A stabilized POD model for turbulent flows over a range of Reynolds numbers: optimal parameter sampling and constrained projection

Lambert Fick[a], Yvon Maday[b,c], Anthony T Patera[d], Tommaso Taddei[b,*]

[a] *Texas A&M University, Department of Nuclear Engineering, College Station, TX 77843, United States*
[b] *Sorbonne Universités, UPMC Univ Paris 06, Laboratoire Jacques-Louis Lions, Paris, F 75005, France*
[c] *Brown University, Division of Applied Mathematics, Providence RI, 02906, United States*
[d] *MIT, Department of Mechanical Engineering, Cambridge MA, 02139, United States*

Abstract

We present a reduced basis technique for long-time integration of parametrized incompressible turbulent flows. The new contributions are threefold. First, we propose a constrained Galerkin formulation that corrects the standard Galerkin statement by incorporating prior information about the long-time attractor. For explicit and semi-implicit time discretizations, our statement reads as a constrained quadratic programming problem where the objective function is the Euclidean norm of the error in the reduced Galerkin (algebraic) formulation, while the constraints correspond to bounds for the maximum and minimum value of the coefficients of the N-term expansion. Second, we propose an *a posteriori* error indicator, which corresponds to the dual norm of the residual associated with the time-averaged momentum equation. We demonstrate that the error indicator is highly-correlated with the error in mean flow prediction, and can be efficiently computed through an offline/online strategy. Third, we propose a Greedy algorithm for the construction of an approximation space/procedure valid over a range of parameters; the Greedy is informed by the *a posteriori* error indicator developed in this paper. We illustrate our approach and we demonstrate its effectiveness by studying the dependence of a two-dimensional turbulent lid-driven cavity flow on the Reynolds number.

Keywords: model order reduction, reduced basis method, CFD, proper orthogonal decomposition, *a posteriori* error estimation

*Corresponding author
Email addresses: lambert.fick@tamu.edu (Lambert Fick), maday@ann.jussieu.fr (Yvon Maday), patera@mit.edu (Anthony T Patera), taddei@ljll.math.upmc.fr (Tommaso Taddei)

1. Introduction

For turbulent flows, estimation of the entire solution trajectory through a low-dimensional Reduced Order Model (ROM) is infeasible due to the slow decay of the Kolmogorov N-width, and due to the sensitivity of the dynamical system to perturbations. Nevertheless, it might still be possible to estimate various moments of the solution associated to a Direct Numerical Simulation (DNS). The goal of this work is to develop a Reduced Basis (RB) technique for long-time integration of turbulent flows. Our equations of interest are the unsteady incompressible Navier-Stokes equations for high-Reynolds number flows with no-slip boundary conditions:

$$\begin{cases} \partial_t u + (u \cdot \nabla)u - \frac{1}{\mathrm{Re}}\Delta u + \nabla p = f & \text{in } \Omega \times \mathbb{R}_+, \\ \nabla \cdot u = 0 & \text{in } \Omega \times \mathbb{R}_+, \\ u = g & \text{on } \partial\Omega \times \mathbb{R}_+, \\ u = u_0 & \text{on } \Omega \times \{0\}, \end{cases} \tag{1}$$

where $\Omega \subset \mathbb{R}^d$, and f, g, u_0 are suitable fields. We denote by $\mu \in \mathcal{P} \subset \mathbb{R}^P$ the set of parameters associated with the equations.

We consider two separate problems: the solution reproduction problem, and the parametric problem. In the *solution reproduction problem*, given the velocity DNS data $\{u(t_s^k;\mu)\}_{k=1}^K$, and possibly the pressure DNS data $\{p(t_s^k;\mu)\}_{k=1}^K$, at the sampling times $\{t_s^k\}_{k=1}^K \subset \mathbb{R}_+$, we wish to construct a ROM that approximates — in a sense that will be defined soon — the original DNS data for the same value of the parameter μ. In the *parametric problem*, we wish to construct a ROM that approximates the DNS data for all values of μ in a prescribed parameter range $\mathcal{P} \subset \mathbb{R}^P$. For the parametric problem, we wish to control the offline costs associated with the construction of the reduced space: this implicates a Greedy (rather than POD) strategy in parameter. Although the solution reproduction problem might be of limited interest in practice, it represents the first step towards the development of a ROM for the parametric problem.

Following [1], we quantify the accuracy of the ROM by computing the error in the long-time average $\langle u \rangle(x) := \lim_{T\to\infty} \frac{1}{T}\int_0^T u(x,t)\,dt$, and the error in the turbulent kinetic energy $\mathrm{TKE}(t) = \frac{1}{2}\int_\Omega \|u(x,t) - \langle u \rangle(x)\|_2^2\,dx$ where $\|\cdot\|_2$ is the Euclidean norm. We remark that, at present, there is no universally-accepted notion of ROM accuracy for turbulent flows. In [2], the authors evaluate different ROMs based on five different criteria: the kinetic energy spectrum, the mean velocity components, the Reynolds stresses, the root mean square values of the velocity fluctuations, and the time evaluations of the POD coefficients. In [3], the authors consider just the time evaluations and the power spectra of selected POD coefficients. From an engineering perspective, the definition of accuracy is entirely determined by the particular quantity of interest we wish to predict: for

this reason, we envision that for several applications accurate estimates of the long-time averages and possibly of the turbulent kinetic energy might suffice.

The most popular approach for the solution reproduction problem is the so-called POD-Galerkin method [4, 5, 6, 7]: first, we generate a reduced space $\mathcal{Z}^{\mathrm{u}} = \{\zeta_n\}_{n=1}^N$ for the velocity field by applying the Proper Orthogonal Decomposition (POD, [8, 9, 10]) in the L^2 inner product; then, we estimate the velocity field for each time-step $t_{\mathrm{g}}^1, \ldots, t_{\mathrm{g}}^J$ as $\hat{u}^j(\cdot) = \sum_{n=1}^N a_n^j \, \zeta_n(\cdot)$ where the coefficients $\mathbf{a}^j = [a_1^j, \ldots, a_N^j]$ are computed by projecting the momentum equation onto the space $\mathcal{Z}^{\mathrm{u}}$. Since all DNS data for the velocity are divergence-free, it is straighforward to verify that POD modes $\zeta_1, \ldots, \zeta_N$ are divergence-free, and so is the ROM solution.

As observed by several authors, ROMs based on L^2 POD-Galerkin are prone to instabilities [11, 12, 13]. This can be explained through a physical argument. In the limit of high-Reynolds numbers, large-scale flow features are broken down into smaller and smaller scales until the scales are fine enough that viscous forces can dissipate their energy ([14]). This implies that small-scale modes have significant influence on the dynamics. POD modes based on the L^2 inner product are biased toward large, high-energy scales: since large scales are not endowed with the natural energy dissipation tendency of the smaller lower-energy viscous scales, this leads to instabilities and/or large errors in the estimate of the turbulent kinetic energy.

To address the issue of stability, several strategies have been proposed: (i) including dissipation via a closure model, (ii) modifying the POD basis by including functions that resolve a range of scales, (iii) employing a minimum residual formulation, (iv) employing stabilizing inner products, (v) calibration methods, and (vi) generating the reduced space through Dynamic Mode Decomposition. Below, we briefly describe each strategy, and we provide some references; we remark that most of the works presented below are restricted to either laminar flows or short-time integration; therefore, they do not directly address the problem of interest (the long-time integration of fully-turbulent flows). We also recall that other topics are treated in the literature: in particular, Noack et al. [15] proposed to incorporate pressure in the ROM for cases with other than no-slip boundary conditions. For the problem considered in this paper, the discussion of [15] is not relevant, and is here omitted.

(i) Starting with the pioneering work in [16], several authors have proposed to include dissipation through the vehicle of a closure model. A first class of models is designed and motivated by analogy with Large Eddy Simulation (see [17] for an introduction to LES): in this respect, Couplet et al. ([18]) and Noack et al. ([19, section 4.2]) provide theoretical and numerical evidence that the energy transfer among L^2-based POD modes is similar to the energy transfer among Fourier modes, and for this reason LES ideas based on the energy cascade concept might be promising for POD-ROMs. Closure models belonging to this class are inspired by detailed analyses of the physical reasons of the fragility of POD models — mode deformation [13], departures from energy preservation of quadratic terms [3], and need

for nonlinear eddy viscosity [20]. We remark that in [18, 19, 3, 20] the POD space is built for the fluctuating field $u_{\mathrm{f}} = u - \langle u \rangle$ for a fixed value of the parameters (solution reproduction problem) based on the L^2 inner product: for this reason, it appears difficult to rigorously apply these ideas in the parametric setting in which we must combine modes associated with different parameters. Another class of closure models is based on the extension of stabilization techniques originally introduced in the Finite Element or Spectral framework: two notable examples are the Spectral Vanishing Viscosity Model (SVVM, [21], see also [22]) originally presented by Tadmor in [23] for spectral discretization of nonlinear conservation laws for controlling high-wave number oscillations, and the SUPG stabilization discussed in [24]. Östh et al. [20] and Cordier et al. [3] consider a nonlinear extension of the SVVM for a high-Reynolds number incompressible flow over an Ahmed body, and for a compressible 2D isothermal mixing layer, respectively. We refer to [2] for a numerical comparison of four closure models for incompressible Navier-Stokes equations: the mixing-length model, the Smagorinsky model, the variational multiscale model, and the dynamic subgrid-scale model. We further refer to [25] for another POD closure model based on approximate deconvolution, and we refer to [26] for a numerical comparison of several closure models for the Burgers' equation. Finally, we mention the nonlinear Galerkin method proposed by Marion and Temam in [27], and applied to the simulation of turbulent flows in [28]. As the above-mentioned variational multiscale method, this approach corrects the standard Galerkin model by exploiting the separation between large-scale and small-scale modes. The nonlinear Galerkin method has been applied in the model reduction framework in [29].

(ii) Another approach is based on including in the POD basis functions that resolve a range of scales. Bergmann et al. in [30] (see also [31]) proposed to augment the original POD basis with a second POD performed on the residuals of the momentum equation (and of the mass equation in case pressure is modelled by the ROM); on the other hand, Balajewicz and Dowell proposed a Greedy technique to include in the basis random linear combinations of low-energy POD modes associated with the L^2 inner product.

(iii) Minimum residual formulation was first introduced in the reduced basis framework in [32] for linear noncoercive problems, and then extended to fluid problems in [33, 34, 11, 35]. Given the reduced space $\mathcal{Z}^{\mathrm{u}}$ for velocity (and possibly the reduced space $\mathcal{Z}^{\mathrm{p}}$ for pressure), after having discretized the equation in time, the latter approach computes the solution in $\mathcal{Z}^{\mathrm{u}}$ (or $\mathcal{Z}^{\mathrm{u}} \times \mathcal{Z}^{\mathrm{p}}$) that minimizes a suitable dual residual at each time-step. We remark that for problems with quadratic nonlinearities minimum residual ROMs require $\mathcal{O}(N^4)$ storage and the online cost for each time-step is $\mathcal{O}(N^4)$ for semi-implicit/explicit time-discretizations — as opposed to $\mathcal{O}(N^3)$ for standard POD-Galerkin ROMs. For this reason,

hyper-reduction techniques are employed to reduce the online cost and the memory constraints ([33, 34, 11]).

(iv) Iollo et al. in [36] proposed to employ the H^1 inner product rather than the more standard L^2 inner product to generate the POD modes. This choice is motivated by dynamic considerations: since small-scale modes have relatively large H^1 norm compared to their L^2 norm, and recalling that small scales are responsible for energy dissipation, the use of the H^1 inner product leads to a more dissipative reduced order model. We remark that several other authors proposed to not employ the standard L^2 inner product ([37, 38, 39]); however, their choices were not motivated by long-time stability considerations.

(v) If we denote by $\dot{\mathbf{a}} = \mathcal{F}(\mathbf{a})$ the ROM for the coefficients of the POD expansion, calibration techniques aim to identify the model $\mathcal{F}$ based on data. To provide a concrete point of reference, Couplet et al. [40] proposed to calibrate the coefficients of $\mathcal{F}$ based on DNS data, under the assumption that $\mathcal{F}$ is a polynomial of degree 2 in $\mathbf{a}$. In recent years, calibration methods have received significant attention in the Fluid Mechanics literature, particularly for system identification based on experimental data; we refer to [41] for a thorough review. We also refer to [42] for a recent application of related ideas in the MOR context. We observe that the ROM $\mathcal{F}$ depends on the particular POD basis selected; for large values of N (dimension of the POD space), calibration might require a substantial number of DNS snapshots. Furthermore, calibration requires dense (in time) DNS data.

(vi) Dynamic Mode Decomposition (DMD) was first proposed by Schmid in [43]; as shown by Rowley et al. in [44] DMD can be interpreted as an algorithm for finding the Koopman modes associated with the nonlinear discrete dynamical system obtained from the discretization of the Navier-Stokes equations. Although several authors have proven the effectiveness of DMD for the extraction of physically-relevant time scales and their associated spatial structures ([44, 43, 45]), the work by Alla and Kutz [46] represents one of the few examples of application of DMD within the Galerkin framework.

Despite these advances, the solution reproduction problem remains an open issue, particularly for turbulent flows. By performing a detailed analysis of the performance of the POD-Galerkin approach, we empirically demonstrate that in the case of turbulent flows POD-Galerkin ROMs might exhibit other spurious effects such as false stable steady flows. This demonstrates the need for a more fundamental correction to the POD-Galerkin formulation. We remark that a similar issue has been observed in [21] by Sirisup and Karniadakis for long-time integration of a POD-Galerkin ROM for a laminar flow past a cylinder, and — in a different context — by Curry et al. in [47] for highly-truncated spectral approximations to turbulent flows.

To our knowledge, there are very few works that systematically address the parametric problem. Ma and Karniadakis ([6]), Galletti et al. ([7]), and Stabile

et al. ([48]) developed a reduced order model based on POD-Galerkin for the flow past a cylinder for a wide range of Reynolds numbers in the laminar regime. In these papers, the authors use DNS data for pre-selected Reynolds numbers to generate reduced spaces for velocity ([6, 7]), and for velocity and pressure ([48]). The choice of the parameters for which the DNS data are computed is performed *a priori*. Non-adaptive explorations of the parameter space typically require a large number of offline evaluations of the Full Order Model (FOM); for this reason, they might not be practical in our context.

The goal of this work is to develop a Model Order Reduction (MOR) procedure for the parametrized incompressible Navier-Stokes equations. The three key pieces of our MOR technique are (i) a reduced formulation for the computation of the reduced-order solution, (ii) an *a posteriori* indicator for the error in the prediction of the mean flow, and (iii) a H^1-POD-hGreedy strategy for the construction of the reduced space informed by the above-mentioned *a posteriori* indicator.

(i) Our reduced formulation is based on a constrained Galerkin formulation. The approach is designed to correct the standard Galerkin formulation, especially for moderate values of N. For explicit and semi-implicit time discretizations the formulation reads as a quadratic programming problem where the objective function corresponds to the Euclidean norm of the error in the reduced Galerkin (algebraic) formulation, while the constraints correspond to bounds for the maximum and minimum value of the coefficients $\{a_n^j\}_{n=1}^N \subset \mathbb{R}$ of the expansion. We discuss an actionable procedure to estimate the lower and upper bounds associated with each coefficient of the reduced expansion based on DNS data.

(ii) Our error indicator corresponds to the dual norm of the residual associated with the time-averaged momentum equation. Time-averaging is here motivated by the chaotic behavior in time of the velocity field. We verify that the error indicator can be efficiently computed through an offline/online strategy; furthermore, we numerically demonstrate that the indicator is highly-correlated with the error in the mean flow prediction: therefore, it is well-suited to drive the Greedy procedure for the generation of the ROM.

(iii) As in the seminal work by Haasdonk and Ohlberger [49], our POD- hGreedy algorithm combines POD in time with Greedy in parameter. The procedure is a simplified version of the h-type Greedy proposed in [50]. Given $\mu^1 \in \mathcal{P}$, we generate the DNS data for μ^1, we apply POD — based on the H^1 inner product — to generate the reduced space $\mathcal{Z}_1^{\mathrm{u}}$, we build the POD-ROM, and we evaluate the error indicator $\Delta_1^{\mathrm{u}}(\mu)$ for all $\mu \in \mathcal{P}_{\mathrm{train}} \subset \mathcal{P}$. Then, we select μ^2 that maximizes the error estimate Δ_1^{u} over the training set $\mathcal{P}_{\mathrm{train}}$. During the second iteration, we perform the same steps as before for μ^2 (generation of DNS data, POD, construction of the ROM, estimate of the error). Finally, we select μ^3 that maximizes $\Delta_{1,2}^{\mathrm{u}}(\mu) := \min\{\Delta_1^{\mathrm{u}}(\mu), \Delta_2^{\mathrm{u}}(\mu)\}$ over $\mathcal{P}_{\mathrm{train}}$. We then proceed to generate $\mu^4, \ldots, \mu^L$.

At the end of the offline stage, the procedure produces L different ROMs; during the online stage, given a new value of $\mu \in \mathcal{P}$, we first evaluate the ROMs associated with the n_{cand} nearest anchor points, and then we select the ROM that minimizes the error indicator.

We observe that in this work we restrict ourselves to linear approximation spaces $\mathcal{Z}^{\text{u}}$ that do not depend on time: this greatly simplifies the implementation, and reduces the memory constraints for long-time integration. We refer to [51, 52, 53] for MOR strategies based on nonlinear approximation spaces for unsteady problems. On the other hand, we refer to [54, 55] for space-time approximations of linear and nonlinear parabolic problems.

The idea of employing a constrained formulation is new in the MOR framework for turbulent flows. For system identification, Colin et al. [56] proposed a Galerkin model with constraints on the maximum and minimum value attained by the POD coefficients for a tumor growth model; for steady-state data assimilation, Argaud et al. [57] proposed a similar constrained formulation. As in our work, the constraints in [56, 57] provide further information about the solution manifold; however, while in our work the constraints are designed to compensate for the effect of the unmodelled dynamics and ultimately improve long-time stability, in [56, 57], constraints are designed to limit the effect of data scarcity and/or experimental noise. Related constrained formulations with different sets of constraints have also been proposed in [58] for system identification, and in [59] for the reduction of unsteady PDEs.

As for calibration techniques and POD-Galerkin closure models, DNS data are here used twice: first, to build the space $\mathcal{Z}^{\text{u}}$; second, to estimate the hyper-parameters of the ROM. However, while the hyper-parameters of calibrated ROMs and closure models are associated with the dynamic response of the reduced system and admit a dynamical interpretation (e.g., a set of eddy viscosities associated with different modes, [20]), the hyper-parameters of our constrained formulation (the lower and upper bounds for the coefficients of the expansion) are directly associated with the solution manifold. As a result, the latter can be tuned directly through sparse DNS data, for an arbitrary reduced space $\mathcal{Z}^{\text{u}}$, without having to evaluate the ROM for several tentative candidates (solution matching). This feature of the approach greatly simplifies the implementation of the method, and in practice reduces the offline costs. We remark that in [20] the authors propose a method to tune linear and nonlinear eddy viscosity models based on a modal power balance that does not require solution matching; however, the approach is restricted to L^2 POD spaces for the fluctuating field.

Another important feature — which we demonstrate empirically in section 3.2.2 — of our constrained ROM is that the box constraints do not necessarily lead to a reduction of the TKE. This is in contrast with POD closure models, which effectively add dissipation to the reduced system.

The time-averaged error indicator is also new. In [49], the authors employ a residual estimator that measures the error in the entire trajectory: for turbulent flows, this metric is not appropriate due to the chaotic nature of the dynamical

system. This explains the importance of our new error indicator for the problem at hand.

The POD-Greedy algorithm was first proposed in [49], and then analyzed in [60]. The algorithm in [49] combines data from different parameters to generate a single reduced space for the entire parameter space $\mathcal{P}$. On the other hand, in our approach we build a reduced space for each of them. Recalling the definitions of [50], the algorithm of [49] corresponds to a POD-pGreedy, while our approach corresponds to a POD-hGreedy. For turbulent flows, we empirically show in AppendixF that combining modes associated with different values of the parameters might lead to poor performance. On the other hand, h-refinement leads to more accurate and stable ROMs.

We finally remark that the objective of this paper is to develop a projection-based intrusive ROM. Several authors have observed that the use of intrusive ROMs might be cumbersome, particularly for industrial applications. For this reason, non-intrusive approaches based on interpolation or regression (see, e.g., [61, 62, 63]) have been proposed for the estimate of (time-averaged) quantities of interest. Although non-intrusive approaches are significantly less expensive to evaluate and easier to integrate with existing DNS codes, they typically require a larger amount of offline data compared to projection-based ROMs, particularly for high-dimensional parameterizations and/or when parameter changes induce a bifurcation or a substantial change in the form of the attractor. Furthermore, non-intrusive approaches are typically restricted to the estimate of steady quantities of interest.

