A rank-adaptive robust integrator for dynamical low-rank approximation

Gianluca Ceruti, Jonas Kusch, Christian Lubich

KARLSRUHE INSTITUTE OF TECHNOLOGY

CRC 1173

Participating universities

Universität Stuttgart

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

TU
WIEN

Funded by

DFG

A rank-adaptive robust integrator for dynamical low-rank approximation

Gianluca Ceruti, Jonas Kusch, and Christian Lubich

Abstract A rank-adaptive integrator for the dynamical low-rank approximation of matrix and tensor differential equations is presented. The fixed-rank integrator recently proposed by two of the authors is extended to allow for an adaptive choice of the rank, using subspaces that are generated by the integrator itself. The integrator first updates the evolving bases and then does a Galerkin step in the subspace generated by both the new and old bases, which is followed by rank truncation to a given tolerance. It is shown that the adaptive low-rank integrator retains the exactness, robustness and symmetry-preserving properties of the previously proposed fixed-rank integrator. Beyond that, up to the truncation tolerance, the rank-adaptive integrator preserves the norm when the differential equation does, it preserves the energy for Schrödinger equations and Hamiltonian systems, and it preserves the monotonic decrease of the functional in gradient flows. Numerical experiments illustrate the behaviour of the rank-adaptive integrator.

Keywords dynamical low-rank approximation · rank adaptivity · structure-preserving integrator · matrix and tensor differential equations

Mathematics Subject Classification (2000) 65L05 · 65L20 · 65L70 · 15A69

1 Introduction

In [1], a robust integrator for dynamical low-rank approximation of large matrix and tensor differential equations was proposed and analysed. As we show in this paper, that integrator allows for a remarkably simple extension to determine the rank adaptively, while retaining its favourable properties. In a step of the integrator of [1], we first update the bases and then do a Galerkin step in the subspace generated by the *new* bases. In the rank-adaptive integrator proposed here, we do instead a Galerkin step in the larger subspace generated by both the *new and old* bases and

G. Ceruti and Ch. Lubich
Mathematisches Institut, Universität Tübingen, Auf der Morgenstelle 10, D-72076 Tübingen, Germany. E-mail: {ceruti, lubich}@na.uni-tuebingen.de
J. Kusch
Karlsruhe Institute of Technology, Kaiserstr. 12, 76131, Karlsruhe, Germany. E-mail: jonas.kusch@kit.edu

then truncate to a given tolerance. This approach overcomes the problem of how to augment the bases to increase the rank, both in a very simple and very effective way, as will be demonstrated by our numerical experiments.

Strategies of rank adaptivity for dynamical low-rank approximation in the context of the projector-splitting integrator of [18,19] have recently been proposed by Dektor, Rodgers & Venturi [4] and Yang & White [30], and we are aware of ongoing work by Schrammer [28]. Those approaches are substantially different from what is proposed here.

In Section 2 we present the rank-adaptive integrator for matrix differential equations and show that it has the same exactness property and robustness to small singular values as the fixed-rank integrator of [1]. It also preserves symmetry and skew-symmetry of the matrix when the matrix differential equation does.

In Section 3 we show further interesting features that are not available with the integrator of [1]. The following remarkable properties are satisfied in each step *up to the truncation tolerance* or a moderate multiple of it:

- The rank-adaptive integrator applied to a gradient system decreases the functional to be minimized.
- The integrator preserves the norm when the differential equation does.
- The integrator applied to a matrix or tensor Schrödinger equation preserves the energy.
- The integrator applied to a Hamiltonian system (in an appropriate way) preserves the energy.

These near-conservation properties come about by the Galerkin approach and the fact that the projection of the initial value to the augmented bases coincides with the original initial value.

In Section 4 we extend the rank-adaptive integrator to tensor differential equations whose solutions are approximated by Tucker tensors of varying multilinear rank. This is an extension of the fixed-rank tensor integrator of [1] that is analogous to the extension from fixed rank to adaptive rank in the matrix case.

In Section 5 we present numerical experiments with examples from the fields of kinetic equations and uncertainty quantification, where dynamical low-rank approximation has recently found much interest, e.g. in [6–8,25] and [9,22,23,27], beyond the original application area of quantum dynamics, e.g. [21,20] and [11,12].

We describe the integrator for *real* matrices and tensors, but the algorithm and its properties extend in a straightforward way to *complex* matrices and tensors. This only requires care in using transposes $\mathbf{U}^\top$ versus adjoints $\mathbf{U}^* = \overline{\mathbf{U}}^\top$.

Throughout the paper, matrices are written in boldface capital letters and tensors in italic capital letters.

2 A rank-adaptive robust low-rank matrix integrator

Dynamical low-rank approximation of time-dependent matrices [15] approximates the solution $\mathbf{A}(t) \in \mathbb{R}^{m \times n}$ of a large (or too large) matrix differential equation

$$\dot{\mathbf{A}}(t) = \mathbf{F}(t, \mathbf{A}(t)), \qquad \mathbf{A}(t_0) = \mathbf{A}_0 \qquad (1)$$

by evolving matrices $\mathbf{Y}(t) \in \mathbb{R}^{m \times n}$ of low rank, which are computed directly without first computing an approximation to the solution $\mathbf{A}(t)$. The initial low-rank matrix $\mathbf{Y}_0$ is typically obtained from a truncated singular value decomposition (SVD) of $\mathbf{A}_0$. Rank-r matrices are represented in a non-unique factorized SVD-like form

$$\mathbf{Y} = \mathbf{U}\mathbf{S}\mathbf{V}^\top, \tag{2}$$

where the slim matrices $\mathbf{U} \in \mathbb{R}^{m \times r}$ and $\mathbf{V} \in \mathbb{R}^{n \times r}$ each have r orthonormal columns, and the small matrix $\mathbf{S} \in \mathbb{R}^{r \times r}$ is invertible (but not necessarily diagonal).

We present a modification of the fixed-rank integrator of [1] that retains its favourable properties but chooses the rank adaptively. This new integrator computes approximations $\mathbf{Y}_n = \mathbf{U}_n\mathbf{S}_n\mathbf{V}_n^\top \approx \mathbf{A}(t_n)$ of an adaptively determined rank r_n at discrete times t_n $(n = 0, 1, 2, \dots)$. The stepsizes $h_n = t_{n+1} - t_n$ may also vary, but as we will not discuss stepsize selection in this paper, we take a constant stepsize $h > 0$ for notational simplicity.

2.1 Formulation of the algorithm

One time step of integration from time t_0 to $t_1 = t_0 + h$ starting from a factored rank-r_0 matrix $\mathbf{Y}_0 = \mathbf{U}_0\mathbf{S}_0\mathbf{V}_0^\top$ computes an updated factorization $\mathbf{Y}_1 = \mathbf{U}_1\mathbf{S}_1\mathbf{V}_1^\top$ of rank $r_1 \leq 2r_0$. In the following algorithm we let $r = r_0$ and we put a hat on quantities related to rank $2r$.

1. Compute augmented basis matrices $\widehat{\mathbf{U}} \in \mathbb{R}^{m \times 2r}$ and $\widehat{\mathbf{V}} \in \mathbb{R}^{n \times 2r}$ (in parallel):

   K-step: Integrate from $t = t_0$ to t_1 the $m \times r$ matrix differential equation

   $$\dot{\mathbf{K}}(t) = \mathbf{F}(t, \mathbf{K}(t)\mathbf{V}_0^\top)\mathbf{V}_0, \qquad \mathbf{K}(t_0) = \mathbf{U}_0\mathbf{S}_0. \tag{3}$$

   Determine the columns of $\widehat{\mathbf{U}} \in \mathbb{R}^{m \times 2r}$ as an orthonormal basis of the range of the $m \times 2r$ matrix $(\mathbf{K}(t_1), \mathbf{U}_0)$ (e.g. by QR decomposition) and compute the $2r \times r$ matrix $\widehat{\mathbf{M}} = \widehat{\mathbf{U}}^\top\mathbf{U}_0$.

   L-step: Integrate from $t = t_0$ to t_1 the $n \times r$ matrix differential equation

   $$\dot{\mathbf{L}}(t) = \mathbf{F}(t, \mathbf{U}_0\mathbf{L}(t)^\top)^\top\mathbf{U}_0, \qquad \mathbf{L}(t_0) = \mathbf{V}_0\mathbf{S}_0^\top. \tag{4}$$

   Determine the columns of $\widehat{\mathbf{V}} \in \mathbb{R}^{n \times 2r}$ as an orthonormal basis of the range of the $n \times 2r$ matrix $(\mathbf{L}(t_1), \mathbf{V}_0)$ (e.g. by QR decomposition) and compute the $2r \times r$ matrix $\widehat{\mathbf{N}} = \widehat{\mathbf{V}}^\top\mathbf{V}_0$.

