponds to the shift mode [77]. This mode captures the distortion between the base flow and the mean flow (Figure 9.6b) and accounts for 46 % of the whole kinetic energy in our snapshots data set. Considering the second and third POD modes, related to the vortex shedding (Figure 9.6c), 97.7 % of the total kinetic energy is captured. Finally, less than 1 % of the kinetic energy is discarded if one considers the first five POD modes, and less than 0.1 % if the first



**Figure 9.5:** Schematic representation of the low-rank approximation of the data matrix  $\mathbf{Q}$  by means of singular value decomposition. Each column of  $\mathbf{Q}$  contains one snapshot obtained from direct numerical simulation. The matrix  $\mathbf{U}$  contains the space-dependent POD modes  $\mathbf{u}_i(\mathbf{x})$  while  $\mathbf{V}$  contains the associated temporal evolutions, with superscript  $H$  denoting the Hermitian (i. e., complex conjugate transpose) operation. Finally, the diagonal matrix  $\mathbf{\Sigma}$  contains the singular values whose square characterizes the amount of variance explained by the associated singular pairs.



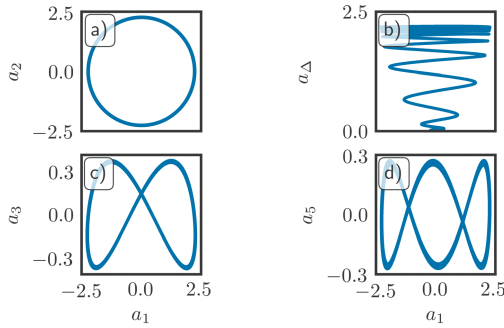
(a)

(b)  $\mathbf{u}_\Delta$ (c)  $\mathbf{u}_1$ (d)  $\mathbf{u}_3$ (e)  $\mathbf{u}_5$ 

**Figure 9.6:** (a) Fraction of the total variance (●) explained by each POD mode and the corresponding cumulative variance (×). This POD analysis has been performed using base flow-subtracted snapshots collected during the gray shaded window in Figure 9.2. The zeroth POD eigenvalue in this plot is associated to the shift mode  $\mathbf{u}_\Delta$ . Figures (b) to (e) depict the vorticity distribution of the shift mode and the first, third, and fifth POD modes, respectively. Only a subset of the whole computational domain is depicted.

seven ones are considered. For the sake of completeness, the vorticity field of selected POD modes are shown in Figure 9.6b–e.

Figure 9.7 depicts the phase plots of these various POD modes. For Figure 9.7a–c, only the evolution of the flow once it has reached the limit cycle is shown. It can be seen that, within the  $(a_1, a_2)$ -plane, the evolution of the flow traces a perfect circle underlining the periodic nature of the saturated vortex shedding for the Reynolds number considered. Additionally, the phase plots shown in Figure 9.7 highlight that the third and fourth POD modes correspond to the second harmonics of the vortex shedding, while the fifth and sixth modes capture its third harmonics. Finally, Figure 9.7d shows the whole evolution of the system, from the base flow to the mean flow, projected onto the  $(a_1, a_\Delta)$ -plane. As expected, one recovers the well-known low-dimensional parabolic manifold [77] characteristic of a large number of wake flows. It is these dynamics that we wish to capture in Section 9.3.2 using a POD-Galerkin reduced-order model.



**Figure 9.7:** Phase plots of various POD modes. For (a), (c), and (d), only the evolution once the flow has reached the limit cycle is depicted. In (b), the whole evolution is shown, from the linearly unstable base flow to the marginally stable mean flow.

### 9.3.2 Reduced-order modeling strategy – Galerkin projection

The POD analysis performed in the previous section has revealed that close to 97.5% of the base flow subtracted fluctuation’s kinetic energy is captured by considering only the shift mode and the first pair of POD modes. Starting from this observation, it thus appears reasonable to approximate the velocity field  $\mathbf{u}(\mathbf{x}, t)$  using the following Galerkin expansion:

$$\mathbf{u}(\mathbf{x}, t) \simeq \mathbf{u}_b(\mathbf{x}) + \mathbf{u}_\Delta(\mathbf{x})a_\Delta(t) + \mathbf{u}_1(\mathbf{x})a_1(t) + \mathbf{u}_2(\mathbf{x})a_2(t), \quad (9.4)$$

where  $\mathbf{u}_b(\mathbf{x})$  is the linearly unstable fixed point of the Navier–Stokes equations, while  $\mathbf{u}_\Delta(\mathbf{x})$ ,  $\mathbf{u}_1(\mathbf{x})$  and  $\mathbf{u}_2(\mathbf{x})$  are the velocity fields associated with the shift mode and the first

two POD modes, respectively. Starting from the Navier–Stokes equations, our goal is thus to derive a low-dimensional system of nonlinearly coupled ordinary differential equations governing the evolution of the POD modes' amplitudes  $a_i(t)$ . Introducing our Galerkin expansion ansatz into the Navier–Stokes equations and projecting the latter onto the span of our POD basis (this process is known as *Galerkin projection*), we obtain evolution equations for each amplitude  $a_i(t)$  of the form

$$\frac{da_i}{dt} = \sum_j L_{ij} a_j + \sum_j \sum_k Q_{ijk} a_j a_k, \quad (9.5)$$

with  $i, j, k = \Delta, 1, 2$ . By convention, the coefficient  $a_0$  associated to the base flow  $\mathbf{u}_b(\mathbf{x})$  is set to  $a_0 = 1$ . In the above equation, the linear term is given by

$$L_{ij} = \left\langle \mathbf{u}_i \left| -\nabla \cdot (\mathbf{u}_b \otimes \mathbf{u}_j + \mathbf{u}_j \otimes \mathbf{u}_b) + \frac{1}{\text{Re}} \nabla^2 \mathbf{u}_j \right. \right\rangle,$$

while the quadratic one is

$$Q_{ijk} = -\langle \mathbf{u}_i | \nabla \cdot (\mathbf{u}_j \otimes \mathbf{u}_k) \rangle,$$

where  $\langle \mathbf{a} | \mathbf{b} \rangle$  denotes the inner product

$$\langle \mathbf{a} | \mathbf{b} \rangle = \int_{\Omega} \mathbf{a} \cdot \mathbf{b} \, d\Omega.$$

Note that, as in [77], we did not explicitly account for the pressure term. For the present case, this omission however hardly changes the prediction of the reduced-order model. For a detailed discussion about the importance (or insignificance) of the pressure term in POD-Galerkin projection reduced-order models, interested readers are referred to [79].

### 9.3.3 Does the model capture the key physics?

Before discussing whether the reduced-order model derived by POD-Galerkin projection is accurate or not, let us first investigate whether it captures the key physics of the problem. In the present case, this would imply that:

1. The reduced-order model has a single fixed point located at  $\mathbf{a} = 0$ .
2. The unstable subspace  $E^u$  of the reduced-order model linearized in the vicinity of  $\mathbf{a} = 0$  is two-dimensional and associated with a complex-conjugate eigenpair.
3. As  $t \rightarrow \infty$ , the system eventually evolves toward a structurally stable limit cycle.

It must be emphasized that if the reduced-order model fails to comply with any of these requirements, then it fails at capturing the key physics of the problem.

Given the low-dimensionality of the present model, condition 1 can easily be (and has been) checked by performing an extensive Newton search. As expected, the only fixed point admitted by our reduced-order model is  $\mathbf{a} = \mathbf{0}$ . The linearization of our model in the vicinity of this fixed point is given by

$$\frac{d\mathbf{a}}{dt} = \mathbf{L}\mathbf{a},$$

with  $\mathbf{a} = [a_1 \ a_2 \ a_\Delta]^T$  and

$$\mathbf{L} = \begin{bmatrix} 0.042 & -0.986 & 0 \\ 0.959 & 0.046 & 0 \\ 0 & 0 & -0.047 \end{bmatrix}.$$

Spectral decomposition of this matrix reveals that its eigenvalues are

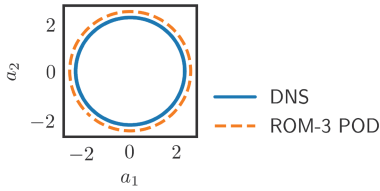
$$\Lambda = \{\lambda_1 = 0.044 + i0.972, \lambda_2 = 0.044 - i0.972, \lambda_\Delta = -0.047\}, \quad (9.6)$$

while the corresponding set of eigenvectors is

$$E_\Lambda = \left\{ \hat{\mathbf{a}}_1 = \begin{bmatrix} 1 \\ -i \\ 0 \end{bmatrix}, \hat{\mathbf{a}}_2 = \begin{bmatrix} 1 \\ i \\ 0 \end{bmatrix}, \hat{\mathbf{a}}_\Delta = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}. \quad (9.7)$$

Looking at these eigenpairs, it is clear that, as for the original Navier–Stokes equations, the fixed point  $\mathbf{a} = \mathbf{0}$  of our POD–Galerkin reduced-order model is linearly unstable. Moreover, its unstable subspace  $E^u$  is also two-dimensional and associated with complex–conjugate eigenvalues and eigenvectors corresponding to oscillatory dynamics in the  $(a_1, a_2)$ -plane while it is stable along the direction corresponding to the shift mode. Condition 2 is thus also fulfilled.

The last condition that needs to be checked is whether or not the system naturally evolves toward a stable limit cycle as  $t \rightarrow \infty$ . To do so, we integrate in time our reduced-order model using a fourth-order accurate Runge–Kutta scheme. Figure 9.8 depicts the predicted asymptotic evolution. As can be observed, this reduced-order model does evolve toward a stable limit cycle, although its amplitude is slightly larger than the amplitude of the limit cycle obtained from direct numerical simulation of the Navier–Stokes equations. Our reduced-order model thus fulfills all three necessary conditions we stated at the beginning of this section and, as such, captures qualitatively the key physics of the two-dimensional cylinder flow. Consequently, the only question that remains to be answered is the following: How accurate is this reduced-order model? The answer to this question is the subject of Section 9.3.4.



**Figure 9.8:** Comparison of the limit cycles observed in DNS (—) and predicted by the three-POD mode reduced-order model (orange - -).

### 9.3.4 How accurate is it?

We have shown in the previous section that a reduced-order model derived from the Navier–Stokes equations by means of a POD–Galerkin projection procedure qualitatively captures the key physics of the problem considered, namely:

Property 1: It has a single fixed point at the origin.

Property 2: This fixed point is linearly unstable and the associated unstable subspace is two-dimensional.

Property 3: As  $t \rightarrow \infty$ , the reduced-order model predicts that the system naturally evolves toward a periodic limit cycle.

Let us now try to further characterize the accuracy of said reduced-order model. In particular, we will focus our attention on two critical aspects:

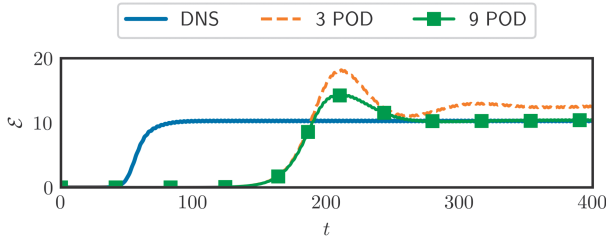
1. Does it appropriately capture the transient dynamics of the flow as it evolves from the linearly stable base flow to the marginally stable mean flow?
2. How good are its flow reconstruction capabilities?

