Numer. Math. (2008) 110:49–81
DOI 10.1007/s00211-008-0155-9

Numerische Mathematik

Error estimates of proper orthogonal decomposition eigenvectors and Galerkin projection for a general dynamical system arising in fluid models

Pedro Galán del Sastre · Rodolfo Bermejo

Received: 9 July 2007 / Published online: 10 June 2008

Abstract We study estimates for proper orthogonal decomposition eigenvectors and eigenvalues as well as error estimates between the exact solution of a 2D Navier–Stokes model and the numerical approach when the proper orthogonal decomposition method is considered. These estimates are also extended when bifurcation diagram are calculated using the so called p-POD or SPOD methods with a new cut-off criterion to minimize noisy modes produced by the p-POD method.

Mathematics Subject Classification (2000) 65M60 · 65L60 · 65N30 · 76M10 · 76M25 · 37N10 · 65F15

1 Introduction

This work is devoted to the theoretical study of the proper orthogonal decomposition method, hereafter the POD method, as well as some new improvements of the method when bifurcation diagrams are required. When one is interested in applying the POD method to study a dynamical system, one uses the *snapshots method* [19]. In short, the strategy to compute bifurcation diagrams with the POD technique consists of calculating a finite dimensional basis to generate a subspace on which to obtain a reduced dynamical system via Galerkin projection of the infinite dimensional dynamical

P. Galán del Sastre (✉)
Departamento de Matemática Aplicada al Urbanismo, a la Edificación y al Medio Ambiente, E.T.S.A.M., Universidad Politécnica de Madrid, Avda. Juan de Herrera 4, 28040 Madrid, Spain
e-mail: pedro.galan@upm.es

R. Bermejo
Departamento de Matemática Aplicada, E.T.S.I.I., Universidad Politécnica de Madrid, C/ José Gutiérrez Abascal 2, 28006 Madrid, Spain
e-mail: rbermejo@etsii.upm.es

system. In doing so, it is of interest to analyze the rate of convergence of the solution of the reduced dynamical system to the solution of the infinite dimensional system.

Some authors have already provided error analysis for POD method. For instance, in [14,15] an error analysis for some time discretization algorithms of the Galerkin projection is developed for parabolic problems as well as for some general models arising in fluid dynamics. Nevertheless, in these works, the authors always suppose the snapshots come from the exact solution sampled at different time instants, instead of a numerical solution as it is usually done in applications. Moreover, the hypotheses are improved to calculate the correlation matrix in a stronger space such as H_0^1 and, therefore, to obtain better estimates.

In this work, we improve such estimates assuming that the POD basis is computed by the snapshot method proposed by Sirovich [19], sampling in a discrete set of time instants a numerical solution to the problem. Furthermore, we perform an error analysis for the Galerkin projection over the space spanned by the POD basis.

We recall the snapshot method. First, one supposes $u(t_i) \in C\left(\bar{\Omega}\right)$ for the time partition $\{t_i\}_{1\leq i\leq n}$, with $t_{i+1} - t_i = \Delta t$, and $u(t_i)$ defined as a discrete approximation in $L^2(\Omega)$. Usually, $u(t_i)$ describes a periodic attractor in (t_1, t_n). The idea of the POD method is to find an orthonormal set in $L^2(\Omega)$, $\left\{\phi_j\right\}_{1\leq j\leq m}$, such that it minimizes

$$\sum_{i=1}^{n} \left\| u(t_i) - \sum_{j=1}^{m} \left(u(t_i), \phi_j\right)_{L^2(\Omega)} \phi_j \right\|^2_{L^2(\Omega)} \tag{1}$$

for any $m \leq n$. In doing so, one has to compute the symmetric and positive semidefinite correlation matrix $K = \left(\left(k_{ij}\right)\right)_{1\leq i,j\leq n}$,

$$k_{ij} = \left(u(t_i), u\left(t_j\right)\right)_{L^2(\Omega)}, \tag{2}$$

and calculate its eigenvectors $\{a_i\}_{1\leq i\leq n} \subset \mathbb{R}^n$. We remark that other method to compute the POD basis is the Sigular Value Decomposition (SVD). Since K is symmetric, one can assume that $\left(a_i, a_j\right)_{\mathbb{R}^n} = \delta_{ij}$, and using this fact, the orthonormal basis is computed as

$$\hat{\phi}_j(x) = \sum_{i=1}^{n} u(t_i, x)\, a_{j,i}, \tag{3}$$

for all $x \in \bar{\Omega}$ with

$$\phi_j = \frac{\hat{\phi}_j}{\left\|\hat{\phi}_j\right\|_{L^2(\Omega)}}, \tag{4}$$

where $a_j = \left(a_{j,1}, a_{j,2}, \ldots, a_{j,n}\right)$. Usually, $\hat{\phi}_j(x)$ are only calculated for the nodes, $x = x_k$, of the space discretization. For further details on the properties of the basis

and the optimality condition (1), see for instance [11] or [3]. In this work, we develop an abstract formulation to obtain both a better understanding of the method and its theoretical results.

On the other hand, a great effort has been made to use the POD method to compute bifurcation diagrams. A strategy employed by some authors consists of computing a POD basis for some specific value of the bifurcation parameter of the dynamical system and then calculating the bifurcation diagram of the Galerkin projection, see for example [1,8]. Results show that when the values of the control parameter are not close to the one used to compute the POD basis, the dynamics may differ significantly. To improve this strategy other authors compute a global basis using attractors corresponding to different values of the bifurcation parameter. The p-POD [6] and SPOD [13] were introduced and applied when the first Hopf bifurcation computation was required.

In this work we introduce an improvement to the cut-off criterion of the SPOD method that minimizes noisy modes arising when the SPOD method is computed as in [13] or [6]. In doing so, we also prove that the method can be applied to compute branches of periodic solutions in the bifurcation diagram. Furthermore, theoretical estimates on the accuracy of the method are also provided.

The paper is organized as follows. In Sect. 2 the general dynamical system is introduced in abstract formulation. The idea is to introduce a general framework to accommodate the formulation of Navier–Stokes type problems. In Sect. 3 some estimates about compact, positive and self-adjoint operators are derived. These estimates will be used later on for the error analysis of the POD method which is fully described in an abstract framework in Sect. 4. Relations between this abstract formulation and the method of snapshots described in the literature are also shown in this section. The bifurcation diagram computed via SPOD is explained in Sect. 5. Section 6 is devoted to the presentation of the Galerkin projection when using POD and SPOD method is described and the main result of error analysis. Finally in Sect. 7 two examples arising from fluid dynamics are analyzed.

2 General dynamical system

We set up in this section our functional analytic setting that is the same as the one used in [20] and [15]. We denote by $(\cdot,\cdot)_X$ and $\|\cdot\|_X$ the inner product and norm respectively in a Hilbert space X, and by $\langle\cdot,\cdot\rangle_{X'\times X}$ the duality pairing between X and its dual X'. Let H and V be two separable and Hilbert spaces such that $V \subset H$ with dense and compact inclusion. Let $a_v : V \times V \longrightarrow \mathbb{R}$ be a bilinear, symmetric, continuous and coercive form that defines the inner product in V, i.e.,

$$(\phi,\varphi)_V = a_v(\phi,\varphi) \text{ for all } \phi,\varphi \in V.$$

Due to the continuous inclusion of V in H, there exists a constant $c_v > 0$ such that

$$\|\varphi\|_H \leq c_v \|\varphi\|_V \text{ for all } \varphi \in V. \tag{5}$$

The bilinear form a_ν defines a linear operator $A_\nu : V \longrightarrow V'$ such that for any $\varphi \in V$

$$\langle A_\nu \varphi, \phi \rangle_{V' \times V} = a_\nu (\varphi, \phi) \text{ for all } \phi, \varphi \in V.$$

Moreover, operator A_ν defines an isomorphism between V and its dual V'. Usually, A_ν is also considered as an unbounded operator in H with $D(A_\nu) = \{\varphi \in V : A_\nu \varphi \in H\}$. By identifying H and its dual H' it follows that

$$D(A) \hookrightarrow V \hookrightarrow H = H' \hookrightarrow V',$$

each embedding being continuous and dense when $D(A)$ is endowed with the graph norm of A.

Next, we introduce a continuous operator $Q : V \longrightarrow V'$ such that

$$\begin{aligned}
&Q\varphi \in H && \text{for all } \varphi \in D(A_\nu), \\
&\|Q\varphi\| \leq c_Q \|\varphi\|_V^{1-\gamma_1} \|\varphi\|_H^{\gamma_1} && \text{for all } \varphi \in D(A_\nu), \\
&\left|\langle Q\varphi, \varphi \rangle_{V' \times V}\right| \leq c_Q \|\varphi\|_V^{1+\gamma_2} \|\varphi\|_H^{1-\gamma_2} && \text{for all } \varphi \in V, \\
&a_\nu(\varphi, \varphi) + \langle Q\varphi, \varphi \rangle_{V' \times V} \geq \zeta \|\varphi\|_V^2 && \text{for all } \varphi \in V,
\end{aligned} \tag{6}$$

with $c_Q, \zeta > 0$ and $\gamma_1, \gamma_2 \in [0, 1)$.

To consider non-linear dynamical systems (as for instance, the Navier–Stokes equations), we also introduce the bilinear continuous operator $B : V \times V \longrightarrow V'$ such that for all $\varphi, \phi, \psi \in V$ and $\chi \in D(A)$,

$$\begin{aligned}
&B(\varphi, \phi) \in H, && \text{if } \varphi, \phi \in D(A), \\
&\langle B(\varphi, \phi), \phi \rangle_{V' \times V} = 0, \\
&\left|\langle B(\varphi, \phi), \psi \rangle_{V' \times V}\right| \leq c_B \|\varphi\|_H^{\gamma_3} \|\varphi\|_V^{1-\gamma_3} \|\phi\|_V \|\psi\|_V^{\gamma_3} \|\psi\|_H^{1-\gamma_3}, \\
&\|B(\varphi, \chi)\|_H + \|B(\chi, \varphi)\|_H \leq c_B \|\varphi\|_V \|\chi\|_V^{1-\gamma_4} \|A_\nu \chi\|_H^{\gamma_4}, \\
&\|B(\varphi, \chi)\|_H \leq c_B \|\varphi\|_H^{\gamma_5} \|\varphi\|_V^{1-\gamma_5} \|\chi\|_V^{1-\gamma_5} \|A_\nu \chi\|_H^{\gamma_5},
\end{aligned} \tag{7}$$

where $\gamma_3, \gamma_4, \gamma_5 \in [0, 1)$. We use the shorthand notation $B\varphi$ to denote $B(\varphi, \varphi)$ for any $\varphi \in V$.

Consider the dynamical system such that for all $v \in V$

$$\begin{cases} \left(\dfrac{\partial u}{\partial t}(t), v\right)_H + a_\nu(u(t), v) + \langle Bu(t) + Qu(t), v \rangle_{V' \times V} = (f, v)_H, \\ u(0) = u_0 \in H, \end{cases} \tag{8}$$

with $f \in H$. The next theorem states the existence and uniqueness of the solution for (8) as well as some regularity results.

Theorem 1 *Suppose that* (6), (7) *and the assumptions for* f *and* u_0 *hold. Then, there exists a unique solution of* (8) *such that*

$$u \in C([0, T]; H) \cap L^2(0, T; V) \cap H^1(0, T; H).$$

Furthermore, u is analytic in t with values in $D(A_\nu)$ for $t>0$ and the mapping $u_0 \to u(t)$ is continuous from H into $D(A_\nu)$, $\forall t > 0$. Finally, if $u_0 \in V$, then

$$u \in C([0,T];V) \cap L^2(0,T;D(A_\nu)) \cap H^1(0,T;H).$$

Proof See [20]. □

3 Preliminaries

We first introduce some results related to estimates of compact, positive and self-adjoint operators defined on the space H of Sect. 2.

