SIAM J. NUMER. ANAL.
Vol. 57, No. 4, pp. 1947–1966

CONVERGENCE OF A LOW-RANK LIE–TROTTER SPLITTING FOR STIFF MATRIX DIFFERENTIAL EQUATIONS*

ALEXANDER OSTERMANN†, CHIARA PIAZZOLA†, AND HANNA WALACH‡

Abstract. We propose a numerical integrator for determining low-rank approximations to solutions of large-scale matrix differential equations. The considered differential equations are semilinear and stiff. Our method consists of first splitting the differential equation into a stiff and a nonstiff part, respectively, and then following a dynamical low-rank approach. We conduct an error analysis of the proposed procedure, which is independent of the stiffness and robust with respect to possibly small singular values in the approximation matrix. Following the proposed method, we show how to obtain low-rank approximations for differential Lyapunov and for differential Riccati equations. Our theory is illustrated by numerical experiments.

Key words. matrix differential equation, differential Lyapunov equation, differential Riccati equation, dynamical low-rank approximation, low-rank splitting

AMS subject classifications. 65L04, 65L20, 65M12, 65F30, 49J20

DOI. 10.1137/18M1177901

1. Introduction.

Dynamical low-rank approximations of matrices are widely used for reducing models of large size. Such an approach has a broad variety of application areas, such as control theory, computer algebra, signal processing, machine learning, image compression, and quantum molecular systems; see, e.g., [5, 37, 29, 22] and references therein. We are interested here in particular in computing low-rank approximations to solutions of large-scale matrix differential equations.

In this paper we consider a class of semilinear stiff matrix differential equations of the form

$$\dot{X}(t) = AX(t) + X(t)A^* + G(t, X(t)), \qquad X(t_0) = X^0,$$

where $X(t) \in \mathbb{C}^{m\times m}$, $G : [t_0, \infty) \times \mathbb{C}^{m\times m} \to \mathbb{C}^{m\times m}$ is nonlinear, and $A \in \mathbb{C}^{m\times m}$ is time invariant. In many applications, the matrix A arises from the spatial discretization of a differential operator. Therefore, it gives rise to a stiff term. The nonlinearity G, however, is assumed to be nonstiff. The objective of this paper is to determine a low-rank approximation to the solution of the given matrix differential equation.

A possible method for obtaining low-rank approximations to solutions of matrix differential equations is the dynamical low-rank approximation proposed in [19]. This approach yields a differential equation for the approximation matrix on the low-rank manifold. Recently, an efficient integrator, the so-called projector-splitting integrator, was proposed in [23] for computing the solution numerically. A comprehensive error analysis for this integration method is given in [18]. Note that the error bounds in this analysis depend on the Lipschitz constant of the right-hand side of the considered matrix differential equation, amongst others. Therefore, this proof does not extend to the present situation in an obvious way.

*Received by the editors March 28, 2018; accepted for publication (in revised form) May 28, 2019; published electronically August 13, 2019.
https://doi.org/10.1137/18M1177901

†Institut für Mathematik, Universität Innsbruck, Austria (alexander.ostermann@uibk.ac.at, chiara.piazzola@uibk.ac.at).

‡Mathematisches Institut, Universität Tübingen, Germany (walach@na.uni-tuebingen.de).

In this work we propose a novel approach, which yields low-rank approximations for stiff matrix differential equations. Our method is derived in two steps. To handle the difficulty with the stiffness, we first split the matrix differential equation into its stiff part $AX + XA^*$ and the nonstiff nonlinearity G. Second, we follow the concept of the dynamical low-rank approximation for both arising subproblems. The linear subproblem can be solved exactly and efficiently by means of exponential integrators and the rank of this solution is preserved. The nonlinearity G is integrated with the projector-splitting integrator [23]. We conduct a convergence analysis for the proposed method where, because of the beneficial way of splitting, we succeed in giving error bounds which are independent of the norm of A.

Moreover, our integration method is independent of small singular values, which might appear in the approximation matrix. When following the original approach of [19], one would have to solve a modified differential equation whose right-hand side has a Lipschitz constant inversely proportional to the smallest singular value of the approximation matrix. This leads to computational difficulties; see [19, 18]. The robustness of our integrator with respect to the presence of small singular values is inherited from the projector-splitting integrator (see [18]) and could be exploited to change the rank adaptively. This would require an appropriate error monitor with respect to the choice of the approximation rank. Such strategy is feasible, but we do not address this matter here.

It is possible to extend our approach to tensor differential equations. There, the dynamical low-rank approximation to tensors of different formats, such as tensor trains [24], Tucker tensors [26], or hierarchical Tucker tensors [25] can be applied.

The paper is structured as follows. In section 2 we illustrate the quality of the proposed approach with the help of a numerical example. In section 3, we derive our method in detail. In sections 4 and 5 we conduct a comprehensive error analysis and give error bounds, which are independent of the norm of A and of small singular values, which might appear in the approximation matrix. Some extensions and further convergence results are given in section 6. After having presented our approach and its convergence analysis, we show how to apply the method to two essential representatives of this class of matrix differential equations: differential Lyapunov equations (DLEs), and differential Riccati equations (DREs). Finally, we illustrate our theoretical result by a numerical experiment. Some complementary numerical experiments can be found in [28].

2. A motivating example. In this work we are interested in low-rank approximations to solutions of semilinear stiff differential equations. The integrator we propose here is a first-order method based on a splitting, which separates the stiff linear part from the nonstiff nonlinear one. The solutions of the two arising subproblems are approximated by low-rank matrices. The linear subproblem can be integrated efficiently by an exponential integrator, whereas the solution of the nonlinear differential equation is approximated by the dynamical low-rank method [19]. The arising differential equation for the nonlinearity on the low-rank manifold is finally integrated by the projector-splitting integrator [23].

The main advantage of the integration method we propose is its insensitivity to stiffness. We illustrate this favorable behavior with the help of an example.

Consider the following two-dimensional partial differential equation in the variable $v(t, x, y)$:

$$\partial_t v = \alpha \Delta v + v^3, \qquad v(0, x, y) = 16\, x(1 - x)y(1 - y),$$

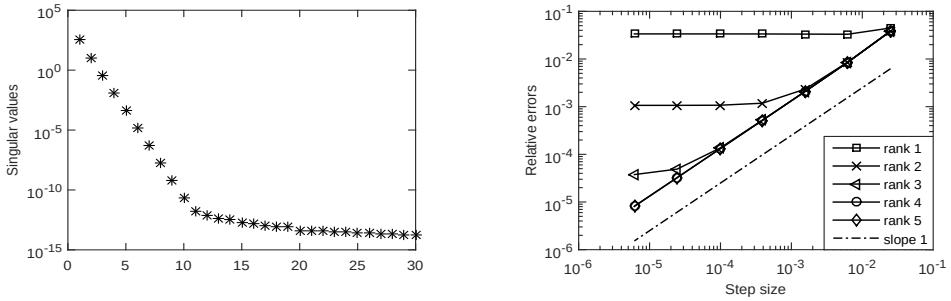

FIG. 1. *Results for the solution of the considered partial differential equation at $T = 0.5$. Left: First 30 singular values of the reference solution computed with DOPRI5. Right: Error of our proposed first-order splitting as a function of the step size and the approximation rank.*

where $\alpha = 1/50$. We solve this problem on the spatial domain $\Omega = [0, 1]^2$, subject to homogeneous Dirichlet boundary conditions, for times $0 \leq t \leq T$. We discretize this partial differential equation in space with m inner points in each direction and denote the grid size by h, which is $h = \frac{1}{m+1}$. The inner grid points in the x and y directions are denoted by

$$x_i = ih \quad \text{and} \quad y_j = jh \quad \text{for} \quad 1 \leq i, j \leq m,$$

respectively. The differential operator is discretized by means of second-order standard finite differences. Denoting the one-dimensional stencil matrices in the x and y directions by A_x and A_y, respectively, this results in the matrix differential equation

$$\dot{U}(t) = \alpha A[U(t)] + U(t)^3, \qquad U(0) = U_0,$$

where $A[U(t)] = A_x U(t) + U(t)A_y$ and $U(t) \in \mathbb{R}^{m \times m}$. The component $U_{ij}(t)$ is the sought after approximation of $v(t, x_i, y_j)$, $1 \leq i, j \leq m$. The nonlinearity is realized by an entrywise product.

In our numerical experiment we choose $m = 500$. The reference solution is computed with DOPRI5 applied to the equivalent vector-valued equation. It is a Runge–Kutta method of order 5 with adaptive step size strategy [13] with high precision.

