



# Tensor-based computation of metastable and coherent sets

Feliks Nüske<sup>a,b,\*</sup>, Patrick Gelß<sup>c</sup>, Stefan Klus<sup>d</sup>, Cecilia Clementi<sup>a,e</sup>

<sup>a</sup> Center for Theoretical Biological Physics, and Department of Chemistry, Rice University, Houston, TX, 77005, United States

<sup>b</sup> Institute of Mathematics, Paderborn University, Paderborn 33100, Germany

<sup>c</sup> Department of Mathematics and Computer Science, Freie Universität Berlin, Berlin 14195, Germany

<sup>d</sup> Department of Mathematics, University of Surrey, Guildford, GU2 7XH, UK

<sup>e</sup> Department of Physics, Freie Universität Berlin, Berlin 14195, Germany

## ARTICLE INFO

### Article history:

Received 27 March 2020

Received in revised form 10 August 2021

Accepted 17 August 2021

Available online 6 September 2021

Communicated by V.M. Perez-Garcia

### Keywords:

Koopman operator

Extended dynamic mode decomposition

Tensor networks

Tensor-train format

Dynamical systems

Molecular dynamics

## ABSTRACT

Recent years have seen rapid advances in the data-driven analysis of dynamical systems based on Koopman operator theory and related approaches. On the other hand, low-rank tensor product approximations – in particular the tensor train (TT) format – have become a valuable tool for the solution of large-scale problems in a number of fields. In this work, we combine Koopman-based models and the TT format, enabling their application to high-dimensional problems in conjunction with a rich set of basis functions or features. We derive efficient algorithms to obtain a reduced matrix representation of the system's evolution operator starting from an appropriate low-rank representation of the data. These algorithms can be applied to both stationary and non-stationary systems. We establish the infinite-data limit of these matrix representations, and demonstrate our methods' capabilities using several benchmark data sets.

© 2021 Elsevier B.V. All rights reserved.

## 1. Introduction

The data-driven analysis of high-dimensional dynamical systems has been a highly successful research field for several years, with applications in fluid dynamics, control theory, molecular dynamics, and many others. Much of the work along these lines has focused on the infinite-dimensional description of a system using *transfer operators* or *Koopman operators*, see [1–5]. For non-stationary or time-dependent systems, it is often advantageous to consider the *forward-backward operator*. Its spectral analysis leads to the detection of *coherent sets* [6,7]. We will summarily refer to all of these operators as *evolution operators* in this paper. A host of different methods for the numerical approximation of evolution operators from simulation or measurement data have been developed. These include *extended dynamic mode decomposition* (EDMD) [5,8], the *variational approach to conformational dynamics* (VAC) [9,10] and its generalization *variational approach to Markov processes* (VAMP) [11], as well as *canonical correlation analysis* (CCA) [12,13]. All of these methods are closely related, and revolve, in one way or another, around data-based approximations to the Koopman operator on (usually finite-dimensional) subspaces. For a detailed review and comparison, see [14].

Much of the appeal of these techniques is due to their formulation as data-driven regression problems, which paves the way for the application of modern machine learning techniques. Examples include kernel-based formulations [13,15–17] and combinations with deep learning [18,19]. A different avenue towards the solution of high-dimensional problems are tensor products, where functions and operators on high-dimensional spaces are approximated in linear spaces of products of simple (often univariate) functions. The expansion coefficients of such a function form a multi-dimensional array, called a *tensor*. As the size of a tensor grows exponentially with the dimension, *low-rank formats* requiring only a manageable number of parameters need to be used. Important examples include the *canonical format* [20], the *Tucker format* [21], and the *hierarchical Tucker format* [22], with the *tensor train (TT) format* [23,24] as an important special case of the latter. The common idea behind these formats is to decompose a high-dimensional tensor into a network of lower-dimensional tensors. Several applications of tensor decompositions have shown that it is possible to tackle large-scale problems which cannot be solved using conventional numerical methods, see, e.g., [25–29], and especially [30–32] for quantum chemistry applications.

\* Correspondence to: Paderborn University, Technologiepark 21, 33100 Paderborn, Germany.  
E-mail address: [feliks.nueske@uni-paderborn.de](mailto:feliks.nueske@uni-paderborn.de) (F. Nüske).

**Table 1**  
Notation used in this work.

Symbol	Description
$\mathbb{H}, L(\mathbb{H})$	General Hilbert space, space of bounded linear operators
$L^2_{\rho_0}, L^2_{\rho_1}, L^2_{\rho_\mu}$	Weighted $L^2$ -spaces
$\mathbb{V}, \mathbb{W}$	Finite-dimensional Hilbert spaces or subspaces
$\psi, \phi; \eta, \zeta$	(Orthonormal) bases of $\mathbb{V}$ and $\mathbb{W}$ , respectively
$\mathcal{P}_\mathbb{V}$	Orthogonal projection onto $\mathbb{V}$
$\mathcal{T}_\tau, \mathcal{K}_\tau, \mathcal{F}_\tau$	Perron–Frobenius, Koopman, and forward–backward operator
$\mathcal{T}_\tau(\mathbb{V}, \mathbb{W}), \mathcal{K}_\tau(\mathbb{V}, \mathbb{W}), \mathcal{F}_\tau(\mathbb{V}, \mathbb{W})$	Galerkin projections onto finite-dimensional subspaces
$T_\tau(\psi, \phi), K_\tau(\psi, \phi), F_\tau(\psi, \phi)$	Matrix representations of Galerkin projections,
$\Psi(\cdot)$	Transformed data matrix
$C(\cdot), A(\cdot, \cdot)$	Covariance and cross-covariance matrices
$\mathbb{G}_r, \mathbb{B}_r$	Spectral subspace and associated coefficient vector space
$\mathbf{T} = \left[ \mathbf{T}^{(1)} \right] \otimes \dots \otimes \left[ \mathbf{T}^{(p)} \right]$	Tensor train of order $p$
$\mathbf{T} _k = \mathbf{T}_{n_1, \dots, n_k}^{n_{k+1}, \dots, n_p}$	Mode- $k$ unfolding of $\mathbf{T}$
$\Psi(\cdot)$	Transformed data tensor
$\mathbf{C}(\cdot), \mathbf{A}(\cdot, \cdot)$	Covariance and cross-covariance tensors
$\hat{\cdot}$	Data-driven estimates

Previous work on low-rank approximation in the context of Koopman operator modeling includes [29,33–36]. Importantly, it was shown in [33] that the matrices required for data-driven approximations to the Perron–Frobenius or Koopman operator based on a product basis can directly be written in the canonical format. The resulting generalized eigenvalue problems can be converted to TT format and then solved with the aid of power iteration methods. This, however, requires repeated rank reductions and appropriate estimates of the eigenvalues so that power iteration converges. In [29,35], a TT representation of the *data tensor*, which contains the evaluations of a product basis at all data points, was introduced, accompanied by a method to compute an approximate singular value decomposition (global SVD) of the data tensor.

In this study, we build on these results to derive and analyze new tensor decomposition methods for Koopman operator approximation. Our methods proceed by first computing a (compressed) representation of the data tensor. Subsequently, a tensor network needs to be contracted to arrive at a reduced matrix approximation for the Koopman operator. Importantly, the calculation or inversion of Gramian matrices is entirely avoided. The detailed contributions of our study are as follows:

- First, we present a multi-linear version of the *AMUSE* algorithm [37] for Koopman approximation, which we call *AMUSEt*. This method uses the global SVD to compute an approximate multi-linear singular value decomposition of the data tensor. By contraction of an appropriate tensor network, a matrix representation of the Koopman operator can then be obtained.
- Second, we analyze *AMUSEt* by showing that the final matrix approximation is indeed the data-driven representation of the Koopman operator on a data-dependent, finite-dimensional subspace. We establish convergence of this representation in the limit of infinite data.
- Third, we present an alternative tensor train decomposition of the data tensor corresponding to a product basis. It is based on a method outlined in [38], which provides a *higher-order CUR decomposition*. A detailed description of the latter method’s algorithmic realization, including several enhancements, is provided.
- Finally, we demonstrate the capabilities of the proposed methods by analyzing benchmark molecular dynamics and fluid dynamics data sets.

The remainder of this work is structured as follows: In Section 2, we introduce the required notation and concepts regarding evolution operators, their numerical approximation, tensor decompositions, and the application of tensor methods in the context of evolution operators. The *AMUSEt* algorithm is introduced in Section 3, while the novel *HOCUR*-based algorithm is presented in Section 4. The theoretical analysis of *AMUSEt* follows in Section 5, numerical results for benchmark problems are then shown in Section 6. Concluding remarks and open problems follow in Section 7.

## 2. Basic concepts

We first recapitulate basic concepts from dynamical systems theory, especially evolution operators and their Galerkin approximation in Section 2.1. Afterwards, we change topics and discuss low-rank approximations of tensors in the tensor train format and how these concepts can be applied in the context of dynamical systems in Section 2.2. Table 1 summarizes the notation that is used throughout the paper.

### 2.1. Evolution operators and their approximation

#### 2.1.1. Evolution operators

The main concern of this study is the analysis of dynamical systems using evolution operators. Let  $\mathcal{X}_t \in \mathbb{R}^d$  be a deterministic or stochastic dynamical system. For a positive lag time  $\tau$ , assume that the densities of  $\mathcal{X}_t$  at times  $t = 0$  and  $t = \tau$  are given by  $\rho_0$  and  $\rho_1$ ,

respectively. The Perron–Frobenius operator  $\mathcal{T}_\tau: L^2_{\rho_0} \rightarrow L^2_{\rho_1}$ , and its adjoint, the Koopman operator  $\mathcal{K}_\tau = \mathcal{T}_\tau^*: L^2_{\rho_1} \rightarrow L^2_{\rho_0}$ , are defined by

$$\mathcal{T}_\tau f(y) = \frac{1}{\rho_1(y)} \int f(x) \rho_0(x) p^\tau(x, y) dx, \quad \mathcal{K}_\tau f(x) = \int p^\tau(x, y) f(y) dy,$$

where  $p^\tau$  is the stochastic transition kernel associated with the process  $\mathcal{X}_t$ , see [1–5] for more details. If the process is stationary and admits an invariant distribution  $\mu$ , it is convenient to choose  $\rho_0 = \rho_1 = \mu$ . In particular, if the dynamics  $\mathcal{X}_t$  are reversible with respect to  $\mu$ , the operator  $\mathcal{T}_\tau$  is self-adjoint on  $L^2_\mu$ , and hence identical to  $\mathcal{K}_\tau$ . For non-reversible or time-dependent systems, it is in general more appropriate to consider the forward–backward operator

$$\mathcal{F}_\tau = \mathcal{T}_\tau^* \mathcal{T}_\tau : L^2_{\rho_0} \rightarrow L^2_{\rho_0}.$$

One is often interested in certain spectral components of these operators, especially those which are largest in magnitude (also called *leading/dominant eigenvalues*), as they can be used to determine dynamically long-lived structures. For stationary systems, leading eigenpairs of  $\mathcal{K}_\tau$  or  $\mathcal{T}_\tau$  (which will be close to one in absolute value), can be used for metastability analysis, especially if the system is reversible [39]. If the system is non-stationary, leading singular values and functions of the Koopman operator are typically used instead. If they exist, left singular functions are automatically eigenfunctions of  $\mathcal{F}_\tau$ , with eigenvalue equal to the square of the singular value. These functions can be used to determine finite-time coherent sets [6,7,13,40].

### 2.1.2. Galerkin projection and dimensionality reduction

For the purposes of numerical analysis, the linear operators above must be represented on finite-dimensional subspaces. To introduce the notation, let  $\mathbb{H}$  be a Hilbert space of functions on  $\mathbb{R}^d$  with inner product  $\langle \cdot, \cdot \rangle_{\mathbb{H}}$ . Let  $\mathbb{V}$  be a finite-dimensional subspace of dimension  $n$ , with a basis  $\psi = (\psi_1, \dots, \psi_n)^\top$ . The orthogonal projector onto  $\mathbb{V}$  is denoted by  $\mathcal{P}_{\mathbb{V}}$ . Any function in  $\mathbb{V}$  can be represented uniquely by a vector of expansion coefficients with respect to  $\psi$  in  $\mathbb{R}^n$ . More generally, any  $r$ -dimensional subspace  $\mathbb{F}$  of  $\mathbb{V}$ , spanned by functions  $\theta = (\theta_1, \dots, \theta_r)^\top$ , is encoded by a matrix  $A \in \mathbb{R}^{n \times r}$ , the columns of which contain the expansion coefficients of  $\theta$  with respect to the basis  $\psi$ :

$$\theta(x)^\top = \psi(x)^\top A.$$

In this context, we will call the column space of  $A$  in  $\mathbb{R}^n$  the *coefficient vector space* corresponding to  $\mathbb{F}$ . Finally, the Gramian matrix of a finite basis is denoted by

$$C(\psi) := (\langle \psi_i, \psi_j \rangle_{\mathbb{H}})_{ij}.$$

Returning to the analysis of evolution operators, let finite-dimensional spaces  $\mathbb{V} \subset L^2_{\rho_0}$  and  $\mathbb{W} \subset L^2_{\rho_1}$ , with bases  $\psi = (\psi_1, \dots, \psi_n)^\top$  and  $\phi = (\phi_1, \dots, \phi_r)^\top$ , be given. We then consider the Galerkin projections

$$\begin{aligned} \mathcal{K}_\tau(\mathbb{V}, \mathbb{W}) &= \mathcal{P}_{\mathbb{V}} \mathcal{K}_\tau \mathcal{P}_{\mathbb{W}}, \\ \mathcal{T}_\tau(\mathbb{V}, \mathbb{W}) &= \mathcal{P}_{\mathbb{W}} \mathcal{T}_\tau \mathcal{P}_{\mathbb{V}}, \\ \mathcal{F}_\tau(\mathbb{V}, \mathbb{W}) &= \mathcal{K}_\tau(\mathbb{V}, \mathbb{W}) \mathcal{T}_\tau(\mathbb{V}, \mathbb{W}). \end{aligned} \tag{1}$$

The matrix representations of these operators with respect to the bases  $\psi$  and  $\phi$  are given by

$$\begin{aligned} K_\tau(\psi, \phi) &= (C(\psi))^{-1} A(\psi, \phi), \\ T_\tau(\psi, \phi) &= (C(\phi))^{-1} A(\psi, \phi)^\top, \\ F_\tau(\psi, \phi) &= K_\tau(\psi, \phi) T_\tau(\psi, \phi), \end{aligned} \tag{2}$$

where  $C(\psi)$  and  $C(\phi)$  are the Gramians in  $L^2_{\rho_0}$  and  $L^2_{\rho_1}$ , respectively, and the matrix  $A(\psi, \phi)$  satisfies

$$A(\psi, \phi)_{ij} = \langle \psi_i, \mathcal{K}_\tau \phi_j \rangle_{\rho_0} = \langle \mathcal{T}_\tau \psi_i, \phi_j \rangle_{\rho_1},$$

see [5,8,14]. We will frequently use orthonormal bases to represent the operators introduced above. If  $\eta$  and  $\zeta$  are orthonormal bases for  $\mathbb{V}$  and  $\mathbb{W}$ , the corresponding matrix representations reduce to

$$K_\tau(\eta, \zeta) = A(\eta, \zeta) = T_\tau(\eta, \zeta)^\top, \quad F_\tau(\eta, \zeta) = A(\eta, \zeta) A(\eta, \zeta)^\top.$$

We observe that the matrix representation  $K_\tau(\eta, \zeta) = A(\eta, \zeta)$  of the projected Koopman operator for orthonormal basis sets serves as a baseline model from which most relevant quantities can be calculated directly, such as eigenpairs in the stationary case, or singular pairs and forward–backward eigenpairs in the non-stationary case. Therefore, representations of this type will be particularly interesting in what follows. Arbitrary bases can always be transformed into orthonormal ones using spectral decompositions of the Gramians: If  $C(\psi) = U_\psi \Sigma_\psi^2 U_\psi^\top$ ,  $C(\phi) = U_\phi \Sigma_\phi^2 U_\phi^\top$ , then we arrive at the following orthonormal bases (*whitening transformation*):

$$\eta(x)^\top = \psi(x)^\top U_\psi \Sigma_\psi^{-1}, \quad \zeta(x)^\top = \phi(x)^\top U_\phi \Sigma_\phi^{-1}.$$

By additionally truncating the spectral decompositions after their first  $r \leq n$  components, reduced orthonormal bases  $\eta_r^\top = \psi^\top (U_{\psi,r} \Sigma_{\psi,r}^{-1})$  and  $\zeta_r^\top = \phi^\top (U_{\phi,r} \Sigma_{\phi,r}^{-1})$  can be obtained. The Koopman matrix corresponding to these bases can be calculated according to

$$M_{\tau,r} := K_\tau(\eta_r, \zeta_r) = (\Sigma_{\psi,r}^{-1} U_{\psi,r}^\top) A(\psi, \phi) (U_{\phi,r} \Sigma_{\phi,r}^{-1}).$$

The matrix  $M_{\tau,r}$  is then typically used as an approximate representation of  $\mathcal{K}_\tau(\mathbb{V}, \mathbb{W})$ . Quantities of interest can be expressed with respect to the full basis by means of the transformations  $(U_{\psi,r} \Sigma_{\psi,r}^{-1})$  and  $(U_{\phi,r} \Sigma_{\phi,r}^{-1})$ . For instance, if  $(\sigma_i, v_i, w_i)$  is a singular triplet of  $M_{\tau,r}$ , then the corresponding singular functions can be expressed in the original bases by the coefficient vectors  $\xi_i = (U_{\psi,r} \Sigma_{\psi,r}^{-1}) v_i$  and  $\chi_i = (U_{\phi,r} \Sigma_{\phi,r}^{-1}) w_i$ . However, it should be kept in mind that  $M_{\tau,r}$  really just represents the Koopman operator on the reduced subspaces spanned by  $\eta_r, \zeta_r$ , which will be analyzed in more detail in Section 5.