As to answer to these questions, the reduced-order model is fed with a random initial condition having the same initial energy as that used in the direct numerical simulation described in Section 9.2, i. e.,

$$\mathbf{a}(0) = \alpha \hat{\mathbf{a}}_1 + \beta \hat{\mathbf{a}}_2,$$

such that  $\|\mathbf{a}(0)\|_2^2 = 10^{-6}$ . Figure 9.9 depicts the evolution of the fluctuation’s kinetic energy as a function of time obtained from direct numerical simulation and predicted by our POD–Galerkin reduced-order model. Although our low-order model qualitatively captures the transient dynamics of the flow, i. e., a period of exponential growth followed by nonlinear saturation, it is clear that it largely overestimates the transients duration. Moreover, as nonlinear saturation occurs, the reduced-order model predicts an energy overshoot before it saturates at a level higher than that observed in DNS. These two observations put in the limelight two critical issues of a large number of reduced-order models derived from the Navier–Stokes equations by a POD–Galerkin procedure.





**Figure 9.9:** Evolution as a function of time of the kinetic energy  $\mathcal{E}(t)$  of the base flow-subtracted fluctuation for the DNS and two Galerkin projection reduced-order models using either the first three or the first nine POD modes.

Let us first consider the problem of the overestimation of the transients duration. This problem finds its roots in the major difference that exists between the POD modes associated with the first harmonics of the vortex shedding and the eigenmodes of the linearized Navier–Stokes operator. Looking at Figure 9.3b and c, it can be seen that the POD modes are located further upstream compared to the instability modes. Consequently, while the projection of the linearized Navier–Stokes operator onto the span of the POD modes reasonably approximates the dynamics of the system in the vicinity of the mean flow, it provides a very crude approximation of the dynamics of the system when close to its fixed point, notably in terms of the instability growth rate. This is a structural problem of POD-Galerkin reduced-order models. Indeed, from a physical point of view, the instability modes continuously deform into the POD modes as the amplitude of the fluctuation grows. However, fixing the projection basis a priori using solely the POD modes prevents the reduced-order model from being able to capture this mode deformation and the continuous change of dynamics associated with it. As to alleviate this problem, [77] explicitly included the instability modes into the projection basis. Although this trick partially solves the problem, it unnecessarily increases the dimensionality of the reduced-order model.

The second problem of the present low-dimensional model is the energy overshoot and the subsequent saturation to a higher level than the one observed in DNS. This problem arises from the projection of the Navier–Stokes equations onto a finite number of basis vectors and thus from the chosen truncation of the POD basis. In the present case, our projection basis consists only of the shift mode (quantifying the distortion between the base flow and the mean flow) and the POD modes associated with the first harmonics of the vortex shedding. Because of this choice, the energy cascade from the large scales to the small scales is truncated early on. As a consequence, the energy extracted by the leading POD modes from the underlying unstable base flow cannot be transferred correctly to smaller-scale structures, hence growing beyond their expected amplitudes and causing the energy overshoot observed in Figure 9.9. This excess energy is eventually absorbed by the mean flow distortion until an equilibrium is reached, even though the final kinetic energy of the reduced-order

model nonetheless saturates at a higher level than the one observed in DNS. A naive approach to fix this issue would be to include more POD modes in the projection basis. This is illustrated in Figure 9.9, where the evolution of the kinetic energy predicted by a reduced-order model derived using a projection basis that includes the POD modes associated with the second, third, and fourth harmonics of the vortex shedding is also shown. Although increasing the rank of the POD basis from 3 to 9 mitigates the problem, the energy overshoot still exists. Moreover, including these higher-order modes in the projection basis also modifies the properties of the linearized dynamics in the vicinity of the fixed point. In the present case, including the POD modes associated with the second harmonics of the vortex shedding actually increases the dimensionality of the unstable subspace  $E^u$  from 2 to 4. In the vicinity of the fixed point, the properties of the linearized reduced-order model thus become inconsistent with those of the linearized Navier–Stokes operator.

### 9.3.5 Limitations of this approach

Although the POD-Galerkin approach to reduced-order modeling has had considerable success over the years, it nonetheless suffers from major limitations, even for a flow configuration as simple as the two-dimensional cylinder flow. For the case considered herein, four major limitations can be listed:

1. In order to accurately capture the dynamics of the system once on the limit cycle, the projection basis had to include a relatively large number of modes (i. e., eight) despite the simplicity of the dynamics, including very low energy modes.
2. The low-dimensional system tends to exhibit an energy overshoot as nonlinear saturation occurs because of the truncation of the energy cascade. This truncation of the energy cascade results from the projection of the nonlinear partial differential equations onto a finite set of basis vectors.
3. Because of the difference between the linear instability and the POD modes obtained from the limit cycle, the reduced-order model largely overestimates the transients duration unless the instability modes are explicitly included into the projection basis.
4. Finally, it can hardly account for the continuous mode deformation taking place as the flow evolves from the vicinity of the linearly unstable base flow to that of the marginally stable mean flow. A similar problem arises if one varies the Reynolds number slowly in time.

Since the generalized mean field model of Noack et al. [77], various attempts have been made to limit these shortcomings. For instance, [99] and [113] used eddy viscosity models to account for the added diffusion induced by the truncated modes, while [75] and [103] used linear interpolation to partially capture the continuous mode deformation. Recently, [34] have used *sparse coding* to obtain a nonorthonormal projection basis

for the turbulent lid-driven cavity flow that nonetheless included some of the small-scale structures needed for the energy cascade, while [40] combined POD-Galerkin projection with constrained convex optimization techniques to ensure that the statistical properties of the POD amplitudes predicted by the reduced-order model were consistent with those obtained from direct numerical simulations. These works however still had to include dozens of POD modes for numerical stability although the dynamics of the system are lower-dimensional. Despite all these attempts to increase the range of validity of the POD-Galerkin projection approach, one must not forget that it still suffers from one critical limitation that cannot be overcome within this particular framework: The governing equations of the high-dimensional system (in our case the Navier–Stokes equations) need to be known before one even tries to perform model reduction.

## 9.4 Manifold model

The approach described in the previous section can be understood as a semi-empirical or partially data-driven approach. Indeed, while on the one hand the projection basis is obtained via POD of a snapshots data matrix, the Galerkin projection procedure relies on a priori knowledge of the high-dimensional system's governing equations. Let us now consider a fully data-driven model of the flow that leverages the existence of a low-dimensional nonlinear manifold. Starting from the POD analysis presented in the previous section, Section 9.4.1 illustrates how one can further reduce the dimensionality of the problem by considering the nonlinear correlations existing between the various POD mode amplitudes. As a second step, a low-dimensional system is obtained using recent system identification techniques in Section 9.4.2. Finally, given that the system under consideration evolves on a low-dimensional manifold, Section 9.4.5 highlights how one can use Grassmannian manifolds to solve the continuous mode deformation problem when reconstructing the high-dimensional state vector of the full-order model, while Section 9.4.6 discusses some of the limitations of the approach proposed herein.

### 9.4.1 Looking for nonlinear correlations

PCA (equivalent to POD in mechanical engineering) is one of the most popular dimensionality reduction techniques. One of the key reasons for this widespread usage is that PCA finds its root in statistics. Moreover, when formulated as a singular value decomposition, PCA can be understood as an optimal low-rank matrix approximation and can thus leverage highly performing and scalable algorithms to handle extremely large data sets. Considering only the first few principal components (i. e.,

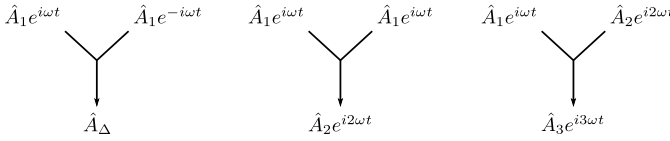
the leading left singular vectors of the data matrix), one can define an optimal linear subspace onto which the data can be orthogonally projected while minimizing (and quantifying) the amount of information lost in the process. From a statistical point of view, this orthogonal projection provides linearly uncorrelated features. Despite its optimality properties, PCA unfortunately cannot unravel nonlinear correlations in the data and postanalyses are thus required. Accounting for such nonlinear correlations may however be beneficial to further reduce the dimensionality of the problem.

Over the years, various alternatives have been proposed to overcome this major limitation in order to be able to capture nonlinear manifolds. One can cite for instance kernel PCA (kPCA) [92], Isomap [107], locally linear embedding (LLE) and its variants [84, 121, 35], spectral embedding [10], multidimensional scaling (MDS) [17], or all the variants of autoencoders recently reviewed in [13]. All these techniques are part of a domain now known as *manifold learning* or *representation learning*. However, for the particular problem considered herein, the dynamics are sufficiently simple so that we can assess the existence of nonlinear correlations directly from time series of POD modes' amplitudes. From a practical point of view, the existence of a clear pattern in a phase plot ( $a_i$ - $a_j$ ) implies the existence of such nonlinear correlations (see Figure 9.7 for examples).

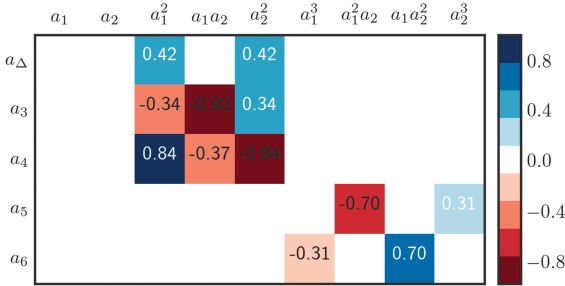
The POD analysis performed in Section 9.3.1 has revealed that less than 0.1% of the total kinetic energy in our training data set is discarded if we only consider the shift mode and the first six POD modes. Given the Fourier-like nature of the POD coefficients once the flow evolves on the limit cycle, these can be approximated by

$$\begin{aligned} a_{\Delta}(t) &\simeq \hat{A}_{\Delta}, \\ a_1(t) \pm ia_2(t) &\simeq \hat{A}_1 e^{\pm i\omega t}, \\ a_3(t) \pm ia_4(t) &\simeq \hat{A}_2 e^{\pm i2\omega t}, \\ a_5(t) \pm ia_6(t) &\simeq \hat{A}_3 e^{\pm i3\omega t}, \end{aligned}$$

where  $\omega$  is the fundamental frequency of the vortex shedding,  $\hat{A}_{\Delta}$  is the amplitude of the shift mode in the saturated stage, and  $\hat{A}_1$ ,  $\hat{A}_2$ , and  $\hat{A}_3$  are the amplitudes of the first, second, and third pairs of POD modes, respectively. Guided by physical intuition, Figure 9.10 summarizes some of the possible triadic interactions arising from the nonlinear convective term  $\nabla \cdot (\mathbf{u} \otimes \mathbf{u})$  of the Navier–Stokes equations. Looking at these triadic interactions, it thus appears that the dynamics of the shift mode and of the second pair of POD modes both result from quadratic interactions of the first pair of POD modes with itself. Similarly, the dynamics of the third pair of POD modes result from the interaction of the first pair with the second pair of modes. Alternatively, this last quadratic interaction can also be understood as a cubic interaction of the first pair with itself. These intuitions are further confirmed by looking at the correlation matrix depicted in Figure 9.11.



