

# Lawrence Berkeley National Laboratory

## LBL Publications

### Title

Collocation methods for nonlinear differential equations on low-rank manifolds

### Permalink

<https://escholarship.org/uc/item/1sz9921z>

### Author

Dektor, Alec

### Publication Date

2025

### DOI

10.1016/j.laa.2024.11.001

### Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at

<https://creativecommons.org/licenses/by/4.0/>

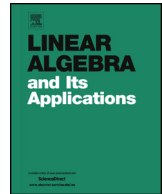
Peer reviewed



Contents lists available at ScienceDirect

# Linear Algebra and its Applications

journal homepage: [www.elsevier.com/locate/laa](http://www.elsevier.com/locate/laa)



## Collocation methods for nonlinear differential equations on low-rank manifolds



Alec Dektor

*Applied Mathematics & Computational Research Division, Lawrence Berkeley National Laboratory, Berkeley (CA) 94720, USA*

### ARTICLE INFO

#### *Article history:*

Received 7 June 2024

Received in revised form 30 October 2024

Accepted 1 November 2024

Available online 6 November 2024

Submitted by D. Kressner

#### *MSC:*

15A69

65F99

65L05

90C06

#### *Keywords:*

Low-rank approximation

Time-dependent tensors

Tensor differential equations

Tensor cross approximation

Tensor train format

### ABSTRACT

We introduce new methods for integrating nonlinear differential equations on low-rank manifolds. These methods rely on interpolatory projections onto the tangent space, enabling low-rank time integration of vector fields that can be evaluated entry-wise. A key advantage of our approach is that it does not require the vector field to exhibit low-rank structure, thereby overcoming significant limitations of traditional dynamical low-rank methods based on orthogonal projection. To construct the interpolatory projectors, we develop a sparse tensor sampling algorithm based on the discrete empirical interpolation method (DEIM) that parameterizes tensor train manifolds and their tangent spaces with cross interpolation. Using these projectors, we propose two time integration schemes on low-rank tensor train manifolds. The first scheme integrates the solution at selected interpolation indices and constructs the solution with cross interpolation. The second scheme generalizes the well-known orthogonal projector-splitting integrator to interpolatory projectors. We demonstrate the proposed methods with applications to several tensor differential equations arising from the discretization of partial differential equations.

© 2024 The Author. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

*E-mail address:* [adektor@lbl.gov](mailto:adektor@lbl.gov).

<https://doi.org/10.1016/j.laa.2024.11.001>

0024-3795/© 2024 The Author. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Consider the initial value problem

$$\frac{\partial u(\mathbf{x}, t)}{\partial t} = \mathcal{G}(u, \mathbf{x}, t), \quad u(\mathbf{x}, 0) = u_0(\mathbf{x}), \quad (1)$$

governing the time evolution of a quantity of interest  $u : \Omega \times [0, T] \rightarrow \mathbb{R}$ , where  $\Omega$  is a subset of  $\mathbb{R}^d$  ( $d \gg 1$ ) and  $\mathcal{G}$  is a nonlinear operator that may depend on  $\mathbf{x}$  and  $t$ . Equations of the form (1) are found in many areas of physical sciences, engineering, and mathematics. For example, in applications of kinetic theory such as the Fokker–Planck equation [41] and the Boltzmann equation [4], in optimal mass transport [19], and as finite-dimensional approximations of functional differential equations [53,52]. Discretizing (1) with a method of lines yields the tensor differential equation

$$\frac{dX(t)}{dt} = G(X(t), t), \quad X(0) = X_0, \quad (2)$$

where  $X(t) : [0, T] \rightarrow \mathbb{R}^{n_1 \times \dots \times n_d}$  is the time-dependent solution tensor and  $G : \mathbb{R}^{n_1 \times \dots \times n_d} \times [0, T] \rightarrow \mathbb{R}^{n_1 \times \dots \times n_d}$  is a discrete form of the operator  $\mathcal{G}$ . At any time  $t$ , the solution tensor  $X(t)$  has  $\mathcal{O}(n^d)$  degrees of freedom that make its computation and storage prohibitively expensive, even for small  $d$ .

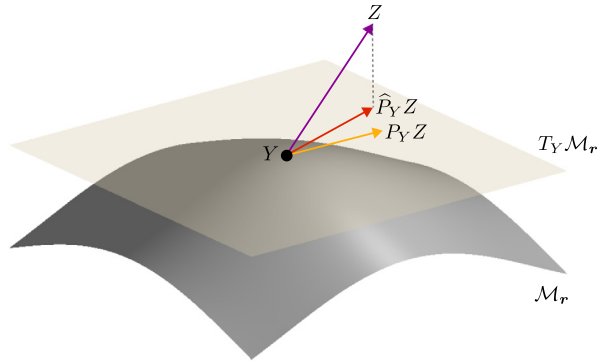
Several algorithms based on tensor networks have recently been proposed to reduce the number of degrees of freedom in the solution tensor  $X(t)$  and integrate (2) at a reasonable computational cost. A tensor network is a factorization of a high-dimensional tensor, such as  $X(t)$ , into a network of low-dimensional tensors with significantly fewer degrees of freedom. The number of degrees of freedom in a network depends on the chosen tensor format, e.g., tensor train (TT) [38], Tucker [34,13], Hierarchical Tucker [22,33,23] or canonical polyadic (CP) [29], and the tensor rank. For instance, tensors in the TT format with rank  $r$  can be parameterized with  $\mathcal{O}(dnr^2)$  degrees of freedom, a significant reduction from  $\mathcal{O}(n^d)$  when the rank  $r$  is sufficiently small. The set of all tensors in a chosen format with fixed rank forms a smooth manifold on which the solution to (2) can be integrated.

Two classes of algorithms for integrating (2) on smooth tensor manifolds are step-truncation methods [42] and dynamical low-rank methods [10]. Step-truncation methods allow the solution rank to naturally increase in a controlled manner during a time step before truncating back to the desired rank. Dynamical low-rank methods integrate the solution on a fixed-rank manifold by projecting  $G(X, t)$  onto a tangent space of the manifold at each time  $t$ . Both methods aim to efficiently compute the best approximate solution to (2) on a fixed-rank tensor manifold at each time  $t$  and are consistent with each other as the temporal step-size approaches zero (see, e.g., [9, Section 3.3]). These methods have been utilized for several applications including uncertainty quantification [2,45], plasma physics [54,17], numerical approximation of functional differential equations [53,44] and

machine learning [48,46], and substantial research efforts have recently been made to improve their accuracy, efficiency, and robustness. These efforts have resulted in several innovations including rank-adaptive integrators [9,5,42], implicit low-rank methods [43,36,50], conservative low-rank methods [25,3,18,16] and coordinate-adaptive low-rank methods [11,12].

Despite these recent advancements, existing low-rank time integration schemes still have significant limitations in their applicability. Most notably, they require the tensor-valued map  $G$  defining the differential equation (2) to have low-rank structure complementary to that of the solution. Such low-rank structure is key to increasing the solution rank in a controlled manner for step-truncation methods or efficiently computing the orthogonal projection of  $G(X, t)$  onto the tangent space for dynamical low-rank methods. In either case, the low-rank structure of  $G$  is crucial for obtaining practical time integration schemes with computational cost and storage requirements comparable to the storage cost of the chosen tensor format. However, many instances of  $G$  that arise from discretizing (1) lack low-rank structure. For example, when  $G$  includes a polynomial nonlinearity computing a low-rank representation of  $G(X, t)$  or its orthogonal projection onto the tangent space is expensive due to the non-optimal rank that results from multiplying low-rank tensors. The situation is worse for other common nonlinearities, such as exponential or fractional, as there are currently no reliable algorithms for performing these nonlinear arithmetic operations with low-rank tensors. In such cases, (2) may admit an approximate low-rank solution. However, existing step-truncation and dynamical low-rank methods cannot efficiently compute it.

In this paper, we introduce a new class of dynamical low-rank methods that can efficiently integrate the solution to (2) on a low-rank tensor manifold even when  $G$  does not have low-rank structure. Our proposed methods rely on a new class of oblique projectors onto low-rank tangent spaces with a cross interpolation property. In the context of dynamical low-rank approximation, these projectors collocate (2) on a tensor manifold and yield equations of motion on the manifold that are efficient to integrate whenever it is possible to evaluate  $G$  entry-wise. The oblique projectors are defined by sets of multi-indices that identify tensor fibers along which the projector interpolates. To select these multi-indices, we introduce a new algorithm, based on the discrete empirical interpolation method (DEIM), to efficiently compute indices that parameterize tensor manifolds with cross interpolation [37]. Using these projectors we propose two low-rank time integration schemes. The first integrates subtensors of the solution defining a tensor cross interpolant and then constructs the low-rank solution later in time with tensor cross interpolation. The second we obtain by applying a splitting scheme to the oblique tangent space projector. Splitting schemes for orthogonal tangent space projectors were introduced in [31] for matrices and were subsequently generalized to TTs [32], Tucker tensors [30] and tree tensor networks [7]. Our method directly generalizes these orthogonal projector-splitting schemes to oblique projectors in the TT format. Related time integration schemes based on oblique projections onto the low-rank manifold (instead



**Fig. 1.** A sketch of the low-rank manifold  $\mathcal{M}_r$  and its tangent space  $T_Y \mathcal{M}_r$  at the point  $Y \in \mathcal{M}_r$ . Also depicted is  $Z \in \mathbb{R}^{n_1 \times \dots \times n_d}$  and its orthogonal projection  $\hat{P}_Y Z$  and oblique projection  $P_Y Z$  onto the tangent space  $T_Y \mathcal{M}_r$ . The orthogonal projection is the best approximation of  $Z$  on the tangent space with respect to the Frobenius norm but is impractical compute when  $G$  lacks rank structure. The oblique projection is a quasi-optimal approximation of  $Z$  on the tangent space that is efficient to compute for any  $G$  that can be evaluated entry-wise. Such oblique projectors allow us to efficiently apply dynamical low-rank methods to a broad class of nonlinear differential equations.

of its tangent space) were recently proposed for computing low-rank approximations to matrix differential equations [15,35] and tensor differential equations [21,20].

The rest of this paper is organized as follows. In Section 2 we introduce interpolatory tangent space projectors and the proposed dynamical low-rank methods on matrix manifolds ( $d = 2$ ). In Section 3 we briefly recall the TT format and orthogonalization of tensors in the TT format. In Section 4 we recall the orthogonal projector onto the TT tangent space and introduce new oblique projectors onto the tangent space. Then we describe a special class of oblique projectors with a cross interpolation property. In Section 5 we present a new index selection algorithm, referred to as TT-cross-DEIM, for constructing oblique projectors onto the TT tangent space. We show that indices obtained with the TT-cross-DEIM algorithm define oblique tangent space projectors and parameterize TT manifolds with cross interpolation. In Section 6 we introduce new dynamical low-rank time integration schemes for (2) using oblique tangent space projectors. In Section 7 we demonstrate the proposed dynamical low-rank methods and compare the results with existing time integration methods on low-rank tensor manifolds. The main findings are summarized in Section 8.

## 2. Dynamical low-rank matrix approximation

Before introducing dynamical low-rank tensor approximation, we describe the proposed low-rank methods for matrices ( $d = 2$ ). The goal is to find an approximate solution  $Y(t)$  to (2) that lies on the manifold  $\mathcal{M}_r$  of rank- $r$  matrices for all time  $t$ . A rank- $r$  matrix can be expressed through left and right factor matrices with dimensions  $n_1 r$  and  $n_2 r$ , respectively. Thus approximating the solution to (2) with  $Y(t) \in \mathcal{M}_r$  reduces the storage cost from  $n_1 n_2$  to  $(n_1 + n_2)r$ . Dynamical low-rank methods integrate the approximate solution  $Y(t)$  on the manifold  $\mathcal{M}_r$  by projecting (2) onto a tangent space of  $\mathcal{M}_r$

at each time  $t$ . Assuming the initial condition  $X_0$  belongs to  $\mathcal{M}_r$ , this procedure yields the evolution equation

$$\frac{dY(t)}{dt} = P_{Y(t)}G(Y(t), t), \quad Y(0) = X_0, \tag{3}$$

where  $P_{Y(t)} : \mathbb{R}^{n_1 \times n_2} \rightarrow T_{Y(t)}\mathcal{M}_r$  projects onto the tangent space of  $\mathcal{M}_r$  at  $Y(t)$ . The solution to (3) remains on  $\mathcal{M}_r$  for all  $t \geq 0$  and serves as an approximate solution to (2). Classical dynamical low-rank methods use an orthogonal tangent space projector to minimize the error of the approximation in the Frobenius norm at each time  $t$ . Let  $U(t)$  and  $V(t)$  be matrices whose columns form orthonormal bases for the range and co-range of  $Y(t)$ , respectively. Such matrices can be obtained, for example, from the SVD of  $Y(t)$ . Then the orthogonal tangent space projector can be expressed as [31]

$$\widehat{P}_Y Z = ZVV^T - UU^T ZVV^T + UU^T Z, \tag{4}$$

where  $Z \in \mathbb{R}^{n_1 \times n_2}$  and we suppressed dependence on  $t$  for simplicity. When  $Z = G(Y(t), t)$  computing the projection (5) at each time step to integrate (3) can be computationally expensive, especially if  $G$  does not have low-rank structure. For most nonlinear functions  $G$ , computing such orthogonal projection has computational cost scaling as  $n_1 n_2$ , making the integration of the approximate solution  $Y(t) \in \mathcal{M}_r$  as expensive as solving for the full solution  $X(t)$  using standard methods.

### 2.1. Interpolatory dynamical low-rank approximation

---

**Algorithm 1** DEIM index selection (adapted from [49]).

---

**Require:**  $V \in \mathbb{R}^{n \times r}$  with  $n \geq r$   
**Ensure:**  $l$ , a vector with  $r$  distinct indices from  $\{1, \dots, n\}$

- 1:  $v = V(:, 1)$
- 2:  $[l, l_1] = \max(|v|)$
- 3: **for**  $j = 2, 3, \dots, r$  **do**
- 4:      $v = V(:, j)$
- 5:      $c = V(l, 1 : j - 1)^{-1} v(l)$
- 6:      $r = v - V(:, 1 : j - 1)c$
- 7:      $[l, l_j] = \max(|r|)$
- 8:      $l = [l; l_j]$
- 9: **end for**

---

To develop efficient dynamical low-rank integrators for nonlinear  $G$  we propose a new interpolatory tangent space projector  $P_Y$  to replace the orthogonal projector. This interpolatory projector is a specialized oblique tangent space projector, obtained by replacing the orthogonal projectors  $UU^T$  and  $VV^T$  in (4) with oblique projectors onto the same spaces. A general form of these oblique projectors is  $U(A^T U)^{-1} A^T$  and  $V(B^T V)^{-1} B^T$  where  $A \in \mathbb{R}^{n_1 \times r}$  and  $B \in \mathbb{R}^{n_2 \times r}$  are any matrices such that  $(A^T U)$  and  $(B^T V)$  are invertible. Interpolatory projectors onto the columns of  $U$  and  $V$  are obtained by selecting  $A$  and  $B$  as specific columns of the identity matrix with appropriate dimensions.

Specifically  $A = I_{n_1}(:, \mathcal{I})$  and  $B = I_{n_2}(:, \mathcal{J})$  where  $\mathcal{I}$  contains  $r$  indices from  $\{1, \dots, n_1\}$  and  $\mathcal{J}$  contains  $r$  indices from  $\{1, \dots, n_2\}$ . The matrix  $A$  can extract  $r$  rows from  $Z$  via left multiplication  $A^T Z = Z(\mathcal{I}, :)$  and the matrix  $B$  can extract  $r$  columns from  $Z$  via right multiplication  $ZB = Z(:, \mathcal{J})$ . Moreover,  $A$  can be defined by choosing  $r$  indices in the set  $\mathcal{I}$ , and similarly,  $B$  can be defined by choosing  $r$  indices in the set  $\mathcal{J}$ . Such indices are to be chosen so that  $A^T U = U(\mathcal{I}, :)$  and  $B^T V = V(\mathcal{J}, :)$  are invertible, and ideally with a small condition number. This can be achieved with a sparse sampling algorithm such as the discrete empirical interpolation method (DEIM). The DEIM, summarized in Algorithm 1, is a greedy algorithm that selects an index for each column of  $U$  or  $V$  to minimize the condition number of the interpolatory projector as much as possible. Other sparse sampling strategies, such as Q-DEIM or oversampling methods, can also be used and may yield better-conditioned interpolatory projectors than DEIM in certain cases. Replacing the orthogonal projectors in (4) with interpolatory projectors onto the columns of  $U$  and  $V$  results in an interpolatory projector onto the tangent space  $T_Y \mathcal{M}$ , of the form

$$P_Y Z = Z(:, \mathcal{J})V(\mathcal{J}, :)^{-T}V^T - UU(\mathcal{I}, :)^{-1}Z(\mathcal{I}, \mathcal{J})V(\mathcal{J}, :)^{-T}V^T + UU(\mathcal{I}, :)^{-1}Z(\mathcal{I}, :). \tag{5}$$

It is easy to verify that  $(P_Y Z)(i, j) = Z(i, j)$  whenever  $i \in \mathcal{I}$  or  $j \in \mathcal{J}$ . In other words, the projection  $P_Y Z$  interpolates  $Z$  along rows and columns specified by the index sets  $\mathcal{I}$  and  $\mathcal{J}$ .

2.1.1. Matrix cross integrator

A straightforward low-rank time integration scheme can be derived by evaluating the dynamical low-rank evolution equation (3) at the sampled indices  $\mathcal{I}$  and  $\mathcal{J}$  and leveraging the interpolation property. This leads to the system of evolution equations

$$\begin{aligned} \frac{dY(\mathcal{I}(t), :, t)}{dt} &= G_Y(\mathcal{I}(t), :, t), \\ \frac{dY(:, \mathcal{J}(t), t)}{dt} &= G_Y(:, \mathcal{J}(t), t), \end{aligned} \tag{6}$$

where  $G_Y(t) = G(Y(t), t)$ . Equation (6) consists of  $(n_1 + n_2)r$  coupled nonlinear differential equations governing the evolution of a subset of the entries in the approximate solution, which can be integrated using standard explicit or implicit methods. The indices  $\mathcal{I}(t)$  and  $\mathcal{J}(t)$  that define the interpolatory projector (5) are chosen at each time  $t$  to ensure that the projector remains well-defined during time integration. If  $G$  arises from the spatial discretization of a PDE (1) involving differential operators, evaluating  $G_Y(\mathcal{I}, :, t)$  and  $G_Y(:, \mathcal{J}, t)$  requires entries of  $Y$  at indices adjacent to  $\mathcal{I}$  and  $\mathcal{J}$ . The values of  $Y$  at adjacent indices can always be obtained by constructing the solution  $Y(t)$  as a low-rank matrix using CUR decomposition. For example given  $Y(\mathcal{I}, :)$  and  $Y(:, \mathcal{J})$  at any time  $t$  the approximate solution  $Y$  can be obtained as

$$Y = Y(:, \mathcal{J})Y(\mathcal{I}, \mathcal{J})^{-1}Y(\mathcal{I}, :). \tag{7}$$

If  $Y = U\Sigma V^T$  is the SVD and the indices  $\mathcal{I}$  and  $\mathcal{J}$  defining the interpolatory projector are chosen so that  $U(\mathcal{I}, :)$  and  $V(\mathcal{J}, :)$  are well-conditioned, then constructing  $Y$  using (7) involves  $Y(\mathcal{I}, \mathcal{J})^{-1} = V(\mathcal{J}, :)^{-T}\Sigma^{-1}U(\mathcal{I}, :)^{-1}$ . Thus the condition number of the middle matrix is inversely proportional to the smallest singular value of  $Y$ . Instead of constructing  $Y$  in this way, if we first take a QR decomposition  $Y(:, \mathcal{J}) = QR$ , then we can write  $Y(\mathcal{I}, \mathcal{J}) = Q(\mathcal{I}, :)R$  and the CUR formula (7) becomes

$$Y = QQ(\mathcal{I}, :)^{-1}Y(\mathcal{I}, :), \tag{8}$$

which is an interpolatory projection of  $Y(\mathcal{I}, :)$  onto the orthonormal basis  $Q$ . In particular the stability of constructing the solution using (8) depends on the condition of the interpolatory projector and is independent of the singular values of  $Y$ . Computing the right hand side of (6), and hence integrating the system, is efficient for any nonlinear  $G$  that can be evaluated entry-wise, regardless of its low-rank structure. This enables us to efficiently compute dynamical low-rank approximations for problems where orthogonal tangent space projection is too expensive.