### 2.1.3. Data-driven approximation

Usually, the integrals required for the matrices in (2) cannot be computed analytically, and are estimated from data instead. Assume we have a pair of  $\mathbb{R}^d$ -valued random variables  $x, y$  on a probability space  $(\Omega, \mathcal{B}, \mathbb{P})$  such that the joint distribution of  $(x, y)$  on  $\mathbb{R}^d \times \mathbb{R}^d$  is  $\vartheta(x, y) = \rho_0(x)p^\tau(x, y)$ . For Hilbert space-valued,  $\vartheta$ -integrable functions  $f: \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{H}$ , the strong law of large numbers [41] then implies that for almost surely any sequence of i.i.d. samples  $(x_k, y_k) \in \mathbb{R}^d \times \mathbb{R}^d$ , we have

$$\lim_{m \rightarrow \infty} \frac{1}{m} \sum_{k=1}^m f(x_k, y_k) = \mathbb{E}^\vartheta [f]. \quad (3)$$

For such an i.i.d. sequence, we can assemble all pairs into data matrices  $X, Y \in \mathbb{R}^{d \times m}$ , where  $X = [x_1, \dots, x_m]$  and  $Y = [y_1, \dots, y_m]$ . For finite-dimensional subspaces  $\mathbb{V} \subset L^2_{\rho_0}, \mathbb{W} \subset L^2_{\rho_1}$  as above, we then define the *transformed data matrices* in  $\mathbb{R}^{n \times m}$  by

$$\Psi(X) = [\psi(x_1) \ \dots \ \psi(x_m)] \quad \text{and} \quad \Phi(Y) = [\phi(y_1) \ \dots \ \phi(y_m)]. \quad (4)$$

These give rise to the following empirical estimates of the matrices  $C(\psi), C(\phi), A(\psi, \phi)$ :

$$\widehat{C}(\psi) = \frac{1}{m} \Psi(X)\Psi(X)^\top, \quad \widehat{C}(\phi) = \frac{1}{m} \Phi(Y)\Phi(Y)^\top, \quad \widehat{A}(\psi, \phi) = \frac{1}{m} \Psi(X)\Phi(Y)^\top. \quad (5)$$

One of the central results of the Koopman approach is [5,8,14]:

**Proposition 1.** *If (3) holds, then almost surely*

$$C(\psi) = \lim_{m \rightarrow \infty} \widehat{C}(\psi), \quad C(\phi) = \lim_{m \rightarrow \infty} \widehat{C}(\phi), \quad A(\psi, \phi) = \lim_{m \rightarrow \infty} \widehat{A}(\psi, \phi). \quad (6)$$

**Remark 1.**

- (i) A standard way to generate the samples  $(x_k, y_k)$  is to draw  $x_k$  i.i.d. from  $\rho_0$ , and  $y_k$  is then obtained by integrating the dynamics  $\mathcal{X}_t$  over time  $\tau$ , starting from  $x_k$ .
- (ii) By ergodic theory [41], the conclusions of Proposition 1 also hold if  $\mathcal{X}_t$  is stationary with unique invariant density  $\mu = \rho_0 = \rho_1$ . In this case,  $x_k$  is chosen as the  $k$ th step of any discretized trajectory of  $\mathcal{X}_t$ , and  $y_k$  is obtained as the  $k$ th step of the same trajectory shifted by  $\tau$ .

By means of the empirical estimates (5), we can obtain data-based projections  $\widehat{\mathcal{K}}_\tau(\mathbb{V}, \mathbb{W}), \widehat{\mathcal{T}}_\tau(\mathbb{V}, \mathbb{W})$  and  $\widehat{\mathcal{F}}_\tau(\mathbb{V}, \mathbb{W})$  of the evolution operators. Their matrix representations with respect to  $\psi, \phi$  are the same as (2), only using empirical estimates, i.e.,

$$\begin{aligned} \widehat{K}_\tau(\psi, \phi) &= (\widehat{C}(\psi))^{-1} \widehat{A}(\psi, \phi), \\ \widehat{T}_\tau(\psi, \phi) &= (\widehat{C}(\phi))^{-1} \widehat{A}(\psi, \phi)^\top, \\ \widehat{F}_\tau(\psi, \phi) &= \widehat{K}_\tau(\psi, \phi) \widehat{T}_\tau(\psi, \phi). \end{aligned} \quad (7)$$

Just as we did for the analytical Galerkin projections in Section 2.1.2, we can use empirically orthonormal bases  $\widehat{\eta}, \widehat{\zeta}$  (i.e.,  $\widehat{C}(\widehat{\eta}) = \text{Id}$  and  $\widehat{C}(\widehat{\zeta}) = \text{Id}$ ) to simplify these matrix approximations, obtaining the empirical Koopman matrix as  $\widehat{K}_\tau(\widehat{\eta}, \widehat{\zeta}) = \widehat{A}(\widehat{\eta}, \widehat{\zeta})$ . Spectral decompositions of the empirical Gramians can be used to find orthonormal bases just as described above, and truncations of the spectral decompositions lead to appropriate reduced matrices  $\widehat{M}_{\tau,r}$ . The data-driven matrices just introduced form the basis for a number of well-known numerical methods to analyze evolution operators. In particular, *extended dynamic mode decomposition (EDMD)* [5,8] and the *variational approach to conformational dynamics (VAC)* [9,10] apply to the stationary case, while *canonical correlation analysis* [12,13] has been formulated for the non-stationary setting. See also [14] for an overview of the nomenclature.

### 2.1.4. The AMUSE algorithm

In the data-driven setting, calculation of the empirical Gramians can be entirely avoided. Let rank- $r$  singular value decompositions (SVDs) of the transformed data matrices be given by

$$\Psi(X) = \widehat{U}_{X,r} \widehat{\Sigma}_{X,r} \widehat{V}_{X,r}^\top + \widehat{E}_{X,r}, \quad \Phi(Y) = \widehat{U}_{Y,r} \widehat{\Sigma}_{Y,r} \widehat{V}_{Y,r}^\top + \widehat{E}_{Y,r},$$

where  $\widehat{E}_{X,r}, \widehat{E}_{Y,r}$  are the errors resulting from truncation of the SVD at rank  $r \leq n$ . Because of (5), the basis sets  $\widehat{\eta}_r^\top = \sqrt{m} \psi^\top (\widehat{U}_{X,r} \widehat{\Sigma}_{X,r}^{-1})$ ,  $\widehat{\zeta}_r^\top = \sqrt{m} \phi^\top (\widehat{U}_{Y,r} \widehat{\Sigma}_{Y,r}^{-1})$  are empirically orthonormal. Moreover, their empirical Koopman matrix can be obtained directly from the above SVDs, by observing that

$$\widehat{M}_{\tau,r} = \widehat{A}(\widehat{\eta}_r, \widehat{\zeta}_r) = (\sqrt{m} \widehat{\Sigma}_{X,r}^{-1} \widehat{U}_{X,r}^\top) \frac{1}{m} \Psi(X) \Phi(Y)^\top \sqrt{m} (\widehat{U}_{Y,r} \widehat{\Sigma}_{Y,r}^{-1}) = \widehat{V}_{X,r}^\top \widehat{V}_{Y,r}.$$

Quantities of interest, such as singular triplets, can be calculated directly from  $\widehat{M}_{\tau,r}$ . The method is summarized in Algorithm 1 [11,14,37]. It should be kept in mind, though, that  $\widehat{M}_{\tau,r}$  only serves as empirical approximation of the Koopman operator on the subspaces spanned by  $\widehat{\eta}_r, \widehat{\zeta}_r$ . Further below, these spaces will be analyzed in more detail.

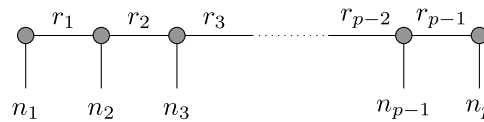
If the subspaces  $\mathbb{V}, \mathbb{W}$  are identical, it is often desirable to use a single reduced basis in Algorithm 1. The resulting modification using just an SVD of  $\Psi(X)$  is shown in Algorithm 2. If the process  $\mathcal{X}_t$  is stationary ( $\rho_0 = \rho_1 = \mu$ ), then the reduced subspaces based on SVDs of  $\Psi(X)$  and  $\Psi(Y)$  converge to the same limit with infinite data, hence Algorithms 1 and 2 are asymptotically equivalent. For finite data however, their outputs will generally be different. We also note that the computation of the reduced matrix in Algorithm 2 does not break down to a single matrix product, which will be important when comparing the tensor-based versions of both algorithms in Section 3.

**Algorithm 1:** AMUSE.

- 
- Input:** transformed data matrices  $\Psi(X)$  and  $\Phi(Y)$   
**Output:** reduced Koopman matrix  $\widehat{M}_{\tau,r} = \widehat{K}_{\tau}(\widehat{\eta}_r, \widehat{\zeta}_r)$ ,  
 approximate singular values  $\widehat{\sigma}_i$  and singular vectors  $\widehat{\xi}_i, \widehat{\chi}_i$  of  $\widehat{K}_{\tau}(\mathbb{V}, \mathbb{W})$ .
- 
- 1: Compute reduced SVDs of  $\Psi(X)$  and  $\Phi(Y)$ , i.e.,  $\Psi(X) \approx \widehat{U}_{X,r} \widehat{\Sigma}_{X,r} \widehat{V}_{X,r}^{\top}$  and  $\Phi(Y) \approx \widehat{U}_{Y,r} \widehat{\Sigma}_{Y,r} \widehat{V}_{Y,r}^{\top}$ .
  - 2: Compute  $\widehat{M}_{\tau,r} = \widehat{V}_{X,r}^{\top} \widehat{V}_{Y,r}$ .
  - 3: Compute singular values  $\widehat{\sigma}_i$  of  $\widehat{M}_{\tau,r}$ , and left and right singular vectors  $\widehat{v}_i, \widehat{w}_i$ .
  - 4: Express singular vectors with respect to original bases:  $\widehat{\xi}_i = \widehat{U}_{X,r} \widehat{\Sigma}_{X,r}^{-1} \widehat{v}_i, \widehat{\chi}_i = \widehat{U}_{Y,r} \widehat{\Sigma}_{Y,r}^{-1} \widehat{w}_i$ .
- 

**Algorithm 2:** Single Basis AMUSE.

- 
- Input:** transformed data matrices  $\Psi(X)$  and  $\Psi(Y)$   
**Output:** reduced Koopman matrix  $\widehat{M}_{\tau,r} = \widehat{K}_{\tau}(\widehat{\eta}_r, \widehat{\eta}_r)$ ,  
 approximate singular triplets  $(\widehat{\sigma}_i, \widehat{\xi}_i, \widehat{\chi}_i)$  or eigenpairs  $(\widehat{\lambda}_i, \widehat{\xi}_i)$  of  $\widehat{K}_{\tau}(\mathbb{V}, \mathbb{V})$ .
- 
- 1: Compute reduced SVD of  $\Psi(X)$ , i.e.,  $\Psi(X) \approx \widehat{U}_{X,r} \widehat{\Sigma}_{X,r} \widehat{V}_{X,r}^{\top}$ .
  - 2: Compute  $\widehat{M}_{\tau,r} = \widehat{V}_{X,r}^{\top} \Psi(Y)^{\top} \widehat{U}_{X,r} \widehat{\Sigma}_{X,r}^{-1}$ .
  - 3: Compute singular triplets  $(\widehat{\sigma}_i, \widehat{v}_i, \widehat{w}_i)$  or eigenpairs  $(\widehat{\lambda}_i, \widehat{w}_i)$  of  $\widehat{M}_{\tau,r}$ .
  - 4: Express singular vectors or eigenvectors w.r.t. original basis as in Algorithm 1.
- 



**Fig. 1.** Graphical representation of a tensor train: A core is depicted by a circle with different arms indicating the modes of the tensor and the rank indices. The first and the last TT core are regarded as matrices due to the fact that  $r_0 = r_p = 1$ .

2.2. Low-rank tensor representations

Tensors are multi-dimensional arrays  $\mathbf{T} \in \mathbb{R}^N$ , where  $N = n_1 \times \dots \times n_p$ . Here,  $p$  is called the order of a tensor, while the dimensions of the elementary vector spaces (the so-called modes) are  $n_k, k = 1, \dots, p$ . Tensor entries are sometimes represented by multi-indices  $\mathbf{i} = (i_1, \dots, i_p)$  with  $i_k \in \{1, \dots, n_k\}$ , i.e.,  $\mathbf{T}_{\mathbf{i}} = \mathbf{T}_{i_1, \dots, i_p}$ . The single-index representation of the multi-index  $\mathbf{i}$  is denoted by  $\bar{\mathbf{i}} \in \{1, \dots, \prod_{k=1}^p n_k\}$ . Conversely, the multi-index representation of the single-index  $i \in \mathbb{N}$  is represented by  $\underline{i}$ . The tensor product is denoted by  $\otimes$ . We use bold capital letters ( $\mathbf{T}, \mathbf{U}$ , etc.) to denote tensors, capital letters ( $U, V$ , etc.) for matrices, and vectors are represented by lower case letters ( $x, x_k$ , etc.). For  $1 \leq k \leq p - 1$ , the *mode- $k$  unfolding* of the tensor  $\mathbf{T}$  is the matrix

$$\mathbf{T}_{n_1, \dots, n_k}^{n_{k+1}, \dots, n_p} \in \mathbb{R}^{(n_1 \dots n_k) \times (n_{k+1} \dots n_p)}.$$

In short, we use the notation

$$\mathbf{T}|_k = \mathbf{T}_{n_1, \dots, n_k}^{n_{k+1}, \dots, n_p}$$

if the modes of  $\mathbf{T}$  are clear.

2.2.1. Tensor train format

We start by introducing the tensor train (TT) format, where a high-dimensional tensor is represented by the contraction of multiple low-dimensional tensors [23,24].

**Definition 1.** A tensor  $\mathbf{T} \in \mathbb{R}^N$  is said to be in the *TT format* if

$$\mathbf{T} = \sum_{l_0=1}^{r_0} \dots \sum_{l_p=1}^{r_p} \bigotimes_{k=1}^p \mathbf{T}_{l_{k-1}, : , l_k}^{(k)} = \sum_{l_0=1}^{r_0} \dots \sum_{l_p=1}^{r_p} \mathbf{T}_{l_0, : , l_1}^{(1)} \otimes \dots \otimes \mathbf{T}_{l_{p-1}, : , l_p}^{(p)}.$$

The tensors  $\mathbf{T}^{(k)} \in \mathbb{R}^{r_{k-1} \times n_k \times r_k}$  of order 3 are called *TT cores* and the numbers  $r_k$  are called *TT ranks*. It holds that  $r_0 = r_p = 1$  and  $r_k \geq 1$  for  $k = 1, \dots, p - 1$ .

The TT ranks  $r_0, \dots, r_p$  have a strong influence on the capability of representing a given tensor as a tensor train and determine the storage consumption of a tensor in the TT format. Fig. 1 shows the graphical representation of a tensor train, which is also called Penrose notation, see [42].

We also represent TT cores as two-dimensional arrays containing vectors as elements. For a given tensor train  $\mathbf{T} \in \mathbb{R}^N$  with cores  $\mathbf{T}^{(k)} \in \mathbb{R}^{r_{k-1} \times n_k \times r_k}$ , a single core is written as

$$\llbracket \mathbf{T}^{(k)} \rrbracket = \begin{bmatrix} \mathbf{T}_{1, \cdot, 1}^{(k)} & \cdots & \mathbf{T}_{1, \cdot, r_k}^{(k)} \\ \vdots & \ddots & \vdots \\ \mathbf{T}_{r_{k-1}, \cdot, 1}^{(k)} & \cdots & \mathbf{T}_{r_{k-1}, \cdot, r_k}^{(k)} \end{bmatrix}.$$

We then use the notation  $\mathbf{T} = \llbracket \mathbf{T}^{(1)} \rrbracket \otimes \cdots \otimes \llbracket \mathbf{T}^{(p)} \rrbracket$  for representing tensor trains  $\mathbf{T}$ , cf. [27,29,43]. This notation can be regarded as a generalization of the standard matrix multiplication. The difference is that we here compute the tensor products of the corresponding elements – which are vectors instead of scalar values – and then sum over the columns and rows, respectively. A core of a tensor train is *left-orthonormal* if

$$(\mathbf{T}^{(k)}|_{r_{k-1}, n_k}^{r_k})^\top \cdot (\mathbf{T}^{(k)}|_{r_{k-1}, n_k}^{r_k}) = \text{Id} \in \mathbb{R}^{r_k \times r_k}.$$

### 2.2.2. Basis decompositions

Within the context of data-driven approximation of evolution operators, see Section 2.1.3, tensors arise if trial spaces are chosen as tensor products of elementary function spaces. We consider a data matrix  $X \in \mathbb{R}^{d \times m}$  originating from a stochastic process  $\mathcal{X}_t$ , and a set of basis functions  $\psi_1, \dots, \psi_p$  with  $\psi_k: \mathbb{R}^d \rightarrow \mathbb{R}^{n_k}$  where  $n_k \in \mathbb{N}$  for  $k = 1, \dots, p$ . Let  $\mathbb{V}^k = \text{span}\{\psi_{k,1}, \dots, \psi_{k,n_k}\}$  denote the  $n_k$ -dimensional subspaces spanned by the elementary basis functions  $\psi_{k,1}, \dots, \psi_{k,n_k} \in L^2_{\rho_0}$ . We consider the Galerkin projection (1) on the tensor product  $\mathbb{V} := \mathbb{V}^1 \otimes \cdots \otimes \mathbb{V}^p \subset L^2_{\rho_0}$ , which is a subspace of dimension at most  $n_1 \cdot \dots \cdot n_p$ . Equivalently, choosing elementary basis functions in  $L^2_{\rho_1}$  for a data matrix  $Y$  yields a tensor space  $\mathbb{W} \subset L^2_{\rho_1}$ .

The tensor-based counterparts of the transformed data matrices given in (4) are denoted by  $\Psi(X)$  and  $\Phi(Y)$ , respectively. These *transformed data tensors* can then be used to obtain empirical estimates of the Galerkin tensors for  $\mathbb{V}, \mathbb{W}$ :

$$\widehat{\mathbf{C}}(\Psi) = \frac{1}{m} \Psi(X) \Psi(X)^\top, \quad \widehat{\mathbf{C}}(\Phi) = \frac{1}{m} \Phi(Y) \Phi(Y)^\top, \quad \widehat{\mathbf{A}}(\Psi, \Phi) = \frac{1}{m} \Psi(X) \Phi(Y)^\top,$$

where the transposes of  $\Psi(X)$  and  $\Phi(Y)$  result from index permutations such that  $\Psi(X)^\top, \Phi(Y)^\top \in \mathbb{R}^{m \times n_1 \times \cdots \times n_p}$ . The multiplication above then denotes the contraction of the last mode of the first tensor with the first mode of the second tensor. As we have shown in [29], the tensor train format can be used to represent transformed data tensors. In what follows, we will focus on the construction of  $\Psi(X)$ , the case for  $\Phi(Y)$  is analogous. We start by considering rank-one tensors of the form

$$\Psi(x) = \psi_1(x) \otimes \cdots \otimes \psi_p(x) = \begin{bmatrix} \psi_{1,1}(x) \\ \vdots \\ \psi_{1,n_1}(x) \end{bmatrix} \otimes \cdots \otimes \begin{bmatrix} \psi_{p,1}(x) \\ \vdots \\ \psi_{p,n_p}(x) \end{bmatrix} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_p}. \tag{8}$$

Note that the so-called *coordinate-major* and *function-major* basis decompositions, introduced in [29], are special cases of the more general decomposition given in (8). The transformed data tensor  $\Psi(X) \in \mathbb{R}^{n_1 \times \cdots \times n_p \times m}$  is then given by adding the rank-one decompositions (8) for all vectors  $x_1, \dots, x_m$  and taking the tensor product with an additional unit vector. The result is the following TT decomposition:

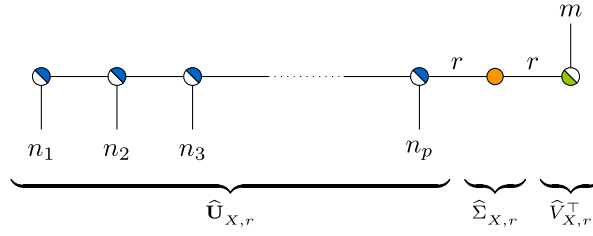
$$\begin{aligned} \Psi(X) &= \sum_{k=1}^m \Psi(x_k) \otimes e_k \\ &= \sum_{k=1}^m \psi_1(x_k) \otimes \cdots \otimes \psi_p(x_k) \otimes e_k \\ &= \llbracket \psi_1(x_1) \ \cdots \ \psi_1(x_m) \rrbracket \otimes \begin{bmatrix} \psi_2(x_1) & & 0 \\ & \ddots & \\ 0 & & \psi_2(x_m) \end{bmatrix} \otimes \cdots \otimes \begin{bmatrix} \psi_p(x_1) & & 0 \\ & \ddots & \\ 0 & & \psi_p(x_m) \end{bmatrix} \otimes \begin{bmatrix} e_1 \\ \vdots \\ e_m \end{bmatrix} \\ &=: \llbracket \Psi^{(1)}(X) \rrbracket \otimes \llbracket \Psi^{(2)}(X) \rrbracket \otimes \cdots \otimes \llbracket \Psi^{(p)}(X) \rrbracket \otimes \llbracket \Psi^{(p+1)}(X) \rrbracket, \end{aligned} \tag{9}$$

where  $e_k, k = 1, \dots, m$ , denote the unit vectors of the standard basis in the  $m$ -dimensional Euclidean space. The matrix-based counterpart of  $\Psi(X)$ , see (4), would be given by the mode- $p$  unfolding

$$\Psi(X) = \Psi(X)|_p = \Psi(X)|_{n_1, \dots, n_p}^m, \tag{10}$$

that is, modes  $n_1, \dots, n_p$  represent row indices of the unfolding, and mode  $m$  is the column index.<sup>1</sup>

<sup>1</sup> The relation (10) would still be satisfied if the position of the unit vectors in (9) is changed. Moreover, contracting the TT core of unit vectors with any other core results in the so-called *block TT format* [44]. For the application of HOCUR and HOSVD, however, we stick to the decomposition scheme given in (9).



