Long-time behavior of the POD method

Jens Kemper*

July 2, 2010

Abstract

We present explicit error bounds concerning the behavior of the proper orthogonal decomposition (POD) method when the data is drawn from long trajectories. We express the error of the POD method in terms of the canonical angle for systems with exponentially decaying behavior. We test our theoretical bounds numerically using a linear parabolic equation. The considerations are motivated by a subdivision algorithm for the computation of invariant measures in discrete dynamical systems using the POD method as a model reduction tool.

Keywords: Model reduction, Singular value decomposition, Dynamical systems, Parabolic equations, Finite Elements.

1 Introduction

The proper orthogonal decomposition (POD), also known as principal component analysis or Karhunen-Loeve decomposition, is a model reduction method that has reached great influence in the theory of dynamical systems in recent years.

The idea of the POD method is to determine a nested family of subspaces in the original state space that optimally span the data consisting of given snapshots. Common applications are given in fluid dynamics and control theory ([KV02], [RCM04], [ASG01]). But also in the theory of parabolic problems the application of POD is investigated in recent years, [KV01].

The method is described in more detail in [HLB96]. Also [Ant05] gives a nice introduction into the power of the method where particular attention is paid to the balanced truncation ansatz. See also [BQO05] for the combination of the POD method with a balanced truncation ansatz.

We introduce the method in a simple, abstract setting. The definition is as follows:

Definition 1.1. *Let H be a separable real Hilbert space and a collection of snapshots in H be given by*

$$\{y_i : i = 1, \dots, m\}.$$

An ℓ-dimensional orthonormal system $\{w_k\}_{k=1}^{\ell}$ is called proper orthogonal decomposition basis of rank ℓ corresponding to $\{y_i\}_i$ *if it solves the minimization problem*

$$E_{\text{pod}}(\{\psi_k\}_{k=1}^{\ell}) := \frac{1}{m}\sum_{j=1}^{m} \|y_j - P_\psi y_j\|^2 \stackrel{!}{=} min \tag{1}$$

where $P_\psi = \sum_{k=1}^{\ell} \psi_k \psi_k^T : H \to H$ is the orthogonal projection onto $\text{span}\{\psi_1, \dots, \psi_\ell\}$.

*Supported by CRC 701 "Spectral Analysis and Topological Methods in Mathematics"

We want to focus on the behavior of the POD method in parabolic problems of the form

$$\begin{aligned} \frac{\partial u}{\partial t} &= f(u) - Au =: F(u) \quad && \text{in } \Omega \\ u &= 0 && \text{on } \partial\Omega \\ u(x,0) &= u_0(x), && x \in \Omega \end{aligned} \tag{2}$$

where $A : V \to H$ is a linear positive operator, $f \in C(V, H)$ with real separable Hilbert spaces V, H defining a so-called *Gelfand triple*, i.e. V is dense in H with continuous injection and, by identifying H and its dual H^*, we get dense embeddings

$$V \subset H = H^* \subset V^*.$$

Typical error results for the POD method in discretized parabolic problems cover the case of a single, finite-time trajectory. Typically, the snapshots are chosen along trajectories of the system. Then, the POD modes span a low-dimensional subspace. By a Galerkin ansatz with respect to the POD modes we get a dynamical system in the low-dimensional POD subspace of the form

$$\begin{aligned} \frac{\partial v}{\partial t} &= F_{\text{red}}(v) \quad F_{\text{red}} : \mathbb{R}^\ell \to \mathbb{R}^\ell \\ v(0) &= P_\psi u_0 \end{aligned} \tag{3}$$

The question arises, how the original and the reduced system (3) relate to each other. We will recall an error result from [KV01] for the resulting POD trajectory in Section 2.

The error bounds in [KV01] strongly depend on the length T_e of the sample trajectory. In this paper we analyze the long time behavior of the POD method, i.e. the case when T_e grows. This problem is motivated by some set-oriented algorithms for the approximation of invariant measures in high-dimensional systems that we developed in recent years, see [Kem10]. These algorithms combine subdivision techniques used by Dellnitz, Junge et al. ([DFJ01],[DJ98], [DJ99]) for computing attractors and invariant measures (cf. the software package GAIO) with the POD method as a model reduction method. Since these algorithms analyze the long time behavior of dynamical systems we are also interested in the long time behavior of the POD method or, to be more precise, in the behavior of the POD method when the data is drawn from long trajectories of the original system.

Another question is also motivated by the subdivision algorithms: Should one rather take one long trajectory or several short ones to set up a collection of snapshots? Since the algorithms in GAIO are based on the computation of short time trajectories, it is reasonable to use a large number of trajectories. Indeed, numerical experiments suggest a trade-off between the number and the length of trajectories used for the collection of snapshots. We will state some theoretical and numerical results concerning this question in Sections 3.4 and 4.3.

Let us look at the problem in more detail. As suggested above we consider the case of m snapshots along a trajectory with time spacing $T = \frac{T_e}{m}$. Using these snapshots we derive an ℓ-dimensional POD basis as explained above. If we compare the POD and the original trajectory various limit problems occur, e.g.

- ℓ, m fixed, T growing to infinity,
- ℓ, T fixed, m growing to infinity,
- m, T fixed, ℓ either growing to infinity or to the dimension N of the spatial discretization of (2).

We will mainly derive explicit error bounds describing the dependence on m, T where ℓ is chosen adequately with respect to the dynamics of the model systems.

We will see in Section 2 that the POD vectors to snapshots $\{y_j\}_j$ are given by the singular vectors of a linear operator Y corresponding to $\{y_j\}_j$. If we consider a linearized system

$$\frac{\partial u}{\partial t} = Au \tag{4}$$

with snapshots along a trajectory, it is easy to see that the POD basis strongly depends on the singular value decomposition of A. On the other hand it is well-known that the dynamical behavior of (4) is described by the eigenvalues and eigenvectors of A. Observe that for arbitrary matrices the singular value decomposition tells us nothing about the structure of the eigenvectors and vice versa. As an add-on we will see in the following that for special systems with exponentially decaying behavior the singular vectors are ruled by the eigenvalue decomposition. By that we get proper error bounds for the POD method.

The outline of this paper is as follows: In Section 2 we will describe the relation between the proper orthogonal decomposition and the singular value decomposition in detail. An explicit formulation of the error term in (1) immediately follows from these considerations. After that we recall a finite-time POD error estimate by Kunisch and Volkwein to see the structure of typical POD error results. For our long-time error estimates we use the well-known concept of canonical angles which we recall in the end of Section 2.

In Section 3 we present our main results. We consider exponentially decaying dynamical behavior. In Section 3.1 the corresponding POD basis for a trajectory converging to an asymptotically stable fixed point is analyzed. In Section 3.2 we derive further estimates by a deeper analysis of the different eigendirections of a linearized system around an asymptotically stable fixed point. As a slight generalization of the fixed point case we consider an attracting invariant subspace after that. Finally, a result for the POD method with snapshots from more than one trajectory is presented in Section 3.4

In Section 4 several of the theoretical bounds derived in Section 3 are illustrated by numerical experiments. We will see that the theoretical bounds for the case of different speeds of convergence describe the numerical results rather well. For the case of snapshots taken from many trajectories instead of one, the numerical experiments suggest a trade-off between the number of trajectories and their length. A deeper analysis of this phenomenon is part of future work.

2 Theoretical background and existing theory

2.1 Singular Value decomposition of compact operators

The concept of POD is strictly related to the singular value decomposition of compact operators.

Theorem 2.1 (see e.g. [Kir96]). *Let G, H be arbitrary Hilbert spaces and $Y : H \to G$ a compact operator with adjoint operator Y^*. Let the eigenvalues of the self-adjoint positive semidefinite operator $K = Y^*Y : H \to H$ be denoted by $\{\lambda_i\}_{i\in I}$ with*

$$\lambda_1 \geq \lambda_2 \geq \ldots \geq 0$$

where I may be finite ($I \subset \mathbb{N}$) or infinite ($I = \mathbb{N}$). Then there are orthonormal systems $\{u_i\}_{i\in I} \subset H$, $\{v_i\}_{i\in I} \subset G$ with

$$Yu_i = \sigma_i v_i \quad \text{and} \quad Y^*v_i = \sigma_i u_i$$

where $\sigma_i = \sqrt{\lambda_i}$, $i \in I$. For every $u \in H$ we get the expression

$$u = u_0 + \sum_{i\in I} \langle u, u_i\rangle u_i$$

with $u_0 \in ker(K)$ and

$$Ku = \sum_{i\in I} \sigma_i \langle u, u_i\rangle v_i. \tag{5}$$

Definition 2.2. *For a compact operator $Y : H \to G$ between Hilbert spaces as above the* singular value decomposition *of Y is defined by $\{\sigma_i, u_i, v_i)\}_{i\in I}$ as given by Theorem 2.1. More precisely, we call σ_i the* singular values*, u_i the* right *and v_i the* left singular vectors*.*

The POD method in a Hilbert space H can be described in terms of the singular value decomposition of a suitable linear operator. Therefore we consider the linear operator $Y : \mathbb{R}^m \to H$ defined for the given collection of snapshots $\{y_i\}_{i\leq m}$ by

$$Y(w) = \frac{1}{\sqrt{m}} \sum_{i=1}^{m} w_i y_i. \tag{6}$$

Observe that the range of Y is finite, hence Y is a compact operator and we can use the singular value decomposition of Y. The adjoint is given by $Y^* : H \to \mathbb{R}^m$ with

$$Y^*(\varphi) = \frac{1}{\sqrt{m}}(\langle \varphi, y_1\rangle, \dots, \langle \varphi, y_m\rangle)^T. \tag{7}$$

In our case the self-adjoint operator $K = Y^*Y : \mathbb{R}^m \to \mathbb{R}^m$ can be identified with the matrix

$$K = \frac{1}{m}(\langle y_j, y_i\rangle)_{ij} \in \mathbb{R}^{m,m}.$$

K is called *correlation matrix*. The right singular vectors $v_k \in \mathbb{R}^m$ defined in Theorem 2.1 are the eigenvectors of the correlation matrix. Using the fact

$$Yv_k = \sigma_k w_k,$$

the left singular vectors w_k can be expressed by

$$w_k = \frac{1}{\sqrt{m}\sigma_k} \sum_{i=1}^{m} (v_k)_i y_i \in H. \tag{8}$$

Theorem 2.3. *Let a collection $\{y_i\}_{i=1}^m$ of snapshots be given in a separable Hilbert space H. The corresponding linear operator $Y : \mathbb{R}^m \to H$ shall be given by (6). Let $d \leq m$ be the dimension of the subspace spanned by the collection of snapshots* $\operatorname{span}\{y_1, \dots, y_m\}$*. Then the POD basis of rank $\ell \leq d$ is given by the left singular vectors $\{w_k\}_{k=1}^{\ell}$ of Y.*

The approximation error is given by the singular values of Y:

$$E_{\mathrm{pod}}(\{w_k\}_{k=1}^{\ell}) = \sum_{k=\ell+1}^{d} \sigma_k^2.$$

In the finite-dimensional case $H = \mathbb{R}^N$ the operator $Y : \mathbb{R}^m \to \mathbb{R}^N$ is given by a matrix and we get the following result

Corollar 2.4. *Let a collection $\{y_i\}_{i=1}^m$ of snapshots be given in $H = \mathbb{R}^N$. Then the POD basis of rank ℓ is given by the singular vectors $w_k \in \mathbb{R}^N$ of*

$$Y = \frac{1}{\sqrt{m}} col(y_1, \dots, y_m) \in \mathbb{R}^{N,m}$$

where $\operatorname{col}(y_1, \dots, y_m)$ *denotes the matrix with columns $y_1, \dots, y_m$. The approximation error is determined by the small singular values of Y:*

$$E_{\mathrm{pod}}(\{w_k\}_{k=1}^{\ell}) = \sum_{k=\ell+1}^{d} \sigma_k^2.$$

2.2 A brief review of finite-time POD error estimates

We give a brief overview of the existing convergence theory of POD methods for parabolic problems developed by Kunisch, Volkwein et al. (see [KV01], [KV02]). We start with equation (2) defining an abstract parabolic system and consider the variational formulation of it. Assuming that A induces a V-elliptic continuous bilinear form $a : V \times V \to \mathbb{R}$, we see that for given $T > 0$ the problem transforms to

$$\begin{aligned} \frac{d}{dt}(u(t), \varphi)_H + a(u(t), \varphi) &= (f(u(t)), \varphi)_h \quad \text{for all } \varphi \in V, t \in (0, T) \\ (u(0), \chi)_H &= (u_0, \chi)_H \quad \text{for all } \chi \in H. \end{aligned} \tag{9}$$

Under proper conditions for the nonlinearity f there is a unique and continuous solution to (9) on a finite time interval (0,T), the so-called weak solution of the PDE (2). We denote this solution by $u(t) = u(t; T, u_0) \in C([0, T], V)$.