Lemma 1 *Let $S, S_h : H \longrightarrow H$ be two compact, positive and self-adjoint operators with orthonormal eigenvector basis $\{\varphi_j\}_{j\in J}, \{\varphi_{hj}\}_{j\in J} \subset H$ respectively, i.e.,*

$$S\varphi_j = \mu_j\varphi_j \text{ and } S_h\varphi_j = \mu_{hj}\varphi_{hj}$$

for all $j \in J = \{1, 2, \ldots, \dim H\}$. Let $\delta > 0$ be such that $\|S - S_h\| < \delta$. Assume that: (i) *$\mu_j \geq \mu_{j+1}$ and $\mu_{hj} \geq \mu_{h,j+1}$ for all $j \in J$; and* (ii) *for some $r_1 \in J$, $\mu_j = \mu_1$ for all $j \leq r_1$ and $\mu_{r_1} > \mu_{r_1+1}$. Then, for $1 \leq k \leq r_1$*

$$1 - \sum_{i=1}^{r_1} (\varphi_{hk}, \varphi_i)_H^2 < \left(\frac{\delta + |\mu_1 - \mu_{hk}|}{\mu_1 - \mu_{r_1+1}}\right)^2. \tag{9}$$

Furthermore, there exists $\hat{\varphi}_1 \in H$ such that $S\hat{\varphi}_1 = \mu_1\hat{\varphi}_1$ and

$$\left\|\hat{\varphi}_1 - \varphi_{h1}\right\|_H^2 < 2\left(1 - \sqrt{1 - \left(\frac{2\delta}{\mu_1 - \mu_{r_1+1}}\right)^2}\right). \tag{10}$$

Moreover, if $\hat{H}_1 = span\{\hat{\varphi}_1\}$, $H_{h1} = span\{\varphi_{h1}\}$, and P_Y denotes the orthonormal projection over any subspace $Y \subset H$, then

$$\left\|P_{\hat{H}_1} - P_{H_{h1}}\right\|^2 < 8\left(1 - \sqrt{1 - \left(\frac{2\delta}{\mu_1 - \mu_{r_1+1}}\right)^2}\right) \tag{11}$$

and

$$|\mu_1 - \mu_{h1}| < \delta. \tag{12}$$

Proof A simple remark can be used to prove (12),

$$\delta > \|S - S_h\| \geq |\|S\| - \|S_h\|| = |\mu_1 - \mu_{h1}|.$$

Since $\{\varphi_j\}_{j\in J}$ is an orthonormal basis in H, then

$$\varphi_{hk} = \sum_{i\geq 1} (\varphi_{hk}, \varphi_i)_H \, \varphi_i .$$

Define $\psi \in H$ as

$$\psi = \varphi_{hk} - \sum_{i=1}^{r_1} (\varphi_{hk}, \varphi_i)_H \, \varphi_i = \sum_{i>r_1} (\varphi_{hk}, \varphi_i)_H \, \varphi_i ,$$

or equivalently

$$\sum_{i=1}^{r_1} (\varphi_{hk}, \varphi_i)_H \, \varphi_i = \varphi_{hk} - \psi .$$

Then,

$$\begin{aligned} S_h\varphi_{hk} - S\varphi_{hk} &= \mu_{hk}\varphi_{hk} - \sum_{i\geq 1} \mu_i \, (\varphi_{hk}, \varphi_i)_H \, \varphi_i \\ &= \mu_{hk}\varphi_{hk} - \mu_1 \sum_{i=1}^{r_1} (\varphi_{hk}, \varphi_i)_H \, \varphi_i - \sum_{i>r_1} \mu_i \, (\varphi_{hk}, \varphi_i)_H \, \varphi_i \\ &= \mu_{hk}\varphi_{hk} - \mu_1 \, (\varphi_{hk} - \psi) - S\psi = (\mu_{hk} - \mu_1) \, \varphi_{hk} + \mu_1\psi - S\psi . \end{aligned}$$

So that,

$$\begin{aligned} \delta &> \|S_h\varphi_{hk} - S\varphi_{hk}\|_H \geq \mu_1 \, \|\psi\|_H - \|S\psi\|_H - |\mu_{hk} - \mu_1| \\ &\geq \mu_1 \, \|\psi\|_H - \mu_{r_1+1} \, \|\psi\|_H - |\mu_{hk} - \mu_1| , \end{aligned}$$

i.e.,

$$\|\psi\|_H < \frac{\delta + |\mu_{hk} - \mu_1|}{\mu_1 - \mu_{r_1+1}} .$$

Using the fact that

$$\|\psi\|_H^2 = \sum_{i>r_1} (\varphi_{hk}, \varphi_i)_H^2 = 1 - \sum_{i=1}^{r_1} (\varphi_{hk}, \varphi_i)_H^2 ,$$

(9) follows.

Now, we proceed with (10) and (11). First, we point out that estimate (9) may be thought as an estimate of the projection of φ_{hk} onto the subspace spanned by $\{\varphi_j\}_{j=1}^{r_1}$; so that, to choose properly $\hat{\varphi}_1$ one can try with the projection of φ_{h1} onto this subspace. Since for any $\hat{\varphi}_1 \in \text{span}\,\{\varphi_j\}_{j=1}^{r_1}$, $S\hat{\varphi}_1 = \mu_1\hat{\varphi}_1$, then if one sets

$$\hat{\varphi}_1 = \alpha \sum_{i=1}^{r_1} (\varphi_{h1}, \varphi_i)_H \, \varphi_i ,$$

with α such that $\left\|\hat{\varphi}_1\right\|_H = 1$, i.e.,

$$\frac{1}{\alpha^2} = \sum_{i=1}^{r_1} (\varphi_{h1}, \varphi_i)_H^2 ,$$

it follows that

$$\left\|\hat{\varphi}_1 - \varphi_{h1}\right\|_H^2 = \left\|\hat{\varphi}_1\right\|_H^2 + \|\varphi_{h1}\|_H^2 - 2\left(\hat{\varphi}_1, \varphi_{h1}\right)_H = 2\left(1 - \left(\hat{\varphi}_1, \varphi_{h1}\right)_H\right).$$

Note that using (9) we have that

$$\left(\hat{\varphi}_1, \varphi_{h1}\right)_H = \left(\sum_{i=1}^{r_1} (\varphi_{h1}, \varphi_i)_H^2\right)^{1/2} > \left(1 - \left(\frac{\delta + |\mu_{h1} - \mu_1|}{\mu_1 - \mu_{r_1+1}}\right)^2\right)^{1/2}; \quad (13)$$

so that, taking into account (12)

$$\left(\hat{\varphi}_1, \varphi_{h1}\right)_H > \left(1 - \left(\frac{\delta + |\mu_{h1} - \mu_1|}{\mu_1 - \mu_{r_1+1}}\right)^2\right)^{1/2} > \left(1 - \left(\frac{2\delta}{\mu_1 - \mu_{r_1+1}}\right)^2\right)^{1/2}$$

and (10) follows. Finally, let $x \in H$, then

$$\begin{aligned}\left\| P_{\hat{H}_1} x - P_{H_{h1}} x \right\|_H &= \left\| \left(x, \hat{\varphi}_1\right)_H \hat{\varphi}_1 - (x, \varphi_{h1})_H \varphi_{h1} \right\|_H \\ &\le \left\| \left(x, \hat{\varphi}_1\right)_H \hat{\varphi}_1 - \left(x, \hat{\varphi}_1\right)_H \varphi_{h1} \right\|_H \\ &\quad + \left\| \left(x, \hat{\varphi}_1\right)_H \varphi_{h1} - (x, \varphi_{h1})_H \varphi_{h1} \right\|_H \\ &\le 2 \|x\|_H \left\| \hat{\varphi}_1 - \varphi_{h1} \right\| .\end{aligned}$$

From this estimate it follows (11). □

Note that Lemma 1 gives an estimate about the projection of the first r_1 eigenvectors of S_h onto the subspace spanned by the first eigenvectors associated to the first eigenvalue of S. Such a bound is useful to obtain the next results to estimate the differences among all the eigenvectors and eigenvalues of S and S_h as well as the subspaces spanned by them. We further go on with another lemma about the eigenvectors associated to μ_1.

Lemma 2 *Assume that the hypotheses of Lemma 1 hold. Then, for any $j \in J$ such that $1 \le j \le r_1$, there exists $\hat{\varphi}_j \in span\{\varphi_i\}_{i=1}^m$ satisfying*

$$\left\|\hat{\varphi}_j - \varphi_{hj}\right\|_H < \hat{\delta}_j, \quad (14)$$

with $\left(\hat{\varphi}_i, \hat{\varphi}_j\right)_H = \delta_{ij}$. *Furthermore,*

$$\left\| P_{\hat{H}_j} - P_{H_{hj}} \right\| < 2 \sum_{i=1}^{j} \hat{\delta}_i \tag{15}$$

$$\left\| \mu_j - \mu_{hj} \right\| < \delta_j \tag{16}$$

where $\delta_1 = \delta$, $\hat{\delta}_j^2 = 2\left(1 - \sqrt{1 - \left(\frac{2\delta_j}{\mu_1 - \mu_{r_1+1}}\right)^2}\right)$, $\delta_j = \delta + 2\mu_1 \sum_{i=1}^{j-1} \hat{\delta}_i$ *and the subspaces* $\hat{H}_j$, H_{hj} *are defined by*

$$\hat{H}_j = span\left\{\hat{\varphi}_i\right\}_{i=1}^{j} \text{ and } H_{hj} = span\left\{\varphi_{hi}\right\}_{i=1}^{j} \text{ for } 1 \leq j \leq r_1.$$

Proof By Lemma 1, (14), (15) and (16) are proven for $j = 1$. We proceed by induction for $j > 1$. First, one can define the operator S_j and S_{hj}, for $j > 1$, as follows:

$$S_j = \left(I - P_{\hat{H}_{j-1}}\right) S \text{ and } S_{hj} = \left(I - P_{H_{h,j-1}}\right) S_h.$$

Then,

$$\begin{aligned}
\left\| S_j - S_{hj} \right\| &= \left\| \left(I - P_{\hat{H}_{j-1}}\right) S - \left(I - P_{H_{h,j-1}}\right) S_h \right\| \\
&\leq \left\| \left(I - P_{\hat{H}_{j-1}}\right) S - \left(I - P_{H_{h,j-1}}\right) S \right\| \\
&\quad + \left\| \left(I - P_{H_{h,j-1}}\right) S - \left(I - P_{H_{h,j-1}}\right) S_h \right\| \\
&\leq \left\| P_{\hat{H}_{j-1}} - P_{H_{h,j-1}} \right\| \|S\| + \left\| I - P_{H_{h,j-1}} \right\| \|S - S_h\| \\
&< 2\mu_1 \sum_{i=1}^{j-1} \hat{\delta}_i + \delta = \delta_j.
\end{aligned}$$

Now, we can point out that S and S_j have the same eigenvectors and so is for S_h and S_{hj}, with the same eigenvalues associated for all $i \geq j$ and eigenvalue zero for $i < j$. Then, using Lemma 1 with S_j and S_{hj}, one can prove the existence of $\hat{\varphi}_j$ such that $S_j \hat{\varphi}_j = \mu_j \hat{\varphi}_j$ and

$$\left\| \hat{\varphi}_j - \varphi_{hj} \right\|_H < \hat{\delta}_j.$$

To prove that $\left\{\hat{\varphi}_i\right\}_{i=1}^{j}$ is an orthonormal set, we note that $\left(I - P_{\hat{H}_{j-1}}\right) S\hat{\varphi}_j = S_j \hat{\varphi}_j = \mu_j \hat{\varphi}_j$ and $S_j \varphi \perp \hat{H}_{j-1}$ for all $\varphi \in H$, then

$$0 = \left(S_j \hat{\varphi}_j, \hat{\varphi}_i\right)_H = \left(\mu_j \hat{\varphi}_j, \hat{\varphi}_i\right)_H$$

for all $i < j$. Moreover, Lemma 1 ensures that $\left|\mu_1 - \mu_{hj}\right| < \delta_j$ and

$$\left\| P_{span\{\hat{\varphi}_j\}} - P_{span\{\varphi_{hj}\}} \right\| < 2\hat{\delta}_j.$$

Finally, to prove (15) we set

$$\begin{aligned}\left\| P_{\hat{H}_j} - P_{H_{hj}} \right\| &= \left\| \left(P_{\hat{H}_{j-1}} + P_{span\{\hat{\varphi}_j\}} \right) - \left(P_{H_{h,j-1}} + P_{span\{\varphi_{hj}\}} \right) \right\| \\ &\leq \left\| P_{\hat{H}_{j-1}} - P_{H_{h,j-1}} \right\| + \left\| P_{span\{\hat{\varphi}_j\}} - P_{span\{\varphi_{hj}\}} \right\| \\ &\leq 2\sum_{i=1}^{j-1} \hat{\delta}_i + 2\hat{\delta}_j = 2\sum_{i=1}^{j} \hat{\delta}_i.\end{aligned}$$

□

Now, we can establish a general result for operators under the hypotheses of Lemma 1.

Theorem 2 *Assume that the hypotheses of Lemma 1 hold and suppose that* $\mu_j = \mu_{r_k+1}$ *for all* $r_k < j \leq r_{k+1}$. *Then, there is a global orthonormal basis* $\left\{\hat{\varphi}_j\right\}_{j\in J}$ *of* H *such that* $S\hat{\varphi}_j = \mu_j\hat{\varphi}_j$ *for all* $j \in J$ *and*:

$$\begin{aligned}\left|\mu_j - \mu_{hj}\right| &< \delta_j, \\ \left\|\hat{\varphi}_j - \varphi_{hj}\right\|_H &< \hat{\delta}_j, \\ \left\| P_{\hat{H}_j} - P_{H_{hj}} \right\| &< 2\sum_{i=1}^{j} \hat{\delta}_i;\end{aligned}$$

where, for $r_k < j \leq r_{k+1}$ *and* $r_0 = 0$, $\delta_1 = \delta$, $\delta_j = \delta + 2\mu_1\sum_{i=1}^{j-1} \hat{\delta}_i$ *for all* $j > 1$ *and* $\hat{\delta}_j^2 = 2\left(1 - \sqrt{1 - \left(\frac{2\delta_j}{\mu_{r_k+1} - \mu_{r_{k+1}+1}}\right)^2}\right)$.