**Fig. 2.** Global SVD in TT format: Global SVD of a transformed data tensor  $\Psi(X)$  with components  $\widehat{\mathbf{U}}_{X,r}$  (half-filled circles in blue),  $\widehat{\Sigma}_{X,r}$  (orange circle), and  $\widehat{\mathbf{V}}_{X,r}$  (half-filled circle in green). The first  $p$  modes depict the row indices while the last mode depicts the column index. We transpose the last TT core in order to account for the different row and column dimensions of the transformed data matrices, cf. Section 2.2.2. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

### 2.2.3. Global SVD

In order to apply the Algorithms 1 and 2 to tensor product bases, SVDs of the mode- $p$  unfoldings of  $\Psi(X)$  and  $\Phi(Y)$  are required. A multi-linear analogue of the standard SVD was presented for the tensor train format in [35]. For a given tensor train  $\Psi(X)$ , the method provides an orthonormal, tensor-train structured segment  $\widehat{\mathbf{U}}_{X,r} \in \mathbb{R}^{n_1 \times \dots \times n_p \times r}$ , a diagonal coupling matrix  $\widehat{\Sigma}_{X,r} \in \mathbb{R}^{r \times r}$ , and an orthonormal matrix  $\widehat{\mathbf{V}}_{X,r} \in \mathbb{R}^{m \times r}$  representing the last core. The dimension  $r$  is the TT rank between the two last cores. Just like a standard SVD of a matrix,  $\widehat{\mathbf{U}}_{X,r}$ ,  $\widehat{\Sigma}_{X,r}$ , and  $\widehat{\mathbf{V}}_{X,r}$  then satisfy the following properties:

- (i)  $\Psi(X) = \widehat{\mathbf{U}}_{X,r} \widehat{\Sigma}_{X,r} \widehat{\mathbf{V}}_{X,r}^T$ ,
- (ii)  $(\widehat{\mathbf{U}}_{X,r}|_p)^T \widehat{\mathbf{U}}_{X,r}|_p = \widehat{\mathbf{V}}_{X,r}^T \widehat{\mathbf{V}}_{X,r} = \text{Id} \in \mathbb{R}^{r \times r}$ ,
- (iii)  $\widehat{\Sigma}_{X,r} \in \mathbb{R}^{r \times r}$  is a diagonal matrix.

#### Algorithm 3: Global SVD.

**Input:** transformed data tensor  $\Psi(X) \in \mathbb{R}^{n_1 \times \dots \times n_p \times m}$  in TT format

**Output:** global SVD of  $\Psi(X)$  in the form of  $\widehat{\mathbf{U}}_{X,r} \widehat{\Sigma}_{X,r} \widehat{\mathbf{V}}_{X,r}^T$

- 1: **for**  $k = 1, \dots, p - 1$  **do**
- 2:   Compute (truncated) SVD of  $\Psi^{(k)}(X)|_2$ , i.e.,  $\Psi^{(k)}(X)|_2 = U \Sigma V^T + E$ ,  $\Sigma \in \mathbb{R}^{r_k \times r_k}$ .
- 3:   Set  $\Psi^{(k)}(X)$  to a reshaped version of  $U$ .
- 4:   Set  $\Psi^{(k+1)}(X)$  to a reshaped version of  $\Sigma V^T \Psi^{(k+1)}(X)|_1$ .
- 5: Compute (truncated) SVD of  $\Psi^{(p)}(X)|_2$ , i.e.,  $\Psi^{(p)}(X)|_2 = U \Sigma V^T + E$ ,  $\Sigma \in \mathbb{R}^{r_p \times r_p}$ .
- 6: Set  $\Psi^{(p)}(X)$  to a reshaped version of  $U$ .
- 7: Set  $\Psi^{(p+1)}(X)$  to  $V^T$ .
- 8: Define  $r = r_p$ ,  $\widehat{\mathbf{U}}_{X,r} = \llbracket \Psi^{(1)}(X) \rrbracket \otimes \dots \otimes \llbracket \Psi^{(p)}(X) \rrbracket$ ,  $\widehat{\Sigma}_{X,r} = \Sigma$ , and  $\widehat{\mathbf{V}}_{X,r} = V$ .

The method described in Algorithm 3 proceeds as follows: Similarly to the TT-SVD algorithm proposed in [24], we left-orthonormalize the TT cores  $\Psi^{(1)}(X), \dots, \Psi^{(p-1)}(X)$  using (truncated) SVDs. Then, we decompose the TT core  $\Psi^{(p)}(X)$ , but this time retain the diagonal matrix containing the singular values and only shift the right-orthonormal matrix to the last core. This provides the components of the global SVD as shown in Fig. 2. Note that we do not require any right-orthonormalization of the last core as stated in [35] since  $\Psi^{(p+1)}(X)$  is simply a reshaped identity matrix.

### 3. AMUSE on tensors

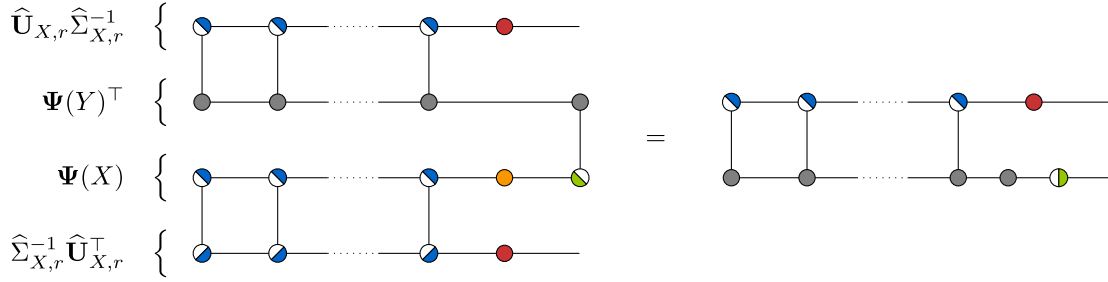
In this section, we combine the ideas of the AMUSE algorithm developed in Section 2.1.4 with the global SVD for tensor-structured bases. Recall that the idea of AMUSE was to use truncated SVDs of the data matrices in order to determine empirically orthonormal bases of reduced subspaces of the trial spaces  $\mathbb{V}, \mathbb{W}$ . We have seen that a matrix representation of the empirical Koopman operator on these spaces can then be found without computing or inverting any of the Gramian matrices. If  $\mathbb{V}$  and  $\mathbb{W}$  are tensor product spaces, this procedure rapidly becomes infeasible, as we would have to calculate SVDs of the mode- $p$  unfoldings of  $\Psi(X), \Phi(Y)$ , which grow exponentially in size. However, we have seen in Section 2.2.3 that an approximation to these SVDs can be obtained by the multi-linear global SVD algorithm.

In the following section, we derive a corresponding multi-linear AMUSE algorithm, which we will call AMUSEt (*AMUSE on tensors*). This method only requires operations on individual cores of the TT representation (9) and the contraction of a tensor network. The analysis of this method will then be presented in Section 5.

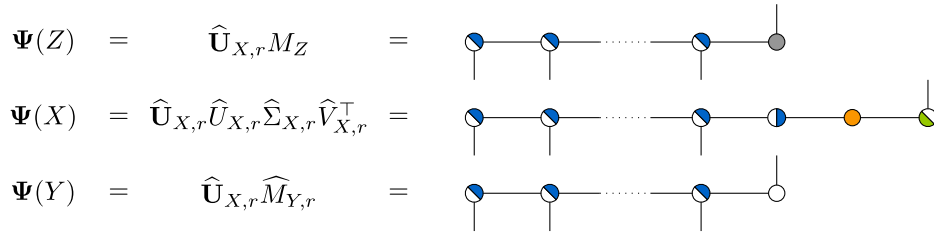
#### 3.1. The AMUSEt algorithm

Suppose we have given TT representations of the transformed data tensors  $\Psi(X)$  and  $\Phi(Y)$ . In order to construct the reduced matrix  $\widehat{M}_{r,r}$  in Algorithm 1, we first compute global SVDs of  $\Psi(X)$  and  $\Phi(Y)$ , i.e.,

$$\Psi(X) = \widehat{\mathbf{U}}_{X,r} \widehat{\Sigma}_{X,r} \widehat{\mathbf{V}}_{X,r}^T \quad \text{and} \quad \Phi(Y) = \widehat{\mathbf{U}}_{Y,r} \widehat{\Sigma}_{Y,r} \widehat{\mathbf{V}}_{Y,r}^T,$$



**Fig. 3.** Graphical representation of the reduced matrix: Given  $\Psi(X)$  (in form of a global SVD) and  $\Psi(Y)$ , the reduced matrix  $\widehat{M}_{\tau,r} \in \mathbb{R}^{r \times r}$  is computed by contracting the above tensor network. Half-filled circles in blue depict the cores of  $\widehat{U}_{X,r}$ ,  $\widehat{\Sigma}_{X,r}$  and  $\widehat{\Sigma}_{X,r}^{-1}$  are represented by orange and red circles, respectively, and the half-filled circle in green depicts the matrix  $\widehat{V}_{X,r}$  of the global SVD. Since we here do not assume any further properties on  $\Psi(Y)$ , we simply represent its TT cores by gray circles. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



**Fig. 4.** Construction of transformed data tensors: The tensor trains  $\Psi(X)$  and  $\Psi(Y)$  are extracted from the left-orthonormalized tensor train  $\Psi(Z)$ . Additionally,  $\Psi(X)$  is represented by its global SVD. Again, the cores of  $\widehat{U}_{X,r}$  (as well as  $\widehat{U}_{X,r}$ ) are represented by half-filled circles in blue,  $\widehat{\Sigma}_{X,r}$  by an orange circle, and  $\widehat{V}_{X,r}$  by a half-filled green circle. The last cores of  $\Psi(Z)$  and  $\Psi(Y)$  are depicted by a gray and a white circle, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

by applying Algorithm 3. Analogously to the matrix case, the reduced matrix can simply be written as  $\widehat{M}_{\tau,r} = \widehat{V}_{X,r}^T \widehat{V}_{Y,r}$ , see line 2 of Algorithm 1. In what follows, we therefore focus on single basis AMUSet. The construction of  $\widehat{M}_{\tau,r}$  as in line 2 of Algorithm 2 leads to the tensor network shown in Fig. 3. Note that here only the contractions of the TT cores of  $\widehat{U}_{X,r}$  and of  $\widehat{\Sigma}_{X,r}$  with its inverse cancel out.

The complexity of AMUSet is mainly determined by the computational cost of the global SVD, which can be estimated as  $O(p \min\{rm^2, r^2 n^2 m\})$ , where  $n$  is the maximum mode size. However, the complexity of AMUSet can be reduced in many cases. Oftentimes, we consider snapshot matrices  $X, Y \in \mathbb{R}^{d \times m}$ , which are extracted from a trajectory data matrix  $Z \in \mathbb{R}^{d \times \tilde{m}}$ ,  $\tilde{m} > m$ , and share a large number (i.e., close to  $m$ ) of common snapshot vectors. Instead of constructing the transformed data tensors separately, we can construct the TT decomposition of  $\Psi(Z)$  and then simply restrict the last TT core to the respective time steps in order to obtain the representations for  $\Psi(X)$  and  $\Psi(Y)$ . That is, given the data matrix  $Z$ , and index sets  $I_X, I_Y$  such that  $X = Z_{:,I_X}$  and  $Y = Z_{:,I_Y}$ , we directly form  $\Psi(Z)$  according to (9):

$$\Psi(Z) = \left[ \Psi^{(1)}(Z) \right] \otimes \left[ \Psi^{(2)}(Z) \right] \otimes \dots \otimes \left[ \Psi^{(p)}(Z) \right] \otimes \left[ \Psi^{(p+1)}(Z) \right].$$

Then, the tensor trains  $\Psi(X)$  and  $\Psi(Y)$  are given by

$$\begin{aligned} \Psi(X) &= \left[ \Psi^{(1)}(Z) \right] \otimes \left[ \Psi^{(2)}(Z) \right] \otimes \dots \otimes \left[ \Psi^{(p)}(Z) \right] \otimes \left[ \left( \Psi^{(p+1)}(Z) \right)_{:,I_X,1} \right] \\ \Psi(Y) &= \left[ \Psi^{(1)}(Z) \right] \otimes \left[ \Psi^{(2)}(Z) \right] \otimes \dots \otimes \left[ \Psi^{(p)}(Z) \right] \otimes \left[ \left( \Psi^{(p+1)}(Z) \right)_{:,I_Y,1} \right]. \end{aligned}$$

In order to arrive at a global SVD of  $\Psi(X)$ , we first left-orthonormalize the tensor train  $\Psi(Z)$ , from which  $\Psi(X)$  and  $\Psi(Y)$  can still be recovered by restricting the last core to the corresponding indices. Additionally, we compute a rank- $r$  SVD of only the last core of  $\Psi(X)$ , thereby completing the global SVD of  $\Psi(X)$  in the form  $\widehat{U}_{X,r} \widehat{U}_{X,r} \widehat{\Sigma}_{X,r} \widehat{V}_{X,r}^T$ . These transformations are visualized in Fig. 4.

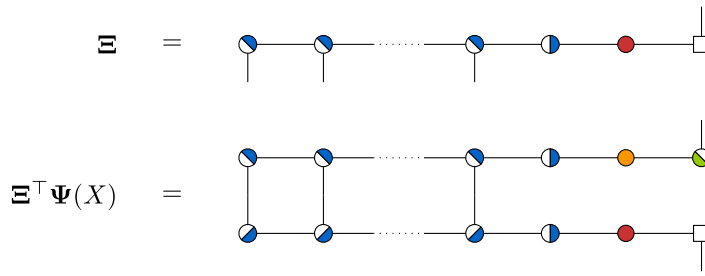
By construction,  $\Psi(X)$  and  $\Psi(Y)$  share the same segment  $\widehat{U}_{X,r}$ . Since the cores of  $\widehat{U}_{X,r}$  (as well as  $\widehat{U}_{X,r}$ ) are left-orthonormal, most of the contractions cancel out, and only four matrices remain, namely  $\widehat{V}_{X,r}$ ,  $\widehat{M}_{Y,r}$ ,  $\widehat{U}_{X,r}$ , and  $\widehat{\Sigma}_{X,r}^{-1}$ . Thus, the reduced matrix is then simply given by

$$\widehat{M}_{\tau,r} = \widehat{\Sigma}_{X,r}^{-1} \widehat{U}_{X,r}^T \widehat{U}_{X,r} \Psi(X) \Psi(Y)^T \widehat{U}_{X,r} \widehat{U}_{X,r} \widehat{\Sigma}_{X,r}^{-1} = \widehat{V}_{X,r}^T \widehat{M}_{Y,r} \widehat{U}_{X,r} \widehat{\Sigma}_{X,r}^{-1}. \quad (11)$$

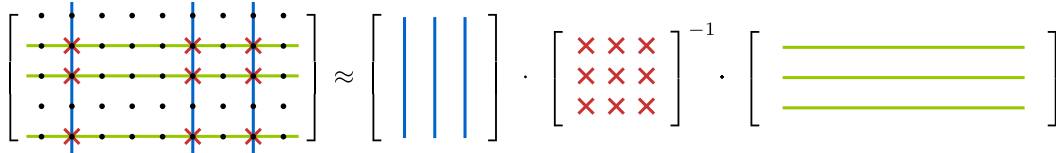
Finally, let us discuss the expression of singular vectors or eigenvectors computed from  $\widehat{M}_{\tau,r}$  with respect to the full tensor product basis. For instance, given the  $q$  leading eigenvectors of the reduced matrix  $\widehat{M}_{\tau,r}$  in form of a matrix  $\widehat{W} = [\widehat{w}_1, \dots, \widehat{w}_q]$ , the approximate eigentensors of  $\widehat{\mathcal{K}}_{\tau}(\mathbb{V}, \mathbb{V})$  can be expressed as a tensor train  $\mathcal{E}$  with  $\mathcal{E} = \widehat{U}_{X,r} \widehat{U}_{X,r} \widehat{\Sigma}_{X,r}^{-1} \widehat{W}$ , see line 2 of Algorithm 2. The evaluations of the associated eigenfunctions at all snapshots are then given by the matrix  $\mathcal{E}^T \Psi(X)$ , see [14]. The corresponding tensor network also breaks down to a simple matrix product, as is shown in Fig. 5.

**Remark 2.** Note that we do not need the cores of  $\widehat{U}_{X,r}$  after the orthonormalization procedure if we are only interested in the approximated eigenfunctions. For both the construction of the reduced matrix  $\widehat{M}_{\tau,r}$  as well as the eigenfunction evaluations  $\mathcal{E}^T \Psi(X)$ , see Fig. 5, the TT segment  $\widehat{U}_{X,r}$  is not required due to its orthonormality. This significantly reduces the storage consumption since we





**Fig. 5.** Graphical representation of the eigentensors and eigenfunctions: The tensor train  $\Xi$  is built by the contraction of  $\widehat{\mathbf{U}}_{X,r}$ ,  $\widehat{\mathbf{U}}_{X,r}$ ,  $\widehat{\Sigma}_{X,r}^{-1}$ , and  $\widehat{\mathbf{W}}$  (depicted by the square). The matrix  $\Xi^T \Psi(X)$  comprising the evaluations of the eigenfunctions at the given snapshots is constructed by multiplying the tensors  $\Psi(X)$  and  $\Xi$ . Similar to the construction of the reduced matrix only two cores remain since the orthonormal cores cancel out and  $\widehat{\Sigma}_{X,r}$  is multiplied by its inverse.