2.1.2. Projector-splitting integrator

An alternative approach to the matrix-cross integrator presented above for integrating (3) is to apply a standard splitting method, similar to the integrators proposed for orthogonal projectors in [31]. Since (3) is a sum of three terms, applying Lie-Trotter splitting yields three substeps commonly referred to as **K**-, **S**-, and **L**-step. Beginning from the rank- $r$  decomposition of the approximate solution  $Y(t_0) = U(t_0)S(t_0)V^T(t_0)$  at time  $t_0$ , each step updates a single factor matrix. After all three steps we obtain the low-rank factors for the approximate solution  $Y(t_1) = U(t_1)S(t_1)V^T(t_1)$  at time  $t_1 = t_0 + \Delta t$ . The substeps for interpolatory projector-splitting integrator are as follows. We denote by  $\text{DEIM}(\cdot)$  a subroutine that takes a matrix of size  $n \times r$  as input and outputs a collection of  $r$  indices computed with the DEIM index selection (Algorithm 1).

1. **K-step**: update  $U(t_0) \rightarrow U(t_1)$  and  $S(t_0) \rightarrow R(t_1)$ .

Compute interpolation indices  $\mathcal{J} = \text{DEIM}(V(t_0))$ . Then integrate the  $n_1 \times r$  differential equation

$$\frac{dK(t)}{dt} = G_K(:, \mathcal{J}, t) [V(\mathcal{J}, :, t_0)]^{-T}, \quad K(t_0) = U(t_0)S(t_0), \tag{9}$$

where  $G_K(t) = G(K(t)V(t_0)^T, t)$  from  $t_0$  to  $t_1$ , and perform a QR-decomposition  $K(t_1) = U(t_1)R(t_1)$ .

2. **S-step**: update  $R(t_1) \rightarrow \tilde{S}(t_1)$ .

Compute interpolation indices  $\mathcal{I} = \text{DEIM}(U(t_1))$ . Then integrate the  $r \times r$  matrix differential equation



$$\frac{d\tilde{S}(t)}{dt} = - [U(\mathcal{I}, :, t_1)]^{-1} G_S(\mathcal{I}, \mathcal{J}, t) [V(\mathcal{J}, :, t_0)]^{-T}, \quad \tilde{S}(t_0) = R(t_1), \quad (10)$$

where  $G_S(t) = G(U(t_1)S(t)V(t_0)^T, t)$  from  $t_0$  to  $t_1$ .

- 3. **L-step:** update  $V(t_0) \rightarrow V(t_1)$  and  $\tilde{S}(t_1) \rightarrow S(t_1)$ .

Integrate the  $n_2 \times r$  matrix differential equation

$$\frac{dL(t)}{dt} = G_L(\mathcal{I}, :, t)^T [U(\mathcal{I}, :, t_1)]^{-T}, \quad L(t_0) = V(t_0)\tilde{S}(t_1)^T, \quad (11)$$

where  $G_L(t) = G(U(t_1)L(t)^T, t)$  from time  $t_0$  to  $t_1$ , and perform a QR-decomposition  $L(t_1) = V(t_1)S(t_1)^T$ .

Similar to the matrix-cross integrator, the interpolatory projector-splitting integrator requires evaluating the output of  $G$  at only a subset of  $nr$  indices in the **K-** and **L-step** and  $r^2$  indices in the **S-step**. These evaluations can be performed efficiently for any  $G$  that can be evaluated entry-wise. Meanwhile the corresponding steps of the orthogonal projector-splitting integrator (summarized in [6]) involve inner products involving the output of  $G$  that are only efficient to compute when  $G$  has low-rank structure. The difference between the interpolatory projector-splitting integrator and the matrix cross integrator is the order in which interpolatory projection and time integration are performed. The former applies the interpolatory projection on the vector-field and then integrates the factor matrices of the solution. The latter integrates the solution at the specified indices and then performs an interpolatory projection onto the updated basis  $Q$  in (8).

### 3. Tensor train (TT) format

In the following sections we propose dynamical low-rank approximation with interpolatory projections for tensors ( $d \geq 2$ ) in the tensor train (TT) format. When  $d = 2$  the TT algorithms described hereafter reduce to the matrix algorithms discussed in Section 2. We begin with a brief review of the TT format and orthogonal TT representations. For a more detailed introduction to the TT format we refer the reader to [38]. Throughout the remainder of this paper, matrices are denoted by boldface letters, while tensors are denoted by regular (non-bold) letters. The  $k$ th unfolding of a tensor  $Y \in \mathbb{R}^{n_1 \times \dots \times n_d}$  is the matrix  $\mathbf{Y}^{(k)} \in \mathbb{R}^{(n_1 \dots n_k) \times (n_{k+1} \dots n_d)}$  with rows and columns indexed colexicographically. The TT-rank of  $Y$  is defined as the vector  $\mathbf{r} = (1, r_1, \dots, r_d, 1)$  where  $r_k$  is the rank of the unfolding matrix  $\mathbf{Y}^{(k)}$ . Any tensor  $Y$  with TT-rank  $\mathbf{r}$  can be represented in the TT format as

$$Y(i_1, i_2, \dots, i_d) = C_1(i_1)C_2(i_2) \cdots C_d(i_d), \quad (12)$$

where each  $C_k$  is a  $r_{k-1} \times n_k \times r_k$  tensor referred to as a TT-core and  $C_k(i_k)$  is a  $r_{k-1} \times r_k$  matrix for a fixed index  $i_k$ . Each TT-core has a left unfolding matrix  $\mathbf{C}_k^{(l)} \in \mathbb{R}^{r_{k-1} n_k \times r_k}$  and a right unfolding  $\mathbf{C}_k^{(r)} \in \mathbb{R}^{r_{k-1} \times n_k r_k}$  obtained by reshaping the elements of  $C_k$

$$C_k^{(l)}(\alpha_{k-1}i_k, \alpha_k) = C_k^{(r)}(\alpha_{k-1}, i_k\alpha_k) = C_k(\alpha_{k-1}, i_k, \alpha_k), \quad k = 1, 2, \dots, d.$$

The left and right unfoldings of each TT-core are full rank whenever  $Y$  has TT-rank  $\mathbf{r}$ . To simplify notation of tensors in the TT format we often omit indices so that (12) is replaced by  $Y = C_1C_2 \cdots C_d$ . To obtain even more compact representations, we define partial products of TT-cores  $C_{\leq k} \in \mathbb{R}^{n_1 \times \cdots \times n_k \times r_k}$  and  $C_{>k} \in \mathbb{R}^{r_k \times n_{k+1} \times \cdots \times n_d}$  with entries

$$\begin{aligned} C_{\leq k}(i_1, \dots, i_k, \cdot) &= C_1(i_1) \cdots C_k(i_k, \cdot), \\ C_{>k}(\cdot, i_{k+1}, \dots, i_d) &= C_{k+1}(\cdot, i_{k+1}) \cdots C_d(i_d), \end{aligned} \tag{13}$$

so that  $Y = C_{\leq k}C_{>k}$ . We also define certain unfolding matrices of these partial product tensors

$$\begin{aligned} C_{\leq k}(i_1 \cdots i_k, \cdot) &= C_{\leq k}(i_1, \dots, i_k, \cdot), \\ C_{>k}(i_{k+1} \cdots i_d, \cdot) &= C_{>k}(\cdot, i_{k+1}, \dots, i_d), \end{aligned} \tag{14}$$

where  $C_{\leq k} \in \mathbb{R}^{(n_1 \cdots n_k) \times r_k}$  and  $C_{>k} \in \mathbb{R}^{(n_{k+1} \cdots n_d) \times r_k}$ , which allows us to write the  $k$ th unfolding matrix of  $Y$  as  $\mathbf{Y}^{(k)} = C_{\leq k}C_{>k}^T$ .

### 3.1. Orthogonalization of tensor trains

Orthogonal TT representations are fundamental for executing many operations in the TT format. We will use them in this paper to obtain projectors onto the tangent spaces of TT manifolds. Hereafter, we recall an algorithm for orthogonalizing TTs by recursively applying QR-decomposition to TT-core unfoldings. Begin by taking a QR-decomposition of the left unfolding of  $C_1$

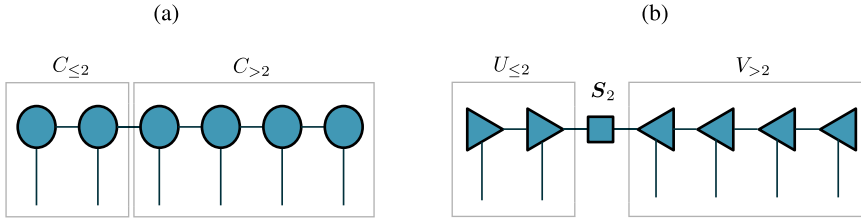
$$C_1^{(l)} = U_1^{(l)} R_1,$$

to obtain the matrix  $R_1 \in \mathbb{R}^{r_1 \times r_1}$  and the new TT-core  $U_1 \in \mathbb{R}^{r_0 \times n_1 \times r_1}$  defined by its left unfolding. The new TT-core is called left-orthogonal because it satisfies

$$\left[ U_1^{(l)} \right]^T U_1^{(l)} = I_{r_1},$$

where  $I_{r_1}$  denotes the  $r_1 \times r_1$  identity matrix. Next define a new second core  $\widehat{C}_2(i_2) = R_1 C_2(i_2)$  to obtain the TT representation  $Y = U_1 \widehat{C}_2 C_{>2}$  of the tensor (12). Then take a QR-decomposition of the left unfolding of the second core

$$\widehat{C}_2^{(l)} = U_2^{(l)} R_2,$$



**Fig. 2.** Tensor network diagrams of a  $d = 6$  dimensional tensor in the TT format. (a) No orthogonalization with the TT-core partial products  $C_{\leq 2}$  and  $C_{> 2}$  indicated. (b) Orthogonalized TT (15) with  $k = 2$  and left orthogonal TT-cores  $U_{\leq 2}$  and right orthogonal TT-cores  $V_{> 2}$  indicated.

to obtain  $\mathbf{R}_2 \in \mathbb{R}^{r_2 \times r_2}$  and the left-orthogonal TT-core  $U_2 \in \mathbb{R}^{r_1 \times n_2 \times r_2}$ . Define a new third core  $\hat{C}_3(i_3) = \mathbf{R}_2 C_3(i_3)$  to write  $Y = U_{\leq 2} \hat{C}_3 C_{> 3}$  where now the first two cores are left orthogonal. Proceeding recursively in this way we obtain the TT representation

$$Y = U_{\leq k} \mathbf{R}_k C_{> k},$$

where  $U_j$  is left-orthogonal for  $j = 1, 2, \dots, k$  and  $\mathbf{R}_k \in \mathbb{R}^{r_k \times r_k}$ . Similar to orthogonalizing cores from left to right, we can also orthogonalize cores from right to left by recursively performing QR-decompositions on right unfoldings (see [38, Section 3]) to obtain

$$Y = U_{\leq k} \mathbf{R}_k \mathbf{R}_{k+1} V_{> k},$$

where  $\mathbf{R}_{k+1} \in \mathbb{R}^{r_k \times r_k}$  and the  $V_j$  are right-orthogonal, i.e.,

$$\mathbf{V}_j^{(r)} \left[ \mathbf{V}_j^{(r)} \right]^T = \mathbf{I}_{r_{j-1}}, \quad j = k + 1, \dots, d.$$

Letting  $\mathbf{S}_k = \mathbf{R}_k \mathbf{R}_{k+1}$  we obtain the orthogonalized TT representation

$$Y = U_{\leq k} \mathbf{S}_k V_{> k}. \tag{15}$$

Utilizing the unfolding matrices of partial products (14) we also have a decomposition of the  $k$ th unfolding matrix

$$\mathbf{Y}^{(k)} = \mathbf{U}_{\leq k} \mathbf{S}_k \mathbf{V}_{> k}^T. \tag{16}$$

It follows from the left orthogonality of  $U_j$  and right orthogonality of  $V_j$  that the columns of  $\mathbf{U}_{\leq k}$  and  $\mathbf{V}_{> k}$  are orthonormal, i.e.,  $\mathbf{U}_{\leq k}^T \mathbf{U}_{\leq k} = \mathbf{V}_{> k}^T \mathbf{V}_{> k} = \mathbf{I}_{r_k}$ . The decomposition (16) resembles a SVD however  $\mathbf{S}_k$  is not necessarily diagonal. If the TT-rank of  $Y$  is  $r$  then  $\mathbf{S}_k$  is invertible. What is important for the projectors defined in the subsequent section is that the columns of  $\mathbf{U}_{\leq k}$  and  $\mathbf{V}_{> k}$  form orthonormal bases for the range and co-range of  $\mathbf{Y}^{(k)}$  respectively. In Fig. 2 we summarize the TT format and TT orthogonalizations using tensor network diagrams.

#### 4. Projections onto the TT tangent space

It is well-known that the collection of all rank- $\mathbf{r}$  TTs

$$\mathcal{M}_{\mathbf{r}} = \{Y \in \mathbb{R}^{n_1 \times \dots \times n_d} \mid \text{TT-rank}(Y) = \mathbf{r}\}, \tag{17}$$

is a smooth embedded submanifold of  $\mathbb{R}^{n_1 \times \dots \times n_d}$  [27]. Hence for any tensor  $Y \in \mathcal{M}_{\mathbf{r}}$  we can define the tangent space  $T_Y \mathcal{M}_{\mathbf{r}}$ , which is a vector subspace of  $\mathbb{R}^{n_1 \times \dots \times n_d}$  that linearizes the manifold  $\mathcal{M}_{\mathbf{r}}$  around  $Y$ . Given a TT representation (12) of  $Y$ , any element of the tangent space can be written (non-uniquely) as

$$\delta Y = \delta C_1 C_2 \dots C_d + C_1 \delta C_2 C_3 \dots C_d + \dots + C_1 \dots C_{d-1} \delta C_d, \tag{18}$$

where  $\delta C_k \in \mathbb{R}^{r_{k-1} \times n_k \times r_k}$  are first order variations of the TT-cores.

##### 4.1. The orthogonal tangent space projector

The orthogonal projector  $\widehat{P}_Y : \mathbb{R}^{n_1 \times \dots \times n_d} \rightarrow T_Y \mathcal{M}_{\mathbf{r}}$  onto the tangent space considered in [32] determines the best approximation of a given tensor  $Z \in \mathbb{R}^{n_1 \times \dots \times n_d}$  in the tangent space relative to the Frobenius norm. Such orthogonal projector can be constructed from the orthogonal projectors

$$\widehat{P}_{\leq k} = \mathbf{U}_{\leq k} \mathbf{U}_{\leq k}^T, \quad \widehat{P}_{> k} = \mathbf{V}_{> k} \mathbf{V}_{> k}^T, \quad k = 1, 2, \dots, d - 1, \tag{19}$$

onto the range and co-range of  $\mathbf{Y}^{(k)}$ . The orthogonal bases  $\mathbf{U}_{\leq k}$  and  $\mathbf{V}_{> k}$  for the range and co-range of  $\mathbf{Y}^{(k)}$  can be obtained from the TT-orthogonalization procedure described in Section 3.1. These projectors act on the matrix space  $\mathbb{R}^{(n_1 \dots n_k) \times (n_{k+1} \dots n_d)}$ . To construct the orthogonal tangent space projector  $\widehat{P}_Y$  it is convenient to define projectors corresponding to (19) that act on the tensor space  $\mathbb{R}^{n_1 \times \dots \times n_d}$  by

$$\widehat{P}_{\leq k} Z = \text{Ten}_k \left[ \widehat{P}_{\leq k} \mathbf{Z}^{(k)} \right], \quad \widehat{P}_{> k} Z = \text{Ten}_k \left[ \mathbf{Z}^{(k)} \widehat{P}_{> k} \right], \tag{20}$$

where  $\text{Ten}_k$  denotes the tensorization operator that is the inverse of the  $k$ th unfolding, i.e.,  $\text{Ten}_k \left( \mathbf{Z}^{(k)} \right) = Z$ . The projectors  $\widehat{P}_{\leq j}, \widehat{P}_{> k}$  commute whenever  $j \leq k$  and can be used to construct the orthogonal projector onto the tangent space [32, Corollary 3.2]

$$\widehat{P}_Y = \sum_{k=1}^{d-1} \widehat{P}_{\leq k-1} \widehat{P}_{> k} - \widehat{P}_{\leq k} \widehat{P}_{> k} + \widehat{P}_{\leq d-1}, \tag{21}$$

where we set  $\widehat{P}_{\leq 0} = 1$ .

4.2. Oblique tangent space projectors

The computational cost of dynamical low-rank approximation of (2) with orthogonal tangent space projections can scale as  $\mathcal{O}(n^d)$  when  $G$  lacks low-rank structure. In this case, the orthogonal dynamical low-rank method is impractical as its computational cost is comparable to solving (2) without low-rank compression. To enable efficient dynamical low-rank approximation in such cases, we introduce oblique projections onto the TT tangent space that can be computed in only  $\mathcal{O}(dnr^3)$  operations for many applications where orthogonal projections require  $\mathcal{O}(n^d)$ . We construct such oblique projectors by replacing the orthogonal projectors defined in (19) with oblique projectors onto the same spaces

$$P_{\leq k} = U_{\leq k} \left( X_{\leq k}^T U_{\leq k} \right)^{-1} X_{\leq k}^T, \quad P_{> k} = X_{> k} \left( X_{> k}^T V_{> k} \right)^{-T} V_{> k}^T, \quad (22)$$

defined for any matrices  $X_{\leq k} \in \mathbb{R}^{(n_1 \cdots n_k) \times r_k}$  and  $X_{> k} \in \mathbb{R}^{(n_{k+1} \cdots n_d) \times r_k}$  such that  $\left( X_{\leq k}^T U_{\leq k} \right)$  and  $\left( X_{> k}^T V_{> k} \right)$  are invertible. Just as before, it is convenient to define oblique projectors corresponding to (22) that act on the tensor space  $\mathbb{R}^{n_1 \times \cdots \times n_d}$  by

$$P_{\leq k} Z = \text{Ten}_k \left[ P_{\leq k} Z^{(k)} \right], \quad P_{> k} Z = \text{Ten}_k \left[ Z^{(k)} P_{> k} \right]. \quad (23)$$

With these projectors we can construct an oblique tangent space projector with the same form as the orthogonal tangent space projector (21).

