



# Data-Driven Reduced Order Models Using Invariant Foliations, Manifolds and Autoencoders

Robert Szalai<sup>1</sup>

Received: 24 June 2022 / Accepted: 2 June 2023 / Published online: 18 June 2023  
© The Author(s) 2023

## Abstract

This paper explores how to identify a reduced order model (ROM) from a physical system. A ROM captures an invariant subset of the observed dynamics. We find that there are four ways a physical system can be related to a mathematical model: invariant foliations, invariant manifolds, autoencoders and equation-free models. Identification of invariant manifolds and equation-free models require closed-loop manipulation of the system. Invariant foliations and autoencoders can also use off-line data. Only invariant foliations and invariant manifolds can identify ROMs, and the rest identify complete models. Therefore, the common case of identifying a ROM from existing data can only be achieved using invariant foliations. Finding an invariant foliation requires approximating high-dimensional functions. For function approximation, we use polynomials with compressed tensor coefficients, whose complexity increases linearly with increasing dimensions. An invariant manifold can also be found as the fixed leaf of a foliation. This only requires us to resolve the foliation in a small neighbourhood of the invariant manifold, which greatly simplifies the process. Combining an invariant foliation with the corresponding invariant manifold provides an accurate ROM. We analyse the ROM in case of a focus type equilibrium, typical in mechanical systems. The nonlinear coordinate system defined by the invariant foliation or the invariant manifold distorts instantaneous frequencies and damping ratios, which we correct. Through examples we illustrate the calculation of invariant foliations and manifolds and at the same time show that Koopman eigenfunctions and autoencoders fail to capture accurate ROMs under the same conditions.

**Keywords** Invariant foliation · Invariant manifold · Reduced order model · Instantaneous frequency · Tensor approximation · Machine learning

**Mathematics Subject Classification** 37M10 · 37M21 · 37C86 · 68T09

---

Communicated by Paul Newton.

---

✉ Robert Szalai  
r.szalai@bristol.ac.uk

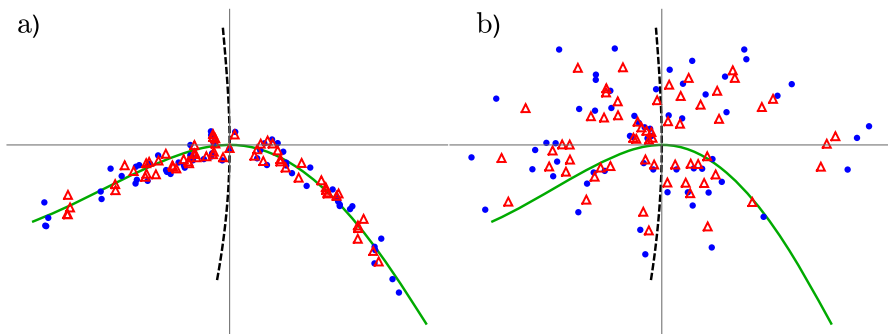
<sup>1</sup> Department of Engineering Mathematics, University of Bristol, Bristol, UK

## 1 Introduction

There is a great interest in the scientific community to identify explainable and/or parsimonious mathematical models from data. In this paper we classify these methods and identify one concept that is best suited to accurately calculate reduced order models (ROM) from off-line data. A ROM must track some selected features of the data over time and predict them into the future. We call this property of the ROM *invariance*. A ROM may also be *unique*, which means that barring a (nonlinear) coordinate transformation, the mathematical expression of the ROM is independent of who and when obtained the data as long as the sample size is sufficiently large and the distribution of the data satisfies some minimum requirements.

Not all methods that identify low-dimensional models produce ROMs. In some cases the data lie on a low-dimensional manifold embedded in a high-dimensional, typically Euclidean, space as in Fig. 1a. In this case the task is to parametrise the low-dimensional manifold and fit a model to the data in the coordinates of the parametrisation. The choice of parametrisation influences the form of the model. It is desired to use a parametrisation that yields a model with the least number of parameters. Approaches to reduce the number of parameters include compressed sensing (Billings 2013; Brunton et al. 2014, 2016; Champion et al. 2019; Donoho 2006) and normal form methods (Cenedese et al. 2022; Read and Ray 1998; Yair et al. 2017). The methods to obtain a parametrisation of the manifold include diffusion maps (Coifman and Lafon 2006), isomaps (Tenenbaum et al. 2000), autoencoders (Cenedese et al. 2022; Champion et al. 2019; Kalia et al. 2021; Kramer 1991) and equation-free models (Kevrekidis et al. 2003; Kevrekidis and Samaey 2009).

The main focus of this paper is genuine ROMs, where we need to find structure in a cloud of data as illustrated in Fig. 1b. An invariant manifold provides such structure, since all trajectories starting from the manifold stay on the manifold for all times. However an invariant manifold is not defined by the dynamics on it but by the dynamics in its neighbourhood (Cabr e et al. 2003; de la Llave 1997; Fenichel 1972). This



**Fig. 1** Two cases of data distribution. The blue dots (initial conditions) are mapped into the red triangles by the nonlinear map  $F$  given by Eq. (36). **a** Data points are distributed in the neighbourhood of the solid green curve, which is identified as an approximate invariant manifold. **b** Initial conditions are well-distributed in the neighbourhood of the steady state and there is no obvious manifold structure. For completeness, the second invariant manifold is denoted by the dashed line (Color figure online)

means that identifying a manifold must also involve identifying the dynamics in its neighbourhood, which is regularly omitted, such as in Cenedese et al. (2022). We find that resolving the nearby dynamics is the same as identifying an invariant foliation, as explained in Sect. 2.2.

Invariant foliations (Szalai 2020) can also be used to find structure in data like in Fig. 1b. An invariant foliation consists of a family of leaves (manifolds) that map onto each other under the dynamics of our system as in Fig. 4a. Each leaf of the foliation has a parameter from a space which has the dimensionality of the phase space minus the dimensionality of a leaf. As the leaves map onto each other, so do the parameters, which then defines a low-dimensional map or ROM. A leaf that maps onto itself is an invariant manifold. Invariant foliations are also useful to find suitable initial conditions for ROMs (Roberts 1989).

A natural question is if we have exhausted all possible concepts that identify structure in dynamic data. To address this, we systematically explore how a physical (or any other data producing) system can be related to a mathematical model. This allows us to categorise ROM concepts and choose the most suitable one for a given purpose. Through this process we arrive at the definitions of four cases: invariant foliations, invariant manifolds, autoencoders and equation-free models, and uncover their relations to each other.

We find that not all methods are applicable to off-line produced data. Indeed, calculating an invariant manifold requires actively probing the mathematical model (Haller and Ponsioen 2016). If a model is not available, the physical system must be placed under *closed-loop* control, such as in control-based continuation (Barton 2017; Sieber and Krauskopf 2008), which can currently identify equilibria or periodic orbits, but potentially it could also be used to find invariant manifolds of those equilibria and periodic orbits. Equation-free models were developed to recover inherently low-order dynamics from large systems by systematically probing the system input (Kevrekidis et al. 2003; Kevrekidis and Samaey 2009). However, not all systems can be put under closed-loop control, mainly because either setting up the required high-speed control loop is too costly or time consuming, or the system is simply inaccessible to the data analyst. In this case, the data collection is carried out separately from the data analysis without instant feedback to the system. We call this case *open-loop* or *off-line* data collection. We conclude that invariant foliations and autoencoders can be fitted to off-line data, but only invariant foliations produce genuine ROMs.

Surprisingly, invariant foliations were not explored for ROM identification before paper (Szalai 2020). Here, we propose a two-stage process, which finds an invariant foliation, that captures a low-order model and then finds a transverse and locally defined invariant foliation, whose fixed leaf is the invariant manifold with the same dynamics as the globally defined invariant foliation. This ensures that we take into account all data when the low-order dynamics is uncovered, and at the same time also produce a familiar structure, which is the corresponding invariant manifold. As per remark 15, this process may be simplified at the possible cost of losing some accuracy.

Another contribution of the paper is that we resolve the problem with high-dimensional data, which requires the approximation of high-dimensional functions, when invariant foliations are identified. We use polynomials that have compressed tensor coefficients (Grasedyck et al. 2013), and whose complexity scales linearly,

instead of combinatorially with the underlying dimension. Compressed tensors in the hierarchical Tucker format (HT) (Hackbusch and Kühn 2009) are amenable to singular value decomposition (Grasedyck 2010), hence by calculating the singular values we can check the accuracy of the approximation. In addition, compressed tensors can be truncated if some singular values are near zero without losing their accuracy. However for our purpose, the most important property is that within an optimisation framework a HT tensor is linear in each of its parameter matrices (when the rest of the parameters are fixed), which gives us a convex cost function. Indeed, when solving the invariance equation of the foliation, we use a block coordinate descent method, the Gauss–Southwell scheme (Nutini et al. 2015), which significantly speeds up the solution process. We also note that optimisation of each coefficient matrix of a HT tensor is constrained to a matrix manifold; hence, we use a trust-region method designed for matrix manifolds (Boumal 2022; Conn et al. 2000) when solving for individual matrix components.

A ROM is usually represented in a nonlinear coordinate system; hence, the quantities predicted by the ROM may not behave the same way as in Euclidean frames. This is particularly true for instantaneous frequencies and damping ratios of decaying vibrations. In Sect. 3, we take the distortion of the nonlinear coordinate system into account and derive correct values of instantaneous frequencies and damping ratios.

The structure of the paper is as follows. We first discuss the type of data assumed for ROM identification. Then, we define what a ROM is and go through all possible connections between a data producing system and a ROM. We then classify the uncovered conceptual connections along two properties: whether they are applicable to off-line data or produce genuine ROMs. Next, we discuss instantaneous frequencies and damping ratios in nonlinear frames. In Sect. 4, we summarise the proposed and tested numerical algorithms and describe their implementation details. Finally, we discuss three example problems. We start with a conceptual model that illustrates the non-applicability of autoencoders to genuine model order reduction. We then use a nonlinear ten-dimensional mathematical model to create synthetic data sets. Here we demonstrate the accuracy of our method to full state-space data, but also highlight problems with phase-space reconstruction from scalar signals. Finally, we analyse vibration data from a jointed beam, for which there is no accurate physical model due to the frictional interface in the joint.

## 1.1 Set-Up

The first step on our journey is to characterise the type of data we are working with. We assume a real  $n$ -dimensional Euclidean space, denoted by  $X$  (a vector space with an inner product  $\langle \cdot, \cdot \rangle_X : X \times X \rightarrow \mathbb{R}$ ), which contains all our data. We further assume that the data are produced by a deterministic process, which is represented by an unknown map  $F : X \rightarrow X$ . In particular, the data are organised into  $N \in \mathbb{N}^+$  pairs of vectors from space  $X$ , that is

$$(\mathbf{x}_k, \mathbf{y}_k), \quad k = 1, \dots, N$$

that satisfy

$$y_k = F(x_k) + \xi_k, \quad k = 1, \dots, N, \tag{1}$$

where  $\xi_k \in X$  represents a small measurement noise, which is sampled from a distribution with zero mean. Equation (1) describes pieces of trajectories if for some  $k \in \{1, \dots, N\}$ ,  $x_{k+1} = y_k$ . The state of the system can also be defined on a manifold, in which case  $X$  is chosen such that the manifold is embedded in  $X$  according to Whitney’s embedding theorem (Whitney 1936). It is also possible that the state cannot be directly measured, in which case Takens’ delay embedding technique (Takens 1981) can be used to reconstruct the state, which we will do subsequently in an optimal manner (Casdagli 1989). As a minimum, we require that our system is observable (Hermann and Krener 1977).

For the purpose of this paper we also assume a fixed point at the origin, such that  $F(0) = 0$  and that the domain of  $F$  is a compact and connected subset  $G \subset X$  that includes a neighbourhood of the origin.

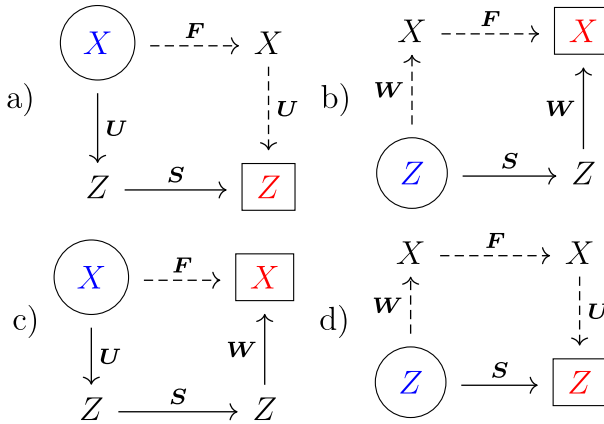
## 2 Reduced Order Models

We now describe a weak definition of a ROM, which only requires invariance. There are two ingredients of a ROM: a function connecting vector space  $X$  to another vector space  $Z$  of lower dimensionality and a map on vector space  $Z$ . The connection can go two ways, either from  $Z$  to  $X$  or from  $X$  to  $Z$ . To make this more precise, we also assume that  $Z$  has an inner product  $\langle \cdot, \cdot \rangle_Z : Z \times Z \rightarrow \mathbb{R}$  and that  $\dim Z < \dim X$ . The connection  $U : X \rightarrow Z$  is called an *encoder* and the connection  $W : Z \rightarrow X$  is called a *decoder*. We also assume that both the encoder and the decoder are continuously differentiable and their Jacobians have full rank. Our terminology is borrowed from computer science (Kramer 1991), but we can also use mathematical terms that calls  $U$  a manifold submersion (Lawson 1974) and  $W$  a manifold immersion (Lang 2012). In accordance with our assumption that  $F(0) = 0$ , we also assume that  $U(0) = 0$  and  $W(0) = 0$ .

**Definition 1** Assume two maps  $F : X \rightarrow X$ ,  $S : Z \rightarrow Z$  and an encoder  $U : X \rightarrow Z$  or a decoder  $W : Z \rightarrow X$ .

1. The encoder-map pair  $(U, S)$  is a *ROM* of  $F$  if for all initial conditions  $x_0 \in G \subset X$ , the trajectory  $x_{k+1} = F(x_k)$  and for initial condition  $z_0 = U(x_0)$  the second trajectory  $z_{k+1} = S(z_k)$  are connected such that  $z_k = U(x_k)$  for all  $k > 0$ .
2. The decoder-map pair  $(W, S)$  is a *ROM* of  $F$  if for all initial conditions  $z_0 \in H = \{z \in Z : W(z) \in G\}$ , the trajectory  $z_{k+1} = S(z_k)$  and for initial condition  $x_0 = W(z_0)$  the second trajectory  $x_{k+1} = F(x_k)$  are connected such that  $x_k = W(z_k)$  for all  $k > 0$ .

**Remark 2** In essence, a ROM is a model whose trajectories are connected to the trajectories of our system  $F$ . We call this property *invariance*. It is possible to define a weaker ROM, where the connections  $z_k = U(x_k)$  or  $x_k = W(z_k)$  are only approximate, which is not discussed here.



**Fig. 2** Commutative diagrams of **a** invariant foliations, **b** invariant manifolds, and connection diagrams of **c** autoencoders and **d** equation-free models. The dashed arrows denote the chain of function composition(s) that involves  $F$ , the continuous arrows denote the chain of function composition(s) that involves map  $S$ . The encircled vector space denotes the domain of the invariance equation, and the boxed vector space is the target of the invariance equation

The following corollary can be thought of as an equivalent definition of a ROM.

**Corollary 3** *The encoder-map pair  $(U, S)$  or decoder-map pair  $(W, S)$  is a ROM if and only if either invariance equation*

$$S(U(x)) = U(F(x)), \quad x \in G \quad \text{or} \quad (2)$$

$$W(S(z)) = F(W(z)), \quad z \in H, \quad (3)$$

hold, where  $H = \{z \in Z : W(z) \in G\}$ .

**Proof** Let us assume that (2) holds, and choose an  $x_0 \in X$  and let  $z_0 = U(x_0)$ . First we check whether  $z_k = U(x_k)$ , if  $x_k = F^k(x_0)$ ,  $z_k = S^k(z_0)$  and (2) hold. Substituting  $x = F^k(x_0)$  into (2) yields

$$S^k(z_0) = U(F^k(x_0))$$

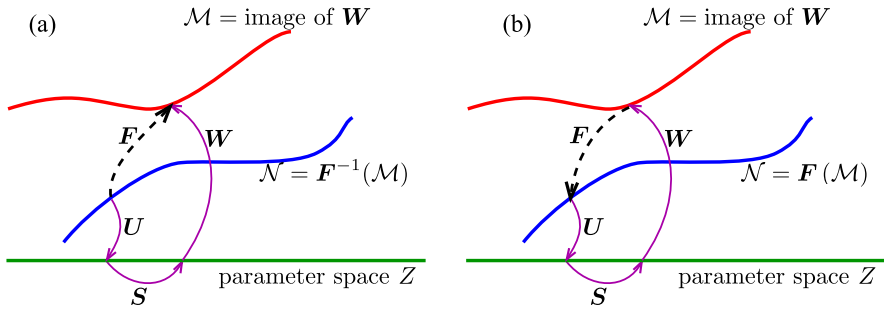
$$z_k = U(x_k).$$

Now in reverse, assuming that  $z_k = U(x_k)$ ,  $z_k = S^k(z_0)$ ,  $x_k = F^k(x_0)$  and setting  $k = 1$  yields that

$$z_1 = S(z_0) = U(x_1),$$

$$S(U(x_0)) = U(F(x_0)),$$

which is true for all  $x_{k-1} \in G$ , hence (2) holds. The proof for the decoder is identical, except that we swap  $F$  and  $S$  and replace  $U$  with  $W$ . □



**Fig. 3** **a** Autoencoder. Encoder  $U$  maps data to parameter space  $Z$ , then the low-order map  $S$  takes the parameter forward in time and finally the decoder  $W$  brings the new parameter back to the state space. This chain of maps denoted by solid purple arrows must be the same as map  $F$  denoted by the dashed arrow. The two results can only match between two manifolds  $\mathcal{M}$  and  $\mathcal{N}$  and not elsewhere in the state space. Manifold  $\mathcal{M}$  is invariant if and only if  $\mathcal{M} \subset \mathcal{N}$ , which is not guaranteed by this construction. **b** Equation-free model. The only difference from the autoencoder is the reversal of the arrow under  $F$ , hence  $S(z) = U(F(W(z)))$  (Color figure online)

Now the question arises if we can use a combination of an encoder and a decoder? This is the case of the autoencoder or nonlinear principal component analysis (Kramer 1991) and the equation-free model (Kevrekidis et al. 2003; Kevrekidis and Samaey 2009). Indeed, there are four combinations of encoders and decoders, which are depicted in diagrams Fig. 2a–d. We name these four scenarios as follows.

**Definition 4** We call the connections displayed in Fig. 2a–d, invariant foliation, invariant manifold, autoencoder and equation-free model, respectively.

As we walk through the dashed and solid arrows in diagram Fig. 2a, we find the two sides of the invariance Eq. (2). If we do the same for diagram Fig. 2b, we find Eq. (3). This implies that invariant foliations and invariant manifolds are ROMs. The same is not true for autoencoders and equation-free models. Reading off diagram Fig. 2c, the autoencoder must satisfy

$$W(S(U(x))) = F(x). \tag{4}$$

Equation (4) is depicted in Fig. 3a. Since the decoder  $W$  maps onto a manifold  $\mathcal{M}$ , Eq. (4) can only hold if  $x$  is chosen from the preimage of  $\mathcal{M}$ , that is,  $x \in \mathcal{N} = F^{-1}(\mathcal{M})$ . For invariance, we need  $F(\mathcal{M}) \subset \mathcal{M}$ , which is the same as  $\mathcal{M} \subset \mathcal{N}$  if  $F$  is invertible. However, the inclusion  $\mathcal{M} \subset \mathcal{N}$  is not guaranteed by Eq. (4). The only way to guarantee  $\mathcal{M} \subset \mathcal{N}$ , is by stipulating that the function composition  $W \circ U$  is the identity map on the data. A trivial case is when  $\dim Z = \dim X$  or, in general, when all data fall onto a  $\dim Z$ -dimensional submanifold of  $X$ . Indeed, the standard way to find an autoencoder (Kramer 1991) is to solve

$$\arg \min_{U, W} \sum_{k=1}^N \|W(U(x_k)) - x_k\|^2. \tag{5}$$

Unfortunately, if the data are not on a  $\dim Z$ -dimensional submanifold of  $X$ , which is the case of genuine ROMs, the minimum of  $\sum_{k=1}^N \|W(U(x_k)) - x_k\|^2$  will be far from zero and the location of  $\mathcal{M}$  as the solution of (5) will only indicate where the data are in the state space. It is customary to seek a solution to (5) under the normalising condition that  $U \circ W$  is the identity and hence  $S = U \circ F \circ W$ .