**Fig. 6.** CUR decomposition: The matrix on the left-hand side is approximated by the matrix product  $C \cdot U^{-1} \cdot R$ , where  $C$  (blue lines) is a column subset,  $R$  (green lines) is a row subset, and  $U$  (red crosses) is the intersection matrix. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

are able to construct the left-orthonormalized version of  $\Psi(X)$  step by step, i.e., we only need to store two TT cores in memory at the same time.

#### 4. HOCUR-based approach

In practice, the direct construction of transformed data tensors as described in Section 2.2.2 may be infeasible due to a large number of basis functions or snapshots. Thus, an alternative isolation technique for the modes of  $\Psi(X)$  and  $\Phi(Y)$  has to be used in order to apply AMUSEt. Our idea to circumvent this problem is a combination of different techniques from [38,45,46], specifically adapted to transformed data tensors as described in Section 2.2.2. Based on so-called *CUR decompositions*, i.e., representing a matrix in terms of appropriate row and column subsets, we propose an iterative technique in Section 4.2. The aim is to construct a low-rank TT decomposition of the transformed data tensor without storing the complete representation of  $\Psi(X)$  as given in (9).

##### 4.1. Higher-order CUR decomposition

For a matrix  $M \in \mathbb{R}^{m \times n}$ , a CUR decomposition consists of index sets  $I, J$ , as well as submatrices  $C = M_{:,J}$ ,  $U = M_{I,J}$ , and  $R = M_{I,:}$ , such that  $M \approx C \cdot U^{-1} \cdot R$ , see Fig. 6.

There are different methods to find optimal sets of rows and columns, cf. [47]. An important subproblem is the following: given a set of column indices  $J = \{j_1, \dots, j_r\}$  with  $r \leq \min(m, n)$  (and  $M_{:,j}$  having full column rank), find an optimal subset of row indices  $I = \{i_1, \dots, i_r\}$ . This NP-hard problem can be approximately solved by applying the *maximum-volume principle* to  $M_{:,j}$ , so that the infinity norm of  $M - M_{:,j} \cdot M_{I,j}^{-1} \cdot M_{I,:}$  is minimized over  $I$ , see [45,46]. We refer to this algorithm as Maxvol from now on.

As described in [38], the CUR decomposition can be generalized to a tensor if its mode- $k$  unfoldings are successively decomposed using CURs. This method presents an alternative to the decomposition (9) if applied to the transformed data tensor  $\Psi(X)$ . However, the procedure requires pre-defined row and column subsets for each unfolding. The authors of [38] also suggested an iterative algorithm to circumvent this problem: after initializing row and column subsets in some way, the method alternates between updating the column subsets while all row sets are fixed, and vice versa. This algorithm is the basis of the method we will present in Section 4.2.

##### 4.2. HOCUR for transformed data tensors

As suggested in [38], Algorithm 4 successively updates the row sets of the unfolded residual tensors during a forward loop, while all column sets are fixed. Then, column sets are updated during a backward loop, with all row sets fixed, and the entire procedure is repeated until convergence. The key insight, used in lines 5–9 and 13–14, is that each update only operates on a small subtensor which is easily evaluated. Assume we are given a row set  $\mathbf{I} = \mathbf{I}_q = \{\mathbf{i}_1, \dots, \mathbf{i}_{r_q}\}$  of multi-indices comprising modes  $n_1, \dots, n_q$ , and a column set  $\mathbf{J} = \mathbf{J}_{q+2} = \{\mathbf{j}_1, \dots, \mathbf{j}_s\}$  of multi-indices comprising modes  $n_{q+2}, \dots, n_p, m$ :

$$\begin{aligned} \mathbf{i}_1, \dots, \mathbf{i}_{r_q} &\in \{1, \dots, n_1\} \times \dots \times \{1, \dots, n_q\}, \\ \mathbf{j}_1, \dots, \mathbf{j}_s &\in \{1, \dots, n_{q+2}\} \times \dots \times \{1, \dots, n_p\} \times \{1, \dots, m\}. \end{aligned}$$

Then, a new extended row set comprising the first  $q + 1$  modes

$$\mathbf{i}_1, \dots, \mathbf{i}_{r_{q+1}} \in \{1, \dots, n_1\} \times \dots \times \{1, \dots, n_{q+1}\},$$

can be obtained by applying Algorithm Maxvol to the submatrix  $\Psi(X)_{\mathbf{I},\mathbf{J}} \in \mathbb{R}^{r_q \cdot n_{q+1} \times s}$ , given by

$$\Psi(X)_{\mathbf{I},\mathbf{J}} = \begin{bmatrix} \Psi(X)_{i_1,:j_1} & \cdots & \Psi(X)_{i_1,:j_s} \\ \vdots & \ddots & \vdots \\ \Psi(X)_{i_q,:j_1} & \cdots & \Psi(X)_{i_q,:j_s} \end{bmatrix}. \quad (12)$$

This matrix is easily set up using basis function evaluations. More precisely, given multi-indices  $\mathbf{i} = (i_1, \dots, i_q) \in \mathbf{I}$  and  $\mathbf{j} = (j_1, \dots, j_s) \in \mathbf{J}$ , entries  $\Psi(X)_{\mathbf{i},\mathbf{j}}$  of  $\Psi(X)_{\mathbf{I},\mathbf{J}}$  are given by

$$\begin{aligned} \Psi(X)_{\mathbf{i},\mathbf{j}} &= \Psi(X)_{i_1, \dots, i_q, :i_{q+2}, \dots, i_p, k} \\ &= \underbrace{\psi_{1,i_1}(x_k) \cdot \dots \cdot \psi_{q,i_q}(x_k)}_{\in \mathbb{R}} \cdot \underbrace{\psi_{q+1}(x_k)}_{\in \mathbb{R}^{n_{q+1}}} \cdot \underbrace{\psi_{q+2,i_{q+2}}(x_k) \cdot \dots \cdot \psi_{p,i_p}(x_k)}_{\in \mathbb{R}}. \end{aligned}$$

Note that the last entry of the column index  $\mathbf{j}$  determines the snapshot  $x_k$  where the product is evaluated.

**Algorithm 4:** Higher-order CUR decomposition.

**Input:** data matrix  $X = [x_1, \dots, x_m] \in \mathbb{R}^{d \times m}$ , basis functions  $\psi_{i,j_i}$ ,  $i = 1, \dots, p$ ,  
 $j_i = 1, \dots, n_i$ , maximum ranks  $r_1, \dots, r_p$  with  $r_q \leq n_{q+1} \cdot r_{q+1}$ ,  
 number of iterations  $N$ , multiplier  $\alpha > 1$

**Output:** TT approximation of the transformed data tensor  $\Psi(X)$

- 1: Set  $n_{p+1} = m$ ,  $r_0 = r_{p+1} = 1$ , and  $\mathbf{I}_0 = \{\emptyset\}$ .
- 2: Define initial multi-index column sets  $\mathbf{J}_2, \dots, \mathbf{J}_{p+2}$ .
- 3: **for**  $k = 1, \dots, N$  **do**
- 4:     **for**  $l = 1, \dots, p$  **do** (First half sweep)
- 5:         Extract submatrix  $M = \Psi(X)_{\mathbf{I}_{l-1}, \mathbf{J}_{l+1}}$ , see (12).
- 6:         **if**  $k = 1$  **then**
- 7:             Find set of linearly independent columns  $J$  of  $M$  with  $|J| \leq r_l$ .
- 8:             Set  $M$  to  $M_{:,J}$  and  $r_l$  to  $|J|$ .
- 9:             Apply Algorithm Maxvol to  $M$  to extract row set  $I$ .
- 10:             Compute multi-index row set  $\mathbf{I}_l$  from  $\mathbf{I}_{l-1}$  and  $I$ , see (13).
- 11:             Define core  $\Psi(X)^{(l)}$  as  $M \cdot M_{I,:}^{-1}$  reshaped as  $\mathbb{R}^{r_{l-1} \times n_l \times r_l}$ .
- 12:     **for**  $l = p + 1, \dots, 2$  **do** (Second half sweep)
- 13:         Extract submatrix  $M = \Psi(X)_{\mathbf{I}_{l-1}, \mathbf{J}_{l+1}}$  and reshape as  $\mathbb{R}^{r_{l-1} \times n_l \times r_l}$ .
- 14:         Apply Algorithm Maxvol to  $M^T$  to extract column set  $J$  and set  $r_{l-1} = |J|$ .
- 15:         Compute multi-index column set  $\mathbf{J}_l$  from  $J$  and  $\mathbf{J}_{l+1}$ .
- 16:         Define core  $\Psi(X)^{(l)}$  as  $M_{:,J}^{-1} \cdot M$  reshaped as  $\mathbb{R}^{r_{l-1} \times n_l \times r_l}$ .
- 17: Define first core  $\Psi(X)^{(1)}$  as  $\Psi(X)_{\mathbf{I}_0, \mathbf{J}_2}$  reshaped as  $\mathbb{R}^{1 \times n_1 \times r_1}$ .

Let us elaborate on a few more details of Algorithm 4. First, note that multi-index sets for the construction of the submatrices  $\Psi(X)_{\mathbf{I},\mathbf{J}}$  are nested sets by construction. In line 10, after each application of Algorithm Maxvol, the resulting single-index set  $I = \{i_1, \dots, i_{r_{q+1}}\}$  needs to be converted into a multi-index row set for modes  $n_1, \dots, n_{q+1}$ . Given a multi-index row set  $\mathbf{I}_q = \{\mathbf{i}_1, \dots, \mathbf{i}_{r_q}\}$  as above, each row of  $\Psi(X)_{\mathbf{I}_q, \mathbf{J}_{q+2}}$  can naturally be associated with a multi-index  $(k_1, k_2) \in \{1, \dots, r_q\} \times \{1, \dots, n_{q+1}\}$ . Hence, we map each single-index  $i_k \in I$  to a multi-index  $\hat{i}_k = (i_{k,1}, i_{k,2})$  in  $\{1, \dots, r_q\} \times \{1, \dots, n_{q+1}\}$ , and then define the extended multi-index row set

$$\mathbf{I}_{q+1} = \{(\hat{i}_{1,1}, \hat{i}_{1,2}), \dots, (\hat{i}_{r_{q+1},1}, \hat{i}_{r_{q+1},2})\}. \quad (13)$$

Column sets are updated analogously in line 15 of Algorithm 4.

Second, the algorithm requires initial column sets which are generated in line 2. While it was suggested in [38] to pick these columns at random, we build them up recursively to ensure the column sets are also nested. Starting from  $\mathbf{J}_{p+2} = \{\emptyset\}$ , column set  $\mathbf{J}_q$  is obtained by simply selecting the first  $\min(\alpha \cdot r_q, n_{q+1} \cdot r_{q+1})$  indices out of the index set  $\{1, \dots, n_{q+1} \cdot r_{q+1}\}$ , and then joining them with multi-index column set  $\mathbf{J}_{q+1}$  as described above in (13). In practice, we found it helpful to select a rather large number of columns at this point, as these initial columns would often be highly redundant. The parameter  $\alpha$  can be tuned to ensure enough columns are selected during the initialization stage.

Third, we need to find index sets of linearly independent columns of the matrices  $\Psi(X)_{\mathbf{I},\mathbf{J}}$  during the first iteration, see line 7. This can again be done by applying QR decompositions with column pivoting. And finally, the cores of the TT approximation of  $\Psi(X)$  are updated in lines 11 and 16, by multiplication of parts of the determined CUR decomposition. In the notation used in Section 4, the updated cores are given by tensor foldings of  $C \cdot U^{-1}$  and  $U^{-1} \cdot R$ , respectively.

## 5. Analysis of AMUSE and AMUSET

The goal of this section is to show that AMUSET as introduced in Section 3.1 produces an empirical matrix representation of the Koopman operator on data-dependent subspaces of the tensor spaces  $\mathbb{V}, \mathbb{W}$ , and to establish the convergence of this representation in the limit of infinite data. We first investigate the standard case in Section 5.1. We define the spectral subspaces and their empirical counterparts. If carried out at fixed prescribed SVD rank  $r$ , the standard AMUSE algorithm amounts to approximating the Koopman

operator on empirical spectral subspaces. Convergence of these spaces and the corresponding Koopman operator representation is then established. This analysis can be carried over to the tensor case. To this end, we first provide a multi-linear analogue of (empirical) spectral subspaces in Section 5.2, and also establish convergence in the limit of infinite data. In Section 5.3, we then show that, analogous to the standard AMUSE algorithm, AMUSEt is indeed the algorithmic framework to compute projections of evolution operators on multi-linear spectral subspaces.

In what follows, the space of bounded linear operators on a Hilbert space  $\mathbb{H}$ , equipped with the standard operator norm, is denoted by  $L(\mathbb{H})$ . The orthogonal projector onto a finite-dimensional subspace  $\mathbb{V}$  is labeled  $\mathcal{P}_{\mathbb{V}}$ . The distance between subspaces  $\mathbb{V}$  and  $\mathbb{W}$  of the same dimension  $n$  is given by

$$d(\mathbb{V}, \mathbb{W}) = \sup_{v \in \mathbb{V}, \|v\|=1} \inf_{w \in \mathbb{W}} \|w - v\| = \|(\text{Id} - \mathcal{P}_{\mathbb{W}})\mathcal{P}_{\mathbb{V}}\|_{L(\mathbb{H})} = \|\mathcal{P}_{\mathbb{V}} - \mathcal{P}_{\mathbb{W}}\|_{L(\mathbb{H})}.$$

Finally, consider the situation that  $\mathbb{V}$  is a finite-dimensional Hilbert space of functions on  $\mathbb{R}^d$ , and  $x_k, k \in \mathbb{N}$ , is a sequence of  $\mathbb{R}^d$ -valued random variables such that the first equation in (6) holds true almost surely. Then for  $m$  large enough, the bilinear form

$$\langle \psi, \tilde{\psi} \rangle_{\mathbb{V}}^{\wedge} := \frac{1}{m} \sum_{k=1}^m \psi(x_k) \tilde{\psi}(x_k), \quad \psi, \tilde{\psi} \in \mathbb{V},$$

is an inner product on  $\mathbb{V}$ . Orthogonal projections with respect to this empirical inner product will be labeled  $\widehat{\mathcal{P}}$ , accordingly.

### 5.1. Spectral subspaces and AMUSE

We begin this section by explicitly stating the definition of the empirical counterpart of the projected Koopman operator  $\mathcal{K}_{\tau}(\mathbb{V}, \mathbb{W})$ , independently of the basis sets used for  $\mathbb{V}, \mathbb{W}$ , see [48]. Analogous results can be obtained for the Perron–Frobenius and forward–backward operators:

**Proposition 2.** *Let (3) hold, and let  $\mathbb{V}, \mathbb{W}$  be finite-dimensional subspaces of  $L^2_{\rho_0}$  and  $L^2_{\rho_1}$ . For almost all sequences  $(x_k, y_k)$ , and  $m$  large enough, there is a linear operator  $\widehat{\mathcal{K}}_{\tau}(\mathbb{V}, \mathbb{W}): \mathbb{W} \rightarrow \mathbb{V}$ , satisfying*

$$\langle \psi, \widehat{\mathcal{K}}_{\tau}(\mathbb{V}, \mathbb{W})\phi \rangle_{L^2_{\rho_0}}^{\wedge} = \frac{1}{m} \sum_{k=1}^m \psi(x_k)\phi(y_k) \quad \forall \psi \in \mathbb{V}. \tag{14}$$

Its matrix representation with respect to  $\psi, \phi$  is  $\widehat{K}_{\tau}(\psi, \phi) = (\widehat{C}(\psi))^{-1}\widehat{A}(\psi, \phi)$ . Also,

$$\|\widehat{K}_{\tau}(\mathbb{V}, \mathbb{W}) - \mathcal{K}_{\tau}(\mathbb{V}, \mathbb{W})\|_{L(\mathbb{W}, \mathbb{V})} \rightarrow 0$$

for  $m \rightarrow \infty$ , where the topology on  $L(\mathbb{W}, \mathbb{V})$  is induced by the standard inner products on  $L^2_{\rho_0}, L^2_{\rho_1}$ .

**Proof.** The right-hand side of (14) is a linear functional on the finite-dimensional space  $\mathbb{V}$  with empirical inner product, which ensures existence of the operator  $\widehat{K}_{\tau}(\mathbb{V}, \mathbb{W})$ . The matrix representation can be directly verified, and by (3), this representation converges to that of  $\mathcal{K}_{\tau}(\mathbb{V}, \mathbb{W})$  in any matrix norm, which proves the last statement.  $\square$

We also observe that for subspaces  $\mathbb{F}_1 \subset \mathbb{V}$  and  $\mathbb{F}_2 \subset \mathbb{W}$ , we have  $\widehat{K}_{\tau}(\mathbb{F}_1, \mathbb{F}_2) = \widehat{\mathcal{P}}_{\mathbb{F}_1}\widehat{K}_{\tau}(\mathbb{V}, \mathbb{W})\widehat{\mathcal{P}}_{\mathbb{F}_2}$ . Next, we introduce the family of subspaces which serve as reduced trial spaces for the Koopman operator if the standard AMUSE algorithm is employed. We have briefly encountered these spaces before in Section 2.1.2.

**Definition 2 (Spectral Subspace).** Let  $\mathbb{V}$  be a finite-dimensional Hilbert space of functions with basis  $\psi = \{\psi_j\}_{j=1}^n$ . Denote the spectral decomposition of the Gramian matrix by  $C(\psi) = U_{\psi}\Sigma_{\psi}^2U_{\psi}^{\top}$ , with eigenvalues arranged in decreasing order. For  $r \leq n$  such that  $\sigma_r > \sigma_{r+1}$ , denote the first  $r$  columns of  $U_{\psi}$  by  $U_{\psi,r}$ , and the upper  $r \times r$ -block of the diagonal matrix by  $\Sigma_{\psi,r}$ . Then the space  $\mathbb{G}_r \subset \mathbb{V}$  spanned by orthonormal functions  $\eta_r^{\top} = \psi^{\top}(U_{\psi,r}\Sigma_{\psi,r}^{-1})$  is called *spectral subspace (of order  $r$ )* of  $\mathbb{V}$ . The coefficient vector space associated with  $\mathbb{G}_r$  is denoted by  $\mathbb{B}_r \subset \mathbb{R}^n$ .

#### Remark 3.

- (i) The condition  $\sigma_r > \sigma_{r+1}$  is necessary for  $\mathbb{G}_r$  to be well-defined, otherwise it would be unclear how to break up the eigenspace corresponding to  $\sigma_r = \sigma_{r+1}$ .
- (ii) The spectral subspace depends on the basis chosen for the Hilbert space  $\mathbb{V}$ . We will emphasize this dependence by writing  $\mathbb{G}_r(\psi), \mathbb{B}_r(\psi)$  whenever necessary.
- (iii) If an orthogonal change of basis is used, that is, the basis  $\psi$  changes to  $\tilde{\psi}^{\top} = \psi^{\top}Q$ , with  $Q^{\top}Q = \text{Id}_n$ , then the Gramian matrix changes to  $C(\tilde{\psi}) = Q^{\top}C(\psi)Q$ . Hence, the spectral subspaces  $\mathbb{G}_r$  remain the same, the associated coefficient vector spaces are  $\mathbb{B}_r = \text{span}(Q^{\top}U_{\psi,r})$ .
- (iv) Clearly, empirical spectral subspaces  $\widehat{\mathbb{G}}_r(\psi), \widehat{\mathbb{B}}_r(\psi)$  can be defined in the same way using the empirical Gramian matrix.

