*l*₁-based calibration of POD-Galerkin models of two-dimensional unsteady flows

Riccardo Rubini^{*}, Davide Lasagna and Andrea Da Ronch University of Southampton, Faculty of Engineering and Physical Sciences, United Kingdom

This paper discusses a physics-informed methodology aimed at reconstructing efficiently the fluid state of a system. Herein, the generation of an accurate reduced order model of two-dimensional unsteady flows from data leverages on sparsity-promoting statistical learning techniques. The cornerstone of the approach is l_1 regularised regression, resulting in sparselyconnected models where only the important quadratic interactions between modes are retained. The original dynamical behaviour is reproduced at low computational costs, as few quadratic interactions need to be evaluated. The approach has two key features. First, interactions are selected systematically as a solution of a convex optimisation problem and no a priori assumptions on the physics of the flow are required. Second, the presence of a regularisation term improves the predictive performance of the original model, generally affected by noise and poor data quality. Test cases are for two-dimensional lid-driven cavity flows, at three values of the Reynolds number for which the motion is chaotic and energy interactions are scattered across the spectrum. It is found that: a) the sparsification generates models maintaining the original accuracy level but with a lower number of active coefficients; this becomes more pronounced for increasing Reynolds numbers suggesting that extension of these techniques to real-life flow configurations is possible; b) sparse models maintain a good temporal stability for predictions. The methodology is ready for more complex applications without modifications of the underlying theory, and the integration into a cyber-physical model is feasible.

I. Introduction

Future air and sea vehicles will be self-aware. The concept of self-awareness [1] requires the ability to monitor the vehicle internal state, to sense the surrounding environment, and to assess its capabilities currently and to project them into the future. This requires integration of multiple individual technologies and development of integrated technologies [2]. Self-aware vehicles have immediate and future applications, i.e. on-demand mobility, and include the full spectrum of aircraft, from small unmanned vehicles to large transport aircraft. Among the required technologies identified in [1] for a self-aware aircraft concept, the present work focuses on the development of physics-informed models of the current system states and the capability to forecast the temporal evolution of the system at low computational costs for on-board, real-time applications. Generally, the development of required technologies for self-aware vehicles is driven by nature, as discussed thoroughly in the review work [2]. A recent and excellent contribution on bio-inspired concepts and existing flying prototypes is presented in [3]. Despite the need to advance a multitude of individual technologies, morphing continues to be one of the most popular research venues [4]. For example, Ref. [5] presents a numerical study of morphing wings that are compared to a conventional wing in terms of aerodynamic performance. On a different but complementary subject, Ref. [6] proposed a novel load prediction tool that extends analytical theories. Predicted loads are parametrised by wing planform geometry, making the tool suitable for exploring a large design space in topology optimisation studies or for predicting the system response (aerodynamics, flight dynamics) when damage characteristics are detected in real-time by sensor networks. The seminal work in Ref. [7] optimised the external aerodynamic surface of an aircraft over an entire flight trajectory, going beyond state-of-the-art single or multi-point optimisation. The optimisation leverages on morphing technologies that are coordinated with the appropriate stage of the flight path, providing superior performances compared to a fixed wing aircraft.

This paper provides an overview of a recent technology development effort aimed at identifying the aerodynamic / fluid states of a system with no a-priori aerodynamic knowledge. The approach is general in its formulation and shall

^{*}PhD candidate, email: r.rubini@soton.ac.uk

see applications in different fields than those documented herein. To demonstrate the proposed approach, a canonical fluid mechanics problem is used as test case. The extension to a fully three–dimensional problem, i.e. flow around a membrane wing, or to a coupled problem, i.e. fluid–structure interaction, is straightforward and entails no modifications to the underlying theory.

Reduced order modelling (ROM) techniques have gained significant importance both as analysis and prediction tools [8–11]. This increase in popularity can be attributed to the availability of large amount of data from experimental and numerical sources [12, 13]. A well established approach consists in projecting the governing equations onto the the subspace spanned by the first N modes obtained from a suitable modal decomposition [14, 15]. The choice of the modal decomposition is not unique [15, 16], but depends upon the aspects of the flow one wants to highlight. One of the most popular choices is Proper Orthogonal Decomposition (POD), mainly for its properties of energetic optimality and the consequent ability of generating compact and parsimonious representations of the original dataset [15, 17, 18].

Galerkin projection of the governing equations onto the POD subspace results in a system of coupled ordinary differential equations (ODEs) with a quadratic nonlinearity described by a third order tensor Q_{ijk} , representing the quadratic interaction between pairs of different modes. Due to the non-local nature of the spatial modes and no *a priori* assumptions on the flow field all entries of Q_{ijk} are generally non-zero. This results in a computational cost scaling as $O(N^3)$ for a model constructed with N modes. Models that include all possible quadratic interactions are referred to as dense or densely connected.

When the Reynolds number grows and the range of energy-relevant spatial scales widens [19], a larger number of modes is required to reach a satisfactory energy resolution. As a consequence, a complete evaluation of the nonlinear interactions can quickly become prohibitive. This limits the use of Galerkin models in real life applications, where the number of modes needed to reconstruct a satisfactory amount of kinetic energy might be large. A possible solution is to construct a reduced order model considering only a small amount of modes. If the POD is used this is equivalent to simulate the large scales of the flow and models the smallest ones. Commonly, this approach results in lack of accuracy and temporal stability of the corresponding ROM [9]. The main problem arising from the truncation process is the elimination of the smallest and less energetic scales of the flow [20] responsible of the dissipation of the kinetic energy generated at the larger scales. This is a well studied issue and it is usually tackled with the introduction of a sub-grid dissipation model [9, 14]. A second aspect undermining the accuracy of reduced order models is the lack of quality of the numerical or experimental data the models are generated from. Often times, numerical simulations are not carried out on complete dynamical representations of the flows (subgrid modelling in LES simulations, for example) while experimental results can be corrupted or affected by measurement errors [12]. The joint effect of truncation of poor data quality can make obtaining accurate short time predictions a challenging task.

