BIT

Dynamical low-rank integrators for second-order matrix differential equations

Marlis Hochbruck[1] · **Markus Neher**[1] · **Stefan Schrammer**[1]

Abstract

In this paper, we construct and analyze a new dynamical low-rank integrator for second-order matrix differential equations. The method is based on a combination of the projector-splitting integrator introduced in Lubich and Oseledets (BIT 54(1):171–188, 2014. https://doi.org/10.1007/s10543-013-0454-0) and a Strang splitting. We also present a variant of the new integrator which is tailored to semilinear second-order problems.

Keywords Dynamical low-rank approximation · Matrix differential equations

Mathematics Subject Classification 65L04 · 65L05 · 65L70

1 Introduction

Dynamical low-rank integrators [10] have been introduced for the approximation of large, time-dependent matrices which are solutions to first-order matrix differential equations that can be well approximated by low-rank matrices. Typically, such matrix differential equations stem from spatially discretized PDEs. The idea is to project the right-hand side of the problem onto the tangent space of the manifold of matrices with small, fixed rank. It was shown in [10], that this ansatz yields differential equations

for the factors of a low-rank decomposition resembling a singular value decomposition. Compared to the approximation of the full matrix solution, working only with the factors of the low-rank decomposition significantly reduces the computational costs and the required storage. Unfortunately, the integrator of [10] suffers from ill-conditioning in the presence of small singular values, a situation which is called over-approximation, i.e., the rank chosen within the method exceeds the rank of the actual solution. A projector-splitting integrator which is robust in the case of over-approximation was introduced in [12] by Lubich and Oseledets. It is based on a clever Lie–Trotter splitting of the projected right-hand side, by which the subproblems can be solved exactly. Another dynamical low-rank integrator for first-order matrix differential equations is the unconventional robust integrator recently presented in [2]. It is especially suited for strongly dissipative problems. Although its derivation is rather different from the construction of the projector-splitting integrator, both methods share their robustness in the case of over-approximation and they satisfy the same error bounds.

Both the projector-splitting integrator and the unconventional integrator have been applied to a variety of first-order matrix differential equations, e.g., for Schrödinger equations in [2], for the Vlasov–Poison equation in [4], for Vlasov-Maxwell equations in [5], and for Burgers' equation with uncertainty in [11]. However, to the best of our knowledge, for second-order matrix differential equations of the form

$$A''(t) = F(A(t)), \quad A(t) \in \mathbb{C}^{m \times n}, \qquad A(0) = A_0, \quad A'(0) = B_0, \qquad (1)$$

with large m, n, such integrators have not been considered so far. The obvious technique of reformulating (1) into a first-order system and applying the projector-splitting integrator [12] for first-order matrix differential equations behaved poorly in our numerical experiments. The reason might be that the inherent structure of the second-order problem is ignored by this procedure, causing the approximation quality to deteriorate.

Therefore, we propose to combine the projector-splitting integrator with a Strang splitting. The resulting method is closely related to the leapfrog scheme and called LRLF method in the following. It is a robust and reliable dynamical low-rank integrator for second-order equations of type (1), provided that the exact solution $A(t)$ and its derivative $A'(t)$ can be well approximated by matrices of low rank. When the exact flows used within the Strang splitting preserve the low rank of the previous approximation, our integrator reduces to the leapfrog scheme if the rank chosen in the method is sufficiently large. We also develop a variant of the scheme which is tailored to semilinear second-order equations. For this, we combine our newly developed scheme with the ideas in [15], where a dynamical low-rank integrator for stiff semilinear first-order equations was derived based on the projector-splitting integrator.

For the projector-splitting integrator, a detailed error analysis was provided in [9]. It relies on an exactness property of the integrator, namely that it provides the exact solution if this solution preserves the (low) rank of the initial value for all times and the exact initial value is used to start the integrator. Unfortunately, this is no longer true for the LRLF scheme. Nevertheless, we will provide error bounds under similar assumptions as in [9].

The paper is organized as follows: In Sect. 2, we briefly recall the projector-splitting integrator introduced in [12] by Lubich and Oseledets. The construction of the LRLF scheme is presented in Sect. 3 and its error analysis in Sect. 4. A modification of the LRLF scheme which is tailored to semilinear second-order problems is developed in Sect. 5.

Throughout this paper, m, n, and r are natural numbers, where w.l.o.g. $m \geq n \gg r$. If $n > m$, we consider the equivalent differential equation for the transpose. By $\mathscr{M}_r$ we denote the manifold of complex $m \times n$ matrices with rank r,

$$\mathscr{M}_r = \{\widehat{Y} \in \mathbb{C}^{m \times n} \mid \mathrm{rank}(\widehat{Y}) = r\}.$$

The Stiefel manifold of $m \times r$ unitary matrices is denoted by

$$\mathscr{V}_{m,r} = \{U \in \mathbb{C}^{m \times r} \mid U^H U = I_r\},$$

where I_r is the identity matrix of dimension r and U^H is the conjugate transpose of U.

The singular value decomposition of a matrix $Y \in \mathbb{C}^{m \times n}$ is given by

$$Y = U \Sigma V^H, \quad U \in \mathscr{V}_{m,m}, \quad V \in \mathscr{V}_{n,n}, \quad \Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_n) \in \mathbb{C}^{m \times n},$$

where $\sigma_1 \geq \ldots \geq \sigma_n \geq 0$ are its singular values. It is well known that for $r < n$, the rank-r best-approximation to Y w.r.t. the Frobenius norm $\|\cdot\|$ is given by

$$\widehat{Y}^{\mathrm{best}} = U \widetilde{\Sigma} V^H = \widehat{U} \widehat{\Sigma} \widehat{V}^H,$$

where $\widetilde{\Sigma} = \mathrm{diag}(\sigma_1, \ldots, \sigma_r, 0, \ldots, 0)$ and

$$\widehat{U} = U[I_r \; 0] \in \mathscr{V}_{m,r}, \quad \widehat{V} = V[I_r \; 0] \in \mathscr{V}_{n,r}, \quad \widehat{\Sigma} = \mathrm{diag}(\sigma_1, \ldots, \sigma_r).$$

For a given step size τ we use the notation $t_k = k\tau$ for any k with $2k \in \mathbb{N}_0$.

2 The projector-splitting integrator

In this section, we briefly review dynamical low-rank approximations as introduced in [10, 12]. We start with the following problem: Given some time-dependent matrix $A(t)$, find a low-rank approximation $\widehat{A}(t) \in \mathscr{M}_r$ such that

$$\widehat{A}(t) \approx A(t) \quad \text{for all} \quad t \in [0, T].$$

In [10], this is done by imposing that $\widehat{A}'(t)$, which is contained in the tangent space $\mathscr{T}_{\widehat{A}(t)} \mathscr{M}_r$ to $\mathscr{M}_r$ at $\widehat{A}(t)$, satisfies

$$\|\widehat{A}'(t) - A'(t)\| = \min! \tag{2}$$

For an initial value $\widehat{A}(0) = \widehat{A}_0 \in \mathscr{M}_r$, condition (2) is equivalent to a Galerkin condition. In fact, then $\widehat{A}$ solves the evolution equation

$$\widehat{A}'(t) = P\big(\widehat{A}(t)\big)A'(t), \qquad \widehat{A}(0) = \widehat{A}_0 \in \mathscr{M}_r, \tag{3}$$

where $P\big(\widehat{A}(t)\big)$ denotes the orthogonal projection onto the tangent space $\mathscr{T}_{\widehat{A}(t)}\mathscr{M}_r$. A natural choice for the initial value $\widehat{A}_0$ of (3) is the rank-r best approximation to $A(0)$.