2. Augment and update $\mathbf{S}_0 \to \widehat{\mathbf{S}}(t_1)$:

   S-step: Integrate from $t = t_0$ to t_1 the $2r \times 2r$ matrix differential equation

   $$\dot{\widehat{\mathbf{S}}}(t) = \widehat{\mathbf{U}}^\top\mathbf{F}(t, \widehat{\mathbf{U}}\widehat{\mathbf{S}}(t)\widehat{\mathbf{V}}^\top)\widehat{\mathbf{V}}, \qquad \widehat{\mathbf{S}}(t_0) = \widehat{\mathbf{M}}\mathbf{S}_0\widehat{\mathbf{N}}^\top. \tag{5}$$

3. **Truncation**: Compute the SVD $\widehat{\mathbf{S}}(t_1) = \widehat{\mathbf{P}}\widehat{\mathbf{\Sigma}}\widehat{\mathbf{Q}}^\top$ with $\widehat{\mathbf{\Sigma}} = \mathrm{diag}(\sigma_j)$ and truncate to the tolerance ϑ: Choose the new rank $r_1 \leq 2r$ as the minimal number r_1 such that

   $$\left(\sum_{j=r_1+1}^{2r} \sigma_j^2 \right)^{1/2} \leq \vartheta.$$

Compute the new factors for the approximation of $\mathbf{Y}(t_1)$ as follows: Let $\mathbf{S}_1$ be the $r_1 \times r_1$ diagonal matrix with the r_1 largest singular values and let $\mathbf{P}_1 \in \mathbb{R}^{2r \times r_1}$ and $\mathbf{Q}_1 \in \mathbb{R}^{2r \times r_1}$ contain the first r_1 columns of $\widehat{\mathbf{P}}$ and $\widehat{\mathbf{Q}}$, respectively. Finally, set $\mathbf{U}_1 = \widehat{\mathbf{U}}\mathbf{P}_1 \in \mathbb{R}^{m \times r_1}$ and $\mathbf{V}_1 = \widehat{\mathbf{V}}\mathbf{Q}_1 \in \mathbb{R}^{n \times r_1}$.

The approximation after one time step is given by

$$\mathbf{Y}_1 = \mathbf{U}_1 \mathbf{S}_1 \mathbf{V}_1^\top \approx \mathbf{Y}(t_1). \tag{6}$$

Then, $\mathbf{Y}_1$ is taken as the starting value for the next step, which computes $\mathbf{Y}_2$ in factorized form, etc.

The $m \times r$, $n \times r$ and $2r \times 2r$ matrix differential equations in the substeps are solved approximately using a standard integrator, e.g., an explicit or implicit Runge–Kutta method or an exponential integrator when $\mathbf{F}$ is predominantly linear.

The S-step is a Galerkin method for the differential equation (1) in the space of matrices $\widehat{\mathbf{U}}\mathbf{S}\widehat{\mathbf{V}}^\top$ generated by the extended basis matrices $\widehat{\mathbf{U}}$ and $\widehat{\mathbf{V}}$. Note that for $\mathbf{Y}_0 = \mathbf{U}_0\mathbf{S}_0\mathbf{V}_0^\top$, we have the projected starting value $\widehat{\mathbf{U}}\widehat{\mathbf{U}}^\top\mathbf{Y}_0\widehat{\mathbf{V}}\widehat{\mathbf{V}}^\top = \widehat{\mathbf{U}}\widehat{\mathbf{S}}(t_0)\widehat{\mathbf{V}}^\top$. The Galerkin step yields the rank-$2r$ approximation

$$\widehat{\mathbf{Y}}_1 = \widehat{\mathbf{U}}\widehat{\mathbf{S}}(t_1)\widehat{\mathbf{V}}^\top \approx \mathbf{Y}(t_1). \tag{7}$$

The above algorithm differs from the integrator in [1] in that the basis matrix $\widehat{\mathbf{U}}$ not only contains an orthonormal basis of the range of $\mathbf{K}(t_1)$ but is extended to cover also the range of the initial basis $\mathbf{U}_0$, and this is done analogously for $\widehat{\mathbf{V}}$. This extension allows us to increase the rank in a simple and effective way, while we retain the favourable properties of the integrator of [1].

2.2 Exactness property and robust error bound

The new adaptive integrator shares very favourable properties with the low-rank matrix integrators of [1] and [18]. First, it reproduces rank-r matrices exactly.

Theorem 1 (Exactness) *Let $\mathbf{A}(t) \in \mathbb{R}^{m \times n}$ be of rank r for $t_0 \leq t \leq t_1$, so that $\mathbf{A}(t)$ has a factorization (2), $\mathbf{A}(t) = \mathbf{U}(t)\mathbf{S}(t)\mathbf{V}(t)^\top$. Moreover, assume that the $r \times r$ matrices $\mathbf{U}(t_1)^\top\mathbf{U}(t_0)$ and $\mathbf{V}(t_1)^\top\mathbf{V}(t_0)$ are invertible. With $\mathbf{Y}_0 = \mathbf{A}(t_0)$, the rank-adaptive integrator for $\mathbf{F}(t, \mathbf{Y}) = \dot{\mathbf{A}}(t)$ is then exact: $\mathbf{Y}_1 = \mathbf{A}(t_1)$, provided that the truncation tolerance ϑ is smaller than the r-th singular value of $\mathbf{A}(t_1)$.*

Proof We show that the arguments of the exactness proof of [1, Theorem 3] apply with small modifications. As $\widehat{\mathbf{S}}(t_0) = \widehat{\mathbf{U}}^\top\mathbf{U}_0\mathbf{S}_0\mathbf{V}_0^\top\widehat{\mathbf{V}} = \widehat{\mathbf{U}}^\top\mathbf{A}(t_0)\widehat{\mathbf{V}}$, we have by (5)

$$\widehat{\mathbf{S}}(t_1) = \widehat{\mathbf{S}}(t_0) + \int_{t_0}^{t_1} \widehat{\mathbf{U}}^\top\dot{\mathbf{A}}(t)\widehat{\mathbf{V}}\,dt = \widehat{\mathbf{S}}(t_0) + \widehat{\mathbf{U}}^\top(\mathbf{A}(t_1) - \mathbf{A}(t_0))\widehat{\mathbf{V}} = \widehat{\mathbf{U}}^\top\mathbf{A}(t_1)\widehat{\mathbf{V}}.$$

We observe that we can choose $\widehat{\mathbf{U}} = (\widetilde{\mathbf{U}}_1, \widetilde{\mathbf{U}}_0)$, where $\widetilde{\mathbf{U}}_1$ is the orthogonal factor in the QR-decomposition of $\mathbf{K}(t_1)$, and $\widetilde{\mathbf{U}}_0^\top\widetilde{\mathbf{U}}_1 = 0$ by orthogonality. Lemma 1 of [1] shows that $\widetilde{\mathbf{U}}_1$ and $\mathbf{A}(t_1)$ have the same range, or equivalently,

$$\widetilde{\mathbf{U}}_1\widetilde{\mathbf{U}}_1^\top\mathbf{A}(t_1) = \mathbf{A}(t_1).$$

We then also have
$$\widehat{\mathbf{U}}\widehat{\mathbf{U}}^\top \mathbf{A}(t_1) = \mathbf{A}(t_1), \tag{8}$$

because the above equation together with $\widetilde{\mathbf{U}}_0^\top \widetilde{\mathbf{U}}_1 = 0$ implies

$$\begin{aligned}
\widehat{\mathbf{U}}\widehat{\mathbf{U}}^\top \mathbf{A}(t_1) &= \widetilde{\mathbf{U}}_1\widetilde{\mathbf{U}}_1^\top \mathbf{A}(t_1) + \widetilde{\mathbf{U}}_0\widetilde{\mathbf{U}}_0^\top \mathbf{A}(t_1) \\
&= \mathbf{A}(t_1) + \widetilde{\mathbf{U}}_0(\widetilde{\mathbf{U}}_0^\top \widetilde{\mathbf{U}}_1)\widetilde{\mathbf{U}}_1^\top \mathbf{A}(t_1) = \mathbf{A}(t_1).
\end{aligned}$$

We note that (8) still holds true for a different choice of orthonormal basis $\widehat{\mathbf{U}}$, since $\widehat{\mathbf{U}}\widehat{\mathbf{U}}^\top$ is the orthogonal projection onto the range of $(\mathbf{K}(t_1), \mathbf{U}_0)$, which does not depend on the particular choice of the orthonormal basis. In the same way we obtain

$$\mathbf{A}(t_1)\widehat{\mathbf{V}}\widehat{\mathbf{V}}^\top = \mathbf{A}(t_1). \tag{9}$$

By (8) and (9) we then have

$$\widehat{\mathbf{U}}\widehat{\mathbf{S}}(t_1)\widehat{\mathbf{V}}^\top = \widehat{\mathbf{U}}\widehat{\mathbf{U}}^\top \mathbf{A}(t_1)\widehat{\mathbf{V}}\widehat{\mathbf{V}}^\top = \mathbf{A}(t_1).$$

As this matrix is of rank r, its truncation to rank r leaves the result unchanged, and so we obtain the stated exactness result.   □

More importantly, the algorithm is robust to the presence of small singular values of the solution or its approximation, as opposed to standard integrators applied to the differential equations for the factors $\mathbf{U}(t)$, $\mathbf{S}(t)$, $\mathbf{V}(t)$, which contain a factor $\mathbf{S}(t)^{-1}$ on the right-hand sides [15, Prop. 2.1]. The appearance of small singular values is to be expected in an adaptive low-rank approximation, because the smallest singular value retained in the approximation cannot be expected to be much larger than the small truncation tolerance ϑ. The following error bound is independent of small singular values. Such a robust error bound was first shown in [14, Theorem 2.1] for the projector-splitting operator of [18] and subsequently, based on that result, in [1, Theorem 4] for the integrator of [1].