The paper is organized as follows. In section 2, we introduce the model problem considered in this work. In section 3, we consider the solution reproduction problem. First, we consider the POD-Galerkin approach: we introduce the formulation, and we assess the numerical performance. Then, we present our constrained POD-Galerkin approach: as in the previous case, we discuss the formulation, and then we numerically assess the performance. In section 4, we consider the parametric problem: first, we present the POD-hGreedy approach; second, we discuss how to adapt the constrained Galerkin formulation to the parametric setting; third, we propose the time-averaged error indicator; and fourth, we present the numerical assessment. In section 5, we offer some concluding remarks, and we discuss potential extensions of the current approach. A number of appendices provide further analysis and numerical investigations: in AppendixA we provide an analysis of the model problem considered; in AppendixB we discuss the selection of the sampling times $\{t_s^k\}_{k=1}^K$; in AppendixC we propose a suitable definition of stability for ROMs; in AppendixD, we derive a bound for the maximum norm of the increment at each time step for the solution to the constrained Galerkin ROM; in AppendixE we investigate the robustness of the constrained formulation proposed in this paper; in AppendixF we illustrate the problem of p-refinement for the parametric case; in AppendixG we describe the offline/online strategy employed to compute the error indicator; and in AppendixH we provide computational timings associated with the overall procedure. Figure 1 summarizes the key elements of the present work.

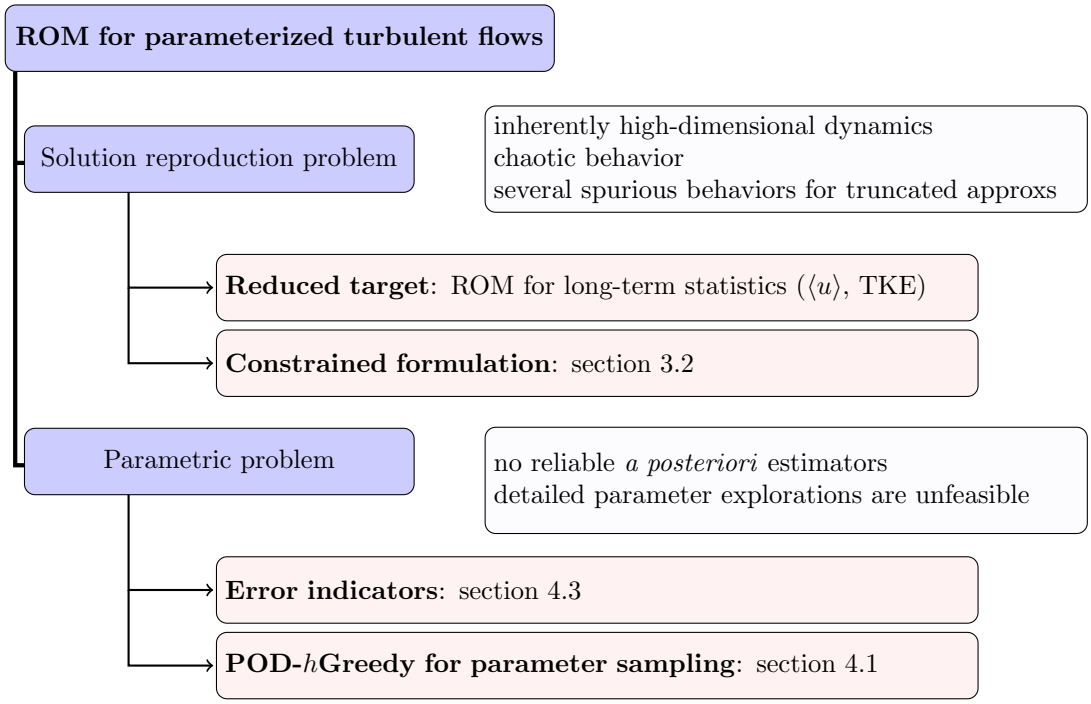


Figure 1: Conceptual flow chart of the paper. Challenges for the solution reproduction problem are a subset of the challenges for the parametric problem.

2. A lid-driven cavity problem

We consider the following unsteady lid-driven cavity problem:

$$\begin{cases} \partial_t u + (u \cdot \nabla)u - \nu(\mathrm{Re})\Delta u + \nabla p = 0 & \text{in } \Omega \times \mathbb{R}_+, \\ \nabla \cdot u = 0 & \text{in } \Omega \times \mathbb{R}_+, \\ u = g(x) & \text{on } \Gamma_{\text{top}} \times \mathbb{R}_+, \\ u = 0 & \text{on } \partial\Omega \setminus \Gamma_{\text{top}} \times \mathbb{R}_+, \\ u = 0 & \text{on } \Omega \times \{0\}, \end{cases} \tag{2a}$$

where the velocity $u : \Omega \times \mathbb{R}_+ \to \mathbb{R}^2$ is a two-dimensional vector field, the pressure $p : \Omega \times \mathbb{R}_+ \to \mathbb{R}$ is a scalar field, $\nu(\mathrm{Re}) = \frac{1}{\mathrm{Re}}$, $\Omega = (-1, 1)^2$, $\Gamma_{\text{top}} = \{x \in \bar{\Omega} : x_2 = 1\}$, the Dirichlet datum is given by

$$g(x) = \begin{bmatrix} (1 - x_1^2)^2 \\ 0 \end{bmatrix}, \tag{2b}$$

and the Laplacian Δ should be interpreted as component-wise. We remark that in (2) time is non-dimensionalized by the convective scaling (i.e., dimensional boxside half-length divided by dimensional maximum lid velocity). The problem corresponds to a isothermal, incompressible, two-dimensional flow inside a square cavity driven by a prescribed lid velocity. The problem is a well-known prototypical example used to validate numerical schemes and reduced order models ([1, 64, 65, 66, 67, 68]); unlike in the more standard lid-driven cavity problem with $g(x) = [1, 0]$, here we regularize the singularity near the upper corners of the cavity.

In this paper, we study the dependence of the flow on the Reynolds number, that is $\mu = \mathrm{Re}$. It is well-known ([66]) that the flow exhibits a long-time unsteady but stationary solution for $\mathrm{Re} > \mathrm{Re}_c$ ([66]); here stationarity implies that all statistics are invariant under a shift in time ([69]). Since we are interested in long-time unsteady flows, we here consider $\mathrm{Re} \in \mathcal{P} = [\mathrm{Re}_{\mathrm{LB}}, \mathrm{Re}_{\mathrm{UB}}] = [15000, 25000]$: for all values of Re in $\mathcal{P}$ the flow is asymptotically statistically stationary. Balajewicz and Dowell considered the same problem — for a single value of Re — in [1]; we remark that they define the viscosity as $\nu(\mathrm{Re}) = \frac{2}{\mathrm{Re}}$, and they consider the case $\mathrm{Re} = 30000$.

In view of the development of the ROM for (2) it is convenient to consider the lifted equations. If we denote by R_g the two-dimensional vector field defined as the solution to the following Stokes problem:

$$\begin{cases} -\Delta R_g + \nabla \lambda = 0 & \text{in } \Omega, \\ \nabla \cdot R_g = 0 & \text{in } \Omega, \\ R_g = g & \text{on } \Gamma_{\text{top}}, \\ R_g = 0 & \text{on } \partial\Omega \setminus \Gamma_{\text{top}}, \end{cases} \tag{3a}$$

we can define the lifted velocity solution $\mathring{u} = u - R_g$ as the solution to:

$$\begin{cases} \partial_t \mathring{u} + ((\mathring{u} + R_g) \cdot \nabla)(\mathring{u} + R_g) - \frac{1}{\mathrm{Re}} \Delta (\mathring{u} + R_g) + \nabla p = 0 & \text{in } \Omega \times \mathbb{R}_+, \\ \nabla \cdot \mathring{u} = 0 & \text{in } \Omega \times \mathbb{R}_+, \\ \mathring{u} = 0 & \text{on } \partial\Omega \times \mathbb{R}_+, \\ \mathring{u}(t=0) = -R_g & \text{on } \Omega \times \{0\}. \end{cases} \tag{3b}$$

Then, if we introduce the spaces $V := [H_0^1(\Omega)]^2$, and $Q = \{q \in L^2(\Omega) : \int_\Omega q \, dx = 0\}$, we can define the weak form associated with (3): find $(\mathring{u}, p) \in \mathcal{V} \times \mathcal{Q}$ such that for a.e. $t > 0$

$$\begin{cases} \langle \partial_t \mathring{u}(t), v \rangle_\star + \frac{1}{\mathrm{Re}} (\mathring{u}(t) + R_g, v)_V + c(\mathring{u}(t) + R_g, \mathring{u}(t) + R_g, v) + b(v, p(t)) = 0 \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \forall\, v \in V, \\ b(\mathring{u}(t), q) = 0 \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \forall\, q \in Q, \end{cases} \tag{4a}$$

where $\mathcal{V} = \{v \in L^2_{\mathrm{loc}}(\mathbb{R}_+; V) : \partial_t v \in L^2_{\mathrm{loc}}(\mathbb{R}_+; V')\}$, $\mathcal{Q} = L^2_{\mathrm{loc}}(\mathbb{R}_+; Q)$, $\langle \cdot, \cdot \rangle_\star$ denotes the pairing between V' and V which (for our smoothness assumptions and numerical approximations) can be evaluated in terms of the pivot space L^2, $(w, v)_V = \int_\Omega \nabla w : \nabla v \, dx$ is the inner product associated with V, and

$$c(w, u, v) = \int_\Omega (w \cdot \nabla) u \cdot v \, dx, \quad b(v, q) = -\int_\Omega (\nabla \cdot v) q \, dx. \tag{4b}$$

We resort to a $\mathbb{Q}_M - \mathbb{Q}_{M-2}$ spectral element ([70]) discretization in space, and to an explicit three-step Adams-Bashforth (AB3)/ implicit two-step Adams-Moulton (AM2) discretization in time. DNS simulations are performed using the open-source software `nek5000` ([71]). We refer to the spectral element literature (see, e.g., [72, 73, 74, 75]) for further details about the spectral element method and its implementation for fluid dynamics problems. More in detail, we consider a 16 by 16 structured quadrilateral mesh, we consider $M = 8$, and we resort to an equispaced time grid $\{t_{\mathrm{g}}^j = j\Delta t\}_{j=0}^J$, with[1] $\Delta t = 2 \cdot 10^{-3}$. The number of spatial degrees of freedom for velocity is equal to $\mathcal{N} = 25538$. We estimate the long-time averaged velocity field as[2]:

$$\langle u \rangle_{\mathrm{g}} = \frac{\Delta t}{T - T_0} \sum_{j=J_0+1}^{J} u^j, \tag{5}$$

where $T_0 = 500$, $T = t_{\mathrm{g}}^J$, and J_0 is such that $t_{\mathrm{g}}^{J_0} = T_0 = 500$. Consequently, we

[1] The time step is required to ensure stability of the run and maintain CFL $\sim 0.5 - 0.6$.

[2] In the current implementation, $\langle u \rangle_{\mathrm{g}}$ is computed inside the time integration loop of the Full Order Model.

estimate the instantaneous turbulent kinetic energy as

$$\mathrm{TKE}^j := \frac{1}{2}\int_\Omega \|u^j - \langle u\rangle_{\mathrm{g}}\|_2^2\, dx. \tag{6}$$

We refer to (5) and (6) as *truth* approximations of mean flow and instantaneous TKE. In AppendixA, we provide a detailed analysis of the solution to the lid-driven cavity problem (2).

In order to generate (and, later, assess) the ROM, we collect data at the sampling times $\{t_{\mathrm{s}}^k = T_0 + \Delta t_{\mathrm{s}} k\}_{k=1}^K$ with $\Delta t_{\mathrm{s}} = 1$. We observe that $\{t_{\mathrm{s}}^k\}_{k=1}^K \subset \{t_{\mathrm{g}}^j\}_{j=J_0}^J$, and $K \ll J$: this is dictated by memory constraints. We further observe that we do not collect data in the transient region: this is motivated by the fact that we are here ultimately interested in the long-time dynamics. In the remainder of the paper, we use the subscript "s" to indicate the sampling times, and the subscript "g" to indicate the time discretization. Furthermore, we use the symbol $\langle\cdot\rangle_{\mathrm{s}}$ to indicate time averages performed based on the sampling times, and the symbol $\langle\cdot\rangle_{\mathrm{g}}$ to indicate time averages performed based on the time grid $\{t_{\mathrm{g}}^j\}_{j=J_0}^J$. In particular, we define the truth estimate of the mean TKE as

$$\langle \mathrm{TKE}\rangle_{\mathrm{s}} = \frac{1}{2K}\sum_{k=1}^K \|u^k - \langle u\rangle_{\mathrm{g}}\|_{L^2(\Omega)}^2. \tag{7}$$

In AppendixB, we comment on the choice of Δt_{s} and K.

3. The solution reproduction problem

In this section, we propose a MOR procedure for the solution reproduction problem. As explained in the introduction, the solution reproduction problem is of limited practical interest; however, it represents a key intermediate step towards the development of a MOR procedure for the parametric problem. Algorithm 1 outlines the general offline/online paradigm for the solution reproduction problem. We recall that the offline stage is expected to be expensive and is performed once, while the online stage should be inexpensive and is performed many times — this distinction is of little relevance here, but will be crucial in section 4 for the parametric problem.

As anticipated in section 2, we here generate a ROM for the lifted velocity field $\mathring{u} = u - R_g$, where R_g is the solution to the Stokes problem (3a). Reduction of the lifted equations is preferable from the MOR perspective since it greatly simplifies the imposition of essential (Dirichlet) inhomogenous boundary conditions. We observe that in the Fluid Mechanics literature many authors consider $R_g = \langle u\rangle_{\mathrm{g}}$; however, the latter choice of the lift cannot be extended to the parametric case.

This section is organized as follows. In section 3.1, we present the POD-Galerkin ROM. We first introduce the formulation, we review Proper Orthogonal Decomposition for the generation of the reduced space, and then we present numerical results that highlight the limitations of the approach. In section 3.2,

we present the constrained POD-Galerkin ROM proposed in this paper. As for POD-Galerkin, we first present and motivate the mathematical statement, and then we present a number of numerical results to motivate the approach.

Algorithm 1 Offline/online paradigm for the solution reproduction problem

Task: find an estimate of $\mathring{u} = \mathring{u}(x,t)$ of the form $\hat{u}(x,t) = \sum_{n=1}^{N} a_n(t)\,\zeta_n(x)$.

Offline stage

1: Generate the DNS data $\{\mathring{u}^k := \mathring{u}(t_{\mathrm{s}}^k)\}_{k=1}^{K} \subset V$.
2: Generate the reduced space $\mathcal{Z}^{\mathrm{u}} = \mathrm{span}\{\zeta_n\}_{n=1}^{N}$.
3: Formulate the Reduced Order Model.

Online stage

1: Estimate the coefficients $\{a_n^j = a_n(t_{\mathrm{g}}^j)\}_{n=1}^{N}$ for $j = 0, 1, \ldots, J$.
2: Compute the QOIs (e.g., mean flow, TKE,...)

3.1. The POD-Galerkin ROM

3.1.1. The Galerkin formulation

Given the reduced space $\mathcal{Z}^{\mathrm{u}} = \mathrm{span}\{\zeta_n\}_{n=1}^{N} \subset V_{\mathrm{div}} = \{v \in V : \nabla \cdot v = 0\}$, we seek $\hat{u} \in \mathcal{V}_N := H^1_{\mathrm{loc}}(\mathbb{R}_+; \mathcal{Z}^{\mathrm{u}})$ such that

$$\begin{cases} \dfrac{d}{dt}(\hat{u}(t), v)_{L^2(\Omega)} + \dfrac{1}{\mathrm{Re}}(\hat{u}(t) + R_g, v)_V + c(\hat{u}(t) + R_g, \hat{u}(t) + R_g, v) = 0 \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \forall\, v \in \mathcal{Z}^{\mathrm{u}}, \\ \hat{u}(0) = -\Pi^{L^2}_{\mathcal{Z}^{\mathrm{u}}} R_g, \end{cases} \tag{8}$$

where $\Pi^{L^2}_{\mathcal{Z}^{\mathrm{u}}} : [L^2(\Omega)]^2 \to \mathcal{Z}^{\mathrm{u}}$ is the $L^2(\Omega)$-projection operator on $\mathcal{Z}^{\mathrm{u}}$, and $(\cdot,\cdot)_{L^2(\Omega)}$ is the $L^2(\Omega)$ inner product. If we employ a semi-implicit time discretization, we obtain:

$$\left(\frac{\hat{u}^{j+1} - \hat{u}^j}{\Delta t}, v\right)_{L^2(\Omega)} + \frac{1}{\mathrm{Re}}(\hat{u}^{j+1} + R_g, v)_V + c(\hat{u}^j + R_g, \hat{u}^{j+1} + R_g, v) = 0 \qquad \forall\, v \in \mathcal{Z}^{\mathrm{u}},\ \ j = 0, 1, \ldots, \tag{9}$$

where $\Delta t = t_{\mathrm{g}}^{j+1} - t_{\mathrm{g}}^j$. We remark that the time scheme is not the same used by `nek5000`. The Galerkin formulation (9) leads to the following algebraic system for the coefficients $\{\mathbf{a}^j\}_{j=0}^{J}$ of the N-term expansion:

$$\mathbb{A}(\mathbf{a}^j; \mathrm{Re})\, \mathbf{a}^{j+1} = \mathbf{F}(\mathbf{a}^j; \mathrm{Re}), \qquad j = 0, 1, \ldots, \tag{10a}$$

where $\mathbb{A}(\mathbf{a}^j;\mathrm{Re}) := \mathbb{A}_1 + \frac{1}{\mathrm{Re}}\mathbb{A}_2 + \mathbb{C}(\mathbf{a}^j)$, $\mathbf{F}(\mathbf{a}^j;\mathrm{Re}) := \mathbb{E}\mathbf{a}^j - \frac{1}{\mathrm{Re}}\mathbf{G}$, with

$$\begin{aligned} &(\mathbb{A}_1)_{m,n} = \frac{1}{\Delta t}(\zeta_n,\zeta_m)_{L^2(\Omega)} + c(R_g,\zeta_n,\zeta_m), \quad (\mathbb{A}_2)_{m,n} = (\zeta_n,\zeta_m)_V, \\ &(\mathbb{C}(\mathbf{w}))_{m,n} = \textstyle\sum_{i=1}^N w_i\, c(\zeta_i,\zeta_n,\zeta_m), \end{aligned} \tag{10b}$$

and

$$\mathbf{G}_m = (R_g,\zeta_m)_V, \quad (\mathbb{E})_{m,n} = \frac{1}{\Delta t}(\zeta_n,\zeta_m)_{L^2(\Omega)} - c(\zeta_n,R_g,\zeta_m). \tag{10c}$$

We observe that the Galerkin model for the velocity field does not contain the pressure field. This follows from (i) the fact that the ROM is derived from the weak form of the equations, (ii) the particular boundary conditions prescribed, and (iii) the absence of parameters in the form $b(\cdot,\cdot)$. We have indeed that for certain choices of the boundary conditions the ROM should be obtained from the strong form of the Navier-Stokes equations: in this respect, we recall that in [15] a Galerkin ROM is derived from the strong form for a laminar flow problem with convective boundary condition ([76]) at the outflow. In the parametric case it is possible to derive a ROM that does not contain the pressure field if the form $b(\cdot,\cdot)$ in (4) is parameter-independent; otherwise, it is not possible in general to generate a space $\mathcal{Z}^{\mathrm{u}}$ such that $b(z,\cdot) \equiv 0$ for all $z \in \mathcal{Z}^{\mathrm{u}}$ and for all values of the parameters. Since the bilinear form $b(\cdot,\cdot)$ in (3) does not depend on the Reynolds number, we will be able in section 4 to generate a ROM for the velocity only. We remark that the case of parametrized b form corresponds to the case of geometric parametrizations, which is of particular interest for applications. A potential strategy to handle this issue is to resort to the Piola's transform (see [77]). We refer to a future work for a detailed discussion of this case. We also refer to the Reduced Basis literature ([78, 79, 80, 81]) for a thorough discussion about fluid problems in parametrized domains for low-to-moderate Reynolds number flows.