For all $\widehat{A} \in \mathscr{M}_r$ there is a non-unique low-rank factorization

$$\widehat{A} = \widehat{U}\widehat{S}\widehat{V}^H, \qquad \widehat{U} \in \mathscr{V}_{m,r}, \quad \widehat{V} \in \mathscr{V}_{n,r}, \quad \widehat{S} \in \mathbb{C}^{r\times r} \text{ invertible},$$

which allows us to write the orthogonal projector $P(\widehat{A})$ onto $\mathscr{T}_{\widehat{A}(t)}\mathscr{M}_r$ as

$$P(\widehat{A})Z = Z\widehat{V}\widehat{V}^H - \widehat{U}\widehat{U}^H Z\widehat{V}\widehat{V}^H + \widehat{U}\widehat{U}^H Z, \tag{4}$$

cf., [10, Lemma 4.1]. The dynamical low-rank integrator constructed in [12] is based on a Lie–Trotter splitting, applied to the differential equation (3) with $P(\widehat{A})$ given in (4). Given a step size $\tau > 0$, the first integration step consists of solving the three subproblems

$$Y'_\alpha = A'\widehat{V}_\alpha \widehat{V}_\alpha^H, \qquad\qquad Y_\alpha(0) = \widehat{A}_0, \tag{5a}$$
$$Y'_\beta = -\widehat{U}_\beta\widehat{U}_\beta^H A'\widehat{V}_\beta\widehat{V}_\beta^H, \qquad\qquad Y_\beta(0) = Y_\alpha(\tau), \tag{5b}$$
$$Y'_\gamma = \widehat{U}_\gamma\widehat{U}_\gamma^H A', \qquad\qquad Y_\gamma(0) = Y_\beta(\tau). \tag{5c}$$

As shown in [12, Lemma 3.1], all subproblems (5) can be solved exactly on $\mathscr{M}_r$ when the additional conditions

$$\widehat{V}'_\alpha(t) = 0, \qquad \widehat{U}'_\beta(t) = 0, \ \widehat{V}'_\beta(t) = 0, \qquad \widehat{U}'_\gamma(t) = 0,$$

are imposed. Then, the solutions

$$Y_\eta(t) = \widehat{U}_\eta(t)\widehat{S}_\eta(t)\widehat{V}_\eta(t)^H, \qquad \eta \in \{\alpha, \beta, \gamma\},$$

can be written in terms of the increment $\Delta A = A(\tau) - A(0)$,

$$\widehat{U}_\alpha(t)\widehat{S}_\alpha(t) = \widehat{U}_\alpha(0)\widehat{S}_\alpha(0) + \Delta A\widehat{V}_\alpha(0), \qquad \widehat{V}_\alpha(t) = \widehat{V}_\alpha(0),$$
$$\widehat{S}_\beta(t) = \widehat{S}_\beta(0) - \widehat{U}_\beta(0)^H \Delta A\widehat{V}_\beta(0), \qquad \widehat{U}_\beta(t) = \widehat{U}_\beta(0), \quad \widehat{V}_\beta(t) = \widehat{V}_\beta(0),$$
$$\widehat{V}_\gamma(t)\widehat{S}_\gamma(t)^H = \widehat{V}_\gamma(0)\widehat{S}_\gamma(0)^H + \Delta A^H\widehat{U}_\gamma(0), \qquad \widehat{U}_\gamma(t) = \widehat{U}_\gamma(0).$$

The integration process is continued with initial value $Y_\gamma(\tau) \approx \widehat{A}(\tau)$ in the next time step. A single time step of the resultant *projector-splitting integrator* is presented in Algorithm 1, version (α).

The above approach is also suitable for the construction of low-rank approximations to the unknown solution $A(t)$ of the first-order differential equation

$$A'(t) = F(A(t)), \qquad A(0) = A_0 \in \mathbb{C}^{m \times n}, \qquad t \in [0, T]. \tag{6}$$

As explained in [12, Section 3.4], the only change affects the replacement of the increment ΔA in Algorithm 1 in a way resembling the explicit Euler method,

$$\Delta A = \tau F(\widehat{A}_0),$$

see version (β). The global error of this first-order scheme depends on the quality of the approximation of the exact solution $A(t)$ of (6) by a low-rank matrix (for $t \in [0, T]$) and on properties of the right-hand side F, cf. [9]. Alternatively, any other explicit scheme, e.g., a fourth order Runge-Kutta method can be used to replace the increment ΔA. We denote the latter method by PSI-RK.

Algorithm 1 Projector-splitting integrator for low-rank approximations to (α) given time-dependent matrices $A(t)$ or (β) the solution of (6), single time step

---

1: **function** PSI($\widehat{U}, \widehat{S}, \widehat{V}, r, \Delta A$)
2:     {**input:** factors $\widehat{U}, \widehat{S}, \widehat{V}$ of rank-r approximation $\widehat{A} = \widehat{U}\widehat{S}\widehat{V}^H \approx A(t)$ with $\widehat{U} \in \mathscr{V}_{m,r}, \widehat{V} \in \mathscr{V}_{n,r}$,
3:         $\widehat{S} \in \mathbb{C}^{r \times r}$, functions for matrix-vector multiplication with ΔA and ΔA^H,
4:         (α) $\Delta A = A(t+\tau) - A(t)$,
5:         (β) $\Delta A = \tau F(\widehat{A})$ }
6:
7:     $\widetilde{K} = \Delta A \widehat{V}$
8:     $\widehat{K} = \widehat{U}\widehat{S} + \widetilde{K}$
9:     compute QR-decomposition $\widehat{U}\widehat{S} = \widehat{K}$
10:    $\widehat{S} = \widehat{S} - \widehat{U}^H \widetilde{K}$
11:    $\widehat{L} = \widehat{V}\widehat{S}^H + \Delta A^H \widehat{U}$
12:    compute QR-decomposition $\widehat{V}\widehat{S}^H = \widehat{L}$
13:
14:    **return** $\widehat{U}, \widehat{S}, \widehat{V}, \widehat{L}$
15:    {**output:** factors $\widehat{U}, \widehat{S}, \widehat{V}$ of rank-r approximation $\widehat{A} = \widehat{U}\widehat{S}\widehat{V}^H \approx A(t+\tau)$ and $\widehat{L} = \widehat{V}\widehat{S}^H$
16:    (optional), with $\widehat{U} \in \mathscr{V}_{m,r}, \widehat{V} \in \mathscr{V}_{n,r}, \widehat{S} \in \mathbb{C}^{r \times r}$}
17: **end function**

Remark 1 The computational complexity of Algorithm 1 is dominated by the two products $\Delta A \widehat{V}$ and $\Delta A^H \widehat{U}$ in lines 7 and 11, respectively, the two QR-decompositions of matrices of dimension $m \times r$ and $n \times r$, respectively, and the three matrix-matrix products in lines 8, 10, and 11. For an efficient implementation, it is important to compute the products $\Delta A \widehat{V}$ and $\Delta A^H \widehat{U}$ without computing ΔA explicitly.

3 Dynamical low-rank approximation of second-order matrix ODEs

Next, we devise a low-rank integrator for second-order matrix differential equations of the form (1). A straightforward practice would be to rewrite (1) as a first-order system

$$\begin{bmatrix} A \\ B \end{bmatrix}' = \begin{bmatrix} B \\ F(A) \end{bmatrix}, \qquad \begin{bmatrix} A(0) \\ B(0) \end{bmatrix} = \begin{bmatrix} A_0 \\ B_0 \end{bmatrix}, \tag{7}$$

and to apply Algorithm 1. However, numerical tests showed that the quality of the numerical solutions deteriorates over time, possibly caused by the neglection of the structure of the the right-hand side of (7).

We thus use a different ansatz and first split (7) into

$$\begin{bmatrix} A \\ B \end{bmatrix}' = \begin{bmatrix} 0 \\ F(A) \end{bmatrix} + \begin{bmatrix} B \\ 0 \end{bmatrix}, \tag{8}$$

and then apply a standard Strang splitting. Solving the subproblems exactly leads to the well-known leapfrog or Störmer-Verlet scheme,

$$B_{k+\frac{1}{2}} = B_k + \frac{\tau}{2} F(A_k), \tag{9a}$$

$$A_{k+1} = A_k + \tau B_{k+\frac{1}{2}}, \tag{9b}$$

$$B_{k+1} = B_{k+\frac{1}{2}} + \frac{\tau}{2} F(A_{k+1}), \tag{9c}$$

cf. [6, Section 1.5]. If approximations to $A' = B$ are not required at full time steps, then the most economic implementation of the leapfrog scheme is to combine (9a) and (9c) via

$$B_{k+\frac{1}{2}} = B_{k-\frac{1}{2}} + \tau F(A_k), \quad k \geq 1. \tag{9d}$$

For $k = 0$, $B_{\frac{1}{2}}$ is computed from (9a).