Theorem 2 (Robust error bound) *Let $\mathbf{A}(t)$ denote the solution of the matrix differential equation* (1). *Assume that the following conditions hold in the Frobenius norm $\|\cdot\| = \|\cdot\|_F$:*

1. $\mathbf{F}$ is Lipschitz-continuous and bounded: for all $\mathbf{Y}, \widetilde{\mathbf{Y}} \in \mathbb{R}^{m\times n}$ and $0 \le t \le T$,

$$\|\mathbf{F}(t,\mathbf{Y}) - \mathbf{F}(t,\widetilde{\mathbf{Y}})\| \le L\|\mathbf{Y} - \widetilde{\mathbf{Y}}\|, \qquad \|\mathbf{F}(t,\mathbf{Y})\| \le B .$$

2. The normal part of $\mathbf{F}(t,\mathbf{Y})$ is ε-small at rank r_n for $\mathbf{Y}$ near $\mathbf{A}(t_n)$ and t near t_n: With $\mathrm{P}_{r_n}(\mathbf{Y})$ denoting the orthogonal projection onto the tangent space of the manifold $\mathcal{M}_{r_n}$ of rank-r_n matrices at $\mathbf{Y} \in \mathcal{M}_{r_n}$, it is assumed that

$$\|(\mathbf{I} - \mathrm{P}_{r_n}(\mathbf{Y}))\mathbf{F}(t,\mathbf{Y})\| \le \varepsilon$$

for all $\mathbf{Y} \in \mathcal{M}$ in a neighbourhood of $\mathbf{A}(t_n)$ and t near t_n.
3. The error in the initial value is δ-small:

$$\|\mathbf{Y}_0 - \mathbf{A}_0\| \le \delta.$$

Let $\mathbf{Y}_n$ denote the low-rank approximation to $\mathbf{A}(t_n)$ at $t_n = nh$ obtained after n steps of the adaptive integrator with step-size $h > 0$. Then, the error satisfies for all n with $t_n = nh \leq T$

$$\|\mathbf{Y}_n - \mathbf{A}(t_n)\| \leq c_0\delta + c_1\varepsilon + c_2 h + c_3 n\vartheta,$$

where the constants c_i only depend on $L, B,$ and T. In particular, the constants are independent of singular values of the exact or approximate solution.

The proof of this error bound follows the lines of the proof of [1, Theorem 4] (with modifications of the same type as in the proof of Theorem 1) and is therefore omitted. The result is not fully satisfactory, as it does not show that the error improves when the truncation tolerance is made smaller, as is observed in numerical experiments. To show this, a proportionality relation between ε and ϑ would be needed, which is not available to us. Some of this effect becomes qualitatively plausible by noting that decreasing ϑ increases the rank, which decreases ε.

As in [14, Section 2.6.3], an inexact solution of the matrix differential equations in the rank-adaptive integrator leads to an additional error that is bounded in terms of the local errors in the inexact substeps, again with constants that do not depend on small singular values.

2.3 Symmetric and skew-symmetric low-rank matrices

We now assume that the right-hand side function in (1) is such that one of the following conditions holds,

$$\mathbf{F}(t, \mathbf{Y})^\top = \mathbf{F}(t, \mathbf{Y}^\top) \qquad \text{for all} \quad \mathbf{Y} \in \mathbb{R}^{n \times n} \tag{10}$$

or

$$\mathbf{F}(t, \mathbf{Y})^\top = -\mathbf{F}(t, -\mathbf{Y}^\top) \qquad \text{for all} \quad \mathbf{Y} \in \mathbb{R}^{n \times n}. \tag{11}$$

Under these conditions, solutions to (1) with symmetric or skew-symmetric initial data remain symmetric or skew-symmetric, respectively, for all times.

Theorem 3 *Let $\mathbf{Y}_0 = \mathbf{U}_0\mathbf{S}_0\mathbf{U}_0^\top \in \mathbb{R}^{n \times n}$ be symmetric or skew-symmetric and assume that the function $\mathbf{F}$ satisfies property (10) or (11), respectively. Then, the approximation $\mathbf{Y}_1$ obtained after one time step of the new integrator is symmetric or skew-symmetric, respectively.*

The proof is the same as in [1, Theorem 5].

3 Structure preservation up to the truncation tolerance

3.1 Starting value of the Galerkin step

The following relation, which is not satisfied for the integrator of [1], will be essential in the following subsections. Here we use the notation of (5).

Lemma 1 *Let $\mathbf{Y}_0 = \mathbf{U}_0\mathbf{S}_0\mathbf{V}_0^\top$ and $\widehat{\mathbf{Y}}_0 = \widehat{\mathbf{U}}\widehat{\mathbf{S}}(t_0)\widehat{\mathbf{V}}^\top$. Then, $\widehat{\mathbf{Y}}_0 = \mathbf{Y}_0$.*

Proof We note that by the definition of $\widehat{\mathbf{S}}(t_0)$, we have $\widehat{\mathbf{Y}}_0 = \widehat{\mathbf{U}}\widehat{\mathbf{U}}^\top \mathbf{U}_0 \mathbf{S}_0 \mathbf{V}_0^\top \widehat{\mathbf{V}}\widehat{\mathbf{V}}^\top$. Here, $\widehat{\mathbf{U}}\widehat{\mathbf{U}}^\top$ is the orthogonal projection onto the range of $\widehat{\mathbf{U}}$, which by definition equals the range of $(\mathbf{K}(t_1), \mathbf{U}_0)$. In particular, the columns of $\mathbf{U}_0$ are in the range of $\widehat{\mathbf{U}}$, and hence $\widehat{\mathbf{U}}\widehat{\mathbf{U}}^\top \mathbf{U}_0 = \mathbf{U}_0$. In the same way we also have $\widehat{\mathbf{V}}\widehat{\mathbf{V}}^\top \mathbf{V}_0 = \mathbf{V}_0$. So we obtain $\widehat{\mathbf{Y}}_0 = \mathbf{Y}_0$. $\quad\square$

3.2 Norm preservation

We now turn to a near-conservation property that is not satisfied with the integrator of [1]. If the function $\mathbf{F}$ satisfies

$$\langle \mathbf{Y}, \mathbf{F}(t, \mathbf{Y}) \rangle = 0 \qquad \text{for all } \mathbf{Y} \in \mathbb{R}^{m \times n} \text{ and all } t, \tag{12}$$

then solutions of (1) preserve the Frobenius norm, i.e. $\|\mathbf{A}(t)\| = \|\mathbf{A}(0)\|$ for all t. For the proposed integrator, we show that each step preserves the norm up to the truncation tolerance ϑ, independently of the stepsize h.

Theorem 4 *If* $\mathbf{F}$ *satisfies* (12), *then the numerical result* $\mathbf{Y}_1$ *obtained after a step of the adaptive integrator with the truncation tolerance* ϑ *satisfies*

$$\big| \|\mathbf{Y}_1\| - \|\mathbf{Y}_0\| \big| \le \vartheta.$$

Proof We show that the non-truncated result $\widehat{\mathbf{Y}}_1$ of (7) has the same norm as $\mathbf{Y}_0$. The stated bound then follows from $\big| \|\mathbf{Y}_1\| - \|\widehat{\mathbf{Y}}_1\| \big| \le \|\mathbf{Y}_1 - \widehat{\mathbf{Y}}_1\| \le \vartheta$. We begin by noting that $\|\widehat{\mathbf{Y}}_1\| = \|\widehat{\mathbf{S}}(t_1)\|$ and $\|\mathbf{Y}_0\| = \|\mathbf{S}_0\|$, and further (omitting the argument t after the first equality)

$$\frac{1}{2}\frac{d}{dt}\|\widehat{\mathbf{S}}(t)\|^2 = \langle \widehat{\mathbf{S}}, \dot{\widehat{\mathbf{S}}} \rangle = \langle \widehat{\mathbf{S}}, \widehat{\mathbf{U}}^\top \mathbf{F}(t, \widehat{\mathbf{U}}\widehat{\mathbf{S}}\widehat{\mathbf{V}}^\top)\widehat{\mathbf{V}} \rangle = \langle \widehat{\mathbf{U}}\widehat{\mathbf{S}}\widehat{\mathbf{V}}^\top, \mathbf{F}(t, \widehat{\mathbf{U}}\widehat{\mathbf{S}}\widehat{\mathbf{V}}^\top) \rangle = 0,$$

where we used (12) in the last equality. This yields $\|\widehat{\mathbf{S}}(t_1)\| = \|\widehat{\mathbf{S}}(t_0)\|$. Furthermore, we have $\|\widehat{\mathbf{S}}(t_0)\| = \|\widehat{\mathbf{Y}}_0\|$, which by Lemma 1 equals $\|\mathbf{Y}_0\|$.