Volkwein and Kunisch consider a POD basis derived from such a trajectory $u(t)$ at given time steps $t_j = j\Delta t$, $j = 1, \ldots, m$. To achieve a better error constant, they also include the corresponding finite difference quotients

$$\bar{\partial} u(t_k) = \frac{u(t_k) - u(t_{k-1})}{\Delta t}$$

to obtain a collection of snapshots

$$\begin{aligned} y_j &= u(t_{j-1}),\ j = 1, \ldots, m+1, \\ y_{m+1+j} &= \bar{\partial} u(t_{j-1}),\ j = 1, \ldots, m+1. \end{aligned}$$

For these snapshots, let the POD basis of rank ℓ be given by $\{\psi_1, \ldots, \psi_\ell\}$ and denote the ℓ-dimensional POD space by

$$V^\ell := \operatorname{span}\{\psi_1, \ldots, \psi_\ell\}.$$

Now the reduced-order system corresponding to (2) using the backward Euler-Galerkin scheme is given by

$$\begin{aligned} (\bar{\partial} U_k, \psi)_H + a(U_k, \psi) &= (f(U_k), \psi)_H \quad \text{for all } \psi \in V^\ell, \\ (U_0, \psi)_H &= (u_0, \psi)_H \qquad \text{for all } \psi \in V^\ell. \end{aligned} \tag{10}$$

Here, $\bar{\partial} U_k$ denotes the analogue to $\bar{\partial} u(t_k)$ for the Euler sequence in $(U_k)_k$ in V^ℓ:

$$\bar{\partial} U_k = \frac{U_k - U_{k-1}}{\Delta t}.$$

Kunisch and Volkwein proved the following relation between the exact weak solution and the POD solution for a given initial value $u_0 \in H$:

Theorem 2.5. *Assume that* (9) *has a unique solution* $u \in C([0, T], V)$ *with* $u \in W^{2,2}([0, T]; H)$ *and that* $\{U_k\}_{k=0}^m$ *is the unique solution to* (10) *satisfying*

$$\max_{0 \le k \le m} \|U_k\|_H \le \tilde{C}$$

for a constant $\tilde{C} > 0$ *independent of* m. *If* f *is locally Lipschitz on* H *and* Δt *sufficiently small, then there exists a constant* $C > 0$ *independent of* ℓ, m *such that*

$$\frac{1}{m} \sum_{j=1}^{m} \|u(t_j) - U_j\|_H^2 \le C \left(\|u_0 - P^\ell u_0\|_H^2 + \sum_{k=\ell+1}^{d} \sigma_k^2 + (\Delta t)^2 \right).$$

Here, the Ritz projector $P^\ell : H \to V^\ell$ *is defined by*

$$a(P^\ell u, \psi) = a(u, \psi) \quad \textit{for all } \psi \in V^\ell$$

and σ_k *are the singular values of* Y *defined by* (6).

2.3 Perturbation theory for the singular value decomposition

We introduce the concept of canonical angles described in [SS90]. By this notion, we will compare the POD modes with the eigendirections of dynamical systems in Section 3. The perturbation result derived by Stewart and Sun in [SS90] will be the key result that we will use in the next section to derive error results for the long time behavior of the POD method.

For the definition of canonical angles we state the following theorem:

Theorem 2.6 (see [SS90]). *Let $X, Y \in \mathbb{C}^{n,\ell}$ be column orthonormal: $X^H X = Y^H Y = I_\ell$. Then it holds*

- *for $2\ell \leq n$: There are unitary $Q \in \mathbb{C}^{n,n}, U, V \in \mathbb{C}^{\ell,\ell}$ with*

$$QXU = \begin{pmatrix} I_\ell \\ 0 \\ 0 \end{pmatrix}, \quad QY_1V = \begin{pmatrix} \Gamma \\ \Sigma \\ 0 \end{pmatrix} \tag{11}$$

  and $\Gamma = \operatorname{diag}(\gamma_1, \ldots, \gamma_\ell)$, $\Sigma = \operatorname{diag}(\sigma_1, \ldots, \sigma_\ell)$, $0 \leq \gamma_1 \leq \ldots \leq \gamma_\ell$, $\sigma_1 \geq \ldots \geq \sigma_\ell \geq 0$, $\sigma_i^2 + \gamma_i^2 = 1$, $i = 1, \ldots, \ell$.
- *for $2\ell > n$: There are unitary $Q \in \mathbb{C}^{n,n}, U, V \in \mathbb{C}^{\ell,\ell}$ with*

$$QXU = \begin{pmatrix} I_{n-\ell} & 0 \\ 0 & I_{2\ell-n} \\ 0 & 0 \end{pmatrix}, \quad QY_1V = \begin{pmatrix} \Gamma & 0 \\ 0 & I_{2\ell-n} \\ \Sigma & 0 \end{pmatrix} \tag{12}$$

  and $\Gamma = \operatorname{diag}(\gamma_1, \ldots, \gamma_{n-\ell})$, $\Sigma = \operatorname{diag}(\sigma_1, \ldots, \sigma_{n-\ell})$, γ_i, σ_i as above.

This theorem allows us to define canonical angles between subspaces.

Definition 2.7. *For $X = \operatorname{col}(x_1, \ldots, x_\ell)$ and $Y = \operatorname{col}(y_1, \ldots, y_\ell)$ with orthonormal columns we define the* canonical angle *between the subspaces $R(X)$ and $R(Y)$ given by X and Y as the matrix*

$$\angle(X, Y) := \sin^{-1} \Sigma \tag{13}$$

where $\Sigma \in \mathbb{C}^{\ell,\ell}$ and $\Sigma \in \mathbb{C}^{n-\ell,n-\ell}$ are the diagonal matrices as defined in (11) and (12), respectively.

Now, let the singular value decompositions of A and $A + E \in \mathbb{C}^{m,n}$, $m \geq n$ be given by

$$\begin{pmatrix} W_1 & W_2 & W_3 \end{pmatrix}^H A \begin{pmatrix} V_1 & V_2 \end{pmatrix} = \begin{pmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \\ 0 & 0 \end{pmatrix}$$

$$\begin{pmatrix} \tilde{W}_1 & \tilde{W}_2 & \tilde{W}_3 \end{pmatrix}^H (A + E) \begin{pmatrix} \tilde{V}_1 & \tilde{V}_2 \end{pmatrix} = \begin{pmatrix} \tilde{\Sigma}_1 & 0 \\ 0 & \tilde{\Sigma}_2 \\ 0 & 0 \end{pmatrix}$$

where $\Sigma_1, \tilde{\Sigma}_1 \in \mathbb{R}^{\ell,\ell}$, $\Sigma_2, \tilde{\Sigma}_2 \in \mathbb{C}^{n-\ell,n-\ell}$, $W_1, \tilde{W}_1 \in \mathbb{C}^{m,\ell}$, $W_2, \tilde{W}_2 \in \mathbb{C}^{m,n-\ell}$, $W_3, \tilde{W}_3 \in \mathbb{C}^{m,m-n}$, $V_1, \tilde{V}_1 \in \mathbb{C}^{n,\ell}$ and $V_2, \tilde{V}_2 \in \mathbb{C}^{n,n-\ell}$.

Theorem 2.8 (see [SS90]). *Let $\alpha, \delta > 0$ be given with*

$$\min \sigma(\tilde{\Sigma}_1) \geq \alpha + \delta \text{ and } \max \sigma(\Sigma_2) \leq \alpha$$

Then we have

$$\max(\|\sin \angle(W_1, \tilde{W}_1)\|_2, \|\sin \angle(V_1, \tilde{V}_1)\|_2) \leq \frac{\max\{\|R\|_2, \|S\|_2\}}{\delta},$$

where R and S are the residuals

$$R := A\tilde{V}_1 - \tilde{W}_1\tilde{\Sigma}_1 \quad \text{and } S := A^H\tilde{W}_1 - \tilde{V}_1\tilde{\Sigma}_1.$$

The following corollary immediately follows by the well-known Theorem of Mirsky.

Corollar 2.9. *With the notation as above let the singular values be arranged in a descending order in Σ_1, Σ_2. Denote by γ the spectral gap between Σ_1 and Σ_2:*

$$\gamma := \min\{\sigma_1 - \sigma_2 : \sigma_1 \in \sigma(\Sigma_1), \sigma_2 \in \sigma(\Sigma_2)\} = \min \sigma(\Sigma_1) - \max \sigma(\Sigma_2).$$

Then it holds for $\gamma > \|E\|_2$:

$$\max(\|\sin \angle(W_1, \tilde{W}_1)\|_2, \|\sin \angle(V_1, \tilde{V}_1)\|_2) \leq \frac{\|E\|_2}{\gamma - \|E\|_2}.$$

3 Long-time behavior of POD solutions

To our knowledge, error estimates concerning POD modes exist only for finite time intervals as in Theorem 2.5.

As explained above, our work is motivated by algorithms analyzing the long time properties of dynamical systems. Therefore we analyze how aspects of the long-time behavior transfer to reduced order systems if we use POD-based model reduction. A detailed description of every possible dynamical behavior seems impossible. Hence we restrict ourselves to dynamical systems with exponentially decaying behavior. We will see that already in these cases the results are getting quite complex.

A central error bound in the convergence theory of POD methods is given by the behavior of the remaining singular values

$$\sum_{k=\ell+1}^{d} \lambda_k$$

of the operator Y given by the snapshots. We have a closer look at the time dependence of this error bound in the following cases. Therefore, we consider the ordinary differential equation

$$\begin{aligned} u_t &= f(u), \quad f : \mathbb{R}^N \to \mathbb{R}^N \\ u(0) &= u_0. \end{aligned} \tag{14}$$

Wherever possible, we will make generalizations to the case of an infinite dimensional state space.

3.1 Asymptotically stable fixed point

We assume that there exists an asymptotically stable fixed point in the ODE (14). We consider snapshots based on a trajectory converging to that fixed point. As we expect, the resulting POD basis converges to the same fixed point. In detail, the following holds:

Theorem 3.1. *Let $\bar{u}$ be an asymptotically stable fixed point of (14), i.e. $f(\bar{u}) = 0$ and*

$$\sigma(Df(\bar{u})) \subset \mathbb{C}_- := \{z \in \mathbb{C}; \mathbb{R}ez < 0\}. \tag{15}$$

Let $u(t)$ be a trajectory of (14) with initial value u_0 near $\bar{u}$ such that

$$\|u(t) - \bar{u}\|_2 \leq e^{-\alpha t}\|u_0 - \bar{u}\|_2 \quad \text{for all } t > 0. \tag{16}$$

and some $\alpha > 0$. Then the POD basis of rank 1 for the snapshots

$$y_j = u(t_j) = u(jT), \quad j = 1, \ldots, m$$

with stepsize $T \geq T_0 > 0$ is given by some $\{w_1\}$ where the error of the POD method (see Corollary 2.4) is given by

$$E_{\text{pod}}(\{w_1\}) \leq C_1 \frac{e^{-\alpha 2T}}{m} \tag{17}$$

and the angle between the POD mode and the fixed point satisfies

$$|\sin(\angle(w_1, \bar{u}))| \leq C_2 \frac{e^{-\alpha T}}{\sqrt{m}} \tag{18}$$

with C_1, C_2 independent of T, m.

Remarks 3.2.

- *In (18) we use the notion of canonical angles introduced in section 2.3. Observe that for two vectors $u, v \in \mathbb{R}^N$ the canonical angle is given explicitly by*

$$\angle(u, v) = cos^{-1}\left(\frac{u^T v}{\|u\|_2 \|v\|_2}\right).$$

- *The existence of a trajectory with (16) follows by standard theory from (15) if we choose $\alpha > 0$ such that $\mathbb{R}e\sigma(Df(\bar{u})) < -\alpha$.*

Proof. We consider the matrix $Y = \frac{1}{\sqrt{m}} \operatorname{col}(y_1, \ldots, y_m)$ and decompose it in the form

$$Y = Y_0 + E, \quad Y_0 = \frac{1}{\sqrt{m}} \operatorname{col}(\bar{u}, \ldots, \bar{u}) \in \mathbb{R}^{N,m}.$$

The singular value decomposition of Y_0 is given by

$$W_0^T Y_0 V_0 = \begin{pmatrix} \|\bar{u}\|_2 & 0 \\ 0 & 0 \end{pmatrix}$$

where $W_0 = \operatorname{col}(w_1, \ldots, w_N)$ with

$$w_1 = \frac{\bar{u}}{\|\bar{u}\|_2}.$$

The singular values $\{\sigma_i^0\}_i$ of Y_0 are given by

$$\sigma_1^0 = \|\bar{u}\|_2, \ \sigma_i^0 = 0, \quad i \geq 2.$$

Let the singular value decomposition of Y be given by

$$W^T Y V = \begin{pmatrix} \operatorname{diag}(\sigma_1, \ldots, \sigma_m) \\ 0 \end{pmatrix}.$$