The equation-free model in diagram Fig. 2d is identical to the autoencoder Fig. 2c if we replace  $F$  with  $F^{-1}$  and  $S$  with  $S^{-1}$ . Reading off diagram Fig. 2d, the equation-free model must satisfy

$$S(z) = U(F(W(z))). \tag{6}$$

Equation (6) immediately provides  $S$ , for any  $U, W$ , without identifying any structure in the data.

Only invariant foliations, through Eq. (2) and autoencoders through Eqs. (4) and (5) can be fitted to off-line data. Invariant manifolds and equation-free models require the ability to manipulate the input of our system  $F$  during the identification process. Indeed, for off-line data, produced by Eq. (1), the foliation invariance Eq. (2) turns into an optimisation problem

$$\arg \min_{S,U} \sum_{k=1}^N \|x_k\|^{-2} \|S(U(x_k)) - U(y_k)\|^2, \tag{7}$$

and the autoencoder Eq. (4) turns into

$$\arg \min_{S,U,W} \sum_{k=1}^N \|x_k\|^{-2} \|W(S(U(x_k))) - y_k\|^2. \tag{8}$$

For the optimisation problem (7) to have a unique solution, we need to apply a constraint to  $U$ . One possible constraint is explained in remark 7. In case of the autoencoder (8), the constraint, in addition to the one restricting  $U$ , can be that  $U \circ W$  is the identity, as also stipulated in Cenedese et al. (2022).

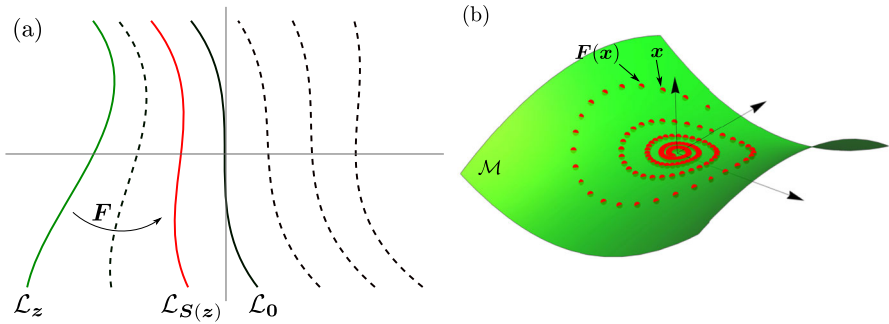
### 2.1 Invariant Foliations and Invariant Manifolds

An encoder represents a family of manifolds, called *foliation*, which is the set of constant level surfaces of  $U$ . A single level surface is called a *leaf* of the foliation; hence, the foliation is a collection of leaves. In mathematical terms, a leaf with parameter  $z \in Z$  is denoted by

$$\mathcal{L}_z = \{x \in G \subset X : U(x) = z\}. \tag{9}$$

All leaves are  $\dim X - \dim Z$ -dimensional differentiable manifolds, because we assumed that the Jacobian  $DU$  has full rank (Lawson 1974). The collection of all leaves is a *foliation*, denoted by





**Fig. 4** **a** Invariant foliation: a leaf  $\mathcal{L}_z$  is being mapped onto leaf  $\mathcal{L}_{S(z)}$ . Since the origin  $\mathbf{0} \in Z$  is a steady state of  $S$ , the leaf  $\mathcal{L}_0$  is mapped into itself and therefore it is an invariant manifold. **b** Invariant manifold: a single trajectory, represented by red dots, starting from the invariant manifold  $\mathcal{M}$  remains on the green invariant manifold (Color figure online)

$$\mathcal{F} = \{ \mathcal{L}_z : z \in H \},$$

where  $H = U(G)$ . The foliation is characterised by its co-dimension, which is the same as  $\dim Z$ . Invariance Eq. (2) means that each leaf of the foliation is mapped onto another leaf, in particular the leaf with parameter  $z$  is mapped onto the leaf with parameter  $S(z)$ , that is

$$F(\mathcal{L}_z) \subset \mathcal{L}_{S(z)}.$$

Due to our assumptions, leaf  $\mathcal{L}_0$  is an invariant manifold, because  $F(\mathcal{L}_0) \subset \mathcal{L}_0$ . This geometry is illustrated in Fig. 4a.

We now characterise the existence and uniqueness of invariant foliations about a fixed point. We assume that  $F$  is a  $C^r$ ,  $r \geq 2$  map and that the Jacobian matrix  $DF(\mathbf{0})$  has eigenvalues  $\mu_1, \dots, \mu_n$  such that  $|\mu_i| < 1$ ,  $i = 1, \dots, n$ . To select the invariant manifold or foliation we assume two  $\nu$ -dimensional linear subspaces  $E$  of  $X$  and  $E^*$  of  $X$  corresponding to eigenvalues  $\mu_1, \dots, \mu_\nu$  such that  $DF(\mathbf{0})E \subset E$  and for the adjoint map  $(DF(\mathbf{0}))^*E^* \subset E^*$ .

**Definition 5** The number

$$\beth_{E^*} = \frac{\min_{k=1 \dots \nu} \log |\mu_k|}{\max_{k=1 \dots n} \log |\mu_k|}$$

is called the spectral quotient of the left-invariant linear subspace  $E^*$  of  $F$  about the origin.

**Theorem 6** Assume that  $DF(\mathbf{0})$  is semisimple and that there exists an integer  $\sigma \geq 2$ , such that  $\beth_{E^*} < \sigma \leq r$ . Also assume that

$$\prod_{k=1}^n \mu_k^{m_k} \neq \mu_j, \quad j = 1, \dots, \nu \tag{10}$$

for all  $m_k \geq 0$ ,  $1 \leq k \leq n$  with at least one  $m_l \neq 0$ ,  $v + 1 \leq l \leq n$  and with  $\sum_{k=0}^n m_k \leq \sigma - 1$ . Then in a sufficiently small neighbourhood of the origin there exists an invariant foliation  $\mathcal{F}$  tangent to the left-invariant linear subspace  $E^*$  of the  $C^r$  map  $F$ . The foliation  $\mathcal{F}$  is unique among the  $\sigma$  times differentiable foliations and it is also  $C^r$  smooth.

**Proof** The proof is carried out in Szalai (2020). Note that the assumption of  $DF(\mathbf{0})$  being semisimple was used to simplify the proof in Szalai (2020), therefore it is unlikely to be needed.  $\square$

**Remark 7** Theorem 6 only concerns the uniqueness of the foliation, but not the encoder  $U$ . However, for any smooth and invertible map  $R : Z \rightarrow Z$ , the encoder  $\tilde{U} = R \circ U$  represents the same foliation and the nonlinear map  $S$  transforms into  $\tilde{S} = R \circ S \circ R^{-1}$ . If we want to solve the invariance Eq. (2), we need to constrain  $U$ . The simplest such constraint is that

$$U(W_1 z) = z, \tag{11}$$

where  $W_1 : Z \rightarrow X$  is a linear map with full rank such that  $E^* \cap \ker W_1^* = \{\mathbf{0}\}$ . To explain the meaning of Eq. (11), we note that the image of  $W_1$  is a linear subspace  $\mathcal{H}$  of  $X$ . Equation (11) therefore means that each leaf  $\mathcal{L}_z$  must intersect subspace  $\mathcal{H}$  exactly at parameter  $z$ . The condition  $E^* \cap \ker W_1^* = \{\mathbf{0}\}$  then means that the leaf  $\mathcal{L}_0$  has a transverse intersection with subspace  $\mathcal{H}$ . This is similar to the graph-style parametrisation of a manifold over a linear subspace.

**Remark 8** Eigenvalues and eigenfunctions of the Koopman operator (Mezić 2005, 2021) are invariant foliations. Indeed, the Koopman operator is defined as  $(\mathcal{K}u)(x) = u(F(x))$ . If we assume that  $U = (u_1, \dots, u_v)$  is a collection of functions  $u_j : X \rightarrow \mathbb{R}$ ,  $Z = \mathbb{R}^v$ , then  $U$  spans an invariant subspace of  $\mathcal{K}$  if there exists a linear map  $S$  such that  $\mathcal{K}(U) = SU$ . Expanding this equation yields  $U(F(x)) = SU(x)$ , which is the same as the invariance Eq. (2), except that  $S$  is linear. The existence of linear map  $S$  requires further non-resonance conditions, which are

$$\prod_{k=1}^v \mu_k^{m_k} \neq \mu_j, \quad j = 1, \dots, v \tag{12}$$

for all  $m_k \geq 0$  such that  $\sum_{k=0}^n m_k \leq \sigma - 1$ . Equation (12) is referred to as the set of *internal non-resonance* conditions, because these are intrinsic to the invariant subspace  $E^*$ . In many cases  $S$  represents the slowest dynamics, hence even if there are no internal resonances, the two sides of (12) will be close to each other for some set of  $m_1, \dots, m_v$  exponents and that causes numerical issues leading to undesired inaccuracies. We will illustrate this in Sect. 5.2.

Now we discuss invariant manifolds. A decoder  $W$  defines a differentiable manifold

$$\mathcal{M} = \{W(z) : z \in H\},$$

where  $H = \{z \in Z : W(z) \in G\}$ . Invariance Eq. (3) is equivalent to the geometric condition that  $F(\mathcal{M}) \subset \mathcal{M}$ . This geometry is shown in Fig. 4b, which illustrates that if a trajectory is started on  $\mathcal{M}$ , all subsequent points of the trajectory stay on  $\mathcal{M}$ .

Invariant manifolds as a concept cannot be used to identify ROMs from off-line data. As we will see below, invariant manifolds can still be identified as a leaf of an invariant foliation, but not through the invariance Eq. (3). Indeed, it is not possible to guess the manifold parameter  $z \in Z$  from data. Introducing an encoder  $U : X \rightarrow Z$  to calculate  $z = U(x)$ , transforms the invariant manifold into an autoencoder, which does not guarantee invariance.

We now state the conditions of the existence and uniqueness of an invariant manifold.

**Definition 9** The number

$$\aleph_E = \frac{\min_{k=v+1\dots n} \log |\mu_k|}{\max_{k=1\dots v} \log |\mu_k|}$$

is called the spectral quotient of the right-invariant linear subspace  $E$  of map  $F$  about the origin.

**Theorem 10** Assume that there exists an integer  $\sigma \geq 2$ , such that  $\aleph_E < \sigma \leq r$ . Also assume that

$$\prod_{k=1}^v \mu_k^{m_k} \neq \mu_j, \quad j = v + 1, \dots, n \tag{13}$$

for all  $m_k \geq 0$  such that  $\sum_{k=0}^v m_k \leq \sigma - 1$ . Then, in a sufficiently small neighbourhood of the origin there exists an invariant manifold  $\mathcal{M}$  tangent to the invariant linear subspace  $E$  of the  $C^r$  map  $F$ . The manifold  $\mathcal{M}$  is unique among the  $\sigma$ -times differentiable manifolds and it is also  $C^r$  smooth.

**Proof** The theorem is a subset of theorem 1.1 in Cabré et al. (2003). □

**Remark 11** To calculate an invariant manifold with a unique representation, we need to impose a constraint on  $W$  and/or  $S$ . The simplest constraint is imposed by

$$U_1 W(z) = z, \tag{14}$$

where  $U_1 : Z \rightarrow X$  is a linear map with full rank such that  $E \cap \ker U_1^* = \{0\}$ . This is similar to a graph-style parametrisation (akin to theorem 1.2 in Cabré et al. (2003)), where the range of  $U_1$  must span the linear subspace  $E$ . Constraint (14) can break down for large  $\|z\|$ , when  $U_1 DW(z)$  does not have full rank. A globally suitable constraint is that  $DW^*(z) DW(z) = I$ .

For linear subspaces  $E$  and  $E^*$  with eigenvalues closest to the complex unit circle (representing the slowest dynamics),  $\beth_E = 1$  and  $\aleph_E$  is maximal. Therefore the foliation corresponding to the slowest dynamics requires the least smoothness, while the invariant manifold requires the maximum smoothness for uniqueness.

**Table 1** Comparison of invariant foliations, invariant manifolds and autoencoders

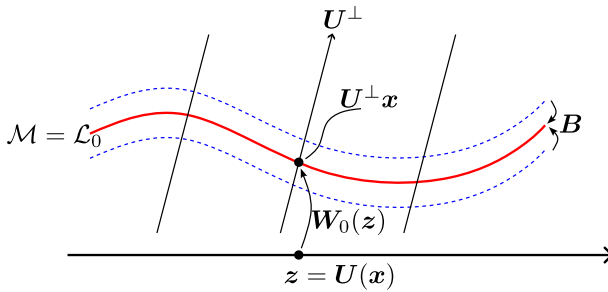
|                    | Invariant foliation   | Invariant manifold   | Autoencoder   | Eq.-free model   |
|--------------------|---|--|---|--|
| Usable closed-loop | YES   | YES  | YES   | YES  |
| Usable open-loop   | YES   | NO   | YES   | NO   |
| Obtains a ROM      | YES   | YES  | NO  | NO   |
| Uniqueness         | Slowest most unique   | Slowest least unique   | NO  | NO   |
| References         | Mezić (2005), Roberts (1989), Roberts (1992), Szalai (2020) | Cabré et al. (2003), Cabré et al. (2005), de la Llave (1997), Haller and Ponsioen (2016), Shaw and Pierre (1994), Szalai et al. (2017), Vizzaccaro et al. (2021) | Cenedese et al. (2022), Champion et al. (2019), Kalia et al. (2021) | Kevrekidis et al. (2003), Kevrekidis and Samaey (2009) |

Table 1 summarises the main properties of the three conceptually different model identification techniques. Ultimately, in the presence of off-line data, only invariant foliations can be fitted to the data and produce a ROM at the same time.

## 2.2 Invariant Manifolds Represented by Locally Defined Invariant Foliations

As discussed before, we cannot fit invariant manifolds to data, instead we can fit an invariant foliation that contains our invariant manifold (which is the leaf containing the fixed point  $\mathcal{L}_0$  at the origin). This invariant foliation only needs to be defined near the invariant manifold and therefore we can simplify the functional representation of the encoder that defines the foliation. In this section we discuss this simplification.

To begin with, assume that we already have an invariant foliation  $\mathcal{F}$  with an encoder  $U$  and nonlinear map  $S$ . Our objective is to find the invariant manifold  $\mathcal{M}$ , represented by decoder  $W$  that has the same dynamics  $S$  as the foliation. This is useful if we want to know quantities that are only defined for invariant manifolds, such as instantaneous frequencies and damping ratios. Formally, we are looking for a simplified invariant foliation  $\hat{\mathcal{F}}$  with encoder  $\hat{U} : X \rightarrow \hat{Z}$  that together with  $\mathcal{F}$  form a coordinate system in  $X$ . (Technically speaking,  $(U, \hat{U}) : X \rightarrow Z \times \hat{Z}$  must be a isomorphism.) In this case our invariant manifold is the zero level surface of encoder  $\hat{U}$ , i.e.  $\mathcal{M} = \hat{\mathcal{L}}_0 \in \hat{\mathcal{F}}$ . Naturally, we must have  $\dim Z + \dim \hat{Z} = \dim X$ .



**Fig. 5** Approximate invariant foliation. The linear map  $U^\perp$  and the nonlinear map  $U$  create a coordinate system, so that in this frame the invariant manifold  $\mathcal{M}$  is given by  $U^\perp x = W_0(z)$ , where  $z = U(x)$ . We then allow to shift manifold  $\mathcal{M}$  in the  $U^\perp$  direction such that  $U^\perp x - W_0(z) = \hat{z}$ , which creates a foliation parametrised by  $\hat{z} \in \hat{Z}$ . If this foliation satisfies the invariance equation in a neighbourhood of  $\mathcal{M}$  with a linear map  $B$  as per Eq. (16), then  $\mathcal{M}$  is an invariant manifold. In other words, nearby blue dashed leaves are mapped towards  $\mathcal{M}$  by linear map  $B$  the same way as the underlying dynamics does (Color figure online)

The sufficient condition that  $\mathcal{F}$  and  $\hat{\mathcal{F}}$  form a coordinate system locally about the fixed point is that the square matrix

$$Q = \begin{pmatrix} DU(0) \\ D\hat{U}(0) \end{pmatrix}$$

is invertible. Let us represent our approximate (or locally defined) encoder by

$$\hat{U}(x) = U^\perp x - W_0(U(x)), \tag{15}$$

where  $W_0 : Z \rightarrow \hat{Z}$  is a nonlinear map with  $DW_0(0) = 0$ ,  $U^\perp : X \rightarrow \hat{Z}$  is an orthogonal linear map,  $U^\perp (U^\perp)^* = I$  and  $U^\perp W_0(z) = 0$ . Here, the linear map  $U^\perp$  measures coordinates in a transversal direction to the manifold and  $W_0$  prescribes where the actual manifold is along this transversal direction, while  $U$  provides the parametrisation of the manifold. All other leaves of the approximate foliation  $\hat{\mathcal{F}}$  are shifted copies of  $\mathcal{M}$  along the  $U^\perp$  direction as displayed in Fig. 5. A locally defined foliation means that  $\hat{z} = \hat{U}(x) \in \hat{Z}$  is assumed to be small; hence, we can also assume linear dynamics among the leaves of  $\hat{\mathcal{F}}$ , which is represented by a linear operator  $B : \hat{Z} \rightarrow \hat{Z}$ . Therefore, the invariance Eq. (2) becomes

$$B\hat{U}(x) = \hat{U}(F(x)). \tag{16}$$

Once  $B$ ,  $U^\perp$  and  $W_0$  are found, the final step is to reconstruct the decoder  $W$  of our invariant manifold  $\mathcal{M}$ .

**Proposition 12** *The decoder  $W$  of the invariant manifold  $\mathcal{M} = \{x \in X : \hat{U}(x) = 0\}$  is the unique solution of the system of equations*

$$\left. \begin{aligned} U(W(z)) &= z \\ U^\perp W(z) &= W_0(z) \end{aligned} \right\}. \tag{17}$$

**Proof** First we show that conditions (17) imply  $\hat{U}(W(x)) = \mathbf{0}$ . We expand our expression using (15) into  $\hat{U}(W(x)) = U^\perp W(z) - W_0(U(W(z)))$ , then use Eq. (17), which yields  $\hat{U}(W(x)) = \mathbf{0}$ . To solve Eq. (17) for  $W$ , we decompose  $U, W$  into linear and nonlinear components, such that  $U(x) = DU(0)x + \tilde{U}(x)$  and  $W(z) = DW(0)z + \tilde{W}(z)$ . Expanding Eq. (17) with the decomposed  $U, W$  yields

$$\left. \begin{aligned} DU(0)(DW(0)z + \tilde{W}(z)) + \tilde{U}(DW(0)z + \tilde{W}(z)) &= z \\ U^\perp(DW(0)z + \tilde{W}(z)) &= W_0(z) \end{aligned} \right\}. \tag{18}$$

The linear part of Eq. (18) is

$$QDW(0) = \begin{pmatrix} I \\ \mathbf{0} \end{pmatrix},$$

hence  $DW(0) = Q^{-1} \begin{pmatrix} I \\ \mathbf{0} \end{pmatrix}$ . The nonlinear part of (18) is

$$Q\tilde{W}(z) = \begin{pmatrix} -\tilde{U}(DW(0)z + \tilde{W}(z)) \\ W_0(z) \end{pmatrix},$$

which can be solved by the iteration

$$\tilde{W}_{k+1}(z) = Q^{-1} \begin{pmatrix} -\tilde{U}(DW(0)z + \tilde{W}_k(z)) \\ W_0(z) \end{pmatrix}, \quad \tilde{W}_1(z) = \mathbf{0}, \quad k = 1, 2, \dots \tag{19}$$

Iteration (19) converges for  $|z|$  sufficiently small, due to  $\tilde{U}(x) = \mathcal{O}(|x|^2)$  and  $W_0(z) = \mathcal{O}(|z|^2)$ .  $\square$

As we will see in Sect. 5, this approach provides better results than using an autoencoder. Here we have resolved the dynamics transversal to the invariant manifold  $\mathcal{M}$  up to linear order. It is essential to resolve this dynamics to find invariance, not just the location of data points.