**Proposition 4.1.** *Let  $Y \in \mathcal{M}_r$  with orthogonal decompositions of its unfolding matrices given in (16) and suppose  $X_{\leq k}, X_{> k}$  define oblique projectors (22). Then the map*

$$P_Y = \sum_{k=1}^{d-1} P_{\leq k-1} P_{> k} - P_{\leq k} P_{> k} + P_{\leq d-1}, \quad (24)$$

with  $P_{\leq 0} = 1$ , defined for any tensor  $Z \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ , is an oblique projector onto the tangent space  $T_Y \mathcal{M}_r$ .

**Proof.** First we show that the image of  $P_Y$  is contained in  $T_Y \mathcal{M}_r$ . Note that  $T_Y \mathcal{M}_r$  is a linear space and thus it is sufficient to show that the image of each term in (24) belongs to  $T_Y \mathcal{M}_r$ . To do so we utilize [32, Corollary 3.2] which shows that  $P_{\leq j}, P_{> k}$  commute whenever  $j \leq k$  and

$$P_{\leq k-1} P_{> k} Z = \text{Ten}_k \left\{ [I_{n_k} \otimes P_{\leq k-1}] Z^{(k)} P_{> k} \right\}. \quad (25)$$

Note that the result is stated for orthogonal projectors  $\widehat{P}_{\leq k}, \widehat{P}_{> k}$  however the proof does not require orthogonality of the projectors and thus holds for oblique projectors  $P_{\leq k}, P_{> k}$ . Inserting the oblique projectors (22) into (25) we obtain

$$\begin{aligned}
 & P_{\leq k-1} P_{> k} Z \\
 = & \text{Ten}_k \left\{ \left[ \mathbf{I}_{n_k} \otimes \mathbf{U}_{\leq k-1} \left( \mathbf{X}_{\leq k-1}^T \mathbf{U}_{\leq k-1} \right)^{-1} \mathbf{X}_{\leq k-1}^T \right] \mathbf{Z}^{(k)} \left[ \mathbf{X}_{> k} \left( \mathbf{X}_{> k}^T \mathbf{V}_{> k} \right)^{-T} \mathbf{V}_{> k}^T \right] \right\} \\
 = & \text{Ten}_k \left\{ \left[ \mathbf{I}_{n_k} \otimes \mathbf{U}_{\leq k-1} \right] \right. \\
 & \left. \left[ \mathbf{I}_{n_k} \otimes \left( \mathbf{X}_{\leq k-1}^T \mathbf{U}_{\leq k-1} \right)^{-1} \mathbf{X}_{\leq k-1}^T \mathbf{Z}^{(k)} \mathbf{X}_{> k} \left( \mathbf{X}_{> k}^T \mathbf{V}_{> k} \right)^{-T} \right] \mathbf{V}_{> k}^T \right\} \\
 = & \text{Ten}_k \left\{ \left[ \mathbf{I}_{n_k} \otimes \mathbf{U}_{\leq k-1} \right] \delta \mathbf{C}_k^{(l)} \mathbf{V}_{> k}^T \right\}
 \end{aligned} \tag{26}$$

where we used the mixed product property of the Kronecker product to obtain the second equality and defined  $\delta \mathbf{C}_k$  as the  $r_{k-1} \times n_k \times r_k$  TT-core with left unfolding

$$\delta \mathbf{C}_k^{(l)} = \mathbf{I}_{n_k} \otimes \left( \mathbf{X}_{\leq k-1}^T \mathbf{U}_{\leq k-1} \right)^{-1} \mathbf{X}_{\leq k-1}^T \mathbf{Z}^{(k)} \mathbf{X}_{> k} \left( \mathbf{X}_{> k}^T \mathbf{V}_{> k} \right)^{-T}. \tag{27}$$

Using the identity

$$\mathbf{C}_{\leq k} = \left( \mathbf{I}_{n_k} \otimes \mathbf{C}_{\leq k-1} \right) \mathbf{C}_k^{(l)}, \tag{28}$$

and (14) we can write (26) in TT format

$$P_{\leq k-1} P_{> k} Z = U_{\leq k-1} \delta \mathbf{C}_k V_{> k}, \tag{29}$$

which has the form (18) and thus belongs to the tangent space  $T_Y \mathcal{M}_r$ . For the  $P_{\leq k} P_{> k}$  terms we have

$$\begin{aligned}
 P_{\leq k} P_{> k} Z &= \text{Ten}_k \left\{ P_{\leq k} \mathbf{Z}^{(k)} P_{> k} \right\} \\
 &= \text{Ten}_k \left\{ U_{\leq k} \left[ \left( \mathbf{X}_{\leq k}^T U_{\leq k} \right)^{-1} \mathbf{X}_{\leq k}^T \mathbf{Z}^{(k)} \mathbf{X}_{> k} \left( \mathbf{X}_{> k}^T \mathbf{V}_{> k} \right)^{-T} \right] \mathbf{V}_{> k}^T \right\},
 \end{aligned} \tag{30}$$

which we write in the TT format as

$$P_{\leq k} P_{> k} Z = U_{\leq k} \delta \mathbf{S}_k V_{> k}, \tag{31}$$

where

$$\delta \mathbf{S}_k = \left( \mathbf{X}_{\leq k}^T U_{\leq k} \right)^{-1} \mathbf{X}_{\leq k}^T \mathbf{Z}^{(k)} \mathbf{X}_{> k} \left( \mathbf{X}_{> k}^T \mathbf{V}_{> k} \right)^{-T}, \tag{32}$$

belongs to  $\mathbb{R}^{r_k \times r_k}$ . Absorbing  $\delta \mathbf{S}_k$  into its left neighboring core  $U_k$  we rewrite (31) as

$$P_{\leq k} P_{> k} Z = U_{\leq k-1} \delta \overline{\mathbf{C}}_k V_{> k}, \tag{33}$$

where  $\delta\overline{\mathcal{C}}_k(i_k) = U_k(i_k)\delta\mathbf{S}_k$ , which has the form (18) and therefore belongs to the tangent space  $T_Y\mathcal{M}_r$ .

It remains to show that  $P_Y$  is idempotent, which we verify by checking that  $P_{\leq k}$  and  $P_{> k}$  are idempotent for all  $k = 1, \dots, d - 1$ . For  $P_{\leq k}$  we have

$$\begin{aligned} P_{\leq k}^2 Z &= P_{\leq k} \text{Ten}_k \left[ P_{\leq k} \mathbf{Z}^{(k)} \right] \\ &= \text{Ten}_k \left[ P_{\leq k}^2 \mathbf{Z}^{(k)} \right] \\ &= \text{Ten}_k \left[ P_{\leq k} \mathbf{Z}^{(k)} \right] \\ &= P_{\leq k} Z, \end{aligned} \tag{34}$$

for all  $k = 1, \dots, d - 1$ . The idempotence of  $P_{> k}$  for  $k = 1, \dots, d - 1$  is also easy to verify directly.  $\square$

We remark that the orthogonality of the bases  $\mathbf{U}_{\leq k}$  and  $\mathbf{V}_{> k}$  is not necessary for the construction of oblique tangent space projectors (24). However, imposing such orthogonality is advantageous for the numerical stability and accuracy of the projectors.

### 4.3. Interpolatory tangent space projectors

Next we introduce a special class of oblique tangent space projectors with a cross interpolation property that enables efficient dynamical low-rank approximation. Such oblique projectors are obtained by selecting  $\mathbf{X}_{\leq k}$  and  $\mathbf{X}_{> k}$  in (24) as  $r_k$  columns of the identity matrix with appropriate dimension

$$\mathbf{X}_{\leq k} = \mathbf{I}_{n_1 \dots n_k} (:, I^{\leq k}), \quad \mathbf{X}_{> k} = \mathbf{I}_{n_{k+1} \dots n_d} (:, I^{> k}). \tag{35}$$

Here,  $I^{\leq k}$  contains  $r_k$  indices of the form  $i_1 \dots i_k$  and  $I^{> k}$  contains  $r_k$  indices of the form  $i_{k+1} \dots i_d$ . The matrix  $\mathbf{X}_{\leq k}$  can extract  $r_k$  rows determined by the index sets  $I^{\leq k}$  from a tensor  $\mathbf{Z}^{(k)}$  with matrix multiplication from the left  $\mathbf{X}_{\leq k}^T \mathbf{Z}^{(k)} = \mathbf{Z}^{(k)}(I^{\leq k}, :)$ . Similarly, the matrix  $\mathbf{X}_{> k}$  can extract  $r_k$  columns from  $\mathbf{Z}^{(k)}$  with matrix multiplication from the right  $\mathbf{Z}^{(k)} \mathbf{X}_{> k} = \mathbf{Z}^{(k)}(:, I^{> k})$ . Moreover, when  $\mathbf{X}_{\leq k}, \mathbf{X}_{> k}$  are defined as in (35) the oblique projectors in (22) become interpolatory projectors

$$\left[ P_{\leq k} \mathbf{Z}^{(k)} \right] (I^{\leq k}, :) = \mathbf{Z}^{(k)} (I^{\leq k}, :), \quad \left[ \mathbf{Z}^{(k)} P_{> k} \right] (:, I^{> k}) = \mathbf{Z}^{(k)} (:, I^{> k}). \tag{36}$$

Since each index  $i_1 \dots i_k$  corresponds to a multi-index  $(i_1, \dots, i_k)$  and  $i_{k+1} \dots i_d$  corresponds to a multi-index  $(i_{k+1}, \dots, i_d)$ , we identify the sets of  $r_k$  indices  $I^{\leq k}$  and  $I^{> k}$  with sets of  $r_k$  multi-indices  $\mathcal{I}^{\leq k}$  and  $\mathcal{I}^{> k}$ . With such identification the matrix interpolation property (36) is equivalent to an interpolation property of the corresponding tensor operators (23)

$$\begin{aligned}
 [P_{\leq k}Z] (\mathcal{I}^{\leq k}, i_{k+1}, \dots, i_d) &= Z (\mathcal{I}^{\leq k}, i_{k+1}, \dots, i_d), \\
 [P_{> k}Z] (i_1, \dots, i_k, \mathcal{I}^{> k}) &= Z (i_1, \dots, i_k, \mathcal{I}^{> k}).
 \end{aligned}
 \tag{37}$$

In order to obtain an oblique tangent space projectors (24) that interpolate, we consider multi-indices that satisfy the nested conditions

$$\mathcal{I}^{\leq k} \subset \mathcal{I}^{\leq k-1} \times \{1, \dots, n_k\}, \quad \mathcal{I}^{> k} \subset \{1, \dots, n_k\} \times \mathcal{I}^{> k+1}, \quad k = 1, 2, \dots, d-1.
 \tag{38}$$

To prove that (38) is sufficient for the oblique tangent space projector (24) to interpolate, we have the following Lemma.

**Lemma 4.1.** *For any nested indices (38) defining oblique projectors (23) and  $k = 1, 2, \dots, d - 1$ , the projector  $P_{\leq k-1}P_{> k}$  satisfies*

$$[P_{\leq k-1}P_{> k}Z] (\mathcal{I}^{\leq j-1}, i_j, \mathcal{I}^{> j}) = \begin{cases} Z (\mathcal{I}^{\leq j-1}, i_j, \mathcal{I}^{> j}), & j = k, \\ [P_{\leq k}P_{> k}Z] (\mathcal{I}^{\leq j-1}, i_j, \mathcal{I}^{> j}), & j > k, \\ [P_{\leq k-1}P_{> k-1}Z] (\mathcal{I}^{\leq j-1}, i_j, \mathcal{I}^{> j}), & j < k. \end{cases}
 \tag{39}$$

**Proof.** The case  $j = k$  follows directly from the interpolation property (37). For  $j > k$  the nesting condition (38) ensures that the first  $k - 1$  indices of each multi-index in  $\mathcal{I}^{\leq j-1}$  are a multi-index in  $\mathcal{I}^{\leq k-1}$  and that the first  $k$  indices of each multi-index in  $\mathcal{I}^{\leq j-1}$  are a multi-index in  $\mathcal{I}^{\leq k}$ . Therefore we can use the interpolation property (37) to obtain

$$[P_{\leq k-1}Z] (\mathcal{I}^{\leq j-1}, i_j, \dots, i_d) = [P_{\leq k}Z] (\mathcal{I}^{\leq j-1}, i_j, \dots, i_d) = Z (\mathcal{I}^{\leq j-1}, i_j, \dots, i_d).
 \tag{40}$$

Since  $P_{\leq k}$  and  $P_{\leq k-1}$  act only on the first  $k$  dimensions of  $Z$  and  $P_{> k}$  acts only on dimensions  $k + 1, \dots, d$  we can use the preceding equation to obtain

$$[P_{\leq k-1}P_{> k}Z] (\mathcal{I}^{\leq j-1}, i_j, \mathcal{I}^{> j}) = [P_{\leq k}P_{> k}Z] (\mathcal{I}^{\leq j-1}, i_j, \mathcal{I}^{> j}),$$

establishing the case  $j > k$ . When  $j < k$  the nested condition ensures that the  $d - j$  indices of each multi-index in  $\mathcal{I}^{> j}$  appear in multi-indices belonging to  $\mathcal{I}^{> k}$  and  $\mathcal{I}^{> k-1}$ . Therefore from the interpolation property (37) we have

$$[P_{> k}Z] (i_1, \dots, i_j, \mathcal{I}^{> j}) = [P_{> k-1}Z] (i_1, \dots, i_j, \mathcal{I}^{> j}) = Z (i_1, \dots, i_j, \mathcal{I}^{> j}),
 \tag{41}$$

from which we obtain the result for  $j < k$ .  $\square$

**Theorem 4.1.** *For any  $Y \in \mathcal{M}_r$  and  $\{\mathcal{I}^{\leq j}, \mathcal{I}^{> j}\}$  nested multi-indices defining interpolatory projectors (22) the oblique tangent space projector (24) has the cross interpolation property*

$$[P_YZ] (\mathcal{I}^{\leq j-1}, i_j, \mathcal{I}^{> j}) = Z (\mathcal{I}^{\leq j-1}, i_j, \mathcal{I}^{> j}), \quad j = 1, 2, \dots, d.
 \tag{42}$$



**Proof.** Rearrange the terms in (24) to write

$$[P_Y Z] = \sum_{k=1}^d P_{\leq k-1} P_{>k} Z - \sum_{k=1}^{d-1} P_{\leq k} P_{>k} Z, \tag{43}$$

with  $P_{>d} = 1$  and evaluate at the indices  $(\mathcal{I}^{\leq j-1}, i_j, \mathcal{I}^{>j})$

$$[P_Y Z] (\mathcal{I}^{\leq j-1}, i_j, \mathcal{I}^{>j}) = \sum_{k=1}^d [P_{\leq k-1} P_{>k} Z] (\mathcal{I}^{\leq j-1}, i_j, \mathcal{I}^{>j}) - \sum_{k=1}^{d-1} [P_{\leq k} P_{>k} Z] (\mathcal{I}^{\leq j-1}, i_j, \mathcal{I}^{>j}). \tag{44}$$

Applying Lemma 4.1 to each term in the first summation in (44) we obtain

$$\begin{aligned} \sum_{k=1}^d [P_{\leq k-1} P_{>k} Z] (\mathcal{I}^{\leq j-1}, i_j, \mathcal{I}^{>j}) &= \\ & \left( \sum_{k=1}^{j-1} [P_{\leq k} P_{>k} Z] (\mathcal{I}^{\leq j-1}, \cdot, \mathcal{I}^{>j}) \right) + Z (\mathcal{I}^{\leq j-1}, i_j, \mathcal{I}^{>j}) \\ & + \left( \sum_{l=j+1}^d [P_{\leq l-1} P_{>l-1} Z] (\mathcal{I}^{\leq j-1}, i_j, \mathcal{I}^{>j}) \right). \end{aligned} \tag{45}$$

Re-indexing the final summation in (45) with  $k = l - 1$  and combining the result with the first summation in (45) yields

$$\begin{aligned} \sum_{k=1}^d [P_{\leq k-1} P_{>k} Z] (\mathcal{I}^{\leq j-1}, i_j, \mathcal{I}^{>j}) \\ = Z (\mathcal{I}^{\leq j-1}, i_j, \mathcal{I}^{>j}) + \sum_{k=1}^{d-1} [P_{\leq k} P_{>k} Z] (\mathcal{I}^{\leq j-1}, i_j, \mathcal{I}^{>j}). \end{aligned} \tag{46}$$

Finally substituting (46) into (44) the two summations cancel and the proof is complete.  $\square$

The cross interpolation property (42) of the tangent space projector resembles the interpolation property of TT-cross approximation [47] with nested indices. Next, we describe a novel index selection algorithm for constructing interpolatory projectors onto the tangent space  $T_Y \mathcal{M}_r$ .

### 5. Index selection for oblique projectors and cross interpolation

In the preceding section we constructed tangent space projectors (24) with the cross interpolation property (42) from oblique projectors (22) onto the bases  $U_{\leq k}$  and  $V_{>k}$ .

We now devise an efficient algorithm based on the discrete empirical interpolation method (DEIM) for computing indices  $\{I^{\leq k}, I^{>k}\}$ , or equivalently multi-indices  $\{\mathcal{I}^{\leq k}, \mathcal{I}^{>k}\}$ , that yield well-defined interpolatory projectors (22) defined by the matrices (35). The DEIM (recalled in Algorithm 1) greedily selects indices to minimize the condition number of interpolatory projectors onto a given basis, e.g.,  $\mathbf{U}_{\leq k}$  or  $\mathbf{V}_{>k}$ , as much as possible. However we can not apply such algorithm directly to the matrices  $\mathbf{U}_{\leq k}, \mathbf{V}_{>k}$  as they have dimensions  $(n_1 \cdots n_k) \times r_k$  and  $(n_{k+1} \cdots n_d) \times r_k$ , respectively, which is too large to store in memory and process with the DEIM algorithm. To address the problem of memory, we do not store the matrices  $\mathbf{U}_{\leq k}$  and  $\mathbf{V}_{>k}$  directly but rather the left orthogonal cores  $U_j$  and right orthogonal cores  $V_j$  that can be used to construct these matrices

$$\begin{aligned} \mathbf{U}_{\leq k}(i_1 \cdots i_k, :) &= U_1(i_1) \cdots U_k(i_k, :), \\ \mathbf{V}_{>k}(i_{k+1} \cdots i_d, :) &= V_{k+1}(:, i_{k+1}) \cdots V_d(i_d), \quad k = 1, 2, \dots, d - 1. \end{aligned} \tag{47}$$

To address the computational cost we propose an algorithm that only samples from a small subset of entries of the matrices  $\mathbf{U}_{\leq k}, \mathbf{V}_{>k}$ . The key idea is to compute the indices  $I^{\leq k}$  recursively for  $k = 1, 2, \dots, d - 1$  by sampling from  $\mathbf{U}_{\leq k}$  only indices corresponding to multi-indices  $\mathcal{I}^{\leq k}$  that are nested (38). By considering only nested indices we reduce the number of possible indices  $i_1 \cdots i_k \in I^{\leq k}$  from  $n_1 \cdots n_k$  to  $r_{k-1}n_k$ . We use the same idea to construct the nested index sets  $\mathcal{I}^{>k}$  sequentially for  $k = d - 1, d - 2, \dots, 1$ . Since the indices obtained from this sampling approach are nested by construction, the resulting tangent space projector (24) has the cross interpolation property (42). Such cross interpolation property will be useful for the dynamical low-rank approximation schemes presented in Section 6. Hereafter we present the nested index selection algorithm in detail and then show in Theorem 5.1 that this nested sampling method always produces well-defined interpolatory projectors.