This work aims to extend the current literature in term of calibration techniques for reduced order models. Recently, various approaches have been proposed to calibrate and improve the accuracy of the reduced order models [12, 21–23]. They all share the idea of leveraging the knowledge of the true temporal dynamics of the flow in order to adjust the numerical value of the coefficients in the ODEs system through an optimisation procedure. A simple approach consists in performing least squares regression on the system coefficients. This approach can provide satisfactory results in case of small ROMs and good quality data to train the algorithm. In real life configurations, as observed by [12, 24], small and poor quality datasets are very common. This makes the use of a pure least squares regression technique without any regularisation term strongly discouraged. Cordier and coworkers [12, 25] proposed an approach based on a Tikhonov penalisation term to control the numerical values of the coefficients in the reduced order model. This approach is a well established statistical learning technique used to regularise least squares regression problems where the database or feature matrix has poor numerical conditioning [24, 26]. Although the Tikhonov regularisation has been proved successful for relatively small models a possible drawback is that this regularisation preserves in the tuned model all the coefficients present in the original one. Since for POD-based Galerkin models the number of coefficients grows as $O(N^3)$, the resulting least squares problem might require a prohibitive number of snapshots to perform a meaningful regression without incurring in overfitting problems. The proposed solution to overcome this limitations exploits the observation that the dynamics of the flow is not equally influenced by all the coefficients contained in Q_{ijk} but, on the contrary, it is mainly governed by a small number of flow structures generating a reduced order model where only a subset of interactions is relevant for the dynamics [27–29, 29, 30]. Therefore, an algorithm able to automatically identify the subset of relevant interactions could be beneficial to improve the computational performances and facilitate the calibration of ROMs of complex flows.

Work in this direction has been recently done by Brunton and coworkers [31], who developed the SINDy framework to identify equations generated by a system that already posses a sparse structure. A further extension of the SINDy has been proposed by [32] introducing energy preserving constraints to improve the prediction accuracy of POD-based

Galerkin models without the need of high-fidelity solver to project the Navier–Stokes equations. In this paper, the aim is to extend these ideas using the l_1 sparsity promoting regression to systematically identify the leading coefficients in *a priori* dense Galerkin systems. To this end, we identify the coefficients associated with the interactions between modal structures that play the dominant role in the dynamics. Once this subset is identified, tuning the numerical values of the coefficients to improve the performances of the model is more convenient. The main advantage of the proposed approach is that only a small subset of coefficients is retained after sparsification and, as a result, the tuning of the coefficients requires a considerably smaller dataset.

The present paper is structured as follows. The first part of section II summarises the methodology utilised to generate Galerkin reduced order models. Subsequently, the l_1 based regression is outlined and a discussion on how it can be leveraged to generate sparse representation of originally dense models is presented. In section III, we demonstrate this approach on a family of large POD-based reduced order models of chaotic solutions of lid-driven cavity flows at different Reynolds numbers for which direct numerical simulation (DNS) is performed. Results are divided into two main sections. In the first, the focus is on the effects of the complexity of the reduced order model on the sparsification. Second, the predictive capabilities of the sparse models are compared with their dense counterparts and with DNS simulations.

II. Methodology

A. Galerkin-based reduced order modelling

A modal decomposition represents the fluctuating part of the flow field $\mathbf{u}'(t, \mathbf{x}) = \mathbf{u}(t, \mathbf{x}) - \bar{\mathbf{u}}(\mathbf{x})$, where the $\bar{\mathbf{u}}(\mathbf{x})$ is the mean flow, as a finite sum of *N* temporal and spatial modes as

$$\mathbf{u}(t,\mathbf{x}) = \bar{\mathbf{u}}(\mathbf{x}) + \sum_{i=1}^{N} a_i(t)\boldsymbol{\phi}_i(\mathbf{x}).$$
(1)

The coefficients $a_i(t)$ and $\phi_i(\mathbf{x})$, i = 1, ..., N are the temporal and spatial modes, respectively. If the modes form an orthonormal basis set, the fluctuating kinetic energy is defined as

$$E(t) = \frac{1}{2} \int_{\Omega} \mathbf{u}'(t, \mathbf{x})^2 d\Omega = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} a_i(t) a_i(t).$$
(2)

Since the spatial modes incorporate automatically the boundary conditions of the problem, this provides a suitable basis to generate reduced order models of flows in arbitrary geometries. Reduced order models are then derived by projecting the governing equations onto the subspace defined by the first N modes. Assuming appropriate boundary conditions [14] and orthogonality of the spatial modes, this results in a system of coupled nonlinear ODEs

$$\dot{a}_i(t) = C_i + \sum_{j=1}^N L_{ij} a_j(t) + \sum_{j=1}^N \sum_{k=1}^N Q_{ijk} a_j(t) a_k(t), \qquad i = 1 \dots, N,$$
(3)

defining the temporal evolution of the coefficients $a_i(t)$. The tensors of coefficients C_i , L_{ij} and Q_{ijk} are defined in Noack et al. [14] and describe the constant, linear and quadratic interactions between modal structures. Here, we focus on the nonlinear interactions term in (3), $Q_{ijk}a_j(t)a_k(t)$. Without any *a priori* assumption on the flow physics and on the boundary conditions, due to the global nature of the spatial basis functions $\phi_i(\mathbf{x})$ all entries of Q_{ijk} are different from zero. This means that the evaluation of \dot{a}_i at any instant of time requires the evaluation of all the linear and quadratic interactions between the mode itself $a_i(t)$ and the other modes $a_j(t)$ with j = 1... N. Due to the mathematical structure, the number of coefficients evaluated at every time step and, consequently, the computational complexity of the model grows with the number of modes as $O(N^3)$.