In Figure 1 (left), we plot the first 30 singular values of the reference solution at $T = 0.5$. We observe that the singular values decay quite fast. Figure 1 (right) shows the errors of our proposed method for different approximation ranks. The error is measured in the Frobenius norm. The figure suggests an explicit dependence of the error on the rank and on the step size. If the approximation rank is chosen sufficiently large (ranks 4 and 5) we solely observe the first-order error due to the splitting into the linear and the nonlinear subproblems. On the other hand, a bad choice of the approximation rank (ranks 1, 2, and 3) leads to a stagnation of the error, independently of the refinement of the time step size.

Note that standard explicit integrators would have to satisfy a CFL-like condition of the form $\tau L \leq 1$, where L is the Lipschitz constant of the right-hand side of the matrix differential equation for $U(t)$ and τ is the time step size. For our choice of the parameters the time step size would be limited by the condition $\tau \leq L^{-1} \approx 2 \cdot 10^{-4}$. Our integrator, on the other hand, works fine for much larger time step sizes.

Higher-dimensional partial differential equations lead to *tensor differential equations* for the approximation *tensor* $U(t)$. The differential operator has to be appropriately discretized, e.g., for the three-dimensional Laplacian the seven-point stencil

tensor is obtained. We determine a low-rank solution by solving the linear part exactly and, afterwards, depending on the underlying low-rank format of the approximation tensor, such as tensor trains [31, 30] or Tucker tensors [40], we apply the corresponding projector-splitting integrator [24] or [26].

3. A low-rank approximation of stiff matrix differential equations. We consider the following matrix differential equation

$$\dot{X}(t) = AX(t) + X(t)A^* + G(t, X(t)), \qquad X(t_0) = X^0, \tag{1}$$

where $X(t) \in \mathbb{C}^{m\times m}$ and $G : [t_0, \infty) \times \mathbb{C}^{m\times m} \to \mathbb{C}^{m\times m}$. The matrix $A \in \mathbb{C}^{m\times m}$ and its conjugate transpose, denoted by A^*, are time independent. We restrict our attention here to parabolic partial differential equations; other settings are described in section 6. For the moment, the matrix A is typically the spatial discretization of an elliptic differential operator. Therefore, the stiffness of (1) is induced by the matrix A. The nonlinearity G, however, is assumed to be nonstiff. The exact full-rank solution of the above differential equation can be represented by the variation-of-constants formula as

$$X(t) = \mathrm{e}^{(t-t_0)A} X(t_0) \mathrm{e}^{(t-t_0)A^*} + \int_{t_0}^{t} \mathrm{e}^{(t-s)A} G(s, X(s)) \mathrm{e}^{(t-s)A^*} \, \mathrm{d}s.$$

The aim of this work is to compute approximate solutions to $X(t)$, which are of low rank r with $r \ll m$. We propose an integrator based on splitting methods.

The construction of our integrator is described in the following sections. The properties of the scheme are also illustrated.

3.1. Splitting into two subproblems. The structure of the matrix differential equation (1) motivates us to use splitting methods. For an introduction to this class of numerical integrators, we refer to [12] and [27]. The idea behind the proposed splitting method is to benefit from the independent integration of the two arising subproblems.

Now, splitting (1) into a stiff and a nonstiff part yields the following two subproblems:

$$\dot{X}_1(t) = AX_1(t) + X_1(t)A^*, \qquad X_1(t_0) = X_1^0 \tag{2}$$

and

$$\dot{X}_2(t) = G(t, X_2(t)), \qquad X_2(t_0) = X_2^0. \tag{3}$$

We denote the solutions to the subproblems (2) and (3) at time $t_0 + \tau$ with initial values X_1^0 and X_2^0 by $\Phi_\tau^A(X_1^0)$ and $\Phi_\tau^G(X_2^0)$, respectively. Our strategy is to solve the differential equations for $X_2(t)$ and $X_1(t)$. An approximate solution of (1) is then obtained by applying the Lie–Trotter splitting scheme with step size τ:

$$\mathcal{L}_\tau := \Phi_\tau^A \circ \Phi_\tau^G. \tag{4}$$

Note, that we will refer to this scheme as *full-rank Lie–Trotter splitting*. It results in an approximation X^1 of the solution $X(t)$ of (1) at $t = t_0+\tau$. Starting with $X^0 = X_2^0$, we obtain

$$X^1 = \mathcal{L}_\tau(X^0) = (\Phi_\tau^A \circ \Phi_\tau^G)(X^0).$$

Note that the numerical solution at time $t_k = t_0 + k\tau$ is $X^k = \mathcal{L}_\tau^k(X^0)$. The exact solution of the homogeneous problem (2) is given by

$$X_1(t_0 + \tau) = \mathrm{e}^{\tau A} X_1^0 \mathrm{e}^{\tau A^*}.$$

Since X_1^0 is typically given in low-rank factorized form (see subsection 3.3 below), X_1 is the result of the action of a matrix exponential. Therefore, it can also be efficiently computed for large step sizes τ. Methods of choice are Taylor interpolation [2], interpolation at Leja points [6], and Krylov subspace methods [36]. Moreover, efficient implementations on GPUs are possible; see, e.g., [9].

The approximate solution X^1 is a full-rank matrix approximation to $X(t_1)$ after one time step. Since we aim to compute rank-r approximations to $X(t)$ at the time grid points, we next determine low-rank solutions of (2) and (3).

3.2. The low-rank integrator. Denoting the manifold of rank-r matrices by

$$\mathcal{M} := \left\{ Y(t) \in \mathbb{C}^{m \times m} : \text{rank } Y(t) = r \right\},$$

we seek a low-rank approximation $Y \in \mathcal{M}$ to the solution of (1). In subsection 3.1, we have already shown how to split the differential equation into (2) and (3). Now, it is the objective to determine low-rank approximations $Y_1 \in \mathcal{M}$ and $Y_2 \in \mathcal{M}$ to X_1 and X_2, respectively. To this end, we denote by $\mathcal{T}_Y\mathcal{M}$ the tangent space of the low-rank manifold $\mathcal{M}$ at a rank-r matrix Y.

We first consider the stiff subproblem (2). We observe that for any $Y \in \mathcal{M}$, $AY + YA^* \in \mathcal{T}_Y\mathcal{M}$ and, thus, (2) defines a vector field on the low-rank manifold $\mathcal{M}$. Hence for an initial value on the low-rank manifold $\mathcal{M}$, the solution of (2) stays in $\mathcal{M}$; see [15]. This means that subproblem (2) is rank preserving and so starting with a rank-r initial value Y_1^0, the solution of

$$\dot{Y}_1(t) = AY_1(t) + Y_1(t)A^*, \qquad Y_1(t_0) = Y_1^0, \tag{5}$$

remains of rank-r for all times.

For the second subproblem (3) we employ the dynamical low-rank approach [19]. There, a rank-r solution $Y_2(t)$ is determined by requiring

$$\dot{Y}_2(t) \in \mathcal{T}_{Y_2(t)}\mathcal{M}, \qquad \|\dot{Y}_2(t) - \dot{X}_2(t)\| = \min,$$

where $\mathcal{T}_{Y_2(t)}\mathcal{M}$ is the tangent space of the low-rank manifold $\mathcal{M}$ at the current approximation $Y_2(t)$. The above condition is equivalent to orthogonally projecting the right-hand side of (3) onto the tangent space $\mathcal{T}_{Y_2(t)}\mathcal{M}$. This results in an evolution equation for $Y_2(t)$, which is of the form

$$\dot{Y}_2(t) = P(Y_2(t))G(t, Y_2(t)), \qquad Y_2(t_0) = Y_2^0, \tag{6}$$

where the initial value Y_2^0 is on the low-rank manifold $\mathcal{M}$. The orthogonal projection is denoted by P. This differential equation needs to be solved numerically. A favorable integration scheme is the so-called projector-splitting integrator [23] which will be described in detail in subsection 3.3. The authors of [18] have proved that this integrator is robust with respect to the presence of small singular values. This is a crucial property, since in most applications the rank might not be known in advance. For accuracy reasons the rank is often overapproximated and small singular values enter in the approximation matrices. Solving the system of differential equations for

the low-rank factors of the solution, as proposed in [19], becomes cumbersome. Standard integrators such as explicit and implicit Runge–Kutta methods suffer from the possible ill-conditioning of the arising matrices. For further details we refer to the discussions in [23, 18].