**Figure 9.10:** Some of the possible triadic interactions arising from the nonlinear convective term  $\nabla \cdot (\mathbf{u} \otimes \mathbf{u})$  of the Navier–Stokes equations. These triadic interactions will guide us to determine the form of nonlinear correlations existing between the amplitudes of the various POD modes considered.

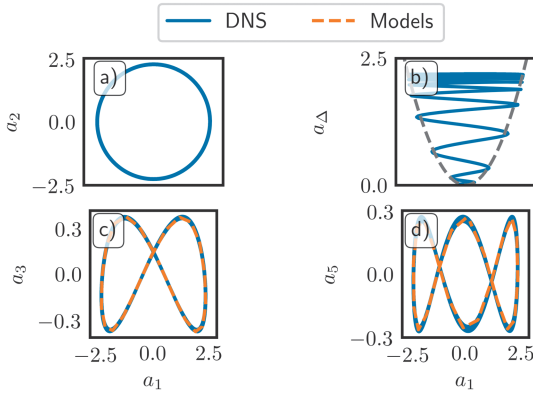


**Figure 9.11:** Pearson’s  $\rho$  correlation coefficient between various monomials of  $a_1$  and  $a_2$  and the amplitude  $a_\Delta$  of the shift mode or the amplitudes  $a_3$  to  $a_6$  of the higher-order POD modes. Blue denotes strong positive linear correlation, red denotes strong negative correlation, and white implies no linear correlation between the two variables considered.

The exact form of these nonlinear correlations can be unraveled by polynomial regression. Doing so, we obtain the following relationships:

$$\begin{aligned}
 a_\Delta &= 0.41(a_1^2 + a_2^2), \\
 a_3 &= -0.028(a_1^2 - a_2^2) - 0.13a_1a_2, \\
 a_4 &= 0.065(a_1^2 - a_2^2) - 0.056a_1a_2, \\
 a_5 &= -0.065a_1^2a_2 + 0.022a_2^3, \\
 a_6 &= -0.021a_1^3 + 0.066a_2^2a_1.
 \end{aligned} \tag{9.8}$$

Figure 9.12 provides a comparison of the evolution of the various POD modes’ amplitudes obtained from DNS and the ones predicted by the nonlinear correlations identified. As can be observed, these quadratic and cubic correlations accurately capture the evolution of the higher-order POD modes as well as the existence of the paraboloid manifold. Hence, it is clear that, although POD analysis reveals that seven POD modes need to be considered to accurately reconstruct the flow, only two of these modes are actual degrees of freedom of the system while the rest of them are entirely slaved to these two. This observation is consistent with the fact that, as shown in Section 9.2, the unstable subspace of the Navier–Stokes operator linearized in the vicinity of the



**Figure 9.12:** Same as Figure 9.7. The evolution of the coefficients  $a_3$  and  $a_5$  predicted by the nonlinear correlation models is also reported. In (d), only the parabola  $a_\Delta = 0.41a_1^2$  (i. e., a slice of the paraboloid manifold in the  $a_2 = 0$  plane) is shown.

unstable fixed point is only two-dimensional. The coming section is then devoted to the identification of the dynamical system governing the dynamics of  $a_1$  and  $a_2$ .

### 9.4.2 Low-dimensional system identification – SINDy

Advanced regression methods from statistics, such as genetic programming or sparse regression, are driving new algorithms that identify parsimonious nonlinear dynamics from measurements of complex systems. Bongard and Lipson [16] and Schmidt and Lipson [91] introduced nonlinear system identification based on genetic programming, which has been used in numerous practical applications in aerospace engineering, the petroleum industry, and finance. More recently, Brunton et al. [22] have proposed a system identification approach based on sparse regression known as *sparse identification of nonlinear dynamics* (SINDy). Following the principle of Ockham’s razor, SINDy rests on the assumption that there are only a few important terms that govern the dynamics of a given system so that the equations are sparse in the space of possible functions. Sparse regression is then used to determine the fewest terms in a dynamical system required to accurately represent the data. The resulting models are parsimonious, balancing model complexity with descriptive power while avoiding overfitting and remaining interpretable. For more details about SINDy, interested readers are referred to Chapter 12 of Volume 1 of the present book series as well as to the increasing body of literature on the subject [22, 66, 23, 87, 101, 32, 89, 88, 67, 61, 63, 52, 24, 48].

The nonlinear correlation analysis conducted in the previous section has revealed that the only true degrees of freedom of the system are the POD amplitudes  $a_1$  and  $a_2$ .

Thus, we now aim to find a nonlinear dynamical system

$$\begin{aligned}\frac{da_1}{dt} &= f_1(a_1, a_2), \\ \frac{da_2}{dt} &= f_2(a_1, a_2),\end{aligned}\tag{9.9}$$

where  $f_1 : \mathbb{R}^2 \rightarrow \mathbb{R}$  and  $f_2 : \mathbb{R}^2 \rightarrow \mathbb{R}$  are two unknown functions to be identified with SINDy. For the sake of simplicity, we will assume that these two functions are polynomial functions of  $a_1$  and  $a_2$ . In general, any basis functions may be used in the SINDy library, although polynomials appear to be a reasonable choice for fluid systems, based on the quadratic nonlinearity in the Navier–Stokes equations. Given time series of  $a_1$  and  $a_2$ , we thus define a library of candidate atoms

$$\Theta(a_1, a_2) = [1 \quad a_1 \quad a_2 \quad a_1^2 \quad a_1 a_2 \quad a_2^2 \quad a_1^3 \quad a_1^2 a_2 \quad a_1 a_2^2 \quad a_2^3]$$

so that the unknown system can be recast as

$$\begin{aligned}\frac{da_1}{dt} &= \Theta(a_1, a_2)\boldsymbol{\xi}_1, \\ \frac{da_2}{dt} &= \Theta(a_1, a_2)\boldsymbol{\xi}_2,\end{aligned}\tag{9.10}$$

where  $\boldsymbol{\xi}_1$  and  $\boldsymbol{\xi}_2$  are the solutions of a sparsity-promoting regression problem. After some cross-validation, the following system has been identified:

$$\begin{aligned}\frac{da_1}{dt} &= 0.09a_1 - 0.77a_2 - 0.016(a_1^2 + a_2^2)a_1 - 0.07(a_1^2 + a_2^2)a_2, \\ \frac{da_2}{dt} &= 0.8a_1 + 0.18a_2 + 0.06(a_1^2 + a_2^2)a_1 - 0.03(a_1^2 + a_2^2)a_2.\end{aligned}\tag{9.11}$$

As for the POD-Galerkin reduced-order model derived in Section 9.3, let us first investigate whether the identified model captures the key physics of the problem before discussing its accuracy.

### 9.4.3 Does the model capture the key physics?

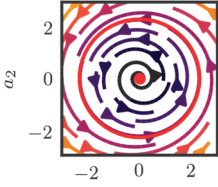
In order to capture the key physics, the identified model (9.11) needs to fulfill the same conditions as those fulfilled by the POD-Galerkin reduced-order model, namely:

Property 1: The model has a single fixed point located at  $\mathbf{a} = \mathbf{0}$ .

Property 2: The unstable subspace  $E^u$  of the model linearized in the vicinity of  $\mathbf{a} = \mathbf{0}$  is two-dimensional and associated to a complex-conjugate eigenpair.

Property 3: As  $t \rightarrow \infty$ , the system eventually evolves toward a structurally stable limit cycle.

Anyone familiar with dynamical system theory might recognize that the model (9.11) identified with SINDy corresponds to the normal form of a supercritical Andronov–Poincaré–Hopf bifurcation whose phase portrait is depicted in Figure 9.13. As such, the identified model fulfills all three conditions at once and thus captures the key physics of the problem. Identifying such a normal form is consistent with earlier works on the same flow configuration [102, 94, 123, 77].



**Figure 9.13:** Phase plane of the low-order model identified using SINDy. The red dot indicates the linearly unstable fixed point while the red circle highlights the attracting limit cycle.

Before discussing its accuracy, let us make use of the nonlinear correlations identified in Section 9.4.1 to recast the present model as

$$\begin{aligned} \frac{d}{dt} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} &= \begin{bmatrix} 0.09(1 - 0.19a_\Delta) & -0.77(1 + 0.09a_\Delta) \\ 0.8(1 + 0.07a_\Delta) & 0.18(1 - 0.18a_\Delta) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}, \\ a_\Delta &= 0.41(a_1^2 + a_2^2). \end{aligned} \quad (9.12)$$

In this form, the identified model strongly underlines the nonlinear feedback mechanism existing between the vortex shedding described by  $a_1$  and  $a_2$  and the induced distortion characterized by  $a_\Delta$ . It can moreover be understood as a low-dimensional counterpart of the self-consistent model proposed by Mantič-Lugo et al. [68] wherein the “instantaneous” mean flow  $\bar{\mathbf{u}}$  is governed by

$$\nabla \cdot (\bar{\mathbf{u}} \otimes \bar{\mathbf{u}}) + \nabla \bar{p} - \frac{1}{\text{Re}} \nabla^2 \bar{\mathbf{u}} = -\nabla \cdot (\overline{\mathbf{u}' \otimes \mathbf{u}'}),$$

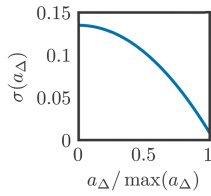
with  $\overline{\mathbf{u}' \otimes \mathbf{u}'}$  being the fluctuation’s Reynolds stress tensor, while the fluctuation itself is governed by the Navier–Stokes equations linearized in the vicinity of the “instantaneous” mean flow

$$\frac{\partial \mathbf{u}'}{\partial t} + \nabla \cdot (\bar{\mathbf{u}} \otimes \mathbf{u}' + \mathbf{u}' \otimes \bar{\mathbf{u}}) = -\nabla p' + \frac{1}{\text{Re}} \nabla^2 \mathbf{u}'.$$

Comparing these two models, it is quite striking that they have a similar structure and thus both describe the same physics. If one considers an infinitesimal perturbation  $\mathbf{u}'$ , its Reynolds stresses become negligible and the instantaneous mean flow  $\bar{\mathbf{u}}$  is nothing but the linearly unstable base flow  $\mathbf{u}_b$ . However, as the amplitude of the fluctuation



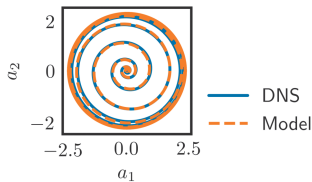
grows, so do its Reynolds stresses, causing the instantaneous mean flow  $\bar{\mathbf{u}}$  to slowly deviate from the base flow  $\mathbf{u}_b$ . Concurrently, this distortion impacts the dynamics of the fluctuation through the linearized convective term  $\nabla \cdot (\bar{\mathbf{u}} \otimes \mathbf{u}' + \mathbf{u}' \otimes \bar{\mathbf{u}})$ . This process then continues until the distortion  $\bar{\mathbf{u}} - \mathbf{u}_b$  is such that the instantaneous growth rate of the fluctuation is zero (i. e., the amplitude of the fluctuation no longer grows), hence resulting in the marginally stable mean flow. Using the identified model, this evolution of the instantaneous growth rate of the instability as a function of the distortion is illustrated in Figure 9.14.