Then by the Theorem of Mirsky

$$\begin{aligned}
E_{\text{pod}}(\{w_1\}) &= \sum_{i=2}^{m} \sigma_i^2 \leq (\sigma_1 - \|\bar{u}\|_2)^2 + \sum_{i=2}^{m} \sigma_i^2 \\
&= \sum_{i=1}^{m} (\sigma_i - \sigma_i^0)^2 \\
&\overset{\text{Mirsky}}{\leq} \frac{1}{m} \sum_{i=1}^{m} \|y_i - \bar{u}\|_2^2 \\
&\leq \frac{1}{m} \sum_{i=1}^{m} \left(e^{-2\alpha T}\right)^i \|u_0 - \bar{u}\|_2^2 \\
&\leq \frac{1}{m} e^{-\alpha 2T} \frac{1}{1 - e^{-2\alpha T_0}} \|u_0 - \bar{u}\|_2^2.
\end{aligned} \tag{19}$$

We use Corollary 2.9 to get the following estimate for the canonical angle between the singular vector w_1 of Y and $\bar{u}$ of Y_0:

$$|\sin\angle(w_1, \bar{u})| \leq \frac{\|E\|_2}{\gamma - \|E\|_2} \tag{20}$$

where the spectral gap γ in this example is given by

$$\gamma = \sigma_1^0 - \sigma_2^0 = \|\bar{u}\|_2.$$

For $E = Y - Y_0 = \mathrm{col}(y_1 - \bar{u}, \ldots, y_m - \bar{u})$ we use the approximation

$$\|E\|_2 \leq \|E\|_F = \left(\frac{1}{m}\sum_{i=1}^{m} \|y_i - \bar{u}\|_2^2\right)^{1/2}.$$

If we take m or T large enough such that

$$\|E\|_F \leq \frac{\|\bar{u}\|_2}{2}$$

we get

$$|\sin(\angle(w_1, \bar{u}))| \leq \frac{2\|E\|_F}{\|\bar{u}\|_2} \overset{(19)}{\leq} \frac{2C}{\sqrt{m}}\|\bar{u} - u_0\|e^{-\alpha T}\sqrt{\frac{1}{1 - e^{-2\alpha T_0}}}.$$

□

3.2 Different speeds of convergence to a fixed point

In the next theorem we deal with the situation of an asymptotically stable fixed point in detail. Therefore we have a closer look at the linearized system around a stable fixed point $\bar{u}$:

$$\begin{aligned} w_t &= Df(\bar{u})w \\ w(0) &= v_0 := u_0 - \bar{u}. \end{aligned} \tag{21}$$

Assume $A := Df(\bar{u})$ to be diagonalizable:

$$VAV^{-1} = \mathrm{diag}(\lambda_1, \ldots, \lambda_N), \quad V = \mathrm{col}(v_1, \ldots, v_N).$$

For the solution $w(t)$ of (21) with initial value $u_0 - \bar{u}$, this yields

$$w(t) = \sum_{i=1}^{N} (w_0)_i e^{-\lambda_i t} v_i$$

where $w_0 = V - 1v_0$. Assuming a gap in the spectrum $\{\lambda_i\}_i$ we analyze the behavior of the POD modes for the linearized system.

Theorem 3.3. *Let a trajectory converge to the fixed point* 0 *in the following way:*

$$u(t) = e^{-\lambda_1 t}\, v + e^{-\lambda_2 t}\, w \tag{22}$$

where $\lambda_2 = \lambda_1 + \delta$. *Then the resulting singular values controlling the POD expansion for the snapshots* $y_j = u(jT)$, $j = 1, \ldots, m$, *have a gap*

$$\sigma_1 \in \left[\left(1 - \gamma\frac{\|w\|_2}{\|v\|_2}\right)\sigma_1^0, \left(1 + \gamma\frac{\|w\|_2}{\|v\|_2}\right)\sigma_1^0\right], \tag{23}$$

$$|\sigma_2| \leq \gamma\frac{\|w\|_2}{\|v\|_2}\sigma_1^0. \tag{24}$$

with $\gamma = e^{-\delta T}$ *and* $\sigma_1^0 = \frac{1}{\sqrt{m}}\|v\|_2\|a\|_2$. *This leads to the following estimate for the angle between the first POD mode* w_1 *of* Y *and the direction of slowest attraction* v

$$|\sin(\angle(v, w_1))| \leq \frac{\gamma}{\frac{\|v\|_2}{\|w\|_2} - \gamma} \tag{25}$$

if $\frac{\|v\|_2}{\|w\|_2} > \gamma$.

Proof. As before, let $Y = \frac{1}{\sqrt{m}}\operatorname{col}(y_1, \ldots, y_m)$ be the matrix of snapshots. It can be written as a sum of two rank-1-matrices

$$Y = Y_0 + E, \quad Y_0 = \frac{1}{\sqrt{m}} v a^T, \; E = \frac{1}{\sqrt{m}} w b^T$$

where $a = (e^{-\lambda_1 T}, \ldots, e^{-\lambda_1 mT})^T$, $b = (e^{-\lambda_2 T}, \ldots, e^{-\lambda_2 mT})^T$. With $D = \operatorname{diag}(\gamma, \ldots, \gamma^m)$, $\gamma = e^{-\delta T}$, one can also write b as $b = Da$.

The singular values $\{\sigma_i^0\}_i$ of Y_0 are zero except for

$$\sigma_1^0 = \frac{1}{\sqrt{m}}\|v\|_2\|a\|_2.$$

The same holds for the singular values $\{\sigma_i^1\}_i$ of E with

$$\begin{aligned} \|E\|_2 = \sigma_1^1 &= \frac{1}{\sqrt{m}}\|w\|_2\|Da\|_2 \\ &\leq \frac{\gamma}{\sqrt{m}}\|w\|_2\|a\|_2 \quad = \gamma\frac{\|w\|_2}{\|v\|_2}\sigma_1^0. \end{aligned} \tag{26}$$

We get the following relative error bounds for the singular values of Y:

$$|\sigma_1 - \sigma_1^0|, |\sigma_2| \leq \|E\|_2 \leq \gamma\frac{\|w\|_2}{\|v\|_2}\sigma_1^0$$

which leads to (23) and (24).

Considering the estimate of the POD mode w_1 note that the singular vector of Y_0 is given by $\frac{v}{\|v\|_2}$. We use (20) to get an estimate for the angle between the first singular vector w_1 of Y and v:

$$|\sin(\angle(v, w_1))| \leq \frac{\|E\|_2}{\delta_Y - \|E\|_2}$$

with spectral gap

$$\delta_Y = \sigma_1^0 - \sigma_2^0 = \sigma_1^0 = \frac{1}{\sqrt{m}}\|v\|_2\|a\|_2$$

and the norm of E already computed in (26). Together this yields

$$|\sin(\angle(v, w_1))| \leq \frac{\gamma\frac{\|w\|_2}{\|v\|_2}\sigma_1^0}{\sigma_1^0 - \gamma\frac{\|w\|_2}{\|v\|_2}\sigma_1^0} = \frac{\gamma}{\frac{\|v\|_2}{\|w\|_2} - \gamma}.$$

□

This theorem shows that the first POD vector approximates the direction of the slower attraction quite well for a two-dimensional linear system. However, in the more realistic case of N different directions of attraction, the result is less satisfying. We will deal with this situation now.

Thinking of the situation in the linear system (21), we look for a more general result of N different directions. For this, we need a norm estimate for the inverse of a Vandermonde matrix.

Theorem 3.4 ([Gau62]). *For distinct numbers $x_i \in \mathbb{C}$, $i = 1, \ldots, n$ define the Vandermonde matrix $V = \mathrm{Vand}(\{x_i\}_{i=1}^n) \in \mathbb{C}^{n,n}$ by*

$$V = \begin{pmatrix} 1 & x_1 & \ldots & x_1^{n-1} \\ 1 & x_2 & \ldots & x_2^{n-1} \\ \vdots & \vdots & & \vdots \\ 1 & x_n & \ldots & x_n^{n-1} \end{pmatrix}$$

Then we have

$$\|V^{-1}\|_1 \leq \max_{j=1,\ldots,n} \prod_{\substack{k=1 \\ k\neq j}}^{n} \frac{1+|x_k|}{|x_k - x_j|} \tag{27}$$

where $\|\cdot\|_1$ is the usual 1-norm induced by the L_1-norm.

$$\|A\|_1 = \max_{j=1,\ldots,n} \sum_{i=1}^{n} |a_{ij}|, \ A \in \mathbb{C}^{n,n}.$$

If the x_j are located on the same ray through the origin, i.e.

$$x_j = |x_j| e^{i\varphi}$$

for a fixed $\varphi \in [0, 2\pi)$, then (27) is an equality.

Using this technical result we are able to state and prove the following main result of this section.

Theorem 3.5. *Let a trajectory of (21) be given by*

$$u(t) = \sum_{j=1}^{N} e^{-\lambda_j t} v_j \tag{28}$$

where $\{v_1, \ldots, v_N\}$ are linearly independent and

$$0 < \lambda_1 < \ldots < \lambda_\ell < \lambda_\ell + \delta = \lambda_{\ell+1} \leq \ldots \leq \lambda_N. \tag{29}$$

Collect the directions v_i in matrices $V_0 = \mathrm{col}(v_1, \ldots, v_\ell)$ and $V_1 = \mathrm{col}(v_{\ell+1}, \ldots, v_N)$. Consider snapshots $y_i = u(iT)$, $i = 1, \ldots, m$, to a given stepsize $T \geq 1/\delta$. Then the corresponding POD basis of rank ℓ given by $W_\ell = \mathrm{col}(w_1, \ldots, w_\ell)$ satisfies

$$\|\sin(\angle(V_0, W_\ell))\|_2 \leq \frac{\gamma}{M - \gamma} \tag{30}$$

if $M > \gamma := e^{-\delta T}$. Here, $M > 0$ is given by

$$M = \begin{cases} \frac{\|v_1\|_2}{\sqrt{2(N-1)}\sigma_1(V_1)}, & \ell = 1 \\ \sqrt{\frac{\ell}{2(N-\ell)}} \left(\frac{\mathrm{gap}_\alpha}{2}\right)^\ell \frac{\sigma_\ell(V_0)}{\sigma_1(V_1)}, & \ell \geq 2 \end{cases} \tag{31}$$

with $\alpha_i = e^{-\lambda_i T}$ and $\mathrm{gap}_\alpha = \min\limits_{1\leq i<j\leq \ell} |\alpha_i - \alpha_j|$.

If the eigenvalues are well-separated, i.e.

$$\lambda_j - \lambda_{j-1} \geq \delta \quad \text{for all } j = \ell + 1, \ldots, N \tag{32}$$

and $T \geq 2/\delta$, then the bound (30) holds for the larger constant

$$M = \begin{cases} \frac{\|v_1\|_2}{\sqrt{2}\sigma_1(V_1)}, & \ell = 1 \\ \sqrt{\frac{\ell}{2}} \left(\frac{\mathrm{gap}_\alpha}{2}\right)^\ell \frac{\sigma_\ell(V_0)}{\sigma_1(V_1)}, & \ell \geq 2. \end{cases} \tag{33}$$

Proof. As before we define the matrix of snapshots $Y = \frac{1}{\sqrt{m}} \operatorname{col}(y_1, \dots, y_m)$. Since we have

$$y_i = u(iT) = \sum_{j=1}^{N} e^{-\lambda_j iT} v_j, \quad i = 1, \dots, m,$$

we can write Y as

$$Y = \frac{1}{\sqrt{m}} V A^T, \ A = (\alpha_j^i)_{ij} \in \mathbb{R}^{m,N}.$$

As before, we describe the POD space of rank ℓ as a perturbation of the first ℓ eigendirections v_i, $i = 1, \dots, \ell$. The POD space $W_\ell = \operatorname{col}(w_1, \dots, w_\ell)$ is given by the singular value decomposition of Y:

$$W^T Y U = \begin{pmatrix} \operatorname{diag}(\sigma_1, \dots, \sigma_N) & 0 \end{pmatrix}$$

with $W = \operatorname{col}(w_1, \dots, w_N) \in \mathbb{R}^{N,N}$, $U \in \mathbb{R}^{m,m}$ orthogonal, $\sigma_1 \geq \dots \geq \sigma_N \geq 0$.

We decompose Y as

$$Y = Y_0 + E, \ Y_0 = \frac{1}{\sqrt{m}} V_0 A_0^T, \ E = \frac{1}{\sqrt{m}} V_1 A_1^T$$

with $A = \begin{pmatrix} A_0 & A_1 \end{pmatrix}$, $V = \begin{pmatrix} V_0 & V_1 \end{pmatrix}$, $A_0 = (\alpha_j^i)_{ij} \in \mathbb{R}^{m,\ell}$, $V_0 = \operatorname{col}(v_1, \dots, v_\ell) \in \mathbb{R}^{N,\ell}$. The singular value decomposition of the rank-ℓ matrix Y_0 is given by

$$W_0^T Y_0 U_0 = D := \begin{pmatrix} \operatorname{diag}(\sigma_1^0, \dots, \sigma_\ell^0) & 0 \\ 0 & 0 \end{pmatrix} \in \mathbb{R}^{m,N} \tag{34}$$

with $W_0 = \operatorname{col}(w_1^0, \dots, w_N^0) \in \mathbb{R}^{N,N}$, $U_0 \in \mathbb{R}^{m,m}$ orthogonal and $\sigma_1^0 \geq \dots \geq \sigma_\ell^0 > 0$. Since $Y_0 = \frac{1}{\sqrt{m}} V_0 A_0^T$, we get $R(Y_0) = R(V_0)$. On the other hand by (34) we have

$$Y_0 U_0 = W_0 D_0 = \operatorname{col}(\sigma_1 w_1^0, \dots, \sigma_\ell w_\ell^0, 0, \dots, 0)$$

and hence $R(V_0) = R(Y_0) = \operatorname{span}\{w_1^0, \dots, w_\ell^0\}$. As before, we can use (20) to get an estimate for the angle between the POD basis given by $W_\ell \in \mathbb{R}^{N,\ell}$ and the eigendirections collected in $V_0 \in \mathbb{R}^{N,\ell}$:

$$\| \sin(\angle(V_0, W_\ell)) \|_2 \leq \frac{\|E\|_2}{\delta_Y - \|E\|_2}$$

where $\delta_Y = \sigma_\ell^0 - \sigma_{\ell+1}^0 = \sigma_\ell^0 = \sigma_\ell(Y_0)$.