After having applied the projector-splitting integrator to (6), the resulting low-rank approximation of $X_2(t)$ at $t_0 + \tau$ is

$$Y_2^1 = \widetilde{\Phi}_\tau^G(Y_2^0),$$

where $\widetilde{\Phi}_\tau^G$ denotes the approximated solution operator of (6). In a nutshell, the integration method we propose consists of first splitting the matrix differential equation (1) and then approximating the subproblems (2) and (3) with respect to low rank. Hence combining the flow $\widetilde{\Phi}_\tau^G$ of the low-rank solution of (3) with the exact flow Φ_τ^A of (2), which is of low rank when starting with low-rank initial data, yields the desired approximation matrix $Y(t)$. We call this procedure *low-rank Lie–Trotter splitting* and denote it by

$$\mathcal{I}_\tau := \Phi_\tau^A \circ \widetilde{\Phi}_\tau^G. \tag{7}$$

Thus, starting with $Y^0 = Y_2^0$, we obtain the rank-r approximation of the solution of (1) at time $t_0 + \tau$, i.e.,

$$Y^1 = \mathcal{I}_\tau(Y^0) = (\Phi_\tau^A \circ \widetilde{\Phi}_\tau^G)(Y^0), \tag{8}$$

where we assume Y^0 to be a rank-r approximation to X^0. At $t_k = t_0 + k\tau$ we obtain $Y^k = \mathcal{I}_\tau^k(Y^0)$.

3.3. The projector-splitting integrator. The low-rank approximation Y_2 is not computed directly from (6) by applying a standard integration method, but by an efficient integrator, which benefits from the underlying low-rank format for matrices. In the following, we drop the subscript in Y_2 and describe the integrator for any $Y \in \mathcal{M}$ and any problem of the form

$$\dot{Y}(t) = P(Y(t))G(t, Y(t)), \qquad Y(t_0) = Y^0 \in \mathcal{M}.$$

The projector-splitting integrator introduced in [23] is based on the observation that every rank-r matrix $Y(t) \in \mathbb{C}^{m\times m}$ can be represented as

$$Y(t) = U(t)S(t)V(t)^*,$$

where $U(t), V(t) \in \mathbb{C}^{m\times r}$ have orthonormal columns. The square matrix $S(t) \in \mathbb{C}^{r\times r}$ is invertible and has the same nonzero singular values as $Y(t)$. In contrast to the singular value decomposition (SVD), this nonunique factorization does not require $S(t)$ to be diagonal. From the computational perspective, this representation has the advantage of a significant reduction in memory requirements and computational cost if $r \ll m$.

The projector-splitting integrator makes use of this SVD-like factorization, in the sense that the time integration is performed only on the low-rank factors. It is based on splitting the projection $P(Y)$ onto the tangent space $\mathcal{T}_Y\mathcal{M}$ of the low-rank manifold $\mathcal{M}$. Following [19, Lemma 4.1], the orthogonal projection $P(Y)$ at the current approximation matrix $Y = USV^* \in \mathcal{M}$ can be written as

$$\begin{aligned} P(Y)G(t,Y) &= UU^*G(t,Y) - UU^*G(t,Y)VV^* + G(t,Y)VV^* \\ &=: P^a(Y)G(t,Y) - P^b(Y)G(t,Y) + P^c(Y)G(t,Y). \end{aligned} \tag{9}$$

Further, UU^* and VV^* are orthogonal projections onto the spaces spanned by the range and corange of Y, respectively. One time step from $t_0 \to t_1 = t_0 + \tau$ of the first-order integrator consists of solving the evolution equations

$$\begin{aligned}
\dot{Y}^a(t) &= P^a(Y)G(t,Y), \qquad & Y^a(t_0) &= Y^0,\\
\dot{Y}^b(t) &= -P^b(Y)G(t,Y), \qquad & Y^b(t_0) &= Y^a(t_1),\\
\dot{Y}^c(t) &= P^c(Y)G(t,Y), \qquad & Y^c(t_0) &= Y^b(t_1),
\end{aligned}$$

consecutively, where $Y^c(t_1)$ is the approximate solution to $Y(t_1)$. In practice, these differential equations have to be solved approximately using a numerical method, e.g., a Runge–Kutta method. Higher-order methods can be obtained from the first-order scheme by employing the standard composition techniques; see [23, 18].

4. The main convergence result. In this section we describe the framework in which the convergence proof can be carried out and formulate the main convergence result. Further, we give an outline of the proof. The technical details are postponed to section 5. It is worth remarking that the convergence analysis of the low-rank Lie–Trotter splitting (7) is performed without introducing Lipschitz conditions of the full right-hand side of (1) nor of the stiff subproblem (2).

Let us consider the Hilbert space $\mathbb{C}^{m\times m}$, endowed with the Frobenius norm $\|\cdot\|$. Let $A \in \mathbb{C}^{m\times m}$ and $G : [t_0, T] \times \mathbb{C}^{m\times m} \to \mathbb{C}^{m\times m}$. In the following, we are given an initial datum X^0 and a final integration time T such that the matrix differential equation (1) has a solution $X(t)$ for $t_0 \le t \le T$. We assume that, given a rank-r approximation Y^0 of the initial value X^0 such that

$$\|X^0 - Y^0\| \le \delta$$

for some $\delta \ge 0$, the exact rank-r solution

$$Y(t) = \mathrm{e}^{(t-t_0)A} Y^0 \mathrm{e}^{(t-t_0)A^*} + \int_{t_0}^{t} \mathrm{e}^{(t-s)A} P(Y(s)) G(s, Y(s)) \mathrm{e}^{(t-s)A^*}\, \mathrm{d}s$$

of the matrix differential equation (1) exists for $t_0 \le t \le T$.

For proving convergence, we further need the following assumption.

Assumption 1. We assume that the following properties hold.

(a) There exists $\omega \in \mathbb{R}$ and $C_s > 0$, such that the matrix A satisfies

$$\|\mathrm{e}^{tA} Z \mathrm{e}^{tA^*}\| \le \mathrm{e}^{t\omega} \|Z\|, \tag{10}$$

$$\|\mathrm{e}^{tA}(AZ + ZA^*)\mathrm{e}^{tA^*}\| \le \frac{1}{t} C_s \mathrm{e}^{t\omega} \|Z\| \tag{11}$$

for all $t > 0$ and all $Z \in \mathbb{C}^{m\times m}$.

(b) G is continuously differentiable in a neighborhood of the exact solution.

(c) There exists $\varepsilon > 0$ such that for all $t_0 \le t \le T$

$$G(t, Y(t)) = M(t, Y(t)) + R(t, Y(t)),$$

where $M(t, Y(t)) \in \mathcal{T}_{Y(t)}\mathcal{M}$ and $\|R(t, Y(t))\| \le \varepsilon$.

The above assumptions require some explanation and discussion. Moreover, we need to specify some crucial properties for the proof of the error bounds given in Theorem 1.

(a) Matrix differential equations of the form (1) typically stem from parabolic partial differential equations. We refer to, e.g., the example in section 2 and to the discussion about DLEs and DREs in sections 7 and 8, respectively. The matrix operator F, given by

$$F(X) = AX + XA^*, \quad X \in \mathbb{C}^{m\times m},$$

is equivalent to the operator $\mathcal{F}$,

$$\mathcal{F}(x) = \mathcal{A}x = (I_m \otimes A + A \otimes I_m)x, \quad x = \mathrm{vec}(X) \in \mathbb{C}^{m^2},$$

where we denote by $\otimes$ the Kronecker product and by $\mathrm{vec}(\cdot)$ the columnwise vectorization of a matrix into a column vector. Then, the bounds (10) and (11) can be translated using the vector 2-norm $\|\cdot\|_2$ as

$$\|\mathrm{e}^{t\mathcal{A}}z\|_2 \le \mathrm{e}^{t\omega}\|z\|_2,$$
$$\|\mathrm{e}^{t\mathcal{A}}\mathcal{A}z\|_2 \le \frac{1}{t}C_s\mathrm{e}^{t\omega}\|z\|_2$$

for all $t > 0$ and all $z = \mathrm{vec}(Z) \in \mathbb{C}^{m^2}$. These properties are well known in the context of semigroup theory for strongly elliptic operators; see, e.g., [11], [32]. In particular, the Lumer–Phillips theorem [35, sect. 12] provides a practical criterion for generators of quasicontraction semigroups in Hilbert spaces. This theorem also applies to standard space discretizations of such operators and in particular shows that the constants ω and C_s can be chosen independently of m. In particular, they are independent of the problem's stiffness.

(b) As a consequence of Assumption 1(b), the function G is locally Lipschitz continuous with constant L, and G is bounded by B in a neighborhood of the solution $X(t)$, i.e.,

$$\begin{aligned} \|G(t,\widehat{X}) - G(t,\widetilde{X})\| &\le L\|\widehat{X} - \widetilde{X}\|, \\ \|G(t,\bar{X})\| &\le B, \end{aligned} \tag{12}$$

as long as $\|\widehat{X} - X(t)\| \le \gamma$, $\|\widetilde{X} - X(t)\| \le \gamma$, and $\|\bar{X} - X(t)\| \le \gamma$ for $t_0 \le t \le T$ for given $\gamma > 0$. The constants L and B depend on γ.