Altogether, we then have

$$\|\widehat{\mathbf{Y}}_1\| = \|\widehat{\mathbf{S}}(t_1)\| = \|\widehat{\mathbf{S}}(t_0)\| = \|\widehat{\mathbf{Y}}_0\| = \|\mathbf{Y}_0\|,$$

which yields the result. $\quad\square$

3.3 Gradient systems

Consider a function $f : \mathbb{R}^{m \times n} \to \mathbb{R}$ that is to be minimized. Along every path $\mathbf{A}(t)$ of matrices, we have

$$\frac{d}{dt} f(\mathbf{A}(t)) = \langle \mathbf{G}(\mathbf{A}(t)), \dot{\mathbf{A}}(t) \rangle,$$

where $\langle \mathbf{A}, \mathbf{B} \rangle = \sum_{i,j} a_{ij} b_{ij}$ denotes the inner product that induces the Frobenius norm $\|\cdot\| = \|\cdot\|_F$, and $\mathbf{G}(\mathbf{A}) = \boldsymbol{\nabla} f(\mathbf{A}) \in \mathbb{R}^{m \times n}$ is the gradient. Clearly, f decreases monotonically if $\mathbf{A}(t)$ is a solution of the gradient system

$$\dot{\mathbf{A}}(t) = -\mathbf{G}(\mathbf{A}(t)).$$

When we use the rank-adaptive dynamical low-rank integrator on the gradient system, then the function f decreases along the low-rank approximations $\mathbf{Y}_n$, up to terms of the order of the truncation tolerance ϑ times the gradient norm and errors made in the numerical integration of the S-step (5). More precisely, we show the following.

Theorem 5 *The result $\mathbf{Y}_1$ obtained after a step of the rank-adaptive integrator with the truncation tolerance ϑ applied to the gradient system for the function f satisfies for some $\alpha, \beta \geq 0$*

$$f(\mathbf{Y}_1) \leq f(\mathbf{Y}_0) - \alpha^2 h + \beta\vartheta.$$

For f with a Lipschitz gradient, we have $\alpha = \|\widehat{\mathbf{U}}^\top \mathbf{G}(\mathbf{Y}_0)\widehat{\mathbf{V}}\| + O(h)$ and $\beta = \|\mathbf{G}(\mathbf{Y}_1)\| + O(\vartheta)$.

Proof Along the solution of the differential equation (5) with $\mathbf{F}(\mathbf{Y}) = -\mathbf{G}(\mathbf{Y})$ we have, with $\widehat{\mathbf{Y}}(t) = \widehat{\mathbf{U}}\widehat{\mathbf{S}}(t)\widehat{\mathbf{V}}^\top$,

$$\frac{d}{dt} f(\widehat{\mathbf{Y}}(t)) = \langle \mathbf{G}(\mathbf{Y}(t)), \widehat{\mathbf{U}}\dot{\widehat{\mathbf{S}}}(t)\widehat{\mathbf{V}}^\top \rangle = \langle \widehat{\mathbf{U}}^\top \mathbf{G}(\mathbf{Y}(t))\widehat{\mathbf{V}}, \dot{\widehat{\mathbf{S}}}(t) \rangle$$

$$= \langle \widehat{\mathbf{U}}^\top \mathbf{G}(\mathbf{Y}(t))\widehat{\mathbf{V}}, -\widehat{\mathbf{U}}^\top \mathbf{G}(\mathbf{Y}(t))\widehat{\mathbf{V}} \rangle = -\|\widehat{\mathbf{U}}^\top \mathbf{G}(\mathbf{Y}(t))\widehat{\mathbf{V}}\|^2 \leq -\alpha^2$$

with $\alpha = \min_{0 \leq \tau \leq 1} \|\widehat{\mathbf{U}}^\top \mathbf{G}(\mathbf{Y}(t_0 + \tau h))\widehat{\mathbf{V}}\|$. Since $\widehat{\mathbf{Y}}(t_0) = \mathbf{Y}_0$ by Lemma 1 and $\widehat{\mathbf{Y}}_1 = \widehat{\mathbf{Y}}(t_1)$, we obtain

$$f(\widehat{\mathbf{Y}}_1) \leq f(\widehat{\mathbf{Y}}_0) - \alpha^2 h.$$

Since the truncation is such that $\|\mathbf{Y}_1 - \widehat{\mathbf{Y}}_1\| \leq \vartheta$, we have

$$f(\mathbf{Y}_1) \leq f(\widehat{\mathbf{Y}}_1) + \beta\vartheta$$

with $\beta = \max_{0 \leq \tau \leq 1} \|\mathbf{G}(\tau\mathbf{Y}_1 + (1-\tau)\widehat{\mathbf{Y}}_1)\|$, and so we obtain the stated result. $\square$

The above result does not include the error made in solving the differential equation (5) for $\mathbf{S}$ only approximately. If a step with the implicit Euler method or a discrete gradient method is made, then f still decreases along the numerical solution of this differential equation; cf. e.g. [13]. Alternatively, a higher-order explicit method may give an accurate approximation to $\widehat{\mathbf{S}}(t_1)$ and thus ensure a decrease in f.

3.4 Schrödinger equations

We now turn to the low-rank approximation of the matrix Schrödinger equation

$$\mathrm{i}\dot{\mathbf{A}}(t) = \mathbf{H}[\mathbf{A}(t)]. \tag{13}$$

The Hamiltonian $\mathbf{H} : \mathbb{C}^{m \times n} \to \mathbb{C}^{m \times n}$ is a linear map that is self-adjoint, i.e.,

$$\langle \mathbf{H}[\mathbf{Y}], \mathbf{Z} \rangle = \langle \mathbf{Y}, \mathbf{H}[\mathbf{Z}] \rangle \qquad \text{for all } \mathbf{Y}, \mathbf{Z} \in \mathbb{C}^{m \times n},$$

where the complex inner product is given as $\langle \mathbf{A}, \mathbf{B} \rangle = \sum_{i,j} \overline{a}_{ij} b_{ij}$ so that the induced norm $\|\cdot\|$ is the Frobenius norm of complex matrices. The energy of a state (here: matrix) $\mathbf{Y}$ of norm 1 is

$$E(\mathbf{Y}) = \langle \mathbf{Y}, \mathbf{H}[\mathbf{Y}] \rangle.$$

We are now in a complex setting to which the real rank-adaptive integrator is readily extended: the transposes in the algorithm are replaced by conjugate transposes. The result on norm preservation of the previous subsection applies also here, with essentially the same proof. Remarkably, we also have energy preservation up to the order of the truncation tolerance.

Theorem 6 *The numerical result* $\mathbf{Y}_1$ *obtained after a step of the adaptive integrator with the truncation tolerance* ϑ *applied to the matrix Schrödinger equation* (13) *with* $\mathbf{Y}_0$ *of norm 1 satisfies*

$$\left| E(\mathbf{Y}_1) - E(\mathbf{Y}_0) \right| \leq \gamma \vartheta$$

with $\gamma = \|\mathbf{H}[\mathbf{Y}_1] + \mathbf{H}[\widehat{\mathbf{Y}}_1]\|$.

Proof The proof is similar to that of Theorem 5. Along the solution of the differential equation (5) with $\mathbf{F}(\mathbf{Y}) = -\mathrm{i}\mathbf{H}[\mathbf{Y}]$ we have, with $\widehat{\mathbf{Y}}(t) = \widehat{\mathbf{U}}\widehat{\mathbf{S}}(t)\widehat{\mathbf{V}}^*$,

$$\frac{d}{dt}E(\widehat{\mathbf{Y}}(t)) = 2\operatorname{Re}\langle \mathbf{H}[\widehat{\mathbf{Y}}(t)], \widehat{\mathbf{U}}\dot{\widehat{\mathbf{S}}}(t)\widehat{\mathbf{V}}^*\rangle = 2\operatorname{Re}\langle \widehat{\mathbf{U}}^*\mathbf{H}[\widehat{\mathbf{Y}}(t)]\widehat{\mathbf{V}}, \dot{\widehat{\mathbf{S}}}(t)\rangle$$

$$= 2\operatorname{Re}\langle \widehat{\mathbf{U}}^*\mathbf{H}[\widehat{\mathbf{Y}}(t)]\widehat{\mathbf{V}}, -\mathrm{i}\,\widehat{\mathbf{U}}^*\mathbf{H}[\widehat{\mathbf{Y}}(t)]\widehat{\mathbf{V}}\rangle$$

$$= 2\operatorname{Re}(-\mathrm{i})\|\widehat{\mathbf{U}}^*\mathbf{H}[\widehat{\mathbf{Y}}(t)]\widehat{\mathbf{V}}\|^2 = 0.$$

Since $\widehat{\mathbf{Y}}(t_0) = \mathbf{Y}_0$ by Lemma 1 and $\widehat{\mathbf{Y}}_1 = \widehat{\mathbf{Y}}(t_1)$, we obtain

$$E(\widehat{\mathbf{Y}}_1) = E(\mathbf{Y}_0).$$

Since the truncation is such that $\|\mathbf{Y}_1 - \widehat{\mathbf{Y}}_1\| \leq \vartheta$, we have by the Cauchy–Schwarz inequality

$$|E(\mathbf{Y}_1) - E(\widehat{\mathbf{Y}}_1)| = |\langle \mathbf{Y}_1 - \widehat{\mathbf{Y}}_1, \mathbf{H}[\mathbf{Y}_1 + \widehat{\mathbf{Y}}_1]\rangle| \leq \vartheta\,\|\mathbf{H}[\mathbf{Y}_1] + \mathbf{H}[\widehat{\mathbf{Y}}_1]\|,$$

which yields the stated result. $\quad\square$

3.5 Hamiltonian systems

Given a smooth Hamilton function $H : \mathbb{R}^{m\times n} \times \mathbb{R}^{m\times n} \to \mathbb{R}$, we consider the corresponding Hamiltonian differential equations

$$\dot{\mathbf{Q}} = \boldsymbol{\nabla}_{\mathbf{P}}H(\mathbf{Q},\mathbf{P}), \quad \dot{\mathbf{P}} = -\boldsymbol{\nabla}_{\mathbf{Q}}H(\mathbf{Q},\mathbf{P}). \tag{14}$$