**Remark 13** More consideration is needed in case  $\hat{U}(x_k)$  assumes large values over some data points. This either requires to replace  $B$  with a nonlinear map or we need to filter out data points that are not in a small neighbourhood of the invariant manifold  $\mathcal{M}$ . Due to  $B$  being high-dimensional, replacing it with a nonlinear map leads to numerical difficulties. Filtering data is easier. For example, we can assign weights to each term in our optimisation problem (7) depending on how far a data point is from the predicted manifold, which is the zero level surface of  $\hat{U}$ . This can be done using the optimisation problem

$$\arg \min_{B, U^\perp, W_0} \sum_{k=1}^N \|x_k\|^{-2} \phi_k \left( \left\| \hat{U}(x_k) \right\|^2 \right) \left\| B\hat{U}(x_k) - \hat{U}(y_k) \right\|^2, \tag{20}$$

where

$$\phi_\kappa(x) = \begin{cases} \exp \frac{x}{x-\kappa^2} & 0 \leq x < \kappa^2 \\ 0 & x \geq \kappa^2 \end{cases}$$

is the bump function and  $\kappa > 0$  determines the size of the neighbourhood of the invariant manifold  $\mathcal{M}$  that we take into account.

**Remark 14** The approximation (16) can be made more accurate if we allow matrix  $U^\perp$  to vary with the parameter of the manifold  $\mathbf{z} = U(\mathbf{x})$ . In this case the encoder becomes

$$\check{U}(\mathbf{x}) = U^\perp(U(\mathbf{x}))\mathbf{x} - W_0(U(\mathbf{x})).$$

This does increase computational costs, but not nearly as much as if we were calculating a globally accurate invariant foliation. For our example problems, we find that this extension is not necessary.

**Remark 15** It is also possible to eliminate a-priori calculation of  $U$ . We can assume that  $U$  is a linear map, such that  $UU^* = I$  and treat it as an unknown in representation (15). The assumption that  $U$  is linear makes sense if we limit ourselves to a small neighbourhood of the invariant manifold  $\mathcal{M}$  by setting  $\kappa < \infty$  in (20), as we have already assumed a linear dynamics among the leaves of the associated foliation  $\hat{\mathcal{F}}$  given by linear map  $B$ . Once  $B$ ,  $U$ ,  $U^\perp$  and  $W_0$  are found, map  $S$  can also be fitted to the invariance Eq. (2). The equation to fit  $S$  to data is

$$\arg \min_S \sum_{k=1}^N \|\mathbf{x}_k\|^{-2} \phi_\kappa \left( \|\hat{U}(\mathbf{x}_k)\|^2 \right) \|U\mathbf{y}_k - S(U\mathbf{x}_k)\|^2,$$

which is a straightforward linear least squares problem, if  $S$  is linear in its parameters. This approach will be further explored elsewhere.

### 3 Instantaneous Frequencies and Damping Ratios

Instantaneous damping ratios and frequencies are usually defined with respect to a model that is fitted to data (Jin et al. 2020). Here we take a similar approach and stress that these quantities only make sense in an Euclidean frame and not in the nonlinear frame of an invariant manifold or foliation. The geometry of a manifold or foliation depends on an arbitrary parametrisation; hence, uncorrected results are not unique. Many studies mistakenly use nonlinear coordinate systems, for example one by the present author (Szalai et al. 2017) and colleagues (Breunung and Haller 2018; Ponsioen et al. 2018). Such calculations are only asymptotically accurate near the equilibrium. Here we describe how to correct this error.

We assume a two-dimensional invariant manifold  $\mathcal{M}$ , parametrised by a decoder  $W$  in polar coordinates  $r, \theta$ . The invariance Eq. (3) for the decoder  $W$  can be written as

$$\mathbf{W}(R(r), \theta + T(r)) = \mathbf{F}(\mathbf{W}(r, \theta)). \quad (21)$$

Without much thinking (as described in (Szalai et al. 2017; Szalai 2020)), the instantaneous frequency and damping could be calculated as

$$\omega(r) = T(r) \quad [\text{rad/step}], \quad (22)$$

$$\zeta(r) = -\frac{\log \frac{R(r)}{r}}{\omega(r)} \quad [-], \quad (23)$$

respectively. The instantaneous amplitude is a norm  $A(r) = \|\mathbf{W}(r, \cdot)\|_A$ , for example

$$\|\mathbf{f}\|_A = \sqrt{\langle \mathbf{f}, \mathbf{f} \rangle_A}, \quad \langle \mathbf{f}, \mathbf{f} \rangle_A = \frac{1}{2\pi} \int_0^{2\pi} \langle \mathbf{f}(\theta), \mathbf{f}(\theta) \rangle_X d\theta, \quad (24)$$

where  $\langle \cdot, \cdot \rangle_X$  is the inner product on vector space  $X$ .

The frequency and damping ratio values are only accurate if there is a linear relation between  $\|\mathbf{W}(r, \cdot)\|_A$  and  $r$ , for example

$$\|\mathbf{W}(r, \cdot)\|_A = r \quad (25)$$

and the relative phase between two closed curves satisfies

$$\arg \min_{\gamma} \|\mathbf{W}(r_1, \cdot) - \mathbf{W}(r_2, \cdot + \gamma)\|_A = 0. \quad (26)$$

Equation (25) means that the instantaneous amplitude of the trajectories on manifold  $\mathcal{M}$  is the same as parameter  $r$ , hence the map  $r \mapsto R(r)$  determines the change in amplitude. Equation (26) stipulates that the parametrisation in the angular variable is such that there is no phase shift between the closed curves  $\mathbf{W}(r_1, \cdot)$  and  $\mathbf{W}(r_2, \cdot)$  for  $r_1 \neq r_2$ . If there would be a phase shift  $\gamma$ , a trajectory that within a period moves from amplitude  $r_1$  to  $r_2$ , would misrepresent its instantaneous period of vibration by phase  $\gamma$ , hence the frequency given by  $T(r)$  would be inaccurate. In fact, one can set a continuous phase shift  $\gamma(r)$  among the closed curves  $\mathbf{W}(r, \cdot)$ , such that the frequency given by  $T(r)$  has a prescribed value. The following result provides accurate values for instantaneous frequencies and damping ratios.

**Proposition 16** *Assume a decoder  $\mathbf{W} : [0, r_1] \times [0, 2\pi] \rightarrow X$  and functions  $R, T : [0, r_1] \rightarrow \mathbb{R}$  such that they satisfy invariance Eq. (21).*

1. *A new parametrisation  $\tilde{\mathbf{W}}$  of the manifold generated by  $\mathbf{W}$  that satisfies the constraints (25) and (26) is given by*

$$\tilde{\mathbf{W}}(r, \theta) = \mathbf{W}(t, \theta + \gamma(t)), \quad t = \kappa^{-1}(r),$$



where

$$\gamma(r) = - \int_0^r \frac{\int_0^{2\pi} \langle D_1 \mathbf{W}(\rho, \theta), D_2 \mathbf{W}(\rho, \theta) \rangle_X d\theta}{\int_0^{2\pi} \langle D_2 \mathbf{W}(\rho, \theta), D_2 \mathbf{W}(\rho, \theta) \rangle_X d\theta} d\rho, \tag{27}$$

$$\kappa(r) = \sqrt{\frac{1}{2\pi} \int_0^{2\pi} \langle \mathbf{W}(r, \theta), \mathbf{W}(r, \theta) \rangle_X d\theta} \tag{28}$$

and  $\langle \cdot, \cdot \rangle_X$  is the inner product on vector space  $X$ .

2. The instantaneous natural frequency and damping ratio are calculated as

$$\omega(r) = T(t) + \gamma(t) - \gamma(R(t)) \quad [\text{rad/step}], \tag{29}$$

$$\zeta(r) = - \frac{\log[r^{-1} \kappa(R(t))]}{\omega(r)} \quad [-]. \tag{30}$$

where  $t = \kappa^{-1}(r)$ .

**Proof** A proof is given in ‘‘Appendix C’’. □

**Remark 17** The transformed expressions (29), (30) for the instantaneous frequency and damping ratio show that any instantaneous frequency can be achieved for all  $r > 0$  if  $R(r) \neq r$  by choosing appropriate functions  $\rho, \gamma$ . For example, zero frequency is achieved by solving

$$\gamma(r) = \gamma(R(r)) - T(r), \tag{31}$$

which is a functional equation. For an  $\epsilon > 0$ , fix  $\gamma(R(\epsilon)) = 0, \gamma(\epsilon) = T(\epsilon)$  and some interpolating values in the interior of the interval  $[R(\epsilon), \epsilon]$ , then use contraction mapping to arrive at a unique solution for function  $\gamma$ .

**Remark 18** The same calculation applies to vector fields,  $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x})$ , but the final result is somewhat different. Assume a decoder  $\mathbf{W} : [0, r_1] \times [0, 2\pi] \rightarrow X$  and functions  $R, T : [0, r_1] \rightarrow \mathbb{R}$  such that they satisfy the invariance equation

$$D_1 \mathbf{W}(r, \theta) R(r) + D_2 \mathbf{W}(r, \theta) T(r) = \mathbf{f}(\mathbf{W}(r, \theta)). \tag{32}$$

The instantaneous natural frequency and damping ratio is calculated by

$$\begin{aligned} \omega(r) &= T(t) - \dot{\gamma}(t) R(t) && [\text{rad/unit time}], \\ \zeta(r) &= - \frac{\dot{\kappa}(t) R(t)}{r\omega(r)} && [-], \end{aligned}$$

where  $t = \kappa^{-1}(r)$ . All other quantities are as in proposition 16. A proof is given in ‘‘Appendix C’’.

**Remark 19** Note that proposition 16 also applies if we create a different measure of amplitude, for example  $\hat{\mathbf{W}}(r, \theta) = \mathbf{w}^* \cdot \mathbf{W}(r, \theta)$ , where  $\mathbf{w}^*$  is a linear map. Indeed, in the proof we did not use that  $\mathbf{W}$  is a manifold immersion, it only served as Euclidean coordinates of points on the invariant manifold. Hence, the transformed function  $\hat{\mathbf{W}}$  gives us the same frequencies and damping ratios as  $\mathbf{W}$ , but at linearly scaled amplitudes.

The following example illustrates that if a system is linear in a nonlinear coordinate system, we can recover the actual instantaneous damping ratios and frequencies using proposition 16. This is the case of Koopman eigenfunctions (as in remark 8), or normal form transformations where all the nonlinear terms are eliminated.

**Example 20** Let us consider the linear map

$$\begin{aligned} r_{k+1} &= e^{\zeta_0 \omega_0} r_k \\ \theta_{k+1} &= \theta_k + \omega_0 \end{aligned} \quad (33)$$

and the corresponding nonlinear decoder

$$\mathbf{W}(r, \theta) = \begin{pmatrix} r \cos \theta - \frac{1}{4} r^3 \cos^3 \theta \\ r \sin \theta + \frac{1}{2} r^3 \cos^3 \theta \end{pmatrix}. \quad (34)$$

In terms of the polar invariance Eq. (21), the linear map (33) translates to  $T(r) = \omega_0$  and  $R(r) = e^{\zeta_0 \omega_0} r$ . If we disregard the nonlinearity of  $\mathbf{W}$ , the instantaneous frequency and the instantaneous damping ratio of our hypothetical system would be constant, that is  $\omega(r) = \omega_0$  and  $\zeta(r) = \zeta_0$ . Using proposition 16, we calculate the effect of  $\mathbf{W}$  and find that

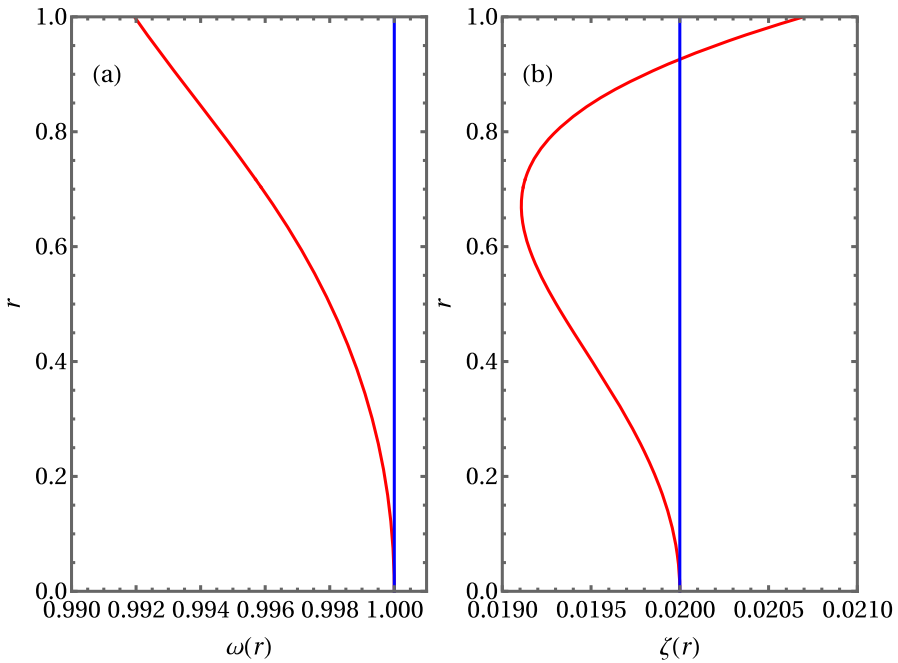
$$\begin{aligned} \gamma(r) &= -\frac{2}{\sqrt{19}} \left( \tan^{-1} \left( \frac{15r^2 - 8}{8\sqrt{19}} \right) + \cot^{-1}(\sqrt{19}) \right), \\ \kappa(r) &= r^2 - \frac{3r^4}{16} + \frac{25r^6}{256}. \end{aligned}$$

Finally, we plot expressions (29) and (30) in Fig. 6a, b, respectively. It can be seen that frequencies and damping ratios change with the vibration amplitude (red lines), but they are constant without taking the decoder (34) into account (blue lines).

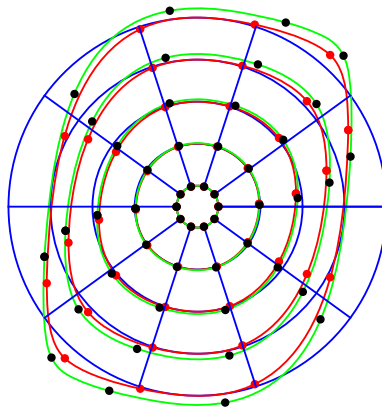
The geometry of the re-parametrisation is illustrated in Fig. 7.

## 4 ROM Identification Procedures

Here we describe our methodology of finding invariant foliations, manifolds and autoencoders. These steps involve methods described so far and further methods from the appendices. First we start with finding an invariant foliation together with the invariant manifold that has the same dynamics as the foliation.



**Fig. 6** The instantaneous frequency **a** and instantaneous damping of system (33) together with the decoder (34). The blue line represents the naive calculation just from system (33), and the red lines represent the corrected values (29) and (30) (Color figure online)



**Fig. 7** Geometry of the decoder (34). The blue grid represents the polar coordinate system with maximum radius  $r = 1$ . The red curves are the images of the blue concentric circles under the decoder  $\mathcal{W}$ . The red dots are images of the intersection points of the blue circles with the blue radial lines under  $\mathcal{W}$ . The red dots do not fall onto the radial blue lines, which indicates a phase shift. The green curves correspond to the images of the blue concentric circles under the re-parametrised decoder  $\tilde{\mathcal{W}}$  for the same parameters as the red curves. The amplitudes of the green curves are now the same as the amplitude of the corresponding blue curves. Due to the re-parametrisation, there is no phase shift between the black dots and the blue radial lines, on average (Color figure online)

- F1. If the data are not from the full state space, especially when it is a scalar signal, we use a state-space reconstruction technique as described in ‘‘Appendix B’’. At this point we have data points  $\mathbf{x}_k, \mathbf{y}_k \in X, k = 1, \dots, N$ .
- F2. To ensure that the solution of (7) converges to the desired foliation, we calculate a linear approximation of the foliation. We only consider a small neighbourhood of the fixed point, where the dynamics is nearly linear. Hence, we define the index set  $\mathcal{I}_\rho = \{k \in \{1, \dots, N\} : |\mathbf{x}_k| < \rho\}$  with  $\rho$  sufficiently small, but large enough that it encompasses enough data for linear parameter estimation.
1. First we fit a linear model to the data using a create a least-squares method (Boyd and Vandenberghe 2018). The linear map is assumed to be  $\mathbf{y}_k = \mathbf{A}\mathbf{x}_k$ , where the coefficient matrix is calculated as

$$\begin{aligned} \mathbf{K} &= \sum_{k \in \mathcal{I}_\rho} |\mathbf{x}_k|^{-2} \mathbf{x}_k \otimes \mathbf{x}_k^T, \\ \mathbf{L} &= \sum_{k \in \mathcal{I}_\rho} |\mathbf{x}_k|^{-2} \mathbf{y}_k \otimes \mathbf{x}_k^T, \\ \mathbf{A} &= \mathbf{L}\mathbf{K}^{-1}. \end{aligned}$$

Matrix  $\mathbf{A}$  approximates the Jacobian  $D\mathbf{F}(\mathbf{0})$ , which then can be used to identify the invariant subspaces  $E^*$  and  $E$ .