(c) We assume that $G(t,Y)$ consists of a tangential part $M(t,Y)$ and a small perturbation term $R(t,Y)$. This means that G, when evaluated along the low-rank solution, is in the tangent space up to a small remainder of size ε. This assumption is crucial in order to have a good low-rank approximation, since if the remainder is large, low-rank approximation is inappropriate.

Having clarified the assumption, we are now in a position to state the main result of this paper.

THEOREM 1 (global error of the low-rank Lie–Trotter splitting integrator). *Under Assumption* 1, *there exists* $\tau_0 > 0$ *such that for all step sizes* $0 < \tau \le \tau_0$ *the error of the low-rank Lie–Trotter splitting integrator* (7) *is uniformly bounded on* $t_0 \le t_0 + n\tau \le T$ *by*

$$\|X(t_0 + n\tau) - \mathcal{I}_\tau^n(Y^0)\| \le c_0\tau(1 + |\log\tau|) + c_1\delta + c_2\varepsilon,$$

where c_0, c_1, *and* c_2 *depend on* ω, C_s, L, B, *and* T, *but are independent of* τ *and* n.

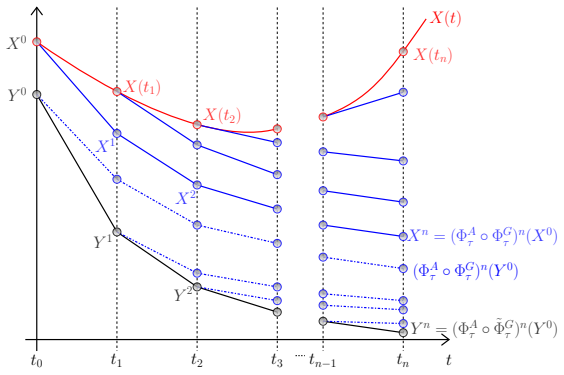

FIG. 2. *Schematic illustration of the convergence analysis. The uppermost curve (in red) depicts the exact solution $X(t)$ of* (1)*, whereas the lowermost (in black) shows the solution obtained by the low-rank Lie–Trotter splitting* (7)*. All other lines (in blue) represent the auxiliary values obtained by the application of the full-rank Lie–Trotter splitting* (4) *either to a full-rank initial datum (continuous lines) or to a low-rank initial datum (dash-dotted lines).*

Note that τ_0 depends only on the size of the Lipschitz constant of G.

In order to facilitate the analysis of (7), we study the global error by introducing auxiliary quantities. The construction of the method already suggests that the global error is composed of the following terms:

(i) The global error of the full-rank Lie–Trotter splitting (4), applied to (2) and (3):

$$E_{sp}^n = X(t_0 + n\tau) - (\Phi_\tau^A \circ \Phi_\tau^G)^n(X^0).$$

(ii) The propagation of the difference between the full-rank initial datum X^0 and its low-rank approximation Y^0 by the full-rank Lie–Trotter splitting (4):

$$E_\delta^n = (\Phi_\tau^A \circ \Phi_\tau^G)^n(X^0) - (\Phi_\tau^A \circ \Phi_\tau^G)^n(Y^0).$$

(iii) The difference between the full-rank Lie–Trotter splitting (4) and the low-rank Lie–Trotter splitting (7) applied to Y^0:

$$E_{lr}^n = (\Phi_\tau^A \circ \Phi_\tau^G)^n(Y^0) - (\Phi_\tau^A \circ \widetilde{\Phi}_\tau^G)^n(Y^0).$$

Hence, the global error in Theorem 1 is obtained as the sum of E_{sp}^n, E_δ^n, and E_{lr}^n as illustrated in Figure 2. Those three contributions are studied in detail in the following section.

5. Detailed convergence analysis. The aim of this section is to provide a convergence analysis of the low-rank Lie–Trotter splitting (7). We give a detailed proof of Theorem 1 and in particular we state and prove error bounds for the three contributions listed above. First, we prove the error bound of the full-rank Lie–Trotter splitting in subsection 5.1, followed by the error estimate for the low-rank Lie–Trotter splitting in subsection 5.2. The propagation of the difference between the full and low-rank initial datum requires just the stability of the full-rank Lie–Trotter splitting. This is shown in subsection 5.3.

Note that our proofs rely on the constants L and B in Assumption 1(b). In order to bound these constants, we have to ensure that the numerical approximations stay in a fixed compact neighborhood $\mathcal{U}$ of the exact solution. This follows (by recursion)

from the given proofs, taking into account that the arising constants can be controlled in terms of L, B, and the final time T. An appropriate choice of the maximum step size τ_0 finally guarantees that all considered approximations stay in $\mathcal{U}$.

5.1. The error of the full-rank Lie–Trotter splitting. The convergence of the full-rank splitting scheme (4) is stated in the following proposition. The ideas in the proof can be traced back to, e.g., [10] and [17].

PROPOSITION 1 (global error of the full-rank Lie–Trotter splitting). *Under Assumption* 1, *the full-rank Lie–Trotter splitting* (4) *is first-order convergent, i.e., the error bound*

$$\|X(t_0 + n\tau) - (\Phi_\tau^A \circ \Phi_\tau^G)^n(X^0)\| \le C\tau(1 + |\log \tau|)$$

holds uniformly on $t_0 \le t_0 + n\tau \le T$. *The constant* C *depends on* ω, C_s, L, B, *and* T, *but is independent of* τ *and* n.

Proof. The solution of the matrix differential equation (1) can be expressed by means of the variation-of-constants formula. Given the initial value $X(t_{k-1}) = Z$, the solution at time $t_k = t_{k-1} + \tau$ with step size $\tau > 0$ is

$$X(t_k) = \mathrm{e}^{\tau A} Z \mathrm{e}^{\tau A^*} + \int_0^\tau \mathrm{e}^{(\tau-s)A} G(t_{k-1} + s, X(t_{k-1} + s)) \mathrm{e}^{(\tau-s)A^*} \,\mathrm{d}s.$$

The exact solution of the first full-rank subproblem (2) at t_k with initial value $X_1(t_{k-1}) = X_2(t_k)$ is given by

$$\Phi_\tau^A(X_2(t_k)) = X_1(t_k) = \mathrm{e}^{\tau A} X_2(t_k) \mathrm{e}^{\tau A^*}, \tag{13}$$

whereas the exact solution of the second full-rank subproblem (3) with initial value $X_2(t_{k-1}) = Z$ can be expressed as

$$\Phi_\tau^G(Z) = X_2(t_k) = Z + \tau G(t_{k-1}, Z) + \int_0^\tau (\tau - s) \ddot{X}_2(t_{k-1} + s) \,\mathrm{d}s. \tag{14}$$

Composing (13) with (14) gives the full-rank Lie–Trotter splitting solution

$$\mathcal{L}_\tau(Z) = \mathrm{e}^{\tau A} Z \mathrm{e}^{\tau A^*} + \tau \mathrm{e}^{\tau A} G(t_{k-1}, Z) \mathrm{e}^{\tau A^*} + \int_0^\tau (\tau - s) \mathrm{e}^{\tau A} \ddot{X}_2(t_{k-1} + s) \mathrm{e}^{\tau A^*} \,\mathrm{d}s. \tag{15}$$

The local error of the method at t_k is

$$\begin{aligned} e_{sp}^k &= X(t_k) - \mathcal{L}_\tau(X(t_{k-1})) \\ &= \int_0^\tau \mathrm{e}^{(\tau-s)A} G(t_{k-1} + s, X(t_{k-1} + s)) \mathrm{e}^{(\tau-s)A^*} \,\mathrm{d}s \\ &\quad - \tau \mathrm{e}^{\tau A} G(t_{k-1}, X(t_{k-1})) \mathrm{e}^{\tau A^*} - \int_0^\tau (\tau - s) \mathrm{e}^{\tau A} \ddot{X}_2(t_{k-1} + s) \mathrm{e}^{\tau A^*} \,\mathrm{d}s. \end{aligned}$$

Let $f(s) = \mathrm{e}^{(\tau-s)A} G(t_{k-1} + s, X(t_{k-1} + s)) \mathrm{e}^{(\tau-s)A^*}$. Then the first integral above can be rewritten as

$$\int_0^\tau f(s) \,\mathrm{d}s = \int_0^\tau \left[f(0) + \int_0^s \dot{f}(r) \,\mathrm{d}r \right] \mathrm{d}s.$$

Using the fact that a matrix commutes with its exponential, the derivative of f is

$$\dot{f}(s) = -\mathrm{e}^{(\tau-s)A} \left(AG + GA^* - \frac{\mathrm{d}G}{\mathrm{d}s} \right) \mathrm{e}^{(\tau-s)A^*}.$$