Since we know that flow structures associated to $a_j(t)a_k(t)$ do not interact in an arbitrary way but some interactions are favoured over others, we assume that not all the entries of Q_{ijk} have the same importance in describing the evolution of \dot{a}_i . Therefore, we aim at developing a systematic procedure to enable us identify a sparse tensor, Q_{ijk}^s , that is a good approximation of the original tensor Q_{ijk} , in the sense that at any time

$$Q_{ijk}^{s}a_{i}(t)a_{j}(t) \simeq Q_{ijk}a_{i}(t)a_{j}(t).$$

$$\tag{4}$$

The identification of the subset of most relevant interactions is beneficial allowing to reduce the number of active coefficients in the system (3) without affecting the accuracy of the system itself. We will refer to this new set of coefficients as sparse or sparsely connected reduced order model.

B. Sparse regression

We use sparse regression to identify the most important quadratic interactions inside the set of equations (3). Assuming N_T snapshots of the velocity field are available from simulation, we arrange the projections $a_i(t_j) = a_i^j$, with i = 1, ..., N and $j = 1, ..., N_T$, into the data matrix $A \in \Re^{N_T \times N}$, where the *i*-th column represents the evolution of the *i*-th mode, while the *j*-th row contains the state vector at time t_j . We exploit the polynomial structure of the Galerkin system (3) to construct the database matrix $\Theta(A) \in \Re^{N_T \times q}$

$$A = \begin{pmatrix} a_1^1 & \dots & a_N^1 \\ \vdots & \ddots & \vdots \\ a_1^{N_T} & \dots & a_N^{N_T} \end{pmatrix} \qquad \Theta(A) = \begin{pmatrix} 1 & a_1^1 & \dots & a_N^1 & a_1^1 a_1^1 & \dots & a_n^1 a_n^1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & a_1^{N_T} & \dots & a_N^{N_T} & a_1^{N_T} a_1^{N_T} & \dots & a_N^{N_T} a_N^{N_T} \end{pmatrix},$$
(5)

where $q = (N + 1) + N \times (N + 1)/2$ is the total number of interactions present in the model (3), the sum of constant, linear and quadratic interactions. The number of quadratic coefficients is N(N + 1)/2 instead of N^2 since the interaction between the mode *i* and *j* is considered only once in the definition of $\Theta(A)$. This avoids generating columns of $\Theta(A)$ that are linearly dependent, which would result in numerical issues in the solution of the regression (see e.g. Perret et al. [33]). Similarly, we construct the modal acceleration matrix $\dot{A} \in \Re^{N_T \times N}$, containing the time derivative of the temporal coefficients by projecting the modes $\phi_i(x)$ on the corresponding snapshots of the Eulerian acceleration field $\partial_t \mathbf{u}(t_j, \mathbf{x})$. Arranging the coefficients tensors C_i , L_{ij} and Q_{ijk} of the *i*-th mode into a vector $\beta_i \in \Re^q$, the *i*-th equation of system (3) is rewritten as

$$\hat{A}_i = \Theta(A)\beta_i,\tag{6}$$

where the dynamics associated with the mode \dot{a}_i is contained in the *i*-th column of the matrix \dot{A}_i .

The assumption that not all interactions in the nonlinear term $Q_{ijk}a_ja_k$ have the same importance is equivalent to assuming that some coefficients in β_i associated with the columns of $\Theta(A)$ can be set to zero without affecting the value of the corresponding column \dot{a}_i . Thus, the idea is to prune some of the entries of β_i , corresponding to columns of Θ , by solving the optimisation problem

$$\hat{\beta}_{i} = \underset{\beta_{i}}{\operatorname{argmin}} \left\{ \underbrace{\frac{||\Theta(A)\beta_{i} - \dot{A}_{i}||_{2}^{2}}{\operatorname{Least-Squares}} + \underbrace{\gamma ||\beta_{i}||_{1}}_{\operatorname{Penalization}} \right\},$$
(7)

depending on the regularisation weight γ . This is a l_1 regression problem also known as LASSO (Least Absolute Shrinkage Selection Operator) regression [34]. It is a well studied convex optimisation problem for which numerous efficient routines have been developed. The objective function in (7) is composed of two terms. The first is a least-squares term that penalises the distance in modal space between the acceleration of the reconstructed system, the term $\Theta(A)\beta_i$, and the objective \dot{A}_i . The second term is a penalisation on the regression coefficients weighted by the variable coefficient γ . It encourages sparsity in the solution shrinking to zero the entries of β_i that have the least contribution to the dynamics of $\dot{a}_i(t)$. More specifically, the l_1 norm is proven to be the best convex approximation of the cardinality operator that counts the nonzero entries of β_i [35].

After the solution of (7) is obtained, we define the relative reconstruction error ϵ as

$$\epsilon = \sum_{i=1}^{N} \frac{||\Theta(A)\beta_i - \dot{A}_i||_2^2}{||\dot{A}_i||_2^2},$$
(8)

where the reconstruction error is normalised with respect to the corresponding amplitude in order to balance the error across the spectrum of modes. We also define the global density as

$$\rho = \frac{1}{Nq} \sum_{i=1}^{N} card(\beta_i).$$
(9)



Fig. 1 Lid-driven cavity flow: snapshots of the vorticity field for different Reynolds numbers. In (a), $Re = 1 \times 10^4$; in (b), $Re = 2 \times 10^4$; and in (c), $Re = 5 \times 10^4$.

where the cardinality operator $card(\beta_i)$ counts the number of entries of β_i different from zero. The cardinality of β_i can be also defined in term of l_0 norm $||\beta_i||_0$ providing a formal definition of sparsity of the vector β_i . A family of models can be generated by varying the regularisation weight γ . For small weights, dense models with good prediction accuracy are obtained and *vice versa* for large weights. We can display this family of models on a $\rho - \epsilon$ plane, producing a curve referred to as sparsification or $\rho - \epsilon$ curve. Since only a subset of interactions is relevant, many of the entries of Q_{ijk} will be shrunk to zero when increasing the regularisation weight, without affecting significantly the error ϵ . This produces a sweet spot in the sparsification curve, defining an 'optimal' value of γ and a consequent optimal value of the model density $\rho = \rho_{opt}$, corresponding to models with low density and low reconstruction error where only the relevant interactions for the flow physics are preserved.