**Figure 9.14:** Evolution of the instantaneous growth rate  $\sigma$  as a function of the distortion  $a_\Delta$ . As the distortion increases, the flow evolves from the linearly unstable base flow to the marginally stable mean flow.

#### 9.4.4 How accurate is it?

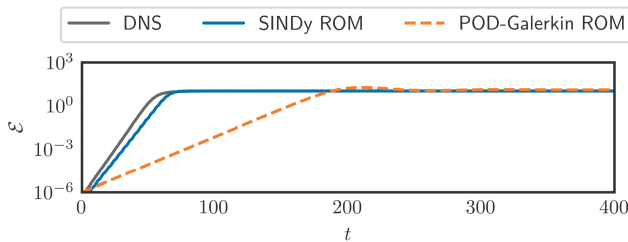
Let us now assess the accuracy of the identified model compared to direct numerical simulation. The initial velocity field used in our DNS is first projected onto the span of the leading POD modes. The corresponding POD coefficients  $a_1(0)$  and  $a_2(0)$  are then used as the initial condition for our reduced-order model. Figure 9.15 provides a comparison of the trajectory of the system in the phase plane  $(a_1, a_2)$  obtained from direct numerical simulation (—) and predicted by our reduced-order model (---). Surprisingly, an almost perfect agreement is obtained. Note however that this is not overfitting. Indeed, even though the two trajectories overlap in the  $(a_1, a_2)$ -plane, the corresponding temporal evolutions slightly differ due to a small underestimation of the instability growth rate as discussed shortly.



**Figure 9.15:** Comparison of the evolution of  $a_1$  and  $a_2$  obtained from direct numerical simulation (—) and predicted by the identified low-order model (---).

Since both the identified model (9.11) and the nonlinear correlations (9.8) are solely defined in terms of the POD coefficients, it is thus quite straightforward to reconstruct an estimate of the flow field as done for the POD-Galerkin reduced-order model. Figure 9.16 depicts the evolution of the base flow-subtracted fluctuation's kinetic energy as a function of time observed in direct numerical simulation as well as the evolution predicted by the POD-Galerkin reduced-order model derived in Section 9.3 and by the present combination of the manifold model and associated nonlinear correlations. Quite clearly, the accuracy of the model proposed in the present section largely outperforms that of the classical POD-Galerkin reduced-order model. In particular, our model does not suffer from the energy overshoot as nonlinear saturation occurs nor does it display the saturation to a higher energy level once the system evolves onto the final limit cycle. However, because we use POD modes computed from the limit cycle dynamics, the flow reconstructed in the vicinity of the fixed point actually differs from the true one since these POD modes provide only a crude approximation of the instability modes. This continuous mode deformation problem can however be solved using Grassmann manifold interpolation techniques discussed in the upcoming section. Finally, Figure 9.16 also highlights that the growth rate of the instability is slightly underestimated by our model, although nothing comparable to the underestimation of the POD-Galerkin ROM. Two different approaches can be used to correct this minor flaw:

1. Instead of restricting ourselves to cubic monomials in  $a_1$  and  $a_2$ , one can include up to seventh-order monomials in the library  $\Theta(a_1, a_2)$  used for the system identification. The resulting model then corresponds to a higher-order expansion of the supercritical Hopf bifurcation normal form.
2. Alternatively, if the growth rate of the instability is known a priori, one can force the linearized low-dimensional operator to have the same eigenvalues as its high-dimensional counterpart. Such an approach then relies on constrained optimization techniques discussed in [61] and [63].



**Figure 9.16:** Evolution as a function of time of the base flow-subtracted fluctuation's kinetic energy  $\mathcal{E}(t)$  for the DNS, the POD-Galerkin ROM derived in Section 9.3, and the model identified using SINDy. Note that, for the latter, the model predicts only the evolution of the  $a_1$  and  $a_2$  POD coefficients. The other coefficients ( $a_\Delta$ ,  $a_3$ , and  $a_4$ ) are then reconstructed using the nonlinear correlations identified previously.

Although not discussed herein, both approaches have been tested and are illustrated in the accompanying Jupyter Notebook. Both of them result in a more accurate low-order model even though the resulting model is either more complex (i. e., includes higher-order terms) or requires more advanced computational techniques for the identification (i. e., constrained  $\ell_1$ -penalized regression).

### 9.4.5 Solving the continuous mode deformation problem: Grassmann manifold interpolation

The previous section highlighted how the transient and posttransient dynamics of the two-dimensional cylinder flow could be modeled by a simple self-exciting self limiting quasi-harmonic oscillator whose degrees of freedom correspond to the amplitudes  $a_1(t)$  and  $a_2(t)$  of the two leading POD modes. If one considers only the shift mode and the first two pairs of POD modes computed from the limit cycle dynamics, the instantaneous fluctuating velocity field  $\mathbf{u}'(\mathbf{x}, t)$  is then approximated by

$$\mathbf{u}'(\mathbf{x}, t) \simeq \mathbf{u}_{\text{pod}}(\mathbf{x}, t) = \mathbf{u}_{\Delta}(\mathbf{x})a_{\Delta}(t) + \sum_{i=1}^4 \mathbf{u}_i(\mathbf{x})a_i(t). \quad (9.13)$$

It must be noted, however, that while the above Galerkin expansion provides a highly accurate approximation of the velocity field once the flow evolves onto the limit cycle, it poorly approximates the fluctuation's velocity field during the phase of exponential growth. This is illustrated in Figure 9.19, which depicts the instantaneous relative error

$$\text{Err}(t) = \frac{\|\mathbf{u}'(\mathbf{x}, t) - \mathbf{u}_{\text{pod}}(\mathbf{x}, t)\|^2}{\|\mathbf{u}'(\mathbf{x}, t)\|^2}.$$

As shown, the relative error for the POD reconstruction during the initial stage of transition is of the order of 50 %. This mismatch results from the inability of the Galerkin expansion (9.13) to capture the continuous mode deformation taking place as the system evolves from the vicinity of the base flow to that of the mean flow.

One way to circumvent this issue is to reconstruct the flow field based on the following parameterized Galerkin expansion

$$\mathbf{u}'(\mathbf{x}, t) \simeq \mathbf{u}_{\mathcal{G}}(\mathbf{x}, t) = \mathbf{u}_{\Delta}(\mathbf{x}, a_{\Delta})a_{\Delta}(t) + \sum_{i=1}^4 \mathbf{u}_i(\mathbf{x}, a_{\Delta})a_i(t). \quad (9.14)$$

In [75, 57, 103], the parameterized expansion modes were computed simply by linearly interpolating between the instability modes obtained from linear stability analysis and the POD modes from the limit cycle dynamics. Although extremely simple to implement, the elements of the resulting reduced-order basis unfortunately do not form in general an orthonormal set of vectors. Taking into account the fact that the instability modes continuously deform into the POD modes as the system evolves onto the

low-dimensional manifold structuring its phase space, a better reduced-order basis can however be obtained using so-called *Grassmann manifold interpolation*. Such an interpolation technique has been used in [3, 2] to derive linear parameterized reduced-order models for aeroelastic problems. Detailed mathematical derivation of the interpolation scheme is beyond the scope of the present contribution and only the resulting algorithmic implementation will be described hereafter. Interested readers are referred to the PhD thesis of Amsellem [2] for more details. Note moreover that Grassmann manifold interpolation is also covered in Chapter 9 of Volume 1 of the present book series.

Let us consider the linearly unstable base flow and the marginally stable mean flow as two different operating points of the same system parameterized by the relative distortion  $s = \frac{a_\Lambda}{\max a_\Lambda}$ . The base flow thus corresponds to  $s_0 = 0$ , while the mean flow corresponds to  $s_1 = 1$ . Furthermore, let us denote by  $\Phi_0 \in \mathbb{R}^{n \times 5}$  a basis of POD modes computed from the snapshots taken during the phase of exponential growth (hereafter denoted as *weakly nonlinear POD modes*, see the first row of Figure 9.18), while the POD basis computed from the mean flow will be denoted as  $\Phi_1 \in \mathbb{R}^{n \times 5}$ . Finally, let us introduce the Grassmann manifold of  $n \times 5$  orthonormal matrices  $\mathcal{G}(n, 5)$  and denote by  $\phi_0$  and  $\phi_1$  the coordinates associated with our two previous bases on this manifold. Given  $\Phi_0$  and  $\Phi_1$ , our goal is thus to compute  $\Phi(s)$ , i. e., the reduced-order basis for  $s \in [0, 1]$ , under the constraint that it has to live onto  $\mathcal{G}(n, 5)$ . A simple three-step procedure has been derived by [3] for that purpose:

1. Compute the projection of  $\Phi_1$  onto the tangent space of the Grassmann manifold  $\mathcal{G}(n, 5)$  at the point  $\phi_0$ . This projection onto the tangent space is given by the so-called logarithmic operator at point  $\phi_0$

$$\begin{aligned} (\mathcal{I} - \Phi_0 \Phi_0^T) \Phi_1 (\Phi_0^T \Phi_1)^{-1} &= U \Sigma V^T, \\ \Gamma &= U \tan^{-1}(\Sigma) V^T, \end{aligned} \quad (9.15)$$

with  $\Gamma$  being the projection of  $\Phi_1$  onto the tangent space considered.

2. Because this tangent space is flat, one can use simple linear interpolation to obtain  $\Gamma(s)$ , i. e., the projection of the yet-unknown basis  $\Phi(s)$  onto the tangent space of the Grassmann manifold at  $\phi_0$ . We then have

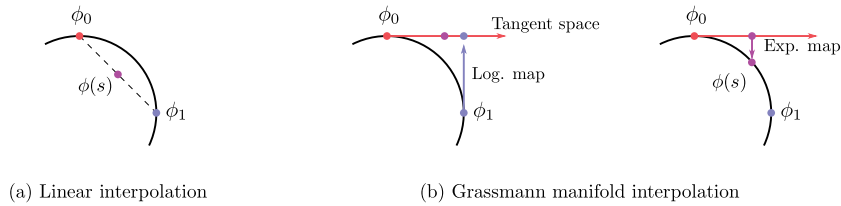
$$\Gamma(s) = U(s \tan^{-1}(\Sigma)) V^T. \quad (9.16)$$

Note that, by construction,  $\Gamma(0) = \mathbf{0}$ .