In this notation, we can estimate the canonical angle between the eigendirections and the POD modes in terms of singular values of Y_0 and E:

$$\| \sin(\angle(V_0, W_\ell)) \|_2 \leq \frac{\sigma_1(E)}{\sigma_\ell(Y_0) - \sigma_1(E)}.$$

We will discuss these values in the following.

1. $\sigma_\ell(Y_0)$: We use the characterization $Y_0 = \frac{1}{\sqrt{m}} V_0 A_0^T$ to get a lower bound for the smallest nonzero singular value $\sigma_\ell(Y_0)$. We use the submultiplicativity of the singular values (see [SS90], Theorem I.4.5 and Exercise I.4.6) to get

   $$\sigma_\ell(Y_0) \geq \frac{1}{\sqrt{m}} \sigma_\ell(V_0) \sigma_\ell(A_0).$$

   For $\ell = 1$, the matrix $A_0 \in \mathbb{R}^{m,\ell}$ is given by a column vector such that

   $$\sigma_\ell(A_0) = \sigma_1(A_0) = \|A_0\|_2 = \|(\alpha_1, \alpha_1^2, \dots, \alpha_1^m)^T\|_2 \geq \alpha_1$$

and with $\sigma_\ell(V_0) = \sigma_1(V_0) = \|v_1\|_2$, we get

$$\sigma_\ell(Y_0) \geq \frac{1}{\sqrt{m}} \alpha_1 \|v_1\|_2. \tag{35}$$

For $\ell \geq 2$ we decompose $A_0 = \begin{pmatrix} A_0^{(1)} \\ A_0^{(2)} \end{pmatrix}$, $A_0^{(1)} \in \mathbb{R}^{\ell,\ell}$, and get a lower bound by the interlacing property of singular values (see [GvL96]):

$$\sigma_\ell(A_0) = \sigma_\ell(A_0^T) = \sigma_\ell(\begin{pmatrix} {A_0^{(1)}}^T & {A_0^{(2)}}^T \end{pmatrix}) \geq \sigma_\ell(A_0^{(1)}).$$

$A_0^{(1)} \in \mathbb{R}^{\ell,\ell}$ can be expressed as

$${A_0^{(1)}}^T = V_\alpha \operatorname{diag}(\alpha_1, \ldots, \alpha_\ell)$$

where $V_\alpha = \operatorname{Vand}(\{\alpha_j\}_{j=1}^\ell)$ is the Vandermonde-Matrix as defined in Theorem 3.4. Together with the submultiplicativity of the singular values we get

$$\begin{aligned} \sigma_\ell(A_0) \geq \sigma_\ell(A_0^{(1)}) \geq \alpha_\ell \sigma_\ell(V_\alpha) = \alpha_\ell \sigma_1(V_\alpha^{-1})^{-1} = \alpha_\ell \, \|V_\alpha^{-1}\|_2^{-1} \geq \alpha_\ell \sqrt{\ell} \, \|V_\alpha^{-1}\|_1^{-1} \\ = \alpha_\ell \sqrt{\ell} \left(\max_{j=1,\ldots,\ell} \prod_{\substack{k=1 \\ k \neq j}}^{\ell} \frac{1+\alpha_k}{|\alpha_k - \alpha_j|} \right)^{-1} \geq \alpha_\ell \sqrt{\ell} \left(\frac{\operatorname{gap}_\alpha}{2} \right)^\ell \end{aligned}$$

since $1 + \alpha_k \leq 2$ and by definition $|\alpha_k - \alpha_j| \geq \operatorname{gap}_\alpha$ for all $k, j \in \{1, \ldots, \ell\}$, $k \neq j$. Another application of the submultiplicativity of the singular values implies

$$\sigma_\ell(Y_0) = \sigma_\ell(\frac{1}{\sqrt{m}} V_0 A_0^T) \geq \alpha_\ell \frac{\sqrt{\ell}}{\sqrt{m}} \left(\frac{\operatorname{gap}_\alpha}{2} \right)^\ell \sigma_\ell(V_0). \tag{36}$$

2. $\|E\|_2 = \sigma_1(E)$: To approximate the first singular value of $E = \frac{1}{\sqrt{m}} V_1 A_1^T$ we use the submultiplicativity of the singular values to get

$$\sigma_1(E) = \sigma_1(\frac{1}{\sqrt{m}} V_1 A_1^T) \leq \frac{1}{\sqrt{m}} \sigma_1(V_1) \sigma_1(A_1^T).$$

For the largest singular value of A_1, recall the structure of the matrix

$$A_1 = (\alpha_j^i)_{ij} \in \mathbb{R}^{m, N-\ell}$$

with $\alpha_j = e^{-\lambda_j T} \in (0,1)$, $j = \ell+1, \ldots, N$ ordered by $\alpha_{\ell+1} \geq \ldots \geq \alpha_N$. We use the well-known matrix norm inequality $\|B\|_2^2 \leq \|B\|_1 \|B\|_\infty$ (see e.g. [GvL96], Corollary 2.3.2) and get

$$\begin{aligned} \sigma_1(A_1)^2 = \|A_1\|_2^2 &\leq \|A_1\|_1 \, \|A_1\|_\infty \qquad (37) \\ &= \left(\max_{j=\ell+1,\ldots,N} \sum_{i=1}^m \alpha_j^i \right) \left(\max_{i=1,\ldots,m} \sum_{j=\ell+1}^N \alpha_j^i \right) \\ &= \left(\sum_{i=1}^m \alpha_{\ell+1}^i \right) \left(\sum_{j=\ell+1}^N \alpha_j \right). \qquad (38) \end{aligned}$$

By assumption γ is bounded by

$$\gamma = e^{-\delta T} < 2^{-\delta T} \leq 1/2$$

and the first sum can be approximated by

$$\begin{aligned}\sum_{i=1}^{m}\alpha_{\ell+1}^{i} &= \alpha_{\ell+1}\frac{1-\alpha_{\ell+1}^{m}}{1-\alpha_{\ell+1}} \le \alpha_{\ell+1}\frac{1}{1-\alpha_{\ell+1}} \\ &= \gamma\alpha_\ell(1-\gamma\alpha_\ell)^{-1} \le \gamma\alpha_\ell(1-\gamma)^{-1} \qquad (39)\\ &\le 2\gamma\alpha_\ell. \qquad (40)\end{aligned}$$

For the second sum in (38) a weak estimate is always given by the monotonicity of α_j, $j = \ell+1,\dots,N$:

$$\sum_{j=\ell+1}^{N}\alpha_j \le \sum_{j=\ell+1}^{N}\alpha_{\ell+1} = (N-\ell)\alpha_{\ell+1} = (N-\ell)\,\gamma\,\alpha_\ell. \qquad (41)$$

Combining (40), (41) and (38), we get the following bound for $\sigma_1(E)$:

$$\sigma_1(E) \le \frac{1}{\sqrt{m}}\,\sigma_1(A_1)\sigma_1(V_1) \le \sqrt{\frac{2(N-\ell)}{m}}\gamma\,\alpha_\ell\,\sigma_1(V_1). \qquad (42)$$

If we have $\lambda_j - \lambda_{j-1} \ge \delta$, $j = \ell+1,\dots,N$, we get a bound for the second sum by

$$\begin{aligned}\sum_{j=\ell+1}^{N}\alpha_j &= \alpha_{\ell+1} + \sum_{j=\ell+2}^{N} e^{-\lambda_j T} = \alpha_{\ell+1} + \sum_{j=\ell+2}^{N} e^{-\frac{\lambda_j}{\delta}\delta T}\cdot 1 \\ &\le \alpha_{\ell+1} + \sum_{j=\ell+2}^{N} e^{-\frac{\lambda_j}{\delta}\delta T}\left(\frac{\lambda_j}{\delta} - \frac{\lambda_{j-1}}{\delta}\right) \\ &\le \alpha_{\ell+1} + \int_{\frac{\lambda_{\ell+1}}{\delta}}^{\frac{\lambda_N}{\delta}} e^{-s\delta T}\,ds = \alpha_{\ell+1} + \frac{1}{\delta T}\left(e^{-\lambda_{\ell+1}T} - e^{\lambda_N T}\right) \\ &\le \alpha_{\ell+1} + \frac{\alpha_{\ell+1}}{\delta T} = \gamma\alpha_\ell\left(1+\frac{1}{\delta T}\right). \qquad (43)\end{aligned}$$

With the assumption $\delta T \ge 2$ and by combining (39), (43) and (38), we obtain

$$\begin{aligned}\sigma_1(A_1)^2 &\le \gamma\alpha_\ell(1-\gamma)^{-1}\gamma\alpha_\ell\left(1+\frac{1}{\delta T}\right) = \gamma^2\alpha_\ell^2(1-e^{-\delta T})^{-1}\left(1+\frac{1}{\delta T}\right) \\ &\le \gamma^2\alpha_\ell^2(1-2^{-2})^{-1}\left(1+\frac{1}{2}\right) \;=\; 2\gamma^2\alpha_\ell^2. \qquad (44)\end{aligned}$$

By the submultiplicativity of the singular values we have

$$\sigma_1(E) \le \frac{1}{\sqrt{m}}\,\sigma_1(A_1)\sigma_1(V_1) \le \sqrt{\frac{2}{m}}\,\gamma\,\alpha_\ell\,\sigma_1(V_1). \qquad (45)$$

Together for $\ell = 1$, it follows

$$\begin{aligned}\|\sin(\angle(v_1,w_1))\|_2 &\le \frac{\sigma_1(E)}{\sigma_1(Y_0)-\sigma_1(E)} \\ &\overset{(35)}{\underset{(42)}{\le}} \frac{\sqrt{2(N-1)}\sigma_1(V_1)\gamma}{\|v_1\|_2 - \sqrt{2(N-1)}\sigma_1(V_1)\gamma} = \frac{\gamma}{\frac{\|v_1\|_2}{\sqrt{2(N-1)}\sigma_1(V_1)} - \gamma}\end{aligned}$$

provided the right hand side is positive.

For $\ell \geq 2$, the estimates (36), (42) imply

$$\begin{aligned} \|\sin(\angle(V_0, W_\ell))\|_2 &\leq \frac{\sigma_1(E)}{\sigma_\ell(Y_0) - \sigma_1(E)} \\ &\overset{(36)}{\underset{(42)}{\leq}} \frac{\sqrt{2(N-\ell)}\sigma_1(V_1)\gamma}{\sqrt{\ell}\left(\frac{\mathrm{gap}_\alpha}{2}\right)^\ell \sigma_\ell(V_0) - \sqrt{2(N-\ell)}\sigma_1(V_1)\gamma} \\ &= \frac{\gamma}{\sqrt{\frac{\ell}{2(N-\ell)}}\left(\frac{\mathrm{gap}_\alpha}{2}\right)^\ell \frac{\sigma_\ell(V_0)}{\sigma_1(V_1)} - \gamma} \end{aligned}$$

provided the right hand side is positive. Similar bounds (but independent on N) hold for the case of separated eigenvalues as noted in (33). In that case we use (45) instead of (42). □

Remark 3.6. *The 2-speed case analyzed in Theorem 3.3 satisfies the assumptions of Theorem 3.5 with $\ell = 1$, $N = 2$. Hence, equation* (30) *should give the same estimate as* (25)*. However, the estimates differ by a factor of $\sqrt{2}$:*

$$\angle(v, w_1) \overset{(30)}{\leq} \frac{\gamma}{M - \gamma} \overset{(33)}{\leq} \frac{\gamma}{\frac{\|v\|_2}{\sqrt{2}\,\|w_1\|_2} - \gamma}.$$

This is one of the reasons why the 2-speed case was done separately.

Observe that $\gamma = \gamma(T)$ is antitone in T. Hence at least for fixed dimensions of the state space, the right hand side tends to zero if the time interval T between the snapshots is increasing—as in the 2-speed case.

In Theorem 3.5 we state two different error bounds for the canonical angle. In the case of well/separated eigenvalues, i.e. (32) holds, the error bound is independent on the state space dimension N as we would expect from the theory of discretization schemes. Fortunately, in most numerical applications the eigenvalues decay very fast, so that the stronger assumption (32) holds. Additionally, the higher modes v_j, $j \gg 1$, usually also decay in norm.