III. Results

In this section this methodology is applied on the two dimensional flow developing inside a lid driven square cavity. First, the flow characteristics and the modal decomposition is presented. Second, the different dense ROMs are generated by Galerkin projection. Lastly, the performance of the sparse models compare with the correspondent dense model and the baseline DNS simulation are analysed.

A. Test case and modal decomposition

The Reynolds number is defined as Re = LU/v, where L and U are, respectively, the cavity edge and the lid velocity. The kinematic viscosity is denoted by v. In this work the edge length and the lid velocity are set equal to 1 in non-dimensional units and different values of Re are obtained varying the values of v. More specifically, three different values of Reynolds numbers are considered. Namely, 10^4 , 2×10^4 and 5×10^4 . The dynamics in these conditions is chaotic with increasing complexity for increasing Reynolds number, as shown by Auteri et al. [36]. The domain is defined in nondimensional Cartesian coordinates $\mathbf{x} = (x, y)$ and the velocity field is defined by the components $\mathbf{u} = (u, v)$. The simulations are performed in OpenFOAM using a modified version of the solver i coFOAM that also outputs the snapshots of the Eulerian acceleration $\partial_t \mathbf{u}(t, \mathbf{x})$ to compute the modal acceleration \dot{a}_i needed in (7) more accurately. The convective and viscous terms are spatially discretised with a second order finite volume technique and the temporal term with a semi-implicit Crank-Nicholson scheme.

Figure 1 shows the vorticity field $\omega = (\nabla \times \mathbf{u})_z$ at the same time instant for the three Reynolds numbers considered. For the lowest Reynolds number, a strong shear layer separates the main vortex from the recirculation areas in the cavity corners. As the Reynolds number increases, the vortex on the lower right corner is shed and advected by the mean flow along the shear layer. This phenomenon is accompanied by strong, quasi-periodic bursts in the turbulent kinetic energy. For the highest value of Reynolds number, the set of vortices breaks the shear layer while being transported downstream by the mean flow. A preliminary study was carried out to ensure that the average flow quantities reached statistical convergence. This is an important aspect to satisfy to obtain reliable cross-validated results in the regression. To this end, the eigenvalues of the POD decomposition were monitored. It was found that a database of length T = 100 time units is adequate for the case at $Re = 1 \times 10^4$, increasing to T = 200 time units at $Re = 2 \times 10^4$, and finally to T = 300



Fig. 2 The first 50 eigenvalues of the POD decomposition in (a); the vorticity fields of the first and most energetic spatial POD mode at $Re = 10^4$ in (b), $Re = 2 \times 10^4$ in (c), and $Re = 5 \times 10^4$ in (d).

time units at $Re = 5 \times 10^4$.

Figure 2-(a) reports the first 50 eigenvalues of the POD decomposition of the three test cases already presented in figure 1. It is observed that the eigenvalues increase monotonically for increasing Reynolds number, as turbulent fluctuations contain increasingly more energy. One also observes that, for the two lowest Reynolds numbers, consecutive eigenvalues appear in pairs with similar value. This feature reflects the presence of an unsteady motion in a quasi-periodic regime, characterised by pairs of spatially-similar structures shifted in space, describing the evolution of a wave travelling along the shear layer. Evidence of this feature can be found in figure 2-(b) and (c). Similar structures have already been observed in [37] for Koopman eigenmodes of cavity flow at similar flow conditions. At the highest Reynolds number, $Re = 5 \times 10^4$, the flow contains a wider hierarchy of modes. The eigenvalue spectrum (blue triangles) has a slower decreasing trend, without pairs of adjacent eigenvalues. The most energetic spatial eigenmode in figure 2-(d) captures the generation and shedding of a vortex on the lower right corner, with no significant activity in the shear layer. This observation is confirmed by inspecting the temporal evolution of the velocity field, revealing the formation of a high energy vortex on the lower right corner transported downstream by the mean flow. First, the temporal correlation matrix between two snapshots defined at two instants of time, t_i and t_j , defined as $S_{ij} \in \mathbb{R}^{N_t,N_t}$, with entries

$$S_{ij} = \frac{1}{N_t} \left(\mathbf{u}'(t_i, \mathbf{x}), \mathbf{u}'(t_j, \mathbf{x}) \right), \tag{10}$$

where the brackets (\cdot, \cdot) denote the inner product defined into the space. To classify reduced order models on the basis of the resolved turbulent kinetic energy for different flows we define the normalised cumulative sum of the eigenvalues of the correlation matrix λ_i as

$$e(n) = \frac{\sum_{i=1}^{n} \lambda_i}{\sum_{i=1}^{N_T} \lambda_i} \tag{11}$$

describing the fraction of the kinetic energy captured by the first *n* term of the expression (1). The increasing complexity of the flow is directly reflected to the fact that the number of modes required to reconstruct the same amount of kinetic energy increases non linearly. Here, we consider three families of models for every test case, resolving 90%, 95% and 99% of the kinetic energy, respectively. The number of modes required to reach the requested energy resolution is shown in Table 1.

e(n)	0.9	0.95	0.99
$Re = 10^{4}$	8	12	36
$Re = 2 \times 10^4$	20	35	75
$Re = 5 \times 10^4$	60	112	290

Table 1 Number of modes required as a function of Reynolds number and energy resolution.

Once the coefficients C_i , L_{ij} and Q_{ijk} (3) are generated by Galerkin projection, the resulting system of ODEs can be integrated in time given one initial condition. In this work, an explicit time integration scheme with a time step of



Fig. 3 Time evolution of turbulent kinetic energy (as in Eq. (2)) at $Re = 2 \times 10^4$ for reduced order models with different energy resolution and form the DNS reference data.