2. The linearised version of foliation invariance (2) and manifold invariance (3) are

$$\begin{aligned} \mathbf{S}_1 \mathbf{U}_1 &= \mathbf{U}_1 \mathbf{A} \text{ and} \\ \mathbf{W}_1 \mathbf{S}_1 &= \mathbf{A} \mathbf{W}_1, \end{aligned}$$

respectively. We also need to calculate the linearised version of our locally defined foliation (16), that is

$$\mathbf{B}\mathbf{U}_1^\perp = \mathbf{U}_1^\perp \mathbf{A}.$$

To find the unknown linear maps  $\mathbf{U}_1, \mathbf{W}_1, \mathbf{U}_1^\perp, \mathbf{S}_1$  and  $\mathbf{B}$  from  $\mathbf{A}$ , let us calculate the using real Schur decomposition of  $\mathbf{A}$ , that is  $\mathbf{A}\mathbf{Q} = \mathbf{Q}\mathbf{H}$  or  $\mathbf{Q}^T \mathbf{A} = \mathbf{H}\mathbf{Q}^T$ , where  $\mathbf{Q}$  is unitary and  $\mathbf{H}$  is in an upper Hessenberg matrix, which is zero below the first subdiagonal. The Schur decomposition is calculated (or rearranged) such that the first  $\nu$  column vectors,  $\mathbf{q}_1, \dots, \mathbf{q}_\nu$  of  $\mathbf{Q}$  span the required right invariant subspace  $E$  and correspond to eigenvalues  $\mu_1, \dots, \mu_\nu$ . Therefore we find that  $\mathbf{W}_1 = [\mathbf{q}_1, \dots, \mathbf{q}_\nu]$ . In addition, the last  $n - \nu$  column vectors of  $\mathbf{Q}$ , when transposed, define a left-invariant subspace of  $\mathbf{A}$ . Therefore we define  $\mathbf{U}_1^\perp = [\mathbf{q}_{\nu+1}, \dots, \mathbf{q}_n]^T$  and  $\mathbf{B}_1 = \mathbf{U}_1^\perp \mathbf{A} \mathbf{U}_1^{\perp T}$ , which provides the initial guesses  $\mathbf{U}^\perp \approx \mathbf{U}_1^\perp$  and  $\mathbf{B} \approx \mathbf{B}_1$  in optimisation (20). In order to find the left-invariant subspace of  $\mathbf{A}$  corresponding to the selected eigenvalues, we rearrange the Schur decomposition into  $\hat{\mathbf{Q}}^T \mathbf{A} = \hat{\mathbf{H}} \hat{\mathbf{Q}}^T$ , where

$\hat{Q} = [\hat{q}_1, \dots, \hat{q}_n]$  (using `ordschur` in Julia or Matlab) such that  $\mu_1, \dots, \mu_\nu$  now appear last in the diagonal of  $\hat{H}$  at indices  $\nu + 1, \dots, n$ . This allows us to define  $U_1 = [\hat{q}_{n-\nu}, \dots, \hat{q}_n]^T$  and  $S_1 = U_1 A U_1^T$  which provides the initial guess for  $DU(\mathbf{0}) = U_1$  and  $DS(\mathbf{0}) = S_1$ .

- F3. We solve the optimisation problem (7) with the initial guess  $DU(\mathbf{0}) = U_1$  and  $DS(\mathbf{0}) = S_1$  as calculated in the previous step. We also prescribe the constraints that  $DU(\mathbf{0})(DU(\mathbf{0}))^T = I$  and that  $U(W_1 z)$  is linear in  $z$ . The numerical representation of the encoder  $U$  is described in “Appendix A.3”.
- F4. We perform a normal form transformation on map  $S$  and transform  $U$  into the new coordinates. The normal form  $S_n$  satisfies the invariance equation  $U_n \circ S = S_n \circ U_n$  (c.f. Equation (2)), where  $U_n : Z \rightarrow Z$  is a nonlinear map. We then replace  $U$  with  $U_n \circ U$  and  $S$  with  $S_n$  as our ROM. This step is optional and only required if we want to calculate instantaneous frequencies and damping ratios. In a two-dimensional coordinate system, where we have a complex conjugate pair of eigenvalues  $\mu_1 = \bar{\mu}_2$ , the real valued normal form is

$$\begin{pmatrix} z_1 \\ z_2 \end{pmatrix}_{k+1} = \begin{pmatrix} z_1 f_r(z_1^2 + z_2^2) - z_2 f_i(z_1^2 + z_2^2) \\ z_1 f_i(z_1^2 + z_2^2) + z_2 f_r(z_1^2 + z_2^2) \end{pmatrix}, \tag{35}$$

which leads to the polar form (21) with  $R(r) = r\sqrt{f_r^2(r^2) + f_i^2(r^2)}$  and  $T(r) = \tan^{-1} \frac{f_i(r^2)}{f_r(r^2)}$ . This normal form calculation is described in Szalai (2020).

- F5. To find the invariant manifold, we calculate a locally defined foliation (15) and solve the optimisation problem (20) to find the decoder  $W$  of invariant manifold  $\mathcal{M}$ . The initial guess in problem (20) is such that  $U^\perp = U_1^\perp$  and  $B = B_1$ . We also need to set a  $\kappa$  parameter, which is assumed to be  $\kappa = 0.2$  throughout the paper. We have found that results are not sensitive to the value of kappa except for extreme choices, such as 0,  $\infty$ .
- F6. In case of an oscillatory dynamics in a two-dimensional ROM, we recover the actual instantaneous frequencies and damping ratios using proposition (16).

The procedure for the Koopman eigenfunction calculation is the same as steps F1-F6, except that  $S$  is assumed to be linear. To identify an autoencoder, we use the same setup as in Cenedese et al. (2022). The numerical representation of the autoencoder is described in “Appendix A.2”. We carry out the following steps

- AE1. We identify  $W_1, S_1$  as in step F2 and set  $U = W_1^T$  and  $DS(\mathbf{0}) = S_1$
- AE2. Solve the optimisation problem

$$\arg \min_{U, W_{nl}} \sum_{k=1}^N \|y_k\|^{-2} \|W(U y_k) - y_k\|^2,$$

which tries to ensure that  $W \circ U$  is the identity, and finally solve

$$\arg \min_S \sum_{k=1}^N \|x_k\|^{-2} \|W(S(U(x_k))) - y_k\|^2.$$

- AE3. Perform a normal form transformation on  $S$  by seeking the simplest  $S_n$  that satisfies  $W_n \circ S_n = S \circ W_n$ . This is similar to step F4, except that the normal form is in the style of the invariance equation of a manifold (3).
- AE4. Same as step F6, but applied to nonlinear map  $S_n$  and decoder  $W \circ W_n$ .

## 5 Examples

We are considering three examples. The first is a caricature model to illustrate all the techniques discussed in this paper and why certain techniques fail. The second example is a series of synthetic data sets with higher dimensionality, to illustrate the methods in more detail using a polynomial representations of the encoder with HT tensor coefficients (see ‘‘Appendix A.3’’). This example also illustrates two different methods to reconstruct the state space of the system from a scalar measurement. The final example is a physical experiment of a jointed beam, where only a scalar signal is recorded and we need to reconstruct the state space with our previously tested technique.

### 5.1 A Caricature Model

To illustrate the performance of autoencoders, invariant foliations and locally defined invariant foliations, we construct a simple two-dimensional map  $F$  with a node-type fixed point using the expression

$$F(x) = V(AV^{-1}(x)), \tag{36}$$

where

$$A = \begin{pmatrix} \frac{9}{10} & 0 \\ 0 & \frac{4}{5} \end{pmatrix},$$

the near-identity coordinate transformation is

$$V(x) = \begin{pmatrix} x_1 + \frac{1}{4}(x_1^3 - 3(x_1 - 1)x_2x_1 + 2x_2^3 + (5x_1 - 2)x_2^2) \\ x_2 + \frac{1}{4}(2x_2^3 + (2x_1 - 1)x_2^2 - x_1^2(x_1 + 2)) \end{pmatrix},$$

and the state vector is defined as  $x = (x_1, x_2)$ . In a neighbourhood of the origin, transformation  $V$  has a unique inverse, which we calculate numerically. Map  $F$  is constructed such that we can immediately identify the smoothest (hence unique) invariant manifolds corresponding to the two eigenvalues of  $DF(0)$  as

$$\begin{aligned} \mathcal{M}_{\frac{9}{10}} &= \{V(z, 0), z \in \mathbb{R}\}, \\ \mathcal{M}_{\frac{4}{5}} &= \{V(0, z), z \in \mathbb{R}\}. \end{aligned}$$

We can also calculate the leaves of the two invariant foliations as

$$\mathcal{L}_{\frac{9}{10},z} = \{V(z, \bar{z}), \bar{z} \in \mathbb{R}\}, \tag{37}$$

$$\mathcal{L}_{\frac{4}{5},z} = \{V(\bar{z}, z), \bar{z} \in \mathbb{R}\}. \tag{38}$$

To test the methods, we created 500 times 30 points long trajectories with initial conditions sampled from a uniform distribution over the rectangle  $[-\frac{4}{5}, \frac{4}{5}] \times [-\frac{1}{4}, -\frac{1}{4}]$  to fit our ROM to.

We first attempt to fit an autoencoder to the data. We assume that the encoder, decoder and the nonlinear map  $S$  are

$$U(x_1, x_2) = x_1, \tag{39}$$

$$W(z) = (z, h(z))^T, \tag{40}$$

$$S(z) = \lambda(z), \tag{41}$$

where  $\lambda, h$  are polynomials of order-5. Our expressions already contain the invariant subspace  $E = \text{span}(1, 0)^T$ , which should make the fitting easier. Finally, we solve the optimisation problem (8). The result of the fitting can be seen in Fig. 8a as depicted by the red curve. The fitted curve is more dependent on the distribution of data than the actual position of the invariant manifold, which is represented by the blue dashed line in Fig. 8a. Various other expressions for  $U$  and  $W$  were also tried that do not assume the direction of the invariant subspace  $E$  with similar results.

To calculate the invariant foliation in the horizontal direction, we assume that

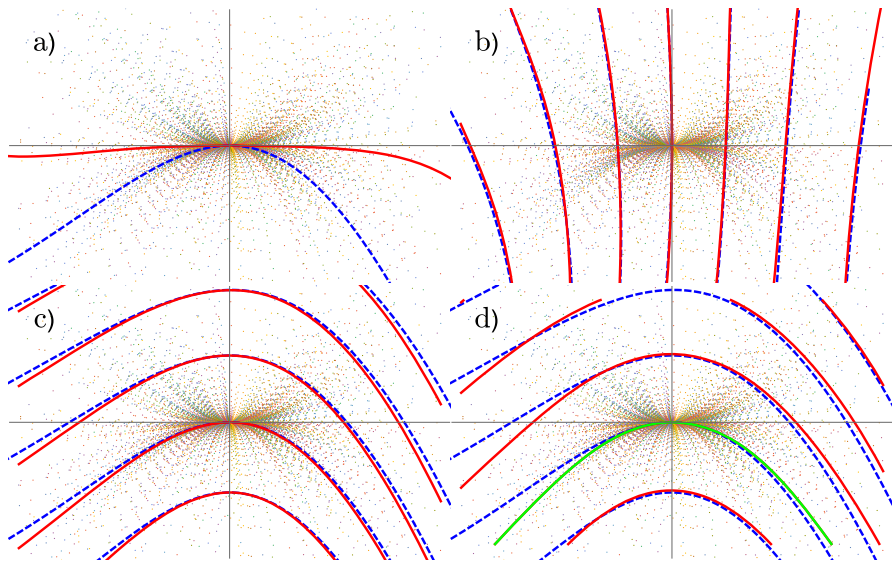
$$\left. \begin{aligned} U(x_1, x_2) &= x_1 + u(x_1, x_2) \\ S(z) &= \lambda z \end{aligned} \right\}, \tag{42}$$

where  $u$  is an order-5 polynomial which lacks the constant and linear terms. The exact expression of  $U$  is not a polynomial, because it is the second coordinate of the inverse of function  $V$ . The fitting is carried out by solving the optimisation problem (7). The result can be seen in Fig. 8b, where the red curves are contour plots of the identified encoder  $U$  and the dashed blue lines are the leaves as defined by Eq. (37). Figure 8c is produced in the same way as Fig. 8b, except that the encoder is defined as  $U(x_1, x_2) = x_2 + h(x_1, x_2)$  and the blue lines are the leaves given by (38).

As we have discussed in Sect. 2.2, a locally defined encoder can also be constructed from a decoder. In the expression of the encoder (15) we take

$$W_0(z) = h(z)$$

and  $U^\perp = (0, 1)$ , where  $h$  is an order-9 polynomial without constant and linear terms. The expressions for  $U$  and  $S$  were already found as (42), hence our approximate encoder becomes



**Fig. 8** Identifying invariant objects in Eq. (36). The data contain 500 trajectories of length 30 with initial conditions picked from a uniform distribution; **a** fitting an autoencoder (red continuous curve) does not reproduce the invariant manifold (blue dashed curve), instead it follows the distribution of data; **b** invariant foliation in the horizontal direction; **c** invariant foliation in the vertical direction; **d** an invariant manifold is calculated as the leaf of a locally defined invariant foliation. The green line is the invariant manifold. Each diagram represents the box  $\left[-\frac{4}{5}, \frac{4}{5}\right] \times \left[-\frac{1}{4}, -\frac{1}{4}\right]$ ; the axes labels are intentionally hidden (Color figure online)

$$\hat{U}(\mathbf{x}) = x_2 - h(x_1 + u(x_1, x_2)).$$

We solve the optimisation problem (20) with  $\kappa = 0.13$ . We do not reconstruct the decoder  $W$ , as it is straightforward to plot the level surfaces of  $\hat{U}$  directly. The result can be seen in Fig. 8d, where the green line is the approximate invariant manifold (the zero level surface of  $\hat{U}$ ) and the red lines are other level surfaces of  $\hat{U}$ .

In conclusion, this simple example shows that only invariant foliations can be fitted to data and autoencoders give spurious results.

## 5.2 A Ten-Dimensional System

To create a numerically challenging example, we construct a ten-dimensional differential equation from five decoupled second-order nonlinear oscillators using two successive coordinate transformations. The system of decoupled oscillators is denoted by  $\dot{\mathbf{x}} = \mathbf{f}_0(\mathbf{x})$ , where the state variable is in the form of

$$\mathbf{x} = (r_1, \dots, r_5, \theta_1, \dots, \theta_5)$$



and the dynamics is given by

$$\begin{aligned}
 \dot{r}_1 &= -\frac{1}{500}r_1 + \frac{1}{100}r_1^3 - \frac{1}{10}r_1^5, & \dot{\theta}_1 &= 1 + \frac{1}{4}r_1^2 - \frac{3}{10}r_1^4, \\
 \dot{r}_2 &= -\frac{e}{500}r_2 - \frac{1}{10}r_2^5, & \dot{\theta}_2 &= e + \frac{3}{20}r_2^2 - \frac{1}{5}r_2^4, \\
 \dot{r}_3 &= -\frac{1}{50}\sqrt{\frac{3}{10}}r_3 + \frac{1}{100}r_3^3 - \frac{1}{10}r_3^5, & \dot{\theta}_3 &= \sqrt{30} + \frac{9}{50}r_3^2 - \frac{19}{100}r_3^4, \\
 \dot{r}_4 &= -\frac{1}{500}\pi^2r_4 + \frac{1}{100}r_4^3 - \frac{1}{10}r_4^5, & \dot{\theta}_4 &= \pi^2 + \frac{4}{25}r_4^2 - \frac{17}{100}r_4^4, \\
 \dot{r}_5 &= -\frac{13}{500}r_5 + \frac{1}{100}r_5^3, & \dot{\theta}_5 &= 13 + \frac{4}{25}r_5^2 - \frac{9}{50}r_5^4.
 \end{aligned}
 \tag{43}$$

The first transformation brings the polar form of Eq. (43) into Cartesian coordinates using the transformation  $\mathbf{y} = \mathbf{g}(\mathbf{x})$ , which is defined by  $y_{2k-1} = r_k \cos \theta_k$  and  $y_{2k} = r_k \sin \theta_k$ . Finally, we couple all variables using the second nonlinear transformation  $\mathbf{y} = \mathbf{h}(\mathbf{z})$ , which reads

$$\begin{aligned}
 y_1 &= z_1 + z_3 - \frac{1}{12}z_3z_5, & y_2 &= z_2 - z_3, \\
 y_3 &= z_3 + z_5 - \frac{1}{12}z_5z_7, & y_4 &= z_4 - z_5, \\
 y_5 &= z_5 + z_7 + \frac{1}{12}z_7z_9, & y_6 &= z_6 - z_7, \\
 y_7 &= z_7 + z_9 - \frac{1}{12}z_1z_9, & y_8 &= z_8 - z_9, \\
 y_9 &= z_9 + z_1 - \frac{1}{12}z_3z_1, & y_{10} &= z_{10} - z_1,
 \end{aligned}
 \tag{44}$$

and where  $\mathbf{y} = (y_1, \dots, y_{10})$  and  $\mathbf{z} = (z_1, \dots, z_{10})$ . The two transformations give us the differential equation  $\dot{\mathbf{z}} = \mathbf{f}(\mathbf{z})$ , where

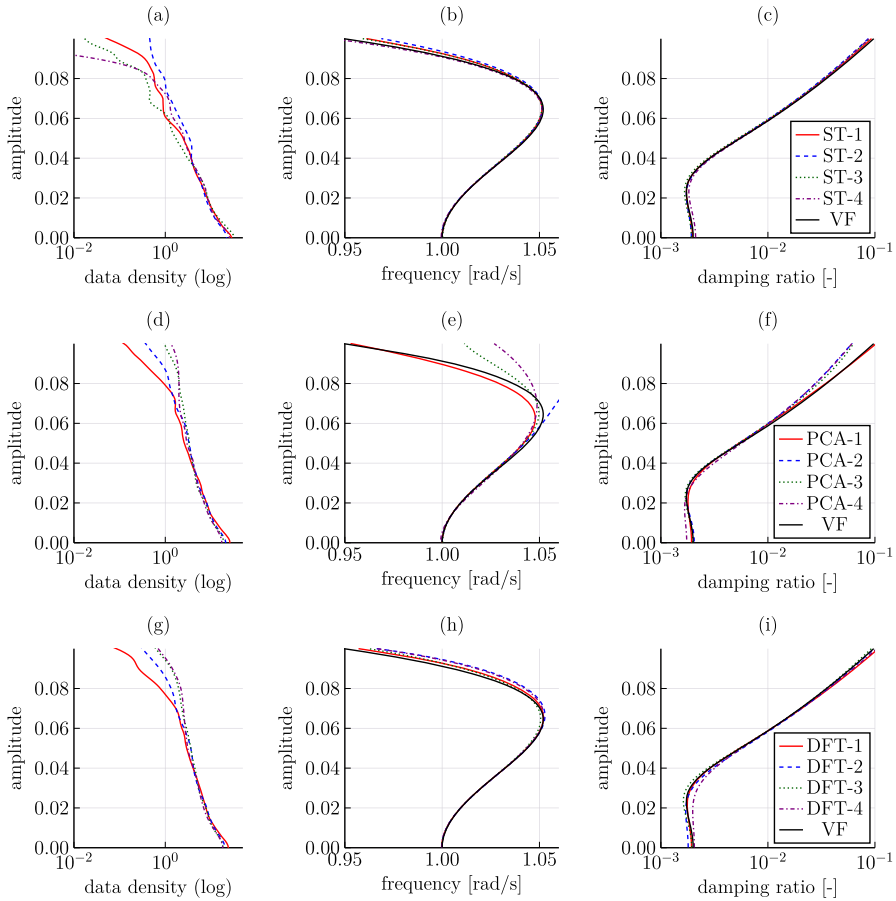
$$\mathbf{f}(\mathbf{z}) = \left[ D\mathbf{g}^{-1}(\mathbf{h}(\mathbf{z})) D\mathbf{h}(\mathbf{z}) \right]^{-1} \mathbf{f}_0\left(\mathbf{g}^{-1}(\mathbf{h}(\mathbf{z}))\right).
 \tag{45}$$

The natural frequencies of our system at the origin are

$$\omega_1 = 1, \omega_2 = e, \omega_3 = \sqrt{30}, \omega_4 = \pi^2, \omega_5 = 13$$

and the damping ratios are the same  $\zeta_1 = \dots = \zeta_5 = 1/500$ .