A low-rank integrator for the second-order equation (1) is now designed by approximating the exact flows of the subproblems in (8) by their respective low-rank flows using the projector-splitting method [12]. First, we determine initial values $\widehat{A}_0$ and $\widehat{B}_0$ as rank-r_A and rank-r_B best-approximations to A_0 and B_0, respectively. After k time steps, the low-rank approximations $\widehat{A}_k \approx A(t_k)$ and $\widehat{B}_{k-\frac{1}{2}} \approx B(t_{k-\frac{1}{2}})$ are represented by (non-unique) decompositions

$$\widehat{A}_k = \widehat{U}_k \widehat{S}_k \widehat{V}_k^H \in \mathscr{M}_{r_A}, \qquad \widehat{B}_{k-\frac{1}{2}} = \widehat{T}_{k-\frac{1}{2}} \widehat{R}_{k-\frac{1}{2}} \widehat{W}_{k-\frac{1}{2}}^H \in \mathscr{M}_{r_B},$$

where

$$\widehat{U}_k \in \mathscr{V}_{m,r_A}, \ \widehat{V}_k \in \mathscr{V}_{n,r_A}, \ \widehat{S}_k \in \mathbb{C}^{r_A \times r_A}, \qquad \widehat{T}_{k-\frac{1}{2}} \in \mathscr{V}_{m,r_B}, \ \widehat{W}_{k-\frac{1}{2}} \in \mathscr{V}_{n,r_B}, \ \widehat{R}_{k-\frac{1}{2}} \in \mathbb{C}^{r_B \times r_B}.$$

The low-rank matrices $\widehat{A}_{k+1} \approx A(t_{k+1})$ and $\widehat{B}_{k+\frac{1}{2}} \approx B(t_{k+\frac{1}{2}})$ are obtained by approximating the solutions of (8) by applying Algorithm 1 to

$$\widetilde{B}'_{k-\frac{1}{2}}(\sigma) = F(\widehat{A}_k), \qquad \widetilde{B}_{k-\frac{1}{2}}(0) = \widehat{B}_{k-\frac{1}{2}}, \qquad \sigma \in [0, \tau], \qquad k \geq 1, \tag{10a}$$

$$\widetilde{A}'_k(\sigma) = \widehat{B}_{k+1/2}, \qquad \widetilde{A}_k(0) = \widehat{A}_k, \qquad \sigma \in [0, \tau], \qquad k \geq 0, \tag{10b}$$

where for $k = 0$, we have

$$\widetilde{B}_0'(\sigma) = F(\widehat{A}_0), \qquad \widetilde{B}_0(0) = \widehat{B}_0, \qquad \sigma \in \left[0, \frac{\tau}{2}\right]. \tag{10c}$$

Since the exact solutions of (10) read

$$\widetilde{B}_{k-\frac{1}{2}}(\tau) = \widehat{B}_{k-\frac{1}{2}} + \tau F(\widehat{A}_k), \tag{11a}$$

$$\widetilde{A}_k(\tau) = \widehat{A}_k + \tau \widehat{B}_{k+\frac{1}{2}}, \tag{11b}$$

$$\widetilde{B}_0\left(\frac{\tau}{2}\right) = \widehat{B}_0 + \frac{\tau}{2} F(\widehat{A}_0), \tag{11c}$$

the increments $\Delta B_{k-\frac{1}{2}}$ and ΔA_k are given explicitly as

$$\Delta B_{k-\frac{1}{2}} = \widetilde{B}_{k-\frac{1}{2}}(\tau) - \widetilde{B}_{k-\frac{1}{2}}(0) = \tau F(\widehat{A}_k), \qquad k \geq 1,$$

$$\Delta A_k = \widetilde{A}_k(\tau) - \widetilde{A}_k(0) = \tau \widehat{B}_{k+\frac{1}{2}}, \qquad k \geq 0,$$

$$\Delta B_0 = \widetilde{B}_0\left(\frac{\tau}{2}\right) - \widetilde{B}_0(0) = \frac{\tau}{2} F(\widehat{A}_0).$$

The resulting dynamical low-rank integrator for second-order matrix ODEs will be named LRLF method, for low-rank leapfrog integrator. It is presented in Algorithm 2.

There is a close relationship between the LRLF and the leapfrog schemes:

Theorem 2 *If for all $t \in [0, T]$ the exact solutions $A(t)$ and $B(t)$ of the subproblems* (8) *with* rank $A_0 = r_A$, rank $B_0 = r_B$ *stay of rank r_A and r_B, respectively, then the solutions of the* LRLF *and leapfrog schemes coincide.*

Proof The LRLF scheme is derived by exchanging the exact flows of the leapfrog scheme by their corresponding low-rank flows. Since the exact flows preserve the rank by assumption, the application of the exactness property of the projector-splitting integrator [12, Theorem 4.1] yields the desired result. □

Clearly, in general the LRLF scheme does not inherit the exactness property of the projector-splitting integrator, because of the splitting error. A detailed error analysis of the LRLF iteration will be presented in Sect. 4.

In the same way, a variant of the LRLF method for the second-order matrix differential equation

$$A'' = \omega^2 F(A), \quad t \in [0, T], \qquad A(0) = A_0, \quad A'(0) = \omega^2 B_0, \tag{12}$$

or equivalently

$$\begin{bmatrix} A \\ B \end{bmatrix}' = \begin{bmatrix} \omega^2 B \\ F(A) \end{bmatrix}, \quad \begin{bmatrix} A(0) \\ B(0) \end{bmatrix} = \begin{bmatrix} A_0 \\ B_0 \end{bmatrix},$$

Algorithm 2 DLR integrator for second-order ODEs (1), LRLF scheme, single time step

---

1: **function** LRLF($\tau, F, \widehat{U}, \widehat{S}, \widehat{V}, \widehat{T}, \widehat{R}, \widehat{W}, r_A, r_B$)
2:    {**input:** step size τ, right-hand side F,
3:    factors $\widehat{U}, \widehat{S}, \widehat{V}$ of rank-r_A approximation $\widehat{A} = \widehat{U}\widehat{S}\widehat{V}^H \approx A(t)$ with $\widehat{U} \in \mathscr{V}_{m,r_A}$, $\widehat{V} \in \mathscr{V}_{n,r_A}$,
4:    $\widehat{S} \in \mathbb{C}^{r_A \times r_A}$,
5:    factors $\widehat{T}, \widehat{R}, \widehat{W}$ of rank-r_B approximation $\widehat{B} = \widehat{T}\widehat{R}\widehat{W}^H \approx A'(t - \frac{\tau}{2})$ with $\widehat{T} \in \mathscr{V}_{m,r_B}$,
6:    $\widehat{W} \in \mathscr{V}_{n,r_B}$, $\widehat{R} \in \mathbb{C}^{r_B \times r_B}$ }
7:
8:    $\widehat{B}$-*step:* $\widehat{T}, \widehat{R}, \widehat{W}, \widehat{L} = \mathrm{PSI}(\widehat{T}, \widehat{R}, \widehat{W}, r_B, \Delta B)$ where $\Delta B = \tau F(\widehat{U}\widehat{S}\widehat{V}^H)$
9:
10:    $\widehat{A}$-*step:* $\widehat{U}, \widehat{S}, \widehat{V} = \mathrm{PSI}(\widehat{U}, \widehat{S}, \widehat{V}, r_A, \Delta A)$ where $\Delta A = \tau \widehat{T}\widehat{L}^H$
11:
12:    **return** $\widehat{U}, \widehat{S}, \widehat{V}, \widehat{T}, \widehat{R}, \widehat{W}$
13:    {**output:** factors $\widehat{U}, \widehat{S}, \widehat{V}$ of rank-r_A approximation $\widehat{A} = \widehat{U}\widehat{S}\widehat{V}^H \approx A(t + \tau)$ with $\widehat{U} \in \mathscr{V}_{m,r_A}$,
14:    $\widehat{V} \in \mathscr{V}_{n,r_A}$, $\widehat{S} \in \mathbb{C}^{r_A \times r_A}$,
15:    factors $\widehat{T}, \widehat{R}, \widehat{W}$ of rank-r_B approximation $\widehat{B} = \widehat{T}\widehat{R}\widehat{W}^H \approx A'(t + \frac{\tau}{2})$ with $\widehat{T} \in \mathscr{V}_{m,r_B}$,
16:    $\widehat{W} \in \mathscr{V}_{n,r_B}$, $\widehat{R} \in \mathbb{C}^{r_B \times r_B}$ }
17: **end function**

Algorithm 3 DLR integrator for second-order ODEs (12), LRLF(ω) scheme, single time step