 $\Delta t = 10^{-3}$ time units has been used. All the systems have been integrated for a time span equal to T = 100 time units, starting from an initial condition obtained from the DNS choosing a snapshot in the region of statistical convergence of the solution. Figure 3 shows the temporal evolution of the turbulent kinetic energy as defined in (2) for three different ROMs resolving different amounts of kinetic energy and the baseline DNS. As expected, it is observed that the models reconstructing a lower amount of kinetic energy tends to overshoot the reference value of the turbulent fluctuations. This is a well know issue of reduced order modelling already discussed in [14, 20] for POD-based models and is a direct consequence of the impossibility of dissipating the kinetic energy produced at the large scales in the small scales eliminated in the truncation process. As a result, the solution is not able to satisfy the energy balance of the original flow, presenting a higher energy in the first few modes associated with larger flow scales. More specifically, figure 3 shows that only the e(n) = 0.99 model contains a sufficient number of modes to provide predictions comparable with the DNS. These results highlight that, in order to obtain reliable reduced order models, a large number of modes needs to be included, generating complex and computationally expensive models. The idea of sparsification consists in pruning the least important coefficients of the tensor Q_{ijk} without truncate any mode from the original system. The desired outcome is to obtain accurate but computationally efficient models constructed by a large amount of modes (spatial scales) but containing only the relevant interactions.

B. Sparse Regression

The sparse regression methodology was applied to the reduced order models with e(n) = 0.9, 0.95 and 0.99, for the three Reynolds numbers. Since the relationship between number of modes and energy resolution is not linear and the size of the database matrix, $\Theta(A)$, increases quadratically with *n*, the number of possible interactions *q* can quickly become larger than the number of available snapshots, N_T . This is a well–known problem in statistical learning and requires cross–validation methods to avoid data overfitting [24]. This work employs the *K*–fold validation procedure available in the *sklearn* library [38]. The database is divided into K = 10 folds. The model is initially trained on a subset of the entire database, obtained by removing the *k*–th fold. Subsequently, the model is tested on the *k*–th fold which was excluded for the model generation. The procedure is repeated *K* times. To quantify the quality of the model predictions, the mean and standard deviation of the cross–validated reconstruction error, Eq. (8), are calculated over the folds.

The sparsification procedure consists in solving problem (7) for increasing values of the regularisation weight γ . Then the corresponding value of the density ρ and reconstruction error ϵ as defined in (9),(8) is computed and displayed on the $\rho - \epsilon$ plane. Figure 4 summarises the results of the sparsification procedure applied to the nine models considered. In each panel, the value of the mean cross validated error ϵ against the density ρ is displayed for the three Reynolds numbers. The value of the resolved kinetic energy increases from left to right showing the results for e(n) = 0.9, 0.95and 0.99 in panel (a,b,c), respectively. When low weights are used (points in the right part of the graphs), dense systems with low prediction accuracy are obtained. The opposite is true for large weights, identifying points in the left part of the graphs. As postulated, since a set of coefficients in the tensor Q_{ijk} is predominant, the curves present an initial plateau for high densities where it is possible to remove coefficients from the system (3) without significantly affecting the reconstruction error ϵ .



Fig. 4 Mean of the cross-validated reconstruction error against the density of the system, ρ : in (a), e(n) = 0.90; in (b), e(n) = 0.95; and in (c), e(n) = 0.99.



Fig. 5 In (a), time evolution of turbulent kinetic energy for the three sparse models highlighted by red circles in Figure 4-(b) compared to reference DNS; in (b), optimal density for all models considered in this study.

Generally, the sparsification curves in figure 4 shown two emerging trends. First, the value of the reconstruction error ϵ in the plateau decreases monotonically as the energy resolution increases, since more modes participates in capturing the dynamics of the fluctuations. Secondly, and more importantly, the optimal density ρ_{opt} decreases when the Reynolds number is increased. This can be observed qualitatively in figure 4 and more quantitatively later on in figure 5-(b) showing the trends of ρ_{opt} against the value of the Reynolds number. These results show that more complex models can be more efficiently sparsified. There are two further considerations. For the model with e(n) = 0.9 at $Re = 1 \times 10^4$, no evident plateau exists in the ϵ - ρ curve, indicating that the reconstruction error increases as soon as coefficients are removed from the system. This phenomenon was also observed in [31], and alerts that a successful sparsification may depend on the basis used in the model generation as well as the mathematical structure of the model. The second consideration relates to Figure 4-(c). The curve for the model with e(n) = 0.99 reveals a sharp elbow point, with an optimal density $\rho_{opt} \sim 0.1$. After a short plateau, the reconstruction error increases, worsening predictions for denser models. This behaviour is due to the small amount of data confronted with the complexity of the system, and shows the importance of using cross-validation for systems of high complexity. It is important to underline that this effect is present only for the case at higher Reynolds number and energy resolution. Increasing the size of the dataset for the other models does not change significatively the shape of the curves.

To understand the effect of the sparsification on the temporal characteristics of the models, we analyse the temporal evolution of Eq. (2) for three different models chosen along the curves corresponding to e(n) = 0.95 and $Re = 2 \times 10^4$, shown in Figure 5-(a). The thee models corresponds to the three red circles in Figure 4-(b) corresponding to the full system, a system located near the elbow point and one located in an area of very low density and with a higher error. Panel (a) shows that the behaviour of the dense system is qualitatively similar to that obtained by projection shown



Fig. 6 Time evolution of the fluctuation kinetic energy E(t), (left) and spectral density function of E(t) reported against the nondimensional frequency St = fL/U, (right). In (a, b), $Re = 1 \times 10^4$; in (c, d), $Re = 2 \times 10^4$; and in (e, f), $Re = 5 \times 10^4$.

in figure 3, leading to an overestimation of the kinetic energy of the flow. This result is expected since for $\rho = 1$ the problem (7) is reduced to a pure least squares regression. More interestingly, decreasing the density of the model moving towards the sweet spot in the sparsification curves, the models' predictions move closer to the DNS reference. Both sparse models seem to predict quite accurately the average kinetic energy level. However, a visual examination of the reconstructed vorticity field shows that for $\rho = 0.1$ the main features of the flow are not reconstructed faithfully. Hereafter, we will only consider the optimally sparse model located in the proximity of the elbow point. Lastly, we observe that the value of ρ_{opt} decreases as the Reynolds number increases. This is quantitatively shown in figure 5-(b) reporting the value of ρ_{opt} as a function of the Reynolds number. This effect is due to the broader range of scales in a model for increasing Reynolds number, and the weaker interactions among modes that are far in the spectrum. A qualitatively similar behaviour is observed increasing the energy resolution of the system. More importantly, this general trend shows that the effectiveness of the sparsification increases as the complexity of the model increases.