Recall that the function G is assumed to be continuously differentiable in a neighborhood of the exact solution. Hence, employing the boundedness of

$$\ddot{X}_2(t) = \frac{\mathrm{d}G}{\mathrm{d}t}(t) = \partial_t G(t, X_2(t)) + \partial_X G(t, X_2(t)) G(t, X_2(t))$$

we are left with a simpler form of the local error:

$$e_{sp}^k = -\int_0^\tau \int_0^s \mathrm{e}^{(\tau-r)A} \left(AG + GA^*\right) \mathrm{e}^{(\tau-r)A^*} \,\mathrm{d}r \,\mathrm{d}s + \mathcal{O}(\tau^2). \tag{16}$$

Due to the presence of the matrix A, we do not bound the local error (16) directly. Instead, we solve the error recursion first. Recalling that $X^{n-1} = \mathcal{L}_\tau^{n-1}(X^0)$, we write the global error of the Lie–Trotter splitting as

$$E_{sp}^n = \mathcal{L}_\tau(X(t_{n-1})) - \mathcal{L}_\tau(X^{n-1}) + e_{sp}^n,$$

where the first two terms represent the propagation of E_{sp}^{n-1} by the numerical method $\mathcal{L}_\tau$, which is nonlinear. Making use of formula (15), we write

$$\mathcal{L}_\tau(X(t_{n-1})) - \mathcal{L}_\tau(X^{n-1}) = \mathrm{e}^{\tau A} E_{sp}^{n-1} \mathrm{e}^{\tau A^*} + \mathrm{e}^{\tau A} H(X(t_{n-1}), X^{n-1}) \mathrm{e}^{\tau A^*}, \tag{17}$$

where

$$\begin{aligned} H(X(t_{n-1}), X^{n-1}) = {} & \tau \left[G(t_{n-1}, X(t_{n-1})) - G(t_{n-1}, X^{n-1}) \right] \\ & + \int_0^\tau (\tau - s) \left[\ddot{X}_2(t_{n-1} + s) - \ddot{\widetilde{X}}_2(t_{n-1} + s) \right] \mathrm{d}s. \end{aligned}$$

The functions X_2 and $\widetilde{X}_2$ are the solutions of the second subproblem (3) with initial values $X(t_{n-1})$ and X^{n-1}, respectively. Starting from $X(t_0)$ and X^0 and using expression (17) for their propagation by the Lie–Trotter splitting method, we rewrite the global error as

$$\begin{aligned} E_{sp}^n = {} & \mathrm{e}^{n\tau A} E_{sp}^0 \mathrm{e}^{n\tau A^*} + \underbrace{\sum_{k=0}^{n-1} \mathrm{e}^{(n-k)\tau A} H(X(t_k), X^k) \mathrm{e}^{(n-k)\tau A^*}}_{=:D_1} \\ & + \underbrace{\sum_{k=1}^{n} \mathrm{e}^{(n-k)\tau A} e_{sp}^k \mathrm{e}^{(n-k)\tau A^*}}_{=:D_2}. \end{aligned} \tag{18}$$

By the choice of the initial value $X(t_0) = X^0$ we have $\|E_{sp}^0\| = 0$. Since the expression H mainly consists of the nonlinear function G, which by Assumption 1(b) is Lipschitz continuous, and of its derivative $\ddot{X}_2$, which is continuous, we have the bound

$$\|H(X(t_k), X^k)\| \le C(\tau \|E_{sp}^k\| + \tau^2).$$

Hence, property (10) yields the following bound for the second term in the representation of the global error (18):

$$\|D_1\| \le C \sum_{k=0}^{n-1} \mathrm{e}^{(n-k)\tau\omega} \left(\tau \|E_{sp}^k\| + \tau^2\right) \le C\tau \left(\sum_{k=0}^{n-1} \|E_{sp}^k\| + 1 \right). \tag{19}$$

Now, in order to bound D_2, we have to consider the form of the local error in (16). We have

$$\|D_2\| \leq \sum_{k=1}^{n} \left\| \int_0^\tau \int_0^s \mathrm{e}^{(\tau-r)A} \left(\mathrm{e}^{(n-k)\tau A}(AG + GA^*)\mathrm{e}^{(n-k)\tau A^*} \right) \mathrm{e}^{(\tau-r)A^*} \, \mathrm{d}r \, \mathrm{d}s \right\|.$$

The quantity in the parentheses can be bounded by means of assumption (11). Further, employing assumption (10) we obtain

$$\|D_2\| \leq C \sum_{k=1}^{n-1} \frac{1}{(n-k)\tau} \mathrm{e}^{(n-k)\tau\omega} \int_0^\tau \int_0^s \mathrm{e}^{(\tau-r)\omega} \, \mathrm{d}r \, \mathrm{d}s + C\tau$$

and achieve the following bound:

$$\|D_2\| \leq C\tau^2 \sum_{k=1}^{n-1} \frac{1}{k\tau} + C\tau. \tag{20}$$

Now, collecting (18), (19), and (20) yields the error bound

$$\|E_{sp}^k\| \leq C\tau \sum_{k=0}^{n-1} \|E_{sp}^k\| + C\tau \log n + C\tau.$$

The global error bound now follows from a discrete Gronwall inequality; see, e.g., [8]. ∎

5.2. The low-rank Lie–Trotter splitting. In this section we compare the full-rank Lie–Trotter splitting (4) and the low-rank Lie–Trotter splitting. We recall that the solution obtained with the latter is given by $Y^n = (\Phi_\tau^A \circ \widetilde{\Phi}_\tau^G)^n(Y^0)$. The following proposition states the error bound. Its proof is given at the end of this section.

PROPOSITION 2. *Under Assumption* 1, *the difference* $E_{lr}^n = (\Phi_\tau^A \circ \Phi_\tau^G)^n(Y^0) - (\Phi_\tau^A \circ \widetilde{\Phi}_\tau^G)^n(Y^0)$ *is uniformly bounded on* $t_0 \leq t_0 + n\tau \leq T$ *as*

$$\|E_{lr}^n\| \leq c_2\varepsilon + c_3\tau,$$

where the constants c_2 *and* c_3 *depend on* ω, L, B, *and* T, *but are independent of* τ *and* n.

The low-rank Lie–Trotter splitting defined in (7) with initial value Y^0 results, after one time step, in $Y^1 = (\Phi_\tau^A \circ \widetilde{\Phi}_\tau^G)(Y^0)$. It consists of first applying the projector-splitting integrator to the evolution equation (6) for the nonlinearity on the tangent space $\mathcal{T}_{Y_2}\mathcal{M}$ and then solving exactly the first subproblem (5) with initial value $\widetilde{\Phi}_\tau^G(Y^0)$. We start with the following preliminary result.

LEMMA 1. *Under Assumption* 1, *the following bound holds uniformly for each* $n \geq 1$ *satisfying* $t_0 \leq t_0 + n\tau \leq T$,

$$\|(\Phi_\tau^A \circ \Phi_\tau^G)(Y^{n-1}) - (\Phi_\tau^A \circ \widetilde{\Phi}_\tau^G)(Y^{n-1})\| \leq b_1\varepsilon\tau + b_2\tau^2,$$

as long as $\|Y^{n-1} - Y(t_{n-1})\| \leq \gamma$ *for given* $\gamma > 0$; *see* (12). *The constants* b_1 *and* b_2 *depend on* ω, L, B, *and* T, *but are independent of* τ *and* n.

Proof. We observe that

$$\begin{aligned}\|(\Phi_\tau^A \circ \Phi_\tau^G)(Y^{n-1}) - (\Phi_\tau^A \circ \widetilde{\Phi}_\tau^G)(Y^{n-1})\| &= \|(\Phi_\tau^A \circ (\Phi_\tau^G - \widetilde{\Phi}_\tau^G))(Y^{n-1})\| \\ &\le \mathrm{e}^{\tau\omega}\|(\Phi_\tau^G - \widetilde{\Phi}_\tau^G)(Y^{n-1})\|,\end{aligned}$$

where in the last step we employ bound (10) for the matrix exponential operator. For estimating the remaining local error $(\Phi_\tau^G - \widetilde{\Phi}_\tau^G)(Y^{n-1})$ of the projector-splitting integrator, we mainly refer to the error analysis in [18]. Let us consider the nonlinear subproblem (6), which by Assumption 1(c) can be written as

$$\dot{Y}_2(t) = M(t, Y_2(t)) + P(Y_2(t))R(t, Y_2(t)), \quad Y_2(t_{n-1}) = Y^{n-1},$$

for all $n \ge 1$ satisfying $t_0 \le t_0 + n\tau \le T$. Now, dropping the perturbation term yields the auxiliary problem

$$\dot{W}(t) = M(t, W(t)), \quad W(t_{n-1}) = W^{n-1}.$$

Following [18, Lemma 2.2], there exists W^{n-1} such that $\|Y^{n-1} - W^{n-1}\| \le \tau(4BL\tau + 2\varepsilon)$ and the following bound holds:

$$\|\widetilde{\Phi}_\tau^G(Y^{n-1}) - W(t_n)\| \le \tau(9BL\tau + 4\varepsilon).$$

Moreover, by the bound of the perturbation term R and the Lipschitz constant of G, we obtain by a Gronwall inequality

$$\|\Phi_\tau^G(Y^{n-1}) - W(t_n)\| \le \mathrm{e}^{L\tau}\tau(4BL\tau + 3\varepsilon).$$

Collecting those two error estimates results in the local error

$$\|(\Phi_\tau^G - \widetilde{\Phi}_\tau^G)(Y^{n-1})\| \le (4BL\mathrm{e}^{L\tau} + 9BL)\tau^2 + (3\mathrm{e}^{L\tau} + 4)\varepsilon\tau,$$

which proves the stated local error bound for $b_1 = \mathrm{e}^{\tau\omega}(3\mathrm{e}^{L\tau} + 4)$ and $b_2 = \mathrm{e}^{\tau\omega}(4BL\mathrm{e}^{L\tau} + 9BL)$. ∎

With this local error estimate at hand, we are now in a position to prove the bound for E_{lr}^n.