The algebraic formulation (10) is the starting point for the development of the offline/online decomposition. The matrices $\mathbb{A}_1$, $\mathbb{A}_2$, $\mathbb{E}$, the third-order tensor $\mathbb{C}$ and the vector $\mathbf{G}$ are pre-computed during the offline stage. Therefore, during the online stage, the method only requires $\mathcal{O}(N^3)$ storage, and the online cost is $\mathcal{O}(N^3 J)$. Provided that N is much smaller than the spatial mesh-size $\mathcal{N}$, the Galerkin ROM is significantly less expensive and less memory-demanding than the FOM. Other choices of the time discretization lead to similar reduced systems that allow the same offline/online decomposition. The flow chart in Figure 2 summarizes the offline development of the POD-Galerkin ROM.

3.1.2. Construction of the reduced space: Proper Orthogonal Decomposition

We employ Proper Orthogonal Decomposition (POD, [8, 9, 82]) to generate the reduced space $\mathcal{Z}^{\mathrm{u}}$. Below we briefly review the numerical strategy — known as *method of snapshots* ([82]) — employed for the computation of the POD modes. We refer to [83] for a review of the theoretical results concerning the optimality properties of POD.

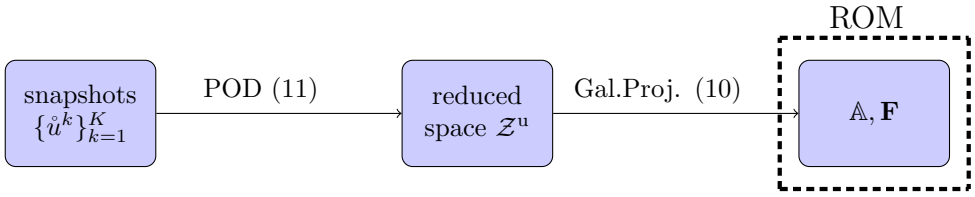


Figure 2: Flow chart of the POD-Galerkin process

Given the snapshot set $\{\mathring{u}^k\}_{k=1}^K$, we assemble the Gramian $\mathbb{U} \in \mathbb{R}^{K,K}$ $\mathbb{U}_{k,k'} = (\mathring{u}^k, \mathring{u}^{k'})_\star$ where $(\cdot,\cdot)_\star$ is a suitable inner product that will be introduced soon; then, we compute the first N eigenmodes of the symmetric matrix $\mathbb{U}$:

$$\mathbb{U}\boldsymbol{\zeta}_n = \lambda_n \boldsymbol{\zeta}_n, \qquad \lambda_1 \geq \ldots \geq \lambda_K \geq 0; \tag{11a}$$

finally, we define the POD modes as

$$\zeta_n := \sum_{k=1}^{K} (\boldsymbol{\zeta}_n)_k \; \mathring{u}^k, \qquad n = 1, \ldots, N. \tag{11b}$$

It is easy to show that $\zeta_1, \ldots, \zeta_N$ can be chosen to be orthogonal in the $(\cdot,\cdot)_\star$-inner product; for stability reasons, we also orthonormalize the POD modes so that $(\zeta_n, \zeta_{n'})_\star = \delta_{n,n'}$, $n, n' = 1, \ldots, N$. In AppendixB, we discuss the choice of the sampling times $\{t_{\mathrm{s}}^k\}_k$, and we propose a numerical technique to assess the accuracy of the POD space for the full trajectory.

In this work, we employ the $H_0^1(\Omega)$ inner product:

$$(w, v)_\star = (w, v)_V = \int_\Omega \nabla w : \nabla v \, dx. \tag{12}$$

As explained in the introduction, this choice is motivated by dynamic considerations. Since small-scale modes have relatively large H^1 norm compared to their L^2 norm, and recalling that small scales are responsible for energy dissipation, the use of the H_0^1 inner product leads to a more dissipative reduced order model ([36]).

3.1.3. Performance of the POD-Galerkin ROM

We assess the numerical performance of the POD-Galerkin ROM presented in this section. We here consider the lid-driven cavity problem (2) for Re = 15000. We consider the time grid[3] $\{t_{\mathrm{g}}^j = \Delta t j\}_{j=0}^J$ with $\Delta t = 5 \cdot 10^{-3}$ and $J = 2 \cdot 10^5$ ($T = t_{\mathrm{g}}^J = 10^3$), and we acquire the snapshots $\{\mathring{u}^k = \mathring{u}(t_{\mathrm{s}}^k)\}_{k=1}^K$ where $t_{\mathrm{s}}^k = 500 + k$ and $K = 500$. The truth long-time averaged velocity field $\langle u \rangle_{\mathrm{g}}$ and the

[3] We emphasize that we consider a larger time step compared to the one employed in the DNS solver, $\Delta t > \Delta t^{\mathrm{FOM}} = 2 \cdot 10^{-3}$. In our experience, refinement of the time grid does not substantially influence performance of the ROM.

mean truth TKE are estimated through (5) and (7), respectively. Assembling and integration of the Reduced Order Model are performed in `Matlab` [84].

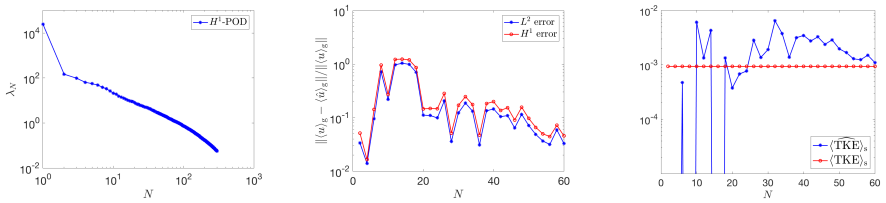
Figure 3(a) shows the behavior of the eigenvalues $\{\lambda_n\}_{n=1}^K$. The first eigenmode is roughly proportional to $\langle u\rangle_{\rm g} - R_g$; provided that the estimate of the coefficients is accurate, it does not contribute to the fluctuating field. Therefore, we can identify the ratio

$$r_N = \frac{\sum_{n=2}^N \lambda_n}{\sum_{n=2}^K \lambda_n}$$

as the portion of H_0^1 energy of the fluctuating field associated with the reduced POD space of dimension N. We find that $r_N = 0.165$ for $N = 2$, $r_N = 0.731$ for $N = 20$, $r_N = 0.797$ for $N = 30$, and $r_N = 0.87$ for $N = 50$. We observe that the decay with N is rather slow; this suggests that accurate estimates of the entire system dynamics are out of reach for fully turbulent flows. Figure 3(b) shows the behavior with N of the relative error in the mean flow prediction:

$$E_N^0 = \frac{\|\langle u - \hat{u}\rangle_{\rm g}\|_{L^2(\Omega)}}{\|\langle u\rangle_{\rm g}\|_{L^2(\Omega)}}, \qquad E_N^1 = \frac{\|\langle u - \hat{u}\rangle_{\rm g}\|_{H_0^1(\Omega)}}{\|\langle u\rangle_{\rm g}\|_{H_0^1(\Omega)}};$$

while Figure 3(c) shows the behavior with N of the mean TKE predicted by the ROM: $\langle\widehat{\rm TKE}\rangle_{\rm s}$. We observe that for small values of N, we predict a false stable steady flow, while for moderate values of N we substantially overestimate the TKE. Finally, for $N \gtrsim 50$ we observe a slow convergence of the Galerkin ROM to the mean values (5) and (7).



(a) (b) (c)

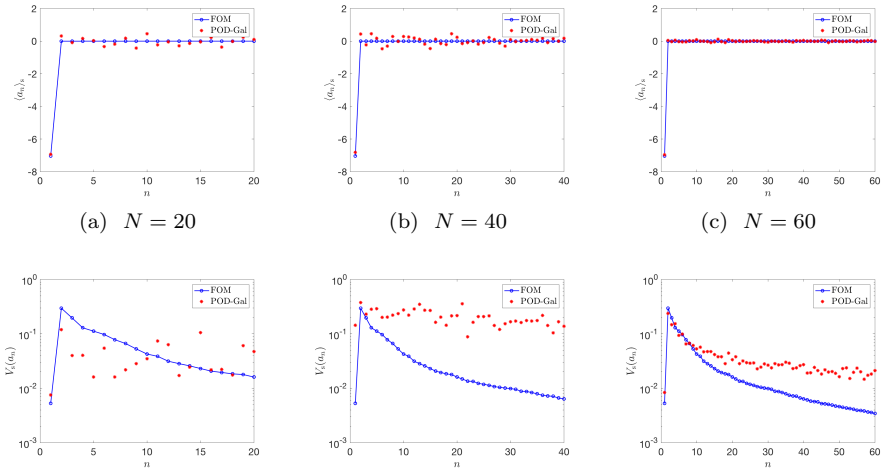
Figure 3: The solution reproduction problem; POD-Galerkin. Figure (a): POD eigenvalues. Figure (b): behavior of the relative L^2 and H^1 errors in mean flow prediction with N. Figure (c): TKE prediction $\langle\widehat{\rm TKE}\rangle_{\rm s}$ vs. truth estimate (7) for multiple values of N. (Re = 15000).

Figure 4 shows the behavior for different values of N of the sample mean and sample variance of the coefficients $\{a_n^j\}_j$:

$$\langle a_n\rangle_{\rm s} = \frac{1}{K}\sum_{k=1}^K a_n(t_{\rm s}^k), \qquad V_{\rm s}(a_n) = \frac{1}{K-1}\sum_{k=1}^K \left(a_n(t_{\rm s}^k) - \langle a_n\rangle_{\rm s}\right)^2,$$

for the FOM and for the POD Galerkin ROM (POD-Gal). Figure 5 shows the behavior of the TKE as a function of time for three values of N; predictions

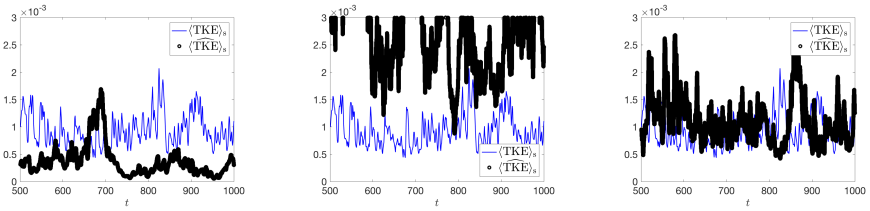
of first and second order moments — based on sampling times — are reported in the caption of the Figure. Results are consistent with the results in Figure 3. For small-to-moderate values of N, we observe several spurious behaviors, namely convergence to false stable steady flows, and overly unstable flows. As N increases, the accuracy of the Galerkin ROM appears to increase.



(a) $N = 20$ (b) $N = 40$ (c) $N = 60$

(d) $N = 20$ (e) $N = 40$ (f) $N = 60$

Figure 4: The solution reproduction problem; POD-Galerkin. Behavior of the sample mean and sample variance of the coefficients $\{a_n^j\}_j$. (Re = 15000).



(a) $N = 20$ (b) $N = 40$ (c) $N = 60$

Figure 5: The solution reproduction problem; POD-Galerkin. TKE prediction vs. truth estimate (6) as a function of time for three values of N. $\langle \widehat{\mathrm{TKE}} \rangle_s = 3.8 \cdot 10^{-4}$ ($N = 20$), $3.5 \cdot 10^{-3}$ ($N = 40$), $1.1 \cdot 10^{-3}$ ($N = 60$). $V_s(\widehat{\mathrm{TKE}}) = 8.8 \cdot 10^{-8}$ ($N = 20$), $6.5 \cdot 10^{-6}$ ($N = 40$), $1.9 \cdot 10^{-7}$ ($N = 60$). ($\langle \mathrm{TKE} \rangle_s = 9.4 \cdot 10^{-4}$, $V_s(\mathrm{TKE}) = 8.5 \cdot 10^{-8}$) (Re = 15000).

Interestingly, the behavior of the ROM observed here is qualitatively similar to the one observed for highly-truncated spectral approximations to turbulent flows ([47]). We argue that the need for large reduced spaces might greatly reduce the benefit of Model Reduction: if the value of N required to obtain sufficiently accurate results is too large, the resulting ROM might not lead to

significant computational speed-ups, and might also not be beneficial in terms of memory. This observation motivates the correction to the Galerkin formulation proposed in the next section. We finally remark that the results shown in this section suggest the need for a pragmatic definition of long-time stability: we address this issue in AppendixC.

3.2. The constrained POD-Galerkin formulation

3.2.1. Formulation

Given the reduced space $\mathcal{Z}^{\mathrm{u}} = \mathrm{span}\{\zeta_n\}_{n=1}^N \subset V_{\mathrm{div}}$, and the time grid $\{t_{\mathrm{g}}^j\}_{j=0}^J$, we seek the coefficients $\{\mathbf{a}^j\}_{j=0}^J \subset \mathbb{R}^N$ such that

$$\mathbf{a}^{j+1} := \arg\min_{\mathbf{a}\in\mathbb{R}^N} \|\mathbb{A}(\mathbf{a}^j;\mathrm{Re})\mathbf{a} - \mathbf{F}(\mathbf{a}^j;\mathrm{Re})\|_2^2, \quad \text{s.t. } \alpha_n \le a_n \le \beta_n, \;\; n = 1,\dots,N; \tag{13}$$

where $\{\alpha_n\}_{n=1}^N$ and $\{\beta_n\}_{n=1}^N$ are suitable hyper-parameters that will be specified later, and $\mathbb{A}(\cdot;\mathrm{Re})$, $\mathbf{F}(\cdot;\mathrm{Re})$ are defined in (10). Formulation (13) reads as a constrained quadratic programming problem where the objective function corresponds to the Euclidean norm of the error in the reduced Galerkin formulation, while the constraints impose that each coefficient of the N-term expansion remains in the interval $[\alpha_n, \beta_n]$, $n = 1,\dots,N$. We refer to (13) as *constrained (POD-)Galerkin formulation*.

The hyper-parameters $\{\alpha_n\}_n$ and $\{\beta_n\}_n$ are designed to compensate for the effects of the unmodelled physics by embedding in the ROM formulation information about the variation in time of the coefficients $\{a_n^j\}_j$, for $n = 1,\dots,N$. We observe that the solution coefficients $\{\mathbf{a}^j\}_j$ associated with (13) belong to the region $[\alpha_1,\beta_1]\times\ldots\times[\alpha_N,\beta_N]$ for all j: if $\max_n|\alpha_n|, \max_n|\beta_n| < \infty$, the reduced solution is guaranteed to be bounded for all time steps. We further observe that, for each value of n, if we introduce the projection[4] of the lifted field on the n-th POD mode at time t_{g}^j, $a_n^{\mathrm{FOM,j}} := (\mathring{u}^j, \zeta_n)_V$, we can interpret α_n and β_n as lower and upper bounds for the sequence $\{a_n^{\mathrm{FOM,j}}\}_{j=J_0}^J$, where $J_0 > 0$ is introduced in (5) to discard the transient dynamics. The hyper-parameters α_n and β_n are not directly related to the POD eigenvalues λ_n: the latter are — up to a multiplicative constant — estimates of the squared ℓ^2-norm of the coefficients, $\lambda_n = \sum_k \left(a_n^{\mathrm{FOM,k}}\right)^2 \approx \frac{K}{J}\sum_j \left(a_n^{\mathrm{FOM,j}}\right)^2$.

Based on the interpretation of the hyper-parameters, we propose to estimate $\{\alpha_n\}_{n=1}^N$ and $\{\beta_n\}_{n=1}^N$ based on the sample minima and the sample maxima associated with the snapshots $\{\mathring{u}^k\}_{k=1}^K$:

$$\alpha_n := m_n^{\mathrm{u}} - \epsilon\Delta_n^{\mathrm{u}}, \qquad \beta_n := M_n^{\mathrm{u}} + \epsilon\Delta_n^{\mathrm{u}}, \tag{14a}$$

where m_n^{u} and M_n^{u} are sample minimum and sample maximum associated with the projection of the lifted field on the n-th POD mode,

$$m_n^{\mathrm{u}} = \min_k a_n^{\mathrm{FOM,k}} := (\mathring{u}^k, \zeta_n)_V, \quad M_n^{\mathrm{u}} = \max_k a_n^{\mathrm{FOM,k}}; \tag{14b}$$

[4] We assume here that the POD eigenmodes $\{\zeta_n\}_{n=1}^N$ are orthonormalized.

Δ_n^{u} is the sample estimate of the difference between maximum and minimum,

$$\Delta_n^{\mathrm{u}} := M_n^{\mathrm{u}} - m_n^{\mathrm{u}}; \tag{14c}$$

and the constant $\epsilon > 0$ takes into account the fact that sample minima and sample maxima in (14b) are upper and lower bounds for the true minima and true maxima, respectively. We emphasize that in our framework $K \ll J$; therefore, $\{a_n^{\mathrm{FOM,k}}\}_{k=1}^K$ should be interpreted as a (deterministic) sample from the population $\{a_n^{\mathrm{FOM,j}}\}_{j=J_0}^J$. Given the special features of the learning task at hand — the estimation of minima and maxima of a population — we expect that we can estimate the hyper-parameters based on *sparse* DNS data (i.e., data that are not dense in any specific region of the time interval).

Accurate estimates of the hyper-parameters of the formulation based on sparse DNS data represent an important feature of our constrained formulation. As observed by many authors, low-frequency features of the turbulent flow — which largely contribute to long-time flow averages — are well-represented by the snapshots $\{\mathring{u}^k\}_{k=1}^K$ and consequently by the POD space only if the sampling times $\{t_{\mathrm{s}}^k\}_{k=1}^K$ are not clustered in any specific region of the time interval. This implies that both the ingredients of the ROM — the space $\mathcal{Z}^{\mathrm{u}}$ and the hyper-parameters $\{\alpha_n\}_n$ and $\{\beta_n\}_n$ — require the same sampling strategy for the construction of the snapshot set. Therefore, the same dataset used to generate the POD space is well-suited to estimate the hyper-parameters of the ROM. This observation allows us to limit the size K of the snapshot set, and ultimately leads to a reduction of the offline memory cost.

Unlike the standard POD-Galerkin ROM, we here use DNS data twice: first, to build the space $\mathcal{Z}^{\mathrm{u}}$; second, to estimate the hyper-parameters $\{\alpha_n\}_{n=1}^N$ and $\{\beta_n\}_{n=1}^N$. Furthermore, while POD-Galerkin is independent of the particular basis $\zeta_1, \ldots, \zeta_N$ chosen for $\mathcal{Z}^{\mathrm{u}}$, the box constraints in (10) depend on the choice of the basis. We emphasize that by choosing $\{\zeta_n\}_n$ as basis of $\mathcal{Z}^{\mathrm{u}}$ we explicitly incorporate (prior) information about the decay of the POD coefficients directly in the formulation. Figure 6 shows the flow chart associated with the offline development of the constrained POD-Galerkin ROM (compare with Figure 2).

We observe that if the solution to Galerkin ROM (10) — $\mathbf{a}_{\mathrm{Gal}}^{j+1} = \mathbb{A}(\mathbf{a}^j; \mathrm{Re})^{-1}$ $\mathbf{F}(\mathbf{a}^j; \mathrm{Re})$ — satisfies the box constraints in (13), then $\mathbf{a}^{j+1} = \mathbf{a}_{\mathrm{Gal}}^{j+1}$. Therefore, our constrained formulation *corrects* the Galerkin formulation only if $\mathbf{a}_{\mathrm{Gal}}^{j+1}$ does not satisfy the prescribed bounds. This represents the main difference between our approach and the other stabilized ROMs proposed in the literature and briefly mentioned in the introduction: rather than introducing artificial dissipation in the Galerkin model, we exploit prior information about the attractor to correct the ROM.

We finally comment on time discretization. In this work, we employ the first-order semi-implicit time-discretization introduced in (9). However, the approach can be trivially extended to other time discretizations: first, we derive the discrete Galerkin ROM from (8), then we substitute the resulting algebraic formulation in the objective function of (13). For explicit and semi-implicit single-step time integrators, the resulting constrained formulation corresponds

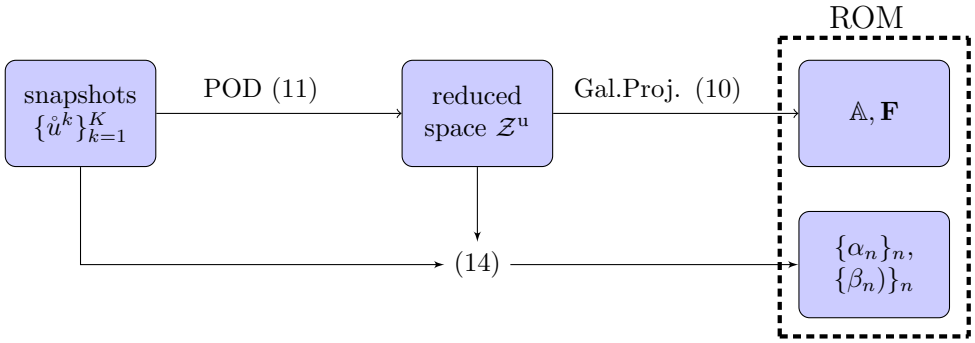


Figure 6: Flow chart of the constrained POD-Galerkin process

to a quadratic programming problem, which can be solved using interior-point methods (see, e.g., [85]). For fully-implicit single-step methods, the constrained formulation reads as a nonlinear constrained optimization problem, which again can be solved using interior-point methods or sequential quadratic programming. We envision that the extension to multistep methods might require some additional care since the solution is not guaranteed to be smooth in time when the constraints are active. A thorough analysis of different time integrators is beyond the scope of this paper: here, we remark that for single-step methods there exists $\Delta t^\star > 0$ such that for all $\Delta t \leq \Delta t^\star$ we have

$$\|\mathbf{a}^{j+1} - \mathbf{a}^j\|_2 \leq C\Delta t, \tag{15}$$

where C depends on the Reynolds number, $\{\zeta_n\}_n$, $\{\alpha_n\}_n$, $\{\beta_n\}_n$ and R_g. We defer the proof to AppendixD.