3. Finally, the projection back onto the Grassmann manifold  $\mathcal{G}$  is computed by the so-called exponential operator at point  $\phi_0$  given by

$$\Phi(s) = \Phi_0 V \cos(s \tan^{-1}(\Sigma)) + U \sin(s \tan^{-1}(\Sigma)). \quad (9.17)$$

The overall procedure is schematically represented in Figure 9.17b. Note that, by construction, the reduced-order basis  $\Phi(s)$  is orthonormal and continuously varies from



**Figure 9.17:** Illustration of different reduced-order basis interpolation techniques;  $\phi_0$  denotes our reference point (i. e., the weakly nonlinear POD basis) and  $\phi_1$  corresponds to the mean flow operating condition for which we use the classical POD modes. The parameter  $s$  is the relative amplitude of the distortion for which we want to interpolate the corresponding reduced-order basis  $\phi(s)$ . The black thick line highlights the manifold onto which our reduced-order bases should live.

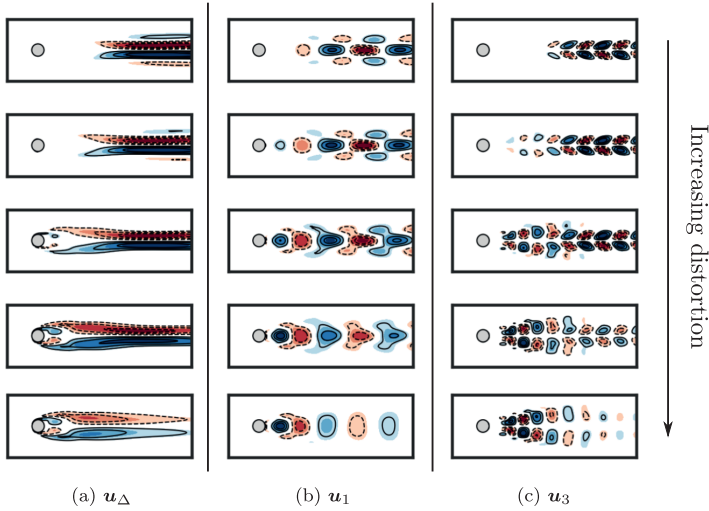
$\Phi_0$  for  $s = 0$  to  $\Phi_1$  for  $s = 1$ . This is illustrated in Figure 9.18 wherein the vorticity field of the instantaneous shift mode and the corresponding first and second harmonics of the vortex shedding are shown for various values of the relative distortion  $s$ , namely,  $s = 0, 0.25, 0.5, 0.75$ , and  $1$ . Finally, Figure 9.19a depicts the evolution as a function of time of the relative projection error

$$\text{Err}(t) = \frac{\|(\mathcal{I} - \Phi\Phi^T)\mathbf{u}'(\mathbf{x}, t)\|^2}{\|\mathbf{u}'(\mathbf{x}, t)\|^2},$$

where  $\Phi$  is either given by the classical POD basis  $\Phi_1$  or the one obtained from Grassmann manifold interpolation  $\Phi(s)$ . Although both bases have the same cardinality, the one parameterized by the instantaneous relative distortion  $s$  largely outperforms the classical POD one in terms of reconstruction accuracy, notably during the phase of exponential growth. This is particularly visible in Figure 9.19b depicting the spatial distribution of the projection error. These results further confirm the inherent low-dimensionality of the problem considered despite the continuous mode deformation occurring as nonlinear saturation takes place.

## 9.4.6 Limitations of the present approach

Although the POD-Galerkin reduced-order model derived in Section 9.3 was able to capture the key physics of the problem investigated, it nonetheless suffers from a number of major limitations listed in Section 9.3.5. On the other hand, the present section illustrated how one could identify a highly accurate and interpretable low-order model of the system by taking into account nonlinear correlations in the POD decomposition and the existence of a low-dimensional manifold. The existence of this low-dimensional manifold moreover enabled us to propose a highly accurate parameterized projection basis largely outperforming classical POD-Galerkin expansion of the velocity, notably in the initial stage of transition where the fluctuation's velocity field is well approximated by the instability modes rather than the POD ones.



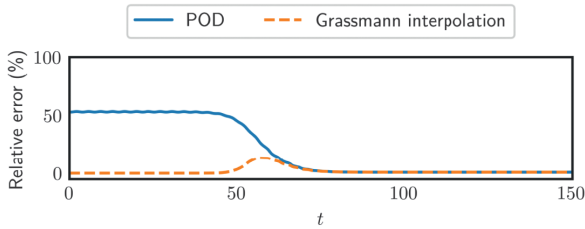
**Figure 9.18:** Evolution of the different POD modes obtained by Grassmann manifold interpolation as the flow evolves from the linearly unstable base flow (top) to the marginally stable mean flow (bottom). The intermediate rows correspond to a relative distortion of 25%, 50%, and 75%, respectively. Column (a) depicts the shift mode  $u_{\Delta}$ , (b) depicts the first harmonics of the vortex shedding, and (c) depicts the second harmonics. Note that, for each value of the relative distortion, these modes form an orthonormal set of vectors.

To the best of our knowledge, the present reduced-order model is the lowest-dimensional and yet most accurate reduced-order model capturing the transient and posttransient dynamics of the two-dimensional cylinder flow. Note moreover that the exact same methodology is likely to be directly applicable to any other flow configuration exhibiting similar dynamics. Despite its impressive accuracy, one must however remain conscious that the methodology proposed herein also has some limitations. First and foremost, the identification of the reduced-order model relied on the existence of a low-dimensional manifold and on our ability to define a corresponding nonlinear embedding of the original high-dimensional data. Although such low-dimensional nonlinear manifolds are likely to exist for a large class of dissipative dynamical systems, they may however be higher-dimensional and/or more complicated to capture. Nonetheless, in such cases one could use advanced techniques from manifold learning such as kPCA [92, 93], Isomap [107], LLE and its variants [84, 121, 35], spectral embedding [10], MDS [17], or autoencoders [13].

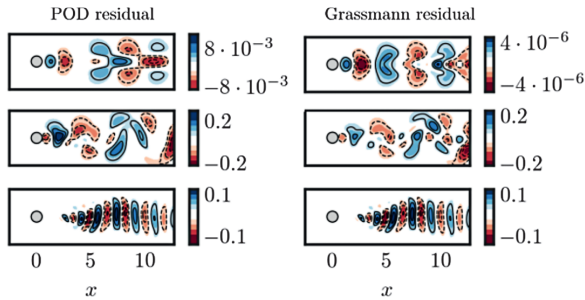
Secondly, we assumed that the right-hand side  $f(\mathbf{a})$  of our low-order model

$$\frac{d\mathbf{a}}{dt} = f(\mathbf{a})$$

could be expressed as a linear combination of monomials in  $a_1$  and  $a_2$ . While this choice may be justified for a large class of dynamical systems, the present choice precludes the identification of systems involving other types of nonlinearities, such as



(a)



(b)

**Figure 9.19:** (a) Comparison of the relative error for the orthogonal projection of the base flow-subtracted fluctuation's velocity field onto either the leading five POD modes (—) extracted from the limit cycle dynamics or the Grassmann interpolated ones (---). The direct numerical simulation has been started from an initial condition close to the linearly unstable base flow. (b) Spatial distribution of the projection error at various times. The vertical velocity component is shown. From top to bottom:  $t = 6$  (exponential growth of the instability),  $t = 60$  (onset of nonlinear saturation), and  $t = 120$  (asymptotic limit cycle).

rational functions. It must be noted however that the SINDy framework is quite extensible and various extensions have been proposed since [22] to enable the identification of dynamical systems with exotic nonlinearities; see for instance [66]. Alternatively, if the dynamics appear to be strongly nonlinear and not expressible in terms of classical analytical functions, one could include wavelets in the library  $\Theta(\mathbf{a})$  used in the identification or turn to a class of neural networks known as *long short-term memory* (LSTM). Although one would sacrifice interpretability by doing so, recent works have shown that such LSTM deep neural networks are able to capture and reproduce the chaotic spatio-temporal dynamics of the Kuramoto-Sivashinsky equation [112, 26].

## 9.5 Good practices

The two-dimensional cylinder flow at  $Re = 100$  is a prototypical example from fluid dynamics capturing the key physics of bluff body flows. Despite the low-dimensionality

of the flow dynamics, it has been shown that a reduced-order model derived from a naive POD-Galerkin projection procedure fails to accurately reproduce the dynamics of the flow, most notably its transient dynamics. The key reasons for this failure, explained in [77], are twofold:

1. Galerkin projection of the Navier–Stokes equations onto the span of a low-dimensional POD basis causes a disruption of the energy cascade, hence giving rise to the energy overshoot illustrated in Figure 9.9.
2. POD modes are classically computed from statistically steady operating conditions. Consequently, this set of modes may provide only a crude approximation of the fluctuation’s velocity field during transient dynamics. As a consequence, the corresponding low-dimensional linear operator obtained from Galerkin projection does not correctly capture the spectral properties of its high-dimensional counterpart.

Recent advances in data-driven techniques and machine learning are likely to help overcoming these limitations. It must be emphasized however that, despite their impressive successes regularly reported in mainstream and scientific media, blindly applying techniques from machine learning (and in particular from deep learning) to fluid dynamics problems may give rise to overly complicated models. The aim of this section is to discuss a set of good practices that, according to the authors, are of crucial importance when it comes to data-driven reduced-order modeling.

### 9.5.1 Dimensionality reduction

The aim of reduced-order modeling is to obtain a low-dimensional representation of the dynamics of the original high-dimensional system. The very first step is thus to apply *dimensionality reduction*. POD, which is discussed at length in this book series, is the standard choice in mechanical engineering due to its ability to rank the modes according to the fraction of the fluctuation’s kinetic energy they capture. Once the POD modes have been computed, most of the reduced-order models proposed in the literature then carry on directly with the derivation of the low-dimensional model governing the dynamics of these modes. It must be noted, however, that, as discussed in the previous section, POD analysis provides a set of modes whose temporal evolutions are only linearly uncorrelated. Hence, truncated POD corresponds simply to an optimal linear embedding of our original high-dimensional data set into a lower linear subspace. While this property might be beneficial for reduced-order models of linear systems, dissipative nonlinear dynamical systems are typically characterized by nonlinear correlations across vastly different ranges of temporal and/or spatial scales. Consequently, if the data turn out to live on a low-dimensional nonlinear manifold, POD analysis would then overestimate the number of dimensions required to describe the dynamics of the system. Accounting for these nonlinear correlations is thus a key



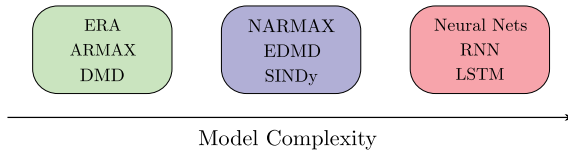
problem for standard reduced-order modeling strategies which is often disregarded by practitioners, although it may cause the identified/derived reduced-order model to be unnecessarily complicated.