However, without knowledge of the distribution of the eigenvalues of higher magnitude, we only get an error bound that grows with increasing state space dimension. We cannot expect a better result, i.e. a result independent of the state space dimension, as we will see in the next theorem. There we consider a 'worst-case scenario' where the eigenvalues of higher absolute value collapse to a multiple eigenvalue and the modes are orthonormal. In this scenario, the error grows indeed with the state space dimension N. In the limit, the POD vector is orthogonal to the first eigendirection, i.e. the direction of slowest attraction is not detected by the POD method.

Proposition 3.7. *Let a trajectory of* (21) *be given by*

$$u(t) = \sum_{j=1}^{N} e^{-\lambda_j t}\, v_j \tag{46}$$

with $V = \mathrm{col}(v_1, \ldots, v_N) \in \mathbb{R}^{N,N}$ orthogonal and eigenvalues

$$0 < \lambda_1 < \lambda_2 = \lambda_1 + \delta = \lambda_3 = \ldots = \lambda_N. \tag{47}$$

As before, we take snapshots along the trajectory denoted by $y_i = u(iT)$, $i = 1, \ldots, m$. Let the POD basis of rank 1 be given by the vector $w_1 \in \mathbb{R}^N$. Then the angle between v_1 and w_1 increases at least with the square root of the space dimension N:

$$\left|\sin\left(\frac{\pi}{2} - \angle(v_1, w_1)\right)\right| = O(N^{-1/2}). \tag{48}$$

Proof. We write the matrix of snapshots again as

$$Y = \frac{1}{\sqrt{m}} V A^T,\ A = (\alpha_j^i)_{ij} \in \mathbb{R}^{m,N}$$

with $\alpha_j = e^{-\lambda_j T}$. In this special case A is a rank-2 matrix of the following form

$$A = \mathrm{col}(a, Da, \ldots, Da)$$

where $D = \mathrm{diag}(\gamma, \ldots, \gamma^m)$, $\gamma = e^{-\delta T}$, and $a = (\alpha_1, \ldots, \alpha_1^m)$.

Observe that, since V is orthogonal, the following holds for the first singular vector w_1 of Y:

Let the singular value decomposition of A^T be given by

$$W_A^T A^T U_A = D_A = \mathrm{diag}(\sigma_1, \sigma_2, 0, \ldots, 0)$$

with $W_A \in \mathbb{R}^{N,N}$, $U_A \in \mathbb{R}^{m,m}$. Now, since V is orthogonal the singular value decomposition of $Y = \frac{1}{\sqrt{m}} V A^T$ can be expressed via the decomposition of A^T:

$$(VW_A)^T Y U_A = \frac{1}{\sqrt{m}} D_A.$$

Hence the first left singular vector w_1 of Y is just given by

$$w_1 = V w_A$$

where w_A is the first column of W_A. Now we consider A as a perturbation of the rank-1 matrix $\bar{A} = \mathrm{col}(Da, \ldots, Da)$. An easy calculation shows that $\overline{\sigma} = \sqrt{N}\|Da\|_2$ is the nontrivial singular value of $\bar{A}^T$ with left and right singular vector

$$\overline{w} = \frac{1}{\sqrt{N}} \mathbb{1}, \quad \overline{u} = \frac{Da}{\|Da\|_2}.$$

As before $V\overline{w}$ is the first left singular vector of $\bar{Y} = \frac{1}{\sqrt{m}} V \bar{A}^T$. Observe that

$$\gamma^{-m}(Da)_i = \left(\frac{1}{\gamma}\right)^{m-i} \alpha_1^i \geq \alpha_1^i \geq (1-\gamma_i)\alpha_1^i = (a - Da)_i \quad \text{for all } 1 \leq i \leq m. \tag{49}$$

For $N \geq N_0 := 4\gamma^{-2m}$ and the angle $\beta := \angle(\overline{w}, w_A)$, we get

$$\begin{aligned} |\sin(\angle(w_1, V\overline{w}))| = |\sin(\beta)| &\leq \frac{\|\bar{A} - A\|_2}{\bar{\sigma}_1 - \|\bar{A} - A\|_2} \\ &= \frac{\|(I-D)a\|_2}{\sqrt{N}\|Da\|_2 - \|(I-D)a\|_2} \\ &\overset{(49)}{\leq} \frac{\gamma^{-m}}{\sqrt{N} - \gamma^{-m}} \leq \frac{2\gamma^{-m}}{\sqrt{N}} \quad \text{for all } N \geq N_0. \end{aligned} \tag{50}$$

Applying the cosine theorem and addition theorems to

$$\angle(v_1, w_1) = \angle(Ve_1, Vw_A) = \angle(e_1, w_A),$$

we get

$$\begin{aligned} |\cos(\angle(v_1, w_1))| - \frac{1}{\sqrt{N}} \leq |\cos(\angle(v_1, w_1)) - \frac{1}{\sqrt{N}}| &= |e_1^T w_A - e_1^T \overline{w}| \\ &= |e_1^T (w_A - \overline{w})| \leq \|e_1\|_2 \|w_A - \overline{w}\|_2 \\ &= \|w_A - \overline{w}\|_2 = \sqrt{(w_A - \overline{w})^T (w_A - \overline{w})} \\ &= \sqrt{2 - 2w_A^T \overline{w}} = \sqrt{2(\cos(0) - \cos(\beta))} \\ &= \sqrt{4\sin^2\left(\frac{\beta}{2}\right)} = 2\sin\left(\frac{\beta}{2}\right). \end{aligned}$$

Since $\sin(h) = h + O(h^3)$ there exists an $N_1 > N_0$ such that the statement follows by (50):

$$|\sin\Big(\frac{\pi}{2} - \angle(v_1, w_1)\Big)| = |\cos(\angle(v_1, w_1))| \leq 2\sin\Big(\frac{\beta}{2}\Big) + \frac{1}{N} \leq \frac{C}{\sqrt{N}} \quad \text{for all } N \geq N_1.$$

□

Remark 3.8. *We note that the special type of trajectories given by* (46) *and* (47) *can also be considered as the 2-speed case of Theorem 3.3 by setting* $v = v_1$ *and* $w = \sum_{j=2}^{N} v_j$ *in* (22)*. This formulation shows that the setting of Proposition 3.7 is the 'worst case' not only considering the distribution of the eigenvalues but also considering the norms of the directions* v *and* w *as above:*

$$1 = \|v\|_2 \ll \|w\|_2 = \Big(\sum_{i=2}^{N} \|v_i\|_2\Big)^{-\frac{1}{2}} = \sqrt{N-1}$$

for large $N \in \mathbb{N}$*. In this example, a reasonable error bound as in Theorem 3.3 can only be achieved for a large integration time* T*. If* T *is chosen too small, the condition* $\frac{\|v\|_2}{\|w\|_2} > \gamma$ *of Theorem 3.3 is not violated:*

$$\frac{1}{\sqrt{N-1}} = \frac{\|v\|_2}{\|w\|_2} > \gamma = e^{-\delta T} \iff T > \frac{\ln\sqrt{N-1}}{\delta}.$$

Indeed, the experiments in Section 4.2 show that the POD method does not detect the first eigendirection in this worst-case scenario.

3.3 Attracting invariant subspace

As a generalization of the situation with an asymptotically stable fixed point we consider a positive invariant subspace that attracts the trajectory defining the snapshots exponentially. This generalization is motivated by the theory of inertial manifolds: We just recall here that one can show for a parabolic equation (2) that a proper spectral gap for A implies the so-called strong squeezing property. By that the existence of an inertial manifold $\mathcal{M}$ follows, i.e. $\mathcal{M}$ is a finite dimensional Lipschitz manifold which is positively invariant and attracts all trajectories exponentially. For details we refer to [Tem97] and [Rob01].

We analyze the case of an attracting invariant subspace as a special case of such an inertial manifold. Roughly speaking the behavior of the POD modes transfers from the case of an asymptotically stable fixed point. Note that we have to make additional regularity assumptions. This is due to the fact that the POD algorithm has too many degrees of freedom if the trajectory does not exhaust the subspace.

Theorem 3.9. *Let* $\mathcal{Z} = \operatorname{span}\{z_1, \ldots, z_\ell\} \subset \mathbb{R}^N$ *be an* ℓ*-dimensional subspace of* $\mathbb{R}^N$ *that is positive invariant under* (14) *and exponentially attracts a trajectory* $u(t)$*:*

$$\operatorname{dist}(u(t), \mathcal{Z}) \leq Ce^{-\alpha t} \operatorname{dist}(u_0, \mathcal{Z}). \tag{51}$$

We choose the snapshots $y_j = u(jT)$*,* $j = 1, \ldots, m$ *along this trajectory. If we assume the trajectory to be regular in the sense that*

$$\operatorname{rank}(ZZ^TY) = \ell \tag{52}$$

with $Z = \operatorname{col}(z_1, \ldots, z_\ell)$*,* $Y = \operatorname{col}(y_1, \ldots, y_m)$*, then for the POD space* $\operatorname{span}\{w_1, \ldots, w_\ell\}$ *of rank* ℓ*, we get the estimate*

$$E_{\text{pod}}(\{w_i\}_i) \leq C_1 \frac{e^{-\alpha 2T}}{m} \operatorname{dist}^2(u_0, \mathcal{Z}).$$

If the trajectory is essential for the subspace $\mathcal{Z}$ *in the sense that*

$$\sigma_\ell(ZZ^TY) \geq C > 0 \tag{53}$$

with C independent of T, m, then in addition we have

$$\|\sin(\angle(Z,W)\|_2 \leq C_2 \frac{e^{-\alpha T}}{\sqrt{m}} \operatorname{dist}(u_0, \mathcal{Z}).$$

As above $C_1, C_2 > 0$ are independent of T, m.

Proof. Consider the orthogonal projection $P_{\mathcal{Z}} : \mathbb{R}^N \to \mathbb{R}^N$ onto $\mathcal{Z}$ that can be expressed by

$$P_{\mathcal{Z}} v = ZZ^T v \in \mathcal{Z}.$$

We can express the Hausdorff distance by

$$\operatorname{dist}^2(u(t), \mathcal{Z}) = \|u(t) - P_{\mathcal{Z}} u(t)\|_2^2 = \|(I - ZZ^T)u(t)\|_2^2.$$

As before, we collect the snapshots $y_j = u(jT)$ along the trajectory u(t) in a matrix $Y = \frac{1}{\sqrt{m}} \operatorname{col}(y_1, \ldots, y_m) \in \mathbb{R}^{N,m}$. We can decompose Y by

$$Y = Y_0 + E,\ Y_0 = ZZ^T Y,\ E = (I - ZZ^T)Y.$$

We have the following singular value decomposition of Y_0

$$W_0^T Y_0 V_0 = \begin{pmatrix} \operatorname{diag}(\sigma_1^0, \ldots, \sigma_\ell^0) & 0 \\ 0 & 0 \end{pmatrix}$$

with positive singular values $\sigma_1^0 \geq \ldots \geq \sigma_\ell^0 > 0$ by assumption (52). By the Theorem of Mirsky, we get an estimate for the singular values $\sigma_1 \geq \ldots \geq \sigma_d > 0$ of Y:

$$\begin{aligned}
\sum_{k=\ell+1}^{d} \sigma_k^2 &\leq \sum_{k=1}^{d} (\sigma_k - \sigma_k^0)^2 \\
&\overset{\text{Mirsky}}{\leq} \frac{1}{m} \sum_{j=1}^{m} \|(I - ZZ^T)y_j\|_2^2 = \frac{1}{m} \sum_{j=1}^{m} \operatorname{dist}^2(y_j, \mathcal{Z}) \\
&\leq \frac{C}{m} \operatorname{dist}^2(u_0, \mathcal{Z}) \sum_{j=1}^{m} e^{-2jT} \\
&\leq \frac{e^{-2\alpha T}}{m} C \frac{1}{1 - e^{-2\alpha T_0}} \operatorname{dist}^2(u_0, \mathcal{Z}).
\end{aligned}$$

For the singular vectors observe that we need a spectral gap to get an estimate for the canonical angles. If the assumption (53) holds, the spectral gap is just given by $C > 0$ and we can use (20) as before to get the result. If we take m or T large enough such that

$$\|E\|_2 \leq \|E\|_F \leq \frac{1}{2} C$$

we get

$$\|\sin(\angle(Z,W))\|_2 \leq \frac{2\|E\|_2}{C} \leq C_2 \frac{e^{-\alpha T}}{\sqrt{m}} \operatorname{dist}(u_0, \mathcal{Z}).$$

□

3.4 Snapshots from different trajectories

As stated in the introduction it is a reasonable approach for systems with more complicated behavior to derive the POD data from more than one trajectory. Therefore we end this section with a first observation of how the POD method behaves if more than one trajectory is used to define snapshots. The next theorem shows that two shadowing trajectories essentially do not generate more information than a single trajectory in the sense that the resulting POD modes are close to each other.