Proof of Proposition 2. Let $\widehat{Y}, \widetilde{Y} \in \mathcal{M}$. Employing bound (10) and the Lipschitz continuity (12) of the nonlinearity G, we obtain

$$\begin{aligned}\|(\Phi_\tau^A \circ \Phi_\tau^G)(\widehat{Y}) - (\Phi_\tau^A \circ \Phi_\tau^G)(\widetilde{Y})\| &= \|\Phi_\tau^A(\Phi_\tau^G(\widehat{Y}) - \Phi_\tau^G(\widetilde{Y}))\| \\ &\le \mathrm{e}^{(L+\omega)\tau}\|\widehat{Y} - \widetilde{Y}\|,\end{aligned}$$

which shows stability of the splitting method $\mathcal{L}_\tau = \Phi_\tau^A \circ \Phi_\tau^G$. Combining the stability with the result of Lemma 1, we obtain the following error recursion:

$$\|E_{lr}^n\| \le b_1\epsilon\tau + b_2\tau^2 + \mathrm{e}^{(L+\omega)\tau}\|E_{lr}^{n-1}\|.$$

The stated bound follows by standard arguments. ∎

5.3. Proof of Theorem 1. Finally, we are in a position to combine the results of the previous sections and prove the main result of this paper.

Proof of Theorem 1. What remains is to give a bound for the propagation E_δ^n of the initial error $\|X^0 - Y^0\|$ by $\mathcal{L}_\tau$. Due to stability of $\mathcal{L}_\tau$, we have the following bound:

$$\|(\Phi_\tau^A \circ \Phi_\tau^G)^n(X^0) - (\Phi_\tau^A \circ \Phi_\tau^G)^n(Y^0)\| \le \mathrm{e}^{(L+\omega)(T-t_0)}\|X^0 - Y^0\|.$$

Combining the three components of the global error results in the stated bound with c_0 containing C and c_3, which come from Propositions 1 and 2, respectively, with c_1, which is the constant of the bound for the propagated initial value, and with c_2, which appears in Proposition 2. ∎

As a remark, we point out that the low-rank Lie–Trotter splitting integrator is not sensitive to the presence of small singular values. The low-rank solution of the linear problem is computed directly by exponential integrators, where possibly appearing small singular values do not cause difficulties. Further, they can also occur in the approximation matrix of the nonlinear subproblem. But since we are applying the projector-splitting integrator, which is robust with respect to small singular values, our integration method inherits this favorable property.

6. Extensions and further convergence results. In this section, we comment on the possible extension of the low-rank Lie–Trotter splitting (7) to a *low-rank Strang splitting*, and we sketch some other situations in which the convergence proof of section 5 also holds.

6.1. The low-rank Strang splitting. The main drawback of the Lie–Trotter splitting scheme in application is its low order. Composing the scheme with its adjoint method, which is again a Lie–Trotter splitting with the order of flows reversed, one obtains the formally second-order Strang splitting. In the low-rank situation, the resulting scheme is given by

$$\Phi_{\tau/2}^A \circ \widetilde{\Phi}_\tau^G \circ \Phi_{\tau/2}^A.$$

This scheme is numerically performing very well in the absence of small nonzero singular values; see [28]. The extension of our convergence proof to this situation, however, is not straightforward. First of all, a second-order scheme needs more regularity of the exact solution, in particular, between the (split) vector fields and the boundary conditions. This was worked out for the full-rank Strang splitting in [10, 17]. The same regularity assumptions and/or modifications are also required here. The numerical example, given in [28, Figure 2] clearly shows that whenever the needed regularity is missing the order is restricted to 1.25 for a formally second-order splitting. The bottleneck, however, is the fact that the projector-splitting Strang scheme is not proven to be second-order convergent in the case of small nonzero singular values; see [18].

6.2. Further convergence results. For the purposes of simplicity and clarity, we have restricted our convergence analysis up to now to parabolic problems and a nonlinearity G that does not necessarily satisfy the boundary conditions of the involved elliptic differential operator. In this case, the quantity $AG + GA^*$ (see (16)) cannot be bounded independently of the spatial grid size. This is the place where the parabolic smoothing property (11) enters the game. A typical instance for such a situation is the following. The matrix A stems from the spatial discretization of an elliptic differential operator subject to homogeneous boundary conditions and $G(t, X(t))$ is a (spatially) smooth function that does not vanish at the boundary.

However, there are interesting situations in which our proof still holds even if (11) does not hold. A typical possibility is the following one. Let the differential operator be of the form $v \cdot \nabla$, where v is a given velocity vector. We thus consider a transport semigroup in a Hilbert space. This is a semigroup of contractions and satisfies (10) with $\omega = 0$.

If this problem is now considered with periodic boundary conditions, the quantity $AG + GA^*$ can be uniformly bounded if $G(t, X(t))$ is smooth in space. In this case, the parabolic smoothing property in not required and low-rank Lie–Trotter splitting is first-order convergent on compact time intervals. As this proof follows from a straightforward modification of the given proof, we do not work out the details.

7. Differential Lyapunov equations. As a special case of the stiff matrix differential equation (1), we consider DLEs, which are of crucial importance in many applications, e.g., Kalman filtering, model reduction of linear time-varying systems, optimal filtering, or numerical simulation of systems governed by stochastic partial differential equations [20].

For DLEs, the term $G(t, X)$ in (1) is solution independent. We denote the resulting time-dependent matrix as $Q(t)$. This gives us the DLE

$$\dot{X}(t) = AX(t) + X(t)A^* + Q(t), \quad X(t_0) = X^0,$$

where $A, Q(t), X(t) \in \mathbb{C}^{m\times m}$. The matrix Q and the initial data X^0 are symmetric and positive semidefinite. Since the DLE is linear, its exact solution exists for all times and is also symmetric and positive semidefinite.

In order to find a low-rank approximation $Y(t) \in \mathcal{M}$ for the solution $X(t)$ of the DLE, we follow the procedure described in section 3. First, we split the DLE into the following two subproblems:

$$\begin{aligned} \dot{X}_1(t) &= AX_1(t) + X_1(t)A^*, & X_1(t_0) &= X_1^0, \\ \dot{X}_2(t) &= Q(t), & X_2(t_0) &= X_2^0. \end{aligned}$$

By $\tilde{\Phi}_\tau^Q$ we denote the flow of the second subproblem approximated by means of the projector-splitting integrator. Then, the low-rank solution is computed by the low-rank Lie–Trotter splitting integrator $\mathcal{I}_\tau = \Phi_\tau^A \circ \tilde{\Phi}_\tau^Q$; see (7) with Q instead of G.

The analysis of the global error of this scheme goes along the proofs in section 5, if the DLE satisfies Assumption 1. Since DLEs are typically stemming from parabolic partial differential equations, we assume that the matrix A satisfies the properties in Assumption 1(a). The inhomogeneity $Q(t)$ is not solution dependent. Thus, we have $L = 0$, and Assumption 1(b) is satisfied. To fulfill Assumption 1(c), we have to assume that $Q(t)$ is in the tangent space $\mathcal{T}_Y\mathcal{M}$ up to a small perturbation $R(t, Y) := Q(t) - P(Y)Q(t)$, i.e., we assume $\|R(t, Y)\| \le \varepsilon$. This assumption is strictly related to the existence of a low-rank structure for the solution of DLEs. Some theoretical results are given in [3, 33, 39].