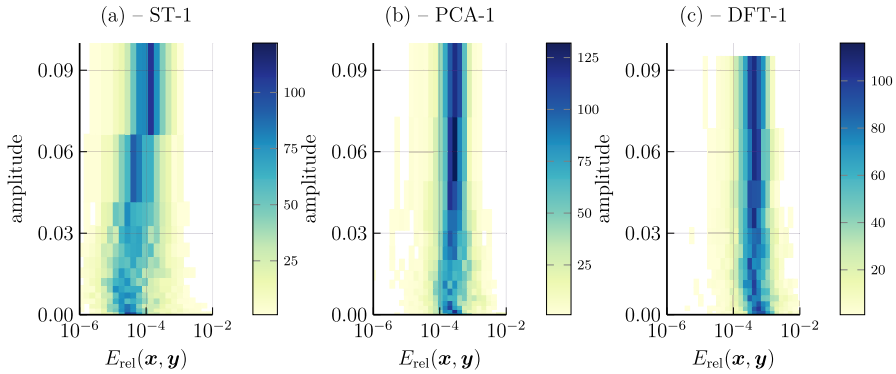
We select the first natural frequency to test various methods. We also test the methods on three types of data. Firstly, full state space information is used, secondly the state space is reconstructed from the signal  $\xi_k = \frac{1}{10} \sum_{j=1}^{10} z_{k,j}$  using principal component analysis (PCA) as described in ‘‘Appendix B.1’’ with 16 PCA components, and finally the state space is reconstructed from  $\xi_k$  using a discrete Fourier transform (DFT) as described in ‘‘Appendix B.2’’. When data are recorded in state space form, 1000 trajectories 16 points long each with time step  $\Delta T = 0.1$  were created by numerically solving (45). Initial conditions were sampled from unit balls of radius 0.8, 1.0, 1.2 and 1.4 about the origin. The Euclidean norm of the initial conditions was uniformly distributed. The four data sets are labelled ST-1, ST-2, ST-3, ST-4 in the diagrams. For state space reconstruction, 100 trajectories, 3000 points each, with time step  $\Delta T = 0.01$  were created by numerically solving (45). The initial conditions



**Fig. 9** Instantaneous frequencies and damping ratios of differential Eq. (45) using invariant foliations. The first column shows the data density with respect to the amplitude  $A(r)$ , the second column is the instantaneous frequency and the third column is the instantaneous damping. The first row corresponds to state space data, the second row shows PCA reconstructed data, and the third row is DFT reconstructed data. The data density is controlled by sampling initial conditions from different sized neighbourhoods of the origin

for these data were similarly sampled from unit balls of radius 0.8, 1.0, 1.2 and 1.4 about the origin, such that the Euclidean norm of the initial conditions are uniformly distributed. The PCA reconstructed data are labelled PCA-1, PCA-2, PCA-3, PCA-4, and the DFT reconstructed data are labelled DFT-1, DFT-2, DFT-3, DFT-4.

The amplitude for each ROM is calculated as  $A(r) = \sqrt{\frac{1}{2\pi} \int (\mathbf{w}^* \cdot \mathbf{W}(r, \theta))^2 d\theta}$ , where  $\mathbf{w}^* = \frac{1}{10}(1, 1, \dots, 1)$  for the state-space data and  $\mathbf{w}^*$  is calculated in “Appendices B.1, B.2” when state-space reconstruction is used. We can also attach an amplitude to each data point  $\mathbf{x}_k$  through the encoder and the decoder. If the ROM assumes the normal form (35), the radial parameter  $r$  is simply calculated as  $r_k = \|\mathbf{U}(\mathbf{x}_k)\|$ , hence the amplitude is  $A(r_k)$ .



**Fig. 10** Histograms of the fitting error (46) for the data sets ST-1, PCA-1 and DFT-1 as a function of vibration amplitude. The distribution appears uniform with respect to the amplitude

Figure 9 shows the result of our calculation for the three types of data. In the first column data density is displayed with respect to amplitude  $A(r)$  in the ROM. Lower amplitudes have higher densities, because trajectories exponentially converge to the origin. In Fig. 9 we also display the identified instantaneous frequencies and damping ratios. The results are then compared to the analytically calculated frequencies and damping ratios labelled by VF.

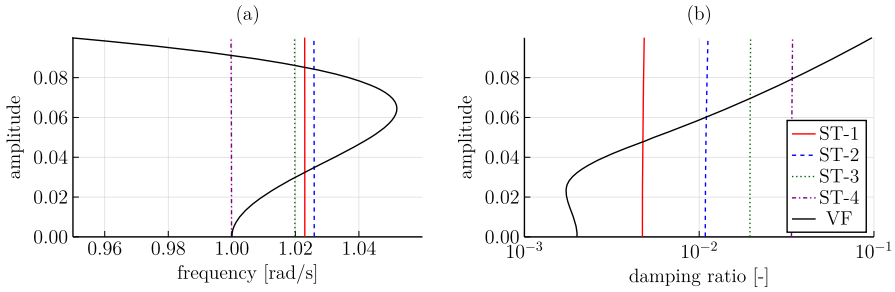
State space data give the closest match to the analytical reference (labelled as VF). We find that the PCA method cannot embed the data in a 10-dimensional space, only an 16-dimensional embedding is acceptable, but still inaccurate. Using a perfect reproducing filter bank (DFT) yields better results, probably because the original signal can be fully reconstructed and we expect a correct state-space reconstruction at small amplitudes. Indeed, the PCA results diverge at higher amplitudes, where the state space reconstruction is no longer valid. The author has also tried non-optimal delay embedding, with inferior results. None of the techniques had any problem with the less the challenging Shaw-Pierre example (Shaw and Pierre 1994; Szalai 2020) (data not shown).

Figure 10 shows the accuracy of the data fit as a function of the amplitude  $A(r)$ . The relative error displayed is defined by

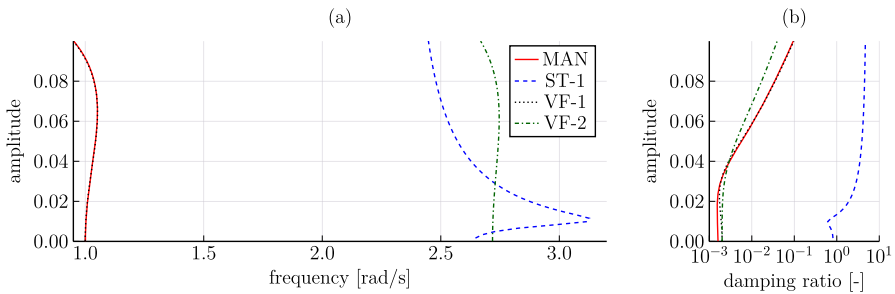
$$E_{rel}(x, y) = \frac{\|S(U(x)) - U(y)\|}{\|U(x)\|}. \tag{46}$$

It turns out that the error is roughly independent of the amplitude, except for data set ST-1, which has lower errors at low amplitudes. The accuracy is the highest for state space data, while the accuracy of the DFT reconstructed data is slightly worse than for the PCA reconstructed data. In contrast, the comparison with the analytically calculated result is worse for the PCA data than for the DFT data. The reason is that PCA reconstruction cannot exactly reproduce the original signal from the identified components while the DFT method can.

When restricting map  $S$  to be linear, we are identifying Koopman eigenfunctions. Despite that linear dynamics is identified we should be able to reproduce the nonlin-



**Fig. 11** ROM by Koopman eigenfunctions. The same quantities are calculated as in Fig. 9, except that map  $S$  is assumed to be linear. There is some variation in the frequencies and damping ratios with respect to the amplitude due to the corrections in Sect. 3, but accurate values could not be recovered as the linear approximation does not account for near internal resonances, as in formula (12)



**Fig. 12** Data analysis by autoencoder. An autoencoder can recover system dynamics if all data are on the invariant manifold. For the solid line MAN, data were artificially forced to be on an a-priori calculated invariant manifold. However if the data are not on an invariant manifold, such as for data set ST-1, the autoencoder calculation is meaningless. The dotted line VF-1 represents the analytic calculation for the first natural frequency, and the dash-dotted VF-2 depicts the analytic calculation for the second natural frequency of vector field (45)

erities as illustrated in Sect. 3. However, we also have near internal resonances as per Eq. (12), which make certain terms of encoder  $U$  large, which are difficult to find by optimisation. The result can be seen in Fig. 11. The identified frequencies and damping ratios show little variation with amplitude and mostly capture the average of the reference values. Fitting the Koopman eigenfunction achieves maximum and average values of  $E_{rel}$  at 8.66% and 0.189% over data set ST-1, respectively. Better accuracy could be achieved using higher rank HT tensor coefficients in the encoder, which would significantly increase the number of model parameters. In contrast, fitting the invariant foliation to the same data set yields maximum and the average values of  $E_{rel}$  at 2.21% and 0.0118%, respectively (also illustrated in Fig. 10a). This better accuracy is achieved with a small number of extra parameters that make the two-dimensional map  $S$  nonlinear.

Knowing that autoencoders are only useful if all the dynamics is on the manifold, we have synthetically created data consisting of trajectories with initial conditions from the invariant manifold of the first natural frequency. We used 800 trajectories, 24 points each with time-step  $\Delta t = 0.2$  starting on the manifold. Fitting an autoencoder

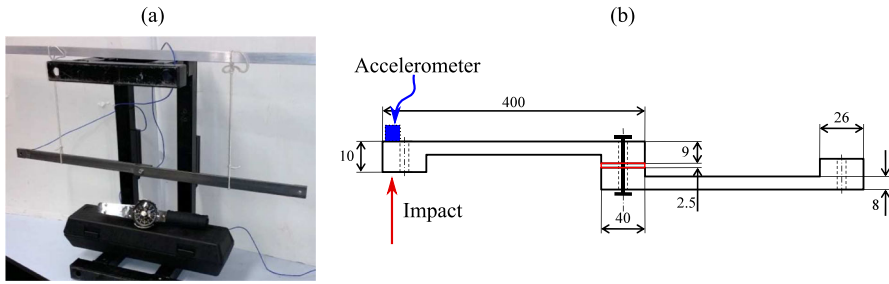
to these data yields a good match in Fig. 12, the corresponding lines are labelled MAN. Then we tried dataset ST-1, that matched the reference best when calculating an invariant foliation. However, our data do not lie on a manifold and it is impossible to make  $W \circ U$  close to the identity on our data. In fact the result (blue dashed line) is closer to the second mode of vibration (green dash-dotted curve), which seems to be the dominant vibration of the system.

### 5.3 Jointed Beam

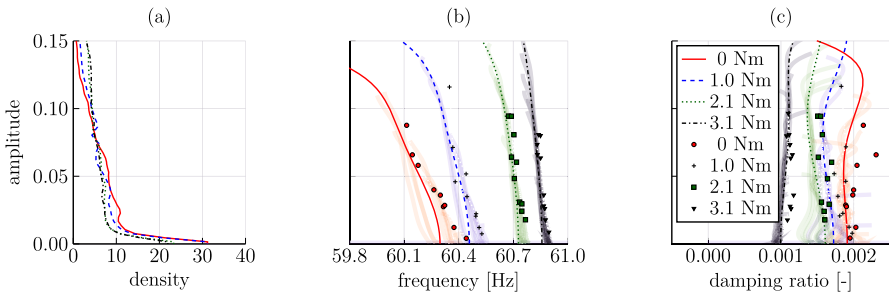
It is challenging to accurately model mechanical friction; hence data-oriented methods can play a huge role in identifying dynamics affected by frictional forces. Therefore we analyse the data published in Titurus et al. (2016). The experimental setup can be seen in Fig. 13. The two halves of the beam were joined together with an M6 bolt. The two interfaces of the beams were sandblasted to increase friction, and a polished steel plate was placed between them; finally the bolt was tightened using four different torques: minimal torque so that the beam does not collapse under its own weight (denoted as 0 Nm), 1 Nm, 2.1 Nm and 3.1 Nm. The free vibration of the steel beam was recorded using an accelerometer placed at the end of the beam. The vibration was initiated using an impact hammer at the position of the accelerometer. Calibration data for the accelerometer are not available. For each torque value a number of 20 s long signals were recorded with sampling frequency of 2048 Hz. The impacts were of different magnitude so that the amplitude dependency of the dynamics could be tracked. In Titurus et al. (2016), a linear model was fitted to each signal and the first five vibration frequencies and damping ratios were identified. These are represented by various markers in Fig. 14. In order to make a connection between the peak impact force and the instantaneous amplitude we also calculated the peak root mean square (RMS) amplitude for signals with the largest impact force for each tightening torque and found that the average conversion factor between the peak RMS amplitude and the peak impact force was 443, which we used to divide the peak force and plot the equivalent peak RMS in Fig. 14. We also band filtered each trajectory with a 511 point FIR filter with 3 dB points at 30 Hz and 75 Hz and estimated the instantaneous frequency and damping ratios from two consecutive vibration cycles, which are then drawn as thick semi-transparent lines in Fig. 14 for each trajectory. It is worth noting that the more friction there is in the system, the less reproducible the frequencies and damping ratios become when using short trajectory segments for estimation.

To calculate the invariant foliation we used a 10-dimensional DFT reconstructed state space, to include all five captured frequencies, as described in “Appendix B.2”. We chose  $\kappa = 0.2$  in the optimisation problem (20) when finding the invariant manifold. The result can be seen in Fig. 14. Since we do not have the ground truth for this system, it is not possible to tell which method is more accurate, especially that our naive alternative calculation (thick semi-transparent lines) displays a wide spread of results.

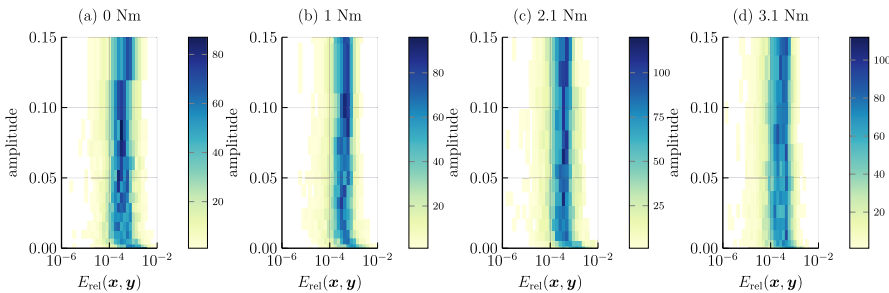
The fitting error to the invariant foliation can be assessed from the histograms in Fig. 15. The distribution of the error is very similar to the synthetic model, which is nearly uniform with respect to the vibration amplitude. It is also clear that there is no



**Fig. 13** **a** Experimental setup. **b** Schematic of the jointed beam. The width of the beam (not shown) is 25 mm. All measurements are in millimetres



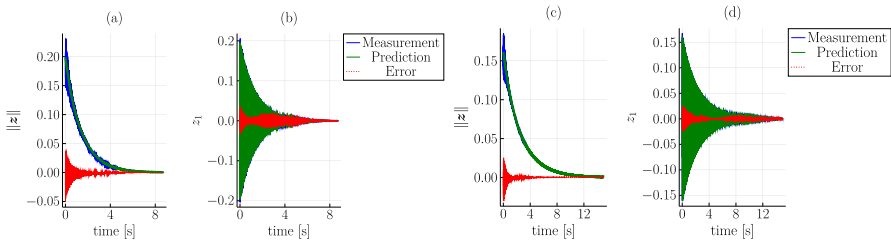
**Fig. 14** Instantaneous frequencies and damping ratios of the jointed beam set-up in Fig. 13. Solid red lines correspond to the invariant foliation of the system with minimal tightening torque, blue dashed lines with tightening torque of 2.1 Nm and green dotted lines with tightening torque of 3.1 Nm. The markers of the same colour show results calculated in Titurus et al. (2016) using curve fitting. **a** Data density with respect to amplitude  $A$  ( $r$ ), **b** instantaneous frequency, **c** instantaneous damping ratio (Color figure online)



**Fig. 15** Histograms of the fitting error (46) for all four tightening torques of the jointed beam. Note that the errors look very similar despite that repeatability for lower tightening torques appears worse in Fig. 14

real difference in the fitting error for different tightening torques, which indicates that frictional dynamics can be accurately characterised using invariant foliations.

As a final test, we also assess whether the measured signal is reproduced by the invariant foliation ( $U, S$ ) in Fig. 16. For this we apply the encoder  $U$  to our original signal  $x_k, k = 1, \dots, N$  and compare this signal to the one produced by the recursion  $z_{k+1} = S(z_k)$ , where  $z_1 = U(x_1)$ . The instantaneous amplitude error in both cases of Fig. 16a, c was mostly due to high frequency oscillations that was not completely



**Fig. 16** Reconstruction error of the foliation. **a, b** Reconstructing the first signal of the series of experimental data with 0 Nm tightening torque. **c, d** Reconstructing the first signal of the series of experimental data with 3.1 Nm tightening torque. The compared values are calculated through the encoder  $z = U(x_k)$  and the reconstructed values are from the iteration  $z_{k+1} = S(z_k)$ . The amplitude error ( $\|z\|$ ) is minimal; however, over time phase error accumulates; hence direct comparison of a coordinate ( $z_1$ ) can look unfavourable

filtered out by the encoder  $U$  at the high amplitudes. The phase error seemed to have only accumulated for the lowest tightening torque (0 Nm) in Fig. 16b. This is to be expected over long trajectories, since the fitting procedure only minimises the prediction error for a single time step. Unfortunately, accumulating phase error is rarely shown in the literature, where only short trajectories are compared, in contrast to the 28685 and 25149 time-steps that are displayed in Fig. 16b, d, respectively.

## 6 Discussion

The main conclusion of this study is that only invariant foliations are suitable for ROM identification from off-line data. Using an invariant foliation avoids the need to use resonance decay (Ehrhardt and Allen 2016), or waiting for the signal to settle near the most attracting invariant manifold (Cenedese et al. 2022), thereby throwing away valuable data. Using invariant foliations can make use of unstructured data with arbitrary initial conditions, such as impact hammer tests. Invariant foliations produce genuine ROMs and not only parametrise a-priori known invariant manifolds, like other methods (Cenedese et al. 2022; Champion et al. 2019; Yair et al. 2017). We have shown that the high-dimensional function required to represent an encoder can be represented by polynomials with compressed tensor coefficients, which significantly reduces the computational and memory costs. Compressed tensors are also amenable to further analysis, such as singular value decomposition (Grasedyck 2010), which gives way to mathematical interpretations of the decoder  $U$ . The low-dimensional map  $S$  is also amenable to normal form transformations, which can be used to extract information such as instantaneous frequencies and damping ratios.

We have tested the related concept of Koopman eigenfunctions, which differs from an invariant foliation in that map  $S$  is assumed to be linear. If there are no internal resonances, Koopman eigenfunctions are theoretically equivalent to invariant foliations. However in numerical settings unresolved near internal resonances become important and therefore Koopman eigenfunctions become inadequate. We have also tried to fit autoencoders to our data (Cenedese et al. 2022), but apart from the artificial case where the invariant manifold was pre-computed, it performed even worse than Koopman eigenfunctions.

Fitting an invariant foliation to data is extremely robust when state space data are available. However, when the state space needed to be reproduced from a scalar signal, the results were not as accurate as we hoped for. While Takens' theorem allows for any generic delay coordinates, in practice a non-optimal choice can lead to poor results. We were expecting that the embedding dimension at least for low-amplitude signals would be the same as the attractor dimension. This is, however, not true if the data also include higher amplitude points. Despite not being theoretically optimal, we have found that perfect reproducing filter banks produce accurate results for low-amplitude signals and at the same time provide a state-space reconstruction with the same dimensionality as that of the attractor. Future work should include exploring various state-space reconstruction techniques in combination with fitting an invariant foliation to data.

We did not fully explore the idea of locally defined invariant foliations in remark 15, which can lead to computationally efficient methods. Further research can also be directed towards cases, where the data are on a high-dimensional submanifold of an even higher-dimensional vector space  $X$ . This is where an autoencoder and invariant foliation may be combined.

**Acknowledgements** I would like to thank Branislav Titurus for supplying the data of the jointed beam. I would also like to thank Sanuja Jayatilake, who helped me collect more data on the jointed beam, which eventually was not needed in this study. Discussions with Alessandra Vizzaccaro, David Barton and his research group provided great inspiration to complete this research. A.V. has also commented on a draft of this manuscript.

**Author Contributions** The author carried out all the work related to the manuscript, including, but not limited to design of the research, writing the paper, preparing figures and writing the necessary computer code.

**Data Availability** The datasets generated and/or analysed during the current study are available in the FMA repository, <https://github.com/rs1909/FMA>.

## Declarations

**Conflict of interest** The author has no relevant financial or non-financial interests to disclose. No funds, grants, or other support was received.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## A Implementation details

In this appendix we deal with how to represent invariant foliations, locally defined invariant foliations and autoencoders. We also describe our specific techniques to carry out optimisation on these representations.