Theorem 3.10. *Consider trajectories $u(t;u_0)$ and $u(t;u_1)$ to different initial values u_0 and u_1 with $\|u_0-u_1\|_2\le\varepsilon$. Assume these trajectories are shadowing each other, i.e.*

$$\|u(t;u_0)-u(t;u_1)\|_2\le\varepsilon,\quad t\ge 0.$$

As before, let a collection of snapshots be given by

$$y_j=u(t_j;u_0)=u(jT;u_0),\quad j=1,\dots,m.$$

Assume a spectral gap $\delta>0$ for the singular values $\{\sigma_i\}_i$ of $Y=\frac{1}{\sqrt{m}}\operatorname{col}(y_1,\dots,y_m)$:

$$\sigma_\ell-\sigma_{\ell+1}=:\delta>0.$$

Another collection of snapshots is built from both trajectories:

$$x_j=y_j=u(jT;u_0),\ x_{m+j}=z_j=u(jT;u_1),\quad j=1,\dots,m.$$

Let $\mathcal{W}^Y$ be the POD space of rank ℓ for the snapshots $\{y_j\}_{j=1}^m$ and let $\mathcal{W}^X$ be the POD space for the snapshots $\{x_j\}_{j=1}^{2m}$. Then,

$$\|\sin(\angle(\mathcal{W}^Y,\mathcal{W}^X))\|_2\le\frac{\varepsilon}{\sqrt{2}\delta-\varepsilon}\le\sqrt{2}\frac{\varepsilon}{\delta}$$

where the second estimate holds for $\varepsilon<\frac{\delta}{\sqrt{2}}$.

Proof. Let the singular value decomposition of $Y=\frac{1}{\sqrt{m}}\operatorname{col}(y_1,\dots,y_m)$ be given by

$$W^TYV=D:=\begin{pmatrix}\operatorname{diag}(\sigma_1,\dots,\sigma_N) & 0\end{pmatrix}$$

with $\sigma_1\ge\ldots\ge\sigma_N\ge 0$. Observe that for the matrix

$$Y_2=\frac{1}{\sqrt{2m}}\operatorname{col}(y_1,\dots,y_m,y_1,\dots,y_m)=\frac{1}{\sqrt{2}}\begin{pmatrix}Y & Y\end{pmatrix}$$

we get the same singular values as for Y. A short calculation shows that the matrix

$$V_2:=\frac{1}{\sqrt{2}}\begin{pmatrix}V & V\\ V & -V\end{pmatrix}\in\mathbb{R}^{2m,2m}$$

is orthogonal. With $W\in\mathbb{R}^{N,N},D\in\mathbb{R}^{N,m}$ as above we get

$$\begin{aligned}W^TY_2V_2&=\frac{1}{2}W^T\begin{pmatrix}Y & Y\end{pmatrix}\begin{pmatrix}V & V\\ V & -V\end{pmatrix}\\&=\frac{1}{2}\begin{pmatrix}W^TY & W^TY\end{pmatrix}\begin{pmatrix}V & V\\ V & -V\end{pmatrix}=\frac{1}{2}\begin{pmatrix}2W^TYV & 0\end{pmatrix}=\begin{pmatrix}D & 0\end{pmatrix}.\end{aligned}$$

If we define $X=\frac{1}{\sqrt{2m}}\operatorname{col}(y_1,\dots,y_m,z_1,\dots,z_m)$, we get the norm-wise estimate

$$\|X-Y_2\|_2\le\|X-Y_2\|_F=\frac{1}{\sqrt{2m}}\left(\sum_{j=1}^m\|y_j-z_j\|_2^2\right)^{1/2}\le\frac{\varepsilon}{\sqrt{2}}.$$

In this way we can compare the POD space $\mathcal{W}^Y$ of rank $\ell \leq m$ associated to the snapshots $\{y_i\}_i$ along one trajectory with the POD space $\mathcal{W}^X$ associated to the snapshots $\{y_i, z_i\}_i$ along two trajectories and get with (20):

$$\|\sin(\angle(\mathcal{W}^Y, \mathcal{W}^X))\|_2 \leq \frac{\|X - Y_2\|_2}{\delta - \|X - Y_2\|_2} \leq \frac{\varepsilon}{\sqrt{2}\delta - \varepsilon}.$$

In the case $\varepsilon < \frac{\delta}{\sqrt{2}}$ we can simplify the denominator to find

$$\|\sin(\angle(\mathcal{W}^Y, \mathcal{W}^X))\|_2 \leq \sqrt{2}\frac{\varepsilon}{\delta}.$$

□

This result gives a first insight into the behavior of the POD method for data drawn from more than one trajectory. As mentioned in the introduction, numerical experiments show a trade-off between the number of trajectories and the length of trajectories in the approximation behavior of the POD method. We will give numerical results indicating such a trade-off in Section 4.3. It is a hard and still open problem to confirm this observation by analytic means.

4 Numerical examples

4.1 Different speeds of convergence: Dependency on N

In Theorem 3.5, we have stated that the POD error bound in the diagonalizable case depends on N. In our first numerical example we analyze this dependency for the worst-case scenario treated in Proposition 3.7. For this, we consider a trajectory given by (46) and (47) with different gaps in the spectrum and plot the angle between the first POD mode w_1 and the first eigendirection v_1 against the dimension of the system.

In detail, we perform a numerical test with the following data: We consider the space $\mathbb{R}^{1\,000}$ with a randomly chosen orthonormal basis $\{v_i\}_{i \leq 1\,000}$. We build the trajectory $u(t) \in \mathbb{R}^{1\,000}$ according to (46) for a given $N \in [2, 1\,000]$ by

$$u(t) = e^{-\lambda_1 t} v_1 + e^{-\lambda_2 t} \sum_{j=2}^{N} v_j, \quad \lambda_1 < \lambda_2.$$

We use time steps $t_i = iT$, $i = 1, \ldots, m$, with $T = 10$ and $m = 100$ for the collection of snapshots $y_i = u(t_i)$, $i = 1, \ldots, m$. We fix m and T and analyze the dependency on N treated in Proposition 3.7.

The result is given by the two plots in Figure 1 which confirm the theoretical result of Proposition 3.7. In the left plot, we see that the angle between the first POD mode and the first eigendirection grows up to the value $\frac{\pi}{2}$, i.e. in the end no information about the first eigendirection is contained in the first POD mode.

In the right plot, the angle is scaled according to (48). We see that, at least for the smaller gaps $\delta = 0.1$ and $\delta = 0.2$, the angle behaves just as the estimate of Theorem 3.7 suggests and the constant can be estimated from the plot. For $\delta = 0.3$ the limit is not reached within the observation interval $1 \leq N \leq 1\,000$ but still we can predict the convergence from the plot.

4.2 Linear parabolic problem: Different speeds of convergence

In a second example we analyze the POD behavior of a linear reaction-diffusion equation. Later on we will focus on nonlinear reaction-diffusion equations, especially the Chafee-Infante problem.

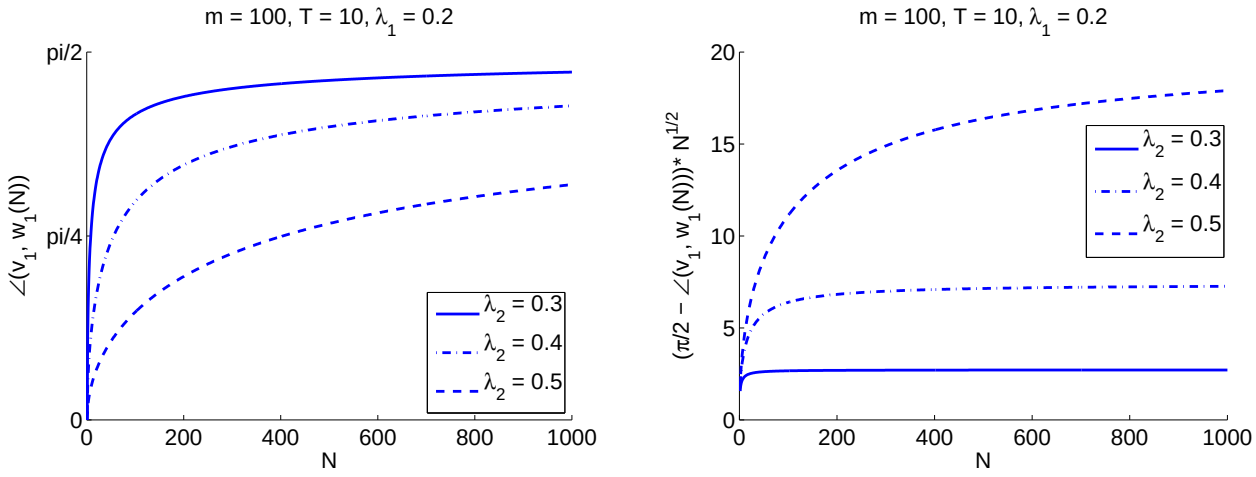

Figure 1: Angle between first eigenvector and POD-vector in a 'worst-case' test system. Different values of λ_2 where $\lambda_1 = 0.2$ is fixed.

This example gives a first insight into parabolic problems. It shows that the 'worst case' considered above is not typical in natural examples. We start with the scalar parabolic problem

$$\begin{aligned} u_t &= u_{xx} + \mu u, \qquad t \geq 0, x \in (0,1) \\ u(0,t) &= u(1,t) = 0, \quad t \geq 0 \\ u(x,0) &= u_0(x), \qquad x \in [0,1] \end{aligned} \tag{54}$$

with parameter $\mu > 0$.

Discretization by FE method

We discretize (54) by the standard finite element method with linear basis functions, see [LT03] for details. We define N equally distributed grid points in the unit interval

$$x_i = ih, \quad i = 1, \ldots, N$$

where $h = \frac{1}{N+1}$ denotes the stepsize. Using piecewise linear basis functions $\Lambda_i : [0,1] \to \mathbb{R}$, called *hat functions* and defined by

$$\Lambda_i(x_j) = \delta_{ij}, \quad i,j = 1, \ldots, N,$$

we write down the weak formulation of (54). If we take the space

$$V_h = \text{span}\{\Lambda_i : i = 1, \ldots, N\}$$

as ansatz and test space, the finite element solution $u_h : \mathbb{R}_+ \to V_h$ is defined by the solution of

$$\begin{aligned} (\frac{d}{dt} u_h(t), \Lambda_j)_2 + a(u_h(t), \Lambda_j) &= \mu (u_h(t), \Lambda_j)_2, \quad 1 \leq j \leq N \\ u_h(0) &= u_{h,0}. \end{aligned} \tag{55}$$

Here, the L_2-product $(\cdot,\cdot)_2$ and the elliptic bilinear form $a(\cdot,\cdot)$ are given by

$$(u,v)_2 = \int_0^1 u(x)\, v(x)\, dx, \quad a(u,v) = \int_0^1 u_x(x) v_x(x)\, dx. \tag{56}$$

The initial value $u_{h,0} \in V_h$ for the finite element system is usually given by a projection of the original initial function u_0. Here we choose the L^2-projection $u_{h,0} = P^h_{L^2} u_0$ defined by

$$(P^h_{L^2} u_0, \Lambda_i)_2 = (u_0, \Lambda_i)_2 \quad i = 1, \ldots, N.$$

For fixed $t > 0$, $u_h(t)$ is an element of V_h. Hence it is represented by a vector $\mathbf{u}(t) \in \mathbb{R}^N$ via the representation

$$u_h(t) = \sum_{i=1}^{N} \mathbf{u}_i(t)\Lambda_i \in V_h, \quad t \geq 0.$$

Using this representation, (55) transforms to

$$\begin{aligned} M_h\mathbf{u}_t + S_h\mathbf{u} &= \mu M_h\mathbf{u} \\ \mathbf{u}(0) &= \mathbf{u}_0 \end{aligned} \tag{57}$$

where the so-called mass and stiffness matrices $M_h, S_h \in \mathbb{R}^{N,N}$ are symmetric matrices with entries $M_h = ((\Lambda_j, \Lambda_i)_2)_{ij}$ and $S_h = (a(\Lambda_j, \Lambda_i))_{ij}$. The vector $\mathbf{u}_0$ corresponds to $u_{h,0}$ as above.

Explicit formula for the FE solution

In the one-dimensional case with linear finite elements that is considered here, M_h and S_h are tridiagonal Toeplitz matrices (i.e. constant along their diagonals) with the following entries:

$$M_h = \frac{h}{6}M,\ M = \begin{pmatrix} 4 & 1 & & \\ 1 & \ddots & \ddots & \\ & \ddots & \ddots & 1 \\ & & 1 & 4 \end{pmatrix}, \quad S = \frac{1}{h}S,\ S = \begin{pmatrix} 2 & -1 & & \\ -1 & \ddots & \ddots & \\ & \ddots & \ddots & -1 \\ & & -1 & 2 \end{pmatrix} \tag{58}$$

The following lemma provides the eigendecompositions of S_h and M_h.