Proof We proceed by induction in k. For $k = 1$, i.e., for $1 \leq j \leq r_1$ the result follows from Lemma 2. For $k > 1$, using the same argument as in Lemma 2 one can define the operators $S_{r_k+1} = \left(I - P_{\hat{H}_{r_k}}\right) S$ and $S_{h,r_k+1} = \left(I - P_{H_{h,r_k}}\right) S_h$. Then,

$$\left\| S_{r_k+1} - S_{h,r_k+1} \right\| \leq \left\| P_{\hat{H}_{r_k}} - P_{H_{h,r_k}} \right\| \|S\| + \|S - S_h\| < \delta + 2\mu_1 \sum_{i=1}^{r_k} \hat{\delta}_i.$$

Thus, applying Lemma 2 to S_{r_k+1} and S_{h,r_k+1} and noting that their first r_k eigenvalues are zero, while for $i \geq r_k + 1$ their eigenvalues are the same, the theorem follows. □

Remark 1 We point out that $\hat{\delta}_j \sim \dfrac{2\delta_j}{\mu_{r_k+1} - \mu_{r_{k+1}+1}}$ when $\delta_j \to 0$, then one can prove by induction that

$$\delta_j \sim \delta \left(1 + \frac{2\mu_1}{\mu_{r_k+1} - \mu_{r_{k+1}+1}}\right)^{j-r_k} \prod_{i=0}^{k-1} \left(1 + \frac{2\mu_1}{\mu_{r_i+1} - \mu_{r_{i+1}}}\right)^{r_{i+1}-r_i}.$$

4 The POD basis

The POD method is a powerful technique which is used to obtain low dimensional models to study different problems. The combination of the POD technique with the Galerkin projection procedure provides a method to generate low dimensional models of dynamical systems that have a very large or even infinite dimensional phase space. And it is in this context that we consider the POD method. Usually, one can consider that the solution of the dynamical system of interest is a function $u \in L^2((0,T) \times \Omega)^M$, with $\Omega \subset \mathbb{R}^d$, $d \in \{1,2,3\}$, a bounded open domain and $M \geq 1$. Thus, $u \in L^2\left(0,T; L^2(\Omega)^M\right)$. In studying the properties of u, one could take advantage if one knows that $u(t) \in Y_m$ for all $t \in (0,T)$, with $Y_m \subset L^2(\Omega)^M$ a finite dimensional linear subspace. If this is the case, then one can take an orthonormal basis in Y_m, say $\{\phi_i\}_{i=1}^m$, thus

$$u(t,x) = \sum_{i=1}^{m} a_i(t)\,\phi_i(x)$$

for $t \in (0,T)$ and $x \in \Omega$, where $a_i(t) = (u(t), \phi_i)_{L^2(\Omega)^M}$. The key idea of the POD method resides in finding such subspace Y_m.

When the solution u does not lie on a finite dimensional subspace of $L^2(\Omega)^M$, then

$$u(t,x) = \sum_{i=1}^{\infty} a_i(t)\,\phi_i(x)$$

with $\{\phi_i\}_{i\in\mathbb{N}}$ being a basis for $L^2(\Omega)^M$. Nevertheless, one must note that

$$\|u\|^2_{L^2\left(0,T;L^2(\Omega)^M\right)} = \sum_{i=1}^{\infty} \|a_i\|^2_{L^2(0,T)},$$

then, if $\sum_{i>m} \|a_i\|^2_{L^2(0,T)} \ll 1$, one obtains

$$\left\| u(t,x) - \sum_{i=1}^{m} a_i(t)\,\phi_i(x) \right\|_{L^2\left(0,T;L^2(\Omega)^M\right)} \ll 1,$$

i.e., the projection of u onto the finite dimension subspace spanned by $\{\phi_i\}_{i=1}^m$ is a good approximation of u. Therefore, it is very important to choose properly the orthonormal basis $\{\phi_i\}_{i\in\mathbb{N}}$. Taking into account all this information, one could look for a basis such that the norm of the coefficients a_i were as large as possible, that is, one could try to maximize $\left\|(u,\phi_i)_{L^2(\Omega)^M}\right\|_{L^2(0,T)}$ for $1 \le i \le m$. This is what the POD method does. To give an abstract formulation of the POD method we assume that $u \in L^2\left(I; H^N\right)$, with $N \in \mathbb{N}$, H a separable Hilbert space and I an open bounded domain in $\mathbb{R}^n$. Below, we will be more specific about I, H, N and u as solution of a dynamical system, but for the time being, it is sufficient for expository purposes to consider $H = L^2(\Omega)^M$. To compute an orthonormal basis of H that maximizes $\left\|(u,\phi_i)_H\right\|_{L^2(I)^N}$ we shall look for elements $\phi \in H$ such that

$$\frac{\left\|(u,\phi)_H\right\|_{L^2(I)^N}}{\|\phi\|_H} = \max_{\varphi\in H} \frac{\left\|(u,\varphi)_H\right\|_{L^2(I)^N}}{\|\varphi\|_H},$$

or equivalently,

$$\left\|(u,\phi)_H\right\|^2_{L^2(I)^N} = \max_{\varphi\in H,\|\varphi\|_H=1} \left\|(u,\varphi)_H\right\|^2_{L^2(I)^N}, \tag{17}$$

where one must note now that $(u,\phi)_H \in L^2(I)^N$. Hence, introducing the functional $J_\lambda : H \longrightarrow \mathbb{R}$,

$$J_\lambda(\phi) = \left\|(u,\phi)_H\right\|^2_{L^2(I)^N} - \lambda\|\phi\|^2_H,$$

we have that the solution $\phi \in H$ in (17) is a critical point of J_λ for some $\lambda \in \mathbb{R}$. Noting that λ is a Lagrange multiplier we characterize the optimal solution to (17) via a first-order necessary optimality conditions. The following proposition ensures the existence of such points for J_λ.

Proposition 1 *Let $J = \{1, 2, \ldots, \dim H\} \subset \mathbb{N}$. Then, there exist $\left\{\left(\lambda_j, \phi_j\right)\right\}_{j\in J} \subset \mathbb{R}^+ \times H$ such that ϕ_j is a critical point of J_{λ_j} for all $j \in J$. Furthermore, $\left\{\phi_j\right\}_{j\in J}$ is an orthonormal basis in H.*

Proof If ϕ is a critical point of J_λ it follows that

$$\lim_{h\to 0} \frac{J_\lambda(\phi + h\varphi) - J_\lambda(\phi)}{h} = 2\left(\left((u,\phi)_H, (u,\varphi)_H\right)_{L^2(I)^N} - \lambda(\phi,\varphi)_H\right) = 0. \tag{18}$$

Now, we define the operator $T : H \longrightarrow H$ as

$$(T\phi,\varphi)_H = \left((u,\phi)_H, (u,\varphi)_H\right)_{L^2(I)^N} \text{ for all } \varphi \in H. \tag{19}$$

It is straightforward to prove that T is a linear, continuous compact positive and self-adjoint operator; so that, there exists a sequence of eigenvalues and associated

eigenvectors of T, $\{(\lambda_j, \phi_j)\}_{j\in J} \subset \mathbb{R}^+ \times H$, $\lambda_j \leq \lambda_{j+1}$ such that $\{\phi_j\}_{j\in J} \subset H$ is an orthonormal basis of H. The proof is finished because (18) implies that ϕ_j is a critical point of J_{λ_j}. □

Some remarks are now in order:

Remark 2 If $\{\phi_j\}_{j\in J}$ is the orthonormal basis in Proposition 1, then

$$u = \sum_{j\in J} (u, \phi_j)_H \phi_j = \sum_{j\in J} a_j \phi_j,$$

with $a_j = (u, \phi_j)_H \in L^2(I)^N$. $\{a_j\}_{j\in J} \subset L^2(I)^N$ is an orthogonal set in $L^2(I)^N$ verifying $\|a_j\|^2_{L^2(I)^N} = \lambda_j$ because

$$\lambda_i \delta_{ij} = (T\phi_i, \phi_j)_H = \left((u, \phi_i)_H, (u, \phi_j)_H\right)_{L^2(I)^N} = (a_i, a_j)_{L^2(I)^N}.$$

Furthermore,

$$\|u\|^2_{L^2(I;H^N)} = \sum_{j\in J} \|a_j\|^2_{L^2(I)^N} \|\phi_j\|^2_H = \sum_{j\in J} \lambda_j$$

and

$$\left\| P_{H_m} u - u \right\|^2_{L^2(I;H^N)} = \sum_{j>m} \lambda_j.$$

Thus, in order to take a "good" finite dimensional subspace, one can define the so called "energy function", depending on the dimension m, as

$$E(m) = \frac{\left\| P_{H_m} u - u \right\|^2_{L^2(I;H^N)}}{\|u\|^2_{L^2(I;H^N)}} = \frac{\sum_{j>m} \lambda_j}{\sum_{j\in J} \lambda_j}, \tag{20}$$

and choose m such that $E(m) < \varepsilon$ for some $\varepsilon > 0$. A typical value of the tolerance parameter ε is 10^{-4} (which is equivalent to 99.99% of the L^2-norm of u).

Remark 3 We note that

$$im\, T \subset \overline{span\,\{u\}}^H \subset H,$$

because if one takes $\varphi \in \left(\overline{span\,\{u\}}^H\right)^{\perp}$, i.e., $(u, \varphi)_H = 0 \in L^2(I)^N$, then,

$$(T\phi, \varphi)_H = \left((u, \phi)_H, (u, \varphi)_H\right)_{L^2(I)^N} = 0$$

for all $\phi \in H$.

The next proposition shows that the orthonormal basis $\{\phi_j\}_{j\in J}$ is optimal in the sense defined above.

Proposition 2 *Let $\{\psi_j\}_{j\in J}$ be an orthonormal basis in H and let $\{\phi_j\}_{j\in J}$ be the basis of Proposition 1. Then, if $m \in J$,*

$$\left\| P_{H_m} u - u \right\|_{L^2(I;H^N)} \leq \left\| P_{\hat{H}_m} u - u \right\|_{L^2(I;H^N)},$$

where $H_m = span\{\phi_j\}_{j=1}^m$, $\hat{H}_m = span\{\psi_j\}_{j=1}^m$ and $P_{H_m}, P_{\hat{H}_m}$ are the orthogonal projection over H_m and $\hat{H}_m$ respectively.

Proof The idea of the proof is in [11] with $H = L^2(\Omega)$ and $I = (0, T)$. □

We turn our attention to the spaces H, I and the integer N considering that $u \in L^2\left(0, T; L^2(\Omega)^M\right)$, this means that $H = L^2(\Omega)^M$, $I = (0, T)$ and $N = 1$. Then, as we show above, the POD method provides a basis $\{\phi_j\}_{j\in J}$ for $L^2(\Omega)^M$, thus

$$u = \sum_{j\in J} a_j \phi_j,$$

with $a_j = \left(u, \phi_j\right)_{L^2(\Omega)^M}$ being an orthogonal set in $L^2(0, T)$ due to Remark 2. On the other hand, if one knows the orthogonal set $\{a_j\}_{j\in J} \subset L^2(0, T)$, one can compute the basis $\{\phi_j\}_{j\in J} \subset L^2(\Omega)^M$ as

$$\phi_j = \left(u, a_j\right)_{L^2(0,T)}.$$

Actually, $\{a_j\}_{j\in J}$ has the same minimization properties as $\{\phi_j\}_{j\in J}$, thus one could compute $\{a_j\}_{j\in J}$ in the same fashion as $\{\phi_j\}_{j\in J}$(i.e., calculating the eigenvectors of the operator T) with $H = L^2(0, T)$, $I = \Omega$ and $N = M$. Note that in this case $u \in L^2\left(\Omega; L^2(0, T)^M\right)$. Let $\{b_j\}_{j\in\mathbb{N}} \subset L^2(0, T)$ be the orthonormal basis computed when we take $H = L^2(0, T)$, $I = \Omega$ and $N = M$ with associated eigenvalues $\{\mu_j\}_{j\in\mathbb{N}}$, then the orthonormal basis for $L^2(\Omega)^M$ is calculated as

$$\psi_j = \frac{1}{\mu_j^{1/2}} \left(u, b_j\right)_{L^2(0,T)} \in L^2(\Omega)^M.$$

Notice that the basis ψ_j becomes orthonormal by virtue of Remark 2.

We have just shown that depending on how one chooses H, I and N the POD method gives two orthonormal bases. The question now is to see if both basis are the same, and if so, which one is more convenient from a computational point of view. The relevance of this question is significant when one applies the POD method to the numerical solution of (8). In such a case one has a finite dimensional space $V_h \subset L^2(\Omega)$ where the discrete solution $u_h : [0, T] \to V_h^M$ is sought. If one chooses

to calculate the eigenvectors and eigenvalues of the discrete operator $T_h : H \to H$ such that

$$(T_h\phi, \varphi)_H = \left((u_h, \phi)_H , (u_h, \varphi)_H\right)_{L^2(I)^N} \text{ for all } \varphi \in H,$$

with $H = V_h^M \subset L^2(\Omega)^M$, $I = [0, T]$ and $N = 1$, then the correlation matrix associated to T_h is a ($M \times$ (*number of mesh points*)) $\times$ ($M \times$ (*number of mesh points*)) matrix; on the other hand, if one chooses $H = L^2(0, T)$, $I = \Omega$ and $M = N$, then the correlation matrix is a $n \times n$ matrix, n being the length of the discrete sample of the numerical solution in the interval $[0, T]$. Usually, in large scale computations the number of mesh points is much larger than n; so that, the second method is computationally more efficient when one applies the POD method in numerical simulations. This was noted first by Sirovich [19] proposing the so called *snapshots method* to calculate the correlation matrix associated to the discrete operator T_h.