Thus, we can simply apply the error analysis given in section 5. The bound of the global error of the full-rank Lie–Trotter splitting integrator stays the same, i.e., $\|E_{sp}^n\| \le C\tau(1 + |\log\tau|)$ with the only difference being that here the constant C does not depend on L. Furthermore, the result given in Proposition 2 becomes $\|E_{lr}^n\| \le c_2\varepsilon$. We observe that, compared to the general case, the constant c_3 drops here. This clearly follows from the fact that the constant b_2 of Lemma 1 is zero here. Also the constant c_1 appearing in the error bound for the propagated difference between the full-rank and the low-rank initial values does not depend on L.

Finally, we remark that the low-rank Lie–Trotter splitting (7) can be tailored to preserve symmetry and positive semidefiniteness of the solution, as explained in [28]. This modification does not introduce any further difficulties in the convergence analysis, since it is only based on a different representation of the solution. A symmetric variant of the SVD-like decomposition in subsection 3.3 is employed in the modified algorithm.

8. Differential Riccati equations. The class of matrix differential equations of the form (1) also includes DREs. They play a crucial role in many applications, such as optimal and robust control problems, optimal filtering, H_∞ control of linear time-varying systems, and differential games; see [1, 16, 34]. Further, several integrators based on low-rank approximations have been proposed in the past years. In particular, we mention methods based on backward differentiation formulas and Rosenbrock methods [4, 5].

For DREs the nonlinearity G in (1) is quadratic and of the form

$$G(t, X) = Q(t) - X(t)KX(t).$$

Thus, we consider here the following initial value problem

$$\dot{X}(t) = AX(t) + X(t)A^* + Q(t) - X(t)KX(t), \quad X(t_0) = X^0, \tag{21}$$

where $A, Q(t), K, X(t) \in \mathbb{C}^{m\times m}$. The matrices Q and K, and the initial value X^0 are symmetric and positive semidefinite. The global existence and positive semidefiniteness of the solution is guaranteed under these conditions; see [7].

As for the case of DLEs, the rather general framework given in Assumption 1 fits to DREs. Condition (a) is fulfilled by assuming that the matrix A is the discretization of a strongly elliptic differential operator. Therefore we restrict ourselves to parabolic problems. Property (b) is a usual requirement in the field of differential equations. On the other hand, condition (c) requires more care. Let Y^0 be a rank-r approximation of X^0. Then the rank-r solution of (21) is given by

$$Y(t) = \mathrm{e}^{(t-t_0)A}Y^0\mathrm{e}^{(t-t_0)A^*} + \int_{t_0}^{t} \mathrm{e}^{(t-s)A}P(Y(s))\left(Q(s) - Y(s)KY(s)\right)\mathrm{e}^{(t-s)A^*}\,\mathrm{d}s$$

for $t_0 \le t \le T$. Assumption 1(c) requires that the nonlinearity G has a particular form when computed along a low-rank solution Y. We can take the tangential part as

$$M(t, Y) = P(Y)Q(t) - YKY,$$

whereas the residual is

$$R(t, Y) = Q(t) - P(Y)Q(t).$$

To verify this, note, that the term $M(t, Y)$ is the sum of two elements of the tangent space. Indeed, $P(Y)Q(t)$ is trivially an element of the tangent space. For YKY we proceed as follows. Making use of the explicit form of the projection recalled in (9), we observe that

$$\begin{aligned} P(Y)(YKY) &= UU^*(USV^*KY) - UU^*(USV^*KUSV^*)VV^* + (YKUSV^*)VV^* \\ &= USV^*KY - USV^*KUSV^* + YKUSV^* \\ &= YKY, \end{aligned}$$

where we have used the fact that U and V have orthonormal columns. Since $\mathcal{T}_Y\mathcal{M}$ is a vector space we conclude that $M(t,Y) \in \mathcal{T}_Y\mathcal{M}$. For the residual we assume $\|R(t,Y)\| \leq \varepsilon$; see [3, 39] for some related theoretical results.

Although we carried out the proof in the matrix setting, DREs can be also studied from a different abstract point of view; see, e.g., [21]. A convergence analysis for a splitting method in the setting of Hilbert–Schmidt operators was proposed in [14]. Moreover, different types of splitting for DREs were proposed in [37, 38].

As for DLEs, the low-rank Lie–Trotter splitting can be tailored to preserve symmetry and positive semidefiniteness of the solution. For an algorithm of our proposed method in the case of DREs, see [28].

9. Numerical results. The aim of this section is to illustrate the numerical behavior of the low-rank Lie–Trotter splitting (7). In particular, we present a numerical example to illustrate the convergence result of Theorem 1.

We study a DRE arising in optimal control for linear quadratic regulator problems. Thus we consider the linear control system

$$\dot{x} = Ax + u, \quad x(0) = x_0,$$

where $A \in \mathbb{R}^{m\times m}$ is the system matrix, $x \in \mathbb{R}^m$ the state variable, and $u \in \mathbb{R}^m$ the control. The functional $\mathcal{J}$ that has to be minimized is given by

$$\mathcal{J}(u,x) = \frac{1}{2}\int_0^T \Big(x(t)^\mathsf{T} C^\mathsf{T} C x(t) + u(t)^\mathsf{T} u(t)\Big)\,\mathrm{d}t,$$

where $C \in \mathbb{R}^{q\times m}$ and $(\cdot)^\mathsf{T}$ denotes the transpose. Further, the optimal control is given in feedback form by $u_{\mathrm{opt}}(t) = -X(t)x(t)$, where $X(t)$ is the solution of the following DRE,

$$\dot{X}(t) = A^\mathsf{T} X(t) + X(t)A + C^\mathsf{T} C - X(t)^2,$$

which is in the form of (21) with $Q = C^\mathsf{T} C$ and $K = I_m$ being the identity matrix.

In order to consider an interesting application, we mainly follow the numerical example presented in [14]. The matrix A arises from the spatial discretization of the diffusion operator

$$\mathcal{D} = \partial_x\left(\alpha(x)\partial_x(\cdot)\right) - \lambda I,$$

defined on the spatial domain $\Omega = (0,1)$ subject to homogeneous Dirichlet boundary conditions. We choose $\alpha(x) = 2+\cos(2\pi x)$ and $\lambda = 1$. The finite difference discretization of the operator $\mathcal{D}$ satisfies Assumption 1(a). Let q be odd. The matrix $C \in \mathbb{R}^{q\times m}$ is defined by taking q independent vectors $\{1, e_1, \ldots, e_{(q-1)/2}, f_1, \ldots, f_{(q-1)/2}\}$, where

$$e_k(x) = \sqrt{2}\cos(2\pi k x) \quad \text{and} \quad f_k(x) = \sqrt{2}\sin(2\pi k x), \quad k = 1, \ldots, (q-1)/2,$$

are evaluated at the grid points $\{x_j\}_{j=1}^m$, where $x_j = \frac{j}{m+1}$. The following results are obtained by choosing the initial value $X^0 = 0$, final time $T = 0.1$, $m = 200$, and $q = 9$.

In Figure 3 (left), we show the rank of the reference solution, which is computed by DOPRI5 [13]. We observe that the effective rank of the solution stays low during the evolution in time. In Figure 3 (right), we plot the first 50 singular values of the solution at the final integration time. In Figure 4, the error behavior of the low-rank Lie–Trotter splitting (7) is illustrated. We observe that the error is composed of two different contributions. The choice of a small approximation rank results in stagnation of the error. On the other hand, if the low-rank error becomes small enough, one observes the usual order of convergence, namely order one for the outer Lie–Trotter splitting. This is consistent with the convergence result given in Theorem 1.

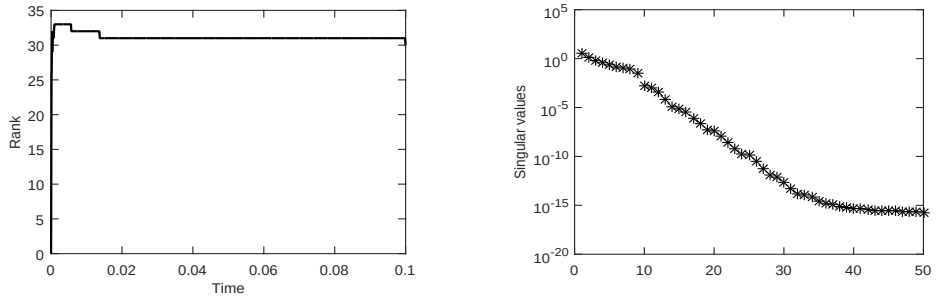

FIG. 3. *Results for the considered DRE for* $m = 200$. *Left: Rank of the reference solution as a function of time. Right: First* 50 *singular values of the reference solution at* $T = 0.1$.