Lemma 4.1 (see e.g. [Ise09], Lemma 10.5)**.** *Let $A = (a_{ij})_{ij}$ be a tridiagonal symmetric Toeplitz matrix with entries*

$$a_{1,1} = a_{i,i} = \alpha,\ a_{i-1,i} = a_{i,i-1} = \beta,\ i = 2, \ldots, N.$$

Then A is diagonalizable. The eigenvalues are given by $D = \operatorname{diag}(\lambda_1, \ldots, \lambda_N) \in \mathbb{R}^{N,N}$ and the corresponding orthonormal basis of eigenvectors by $V = \operatorname{col}(v_1, \ldots, v_N) \in \mathbb{R}^{N,N}$ where

$$\begin{aligned} Av_k &= \lambda_k v_k, \\ \lambda_k &= \alpha + 2\beta\cos(k\pi h), \quad k = 1 \ldots, N, \\ v_{k\ell} &= \sqrt{2h}\sin(k\ell\pi h), \quad k, \ell = 1, \ldots, N. \end{aligned}$$

As before, $h = \frac{1}{N+1}$.

According to Lemma 4.1, the eigendecompositions of M and S are given by

$$M = VD_MV, \quad S = VD_SV$$

with an orthogonal symmetric matrix $V \in \mathbb{R}^{N,N}$ as defined in Lemma 4.1 and diagonal matrices $D_{S,M} = \operatorname{diag}(\lambda_1^{S,M}, \ldots, \lambda_N^{S,M})$ defined by

$$\lambda_k^M = 4 + 2\cos(k\pi h), \quad \lambda_k^S = 2 - 2\cos(k\pi h), \quad k = 1 \ldots, N.$$

Solving equation (57) for $\mathbf{u}_t$ leads to

$$\mathbf{u}_t = V(\mu I_N - \frac{6}{h^2}D_M^{-1}D_S)V\mathbf{u}.$$

Hence the solution of (57) is given by

$$\mathbf{u}(t) = \sum_{k=1}^{N} e^{-\lambda_k t}\,(V\mathbf{u}_0)_k v_k, \quad \lambda_k = \frac{6}{h^2} \cdot \frac{2 - 2\cos(k\pi h)}{4 + 2\cos(k\pi h)} - \mu. \tag{59}$$

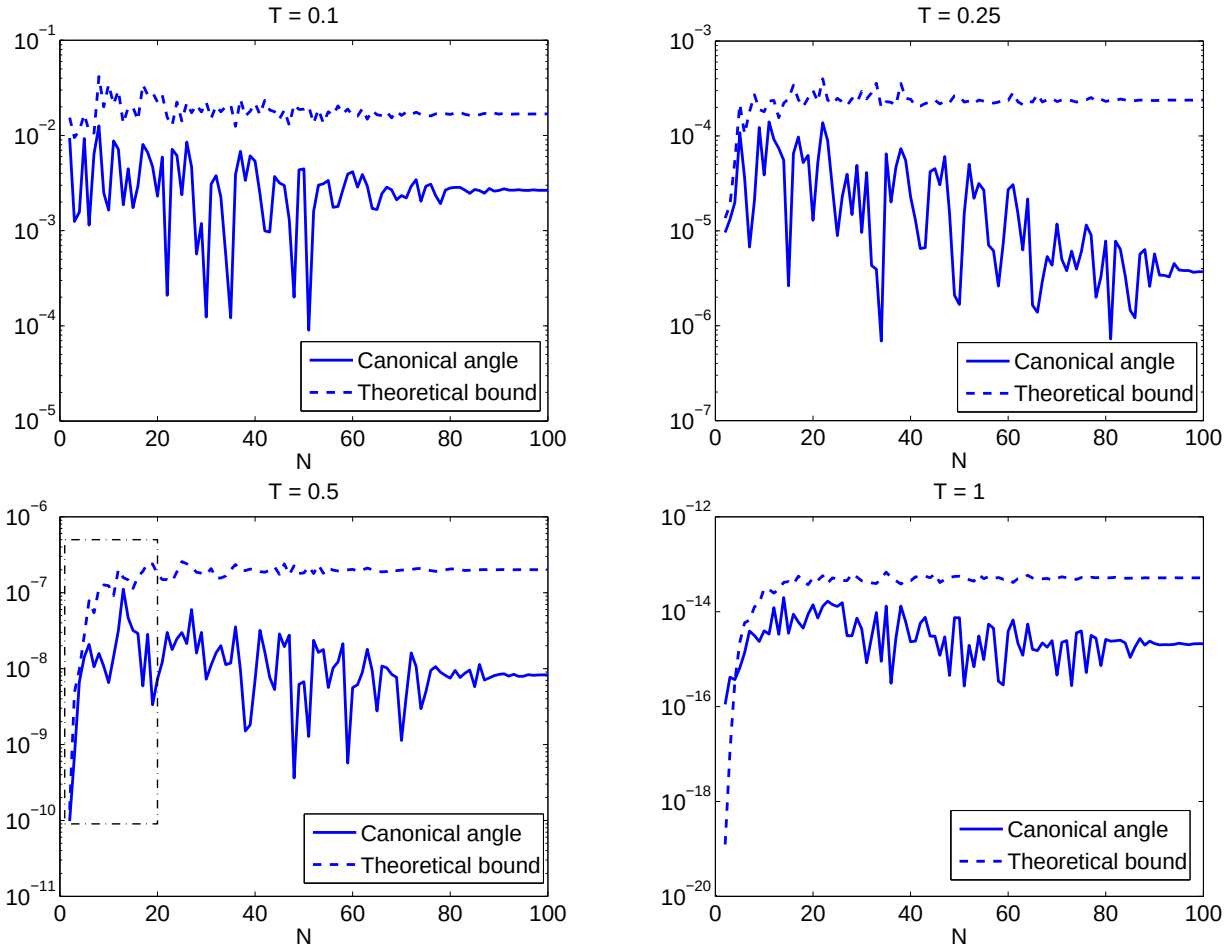

Figure 2: A linear reaction-diffusion system discretized by linear finite elements. Angle between first eigenvector and POD-vector together with the theoretical bound derived from (30). Different values of T.

The solution given by (59) is a trajectory of type (28), i.e. Theorem 3.5 is applicable.

Choice of initial data

To get comparable results for our experiments in different spaces V_h, $h = \frac{1}{N+1}$, $N = 1, \dots, \overline{N}$ we choose the initial vector $\mathbf{u}_{h,0}$ corresponding to the initial function $u_{h,0} \in V_h$ as the L^2-projection of a fixed initial function in $V_{\overline{h}}$, $\overline{h} = \frac{1}{\overline{N}+1}$:

$$u_0 = \sum_{i=1}^{\overline{N}} \alpha_i^0 \Lambda_i \in V_{\overline{h}}, \tag{60}$$

where we choose $\{\alpha_i^0\}_{1 \le i \le \overline{N}}$ at random. Observe that this refers to random initial data which should represent the worst-case behavior of solutions. We will come back to another possible choice of initial data at the end of our analysis.

Application of Theorem 3.5 for $\ell = 1$

A Taylor expansion immediately shows that the eigenvalues in (59) are of order

$$\lambda_k = \lambda_k(h) = k^2\pi^2 + O(h^2).$$

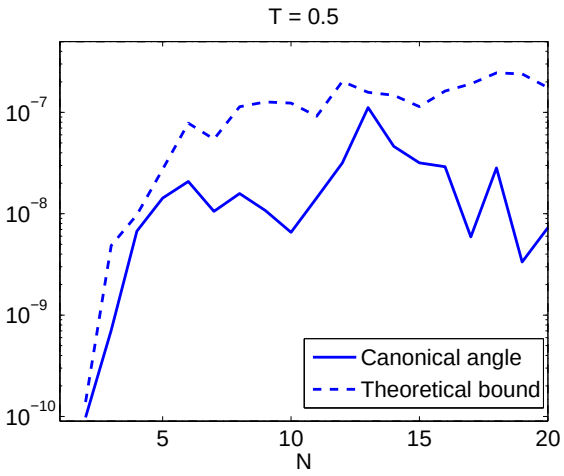

Figure 3: Blowup of the first $N = 20$ experiments of the example as above for $T = 0.5$. Plot of the region marked in Figure 2.

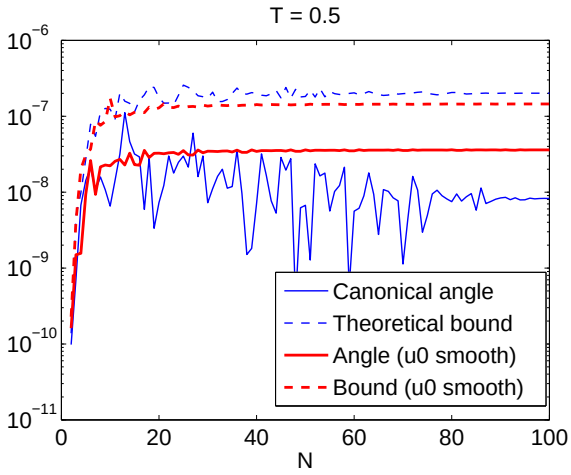

Figure 4: The same system as in figure 2 for $T = 0.5$. Comparison of the results for the stochastic initial data as in Figure 2 vs. smoother initial data.

Hence the stronger assumptions (32) of Theorem 3.5 are satisfied.

Let us first take a look at the case $\ell = 1$, i.e. we analyze the behavior of the first POD vector $w_1 \in \mathbb{R}^N$ measured by the angle between w_1 and the first Fourier mode v_1. Using the notation of Theorem 3.5 we get an explicit bound $\gamma = \gamma(T)$ with

$$\gamma = e^{-\delta T}, \quad \delta = \lambda_2 - \lambda_1 \approx 3\pi^2.$$

Since V is orthonormal, we can derive $M = M_N(\bar{u}_{0,N})$ depending on the initial data $\bar{u}_{0,N}$:

$$M = M_N = \frac{|(V\bar{u}_{0,N})_1|}{\sqrt{2}\max\limits_{2\le k\le N}|(V\bar{u}_{0,N})_k|}.$$

Therefore the bound (30) for the angle is given explicitly by

$$E_N = E_N(T, \bar{u}_{0,N}) = \frac{\gamma(T)}{M_N(\bar{u}_{0,N}) - \gamma(T)}.$$

Figure 2 shows the results of the experiments for $\overline{N} = 100$. We plot the angle between the first eigenvector $v_1 \in \mathbb{R}^N$ and the first POD vector as a function of the spatial dimension $1 \le N \le 100$ together with the theoretical bound E_N for different values of T all satisfying $T \ge 2/\delta$ as in the theorem. We see that for small N, the canonical angle is increasing but converges very quickly. The theoretical bound is confirmed by the experiments but overestimates the real error by a factor of about 10-20. Nevertheless the theoretical bound predicts the right order of magnitude of the angle in all cases.

In Figure 3 we take a closer look to the beginning of the experiments, i.e. at the interval $1 \le N \le 20$. We see that the bound is even sharper for small values of N. We mention a small numerical artefact in the last plot of Figure 2. We see that there is a small interval where the error bound does not hold. This can be explained by roundoff errors since the absolute value of the computed angle is very small.

In Figure 4 we compare the results of Figure 2 for $T = 0.5$ with an experiment for smoother initial data. In detail the initial data is chosen randomly in the space $V_{\frac{1}{11}}$ instead of $V_{\frac{1}{101}}$ as in (60). By that, more or less only the first 10 Fourier modes $v_1, \dots, v_{10}$ are involved in the L^2-projections for $N \ge 10$.