Next theorem states that both bases are the same.

Theorem 3 *Let $u \in L^2((0, T) \times \Omega)^M$, then, with the above notation, $\lambda_j = \mu_j$, $a_j = b_j$ and $\phi_j = \psi_j$ for all $j \in \mathbb{N}$.*

Proof We first note that

$$\sum_{j\in\mathbb{N}} a_j\phi_j = u = \sum_{j\in\mathbb{N}} b_j\psi_j,$$

then, as noticed in Remark 2,

$$\sum_{j\in\mathbb{N}} \lambda_j = \|u\|^2_{L^2(0,T;L^2(\Omega)^M)} = \sum_{j\in\mathbb{N}} \mu_j.$$

Now, one can define

$$u^m = \sum_{j=1}^{m} a_j\phi_j \text{ and } \hat{u}^m = \sum_{j=1}^{m} b_j\psi_j,$$

thus, using Proposition 2 when $H = L^2(\Omega)^M$,

$$\left\|u^m - u\right\|_{L^2(0,T;L^2(\Omega)^M)} \leq \left\|\hat{u}^m - u\right\|_{L^2(0,T;L^2(\Omega)^M)},$$

and when $H = L^2(0, T)$,

$$\left\|\hat{u}^m - u\right\|_{L^2(0,T;L^2(\Omega)^M)} \leq \left\|u^m - u\right\|_{L^2(0,T;L^2(\Omega)^M)}.$$

Then,

$$\sum_{j>m} \lambda_j = \left\|u^m - u\right\|_{L^2(0,T;L^2(\Omega)^M)} = \left\|\hat{u}^m - u\right\|_{L^2(0,T;L^2(\Omega)^M)} = \sum_{j>m} \mu_j,$$

i.e., $\lambda_j = \mu_j$ for all $j \in \mathbb{N}$. Now, it is straight forward to prove that $a_j = b_j$ and $\phi_j = \psi_j$ for all $j \in \mathbb{N}$ if one calculates Ta_j, with the operator T defined by (19) and $H = L^2(0,T)$. □

This theorem states that one can choose I, H and N in two different ways to compute the POD basis. In what follows, $\{\phi_j\}_{j\in\mathbb{N}}, \{\phi_{hj}\}_{j\in\mathbb{N}}$ are the orthonormal basis of $L^2(\Omega)^M$ (for u and u_h respectively) and $\{a_j\}_{j\in\mathbb{N}}, \{a_{hj}\}_{j\in\mathbb{N}}$ are the orthogonal basis of $L^2(0,T)$, no matter how H, I and N are chosen. Again, λ_j and λ_{hj} will denote the eigenvalues associated, thus

$$\|a_j\|^2_{L^2(0,T)} = \lambda_j \text{ and } \|a_{hj}\|^2_{L^2(0,T)} = \lambda_{hj}$$

for all $j \in \mathbb{N}$.

We can now state next theorem.

Theorem 4 *Suppose that* $\|u - u_h\|_{L^2(0,T;L^2(\Omega)^M)} < \varepsilon$. *Then,*

$$\begin{aligned}
|\lambda_j - \lambda_{hj}| &< \delta_j, \\
\|\phi_j - \phi_{hj}\|_{L^2(\Omega)^M} &< \hat{\delta}_j, \\
\|P_{H_{hm}} - P_{H_m}\| &< 2\sum_{j=1}^{m} \hat{\delta}_j, \\
\|P_{H_{hm}} u - u\|_{L^2(0,T;L^2(\Omega)^M)} &< \varepsilon + \sum_{j>m} \lambda_{hj},
\end{aligned}$$

where $H_{hm} = span\{\phi_{hj}\}_{j=1}^{m}$, $\delta = \varepsilon\left(2\|u\|_{L^2(I;H^N)} + \varepsilon\right)$ *and* $\delta_j, \hat{\delta}_j$ *as in Theorem* 2.

Proof The idea of the proof is to apply Theorem 2 to both T and T_h. As we pointed out above, one can choose H, I and N in two different ways. Theorem 3 assures that both ways yield the same basis; so that, to simplify the analysis we shall take $H = L^2(\Omega)^M$, $I = (0,T)$ and $N = M$.

First of all, one needs an estimate for $\|T - T_h\|$. To this end, let $\phi, \varphi \in H$, then

$$\begin{aligned}
(T\phi - T_h\phi, \varphi)_H &= \left((u,\phi)_H, (u,\varphi)_H\right)_{L^2(I)^N} - \left((u_h,\phi)_H, (u_h,\varphi)_H\right)_{L^2(I)^N} \\
&= \left((u-u_h,\phi)_H, (u,\varphi)_H\right)_{L^2(I)^N} \\
&\quad + \left((u_h,\phi)_H, (u-u_h,\varphi)_H\right)_{L^2(I)^N} \\
&\le \|u-u_h\|_{L^2(I;H^N)} \|\phi\|_H \|u\|_{L^2(I;H^N)} \|\varphi\|_H \\
&\quad + \|u_h\|_{L^2(I;H^N)} \|\phi\|_H \|u-u_h\|_{L^2(I;H^N)} \|\varphi\|_H \\
&\le \left(\|u\|_{L^2(I;H^N)} + \|u_h\|_{L^2(I;H^N)}\right)\|u-u_h\|_{L^2(I;H^N)}\|\phi\|_H\|\varphi\|_H \\
&< \left(2\|u\|_{L^2(I;H^N)} + \varepsilon\right)\varepsilon\|\phi\|_H\|\varphi\|_H,
\end{aligned}$$

so, if one sets $\varphi = T\phi - T_h\phi$, then

$$\|T\phi - T_h\phi\|_H < \left(2\|u\|_{L^2(I;H^N)} + \varepsilon\right)\varepsilon\|\phi\|_H,$$

i.e.,

$$\|T - T_h\| < \varepsilon\left(2\|u\|_{L^2(I;H^N)} + \varepsilon\right). \tag{21}$$

From (21) and Theorem 2 the first three estimates follow. Finally, to estimate the difference between u and its projection onto the subspace H_{hm},

$$P_{H_{hm}}u - u = \left(P_{H_{hm}} - I\right)(u - u_h) + P_{H_{hm}}u_h - u_h,$$

thus,

$$\begin{aligned}\left\|P_{H_{hm}}u - u\right\|_{L^2(0,T;L^2(\Omega)^M)} &\leq \left\|P_{H_{hm}} - I\right\|\|u - u_h\|_{L^2(0,T;L^2(\Omega)^M)} + \sum_{j>m}\lambda_{hj} \\ &< \varepsilon + \sum_{j>m}\lambda_{hj}.\end{aligned}$$

□

5 The POD method for computing bifurcation diagrams

The POD method implemented via the snapshot technique is often applied to derive from an infinite-dimensional dynamical system, with solution $u(t)$, a reduced one whose solution presumably has the same properties as $u(t)$. If the dynamical system depends on a parameter, say ν, the behavior of $u(t)$ is controlled by the values of ν, and $u(t)$ will live in different states or attractors as ν varies; so that, to characterize the dynamics of $u(t)$ one computes the so called bifurcation diagram of $u(t)$ as a function of ν. A conventional way to obtain the bifurcation diagram consists of approximating the infinite-dimensional dynamical system by a high-dimension discrete dynamical system using a space-time discretization and then calculating its numerical solution for different values of ν. Clearly, this methodology is very expensive in terms of computing resources; whereas the calculation of the bifurcation diagram may be more affordable if one uses the reduced dynamical system given by the POD method. However, many authors have already reported that a straightforward application of the POD method is not able to compute the bifurcation diagram accurately, i.e., computing the POD basis from a set of snapshots collected from different attractors instead of just one attractor as described in Sect. 4. The reason for this behavior can be understood if one thinks of the POD procedure as a statistical tool, where the elements of the basis are taken for being the most representative elements in a set of data (this was the original description of the POD method, also known as Karhunen-Loeve decomposition [16]). In doing so, *fundamental anomalies* of any specific attractor can be minimized if several attractors are introduced in the set to be studied.

In order to handle this problem several authors have introduced modifications to the POD technique. One of this modifications consists of using the POD method once for each different attractor, generating in this way a basis for each corresponding attractor, and then making use of any orthogonalization technique of the basis to get a finite dimensional subspace.

In doing so, the p-POD and SPOD method was introduced in [6] and [13]. In these works, it is intended to get an optimal procedure so that the dimension of the global basis can be as low as possible. Their idea is to use the POD method for one specific attractor. Then, in order to get the basis for a second attractor, one applies the POD method for the second attractor projected over the orthogonal subspace to the one computed by the POD method. By induction, one obtains a global basis. However, the problem now is that sometimes the bases generated by this methodology may be too large. We propose a cut-off criterion, to be applied at each step, that may partially overcome this problem.

In order to give a general description of the method, suppose one has s attractors $u_i \in L^2\left(I; H_i^N\right)$, for $1 \leq i \leq s$ (note that $I = \Omega$, $H_i = L^2(0, T_i)$ and $N = M$). The procedure to compute the global basis, for given tolerances $\hat{\varepsilon}_i$, can be formulated as follows:

1. Apply the POD method to u_1 and obtain $\left\{\phi_j\right\}_{j=1}^{m_1} \subset H$, with $m_1 \in \mathbb{N}$ such that $E_1(m_1) < \hat{\varepsilon}_1$.
2. For $i = 2, 3, \ldots, s$, define

$$\hat{u}_i = u_i - \sum_{j=1}^{m_{i-1}} \left(u_i, \phi_j\right)_h \phi_j$$

and apply the POD method to $\hat{u}_i$ to get $\left\{\phi_j\right\}_{j=m_{i-1}+1}^{m_i} \subset H$, with $m_i \in \mathbb{N}$ such that $E_i(m_i) < \hat{\varepsilon}_i$.

Note that E_i is defined analogously as the energy function in Remark 2:

$$E_i(m) = \frac{\left\| P_{H_m} u_i - u_i \right\|^2_{L^2(I;H^N)}}{\|u_i\|^2_{L^2(I;H^N)}}, \tag{22}$$

which can not be calculated via eigenvalues for any $i \geq 2$.

Note that in [6] and [13], the p-POD and SPOD method were introduced mainly to compute the first Hopf bifurcation point of a dynamical system, while in the present work, the SPOD method is intended to compute branches in the bifurcation diagram. They also compute the POD basis via the eigenvalues formulae (using (20) that can be calculated using only the eigenvalues), no matter which u_i is being considered, so one gets new elements in the basis without using the information that the other u_j can contribute. With the energy function (22), one can optimize the number of elements in the POD basis.

Next step is the analogously results obtained in Theorem 4 for the global basis obtained with the SPOD method when applied to any approximation set of attractors u_{hi}.

Theorem 5 *Suppose that* $u_i, u_{hi} \in L^2\left(0, T; L^2\left(\Omega\right)^M\right)$*, and suppose that*

$$\|u_i - u_{hi}\|_{L^2\left(0,T;L^2(\Omega)^M\right)} < \varepsilon_i$$

for all $1 \leq i \leq s$*. Then, if* $E_i\left(m_i\right) < \hat{\varepsilon}_i$ *for all* $1 \leq i \leq s$*,*

$$\left\| P_{H_{h,m_s}} u_i - u_i \right\|_{L^2\left(0,T;L^2(\Omega)^M\right)} < \varepsilon_i \left(1 + \hat{\varepsilon}_i\right) + \hat{\varepsilon}_i \left\|u_i\right\|_{L^2\left(0,T;L^2(\Omega)^M\right)}$$

for all $1 \leq i \leq s$*.*

Proof Using the same argument as in Theorem 4,

$$P_{H_{h,m_s}} u_i - u_i = \left(I - P_{H_{h,m_s}}\right)\left(u_i - u_{hi}\right) + P_{H_{h,m_s}} u_{hi} - u_{hi},$$

thus

$$\begin{aligned}\left\| P_{H_{h,m_s}} u_i - u_i \right\|_{L^2\left(0,T;L^2(\Omega)^M\right)} &< \varepsilon_i + \hat{\varepsilon}_i \left\|u_{hi}\right\|_{L^2\left(0,T;L^2(\Omega)^M\right)} \\ &< \varepsilon_i \left(1 + \hat{\varepsilon}_i\right) + \hat{\varepsilon}_i \left\|u_i\right\|_{L^2\left(0,T;L^2(\Omega)^M\right)}.\end{aligned}$$

□

6 Galerkin projection

This section is devoted to give estimates between the solution of (8) and the one obtained via Galerkin projection of (8) into the finite dimensional subspace computed by POD. We have just proven some estimates between any set or solution u and its projection onto the finite dimensional subspace that the POD method can calculate. The idea is to use these estimates to obtain estimates in dynamical systems.