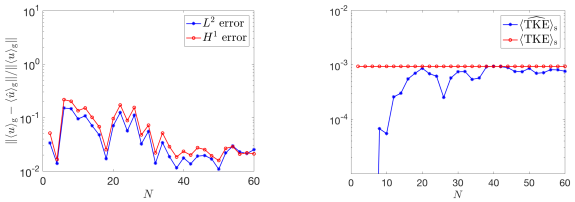
3.2.2. Performance of the constrained POD-Galerkin ROM

We present numerical results for Re = 15000. Time grid $\{t_{\mathrm{g}}^j\}_{j=0}^J$ and sampling times $\{t_{\mathrm{s}}^k\}_{k=1}^K$ are the same considered for POD-Galerkin, if not specified otherwise. As for the previous test the ROM is implemented in Matlab; the quadratic programming problem is solved using the routine `quadprog` based on an interior-point algorithm. We here set $\epsilon = 0.01$ in (14).

Figure 7(a) shows the behavior of the relative L^2 and H^1 errors in the mean flow prediction with N, while Figure 7(b) shows the behavior of the mean predicted TKE with N. We observe that the constrained formulation leads to a substantial improvement in performance compared to the standard POD-Galerkin method (cf. Figures 3(b) and 3(c)): for $N \gtrsim 40$ the relative error in the mean is less than 2%, while the predicted mean TKE is bounded from above by $\langle \mathrm{TKE} \rangle_{\mathrm{s}}$ for all values of N. Furthermore, we observe that the TKE of our constrained Galerkin formulation is larger than the one predicted by the Galerkin ROM for certain values of N, and is smaller for other values of N: this empirically proves that our approach does not necessarily add dissipation to the Galerkin ROM. Finally, we observe a distinctly non-monotonic behavior

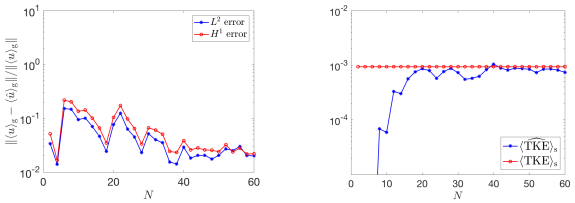
around $N = 18$, which corresponds to a change in the temporal dynamics of the reduced system from chaotic ($N = 18$) to quasiperiodic ($N = 20$): since for $N \approx 20$ the flow is still extremely under-resolved, the constrained ROM does not provide a stable (with N) representation of the dynamics.

In Figure 8, we repeat the tests of Figure 7 for $\Delta t' = 0.5\Delta t = 2.5 \cdot 10^{-3}$. We observe that results are consistent with the results shown in Figure 7: this provides empirical evidence for the stability of our constrained formulation under time-step refinement. Figure 9 shows the behavior of the sample mean and sample variance of the coefficients $\{a_n^j\}_j$ for three different values of N. Also in this case, we observe a substantial improvement in performance compared to POD-Galerkin, particularly for high modes. Finally, Figure 10 shows the behavior of the TKE as a function of time for three values of N. We observe that for $N = 40$ and $N = 60$ the predicted TKE is in good qualitative agreement with (6); in addition, predictions of first- and second-order moments (reported in the caption) are accurate. In AppendixE, we present additional results to demonstrate the efficiency of the constrained formulation, and also the robustness with respect to the choice of ϵ; furthermore, in AppendixH we discuss in detail the computational cost of the procedure.



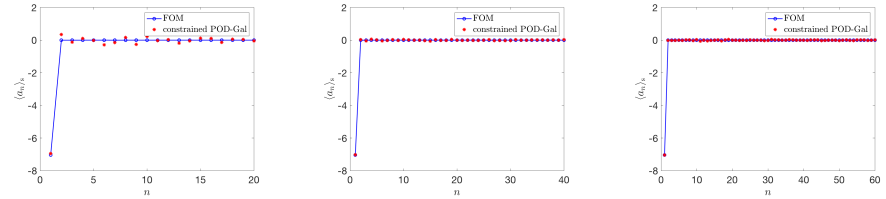
(a) (b)

Figure 7: The solution reproduction problem; constrained POD-Galerkin. Figure (a): behavior of the relative L^2 and H^1 errors in mean flow prediction with N. Figure (b): TKE prediction $\widehat{\langle \mathrm{TKE}\rangle_s}$ vs. truth estimate (7) for multiple values of N. ($\mathrm{Re} = 15000$, $\epsilon = 0.01$).

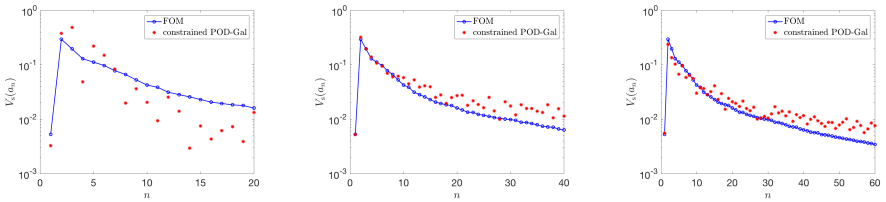


(a) (b)

Figure 8: The solution reproduction problem; constrained POD-Galerkin for a finer time grid. Figure (a): behavior of the relative L^2 and H^1 errors in mean flow prediction with N. Figure (b): TKE prediction $\widehat{\langle \mathrm{TKE}\rangle_s}$ vs. truth estimate (7) for multiple values of N. ($\mathrm{Re} = 15000$, $\epsilon = 0.01$, $\Delta t = 2.5 \cdot 10^{-3}$).

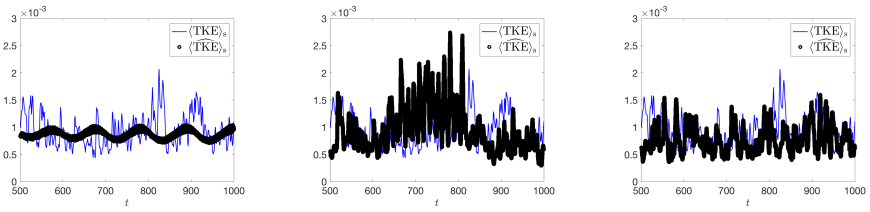


(a) $N = 20$ (b) $N = 40$ (c) $N = 60$



(d) $N = 20$ (e) $N = 40$ (f) $N = 60$

Figure 9: The solution reproduction problem; constrained POD-Galerkin. Behavior of the sample mean and sample variance of the coefficients $\{a_n^j\}_j$. ($\mathrm{Re} = 15000$, $\epsilon = 0.01$).



(a) $N = 20$ (b) $N = 40$ (c) $N = 60$

Figure 10: The solution reproduction problem; constrained POD-Galerkin. TKE prediction vs. truth estimate (6) as a function of time for three values of N. $\langle \widehat{\mathrm{TKE}} \rangle_s = 8.6 \cdot 10^{-4}$ ($N = 20$), $9.4 \cdot 10^{-4}$ ($N = 40$), $7.7 \cdot 10^{-4}$ ($N = 60$). $V_s(\widehat{\mathrm{TKE}}) = 5.5 \cdot 10^{-9}$ ($N = 20$), $1.7 \cdot 10^{-7}$ ($N = 40$), $5.8 \cdot 10^{-8}$ ($N = 60$). ($\langle \mathrm{TKE} \rangle_s = 9.4 \cdot 10^{-4}$, $V_s(\mathrm{TKE}) = 8.5 \cdot 10^{-8}$) ($\mathrm{Re} = 15000$, $\epsilon = 0.01$).

4. The parametric problem

We consider the extension of our MOR approach to the parametric context. For the purpose of exposition, we focus our discussion on the lid-driven cavity problem presented in section 2: we wish to estimate the solution to (2) for $\mathrm{Re} \in \mathcal{P} = [15000, 25000]$. In view of the h-refinement, we introduce the partition of $\mathcal{P}$, $\mathcal{I}_1, \ldots, \mathcal{I}_M$ such that $\bigcup_{m=1}^{M} \mathcal{I}_m = \mathcal{P}$, $\mathcal{I}_m \cap \mathcal{I}_{m'} = \emptyset$. We seek an estimate of the lifted velocity field $\mathring{u} = \mathring{u}(x, t; \mathrm{Re})$ of the form $\hat{u}(x, t; \mathrm{Re}) = \sum_{n=1}^{N} a_n^m(t; \mathrm{Re}) \zeta_n^m(x)$ for all $\mathrm{Re} \in \mathcal{I}_m$. The approach can be trivially extended to other parametric problems that do not involve geometric

variations; as already mentioned in section 3, the extension to the latter case is beyond the scope of the present work. Algorithm 2 summarizes the general offline/online paradigm for the parametric problem. We highlight that, for the sake of generality, in Algorithm 2 we distinguish between L (number of offline solves) and M (number of reduced spaces). However, in this work we consider the case $L = M$.

In order to tackle the parametric problem outlined above, we should address two challenges: first, we should extend the constrained formulation to the parametric case; second, we should develop a Greedy strategy for the proper selection of the parameters $\mathrm{Re}_1^\star, \dots, \mathrm{Re}_L^\star$, and the partition $\{\mathcal{I}_m\}_m$. We emphasize that a proper selection of the parameters reduces the number of offline full order solves, and is thus crucial for the feasibility of the approach. In order to address the first challenge, we propose an actionable procedure for the selection of the hyper-parameters $\{\alpha_n\}_n$ and $\{\beta_n\}_n$ associated with the constrained formulation (13) in the parametric case. On the other hand, the Greedy approach relies on an inexpensive error indicator, which corresponds to the dual norm of the residual associated with the time-averaged momentum equation.

Algorithm 2 Offline/online paradigm for the parametric problem

Task: find an estimate of the lifted velocity field $\mathring{u} = \mathring{u}(x, t; \mathrm{Re})$ of the form $\hat{u}(x, t; \mathrm{Re}) = \sum_{n=1}^{N} a_n^m(t; \mathrm{Re})\zeta_n^m(x)$ for all $\mathrm{Re} \in \mathcal{I}_m$, $m = 1, \dots, M$.

Offline stage

1: Generate the DNS data $\{\mathring{u}^k(\mathrm{Re}_\ell{}^\star) := \mathring{u}(t_{\mathrm{s}}^k; \mathrm{Re}_\ell^\star)\}_{k=1}^K \subset V$, and $\mathrm{Re}_1^\star, \dots, \mathrm{Re}_L^\star \in \mathcal{P}$.

2: Generate the partition $\{\mathcal{I}_m\}_m$ of $\mathcal{P}$, and the reduced spaces $\mathcal{Z}_m^{\mathrm{u}} = \mathrm{span}\{\zeta_n^m\}_{n=1}^N$, $m = 1, \dots, M$.

3: Formulate the Reduced Order Models for each subregion.

Online stage

1: Given $\mathrm{Re} \in \mathcal{P}$, find $m \in \{1, \dots, M\}$ such that $\mathrm{Re} \in \mathcal{I}_m$.

2: Estimate the coefficients $\{a_n^{m,j}(\mathrm{Re}) = a_n^m(t_{\mathrm{g}}^j; \mathrm{Re})\}_{n=1}^N$ for $j = 0, 1, \dots, J$.

3: Compute the QOIs (e.g., mean flow, TKE,...).

The section is organized as follows. In section 4.1 we present the POD-hGreedy approach, in section 4.2 we present the ROM formulation, and we discuss the choice of the hyper-parameters $\{\alpha_n\}_n$ and $\{\beta_n\}_n$. Then, in section 4.3 we propose the time-averaged error indicator. Finally, in section 4.4, we present the numerical results for the lid-driven cavity problem.

4.1. POD-Greedy algorithm

We first present the POD-hGreedy algorithm for the construction of the reduced spaces $\{\mathcal{Z}^{\mathrm{u}}_m\}_{m=1}^M$, and the partition $\{\mathcal{I}_m\}_{m=1}^M$ of $\mathcal{P}$, based on the results of L FOM simulations associated with the parameters $\mathrm{Re}^\star_1, \ldots, \mathrm{Re}^\star_L$. The approach is a simplified version of the h-refinement procedure for parabolic problems proposed in [50]. In particular, we here consider $L = M$: this implies that each reduced space $\mathcal{Z}^{\mathrm{u}}_\ell$ is based on the POD of a single full-order solve.

In view of the presentation of the algorithm, we introduce the discretized parameter space $\mathcal{P}_{\mathrm{train}} = \{\mathrm{Re}_i\}_{i=1}^{n_{\mathrm{train}}}$, $\mathrm{Re}_1 \leq \ldots \leq \mathrm{Re}_{n_{\mathrm{train}}}$, the integers L and N, which fix the maximum number of offline solves and the size of the reduced space $\mathcal{Z}^{\mathrm{u}}$, the integer $n_{\mathrm{cand}} < L$, which is the number of ROM evaluations performed online for a given value of the parameters, and the *a posteriori* error indicator $\Delta^{\mathrm{u}} : \bigotimes_{j=0}^J V \times \mathcal{P} \to \mathbb{R}_+$. The error indicator takes as input a sequence $\{w^j\}_{j=0}^J \subset V$ and the value of the parameter, and returns an estimate of the error in the prediction of the mean flow; we formally present the indicator in section 4.3. We further introduce the functions

$$[\{\zeta_n\}_{n=1}^N] = \texttt{POD}_V\left(\mathcal{S}, N\right); \quad [\{\mathring{u}^k(\mathrm{Re})\}_{k=1}^K] = \texttt{DNS-solver}\left(\mathrm{Re}, \{t^k_{\mathrm{s}}\}_{k=1}^K\right);$$

$$[\{\hat{u}^j(\mathrm{Re})\}_{j=0}^J] = \texttt{ROM-solver}\left(\mathrm{Re}, \mathcal{Z}^{\mathrm{u}}\right).$$

$\texttt{POD}_V$ takes as input the set of snapshots $\mathcal{S} = \{w^i\}_{i=1}^{|\mathcal{S}|}$ and an integer $N > 0$, and returns the orthonormalized first N POD eigenmodes (see section 3.1.2); on the other hand, `DNS-solver` takes as input the value of the Reynolds number and the sampling times $\{t^k_{\mathrm{s}}\}_{k=1}^K$, and returns the instantaneous velocity at times $\{t^k_{\mathrm{s}}\}_{k=1}^K$; finally, `ROM-solver` takes as input the value of the Reynolds number and the reduced space $\mathcal{Z}^{\mathrm{u}}$, and returns the ROM solution $\hat{u}^j = \sum_{n=1}^N a^j_n \zeta_n$ for each time step of the grid $\{t^j_{\mathrm{g}}\}_{j=0}^J$. Algorithm 3 presents the computational procedure for both offline and online stage. With some abuse of notation, we use $\Delta^{\mathrm{u}}_\ell(\cdot)$ to refer to the error estimate associated with the ℓ-th model.

Algorithm 3 combines a POD in time with a Greedy in parameter. As explained in the introduction, Greedy techniques are crucial to allow efficient parameter explorations at an affordable offline computational cost. We emphasize that our approach is different from the POD-Greedy strategy proposed in [49]: rather than building L different reduced spaces, the authors of [49] combine data from different parameters to generate a single reduced space.

We also remark that our definition of the partition — see Algorithm 3, Online stage — can be formally expressed as follows:

$$\mathcal{I}_\ell = \{\mathrm{Re} \in \mathcal{P} : \ \ell \in I(\mathrm{Re}), \ \Delta^{\mathrm{u}}_\ell(\mathrm{Re}) < \Delta^{\mathrm{u}}_{\ell'}(\mathrm{Re}), \ell' \in I(\mathrm{Re}), \ell' \neq \ell\}, \tag{16}$$

where $I(\mathrm{Re}) \subset \{1, \ldots, L\}$ is the set of indices associated with the n_{cand} nearest anchor points to Re.

4.2. Constrained Galerkin formulation

Given the reduced spaces $\{\mathcal{Z}^{\mathrm{u}}_\ell = \mathrm{span}\{\zeta^\ell_n\}_{n=1}^N\}_{\ell=1}^L$, we consider the constrained Galerkin formulation proposed in section 3: given $\mathrm{Re} \in \mathcal{P}$, and the

time grid $\{t_{\mathrm{g}}^j\}_{j=0}^J$, the ℓ-th ROM seeks the coefficients $\{\mathbf{a}_\ell^j\}_{j=0}^J \subset \mathbb{R}^N$ such that

$$\mathbf{a}_\ell^{j+1} := \arg\min_{\mathbf{a}\in\mathbb{R}^N} \|\mathbb{A}_\ell(\mathbf{a}^j;\mathrm{Re})\mathbf{a} - \mathbf{F}_\ell(\mathbf{a}^j;\mathrm{Re})\|_2^2, \quad \text{s.t. } \alpha_n^\ell(\mathrm{Re}) \le a_n \le \beta_n^\ell(\mathrm{Re}), \quad n = 1,\dots,N; \tag{17}$$

where $\mathbb{A}_\ell$ and $\mathbf{F}_\ell$ can be computed by exploiting (10) for the reduced space $\mathcal{Z}_\ell^{\mathrm{u}}$, and the constraints $\{\alpha_n^\ell\}_{n,\ell}$ and $\{\beta_n^\ell\}_{n,\ell}$ are based on the DNS data for the anchor point $\mathrm{Re}_\ell^\star$. In greater detail, given $\ell \in \{1,\dots,L\}$, assuming that $\zeta_1^\ell,\dots,\zeta_N^\ell$ are orthonormal in V, we define $\{\alpha_n^\ell\}_n$ and $\{\beta_n^\ell\}_n$ such that

$$\alpha_n^\ell := m_{n,\ell}^{\mathrm{u}} - \epsilon\Delta_{n,\ell}^{\mathrm{u}}, \quad \beta_n^\ell := M_{n,\ell}^{\mathrm{u}} + \epsilon\Delta_{n,\ell}^{\mathrm{u}}, \tag{18a}$$

Algorithm 3 POD-hGreedy algorithm for the construction of $\{\mathcal{Z}_\ell^{\mathrm{u}}, \mathcal{I}_\ell\}_\ell$

Offline stage: $[\{\mathcal{Z}_\ell^{\mathrm{u}}\}_{\ell=1}^L] = \texttt{Offline}\ (\mathcal{P}_{\mathrm{train}}, N, L, \Delta^{\mathrm{u}}, \{t_{\mathrm{s}}^k\}_{k=1}^K)$.
Inputs: $\mathcal{P}_{\mathrm{train}} = \{\mathrm{Re}_i\}_{i=1}^{n_{\mathrm{train}}}$ = discretized parameter space, N = dimension of each reduced space, L = maximum number of offline solves, Δ^{u} = error indicator, $\{t_{\mathrm{s}}^k\}_{k=1}^K$ = sampling times.
Output: $\{(\mathcal{Z}_\ell^{\mathrm{u}}, \mathrm{Re}_\ell^\star\}_{\ell=1}^L$ = reduced space/anchor point pairs.