### A.1 Dense Polynomial Representation

The nonlinear map  $S$ , the polynomial part of the locally defined invariant foliation  $W_0$  and the decoder part an autoencoder are represented by dense polynomials, that we define here.

A polynomial  $P : Z \rightarrow X$  of order  $d$  is represented as a linear combination of monomials formed from the coordinates of vectors of  $Z$ . First we define what we mean by a monomial and how we order them. Given a basis in  $Z$  with  $v = \dim Z$ , we can represent each vector  $z \in Z$  by  $z = (z_1, z_2, \dots, z_v)$ . Then we define non-negative integer vectors  $m = (m_1, \dots, m_v) \in \mathbb{N}^v$  and finally define a monomial of  $z$  by  $z^m = z_1^{m_1} z_2^{m_2} \dots z_v^{m_v}$ . We also need an ordered set of integer exponents, which is denoted by

$$\mathfrak{M}_{c,d} = \{m \in \mathbb{N}^v : c \leq m_1 + \dots + m_v \leq d\}.$$

The ordering of  $\mathfrak{M}_{c,d}$  is such that  $m < n$  if there exists  $k \leq v$  and  $m_j < n_j$  for all  $j = 1, \dots, k$ . The cardinality of set  $\mathfrak{M}_{c,d}$  is denoted by  $\#(v, c, d)$ . Therefore we can also write that  $\mathfrak{M}_{c,d} = \{m_1, m_2, \dots, m_{\#(v,c,d)}\}$ . Using the ordered notation of monomials, a polynomial  $P$  containing terms at least order  $c$  up to order- $d$  is represented by a matrix  $\underline{P} \in \mathbb{R}^{n \times \#(v,c,d)}$ , such that

$$P(z) = \underline{P} (z^{m_1}, z^{m_2}, \dots, z^{m_{\#(v,c,d)}})^T.$$

We also call such polynomials order- $(c, d)$  polynomials. For optimisation purposes order- $(c, d)$  polynomials form an Euclidean manifold and therefore no constraint is placed on matrix  $\underline{P}$ .

### A.2 Autoencoder Representation

A polynomial autoencoder of order  $d$  (as defined in Cenedese et al. 2022) is represented by an orthogonal matrix  $\underline{U} \in \mathbb{R}^{n \times v}$  ( $\underline{U}^T \underline{U} = \mathbf{I}$ ) and an order- $(2, d)$  polynomial  $\underline{W}$ , represented by matrix  $\underline{W} \in \mathbb{R}^{n \times \#(v,2,d)}$ . The associated encoder is given by  $\underline{U}(x) = \underline{U}^T x$  and the decoder is given by  $\underline{W}(z) = \underline{U}z + \underline{W}(z)$ , which must satisfy the additional constraint  $\underline{U}(\underline{W}(z)) = z$ , that in terms of our matrices means  $\underline{U}^T \underline{W} = \mathbf{0}$ . In summary, we have the constraints

$$\begin{aligned} \underline{U}^T \underline{U} &= \mathbf{I}, \\ \underline{U}^T \underline{W} &= \mathbf{0}, \end{aligned} \tag{47}$$

which turns the admissible set of matrices  $\underline{U}, \underline{W}$  into a matrix manifold. We call this manifold the orthogonal autoencoder manifold and denote it by  $OAE_{n,v,\#(v,2,d)}$ . In paper (Cenedese et al. 2022), the authors also consider the case when the decoder is decoupled from the encoder. In this case  $\underline{W}(z)$  does not include the orthogonal matrix  $\underline{U}$  and therefore  $\underline{W}(z)$  is represented by matrix  $\underline{W} \in \mathbb{R}^{n \times \#(v,1,d)}$ , which also includes

linear terms. The constraint on matrices  $\underline{U}$ ,  $\underline{W}$  is therefore different

$$\begin{aligned}\underline{U}^T \underline{U} &= \mathbf{I}, \\ \underline{U}^T \underline{W} &= \mathbf{M},\end{aligned}\tag{48}$$

where matrix  $\mathbf{M}$  represents the identity polynomial with respect to the monomials  $\mathfrak{M}_{1,d}$ . We call the resulting matrix manifold the generalised orthogonal autoencoder manifold and denote it by  $GOAE_{M,n,v,\#(v,1,d)}$ . Unfortunately, this generalised autoencoder leads to an ill-defined optimisation problem. Indeed, if all the data are on an invariant manifold (the only sensible case), then the directions encoded by  $\underline{U}$  are not defined by the data, because any transversal direction is equally suitable for projecting the data. In our numerical test  $\underline{U}$  quickly degenerated and led to spurious results. Nevertheless, due to the similar formulation we treat both cases at the same time, because  $OAE_{n,m,l}$  is a particular case of  $GOAE_{M,n,m,l}$  with  $\mathbf{M}$  being equal to zero. The details of how projections and retractions are calculated for  $GOAE_{M,n,m,l}$  are presented in Sect. A.5.1.

### A.3 Compressed Polynomial Representation with Hierarchical Tensor Coefficients

To represent encoders of invariant foliations we need to approximate high-dimensional functions. Dense polynomials are not suitable to represent high-dimensional functions, because the number of required parameters increase combinatorially. One approach is to use multi-layer neural networks (Elbrachter et al. 2021; Hornik 1991); however, their training can be problematic (Orr and Müller 2003). Standard training methods take a long time and they rarely reach the global minimum (and there is no free lunch Wolpert and Macready 1997). In addition, the approximation can overfit the data, that is the accuracy on unseen data can be significantly worse than on the training data. In terms of the geometry of neural networks, they do not form a differentiable manifold and therefore parameters may tend to infinity without minimising the loss function (Petersen et al. 2021).

To represent encoders we still use polynomials, except that the polynomial coefficients are represented by compressed tensors in the hierarchical Tucker (HT) format. This tensor format was introduced in Hackbusch and Kühn (2009) and demonstrated to solve many problems (Grasedyck et al. 2013) that would normally wrestle with the 'curse of dimensionality' (Bellman 2015). In chemistry and physics a somewhat different but related format, the matrix product state (Perez-Garcia et al. 2007) is frequently used and in other problems a particular type of HT representation the tensor train format (Oseledets 2011) is used. However, the most important property of HT tensors is that they form a smooth quotient manifold and therefore suitable to use in optimisation problems (Uschmajew and Vandereycken 2013).

To define the format, we partly follow the notation of Uschmajew and Vandereycken (2013) in our definitions. First, we need to define the dimension tree.

**Definition 21** Given an order  $d$ , a dimension tree  $T$  is a non-trivial rooted binary tree whose nodes can be labelled by subsets of  $\{1, 2, \dots, d\}$  such that

1. the root has the label  $t_r = \{1, 2, \dots, d\}$ , and

- every node  $t \in T$ , which is not a leaf has two descendants,  $t_1$  and  $t_2$  that form an ordered partition of  $t$ , that is,

$$t_1 \cup t_2 = t \quad \text{and} \quad \mu < \nu \text{ for all } \mu \in t_1, \nu \in t_2.$$

The set of leaf nodes is denoted by  $L = \{\{1\}, \{2\}, \dots, \{d\}\}$ .

**Definition 22** Let  $T$  be a dimension tree and  $\mathbf{k} = (k_t)_{t \in T}$  a set of positive integers. The hierarchical Tucker format for  $\mathbb{R}^{n_0 \times n_1 \times \dots \times n_d}$  is defined as follows

- For a leaf node  $U_t$  is an orthonormal matrix, such that  $U_t \in \mathbb{R}^{k_t \times n_t}$  with elements  $U_t(i; j)$ , where  $i = 1 \dots k_t, j = 1 \dots n_t$
- For not a leaf node with  $t = t_1 \cup t_2$  we define

$$U_t(p; i_1, \dots, i_{|t_1|}, j_1, \dots, j_{|t_2|}) = \sum_{q=1}^{k_{t_1}} \sum_{r=1}^{k_{t_2}} B_t(p, q, r) U_{t_1}(q; i_1, \dots, i_{|t_1|}) U_{t_2}(r; j_1, \dots, j_{|t_2|}), \quad (49)$$

where  $|t|$  stands for the number of indices in label  $t$ .

The definition is recursive, hence data storage for  $U_t$  is only required for the leaf nodes, and for non-leaf nodes only storing  $B_t$  is sufficient. In other words, the set of matrices

$$\{U_t : t \in L\} \cup \{B_t : t \in T \setminus L\}$$

fully specifies a HT tensor. For each node, apart from the root node  $t_r$ , one can apply a transformation, which does not change the resulting tensor. That is, for  $t \neq t_r$  define  $\tilde{U}_t = A_t U_t$  and  $\tilde{B}_t = B_t \times_{2,3} A_{t_1}^{-1} \otimes A_{t_2}^{-1}$ , where  $A_t \in \mathbb{R}^{k_t \times k_t}$  are invertible matrices. Without restricting generality, we can stipulate that the leaf nodes  $U_t$  and matricisation of  $B_t$  with respect to the first index are also orthogonal matrices, which means that transformations  $A_t$  must be unitary to preserve orthogonality. This identity transformation defines an equivalence relation between parametrisations of a HT tensor and therefore the HT format is a quotient manifold as described in (Boumal, 2022, chapter 7). Using a quotient manifold would prevent us from optimising for the matrices  $U_t, B_t$  individually, as we need to treat the whole HT tensor as a single entity. In addition, we could not find any example in the literature, where a HT tensor was treated as a quotient manifold. Instead, we simply assume a simpler structure, that is,  $B_{t_r}$  is on a Euclidean manifold and the rest of the coefficients  $B_t$  and  $U_t$  are orthogonal matrices and therefore elements of Stiefel manifolds, which is the running example in Boumal (2022). This way we have the HT format as an element of a product manifold, where each coefficient matrix remains independent of each other. It is also possible to calculate singular values of HT tensors. Vanishing singular values indicate that the required quantity is well-approximated and that the HT tensor can be simplified to include less terms. This further adds to our ability to identify parsimonious models. In what follows, we use balanced dimension trees and up to rank six matrices  $U_t, B_t$  for simplicity.

Using notation (49), we can write the encoders as

$$\mathbf{U}(\mathbf{x}) = \mathbf{U}^1 \mathbf{x} + \sum_{d=2}^p \sum_{i_1 \cdots i_d=1}^n U_{i_r}^d(:, i_1, \dots, i_d) x_{i_1} \cdots x_{i_d}, \quad (50)$$

where  $\mathbf{U}^1$  is an orthogonal matrix, similarly element of a Stiefel manifold. We also apply the constraint that  $\mathbf{U}(\mathbf{W}_1 \mathbf{z})$  must be linear. This simply means that for the nonlinear terms of  $\mathbf{U}$  the data must not include components in the  $\mathbf{W}_1$  direction. This can be achieved by projecting our data, that is  $\hat{\mathbf{x}} = \mathbf{U}_1^\perp \mathbf{x}$ , where  $\mathbf{U}_1^\perp$  is obtained in step F2 of the invariant foliation identification process of Sect. 4 and consists of row vectors that are orthogonal to the column vectors of  $\mathbf{W}_1$ . Therefore our constrained encoder assumes the form of

$$\mathbf{U}(\mathbf{x}) = \mathbf{U}^1 \mathbf{x} + \sum_{d=2}^p \sum_{i_1 \cdots i_d=1}^n U_{i_r}^d(:, i_1, \dots, i_d) \hat{x}_{i_1} \cdots \hat{x}_{i_d}.$$

#### A.4 Optimisation Techniques

Given our compressed representation of the encoder in the form (50), we can now discuss how to solve the optimisation problem (7).

HT tensors, which make up the parametrisation of the encoder  $\mathbf{U}$ , depend linearly on each matrix  $B_t^d, U_t^d$  individually. Therefore, it is beneficial to carry out the optimisation for each matrix component individually, and cycle through all matrix components a number of times until convergence is reached. This is called batch coordinate descent (Ortega et al. 1970). The complicating factor in this approach is that the encoder also appears as an inner function of map  $\mathcal{S}$  and therefore the dependence of the objective function becomes nonlinear on matrices  $B_t, U_t$ . For optimisation in each coordinate we use the second order trust-region method (Conn et al. 2000) as we can explicitly calculate the Hessian of the objective function with respect to each matrix  $B_t^d, U_t^d$  and the parameters of  $\mathcal{S}$ . We only take a limited number of steps with the trust region method, as we cycle through all parameter matrices. The algorithm cycles through each tensor and the parameters of  $\mathcal{S}$  in a given order, but for each tensor of order  $d$ , only the coefficient matrix for which the gradient of the objective function is the largest is optimised for. This is a variant of the Gauss–Southwell algorithm (Nutini et al. 2015). We found that this technique, as it eliminates unnecessary optimisation steps for parameters that do not influence the objective function much, converges relatively fast. Attempts to use off-the-shelf optimisers were fruitless, due to computational costs.

To identify locally defined foliations we use the Riemannian trust region method and for identifying autoencoders, we use the Riemannian BFGS quasi-Newton technique from software package (Bergmann 2022).

### A.5 Projections and Retractions on Matrix Manifolds

As we perform optimisation on matrix manifolds, we also require an orthogonal projection from the ambient space to the tangent space to the manifold and a retraction (Absil et al. 2009; Boumal 2022). The advantage of optimising on a manifold instead of using standard constrained optimisation is that retractions and projections are generally easier to evaluate than solving the constraints using generic methods.

Let us assume that our matrix manifold  $\mathcal{M}$  is embedded into the Euclidean space  $\mathcal{E}$ . As computers deal with lists of numbers, we only have a cost function defined on  $\mathcal{E}$ , which we denote by  $\bar{F} : \mathcal{E} \rightarrow \mathbb{R}$ , instead on the manifold  $F : \mathcal{M} \rightarrow \mathbb{R}$ . Carrying out optimisation on  $\mathcal{M}$  therefore requires various corrections when we use cost function  $\bar{F}$ . The gradient  $\nabla F$  on  $\mathcal{M}$  is a map from  $\mathcal{M}$  to the tangent bundle  $T\mathcal{M}$ . Generally  $\bar{F}$  does not map into  $T\mathcal{M}$ , therefore a projection is necessary. In particular, for a specific point  $p \in \mathcal{M}$ , the projection is the linear map  $P_p : T_p\mathcal{E} \rightarrow T_p\mathcal{M}$  and the gradient is calculated as

$$\nabla F(p) = P_p \nabla \bar{F}(p).$$

A simple gradient descent method would move in the direction of the gradient, however that is not necessarily on manifold  $\mathcal{M}$ , so we need to bring the result back to  $\mathcal{M}$  using a retraction. A retraction is a map  $R_p(X) : T\mathcal{M} \rightarrow \mathcal{M}$  for which  $DR_p(\mathbf{0})Y = Y$ . A retraction is supposed to approximate the so-called exponential map on the tangent space, which produces the geodesics along the manifold in the direction of the tangent vector  $X$  up to the length of  $X$ . A second-order approximation of the exponential retraction is the solution of the minimisation problem

$$R_p(X) = \arg \min_{q \in \mathcal{M}} \|p + X - q\|^2. \tag{51}$$

Projection like retractions are explained in detail in Absil and Malick (2012). Using a retraction we can pull back our result onto the manifold. The Hessian of  $F$  can be defined in terms of a second order retraction  $R$ . If we assume that  $\tilde{F}(p, X) = \bar{F}(R_p(P_p X))$ , then the Hessian is

$$D^2 F(p) = D_2^2 \tilde{F}(p, \mathbf{0}).$$

The expression of the Hessian can be simplified in many ways, which is described, e.g. in Boumal (2022).

In what follows we detail two matrix manifolds that do not appear in the literature and are specific to our problems. We also use the Stiefel manifold of orthogonal matrices that is covered in many publications (Boumal 2022). In particular, we use the so-called polar retraction of the Stiefel manifold, which is equivalent to Eq. (51). For our specific matrix manifolds we solve (51).

### A.5.1 Autoencoder Manifold

The two matrices  $\mathbf{p}_1 = \underline{\mathbf{U}} \in \mathbb{R}^{n \times m}$ ,  $\mathbf{p}_2 = \underline{\mathbf{W}} \in \mathbb{R}^{n \times l}$  that satisfy the constraints (47) represent an orthogonal autoencoder form the matrix manifold  $\mathcal{M} = GOAE_{\mathcal{M},n,m,l}$ , which is defined by the zero-level set of submersion

$$\mathbf{h}(\mathbf{p}_1, \mathbf{p}_2) = \begin{pmatrix} \mathbf{p}_1^T \mathbf{p}_1 - \mathbf{I} \\ \mathbf{p}_1^T \mathbf{p}_2 - \mathbf{M} \end{pmatrix}.$$

For  $\mathcal{M}$  to be a manifold the derivative  $D\mathbf{h}$  must have full rank. Indeed, we calculate that

$$D\mathbf{h}(\mathbf{p}_1, \mathbf{p}_2)(X_1, X_2) = \begin{pmatrix} X_1^T \mathbf{p}_1 + \mathbf{p}_1^T X_1 \\ X_1^T \mathbf{p}_2 + \mathbf{p}_1^T X_2 \end{pmatrix},$$

and substitute  $X_1 = \mathbf{p}_1 \mathbf{A}$ ,  $X_2 = \mathbf{p}_1 \mathbf{B}$ , which yields

$$D\mathbf{h}(\mathbf{p}_1, \mathbf{p}_2)(X_1, X_2) = \begin{pmatrix} \mathbf{A}^T + \mathbf{A} \\ \mathbf{B} \end{pmatrix}.$$

Since  $\mathbf{A}$  and  $\mathbf{B}$  are arbitrary matrices, the range of  $D\mathbf{h}$  is the full set of symmetric and general matrices, that is,  $D\mathbf{h}$  has full rank; hence  $\mathcal{M}$  is an embedded manifold.

**Tangent space.** The tangent space  $T_{(\mathbf{p}_1, \mathbf{p}_2)}\mathcal{M}$  is defined as the null space of  $D\mathbf{h}(\mathbf{p})$ , that is

$$T_{(\mathbf{p}_1, \mathbf{p}_2)}\mathcal{M} = \{X \in \mathbb{R}^{n \times m} : D\mathbf{h}(\mathbf{p}_1, \mathbf{p}_2)(X_1, X_2) = (\mathbf{0}, \mathbf{0})\}.$$

We define  $\mathbf{p}^\perp \in \mathbb{R}^{n \times (n-m)}$  by  $\mathbf{p}^{\perp T} \mathbf{p}^\perp = \mathbf{I}$  and  $\mathbf{p}^T \mathbf{p}^\perp = \mathbf{0}$ , hence the direct sum of ranges of  $\mathbf{p}$  and  $\mathbf{p}^\perp$  span  $\mathbb{R}^n$ . To characterise the tangent space we decompose  $X_1 = \mathbf{p}_1 \mathbf{A}_1 + \mathbf{p}_1^\perp \mathbf{B}_1$ ,  $X_2 = \mathbf{p}_1 \mathbf{A}_2 + \mathbf{p}_1^\perp \mathbf{B}_2$ , and find that

$$D\mathbf{h}(\mathbf{p}_1, \mathbf{p}_2)(X_1, X_2) = \begin{pmatrix} \mathbf{A}_1^T + \mathbf{A}_1 \\ \mathbf{B}_1^T \mathbf{p}_1^{\perp T} \mathbf{p}_2 + \mathbf{A}_2 \end{pmatrix}. \tag{52}$$

For (53) to equal zero,  $\mathbf{A}_1$  must be antisymmetric,  $\mathbf{B}_1, \mathbf{B}_2$  can be arbitrary and  $\mathbf{A}_2 = -\mathbf{B}_1^T \mathbf{p}_1^{\perp T} \mathbf{p}_2$ .