It is not advisable to do dynamical low-rank approximation in the usual way directly on these matrix differential equations. What we propose here, is to rewrite the differential equations (14) in the complex variables

$$\mathbf{Z} = \mathbf{Q} + \mathrm{i}\mathbf{P}, \quad \overline{\mathbf{Z}} = \mathbf{Q} - \mathrm{i}\mathbf{P}$$

with the energy function E defined by

$$\tfrac{1}{2}E(\mathbf{Z},\overline{\mathbf{Z}}) = H(\mathbf{Q},\mathbf{P}),$$

which yields the differential equation in Schrödinger form

$$\mathrm{i}\dot{\mathbf{Z}} = \boldsymbol{\nabla}_{\overline{\mathbf{Z}}}E(\mathbf{Z},\overline{\mathbf{Z}}) \quad (= \boldsymbol{\nabla}_{\mathbf{Q}}H + \mathrm{i}\boldsymbol{\nabla}_{\mathbf{P}}H). \tag{15}$$

We then apply the complex version of the rank-adaptive integrator to this differential equation and finally separate real and imaginary parts to obtain approximations to $\mathbf{Q}(t), \mathbf{P}(t)$. With this approach, we obtain energy conservation up to a multiple of the truncation tolerance ϑ, irrespective of the stepsize h.

Theorem 7 *The result $\mathbf{Z}_1 = \mathbf{Q}_1 + i\mathbf{P}_1$ obtained after a step of the rank-adaptive integrator with the truncation tolerance ϑ applied to the complex system* (15) *with initial value $\mathbf{Z}_0 = \mathbf{Q}_0 + i\mathbf{P}_0$ satisfies*

$$\left| H(\mathbf{Q}_1, \mathbf{P}_1) - H(\mathbf{Q}_0, \mathbf{P}_0) \right| \leq \beta\vartheta,$$

where $\beta = \|\boldsymbol{\nabla} H(\mathbf{Q}_1, \mathbf{P}_1)\| + O(\vartheta)$.

Proof The proof is similar to that of Theorems 5 and 6. Along the solution of the differential equation (5) with $\mathbf{F}(\mathbf{Z}) = -i\boldsymbol{\nabla}_{\bar{\mathbf{Z}}} E(\mathbf{Z}, \overline{\mathbf{Z}})$ we have, with $\widehat{\mathbf{Z}}(t) = \widehat{\mathbf{U}}\widehat{\mathbf{S}}(t)\widehat{\mathbf{V}}^*$,

$$\frac{d}{dt} E(\widehat{\mathbf{Z}}(t), \overline{\widehat{\mathbf{Z}}(t)}) = 2\operatorname{Re}\langle \boldsymbol{\nabla}_{\bar{\mathbf{Z}}} E(\widehat{\mathbf{Z}}(t), \overline{\widehat{\mathbf{Z}}(t)}), \widehat{\mathbf{U}}\dot{\widehat{\mathbf{S}}}(t)\widehat{\mathbf{V}}^*\rangle$$

$$= 2\operatorname{Re}\langle \widehat{\mathbf{U}}^* \boldsymbol{\nabla}_{\bar{\mathbf{Z}}} E(\widehat{\mathbf{Z}}(t), \overline{\widehat{\mathbf{Z}}(t)})\widehat{\mathbf{V}}, \dot{\widehat{\mathbf{S}}}(t)\rangle$$

$$= 2\operatorname{Re}\langle \widehat{\mathbf{U}}^* \boldsymbol{\nabla}_{\bar{\mathbf{Z}}} E(\widehat{\mathbf{Z}}(t), \overline{\widehat{\mathbf{Z}}(t)})\widehat{\mathbf{V}}, -i\widehat{\mathbf{U}}^* \boldsymbol{\nabla}_{\bar{\mathbf{Z}}} E(\widehat{\mathbf{Z}}(t), \overline{\widehat{\mathbf{Z}}(t)})\widehat{\mathbf{V}}\rangle$$

$$= 2\operatorname{Re}(-i)\|\widehat{\mathbf{U}}^* \boldsymbol{\nabla}_{\bar{\mathbf{Z}}} E(\widehat{\mathbf{Z}}(t), \overline{\widehat{\mathbf{Z}}(t)})\widehat{\mathbf{V}}\|^2 = 0.$$

Since $\widehat{\mathbf{Z}}(t_0) = \mathbf{Z}_0$ by Lemma 1, we obtain for $\widehat{\mathbf{Z}}_1 = \widehat{\mathbf{Q}}_1 + i\widehat{\mathbf{P}}_1 = \widehat{\mathbf{Z}}(t_1)$ that

$$H(\widehat{\mathbf{Q}}_1, \widehat{\mathbf{P}}_1) = \tfrac{1}{2} E(\widehat{\mathbf{Z}}_1, \overline{\widehat{\mathbf{Z}}_1}) = \tfrac{1}{2} E(\mathbf{Z}_0, \overline{\mathbf{Z}_0}) = H(\mathbf{Q}_0, \mathbf{P}_0).$$

Since the truncation is such that $\|(\mathbf{Q}_1, \mathbf{P}_1) - (\widehat{\mathbf{Q}}_1, \widehat{\mathbf{P}}_1)\| = \|\mathbf{Z}_1 - \widehat{\mathbf{Z}}_1\| \leq \vartheta$, we obtain

$$\left| H(\mathbf{Q}_1, \mathbf{P}_1) - H(\mathbf{Q}_0, \mathbf{P}_0) \right| = \left| H(\mathbf{Q}_1, \mathbf{P}_1) - H(\widehat{\mathbf{Q}}_1, \widehat{\mathbf{P}}_1) \right| \leq \beta\vartheta$$

with $\beta = \max_{0 \leq \tau \leq 1} \|\boldsymbol{\nabla} H(\tau\mathbf{Q}_1 + (1-\tau)\widehat{\mathbf{Q}}_1, \tau\mathbf{P}_1 + (1-\tau)\widehat{\mathbf{P}}_1)\|$. $\square$

4 A rank-adaptive robust low-rank Tucker tensor integrator

The solution $A(t) \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ of a tensor differential equation

$$\dot{A}(t) = F(t, A(t)), \qquad A(0) = A_0 \tag{16}$$

is approximated by evolving tensors $Y(t) \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ of varying multilinear rank $\mathbf{r} = (r_1, \ldots, r_d)$. Such tensors are represented in the Tucker form [2] and are written in a notation following [16]:

$$Y(t) = C(t) \mathsf{X}_{i=1}^d \mathbf{U}_i(t), \tag{17}$$

$$\text{i.e.,} \quad y_{i_1, \ldots, i_d}(t) = \sum_{j_1, \ldots, j_d} c_{j_1, \ldots, j_d}(t)\, u_{i_1, j_1}(t) \ldots u_{i_d, j_d}(t),$$

where the slim basis matrices $\mathbf{U}_i \in \mathbb{R}^{n_i \times r_i}$ have orthonormal columns and the smaller core tensor $C(t) \in \mathbb{R}^{r_1 \times \cdots \times r_d}$ is of full multilinear rank $\mathbf{r}$.

We present a rank-adaptive modification of the fixed-rank Tucker tensor integrator of [1] that retains its favourable properties. The integrator computes approximations $Y^n = C^n \mathsf{X}_{i=1}^d \mathbf{U}_i^n \approx A(t_n)$ of an adaptively determined rank $\mathbf{r}^n = (r_1^n, \ldots, r_d^n)$ at discrete times t_n $(n = 0, 1, 2, \ldots)$.

4.1 Formulation of the algorithm

One time step of integration from time t_0 to $t_1 = t_0 + h$ starting from a Tucker tensor of multilinear rank $\mathbf{r}^0 = (r_1^0, \ldots, r_d^0)$ in factorized form, $Y^0 = C^0 \mathsf{X}_{i=1}^d \mathbf{U}_i^0$, computes an updated Tucker tensor of multilinear rank $\mathbf{r}^1 = (r_1^1, \ldots, r_d^1)$ in factorized form, $Y^1 = C^1 \mathsf{X}_{i=1}^d \mathbf{U}_i^1$. In the following algorithm we let $\mathbf{r} = \mathbf{r}^0$ and we put a hat on quantities related to rank $2\mathbf{r}$.