In order to obtain as much general results as possible, one can compute an m-dimensional subspace $H_{hm} \subset V$ such that

$$\left\| P_{H_{hm}} u_{h\nu} - u_{h\nu} \right\|_{L^2(0,T;H)} < \hat{\varepsilon}_\nu \tag{23}$$

for all $\nu \in \Lambda$, with $\Lambda \subset \mathbb{R}$, u_ν the solution of (8) for a_ν, and $u_{h\nu}$ a numerical approximation of u_ν such that

$$\|u_\nu - u_{h\nu}\|_{L^2(0,T;V)} < \varepsilon_\nu \tag{24}$$

for all $\nu \in \Lambda$.

Assuming that $u_0 \in V$, let u_ν^m be the solution of the Galerkin projection of (8) onto H_{hm}, i.e., u_ν^m is such that for all $v \in H_{hm}$,

$$\begin{cases} \left(\dfrac{\partial u_\nu^m}{\partial t}(t), v\right)_H + a_\nu\left(u_\nu^m(t), v\right) + \left\langle Bu_\nu^m(t) + Qu_\nu^m(t), v\right\rangle_{V'\times V} = (f, v)_H \\ u_\nu^m(0) = P_{H_m} u_0 \in H_{hm} \end{cases} \tag{25}$$

One can also introduce the Ritz projection onto H_{hm} like the operator $R_\nu^m : V \longrightarrow H_{hm}$, H_{hm} equipped with V-norm, such that for any $w \in V$, $R_\nu^m w \in H_{hm}$ and

$$a_\nu\left(R_\nu^m w, v\right) = a_\nu(w, v) \text{ for all } v \in H_{hm}. \tag{26}$$

Furthermore, as it was noted in [15], when the POD basis is computed for any $u_{h\nu}$ obtained numerically, one can easily estimate the constants involved in the equivalence between the norm in H_{hm} when equipped with H-norm or V-norm, for any m. These estimates have a close relation with the stiffness matrix S defined by

$$s_{ij} = a_\nu\left(\phi_i, \phi_j\right) \text{ for } i, j = 1, 2, \ldots, m',$$

where $\left\{\phi_j\right\}_{j=1}^{m'} \subset span\left\{u_{h\nu}\right\}_{\nu\in\Lambda}$ is an orthonormal basis with $\left\{\phi_j\right\}_{j=1}^{m}$ the POD basis and $\left\{\phi_j\right\}_{j=m+1}^{m'}$ a complementary orthonormal set in $span\left\{u_{h\nu}\right\}_{\nu\in\Lambda}$, with $dim\left(span\left\{u_{h\nu}\right\}_{\nu\in\Lambda}\right) = m'$.

Proposition 3 *For all $v \in span\left\{u_{h\nu}\right\}_{\nu\in\Lambda}$,*

$$\frac{1}{\left\|S^{-1}\right\|_2}\|v\|_H^2 \le \|v\|_V^2 \le \|S\|_2\|v\|_H^2,$$

where $\|\cdot\|_2$ is the spectral norm for matrices.

Proof See [15]. □

Theorem 6 *Assuming that (23)–(25) hold, then,*

$$\begin{aligned} \left\|u_\nu^m - u_\nu\right\|_{L^2(0,T;H)} &< \left(c_\nu + C\sqrt{T}\right)\varepsilon_\nu + \left(\left\|S^{-1}\right\|_2^{1/2} + C\sqrt{T}\right)\|S\|_2^{1/2}\hat{\varepsilon}_\nu \\ &+ \sqrt{T}\left(\left\|\left(R^m - P_{H_m}\right)u_0\right\|_H + C\left\|\frac{\partial}{\partial t}\left(I - R^m\right)u_\nu\right\|_{L^2(0,T;H)}\right), \end{aligned}$$

where C is a constant that depends on ζ and $\left\|S^{-1}\right\|_2$ (ζ defined in (6)).

Proof First of all, one can define the error function $e(t)$ as

$$\begin{aligned} e(t) &= u_\nu(t) - u_\nu^m(t) \\ &= \left(u_\nu(t) - R^m u_\nu(t)\right) + \left(R^m u_\nu(t) - u_\nu^m(t)\right) = \rho(t) + \theta(t) \end{aligned}$$

with R^m defined in (26) (note the dependence of e, ρ, θ in ν, omitted for a simplified notation).

Now, one can proceed with estimates for ρ and θ in $L^2(0, T; H)$. Note that

$$\begin{aligned}\|\rho\|_{L^2(0,T;H)} &= \left\|u_\nu - R^m u_\nu\right\|_{L^2(0,T;H)} \\ &\leq \left\|\left(I - R^m\right)(u_\nu - u_{h\nu})\right\|_{L^2(0,T;H)} + \left\|\left(I - R^m\right) u_{h\nu}\right\|_{L^2(0,T;H)}.\end{aligned}$$

The first term can be estimated using (5) and (24),

$$\begin{aligned}\left\|\left(I - R^m\right)(u_\nu(t) - u_{h\nu}(t))\right\|_H &\leq c_\nu \left\|\left(I - R^m\right)(u_\nu(t) - u_{h\nu}(t))\right\|_V \\ &\leq c_\nu \|u_\nu(t) - u_{h\nu}(t)\|_V,\end{aligned}$$

because $\|I - R^m\| < 1$. Thus,

$$\left\|\left(I - R^m\right)(u_\nu - u_{h\nu})\right\|_{L^2(0,T;H)} \leq c_\nu \|u_\nu - u_{h\nu}\|_{L^2(0,T;V)} < c_\nu \varepsilon_\nu.$$

To estimate the last term, one should notice that for any t

$$\begin{aligned}\left\|\left(I - R^m\right) u_{h\nu}(t)\right\|_H^2 &\leq \left\|S^{-1}\right\|_2 \left\|\left(I - R^m\right) u_{h\nu}(t)\right\|_V^2 \\ &\leq \left\|S^{-1}\right\|_2 \left\|\left(I - P_{H_{hm}}\right) u_{h\nu}(t)\right\|_V^2 \\ &\leq \left\|S^{-1}\right\|_2 \|S\|_2 \left\|\left(I - P_{H_{hm}}\right) u_{h\nu}(t)\right\|_H^2,\end{aligned}$$

where the inequality $\|u - R^m u\|_V \leq \|u - w\|_V$ for all $w \in H^m$ has been used. Hence,

$$\begin{aligned}\left\|\left(I - R^m\right) u_{h\nu}\right\|_{L^2(0,T;H)} &\leq \left\|S^{-1}\right\|_2^{1/2} \|S\|_2^{1/2} \left\|\left(I - P_{H_{hm}}\right) u_{h\nu}\right\|_{L^2(0,T;H)} \\ &< \left\|S^{-1}\right\|_2^{1/2} \|S\|_2^{1/2} \hat{\varepsilon}_\nu.\end{aligned}$$

Using both estimates,

$$\|\rho\|_{L^2(0,T;H)} < c_\nu \varepsilon_\nu + \left\|S^{-1}\right\|_2^{1/2} \|S\|_2^{1/2} \hat{\varepsilon}_\nu. \tag{27}$$

In an analogous way, one can get

$$\|\rho\|_{L^2(0,T;V)} < \varepsilon_\nu + \|S\|_2^{1/2} \hat{\varepsilon}_\nu. \tag{28}$$

We turn our attention to θ. Then, for any $v \in H_{hm}$,

$$\begin{aligned}\left(\frac{\partial\theta}{\partial t}(t), v\right)_H + a_\nu(\theta(t), v) &= \left(\frac{\partial R^m u_\nu}{\partial t}(t), v\right)_H + a_\nu(u_\nu(t), v) \\ &\quad + \left\langle Bu_\nu^m(t) + Qu_\nu^m(t), v\right\rangle_{V'\times V} - (f(t), v)_H \\ &= \left(\frac{\partial}{\partial t}\left(R^m u_\nu(t) - u_\nu(t)\right), v\right)_H \\ &\quad + \left\langle Bu_\nu^m(t) + Qu_\nu^m(t), v\right\rangle_{V'\times V} \\ &\quad - \langle Bu_\nu(t) + Qu_\nu(t), v\rangle_{V'\times V} \\ &= -\left(\frac{\partial\rho}{\partial t}(t), v\right)_H + \left\langle Bu_\nu^m(t) - Bu_\nu(t), v\right\rangle_{V'\times V} \\ &\quad - \langle Q(\rho(t) + \theta(t)), v\rangle_{V'\times V},\end{aligned}$$

then, if $v = \theta(t)$,

$$\begin{aligned}&\frac{1}{2}\frac{\partial}{\partial t}\|\theta(t)\|_H^2 + \zeta\|\theta(t)\|_V^2 \le \left\|\frac{\partial\rho}{\partial t}(t)\right\|_H \|\theta(t)\|_H \\ &\quad + \left|\left\langle B\left(u_\nu^m(t)\right) - B(u_\nu(t)), \theta(t)\right\rangle\right| - \langle Q\rho(t), \theta(t)\rangle_{V'\times V}.\end{aligned} \tag{29}$$

We now estimate the nonlinear term. In doing so,

$$\begin{aligned}Bu_\nu^m(t) - Bu_\nu(t) &= -B\left(u_\nu(t), u_\nu(t) - u_\nu^m(t)\right) \\ &\quad + B\left(u_\nu(t) - u_\nu^m(t)\right) - B\left(u_\nu(t) - u_\nu^m(t), u_\nu(t)\right).\end{aligned} \tag{30}$$

The first term can be estimate as follows,

$$\begin{aligned}\left|\left\langle B\left(u_\nu(t), u_\nu(t) - u_\nu^m(t)\right), \theta(t)\right\rangle\right| &\le \left|\langle B(u_\nu(t), \rho(t)), \theta(t)\rangle\right| \\ &\le c_B\|u_\nu(t)\|_H^{\gamma_3}\|u_\nu(t)\|_V^{1-\gamma_3} \\ &\quad \times \|\rho(t)\|_V\|\theta(t)\|_V^{\gamma_3}\|\theta(t)\|_H^{1-\gamma_3} \\ &\le c_B c_\nu^{\gamma_3}\|u_\nu(t)\|_V\|\rho(t)\|_V\|\theta(t)\|_V^{\gamma_3}\|\theta(t)\|_H^{1-\gamma_3} \\ &\le \alpha\|\theta(t)\|_V^2 + \beta\|\theta(t)\|_H^2 + c(\alpha,\beta)\|\rho(t)\|_V^2,\end{aligned}$$

where $\|u_\nu(t)\|_V \le \|u_\nu\|_{L^\infty(0,T;V)}$ for all t has been used as well as Young inequality for $\alpha, \beta > 0$ and $c(\alpha, \beta)$ a constant depending on α and β to be chosen later.

Using the same ideas one can estimate the second and third term in (30). Thus,

$$\begin{aligned}&\left|\left\langle B\left(u_\nu(t) - u_\nu^m(t), u_\nu(t)\right), \theta(t)\right\rangle\right| \\ &\le \left|\left\langle B\left(\rho(t), u_\nu(t)\right), \theta(t)\right\rangle\right| + \left|\left\langle B(\theta(t), u_\nu(t)), \theta(t)\right\rangle\right| \\ &\le c_B\|u_\nu\|_{L^\infty(0,T;V)}\left(c_V^{\gamma_3}\|\rho(t)\|_V\|\theta(t)\|_V^{\gamma_3}\|\theta(t)\|_H^{1-\gamma_3} + \|\theta(t)\|_H\|\theta(t)\|_V\right) \\ &\le \alpha\|\theta(t)\|_V^2 + \beta\|\theta(t)\|_H^2 + c(\alpha,\beta)\|\rho(t)\|_V^2,\end{aligned}$$

and

$$\begin{aligned}
&\left|\left\langle B\left(u_\nu(t)-u_\nu^m(t),u_\nu(t)-u_\nu^m(t)\right),\theta(t)\right\rangle\right| \\
&\quad\le \left|\left\langle B\left(\rho(t),\rho(t)\right),\theta(t)\right\rangle\right|+\left|\left\langle B\left(\theta(t),\rho(t)\right),\theta(t)\right\rangle\right| \\
&\quad\le c_B\|\rho(t)\|_V\|\theta(t)\|_V^{\gamma_3}\|\theta(t)\|_H^{1-\gamma_3}\left(\|\rho(t)\|_H^{\gamma_3}\|\rho(t)\|_V^{1-\gamma_3}+\|\theta(t)\|_H^{\gamma_3}\|\theta(t)\|_V^{1-\gamma_3}\right) \\
&\quad\le c_B\|\rho(t)\|_V\left(c_V^{\gamma_3}\|\rho(t)\|_V\|\theta(t)\|_H^{\gamma_3}\|\theta(t)\|_V^{1-\gamma_3}+\|\theta(t)\|_H\|\theta(t)\|_V\right).
\end{aligned}$$

Noting that u_ν and u_ν^m are continuous functions in V, then ρ is bounded in V. Thus, due to the fact that $\|\rho(t)\|_V \le \|\rho\|_{L^\infty(0,T;V)} < +\infty$, and applying again Young inequality,

$$\begin{aligned}
&\left|\left\langle B\left(u_\nu(t)-u_\nu^m(t),u_\nu(t)-u_\nu^m(t)\right),\theta(t)\right\rangle\right| \\
&\quad\le \alpha\|\theta(t)\|_V^2+\beta\|\theta(t)\|_H^2+c(\alpha,\beta)\|\rho(t)\|_V^2.
\end{aligned}$$

Thus, putting all these estimates together in (29), and if one takes $\alpha=\dfrac{\zeta}{4}$ and $\beta=\dfrac{\zeta}{16}\left\|S^{-1}\right\|_2^{-1}$, then

$$\begin{aligned}
\frac{1}{2}\frac{\partial}{\partial t}\|\theta(t)\|_H^2 &\le \left\|\frac{\partial\rho}{\partial t}(t)\right\|_H\|\theta(t)\|_H-\frac{\zeta}{4}\|\theta(t)\|_V^2+3\beta\|\theta(t)\|_H^2+\frac{C^2}{2}\|\rho(t)\|_V^2 \\
&\le \frac{C^2}{2}\left(\left\|\frac{\partial\rho}{\partial t}(t)\right\|_H^2+\|\rho(t)\|_V^2\right)-\frac{\zeta}{4}\|\theta(t)\|_V^2+\frac{1}{8}\left\|S^{-1}\right\|_2^{-1}\|\theta(t)\|_H^2,
\end{aligned}$$

where $C^2=2\max\left\{3c(\alpha,\beta),\dfrac{4}{\zeta}\left\|S^{-1}\right\|_2\right\}$, and Young inequality has been used.