Figure 6 shows time histories of the turbulent kinetic energy (left panels) and the corresponding power spectral densities obtained by temporal simulation of models at e(n) = 0.95 (right panels) for the three different Reynolds numbers considered. DNS data is used as reference for two reduced order models: an optimally sparse model (black dashed curves), and a dense reduced order model (black curves). It is found that dense models overestimate the turbulent kinetic energy compared to DNS by several orders of magnitude. This is not unexpected because the model, not resolving the full range of scales of the flow, does not correctly model the energy dissipation occurring at the small scales. A confirmation of this observation can be seen on the right panels of Figure 6, where the dense ROMs over-predict the energy content at all scales and for all Reynolds numbers. On the other hand, we observe that for the sparse models the level of the predicted kinetic energy is, on average, in agreement with the reference value.

To better visualise how the sparsification affects the predicted average kinetic energy, the ratio k_r between the average reference kinetic energy and that predicted by the model is considered. The average has been computed from



Fig. 7 The ratio k_r for models with different energy resolutions e(n). The empty symbols represent dense systems while the full symbols represent the optimally sparse systems. Results for different Reynolds numbers $Re = 10^4, 2 \times 10^4$ and 5×10^4 are shown in panels a), b), and c), respectively

 $t \ge 40$, corresponding to the time needed by the solution to reach statistical convergence as can be observed in figure 6. Results are reported in figure 7. Each panel displays data for six models, three dense models obtained by projection (empty triangles), and the corresponding sparse models with $\rho = \rho_{opt}$ (full black circles), as shown in figure 4. The Reynolds number increases moving from panel (a) to panel (c). A common trend for the three different flow conditions is observed. The models with a low energy resolution overestimate the fluctuating kinetic energy by two orders of magnitude. As the energy resolution of the model increases the value of k_r tends to values close to 1. This is an expected result since the number of spatial structures included in the model increases allowing the model to better describe the energy dissipation scales. Interestingly, for all sparsified models, the ratio k_r is close to 1. This means that on average the sparse system predict the right average amount of kinetic energy. In addition, we observe that the ratio k_r is almost constant as the energy resolution increases. This means that regardless the energy resolution chosen the optimisation problem (7) is able to provide correct predictions on the kinetic energy of the system. This is a direct consequence to the fact that the target of the acceleration is the modal accelerations $\dot{a}_i(t)$ computed from DNS simulation.

C. Reconstructed Flow Field

Once a solution of (7) is obtained, rearranging the coefficients vector β_i into a new set of matrices C_i^s, L_{ij}^s and Q_{ijk}^s the sparse representation of (3) is obtained. The new dynamical system can be integrated to obtain the evolution of a new set of $a_i(t)$ that are used to reconstruct the flow field according to Eq. (1) since the spatial modes remain unchanged. Figure 8 shows the reconstructed vorticity field ω for the optimally sparse models with e(n) = 0.95. Every column contains results for a value of Reynolds number increasing from left to right. The top row shows the result of the time integration of the dense system while the bottom row shows the ones of the corresponding sparse model. From a visual analysis, we observe that the overshoot of kinetic energy observed in the dense models in figure 6 corresponds to non-physical large amplitude oscillations in the shear layer, as shown in panels (a,b), similarly for the model at highest Reynolds number the vortex in the lower right corner is amplified as well as shown in panel (c). This behaviour is due to an overestimation of the amplitude corresponding to the large and most energetic modes leading to an exaggeration of the low indexes modes as shown in figure 1-(b,c,d). More specifically, the sparse models reconstruct quite accurately the topology of the original flow field reproducing the shear layer and the vortex detachment for the lowest Reynolds numbers considered in panel (d) and (e). Similarly, for the highest value of the Reynolds number (panel (f)), the sparse model is able to describe the formation and the shedding of the vortex produced in the bottom right corner of the cavity and subsequently shed and advected along the shear layer by the mean flow.

A more in depth analysis on the dynamics of the velocity fluctuations can be done analysing the value of the time average of the Reynolds stress $\overline{u'v'}$. This term is particular relevant for our analysis since the sparsification is performed on the set of equations (3) describing only the fluctuating part of the velocity field, while the mean field is left unchanged. Figure 9 shows the profile of the Reynolds stress, $\overline{u'v'}$, along a vertical line located at x = 0.5 for increasing values of the Reynolds numbers, from panel (a) to (c). The grey continuous line is the reference from DNS while the optimally sparse model is represented as black dashed line. Results of the dense ROMs are reported as black dotted line in the



Fig. 8 Instantaneous snapshots of the vorticity field ω . Each column reports data for a different Reynolds number $(1 \times 10^4, 2 \times 10^4 \text{ and } 5 \times 10^4, \text{ from left to right})$. The upper row reports results for the dense reduced order model, while the lower row for the optimally sparse reduced order model.



Fig. 9 Profiles of the Reynolds stress $\overline{u'v'}$ from simulation and from time integration of the sparse and dense ROMs along the line x = 0.5: in (a), for $Re = 10^4$; in (b), $Re = 2 \times 10^4$; and in (c), $Re = 5 \times 10^4$.

background as reference. Since they are orders of magnitude larger than the ones observed in the DNS we chose a range of values on the *x* axis in order to better visualise the results of the DNS and the sparse ROM.