### 5.1. The TT-cross-DEIM algorithm

To compute the indices  $I^{\leq j}$ , begin by applying the DEIM algorithm to the  $n_1 \times r_1$  matrix  $\mathbf{U}_{\leq 1}$

$$I^{\leq 1} = \text{DEIM}(\mathbf{U}_{\leq 1}). \tag{48}$$

To obtain  $I^{\leq 2}$ , construct the  $r_1 n_2 \times r_2$  matrix

$$\widehat{\mathbf{U}}_{\leq 2}(\alpha_1 i_2, \alpha_2) = \mathbf{U}_{\leq 2}(I_{\alpha_1}^{\leq 1} i_2, \alpha_2), \tag{49}$$

which is the restriction of  $\mathbf{U}_{\leq 2}$  to the indices  $I^{\leq 1}$  with  $I_{\alpha_1}^{\leq 1}$  denoting the  $\alpha_1$ st index in  $I^{\leq 1}$ . Then sample  $r_2$  indices from this restricted matrix

$$\mathbf{l}_{\leq 2} = \text{DEIM}(\widehat{\mathbf{U}}_{\leq 2}). \tag{50}$$

Due to the construction of  $\widehat{\mathbf{U}}_{\leq 2}$  in (49), each index  $\mathbf{l}_{\leq 2}(\alpha_2)$  corresponds to a multi-index  $(\mathbf{p}_{\leq 1}(\alpha_2), \mathbf{p}_2(\alpha_2))$  where  $\mathbf{p}_{\leq 1}(\alpha_2)$  identifies an index in  $\mathcal{I}^{\leq 1}$  and  $\mathbf{p}_2(\alpha_2)$  identifies an index in  $\{1, \dots, n_2\}$ . Construct the multi-index set

$$\mathcal{I}_{\alpha_2}^{\leq 2} = \left( \mathcal{I}_{\mathbf{p}_{\leq 1}(\alpha_2)}^{\leq 1}, \mathbf{p}_2(\alpha_2) \right), \quad \alpha_2 = 1, 2, \dots, r_2, \tag{51}$$

which corresponds to the set of indices  $I^{\leq 2}$ . The remaining sets of indices  $I^{\leq j}$  are obtained inductively with a similar procedure. After computing  $I^{\leq j-1}$ , construct the  $r_{j-1}n_j \times r_j$  matrix

$$\widehat{\mathbf{U}}_{\leq j}(\alpha_{j-1}i_j, \alpha_j) = \mathbf{U}_{\leq j} \left( I_{\alpha_{j-1}}^{\leq j-1} i_j, \alpha_j \right), \tag{52}$$

which is the restriction of  $\mathbf{U}_{\leq j}$  to the indices  $I^{\leq j-1}$ . Then sample  $r_j$  indices from the restricted matrix

$$\mathbf{l}_{\leq j} = \text{DEIM} \left( \widehat{\mathbf{U}}_{\leq j} \right). \tag{53}$$

Due to the construction of  $\widehat{\mathbf{U}}_{\leq j}$  in (52) each index  $\mathbf{l}_{\leq j}(\alpha_j)$  corresponds to a multi-index  $(\mathbf{p}_{\leq j-1}(\alpha_j), \mathbf{p}_j(\alpha_j))$  where  $\mathbf{p}_{\leq j-1}(\alpha_j)$  identifies a multi-index in  $\mathcal{I}^{\leq j-1}$  and  $\mathbf{p}_j(\alpha_j)$  identifies an index in  $\{1, \dots, n_j\}$ . Construct the set  $\mathcal{I}^{\leq j}$  with multi-indices

$$\mathcal{I}_{\alpha_j}^{\leq j} = \left( \mathcal{I}_{\mathbf{p}_{\leq j-1}(\alpha_j)}^{\leq j-1}, \mathbf{p}_j(\alpha_j) \right), \quad \alpha_j = 1, 2, \dots, r_j, \tag{54}$$

which corresponds to  $I^{\leq j}$ . This procedure computes index sets  $I^{\leq j}$  sequentially for  $j = 1, 2, \dots, d-1$  by applying the DEIM sampling algorithm to the restricted matrices  $\widehat{\mathbf{U}}_{\leq j}$  of dimension  $r_{j-1}n_j \times r_j$ . In practice the restricted matrices  $\widehat{\mathbf{U}}_{\leq j}$  are not obtained from  $\mathbf{U}_{\leq j}$  as written in (52) but rather from the low-dimensional tensor cores  $U_j$  that construct  $\mathbf{U}_{\leq j}$  in (47).

We compute the index sets  $I^{> j}$  in a similar manner. First obtain  $I^{> d-1}$  by sampling the  $n_d \times r_{d-1}$  matrix  $\mathbf{V}_{> d}$

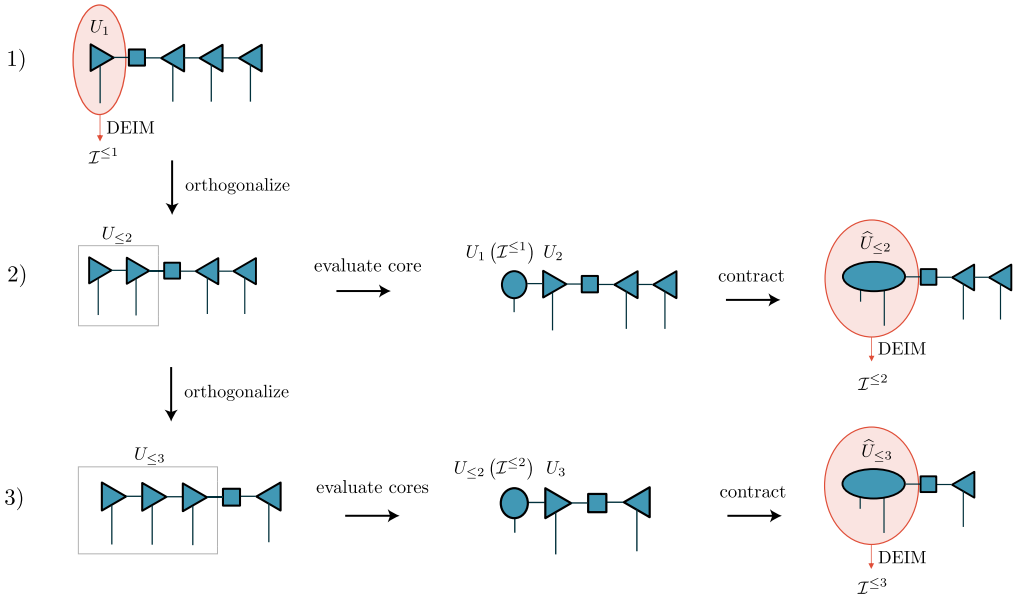
$$I^{> d-1} = \text{DEIM}(\mathbf{V}_{> d-1}). \tag{55}$$

Then construct index sets  $I^{> j}$  inductively for  $j = d-2, d-3, \dots, 1$  as follows. After computing  $I^{> j+1}$ , construct the  $n_{j+1}r_{j+1} \times r_j$  matrix  $\widehat{\mathbf{V}}_{> j}$

$$\widehat{\mathbf{V}}_{> j}(i_{j+1}\alpha_{j+1}, \alpha_j) = \mathbf{V}_{> j} \left( i_{j+1} I_{\alpha_{j+1}}^{> j+1}, \alpha_j \right) \tag{56}$$

which is the restriction of  $\mathbf{V}_{> j}$  to the indices  $\mathcal{I}^{> j+1}$ . Then sample  $r_j$  indices from the restricted matrix

$$\mathbf{l}_{> j} = \text{DEIM} \left( \widehat{\mathbf{V}}_{> j} \right). \tag{57}$$



**Fig. 3.** Illustration of the TT-cross-DEIM left-to-right sweep that computes multi-indices  $\mathcal{I}^{\leq j}$  sequentially for  $j = 1, 2, \dots, d - 1$  with  $d = 4$ .

Due to the construction of  $\widehat{\mathbf{V}}_{>j}$  in (56) each index  $\mathbf{l}_{>j}(\alpha_j)$  corresponds to a multi-index  $(\mathbf{p}_{j+1}(\alpha_j), \mathbf{p}_{>j+1}(\alpha_j))$ , where  $\mathbf{p}_{j+1}(\alpha_j)$  identifies an index in  $\{1, \dots, n_{j+1}\}$  and  $\mathbf{p}_{>j+1}(\alpha_j)$  identifies a multi-index in  $\mathcal{I}^{>j+1}$ . Construct the set  $\mathcal{I}^{>j}$  with multi-indices

$$\mathcal{I}_{\alpha_j}^{>j} = \left( \mathbf{p}_{j+1}(\alpha_j), \mathcal{I}_{\mathbf{p}_{>j+1}(\alpha_j)}^{>j+1} \right), \quad \alpha_j = 1, 2, \dots, r_j, \tag{58}$$

which corresponds to  $I^{>j}$ . Just as in the computation of  $I^{\leq j}$ , the restricted matrices  $\widehat{\mathbf{V}}_{>j}$  are not obtained from  $\mathbf{V}_{>j}$  as written in (56) but rather from the low-dimensional tensor cores  $V_j$  that construct  $\mathbf{V}_{>j}$  in (47).

The entire algorithm is summarized in Algorithm 2 and a tensor network diagram of the left-to-right sweep for computing  $I^{\leq j}$  from the TT-cores  $U_j$  is shown in Fig. 3 with  $d = 4$ . In Algorithm 2 we denote by `ind2sub` the Matlab function that reshapes linear indices to multi-indices. As mentioned above, the multi-index sets  $\mathcal{I}^{\leq k}, \mathcal{I}^{>k}$  obtained in (54) and (58) are nested (38) by construction. Thus the oblique tangent space projector (24) constructed from these indices is a cross interpolant (see Theorem 4.1), provided it is well-defined. It is shown in Section 5.2 that such projector is in fact well-defined.

*Computational cost* To simplify the operation count of the proposed TT-cross-DEIM algorithm, we assume that  $r_k = r$  and  $n_k = n$  for all  $k = 1, 2, \dots, d$ . The Algorithm requires access to all  $d - 1$  orthogonal representations (15) which can be computed in  $\mathcal{O}(dnr^3)$  operations [38]. With these orthogonal representations available, each index set  $I^{\leq k}$  and  $I^{>k}$  for  $k = 1, 2, \dots, d - 1$  is computed by applying DEIM to a matrix of size

---

**Algorithm 2** TT-cross-DEIM index selection.

---

Require:

$\mathbf{U}_{\leq j}, \mathbf{V}_{> j}, j = 1, \dots, d - 1$ , orthogonal bases for range and co-range of  $\mathbf{Y}^{(j)}$  as in (16)

Ensure:

$\{\mathcal{I}^{\leq j}, \mathcal{I}^{> j}\}$ , nested multi-index sets defining oblique projectors (22) via (35)

- 1:  $I^{\leq 1} = \text{DEIM}(\mathbf{U}_{\leq 1})$  ▷ left-to-right sweep: computing  $I^{\leq j}$
  - 2: **for**  $j = 2$  **to**  $d - 1$  **do**
  - 3:      $\widehat{\mathbf{U}}_{\leq j} = \mathbf{U}_{\leq j} (I^{\leq j-1} \cdot, \cdot)$
  - 4:      $\mathbf{l}_{\leq j} = \text{DEIM}(\widehat{\mathbf{U}}_j)$
  - 5:      $\mathbf{p}_{\leq j-1}, \mathbf{p}_j = \text{ind2sub}([r_{j-1}, n_j], \mathbf{l}_{\leq j})$
  - 6:      $\mathcal{I}_{\alpha_j}^{\leq j} = (\mathcal{I}_{\mathbf{p}_{\leq j-1}(\alpha_j)}^{\leq j-1}, \mathbf{p}_j(\alpha_j))$ ,      $\alpha_j = 1, 2, \dots, r_j$
  - 7: **end for**
  - 8:  $I^{> d-1} = \text{DEIM}(\mathbf{V}_{> d-1})$  ▷ right-to-left sweep: computing  $I^{> j}$
  - 9: **for**  $j = d - 2$  **to**  $1$  **do**
  - 10:      $\widehat{\mathbf{V}}_{> j} = \mathbf{V}_{> j} (: I^{> j+1}, \cdot)$
  - 11:      $\mathbf{l}_{> j} = \text{DEIM}(\widehat{\mathbf{V}}_{> j})$
  - 12:      $\mathbf{p}_{j+1}, \mathbf{p}_{> j+1} = \text{ind2sub}([n_j, r_j], \mathbf{l}_{> j})$
  - 13:      $\mathcal{I}_{\alpha_j}^{> j} = (\mathbf{p}_{j+1}(\alpha_j), \mathcal{I}_{\mathbf{p}_{> j+1}(\alpha_j)}^{> j+1})$ ,      $\alpha_j = 1, 2, \dots, r_j$
  - 14: **end for**
- 

$rn \times r$  each of which requires  $\mathcal{O}(nr)$  operations. Therefore the total number of operations in the TT-cross-DEIM algorithm is  $\mathcal{O}(dnr^3)$ .

Note that the operation count of TT-cross-DEIM is dominated by the computation of orthogonal TT representations and thus has the same complexity as many common TT algorithms, e.g., TT rounding.

### 5.2. Condition of the oblique projectors

For the oblique projectors (22) to be defined (and thus for the oblique tangent space projector (24) to be defined) the  $r_j \times r_j$  matrices  $\mathbf{M}_j = \mathbf{X}_{\leq j}^T \mathbf{U}_{\leq j}$  and  $\mathbf{N}_j = \mathbf{X}_{> j}^T \mathbf{V}_{> j}$  must be invertible. For interpolatory projectors, i.e., when  $\widehat{\mathbf{X}}_{\leq j}$  and  $\widehat{\mathbf{X}}_{> j}$  are defined as in (35), these matrices are given entry-wise by

$$\mathbf{M}_j(\alpha_j, \beta_j) = \mathbf{U}_{\leq j} \left( I_{\alpha_j}^{\leq j}, \beta_j \right), \quad \mathbf{N}_j(\alpha_j, \beta_j) = \mathbf{V}_{> j} \left( I_{\alpha_j}^{> j}, \beta_j \right). \tag{59}$$

The following result shows that the TT-cross-DEIM produces indices that yield invertible matrices (59) and therefore define oblique projectors (22) and (24).

**Theorem 5.1.** *If  $Y \in \mathcal{M}_r$  and  $I^{\leq j}, I^{> j}$  are obtained with the TT-cross-DEIM then the  $r_j \times r_j$  matrices (59) are invertible for all  $j = 1, 2, \dots, d - 1$ .*

**Proof.** We prove the result for  $\mathbf{M}_j$  by induction on  $j$ . The  $r_1$  indices  $I^{\leq 1}$  are obtained in (48) from  $\mathbf{U}_{\leq 1}$  with the DEIM. Because  $Y \in \mathcal{M}_r$ , the unfolding matrix  $\mathbf{U}_{\leq 1}$  is full rank. Thus we have from [49, Lemma 3.1] that the  $r_1 \times r_1$  matrix  $\mathbf{M}_1 = \mathbf{U}_{\leq 1} (I^{\leq 1}, \cdot)$  is full rank, establishing the result for  $\mathbf{M}_j$  when  $j = 1$ . Now assume that  $\mathbf{M}_{j-1}$  is full rank. Rewriting (52) using (13)-(14) we have

$$\begin{aligned} \widehat{U}_j(\alpha_{j-1}i_j, \alpha_j) &= U_{\leq j-1} \left( I_{\alpha_{j-1}}^{\leq j-1}, \cdot \right) U_j^{(r)}(\cdot, i_j \alpha_j) \\ &= M_{j-1}(\alpha_{j-1}, \cdot) U_j^{(r)}(\cdot, i_j \alpha_j). \end{aligned} \tag{60}$$

By assumption  $M_{j-1}$  is full rank and since  $Y \in \mathcal{M}_r$  both unfoldings of  $U_j$  are full rank. It follows that  $\widehat{U}_j$  is full rank. In (53) the  $r_j$  indices  $l_{\leq j}$  are obtained from  $\widehat{U}_j$  with the DEIM and invoking [49, Lemma 3.1] we have that the  $r_j \times r_j$  matrix  $\widehat{U}_j(l_{\leq j}, \cdot)$  is full rank. The indices  $\mathcal{I}^{\leq j}$  are obtained from  $l_{\leq j}$  in (54) so that

$$U_{\leq j} \left( I_{\alpha_j}^{\leq j}, \beta_j \right) = \widehat{U}_{\leq j} \left( l_{\leq j}(\alpha_j), \beta_j \right), \tag{61}$$

proving the result for  $M_j$ . The statement for  $N_j$  is proven similarly with induction on  $j = d - 1, d - 2, \dots, 1$ .  $\square$

In addition to generating invertible matrices (59), the TT-cross-DEIM is a greedy algorithm that aims to minimize the condition number of these matrices as much as possible while ensuring the indices remain nested. Indeed, the DEIM algorithm is used to compute  $l_{\leq j}$  in (53) and  $l_{> j}$  in (57), selecting indices greedily to keep the condition numbers of  $\widehat{U}_{\leq j}(l_{\leq j}, \cdot)$  and  $\widehat{V}_{> j}(l_{> j}, \cdot)$  small. Moreover,  $\widehat{U}_{\leq j}$  is the restriction of  $U_{\leq j}$  to the indices  $\mathcal{I}^{\leq j-1}$  (see (52)), and  $\widehat{V}_{> j}$  is the restriction of  $V_{> j}$  to the indices  $\mathcal{I}^{> j+1}$  (see (56)). We also note that other sparse sampling methods can be used in place of DEIM in the TT-cross-DEIM algorithm, e.g., Q-DEIM or oversampling methods, which can yield better conditioned interpolatory projectors in some cases.

### 5.3. Tensor cross interpolation

We have shown above that the multi-indices obtained with the TT-cross-DEIM produce a well-defined interpolatory projector (24) onto the tangent space. Incidentally, we can use the same multi-index sets to parameterize  $Y$  with tensor cross interpolation. Recall that TT-cross approximation [14,37,47,40] is a specific instance of the TT format (12) that generalizes the matrix CUR decomposition to tensors. In this representation the TT-cores are defined by the entries of  $Y$

$$\tilde{Y}(i_1, \dots, i_d) = \prod_{k=1}^{d-1} Y(\mathcal{I}^{\leq k-1}, i_k, \mathcal{I}^{> k}) [Y(\mathcal{I}^{\leq k}, \mathcal{I}^{> k})]^{-1} Y(\mathcal{I}^{\leq d-1}, i_d), \tag{62}$$

where for convenience we set  $\mathcal{I}^{\leq 0} = \emptyset$ . It is well-known that the nested condition (38) is sufficient for the tensor cross approximation (62) to be a tensor cross interpolant [47]

$$\tilde{Y}(\mathcal{I}^{\leq k-1}, i_k, \mathcal{I}^{> k}) = Y(\mathcal{I}^{\leq k-1}, i_k, \mathcal{I}^{> k}), \quad k = 1, 2, \dots, d. \tag{63}$$

We now use the nested multi-index sets constructed by the TT-cross-DEIM to prove that any TT can be exactly represented as a TT-cross interpolant. This result follows as a Corollary of Theorem 5.1.