1. Compute augmented basis matrices $\widehat{\mathbf{U}}_i \in \mathbb{R}^{n_i \times 2r_i}$ for $i = 1, \ldots, d$ (in parallel):

   Perform a QR factorization of the transposed i-mode matricization of the core tensor:
   $$\mathbf{Mat}_i(C^0)^\top = \mathbf{W}_i \mathbf{S}_i^{0,\top}.$$
   With $\mathbf{V}_i^{0,\top} = \mathbf{W}_i^\top \bigotimes_{j \neq i}^d \mathbf{U}_j^{0,\top} \in \mathbb{R}^{r_i \times n_{\neg i}}$ (which yields $\mathbf{Mat}_i(Y_0) = \mathbf{U}_i^0 \mathbf{S}_i^0 \mathbf{V}_i^{0,\top}$)

   and the matrix function $\mathbf{F}_i(t, \cdot) := \mathbf{Mat}_i \circ F(t, \cdot) \circ Ten_i$, integrate from $t = t_0$ to t_1 the $n_i \times r_i$ matrix differential equation
   $$\dot{\mathbf{K}}_i(t) = \mathbf{F}_i(t, \mathbf{K}_i(t)\mathbf{V}_i^{0,\top})\mathbf{V}_i^0, \qquad \mathbf{K}_i(t_0) = \mathbf{U}_i^0 \mathbf{S}_i^0.$$
   Determine the columns of $\widehat{\mathbf{U}}_i \in \mathbb{R}^{n_i \times 2r_i}$ as an orthonormal basis of the range of the $n_i \times 2r_i$ matrix $(\mathbf{K}_i(t_1), \mathbf{U}_i^0)$ (e.g. by QR decomposition) and compute the $2r_i \times r_i$ matrix $\widehat{\mathbf{M}}_i = \widehat{\mathbf{U}}_i^\top \mathbf{U}_i^0$.

2. Augment and update the core tensor $C^0 \to \widehat{C}(t_1)$:

   Integrate from $t = t_0$ to t_1 the $2r_1 \times \cdots \times 2r_d$ tensor differential equation
   $$\dot{\widehat{C}}(t) = F\left(t, \widehat{C}(t) \mathsf{X}_{i=1}^d \widehat{\mathbf{U}}_i\right) \mathsf{X}_{i=1}^d \widehat{\mathbf{U}}_i^\top, \quad \widehat{C}(t_0) = C^0 \mathsf{X}_{i=1}^d \widehat{\mathbf{M}}_i.$$

3. Truncate to the tolerance ϑ (cf. [3]): Set $C_0 = \widehat{C}(t_1)$. For $i = 1, \ldots, d$ (sequentially), set $\widehat{C}_i = C_{i-1}$, compute the SVD
   $$\mathbf{Mat}_i(\widehat{C}_i) = \widehat{\mathbf{P}}_i \widehat{\mathbf{\Sigma}}_i \widehat{\mathbf{Q}}_i^\top$$
   and choose the new rank $r_i^1 \leq 2r_i$ as the minimal number r_i^1 such that
   $$\left(\sum_{j=r_i^1+1}^{2r_i} \sigma_j^2 \right)^{1/2} \leq \vartheta/d.$$

   Let $\mathbf{\Sigma}_i$ be the $r_i^1 \times r_i^1$ diagonal matrix with the r_i^1 largest singular values of $\widehat{\mathbf{\Sigma}}_i$ and let $\mathbf{P}_i^1 \in \mathbb{R}^{2r_i \times r_i^1}$ and $\mathbf{Q}_i^1$ contain the first r_i^1 columns of $\widehat{\mathbf{P}}_i$ and $\widehat{\mathbf{Q}}_i$, respectively. Tensorize $C_i = Ten_i(\mathbf{\Sigma}_i^1 \mathbf{Q}_i^{1,\top}) \in \mathbb{R}^{r_1^1 \times \cdots \times r_i^1 \times 2r_{i+1} \times \cdots \times 2r_d}$ and set $\mathbf{U}_i^1 = \widehat{\mathbf{U}} \mathbf{P}_i^1 \in \mathbb{R}^{n_i \times r_i^1}$.

With $C^1 = C_d$, the approximation after one time step is then given by
$$Y^1 = C^1 \mathsf{X}_{i=1}^d \mathbf{U}_i^1. \tag{18}$$

To continue in time, we take Y^1 as starting value for the next step and do another step of the integrator, and so on.

4.2 Properties

– The exactness property of Theorem 6 in [1] extends to the rank-adaptive Tucker tensor integrator, as can be shown by combining the proof of that theorem with the arguments in the proof of Theorem 1.

– The robust error bound of Theorem 7 in [1] also extends to the rank-adaptive Tucker tensor integrator, with an extra term $c_3\vartheta$ in the error bound as in Theorem 2.

– The preservation of (anti-)symmetry of Theorem 8 in [1] extends likewise.

– So does the conservation of norm up to the truncation tolerance ϑ of Theorem 4

– and the decrease of the functional in gradient systems.

– Also the near-conservation of energy for Schrödinger equations and Hamiltonian systems extends to the rank-adaptive Tucker tensor integrator.

The proofs of these extensions do not require new arguments beyond those of [1] and of Sections 2 and 3, therefore are omitted.

5 Numerical Experiments

In this section, we present results of different numerical experiments. These numerical simulations are implemented using MATLAB R2019b and JULIA 1.5.2.

5.1 Error behaviour comparison

We compare the error behaviour of the "unconventional" matrix integrator of [1] with the rank-adaptive matrix integrator of Section 2.1. The matrix numerical example of [1, Section 6.2] is considered:

$$\dot{\mathbf{Y}}(t) = -\mathbf{H}[\mathbf{Y}(t)], \quad \mathbf{Y}(t_0) = \mathbf{U}_0\mathbf{S}_0\mathbf{V}_0^\top \in \mathbb{R}^{n\times n},$$

where

$$\mathbf{H}[\mathbf{Y}] = \left(\mathbf{V}_{\cos} - \frac{1}{2}\mathbf{D}\right)\mathbf{Y} + \mathbf{Y}\left(\mathbf{V}_{\cos} - \frac{1}{2}\mathbf{D}\right)^\top \in \mathbb{R}^{n\times n},$$

$$\mathbf{D} = \mathtt{tridiag}(-1,2,-1) \in \mathbb{R}^{n\times n},$$

$$\mathbf{V}_{\cos} := \mathrm{diag}\{1 - \cos(\frac{2\pi j}{n})\}, \quad j = -n/2,\ldots,n/2-1 .$$

The diagonal matrix $\mathbf{S}_0 \in \mathbb{R}^{n\times n}$ has elements $(S_0)_{ii} = 10^{-i}$ for $i = 1,\ldots n$ and the orthonormal matrices $\mathbf{U}_0, \mathbf{V}_0 \in \mathbb{R}^{n\times n}$ are randomly generated.

The reference solution is computed with the MATLAB solver `ode45` and tolerance parameters {'RELTOL', 1E-10, 'ABSTOL', 1E-10} . The differential equations appearing in the substeps of the fixed-rank and adaptive-rank matrix integrators are integrated with a second-order explicit Runge–Kutta method.

We choose $n = 100$, ranks $r = 4, 8$ and final time $T = 0.1$. The tolerance parameter selected for this numerical example is $\vartheta = 10^{-6}$.

The absolute errors $\|Y_n - A(t_n)\|_F$ at final time $t_n = T$ of the approximate solutions for different time-step sizes are shown in Figure 1. The figure illustrates that the new rank-adaptive integrator retains first-order behavior in time and improves the error in the final approximation for smaller time-step sizes. The rank-adaptive integrator approximately doubles the initial rank within this time interval.
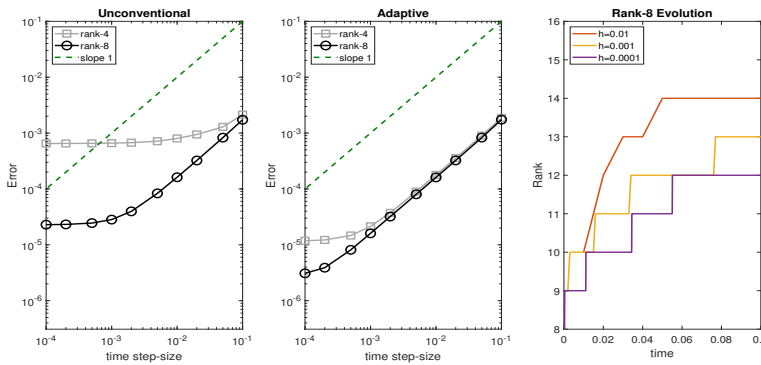
Fig. 1 Comparison of the rank-4 and rank-8 approximations of the fixed-rank integrator of [1] and the new rank-adaptive integrator starting with initial ranks 4 and 8. The error is computed with respect to the reference solution at time $T = 0.1$. The rank evolution of the approximation with initial rank 8 arising from the new matrix adaptive integrator for different time-step sizes is shown to the right.