Empirical spectral subspaces are stable in the limit of infinite data:

**Lemma 1.** *Let  $\mathbb{V}$  be a finite-dimensional Hilbert space of functions on  $\mathbb{R}^d$  with basis  $\psi = \{\psi_j\}_{j=1}^n$ . For fixed  $r \leq n$ , let the spectral subspace  $\mathbb{G}_r$  be well-defined. Let  $x_k, k \in \mathbb{N}$ , be a sequence of  $\mathbb{R}^d$ -valued random variables such that the first equation in (6) holds true almost surely. Then we also have with probability one:*

$$\lim_{m \rightarrow \infty} d(\mathbb{B}_r, \widehat{\mathbb{B}}_r) = 0, \quad \lim_{m \rightarrow \infty} d(\mathbb{G}_r, \widehat{\mathbb{G}}_r) = 0, \quad \lim_{m \rightarrow \infty} \|\widehat{\mathcal{P}}_{\widehat{\mathbb{G}}_r(\psi)} - \mathcal{P}_{\mathbb{G}_r(\psi)}\|_{L(\mathbb{V})} = 0,$$

where the topology on  $L(\mathbb{V})$  is induced by the standard inner product on  $\mathbb{V}$ .

**Proof.** By assumption, we have  $\lim_{m \rightarrow \infty} \|C(\psi) - \widehat{C}(\psi)\|_2 = 0$  almost surely, hence  $\widehat{\mathbb{B}}_r, \widehat{\mathbb{G}}_r$  can be defined with probability one if  $m$  is large enough. As the gap between  $\sigma_r$  and  $\sigma_{r+1}$  is positive, perturbation theory for symmetric matrices [49][Chapter 5, Thm 3.6] ensures that the distance in  $\mathbb{R}^n$  between  $\mathbb{B}_r$  and  $\widehat{\mathbb{B}}_r$  converges to zero. As these are the coefficient vector spaces corresponding to  $\mathbb{G}_r, \widehat{\mathbb{G}}_r$ , Lemma 4 (iii) in Appendix implies the second claim. Finally, we choose an orthonormal basis (ONB)  $U \in \mathbb{R}^{n \times r}$  of  $\mathbb{B}_r(\psi)$ , and invoke Lemma 5 to choose ONBs  $\widehat{U}$  of  $\widehat{\mathbb{B}}_r(\psi)$  such that  $\|\widehat{U} - U\|_2 \rightarrow 0$  holds almost surely. Using the representations in Lemma 4(i) and (ii), this implies

$$\begin{aligned} & \|\widehat{\mathcal{P}}_{\widehat{\mathbb{G}}_r(\psi)} - \mathcal{P}_{\mathbb{G}_r(\psi)}\|_{L(\mathbb{V})} \\ &= \|C(\psi)^{1/2} [\widehat{U}(\widehat{U}^\top \widehat{C}(\psi) \widehat{U})^{-1} \widehat{U}^\top \widehat{C}(\psi) - U(U^\top C(\psi) U)^{-1} U^\top C(\psi)] C(\psi)^{-1/2}\|_2 \rightarrow 0, \end{aligned}$$

using that  $\widehat{C}(\psi) \rightarrow C(\psi)$  and  $\widehat{U} \rightarrow U$ .  $\square$

We can now use the previous result to complete our analysis of AMUSE. We have already seen in Section 2.1.4 that, if the SVD rank  $r$  is fixed, the basis sets  $\widehat{\eta}_r, \widehat{\zeta}_r$  appearing in the standard AMUSE algorithm are spanning the empirical spectral subspaces  $\widehat{\mathbb{G}}_r(\psi), \widehat{\mathbb{G}}_r(\phi)$  in  $\mathbb{V}$  and  $\mathbb{W}$ . We show that the resulting empirical projection of the Koopman operator converges in concert with these spaces.

**Proposition 3.** Let  $\mathcal{X}_t$  be a dynamical system such that (3) holds true. Let  $\mathbb{V} \subset L^2_{\rho_0}$  and  $\mathbb{W} \subset L^2_{\rho_1}$  be  $n$ -dimensional subspaces with bases  $\psi, \phi$ . For  $r \leq n$  such that  $\mathbb{G}_r(\psi)$  and  $\mathbb{G}_r(\phi)$  are both well-defined, we conclude that almost surely:

$$\|\widehat{\mathcal{K}}_\tau(\widehat{\mathbb{G}}_r(\psi), \widehat{\mathbb{G}}_r(\phi)) - \mathcal{K}_\tau(\mathbb{G}_r(\psi), \mathbb{G}_r(\phi))\|_{L(\mathbb{W}, \mathbb{V})} \rightarrow 0.$$

**Proof.** We have already seen that

$$\widehat{\mathcal{K}}_\tau(\widehat{\mathbb{G}}_r(\psi), \widehat{\mathbb{G}}_r(\phi)) = \widehat{\mathcal{P}}_{\widehat{\mathbb{G}}_r(\psi)} \widehat{\mathcal{K}}_\tau(\mathbb{V}, \mathbb{W}) \widehat{\mathcal{P}}_{\widehat{\mathbb{G}}_r(\phi)}, \quad \mathcal{K}_\tau(\mathbb{G}_r(\psi), \mathbb{G}_r(\phi)) = \mathcal{P}_{\mathbb{G}_r(\psi)} \mathcal{K}_\tau(\mathbb{V}, \mathbb{W}) \mathcal{P}_{\mathbb{G}_r(\phi)}.$$

Since  $\widehat{\mathcal{P}}_{\widehat{\mathbb{G}}_r(\psi)} \rightarrow \mathcal{P}_{\mathbb{G}_r(\psi)}$  and  $\widehat{\mathcal{P}}_{\widehat{\mathbb{G}}_r(\phi)} \rightarrow \mathcal{P}_{\mathbb{G}_r(\phi)}$  by Lemma 1, and  $\widehat{\mathcal{K}}_\tau(\mathbb{V}, \mathbb{W}) \rightarrow \mathcal{K}_\tau(\mathbb{V}, \mathbb{W})$  by Proposition 2, the claim follows.  $\square$

### 5.2. Multi-linear spectral subspaces

For the rest of this section, we assume that a vector of fixed ranks  $\mathbf{r} = [r_1, \dots, r_p]$  is given. Also,  $\mathbb{V} = \bigotimes_{k=1}^p \mathbb{V}^k$  is a tensor product space of functions on  $\mathbb{R}^d$ , where  $\mathbb{V}^k = \text{span}\{\psi_{k,i_k}\}_{i_k=1}^{n_k}$ . The full tensor product basis is denoted by  $\Psi$ . We also introduce the symbols  $\mathbb{V}^{:k} = \bigotimes_{l=1}^k \mathbb{V}^l$  for the partial tensor product up to mode  $k$ , the corresponding basis is denoted  $\Psi_{:k}$ . Similarly, if  $\mathbf{T}$  is a tensor train of order  $p$ , the partial tensor train up to mode  $k \leq p$  is denoted  $\mathbf{T}^{:k} = \llbracket \mathbf{T}^{(1)} \rrbracket \otimes \dots \otimes \llbracket \mathbf{T}^{(k)} \rrbracket$ .

The results of the previous section can now be generalized to tensor-structured subspaces. The construction of multi-linear spectral subspaces, to be described below, is inspired by the global SVD. Recall from Section 2.2.3 that the first step of Algorithm 3 applied to  $\Psi(X)$  and  $\Psi(Y)$  is the same as if we were applying standard AMUSE just to  $\mathbb{V}^1$  and  $\mathbb{W}^1$ . For  $k = 1$ , and focusing just on  $\Psi(X)$  for the sake of illustration, the matrix of singular vectors  $U$  in line 2 of Algorithm 3 encodes a basis of the  $r_1$ -dimensional spectral subspace  $\mathbb{G}_{r_1}(\psi_1)$ . As we show in Section 5.3, the next step amounts to computing a basis of the  $r_2$ -dimensional spectral subspace for a specific basis of the tensor product  $\mathbb{G}_{r_1}(\psi_1) \otimes \mathbb{V}^2$ . This procedure is then repeated through all steps of the method. Some care, however, needs to be taken when choosing the basis set for each of these spectral subspaces.

**Definition 3 (Multi-linear Spectral Subspaces).** Define  $\mathbb{G}^{:0} = \text{span}\{1\}$  and  $\theta_0 = 1$ . Recursively for  $k = 1, \dots, p$ , consider the Gramian matrix  $C(\text{vec}(\theta_{k-1} \otimes \psi_k))$  of  $\mathbb{G}^{:k-1} \otimes \mathbb{V}^k$ , and denote its eigenvalues by  $\sigma_{k,l}^2$ ,  $1 \leq l \leq r_{k-1} n_k$ . If  $\sigma_{k,r_k} > \sigma_{k,r_k+1}$ , denote its  $r_k$ -dimensional spectral subspace by  $\mathbb{G}^{:k}$ . Choose any orthonormal basis  $U_k \in \mathbb{R}^{r_{k-1} n_k \times r_k}$  of the associated coefficient vector space in  $\mathbb{R}^{r_{k-1} n_k}$ , and define a basis set for  $\mathbb{G}^{:k}$  as  $(\theta_k)^\top = (\text{vec}(\theta_{k-1} \otimes \psi_k))^\top U_k$ .

The subspaces  $\mathbb{G}^{:k}$  are then called *multi-linear spectral subspaces*. Their associated coefficient vector spaces in  $\mathbb{R}^{\prod_{l=1}^k n_l}$ , with respect to the basis  $\Psi_{:k}$ , are denoted  $\mathbb{B}^{:k}$ .

#### Remark 4.

- (i) Definition 3 can be repeated almost verbatim to define empirical counterparts of all quantities introduced above. As before, we will use hats  $\widehat{\cdot}$  to denote these quantities.
- (ii) Note that the ONBs  $U_k$  used in the construction are coefficient vectors with respect to the previous level of recursion, that is, to a basis of  $\mathbb{G}^{:k-1} \otimes \mathbb{V}^k$ . The coefficient vector space  $\mathbb{B}^{:k}$  refers to the full  $k$ -fold tensor product basis  $\Psi_{:k}$ .

The spaces  $\mathbb{G}^{:k}$  are independent of the specific orthonormal bases used in each step of the construction. In fact, the following results can be proven:

**Lemma 2.** Compile the orthonormal bases  $U_k$  in the construction of Definition 3 into a tensor train  $\mathbf{U}$ , that is:

$$\mathbf{U} = \llbracket \mathbf{U}^{(1)} \rrbracket \otimes \dots \otimes \llbracket \mathbf{U}^{(p)} \rrbracket, \quad \mathbf{U}^{(k)}|_2 = U_k, \quad 1 \leq k \leq p.$$

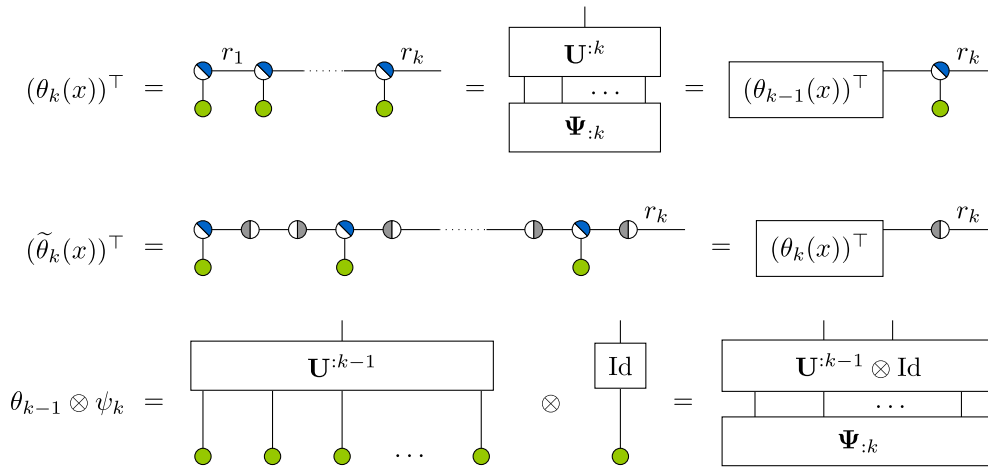
For  $1 \leq k \leq p$ , the partial tensor train  $\mathbf{U}^{:k} \in \mathbb{R}^{n_1 \times \dots \times n_k \times r_k}$  is the coefficient tensor of the basis  $\theta_k$  with respect to  $\Psi_{:k}$ , that is  $(\theta_k)^\top = (\Psi_{:k}|_k)^\top \mathbf{U}^{:k}|_k$ . Moreover, let  $U_k, 1 \leq k \leq p$  be a different sequence of ONBs used in the construction in Definition 3, with resulting tensor train  $\mathbf{U}$  and basis sets  $\tilde{\theta}_k$ . Then there exist orthonormal matrices  $Q_k \in \mathbb{R}^{r_k \times r_k}$ ,  $0 \leq k \leq p$ , such that

$$\widetilde{\mathbf{U}}^{(k)}|_2 = (Q_{k-1} \otimes \text{Id}_{n_k})^\top \mathbf{U}^{(k)}|_2 Q_k \tag{15}$$

for  $1 \leq k \leq p$ . In particular,  $\theta_k$  and  $\tilde{\theta}_k$  are related by

$$(\theta_k)^\top = (\Psi_{:k}|_k)^\top \mathbf{U}^{:k}|_k, \quad (\tilde{\theta}_k)^\top = (\Psi_{:k}|_k)^\top \mathbf{U}^{:k}|_k Q_k = (\theta_k)^\top Q_k, \tag{16}$$

and the spaces  $\mathbb{G}^{:k}$  and  $\mathbb{B}^{:k}$  are the same in both cases.



**Fig. 7.** Relationship between the basis sets  $\theta_k$  and  $\tilde{\theta}_k$  for the construction of multi-linear spectral subspaces. Half-filled blue circles represent the cores of  $\mathbf{U}$ , green circles the corresponding tensor product bases, and half-filled gray circles the orthonormal matrices  $Q_k$ . The third row illustrates the recursive expression of the product basis  $\theta_{k-1} \otimes \psi_k$  with respect to the coefficient tensor  $\mathbf{U}^{k-1}$  of  $\theta_{k-1}$ . (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

**Proof.** The fact that each  $\mathbf{U}^k$  is the coefficient tensor of the basis  $\theta_k$  can be verified directly by an inductive argument. To prove (15)–(16), we proceed by induction over  $k$ . The claims are clearly true for  $k = 1$  with  $Q_0 = 1$ . For general  $k$ , if the claims of the Lemma are true for  $k - 1$ , then (16) for  $k - 1$ , and Remark 3 (iii), imply that the spectral subspaces of  $C(\text{vec}(\theta_{k-1} \otimes \psi_k))$  and  $C(\text{vec}(\tilde{\theta}_{k-1} \otimes \psi_k))$  are the same, and the coefficient vector spaces are related by the transformation  $(Q_{k-1} \otimes \text{Id}_{n_k})^T$ . It follows that there is an orthonormal  $Q_k \in \mathbb{R}^{r_k \times r_k}$  such that

$$\tilde{U}_k = (Q_{k-1} \otimes \text{Id}_{n_k})^T U_k Q_k,$$

as claimed. In turn, this implies that

$$\begin{aligned} \tilde{\mathbf{U}}^k|_k &= (\tilde{\mathbf{U}}^{k-1}|_{k-1} \otimes \text{Id}_{n_k}) \tilde{U}_k \\ &= ((\mathbf{U}^{k-1}|_{k-1} Q_{k-1}) \otimes \text{Id}_{n_k}) (Q_{k-1} \otimes \text{Id}_{n_k})^T U_k Q_k \\ &= (\mathbf{U}^{k-1}|_{k-1} \otimes \text{Id}_{n_k}) U_k Q_k = \mathbf{U}^k|_k Q_k. \end{aligned}$$

This proves (16), and hence  $\theta_k$  and  $\tilde{\theta}_k$  span the same space  $\mathbb{G}^k$  with coefficient vector space  $\mathbb{B}^k$ .  $\square$

A pictorial illustration of the relationship between  $\theta_k$  and  $\tilde{\theta}_k$  is given in the first two rows of Fig. 7. Now that multi-linear spectral subspaces have been shown to be well-defined, it is time to generalize Lemma 1 to the multi-linear case:

**Lemma 3.** Let  $\mathbb{V} = \bigotimes_{k=1}^p \mathbb{V}^k$  be a finite-dimensional tensor space of functions on  $\mathbb{R}^d$  as described above. For a fixed vector of TT ranks  $\mathbf{r}$ , let all multi-linear spectral subspaces  $\mathbb{G}^k$  be well-defined. Let  $x_k$ ,  $k \in \mathbb{N}$  be a sequence of  $\mathbb{R}^d$ -valued random variables such that the first equation in (6) holds true almost surely in  $\mathbb{V}$ . Then we have for all  $1 \leq k \leq p$ :

$$d(\mathbb{B}^k, \hat{\mathbb{B}}^k) \rightarrow 0, \quad d(\mathbb{G}^k, \hat{\mathbb{G}}^k) \rightarrow 0, \quad \|\hat{\mathcal{P}}_{\hat{\mathbb{G}}^k} - \mathcal{P}_{\mathbb{G}^k}\|_{L(\mathbb{V})} \rightarrow 0.$$

**Proof.** The proof is an inductive application of Lemma 1. For  $k = 1$ , the result is the same as that of Lemma 1, so let us assume the claim stands for  $k - 1$ . We fix the coefficient tensors  $\mathbf{U}^k$ ,  $\hat{\mathbf{U}}^k$ . By Lemma 2, the unfoldings  $\mathbf{U}^{k-1}|_{k-1}$ ,  $\hat{\mathbf{U}}^{k-1}|_{k-1}$  provide ONBs of  $\mathbb{B}^{k-1}$ ,  $\hat{\mathbb{B}}^{k-1}$ . We use the induction hypothesis and Lemma 5 to determine an orthonormal  $\hat{Q}_{k-1} \in \mathbb{R}^{r_{k-1} \times r_{k-1}}$  such that  $\|\mathbf{U}^{k-1}|_{k-1} - \hat{\mathbf{U}}^{k-1}|_{k-1} \hat{Q}_{k-1}\|_2 \rightarrow 0$  almost surely. By extension, we also have

$$\|\mathbf{U}^{k-1}|_{k-1} \otimes \text{Id}_{n_k} - (\hat{\mathbf{U}}^{k-1}|_{k-1} \hat{Q}_{k-1}) \otimes \text{Id}_{n_k}\|_2 \rightarrow 0. \tag{17}$$

Denoting the basis encoded by  $(\hat{\mathbf{U}}^{k-1}|_{k-1} \hat{Q}_{k-1})$  by  $\tilde{\theta}_{k-1}$ , the matrices in (17) encode the basis sets  $\text{vec}(\theta_{k-1} \otimes \psi_k)$  and  $\text{vec}(\tilde{\theta}_{k-1} \otimes \psi_k)$  with respect to  $\Psi^{:k}$ , see the third row in Fig. 7. We then find:

$$\begin{aligned} \|C(\text{vec}(\theta_{k-1} \otimes \psi_k)) - \hat{C}(\text{vec}(\tilde{\theta}_{k-1} \otimes \psi_k))\|_2 &= \|(\mathbf{U}^{k-1}|_{k-1} \otimes \text{Id}_{n_k})^T C(\Psi^{:k}) (\mathbf{U}^{k-1}|_{k-1} \otimes \text{Id}_{n_k}) \\ &\quad - ((\hat{\mathbf{U}}^{k-1}|_{k-1} \hat{Q}_{k-1}) \otimes \text{Id}_{n_k})^T \hat{C}(\Psi^{:k}) ((\hat{\mathbf{U}}^{k-1}|_{k-1} \hat{Q}_{k-1}) \otimes \text{Id}_{n_k})\|_2 \rightarrow 0 \quad \text{a.s.}, \end{aligned}$$

where convergence follows from (17) and since  $\hat{C}(\Psi^{:k}) \rightarrow C(\Psi^{:k})$ . Now, we note that the dominant  $r_k$ -dimensional eigenspace of  $C(\text{vec}(\theta_{k-1} \otimes \psi_k))$  is spanned by  $U_k$  (cf. Definition 3), while that of  $\hat{C}(\text{vec}(\tilde{\theta}_{k-1} \otimes \psi_k))$  is spanned by  $(\hat{Q}_{k-1} \otimes \text{Id}_{n_k})^T \hat{U}_k$ . By perturbation theory for singular vectors and Lemma 5, we have found that there is an orthonormal  $\hat{Q}_k \in \mathbb{R}^{r_k \times r_k}$  such that  $\|U_k - (\hat{Q}_{k-1} \otimes \text{Id}_{n_k})^T \hat{U}_k \hat{Q}_k\|_2 \rightarrow 0$ .