**Corollary 5.1.** Any  $Y \in \mathcal{M}_r$  can be exactly represented as a rank- $r$  TT-cross interpolant with nested indices.

**Proof.** Let  $\{I^{\leq j}, I^{> j}\}$  be nested index sets obtained with the TT-cross-DEIM. Using (16) write the  $r_j \times r_j$  matrix  $Y(I^{\leq j}, I^{> j})$  as

$$\begin{aligned} Y(I^{\leq j}, I^{> j}) &= U_{\leq j}(I^{\leq j}, :) \mathbf{S}_j \mathbf{V}_{> j}(I^{> j}, :)^T \\ &= M_j \mathbf{S}_j \mathbf{N}_j^T, \end{aligned} \tag{64}$$

where we used (59) to obtain the second equality. We have shown in Theorem 5.1 that  $M_j$  and  $N_j$  are invertible and since  $Y \in \mathcal{M}_r$ , the matrices  $\mathbf{S}_j$  are also invertible for all  $j = 1, 2, \dots, d - 1$ . Therefore the matrices  $Y(I^{\leq j}, I^{> j})$  are invertible for all  $j = 1, 2, \dots, d - 1$  and hence by [14, Theorem 2] the nested multi-indices  $\{I^{\leq j}, I^{> k}\}$  provide an exact representation of  $Y$  with TT-cross interpolation.  $\square$

Several algorithms for computing tensor cross approximations (62) from black-box tensors based on the maximum volume principle have recently been developed [37,14]. The purpose of our proposed TT-cross-DEIM algorithm is to obtain interpolatory tangent space projections for a tensor  $Y \in \mathcal{M}_r$ , not for black-box tensor approximation. However, it was recently demonstrated in [20] that DEIM-based cross approximation algorithms can be applied iteratively to obtain tensor cross approximations from black-box tensors with comparable performance to the corresponding maximum volume algorithms.

### 6. Time integration on tensor train manifolds

We now consider the dynamical low-rank evolution equation (3) for tensors ( $d \geq 2$ ) using interpolatory projectors (24) onto tangent spaces of TT manifolds. The concept of dynamical low-rank tensor approximation is a natural extension the dynamical low-rank matrix approximation described in Section 2. Similar to the matrix case, classical dynamical low-rank tensor approximation uses the orthogonal projector (21) to obtain the best approximation (in the Frobenius norm) of  $G(Y, t)$  in the tangent space of the TT manifold (see Fig. 1). However, as noted earlier, orthogonal projection onto the TT tangent space can have a computational cost  $\mathcal{O}(n^d)$  when  $G$  lacks low-rank structure. By replacing the orthogonal projector with an interpolatory projector onto the TT tangent space we propose new dynamical low-rank methods with computational cost scaling as  $\mathcal{O}(dnr^3)$  for a large class of nonlinear functions  $G$  that do not have rank structure. In particular, the proposed interpolatory dynamical low-rank tensor methods are efficient whenever it is possible to evaluate the tensor  $G(X, t)$  entry-wise.

We also point out that cross approximation algorithms based on the maximum volume principle developed in [37] are designed to obtain TT approximations of tensors that can be evaluated entry-wise. TT-cross based on maximum volume can be used to obtain a low-rank approximation of the tensor  $G(Y, t)$  at each time step. Such approximation can

then be projected orthogonally onto the tangent space for a dynamical low-rank method or used in a step-truncation scheme. The TT-cross-DEIM index selection strategy developed in the present work differs in that it selects interpolation indices from the solution tensor  $Y(t)$ , not from  $G(Y, t)$ . Such indices are selected so that  $Y$  can be represented using a TT-cross interpolant. More importantly, it allows  $G(Y, t)$  to be interpolated directly onto the tangent space of the TT manifold at  $Y$ , making the TT-cross-DEIM particularly suitable for efficient dynamical low-rank approximation.

Hereafter we propose two time integration schemes for solving the dynamical low-rank equation (3) with interpolatory TT tangent space projectors (24). The first scheme, referred to as TT-cross time integration, extends the matrix cross integrator described in Section 2.1.1 to tensors in the TT format. This method integrates forward in time the entries of the solution tensor  $Y(t)$  required to construct the TT-cross interpolant (62) at any time  $t$ . The second scheme extends the projector-splitting scheme described in Section 2.1.2 to tensors in the TT format. It is a direct generalization of the projector-splitting integrator for orthogonal dynamical low-rank approximation introduced in [31] for matrices and subsequently generalized to TTs [32], Tucker tensors [30] and tree tensor networks [7], to interpolatory tangent space projectors in the TT format.

### 6.1. TT-cross integrator

The time-dependent interpolatory TT tangent space projector (24) in (3) is defined at each time  $t$  by a set of time-dependent multi-indices  $\{\mathcal{I}^{\leq k}(t), \mathcal{I}^{> k}(t)\}$ . Selecting such with the TT-cross-DEIM ensures that they are nested (38) at each time  $t$ . Hence the tangent space projector has the cross interpolation property (42) at each  $t$ . Evaluating (3) at the multi-indices  $\{\mathcal{I}^{\leq k}(t), \mathcal{I}^{> k}(t)\}$  and utilizing the cross interpolation property yields evolution equations for the entries of  $Y(t)$  defining a TT-cross interpolant

$$\frac{dY(\mathcal{I}^{\leq k-1}(t), :, \mathcal{I}^{> k}(t), t)}{dt} = G_Y(\mathcal{I}^{\leq k-1}(t), :, \mathcal{I}^{> k}(t), t), \quad k = 1, 2, \dots, d, \quad (65)$$

where we defined the tensor  $G_Y(t) = G(Y(t), t)$ . Equation (65) consists of  $\sum_{k=1}^d r_{k-1}n_k r_k$  coupled nonlinear differential equations governing the evolution of a subset of entries in the approximate solution  $Y(t) \in \mathcal{M}_r$ , which can be integrated using standard methods. If  $G$  arises from the spatial discretization of a PDE (1) involving differential operators then evaluating  $G_Y(\mathcal{I}^{\leq k-1}, :, \mathcal{I}^{> k})$  requires entries of  $Y$  at indices adjacent to  $\{\mathcal{I}^{\leq k}, \mathcal{I}^{> k}\}$ . Letting  $\mathcal{I}_{(a)}^{\leq k}, \mathcal{I}_{(a)}^{> k}$  denote the union of  $\mathcal{I}^{\leq k}, \mathcal{I}^{> k}$  with the required adjacent indices, the right-hand side tensors in (65) are computed by

$$G_Y(\mathcal{I}^{\leq k-1}, :, \mathcal{I}^{> k}, t) = G\left(Y\left(\mathcal{I}_{(a)}^{\leq k-1}, :, \mathcal{I}_{(a)}^{> k}, t\right), t\right), \quad k = 1, 2, \dots, d. \quad (66)$$

The values of  $Y$  at adjacent indices can always be obtained by constructing the low-rank solution  $Y(t)$  in the TT format using TT cross interpolation as described below.



Computing (66) is efficient for any nonlinear  $G$  that we can evaluate entry-wise, regardless of the low-rank structure in  $G$ . This gives the evolution equations (65) a clear computational advantage over the evolution equations of orthogonal dynamical low-rank approximation or step-truncation methods [42], which require a low-rank representation of  $G$  to be practical. We also note that the stiffness of the evolution equations (65) is independent of the singular values of the solution tensor  $Y(t)$ , unlike other dynamical low-rank methods. However, the stability of constructing the solution in TT format using cross interpolation, which is often needed to evaluate the right-hand side of (65), depends on the condition of the interpolatory projectors obtained through the TT-cross-DEIM index selection as shown below.

6.1.1. Constructing the low-rank solution in TT format

We can access entries of the approximate solution  $Y(t)$  at indices other than the interpolation indices by constructing  $Y(t)$  using TT-cross interpolation (62)

$$\begin{aligned}
 & Y(i_1, \dots, i_d, t) \\
 &= \prod_{k=1}^{d-1} Y(\mathcal{I}^{\leq k-1}(t), i_k, \mathcal{I}^{>k}(t), t) [Y(\mathcal{I}^{\leq k}(t), \mathcal{I}^{>k}(t), t)]^{-1} Y(\mathcal{I}^{\leq d-1}(t), i_d, t). \tag{67}
 \end{aligned}$$

Such entries are often needed to evaluate the right-hand side of (65) and the TT-representation (67) is also needed to construct indices for interpolatory projection onto the tangent space. Constructing  $Y(t)$  using (67) can lead to numerical instability as the time-dependent  $r_k \times r_k$  matrices  $Y(\mathcal{I}^{\leq k}, \mathcal{I}^{>k}, t)$  can be ill-conditioned. Hereafter we describe a more robust method for computing  $Y(t)$  by orthogonalization, omitting the dependence on  $t$  to simplify notation. Take QR-decompositions

$$[Y(\mathcal{I}^{\leq k-1}, :, \mathcal{I}^{>k})]^{(l)} = \mathbf{Q}_k^{(l)} \mathbf{R}_k, \quad k = 1, 2, \dots, d-1, \tag{68}$$

to write

$$Y(\mathcal{I}^{\leq k-1}, i_k, \mathcal{I}^{>k}) = Q_k(i_k) \mathbf{R}_k, \quad Y(\mathcal{I}^{\leq k}, \mathcal{I}^{>k}) = \mathbf{Q}_k^{(l)}(\mathbf{l}_{\leq k}, :) \mathbf{R}_k, \tag{69}$$

where  $\mathbf{l}_{\leq k}$  is defined in (53). Then substitute (69) into (67) to obtain

$$\begin{aligned}
 Y(i_1, \dots, i_d) &= \prod_{k=1}^{d-1} Q_k(i_k) \mathbf{R}_k \left[ \mathbf{Q}_k^{(l)}(\mathbf{l}_{\leq k}, :) \mathbf{R}_k \right]^{-1} Y(\mathcal{I}^{\leq d-1}, i_d) \\
 &= \prod_{k=1}^{d-1} Q_k(i_k) \left[ \mathbf{Q}_k^{(l)}(\mathbf{l}_{\leq k}, :) \right]^{-1} Y(\mathcal{I}^{\leq d-1}, i_d). \tag{70}
 \end{aligned}$$

Computing  $Y$  via (70) instead of (67) yields a more stable numerical algorithm as the matrices  $\widehat{\mathbf{Q}}_k = \mathbf{Q}_k^{(l)}(\mathbf{l}_{\leq k}, :)$  are related to the orthogonal bases  $\mathbf{U}_{\leq k}$  from which the multi-index sets  $\mathcal{I}^{\leq k}, \mathcal{I}^{>k}$  were obtained with the TT-cross-DEIM index selection algorithm

and therefore have smaller condition number than  $Y(\mathcal{I}^{\leq k}, \mathcal{I}^> k)$ . The improvement in condition number is verified by our numerical experiments as shown in Fig. 5(d).

6.1.2. Discrete-time TT-cross integrator

Let us describe one step of the TT-cross time integration scheme from time  $t_0$  to  $t_1 = t_0 + \Delta t$  starting from the rank- $r$  TT representation

$$Y(t_0) = C_1(t_0)C_2(t_0) \cdots C_d(t_0). \tag{71}$$

First compute the indices  $\{\mathcal{I}^{\leq k}(t_0), \mathcal{I}^> k(t_0)\}$  for the interpolatory projector defining the dynamical low-rank evolution equation (3) using the TT-cross-DEIM algorithm. Then integrate the evolution equations (65) from time  $t_0$  to  $t_1$  using an explicit time stepping scheme with multi-indices fixed at time  $t_0$ , e.g., Euler forward yields

$$Y(\mathcal{I}^{\leq k-1}(t_0), :, \mathcal{I}^> k(t_0), t_1) = Y(\mathcal{I}^{\leq k-1}(t_0), :, \mathcal{I}^> k(t_0), t_0) + \Delta t G_Y(\mathcal{I}^{\leq k-1}(t_0), :, \mathcal{I}^> k(t_0), t_0), \tag{72}$$

for all  $k = 1, 2, \dots, d$ . Use the result of explicit time integration (72) to construct TT-cores for the solution at time  $t_1$

$$Y(t_1) = C_1(t_1)C_2(t_1) \cdots C_d(t_1), \tag{73}$$

with the QR-stabilized procedure described in (68)-(70), i.e.,

$$C_k(i_k, t_1) = Q_k(i_k, t_1) \left[ \mathbf{Q}_k^{(l)}(\mathcal{I}^{\leq k}(t_0), :, t_1) \right]^{-1}, \quad k = 1, 2, \dots, d-1, \tag{74}$$

$$C_d(t_1) = Y(\mathcal{I}^{\leq d-1}(t_0), :, t_1).$$

This completes one step of the TT-cross time integration scheme.

*Computational cost* To simplify the operation count of one step of the TT-cross integrator, we assume that  $r_k = r$  and  $n_k = n$  for all  $k = 1, 2, \dots, d$ . As shown in Section 5, the TT-cross-DEIM algorithm used to obtain the indices  $\{\mathcal{I}^{\leq k}, \mathcal{I}^> k\}$  requires  $\mathcal{O}(dnr^3)$  operations. Preparing the tensors  $Y(\mathcal{I}^{\leq k-1}, : \mathcal{I}^> k)$  required to take the explicit time step (72) requires  $d - 1$  matrix multiplications with matrices of size  $r \times r$  and  $r \times (nr)$  for a number of operations scaling as  $\mathcal{O}(dnr^3)$ . For many  $G$  that can be evaluated entry-wise (e.g., entry-wise nonlinearities), the cost of evaluating  $G_Y(\mathcal{I}^{\leq k-1}, : \mathcal{I}^> k; t)$  is on the same order of computing the subtensors  $Y(\mathcal{I}^{\leq k-1}, : \mathcal{I}^> k)$ , i.e.,  $\mathcal{O}(dnr^3)$ . Finally reconstructing the tensor cores at time  $t_1$  in (74) requires  $d - 1$  QR-decompositions of matrices with size  $r \times (nr)$  and inverting  $d - 1$  matrices of size  $r \times r$  with total cost scaling as  $\mathcal{O}(dnr^3)$ . Thus the computational cost of one step of the TT-cross time integrator scales as  $\mathcal{O}(dnr^3)$ . We note that it is not strictly necessary to perform these steps at every time step as multi-indices can be reused over many time steps, provided the condition of

$\mathbf{Q}_k^{(l)}(\mathbf{l}_{\leq k}, :, t)$  remains under control. If the computation of new indices is not required then (72) can be iterated for a number of time steps at a cost of  $\mathcal{O}(dnr^2)$  operations before computing new indices.

6.2. Interpolatory projector-splitting integrator

Next, we propose a second time integration scheme by directly applying a splitting integrator to the dynamical low-rank evolution equation (3). This method is a direct generalization of the orthogonal projector-splitting integrator introduced in [32] for TTs. As we will see, the oblique projector-splitting integrator satisfies the same desirable properties: it is robust to small singular values, it exactly reproduces low-rank solutions, and one step of the integrator can be implemented as an efficient sweeping algorithm. The derivation of the integrator and the proofs of these results follow the same steps as the orthogonal projector-splitting integrator. Inserting the oblique tangent space projector (24) into the dynamical low-rank evolution equation (3) we see that the right-hand side is a sum of  $2d - 1$  terms

$$P_Y G(Y, t) = \sum_{j=1}^{d-1} P_j^+ G(Y, t) - P_j^- G(Y, t) + P_d^+ G(Y, t), \tag{75}$$

where  $P_j^+ = P_{\leq j-1} P_{> j}$  and  $P_j^- = P_{\leq j} P_{> j}$ . Integrating (3) from time  $t_0$  to  $t_1 = t_0 + \Delta t$  with first order Lie-Trotter splitting requires solving the  $2d - 1$  substeps

$$\begin{aligned} \frac{dY_1^+(t)}{dt} &= P_1^+ G(Y_1^+, t), & Y_1^+(t_0) &= Y(t_0), \\ \frac{dY_1^-(t)}{dt} &= -P_1^- G(Y_1^-, t), & Y_1^-(t_0) &= Y_1^+(t_1), \\ &\vdots & & \\ \frac{dY_j^+(t)}{dt} &= P_j^+ G(Y_j^+, t), & Y_j^+(t_0) &= Y_{j-1}^-(t_1), \\ \frac{dY_j^-(t)}{dt} &= -P_j^- G(Y_j^-, t), & Y_j^-(t_0) &= Y_j^+(t_1), \\ &\vdots & & \\ \frac{dY_d^+(t)}{dt} &= P_d^+ G(Y_d^+, t), & Y_d^+(t_0) &= Y_{d-1}^-(t_1), \end{aligned} \tag{76}$$

in consecutive order to obtain the approximate solution  $Y(t_1) = Y_d^+(t_1)$  at time  $t_1$ . Note that the projectors  $P_j^+, P_j^-$  depend on solutions to each substep  $Y_j^+(t)$  or  $Y_j^-(t)$  and thus are time-dependent. In the case of orthogonal tangent space projector-splitting it was shown in [32, Theorem 4.1] that  $P_j^+, P_j^-$  can be kept constant during each substep and

each of the differential equations (76) can be solved exactly by updating a single TT-core. We have an analogous result for the oblique projector-splitting integrator. In the following Theorem we suppress the dependence of the multi-indices  $\mathcal{I}^{\leq k}, \mathcal{I}^{\gt k}$  on  $t$  although it is assumed that such indices are selected at each time  $t$  so that the interpolatory tangent space projector (24) is well-defined.

**Theorem 6.1.** *Each split differential equation in (76) is solved exactly using time-independent projectors  $P_j^+$  and  $P_j^-$  at  $Y_j^+(t_0)$  and  $Y_j^-(t_0)$ , respectively. Moreover, if  $Y_j^+(t_0)$  has the TT representation*

$$Y_j^+(t_0) = U_{\leq j-1} [U_j \mathbf{S}_j] V_{> j} \tag{77}$$

then

$$Y_j^+(t) = U_{\leq j-1} K_j(t) V_{> j},$$

where

$$\frac{dK_j(t)}{dt} = [U_{\leq j-1} (\mathcal{I}^{\leq j-1}, :)]^{-1} G_j^+ (\mathcal{I}^{\leq j-1}, :, \mathcal{I}^{\gt j}, t) [V_{> j} (:, \mathcal{I}^{\gt j})]^{-T}, \quad K_j(t_0) = U_j \mathbf{S}_j, \tag{78}$$

and  $G_j^+(t) = G(Y_j^+(t), t)$ .