$\mathrm{Re}_1^\star = \mathrm{rand}(\mathcal{P}_{\mathrm{train}})$
for $\ell = 1,\dots,L$ **do**
  $[\{\mathring{u}^k(\mathrm{Re}_\ell^\star)\}_{k=1}^K] = \texttt{DNS-solver}\left(\mathrm{Re}_\ell^\star, \{t_{\mathrm{s}}^k\}_{k=1}^K\right)$
  $[\{\zeta_n^\ell\}_{n=1}^N] = \texttt{POD}_V\left(\{\mathring{u}^k(\mathrm{Re}_\ell^\star)\}_{k=1}^K, N\right)$
  Define $\mathcal{Z}_\ell^{\mathrm{u}} = \mathrm{span}\{\zeta_n^\ell\}_{n=1}^N$, build the ROM structures (cf. sections 3.1.1 and 4.3).
  for $i = 1,\dots,n_{\mathrm{train}}$ **do**
    $[\{\hat{u}_\ell^j(\mathrm{Re}_i)\}_{j=0}^J] = \texttt{ROM-solver}\,(\mathrm{Re}_i, \mathcal{Z}_\ell^{\mathrm{u}})$;
    Compute the error estimate $\Delta_\ell^{\mathrm{u}}(\mathrm{Re}_i)$
  end for
  $\mathrm{Re}_{\ell+1}^\star = \arg\max_{\mathrm{Re}\in\mathcal{P}_{\mathrm{train}}} \min_{\ell'=1,\dots,\ell} \Delta_{\ell'}^{\mathrm{u}}(\mathrm{Re})$.
end for

Online stage: $[\{\hat{u}^j\}_j] = \texttt{Online}\ (\{(\mathcal{Z}_\ell^{\mathrm{u}}, \mathrm{Re}_\ell^\star\}_{\ell=1}^L, \Delta^{\mathrm{u}}, n_{\mathrm{cand}}, \mathrm{Re})$.
Inputs: $\{(\mathcal{Z}_\ell^{\mathrm{u}}, \mathrm{Re}_\ell^\star\}_{\ell=1}^L$ = reduced space/anchor point pairs, n_{cand} = online ROM evaluations, Re = input parameter.
Output: $\{\hat{u}^j\}_j$ = solution estimate.

Find the n_{cand} nearest anchors to Re: $\mathrm{Re}_{(1)}^\star,\dots,\mathrm{Re}_{(n_{\mathrm{cand}})}^\star$
for $i = 1,\dots,n_{\mathrm{cand}}$ **do**
  $[\{\hat{u}_{(i)}^j\}_{j=0}^J] = \texttt{ROM-solver}\left(\mathrm{Re}, \mathcal{Z}_{(i)}^{\mathrm{u}}\right)$
  Compute the error estimate $\Delta_{(i)}^{\mathrm{u}}(\mathrm{Re})$
end for
Return $\{\hat{u}^j = \hat{u}_{(i^\star)}^j\}_j$, where $i^\star$ is the minimizer of $\{\Delta_{(i)}^{\mathrm{u}}(\mathrm{Re})\}_i$.

where

$$m^{\mathrm{u}}_{n,\ell} := \min_k a^{\mathrm{FOM,k}}_{n,\ell}, \quad M^{\mathrm{u}}_{n,\ell} := \max_k a^{\mathrm{FOM,k}}_{n,\ell}, \quad \Delta^{\mathrm{u}}_{n,\ell} := M^{\mathrm{u}}_{n,\ell} - m^{\mathrm{u}}_{n,\ell}, \tag{18b}$$

and $a^{\mathrm{FOM,k}}_{n,\ell} := (\mathring{u}^k(\mathrm{Re}^\star_\ell), \zeta^\ell_n)_V$. The offline/online decomposition is equivalent to the one described in section 3.1.1. We omit the details.

We observe that our choices of α^ℓ_n and β^ℓ_n correspond to a constant approximation of the functions

$$m^{\mathrm{FOM,u}}_{n,\ell}(\mathrm{Re}) := \min_{j=J_0,\ldots,J} \left(\mathring{u}^j(\mathrm{Re}), \zeta^\ell_n\right)_V, \quad M^{\mathrm{FOM,u}}_{n,\ell}(\mathrm{Re}) := \max_{j=J_0,\ldots,J} \left(\mathring{u}^j(\mathrm{Re}), \zeta^\ell_n\right)_V; \tag{19}$$

where $\mathcal{I}_\ell \subset \mathcal{P}$ is defined in (16). We observe that the piece-wise constant approximations of $m^{\mathrm{FOM,u}}_{n,\ell}$ and $M^{\mathrm{FOM,u}}_{n,\ell}$ are justified by our Greedy algorithm, which adaptively determines the partition of $\mathcal{P}$ based on the error indicator. For practical parametrizations, and practical values of L (i.e., number of offline solves) we expect that accurate estimates of $m^{\mathrm{FOM,u}}_{n,\ell}$ and $M^{\mathrm{FOM,u}}_{n,\ell}$ over $\mathcal{P}$ might be out of reach. Therefore, we here effectively rely on (i) the robustness of our constrained approach to perturbations in the value of the hyper-parameters, and (ii) the weak sensitivity of the functions $m^{\mathrm{ROM,u}}_{n,\ell}$ and $M^{\mathrm{ROM,u}}_{n,\ell}$ with respect to the parameter. For the lid-driven cavity problem considered in this work, we provide numerical evidence to support these two assumptions in AppendixE.

4.3. A time-averaged error indicator

Given the sequence $\{w^j\}_{j=0}^J \subset V$ and $\mathrm{Re} \in \mathcal{P}$, we define the discrete time-averaged residual $\langle R\rangle : \bigotimes_{j=0}^J V \times V_{\mathrm{div}} \times \mathcal{P} \to \mathbb{R}$ associated with (9):

$$\langle R\rangle\left(\{w^j\}_{j=0}^J, v; \mathrm{Re}\right) = \frac{\Delta t}{T - T_0} \sum_{j=J_0}^{J-1} e(\hat{u}^j, \hat{u}^{j+1}, \mathrm{Re}) \tag{20a}$$

where $T = t^J_{\mathrm{g}}$, $T_0 = t^{J_0}_{\mathrm{g}}$, and

$$\begin{aligned} e(\hat{u}^j, \hat{u}^{j+1}, \mathrm{Re}) := &\left(\frac{\hat{u}^{j+1} - \hat{u}^j}{\Delta t}, v\right)_{L^2(\Omega)} + \frac{1}{\mathrm{Re}}(\hat{u}^{j+1} + R_g, v)_V \\ &+ c(\hat{u}^j + R_g, \hat{u}^{j+1} + R_g, v),\ j = J_0, \ldots, J-1 \end{aligned} \tag{20b}$$

Then, we define the error indicator $\Delta^{\mathrm{u}} : \bigotimes_{j=0}^J V \times \mathcal{P} \to \mathbb{R}_+$ as follows:

$$\Delta^{\mathrm{u}}\left(\{w^j\}_{j=0}^J; \mathrm{Re}\right) := \left\| \langle R\rangle\left(\{w^j\}_{j=0}^J, \cdot; \mathrm{Re}\right) \right\|_{V'_{\mathrm{div}}} \tag{21}$$

where $\|\cdot\|_{V'_{\mathrm{div}}}$ denotes the norm of the dual space V'_{div}.

In our numerical tests, as in (5), we consider J_0 such that $t^{J_0}_{\mathrm{g}} = T_0 = 500$: this choice is designed to limit the effect of the transient dynamics. It is easy to verify that the solution to the FOM for any initial condition — provided that the same time discretization is employed — satisfies $\Delta^{\mathrm{u}} \equiv 0$. This implies that

two sequences $\{w^j\}_{j=0}^J, \{\tilde{w}^j\}_{j=0}^J \subset V$ satisfying $\Delta^{\mathrm{u}} \equiv 0$ might be far from each other at each time step (i.e., $\|\tilde{w}^j - w^j\|_V$ is large for any $j \geq 0$). However, for sufficiently large values of J, we expect Δ^{u} to be highly-correlated with the error in the mean flow prediction; for this reason, we can exploit Δ^{u} to guide the Greedy algorithm presented in section 4.1. We empirically investigate the correlation between Δ^{u} and the error in the mean flow prediction in the numerical experiments at the end of the section. A theoretical justification of the error indicator is beyond the scope of the present work.

The error indicator Δ^{u} can be computed efficiently for sequences in $\mathcal{Z}^{\mathrm{u}}$ exploiting an offline/online computational decomposition; the procedure is standard in the Reduced Basis literature, and is reported in AppendixG.

Remark 4.1. *We do not expect that the residual indicator* (21) *is in good quantitative agreement with the error in mean flow prediction* $\|\langle u - \hat{u}\rangle_{\mathrm{g}}\|_V$*. More precisely, if we define the effectivity* $\eta := \frac{\Delta^{\mathrm{u}}}{\|\langle u-\hat{u}\rangle_{\mathrm{g}}\|_V}$ *of the residual error indicator, we do not expect that* η *is close to one.*

In order to obtain a quantitative estimate of the error of the ROM anchored in $\mathrm{Re}^\star$*, we can consider the corrected estimator*

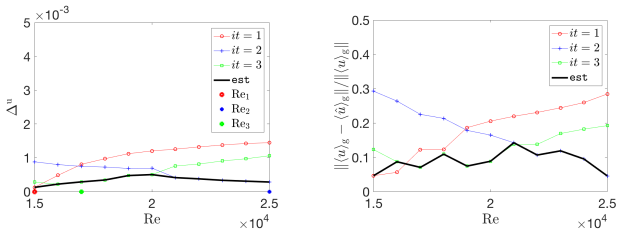
$$\Delta^{\mathrm{u,corr}}(\mathrm{Re}; \mathrm{Re}^\star) := \frac{1}{\eta(\mathrm{Re}^\star)} \Delta^{\mathrm{u}}(\mathrm{Re}), \tag{22}$$

where $\Delta^{\mathrm{u}}(\mathrm{Re})$ *is the error indicator associated with the ROM anchored in* $\mathrm{Re}^\star$*, and* $\eta(\mathrm{Re}^\star)$ *is the effectivity evaluated at* $\mathrm{Re} = \mathrm{Re}^\star$*. Note that the computation of* $\eta(\mathrm{Re}^\star)$ *does not require any additional call to the DNS solver.*

4.4. Numerical results

Figure 11 shows the results of the application of Algorithm 3 for the construction of the ROM for the parametric problem. In order to assess performance, we generate DNS data for $t_{\mathrm{g}}^j \in \{0, \ldots, 1500\}$, $\{t_{\mathrm{s}}^k = 500 + k\}_{k=1}^{K=1000}$ for $\mathrm{Re} = 15000, 16000, \ldots, 25000$ ($n_{\mathrm{train}} = 11$ datapoints). Then, we apply Algorithm 3 with $\mathrm{Re}_1^\star = 15000$, $N = 80$, and $\epsilon = 0.05$. We perform $L = 3$ iterations of the Greedy procedure. Figure 11(a) shows the behavior of Δ^{u} with Re for the three iterations, while Figure 11(b) shows the behavior of the relative H^1 error in mean flow prediction with Re. The black continuous line denotes the performance of the reduced model which minimizes the error indicator, and thus is selected by the Greedy procedure (cf. Algorithm 3, $n_{\mathrm{cand}} = 2$). We observe that the maximum relative error decreases at each iteration, and it is roughly 13% after the third iteration.

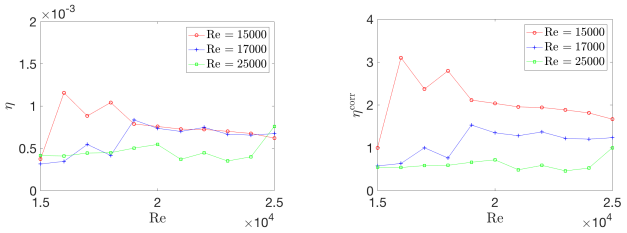
Results of Figure 11 show the importance of the error indicator Δ^{u} in (21) to select the parameters Re_2 and Re_3, and also motivate the choice of the partition $\{\mathcal{I}_\ell\}_\ell$ in (16): after the third iteration, for 10 out of 11 values of the Reynolds number, the reduced model that minimizes the error indicator (over all models) is the same that minimizes the true error. On the other hand, we observe that the indicator is in poor quantitative agreement with the true error: Figure 12(a) shows that the effectivity η of the error indicator is $\mathcal{O}(10^{-3})$ for all three



(a) (b)

Figure 11: The parametric problem; performance of POD-hGreedy. Figure (a): behavior of Δ^{u} with Re for three iterations. Figure (b): behavior of the relative H^1 error in mean flow prediction with Re for three iterations; the black line (`est`) shows the performance of the reduced model which minimizes the error indicator (and thus is selected by the Greedy procedure). ($\epsilon = 0.05$, $N = 80$,$\mathrm{Re}_1 = 15000, \mathrm{Re}_2 = 25000, \mathrm{Re}_3 = 17000$).

reduced order models and for all values of the Reynolds numbers considered. However, Figure 12(b) shows that the correction proposed in Remark 4.1 leads to an indicator that is in reasonable quantitative agreement with the error in mean flow prediction.

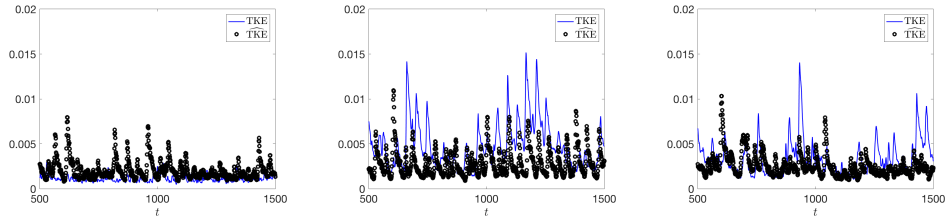


(a) (b)

Figure 12: The parametric problem; error estimator. Figure (a): behavior of the effectivity η of the error indicator Δ^{u} (21) for the three ROMs. Figure (b): behavior of the effectivity η^{corr} of the corrected error indicator $\Delta^{\mathrm{u,corr}}$ (22) for the three ROMs.

Figure 13 shows the behavior of the TKE prediction ($\widehat{\mathrm{TKE}}$) and of the truth estimate (6) for three values of the Reynolds number, Re =16000, 20000, 23000, which have not been selected by the Greedy procedure. Here, predictions are based on the ROM after three iterations of Algorithm 2. For Re = 16000, we observe significantly larger spikes for the ROM than for the FOM. These spikes, which are associated with eddies that are ejected into the core region — see AppendixA for further details — are observed in the DNS data for Re $\geq$ 17000: since the ROM used for Re = 16000 is built using DNS data associated with Re = 17000, modes associated with this flow feature enter in the reduced space and ultimately make the ROM more prone to show this instability. On a positive note, comparing the results of Figure 13(a) and Figure 13(b), we observe that the spikes are significantly more pronounced for Re = 20000 even if the reduced

space and the box constraints employed are the same: this empirically shows that our ROM is sensitive to changes in the Reynolds number.



(a) Re = 16000 (b) Re = 20000 (c) Re = 23000

Figure 13: The parametric problem; TKE prediction ($\widehat{\text{TKE}}$) vs. truth estimate (6) as a function of time for three values of the Reynolds number. ($\epsilon = 0.05$, $N = 80$, $\text{Re}_1 = 15000, \text{Re}_2 = 25000, \text{Re}_3 = 17000$). Predictions for Re_1 and Re_2 rely on the ROM anchored in $\text{Re}^\star = 17000$, while predictions for Re_3 rely on the ROM anchored in $\text{Re}^\star = 25000$.

5. Conclusions

In this paper, we present a Reduced Basis technique for long-time integration of turbulent flows. The three contributions of this work are (i) a constrained Galerkin formulation that corrects the Galerkin statement by incorporating prior information about the long-time attactor, (ii) an inexpensive time-averaged indicator for the error in mean flow prediction, and (iii) a POD-hGreedy technique for the construction of the ROM. In order to assess performance, we apply our approach to a lid-driven cavity problem parametrized with respect to the Reynolds number: first, we consider the solution reproduction problem (non-predictive case) to demonstrate the effectivity of our new constrained formulation; second, we consider the parametric problem (predictive case) to validate our error indicator, and more broadly the POD-Greedy procedure.

Our constrained Galerkin formulation is able to accurately predict mean flow and also the TKE. The error indicator, even if it is not corroborated by a firm theoretical analysis, is found to be highly-correlated with the error in the prediction of the mean flow; hence, it is naturally suited to drive the offline Greedy.

In this paper, we also highlight a number of challenges, which are particularly relevant for turbulent flows, and that should be taken into consideration in the design of MOR strategies for turbulent flows: first, the slow convergence of the Kolmogorov N-width suggested by Figure 3(a) which prevents us from accurately representing the full dynamics; second, the difficulty to combine modes associated with different parameters (cf. AppendixF); third, the large offline costs both in terms of computational time and required storage. In this paper, we propose to address the first challenge by *reducing our goal*: rather than

trying to estimate the full trajectory, we develop a ROM uniquely for the prediction of first and second moments of the long-time dynamics. Furthermore, we propose to address the second challenge by resorting to an h-refinement in parameter. On the other hand, we here postulate that the snapshot set $\{u^k\}_{k=1}^K$ is rich enough to accurately estimate the first N POD modes associated with the full trajectory $\{\mathring{u}^j\}_{j=J_0}^J$, and also that it is possible to compute and store the Riesz representers $\tilde{\xi}_1, \ldots, \tilde{\xi}_M$, $M = N^2 + 3N + 2$, for residual calculations. In AppendixB, we review a computational strategy to assess *a posteriori* the representativity of our snapshot set; on the other hand, we refer to a future work for the development of computational strategies to reduce the offline costs related to residual calculations.

We finally outline a number of potential next steps that we wish to pursue in the future.

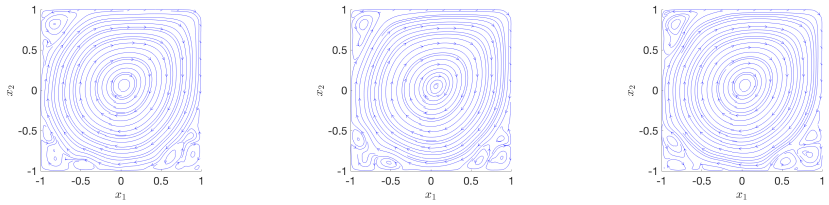
- *Constrained formulation* Our constrained formulation minimizes the ℓ^2 error in the reduced Galerkin statement subject to lower and upper bounds for the coefficients of the N-term expansion. We wish to consider other choices both for the objective functions and for the constraints. In particular, we wish to minimize the residual in a suitable dual norm, and we wish to design other constraints to take into account the properties of the attractor. Furthermore, we also wish to consider the post-processing rectification method proposed in [86] to improve the accuracy of the mean flow. Finally, we wish to consider alternative strategies for writing the nonlinear term in the momentum equation, and also for imposing strong boundary conditions.

- *hp-Greedy* In AppendixF, we discuss the limitation of the traditional POD-(p)Greedy algorithm. However, in order to tackle complex parametrizations, we envision that the h-refinement strategy proposed in this paper might require an unfeasible number of offline simulations. This is why we wish to consider more advanced sampling strategies that combine h-refinement and p-refinement.

- *Extension to more challenging problems* We wish to consider geometry variations, which are particularly relevant for applications. As explained in the body of the paper, this might be accomplished by resorting to the Piola transform, or by considering a two-field (velocity and pressure) formulation. Furthermore, we wish to apply our approach to transient problems: in order to face this task, we envision that time-dependent constraints should be considered, and also the time-averaged error indicator should be modified based on the particular quantity of interest we wish to predict. Finally, we wish to apply our approach to the reduction of LES/URANS flow simulations. This would substantially increase the range of engineering applications we can tackle with our method.

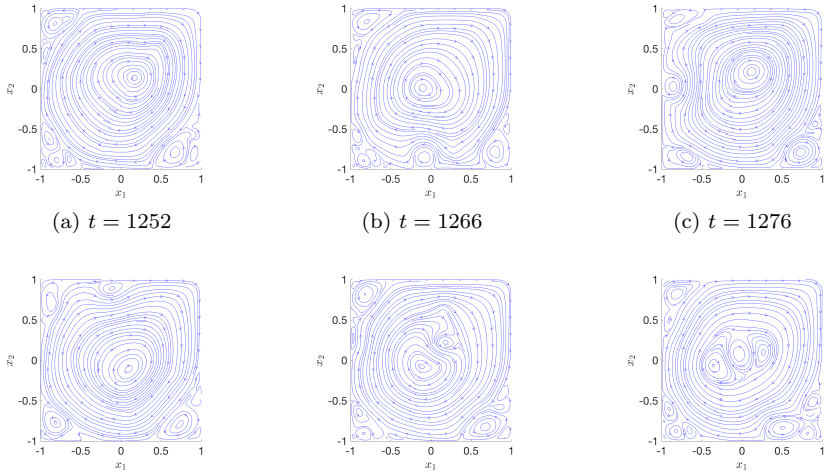
AppendixA. Analysis of the solution to the lid-driven cavity problem

Figure A.14 shows the velocity streamlines for three different times for Re = 15000, while Figure A.15 shows the velocity streamlines for several times for Re = 20000: for the latter value of the Reynolds number, we observe the presence of vortices along the edges and in the center of the cavity. We remark that the same behavior has been observed by Cazemier et al. in [65] (cf. Figure 3, page 1687). This rare behavior makes estimates of long-time averages particularly difficult, especially for the TKE. Figure A.16 shows the behavior of the turbulent kinetic energy TKE with time for three values of the Reynolds number. We observe that for sufficiently large values of the Reynolds number we have significant peaks in the TKE. These peaks correspond to eddies that are ejected into the core region and cross the cavity.