**Normal space.** We use  $X_1 = \mathbf{p}_1 \mathbf{A}_1 + \mathbf{p}_1^\perp \mathbf{B}_1$ ,  $X_2 = -\mathbf{p}_1 \mathbf{B}_1^T \mathbf{p}_1^{\perp T} \mathbf{p}_2 + \mathbf{p}_1^\perp \mathbf{B}_2$  to represent an element of the tangent space and  $\mathbf{Y}_1 = \mathbf{p}_1 \mathbf{K}_1 + \mathbf{p}_1^\perp \mathbf{L}_1$ ,  $\mathbf{Y}_2 = \mathbf{p}_1 \mathbf{K}_2 + \mathbf{p}_1^\perp \mathbf{L}_2$  to represent an arbitrary element of the ambient space. For  $(\mathbf{Y}_1, \mathbf{Y}_2)$  to be a normal vector in  $N_{(\mathbf{p}_1, \mathbf{p}_2)}\mathcal{M}$  equation

$$\langle X_1, \mathbf{Y}_1 \rangle + \langle X_2, \mathbf{Y}_2 \rangle = 0 \tag{53}$$

must hold for all  $\mathbf{B}_1, \mathbf{B}_2$  and  $\mathbf{A}_1$  antisymmetric matrices. Therefore we expand (53) into

$$\langle \mathbf{A}_1, \mathbf{K}_1 \rangle + \langle \mathbf{B}_1, \mathbf{L}_1 \rangle - \langle \mathbf{B}_1^T \mathbf{p}_1^{\perp T} \mathbf{p}_2, \mathbf{K}_2 \rangle + \langle \mathbf{B}_2, \mathbf{L}_2 \rangle = 0. \tag{54}$$

To explore what parameters are allowed, we consider (54) term-by-term. If we set  $\mathbf{B}_1 = \mathbf{0}, \mathbf{B}_2 = \mathbf{0}$ , then what remains is  $\langle \mathbf{A}_1, \mathbf{K}_1 \rangle = 0$  for all  $\mathbf{A}_1$  anti-symmetric. This constraint holds if and only if  $\mathbf{K}_1$  is symmetric. Now we set  $\mathbf{A}_1 = \mathbf{0}, \mathbf{B}_1 = \mathbf{0}$ , which leads to  $\langle \mathbf{B}_2, \mathbf{L}_2 \rangle = 0$  for all  $\mathbf{B}_2$ , hence we must have  $\mathbf{L}_2 = \mathbf{0}$ . Finally we set  $\mathbf{A}_1 = \mathbf{0}, \mathbf{B}_2 = \mathbf{0}$ , which gives us

$$\begin{aligned} \langle \mathbf{B}_1, \mathbf{L}_1 \rangle - \langle \mathbf{B}_1^T \mathbf{p}_1^{\perp T} \mathbf{p}_2, \mathbf{K}_2 \rangle &= 0, \\ \langle \mathbf{B}_1, \mathbf{L}_1 \rangle - \langle \mathbf{p}_2, \mathbf{p}_1^{\perp} \mathbf{B}_1 \mathbf{K}_2 \rangle &= 0 \end{aligned}$$

or in index notation

$$B_{1ij}L_{1ij} - p_{2ji}p_{1jk}^{\perp}B_{1kl}K_{2li} = 0. \tag{55}$$

Now we differentiate (55) with respect to  $\mathbf{B}_1$ , which gives

$$L_{1rs} - p_{2ji}p_{1jr}^{\perp}K_{2si} = 0 \implies \mathbf{L}_1 = \mathbf{p}_1^{\perp T} \mathbf{p}_2 \mathbf{K}_2^T.$$

In conclusion the normal space is given by matrices of the form

$$\begin{aligned} \mathbf{Y}_1 &= \mathbf{p}_1 \mathbf{K}_1 + \mathbf{p}_1^{\perp} \mathbf{p}_1^{\perp T} \mathbf{p}_2 \mathbf{K}_2^T, \\ \mathbf{Y}_2 &= \mathbf{p}_1 \mathbf{K}_2, \end{aligned} \tag{56}$$

where  $\mathbf{K}_1$  symmetric and  $\mathbf{K}_2$  is a general matrix.

**Projection to tangent space.** A projection is an operation that removes a vector from the normal space from any input such that the result becomes a tangent vector. Therefore we need to solve equation

$$Dh(\mathbf{p}_1, \mathbf{p}_2) \begin{pmatrix} X_1 - Y_1 \\ X_2 - Y_2 \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix},$$

where  $(Y_1, Y_2)$  is in the normal space  $N_{(\mathbf{p}_1, \mathbf{p}_2)}\mathcal{M}$ . Using the representation (56) of the normal space, we find that the equation to solve is

$$Dh(\mathbf{p}_1, \mathbf{p}_2) \begin{pmatrix} X_1 - \mathbf{p}_1 \mathbf{K}_1 - \mathbf{p}_1^{\perp} \mathbf{p}_1^{\perp T} \mathbf{p}_2 \mathbf{K}_2^T \\ X_2 - \mathbf{p}_1 \mathbf{K}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}.$$

It remains to evaluate  $Dh$ , which yields

$$\begin{aligned}
 Dh(p_1, p_2) & \begin{pmatrix} X_1 - p_1 K_1 - p_1^\perp p_1^{\perp T} p_2 K_2^T \\ X_2 - p_1 K_2 \end{pmatrix} \\
 & = \begin{pmatrix} X_1^T p_1 + p_1^T X_1 - K_1^T - K_1 \\ X_1^T p_2 + p_1^T X_2 - K_2 (p_2^T p_1^\perp p_1^{\perp T} p_2 + I) \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}. \tag{57}
 \end{aligned}$$

The solution of Eq. (57) is

$$\begin{aligned}
 K_1 & = \frac{1}{2} (X_1^T p_1 + p_1^T X_1) \\
 K_2 & = (X_1 p_2 + p_1^T X_2) (I + p_2^T p_1^\perp p_1^{\perp T} p_2)^{-1}
 \end{aligned}$$

and therefore the required projection is written as

$$\begin{aligned}
 P_{(p_1, p_2)} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} & = \begin{pmatrix} X_1 - \frac{1}{2} p_1 (X_1^T p_1 + p_1^T X_1) - p_1^\perp p_1^{\perp T} p_2 (I + p_2^T p_1^\perp p_1^{\perp T} p_2)^{-1} (p_2^T X_1^T + X_2^T p_1) \\ X_2 - p_1 (X_1 p_2 + p_1^T X_2) (I + p_2^T p_1^\perp p_1^{\perp T} p_2)^{-1} \end{pmatrix}.
 \end{aligned}$$

**Projective retraction.** The retraction we calculate is the orthogonal projection onto the manifold (51). This is a second order retraction according to Absil and Malick (2012). In our case, the projection is defined as

$$P_{\mathcal{M}}(q_1, q_2) = \arg \min_{(p_1, p_2) \in \mathcal{M}} \frac{1}{2} \|q_1 - p_1\|^2 + \frac{1}{2} \|q_2 - p_2\|^2,$$

where we can assume that  $m < l$  and  $m < n$ . Using constrained optimisation we define the auxiliary objective function

$$g(p_1, p_1, \alpha, \beta) = \frac{1}{2} \|q_1 - p_1\|^2 + \langle \alpha, p_1^T p_1 - I \rangle + \frac{1}{2} \|q_2 - p_2\|^2 + \langle \beta, p_1^T p_2 - M \rangle,$$

where  $\alpha \in \mathbb{R}^{m \times m}$ ,  $\beta \in \mathbb{R}^{m \times l}$  are Lagrange multipliers. We can also write the augmented cost function in index notation, that is

$$\begin{aligned}
 g(p_1, p_1, \alpha, \beta) & = \frac{1}{2} q_{1ij} q_{1ij} - q_{1ij} p_{1ij} + \frac{1}{2} p_{1ij} p_{1ij} + \alpha_{ji} p_{1kj} p_{1ki} - \alpha_{ii} \\
 & \quad + \frac{1}{2} q_{2ij} q_{2ij} - q_{2ij} p_{2ij} + \frac{1}{2} p_{2ij} p_{2ij} + \beta_{ji} p_{1kj} p_{2ki} - \beta_{ji} M_{ji}.
 \end{aligned}$$



To find the stationary point of  $g$ , we take the derivatives

$$\begin{aligned} D_1g(\mathbf{p}_1, \mathbf{p}_2, \boldsymbol{\alpha}, \boldsymbol{\beta}) &= -q_{1rs} + p_{1rs} + \alpha_{js}p_{1rj} + \alpha_{si}p_{1ri} + \beta_{si}p_{2ri} \\ &= -\mathbf{q}_1 + \mathbf{p}_1 + \mathbf{p}_1\boldsymbol{\alpha} + \mathbf{p}_1\boldsymbol{\alpha}^T + \mathbf{p}_2\boldsymbol{\beta}^T, \\ D_2g(\mathbf{p}_1, \mathbf{p}_2, \boldsymbol{\alpha}, \boldsymbol{\beta}) &= -q_{2rs} + p_{2rs} + \beta_{js}p_{1rj} = -\mathbf{q}_2 + \mathbf{p}_2 + \mathbf{p}_1\boldsymbol{\beta}, \end{aligned}$$

which must vanish. Let us define  $\boldsymbol{\sigma} = \mathbf{I} + \boldsymbol{\alpha} + \boldsymbol{\alpha}^T$ ; hence the equations to solve become

$$\begin{aligned} \mathbf{p}_1\boldsymbol{\sigma} + \mathbf{p}_2\boldsymbol{\beta}^T &= \mathbf{q}_1, \\ \mathbf{p}_2 + \mathbf{p}_1\boldsymbol{\beta} &= \mathbf{q}_2, \\ \mathbf{p}_1^T\mathbf{p}_1 &= \mathbf{I}, \\ \mathbf{p}_1^T\mathbf{p}_2 &= \mathbf{M}. \end{aligned}$$

We can eliminate unknown variables  $\boldsymbol{\sigma}$ ,  $\boldsymbol{\beta}$  and  $\mathbf{p}_2$  by solving the equations, that is

$$\begin{aligned} \boldsymbol{\beta} &= \mathbf{p}_1^T\mathbf{q}_2 - \mathbf{M}, \\ \boldsymbol{\sigma} &= \mathbf{p}_1^T\mathbf{q}_1 - \mathbf{M}\mathbf{q}_2^T\mathbf{p}_1 + \mathbf{M}\mathbf{M}^T, \\ \mathbf{p}_2 &= \mathbf{q}_2 - \mathbf{p}_1(\mathbf{p}_1^T\mathbf{q}_2 - \mathbf{M}). \end{aligned}$$

The equation for the remaining  $\mathbf{p}_1$  then becomes

$$(\mathbf{p}_1\mathbf{p}_1^T - \mathbf{I})(\mathbf{q}_1 - \mathbf{q}_2\mathbf{q}_2^T\mathbf{p}_1 + \mathbf{q}_2\mathbf{M}^T) = \mathbf{0}. \tag{58}$$

Equation (58) means that  $\mathbf{q}_1 + \mathbf{q}_2\mathbf{M}^T - \mathbf{q}_2\mathbf{q}_2^T\mathbf{p}_1 \in \text{range } \mathbf{p}_1$ , therefore there exists matrix  $\mathbf{a}$  such that

$$\begin{aligned} \mathbf{p}_1\mathbf{a} &= \mathbf{q}_1 + \mathbf{q}_2\mathbf{M}^T - \mathbf{q}_2\mathbf{q}_2^T\mathbf{p}_1 \\ \mathbf{p}_1^T\mathbf{p}_1 &= \mathbf{I}. \end{aligned} \tag{59}$$

The solution of Eq. (59) is not unique, because for any unitary transformation  $\mathbf{T}$ ,  $\mathbf{p}_1 \mapsto \mathbf{p}_1\mathbf{T}$  and  $\mathbf{a} \mapsto \mathbf{T}^T\mathbf{a}$  are also a solution. We use Newton’s method in combination with the Moore–Penrose inverse of the Jacobian to find a solution of (59). Finally, the retraction is set to the solution of (59)

$$\mathbf{R}_{(\tilde{\mathbf{p}}_1, \tilde{\mathbf{p}}_2)} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = (\mathbf{p}_1, \mathbf{p}_2),$$

where  $\mathbf{q}_1 = \tilde{\mathbf{p}}_1 + X_1$ ,  $\mathbf{q}_2 = \tilde{\mathbf{p}}_2 + X_2$ .

## B State-Space Reconstruction

When the full state of the system cannot be measured, but it is still observable from the output, we need to employ state-space reconstruction. We focus on the case of a real valued scalar signal  $z_j \in \mathbb{R}$ ,  $j = 1 \dots, M$ , sampled with frequency  $f_s$ , where  $j$  stands for time. Takens' embedding theorem (Takens 1981) states that if we know the box counting dimension of our attractor, which is denoted by  $n$ , the full state can almost always be observed if we create a new vector  $\mathbf{x}_p = (z_p, z_{p+\tau_1}, \dots, z_{p+\tau_{d-1}}) \in \mathbb{R}^d$ , where  $d > 2n$  and  $\tau_1, \dots, \tau_{d-1}$  are integer delays. The required number of delays is a conservative estimate, in many cases  $n \leq d < 2n + 1$  is sufficient. The general problem, how to select the number of delays  $d$  and the delays  $\tau_1, \dots, \tau_{d-1}$  optimally (Kramer 1991; Pecora et al. 2007), is a much researched subject.

Instead of selecting delays  $\tau_1, \dots, \tau_{d-1}$ , we use linear combinations of all possible delay embeddings, which allows us to consider the optimality of the embedding in a linear parameter space. We create vectors of length  $L$ ,

$$\mathbf{u}_p = (z_{ps+1}, \dots, z_{ps+L}) \in \mathbb{R}^L, \quad p = 0, 1, \dots, \left\lfloor \frac{M-L}{s} \right\rfloor,$$

where  $s$  is the number of samples by which the window is shifted forward in time with increasing  $p$ . Our reconstructed state variable  $\mathbf{x}_p \in \mathbb{R}^d \approx X$  is then calculated by a linear map  $\mathbf{T} : \mathbb{R}^L \rightarrow \mathbb{R}^d$ , in the form  $\mathbf{x}_p = \mathbf{T}\mathbf{u}_p$  such that our scalar signal is returned by  $z_{ps+\ell} \approx \mathbf{v} \cdot \mathbf{x}_p$  for some  $\ell \in \{1, \dots, L\}$ . We assume that the signal has  $m$  dominant frequencies  $f_k$ ,  $k = 1, \dots, m$ , with  $f_1$  being the lowest frequency. We define  $w = \left\lfloor \frac{f_s}{f_1} \right\rfloor$  as an approximate period of the signal, where  $f_s$  is the sampling frequency and set  $L = 2w + 1$ . This makes each  $\mathbf{u}_p$  capture roughly two periods of oscillations of the lowest frequency  $f_1$ .

In what follows we use two approaches to construct the linear map  $\mathbf{T}$ .

### B.1 Principal Component Analysis

In Broomhead and King (1986) it is suggested to use Principal Component Analysis to find an optimal delay embedding. Let us construct the matrix

$$\mathbf{V} = \left( \mathbf{u}_1 \ \mathbf{u}_2 \ \dots \ \mathbf{u}_{\left\lfloor \frac{M-L}{s} \right\rfloor} \right),$$

where  $\mathbf{u}_p$  are the columns of  $\mathbf{V}$ . Then calculate the singular value decomposition (or equivalently Jordan decomposition) of the symmetric matrix in the form

$$\mathbf{V}\mathbf{V}^T = \mathbf{H}\mathbf{\Sigma}\mathbf{H}^T,$$

where  $\mathbf{H}$  is a unitary matrix and  $\mathbf{\Sigma}$  is diagonal. Now let us take the columns of  $\mathbf{H}$  corresponding to the  $d$  largest elements of  $\mathbf{\Sigma}$ , and these columns then become the  $d$  rows of  $\mathbf{T}$ . In this case, we can only achieve an approximate reconstruction of the

signal by defining

$$\mathbf{v} = (T_{1\ell} \ T_{2\ell} \ \cdots \ T_{d\ell}),$$

where  $T_{k\ell}$  is the element of matrix  $\mathbf{T}$  in the  $k$ -th row and  $\ell$ -th column. The dot product  $z_{ps+\ell} \approx \mathbf{v} \cdot \mathbf{x}_p$  approximately reproduces our signal. The quality of the approximation depends on how small the discarded singular values in  $\Sigma$  are.

### B.2 Perfect Reproducing Filter Bank

We can also use discrete Fourier transform (DFT) to create state-space vectors and reconstruct the original signal exactly. Such transformations are called perfect reproducing filter banks (Strang and Nguyen 1996). The DFT of vector  $\mathbf{u}_p$  is calculated by  $\hat{\mathbf{u}}_p = \mathbf{K}\mathbf{u}_p$ , where matrix  $\mathbf{K}$  is defined by its elements

$$\begin{aligned} K_{1,q} &= \frac{\sqrt{2}}{2}, \\ K_{2p,q} &= \cos\left(2\pi p \frac{q-1}{2w+1}\right), \\ K_{2p+1,q} &= \sin\left(2\pi p \frac{q-1}{2w+1}\right), \end{aligned}$$

where  $p = 1, \dots, w$  and  $q = 1, \dots, 2w + 1$ . Note that  $\mathbf{K}^T \mathbf{K} = \mathbf{I} \frac{2w+1}{2}$ , hence the inverse transform is  $\mathbf{K}^{-1} = \frac{2}{2w+1} \mathbf{K}^T$ . Let us denote the  $\ell$ -th column of  $\mathbf{K}$  by  $\mathbf{c}_\ell$ , which creates a delay filter

$$\frac{2}{2w+1} \mathbf{c}_\ell^T \mathbf{K}\mathbf{u}_p = z_{ps+\ell},$$

that delays the signal by  $\ell$  samples. Separating  $\mathbf{c}_\ell$  into components, such that

$$\mathbf{c}_\ell = \mathbf{v}_1 + \mathbf{v}_2 + \cdots + \mathbf{v}_d, \tag{60}$$

we can create a perfect reproducing filter bank, that is the vector

$$\hat{\mathbf{x}}_p = \frac{2}{2w+1} \begin{pmatrix} \mathbf{v}_1^T \mathbf{K}\mathbf{u}_p \\ \mathbf{v}_2^T \mathbf{K}\mathbf{u}_p \\ \vdots \\ \mathbf{v}_{2m}^T \mathbf{K}\mathbf{u}_p \end{pmatrix}$$

when summed over its components, the input  $z_{ps+\ell}$  is recovered. The question is how to create an optimal decomposition (60). As we assumed that our signal has  $m$  frequencies, we can further assume that we are dealing with  $m$  coupled nonlinear oscillators, and therefore we can set  $d = 2m$ . We therefore divide the resolved frequencies of the