Similarly if  $Y_j^-(t_0)$  has the TT representation  $Y_j^-(t_0) = U_{\leq j} \mathbf{S}_j(t_0) V_{> j}$  then

$$Y_j^-(t) = U_{\leq j} \mathbf{S}_j(t) V_{> j},$$

where

$$\frac{d\mathbf{S}_j(t)}{dt} = - [U_{\leq j} (\mathcal{I}^{\leq j}, :)]^{-1} G_j^- (\mathcal{I}^{\leq j}, \mathcal{I}^{\gt j}, t) [V_{> j} (:, \mathcal{I}^{\gt j})]^{-T}, \tag{79}$$

and  $G_j^-(t) = G(Y_j^-(t), t)$ .

**Proof.** The proof follows a similar approach to the analogous proof for the orthogonal projector-splitting integrator. First recall that we have shown in the proof of Proposition 4.1 that  $P_{\leq j-1} P_{> j}$  maps onto a tangent space of  $\mathcal{M}_r$  at each time  $t$ . This ensures that  $Y_j^+(t)$  belongs to  $\mathcal{M}_r$  for all  $t$  and therefore admits a time-dependent orthogonalized rank- $r$  TT decomposition of the form

$$Y_j^+(t) = U_{\leq j-1}(t) K_j(t) V_{> j}(t). \tag{80}$$

Substituting (80) into (76) and using the product rule we obtain

$$\begin{aligned}
 & \frac{dU_{\leq j-1}(t)}{dt} K_j(t) V_{> j}(t) + U_{\leq j-1}(t) \frac{dK_j(t)}{dt} V_{> j}(t) + U_{\leq j-1}(t) K_j(t) \frac{dV_{> j}(t)}{dt} \\
 & = P_{\leq j-1} P_{> j} G(Y_j^+(t), t) \\
 & = U_{\leq j-1}(t) \delta C_j(t) V_{> j}(t),
 \end{aligned} \tag{81}$$

where we used (29) to obtain the third line with

$$\delta C_j(t) = [U_{\leq j-1}(\mathcal{I}^{\leq j-1}, :, t)]^{-1} G_j^+(\mathcal{I}^{\leq j-1}, :, \mathcal{I}^{> j}, t) [V_{> j}(:, \mathcal{I}^{> j}, t)]^{-T} \tag{82}$$

Equation (81) is solved exactly by setting  $dU_{\leq j-1}(t)/dt = 0$ ,  $dV_{> j}(t)/dt = 0$  and  $dK_j(t)/dt = \delta C_j(t)$  and from the initial condition  $Y_j^+(t_0)$  in (77) we obtain

$$U_{\leq j-1}(t) = U_{\leq j-1}, \quad V_{> j}(t) = V_{> j}, \quad K_j(t_0) = U_j S_j, \tag{83}$$

proving the result for  $Y_j^+(t)$ . The proof of the assertion for  $Y_j^-(t)$  is similar.  $\square$

Similar to the TT-cross evolution equations (65), computing the right-hand side of the differential equation (78) requires evaluating  $G_j^+$  at a subset of  $r_{j-1} n_j r_j$  indices and computing the right-hand side of (79) requires evaluating  $G_j^-$  at a subset of  $r_j^2$  indices. These evaluations are efficient for any  $G$  that can be evaluated entry-wise and do not require  $G$  to have any low-rank structure. The differential equations (78) and (79) involve inverses of  $r_j \times r_j$  matrices  $U_{\leq j}(\mathcal{I}^{\leq j}, :)$  and  $V_{> j}(:, \mathcal{I}^{> j})$ . These matrices define the interpolatory projectors (22) and we select the multi-indices  $\mathcal{I}^{\leq j}, \mathcal{I}^{> j}$  with the TT-cross-DEIM at each time  $t$  to keep their condition number is small during time integration.

### 6.2.1. Sweeping algorithm for interpolatory projector-splitting integrator

One complete step of the interpolatory projector-splitting integrator from time  $t_0$  to  $t_1 = t_0 + \Delta t$  can be implemented by sweeping through the cores of  $Y$  updating individual cores from  $t_0$  to  $t_1$ . As we update the TT-cores we also update the multi-index sets  $\{\mathcal{I}^{\leq j}, \mathcal{I}^{> j}\}$  to ensure the interpolatory projectors remain well-defined. We begin with an orthogonal TT representation of solution at time  $t_0$  of the form

$$Y(t_0) = U_1(t_0) S_1(t_0) V_{> 1}(t_0), \tag{84}$$

and the multi-indices  $\mathcal{I}^{> j}(t_0)$  defining interpolatory projectors onto the bases  $V_{> j}(t_0)$  for  $j = 1, 2, \dots, d - 1$ . The sweeping algorithm solves the equations in (76) sequentially by updating the solution TT-cores  $U_j(t_0)$  to  $U_j(t_1)$  and then computes the indices  $\mathcal{I}^{\leq j}(t_1)$  for the oblique projectors (22) onto the updated bases  $U_{\leq j}(t_1)$  required for the next step in the sweep.

To begin we apply Theorem 6.1 to solve the first differential equation in (76) by integrating

$$\frac{dK_1(t)}{dt} = G_1^+(:, \mathcal{I}^{>1}(t_0), t) [V_{>1}(:, \mathcal{I}^{>1}(t_0), t_0)]^{-T}, \quad K_1(t_0) = U_1(t_0)\mathbf{S}_1(t_0), \quad (85)$$

from  $t_0$  to  $t_1$ . The solution  $Y_1^+(t_1) = K_1(t_1)V_{>1}(t_0)$  is the starting value  $Y_1^-(t_0) = Y_1^+(t_1)$  for the second equation in (76). We then prepare  $Y_1^-(t_0)$  for the application of Theorem 6.1 by decomposing  $K_1(t_1) = U_1(t_1)\mathbf{R}_1(t_1)$  to obtain  $Y_1^-(t_0) = U_1(t_1)\mathbf{R}_1(t_1)V_{>1}(t_0)$  where  $U_1(t_1)$  is left-orthogonal and compute indices  $\mathcal{I}^{\leq 1}(t_1) = \text{DEIM}(U_{\leq 1}(t_1))$ . Now we can apply Theorem 6.1 to solve the second differential equation in (76) by integrating

$$\begin{aligned} \frac{d\mathbf{S}_1(t)}{dt} &= - [U_{\leq 1}(\mathcal{I}^{\leq 1}(t_1), :, t_1)]^{-1} G_1^-(\mathcal{I}^{\leq 1}(t_1), \mathcal{I}^{>1}(t_0), t) [V_{>1}(:, \mathcal{I}^{>1}(t_0), t_0)]^{-T}, \\ \mathbf{S}_1(t_0) &= \mathbf{R}_1(t_1), \end{aligned} \quad (86)$$

from time  $t_0$  to  $t_1$  to obtain the solution  $Y_1^-(t_1) = U_1(t_1)\mathbf{S}_1(t_1)V_{>1}(t_0)$ . The algorithm proceeds recursively with step  $j$  of the sweep described below.

*Computation of  $Y_j^+(t_1)$*  The starting value  $Y_j^+(t_0) = Y_{j-1}^-(t_1)$  is available in the form

$$Y_j^+(t_0) = U_{\leq j-1}(t_1)\mathbf{S}_{j-1}(t_1)V_{>j-1}(t_0), \quad (87)$$

from the computation of  $Y_{j-1}^-(t_1)$ , as are the multi-index sets  $\mathcal{I}^{\leq j-1}(t_1)$  and  $\mathcal{I}^{>j}(t_0)$ . To apply Theorem 6.1 we write (87) as

$$Y_j^+(t_0) = U_{\leq j-1}(t_1) [\mathbf{S}_{j-1}(t_1)V_j(t_0)] V_{>j}(t_0), \quad (88)$$

and then integrate

$$\begin{aligned} \frac{dK_j(t)}{dt} &= [U_{\leq j-1}(\mathcal{I}^{\leq j-1}(t_1), :, t_1)]^{-1} G_j^+(\mathcal{I}^{\leq j-1}(t_1), :, \mathcal{I}^{>j}(t_0), t) [V_{>j}(:, \mathcal{I}^{>j}(t_0), t_0)]^{-T}, \\ K_j(t_0) &= \mathbf{S}_{j-1}(t_1)V_j(t_0), \end{aligned} \quad (89)$$

from  $t_0$  to  $t_1$  to obtain the solution  $Y_j^+(t_1) = U_{\leq j-1}(t_1)K_j(t_1)V_{>j}(t_0)$ .

*Computation of  $Y_j^-(t_1)$*  The starting value  $Y_j^-(t_0) = Y_j^+(t_1)$  is available in the form

$$Y_j^-(t_0) = U_{\leq j-1}(t_1)K_j(t_1)V_{>j}(t_0), \quad (90)$$

from the computation of  $Y_j^+(t_1)$  as are the multi-index sets  $\mathcal{I}^{\leq j-1}(t_1)$  and  $\mathcal{I}^{>j}(t_0)$ . We prepare  $Y_j^-(t_0)$  for the application of Theorem 6.1 by decomposing  $K_j(t_1) =$

$U_j(t_1)\mathbf{R}_j(t_1)$  where  $U_j(t_1)$  is left-orthogonal (see Section 3.1) which allows us to write the starting value (90) as

$$Y_j^-(t_0) = U_{\leq j}(t_1)\mathbf{R}_j(t_1)V_{> j}(t_0). \tag{91}$$

Then we obtain the multi-indices  $\mathcal{I}^{\leq j}(t_1)$  from  $\mathcal{I}^{\leq j-1}(t_1)$  and the TT-cores  $U_{\leq j}(t_1)$  with a substep of the TT-cross-DEIM algorithm as described in (52)-(54). Then by Theorem 6.1, integrating

$$\begin{aligned} \frac{d\mathbf{S}_j(t)}{dt} &= - [U_{\leq j}(\mathcal{I}^{\leq j}(t_1), :, t_1)]^{-1} G_j^-(\mathcal{I}^{\leq j}(t_1), \mathcal{I}^{> j}(t_0), t) [V_{> j}(:, \mathcal{I}^{> j}(t_0), t_0)]^{-T}, \\ \mathbf{S}_j(t_0) &= \mathbf{R}_j(t_1), \end{aligned} \tag{92}$$

from time  $t_0$  to  $t_1$  yields the solution  $Y_j^-(t_1) = U_{\leq j}(t_1)\mathbf{S}_j(t_1)V_{> j}(t_0)$ .

Iterating these steps until we obtain  $Y_d^+(t_1) = U_{\leq d-1}(t_1)K_d(t_1) = Y(t_1)$  completes one step of the first-order splitting integrator. To take another time step the TT representation of  $Y(t_1)$  must be orthogonalized from right to left to obtain an orthogonal representation of the solution at time  $t$  in the form of (15) with  $k = 1$ . During this orthogonalization procedure, the indices  $\mathcal{I}^{> j}(t_1)$  can be computed with the right-to-left TT-cross-DEIM sweep as described in Section 5.1. Similar to the orthogonal projector-splitting integrator, obtaining the second-order Strang projector-splitting integrator is straightforward by composing the Lie-Trotter integrator with its adjoint. In this case the forward sweep described above is performed with step-size  $\Delta t/2$  and is then followed by a backward sweep also with step-size  $\Delta t/2$ . The oblique projector-splitting integrator has the same computational complexity as the TT-cross integrator. Just as with the corresponding matrix integrators described in Section 2, the difference between these two integrators is the order in which interpolatory projection and time integration are performed.

### 6.3. Rank-adaptive time integration

The solution to (2) is often not accurately represented on a tensor manifold  $\mathcal{M}_r$  with constant rank for all  $t \in [0, T]$ . Therefore the dynamical low-rank integrators must be able to decrease or increase the solution rank during time integration. To decrease the solution rank we use the TT-SVD truncation algorithm at each time  $t$  which requires  $d-1$  orthogonal representations (15) of the solution. Such orthogonalizations are required for the TT-cross-DEIM index selection algorithm and thus rank decrease can be performed during time integration with either the TT-cross or interpolatory projector-splitting algorithms at no additional computational cost.

To increase the  $k$ th component of the TT solution rank during integration with the TT-cross integrator we modify Algorithm 2 to sample  $\hat{r}_k > r_k$  indices  $\mathbf{l}_{\leq k}$  from the left

singular vectors (53) and  $\hat{r}_k$  indices  $\mathbf{l}_{>k}$  from right singular vectors in (57) by augmenting the DEIM indices with additional indices selected by another sparse index selection algorithm, e.g., GappyPOD+E [39]. From the  $\mathbf{l}_{\leq k}, \mathbf{l}_{>k}$  we construct  $\mathcal{I}^{\leq k}, \mathcal{I}^{>k}$  in (54), (58) each with  $\hat{r}_k$  indices. We then integrate the solution  $Y(t)$  forward in time on the manifold  $\mathcal{M}_{\hat{r}}$  using the equations (65). It is well-known that the solution  $Y(t)$  with rank  $\mathbf{r}$  belongs to the boundary of the higher rank manifold  $\mathcal{M}_{\hat{r}}$  where the tangent space is not well-defined [51]. Nevertheless, the evolution equations (65), which define the interpolatory tangent space projection, are well-defined on the boundary of  $\mathcal{M}_{\hat{r}}$ . These equations allow us to integrate  $Y(t)$  forward in time on  $\mathcal{M}_{\hat{r}}$  thereby increasing the solution rank. To increase the  $k$ th component of the TT solution rank during integration with the projector-splitting integrator we add new (orthogonal) basis vectors to the TT cores with zero singular value and then sample indices from this augmented basis and apply the projector-splitting integrator to the augmented solution. Once again adding new basis vectors with zero singular values places the approximate solution on the boundary of a higher rank manifold  $\mathcal{M}_{\hat{r}}$ . The projector-splitting integrator is robust to zero singular values and allows us to integrate the solution off of the boundary of the low-rank manifold. A simple criterion for determining when to increase the  $k$ th component of the TT-rank vector is based on the singular values  $\{\sigma_k(\alpha_k, t)\}_{\alpha_k=1}^{r_k}$  of the unfolding matrix  $\mathbf{Y}^{(k)}$ . We select the rank to ensure that the relative size of the smallest singular value

$$\epsilon_k(t) = \frac{\sigma_k(r_k, t)}{\sqrt{\sum_{\alpha_k=1}^{r_k} \sigma_k(\alpha_k, t)^2}}, \quad k = 1, 2, \dots, d - 1 \tag{93}$$

remains in a desired range  $\epsilon_l \leq \epsilon_k(t) \leq \epsilon_u$ . This criterion is an adaptation of the rank-adaptive criterion proposed in [15] for matrix differential equations and subsequently generalized to the Tucker format [21], to the TT format.

### 7. Numerical examples

We now apply the proposed dynamical low-rank collocation methods to several tensor differential equations (2) arising from the discretization of partial differential equations (1) and compare the accuracy and efficiency with existing time integration schemes on tensor manifolds. We measure the accuracy of the low-rank approximations  $Y(t)$  to the solution  $X(t)$  of (2) in the relative Frobenius norm

$$E(t) = \frac{\|Y(t) - X(t)\|_F}{\|X(t)\|_F}. \tag{94}$$

We compute a reference solution  $X(t)$  for each application below by integrating the differential equation (2) with the four-stage explicit Runge-Kutta (RK4) method using time step-size  $\Delta t = 10^{-3}$ .



### 7.1. 2D Vlasov-Poisson equation

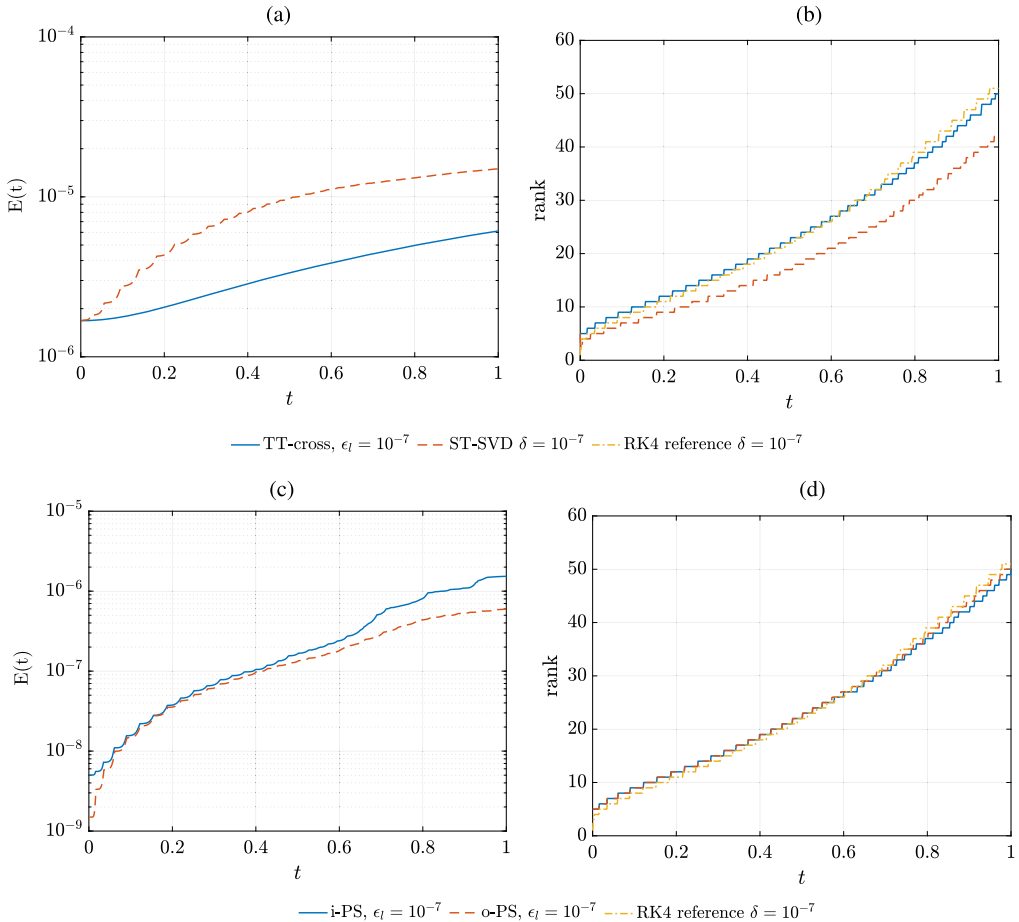
We begin with a two-dimensional example ( $d = 2$ ) demonstrating the proposed methods on low-rank matrix manifolds described in Section 2. We consider the Vlasov-Poisson equation

$$\begin{cases} \frac{\partial u(x, v, t)}{\partial t} + v \frac{\partial u(x, v, t)}{\partial x} + E(x) \frac{\partial u(x, v, t)}{\partial v} = 0 \\ u(x, v, 0) = u_0(x, v), \end{cases} \quad (95)$$

from [24, Example 4.4] with initial condition  $f(x, v, 0) = \exp(-20(x^2 + v^2))$ , electric field  $E(x) = 0.5 \sin(\pi x)$  and  $x \in \Omega_x = [-1, 1]$ ,  $v \in \Omega_v = [-1, 1]$ . Discretizing  $\Omega_x$  and  $\Omega_v$  using  $n = 64$  points and approximating derivatives with a Fourier pseudo-spectral method [26] we obtain a semi-discrete version of the Vlasov-Poisson equation (95) in the form of a differential equation (2) with  $d = 2$ , i.e., a matrix differential equation.