Looking for nonlinear correlations between the various features of a multivariate time series is obviously significantly more complicated than looking for simple linear correlations. Given the quadratic nature of the nonlinear convective term in the Navier–Stokes equations, it seems however reasonable to restrict ourselves to polynomial correlations. Moreover, when the investigated flow exhibits only periodic dynamics as for the one considered herein, one can simply guess a priori the variables involved in the correlations by considering a limited number of triadic interactions. Polynomial regression can then be used to unravel the exact form of these nonlinear correlations. For more complicated flow configurations (e. g., chaotic and/or higher-dimensional dynamics), this task can however quickly become intractable without further preprocessing. Recently, Lopez-Paz et al. [64] proposed a new correlation metric to unravel whether two features of a multivariate time series are nonlinearly correlated or not: the *randomized dependence coefficient* (RDC). Mathematical derivation of this metric is far beyond the scope of this contribution and interested readers are referred to the original paper [64] for more details. Note that this nonlinear correlation metric is extremely simple to use and can be implemented with less than 10 lines of R or Python. Preliminary results on a high Reynolds number shear-driven cavity flow have shown that the shear-layer dynamics and inner-cavity flow were only weakly nonlinearly correlated, thus considerably simplifying the identification of a reduced-order model with only four degrees of freedom. As an element of comparison, a classical POD-Galerkin reduced-order model would involve 12 to 15 degrees of freedom.

Although the combination of POD, RDC analysis, and polynomial regression has now become one of the standard approaches used by the present authors, it must be noted that numerous other alternatives exist to unravel nonlinear correlations. In the field of machine learning, these tools form a subset known as *manifold learning* or *representation learning*. From the authors' point of view, a particularly interesting technique from manifold learning is the use of so-called *autoencoders*. This is the subject of ongoing investigations by the present authors. For more details about autoencoders and manifold learning, please see the excellent review article by Bengio et al. [13].

## 9.5.2 System identification

The field of system identification uses statistical methods to build mathematical models of dynamical systems from measured data. With respect to the classification proposed in [115], system identification enables us to obtain either *gray-box* or *black-box* models. Various methods have been proposed over the years. Some of these are classified in Figure 9.20 depending on the complexity (linear or nonlinear, interpretable or noninterpretable) of the resulting model. While the identification of



**Figure 9.20:** Classification of various system identification techniques based on the complexity of the resulting model. On the left, these techniques and their variants enable the identification of linear input–output models. At the center, NARMAX, EDMD, and SINDy allow one to identify interpretable input–output nonlinear dynamical systems. Finally, on the right, neural networks and their variants give rise to black-box strongly nonlinear models.

a linear time-invariant dynamical system has a plethora of theoretical results, theoretical guarantees for nonlinear system recovery are much more scarce. Like many fields, nonlinear system identification has nonetheless been revolutionized with the popularization of deep learning. It must be noted however that, from the authors’ point of view, a number of recent studies have put too much emphasis on illustrating deep learning techniques while discarding the possibility that the system considered could be modeled using a much simpler approach, notably studies which have used the two-dimensional cylinder flow as an illustration. Following Ockham’s razor, we thus strongly encourage practitioners to try linear system identification first (e. g., ERA, DMD, ARMAX), before moving to interpretable nonlinear system identification (e. g., NARMAX, SINDy) and eventually neural network-based techniques only if the previous two approaches have failed.

## 9.6 Conclusion

This work proposes a new reduced-order modeling procedure for unsteady fluid flows that yields accurate nonlinear models and insight into relevant flow structures. This procedure identifies sparse interpretable nonlinear models, not on the full fluid state, but from time-resolved measurements of the leading POD coefficients that may be realistically obtained in experiments. The sparsity of the model prevents overfitting and uncovers key nonlinear interaction terms. Although models are data-driven, they are interpretable, and it is also possible to incorporate partial prior knowledge of the physics or constraints to improve the models. If the stability modes are also available, it is possible to estimate the full state from the sparse model using Grassmann manifold interpolation: The full state is expanded in terms of a parameterized reduced-order basis, based on the dynamics.

This methodology is illustrated using the canonical two-dimensional cylinder flow at  $Re = 100$ . Despite its simplicity, this flow configuration is a prototypical example capturing the key physics of bluff body flows. Even though this study uses data

from direct numerical simulations, the overall strategy is generally applicable to a real flow experiment with minor modifications. Despite their simplicity, the identified models do not suffer the same drawbacks as reduced-order models obtained from a Galerkin projection procedure, namely, overestimation of the duration of transients and energy overshoots at the onset of nonlinear saturation. Instead, the identified sparse models provide simple explanations for the nonlinear saturation process of globally unstable flows. Moreover, the models are based on sensor measurements, which may include POD coefficients, lift, drag, or pressure measurements that are physically linked to the geometry. Working in these *intrinsic* coordinates has the potential to overcome many of the limitations of classical modal-based projection methods, including mode deformation due to moving geometry and varying parameters.

## 9.7 Perspectives

The effectiveness of the reduced-order models identified and the modularity of the methodology proposed in the present work suggest a number of exciting future directions. There is significant potential for these methods to be applied broadly to obtain interpretable reduced-order models for a range of flow configurations in simulations and experiments. For example, these manifold models may be applied to develop nonlinear unsteady aerodynamic models, generalizing previous linear and linear parameter-varying models [19, 20, 47].

A key motivation in this work is its extension to flow control. Given a feature vector  $\mathbf{a}$  and actuators characterized by a control law  $\mathbf{b}(t)$ , one could use SINDy with control [23, 52] in order to identify low-order models

$$\frac{d\mathbf{a}}{dt} = \mathbf{f}(\mathbf{a}, \mathbf{b})$$

that incorporate the influence of the actuation  $\mathbf{b}$  on the dynamics of  $\mathbf{a}$ . Combining such an approach with *machine learning control* [36] may result in interpretable models of entirely new flow behaviors and previously unobserved flow physics that are discovered in the controlled flow. The identified models can then serve as a low-dimensional representation of the actual system in order to facilitate the computation of nonlinear optimal feedback control laws. This is an area of active research by the authors. In the near future, the authors aim to apply the methodology introduced in the present work to the optimal control of experimental flows.

There are a number of methodological extensions that may improve the performance of this sparse modeling framework. First, it will be important to demonstrate that these methods scale favorably to systems with higher-dimensional attractors. Because the algorithms are based on simple regression and sparse optimization, they

should remain computationally tractable. Next, it may be possible to increase the accuracy of the Grassmann interpolation by building local modal libraries in different dynamic regimes (e. g., linear instability, saturated limit cycle, etc.). The storage requirements may further be reduced using compression techniques and sparse sampling. Finally, it has been demonstrated in [63] how such manifold models could be identified directly from sensor measurements such as the lift and drag coefficients. For the present flow configuration, the present authors identified that the dynamical system governing the dynamics of the lift coefficient  $C_L(t)$  of the form

$$\frac{d^2 C_L}{dt^2} + \left( \sigma - \left[ \alpha C_L^2 + \beta \left( \frac{dC_L}{dt} \right)^2 \right] \right) \frac{dC_L}{dt} + \omega_0^2 C_L = 0.$$

Such sensor-based models are strongly related to the existence of a low-dimensional manifold structuring the phase space of the system investigated and to the strong correlations existing between the various sensor measurements considered and the spatio-temporal coherent structures found in the flow. Our ability to identify such sensor-based manifold models may eventually have a major impact in experimental fluid mechanics and flow control.

A data-driven generalization of manifold models are cluster-based network models, where the snapshots are coarse-grained by centroids and the topology is encoded in a transition model between these centroids [58]. Such models may approximate broadband-frequency wall turbulence for dozens of different wall surface actuations [39]. The price for for this conceptually simple, automatable, and robust reduced-order modeling avenue is that the manifold and sparse dynamics still need to be distilled—if they exist.

## Bibliography

- [1] E. Åkervik, L. Brandt, D. S. Henningson, J. Høpfner, O. Marxen, and P. Schlatter, Steady solutions of the Navier-Stokes equations by selective frequency damping, *Phys. Fluids*, **18** (6) (2006), 068102.
- [2] D. Amsallem, *Interpolation on manifolds of CFD-based fluid and finite element-based structural reduced-order models for on-line aeroelastic predictions*. PhD thesis, Stanford University, 2010.
- [3] D. Amsallem and C. Farhat, Interpolation method for adapting reduced-order models and application to aeroelasticity, *AIAA J.*, **46** (7) (2008), 1803–1813.
- [4] N. Aubry, P. Holmes, J. L. Lumley, and E. Stone, The dynamics of coherent structures in the wall region of a turbulent boundary layer, *J. Fluid Mech.*, **192** (1988), 115–173.
- [5] H. Babaee and T. P. Sapsis, A variational principle for the description of time-dependent modes associated with transient instabilities, *Philos. Trans. R. Soc. Lond.*, **472** (2186) (2016), 20150779.
- [6] S. Bagheri, E. Åkervik, L. Brandt, and D. S. Henningson, Matrix-free methods for the stability and control of boundary layers, *AIAA J.*, **47** (5) (2009), 1057–1068.

- [7] M. Balajewicz, E. H. Dowell, and B. R. Noack, Low-dimensional modelling of high-Reynolds-number shear flows incorporating constraints from the Navier-Stokes equation, *J. Fluid Mech.*, **729** (2013), 285–308.
- [8] D. Barkley and R. D. Henderson, Three-dimensional Floquet stability analysis of the wake of a circular cylinder, *J. Fluid Mech.*, **322** (1) (1996), 215.
- [9] D. Barkley, Linear analysis of the cylinder wake mean flow, *Europhys. Lett.*, **75** (5) (2006), 750–756.
- [10] M. Belkin and P. Niyogi, Laplacian eigenmaps for dimensionality reduction and data representation, *Neural Comput.*, **15** (2003), 1373–1396.
- [11] S. Beneddine, D. Sipp, A. Arnault, J. Dandois, and L. Lesshafft, Conditions for validity of mean flow stability analysis, *J. Fluid Mech.*, **798** (2016), 485–504.
- [12] P. Benner, S. Grivet-Talocia, A. Quarteroni, G. Rozza, W. H. A. Schilders, and L. M. Solveira, *Model Order Reduction. Volume 2: Snapshot-Based Methods and Algorithms*, De Gruyter, 2020.
- [13] Y. Bengio, A. Courville, and P. Vincent, Representation learning: a review and new perspectives, *IEEE Trans. Pattern Anal. Mach. Intell.*, **35** (2013), 1798–1828.
- [14] G. Berkooz, P. Holmes, and J. L. Lumley, The proper orthogonal decomposition in the analysis of turbulent flows, *Annu. Rev. Fluid Mech.*, **25** (1) (1993), 539–575.
- [15] S. A. Billings, *Nonlinear System Identification: NARMAX Methods in the Time, Frequency, and Spatio-Temporal Domains*. Paperbackshop UK Import, 2013.
- [16] J. Bongard and H. Lipson, Automated reverse engineering of nonlinear dynamical systems, *Proc. Natl. Acad. Sci. USA*, **104** (24) (2007), 9943–9948.
- [17] I. Borg and P. Groenen, Modern multidimensional scaling: theory and applications, *J. Educ. Meas.*, **40** (2003), 277–280.
- [18] R. Bourguet, M. Braza, and A. Dervieux, Reduced-order modeling of transonic flows around an airfoil submitted to small deformations, *J. Comput. Phys.*, **230** (2011), 159–184.
- [19] S. L. Brunton, C. W. Rowley, and D. R. Williams, Reduced-order unsteady aerodynamic models at low Reynolds numbers, *J. Fluid Mech.*, **724** (2013), 203–233.
- [20] S. L. Brunton, S. T. M. Dawson, and C. W. Rowley, State-space model identification and feedback control of unsteady aerodynamic forces, *J. Fluids Struct.*, **50** (2014), 253–270.
- [21] S. L. Brunton and B. R. Noack, Closed-loop turbulence control: progress and challenges, *Appl. Mech. Rev.*, **67** (5) (2015), 050801.
- [22] S. L. Brunton, J. L. Proctor, and J. N. Kutz, Discovering governing equations from data by sparse identification of nonlinear dynamical systems, *Proc. Natl. Acad. Sci. USA*, **113** (15) (2016), 3932–3937.
- [23] S. L. Brunton, J. L. Proctor, and J. N. Kutz, Sparse identification of nonlinear dynamics with control (SINDYc), *IFAC-PapersOnLine*, **49** (18) (2016), 710–715.
- [24] S. L. Brunton, B. W. Brunton, J. L. Proctor, E. Kaiser, and J. N. Kutz, Chaos as an intermittently forced linear system, *Nat. Commun.*, **8** (2017), 1.
- [25] S. L. Brunton, B. R. Noack, and P. Koumoutsakos, Machine learning for fluid dynamics, *Annu. Rev. Fluid Mech.*, **52** (2020), 477–508.
- [26] M. A. Bucci, O. Semeraro, A. Allauzen, G. Wisniewski, L. Cordier, and L. Mathelin, Control of chaotic systems by deep reinforcement learning, *Proc. R. Soc. A*, **475** (2231) (2019), 20190351.
- [27] M. Carini, F. Auteri, and F. Giannetti, Centre-manifold reduction of bifurcating flows, *J. Fluid Mech.*, **767** (2015), 109–145.
- [28] K. Carlberg, R. Tuminaro, and P. Boggs, Preserving Lagrangian structure in nonlinear model reduction with application to structural dynamics, *SIAM J. Sci. Comput.*, **37** (2) (2015), B153–B184.