Hence

$$\begin{aligned} \widehat{\mathbf{U}}^k|_k \widehat{\mathbf{Q}}_k &= ((\widehat{\mathbf{U}}^{k-1}|_{k-1} \widehat{\mathbf{Q}}_{k-1}) \otimes \text{Id}_{n_k})(\widehat{\mathbf{Q}}_{k-1} \otimes \text{Id}_{n_k})^\top \widehat{\mathbf{U}}_k \widehat{\mathbf{Q}}_k \\ &\rightarrow (\mathbf{U}^{k-1} \otimes \text{Id}_{n_k}) \mathbf{U}_k = \mathbf{U}^k|_k. \end{aligned}$$

This proves that  $d(\mathbb{B}^k, \widehat{\mathbb{B}}^k) \rightarrow 0$ , and the conclusion about  $\mathbb{G}^k, \widehat{\mathbb{G}}^k$  follows from Lemma 4 (iii). The final statement about projectors can be shown in the same way as in Lemma 1.  $\square$

With this, we obtain the central result of this section, which is the convergence of Koopman operator representations on multi-linear subspaces:

**Theorem 1.** Let  $\mathcal{X}_t$  be a dynamical system such that (3) holds true. Let  $\mathbb{V} = \otimes_{k=1}^p \mathbb{V}^k \subset L^2_{\rho_0}$  and  $\mathbb{W} = \otimes_{k=1}^p \mathbb{W}^k \subset L^2_{\rho_1}$  be tensor product subspaces with bases  $\Psi = \otimes_{k=1}^p \psi_k$  and  $\Phi = \otimes_{k=1}^p \phi_k$ . For a fixed vector of TT ranks  $\mathbf{r}$ , denote the multi-linear spectral subspaces of both bases by  $\mathbb{G}^k(\Psi), \mathbb{G}^k(\Phi)$ , and assume they are all well-defined. Then we have for all  $k \leq p$ :

$$\|\widehat{\mathcal{K}}_\tau(\widehat{\mathbb{G}}^k(\Psi), \widehat{\mathbb{G}}^k(\Phi)) - \mathcal{K}_\tau(\mathbb{G}^k(\Psi), \mathbb{G}^k(\Phi))\|_{L(\mathbb{W}, \mathbb{V})} \rightarrow 0.$$

**Proof.** Since  $\widehat{\mathcal{P}}_{\widehat{\mathbb{G}}^k(\Psi)} \rightarrow \mathcal{P}_{\mathbb{G}^k(\Psi)}$  in  $L(\mathbb{V})$ , and  $\widehat{\mathcal{P}}_{\widehat{\mathbb{G}}^k(\Phi)} \rightarrow \mathcal{P}_{\mathbb{G}^k(\Phi)}$  in  $L(\mathbb{W})$  by Lemma 3, the argument is just the same as in the proof of Proposition 3.  $\square$

### 5.3. Multi-linear spectral subspaces and AMUSET

We complete this section by showing that the global SVD applied to the transformed data tensor (9) is the computational tool to compute bases of multi-linear spectral subspaces, analogous to the standard SVD applied to the data matrix. It will follow that AMUSET provides a representation of the Koopman operator on multi-linear spectral subspaces.

**Proposition 4.** Let  $\mathbb{V} = \otimes_{k=1}^p \mathbb{V}^k$  be a finite-dimensional tensor space of functions on  $\mathbb{R}^d$ , with basis  $\Psi = \otimes_{k=1}^p \psi_k$ . Let  $x_1, \dots, x_m \in \mathbb{R}^d$  be data points such that the empirical spectral subspaces  $\widehat{\mathbb{G}}^k$  are well-defined. For prescribed ranks  $\mathbf{r} = [r_1, \dots, r_p]$ , denote the global SVD of the transformed data tensor (9) by  $\Psi(X) = \widehat{\mathbf{U}}_{X,r} \widehat{\Sigma}_{X,r} \widehat{\mathbf{V}}_{X,r}^\top$ , with  $r = r_p$ . Then, the orthonormal part  $\widehat{\mathbf{U}}_{X,r}$  provides orthonormal coefficient tensors for each of the spaces  $\widehat{\mathbb{B}}^k$ , i.e.  $(\widehat{\theta}_k)^\top = (\Psi_{\cdot k|k})^\top \widehat{\mathbf{U}}_{X,r}^k|_k$  is a basis for  $\widehat{\mathbb{G}}^k$ . The remaining parts of the global SVD contain the time series of the basis for the final space  $\widehat{\mathbb{G}}^p$ , i.e.  $\widehat{\Sigma}_{X,r} \widehat{\mathbf{V}}_{X,r}^\top = \widehat{\theta}_p(X) = (\widehat{\mathbf{U}}_{X,r|p})^\top \Psi(X)|_p$ .

**Proof.** We prove the statement by showing that after  $k \leq p - 1$  iterations in Algorithm 3, the updated core  $\Psi^{(k+1)}(X)$  contains a time series of the product basis  $\text{vec}(\widehat{\theta}_k \otimes \psi_{k+1})(X)$ , where  $\widehat{\theta}_k$  is a basis of  $\widehat{\mathbb{G}}^k$  encoded by the partial tensor train  $\widehat{\mathbf{U}}_{X,r}^k$ . This is clearly true for  $k = 0$ , as  $\Psi^{(1)}(X) = \psi_1(X)$  and  $\widehat{\theta}_0 = 1$ , so we can again resort to an inductive argument. Assume that, after  $k - 1$  iterations,  $\Psi^{(k)}(X)$  is indeed of the form

$$\Psi^{(k)}(X)|_2 = \text{vec}(\widehat{\theta}_{k-1} \otimes \psi_k)(X),$$

where  $(\widehat{\theta}_{k-1})^\top = (\Psi_{\cdot k-1|k-1})^\top \widehat{\mathbf{U}}_{X,r}^{k-1}|_{k-1}$  is a basis of  $\widehat{\mathbb{G}}^{k-1}$ . It follows that the leading  $r_k$  left singular vectors  $U$  of  $\Psi^{(k)}(X)|_2$  (see line 2 of Algorithm 3) are also eigenvectors of the empirical Gramian  $\widehat{\mathbf{C}}(\text{vec}(\widehat{\theta}_{k-1} \otimes \psi_k))$ , as

$$\frac{1}{m} \Psi^{(k)}(X)|_2 (\Psi^{(k)}(X)|_2)^\top = \widehat{\mathbf{C}}(\text{vec}(\widehat{\theta}_{k-1} \otimes \psi_k)).$$

Hence, the basis set

$$(\widehat{\theta}_k)^\top = (\text{vec}(\widehat{\theta}_{k-1} \otimes \psi_k))^\top U = (\Psi_{\cdot k|k})^\top \widehat{\mathbf{U}}_{X,r}^k|_k$$

is, by definition, a basis of  $\widehat{\mathbb{G}}^k$ . It remains to show that the updated core  $\Psi^{(k+1)}(X)$  is of the required form. By inspecting the update formula in line 4, we see that

$$\begin{aligned} (\Sigma V^\top) \Psi^{(k+1)}(X)|_1 &= (U^\top \Psi^{(k)}(X)|_2) \Psi^{(k+1)}(X)|_1 \\ &= (U^\top (\text{vec}(\widehat{\theta}_{k-1} \otimes \psi_k)(X))) \Psi^{(k+1)}(X)|_1 \\ &= \widehat{\theta}_k(X) \Psi^{(k+1)}(X)|_1. \end{aligned}$$

The last expression is a matrix of shape  $r_k \times (n_{k+1}m)$ . Using the basis decomposition (9) for  $\Psi^{(k+1)}(X)$ , we determine its entries as

$$(\widehat{\theta}_k(X) \Psi^{(k+1)}(X)|_1)_{k; i_{k+1}, l} = \sum_{s=1}^m \widehat{\theta}_{k, l_k}(x_s) \psi_{k+1, i_{k+1}}(x_s) \delta_{s, l} = \widehat{\theta}_{k, l_k}(x_l) \psi_{k+1, i_{k+1}}(x_l),$$

which, upon re-shaping, equals  $\text{vec}(\widehat{\theta}_k \otimes \psi_{k+1})(X)$ , as claimed. Finally, the same arguments also show that  $(\widehat{\theta}_p)^\top = (\Psi|_p)^\top \widehat{\mathbf{U}}_{X,r|p}$  is a basis of  $\widehat{\mathbb{G}}^p$ , so  $\widehat{\mathbf{U}}_{X,r|p}$  is a basis of  $\mathbb{B}^p$ , and the time series  $\widehat{\theta}_p(X)$  is simply contained in the remaining components  $\widehat{\Sigma}_{X,r} \widehat{\mathbf{V}}_{X,r}^\top$  of the global SVD.  $\square$

We can now put the pieces together. In AMUSET, with fixed ranks  $\mathbf{r}$ , we separately apply global SVD to  $\Psi(X)$  and  $\Phi(Y)$ . We observe that, for  $r = r_p$ ,

$$(\widehat{\eta}_r)^\top = \sqrt{m} (\Psi|_p)^\top \widehat{\mathbf{U}}_{X,r|p} \widehat{\Sigma}_{X,r}^{-1}, \quad (\widehat{\zeta}_r)^\top = \sqrt{m} (\Phi|_p)^\top \widehat{\mathbf{U}}_{Y,r|p} \widehat{\Sigma}_{Y,r}^{-1}$$

are empirically orthonormal bases of the multi-linear spectral subspaces  $\widehat{\mathbb{G}}^p(\Psi)$  and  $\widehat{\mathbb{G}}^p(\Phi)$ , respectively. Hence, the empirical Koopman operator between these spaces possesses the matrix representation

$$\begin{aligned}\widehat{K}_\tau(\widehat{\eta}_r, \widehat{\zeta}_r) &= \widehat{A}(\widehat{\eta}_r, \widehat{\zeta}_r) = \frac{1}{m} \widehat{\eta}_r(X) \widehat{\zeta}_r(Y)^\top \\ &= \widehat{\Sigma}_{X,r}^{-1} (\widehat{\mathbf{U}}_{X,r|p})^\top (\Psi(X)|_p) (\Phi(Y)|_p)^\top \widehat{\mathbf{U}}_{Y,r|p} \widehat{\Sigma}_{Y,r}^{-1} = \widehat{V}_{X,r}^\top \widehat{V}_{Y,r} = \widehat{M}_{\tau,r},\end{aligned}$$

which is just the reduced matrix described in Section 3.1. In summary, application of AMUSet as outlined in Section 3.1 at fixed ranks provides a representation of the Koopman operator on the empirical multi-linear spectral subspaces  $\widehat{\mathbb{G}}^p(\Psi)$ ,  $\widehat{\mathbb{G}}^p(\Phi)$ , which consistently approximates the corresponding representation on  $\mathbb{G}^p(\Psi)$ ,  $\mathbb{G}^p(\Phi)$ , by Theorem 1.

## 6. Numerical examples

In this section, we provide numerical illustrations of the algorithmic and theoretical results presented in this study. All of the example systems have been analyzed using standard techniques before, which serve as reference results for our experiments. However, we will show that by means of our algorithms we are able to compute approximations to evolution operators on large trial spaces which are not amenable to standard treatment. The construction of these trial spaces requires a varying degree of preprocessing between examples. For the molecular examples, we still cannot do without prior knowledge, but the preprocessing pipeline is conceptually much simpler compared to standard methods.

Our algorithms have been implemented in Python 3.6 and collected in the toolbox Scikit-TT.<sup>2</sup> Furthermore, we used d3s<sup>3</sup>, PyEMMA<sup>4</sup> [50] as well as scikit-image<sup>5</sup> for simulating and analyzing the numerical examples.

### 6.1. Molecular dynamics

We re-analyze two data sets of equilibrium molecular dynamics simulations in explicit water. The first system is the ten residue peptide deca-alanine (see [34] for the simulation setup). After downsampling, this data set comprises  $m = 3 \cdot 10^5$  frames at time spacing of 10 ps. As a reference, we built a *Markov state model* (MSM) using 500 discrete states by following a typical protocol from the literature (linear dimension reduction by TICA, followed by k-means clustering in reduced space), see [51]. The second system is 39 residue protein NTL9. The data were produced by D. E. Shaw Research on the Anton Supercomputer [52]. The downsampled data set comprises approximately 56,000 frames at a time spacing of 50 ns. The MSM analysis presented in [53] serves as reference model.

As both systems are stationary and reversible, we have  $\rho_0 = \rho_1 = \mu$ , and we can use identical trial spaces  $\mathbb{V} = \mathbb{W}$ . Since the Koopman operator is self-adjoint in this case, all evolution operators essentially contain the same information, and singular pairs of  $\mathcal{K}_\tau$  are in fact eigenpairs. Following standard methodology in the field, we convert eigenvalue estimates  $\widehat{\lambda}_i$  into *implied timescales* (ITS) by the formula

$$\widehat{t}_i(\tau) = -\frac{\tau}{\log(\widehat{\lambda}_i)}. \quad (18)$$

Each ITS bears a unit of time and corresponds to the relaxation timescale of the dynamical process associated to eigenvalue  $\lambda_i$ . ITS are typically compared across a range of different lag times  $\tau$  (implied timescale test), as ITS estimates are known to improve with increasing  $\tau$ , and observing a plateau indicates convergence of the Koopman model [51,54,55]. Consequently, we will also use the implied timescale test to evaluate the performance of tensor-based Koopman models.

#### 6.1.1. Deca-alanine

For deca-alanine, we construct the trial space  $\mathbb{V}$  by first choosing  $p = 10$  backbone dihedral angles of the peptide as elementary descriptors, and then defining a subspace  $\mathbb{V}^k$  for each of them as the span of either  $n_k = 3$  or  $n_k = 4$  scalar functions on that dihedral angle. These functions always include the constant and two or three periodic Gaussians of the form

$$\psi_{k,i_k}(x) = \exp \left[ -\frac{1}{2s_{i_k}} \sin^2(0.5(x - c_{i_k})) \right].$$

Their positions and shapes are chosen to align with the typical marginal distribution of protein data along its backbone dihedral angles, see Figs. 8 (a) and (b). The full tensor space  $\mathbb{V}$  is then of dimension  $N = 3^5 \cdot 4^5 \approx 2.5 \cdot 10^5$ , which already exceeds what is typically considered feasible as trial space dimension for standard methods. The construction for  $\mathbb{V}^k$  used here is clearly informed by physical insight, but only at a fairly basic level.

As the system is in equilibrium, we can generate the data matrices  $X$ ,  $Y$  from just a single long trajectory, with  $Y$  being obtained by shifting all time steps in  $X$  by the lag time  $\tau$ . We then use the decomposition (9) and the HOCUR decomposition (Section 4.2) to arrive at TT representations of the transformed data tensor  $\Psi(Z)$ . Subsequently, the procedure outlined in Section 3.1 is applied to obtain of the reduced problem (11). Focusing on just the slowest dynamical process, we monitor the second ITS  $\widehat{t}_2(\tau)$  given by (18), as a function of  $\tau$ .

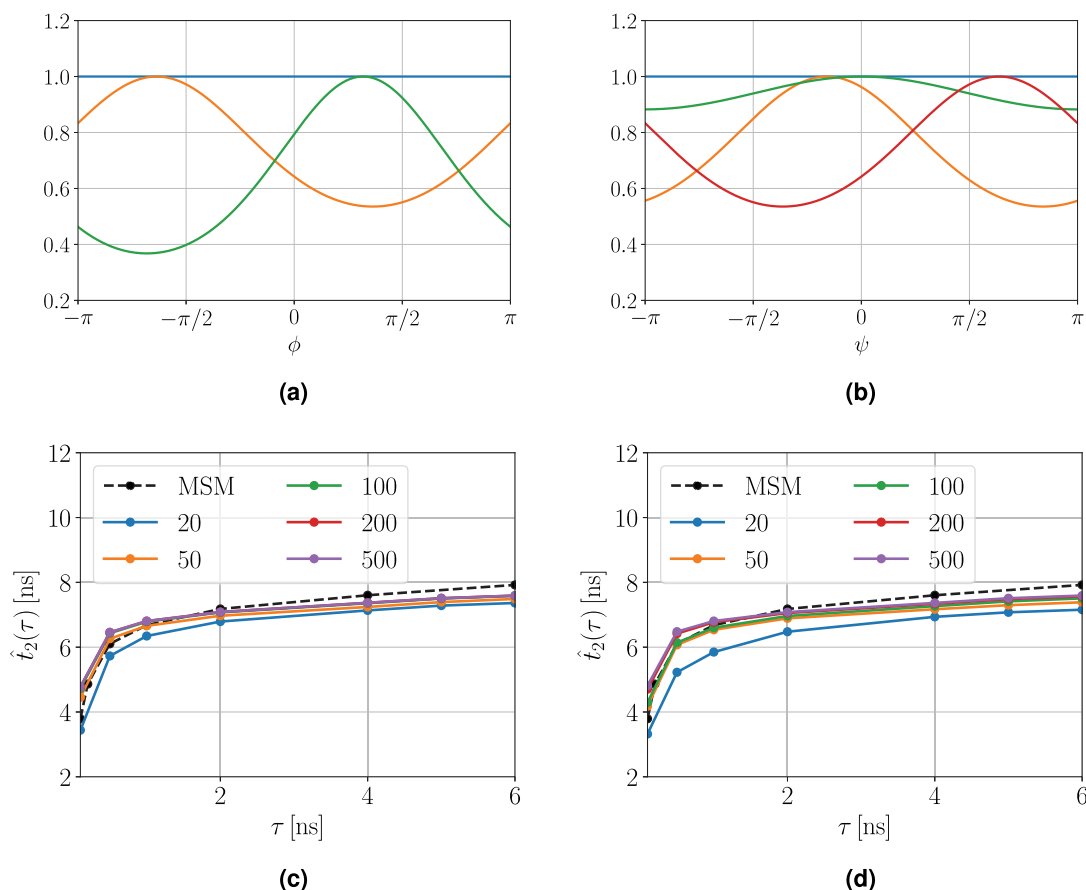
In line with the theoretical results presented in Section 5, we build the direct representation (9) based on varying amounts of data, ranging between  $m = 3000$  and  $m = 3 \cdot 10^5$  data points. However, we find the resulting timescale estimates to be virtually indistinguishable. The quality of approximation seems to depend more critically on the rank of the TT representation for  $\Psi(Z)$ . We analyze this dependence by capping the maximal rank  $r$  allowed either for the global SVD of (9), or for the HOCUR representation of  $\Psi(Z)$ , at different values. The resulting timescale estimates, as a function of the lag time  $\tau$ , are shown in Figs. 8(c)–(d). In both cases, a maximal rank of 50 is sufficient to obtain excellent agreement with the Markov model results.