(a) $t = 501$ (b) $t = 600$ (c) $t = 700$

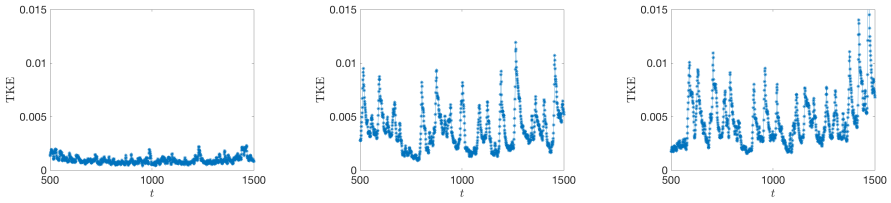
Figure A.14: A lid-driven cavity problem. Velocity streamlines for Re = 15000.



(a) $t = 1252$ (b) $t = 1266$ (c) $t = 1276$

(d) $t = 1286$ (e) $t = 1320$ (f) $t = 1344$

Figure A.15: A lid-driven cavity problem. Velocity streamlines for Re = 20000 for several time steps.



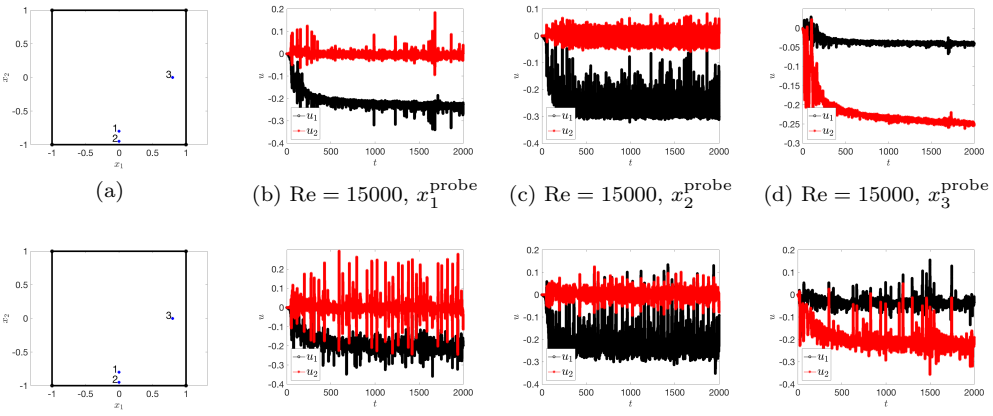
(a) Re = 15000 (b) Re = 20000 (c) Re = 25000

Figure A.16: A lid-driven cavity problem. Behavior of the turbulent kinetic energy TKE with time for three values of the Reynolds number.

Figure A.17 shows the behavior with time of the first and second components of the velocity field at three spatial locations for Re = 15000 and Re = 25000. We observe that for $t \gtrsim T_0 = 500$ the effects of the transient dynamics are negligible. Figure A.18 shows the autocorrelation factors associated with the time series $\{u_i(x_\ell^{\text{probe}}, t^j, \text{Re})\}_{j=J_0}^{J}$ for $i, \ell = 1, 2$ and for Re = 15000, 25000. We here define the autocorrelation factors for a time sequence $\{y^j\}_{j=0}^{J}$ as follows:

$$\rho_{\text{g}}(\tau = \kappa\Delta t) = \frac{\dfrac{1}{J-\kappa-J_0+1}\displaystyle\sum_{j=J_0}^{J-\kappa}\left(y^j - \langle y\rangle_{\text{g}}\right)\left(y^{j+\kappa} - \langle y\rangle_{\text{g}}\right)}{\left\langle \left(y - \langle y\rangle_{\text{g}}\right)^2\right\rangle_{\text{g}}}.$$

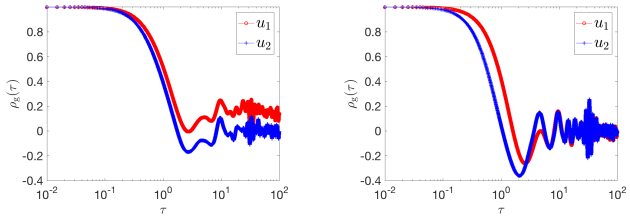
We observe that the autocorrelation factor decreases as τ increases, and is roughly 0.8 for $\tau = 1$, for all probes considered.



(a) (b) Re = 15000, x_1^{probe} (c) Re = 15000, x_2^{probe} (d) Re = 15000, x_3^{probe}

(e) (f) Re = 25000, x_1^{probe} (g) Re = 25000, x_2^{probe} (h) Re = 25000, x_3^{probe}

Figure A.17: A lid-driven cavity problem. Behavior of the velocity components at three spatial locations, for two values of Re ($x_1^{\text{probe}} = [0, -0.8]$, $x_2^{\text{probe}} = [0, -0.95]$, $x_3^{\text{probe}} = [0.8, 0]$).



(a) $\mathrm{Re} = 1510^3$, x_1^{probe} (b) $\mathrm{Re} = 1510^3$, x_2^{probe}

(c) $\mathrm{Re} = 2510^3$, x_1^{probe} (d) $\mathrm{Re} = 2510^3$, x_2^{probe}

Figure A.18: A lid-driven cavity problem. Behavior of the autocorrelation for the velocity components at two spatial locations, for two values of Re ($x_1^{\mathrm{probe}} = [0, -0.8]$, $x_2^{\mathrm{probe}} = [0, -0.95]$).

AppendixB. *A posteriori* assessment of the POD accuracy

As explained in the main body of the paper, POD relies on a snapshot set $\{\mathring{u}^k\}_{k=1}^K$ to generate a N-dimensional approximation space for the full trajectory $\{\mathring{u}^j\}_{j=J_0}^J$ in the limit $J \to \infty$. The snapshot set is associated with the sampling times $\{t_{\mathrm{s}}^k := T_0 + \Delta t_{\mathrm{s}} k\}_{k=1}^K$, where $T_0 = t_{\mathrm{g}}^{J_0}$, K is the cardinality of the snapshot set, and Δt_{s} is the sampling period.

The choices of Δt_{s} and K are a trade-off between (i) information content of the snapshot set, and (ii) computational resources. The snapshot set should be rich enough to accurately estimate the first N POD modes associated with the full trajectory $\{\mathring{u}^j\}_{j=J_0}^J$ in the limit $J \to \infty$. On the other hand, it is well-known that POD suffers from (i) the quadratic growth in K in computational complexity for computing the Gramian, and for computing the symmetric eigen-decomposition; and (ii) the memory requirements related to the storage of the snapshots, which scale linearly with K. Furthermore, by increasing Δt_{s} and K, we ultimately increase the number of time steps performed by the spectral element solver — which is given by $J = (T_0 + \Delta t_{\mathrm{s}} K) / \Delta t$.

In this Appendix, we propose a cross-validation (CV, see, e.g., [87] and [88, Chapter 7.10]) strategy to estimate the ℓ^2-averaged projection error associated with the POD reduced space over the full-trajectory,

$$\mathcal{E}(\{\mathring{u}^j\}_{j=J_0}^J, \mathcal{Z}^{\mathrm{u}}) = \frac{1}{J+1-J_0} \sum_{j=J_0}^{J} \|\mathring{u}^j - \Pi_{\mathcal{Z}^{\mathrm{u}}}^V \mathring{u}^j\|_V^2.$$

Estimates of this quantity might be employed to decide whether or not to acquire new snapshots and/or to increase the dimension N of the reduced space. On the other hand, evaluations of the autocorrelation factor introduced in AppendixA can be used to assess *a posteriori* the amount of redundancy in the snapshot set.

Since in our setting the snapshots are correlated in time (cf. AppendixA), we here rely on the h-block variant proposed in [89] (see also [90]). The approach relies on the assumption that the snapshot set is associated with a stationary process: under this assumption, the covariance matrix between $\mathring{u}^j$ and $\mathring{u}^{j+\kappa}$ is only a function of κ, and approaches 0 as $\kappa \to \infty$. The key idea of h-block CV is to reduce the training set by removing the h observations preceding and following the observation in the test set. In AppendixB.1, we adapt the computational procedure discussed in [89] to the particular learning task of interest; then, in AppendixB.2, we apply the procedure to the case $\mathrm{Re} = 15000$ to support our choice $K = 500$.

Before proceeding with the presentation of the methodology, we remark that, in the statistics literature, several authors have proposed validation techniques to assess the accuracy of POD (or, equivalently, PCA and Karhunen-Loéve) spaces. We refer to [91, Chapter 6] and to the references therein for a number of different proposals. We further recall the work by Chowdhary and Najm [92] that relies on a Bayesian framework to account for inaccuracies due to limited sample size. The approaches presented in [91, 92] aim at generating confidence (credible) regions for the estimate of the POD modes; on the other hand, we are here primarily interested in assessing the *out-of-sample* accuracy of the N-dimensional POD reduced space $\mathcal{Z}^{\mathrm{u}}$ for the full trajectory $\{\mathring{u}^j\}_{j=J_0}^J$. For completeness, we also recall that several authors ([93, 94]) have proposed and analyzed hierarchical POD approaches to reduce the size K of the snapshot set, without significantly compromising the accuracy of the POD space.

AppendixB.1. h-block Cross-Validation

Algorithm 4 summarizes the computational procedure for the estimation of the ℓ^2-averaged projection error $\mathcal{E}(\{\mathring{u}^j\}_{j=J_0}^J, \mathcal{Z}^{\mathrm{u}})$. We observe that for $h = 0$ the procedure reduces to Leave-One-Out-Cross-Validation (LOOCV). We further observe that the approach requires the assembling of the Gramian matrix $\mathbb{U}$, and then the solution to K dense eigenvalue problems of size $K - 2h - 1$: for the particular problem at hand, the computational cost associated with the procedure is negligible compared to the computational cost associated with the solution to the FOM. Finally, we emphasize that the procedure relies on the input parameter h. We here propose to choose h based on the analysis of the autocorrelation factor: recalling the results presented in AppendixA, we consider $h = \frac{4}{\Delta t_s} = 4$.

Algorithm 4 h-block Cross-Validation

$[\widehat{\mathcal{E}}] = \texttt{hblock-CV}\ (\{\mathring{u}^k\}_{k=1}^K, h, N)$
Inputs: $\{\mathring{u}^k\}_{k=1}^K$ = snapshot set, h = correlation parameter, N = size of the POD space.
Output: $\widehat{\mathcal{E}}$ = CV estimate of $\mathcal{E}(\{\mathring{u}\}_{j=J_0}^J, \mathcal{Z}^{\mathrm{u}})$.

1: **for** $k = 1, \ldots, K$ **do**
2: $\quad [\mathcal{Z}^{\mathrm{u},(\mathrm{k})} := \mathrm{span}\{\zeta_n^{(k)}\}_{n=1}^N] = \texttt{POD}_V\left(\{\mathring{u}^1, \ldots, \mathring{u}^{k-h-1}, \mathring{u}^{k+h+1}, \ldots, \mathring{u}^K\}_{k=1}^K, N\right)$
3: **end for**
4: Compute the CV estimate $\widehat{\mathcal{E}}$ as $\widehat{\mathcal{E}} = \frac{1}{K}\sum_{k=1}^K \|\mathring{u}^k - \Pi_{\mathcal{Z}^{\mathrm{u},(\mathrm{k})}}^V \mathring{u}^k\|_V^2$.

AppendixB.2. Results for $\mathrm{Re} = 15000$

Figure B.19 shows the behavior of $\widehat{\mathcal{E}}$ for different values of N for the snapshot set $\{\mathring{u}^k\}_{k=1}^K$ associated with $\{t_{\mathrm{s}}^k = 500+k\}_{k=1}^{K=500}$, and $\mathrm{Re} = 15000$. We compare results with the in-sample estimate

$$\mathcal{E}^{\mathrm{in}} = \frac{1}{K}\sum_{k=1}^{K} \|\mathring{u}^k - \Pi_{\mathcal{Z}^{\mathrm{u}}}^V \mathring{u}^k\|_V^2,$$

and the out-of-sample estimate

$$\mathcal{E}^{\mathrm{out}} = \frac{1}{K}\sum_{k=K+1}^{2K} \|\mathring{u}^k - \Pi_{\mathcal{Z}^{\mathrm{u}}}^V \mathring{u}^k\|_V^2,$$

where $\{\mathring{u}^k\}_{k=K+1}^{2K}$ are associated with the sampling times $\{t_{\mathrm{s}}^k = 500+k\}_{k=K+1}^{2K}$. For visualization purposes, we normalize $\widehat{\mathcal{E}}$, $\mathcal{E}^{\mathrm{in}}$, and $\mathcal{E}^{\mathrm{out}}$ by $\mathcal{E}^{\mathrm{in}}(N=1)$: for $N = 60$, $\widehat{\mathcal{E}} \approx 15\% \times \mathcal{E}^{\mathrm{in}}(N=1)$, $\mathcal{E}^{\mathrm{in}} \approx 10\% \times \mathcal{E}^{\mathrm{in}}(N=1)$, and $\mathcal{E}^{\mathrm{out}} \approx 17\% \times \mathcal{E}^{\mathrm{in}}(N=1)$. We observe that $\widehat{\mathcal{E}}$ is a more accurate estimate of $\mathcal{E}^{\mathrm{out}}$ compared to the in-sample estimate $\mathcal{E}^{\mathrm{in}}$.

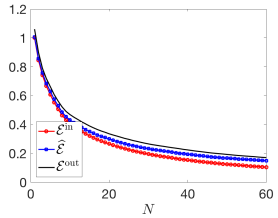


Figure B.19: A Cross-Validation procedure for the *a posteriori* assessment of the POD accuracy. Behavior of $\mathcal{E}^{\mathrm{in}}$, $\widehat{\mathcal{E}}$, and $\mathcal{E}^{\mathrm{out}}$ with N. All quantities are normalized by $\mathcal{E}^{\mathrm{in}}(N=1)$ ($K = 500$, $h = 2$, $\mathrm{Re} = 15000$).

AppendixC. On the definition of ROM stability

Based on the results of section 3, we could take a pragmatic view of the long-time ROM stability. Given the reduced space $\mathcal{Z}^{\mathrm{u}} \subset V_{\mathrm{div}}$, we define the

best-fit errors associated with mean flow and TKE:

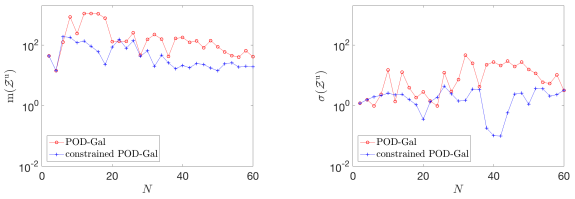
$$e_1^{\text{opt}} := \frac{\|\langle \mathring{u} \rangle_{\text{g}} - \Pi^V_{\mathcal{Z}^{\text{u}}} \langle \mathring{u} \rangle_{\text{g}}\|_V}{\|\langle u \rangle_{\text{g}}\|_V}, \quad e_2^{\text{opt}} := \frac{|\langle \text{TKE} \rangle_{\text{s}} - \langle \text{TKE}^{\text{opt}}(\cdot; \mathcal{Z}^{\text{u}}) \rangle_{\text{s}}|}{\langle \text{TKE} \rangle_{\text{s}}},$$

where $\Pi^V_{\mathcal{Z}^{\text{u}}}$ denotes the projection operator associated with the V inner product on the subspace $\mathcal{Z}^{\text{u}}$, and $\text{TKE}^{\text{opt}}(t; \mathcal{Z}^{\text{u}}) = \frac{1}{2} \int_\Omega \|\Pi^{L^2}_{\mathcal{Z}^{\text{u}}}(\mathring{u}(t) - \langle \mathring{u} \rangle_{\text{g}})\|_2^2 \, dx$. Then, we define the effective stability constants as the ratios between the optimal mean error and the actual error:

$$\text{m}(\mathcal{Z}^{\text{u}}) := \frac{\|\langle \mathring{u} \rangle_{\text{g}} - \langle \hat{u} \rangle_{\text{g}}\|_V}{\|\langle u \rangle_{\text{g}}\|_V e_1^{\text{opt}}}, \qquad \sigma(\mathcal{Z}^{\text{u}}) := \frac{|\langle \text{TKE} \rangle_{\text{s}} - \langle \widehat{\text{TKE}} \rangle_{\text{s}}|}{\langle \text{TKE} \rangle_{\text{s}} e_2^{\text{opt}}}. \tag{C.1}$$

The stability constants m and σ can be used to quantitatively measure the stability of the ROM. We observe that, in the limit $T \to \infty$, our definition of long-time stability for ROMs is independent of transient dynamics. We further observe that a ROM of dimension N is stable if and only if mean and variance of the time coefficients $\{a_n^j\}_j$ are correctly estimated for $n = 1, \ldots, N$. We finally remark that our definition of stability is close to the one proposed in [1]; however, while the definition in [1] is tailored to L^2 POD spaces and $R_g = \langle u \rangle_{\text{g}}$, our definition applies to any reduced space and to any choice of the lift.

Figure C.20 shows the behavior of $\text{m}(\mathcal{Z}^{\text{u}})$ and $\sigma(\mathcal{Z}^{\text{u}})$ defined in (C.1) for POD-Galerkin and constrained POD-Galerkin for $\text{Re} = 15000$: our constrained POD-Galerkin ROM is more stable — according to the definition given in this Appendix — than the standard Galerkin ROM.



(a) (b)

Figure C.20: The solution reproduction problem; behavior of $\text{m}(\mathcal{Z}^{\text{u}})$ and $\sigma(\mathcal{Z}^{\text{u}})$ (C.1) for POD-Galerkin and constrained POD-Galerkin ($\text{Re} = 15000$, $\epsilon = 0.01$).

AppendixD. Proof of (15)

In view of the proof, we introduce $\mathbb{M}_{n,n'} = (\zeta_n, \zeta_{n'})_{L^2(\Omega)}$, and $\widetilde{\mathbb{A}}(\mathbf{w}; \text{Re}) = \mathbb{A}(\mathbf{w}; \text{Re}) - \frac{1}{\Delta t}\mathbb{M}$, $\widetilde{\mathbf{F}}(\mathbf{w}; \text{Re}) = \mathbf{F}(\mathbf{w}; \text{Re}) - \frac{1}{\Delta t}\mathbb{M}\mathbf{w}$, where $\mathbb{A}$ and $\mathbf{F}$ are introduced in (10). We remark that $\widetilde{\mathbb{A}}, \widetilde{\mathbf{F}}$ do not depend on Δt and are smooth functions of $\mathbf{w}$ and Re.