---

1: **function** LRLFOMEGA($\tau, \omega, F, \widehat{U}, \widehat{S}, \widehat{V}, \widehat{T}, \widehat{R}, \widehat{W}, r_A, r_B$)
2:    {**input:** step size τ, weight ω, right-hand side F,
3:    factors $\widehat{U}, \widehat{S}, \widehat{V}$ of rank-r_A approximation $\widehat{A} = \widehat{U}\widehat{S}\widehat{V}^H \approx A(t)$ with $\widehat{U} \in \mathscr{V}_{m,r_A}$, $\widehat{V} \in \mathscr{V}_{n,r_A}$,
4:    $\widehat{S} \in \mathbb{C}^{r_A \times r_A}$,
5:    factors $\widehat{T}, \widehat{R}, \widehat{W}$ of rank-r_B approximation $\widehat{B} = \widehat{T}\widehat{R}\widehat{W}^H \approx A'(t)$ with $\widehat{T} \in \mathscr{V}_{m,r_B}$, $\widehat{W} \in \mathscr{V}_{n,r_B}$,
6:    $\widehat{R} \in \mathbb{C}^{r_B \times r_B}$ }
7:
8:    $\widehat{B}$-*step:* $\widehat{T}, \widehat{R}, \widehat{W}, \widehat{L} = \mathrm{PSI}(\widehat{T}, \widehat{R}, \widehat{W}, r_B, \Delta B)$ where $\Delta B = \frac{\tau}{2} F(\widehat{U}\widehat{S}\widehat{V}^H)$
9:
10:    $\widehat{A}$-*step:* $\widehat{U}, \widehat{S}, \widehat{V} = \mathrm{PSI}(\widehat{U}, \widehat{S}, \widehat{V}, r_A, \Delta A)$ where $\Delta A = \omega^2 \tau \widehat{T}\widehat{L}^H$
11:
12:    $\widehat{B}$-*step:* $\widehat{T}, \widehat{R}, \widehat{W} = \mathrm{PSI}(\widehat{T}, \widehat{R}, \widehat{W}, r_B, \Delta B)$ where $\Delta B = \frac{\tau}{2} F(\widehat{U}\widehat{S}\widehat{V}^H)$
13:
14:    **return** $\widehat{U}, \widehat{S}, \widehat{V}, \widehat{T}, \widehat{R}, \widehat{W}$
15:    {**output:** factors $\widehat{U}, \widehat{S}, \widehat{V}$ of rank-r_A approximation $\widehat{A} = \widehat{U}\widehat{S}\widehat{V}^H \approx A(t + \tau)$ with $\widehat{U} \in \mathscr{V}_{m,r_A}$,
16:    $\widehat{V} \in \mathscr{V}_{n,r_A}$, $\widehat{S} \in \mathbb{C}^{r_A \times r_A}$,
17:    factors $\widehat{T}, \widehat{R}, \widehat{W}$ of rank-r_B approximation $\widehat{B} = \widehat{T}\widehat{R}\widehat{W}^H \approx A'(t + \tau)$ with $\widehat{T} \in \mathscr{V}_{m,r_B}$,
18:    $\widehat{W} \in \mathscr{V}_{n,r_B}$, $\widehat{R} \in \mathbb{C}^{r_B \times r_B}$ }
19: **end function**

---

is based on the non-staggered version of the leapfrog scheme (9), procuring approximations to A' on the same time-grid as approximations to A. Here, the function F has to be evaluated twice in each time step so that the computational effort is larger. The resulting method is called the LRLF(ω) scheme. It is described in Algorithm 3.

Remark 3 A dynamical low-rank integrator for second-order matrix differential equations (1) can also be derived based on the two-step formulation of the leapfrog scheme,

$$A_{k+1} - 2A_k + A_{k-1} = \tau^2 F(A_k), \quad k \geq 1,$$

$$A_1 = A_0 + \tau B_0 + \frac{\tau^2}{2} F(A_0),$$

cf. [16, Section 4.2]. In numerical experiments it was observed that the resulting scheme does not yield better approximations than the LRLF method, and even produces larger errors in some examples.

4 Error analysis of LRLF

In the following, we analyze the error of the LRLF scheme given in Algorithm 2 when applied to (1) with a right-hand side F which is Lipschitz-continuous with a moderate Lipschitz constant L, i.e., F satisfies

$$\|F(Y) - F(\widetilde{Y})\| \le L \|Y - \widetilde{Y}\| \quad \text{for all } Y, \widetilde{Y} \in \mathbb{C}^{m \times n}. \tag{13}$$

Our analysis relies on the error analysis in [9] for the projector-splitting integrator.

Recall that the LRLF scheme is derived from the leapfrog scheme (9), which is stable under the CFL condition, cf. [8],

$$\tau^2 \le \tau_{\text{CFL}}^2 = \frac{4}{L}. \tag{14}$$

We therefore assume that the step size τ always satisfies (14).

Assumption 1 The exact solution $A : [0, T] \to \mathbb{C}^{m \times n}$ of (1) is in $\mathscr{C}^4([0, T])$. Furthermore, there are low-rank approximations $X_A(t) \in \mathscr{M}_{r_A}, X_B(t) \in \mathscr{M}_{r_B}$ such that

$$A(t) = X_A(t) + R_A(t), \qquad \|R_A(0)\| \le \rho_A, \qquad \|R_A'(t)\| \le \rho_A', \tag{15a}$$
$$B(t) = A'(t) = X_B(t) + R_B(t), \qquad \|R_B(0)\| \le \rho_B, \qquad \|R_B'(t)\| \le \rho_B'. \tag{15b}$$

Additionally, there exist sufficiently large constants γ_A and γ_B such that (15) is also satisfied for all $Y_A(t), Y_B(t) \in \mathbb{C}^{m \times n}$ with

$$\|A(t) - Y_A(t)\| \le \gamma_A, \qquad \|B(t) - Y_B(t)\| \le \gamma_B.$$

For the approximations $\widehat{A}_k \approx A(t_k)$ and $\widehat{B}_{k+\frac{1}{2}} \approx B(t_{k+\frac{1}{2}})$ computed with the LRLF scheme and for $\widetilde{A}_k$ and $\widetilde{B}_{k-\frac{1}{2}}$ given in (11), we define

$$E_A^k = \|A(t_k) - \widehat{A}_k\|, \qquad E_B^{k+\frac{1}{2}} = \|B(t_{k+\frac{1}{2}}) - \widehat{B}_{k+\frac{1}{2}}\|, \qquad k \ge 0, \tag{16a}$$

$$\widetilde{E}_A^0 = 0, \qquad \widetilde{E}_B^{\frac{1}{2}} = \|B(t_{\frac{1}{2}}) - \widetilde{B}_0(t_{\frac{1}{2}})\|, \tag{16b}$$

$$\widetilde{E}_A^k = \|A(t_k) - \widetilde{A}_{k-1}(\tau)\|, \qquad \widetilde{E}_B^{k+\frac{1}{2}} = \|B(t_{k+\frac{1}{2}}) - \widetilde{B}_{k-\frac{1}{2}}(\tau)\|, \qquad k \ge 1, \tag{16c}$$

$$\widehat{E}_A^0 = E_A^0, \qquad \widehat{E}_B^{\frac{1}{2}} = \|\widetilde{B}_0(t_{\frac{1}{2}}) - \widehat{B}_{\frac{1}{2}}\|, \tag{16d}$$

$$\widehat{E}_A^k = \|\widetilde{A}_{k-1}(\tau) - \widehat{A}_k\|, \qquad \widehat{E}_B^{k+\frac{1}{2}} = \|\widetilde{B}_{k-\frac{1}{2}}(\tau) - \widehat{B}_{k+\frac{1}{2}}\|, \qquad k \geq 1. \quad (16e)$$

By the triangle inequality, we have

$$E_B^{k+\frac{1}{2}} \leq \widetilde{E}_B^{k+\frac{1}{2}} + \widehat{E}_B^{k+\frac{1}{2}} \quad \text{and} \quad E_A^{k+1} \leq \widetilde{E}_A^{k+1} + \widehat{E}_A^{k+1}, \quad k \geq 0. \quad (17)$$

The analysis of the LRLF scheme is organized in two lemmas and a theorem. Our first result are coupled, recursive inequalities for E_A^{k+1} and $E_B^{k+\frac{1}{2}}$.

Lemma 4 *Let $A : [0, T] \to \mathbb{C}^{m \times n}$ with $A \in \mathscr{C}^4([0, T])$ be the exact solution of* (1) *with initial values $A_0, B_0 \in \mathbb{C}^{m \times n}$ and $B = A'$. Further, denote by $\widehat{B}_{k-\frac{1}{2}}$ and $\widehat{A}_k$ the low-rank approximations obtained by the* LRLF *scheme after k steps started with initial values $\widehat{A}_0 \in \mathscr{M}_{r_A}$, $\widehat{B}_0 \in \mathscr{M}_{r_B}$. Then, the errors introduced in* (16) *satisfy*

$$E_B^{\frac{1}{2}} \leq E_B^0 + \frac{\tau}{2} L E_A^0 + \widehat{E}_B^{\frac{1}{2}} + C_B^{LF} \tau^2, \quad (18a)$$

$$E_A^1 \leq \left(1 + \frac{\tau^2}{2} L\right) E_A^0 + \tau E_B^0 + \tau \widehat{E}_B^{\frac{1}{2}} + \widehat{E}_A^1 + \left(C_A^{LF} + C_B^{LF}\right) \tau^3 \quad (18b)$$

and for $k \in \mathbb{N}$ we have

$$E_B^{k+\frac{1}{2}} \leq E_B^{k-\frac{1}{2}} + \tau L E_A^k + \widehat{E}_B^{k+\frac{1}{2}} + C_B^{LF} \tau^3, \quad (19a)$$

$$E_A^{k+1} \leq \left(1 + \tau^2 L\right) E_A^k + \tau E_B^{k-\frac{1}{2}} + \tau \widehat{E}_B^{k+\frac{1}{2}} + \widehat{E}_A^{k+1} + \left(C_A^{LF} + C_B^{LF} \tau\right) \tau^3. \quad (19b)$$