<sup>2</sup> [https://github.com/PGelss/scikit\\_tt](https://github.com/PGelss/scikit_tt).

<sup>3</sup> <https://github.com/sklus/d3s>.

<sup>4</sup> <http://www.emma-project.org>

<sup>5</sup> <http://www.scikit-image.org/>.



**Fig. 8.** Results for molecular dynamics simulation data of deca-alanine peptide: (a) Univariate basis set used for all  $\phi$ -dihedral angles, comprised of the constant and periodic Gaussians centered at  $c_{i_k} = \{-2, 1\}$ , with  $s_{i_k} = \{0.8, 0.5\}$ . (b) The same for all  $\psi$ -dihedral angles, where periodic Gaussians are centered at  $c_{i_k} = \{-0.5, 0.0, 2.0\}$ , with  $s_{i_k} = \{0.8, 4.0, 0.8\}$ . (c) Slowest timescale  $\hat{t}_2$  obtained from (11) after constructing  $\Psi(Z)$  using the TT decomposition (9). We show results for different values of the maximal rank allowed during the global SVD, and of the lag time  $\tau$ . The reference MSM is represented by the black line. (d) The same if  $\Psi(Z)$  is represented by the HOCUR algorithm, for different values of the maximal rank in Algorithm 4.

### 6.1.2. NTL9

For NTL9, we also follow an established protocol to arrive at a basic set of descriptors. We consider all closest heavy-atom distances between protein residues, and rank these distances by the fraction of simulation time during which a contact between residues was formed (i.e. their distance is smaller than 0.35 nm). For each of these distance features, we define a space  $\mathbb{V}^k$  as the span of the constant and two Gaussian functions, given by

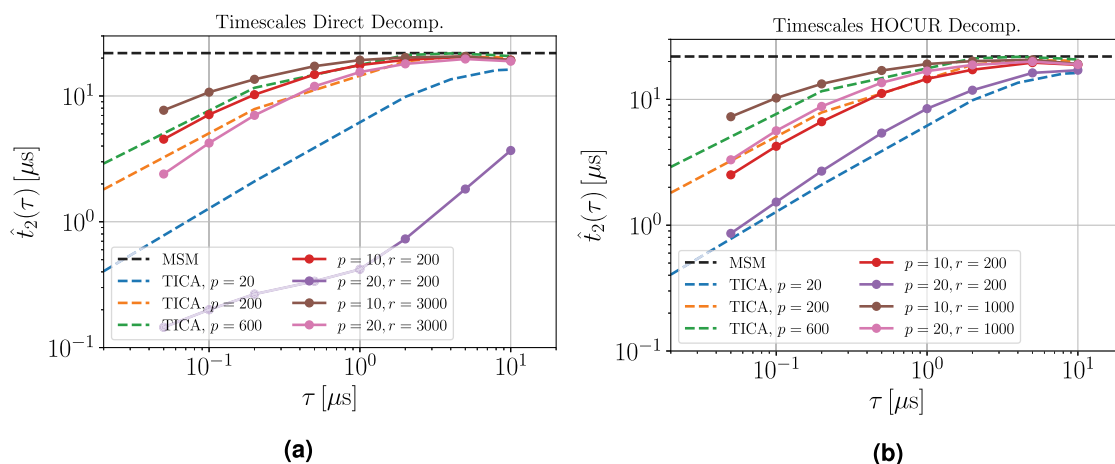
$$\psi_{k,i_k}(x) = \exp\left[-\frac{1}{2s_{i_k}}(x - c_{i_k})^2\right],$$

centered at  $c_{i_k} \in \{0.285, 0.62\}$ , with  $s_{i_k} \in \{0.001, 0.01\}$ . Again, these parameters were selected to make sure that the marginal distribution of the data along each distance can be reproduced by a linear combination of the Gaussians. Just as in the previous example, the construction of the elementary function spaces relies on some degree of preprocessing, but only at a basic level. Below, we use either the first  $p = 10$  or  $p = 20$  distance features to construct the full tensor space, which is therefore of dimension  $N = 3^{10} \approx 6 \cdot 10^4$  or  $N = 3^{20} \approx 3.5 \cdot 10^9$ . Both dimensions are beyond what is considered tractable for standard methods.

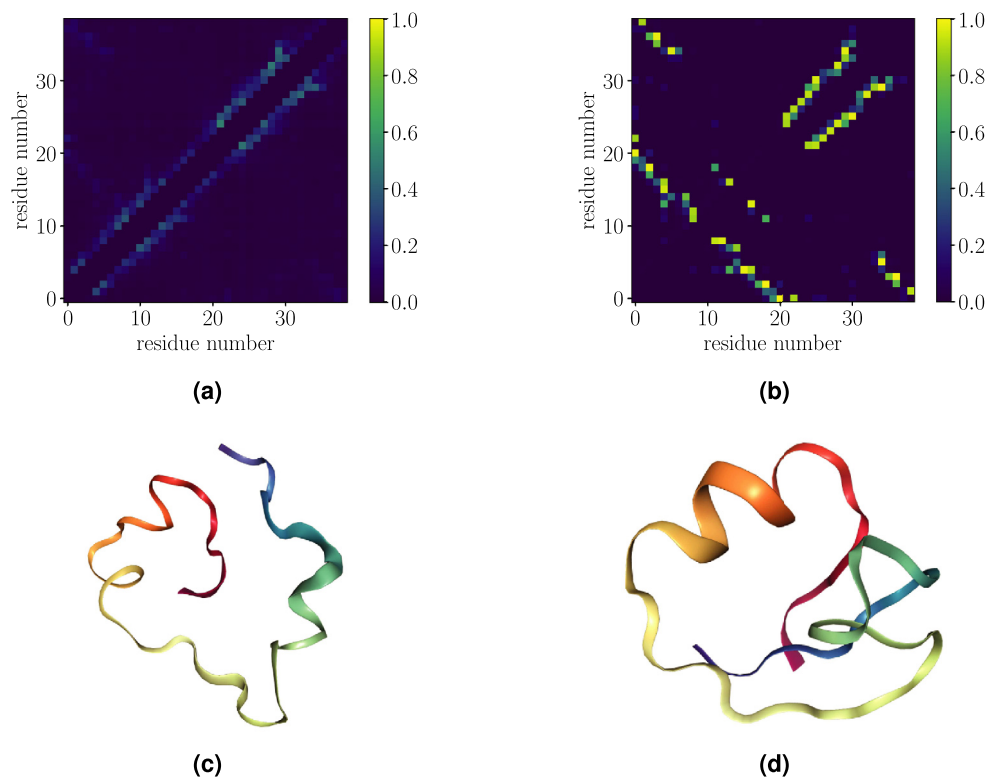
We follow essentially the same protocol as for deca-alanine, using both the direct decomposition (9) and the HOCUR iteration to represent the data tensor  $\Psi(Z)$ . For a series of lag times  $\tau$  and various maximal ranks  $r$ , we apply the procedure outlined in Section 3.1 to obtain the reduced matrix  $\hat{M}_{\tau,r}$  in (11). Estimates for the slowest implied timescale  $\hat{t}_2(\tau)$  are shown in panel (a) of Fig. 9 for the decomposition (9), and for the HOCUR representation in panel (b) of the same figure.

Two observations stand out: the first is, since our focus is to illustrate the advantage of using a large non-linear model class, a natural comparison for our model is the direct application of linear TICA [56], rather than the MSM analysis from [53], which required a significant level of expertise. Linear TICA is the same as applying the standard AMUSE Algorithm 2 to the basis set given by all elementary descriptors (i.e. the identity function on each distance feature in our case). Our results show that a tensor-based model on about ten distances, using a moderate TT rank, provides the same performance as linear TICA on several hundreds of distance features. The second observation is that including additional distance features ( $p = 20$ ) does not improve timescale estimates. Rather, larger ranks are required to ensure the same quality of approximation. The HOCUR-based models seem to be less sensitive to this effect for this example. An intuitive explanation would be that larger ranks are required to pass the relevant information down the chain of TT cores. We conclude that efficient and stable ways to compute low-rank representations of the data tensor  $\Psi(X)$  clearly remain an important topic for future work.





**Fig. 9.** Results for molecular dynamics simulation data of NTL9 protein: (a) Slowest timescale obtained from (11) after representing the data tensor  $\Psi(Z)$  using the exact decomposition (9). We show results for different maximal ranks during the global SVD of  $\Psi(Z)$ , and if either the first  $p = 10$  or  $p = 20$  distance features are used. The MSM-based reference value, as well as the timescales computed by linear TICA are indicated by the dashed lines. (b) The same if  $\Psi(Z)$  is represented by the HOCUR decomposition, with different maximal ranks in Algorithm 4.

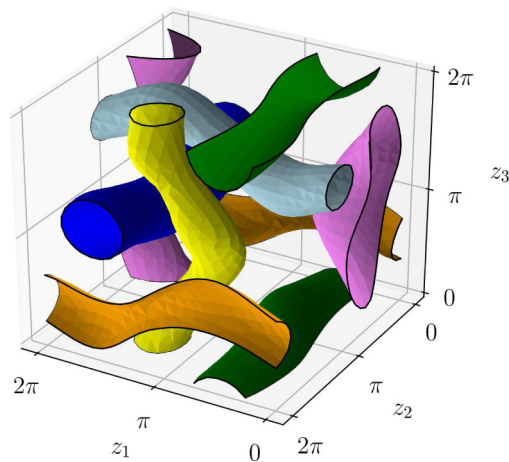


**Fig. 10.** States of NTL9: (a) Contact map for the unfolded state computed from the first two eigenfunctions of the HOCUR model corresponding to  $p = 10$ ,  $r = 200$ , and  $\tau = 2 \mu\text{s}$  (upper left triangle), compared to the corresponding contact map of the reference MSM (lower right triangle). (b) The same for the folded state. (c)/(d) Representative molecular structures for the unfolded and folded state of NTL9.

For completeness, we also verify that the eigenfunctions estimated by AMUSEt correctly encode the folding process of NTL9, which is known as the slowest dynamical process for this system. To this end, we apply PCCA [39] to the time series of the first two eigenfunctions and assign each snapshot to one of two metastable states if the degree of membership exceeds 0.5. We then calculate the contact frequencies separately for each of the two states. The resulting so-called contact maps are shown in the upper left triangles of Figs. 10 (a) and (b). By comparing to the contact maps provided by the reference Markov model (lower right triangles in Figs. 10 (a) and (b)), we see that there is virtually no difference.

## 6.2. ABC flow

Finally, we study a popular toy model for fluid dynamics problems, in order to illustrate the treatment of the non-stationary case and the computation of coherent sets using tensor-based methods. Let us consider the well-known ABC (Arnold–Beltrami–Childress)



**Fig. 11.** Results for the non-stationary ABC flow: six coherent vortices in the domain  $[0, 2\pi]^3$ , obtained from approximate eigenfunctions of the forward-backward operator  $\widehat{\mathcal{F}}_\tau$ , computed by AMUSET with  $10^3$  Gaussian basis functions.

flow, given by the ordinary differential equation

$$\begin{aligned}\dot{z}_1 &= A \sin(z_3) + C \cos(z_2), \\ \dot{z}_2 &= B \sin(z_1) + A \cos(z_3), \\ \dot{z}_3 &= C \sin(z_2) + B \cos(z_1),\end{aligned}$$

with  $A = \sqrt{3}$ ,  $B = \sqrt{2}$ , and  $C = 1$ . The system is defined on the torus, i.e.,  $0 \leq z_i \leq 2\pi$  for  $i = 1, 2, 3$ , see [57] for details. We define the lag time to be  $\tau = 5$  and sample  $25^3$  test points  $x_i$  uniformly in  $[0, 2\pi]^3$ . In order to compute the corresponding points  $y_i$ , we use a standard Runge-Kutta integrator with variable step size. Using a coordinate-major decomposition comprising ten Gaussian functions with variance 1 in each dimension, we arrive at a tensor space  $\mathbb{V} = \mathbb{W}$  spanned by 1000 three-dimensional Gaussians on an equidistant grid. Since  $\rho_0 \neq \rho_1$ , we are interested in eigenpairs of the projected forward-backward operator  $\widehat{\mathcal{F}}_\tau(\mathbb{V}, \mathbb{V})$ . We apply AMUSET as described in Section 3.1, and we use scikit-image in order to extract isosurfaces from the computed eigenfunctions. The coherent sets shown in Fig. 11 are consistent with the results presented in [40].

## 7. Summary

We have presented novel techniques to approximate evolution operators associated with high-dimensional dynamical systems using tensor-structured basis sets. Specifically, we have introduced AMUSET, a multi-linear version of the AMUSE algorithm, which allows us to derive a reduced matrix representation of evolution operators while operating only on the data tensor. For fixed multi-linear ranks, we have established convergence of AMUSET in the limit of infinite data. In addition, we have provided a detailed algorithmic description of a novel iterative method to compute a higher-order CUR decomposition of the data tensor, which only requires evaluations of the basis set on the data. We have also presented successful applications to benchmarking data sets of molecular dynamics simulation and fluid dynamics.

## CRediT authorship contribution statement

**Felix Nüske:** Conceptualization, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing - original draft, Writing - review & editing. **Patrick Gelfß:** Conceptualization, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing - original draft, Writing - review & editing. **Stefan Klus:** Conceptualization, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing - original draft, Writing - review & editing. **Cecilia Clementi:** Funding acquisition, Methodology, Resources, Supervision, Validation, Writing - original draft.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

This research has been funded by the National Science Foundation (CHE-1738990, CHE-1900374, PHY-2019745) [FN, CC], the Welch Foundation (C-1570) [FN, CC], the Rice University Academy of Fellows [FN], Deutsche Forschungsgemeinschaft (CRC 1114, “Scaling Cascades in Complex Systems”, [PG, SK]), (CRC/TRR 186/A12, “Molecular Switches: Spatio-temporal Control of Cellular Signal Transduction”, CRC 1078/C7, “Protonation Dynamics in Protein Function”, GRK 2433, [CC]), and the Einstein Foundation Berlin [CC]. The authors are grateful to D. E. Shaw Research for providing the NTL9 simulation data, and to the Paderborn Center for Parallel Computing for computational resources.

**Appendix. Auxiliary results**

First, we compile some useful results concerning linear operations in finite-dimensional Hilbert spaces:

**Lemma 4.** Let  $\mathbb{V}$  be a finite-dimensional Hilbert space with basis  $\psi = [\psi_1, \dots, \psi_n]^\top$  and Gramian matrix  $C(\psi)$ .

(i) Let  $\theta_1, \theta_2 \in \mathbb{V}$  with coefficient vectors  $a_1, a_2$  with respect to  $\psi$ . Also, let  $\mathcal{B} : \mathbb{V} \mapsto \mathbb{V}$  be a linear operator on  $\mathbb{V}$  with matrix representation  $B$  with respect to  $\psi$ . Then

$$\langle \theta_1, \theta_2 \rangle_{\mathbb{V}} = a_1^\top C(\psi) a_2, \quad \|\mathcal{B}\|_{L(\mathbb{V})} = \|C(\psi)^{1/2} B C(\psi)^{-1/2}\|_2.$$

(ii) Let  $\mathbb{F} \subset \mathbb{V}$  be an  $r$ -dimensional subspace with basis  $\theta$ . Let the matrix of coefficient vectors associated to  $\theta$  be  $A \in \mathbb{R}^{n \times r}$ . The matrix representation of the projector  $\mathcal{P}_{\mathbb{F}}$  with respect to  $\psi$  is

$$P_{\mathbb{F}} = A(A^\top C(\psi) A)^{-1} A^\top C(\psi).$$

(iii) Let  $\mathbb{F}_\nu, \nu = 1, 2, \dots$  and  $\mathbb{F}$  be subspaces of  $\mathbb{V}$  of dimension  $r \leq n$ , with coefficient vector spaces  $\mathbb{A}_\nu, \mathbb{A} \subset \mathbb{R}^n$ . Then

$$d(\mathbb{A}_\nu, \mathbb{A}) \rightarrow 0 \implies d(\mathbb{F}_\nu, \mathbb{F}) \rightarrow 0.$$

**Proof.** (i) and (ii) can be verified directly (see also [58]). To prove (iii), we fix an orthonormal basis  $U \in \mathbb{R}^{n \times r}$  of  $\mathbb{A}$ , and then use auxiliary Lemma 5 to select ONBs  $U_\nu$  of  $\mathbb{A}_\nu$  such that  $\|U_\nu - U\|_2 \rightarrow 0$  is also satisfied. The orthonormal projectors  $\mathcal{P}_{\mathbb{F}_\nu}, \mathcal{P}_{\mathbb{F}}$  then have matrix representations with respect to  $\psi$  given by

$$P_{\mathbb{F}_\nu} = U_\nu (U_\nu^\top C(\psi) U_\nu)^{-1} U_\nu^\top C(\psi), \quad P_{\mathbb{F}} = U (U^\top C(\psi) U)^{-1} U^\top C(\psi).$$

Using part (i), we then find

$$\begin{aligned} \|\mathcal{P}_{\mathbb{F}_\nu} - \mathcal{P}_{\mathbb{F}}\|_{L(\mathbb{V})} &= \|C(\psi)^{1/2} (U_\nu (U_\nu^\top C(\psi) U_\nu)^{-1} U_\nu^\top - U (U^\top C(\psi) U)^{-1} U^\top) C(\psi)^{1/2}\|_2 \\ &\leq \|C(\psi)\|_2 \|U_\nu (U_\nu^\top C(\psi) U_\nu)^{-1} U_\nu^\top - U (U^\top C(\psi) U)^{-1} U^\top\|_2 \rightarrow 0. \quad \square \end{aligned}$$

The following technical result helps us translate convergence of subspaces in Euclidean space into convergence of specific orthonormal bases:

**Lemma 5.** For  $\nu \in \mathbb{N}$ , let  $\mathbb{A}_\nu \subset \mathbb{R}^n$  be a subspace of dimension  $r \leq n$ , each with an orthonormal basis  $U_\nu \in \mathbb{R}^{n \times r}$ . Let  $\mathbb{A} \subset \mathbb{R}^n$  be another subspace of the same dimension, with orthonormal basis  $U \in \mathbb{R}^{n \times r}$ , and assume  $d(\mathbb{A}_\nu, \mathbb{A}) \rightarrow 0$  as  $\nu \rightarrow \infty$ . For each  $\nu$ , define  $\bar{U}_\nu \in \mathbb{R}^{n \times r}$  as the closest orthonormal matrix to  $U$  in the column span of  $U_\nu$ , i.e.,  $\bar{U}_\nu$  is obtained by solving the orthogonal Procrustes problem

$$\bar{U}_\nu = U_\nu R_\nu, \quad R_\nu = \underset{\substack{R \in \mathbb{R}^{r \times r}, \\ R^\top R = \text{Id}}}{\text{arg min}} \|U_\nu R - U\|_F^2.$$

Then we have  $\|U - \bar{U}_\nu\|_2 \rightarrow 0$ .