- [29] K. Carlberg, M. Barone, and H. Antil, Galerkin v. least-squares Petrov–Galerkin projection in nonlinear model reduction, *J. Comput. Phys.*, **330** (2017), 693–734.
- [30] J.-M. Chomaz, Global instabilities in spatially developing flows: non-normality and nonlinearity, *Annu. Rev. Fluid Mech.*, **37** (2005), 357–392.
- [31] L. Cordier, B. R. Noack, G. Daviller, J. Delvile, G. Lehnasch, G. Tissot, M. Balajewicz, and R. K. Niven, Control-oriented model identification strategy, *Exp. Fluids*, **54** (2013), 1580.
- [32] M. Dam, M. Brøns, J. Juul Rasmussen, V. Naulin, and J. S. Hesthaven, Sparse identification of a predator-prey system from simulation data of a convection model, *Phys. Plasmas*, **24** (2) (2017), 022310.
- [33] A. E. Deane, I. G. Kevrekidis, G. E. Karniadakis, and S. A. Orszag, Low-dimensional models for complex geometry flows: application to grooved channels and circular cylinders, *Phys. Fluids A*, **3** (1991), 2337–2354.
- [34] R. Deshmukh, J. J. McNamara, Z. Liang, J. Z. Kolter, and A. Gogulapati, Model order reduction using sparse coding exemplified for the lid-driven cavity, *J. Fluid Mech.*, **808** (2016), 189–223.
- [35] D. L. Donoho and C. Grimes, Hessian eigenmaps: locally linear embedding techniques for high-dimensional data, *Proc. Natl. Acad. Sci. USA*, **100** (2003), 5591–5596.
- [36] T. Duriez, S. L. Brunton, and R. R. Noack, *Machine Learning Control – Taming Nonlinear Dynamics and Turbulence*, Springer International Publishing, 2017.
- [37] W. S. Edwards, L. S. Tuckerman, R. A. Friesner, and D. C. Sorensen, Krylov methods for the incompressible Navier-Stokes equations, *J. Comput. Phys.*, **110** (1) (1994), 82–102.
- [38] N. Fabbiane, O. Semeraro, S. Bagheri, and D. S. Henningson, Adaptive and model-based control theory applied to convectively unstable flows, *Appl. Mech. Rev.* (2014).
- [39] D. Fernex, R. Semaan, M. Albers, P. S. Meysonnat, R. Ishar, E. Kaiser, W. Schröder, and B. R. Noack, Cluster-based network model for drag reduction mechanisms of an actuated turbulent boundary layer, *Proc. Appl. Math. Mech.*, **19**(1) article e201900219 (2019), 1–2.
- [40] L. Fick, Y. Maday, A. T. Patera, and T. Taddei, A stabilized POD model for turbulent flows over a range of Reynolds numbers: optimal parameter sampling and constrained projection, *J. Comput. Phys.*, **371** (2018), 214–243.
- [41] P. F. Fischer, J. W. Lottes, and S. G. Kerkemeir, Nek5000 Web pages, 2008. <http://nek5000.mcs.anl.gov>.
- [42] G. Galletti, C. H. Bruneau, L. Zannetti, and A. Iollo, Low-order modelling of laminar flow regimes past a confined square cylinder, *J. Fluid Mech.*, **503** (2004), 161–170.
- [43] F. Giannetti and P. Luchin, Structural sensitivity of the first instability of the cylinder wake, *J. Fluid Mech.*, **581** (2007), 167.
- [44] B. Glaz, L. Liu, and P. P. Friedmann, Reduced-order nonlinear unsteady aerodynamic modeling using a surrogate-based recurrence framework, *AIAA J.*, **48** (10) (2010), 2418–2429.
- [45] W. R. Graham, J. Peraire, and K. Y. Tang, Optimal control of vortex shedding using low-order models. Part I – Open-loop model development, *Int. J. Numer. Methods Eng.*, **44** (1999), 945–972.
- [46] J. Guckenheimer and J. Holmes, *Nonlinear Oscillations, Dynamical Systems, and Bifurcation of Vector Fields*, Theorem 1.4.2, corrected seventh printing, Springer, 2002.
- [47] M. S. Hemati, S. T. M. Dawson, and C. W. Rowley, Parameter-varying aerodynamics models for aggressive pitching-response prediction, *AIAA J.*, **55** (3) (2017), 693–701.
- [48] M. Hoffmann, C. Fröhner, and F. Noé, Reactive SINDy: discovering governing reactions from concentration data, *J. Chem. Phys.*, **150** (2) (2019), 025101.
- [49] P. Holmes, J. L. Lumley, and G. Berkooz, *Turbulence, Coherent Structures, Dynamical Systems and Symmetry*, Cambridge University Press, 1996.
- [50] P. J. Schmid, Dynamic mode decomposition of numerical and experimental data, *J. Fluid Mech.*, **656** (2010), 5–28.

- [51] J.-N. Juang and R. S. Pappa, An eigensystem realization algorithm for modal parameter identification and model reduction, *J. Guid.*, **8** (5) (1985), 620–627.
- [52] E. Kaiser, J. N. Kutz, and S. L. Brunton, Sparse identification of nonlinear dynamics for model predictive control in the low-data limit, *Proc. R. Soc. A*, **474** (2219) (2019), 20180335.
- [53] E. Kaiser, B. R. Noack, L. Cordier, A. Spohn, M. Segond, M. Abel, G. Daviller, J. Östh, S. Krajnović, and R. K. Niven, Cluster-based reduced-order modelling of a mixing layer, *J. Fluid Mech.*, **754** (2014), 365–414.
- [54] A. Krizhevsky, I. Sutskever, and G. E. Hinton, ImageNet classification with deep convolutional neural networks, *Commun. ACM*, **60** (6) (2017), 84–90.
- [55] J. N. Kutz, S. L. Brunton, B. W. Brunton, and J. L. Proctor, *Dynamic Mode Decomposition: Data-Driven Modeling of Complex Systems*, SIAM-Society for Industrial and Applied Mathematics, 2016.
- [56] C. Lee, J. Kim, D. Babcock, and R. Goodman, Application of neural networks to turbulence control for drag reduction, *Phys. Fluids*, **9** (6) (1997), 1740–1747.
- [57] O. Lehmann, M. Luchtenburg, B. R. Noack, R. King, M. Morzynski, and G. Tadmor, Wake stabilization using POD Galerkin models with interpolated modes, in *Proceedings of the 44th IEEE Conference on Decision and Control*, pp. 500–505, IEEE, 2005.
- [58] H. Li, D. Fernex, R. Semaan, J. Tan, M. Morzyński, and B. R. Noack, Cluster-based network model, *J. Fluid Mech.*, (in print) (2020), 1–41.
- [59] J. Ling, A. Kurzwaski, and J. Templeton, Reynolds averaged turbulence modelling using deep neural networks with embedded invariance, *J. Fluid Mech.*, **807** (2016), 155–166.
- [60] J.-Ch. Loiseau, *Dynamics and global stability analysis of three-dimensional flows*. PhD thesis, Ecole Nationale Supérieure d'Arts et Métiers, 2014.
- [61] J.-Ch. Loiseau and S. L. Brunton, Constrained sparse Galerkin regression, *J. Fluid Mech.*, **838** (2018), 42–67.
- [62] J.-Ch. Loiseau, M. A. Bucci, S. Cherubini, and J.-Ch. Robinet, Time-stepping and Krylov methods for large-scale instability problems, in *Computational Modelling of Bifurcations and Instabilities in Fluid Dynamics*, pp. 33–73, Springer, 2019.
- [63] J.-Ch. Loiseau, B. R. Noack, and S. L. Brunton, Sparse reduced-order modelling: sensor-based dynamics to full-state estimation, *J. Fluid Mech.*, **844** (2018), 459–490.
- [64] D. Lopez-Paz, P. Hennig, and B. Schölkopf, The randomized dependence coefficient, in C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger (eds.), *Advances in Neural Information Processing Systems 26*, pp. 1–9, Curran Associates, Inc., 2013.
- [65] W. V. R. Malkus, Outline of a theory of turbulent shear flow, *J. Fluid Mech.*, **1** (05) (1956), 521.
- [66] N. M. Mangan, S. L. Brunton, J. L. Proctor, and J. N. Kutz, Inferring biological networks by sparse identification of nonlinear dynamics, *IEEE Trans. Mol. Biol. Multi-Scale Commun.*, **2** (1) (2016), 52–63.
- [67] N. M. Mangan, J. N. Kutz, S. L. Brunton, and J. L. Proctor, Model selection for dynamical systems via sparse regression and information criteria, *Proc. R. Soc. A*, **473** (2204) (2017), 20170009.
- [68] V. Mantič-Lugo, V. Arratia, and F. Gallaire, Self-consistent mean flow description of the nonlinear saturation of the vortex shedding in the cylinder wake, *Phys. Rev. Lett.*, **113** (2014), 8.
- [69] V. Mantič-Lugo, C. Arratia, and F. Gallaire, A self-consistent model for the saturation dynamics of the vortex shedding around the mean flow in the unsteady cylinder wake, *Phys. Fluids*, **27** (2015), 074103.
- [70] P. Meliga, Harmonics generation and the mechanics of saturation in flow over an open cavity: a second-order self-consistent description, *J. Fluid Mech.*, **826** (2017), 503–521.