The constants C_A^{LF} and C_B^{LF} are given explicitly as

$$C_A^{LF} = \max_{t \in [0,T]} \frac{1}{24} \|A'''(t)\|, \qquad C_B^{LF} = \max\left\{ \max_{t \in [0, \frac{\tau}{2}]} \frac{1}{8} \|A'''(t)\|, \quad \max_{t \in [0,T]} \frac{1}{24} \|A^{(4)}(t)\| \right\}.$$

Proof By Taylor series expansion, we have

$$\|A(t_{k+1}) - \left(A(t_k) + \tau B(t_{k+\frac{1}{2}})\right)\| \leq C_A^{LF} \tau^3, \qquad k \geq 0, \quad (20a)$$

$$\|B(t_{k+\frac{1}{2}}) - \left(B(t_{k-\frac{1}{2}}) + \tau F(A(t_k))\right)\| \leq C_B^{LF} \tau^3, \qquad k \geq 1, \quad (20b)$$

as well as

$$\|B(t_{\frac{1}{2}}) - \left(B_0 + \frac{\tau}{2} F(A_0)\right)\| \leq C_B^{LF} \tau^2. \quad (20c)$$

Hence for $k = 0$, we have

$$\widetilde{E}_B^{\frac{1}{2}} \leq \|B_0 + \frac{\tau}{2} F(A_0) - (\widehat{B}_0 + \frac{\tau}{2} F(\widehat{A}_0))\| + C_B^{LF} \tau^2$$
$$\leq E_B^0 + \frac{\tau}{2} L E_A^0 + C_B^{LF} \tau^2 \quad (21)$$

by (20c), (11c), and (13). Employing (17) shows (18a). Using (18a) and (20a) yields

$$
\begin{aligned}
\widetilde{E}_A^1 &\leq \| A_0 + \tau B(\tfrac{\tau}{2}) - (\widehat{A}_0 + \tau \widehat{B}_{\frac{1}{2}}) \| + C_A^{\mathrm{LF}} \tau^3 \\
&\leq E_A^0 + \tau E_B^{\frac{1}{2}} + C_A^{\mathrm{LF}} \tau^3 \\
&\leq \left(1 + \frac{\tau^2}{2} L\right) E_A^0 + \tau E_B^0 + \tau \widehat{E}_B^{\frac{1}{2}} + (C_A^{\mathrm{LF}} + C_B^{\mathrm{LF}}) \tau^3.
\end{aligned}
\tag{22}
$$

Together with (17) this proves (18b).

For $k \geq 1$ we follow the same steps. We have by (11), (13), and (20b)

$$
\widetilde{E}_B^{k+\frac{1}{2}} \leq E_B^{k-\frac{1}{2}} + \tau L E_A^k + C_B^{\mathrm{LF}} \tau^3
\tag{23}
$$

and by (17) thus (19a). Lastly, we have by (20a), using again (11), and inserting (19a)

$$
\begin{aligned}
\widetilde{E}_A^{k+1} &\leq E_A^k + \tau E_B^{k+\frac{1}{2}} + C_A^{\mathrm{LF}} \tau^3 \\
&\leq E_A^k + \tau \left(E_B^{k-\frac{1}{2}} + \tau L E_A^k + \widehat{E}_B^{k+\frac{1}{2}} + C_B^{\mathrm{LF}} \tau^3 \right) + C_A^{\mathrm{LF}} \tau^3 \\
&= (1 + \tau^2 L) E_A^k + \tau E_B^{k-\frac{1}{2}} + \tau \widehat{E}_B^{k+\frac{1}{2}} + (C_A^{\mathrm{LF}} + C_B^{\mathrm{LF}} \tau) \tau^3,
\end{aligned}
\tag{24}
$$

which together with (17) completes the proof.　　　　　　　　　$\square$

In [9, Section 2.6.1] it was shown that the error between a time-dependent matrix $A(t)$ satisfying (15a) and the rank-r_A approximation $Y_1 \approx A(\tau)$ computed by Algorithm 1 started from $X_A(0) \in \mathscr{M}_{r_A}$ is bounded by

$$
\| A(\tau) - Y_1 \| \leq \rho_A + 7\tau \rho_A'.
\tag{25}
$$

In the next lemma we eliminate $\widehat{E}_B^{k+\frac{1}{2}}$ and $\widehat{E}_A^{k+1}$ from (19) by using (25).

Lemma 5 *Let the assumptions of Lemma 4 be satisfied. Furthermore, assume that*

$$
E_A^0 \leq \rho_A, \qquad E_B^0 \leq \rho_B.
\tag{26}
$$

If Assumption 1 is fulfilled, then for all k such that $t_{k+4} \leq T$, the errors $E_B^{k+\frac{1}{2}}$ and E_A^{k+1} defined in (16) satisfy

$$
E_B^{k+\frac{1}{2}} \leq \rho_B + 7 t_{k+1} \rho_B' + \tau L \sum_{j=0}^{k} E_A^j + C_B^{LF} \tau^2 (1 + t_k),
\tag{27a}
$$

$$
E_A^{k+1} \leq \rho_A + t_{k+1} \rho_B + \tau L \sum_{j=0}^{k} t_{k+1-j} E_A^j + 7 t_{k+1} \rho_A' + \frac{7}{2} t_{k+1} t_{k+2} \rho_B'
$$

$$+ \left((C_A^{LF} + C_B^{LF})t_{k+1} + \frac{1}{2}t_k t_{k+1} C_B^{LF} \right) \tau^2, \tag{27b}$$

respectively.

Proof The proof is accomplished by induction on k. First, we show that the errors $\widetilde{E}_A^{k+1}$ and $\widetilde{E}_B^{k+\frac{1}{2}}$ are uniformly bounded by suitable constants γ_A and γ_B. Then the auxiliary solutions $\widetilde{A}$ and $\widetilde{B}$ are sufficiently close to the exact solutions A and B, and hence they admit representations like (15) by Assumption 1. Since the approximations $\widehat{A}$ and $\widehat{B}$ are low-rank approximation to $\widetilde{A}$ and $\widetilde{B}$ computed by Algorithm 1 started from initial values of rank r_A and r_B, respectively, the local errors $\widehat{E}_A^{k+1}$ and $\widehat{E}_B^{k+\frac{1}{2}}$ are bounded by

$$\widehat{E}_B^{\frac{1}{2}} \leq \frac{7}{2}\tau\rho'_B, \quad \widehat{E}_B^{k+\frac{1}{2}} \leq 7\tau\rho'_B, \quad k \geq 1, \qquad \widehat{E}_A^{k+1} \leq 7\tau\rho'_A, \quad k \geq 0, \tag{28}$$

cf. (25). The estimate on the global error then follows from (17).

For $k = 0$ we deduce from (21) and (26)

$$\widetilde{E}_B^{\frac{1}{2}} \leq \rho_B + \frac{\tau}{2}L\rho_A + C_B^{LF}\tau^2.$$

By Assumption 1, for $\gamma_B \geq \rho_B + \frac{\tau}{2}L\rho_A + C_B^{LF}\tau^2$, it holds by (18a) and (28) that

$$\begin{aligned} E_B^{\frac{1}{2}} &\leq \rho_B + \frac{\tau}{2}LE_A^0 + 7\frac{\tau}{2}\rho'_B + C_B^{LF}\tau^2 \\ &< \rho_B + \tau L E_A^0 + 7t_1\rho'_B + C_B^{LF}\tau^2, \end{aligned} \tag{29}$$

which is (27a) for $k = 0$.

Likewise, by (22), (26), and (29) we have

$$\widetilde{E}_A^1 \leq \rho_A + \tau^2 L E_A^0 + \tau\rho_B + 7\tau^2\rho'_B + (C_A^{LF} + C_B^{LF})\tau^3. \tag{30}$$

Choosing γ_A as the right-hand side of (30), by Assumption 1 and (28) we deduce

$$E_A^1 \leq \rho_A + \tau^2 L E_A^0 + t_1\rho_B + 7t_1\rho'_A + \frac{7}{2}t_1 t_2\rho'_B + (C_A^{LF} + C_B^{LF})t_1\tau^2,$$

which shows (27b) for $k = 0$.