We first restate (13) as a constrained optimization statement for the increment $\delta\mathbf{a} := \mathbf{a} - \mathbf{a}^j$

$$\min_{\delta\mathbf{a}\in\mathbb{R}^N} \| \left(\frac{1}{\Delta t}\mathbb{M} + \widetilde{\mathbb{A}}_j\right)\delta\mathbf{a} - \widetilde{\mathbf{F}}_j\|_2^2, \qquad \text{s.t. } \tilde{\alpha}_n^j \le \delta a_n \le \tilde{\beta}_n^j, \tag{D.1}$$

where $\tilde{\alpha}_n^j = \alpha_n - a_n^j$, $\tilde{\beta}_n^j = \beta_n - a_n^j$ and $\widetilde{\mathbb{A}}_j = \widetilde{\mathbb{A}}(\mathbf{a}^j;\mathrm{Re})$, $\widetilde{\mathbf{F}}_j = \widetilde{\mathbf{F}}(\mathbf{a}^j;\mathrm{Re})$. Since $\mathbf{a}^j$ is admissible for (13), we find that $\delta\mathbf{a} = \mathbf{0}$ is admissible for (D.1). Therefore, the solution to (D.1) $\delta\mathbf{a}^{j+1} = \mathbf{a}^{j+1} - \mathbf{a}^j$ satisfies:

$$\| \left(\frac{1}{\Delta t}\mathbb{M} + \widetilde{\mathbb{A}}_j\right)\delta\mathbf{a}^{j+1} - \widetilde{\mathbf{F}}_j\|_2^2 \le \|\widetilde{\mathbf{F}}_j\|_2^2.$$

The latter implies:

$$\| \left(\frac{1}{\Delta t}\mathbb{M} + \widetilde{\mathbb{A}}_j\right)\delta\mathbf{a}^{j+1}\|_2^2 \le 2\widetilde{\mathbf{F}}_j^T \left(\frac{1}{\Delta t}\mathbb{M} + \widetilde{\mathbb{A}}_j\right)\delta\mathbf{a}^{j+1} \le 2\|\widetilde{\mathbf{F}}_j\|_2^2 + \frac{1}{2}\| \left(\frac{1}{\Delta t}\mathbb{M} + \widetilde{\mathbb{A}}_j\right)\delta\mathbf{a}^{j+1}\|_2^2;$$

where in the last step we used the Young's inequality $2ab \le \frac{a^2}{\epsilon} + \epsilon b^2$ for $\epsilon = \frac{1}{2}$. Exploiting the inverse triangular inequality, we then find

$$2\|\widetilde{\mathbf{F}}_j\|_2 \ge \frac{1}{\Delta t}\|\mathbb{M}\delta\mathbf{a}^{j+1}\|_2 - \|\widetilde{\mathbb{A}}_j\delta\mathbf{a}^{j+1}\|_2 \ge \left(\frac{1}{\Delta t}\lambda_{\min}(\mathbb{M}) - \sigma_{\max}(\widetilde{\mathbb{A}}_j)\right)\|\delta\mathbf{a}^{j+1}\|_2,$$

where $\lambda_{\min}(\mathbb{M})$ is the minimum eigenvalue of the mass matrix $\mathbb{M}$, and $\sigma_{\max}(\widetilde{\mathbb{A}}_j)$ is the maximum singular value of $\widetilde{\mathbb{A}}_j$. In conclusion, we obtain:

$$\|\delta\mathbf{a}^{j+1}\|_2 \le C_j(\Delta t)\,\Delta t; \qquad C_j(\Delta t) = \frac{2\|\widetilde{\mathbf{F}}_j\|_2}{(\lambda_{\min}(\mathbb{M}) - \Delta t\sigma_{\max}(\widetilde{\mathbb{A}}_j))}.$$

Since $\{\mathbf{a}^j\}_j$ is bounded for all j and $\widetilde{\mathbb{A}}, \widetilde{\mathbf{F}}$ are smooth functions, there exist $C > 0$ and $\Delta t^\star > 0$ such that $C_j(\Delta t) \le C$ for all $\Delta t \le \Delta t^\star$. Thesis follows.

AppendixE. Robustness of the constrained formulation

In this Appendix, we present a number of numerical results that provide further insights about the constrained formulation proposed in this paper. In greater detail, we study the activation rate of the box constraints, the dependence of the solution to the choice of ϵ, and the behavior of $m_n^{\mathrm{FOM,u}}$ and $M_n^{\mathrm{FOM,u}}$ defined in (19) with respect to the Reynolds number Re.

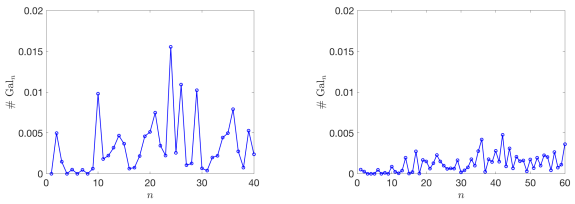
In Figure E.21, we study the behavior of the activation rate of each box constraint for two values of N for $\mathrm{Re} = 15000$, $\{t_{\mathrm{s}}^k = 500 + k\}_{k=1}^{K=500}$. In greater detail, we count how many times the n-th component of the solution to Galerkin satisfies the prescribed constraints:

$$\#\mathrm{Gal}_n := \frac{1}{J - J_0}\sum_{j=J_0+1}^{J} \mathbb{1}\left((\mathbf{a}_{\mathrm{Gal}}^j)_n \in [\alpha_n, \beta_n]\right), \quad n = 1, \dots, N.$$

We observe that the behavior with n of $\#\text{Gal}_n$ is irregular, and strongly depends on N. This suggests that selecting *a priori* the active constraints might be impractical. In Figure E.22, we study the behavior of the relative error in the mean flow prediction, the behavior of the TKE prediction $\langle\widehat{\text{TKE}}\rangle_{\text{s}}$ versus the truth estimate (7), and the behavior of the activation rate of the box constraints

$$\#\text{Gal} := \frac{1}{J-J_0}\sum_{j=J_0+1}^{J} \mathbb{1}\left((\mathbf{a}^j_{\text{Gal}})_n \in [\alpha_n, \beta_n],\, n=1,\dots,N\right),$$

with respect to ϵ, for two values of N. We observe that for $\epsilon \lesssim \bar{\epsilon} = 0.1$ results do not seem to depend on the value of ϵ. This provides evidence that the current approach is robust with respect to the choice of ϵ. We further observe that for all values of ϵ considered $\#\text{Gal}(N=40) \gtrsim 0.85$ and $\#\text{Gal}(N=60) \gtrsim 0.90$. Therefore, our constrained formulation corrects the original formulation only for $10-15\%$ time steps. For this reason, we envision that efficient implementations of the constrained ROM might be nearly as inexpensive as the Galerkin ROM. We further observe that #Gal increases as N increases: this can be explained by observing that the POD-Galerkin ROM becomes more and more accurate as N increases, and thus requires less corrections.



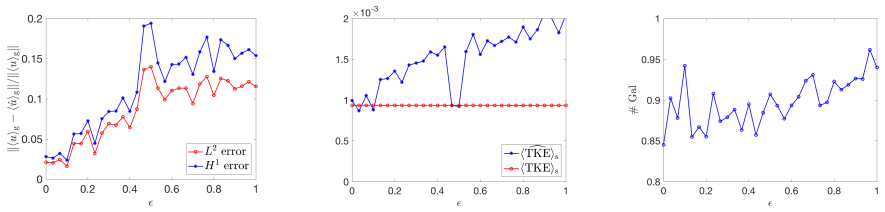
(a) $N = 40$ (b) $N = 60$

Figure E.21: The solution reproduction problem; activity of the box constraints for two values of N. ($\text{Re} = 15000$, $\epsilon = 0.01$).

Figure E.23 investigates the behavior of $m_n^{\text{FOM,u}}$ and $M_n^{\text{FOM,u}}$ defined in (19) with respect to the Reynolds number Re. For this test, we consider the POD space associated with $\text{Re} = 20000$ and the sampling times $\{t_{\text{s}}^k = 500+k\}_{k=1}^{K=1000}$, and we show results for $n = 1,\dots,12$. Results suggest that the sensitivity of $m_n^{\text{FOM,u}}$ and $M_n^{\text{FOM,u}}$ with Re are relatively modest if compared to $M_n^{\text{FOM,u}} - m_n^{\text{FOM,u}}$.

AppendixF. On the problem of *p*-refinement

We here illustrate the major issue associated with the combination of POD modes associated with different values of the parameter. We here simulate the application of the first two iterations of the POD-pGreedy algorithm as proposed in [49]. In more detail, we consider the following test.



(a) $N = 40$ (b) $N = 40$ (c) $N = 40$

(d) $N = 60$ (e) $N = 60$ (f) $N = 60$

Figure E.22: The solution reproduction problem; sensitivity with respect to ϵ for constrained POD-Galerkin for two values of N. Figures (a) and (d): behavior of the relative L^2 and H^1 errors. Figures (b) and (e): TKE prediction $\widehat{\langle \mathrm{TKE} \rangle_s}$ vs. truth estimate (7). Figures (c) and (f): percentage of pure Galerkin solves. (Re = 15000).

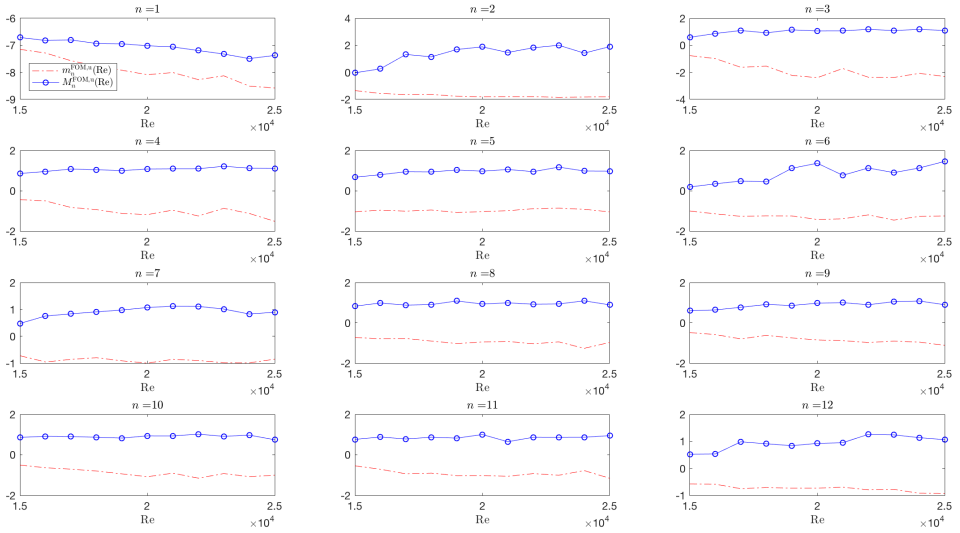


Figure E.23: The parametric problem; behavior of $m_n^{\mathrm{FOM,u}}(\mathrm{Re})$ and $M_n^{\mathrm{FOM,u}}(\mathrm{Re})$ with Re. The POD space is generated from the DNS data for Re = 20000.

1. Generate DNS data for Re = 20000, and use them to build the $N_1 = 60$-dimensional POD space.
2. Generate DNS data for Re = 15000, and build the $N_2 = 60$-dimensional POD space for the set of snapshots $\{(\mathbb{I} - \Pi^V_{\mathcal{Z}^{\mathrm{u}}})\mathring{u}^k(\mathrm{Re} = 15000)\}_{k=1}^K$.
3. Define $\mathcal{Z}^{\mathrm{u}} = \mathrm{span}\{\zeta_n\}_{n=1}^{N_1+N_2}$ where $\zeta_1, \ldots, \zeta_{N_1}$ are associated with Step 1 and $\zeta_{N_1+1}, \ldots, \zeta_{N_2}$ are associated with Step 2.
4. Perform a convergence study in N for Re = 15000 and Re = 20000 for both pure Galerkin and constrained Galerkin ($\epsilon = 0.01$).

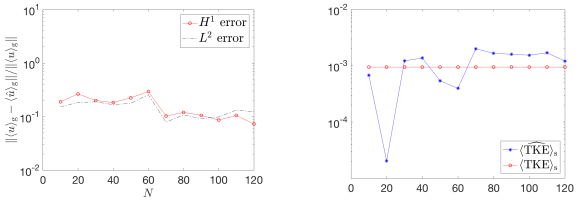
We consider here $T_0 = 500$ and $T = 1500$, $\{t_{\mathrm{s}}^k = 500 + k\}_{k=1}^{K=1000}$. We recall that for Re = 15000 (cf. section 3) we were able to obtain accurate ROMs for $N \gtrsim 40$ both in terms of mean flow prediction and TKE.

Figures F.24 and F.25 show the results of this test for the constrained-Galerkin ROM. Figures F.24(a) and (c) show the behavior of the relative error in mean flow prediction for the constrained formulation, for Re = 15000 and Re = 20000, respectively. We here compute lower and upper bounds $\{\alpha_n\}_n$ and $\{\beta_n\}_n$ using (14) with $\epsilon = 0.01$. Figures F.24(b) and (d) show the behavior of the mean TKE for the same values of the Reynolds number. Similarly, Figures F.25(a) and (b) show the behavior of the TKE in time for $N = 120$. Finally, Figures F.26 (a) and (b) show the behavior of the TKE in time for $N = 120$ for the unconstrained formulation. Results — especially for Re = 15000 — show the key issue of combining modes associated with different parameters. For $N_2 = 60$ (and $N_1 + N_2 = 120$), the error in mean flow prediction is roughly 10%, and we also significantly overestimate the mean and the peaks of the TKE. As expected, these issues are even more severe for the unconstrained formulation: the behavior with time of the TKE predicted by the unconstrained ROM is roughly the same for the two values of the Reynolds number considered.

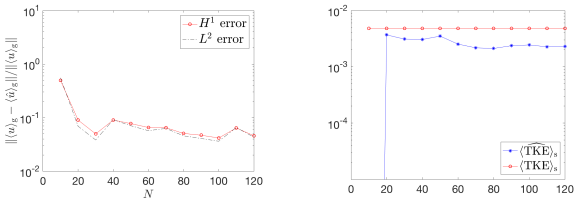
We offer a physical explanation for the poor performance of POD-pGreedy. As observed in AppendixA, for sufficiently large values of Re eddies are ejected into the core region of the cavity. This instability is observed for Re = 20000, but is not observed for Re = 15000. As a result, the ejection of the eddies into the core region of the cavity is well-represented by the POD space associated with Re = 20000, and then, by construction, by the final reduced space $\mathcal{Z}^{\mathrm{u}}$. The presence of modes associated with the core eddies makes the ROM more prone to show this instability even for values of the Reynolds number at which the full-order solution does not show it.

AppendixG. Offline/online computational decomposition for the residual indicator

We here describe the offline/online computational decomposition for the computation of the residual error indicator introduced in this paper. We omit the subscript ℓ associated with the partition of the parameter domain to simplify

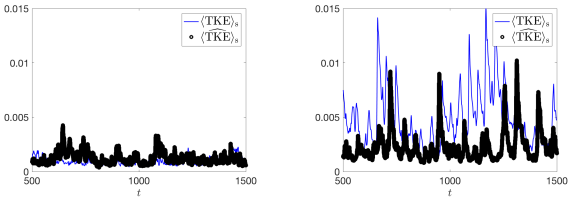


(a) Re = 15000 (b) Re = 15000



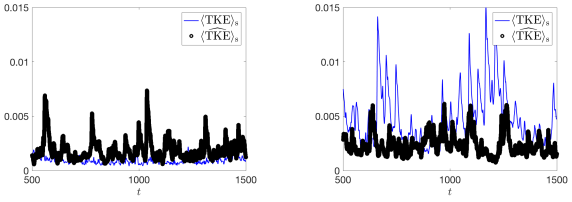
(c) Re = 20000 (d) Re = 20000

Figure F.24: The parametric problem; on the problems of p-refinement. Performance of constrained Galerkin ($\epsilon = 0.01$, $N_1 = 60$, $N_2 = 60$).



(a) Re = 15000, $N = 120$ (b) Re = 20000, $N = 120$

Figure F.25: The parametric problem; on the problems of p-refinement. Behavior of the TKE with time for constrained Galerkin ($\epsilon = 0.01$, $N_1 = 60$, $N_2 = 60$).



(a) Re = 15000, $N = 120$ (b) Re = 20000, $N = 120$

Figure F.26: The parametric problem; on the problems of p-refinement. Behavior of the TKE with time for unconstrained Galerkin ($N_1 = 60$, $N_2 = 60$).

notation. We first introduce the Riesz representers:

$$
\begin{aligned}
&(\xi_n^{\mathrm{m}}, v)_V = (\zeta_n, v)_{L^2(\Omega)}, \quad (\xi_n^{\mathrm{a}}, v)_V = (\zeta_n, v)_V, \quad (\xi_n^{\mathrm{cg}}, v)_V = c(R_g, \zeta_n, v),\\
&(\xi_{m,n}^{\mathrm{c}}, v)_V = c(\zeta_n, \zeta_m, v), \quad (\xi_n^{\mathrm{mg}}, v)_V = c(\zeta_n, R_g, v), \quad (\xi_1^{\mathrm{f}}, v)_V = (R_g, v)_V,\\
&(\xi_2^{\mathrm{f}}, v)_V = c(R_g, R_g, v),
\end{aligned}
\tag{G.1}
$$

for $n = 1, \ldots, N$ and for all $v \in V_{\mathrm{div}}$. Then, it is easy to verify that, if $w^j = \sum_{n=1}^N a_n^j \zeta_n$ for $j = 0, \ldots, J$, we can rewrite $\langle R \rangle$ as follows:

$$
\begin{aligned}
\langle R \rangle \left(\{w^j\}_{j=0}^J, v; \mathrm{Re} \right) = \quad & \Big(\sum_{n=1}^N \left(\xi_n^{\mathrm{m}} \left(\frac{a_n^J - a_n^{J_0}}{T - T_0} \right) + \xi_n^{\mathrm{a}} \left(\frac{1}{\mathrm{Re}} \bar{a}_n^+ \right) + \xi_n^{\mathrm{cg}} \bar{a}_n^+ \right.\\
& \left. + \sum_{m=1}^N \xi_{m,n}^{\mathrm{c}} \bar{c}_{m,n} + \xi_n^{\mathrm{mg}} \bar{a}_n^- \right), v \Big)_V\\
& + \frac{1}{\mathrm{Re}} (\xi_1^{\mathrm{f}}, v)_V + (\xi_2^{\mathrm{f}}, v)_V;
\end{aligned}
\tag{G.2a}
$$

where

$$
\bar{a}_n^+ = \frac{\Delta t}{T - T_0} \sum_{j=J_0+1}^{J} a_n^j, \quad \bar{a}_n^- = \frac{\Delta t}{T - T_0} \sum_{j=J_0}^{J-1} a_n^j, \quad \bar{c}_{m,n} = \frac{\Delta t}{T - T_0} \sum_{j=J_0}^{J-1} a_m^{j+1} a_n^j.
\tag{G.2b}
$$

Equation (G.2) can be rewritten as $\langle R \rangle \left(\{w^j\}_{j=0}^J, v; \mathrm{Re} \right) = \sum_{m=1}^M \Theta_i(\{\mathbf{a}^j\}_j; \mathrm{Re}) \tilde{\xi}_m$ where $M = N^2 + 3N + 2$ and

$$
\begin{aligned}
[\tilde{\xi}_1, \ldots, \tilde{\xi}_M] = [&\xi_1^{\mathrm{m}}, \ldots, \xi_N^{\mathrm{m}}, \xi_1^{\mathrm{a}}, \ldots, \xi_N^{\mathrm{a}}, \xi_1^{\mathrm{cg}}, \ldots, \xi_N^{\mathrm{cg}},\\
&\xi_{1,1}^{\mathrm{c}}, \ldots, \xi_{N,N}^{\mathrm{c}}, \xi_1^{\mathrm{mg}}, \ldots, \xi_N^{\mathrm{mg}}, \xi_1^{\mathrm{f}}, \xi_2^{\mathrm{f}}]
\end{aligned}
$$

$$
\begin{aligned}
[\Theta_1, \ldots, \Theta_M] = \Big[&\frac{a_1^J - a_1^{J_0}}{T - T_0}, \ldots, \frac{a_N^J - a_N^{J_0}}{T - T_0}, \frac{\bar{a}_1^+}{\mathrm{Re}}, \ldots, \frac{\bar{a}_N^+}{\mathrm{Re}}, \bar{a}_1^+, \ldots, \bar{a}_N^+,\\
&\bar{c}_{1,1}, \ldots, \bar{c}_{N,N}, \bar{a}_1^-, \ldots, \bar{a}_N^-, \frac{1}{\mathrm{Re}}, 1 \Big].
\end{aligned}
$$

Therefore, recalling the Riesz representation theorem, we find

$$
\Delta^{\mathrm{u}}(\{w^j\}_j; \mathrm{Re}) = \sqrt{\boldsymbol{\Theta}^T \Sigma^{\mathrm{off}} \boldsymbol{\Theta}}, \qquad \boldsymbol{\Theta} = \boldsymbol{\Theta}(\{\mathbf{a}^j\}_j; \mathrm{Re}),
\tag{G.3}
$$

where $\Sigma_{i,i'}^{\mathrm{off}} = (\tilde{\xi}_i, \tilde{\xi}_{i'})_V$. Equation (G.3) clarifies the offline/online decomposition: during the offline stage, we compute the Riesz representers (G.1) — this corresponds to the solution to M Stokes problems — and we assemble the matrix Σ^{off}; during the online stage, we compute the vector $\boldsymbol{\Theta}$ and we exploit (G.3) to compute the error estimator Δ^{u}. We observe that the online cost is $\mathcal{O}(N^2 J + N^4)$: provided that $N \ll J$ the cost associated with the computation of the error indicator is negligible compared to the cost of computing the solution to the time-integration scheme, $\mathcal{O}(N^3 J)$.

DNS + snapshot acquisition ($\Delta t = 2 \cdot 10^{-3}$, $K = 500$, Re $= 15000$)	4950.46 s (SC)
POD space construction: solution to (11) ($N = 60$, $K = 500$)	0.92 s
construction of Galerkin structures (10b) ($N = 60$)	21.25 s
POD-Gal: online evaluation of (10a) ($N = 60$)	34.90 s
constrained POD-Gal: online evaluation of (13) ($N = 60$, $\epsilon = 0.01$)	209.02 s
error estimation: offline computation of Σ^{off} (G.3) ($N = 80$)	935.26 s
error estimation: online computation of Δ^{u} (G.3) ($N = 80$)	12.77 s

Table H.1: Representative computational times. Evolutionary MOR algorithms are evaluated up to $T = 1000$, for $\Delta t = 5 \cdot 10^{-3}$ and Re $= 15000$. DNS calculations are performed on a supercomputer (SC), MOR calculations are performed on a Desktop computer.