- [71] P. Meliga, E. Boujo, and F. Gallaire, A self-consistent formulation for the sensitivity analysis of finite-amplitude vortex shedding in the cylinder wake, *J. Fluid Mech.*, **800** (2016), 327–357.
- [72] I. Mezić, Spectral properties of dynamical systems, model reduction and decompositions, *Nonlinear Dyn.*, **41** (1–3) (2005), 309–325.
- [73] I. Mezić, Analysis of fluid flows via spectral properties of the Koopman operator, *Annu. Rev. Fluid Mech.*, **45** (1) (2013), 357–378.
- [74] M. Milano and P. Koumoutsakos, Neural network modeling for near wall turbulent flow, *J. Comput. Phys.*, **182** (1) (2002), 1–26.
- [75] M. Morzynski, W. Stankiewicz, B. R. Noack, F. Thiele, R. King, and G. Tadmor, Generalized mean-field model for flow control using a continuous mode interpolation, in *3rd AIAA Flow Control Conference*, American Institute of Aeronautics and Astronautics, 2006.
- [76] A. G. Nair and K. Taira, Network-theoretic approach to sparsified discrete vortex dynamics, *J. Fluid Mech.*, **768** (2015), 549–571.
- [77] B. R. Noack, K. Afanasiev, M. Morzyński, G. Tadmor, and F. Thiele, A hierarchy of low-dimensional models for the transient and post-transient cylinder wake, *J. Fluid Mech.*, **497** (2003), 335–363.
- [78] B. R. Noack, M. Morzynski, and G. Tadmor (eds.), *Reduced-Order Modelling for Flow Control*, Springer Vienna, 2011.
- [79] B. R. Noack, P. Papas, and P. A. Monkewitz, The need for a pressure-term representation in empirical Galerkin models of incompressible shear flows, *J. Fluid Mech.*, **523** (2005), 339–365.
- [80] J. Östh, S. Krajnović, B. R. Noack, D. Barros, and J. Borée, On the need for a nonlinear subscale turbulence term in POD models as exemplified for a high Reynolds number flow over an Ahmed body, *J. Fluid Mech.*, **747** (2014), 518–544.
- [81] O. K. Rediniotis, J. Ko, and A. J. Kurdila, Reduced order nonlinear Navier-Stokes models for synthetic jets, *J. Fluids Eng.*, **124** (2) (2002), 433–443.
- [82] D. Rempfer, On low-dimensional Galerkin models for fluid flow, *Theor. Comput. Fluid Dyn.*, **14** (2000), 75–88.
- [83] D. Rempfer and F. H. Fasel, Dynamics of three-dimensional coherent structures in a flat-plate boundary-layer, *J. Fluid Mech.*, **275** (1994), 257–283.
- [84] S. T. Roweis and L. K. Saul, Nonlinear dimensionality reduction by locally linear embedding, *Science*, **290** (5500) (2000), 2323–2326.
- [85] C. W. Rowley and S. T. M. Dawson, Model reduction for flow analysis and control, *Annu. Rev. Fluid Mech.*, **49** (1) (2017), 387–417.
- [86] C. W. Rowley, I. Mezic, S. Bagheri, and P. Schlatter, Spectral analysis of nonlinear flows, *J. Fluid Mech.*, **641** (2009), 115.
- [87] S. H. Rudy, S. L. Brunton, J. L. Proctor, and J. N. Kutz, Data-driven discovery of partial differential equations, *Sci. Adv.*, **3** (4) (2017), e1602614.
- [88] H. Schaeffer, Learning partial differential equations via data discovery and sparse optimization, *Proc. R. Soc. A*, **473** (2197) (2017), 20160446.
- [89] H. Schaeffer and S. G. McCalla, Sparse model selection via integral terms, *Phys. Rev. E*, **96** (2017), 2.
- [90] M. Schlegel and B. R. Noack, On long-term boundedness of Galerkin models, *J. Fluid Mech.*, **765** (2015), 325–352.
- [91] M. Schmidt and H. Lipson, Distilling free-form natural laws from experimental data, *Science*, **324** (5923) (2009), 81–85.
- [92] B. Schölkopf, A. Smola, and K.-R. Müller, Kernel principal component analysis, in *International Conference on Artificial Neural Networks*, pp. 583–588, 1997.
- [93] B. Schölkopf, A. Smola, and K.-R. Müller, Nonlinear component analysis as a kernel eigenvalue problem, *Neural Comput.*, **10** (5) (1998), 1299–1319.



- [94] M. Schumm, E. Berger, and P. A. Monkewitz, Self-excited oscillations in the wake of two-dimensional bluff bodies and their control, *J. Fluid Mech.*, **271** (1994), 17.
- [95] O. Semeraro, F. Lusseyran, L. Pastur, and P. Jordan, Qualitative dynamics of wave packets in turbulent jets, *Phys. Rev. Fluids*, **2** (2017), 9.
- [96] D. Sipp and A. Lebedev, Global stability of base and mean flows: a general approach and its applications to cylinder and open cavity flows, *J. Fluid Mech.*, **593** (2007).
- [97] D. Sipp, O. Marquet, P. Meliga, and A. Barbagallo, Dynamics and control of global instabilities in open-flows: a linearized approach, *Appl. Mech. Rev.*, **63** (3) (2010), 030801.
- [98] D. Sipp and P. J. Schmid, Linear closed-loop control of fluid instabilities and noise-induced perturbations: a review of approaches and tools, *Appl. Mech. Rev.*, **68** (2) (2016), 020801.
- [99] S. Sirisup and G. E. Kaniadakis, A spectral viscosity method for correcting the long-term behavior of POD models, *J. Comput. Phys.*, **194** (1) (2004), 92–116.
- [100] L. Sirovich, Turbulence and the dynamics of coherent structures. I – Coherent structures. II – Symmetries and transformations. III – Dynamics and scaling, *Q. Appl. Math.*, **45** (1987), 561–571.
- [101] M. Sorokina, S. Sygletos, and S. Turitsyn, Sparse identification for nonlinear optical communication systems: SINO method, *Opt. Express*, **24** (26) (2016), 30433–30443.
- [102] K. R. Sreenivasan, P. J. Strykowski, and D. J. Olinger, Hopf bifurcation, Landau equation, and vortex shedding behind circular cylinders, in *Forum on Unsteady Flow Separation*, vol. 1, pp. 1–13, American Society for Mechanical Engineers, Fluids Engineering Division New York, 1987.
- [103] W. Stankiewicz, M. Morzynski, K. Kotecki, and B. R. Noack, On the need of mode interpolation for data-driven Galerkin models of a transient flow around a sphere, *Theor. Comput. Fluid Dyn.*, **31** (2) (2016), 111–126.
- [104] P. J. Strykowski and K. R. Sreenivasan, On the formation and suppression of vortex ‘shedding’ at low Reynolds numbers, *J. Fluid Mech.*, **218** (1) (1990), 71.
- [105] G. Tadmor, O. Lehmann, B. R. Noack, L. Cordier, J. Delville, J.-P. Bonnet, and M. Morzyński, Reduced order models for closed-loop wake control, *Philos. Trans. R. Soc. A*, **369** (1940) (2011), 1513–1524.
- [106] G. Tadmor, O. Lehmann, B. R. Noack, and M. Morzyński, Mean field representation of the natural and actuated cylinder wake, *Phys. Fluids*, **22** (3) (2010), 034102.
- [107] J. B. Tenenbaum, V. De Silva, and J. C. Langford, A global parametric framework for nonlinear dimensionality reduction, *Science*, **290** (2000), 2319–2323.
- [108] V. Theofilis, Global linear instability, *Annu. Rev. Fluid Mech.*, **43** (2011), 319–352.
- [109] J. H. Tu, C. W. Rowley, D. M. Luchtenburg, S. L. Brunton, and J. N. Kutz, On dynamic mode decomposition: theory and applications, *J. Comput. Dyn.*, **1** (2) (2014), 391–421.
- [110] S. E. Turton, L. S. Tuckerman, and D. Barkley, Prediction of frequencies in thermosolutal convection from mean flows, *Phys. Rev. E*, **91** (4) (2015), 0403009.
- [111] L. Ukeiley, L. Cordier, R. Manceau, J. Delville, J. P. Bonnet, and M. Glauser, Examination of large-scale structures in a turbulent plane mixing layer. Part 2. Dynamical systems model, *J. Fluid Mech.*, **441** (2001), 61–108.
- [112] P. R. Vlachas, W. Byeon, Z. Y. Wan, T. P. Sapsis, and P. Komoutsakos, Data-driven forecasting of high-dimensional chaotic systems with long short-term memory networks, *Proc. R. Soc. A*, **474** (2018), 20170844.
- [113] Z. Wang, I. Akthar, J. Borggaard, and T. Ilescu, Proper orthogonal decomposition closure models for turbulent flows: a numerical comparison, *Comput. Methods Appl. Mech. Eng.*, **237–240** (2012), 10–26.
- [114] M. Wei and C. W. Rowley, Low-dimensional models of a temporally evolving free shear layer, *J. Fluid Mech.*, **618** (2009), 113–134.

- [115] N. Wiener, *Cybernetics or Control and Communication in the Animal and the Machine*, 1st, MIT Press, Boston, 1948.
- [116] M. O. Williams, I. G. Kevrekidis, and C. W. Rowley, A data-driven approximation of the Koopman operator: extending dynamic mode decomposition, *J. Nonlinear Sci.*, **25** (6) (2015), 1307–1346.
- [117] C. H. K. Williamson, Vortex dynamics in the cylinder wake, *Annu. Rev. Fluid Mech.*, **28** (1) (1996), 477–539.
- [118] A. Zebib, Stability of viscous flow past a circular cylinder, *J. Eng. Math.*, **21** (2) (1987), 155–165.
- [119] H.-Q. Zhang, U. Fey, B. R. Noack, M. König, and H. Eckelmann, On the transition of the cylinder wake, *Phys. Fluids*, **7** (4) (1995), 779–794.
- [120] W. Zhang, B. Wang, Z. Ye, and J. Quan, Efficient method for limit cycle flutter analysis based on nonlinear aerodynamic reduced-order models, *AIAA J.*, **50** (5) (2012), 1019–1028.
- [121] Z. Zhang and J. Wang, MLLE: modified locally linear embedding using multiple weights, in B. Schölkopf, J. C. Platt, and T. Hoffman (eds.), *Advances in Neural Information Processing Systems 19*, pp. 1593–1600, MIT Press, 2007.
- [122] Z. J. Zhang and K. Duraisamy, Machine learning methods for data-driven turbulence modeling, in *22nd AIAA Computational Fluid Dynamics Conference*, American Institute of Aeronautics and Astronautics, 2015.
- [123] B. J. A. Zielinska and J. E. Wesfreid, On the spatial structure of global modes in wake flow, *Phys. Fluids*, **7** (6) (1995), 1418–1424.
- [124] P. Benner, S. Grivet-Talocia, A. Quarteroni, G. Rozza, W. H. A. Schilders, and L. M. Solveira, *Model Order Reduction. Volume 1: System- and Data-Driven Methods and Algorithms*, De Gruyter, 2020.