Assuming that (27) holds true for some arbitrary, but fixed $k - 1 \in \mathbb{N}_0$, we now prove (27). Applying the Gronwall-type Lemma 7 below, we find from (27b) for $j = 1, \ldots, k$

$$E_A^j \leq e^{\sqrt{L}t_k} M_A^j, \tag{31}$$

where

$$M_A^j = \rho_A + t_j \rho_B + 7 t_j \rho_A' + \frac{7}{2} t_j t_{j+1} \rho_B' + \left((C_A^{\mathrm{LF}} + C_B^{\mathrm{LF}}) t_j + \frac{1}{2} t_{j-1} t_j C_B^{\mathrm{LF}} \right) \tau^2. \tag{32}$$

By (26), this bound is also valid for $j = 0$. From (23) and (27a) we obtain

$$
\begin{aligned}
\widetilde{E}_B^{k+\frac{1}{2}} &\le E_B^{k-\frac{1}{2}} + \tau L E_A^k + C_B^{\mathrm{LF}} \tau^3 \\
&\le \left(\rho_B + 7 t_k \rho_B' + \tau L \sum_{j=0}^{k-1} E_A^j + C_B^{\mathrm{LF}} \tau^2 (1 + t_{k-1}) \right) + \tau L E_A^k + C_B^{\mathrm{LF}} \tau^3 \\
&= \rho_B + 7 t_k \rho_B' + \tau L \sum_{j=0}^{k} E_A^j + C_B^{\mathrm{LF}} \tau^2 (1 + t_k). \tag{33}
\end{aligned}
$$

Inserting (31) into (33) for $0 \le t_{k+4} \le T$ results in constants $C_B(T), \widetilde{C}_B(T)$ depending on $L, \rho_A, \rho_B, \rho_A', \rho_B', C_A^{\mathrm{LF}}$, and C_B^{LF} such that

$$\widetilde{E}_B^{k+\frac{1}{2}} \le C_B(T) + \tau^2 \widetilde{C}_B(T).$$

By Assumption 1 for $\gamma_B \ge C_B(T) + \tau^2 \widetilde{C}_B(T)$, (28) shows that $\widehat{E}_B^{k+\frac{1}{2}} \le 7 \tau \rho_B'$. Thus we obtain from (33) and (17)

$$E_B^{k+\frac{1}{2}} \le \rho_B + 7 t_{k+1} \rho_B' + \tau L \sum_{j=0}^{k} E_A^j + C_B^{\mathrm{LF}} \tau^2 (1 + t_k),$$

which proves (27a) for all $k \in \mathbb{N}_0$.

Similarly, using (24) and the induction hypothesis, we get

$$
\begin{aligned}
\widetilde{E}_A^{k+1} &\le \left(\rho_A + t_k \rho_B + \tau L \sum_{j=0}^{k-1} t_{k-j} E_A^j + 7 t_k \rho_A' + \frac{7}{2} t_k t_{k+1} \rho_B' \right. \\
&\quad \left. + \left((C_A^{\mathrm{LF}} + C_B^{\mathrm{LF}}) t_{k+1} + \frac{1}{2} t_k t_{k+1} C_B^{\mathrm{LF}} \right) \tau^2 \right) + \tau^2 L E_A^k \\
&\quad + \tau \left(\rho_B + 7 t_k \rho_B' + \tau L \sum_{j=0}^{k-1} E_A^j + C_B^{\mathrm{LF}} \tau^2 (1 + t_{k-1}) \right) \\
&\quad + 7 \tau^2 \rho_B' + (C_A^{\mathrm{LF}} + C_B^{\mathrm{LF}} \tau) \tau^3 \\
&= \rho_A + t_{k+1} \rho_B + \tau L \sum_{j=0}^{k} t_{k+1-j} E_A^j + 7 t_k \rho_A' + \frac{7}{2} t_{k+1} t_{k+2} \rho_B'
\end{aligned}
$$

$$+ \left((C_A^{\mathrm{LF}} + C_B^{\mathrm{LF}})t_{k+1} + \frac{1}{2}t_k t_{k+1} C_B^{\mathrm{LF}} \right) \tau^2. \tag{34}$$

By employing the bound (31) on E_A^j for $0 \le t_{k+4} \le T$, we define constants $C_A(T)$, $\widetilde{C}_A(T)$ depending on L, ρ_A, ρ_B, ρ_A', ρ_B', C_A^{LF}, and C_B^{LF} such that

$$\widetilde{E}_A^{k+1} \le C_A(T) + \tau^2 \widetilde{C}_A(T).$$

Now let Assumption 1 be fulfilled for some $\gamma_A \ge C_A(T) + \tau^2 \widetilde{C}_A(T)$. Then we have $\widehat{E}_A^{k+1} \le 7\tau \rho_A'$ by (28). Finally, we conclude from (34) and (17)

$$E_A^{k+1} \le \rho_A + t_{k+1}\rho_B + \tau L \sum_{j=0}^{k} t_{k+1-j} E_A^j + 7 t_{k+1} \rho_A' + \frac{7}{2} t_{k+1} t_{k+2} \rho_B'$$

$$+ \left((C_A^{\mathrm{LF}} + C_B^{\mathrm{LF}})t_{k+1} + \frac{1}{2}t_k t_{k+1} C_B^{\mathrm{LF}} \right) \tau^2.$$

This completes the proof.   □

We are now able to prove a global error bound.

Theorem 6 *If the assumptions of Lemma 5 are satisfied, then the global errors E_A^{k+1} and $E_B^{k+\frac{1}{2}}$ are bounded by*

$$E_A^{k+1} \le \mathrm{e}^{\sqrt{L} t_{k+1}} M_A^{k+1},$$

where M_A^{k+1} is given in (32), and

$$E_B^{k+\frac{1}{2}} \le \mathrm{e}^{\sqrt{L} t_k} M_B^{k+\frac{1}{2}},$$

for

$$M_B^{k+\frac{1}{2}} = \rho_B + t_{k+\frac{1}{2}} L \rho_A + 7 t_{k+\frac{1}{2}} \rho_B' + \frac{1}{2}t_k t_{k+1} \rho_A' + \left(\frac{1}{2}t_k t_{k+1} C_A^{LF} + (1 + t_k)C_B^{LF} \right) \tau^2,$$

respectively, as long as $t_{k+4} \le T$.

Proof The bound for E_A^{k+1} is a direct consequence of (31) with $j = k+1$. The bound for $E_B^{k+\frac{1}{2}}$ is obtained as for E_A^{k+1} in the proof of Lemma 5, but starting from substituting (19b) into (19a).   □

The error of the LRLF scheme is hence a combination of two error contributions: an error caused by the low-rank approximations, and a time discretization error stemming from the leapfrog scheme. If the low-rank errors ρ_A, ρ_B, ρ_A', and ρ_B' are small, i.e.,

the solutions A, B of (1) are well-approximated by low-rank matrices, the time discretization error dominates.

In the proofs of Lemma 5 and Theorem 6 we used the following Gronwall-type lemma.

Lemma 7 *Let $\tau, L \geq 0$ and $\{M_k\}_{k\geq0}$ a nonnegative, monotonically increasing sequence. If the nonnegative sequence $\{E_k\}_{k\geq0}$ satisfies*

$$E_k \leq M_k + \tau^2 L \sum_{j=0}^{k-1} (k-j) E_j,$$

then

$$E_k \leq M_k \, e^{\tau k \sqrt{L}}.$$

Proof Define $\varepsilon_k := E_k / M_k$ for all $k \geq 0$. The sequence $\{\varepsilon_k\}_{k\geq0}$ is nonnegative and satisfies

$$\varepsilon_k \leq 1 + \tau^2 L \sum_{j=0}^{k-1} (k-j) \varepsilon_j$$

due to the monotonicity of $\{M_k\}_{k\geq0}$. The statement then follows from [1, Lemma 3.8].
$\square$

5 Dynamical low-rank integrator for semilinear second-order matrix differential equations

We now consider semilinear second-order matrix differential equations of the form

$$A'' = -\Omega_1^2 A - A \Omega_2^2 + f(A), \quad t \in [0, T], \quad A(0) = A_0, \quad A'(0) = B_0 \quad (35)$$

with given Hermitian, positive semidefinite matrices $\Omega_1 \in \mathbb{C}^{m \times m}$ and $\Omega_2 \in \mathbb{C}^{n \times n}$ of large norm and a Lipschitz continuous function f with a moderate Lipschitz constant. Our aim is to construct a dynamical low-rank integrator which exploits the semilinear structure of the right-hand side of (35) and yields smaller errors than the LRLF scheme.

A dynamical low-rank integrator for stiff semilinear first-order equations was proposed in [15]. It is based on the important property of linear first-order equations, that the exact solution of an initial value problem stays in $\mathscr{M}_r$ for all times if the initial value is in $\mathscr{M}_r$. This is in general not the case for linear second-order equations, so that additional considerations are required here.