AppendixH. Computational timings

We comment on the computational cost of the procedure. DNS simulations, which rely on the CFD software `nek5000`, are performed on a supercomputer at Argonne National Lab[5]: the cost of a single DNS simulation (elapsed time) for $\Delta t = 2 \cdot 10^{-3}$ and $T = 1000$, is 4950.46 s. Details concerning the numerical scheme and the discretization are provided in section 2. On the other hand, ROM computations, implemented in `Matlab2016a`, are performed using a Desktop computer (RAM 16GB 800 MHz, Processor Intel Xeon 3.60GHz, 8 cores).

Table H.1 lists the computational cost of the offline and online procedures: all Matlab routines are run in serial (i.e., no parallelism). We observe that the speed-up of the constrained POD-Galerkin approach is equal to[6] $4950.46/209.02 = 23.68$ for $N = 60$, $T = 1000$, and $\Delta t = 5 \cdot 10^{-3}$. Note here we equate cost to wall-clock time, however clearly in actual practice a supercomputer second is much more costly in resources than a desktop second. Figure H.27 shows the online costs for POD-Galerkin and constrained POD-Galerkin. We observe that the behavior of the computational cost of the constrained formulation is non-monotonic in N: the reason is that as N increases the number of pure Galerkin solves increases.

Acknowledgements

The authors thank Prof. Paul Fischer (UIUC), and Dr. Elia Merzari (Argonne National Lab) for their support with the software `Nek5000`. We also acknowledge the Argonne National Laboratory Mathematics and Computer Science Division for the use of their computational resources.

[5] ANL MCS Compute001: Mainboard Supermicro X10QBi; RAM 1.5TB DDR3 1600 MHz (96 x 16GB Modules); CPUs 4 x Intel Xeon E7-4820 v2 (8 cores per CPU, i.e. 32 cores total).

[6] In our calculation the online speed-up does not take into account the computation of the error indicator.

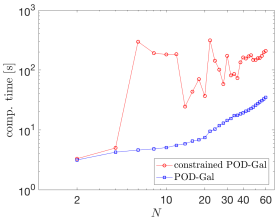


Figure H.27: Online costs for POD-Galerkin and constrained POD-Galerkin (Re = 15000, $T = 1000$, $\epsilon = 0.01$, $\Delta t = 5 \cdot 10^{-3}$).

References

[1] M. Balajewicz, E. Dowell, Stabilization of projection-based reduced order models of the navier–stokes, Nonlinear Dynamics 70 (2) (2012) 1619–1632.

[2] Z. Wang, I. Akhtar, J. Borggaard, T. Iliescu, Proper orthogonal decomposition closure models for turbulent flows: a numerical comparison, Computer Methods in Applied Mechanics and Engineering 237 (2012) 10–26.

[3] L. Cordier, B. R. Noack, G. Tissot, G. Lehnasch, J. Delville, M. Balajewicz, G. Daviller, R. K. Niven, Identification strategies for model-based control, Experiments in fluids 54 (8) (2013) 1–21.

[4] A. Deane, I. Kevrekidis, G. E. Karniadakis, S. Orszag, Low-dimensional models for complex geometry flows: Application to grooved channels and circular cylinders, Physics of Fluids A: Fluid Dynamics (1989-1993) 3 (10) (1991) 2337–2354.

[5] X. Ma, G.-S. Karamanos, G. Karniadakis, Dynamics and low-dimensionality of a turbulent near wake, Journal of Fluid Mechanics 410 (2000) 29–65.

[6] X. Ma, G. E. Karniadakis, A low-dimensional model for simulating three-dimensional cylinder flow, Journal of Fluid Mechanics 458 (2002) 181–190.

[7] B. Galletti, C. Bruneau, L. Zannetti, A. Iollo, Low-order modelling of laminar flow regimes past a confined square cylinder, Journal of Fluid Mechanics 503 (2004) 161–170.

[8] G. Berkooz, P. Holmes, J. L. Lumley, The proper orthogonal decomposition in the analysis of turbulent flows, Annual review of fluid mechanics 25 (1) (1993) 539–575.

[9] K. Kunisch, S. Volkwein, Galerkin proper orthogonal decomposition methods for a general equation in fluid dynamics, SIAM Journal on Numerical analysis 40 (2) (2002) 492–515.

[10] M. Kahlbacher, S. Volkwein, Galerkin proper orthogonal decomposition methods for parameter dependent elliptic systems, Discussiones Mathematicae, Differential Inclusions, Control and Optimization 27 (1) (2007) 95–117.

[11] K. Carlberg, M. Barone, H. Antil, Galerkin v. least-squares petrov–galerkin projection in nonlinear model reduction, Journal of Computational Physics 330 (2017) 693–734.

[12] D. Rempfer, On low-dimensional galerkin models for fluid flow, Theoretical and Computational Fluid Dynamics 14 (2) (2000) 75–88.

[13] B. R. Noack, From snapshots to modal expansions–bridging low residuals and pure frequencies, Journal of Fluid Mechanics 802 (2016) 1–4.

[14] P. Moin, K. Mahesh, Direct numerical simulation: a tool in turbulence research, Annual review of fluid mechanics 30 (1) (1998) 539–578.

[15] B. R. Noack, P. Papas, P. A. Monkewitz, The need for a pressure-term representation in empirical galerkin models of incompressible shear flows, Journal of Fluid Mechanics 523 (2005) 339–365.

[16] N. Aubry, P. Holmes, J. L. Lumley, E. Stone, The dynamics of coherent structures in the wall region of a turbulent boundary layer, Journal of Fluid Mechanics 192 (1988) 115–173.

[17] P. Sagaut, Large eddy simulation for incompressible flows: an introduction, Springer Science & Business Media, 2006.

[18] M. Couplet, P. Sagaut, C. Basdevant, Intermodal energy transfers in a proper orthogonal decomposition–galerkin representation of a turbulent separated flow, Journal of Fluid Mechanics 491 (2003) 275–284.

[19] B. R. Noack, K. Afanasiev, M. Morzynski, G. Tadmor, F. Thiele, A hierarchy of low-dimensional models for the transient and post-transient cylinder wake, Journal of Fluid Mechanics 497 (2003) 335–363.

[20] J. Östh, B. R. Noack, S. Krajnović, D. Barros, 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, Journal of Fluid Mechanics 747 (2014) 518–544.

[21] S. Sirisup, G. Karniadakis, A spectral viscosity method for correcting the long-term behavior of pod models, Journal of Computational Physics 194 (1) (2004) 92–116.

[22] Y. Maday, A. Manzoni, A. Quarteroni, An online intrinsic stabilization strategy for the reduced basis approximation of parametrized advection-dominated problems, Comptes Rendus Mathematique 354 (12) (2016) 1188–1194.

[23] E. Tadmor, Convergence of spectral methods for nonlinear conservation laws, SIAM Journal on Numerical Analysis 26 (1) (1989) 30–44.

[24] S. Giere, T. Iliescu, J. Volker, D. Wells, Supg reduced order models for convection-dominated convection–diffusion–reaction equations, Computer Methods in Applied Mechanics and Engineering 289 (2015) 454–474.

[25] X. Xie, D. Wells, Z. Wang, T. Iliescu, Approximate deconvolution reduced order modeling, Computer Methods in Applied Mechanics and Engineering 313 (2017) 512–534.

[26] O. San, T. Iliescu, Proper orthogonal decomposition closure models for fluid flows: Burgers equation, International Journal of Numerical Analysis & Modeling - Series B 5 (3) (2014) 217–237.

[27] M. Marion, R. Temam, Nonlinear galerkin methods: the finite elements case, Numerische Mathematik 57 (1) (1990) 205–226.

[28] A. Debussche, T. Dubois, R. Temam, The nonlinear galerkin method: a multiscale method applied to the simulation of homogeneous turbulent flows, Theoretical and Computational Fluid Dynamics 7 (4) (1995) 279–315.

[29] A. Steindl, H. Troger, Methods for dimension reduction and their application in nonlinear dynamics, International Journal of Solids and Structures 38 (10) (2001) 2131–2147.

[30] M. Bergmann, C.-H. Bruneau, A. Iollo, Enablers for robust pod models, Journal of Computational Physics 228 (2) (2009) 516–538.

[31] M. Bergmann, Optimisation aérodynamique par réduction de modèle pod et contrôle optimal: application au sillage laminaire d'un cylindre circulaire, Ph.D. thesis, Vandoeuvre-les-Nancy, INPL (2004).

[32] Y. Maday, A. T. Patera, D. V. Rovas, A blackbox reduced-basis output bound method for noncoercive linear problems, in Studies in Mathematics and its Applications, D. Cioranescu and J. L. Lions, eds., Elsevier Science B. V (2001) 533–569.

[33] K. Carlberg, C. Bou-Mosleh, C. Farhat, Efficient non-linear model reduction via a least-squares petrov–galerkin projection and compressive tensor approximations, International Journal for Numerical Methods in Engineering 86 (2) (2011) 155–181.

[34] K. Carlberg, C. Farhat, J. Cortial, D. Amsallem, The gnat method for nonlinear model reduction: effective implementation and application to computational fluid dynamics and turbulent flows, Journal of Computational Physics 242 (2013) 623–647.

[35] A. Tallet, C. Allery, C. Leblond, E. Liberge, A minimum residual projection to build coupled velocity–pressure pod–rom for incompressible navier–stokes equations, Communications in Nonlinear Science and Numerical Simulation 22 (1) (2015) 909–932.

[36] A. Iollo, S. Lanteri, J. A. Désidéri, Stability properties of pod–galerkin approximations for the compressible navier–stokes equations, Theoretical and Computational Fluid Dynamics 13 (6) (2000) 377–396.

[37] N. Aubry, R. Guyonnet, R. Lima, Spatio-temporal symmetries and bifurcations via bi-orthogonal decompositions, Journal of Nonlinear Science 2 (2) (1992) 183–215.

[38] M. F. Barone, I. Kalashnikova, D. J. Segalman, H. K. Thornquist, Stable galerkin reduced order models for linearized compressible flow, Journal of Computational Physics 228 (6) (2009) 1932–1946.

[39] C. W. Rowley, T. Colonius, R. M. Murray, Model reduction for compressible flows using pod and galerkin projection, Physica D: Nonlinear Phenomena 189 (1) (2004) 115–129.

[40] M. Couplet, C. Basdevant, P. Sagaut, Calibrated reduced-order pod-galerkin system for fluid flow modelling, Journal of Computational Physics 207 (1) (2005) 192–220.

[41] T. Duriez, S. L. Brunton, B. R. Noack, Machine Learning Control-Taming Nonlinear Dynamics and Turbulence, Springer, 2017.

[42] X. Xie, M. Mohebujjaman, L. Rebholz, T. Iliescu, Data-driven filtered reduced order modeling of fluid flows, arXiv preprint arXiv:1709.04362.

[43] P. J. Schmid, Dynamic mode decomposition of numerical and experimental data, Journal of fluid mechanics 656 (2010) 5–28.

[44] C. W. Rowley, I. Mezić, S. Bagheri, P. Schlatter, D. S. Henningson, Spectral analysis of nonlinear flows, Journal of fluid mechanics 641 (2009) 115–127.

[45] P. J. Schmid, L. Li, M. P. Juniper, O. Pust, Applications of the dynamic mode decomposition, Theoretical and Computational Fluid Dynamics 25 (1-4) (2011) 249–259.

[46] A. Alla, J. N. Kutz, Nonlinear model order reduction via dynamic mode decomposition, arXiv preprint arXiv:1602.05080.

[47] J. H. Curry, J. R. Herring, J. Loncaric, S. A. Orszag, Order and disorder in two-and three-dimensional benard convection, Journal of Fluid Mechanics 147 (1984) 1–38.

[48] G. Stabile, S. Hijazi, A. Mola, S. Lorenzi, G. Rozza, Advances in reduced order modelling for cfd: vortex shedding around a circular cylinder using a pod-galerkin method, arXiv preprint arXiv:1701.03424.

[49] B. Haasdonk, M. Ohlberger, Reduced basis method for finite volume approximations of parametrized linear evolution equations, ESAIM: Mathematical Modelling and Numerical Analysis-Modélisation Mathématique et Analyse Numérique 42 (2) (2008) 277–302.

[50] J. L. Eftang, D. J. Knezevic, A. T. Patera, An hp certified reduced basis method for parametrized parabolic partial differential equations, Mathematical and Computer Modelling of Dynamical Systems 17 (4) (2011) 395–422.

[51] T. Taddei, S. Perotto, A. Quarteroni, Reduced basis techniques for nonlinear conservation laws, ESAIM: Mathematical Modelling and Numerical Analysis 49 (3) (2015) 787–814.

[52] N. Cagniart, Y. Maday, B. Stamm, Model Order Reduction for Problems with large Convection Effects, working paper or preprint (Oct. 2016).
URL http://hal.upmc.fr/hal-01395571

[53] M. Ohlberger, S. Rave, Nonlinear reduced basis approximation of parameterized evolution equations via the method of freezing, Comptes Rendus Mathematique 351 (23) (2013) 901–906.

[54] K. Urban, A. T. Patera, A new error bound for reduced basis approximation of parabolic partial differential equations, Comptes Rendus Mathematique 350 (3) (2012) 203–207.

[55] M. Yano, A space-time petrov–galerkin certified reduced basis method: Application to the boussinesq equations, SIAM Journal on Scientific Computing 36 (1) (2014) A232–A266.

[56] T. Colin, A. Iollo, D. Lombardi, O. Saut, System identification in tumor growth modeling using semi-empirical eigenfunctions, Mathematical Models and Methods in Applied Sciences 22 (06) (2012) 1250003.

[57] J. Argaud, B. Bouriquet, H. Gong, Y. Maday, O. Mula, Stabilization of (g) eim in presence of measurement noise: application to nuclear reactor physics, arXiv preprint arXiv:1611.02219.

[58] J. Loiseau, S. L. Brunton, Constrained sparse galerkin regression, arXiv preprint arXiv:1611.03271.

[59] S. R. Reddy, B. A. Freno, P. G. Cizmas, S. Gokaltun, D. McDaniel, G. S. Dulikravich, Constrained reduced-order models based on proper orthogonal decomposition, Computer Methods in Applied Mechanics and Engineering 321 (2017) 18–34.

[60] B. Haasdonk, Convergence rates of the pod–greedy method, ESAIM: Mathematical Modelling and Numerical Analysis 47 (03) (2013) 859–873.

[61] H. V. Ly, H. T. Tran, Modeling and control of physical processes using proper orthogonal decomposition, Mathematical and computer modelling 33 (1-3) (2001) 223–236.

[62] T. Franz, R. Zimmermann, S. Görtz, N. Karcher, Interpolation-based reduced-order modelling for steady transonic flows via manifold learning, International Journal of Computational Fluid Dynamics 28 (3-4) (2014) 106–121.

[63] P. Gallinari, Y. Maday, M. Sangnier, O. Schwander, T. Taddei, Reduced basis acquisition by a learning process for rapid on-line approximation of solution to pdes: Laminar flow past a backstep, Archives of Computational Methods in Engineering 25 (1) (2018) 131–141.

[64] M. Balajewicz, A new approach to model order reduction of the navier-stokes equations, Ph.D. thesis, Duke University (2012).

[65] W. Cazemier, R. Verstappen, A. Veldman, Proper orthogonal decomposition and low-dimensional models for driven cavity flows, Physics of Fluids (1994-present) 10 (7) (1998) 1685–1699.

[66] P. Shankar, M. Deshpande, Fluid mechanics in the driven cavity, Annual Review of Fluid Mechanics 32 (1) (2000) 93–136.

[67] F. Terragni, E. Valero, J. M. Vega, Local pod plus galerkin projection in the unsteady lid-driven cavity problem, SIAM Journal on Scientific Computing 33 (6) (2011) 3538–3561.

[68] S. Lorenzi, A. Cammi, L. Luzzi, G. Rozza, Pod-galerkin method for finite volume approximation of navier–stokes and rans equations, Computer Methods in Applied Mechanics and Engineering 311 (2016) 151–179.

[69] S. B. Pope, Turbulent flows, Cambridge University Press, 2000.

[70] A. T. Patera, A spectral element method for fluid dynamics: laminar flow in a channel expansion, Journal of Computational Physics 54 (3) (1984) 468–488.

[71] J. L. PF Fischer, S. Kerkemeier, nek5000 Web page, http://nek5000.mcs.anl.gov (2008).

[72] C. Bernardi, Y. Maday, Spectral methods, Handbook of numerical analysis 5 (1997) 209–485.

[73] C. Canuto, M. Y. Hussaini, A. Quarteroni, T. A. Zang, Spectral methods in fluid dynamics, Springer Science & Business Media, 2012.

[74] G. Karniadakis, S. Sherwin, Spectral/hp element methods for computational fluid dynamics, Oxford University Press, 2013.

[75] M. O. Deville, P. F. Fischer, E. H. Mund, High-order methods for incompressible fluid flow, Vol. 9, Cambridge University Press, 2002.

[76] A. Sohankar, C. Norberg, L. Davidson, Low-reynolds-number flow around a square cylinder at incidence: study of blockage, onset of vortex shedding and outlet boundary condition, International journal for numerical methods in fluids 26 (1) (1998) 39–56.

[77] A. E. Løvgren, Y. Maday, E. M. Rønquist, A reduced basis element method for the steady stokes problem, ESAIM: Mathematical Modelling and Numerical Analysis 40 (3) (2006) 529–552.

[78] G. Rozza, K. Veroy, On the stability of the reduced basis method for stokes equations in parametrized domains, Computer methods in applied mechanics and engineering 196 (7) (2007) 1244–1260.

[79] G. Rozza, D. P. Huynh, A. Manzoni, Reduced basis approximation and a posteriori error estimation for stokes flows in parametrized geometries: roles of the inf-sup stability constants, Numerische Mathematik 125 (1) (2013) 115–152.

[80] T. Lassila, A. Manzoni, A. Quarteroni, G. Rozza, Model order reduction in fluid dynamics: challenges and perspectives, in: Reduced Order Methods for modeling and computational reduction, Springer, 2014, pp. 235–273.

[81] F. Ballarin, G. Rozza, Y. Maday, Reduced-order semi-implicit schemes for fluid-structure interaction problems, in: Model Reduction of Parametrized Systems, Springer, 2017, pp. 149–167.

[82] L. Sirovich, Turbulence and the dynamics of coherent structures. part i: Coherent structures, Quarterly of applied mathematics 45 (3) (1987) 561–571.

[83] S. Volkwein, Model reduction using proper orthogonal decomposition, Lecture Notes, Institute of Mathematics and Scientific Computing, University of Graz. see http://www. uni-graz. at/imawww/volkwein/POD. pdf.

[84] MATLAB, version 9.0 (R2016a), The MathWorks Inc., Natick, Massachusetts, 2016.

[85] J. Nocedal, S. Wright, Numerical optimization, Springer Science & Business Media, 2006.

[86] Y. M. R Chakir, A two-grid finite-element/reduced basis scheme for the approximation of the solution of parametric dependent p.d.e., Comptes Rendus Mathmatique 1513 (7) (2009) 343–462.

[87] R. Kohavi, et al., A study of cross-validation and bootstrap for accuracy estimation and model selection, in: Ijcai, Vol. 14, 1995, pp. 1137–1145.

[88] T. Hastie, R. Tibshirani, J. Friedman, The elements of statistical learning, Vol. 2, Springer, 2009.

[89] P. Burman, E. Chow, D. Nolan, A cross-validatory method for dependent data, Biometrika 81 (2) (1994) 351–358.

[90] J. Racine, Consistent cross-validatory model-selection for dependent data: hv-block cross-validation, Journal of econometrics 99 (1) (2000) 39–61.

[91] I. Jolliffe, Principal component analysis, Wiley Online Library, 2002.

[92] K. Chowdhary, H. N. Najm, Bayesian estimation of karhunen–loève expansions; a random subspace approach, Journal of Computational Physics 319 (2016) 280–293.

[93] A. Paul-Dubois-Taine, D. Amsallem, An adaptive and efficient greedy procedure for the optimal training of parametric reduced-order models, International Journal for Numerical Methods in Engineering 102 (5) (2015) 1262–1292.

[94] C. Himpe, T. Leibner, S. Rave, Hierarchical approximate proper orthogonal decomposition, arXiv preprint arXiv:160705210v3.