Using Proposition 3 in the above equation,

$$\frac{\partial}{\partial t}\|\theta(t)\|_H^2\le C^2\left(\left\|\frac{\partial\rho}{\partial t}(t)\right\|_H^2+\|\rho(t)\|_V^2\right).$$

Finally, using (28),

$$\begin{aligned}
\|\theta(t)\|_H^2 &\le \|\theta(0)\|_H^2+C^2\left(\int_0^t\left\|\frac{\partial\rho}{\partial t}(s)\right\|_H^2ds+\int_0^t\|\rho(s)\|_V^2ds\right) \\
&\le \|\theta(0)\|_H^2+C^2\left(\left\|\frac{\partial\rho}{\partial t}\right\|_{L^2(0,T;H)}^2+\|\rho\|_{L^2(0,T;V)}^2\right) \\
&< \|\theta(0)\|_H^2+C^2\left(\left\|\frac{\partial\rho}{\partial t}\right\|_{L^2(0,T;H)}^2+\left(\varepsilon_\nu+\|S\|_2^{1/2}\hat{\varepsilon}_\nu\right)^2\right).
\end{aligned}$$

Thus, setting the norm in $L^2(0,T;H)$ and using (27),

$$\begin{aligned}\left\|u_v^m - u_v\right\|_{L^2(0,T;H)} &\leq \|\rho\|_{L^2(0,T;H)} + \|\theta\|_{L^2(0,T;H)} \\ &< c_v\varepsilon_v + \left\|S^{-1}\right\|_2^{1/2} \|S\|_2^{1/2}\,\hat{\varepsilon}_v \\ &\quad + \sqrt{T}\left(\|\theta(0)\|_H + C\left(\left\|\frac{\partial\rho}{\partial t}\right\|_{L^2(0,T;H)} + \varepsilon_v + \|S\|_2^{1/2}\,\hat{\varepsilon}_v\right)\right) \\ &= \left(c_v + C\sqrt{T}\right)\varepsilon_v + \left(\left\|S^{-1}\right\|_2^{1/2} + C\sqrt{T}\right)\|S\|_2^{1/2}\,\hat{\varepsilon}_v \\ &\quad + \sqrt{T}\left(\|\theta(0)\|_H + C\left\|\frac{\partial\rho}{\partial t}\right\|_{L^2(0,T;H)}\right),\end{aligned}$$

where $\sqrt{a^2+b^2+c^2} \leq (|a|+|b|+|c|)$ have been used. □

The result of Theorem 6 gives an estimate of the error committed when one approximates the exact solution of (8) by the Galerkin projection onto the finite element space generated by the POD basis calculated via a numerical solution of (8). This result is relevant to estimate the fully discrete error committed when (25) is integrated by a time discretization scheme. The time discretization component of the error will depend on the approximation properties of the time discretization scheme and, in principle, can be estimated by applying conventional techniques.

7 Numerical examples

This section is devoted to some numerical examples related to fluid models. The aim is to prove that the SPOD method with the cut-off criterion explained in Sect. 5 can be used to compute bifurcation diagrams, in which stationary and periodic branches have to be studied.

7.1 Periodic flow past a square obstacle

In this first example we are interested in studying a fluid in a periodic 2D channel with a small square obstacle inside it. The fluid is acted by an external force $\mathbf{f}$ that makes the fluid move from left to right. Specifically, the governing Navier–Stokes equations plus initial and boundary conditions are the following:

$$\begin{cases} \dfrac{\partial \mathbf{u}}{\partial t} + \mathbf{u}\cdot\nabla\mathbf{u} - \nu\Delta\mathbf{u} = -\nabla p + \mathbf{f}\ , \text{ in } (0,T)\times\Omega \\ \mathbf{u}_{|\Gamma_0} = 0, \mathbf{u}_{|\Gamma_1} = \mathbf{u}_{|\Gamma_2} \\ \mathbf{u}(0) = 0 \end{cases} \tag{31}$$

where $\Omega = (-5/2, 5/2)\times(-1,1)\setminus[-1/10, 1/10]^2$, $\Gamma_1 = \{-5/2\}\times[-1,1]$, $\Gamma_2 = \{5/2\}\times[-1,1]$ and $\Gamma_0 = \partial\Omega\setminus(\Gamma_1\cup\Gamma_2)$, see Fig. 1. Finally the external force

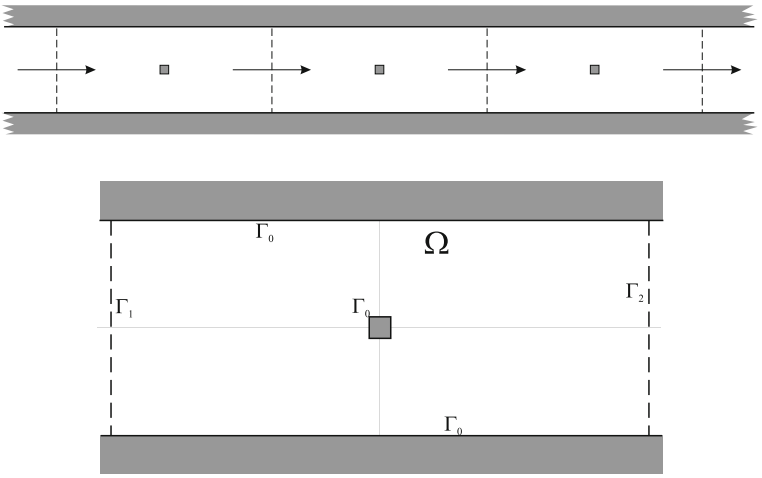

Fig. 1 Flow past a square obstacle in the periodic domain Ω

$\mathbf{f} = (2\nu, 0)$. The *Reynolds number*, in this case $Re = \frac{1}{\nu}$, will play the role of the bifurcation parameter. Dynamical system (31) can be written as (8), see [20].

To calculate the numerical solution to (1) we use Taylor-Hood P_2/P_1 finite elements for space discretization combined with a second order BDF-semi-Lagrangian scheme for time discretization. This leads to solve a linear Stokes problem every time step, which is solved by the Uzawa-preconditioned conjugate gradient algorithm [7]. More details about the numerical algorithm can be found in [10] and [4]. Notice that no splitting method has been used to calculate the numerical solution of (31).

To apply the POD or SPOD methods as described above, we need functions $u_h \in L^2\left(0, T; L^2\left(\Omega\right)^2\right)$. Noting that the finite element solution is continuous in space, then, from the numerical solution $\{u_h\left(t_i\right)\}_i$, we can define u_h as

$$u_h\left(t\right) = u_h\left(t_{s_j}\right) \text{ for all } t \in \left[t_{s_j} - \frac{\Delta t_s}{2}, t_{s_j} + \frac{\Delta t_s}{2}\right) \text{ and for all } j, \tag{32}$$

where $\left\{t_{s_j}\right\}_j \subset \{t_i\}_i$ and $\Delta t_s = t_{s_{j+1}} - t_{s_j}$ for all j, $u_h\left(t_{s_j}\right)$ are the so called snapshots and t_{s_j} are the time instants for the snapshots. Now, the time integrals of $u_h\left(t\right)$ yields Sirovich's formula for the POD (see (1)–(4) in Sect. 1). To compute the POD basis, it is customary to substract a reference state $\bar{u}_h$ from the snapshots $u_h\left(t_{s_j}\right)$. In the numerical results that follow we have chosen $\bar{u}_h$ as the average of the values of the solution u_{h1} of (31) (see Table 1 and the corresponding explanation below).

As usual in fluid models, when one solves numerically the problem, stationary attractors are obtained when Re is small. If one increases Re a first Hopf bifurcation is obtained, $Re = 457 \pm 1$, as well as a periodic attractor branch. In this specific model a new Hopf bifurcation leads to quasi-periodic and aperiodic attractors ($Re = 590 \pm 10$).

Here, three different basis are considered to compute the bifurcation analysis. In Table 1 we summarize the u_{hi} numerical solutions of (31) used in the SPOD

Table 1 The three bases considered to compute the bifurcation diagram. Each i denotes a different attractor in the SPOD method, n_i being the number of snapshots of each attractor (one complete period for periodic attractors for Δt_s) and m_i as defined in Sect. 5

Basis A				Basis B				Basis C			
i	Re	n_i	m_i	i	Re	n_i	m_i	i	Re	n_i	m_i
1	460	47	6	1	460	47	6	1	460	47	6
2	600	200	19	2	600	200	19	2	600	200	19
3	stat. att.	9	21	3	560	49	24	3	560	49	24
				4	500	48	29	4	500	48	29
								5	stat. att.	9	32

method to compute a global basis (each u_{hi} is a numerical attractor, i.e., a function in $L^2\left(0, T_i; L^2(\Omega)^2\right)$). Note that u_{h3} in basis A and u_{h5} in basis C are not periodic attractors of (31). These sets are made of snapshots of stationary attractors of (31) for low Reynolds numbers, specifically for $Re = 400, 410, 420, 430, 440, 450, 452, 454$ and 456. Although POD method is thought as a way to compute an orthonormal basis of a subspace that contains a periodic attractor, in fact, it can also be used to compute the subspace which best approximates any set of data, see [6]. The idea here is that the difference between any stationary attractor and its projection over the subspace computed by the SPOD method is as small as possible.

The tolerances taken for the SPOD method are: $\hat{\varepsilon}_{460} = 10^{-4}$, $\hat{\varepsilon}_{500} = 5 \times 10^{-6}$, $\hat{\varepsilon}_{560} = 10^{-5}$, $\hat{\varepsilon}_{600} = 5 \times 10^{-6}$ and $\hat{\varepsilon}_{s.a.} = 10^{-6}$. The tolerance for $Re = 460$ is the highest one because its dynamics is relatively simple. On the other hand, the attractor for $Re = 600$ has the most complicated behavior due to its non periodic dynamics. The attractors for $Re = 500$ and $Re = 560$ are also periodic, but we keep a similar tolerance to the attractor for $Re = 600$. Note that the attractor for $Re = 560$ has slightly higher tolerance than $Re = 500$. This can be done because when the SPOD method adds more new elements to the basis from $Re = 500$, the subspace will increase and the difference between these attractors and their projection over the subspace will become similar.

In Fig. 2 one can see the energy function defined in (22) for some attractors as well as for the last set in the SPOD method for basis A and C. This figure shows the fast decay of the energy function when m increases. Note that the decay for every attractor is deepest in the range of eigenvalues taken from the SPOD method when the attractor was analyzed. For instance, for $Re = 600$, the range between $m = 7$ and $m = 19$ is the fastest decay while the range between $m = 1$ and $m = 6$ is almost flat, i.e., the first 6 elements of the basis generates a subspace that could not approximate the dynamics of this attractor.

Finally, we also include in Fig. 2 an attractor that was not used to compute the SPOD basis, $Re = 520$. Nevertheless, the basis C can give a good approximation of this attractor what suggests that the range between $Re = 460$ and $Re = 600$ can be properly approximated by this basis (also the reduction in the number of elements added to the basis from u_{500} confirms this idea).