We transform (35) into an equivalent first-order problem and split the right-hand side into two linear and one nonlinear part by introducing weights $\omega_i \geq 0$, $i = 1, 2, 3$,

with $\omega_1^2 + \omega_2^2 + \omega_3^2 = 1$, that is

$$\begin{bmatrix} A \\ B \end{bmatrix}' = \begin{bmatrix} B \\ -\Omega_1^2 A - A\Omega_2^2 + f(A) \end{bmatrix} = \begin{bmatrix} \omega_1^2 B \\ -\Omega_1^2 A \end{bmatrix} + \begin{bmatrix} \omega_2^2 B \\ -A\Omega_2^2 \end{bmatrix} + \begin{bmatrix} \omega_3^2 B \\ f(A) \end{bmatrix}.$$

A natural choice is $\omega_i^2 = 1/3$, $i = 1, 2, 3$, but different weightings are feasible. The split equations can be written as

$$\begin{bmatrix} A \\ B \end{bmatrix}' = \begin{bmatrix} \omega_1^2 B \\ -\Omega_1^2 A \end{bmatrix} = \begin{bmatrix} 0 & \omega_1^2 I \\ -\Omega_1^2 & 0 \end{bmatrix} \begin{bmatrix} A \\ B \end{bmatrix}, \tag{36a}$$

$$\begin{bmatrix} A & B \end{bmatrix}' = \begin{bmatrix} \omega_2^2 B & -A\Omega_2^2 \end{bmatrix} = \begin{bmatrix} A & B \end{bmatrix} \begin{bmatrix} 0 & -\Omega_2^2 \\ \omega_2^2 I & 0 \end{bmatrix}, \tag{36b}$$

$$\begin{bmatrix} A \\ B \end{bmatrix}' = \begin{bmatrix} \omega_3^2 B \\ f(A) \end{bmatrix}. \tag{36c}$$

The solution of the linear problems (36a) and (36b) can be expressed in terms of the matrix exponential

$$\exp\left(t \begin{bmatrix} 0 & \omega_i^2 I \\ -\Omega_j^2 & 0 \end{bmatrix} \right) = \begin{bmatrix} \cos(\omega_i t \Omega_j) & \omega_i^2 t \, \mathrm{sinc}(\omega_i t \Omega_j) \\ -t\Omega_j^2 \, \mathrm{sinc}(\omega_i t \Omega_j) & \cos(\omega_i t \Omega_j) \end{bmatrix}, \quad j = 1, 2.$$

The full splitting scheme reads

$$\begin{bmatrix} \widehat{A}_1 \\ \widehat{B}_1 \end{bmatrix} = \left(\phi_{\frac{\tau}{2}}^{\Omega_1} \circ \phi_{\frac{\tau}{2}}^{\Omega_2} \circ \phi_\tau^{\mathscr{S}} \circ \phi_{\frac{\tau}{2}}^{\Omega_2} \circ \phi_{\frac{\tau}{2}}^{\Omega_1} \right) \begin{bmatrix} \widehat{A}_0 \\ \widehat{B}_0 \end{bmatrix}, \tag{37}$$

where $\phi_{\frac{\tau}{2}}^{\Omega_1}$ and $\phi_{\frac{\tau}{2}}^{\Omega_2}$ denote the numerical flows given by Algorithm 1 with step size $\frac{\tau}{2}$ to the exact solutions of (36a) and (36b), respectively. $\phi_\tau^{\mathscr{S}}$ denotes the numerical flow of the LRLF(ω) scheme, described in Algorithm 3, with right-hand side f. The overall method (37) is called LRLF-semi.

When approximations at full time steps are dispensable, the last half step $\phi_{\frac{\tau}{2}}^{\Omega_1}$ in (37) can be combined with the first one of the next time step.

6 Numerical experiments

In this section, we present numerical experiments to confirm our theoretical findings. We consider the wave equation

$$\frac{\partial^2}{\partial t^2} a(t, x, y) = \Delta a(t, x, y) + \gamma a^3(t, x, y), \qquad t \in [0, T], \quad (x, y) \in \Omega, \tag{38a}$$

$$a(0, x, y) = a_0(x, y), \qquad \frac{\partial}{\partial t} a(0, x, y) = b_0(x, y), \tag{38b}$$

with periodic boundary conditions on a rectangular domain $\Omega \subset \mathbb{R}^2$.

The codes for reproducing the numerical results are available on https://doi.org/10.5445/IR/1000152590.

6.1 Homogeneous wave equation

In our first example we set $\Omega = [-\pi, \pi]^2$ and $\gamma = 0$. With the initial values

$$a_0(x, y) = \frac{1}{2} \sin\left(-2(k_x x + k_y y)\right), \qquad b_0(x, y) = \sqrt{2} \cos\left(-2(k_x x + k_y y)\right),$$

where $k_x, k_y \in \mathbb{R}$, (38) models a planar wave traveling into the direction of $k = (k_x, k_y)$. For the discretization in space we use second-order finite differences with n and m discretization points in x- and y-direction, respectively. This yields the second-order matrix differential equation (35) with $f \equiv 0$, where Ω_1^2 and Ω_2^2 denote the positive semidefinite, symmetric Toeplitz matrices with first rows

$$\left(\frac{n}{2\pi}\right)^2 [2, -1, 0, \dots, -1] \quad \text{and} \quad \left(\frac{m}{2\pi}\right)^2 [2, -1, 0, \dots, -1],$$

respectively.

For $m = n = 512$, $T = 10$ and $2k_x = k_y = 2$, we compute low-rank approximations to the exact solution of (35) with the LRLF and the LRLF-semi schemes and ranks $r_A = r_B = 2$. For the latter method, we use the weight configurations

$$\omega_1^2 = \omega_2^2 = \frac{1}{2}, \quad \text{and} \quad \omega_1^2 = \sin^2 \varphi, \quad \omega_2^2 = \cos^2 \varphi, \quad \varphi = \arctan \frac{k_y}{k_x}.$$

While the first choice treats the two subproblems equally, the second takes the direction of motion into account. Moreover, we compute low-rank approximations with the projector-splitting integrator (Algorithm 1), applied to the equivalent first-order system (7) of (35) and rank $r = 4$. For the inner integration we use the explicit Euler method (PSI) and the classical Runge–Kutta method of order 4 (PSI-RK).

In Fig. 1, the relative global error between the exact solution A and its low-rank approximations $\widehat{A}$ at time $t = T$ is displayed. Temporal convergence of order 2 is observed for both the LRLF and the LRLF-semi scheme. Compared to the LRLF method, the LRLF-semi scheme is more accurate and allows slightly larger step sizes. Choosing the weights ω_i in consideration of the direction of motion improves the approximation substantially. In contrast, the PSI and PSI-RK methods require significantly smaller step sizes τ and come with larger errors than the schemes for second-order matrix differential equations.
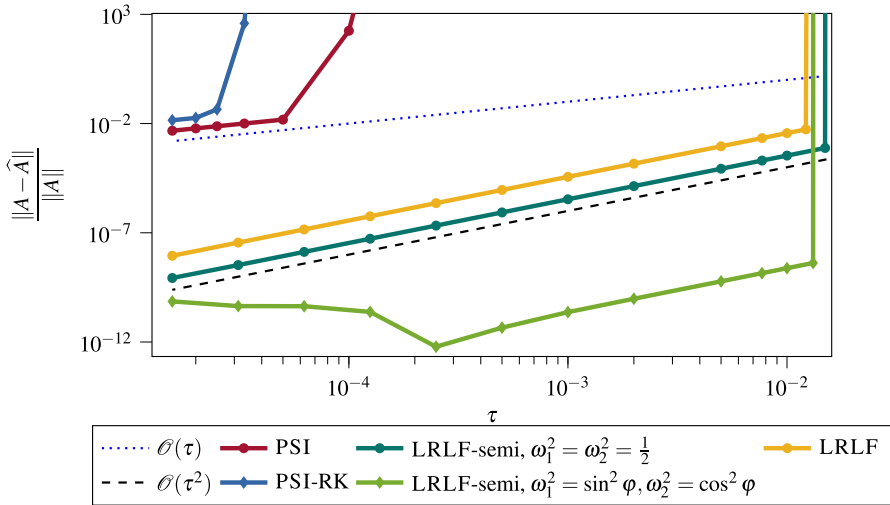
6.2 Semilinear wave equation

Fig. 1 Example from Sect. 6.1: Relative error in Frobenius norm between the exact solution of (35) with $f \equiv 0$ and its low-rank approximations computed with different dynamical low-rank integrators at time $t = 10$

In a second experiment, we set $\Omega = [-\pi, \pi] \times [-2\pi, 2\pi]$ and $\gamma = 0.01$ in (38), and use the initial values