For the two lower Reynolds numbers, panels (a,b), oscillations in the the stress $\overline{u'v'}$ can be observed, becoming more intense as the Reynolds number increases. In addition, outside the shear layer the stress $\overline{u'v'}$ drops to zero except near the cavity lid. This result is expected since from a visual examination of figure 1 and 8 we observe that outside the shear layer the flow is mainly stationary and presents quasi-periodic fluctuations. A different behaviour is shown in panel (c) for the highest Reynolds number. As already observed in figure 1 and 8, in this case the shedding of the bottom-right vortex generates a much more complex flow field, with oscillations extended also near the central area of the cavity. For this flow configuration we observe that although the general distribution is reproduced the values are locally more different with respect to the other two test cases.

IV. Conclusions

In this work a systematic data-driven methodology to perform sparsification and calibration of Galerkin based reduced order models of turbulent flows was utilised. The cornerstone of this approach consists in transforming the sparsification of a ROM into a convex optimisation problem independent from *a priori* knowledge on the flow physics. This methodology was applied to chaotic solutions of lid-driven square cavity flow. A family of POD-based Galerkin models with increasing complexity and spanning a range of Reynolds numbers and different energy resolution was created. Results show that pruning interactions from the original system does not affect the reconstruction error. The $\rho - \epsilon$ plane is used to identify the trade off between reconstruction error ϵ and system density ρ_{opt} decreases monotonically, showing that the sparsification becomes more efficient. Arguably, this behaviour is due to the widening of the range of dynamically active structures, generating a more pronounced decoupling between modes at different spatio-temporal scales.

A second important aspect is how the sparsification affects the temporal stability of the models. Interestingly, sparse models generally outperform the dense ones when it comes to the prediction of the average turbulent kinetic energy and its power spectrum. This improvement is due to the training phase of the sparsification algorithm, where the model is trained with respect to the DNS simulations. This allows the coefficients of the system to be calibrated to better describe the flow dynamics. This correction is more beneficial for models resolving a small amount of kinetic energy. This result was expected since for small ROMs energy dissipating scales are not resolved, generating an over-prediction of turbulent kinetic energy.

Lastly, the reconstructed flow field for the different Reynolds numbers was analysed. As expected, the over-prediction of turbulent kinetic energy is reflected on the flow field as non-physical oscillations associated with the largest and low index modal structures. The sparsification is beneficial and allows reconstructing the flow structures observed in the DNS simulation. This is also confirmed by the distribution of Reynolds stresses in the shear layer; for the lowest Reynolds numbers these are in good agreement with the baseline DNS simulations. Differently, for the case at higher Reynolds number, the sparse model reproduces the magnitude of the Reynolds stresses while the spatial distribution is recovered only qualitatively. A visual examination of the flow fields shows that the key flow features, such as the formation and shedding of corner vortices, are recovered.

Looking further into the future, the proposed methodology serves as a key enabler for autonomous air and sea operations. The physics-informed model may be integrated into a cyber-physical model, for on-board and real-time applications. In this context, the cyber-physical model uses information derived from a network of sensors embedded on the system, at certain spatial locations. The measured information drives the reduced order model which reconstructs the flow state around the system, and feeds the state reconstruction to a mission manager. The mission manager, i.e. the "brain" of the system, commands appropriate actions currently to match desired performance targets in the future using advanced navigation and guidance control laws, i.e. model reference adaptive control. This type of application is feasible in the near future by integrating commercial off-the-shelves hardware components into a flying testbed.

The authors gratefully acknowledge support for this work from the Air Force Office of Scientific Research (Grant No. FA9550-17-1-0324, Program Manager Dr D. Smith).