DFT, which are  $\tilde{f}_k = k \frac{f_s}{2w+1}, k = 0, \dots, w$  into  $m$  bins, which are centred around the frequencies of the signal  $f_1, \dots, f_m$ . We set the boundaries of these bins as

$$b_0 = 0, b_1 = \frac{f_1 + f_2}{2}, b_2 = \frac{f_2 + f_3}{2}, \dots, b_{m-1} = \frac{f_{m-1} + f_m}{2}, b_m = \frac{wf_s}{2w + 1}$$

and for each bin labelled by  $k$ , we create an index set

$$\mathcal{I}_k = \left\{ l \in 1, \dots, w : b_{k-1} < \tilde{f}_l \leq b_k \right\},$$

which ensures that all DFT frequencies are taken into account without overlap, that is  $\cup_k \mathcal{I}_k = \{0, 1, \dots, w\}$  and  $\mathcal{I}_p \cap \mathcal{I}_q = \emptyset$  for  $p \neq q$ . This creates a decomposition (60) in the form of

$$v_{2k-1,2j} = \begin{cases} c_{\ell,2j} & \text{if } j \in \mathcal{I}_k \\ 0 & \text{if } j \notin \mathcal{I}_k \end{cases}, v_{2k,2j+1} = \begin{cases} c_{\ell,2j+1} & \text{if } j \in \mathcal{I}_k \\ 0 & \text{if } j \notin \mathcal{I}_k \end{cases}$$

with the exception of  $v_1$ , for which we also set  $v_{1,1} = c_{\ell,1}$  to take into account the moving average of the signal and to make sure that  $\sum v_k = c_\ell$ . Here  $v_{k,l}$  stands for the  $l$ -th element of vector  $v_k$ . Finally, we define the transformation matrix

$$H = \frac{2}{2w + 1} \begin{pmatrix} \|v_1\|^{-1} v_1^T \\ \vdots \\ \|v_{2m}\|^{-1} v_{2m}^T \end{pmatrix},$$

where each row is normalised such that our separated signals have the same amplitude. The newly created signal is then

$$x_p = T u_p,$$

where  $T = HK$  and the original signal can be reproduced as

$$z_{ps+\ell} = (\|v_1\| \cdots \|v_{2m}\|) x_p.$$

### C Proof of Proposition 16

**Proof of proposition 16** First we explore what happens if we introduce a new parametrisation of the decoder  $W$  that replaces  $r$  by  $\rho(r)$  and  $\theta$  by  $\theta + \gamma(\rho(r))$ , where  $\rho : \mathbb{R}^+ \rightarrow \mathbb{R}^+$  is an invertible function with  $\rho(0) = 0$  and  $\gamma : \mathbb{R}^+ \rightarrow \mathbb{R}$  with  $\gamma(0) = 0$ . This transformation creates a new decoder  $\tilde{W}$  of  $\mathcal{M}$  in the form of

$$\tilde{W}(r, \theta) = W(\rho(r), \theta + \gamma(\rho(r))), \tag{61}$$

$$\tilde{W}(\rho^{-1}(r), \theta - \gamma(r)) = W(r, \theta). \tag{62}$$

Equation (61) is not the only re-parametrisation, but this is the only one that preserves the structure of the polar invariance Eq. (21). After substituting the transformed decoder, invariance Eq. (21) becomes

$$\tilde{W}(\tilde{R}(r), \theta + \tilde{T}(r)) = F(\tilde{W}(r, \theta)),$$

where

$$\begin{aligned} \tilde{R}(r) &= \rho^{-1}(R(\rho(r))), \\ \tilde{T}(r) &= T(\rho(r)) + \gamma(\rho(r)) - \gamma(R(\rho(r))). \end{aligned}$$

In the new coordinates, the instantaneous frequency  $\omega(r)$  and damping ratio  $\zeta(r)$  become

$$\begin{aligned} \omega(r) &= T(\rho(r)) + \gamma(\rho(r)) - \gamma(R(\rho(r))) \quad [\text{rad/step}], \quad (63) \\ \zeta(r) &= -\frac{\log[r^{-1}\rho^{-1}(R(\rho(r)))]}{\omega(r)} \quad [-]. \quad (64) \end{aligned}$$

Before we go further, let us introduce the notation

$$\langle \mathbf{x}, \mathbf{y} \rangle = \frac{1}{2\pi} \int_0^{2\pi} \langle \mathbf{x}(\theta), \mathbf{x}(\theta) \rangle_X d\theta,$$

where  $\mathbf{x}, \mathbf{y} : [0, 2\pi] \rightarrow X$  and  $\langle \cdot, \cdot \rangle_X$  is the inner product on vector space  $X$ . The norm of function  $\mathbf{x}$  is then defined by  $\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$ .

We now turn to the phase constraint given by Eq. (26). We choose a fixed  $\epsilon > 0$ , substitute  $r_1 = \rho(r + \epsilon), r_2 = \rho(r)$  and the phase shift  $\gamma = \gamma(\rho(r + \epsilon)) - \gamma(\rho(r))$  into Eq. (26) and find the objective function

$$J(\gamma(\rho(r + \epsilon))) = \epsilon^{-1} \|\mathbf{W}(\rho(r + \epsilon), \cdot + \gamma(\rho(r + \epsilon))) - \mathbf{W}(\rho(r), \cdot + \gamma(\rho(r)))\|^2,$$

where we also divided by  $\epsilon$ . We then need to find the value of  $\gamma(\rho(r + \epsilon))$  that minimises  $J$  for a fixed  $\epsilon$ . A necessary condition for a local minimum of  $J$  is that the derivative  $J' = 0$ , that is,

$$\begin{aligned} \epsilon^{-1} \langle \mathbf{W}(\rho(r + \epsilon), \cdot + \gamma(\rho(r + \epsilon))) - \mathbf{W}(\rho(r), \cdot + \gamma(\rho(r))), \\ D_2 \mathbf{W}(\rho(r + \epsilon), \cdot + \gamma(\rho(r + \epsilon))) \rangle = 0. \end{aligned}$$

To find a continuous parametrisation, let us now take the limit  $\epsilon \rightarrow 0$  to get

$$\begin{aligned} \dot{\rho}(r) \langle D_1 \mathbf{W}(\rho(r), \cdot + \gamma(\rho(r))) + D_2 \mathbf{W}(\rho(r), \cdot + \gamma(\rho(r))) \dot{\gamma}(\rho(r)), \\ D_2 \mathbf{W}(\rho(r), \cdot + \gamma(\rho(r))) \rangle = 0. \end{aligned}$$

We can also remove the constant phase shift and use  $t = \rho(r)$  to find the *phase condition*

$$\langle D_1 \mathbf{W}(t, \cdot) + D_2 \mathbf{W}(t, \cdot) \dot{\gamma}(t), D_2 \mathbf{W}(t, \cdot) \rangle = 0. \quad (65)$$

Integrating Eq. (65) leads to the phase shift

$$\gamma(t) = - \int_0^t \frac{\langle D_1 \mathbf{W}(r, \cdot), D_2 \mathbf{W}(r, \cdot) \rangle}{\langle D_2 \mathbf{W}(r, \cdot), D_2 \mathbf{W}(r, \cdot) \rangle} dr \quad (66)$$

Note that phase conditions are commonly used in numerical continuation of periodic orbits (Beyn and Thümmler 2007), for slightly different reasons.

The next quantity to fix is the amplitude constraint given by Eq. (25). In the new parametrisation, Eq. (25) can be written as

$$\|\mathbf{W}(\rho(r), \cdot)\|^2 = r^2. \quad (67)$$

If we define  $\kappa(r) = \|\mathbf{W}(r, \cdot)\|$ , then we have

$$\rho(r) = \kappa^{-1}(r). \quad (68)$$

Finally, substituting (66), (68) into Eqs. (63), (64) completes the proof.  $\square$

**Proof of remark 18** In case of a vector field  $f : X \rightarrow X$ , and invariance equation

$$D_1 \mathbf{W}(r, \theta) R(r) + D_2 \mathbf{W}(r, \theta) T(r) = f(\mathbf{W}(r, \theta)),$$

the instantaneous frequency and damping ratios with respect to the coordinate system defined by the decoder  $\mathbf{W}$ , are

$$\left. \begin{aligned} \omega(r) &= T(r), \\ \zeta(r) &= -\frac{R(r)}{r\omega(r)}, \end{aligned} \right\} \quad (69)$$

respectively. However the coordinate system defined by  $\mathbf{W}$  is nonlinear, which needs correcting. As in the proof of proposition 16, we assume a coordinate transformation

$$\tilde{\mathbf{W}}(r, \theta) = \mathbf{W}(\rho(r), \theta + \gamma(\rho(r))), \quad (70)$$

and find that invariance Eq. (32) becomes

$$D_1 \tilde{\mathbf{W}}(r, \theta) \tilde{R}(r) + D_2 \tilde{\mathbf{W}}(r, \theta) \tilde{T}(r) = f(\tilde{\mathbf{W}}(r, \theta)). \quad (71)$$

When substituting (70) into (71) we find

$$\begin{aligned} D_1 \tilde{\mathbf{W}}(r, \theta) \dot{\kappa}(\rho(r)) R(\rho(r)) + D_2 \tilde{\mathbf{W}}(r, \theta) (T(\rho(r)) - \dot{\gamma}(\rho(r)) R(\rho(r))) &= \\ = f(\tilde{\mathbf{W}}(r, \theta)). \end{aligned} \quad (72)$$

Comparing (72) and (32) we extract that

$$\left. \begin{aligned} \tilde{R}(r) &= \dot{\kappa}(r) R(\rho(r)) \\ \tilde{T}(r) &= T(\rho(r)) - \dot{\gamma}(r) R(\rho(r)) \end{aligned} \right\}. \quad (73)$$

Recognising that  $t = \rho(r) = \kappa^{-1}(r)$  and replacing  $T$  with  $\tilde{T}$ ,  $R$  with  $\tilde{R}$  in formulae (69) proves remark 18.  $\square$

## References

- Absil, P.A., Mahony, R., Sepulchre, R.: Optimization Algorithms on Matrix Manifolds. Princeton University Press (2009)
- Absil, P.A., Malick, J.: Projection-like retractions on matrix manifolds. *SIAM J. Optim.* **22**(1), 135–158 (2012)
- Barton, D.A.W.: Control-based continuation: bifurcation and stability analysis for physical experiments. *Mech. Syst. Signal Process.* **84**, 54–64 (2017)
- Bellman, R.E.: Adaptive Control Processes. Princeton University Press (2015)
- Bergmann, R.: Manopt.jl: optimization on manifolds in Julia. *J. Open Source Softw.* **7**(70), 3866 (2022)
- Beyn, W.-J., Thümmler, V.: Phase Conditions, Symmetries and PDE Continuation, pp. 301–330. Springer, Dordrecht (2007)
- Billings, S.A.: Nonlinear System Identification: “NARMAX” Methods in the Time, Frequency, and Spatio-Temporal Domains. Wiley (2013)
- Boumal, N.L.: An introduction to optimization on smooth manifolds. To appear with Cambridge University Press (2022)
- Boyd, S., Vandenberghe, L.: Introduction to Applied Linear Algebra: Vectors, Matrices, and Least Squares. Cambridge University Press (2018)
- Breunung, T., Haller, G.: Explicit backbone curves from spectral submanifolds of forced-damped nonlinear mechanical systems. *Proc. R. Soc. Lond. A Math. Phys. Eng. Sci.* **474**(2213), 20180083 (2018)
- Broomhead, D.S., King, G.P.: Extracting qualitative dynamics from experimental data. *Physica D* **20**(2), 217–236 (1986)
- Brunton, S.L., Tu, J.H., Bright, I., Kutz, J.N.: Compressive sensing and low-rank libraries for classification of bifurcation regimes in nonlinear dynamical systems. *SIAM J. Appl. Dyn. Syst.* **13**(4), 1716–1732 (2014)
- Brunton, S.L., Proctor, J.L., Kutz, J.N., Bialek, W.: Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proc. Natl. Acad. Sci. U.S.A.* **113**(15), 3932–3937 (2016)
- Cabr e, X., Fontich, E., de la Llave, R.: The parameterization method for invariant manifolds I: manifolds associated to non-resonant subspaces. *Indiana Univ. Math. J.* **52**, 283–328 (2003)
- Cabr e, X., Fontich, E., de la Llave, R.: The parameterization method for invariant manifolds iii: overview and applications. *J. Differ. Equ.* **218**(2), 444–515 (2005)
- Casdagli, M.: Nonlinear prediction of chaotic time series. *Physica D* **35**(3), 335–356 (1989)
- Cenedese, M., Ax as, J., B auerlein, B., Avila, K., Haller, G.: Data-driven modeling and prediction of nonlinearizable dynamics via spectral submanifolds. *Nat. Commun.* **13**, 872 (2022)
- Champion, K., Lusch, B., Nathan Kutz, J., Brunton, S.L.: Data-driven discovery of coordinates and governing equations. *Proc. Natl. Acad. Sci. U.S.A.* **116**(45), 22445–22451 (2019)
- Coifman, R.R., Lafon, S.: Diffusion maps. *Appl. Comput. Harmon. Anal.* **21**(1), 5–30 (2006)
- Conn, A.R., Gould, N.I.M., Toint, P.L.: Trust Region Methods. MPS-SIAM Series on Optimization. SIAM (2000)
- de la Llave, R.: Invariant manifolds associated to nonresonant spectral subspaces. *J. Stat. Phys.* **87**(1), 211–249 (1997)
- Donoho, D.L.: Compressed sensing. *IEEE Trans. Inf. Theory* **52**(4), 1289–1306 (2006)
- Ehrhardt, D.A., Allen, M.S.: Measurement of nonlinear normal modes using multi-harmonic stepped force appropriation and free decay. *Mech. Syst. Signal Process.* **76–77**, 612–633 (2016)

- Elbrachter, D., Perekrestenko, D., Grohs, P., Bolcskei, H.: Deep neural network approximation theory. *IEEE Trans. Inf. Theory* **67**(5), 2581–2623 (2021)
- Fenichel, N.: Persistence and smoothness of invariant manifolds for flows. *Indiana Univ. Math. J.* **21**, 193–226 (1972)
- Grasedyck, L.: Hierarchical singular value decomposition of tensors. *SIAM J. Matrix Anal. Appl.* **31**(4), 2029–2054 (2010)
- Grasedyck, L., Kressner, D., Tobler, C.: A literature survey of low-rank tensor approximation techniques. *GAMM-Mitteilungen* **36**(1), 53–78 (2013)
- Hackbusch, W., Kühn, S.: A new scheme for the tensor representation. *J. Fourier Anal. Appl.* **15**, 706–722 (2009)
- Haller, G., Ponsioen, S.: Nonlinear normal modes and spectral submanifolds: existence, uniqueness and use in model reduction. *Nonlinear Dyn.* **86**(3), 1493–1534 (2016)
- Hermann, R., Krener, A.: Nonlinear controllability and observability. *IEEE Trans. Autom. Control* **22**(5), 728–740 (1977)
- Hornik, K.: Approximation capabilities of multilayer feedforward networks. *Neural Netw.* **4**(2), 251–257 (1991)
- Jin, M., Chen, W., Brake, M.R.W., Song, H.: Identification of instantaneous frequency and damping from transient decay data. *J. Vib. Acoust. Trans. ASME* **142**(5), 051111 (2020)
- Kalia, M., Brunton, S.L., Meijer, H.G.E., Brune, C., Kutz, J.N.: Learning normal form autoencoders for data-driven discovery of universal, parameter-dependent governing equations (2021)
- Kevrekidis, I.G., Gear, C.W., Hyman, J.M., Kevrekidis, P.G., Runborg, O., Theodoropoulos, C.: Equation-free, coarse-grained multiscale computation: enabling microscopic simulators to perform system-level analysis. *Commun. Math. Sci.* **1**(4), 715–762 (2003)
- Kevrekidis, I.G., Samaey, G.: Equation-free multiscale computation: algorithms and applications. *Annu. Rev. Phys. Chem.* **60**(1), 321–344 (2009)
- Kramer, M.A.: Nonlinear principal component analysis using autoassociative neural networks. *AIChE J.* **37**(2), 233–243 (1991)
- Lang, S.: *Fundamentals of Differential Geometry*. Graduate Texts in Mathematics. Springer, New York (2012)
- Lawson, H.B., Jr.: Foliations. *Bull. Am. Math. Soc.* **80**, 369–418 (1974)
- Mezić, I.: Spectral properties of dynamical systems, model reduction and decompositions. *Nonlinear Dyn.* **41**(1–3), 309–325 (2005)
- Mezić, I.: Koopman operator, geometry, and learning of dynamical systems. *Not. Am. Math. Soc.* **68**(7), 1087–1105 (2021)
- Nutini, J., Schmidt, M., Laradji, I.H., Friedlander, M., Koepke, H.: Coordinate descent converges faster with the gauss-southwell rule than random selection. In: *Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37, ICML'15*, pp. 1632–1641 (2015)
- Orr, G.B., Müller, K.R.: *Neural Networks: Tricks of the Trade*. Lecture Notes in Computer Science. Springer, Berlin (2003)
- Ortega, J.M., Ortega, J.R.W., Rheinboldt, W.C.: *Iterative Solution of Nonlinear Equations in Several Variables*. Classics in Applied Mathematics. SIAM (1970)
- Oseledets, I.V.: Tensor-train decomposition. *SIAM J. Sci. Comput.* **33**(5), 2295–2317 (2011)
- Pecora, L.M., Moniz, L., Nichols, J., Carroll, T.L.: A unified approach to attractor reconstruction. *Chaos Interdiscip. J. Nonlinear Sci.* **17**(1), 013110 (2007)
- Perez-Garcia, D., Verstraete, F., Wolf, M.M., Cirac, J.I.: Matrix product state representations. *Quantum Info. Comput.* **7**(5), 401–430 (2007)
- Petersen, P., Raslan, M., Voigtlaender, F.: Topological properties of the set of functions generated by neural networks of fixed size. *Found. Comput. Math.* **21**(2), 375–444 (2021)
- Ponsioen, S., Pedergrana, T., Haller, G.: Automated computation of autonomous spectral submanifolds for nonlinear modal analysis. *J. Sound Vib.* **420**, 269–295 (2018)
- Read, N.K., Ray, W.H.: Application of nonlinear dynamic analysis to the identification and control of nonlinear systems - iii. n-dimensional systems. *J. Process Control* **8**(1), 35–46 (1998)
- Roberts, A.J.: Appropriate initial conditions for asymptotic descriptions of the long-term evolution of dynamical systems. *J. Aust. Math. Soc. Ser. B* **31**, 48–75 (1989)
- Roberts, A.J.: Boundary conditions for approximate differential equations. *J. Aust. Math. Soc. Ser. B* **34**, 54–80 (1992)



- Shaw, S.W., Pierre, C.: Normal-modes of vibration for nonlinear continuous systems. *J. Sound Vib.* **169**(3), 319–347 (1994)
- Sieber, J., Krauskopf, B.: Control based bifurcation analysis for experiments. *Nonlinear Dyn.* **51**, 365–377 (2008)
- Strang, G., Nguyen, T.: *Wavelets and Filter Banks*. Wellesley-Cambridge Press (1996)
- Szalai, R.: Invariant spectral foliations with applications to model order reduction and synthesis. *Nonlinear Dyn.* **101**(4), 2645–2669 (2020)
- Szalai, R., Ehrhardt, D., Haller, G.: Nonlinear model identification and spectral submanifolds for multi-degree-of-freedom mechanical vibrations. *Proc. R. Soc. A* **473**, 20160759 (2017)
- Takens, F.: Detecting strange attractors in turbulence. In: Rand, D., Young, L.-S. (eds.) *Dynamical Systems and Turbulence*, Warwick 1980, pp. 366–381. Springer, Berlin (1981)
- Tenenbaum, J.B., De Silva, V., Langford, J.C.: A global geometric framework for nonlinear dimensionality reduction. *Science* **290**(5500), 2319–2323 (2000)
- Titurus, B., Yuan, J., Scarpa, F., Patsias, S., Pattison, S.: Impact hammer-based analysis of nonlinear effects in bolted lap joint. In: *Proceedings of ISMA2016 including USD2016, ISMA2016*, pp. 789–802 (2016)
- Uschmajew, A., Vandereycken, B.: The geometry of algorithms using hierarchical tensors. *Linear Algebra Appl.* **439**(1), 133–166 (2013)
- Vizzaccaro, A., Shen, Y., Salles, L., Blahoš, J., Touzé, C.: Direct computation of nonlinear mapping via normal form for reduced-order models of finite element nonlinear structures. *Comput. Methods Appl. Mech. Eng.* **384**, 113957 (2021)
- Whitney, H.: Differentiable manifolds. *Ann. Math.* **37**(3), 645–680 (1936)
- Wolpert, D.H., Macready, W.G.: No free lunch theorems for optimization. *IEEE Trans. Evol. Comput.* **1**(1), 67–82 (1997)
- Yair, O., Talmon, R., Coifman, R.R., Kevrekidis, I.G.: Reconstruction of normal forms by learning informed observation geometries from data. *Proc. Natl. Acad. Sci. U.S.A.* **114**(38), E7865–E7874 (2017)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.