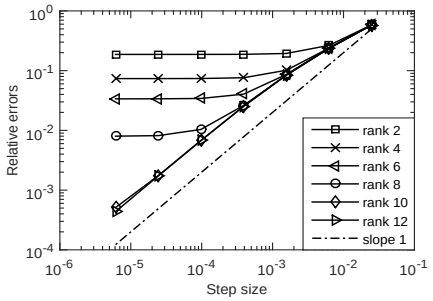

FIG. 4. *Errors of the low-rank Lie–Trotter splitting in the Frobenius norm as a function of step size and rank at* $T = 0.1$ *for the considered DRE for* $m = 200$.

Acknowledgment. We thank the referees for their helpful comments, which improved the presentation of this paper.

REFERENCES

[1] H. Abou-Kandil, G. Freiling, V. Ionescu, and G. Jank, *Matrix Riccati Equations in Control and Systems Theory*, Birkhäuser, Basel, 2003, https://doi.org/10.1007/978-3-0348-8081-7.

[2] A. H. Al-Mohy and N. J. Higham, *Computing the action of the matrix exponential, with an application to exponential integrators*, SIAM J. Sci. Comput., 33 (2011), pp. 488–511, https://doi.org/10.1137/100788860.

[3] A. Antoulas, D. Sorensen, and Y. Zhou, *On the decay rate of Hankel singular values and related issues*, Systems Control Lett., 46 (2002), pp. 323–342, https://doi.org/10.1016/S0167-6911(02)00147-0.

[4] P. Benner and H. Mena, *Rosenbrock methods for solving differential Riccati equations*, IEEE Trans. Automat. Control, 58 (2013), pp. 2950–2956, https://doi.org/10.1109/TAC.2013.2258495.

[5] P. Benner and H. Mena, *Numerical solution of the infinite-dimensional LQR problem and the associated Riccati differential equations*, J. Numer. Math., 26 (2018), pp. 1–20, https://doi.org/10.1515/jnma-2016-1039.

[6] M. Caliari, P. Kandolf, A. Ostermann, and S. Rainer, *The Leja method revisited: Backward error analysis for the matrix exponential*, SIAM J. Sci. Comput., 38 (2016), pp. A1639–A1661, https://doi.org/10.1137/15M1027620.

[7] L. Dieci and T. Eirola, *Positive definiteness in the numerical solution of Riccati differential equations*, Numer. Math., 67 (1994), pp. 303–313, https://doi.org/10.1007/s002110050030.

[8] J. Dixon and S. McKee, *Weakly singular discrete Gronwall inequalities*, ZAMM Z. Angew. Math. Mech., 66 (1986), pp. 535–544, https://doi.org/10.1002/zamm.19860661107.

[9] L. Einkemmer and A. Ostermann, *Exponential integrators on graphic processing units*, in 2013 International Conference on High Performance Computing Simulation (HPCS), IEEE, Piscataway, NJ, 2013, pp. 490–496, https://doi.org/10.1109/HPCSim.2013.6641458.

[10] L. Einkemmer and A. Ostermann, *Overcoming order reduction in diffusion-reaction splitting. Part* 1: *Dirichlet boundary conditions*, SIAM J. Sci. Comput., 37 (2015), pp. A1577–A1592, https://doi.org/10.1137/140994204.

[11] K.-J. Engel and R. Nagel, *A Short Course on Operator Semigroups*, Springer, New York, 2006, https://doi.org/10.1007/0-387-36619-9.

[12] E. Hairer, C. Lubich, and G. Wanner, *Geometric Numerical Integration. Structure-Preserving Algorithms for Ordinary Differential Equations*, 2nd ed., Springer, Berlin, 2000, https://doi.org/10.1007/3-540-30666-8.

[13] E. Hairer, S. P. Nørsett, and G. Wanner, *Solving Ordinary Differential Equations* I: *Nonstiff Problems*, 2nd ed., Springer, New York, 1993, https://doi.org/10.1007/978-3-540-78862-1.

[14] E. Hansen and T. Stillfjord, *Convergence analysis for splitting of the abstract differential Riccati equation*, SIAM J. Numer. Anal., 52 (2014), pp. 3128–3139, https://doi.org/10.1137/130935501.

[15] U. Helmke and J. B. Moore, *Optimization and Dynamical Systems*, Springer, London, 1996, https://doi.org/10.1007/978-1-4471-3467-1.

[16] A. Ichikawa and H. Katayama, *Remarks on the time-varying H_∞ Riccati equations*, Systems Control Lett., 37 (1999), pp. 335–345, https://doi.org/10.1016/S0167-6911(99)00041-9.

[17] T. Jahnke and C. Lubich, *Error bounds for exponential operator splitting*, BIT, 40 (2000), pp. 735–744, https://doi.org/10.1023/A:1022396519656.

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

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

[20] N. Lang, J. Saak, and T. Stykel, *Balanced truncation model reduction for linear time-varying systems*, Math. Comput. Model. Dyn. Syst., 22 (2016), pp. 267–281, https://doi.org/10.1080/13873954.2016.1198386.

[21] I. Lasiecka and R. Triggiani, *Control Theory for Partial Differential Equations: Volume* 1, *Abstract Parabolic Systems: Continuous and Approximation Theories*, Cambridge University Press, Cambridge, 2000, https://doi.org/10.1017/CBO9781107340848.

[22] C. Lubich, *Time integration in the multiconfiguration time-dependent Hartree method of molecular quantum dynamics*, Appl. Math. Res. Express., 2015 (2015), pp. 311–328, https://doi.org/10.1093/amrx/abv006.

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

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

[25] C. Lubich, T. Rohwedder, R. Schneider, and B. Vandereycken, *Dynamical approximation by hierarchical Tucker and tensor-train tensors*, SIAM J. Matrix Anal. Appl., 34 (2013), pp. 470–494, https://doi.org/10.1137/120885723.

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

[27] R. I. McLachlan and G. R. W. Quispel, *Splitting methods*, Acta Numer., 11 (2002), pp. 341–434, https://doi.org/10.1017/S0962492902000053.

[28] H. Mena, A. Ostermann, L.-M. Pfurtscheller, and C. Piazzola, *Numerical low-rank approximation of matrix differential equations*, J. Comput. Appl. Math., 340 (2018), pp. 602–614, https://doi.org/10.1016/j.cam.2018.01.035.

[29] A. Nonnenmacher and C. Lubich, *Dynamical low-rank approximation: Applications and numerical experiments*, Math. Comput. Simul., 79 (2008), pp. 1346–1357, https://doi.org/10.1016/j.matcom.2008.03.007.

[30] I. V. Oseledets, *Tensor-train decomposition*, SIAM J. Sci. Comput., 33 (2011), pp. 2295–2317, https://doi.org/10.1137/090752286.

[31] I. V. Oseledets and E. E. Tyrtyshnikov, *Breaking the curse of dimensionality, or how to use SVD in many dimensions*, SIAM J. Sci. Comput., 31 (2009), pp. 3744–3759, https://doi.org/10.1137/090748330.

[32] A. Pazy, *Semigroups of Linear Operators and Applications to Partial Differential Operators*, Springer, New York, 1983, https://doi.org/10.1007/978-1-4612-5561-1.

[33] T. Penzl, *Eigenvalue decay bounds for solutions of Lyapunov equations: The symmetric case*, Systems Control Lett., 40 (2000), pp. 139–0144, https://doi.org/10.1016/S0167-6911(00)00010-4.

[34] I. R. Petersen, V. A. Ugrinovskii, and A. V. Savkin, *Robust Control Design Using H^∞ Methods*, Springer, London, 2000, https://doi.org/10.1007/978-1-4471-0447-6.

[35] M. Renardy and R. Rogers, *An Introduction to Partial Differential Equations*, 2nd ed., Springer, New York, 2004, https://doi.org/10.1007/b97427.

[36] Y. Saad, *Analysis of some Krylov subspace approximations to the matrix exponential operator*, SIAM J. Numer. Anal., 29 (1992), pp. 209–228, https://doi.org/10.1137/0729014.

[37] T. Stillfjord, *Low-rank second-order splitting of large-scale differential Riccati equations*, IEEE Trans. Automat. Control, 60 (2015), pp. 2791–2796, https://doi.org/10.1109/TAC.2015.2398889.

[38] T. Stillfjord, *Adaptive high-order splitting schemes for large-scale differential Riccati equations*, Numer. Algorithms, 78 (2018), pp. 1129–1151, https://doi.org/10.1007/s11075-017-0416-8.

[39] T. Stillfjord, *Singular value decay of operator-valued differential Lyapunov and Riccati equations*, SIAM J. Control Optim., 56 (2018), pp. 3598–3618, https://doi.org/10.1137/18M1178815.

[40] L. R. Tucker, *Some mathematical notes on three-mode factor analysis*, Psychometrika, 31 (1966), pp. 279–311, https://doi.org/10.1007/BF02289464.