We compared the TT-cross integrator presented in Section 6.1 with a step-truncation method using SVD-based truncation (ST-SVD). For both integrators we used Adams-Bashforth 2 with step-size  $\Delta t = 10^{-3}$ . We utilized the rank-adaptive mechanism described in Section 6.3 for TT-cross with parameter  $\epsilon_l = 10^{-7}$ . For the ST-SVD solution we set relative truncation tolerance  $\delta = 10^{-7}$  at each time step allowing the solution rank to adapt in time accordingly. In Fig. 4(b) we plot the rank of the TT-cross and ST-SVD solutions versus time and the numerical rank of the reference RK4 solution with singular value threshold  $\delta = 10^{-7}$ , i.e., the number of singular values with relative size larger than  $\delta$ . The rank grows rapidly during time integration which allows us to assess the robustness of the rank-adaptive mechanism for the TT-cross integrator. In Fig. 4(a) we plot the relative error of the TT-cross and ST-SVD solutions in the Frobenius norm versus time. We observe that the TT-cross solution is more accurate than the ST-SVD solution due to the TT-cross solution rank being slightly larger than the rank of the ST-SVD solution at each step. The error of the TT-cross solution remains controlled during time integration, demonstrating the effectiveness of the rank-adaptive mechanism.

Next we compared the interpolatory projector-splitting integrator (i-PS) presented in Section 6.2 with the orthogonal projector-splitting integrator (o-PS) introduced in [31]. For the interpolatory and orthogonal projector-splitting integrators we used step-size  $\Delta t = 10^{-3}$  and solved the differential equations in the **K**-, **S**-, and **L**-step with RK4. We also used the rank-adaptive mechanism described in Section 6.3 with parameter  $\epsilon_l = 10^{-7}$  for both solutions. In Fig. 4(b) we plot the solution ranks versus time and the numerical rank of the reference RK4 solution with singular value threshold  $10^{-7}$ . Both solutions have the same rank until approximately  $t = 0.7$  when the i-PS solution rank becomes slightly smaller than the o-PS solution rank. In Fig. 4(a) we plot the relative errors in the Frobenius norm versus time. The error of the i-PS and o-PS solutions is similar until around  $t = 0.4$ , at which point the i-PS solution becomes slightly less accurate. This difference in accuracy is due to the i-PS method computing a quasi-optimal projection



**Fig. 4.** Low-rank approximations to the solution of the two-dimensional Vlasov-Poisson equation (95). (a) Relative error versus time of TT-cross and ST-SVD solutions. The TT-cross solution was computed using rank-adaptive singular value threshold  $\epsilon_t = 10^{-7}$  and the ST-SVD solution was computed using truncation threshold  $\delta = 10^{-7}$ . (b) Rank versus time of the rank-adaptive TT-cross and ST-SVD solutions and the numerical rank of the reference RK4 solution with singular value threshold  $10^{-7}$ . (c) Relative error versus time of solutions computed with interpolatory and orthogonal projector-splitting using rank-adaptive singular value threshold  $\epsilon_t = 10^{-7}$ . (d) Rank versus time of the rank-adaptive solutions computed with interpolatory and orthogonal projector splitting integrators and the numerical rank of the reference RK4 solution with singular value threshold  $10^{-7}$ .

onto the tangent space, while the o-PS method computes the optimal projection at each time step. In addition, the slight difference in rank of the solutions also contributes to the difference in accuracy.

### 7.2. 3D Allen-Cahn equation

The Allen-Cahn equation is a reaction-diffusion PDE that models phase separation in multi-component alloy systems [1,28]. A simple form of such equation features a Laplacian and a cubic non-linearity

$$\begin{cases} \frac{\partial u(\mathbf{x}, t)}{\partial t} = \alpha \Delta u(\mathbf{x}, t) + u(\mathbf{x}, t) - u(\mathbf{x}, t)^3, \\ u(\mathbf{x}, 0) = u_0(\mathbf{x}). \end{cases} \tag{96}$$

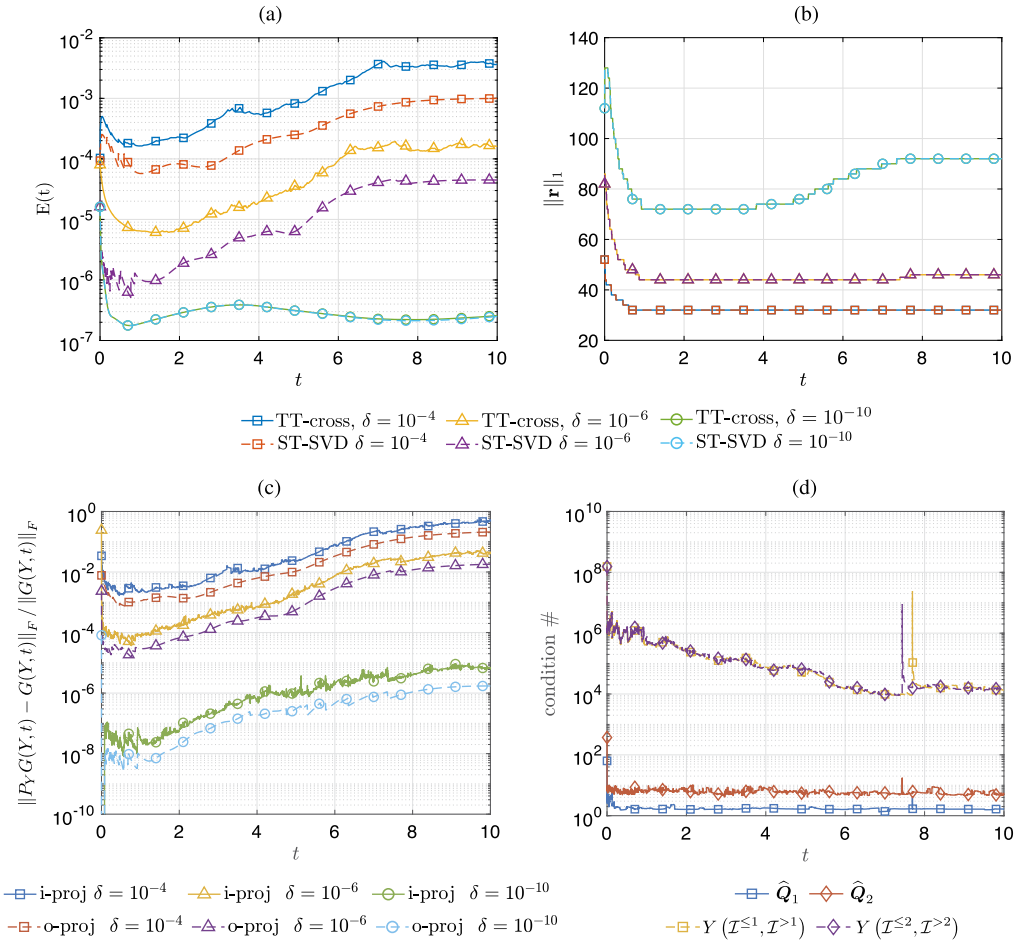
We consider the spatial domain  $\Omega = [0, 2\pi]^3$  with periodic boundary conditions, initial condition  $u_0(x_1, x_2, x_3) = g(x_1, x_2, x_3) - g(2x_1, x_2, x_3) + g(x_1, 2x_2, x_3) - g(x_1, x_2, 2x_3)$  where

$$g(x_1, x_2, x_3) = \frac{(e^{-\tan(x_1)^2} + e^{-\tan(x_2)^2} + e^{-\tan(x_3)^2}) \sin(x_1 + x_2 + x_3)}{1 + e^{|\csc(-x_1/2)|} + e^{|\csc(-x_2/2)|} + e^{|\csc(-x_3/2)|}}, \tag{97}$$

and diffusion parameter  $\alpha = 0.1$ . Discretizing  $\Omega$  using  $n = 64$  points in each dimension and approximating derivatives with a Fourier pseudo-spectral method [26], we obtain a semi-discrete version of the Allen-Cahn equation in the form of (2).

We compared the TT-cross integrator presented in Section 6.1 with the step-truncation SVD (ST-SVD) integrator [42] using different relative truncation tolerances  $\delta = 10^{-3}, 10^{-4}, 10^{-6}, 10^{-10}$  for determining the solution rank at each time step. We set the solution rank in the TT-cross simulations equal to the ranks obtained from the ST-SVD simulations with truncation tolerances in order to compare the methods for solutions computed with the same rank. The rank decrease was performed using TT-SVD truncation and the rank increase by sampling more tensor cross indices than singular vectors using the GappyPOD+E algorithm [39] as described in Section 6.3. Time integration for both ST-SVD and TT-cross was performed with Adams-Bashforth 2 and step-size  $\Delta t = 10^{-3}$ .

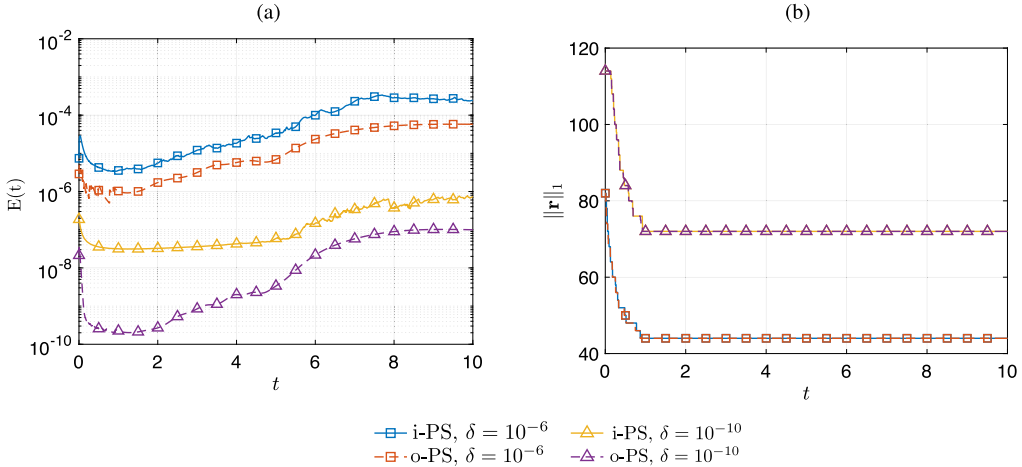
In Fig. 5(b), we plot the 1-norm of the ST-SVD and TT-cross solution ranks. The smoothing effects due to diffusion in the Allen-Cahn equation cause the TT ranks to decay relatively quickly from time  $t = 0$  to time  $t \approx 1.5$ . In Fig. 5(a), we plot the relative error measured in the Frobenius norm of the ST-SVD and TT-cross solutions versus time. The ST-SVD solution is more accurate than the TT-cross solution computed with the same rank, which is expected. Indeed, the ST-SVD method computes the best rank- $r$  projection of the solution onto the low-rank manifold  $\mathcal{M}_r$  at each time step while the TT-cross method computes a quasi-optimal projection onto the tangent space of the manifold at each time step. When the rank of the TT solutions is large enough (in this case corresponding to  $\delta = 10^{-10}$ ), the time integration error dominates the low-rank approximation error and the ST-SVD and TT-cross methods produce solutions with the same accuracy. When the low-rank error dominates the time integration error ( $\delta = 10^{-4}, 10^{-6}$ ) we observe in Fig. 5(a) that the ST-SVD is about half an order of magnitude more accurate than the TT-cross solution of the same rank for all ranks and at each time  $t$ . In Fig. 5(c), we compare the accuracy of the interpolatory projection (i-proj) onto the tangent space computed from the TT-cross solution and the orthogonal projection (o-proj) onto the tangent space computed from the ST-SVD simulation. The orthogonal projection is more accurate than the interpolatory projection by approximately one order of magnitude or less at each time  $t$ . Similar to the difference in error



**Fig. 5.** Low-rank approximations to the solution of the three-dimensional Allen-Cahn equation (96) computed with the TT-cross and ST-SVD methods. The ranks were determined using different truncation tolerances  $\delta = 10^{-4}, 10^{-6}, 10^{-10}$  in the ST-SVD method. (a) Relative error in the Frobenius norm versus time. (b) 1-norm of the TT-rank vector versus time. (c) Relative error of interpolatory (i-proj) and orthogonal (o-proj) projections onto the tensor manifold tangent space versus time. (d) Condition number of the matrices non-orthogonalized matrices in (67) and the corresponding orthogonalized matrices in (70) used to construct the TT-cross solution at each time step.

between the solutions, the difference in error between the i-proj and o-proj is constant over all ranks and for all time  $t$ .

The improved accuracy of the ST-SVD method over the TT-cross method comes at a significant computational cost due to the cubic nonlinearity in the Allen-Cahn equation (96). The reason is that the ST-SVD method requires computing a TT representation of  $G(Y(t), t)$  at each time  $t$ , which is costly. Indeed, recall that standard algorithms for multiplying two TTs  $Y_1$  and  $Y_2$  with ranks  $r_1 = [r_1 \ \cdots \ r_1]$  and  $r_2 = [r_2 \ \cdots \ r_2]$  results in a TT  $Y_1 Y_2$  with rank equal to the Hadamard (element-wise) product of the two ranks  $r_1 \circ r_2$ . These ranks are in general not optimal and to control the TT rank we



**Fig. 6.** Low-rank approximations to the solution of the three-dimensional Allen-Cahn equation (96) computed with the interpolatory and orthogonal projector-splitting integrators. Solutions are truncated at each time  $t$  using tolerances TT-SVD with relative tolerance  $\delta = 10^{-6}, 10^{-10}$ . (a) Relative error in the Frobenius norm versus time. (b) 1-norm of the TT-rank vector versus time.

perform a TT-SVD truncation requiring  $\mathcal{O}(dn(r_1r_2)^3)$  operations. We used two TT-SVD truncations  $\mathfrak{T}_\delta$  with relative accuracy  $\delta$  to compute the cubic term

$$(Y)^3 = \mathfrak{T}_\delta^{\text{svd}} (Y \mathfrak{T}_\delta^{\text{svd}} (YY)), \tag{98}$$

incurring a cost of  $\mathcal{O}(dnr^6)$  operations at each time  $t$ . It is possible to accelerate the computation  $G(Y, t)$  by carrying out sums and products of TTs with approximate low-rank tensor arithmetic, black-box tensor cross approximation [14], or randomized algorithms [8]. However such algorithms introduce additional errors in the low-rank approximation that can be difficult to control. In comparison, the TT-cross integrator does not require  $G(Y, t)$  in a low-rank form and instead evaluates  $G(Y, t)$  at  $\mathcal{O}(dnr^2)$  indices. Thus the computational cost of the cubic nonlinearity for TT-cross is negligible compared to the  $\mathcal{O}(dnr^3)$  cost of the TT-cross-DEIM index selection algorithm and evaluating the subtensors of  $Y(t)$  required to integrate the system of equations (65).

We also compared the interpolatory projector-splitting (i-PS) integrator presented in Section 6.2 with the orthogonal projector-splitting (o-PS) integrator from [32] using two different truncation tolerances  $\delta = 10^{-6}, 10^{-10}$  on the singular values of the solutions. In both cases we used first-order Lie-Trotter splitting with time step-size  $\Delta t = 10^{-3}$  and solved each of the substeps in (76) with RK4. In Fig. 6(a) we plot the error of the solutions computed with the i-PS and o-PS methods versus time. We observe that the i-PS method is less accurate than the o-PS method. This is expected since the i-PS method integrates  $Y(t)$  on  $\mathcal{M}_r$  using a quasi-optimal tensor in the tangent space while the o-PS method integrates uses the optimal tensor in the tangent space. The difference in error is similar to the comparison of TT-cross and ST-SVD except for  $t \in [0, 5]$  in

**Table 1**

CPU-time and accuracy of low-rank methods for integrating the 3D Allen-Cahn equation (96). The ranks were chosen using  $\delta = 10^{-3}$  and  $\delta = 10^{-4}$ .

Method	Average rank $\ \mathbf{r}(t)\ _1$	Runtime (min)	Relative Error ( $t = 10$ )
TT-cross AB2	24.2	4.0	$2.5 \times 10^{-2}$
ST-SVD AB2	24.2	16.7	$7.13 \times 10^{-3}$
i-PS RK4	24.2	23.9	$2.21 \times 10^{-2}$
o-PS RK4	24.2	287.6	$7.13 \times 10^{-3}$
TT-cross AB2	32.4	4.3	$3.6 \times 10^{-3}$
ST-SVD AB2	32.4	27.2	$1.0 \times 10^{-3}$
i-PS RK4	32.4	21.8	$3.6 \times 10^{-3}$
o-PS RK4	32.4	522.1	$1.0 \times 10^{-3}$

the simulations using  $\delta = 10^{-10}$  where the difference in error is significantly larger. In Fig. 6(b) we plot the ranks of the i-PS and o-PS solutions versus time.

In Table 1 we compare the runtime and relative error at time  $t = 10$  of the low-rank solutions computed using existing methods (ST-SVD AB2, o-PS RK4) with the solutions computed using the proposed methods (TT-cross AB2, i-PS RK4). We consider two different rank-adaptive simulations with ranks determined by  $\delta = 10^{-3}, 10^{-4}$  and report the average 1-norm of the rank vector over all time steps. The TT-cross AB2 method with an average rank of 24.2 is approximately 4.2 times faster than the ST-SVD AB2 method at the same rank, while being roughly half an order of magnitude less accurate. With an average rank of 32.4, the TT-cross AB2 method is approximately 6.3 times faster than the ST-SVD AB2 method, while being less than half an order of magnitude less accurate. The speedup observed for the projected RK4 methods is even greater, as these methods require more evaluations of the right-hand side tensor, which includes the cubic nonlinearity. The i-PS RK4 method with an average rank of 24.2 is approximately 12 times faster than the o-PS RK4 method at the same rank, while being roughly half an order of magnitude less accurate. With an average rank of 32.4, the i-PS RK4 method is approximately 24 times faster than the ST-SVD AB2 method, while being less than half an order of magnitude less accurate.