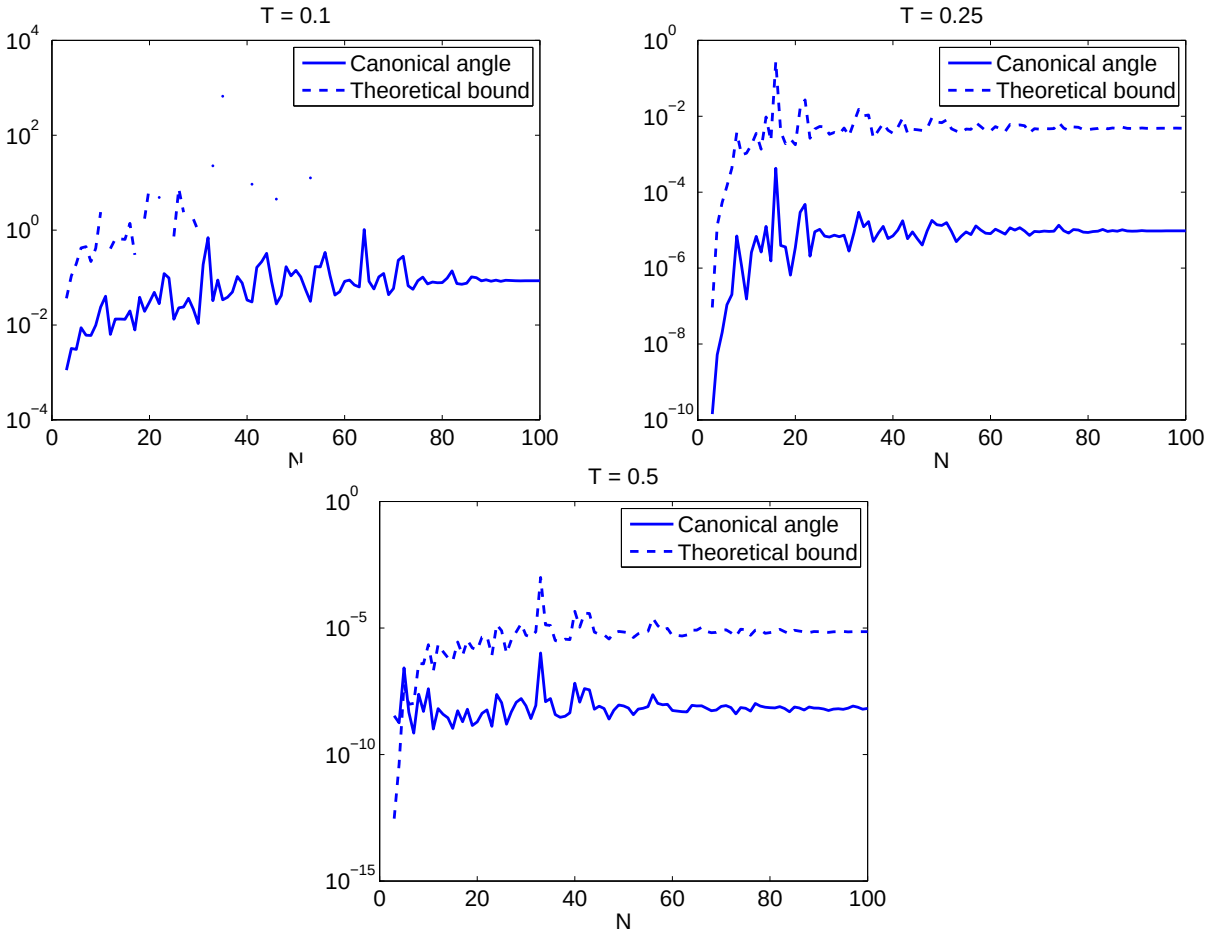

Figure 5: Same example as in Figure 2. Angle between the first $\ell = 2$ eigenvectors and POD-vectors together with the theoretical bound derived from (30). Different values of T.

We see in the plots that, qualitatively, the relation between the canonical angle and the bound given by Theorem 3.5 does not change. Both curves are just getting smoother for $N \geq 10$. This meets our expectations since the higher Fourier modes are negligible in this case.

Application of Theorem 3.5 for $\ell = 2$

We extend our experiments to the case of $\ell = 2$ POD modes to illustrate the bounds of Theorem 3.5 also for $\ell > 1$. With analogous calculations as above we derive the bounds for $\gamma = \gamma(T)$ and $M_N = M_N(\mathbf{u}_{0,N})$ also for the case $\ell = 2$. Observe that we have $\delta = \lambda_3 - \lambda_2 \approx 5\pi^2$ with the exact value given by (59) and further on

$$\begin{aligned} \mathrm{gap}_\alpha &= |\alpha_1 - \alpha_2| = e^{-\lambda_1 T} - e^{-\lambda_2 T}, \\ \sigma_2(V_0) &= \sigma_2((V\mathbf{u}_0)_1 v_1, (V\mathbf{u}_0)_2 v_2) = \min\{(V\mathbf{u}_0)_1, (V\mathbf{u}_0)_2\}, \\ \sigma_1(V_1) &= \sigma_1(\{(V\mathbf{u}_0)_k v_k\}_{k\geq 3}) = \max\{(V\mathbf{u}_0)_k\}_{k\geq 3}. \end{aligned}$$

Again, we can derive the bound $E_N = \frac{\gamma}{M_N - \gamma}$ explicitly with

$$\gamma = e^{-(\lambda_3 - \lambda_2)T}, \quad M_N = \frac{(e^{-\lambda_1 T} - e^{\lambda_2 T})}{4} \frac{\min\{(V\mathbf{u}_0)_1, (V\mathbf{u}_0)_2\}}{\max_{k=3,\ldots,N}(V\mathbf{u}_0)_k}.$$

Numerical experiments show that we have to be careful in choosing the integration time T. If we choose T too large (e.g. $T = 1$), the second singular value almost vanishes. Then the problem of finding a second singular vector becomes ill-posed such that the second POD mode is chosen almost randomly by the algorithm and does not fit to the snapshots.

In Figure 5 we have plotted the experimental data in the same way as in Figure 2 for the case $\ell = 2$ and $T \in \{0.1, 0.25, 0.5\}$.

For $T = 0.1$ the bound exists only for some $N \in \{1, \ldots, \bar{N}\}$. This is due to the fact that in most of the randomly chosen initial data the condition $M_N \geq \gamma$ is violated by choosing T that small. Therefore in the first plot of Figure 5, the theoretical bound is only evaluated in those experiments where $M_N \geq \gamma$ is satisfied.

If we choose a proper T like for example for $T = 0.25$ and $T = 0.5$, also for $\ell = 2$ the bounds of Theorem 3.5 are very suitable. As it is the case for $\ell = 1$, the magnitude of the canonical angles is described quite well by the error bounds. In particular the bounds overestimate the angles only by a factor of about $10 - 20$ as for $\ell = 1$.

4.3 Snapshots from different trajectories

In our concluding example we try to broaden the results of Theorem 3.10 by numerical means. The setting of the following computation is not covered any more by the assumptions of the theorem. Yet we mention it here since it gives a nice prospect for the POD algorithms where POD modes are computed from a whole bundle of short time trajectories. In the computations for Figure 6, we derive the first POD vector for the discretized linear parabolic system (59) with $T = 0.5$ in different ways. We build three different sets of snapshots collected in

$$Y^i = \frac{1}{\sqrt{6m}} \operatorname{col}(y_1^i, \ldots y_{6m}^i), \quad i = 1, 2, 3.$$

We consider trajectories $u^i(t) = u(t, \mathbf{u}_0^i)$ starting in 3 different initial points $\mathbf{u}_0^i$, $i = 1, 2, 3$, with randomly chosen coefficients as above. The first collection of snapshots is built from one trajectory of length $6mT$ as above:

$$y_j^1 = u^1(t_j), \quad j = 1, \ldots, 6m.$$

The second collection is built from two trajectories of half length $3mT$:

$$y_j^2 = u^1(t_j), \quad y_{3m+j}^j = u^2(t_j), \quad j = 1, \ldots, 3m.$$

Finally, the last collection of snapshots is built from three trajectories of length $2mT$:

$$y_j^3 = u^1(t_j), \quad y_{2m+j}^3 = u^2(t_j), \quad y_{4m+j}^3 = u^3(t_j), \quad j = 1, \ldots, 2m.$$

The resulting angles between the first POD mode w_1^i, $i = 1, 2, 3$, and the first Fourier mode in a typical realization of the experiment is plotted in Figure 6. We should mention that the computed angles depend strongly on the choice of the initial data $\mathbf{u}_0^i$. We have shown here a typical curve shape.

Obviously, the first eigendirection is approximated better by two trajectories than by one, at least for large N. Nevertheless for three trajectories in most realizations the error gets worse again, as it is the case in the example shown in Figure 6. It seems that the length of the involved trajectories is of greater advantage in this case than the convergence from more than two initial points in view of the approximation of the eigendirections. A deeper theoretical analysis of this obvious trade-off between the number and the length of trajectories used as data for the POD method is part of a future work.

References

[Ant05] A. C. Antoulas. *Approximation of large-scale dynamical systems*, *Advances in Design and Control*, volume 6. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2005.

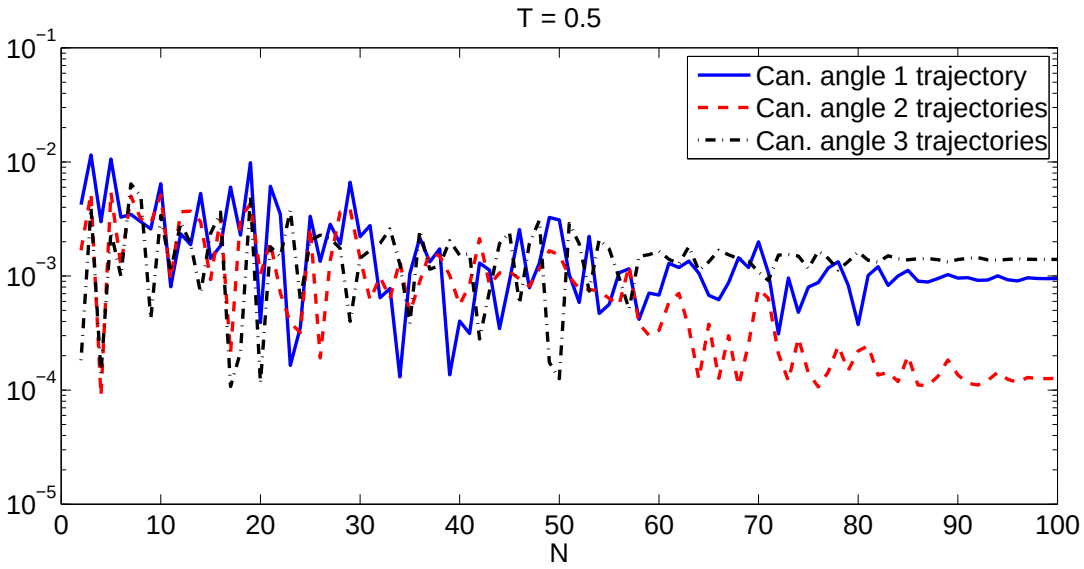

Figure 6: The same linear reaction-diffusion system discretized by linear finite elements as in Figure 2. Angle between the first eigenvector and POD-vector for snapshots taken from one trajectory, two trajectories and three trajectories in the case of $T = 0.5$.

[ASG01] A. C. Antoulas, D. C. Sorensen, and S. Gugercin. A survey of model reduction methods for large-scale systems. In *Structured matrices in mathematics, computer science, and engineering, I (Boulder, CO, 1999)*, *Contemp. Math.*, volume 280, pp. 193–219. Amer. Math. Soc., 2001.

[BQO05] P. Benner and E. S. Quintana-Ortí. Model reduction based on spectral projection methods. In *Dimension reduction of large-scale systems*, *Lect. Notes Comput. Sci. Eng.*, volume 45, pp. 5–48. Springer, Berlin, 2005.

[DFJ01] M. Dellnitz, G. Froyland, and O. Junge. The algorithms behind GAIO-set oriented numerical methods for dynamical systems. In *Ergodic theory, analysis, and efficient simulation of dynamical systems*, pp. 145–174, 805–807. Springer-Verlag, Berlin, 2001.

[DJ98] M. Dellnitz and O. Junge. An adaptive subdivision technique for the approximation of attractors and invariant measures. *Comput. Visual. Sci.*, 1, pp. 63–68, 1998.

[DJ99] M. Dellnitz and O. Junge. On the approximation of complicated dynamical behavior. *SIAM J. Numer. Anal.*, 36(2), pp. 491–515, 1999.

[Gau62] W. Gautschi. On inverses of Vandermonde and confluent Vandermonde matrices. *Numer. Math.*, 4, pp. 117–123, 1962.

[GvL96] G. H. Golub and C. F. van Loan. *Matrix computations*. Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press, Baltimore, third edition, 1996.

[HLB96] P. Holmes, J. L. Lumley, and G. Berkooz. *Turbulence, coherent structures, dynamical systems and symmetry*. Cambridge Monographs on Mechanics. Cambridge University Press, Cambridge, 1996.

[Ise09] A. Iserles. *A first course in the numerical analysis of differential equations*. Cambridge Texts in Applied Mathematics. Cambridge University Press, Cambridge, second edition, 2009.

[Kem10] J. Kemper. *Computation of invariant measures with dimension reduction methods*. Ph.D. thesis, Universität Bielefeld, Berlin, 2010.

[Kir96] A. Kirsch. *An introduction to the mathematical theory of inverse problems*, *Applied Mathematical Sciences*, volume 120. Springer-Verlag, New York, 1996.

[KV01] K. Kunisch and S. Volkwein. Galerkin proper orthogonal decomposition methods for parabolic problems. *Numer. Math.*, 90(1), pp. 117–148, 2001.

[KV02] K. Kunisch and S. Volkwein. Galerkin proper orthogonal decomposition methods for a general equation in fluid dynamics. *SIAM J. Numer. Anal.*, 40(2), pp. 492–515, 2002.

[LT03] S. Larsson and V. Thomée. *Partial differential equations with numerical methods*, *Texts in Applied Mathematics*, volume 45. Springer-Verlag, Berlin, 2003.

[RCM04] C. Rowley, T. Colonius, and R. Murray. Model reduction for compressible flows using POD and Galerkin projection. *Physica D Nonlinear Phenomena*, 189(1-2), pp. 115–129, 2004.

[Rob01] J. C. Robinson. *Infinite-dimensional dynamical systems*. Cambridge Texts in Applied Mathematics. Cambridge University Press, Cambridge, 2001.

[SS90] G. W. Stewart and J. G. Sun. *Matrix perturbation theory*. Computer Science and Scientific Computing. Academic Press Inc., Boston, MA, 1990.

[Tem97] R. Temam. *Infinite-dimensional dynamical systems in mechanics and physics*, *Applied Mathematical Sciences*, volume 68. Springer-Verlag, New York, second edition, 1997.