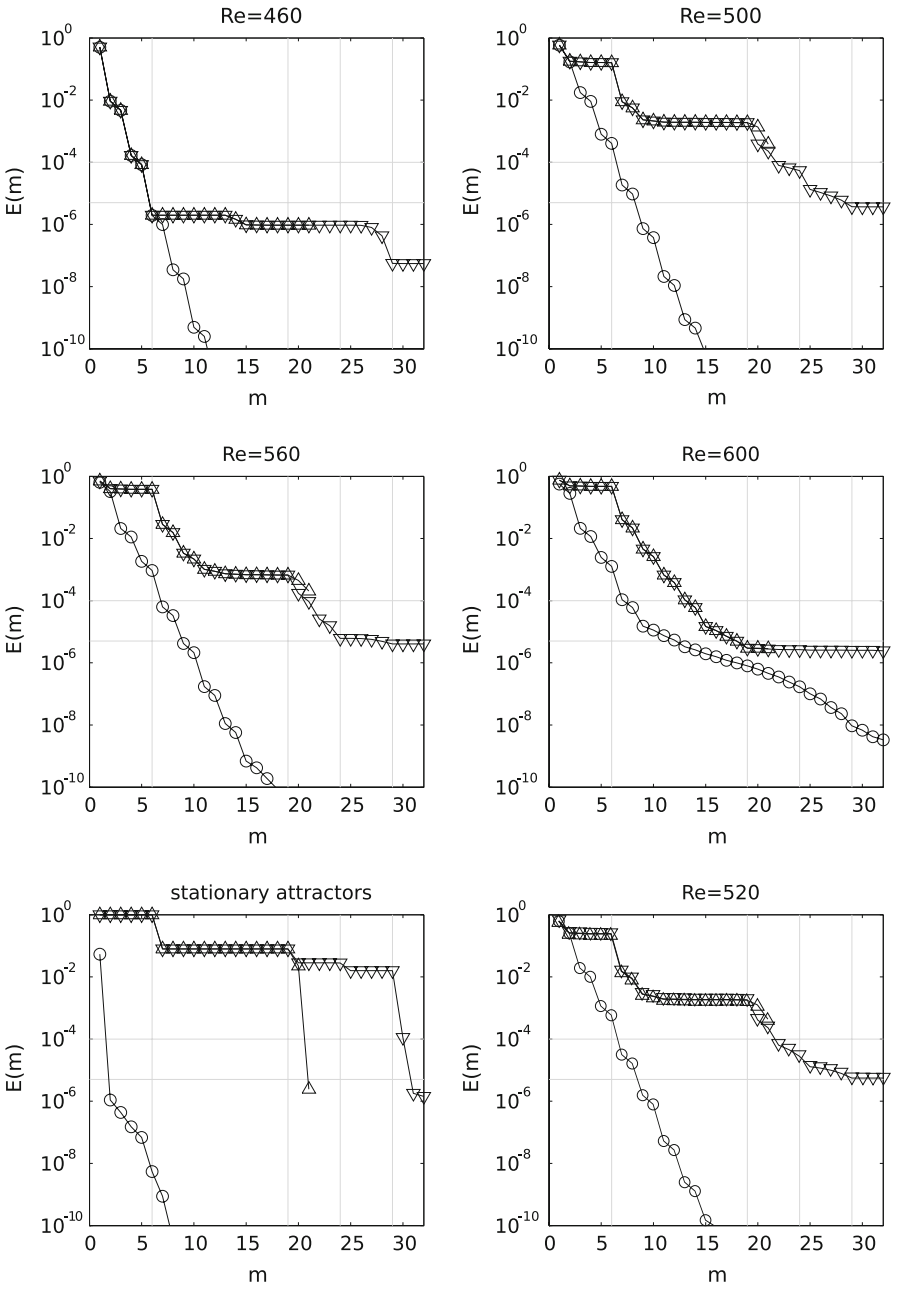

Fig. 2 Energy function (22) for basis A, △, and C, ▽. The energy function for the POD basis computed individually for each attractor is also shown, ○

We also have to point out that with our basis, from Fig. 2, $E_{Re}\,(32) < 5 \times 10^{-6}$ for any Re. If the cut-off criterion is not used to compute the SPOD basis, and any specific attractor is studied independently, the number of elements of the basis would

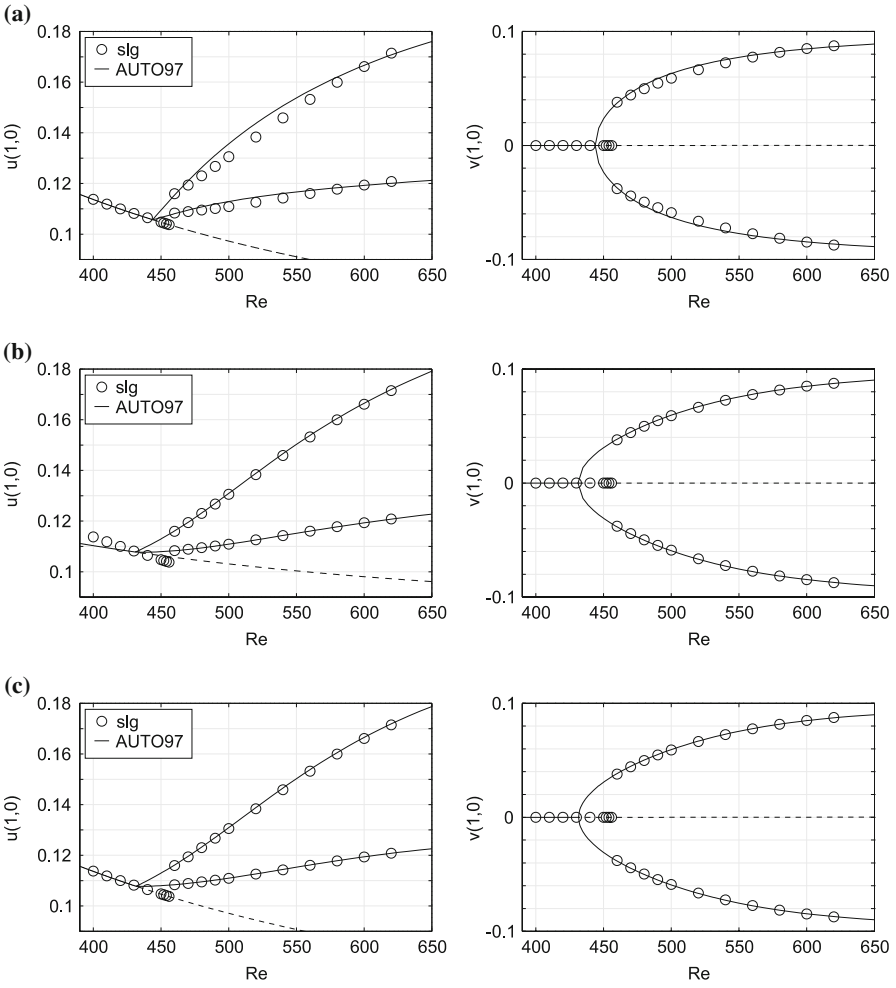

Fig. 3 Bifurcation diagram for the Galerkin projection (–) for: **a** basis A; **b** basis B; **c** basis C. DNS attractors in circles. When periodic attractors, maximum and minimum values of the attractors are shown in different branches

have increased to approximately 40. Furthermore, the basis would have had problems related to lack of orthogonality due to numerical errors and some noisy problems would have appeared as well.

In Fig. 3, we show bifurcation diagrams computed by the Galerkin projection of (31) onto the subspaces spanned by basis A, B and C. These diagrams were computed by AUTO97 [9]. In the figure we show the AUTO97 solution –line–, i.e., attractors, compared to the numerical attractors obtained by semi-Lagrangian method –circle– (usually called DNS, Direct Numerical Solution) for function u and v at the point $(1, 0)$ of the domain Ω. Continuous lines means stable attractor while dashed lines are unstable ones. For periodic attractors, minimum and maximum values of any attractor

are shown, i.e., a periodic attractor branch in the bifurcation diagram is shown as two curves.

As one could expect, basis C gives the best result. The agreement between DNS and SPOD method is very good in the whole range. One can only see differences in the Hopf bifurcation point, which might be due to numerical inaccuracies of the DNS.

The difference between basis A and C is that the first one does not include any attractor between $Re = 460$ and $Re = 600$. In the bifurcation diagram obtained for this basis one can check that the accuracy is similar to the bifurcation diagram for basis C apart from the range of Re between 460 and 600, where both solutions are slightly different, as one might expect. In fact, this means that the dynamical system for values of Re between 460 and 600 can not be interpolated with the solutions for $Re = 460$ and 600.

Differences between basis B and C can be explained in a similar manner. In this case, both bifurcation diagrams differ just in the stationary branch, and basis B differs from basis C exactly in the set of stationary attractors. That is, one can not extrapolate the dynamics of the attractors of the dynamical system, as one could expect.

7.2 A barotropic ocean model

This second example is devoted to a barotropic ocean model. We suppose an idealized ocean enclosed in the domain $[0, L] \times [0, 2L]$, with $L = 1000\, km$, with constant depth of 800 m. and driven by a wind stress $\tau\,(x, y) = (-\tau_0 \cos(\pi y/L), 0)$. This is a simple model in oceanography, but still very interesting because it retains some of the important features in the dynamics of the oceans. In studying this model, one can use the vorticity-stream function formulation:

$$\begin{cases} \dfrac{\partial \omega}{\partial t} + J(\psi, \omega) + \beta \dfrac{\partial \psi}{\partial x} = A_H \Delta \omega - \gamma \omega + \dfrac{rot\, \boldsymbol{\tau}}{H}, \text{ in } (0, T) \times \Omega \\ \Delta \psi = \omega, \text{ in } (0, T) \times \Omega \\ \omega_{|\partial\Omega} = \psi_{|\partial\Omega} = 0 \end{cases} \tag{33}$$

where $\omega = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}$ is the vorticity of the fluid, ψ is the stream function, i.e., $u = -\frac{\partial \psi}{\partial y}$ and $v = \frac{\partial \psi}{\partial x}$, $J(\psi, \omega) = \frac{\partial \psi}{\partial x}\frac{\partial \omega}{\partial y} - \frac{\partial \psi}{\partial y}\frac{\partial \omega}{\partial x}$ and $\mathbf{u} = (u, v)$ is the velocity of the fluid. The parameters in (33) are: H, the constant depth; $\gamma = 10^{-7}$, the bottom friction coefficient; A_H, the horizontal viscosity and bifurcation parameter; and $\beta = 2 \times 10^{-11}$, the first order approximation to the Coriolis force in the β-plane approximation. Discussions about this model can be found in [5,12] or in a different formulation in [17] and [18], for example. The dynamical system (33) can also be written as the abstract formulation (8) [2].

Numerical solutions of this model are computed in the same fashion that the ones from the obstacle problem: finite element method combined with semi-Lagrangian schemes and a second order BDF discretization. The numerical algorithm first solves ω at t_{n+1}, with ψ extrapolated at t_{n+1}, and then ψ at t_{n+1}. Details about the numerical algorithm can also be found in [10] and [4].

Table 2 The basis considered to compute the bifurcation diagram of (31). The notation in the table is analogous to the Table 1

i	A_H	n_i	m_i
1	800	93	12
2	500	266	53
3	600	236	79
4	700	113	82
5	stat. att.	13	87

Model (33) is sometimes called as a double-gyre circulation model in oceanography. The solution shows two gyres close to the western boundary (due to Coriolis force), one in the northern half of the domain, simulating the subpolar gyre, and another one in the southern half, the subtropical gyre. Which one of them is stronger can be important from the oceanography point of view. In studying so, some authors introduce the normalized transport difference,

$$TD = \frac{|\psi_{po}| - |\psi_{tr}|}{\max\limits_{\Omega} |\psi|},$$

where $\psi_{po} = \min\limits_{\Omega} \psi$ and $\psi_{tr} = \max\limits_{\Omega} \psi$ are the maximum transport within the subpolar and subtropical gyres, see [5] for instance. When $TD = 0$, it means that the fluid is symmetric respect $y = L$, the centre of the domain, i.e., there is no mixture between subpolar and subtropical flow. When $TD > 0$, then the subpolar transport becomes larger and the subpolar gyre dominates the subtropical one. When $TD < 0$ the model is in the opposite. Thus, TD can be used as a characteristic of the attractors of (33).

When (33) is solved numerically for different values of A_H, stationary symmetric ($A_H \geq 1262.5 \pm 12.5$) and non-symmetric attractors are obtained ($865 \pm 5 \leq A_H \leq 1262.5 \pm 12.5$). At $A_H = 865 \pm 5$ a Hopf bifurcation occurs and a branch of periodic attractors are found (in fact, two symmetric branches). If A_H is further diminished, a new Hopf bifurcation appears at $A_H = 675 \pm 25$, and the computed attractors become double-periodic. Finally, when A_H is much lower, $A_H \leq 465 \pm 5$, aperiodic attractors appear.

This description of the attractors of (33) comes from many independent numerical simulations of the model for different values of the bifurcation parameter A_H. Nevertheless, this can be done just for some values of A_H, so, if one is interested in studying the attractors in the whole range of A_H, bifurcation diagram must be computed. Again, we use Galerkin projection over a subspace computed by SPOD. In this example, one could think in computing SPOD from (ω, ψ), but the dynamical system, when written in abstract formulation (8), only depends on ω because $\psi = A^{-1}\omega$, with A the Laplacian operator with Dirichlet boundary conditions. Thus, SPOD basis must be computed just from ω.