### 7.3. 4D advection-diffusion-reaction equation

Finally we consider the advection-diffusion-reaction (ADR) equation

$$\begin{cases} \frac{\partial u(\mathbf{x}, t)}{\partial t} = \nabla \cdot (\boldsymbol{\mu}_i(\mathbf{x}, t)u(\mathbf{x}, t)) + \sigma \Delta u(\mathbf{x}, t) + R(u) \\ u(\mathbf{x}, 0) = u_0(\mathbf{x}), \end{cases} \tag{99}$$

where  $R(u)$  is a nonlinear reaction term. We consider the spatial domain  $\Omega = [0, 2\pi]^4$  with periodic boundary conditions and set

$$p_0(\mathbf{x}) = \exp(\sin(x_1) \sin(x_2) \sin(x_3) \sin(x_4)), \tag{100}$$

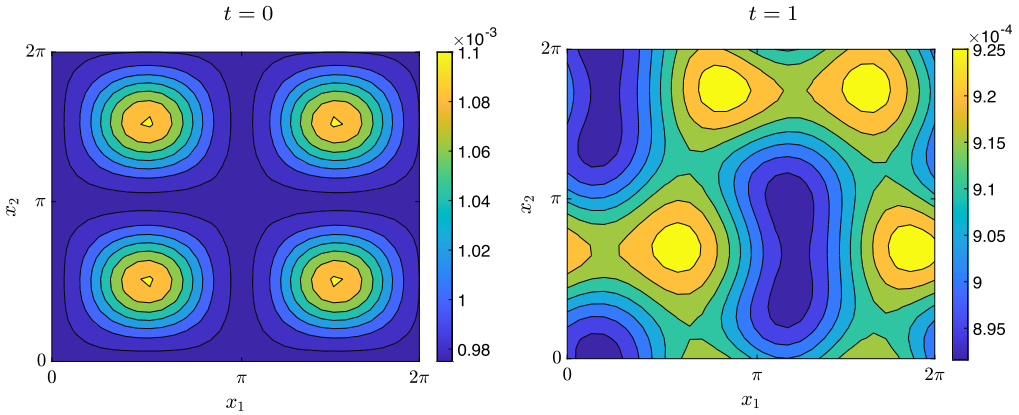


Fig. 7. The  $(x_1, x_2)$ -marginals of the reference solution to the four-dimensional ADR equation (99) at time  $t = 0$  and  $t = 1$ . (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

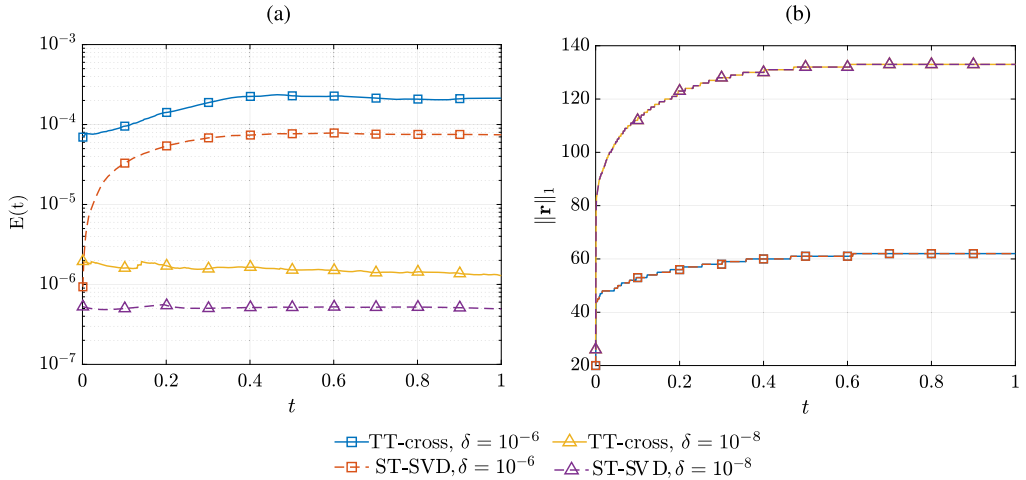
$R(u) = -0.1u/(1 + u^2)$ ,  $\sigma = 1/4$  and

$$\boldsymbol{\mu}(\mathbf{x}) = \frac{1}{2} \begin{bmatrix} g(x_2, x_3) \\ g(x_3, x_4) \\ g(x_4, x_1) \\ g(x_2, x_3) \end{bmatrix}, \tag{101}$$

where  $g(x, y) = \exp(\sin(x) \cos(y))$ . Discretizing  $\Omega$  using  $n = 32$  points in each dimension and approximating derivatives with a Fourier pseudo-spectral method [26] we obtain a semi-discrete version of the ADR equation (99) in the form of (2). In Fig. 7 we plot  $(x_1, x_2)$ -marginals of the RK4 reference solution at time  $t = 0$  and  $t = 1$ .

We computed two approximate low-rank solutions on a TT manifold (17) with the step-truncation SVD method (ST-SVD) [42] using different relative truncation tolerances  $\delta = 10^{-6}, 10^{-8}$ . Computing the ST-SVD solution requires a low-rank approximation of  $G(Y(t), t)$  at each time  $t$ , which is challenging for the nonlinear ADR equation (99) as there are no reliable algorithms available for computing the fractional nonlinearity in the low-rank format. To compute the  $G(Y(t), t)$ , we construct the full tensor representation of the TT-SVD solution with  $n^4$  degrees of freedom, compute the fractional nonlinearity, and then compress the result into a TT with a recursive SVD. This approach is of course not viable in higher dimensions but it allows us to compare our TT-cross solution with the ST-SVD method in this case which computes the best low-rank approximate solution at each time step.

The map  $G$  obtained from discretizing (99) includes four coefficient tensors  $c_1, c_2, c_3, c_4 \in \mathbb{R}^{n \times n \times n \times n}$  (resulting from the discretization of  $g(x, y)$ ) that are not expressed in a low-rank format upon discretization of  $G$ . In order to compute  $G(Y(t), t)$  in low-rank format at each time, we decomposed the four coefficient tensors in  $G$  using TT-SVD compression with relative accuracy  $\delta$ . For  $\delta = 10^{-6}$  and  $\delta = 10^{-8}$  we obtained coefficient tensors of the same rank



**Fig. 8.** Low-rank approximations to the solution of the four-dimensional ADR equation (99) computed with the TT-cross and ST-SVD integrators. The ranks were determined using different truncation tolerances  $\delta = 10^{-6}, 10^{-8}$  in the ST-SVD method. (a) Relative error in the Frobenius norm versus time. (b) 1-norm of the TT-rank vector versus time.

$$\begin{aligned}
 \text{TT-rank}(c_1) &= [1 \quad 1 \quad 12 \quad 1 \quad 1], \\
 \text{TT-rank}(c_2) &= [1 \quad 1 \quad 1 \quad 12 \quad 1], \\
 \text{TT-rank}(c_3) &= [1 \quad 13 \quad 13 \quad 13 \quad 1], \\
 \text{TT-rank}(c_4) &= [1 \quad 1 \quad 12 \quad 1 \quad 1].
 \end{aligned}
 \tag{102}$$

We computed  $G(Y(t), t)$  in the ST-SVD method at each time step by taking products of the low-rank approximate coefficient tensors  $c_k$  with the low-rank solution tensor  $Y$  and then used TT-SVD truncation to compress the product. We then added the TT representation of the reaction term and applied TT-SVD truncation after adding two low-rank tensors in order to control tensor rank when computing  $G(Y(t), t)$  at each time  $t$ . Time integration for the ST-SVD simulation was performed with AB2 and time step-size  $\Delta t = 10^{-3}$ . In Fig. 8(b) we plot the 1-norm of the TT-rank of each ST-SVD solution versus time. We observe that the ranks of both solution increase until around  $t = 0.5$  and then stabilize for  $t \in [0.5, 1]$ .

We then computed two approximate low-rank solutions on the TT manifold  $\mathcal{M}_r$  using the proposed TT-cross integrator. In order to compare the results with the ST-SVD simulations we set the solution ranks in the TT-cross simulations equal to the ranks obtained from the ST-SVD simulations with truncation tolerances  $\delta = 10^{-6}, 10^{-8}$ . We computed the right-hand side of the TT-cross evolution equations (65) by simply evaluating the coefficient tensors at the indices determined by the TT-cross-DEIM Algorithm at each time step. The cost of computing the right hand-hand side for the TT-cross evolution equations is negligible compared to the  $\mathcal{O}(dnr^3)$  cost of the TT-cross-DEIM index selection algorithm and evaluating the subtensors of  $Y(t)$  required to integrate the system of equations (65). In Fig. 8(a) we compare the relative error in the Frobenius



norm of the TT-cross solutions and the ST-SVD solutions. We observe that the TT-cross solutions are less accurate than the ST-SVD solutions of the same rank and the difference in accuracy is constant over all solution ranks and for all time  $t$ . This is expected as the ST-SVD method computes the best rank- $r$  projection of the solution onto the TT manifold  $\mathcal{M}_r$  at each time step while the TT-cross method computes a quasi-optimal projection onto the tangent space of the low-rank manifold at each time step.

## 8. Conclusions

We introduced new general purpose dynamical low-rank methods for solving nonlinear differential equations on low-rank manifolds. The methods rely on a particular class of oblique projectors onto the tangent space of a low-rank manifold with a cross-interpolation property. Such projectors collocate the differential equation on a low-rank tensor manifold and give rise to efficient time integration schemes that allow us to integrate differential equations defined by vector fields that can be evaluated entry-wise. To construct the oblique projections we introduced a new index selection algorithm based on the DEIM for tensors in the TT format. Furthermore, we showed that such index selection algorithm parameterizes low-rank TT manifolds and their tangent spaces with cross interpolation. Our numerical results demonstrate that the oblique projections onto the tangent space yield good approximations on the low-rank manifold in the Frobenius norm that are efficiently computed for problems defined by vector fields without low-rank structure. Our proposed methods thus make dynamical low-rank approximation applicable to a broader class of differential equations and facilitate its use in various practical applications.

## Declaration of competing interest

None.

## Acknowledgements

This material is based upon work supported by the U.S. Department of Energy Office of Science, Office of Advanced Scientific Computing Research, Scientific Discovery through Advanced Computing (SciDAC) program under the contract No. DE-AC02-05CH11231. This research used resources of the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231 using NERSC award ASCR-ERCAP-m1027.

## Data availability

Data will be made available on request.

## References

- [1] S.M. Allen, J.W. Cahn, Ground state structures in ordered binary alloys with second neighbor interactions, *Acta Metall.* 20 (3) (1972) 423–433.
- [2] H. Babae, M. Choi, T.P. Sapsis, G.E. Karniadakis, A robust bi-orthogonal/dynamically-orthogonal method using the covariance pseudo-inverse with application to stochastic flow problems, *J. Comput. Phys.* 344 (2017) 303–319.
- [3] L. Baumann, L. Einkemmer, C. Klingenberg, J. Kusch, Energy stable and conservative dynamical low-rank approximation for the Su–Olson problem, *SIAM J. Sci. Comput.* 46 (2) (2024) B137–B158.
- [4] C. Cercignani, *The Boltzmann Equation and Its Applications*, Springer, 1988.
- [5] G. Ceruti, J. Kusch, C. Lubich, A rank-adaptive robust integrator for dynamical low-rank approximation, *BIT Numer. Math.* 62 (4) (2022) 1149–1174.
- [6] G. Ceruti, C. Lubich, An unconventional robust integrator for dynamical low-rank approximation, *BIT Numer. Math.* 62 (1) (Mar. 2022) 23–44.
- [7] G. Ceruti, C. Lubich, H. Walach, Time integration of tree tensor networks, *SIAM J. Numer. Anal.* 59 (1) (2021) 289–313.
- [8] H.A. Daas, G. Ballard, P. Cazeaux, E. Hallman, A. Miedlar, M. Pasha, T.W. Reid, A.K. Saibaba, Randomized algorithms for rounding in the tensor-train format, *SIAM J. Sci. Comput.* 45 (1) (2023) A74–A95.
- [9] A. Dektor, A. Rodgers, D. Venturi, Rank-adaptive tensor methods for high-dimensional nonlinear PDEs, *J. Sci. Comput.* 88 (36) (2021) 1–27.
- [10] A. Dektor, D. Venturi, Dynamic tensor approximation of high-dimensional nonlinear PDEs, *J. Comput. Phys.* 437 (2021) 110295.
- [11] A. Dektor, D. Venturi, Tensor rank reduction via coordinate flows, *J. Comput. Phys.* 491 (2023) 112378.
- [12] A. Dektor, D. Venturi, Coordinate-adaptive integration of PDEs on tensor manifolds, *Commun. Appl. Math. Comput.* (2024).
- [13] S. Dolgov, D. Kressner, C. Strössner, Functional Tucker approximation using Chebyshev interpolation, *SIAM J. Sci. Comput.* 43 (3) (2021) A2190–A2210.
- [14] S. Dolgov, D. Savostyanov, Parallel cross interpolation for high-precision calculation of high-dimensional integrals, *Comput. Phys. Commun.* 246 (2020) 106869.
- [15] M. Donello, G. Palkar, M.H. Naderi, D.C. Del Rey Fernández, H. Babae, Oblique projection for scalable rank-adaptive reduced-order modelling of nonlinear stochastic partial differential equations with time-dependent bases, *Proc. R. Soc. A, Math. Phys. Eng. Sci.* 479 (2278) (2023) 20230320.
- [16] L. Einkemmer, J. Kusch, S. Schotthöfer, Conservation properties of the augmented basis update & Galerkin integrator for kinetic problems, arXiv preprint arXiv:2311.06399, 2023.
- [17] L. Einkemmer, C. Lubich, A low-rank projector-splitting integrator for the Vlasov–Poisson equation, *SIAM J. Sci. Comput.* 40 (5) (2018) B1330–B1360.
- [18] L. Einkemmer, A. Ostermann, C. Scalone, A robust and conservative dynamical low-rank algorithm, *J. Comput. Phys.* 484 (2023) 112060.
- [19] W. Gangbo, W. Li, S. Osher, M. Puthawala, Unnormalized optimal transport, *J. Comput. Phys.* 399 (2019) 108940.
- [20] B. Ghahremani, H. Babae, Cross interpolation for solving high-dimensional dynamical systems on low-rank Tucker and tensor train manifolds, *Comput. Methods Appl. Mech. Eng.* 432 (2024) 117385.
- [21] B. Ghahremani, H. Babae, A DEIM Tucker tensor cross algorithm and its application to dynamical low-rank approximation, *Comput. Methods Appl. Mech. Eng.* 423 (2024) 116879.
- [22] L. Grasedyck, Hierarchical singular value decomposition of tensors, *SIAM J. Matrix Anal. Appl.* 31 (4) (2010) 2029–2054.
- [23] L. Grasedyck, C. Löbber, Distributed hierarchical svd in the hierarchical Tucker format, *Numer. Linear Algebra Appl.* 25 (6) (2018) e2174.
- [24] W. Guo, J.-M. Qiu, A low rank tensor representation of linear transport and nonlinear Vlasov solutions and their associated flow maps, *J. Comput. Phys.* 458 (2022) 111089.
- [25] W. Guo, J.-M. Qiu, A conservative low rank tensor method for the Vlasov dynamics, *SIAM J. Sci. Comput.* 46 (1) (2024) A232–A263.
- [26] J.S. Hesthaven, S. Gottlieb, D. Gottlieb, *Spectral Methods for Time-Dependent Problems*, Cambridge Monographs on Applied and Computational Mathematics, vol. 21, Cambridge University Press, Cambridge, 2007.
- [27] S. Holtz, T. Rohwedder, R. Schneider, On manifolds of tensors of fixed TT-rank, *Numer. Math.* 120 (4) (2012) 701–731.

- [28] A.-K. Kassam, L.N. Trefethen, Fourth-order time-stepping for stiff PDEs, *SIAM J. Sci. Comput.* 26 (4) (2005) 1214–1233.
- [29] T.G. Kolda, B.W. Bader, Tensor decompositions and applications, *SIAM Rev.* 51 (3) (2009) 455–500.
- [30] C. Lubich, Time integration in the multiconfiguration time-dependent Hartree method of molecular quantum dynamics, *Appl. Math. Res. Express* 2015 (2) (2015) 311–328.
- [31] C. Lubich, I.V. Oseledets, A projector-splitting integrator for dynamical low-rank approximation, *BIT Numer. Math.* 54 (1) (2014) 171–188.
- [32] C. Lubich, I.V. Oseledets, B. Vandereycken, Time integration of tensor trains, *SIAM J. Numer. Anal.* 53 (2) (2015) 917–941.
- [33] C. Lubich, T. Rohwedder, R. Schneider, B. Vandereycken, Dynamical approximation by hierarchical Tucker and tensor-train tensors, *SIAM J. Matrix Anal. Appl.* 34 (2) (2013) 470–494.
- [34] O.A. Malik, S. Becker, Low-rank Tucker decomposition of large tensors using tensorsketch, *Adv. Neural Inf. Process. Syst.* 31 (2018).
- [35] M.H. Naderi, H. Babaei, Adaptive sparse interpolation for accelerating nonlinear stochastic reduced-order modeling with time-dependent bases, *Comput. Methods Appl. Mech. Eng.* 405 (2023) 115813.
- [36] J. Nakao, J. Qiu, L. Einkemmer, Reduced augmentation implicit low-rank (RAIL) integrators for advection-diffusion and Fokker-Planck models, arXiv:2311.15143, 2023, pp. 1–25.
- [37] I. Oseledets, E. Tyrtyshnikov, TT-cross approximation for multidimensional arrays, *Linear Algebra Appl.* 432 (1) (2010) 70–88.
- [38] I.V. Oseledets, Tensor-train decomposition, *SIAM J. Sci. Comput.* 33 (5) (2011) 2295–2317.
- [39] B. Peherstorfer, Z. Drmač, S. Gugercin, Stability of discrete empirical interpolation and gappy proper orthogonal decomposition with randomized and deterministic sampling points, *SIAM J. Sci. Comput.* 42 (5) (2020) A2837–A2864.
- [40] Z. Qin, A. Lidiak, Z. Gong, G. Tang, M.B. Wakin, Z. Zhu, Error analysis of tensor-train cross approximation, *Adv. Neural Inf. Process. Syst.* 35 (2022) 14236–14249.
- [41] H. Risken, *The Fokker-Planck Equation: Methods of Solution and Applications*, second edition, *Mathematics in Science and Engineering*, vol. 60, Springer-Verlag, 1989.
- [42] A. Rodgers, A. Dektor, D. Venturi, Adaptive integration of nonlinear evolution equations on tensor manifolds, *J. Sci. Comput.* 92 (39) (2022) 1–31.
- [43] A. Rodgers, D. Venturi, Implicit integration of nonlinear evolution equations on tensor manifolds, *J. Sci. Comput.* 97 (2) (2023) 33.
- [44] A. Rodgers, D. Venturi, Tensor approximation of functional differential equations, *Phys. Rev. E* 110 (1) (2024) 015310.
- [45] T.P. Sapsis, P.F.J. Lermusiaux, Dynamically orthogonal field equations for continuous stochastic dynamical systems, *Phys. D, Nonlinear Phenom.* 238 (23–24) (2009) 2347–2360.
- [46] D. Savostianova, E. Zangrando, G. Ceruti, F. Tudisco, Robust low-rank training via approximate orthonormal constraints, *Adv. Neural Inf. Process. Syst.* 36 (2024).
- [47] D. Savostyanov, Quasioptimality of maximum-volume cross interpolation of tensors, *Linear Algebra Appl.* 458 (2014) 217–244.
- [48] S. Schotthöfer, E. Zangrando, J. Kusch, G. Ceruti, F. Tudisco, Low-rank lottery tickets: finding efficient low-rank neural networks via matrix differential equations, *Adv. Neural Inf. Process. Syst.* 35 (2022) 20051–20063.
- [49] D.C. Sorensen, M. Embree, A DEIM induced CUR factorization, *SIAM J. Sci. Comput.* 38 (3) (2016) A1454–A1482.
- [50] M. Sutti, B. Vandereycken, Implicit low-rank Riemannian schemes for the time integration of stiff partial differential equations, *J. Sci. Comput.* 101 (1) (2024) 3.
- [51] A. Uschmajew, B. Vandereycken, The geometry of algorithms using hierarchical tensors, *Linear Algebra Appl.* 439 (1) (2013) 133–166.
- [52] D. Venturi, The numerical approximation of nonlinear functionals and functional differential equations, *Phys. Rep.* 732 (2018) 1–102.
- [53] D. Venturi, A. Dektor, Spectral methods for nonlinear functionals and functional differential equations, *Res. Math. Sci.* 8 (27) (2021) 1–39.
- [54] E. Ye, N.F. Loureiro, Quantized tensor networks for solving the Vlasov–Maxwell equations, *J. Plasma Phys.* 90 (3) (2024) 805900301.