**Proof.** From an SVD of  $U_\nu^\top U$ , i.e.  $U_\nu^\top U = V_\nu \Sigma_\nu W_\nu^\top$ , the optimal orthogonal transformation can be obtained as  $R_\nu = V_\nu W_\nu^\top$  [59]. Moreover, from the above SVD we also obtain a compact singular value decomposition of  $U_\nu U_\nu^\top U U^\top$  via

$$U_\nu U_\nu^\top U U^\top = (U_\nu V_\nu) \Sigma_\nu (W_\nu^\top U^\top),$$

as the matrices in brackets are orthonormal and the rank of the left-hand side equals  $r$ . Now, by the convergence of the subspaces  $\mathbb{A}_\nu$  towards  $\mathbb{A}$ , we must have  $U_\nu U_\nu^\top U U^\top \rightarrow U U^\top$ , and as the latter is an orthogonal projector, we conclude  $\Sigma_\nu \rightarrow \text{Id}_r$  with  $\nu \rightarrow \infty$ . Now consider the difference between  $U$  and  $\bar{U}_\nu$ :

$$U - \bar{U}_\nu = U - U_\nu V_\nu W_\nu^\top = U - U_\nu V_\nu \Sigma_\nu^{-1} V_\nu^\top U_\nu^\top U = (\text{Id} - U_\nu V_\nu \Sigma_\nu^{-1} V_\nu^\top U_\nu^\top) U.$$

For each  $\nu$ , the columns of  $U_\nu V_\nu \in \mathbb{R}^{n \times r}$  are orthonormal. Let the columns of  $Y_\nu \in \mathbb{R}^{n \times n-r}$  form an orthonormal basis of the complement of  $\text{span}(U_\nu V_\nu)$ . Denote the orthonormal  $n \times n$ -matrices obtained by padding these two column sets by  $S_\nu = [U_\nu V_\nu \mid Y_\nu]$ . With this, we can further manipulate the above expression as follows:

$$\begin{aligned} U - \bar{U}_\nu &= [S_\nu S_\nu^\top - U_\nu V_\nu \Sigma_\nu^{-1} V_\nu^\top U_\nu^\top] U \\ &= \begin{bmatrix} [U_\nu V_\nu \mid Y_\nu] \begin{bmatrix} (V_\nu^\top U_\nu^\top) \\ Y_\nu^\top \end{bmatrix} - U_\nu V_\nu \Sigma_\nu^{-1} V_\nu^\top U_\nu^\top \end{bmatrix} U \\ &= (U_\nu V_\nu) (\text{Id} - \Sigma_\nu^{-1}) (V_\nu^\top U_\nu^\top) U + Y_\nu Y_\nu^\top U. \end{aligned}$$

As  $Y_\nu Y_\nu^\top$  is a matrix representation of the orthogonal projector onto  $\mathbb{A}_\nu^\perp$ , we must have  $Y_\nu Y_\nu^\top U \rightarrow 0$  by the convergence of the subspaces  $\mathbb{A}_\nu$ . For the first term, we find:

$$\begin{aligned} \|(U_\nu V_\nu) (\text{Id} - \Sigma_\nu^{-1}) (V_\nu^\top U_\nu^\top) U\|_2 &\leq \|(U_\nu V_\nu)\|_2 \|(\text{Id} - \Sigma_\nu^{-1})\|_2 \|(V_\nu^\top U_\nu^\top)\|_2 \|U\|_2 \\ &= \|(\text{Id} - \Sigma_\nu^{-1})\|_2 \rightarrow 0, \end{aligned}$$

by our observations above. This completes the proof.  $\square$

## References

- [1] B. Koopman, Hamiltonian Systems and transformation in Hilbert space, *Proc. Natl. Acad. Sci. U S A* 17 (5) (1931) 315, <http://dx.doi.org/10.1073/pnas.17.5.315>.
- [2] M. Dellnitz, O. Junge, On the approximation of complicated dynamical behavior, *SIAM J. Numer. Anal.* 36 (2) (1999) 491–515, <http://dx.doi.org/10.1137/S0036142996313002>.
- [3] C. Schütte, A. Fischer, W. Huisinga, P. Deuffhard, A direct approach to conformational dynamics based on hybrid Monte Carlo, *J. Comput. Phys.* 151 (1) (1999) 146–168, <http://dx.doi.org/10.1006/jcph.1999.6231>.
- [4] I. Mezić, Spectral properties of dynamical systems, model reduction and decompositions, *Nonlinear Dynam.* 41 (1) (2005) 309–325.
- [5] S. Klus, P. Koltai, C. Schütte, On the numerical approximation of the Perron–Frobenius and Koopman operator, *J. Comput. Dyn.* 3 (1) (2016) 51–79, <http://dx.doi.org/10.3934/jcd.2016003>.
- [6] G. Froyland, N. Santitissadeekorn, A. Monahan, Transport in time-dependent dynamical systems: Finite-time coherent sets, *Chaos* 20 (4) (2010) 043116, <http://dx.doi.org/10.1063/1.3502450>.
- [7] G. Froyland, An analytic framework for identifying finite-time coherent sets in time-dependent dynamical systems, *Physica D* 250 (2013) 1–19, <http://dx.doi.org/10.1016/j.physd.2013.01.013>.
- [8] M. Williams, I. Kevrekidis, C. Rowley, A data-driven approximation of the Koopman operator: Extending dynamic mode decomposition, *J. Nonlinear Sci.* 25 (6) (2015) 1307–1346, <http://dx.doi.org/10.1007/s00332-015-9258-5>.
- [9] F. Noé, F. Nüske, A variational approach to modeling slow processes in stochastic dynamical systems, *Multiscale Model. Simul.* 11 (2013) 635–655, <http://dx.doi.org/10.1137/110858616>.
- [10] F. Nüske, B. Keller, F. Pérez-Hernández, A. Mey, F. Noé, Variational approach to molecular kinetics, *J. Chem. Theory Comput.* 10 (2014) 1739–1752, <http://dx.doi.org/10.1021/ct4009156>.
- [11] H. Wu, F. Noé, Variational approach for learning Markov processes from time series data, *J. Nonlinear Sci.* 30 (1) (2020) 23–66, <http://dx.doi.org/10.1007/s00332-019-09567-y>.
- [12] D. Hardoon, S. Szedmak, J. Shawe-Taylor, Canonical correlation analysis: An overview with application to learning methods, *Neural Comput.* 16 (12) (2004) 2639–2664, <http://dx.doi.org/10.1162/0899766042321814>.
- [13] S. Klus, B. Husic, M. Mollenhauer, F. Noé, Kernel methods for detecting coherent structures in dynamical data, *Chaos* (2019) <http://dx.doi.org/10.1063/1.5100267>.
- [14] S. Klus, F. Nüske, P. Koltai, H. Wu, I. Kevrekidis, C. Schütte, F. Noé, Data-driven model reduction and transfer operator approximation, *J. Nonlinear Sci.* 28 (3) (2018) 985–1010, <http://dx.doi.org/10.1007/s00332-017-9437-7>.
- [15] M. Williams, C. Rowley, I. Kevrekidis, A kernel-based method for data-driven Koopman spectral analysis, *J. Comput. Dyn.* 2 (2) (2015) 247–265, <http://dx.doi.org/10.3934/jcd.2015005>.
- [16] S. Klus, F. Nüske, B. Hamzi, Kernel-based approximation of the Koopman generator and Schrödinger operator, *Entropy* 22 (7) (2020) 722, <http://dx.doi.org/10.3390/e22070722>.
- [17] S. Klus, I. Schuster, K. Muandet, Eigendecompositions of transfer operators in reproducing kernel Hilbert spaces, *J. Nonlinear Sci.* 30 (1) (2020) 283–315, <http://dx.doi.org/10.1007/s00332-019-09574-z>.
- [18] B. Lusch, J. Kutz, S. Brunton, Deep learning for universal linear embeddings of nonlinear dynamics, *Nature Commun.* 9 (1) (2018) 1–10, <http://dx.doi.org/10.1038/s41467-018-07210-0>.
- [19] A. Mardt, L. Pasquali, H. Wu, F. Noé, VAMPnets for deep learning of molecular kinetics, *Nature Commun.* 9 (1) (2018) 5, <http://dx.doi.org/10.1038/s41467-017-02388-1>.
- [20] J. Carroll, J. Chang, Analysis of individual differences in multidimensional scaling via an N-way generalization of ‘Eckart-Young’ decomposition, *Psychometrika* 35 (3) (1970) 283–319, <http://dx.doi.org/10.1007/BF02310791>.
- [21] L. Tucker, The extension of factor analysis to three-dimensional matrices, in: H. Gulliksen, N. Frederiksen (Eds.), *Contributions To Mathematical Psychology*, Holt, Rinehart and Winston, 1964, pp. 110–127.
- [22] W. Hackbusch, S. Kühn, A new scheme for the tensor representation, *J. Fourier Anal. Appl.* 15 (5) (2009) 706–722, <http://dx.doi.org/10.1007/s00041-009-9094-9>.
- [23] I. Oseledets, E. Tyrtshnikov, Breaking the curse of dimensionality, or how to use SVD in many dimensions, *SIAM J. Sci. Comput.* 31 (5) (2009) 3744–3759, <http://dx.doi.org/10.1137/090748330>.
- [24] I. Oseledets, Tensor-train decomposition, *SIAM J. Sci. Comput.* 33 (2011) 2295–2317, <http://dx.doi.org/10.1137/090752286>.
- [25] M. Beck, A. Jäckle, G. Worth, H. Meyer, The multiconfiguration time-dependent hartree (MCTDH) method: a highly efficient algorithm for propagating wavepackets, *Phys. Rep.* 324 (2000) 1–105, [http://dx.doi.org/10.1016/S0370-1573\(99\)00047-2](http://dx.doi.org/10.1016/S0370-1573(99)00047-2).
- [26] S. Dolgov, B. Khoromskij, Simultaneous state-time approximation of the chemical master equation using tensor product formats, *Numer. Linear Algebra Appl.* 22 (2) (2015) 197–219, <http://dx.doi.org/10.1002/nla.1942>.
- [27] P. Gelß, S. Matera, C. Schütte, Solving the master equation without kinetic Monte Carlo, *J. Comput. Phys.* 314 (C) (2016) 489–502, <http://dx.doi.org/10.1016/j.jcp.2016.03.025>.
- [28] Z. Zhang, K. Batselier, H. Liu, L. Daniel, N. Wong, Tensor computation: A new framework for high-dimensional problems in EDA, *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.* 36 (4) (2017) 521–536, <http://dx.doi.org/10.1109/TCAD.2016.2618879>.
- [29] P. Gelß, S. Klus, J. Eisert, C. Schütte, Multidimensional approximation of nonlinear dynamical systems, *J. Comput. Nonlinear Dyn.* 14 (6) (2019) 061006, <http://dx.doi.org/10.1115/1.4043148>.
- [30] I. Affleck, T. Kennedy, E. Lieb, H. Tasaki, Rigorous results on valence-bond ground states in antiferromagnets, *Phys. Rev. Lett.* 59 (7) (1987) 799–802, <http://dx.doi.org/10.1103/PhysRevLett.59.799>.
- [31] S. Östlund, S. Rommer, Thermodynamic limit of density matrix renormalization, *Phys. Rev. Lett.* 75 (19) (1995) 3537, <http://dx.doi.org/10.1103/PhysRevLett.75.3537>.
- [32] S. Szalay, M. Pfeffer, V. Murg, G. Barcza, F. Verstraete, R. Schneider, O. Legeza, Tensor product methods and entanglement optimization for ab initio quantum chemistry, *Int. J. Quantum Chem.* 115 (19) (2015) 1342–1391, <http://dx.doi.org/10.1002/qua.24898>.
- [33] S. Klus, C. Schütte, Towards tensor-based methods for the numerical approximation of the Perron–Frobenius and Koopman operator, *J. Comput. Dyn.* 3 (2) (2016) 139–161, <http://dx.doi.org/10.3934/jcd.2016007>.
- [34] F. Nüske, R. Schneider, F. Vitalini, F. Noé, Variational tensor approach for approximating the rare-event kinetics of macromolecular systems, *J. Chem. Phys.* 144 (5) (2016) 054105, <http://dx.doi.org/10.1063/1.4940774>.
- [35] S. Klus, P. Gelß, S. Peitz, C. Schütte, Tensor-based dynamic mode decomposition, *Nonlinearity* 31 (7) (2018) <http://dx.doi.org/10.1088/1361-6544/aac8f>.
- [36] F. Litzinger, L. Boninsegni, H. Wu, F. Nüske, R. Patel, R. Baraniuk, F. Noé, C. Clementi, Rapid calculation of molecular kinetics using compressed sensing, *J. Chem. Theory Comput.* 14 (5) (2018) 2771–2783, <http://dx.doi.org/10.1021/acs.jctc.8b00089>.
- [37] L. Tong, V. Soon, Y. Huang, R. Liu, AMUSE: a new blind identification algorithm, in: *IEEE International Symposium on Circuits and Systems*, IEEE, 1990, pp. 1784–1787, <http://dx.doi.org/10.1109/ISCAS.1990.111981>.
- [38] I. Oseledets, E. Tyrtshnikov, TT-Cross approximation for multidimensional arrays, *Linear Algebra Appl.* 432 (1) (2010) 70–88, <http://dx.doi.org/10.1016/j.laa.2009.07.024>.
- [39] P. Deuffhard, M. Weber, Robust Perron cluster analysis in conformation dynamics, *Linear Algebra Appl.* 398 (2005) 161–184, <http://dx.doi.org/10.1016/j.laa.2004.10.026>.
- [40] R. Banisch, P. Koltai, Understanding the geometry of transport: Diffusion maps for Lagrangian trajectory data unravel coherent sets, *Chaos* 27 (3) (2017) 035804, <http://dx.doi.org/10.1063/1.4971788>.
- [41] R. Chacon, An ergodic theorem for operators satisfying norm conditions, *J. Appl. Math. Mech.* 11 (1) (1962) 165–172.
- [42] R. Penrose, Applications of negative dimensional tensors, in: *Combinatorial Mathematics and Its Applications*, Academic Press London, 1971, pp. 221–244.
- [43] P. Gelß, S. Klus, S. Matera, C. Schütte, Nearest-neighbor interaction systems in the tensor-train format, *J. Comput. Phys.* 341 (2017) 140–162, <http://dx.doi.org/10.1016/j.jcp.2017.04.007>.

- [44] S. Dolgov, B. Khoromskij, I. Oseledets, D. Savostyanov, Computation of extreme eigenvalues in higher dimensions using block tensor train format, *Comput. Phys. Comm.* 185 (4) (2014) 1207–1216, <http://dx.doi.org/10.1016/j.cpc.2013.12.017>.
- [45] S. Goreinov, I. Oseledets, D. Savostyanov, E. Tyrtyshnikov, N. Zamarashkin, How to find a good submatrix, in: *Matrix Methods: Theory, Algorithms and Applications: Dedicated To the Memory of Gene Golub*, World Scientific, 2010, pp. 247–256, [http://dx.doi.org/10.1142/9789812836021\\_0015](http://dx.doi.org/10.1142/9789812836021_0015).
- [46] S. Goreinov, E. Tyrtyshnikov, The maximal-volume concept in approximation by low-rank matrices, in: *Structured Matrices in Mathematics, Computer Science, and Engineering I*, American Mathematical Society, 2001, pp. 47–51, <http://dx.doi.org/10.1090/conm/280/4620>.
- [47] C. Boutsidis, D. Woodruff, Optimal CUR matrix decompositions, in: *Proceedings of the Forty-Sixth Annual ACM Symposium on Theory of Computing*, in: STOC '14, ACM, 2014, pp. 353–362, <http://dx.doi.org/10.1145/2591796.2591819>.
- [48] M. Korda, I. Mezić, On convergence of extended dynamic mode decomposition to the Koopman operator, *J. Nonlinear Sci.* 28 (2) (2018) 687–710, <http://dx.doi.org/10.1007/s00332-017-9423-0>.
- [49] G. Stewart, J. Sun, *Matrix Perturbation Theory*, Academic Press, 1990.
- [50] M. Scherer, B. Trendelkamp-Schroer, F. Paul, G. Pérez-Hernández, M. Hoffmann, N. Plattner, C. Wehmeyer, J.-H. Prinz, F. Noé, PyEMMA 2: A software package for estimation, validation, and analysis of Markov models, *J. Chem. Theory Comput.* 11 (11) (2015) 5525–5542, <http://dx.doi.org/10.1021/acs.jctc.5b00743>.
- [51] G.R. Bowman, V.S. Pande, F. Noé (Eds.), *An introduction to Markov state models and their application to long timescale molecular simulation*, Springer Netherlands, 2014, <http://dx.doi.org/10.1007/978-94-007-7606-7>.
- [52] K. Lindorff-Larsen, S. Piana, R. Dror, D. Shaw, How fast-folding proteins fold, *Science* 334 (6055) (2011) 517–520, <http://dx.doi.org/10.1126/science.1208351>.
- [53] L. Boninsegna, R. Banisch, C. Clementi, A data-driven perspective on the hierarchical assembly of molecular structures, *J. Chem. Theory Comput.* 14 (1) (2017) 453–460, <http://dx.doi.org/10.1021/acs.jctc.7b00990>.
- [54] M. Sarich, F. Noé, C. Schütte, On the approximation quality of Markov state models, *Multiscale Model. Simul.* 8 (4) (2010) 1154–1177, <http://dx.doi.org/10.1137/090764049>.
- [55] J.-H. Prinz, H. Wu, M. Sarich, B. Keller, M. Senne, M. Held, J. Chodera, C. Schütte, F. Noé, Markov models of molecular kinetics: Generation and validation, *J. Chem. Phys.* 134 (17) (2011) 174105, <http://dx.doi.org/10.1063/1.3565032>.
- [56] G. Pérez-Hernández, F. Paul, T. Giorgino, G. De Fabritiis, F. Noé, Identification of slow molecular order parameters for Markov model construction, *J. Chem. Phys.* 139 (1) (2013) 15102, <http://dx.doi.org/10.1063/1.4811489>.
- [57] G. Froyland, K. Padberg, Almost-invariant sets and invariant manifolds – Connecting probabilistic and geometric descriptions of coherent structures in flows, *Physica D* 238 (2009) 1507–1523, <http://dx.doi.org/10.1016/j.physd.2009.03.002>.
- [58] A. Knyazev, M. Argentati, Principal angles between subspaces in an A-based scalar product: algorithms and perturbation estimates, *SIAM J. Sci. Comput.* 23 (6) (2002) 2008–2040, <http://dx.doi.org/10.1137/S1064827500377332>.
- [59] P. Schönemann, A generalized solution of the orthogonal procrustes problem, *Psychometrika* 31 (1) (1966) 1–10, <http://dx.doi.org/10.1007/BF02289451>.