References

- Gregory, I. M., Leonard, C., and Scotti, S. J., "Self–Aware Vehicles: Mission and Performance Adaptation to System Health," AIAA Paper 2016–3165, 2016.
- [2] Li, D., Zhao, S., Da Ronch, A., Drofelnik, J., Li, Y., Zhang, L., Wu, Y., Kintscher, M., Monner, H. P., Rudenko, A., Guo, S., Yin, W., Kirn, J., Storm, S., De Breuker, R., and Xiang, J., "A Review of Modelling and Analysis of Morphing Wings," *Progress in Aerospace Sciences*, Vol. 100, 2018, pp. 46–62.
- [3] Han, J., Hui, Z., Tian, F., and Chen, G., "Review on bio-inspired flight systems and bionic aerodynamics," *Chinese Journal of Aeronautics*, 2020.
- [4] Xu, D., Hui, Z., Liu, Y., and Chen, G., "Morphing control of a new bionic morphing UAV with deep reinforcement learning," *Aerospace Science and Technology*, Vol. 92, 2019, pp. 232–243.
- [5] Kan, Z., Li, D., Shen, T., Xiang, J., and Zhang, L., "Review on bio-inspired flight systems and bionic aerodynamics," *Chinese Journal of Aeronautics*, 2020.
- [6] Da Ronch, A., Ventura, A., Righi, M., Franciolini, M., Berci, M., and Kharlamov, D., "Extension of analytical indicial aerodynamics to generic trapezoidal wings in subsonic flow," *Chinese Journal of Aeronautics*, Vol. 31, No. 4, 2018, pp. 617–631.
- [7] Chen, X., Li, C., Gong, C., Gu, L., and Da Ronch, A., "A study of morphing aircraft on morphing rules," *Chinese Journal of Aeronautics*, 2020.
- [8] Noack, B. R., Papas, P., and Monkewitz, P. A., "The need for a pressure-term representation in empirical Galerkin models of incompressible shear flows," *Journal of Fluid Mechanics*, Vol. 523, 2005, pp. 339–365.
- [9] Östh, J., Noack, B. R., Krajnović, S., Barros, D., and Borée, J., "On the need for a nonlinear subscale turbulence term in POD models as exemplified for a high-Reynolds-number flow over an Ahmed body," *Journal of Fluid Mechanics*, Vol. 747, 2014, pp. 518–544.
- [10] Rowley, C. W., and Dawson, S. T. M., "Model reduction for flow analysis and control," *Annual Review of Fluid Mechanics*, Vol. 49, 2017, pp. 387–417.
- [11] Da Ronch, A., Badcock, K. J., Wang, Y., Wynn, A., and Palacios, R., "Nonlinear model reduction for flexible aircraft control design," AIAA Paper 2012–4404, 2012.
- [12] Cordier, L., El Majd, B. A., and Favier, J., "Calibration of POD reduced-order models using Tikhonov regularization," *International Journal for Numerical Methods in Fluids*, Vol. 63, No. 2, 2010, pp. 269–296.
- [13] Brunton, S. L., Noack, B. R., and Koumoutsakos, P., "Machine learning for fluid mechanics," *Annual Review of Fluid Mechanics*, Vol. 52, 2019.
- [14] Noack, B., Morzynski, M., and Tadmor, G., *Reduced-order modelling for flow control*, Vol. 528, Springer Science & Business Media, 2011.
- [15] Taira, K., Brunton, S. L., Dawson, S. T. M., Rowley, C. W., Colonius, T., McKeon, B. J., Schmidt, O. T., Gordeyev, S., Theofilis, V., and Ukeiley, L. S., "Modal analysis of fluid flows: An overview," *AIAA Journal*, 2017, pp. 4013–4041.
- [16] Mendez, M. A., Balabane, M., and Buchlin, J.-M., "Multi-scale proper orthogonal decomposition of complex fluid flows," *Journal of Fluid Mechanics*, Vol. 870, 2019, p. 988–1036. https://doi.org/10.1017/jfm.2019.212.
- [17] Lumley, J. L., "Computational modeling of turbulent flows," Advances in applied mechanics, Vol. 18, Elsevier, 1979, pp. 123–176.
- [18] Lumley, J. L., "Stochastic Tools in Turbulence," Tech. rep., 1970.
- [19] Pope, S. B., "Turbulent flows,", 2001.
- [20] Balajewicz, M. J., Dowell, E. H., and Noack, B. R., "Low-dimensional modelling of high-Reynolds-number shear flows incorporating constraints from the Navier–Stokes equation," *Journal of Fluid Mechanics*, Vol. 729, 2013, pp. 285–308.
- [21] Stankiewicz, W., Roszak, R., and Morzyński, M., "Genetic Algorithm-based Calibration of Reduced Order Galerkin Models," *Mathematical Modelling and Analysis*, Vol. 16, No. 2, 2011, pp. 233–247. https://doi.org/10.3846/13926292.2011.579187, URL https://doi.org/10.3846/13926292.2011.579187.

- [22] Amsallem, D., and Farhat, C., "Interpolation method for adapting reduced-order models and application to aeroelasticity," AIAA journal, Vol. 46, No. 7, 2008, pp. 1803–1813.
- [23] Perez, R., Bartram, G., Beberniss, T., Wiebe, R., and Spottswood, S., "Calibration of aero-structural reduced order models using full-field experimental measurements," *Mechanical Systems and Signal Processing*, Vol. 86, 2017, pp. 49–65.
- [24] Friedman, J., Hastie, T., and Tibshirani, R. J., The Elements of Statistical Learning, Springer, 2008.
- [25] Galletti, B., Bruneau, C., Zannetti, L., and Iollo, A., "Low-order modelling of laminar flow regimes past a confined square cylinder," *Journal of Fluid Mechanics*, Vol. 503, 2004, pp. 161–170.
- [26] Hansen, P. C., "The discrete Picard condition for discrete ill-posed problems," *BIT Numerical Mathematics*, Vol. 30, No. 4, 1990, pp. 658–672.
- [27] Rempfer, D., and Fasel, H. F., "Evolution of three-dimensional coherent structures in a flat-plate boundary layer," *Journal of Fluid Mechanics*, Vol. 260, 1994, pp. 351–375.
- [28] Rempfer, D., and Fasel, H. F., "Dynamics of three-dimensional coherent structures in a flat-plate boundary layer," *Journal of Fluid Mechanics*, Vol. 275, 1994, pp. 257–283.
- [29] Couplet, M., Sagaut, P., and Basdevant, C., "Intermodal energy transfers in a proper orthogonal decomposition–Galerkin representation of a turbulent separated flow," *Journal of Fluid Mechanics*, Vol. 491, 2003.
- [30] Jiménez, J., "Coherent structures in wall-bounded turbulence," Journal of Fluid Mechanics, Vol. 842, 2018.
- [31] Brunton, S. L., Proctor, J. L., and Kutz, J. N., "Discovering governing equations from data by sparse identification of nonlinear dynamical systems," *Proceedings of the National Academy of Sciences*, Vol. 113, No. 15, 2016.
- [32] Loiseau, J. C., and Brunton, S. L., "Constrained sparse Galerkin regression," *Journal of Fluid Mechanics*, Vol. 838, 2018, pp. 42–67.
- [33] Perret, L., Collin, E., and Delville, J., "Polynomial identification of POD based low-order dynamical system," *Journal of Turbulence*, , No. 7, 2006, p. N17.
- [34] Tibshirani, R., "Regression shrinkage and selection via the LASSO," *Journal of the Royal Statistical Society, B*, Vol. 58, No. 1, 1996.
- [35] Ramirez, C., Kreinovich, V., and Argaez, M., "Why *l*₁ is a good approximation to *l*₀: A geometric explanation," *Journal of Uncertain Systems*, Vol. 7, No. 3, 2013, pp. 203–207.
- [36] Auteri, F., Parolini, N., and Quartapelle, L., "Numerical investigation on the stability of singular driven cavity flow," *Journal of Computational Physics*, Vol. 183, No. 1, 2002, pp. 1–25.
- [37] Arbabi, H., and Mezić, I., "Study of dynamics in post-transient flows using Koopman mode decomposition," *Physical Review Fluids*, Vol. 2, No. 12, 2017, p. 124402.
- [38] Pedregosa, F., and Varoquaux, G. e. a., "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, Vol. 12, 2011.