5.2 Radiation transport equation

In the following numerical example, we consider a one-dimensional radiation transport equation. This equation is a mesoscopic model for the transport and interaction of radiation particles with a background material. For time $t \in [0, T]$ and particle density (or angular flux) $f = f(t, x, \mu)$, the radiation transport equation with isotropic scattering reads

$$\partial_t f + \mu \partial_x f + \sigma_s f = \frac{\sigma_s}{2} \int_{-1}^1 f \, d\mu, \qquad (x, \mu) \in [a, b] \times [-1, 1],$$

$$f(t_0) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{x^2}{2\sigma^2}\right).$$

(19)

The chosen initial condition is a Gaussian with constant deviation $\sigma = 3 \cdot 10^{-2}$. Hence, particles are initially positioned around $x = 0$ and move into directions $\mu \in [-1, 1]$. The reference solution to this problem is given by the *standard de facto* Ganapol's benchmark test [10] and this problem has been investigated for dynamical low-rank approximations in [25, 24]. As time increases, the scalar flux $\Phi(t, x) = \int_{-1}^1 f(t, x, \mu) \, d\mu$ moves to the left and right side of the spatial domain, showing a discontinuous (or shock) profile at the front. When particles interact with the background material through collisions, which is the case for $\sigma_s > 0$, the shock decreases over time and finally yields a smooth profile. In this work, we choose a scattering cross-section of $\sigma_s = 1$.

The physical domain is discretized with a Lax–Friedrichs method in combination with a Legendre-polynomial expansion in the variable μ.

A number of $N + 1$ Legendre-polynomials is used and we split the space-interval $[a, b]$ in N_x sub-intervals. Time-integration for the sub-steps of the adaptive integrator is performed with a first-order Runge-Kutta method and prescribed CFL number.
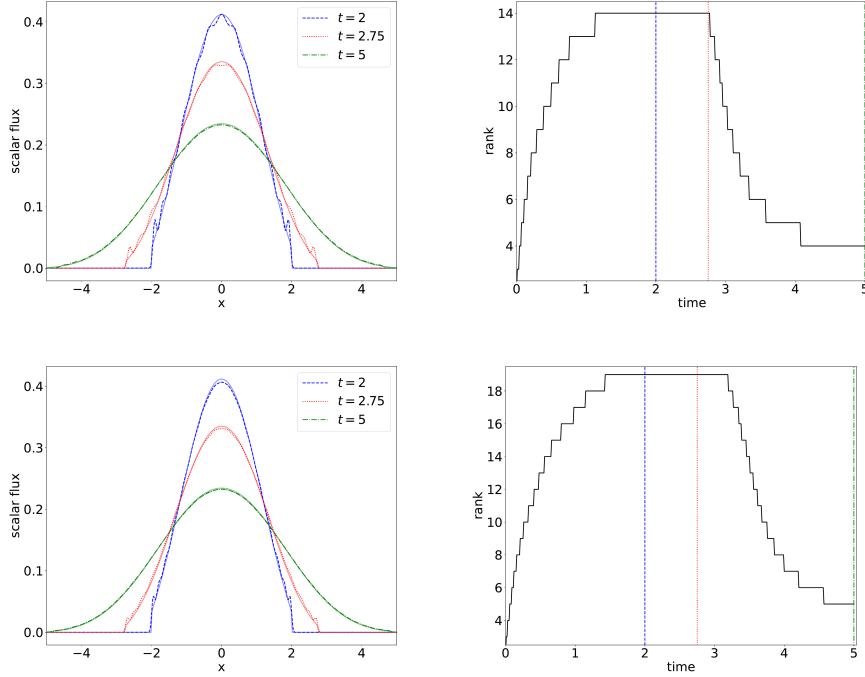
Fig. 2 Top: Tolerance parameter $\vartheta = 10^{-1}\|\hat{\boldsymbol{\Sigma}}\|_2$. Bottom: Tolerance parameter $\vartheta = 5 \cdot 10^{-2}\|\hat{\boldsymbol{\Sigma}}\|_2$. Left: Scalar flux of the reference solution (solid lines) of the radiation transport equation (19) in comparison with the approximation of the new rank-adaptive matrix integrator at different times. Right: Rank evolution of the approximation arising from the rank-adaptive integrator.

The spatial domain has boundaries $a = -5$, $b = 5$ and is discretized with $N_x = 1000$ spatial cells. The polynomial representation of the scalar flux uses $N + 1 = 200$ Legendre polynomials. A time step size is chosen with a CFL number of 0.99. The tolerance parameter is set to $\vartheta = 10^{-1}\|\hat{\boldsymbol{\Sigma}}\|_2$ and $\vartheta = 5 \cdot 10^{-2}\|\hat{\boldsymbol{\Sigma}}\|_2$. Here, the matrix $\hat{\boldsymbol{\Sigma}}$ arises from the SVD-factorization of the solution of the S-step computed with the adaptive matrix unconventional integrator, as illustrated in the last truncation step of the algorithm proposed in Section 2.1.

In the left column of Figure 2, we consider the scalar flux of the reference solution at times $T \in \{2, 2.75, 5\}$ compared to the angular flux of a rank-r approximation generated by the rank-adaptive matrix integrator of Section 2.1. In the right column, the rank evolution is plotted. The top row depicts the scalar flux and rank for a tolerance parameter $\vartheta = 10^{-1}\|\hat{\boldsymbol{\Sigma}}\|_2$ and the bottom row uses $\vartheta = 5 \cdot 10^{-2}\|\hat{\boldsymbol{\Sigma}}\|_2$. Solid lines show the analytic solution computed according to [10]. Our numerical experiment shows that decreasing the tolerance parameter improves the solution quality by increasing the chosen rank. However, the overall characteristics remain unchanged: While the rank of the initial condition is $r = 1$, the adaptive algorithm increases the rank automatically. This increased rank is beneficial to capture the shock profile of the solution in the beginning. As the shock dissolves, the rank is

reduced. We observe that the approximation arising from the adaptive method shows good agreement with the reference solution.

5.3 Uncertainty quantification

The following numerical experiment investigates Burgers' equation with uncertain initial condition

$$
\begin{cases}
\partial_t u(t, x, \boldsymbol{\xi}) + \partial_x \frac{u(t,x,\boldsymbol{\xi})^2}{2} = 0, & (x, \boldsymbol{\xi}) \in [a, b] \times \Theta\,, \\
u(t_0, x, \boldsymbol{\xi}) = u_{\mathrm{IC}}(x, \boldsymbol{\xi})\,.
\end{cases}
\tag{20}
$$

A uniformly distributed random vector $\boldsymbol{\xi} \in \Theta = [-1, 1] \times [0, 1]$ is used to model uncertainties in the initial condition. Since the dynamics of this non-linear hyperbolic model mimics advection effects that arise in gas dynamics, Burgers' equation is a standard model to test numerical methods. Numerical and analytic investigations of Burgers' equation with uncertain initial condition can for example be found in [26, 29,5]. To demonstrate the effects of varying smoothness in time, we choose an initial condition of

$$
u_{\mathrm{IC}}(x, \boldsymbol{\xi}) = \begin{cases}
u_L, & \text{if } x < x_0 + \sigma_1 \xi_1 \\
u_L + \frac{u_R(\xi_2) - u_L}{x_0 - x_1}(x_0 + \sigma_1 \xi_1 - x), & \text{if } x \in [x_0 + \sigma_1 \xi_1, x_1 + \sigma_1 \xi_1] \\
u_R(\xi_2), & \text{else}
\end{cases}\,,
$$

with $u_R(\xi_2) = u_R + \sigma_2 \xi_2$. Note that this intial condition is similar to [26,17], when additionally assuming an uncertain right state u_R to increase computational complexity. At time $t = 0$ the solution is a ramp or forming shock ranging from $x_0 + \sigma_1 \xi_1$ to $x_1 + \sigma_1 \xi_1$. Since the left state u_L moves faster than the right state u_R to the right side of the domain, a shock will form over time. The time at which the shock has fully developed depends on ξ_2 and is given by $t_s = \frac{x_1 - x_0}{u_L - u_R(\xi_2)}$. We use the following parameter values:

$[a, b] = [0, 1]$	range of spatial domain
$T = 0.04$	end time
$N_x = 600$	number of spatial cells
$x_0 = 0.3, x_1 = 0.4, u_L = 12, u_R = 1$	parameters of initial condition
$\sigma_1 = 0.2, \sigma_2 = 5$	parameters of the uncertainty
$\vartheta = \{1, 1.2, 1.5\} \cdot 10^{-2} \|\hat{\boldsymbol{\Sigma}}\|_2$	tolerance parameter

A spatial discretization of equation (20) is performed by a first order finite volume method with Lax–Friedrichs numerical fluxes. As time discretization an explicit Euler method is chosen. The stabilization of the finite volume method is applied in the K, L and S-steps. We do not split the uncertain domain and choose a modal representation of the uncertain basis making use of tensorized Legendre polynomials. For each uncertainty, Legendre polynomials up to degree 19, i.e., $20^2 = 400$ polynomials, are used to represent the uncertain basis.

Numerical results for this testcase are presented in Figure 3. An analytic solution for given values of $\boldsymbol{\xi}$ is determined with characteristics. Expectation and variance are computed by using a fine tensorized quadrature rule with $200 \cdot 200$ Gauss-Legendre quadrature points. The resulting expectation is depicted in red and the corresponding standard deviation is shown in blue. The rank-adaptive method proposed in

Section 2.1 shows satisfactory agreement with the analytic solution, especially for the expectation. Initially, the rank is chosen as 40 and the method reduces this rank after the first time step. As time increases, a shock (or discontinuity) forms in both, the spatial and uncertain domain. The adaptive method captures the growing solution complexity by increasing the rank. After a certain time, the method remains at a fixed higher rank for all chosen tolerance parameters, where the rank depends on the chosen tolerance. Note that this rank will be reached after the shock has fully developed. This is most likely due to the sharpening of the numerical solution which results from increasing the rank. As a result, the singular values of the $\mathbf{S}$ matrix continue to grow.