Then, in this example one can use again the methodology explained in Sect. 5. The attractors used to compute the global basis to be used in the Galerkin projection are summarize in Table 2. Note that we consider again a set of stationary attractors, in this

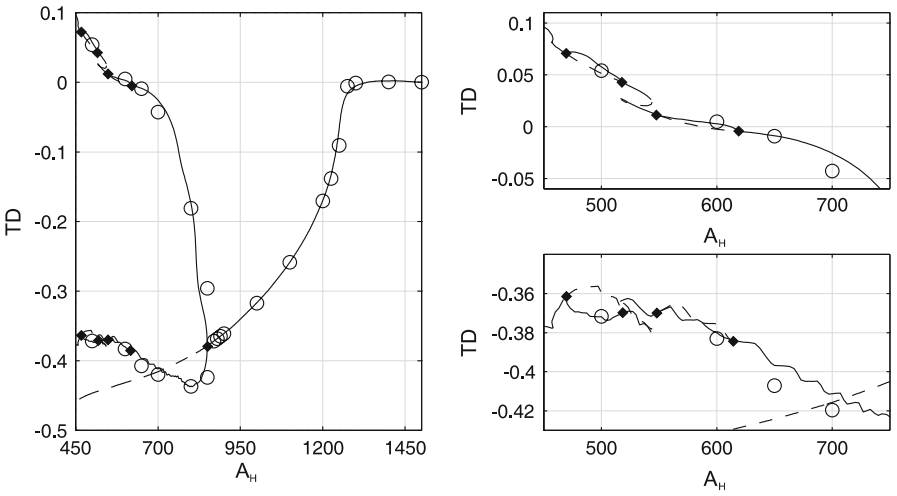

Fig. 4 Same as Fig. 3. *Right panels* are a detail of the *left one*

case for $A_H = 870, 880, 890, 900, 1000, 1100, 1200, 1225, 1250, 1275, 1300, 1400$ and 1500.

Taking into account the same considerations and results as in the obstacle problem, we consider $\hat{\varepsilon}_{800} = 10^{-4}$, $\hat{\varepsilon}_{700} = 10^{-4}$, $\hat{\varepsilon}_{600} = 5 \times 10^{-5}$, $\hat{\varepsilon}_{500} = 5 \times 10^{-5}$ and $\hat{\varepsilon}_{s.t.} = 10^{-6}$. Note that again we require a lower tolerance when the attractor has a more complicated dynamics, in this case, when it has a double period.

In Fig. 4 we show the bifurcation diagram computed by AUTO97 for the Galerkin projection of (33). The same symbols are used in Fig. 4 and Fig. 3. In this case, we also included a ♦ symbol to mark the exact value of the Hopf bifurcation points. One can see that the bifurcation diagram computed using the Galerkin projection agrees properly with the DNS. When A_H is large, there is one stationary attractor that is symmetric. It is possible to check that it is symmetric because TD is close to 0. Approximately at the same value of A_H, both, the DNS and the Galerkin projection, give nonsymmetric stationary attractors, in this case with $TD < 0$. Note that in fact, for model (33), around this value of A_H there is a pitchfork bifurcation point. In the bifurcation diagram only one of the branches can be computed because the SPOD basis was calculated from attractors in the same branch. Nevertheless, one can prove from equations (33) that the symmetric branch also exists.

Next, at $A_H = 850.00$ there is a Hopf bifurcation in the reduced model and, therefore, periodic attractors appear. Note the good agreement between the attractors in the reduced model and those computed in the DNS. In Fig. 5 we show a phase portrait of some attractors for the variables TD and the kinetic energy of the solution:

$$KE = \frac{1}{2} \int_{\Omega} |\mathbf{u}|^2 .$$

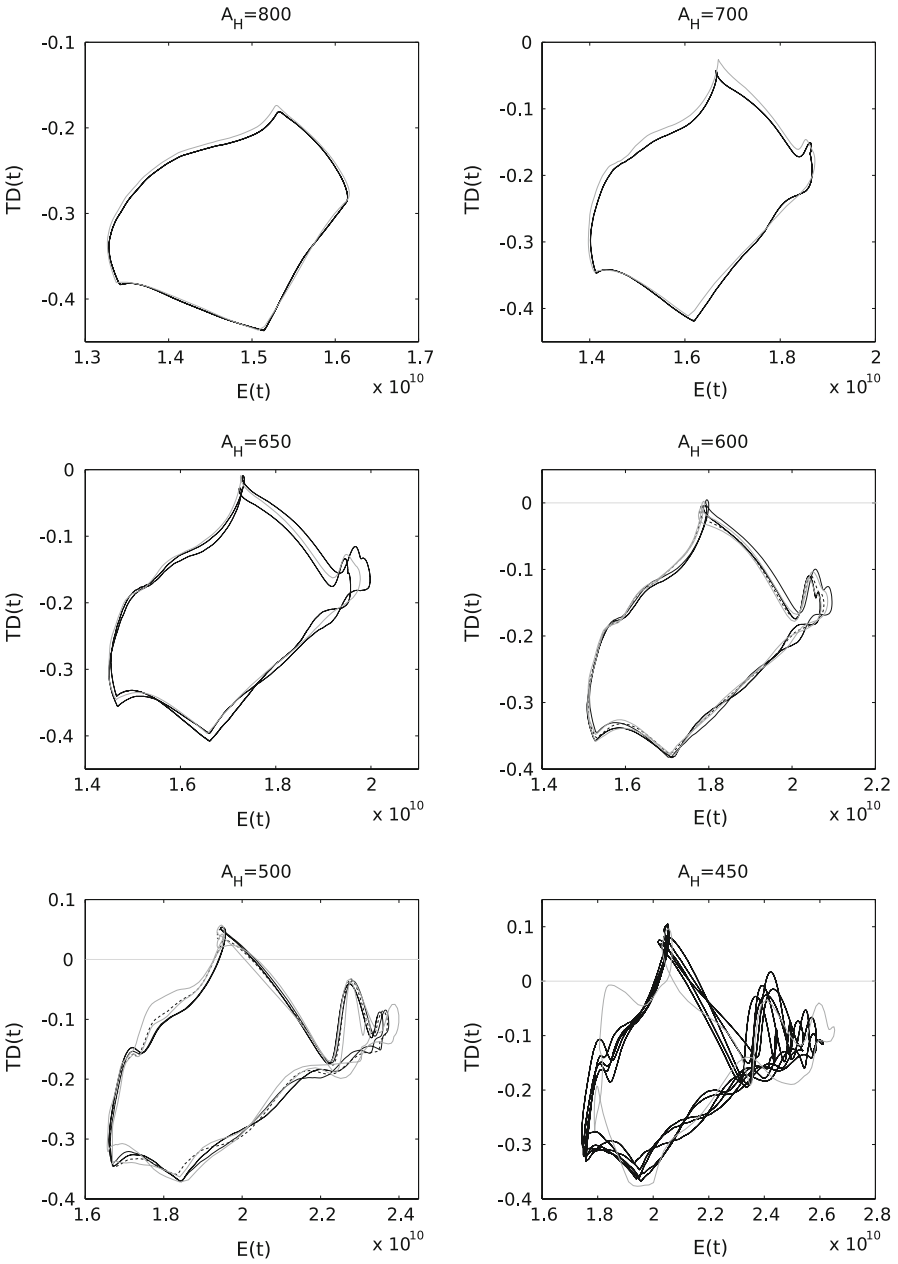

Fig. 5 Phase portrait of DNS (*black*) and Galerkin projection attractors (*gray*). For $A_H = 600$ and 500, periodic unstable solutions are also shown (*dashed*)

One of the differences between both models lies in the next bifurcation points. In the reduced model there are two more bifurcation points. In the first one a new branch of double-periodic attractors appears, while the branch of simple-periodic

attractors remains as unstable periodic solutions. The interval for this breach is $A_H \in (549.41, 618.91)$. The second one is a similar branch of attractors with a double period, now for $A_H \in (469.08, 518.09)$. Moreover, for $A_H \in (516.34, 543.91)$ there are three periodic solution for each value of A_H, two of them stable and the other one unstable.

It is difficult to know exactly how different are both models for $A_H < 700$, because attractors from model (33) can only be computed for some values in the whole range. For $A_H = 650$ the attractor computed by DNS is a double-period attractor, while for the reduced model it is a simple period one. On the other hand, for $A_H = 600$ and 500 attractors from the reduced model and DNS are double period attractors. Nevertheless, these double period attractors are in different branches in the bifurcation diagram of the reduced model, which is not possible to check from DNS. In any case, attractors in both models are qualitatively the same, even for $A_H = 650$, attractor that was not included to compute the SPOD basis.

For $A_H = 450$ the attractors computed by the reduced model and DNS have a different behavior. The main difference is that the DNS attractor is aperiodic and the attractor of the Galerkin projection is periodic. This difference between both models can be explained by noting that the attractors used to compute the SPOD basis include DNS attractors for $A_H \geq 500$, so any attractor for lower values of A_H must be extrapolated. Furthermore, aperiodic attractors can not be handled with POD as noted, for instance, in [1].

Acknowledgments This research has been partially funded by grant REN2002-03276 from Comisión Interministerial de Ciencia y Tecnología and grant CGL2006-11264-C04-02 from Ministerio de Educación y Ciencia de España.

References

1. Bangia, A.K., Batcho, P.F., Kevrekidis, I.G., Karniadakis, G.E.: Unsteady two-dimensional flows in complex geometries: comparative bifurcation studies with global eigenfunction expansions. SIAM J. Sci. Comput. **18**(3), 775–805 (1997)
2. Barcilon, V., Constantin, P., Titi, E.S.: Existence of solutions to the Stommel-Charney model of the Gulf Stream. SIAM J. Math. Anal. **19**(6), 1355–1364 (1988)
3. Berkooz, G., Holmes, P., Lumley, J.L.: The proper orthogonal decomposition in the analysis of turbulent flows. In: Annual Review of Fluid Mechanics, vol. 25, pp. 539–575. Annu. Rev. Palo Alto (1993)
4. Bermejo, R., Galán del Sastre, P.: Long term behavior of the wind driven circulation of a numerical North-Atlantic ocean circulation model. In: ECCOMAS 2004 (2004)
5. Chang, K.I., Ghil, M., Ide, K., Lai, C.C.A.: Transition to aperiodic variability in a wind-driven double-gyre circulation model. J. Phys. Oceanogr. **31**, 1260–1286 (2001)
6. Christensen, E.A., Brøns, M., Sørensen, J.N.: Evaluation of proper orthogonal decomposition-based decomposition techniques applied to parameter-dependent nonturbulent flows. SIAM J. Sci. Comput. **21**(4), 1419–1434 (electronic) (2000)
7. Dean, E.J., Glowinski, R.: On some finite element methods for the numerical simulation of incompressible viscous flow. In: Incompressible Computational Fluid Dynamics, pp. 17–65. Cambridge University Press, Cambridge (1993)
8. Deane, A., Kevrekidis, I.G., Karniadakis, G.E., Orszag, S.: Low-dimensional models for complex geometry flows: Application to grooved channels and circular cylinders. Phys. Fluids A. **3**(10), 2337–2354 (1991)
9. Doedel, E.J., Champneys, A.R., Fairgrieve, T.F., Kuznetsov, Y.A., Sandstede, B., Wang, X.J.: AUTO97: Continuation and bifurcation software for ordinary differential equations. Technical report, Department of Computer Science, Concordia University, Montreal, Canada (1997). (Disponible por FTP desde ftp.cs.concordia.ca en el directorio pub/doedel/auto)

10. Galán del Sastre, P.: Estudio numérico del atractor en ecuaciones de Navier-Stokes aplicadas a modelos de circulación del océano. PhD Thesis, Universidad Complutense de Madrid, Departamento de Matemática Aplicada, Madrid, España(2004)
11. Holmes, P., Lumley, J.L., Berkooz, G.: Turbulence, Coherent Structures, Dynamical Systems and Symmetry. Cambridge Monographs on Mechanics. Cambridge University Press, Cambridge (1996)
12. Jiang, S., Jin, F.F., Ghil, M.: Multiple equilibria, periodic, and aperiodic solutions in a wind-driven, double-gyre, shallow-water model. J. Phys. Oceanogr. **25**, 764–786 (1995)
13. Jørgensen, B.H. : Low-dimensional modeling and dynamics of the flow in a lid driven cavity with a rotating rod. PhD Thesis, Technical University of Denmark, Deparment of Energy Engineering, Fluid Mechanics Section, DK-2800 Lyngby, Denmark (2000)
14. Kunisch, K., Volkwein, S.: Galerkin proper orthogonal decomposition methods for parabolic problems. Numer. Math. **90**(1), 117–148 (2001)
15. Kunisch, K., Volkwein, S.: Galerkin proper orthogonal decomposition methods for a general equation in fluid dynamics. SIAM J. Numer. Anal. **40**(2), 492–515 (electronic) (2002)
16. Loève, M.: Probability Theory, 4th edn. Springer-Verlag, New York,[etc] (1978)
17. Simonnet, E.M.G., Ide, K., Temam, R., Wang, S.: Low-frequency variability in shallow-water models of the wind-driven ocean circulation. Part I: Steady-state solutions. J. Phys. Oceanogr. **33**, 712–728 (2003)
18. Simonnet, E.M.G., Ide, K., Temam, R., Wang, S.: Low-frequency variability in shallow-water models of the wind-driven ocean circulation. Part II: Time-dependent solutions. J. Phys. Oceanogr. **33**, 729–752 (2003)
19. Sirovich, L.: Turbulence and the dynamics of coherent structures. Part I: Coherent structures. Quart. Appl. Math. **45**(3), 561–571 (1987)
20. Temam, R.: Infinite-dimensional dynamical systems in mechanics and physics, Applied Mathematical Sciences, vol. 68. Springer, New York (1988)