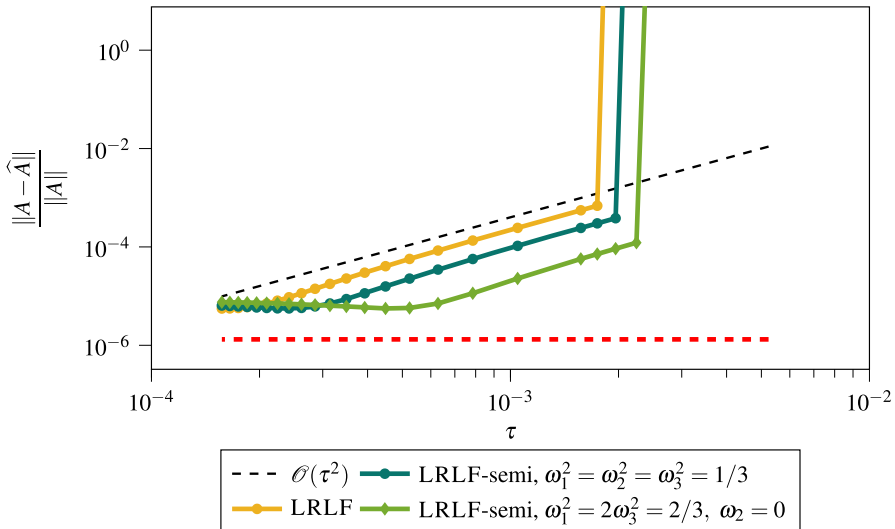$$a_0(x, y) = 0.1 \exp\left(-\frac{y^2}{l_0^2}\right) \exp\left(-\frac{x^2}{w_0^2}\right), \qquad b_0(x, y) = 2\frac{y}{l_0^2} a_0(x, y),$$

Fig. 2 Example from Sect. 6.2: Relative error in the Frobenius norm between the reference solution of (35) with f given in (39) and the low-rank approximations computed by LRLF and LRLF-semi at time $t = 0.5\pi$. The red dashed line shows the relative best-approximation error, cf. (40)

with $l_0 = \pi/30$ and $w_0 = \pi/3$. Due to the strong localization of the initial value, we discretize in space by a pseudospectral method in y-direction and fourth-order finite differences in x-direction. This yields the second-order matrix differential equation (35) with the cubic nonlinearity

$$f(A) = \gamma \, A \bullet A \bullet A, \tag{39}$$

where the symbol $\bullet$ denotes the entrywise matrix product. The matrix $\Omega_1^2 \in \mathbb{R}^{m \times m}$ is given by

$$\Omega_1^2 = \frac{1}{4} \mathscr{F}_m^{-1} \operatorname{diag}\left(0, \ldots, \frac{m}{2} - 1, -\frac{m}{2}, \ldots, -1\right) \mathscr{F}_m,$$

where $\mathscr{F}_m$ denotes the discrete Fourier transformation operator with m modes, and $\Omega_2^2 \in \mathbb{R}^{n \times n}$ denotes the positive semidefinite, symmetric Toeplitz matrix with first row

$$\frac{1}{12}\left(\frac{n}{2\pi}\right)^2 [30, -16, 1, 0, \ldots, 1, -16].$$

For $m = 4096, n = 512, T = 0.5\pi$, and $r_A = r_B = 10$, we compute low-rank approximations to the solution of (35) with the LRLF and the LRLF-semi scheme. A reference solution has been computed with the leapfrog scheme and step size $\tau_0 = 10^{-4}T$. For the LRLF-semi scheme, we again use two weight configurations, where the first treats all subproblems equally and the second takes the direction of propagation into account,

$$\omega_1^2 = \omega_2^2 = \omega_3^2 = \frac{1}{3}, \quad \text{and} \quad \omega_1^2 = \frac{2}{3}, \ \omega_2 = 0, \ \omega_3^2 = \frac{1}{3}.$$

In Fig. 2, the relative global error of the low-rank approximations compared to the reference solution at time $t = T$ is displayed. Both LRLF and LRLF-semi schemes show convergence of order 2. For both weight configurations the LRLF-semi scheme yields smaller errors, and the approximation quality improves if the weights are chosen according to the direction of motion. For all integrators, the error reaches a plateau for small τ. This is caused by the low rank best approximation error to the exact solution. In fact, the rank-r_A best-approximation $\widehat{A}_{r_A}^{\text{best}}$ to the exact solution $A \in \mathbb{C}^{m \times n}$ satisfies

$$\frac{\|A - \widehat{A}_{r_A}^{\text{best}}\|^2}{\|A\|^2} = \frac{\sigma_{r_A+1}^2 + \cdots + \sigma_n^2}{\|A\|^2}, \tag{40}$$

which is a lower bound on the relative error of an arbitrary low-rank approximation.

The error plateau observed in Fig. 2 suggests to choose the rank adaptively such that the temporal convergence order is preserved for smaller step sizes. This issue has been addressed in [7].

7 Conclusion and outlook

In the present paper, we developed and analyzed dynamical low-rank integrators for second-order matrix differential equations of the forms (1) or (35). For LRLF we proved second-order convergence under reasonable assumptions. The theoretical results have been confirmed by numerical experiments.

Both the projector-splitting integrator and the unconventional robust integrator have been successfully adapted to first-order tensor differential equations, cf. [3, 13, 14] and references therein. We are confident that the constructed dynamical low-rank integrators for second-order matrix differential equations can be adapted to the tensor case as well. This is part of ongoing research and will be reported in the future.

Declarations

Conflict of interest Not applicable.

References

1. Carle, C., Hochbruck, M., Sturm, A.: On leapfrog-Chebyshev schemes. SIAM J. Numer. Anal. **58**(4), 2404–2433 (2020). https://doi.org/10.1137/18M1209453
2. Ceruti, G., Lubich, C.: An unconventional robust integrator for dynamical low-rank approximation. BIT **62**(1), 23–44 (2022). https://doi.org/10.1007/s10543-021-00873-0
3. Ceruti, G., Lubich, C., Walach, H.: Time integration of tree tensor networks. SIAM J. Numer. Anal. **59**(1), 289–313 (2021). https://doi.org/10.1137/20M1321838
4. Einkemmer, L., Lubich, C.: A low-rank projector-splitting integrator for the Vlasov–Poisson equation. SIAM J. Sci. Comput. **40**(5), B1330–B1360 (2018). https://doi.org/10.1137/18M116383X
5. Einkemmer, L., Ostermann, A., Piazzola, C.: A low-rank projector-splitting integrator for the Vlasov–Maxwell equations with divergence correction. J. Comput. Phys. **403**, 109,063 (2020)
6. Hairer, E., Lubich, C., Wanner, G.: Geometric numerical integration illustrated by the Störmer–Verlet method. Acta Numer. **12**, 399–450 (2003). https://doi.org/10.1017/S0962492902000144
7. Hochbruck, M., Neher, M., Schrammer, S.: Dynamical low-rank integrators for second-order matrix differential equations. CRC 1173 Preprint 2022/12, Karlsruhe Institute of Technology (2022). https://doi.org/10.5445/IR/1000143198. https://www.waves.kit.edu/downloads/CRC1173_Preprint_2022-13.pdf

8. Joly, P., Rodríguez, J.: Optimized higher order time discretization of second order hyperbolic problems: construction and numerical study. J. Comput. Appl. Math. **234**(6), 1953–1961 (2010). https://doi.org/10.1016/j.cam.2009.08.046

9. Kieri, E., Lubich, C., Walach, H.: Discretized dynamical low-rank approximation in the presence of small singular values. SIAM J. Numer. Anal. **54**(2), 1020–1038 (2016). https://doi.org/10.1137/15M1026791

10. Koch, O., Lubich, C.: Dynamical low-rank approximation. SIAM J. Matrix Anal. Appl. **29**(2), 434–454 (2007). https://doi.org/10.1137/050639703

11. Kusch, J., Ceruti, G., Einkemmer, L., Frank, M.: Dynamical low-rank approximation for Burgers' equation with uncertainty. Int. J. Uncertain. Quantif. **12**(5), 1–21 (2022). https://doi.org/10.1615/int.j.uncertaintyquantification.2022039345

12. Lubich, C., Oseledets, I.V.: A projector-splitting integrator for dynamical low-rank approximation. BIT **54**(1), 171–188 (2014). https://doi.org/10.1007/s10543-013-0454-0

13. Lubich, C., Oseledets, I.V., Vandereycken, B.: Time integration of tensor trains. SIAM J. Numer. Anal. **53**(2), 917–941 (2015). https://doi.org/10.1137/140976546

14. Lubich, C., Vandereycken, B., Walach, H.: Time integration of rank-constrained Tucker tensors. SIAM J. Numer. Anal. **56**(3), 1273–1290 (2018). https://doi.org/10.1137/17M1146889

15. Ostermann, A., Piazzola, C., Walach, H.: Convergence of a low-rank Lie–Trotter splitting for stiff matrix differential equations. SIAM J. Numer. Anal. **57**(4), 1947–1966 (2019). https://doi.org/10.1137/18M1177901

16. Schrammer, S.: On dynamical low-rank integrators for matrix differential equations. Ph.D. thesis, Karlsruher Institut für Technologie (KIT) (2022). https://doi.org/10.5445/IR/1000148853