To point out differences to the fixed-rank integrator, we include a comparison of the rank-adaptive integrator with tolerance parameter $\vartheta = 0.015$ with numerical solutions for fixed rank 9 and 25. These ranks are the minimal and maximal rank chosen by the rank-adaptive integrator during the computation. The numerical solution of the rank-adaptive algorithm shows good agreement with the fixed-rank integrator when using a constant rank of 25. In comparison to a fixed rank of 9, the adaptive method yields a strongly improved solution quality.

References

1. G. Ceruti and C. Lubich. An unconventional robust integrator for dynamical low-rank approximation. *arXiv preprint arXiv:2010.02022*, 2020.
2. L. De Lathauwer, B. De Moor, and J. Vandewalle. A multilinear singular value decomposition. *SIAM J. Matrix Anal. Appl.*, 21(4):1253–1278, 2000.
3. L. De Lathauwer, B. De Moor, and J. Vandewalle. On the best rank-1 and rank-$(R_1, R_2, \cdots, R_N)$ approximation of higher-order tensors. *SIAM J. Matrix Anal. Appl.*, 21(4):1324–1342, 2000.
4. A. Dektor, A. Rodgers, and D. Venturi. Rank-adaptive tensor methods for high-dimensional nonlinear PDEs. *arXiv preprint arXiv:2012.05962*, 2020.
5. B. Després, G. Poëtte, and D. Lucor. Robust uncertainty propagation in systems of conservation laws with the entropy closure method. In *Uncertainty quantification in computational fluid dynamics*, pages 105–149. Springer, 2013.
6. L. Einkemmer, J. Hu, and L. Ying. An efficient dynamical low-rank algorithm for the Boltzmann-BGK equation close to the compressible viscous flow regime. *arXiv preprint arXiv:2101.07104*, 2021.
7. L. Einkemmer and I. Joseph. A mass, momentum, and energy conservative dynamical low-rank scheme for the Vlasov equation. *arXiv preprint arXiv:2101.12571*, 2021.
8. L. Einkemmer and C. Lubich. A low-rank projector-splitting integrator for the Vlasov–Poisson equation. *SIAM J. Sci. Comput.*, 40(5):B1330–B1360, 2018.
9. F. Feppon and P. F. Lermusiaux. Dynamically orthogonal numerical schemes for efficient stochastic advection and Lagrangian transport. *SIAM Rev.*, 60(3):595–625, 2018.
10. B. D. Ganapol. Analytical benchmarks for nuclear engineering applications. *Case Studies in Neutron Transport Theory*, 2008.
11. J. Haegeman, J. I. Cirac, T. J. Osborne, I. Pižorn, H. Verschelde, and F. Verstraete. Time-dependent variational principle for quantum lattices. *Phys. Rev. Letters*, 107(7):070601, 2011.
12. J. Haegeman, C. Lubich, I. Oseledets, B. Vandereycken, and F. Verstraete. Unifying time evolution and optimization with matrix product states. *Phys. Rev. B*, 94(16):165116, 2016.
13. E. Hairer and C. Lubich. Energy-diminishing integration of gradient systems. *IMA J. Numer. Anal.*, 34(2):452–461, 2014.
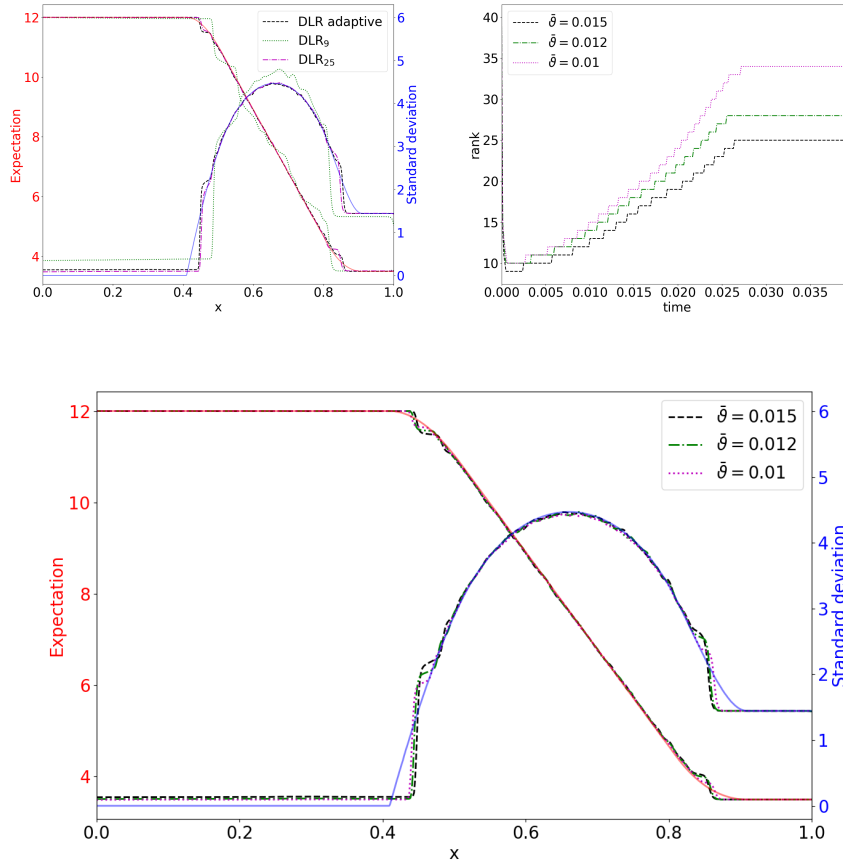
Fig. 3 Top left: Expected value (red) and variance (blue) of the analytic solution compared between the rank-adaptive and fixed-rank integrators. The rank-adaptive integrator uses the truncation tolerance $\vartheta = \bar{\vartheta}\|\hat{\boldsymbol{\Sigma}}\|_2$ where $\bar{\vartheta} = 0.015$. The fixed-rank integrator uses $r = 9$ and $r = 25$. Top right: Rank evolution of the rank-adaptive integrator over time using $\bar{\vartheta} \in \{0.01, 0.012, 0.015\}$. Bottom: Corresponding approximations for expectation and standard deviation.

14. E. Kieri, C. Lubich, and H. Walach. Discretized dynamical low-rank approximation in the presence of small singular values. *SIAM J. Numer. Anal.*, 54(2):1020–1038, 2016.
15. O. Koch and C. Lubich. Dynamical low-rank approximation. *SIAM J. Matrix Anal. Appl.*, 29(2):434–454, 2007.
16. T. G. Kolda and B. W. Bader. Tensor decompositions and applications. *SIAM Rev.*, 51(3):455–500, 2009.
17. J. Kusch, G. W. Alldredge, and M. Frank. Maximum-principle-satisfying second-order intrusive polynomial moment scheme. *SMAI J. Comput. Math.*, 5:23–51, 2019.
18. C. Lubich and I. V. Oseledets. A projector-splitting integrator for dynamical low-rank approximation. *BIT*, 54(1):171–188, 2014.
19. C. Lubich, I. V. Oseledets, and B. Vandereycken. Time integration of tensor trains. *SIAM J. Numer. Anal.*, 53(2):917–941, 2015.
20. H.-D. Meyer, F. Gatti, and G. A. Worth. *Multidimensional quantum dynamics: MCTDH theory and applications.* John Wiley & Sons, 2009.

21. H.-D. Meyer, U. Manthe, and L. S. Cederbaum. The multi-configurational time-dependent Hartree approach. *Chem. Phys. Letters*, 165(1):73–78, 1990.
22. E. Musharbash and F. Nobile. Dual dynamically orthogonal approximation of incompressible Navier–Stokes equations with random boundary conditions. *J. Comput. Phys.*, 354:135–162, 2018.
23. E. Musharbash, F. Nobile, and E. Vidličková. Symplectic dynamical low rank approximation of wave equations with random parameters. *BIT Numer. Math.*, 60:1153–1201, 2020.
24. Z. Peng and R. G. McClarren. A high-order/low-order (holo) algorithm for preserving conservation in time-dependent low-rank transport calculations. *arXiv preprint arXiv:2011.06072*, 2020.
25. Z. Peng, R. G. McClarren, and M. Frank. A low-rank method for two-dimensional time-dependent radiation transport calculations. *J. Comput. Phys.*, 421:109735, 2020.
26. G. Poëtte, B. Després, and D. Lucor. Uncertainty quantification for systems of conservation laws. *J. Comput. Phys.*, 228(7):2443–2467, 2009.
27. T. P. Sapsis and P. F. Lermusiaux. Dynamically orthogonal field equations for continuous stochastic dynamical systems. *Physica D*, 238(23-24):2347–2360, 2009.
28. S. Schrammer. Doctoral thesis in preparation. KIT, 2021.
29. J. Tryoen, O. Le Maitre, M. Ndjinga, and A. Ern. Intrusive Galerkin methods with upwinding for uncertain nonlinear hyperbolic systems. *J. Comput. Phys.*, 229(18):6485–6511, 2010.
30. M. Yang and S. R. White. Time-dependent variational principle with ancillary Krylov subspace. *Phys. Rev. B*, 102(9):094315, 2020.