Low-rank Parareal: a low-rank parallel-in-time integrator

Benjamin Carrel, Martin J. Gander, Bart Vandereycken

Abstract In this work, the Parareal algorithm is applied to evolution problems that admit good low-rank approximations and for which the dynamical low-rank approximation (DLRA) can be used as time stepper. Many discrete integrators for DLRA have recently been proposed, based on splitting the projected vector field or by applying projected Runge–Kutta methods. The cost and accuracy of these methods are mostly governed by the rank chosen for the approximation. These properties are used in a new method, called low-rank Parareal, in order to obtain a time-parallel DLRA solver for evolution problems. The algorithm is analyzed on affine linear problems and the results are illustrated numerically.

Keywords dynamical low-rank approximation · initial value problem · matrix differential equation · parallel algorithm

Mathematics Subject Classification (2000) 65L05 · 65L20 · 65L70 · 68W10 · 65F45 · 65F55

1 Introduction

This work is concerned with the parallel-in-time integration of evolution problems for which the solution can be well approximated by a time-dependent low-rank matrix. In particular, we aim to solve approximately the evolution problem

$$
\begin{aligned}
\dot{X}(t) &= F(t, X(t)), \quad t \in [0, T], \\
X(0) &= X_0,
\end{aligned}
\tag{1}
$$

where $X(t)$ is a matrix of size $m \times m$. When the dimension m is large, the numerical solution of (1) can be very expensive since the matrix $X(t)$ is usually dense. One way to alleviate this curse of dimensionality is to use low-rank approximations where, for every t, we approximate $X(t)$ by $Y(t) \in \mathbb{R}^{m \times m}$ such that $\mathrm{rank}(Y(t)) = r \ll m$. The accuracy of this approximation will depend on the choice of the rank r. Here, $X(t)$ is assumed to be square for notational convenience and all results can be easily formulated for rectangular $X(t)$.

A popular paradigm to solve directly for the low-rank approximation $Y(t)$ is the dynamical low-rank algorithm (DLRA), first proposed in [25]. As defined later in Def. 5, DLRA leads to an evolution problem that is a projected version of (1). In the last decade, many discrete integrators for this projected problem have been proposed. One class of integrators consists in a clever splitting of the projector so that the resulting splitting method can be implemented efficiently. An influential example is the projector-splitting scheme proposed in [29]. Other methods that require integrating parts of the vector field can be found in [22,4]. Another approach, proposed in [9,24,35], is based on projecting standard Runge–Kutta methods (sometimes including their intermediate stages). Most of these methods are formulated for constant rank r. Rank adaptivity can be incorporated without much difficulty for splitting and for projected schemes; see [9,6,35,3]. Finally, given the importance of DLRA in problems from physics (like the Schrödinger and Vlasov

Section of Mathematics, University of Geneva.

Benjamin Carrel
E-mail: benjamin.carrel@unige.ch

Martin J. Gander
E-mail: martin.gander@unige.ch

Bart Vandereycken
E-mail: bart.vandereycken@unige.ch

equation), the integrators in [29, 7] also preserve certain invariants, like energy. However, none of these time integrators consider a parallel-in-time scheme for DLRA, which is particularly interesting in the large-scale setting.

Parallel computing can be very effective and is even necessary to solve huge-scale problems. While parallelization in space is well known, also the time direction can be parallelized to some extent when solving evolution problems. Over the last decades, various parallel-in-time algorithms have been proposed; see, e.g., the overviews [12, 32]. Among these, the Parareal algorithm from [28] is one of the more popular algorithms for time parallelization. It is based on a Newton-like iteration, with inaccurate but cheap corrections performed sequentially, and accurate but expensive solves performed in parallel. This idea of solving in parallel an evolution problem as a nonlinear (discretized) system also appears in related methods like PFASST [8], MGRIT [11] and Space-Time Multi-Grid [17]. Theoretical results and numerical studies on a large numbers of cores, show that these parallel-in-time methods can have good parallel performance for parabolic problems; see, e.g., [36, 21, 16]. So far, these methods did not incorporate a low-rank compression of the space dimension, which is the main topic of this work.

2 Preliminaries and contributions

2.1 The Parareal algorithm

The Parareal iteration in Def. 1 below is given for constant time step h (hence $T = Nh$) and for autonomous F. Both restrictions are not crucial but ease the presentation. The quantity X_n^k is an approximation for $X(t_n)$ at time $t_n = nh$ and iteration k. The accuracy of this approximation is expected to improve with increasing k. Here, and throughout the paper, we denote dependency on the iteration index k as ^{k}, which should not be confused with the kth power.

Definition 1 (Parareal) The Parareal algorithm is defined by the following double iteration on k and n,

$$\text{(Initial value)} \qquad X_0^k = X_0, \tag{2}$$

$$\text{(Initial approximation)} \qquad X_{n+1}^0 = \mathscr{G}^h(X_n^0), \tag{3}$$

$$\text{(Iteration)} \qquad X_{n+1}^{k+1} = \mathscr{F}^h(X_n^k) + \mathscr{G}^h(X_n^{k+1}) - \mathscr{G}^h(X_n^k). \tag{4}$$

Here, $\mathscr{F}^h(X)$ represents a fine (accurate) time stepper applied to the initial value X and propagated until time h. Similarly, $\mathscr{G}^h(X)$ represents a coarse (inaccurate) time stepper.

Given two time steppers, the Parareal algorithm is easy to implement. A remarkable property of Parareal is the convergence in a finite number of steps for $k = n$. It is well known that Parareal works well on parabolic problems but behaves worse on hyperbolic problems; see [18] for an analysis.

2.2 Dynamical low-rank approximation

Let $\mathscr{M}_r$ denote the set of $m \times m$ matrices of rank r, which is a smooth embedded submanifold in $\mathbb{R}^{m \times m}$. Instead of solving (1), the DLRA solves the following projected problem:

Definition 2 (Dynamical low-rank approximation) For a rank r, the dynamical low-rank approximation of problem (1) is the solution of

$$\begin{aligned} \dot{Y}(t) &= \mathscr{P}_{Y(t)} F(t, Y(t)), \\ Y(0) &= Y_0 \in \mathscr{M}_r, \end{aligned} \tag{5}$$

where $\mathscr{P}_Y$ is the l_2-orthogonal projection onto the tangent space $\mathscr{T}_Y \mathscr{M}_r$ of $\mathscr{M}_r$ at $Y \in \mathscr{M}_r$. In particular, $Y(t) \in \mathscr{M}_r$ for every $t \in [0, T]$.

To analyze the approximation error of DLRA, we need the following standard assumptions from [23]. Here, and throughout the paper, $\|\cdot\|$ denotes the Frobenius norm.

Assumption 1 (DLRA assumptions) *The function F satisfies the following properties for all $X, Y \in \mathbb{R}^{m \times m}$:*

- *Lipschitz with constant L: $\|F(X) - F(Y)\| \leq L\|X - Y\|$.*
- *One-sided Lipschitz with constant ℓ: $\langle X - Y, F(X) - F(Y) \rangle \leq \ell \|X - Y\|^2$.*
- *Maps almost to the tangent bundle of $\mathscr{M}_r$: $\|F(Y) - \mathscr{P}_Y F(Y)\| \leq \varepsilon_r$.*

2

In the analysis in Section 3, it is necessary to have $\ell < 0$ for convergence. This holds when F is a discretization of certain parabolic PDEs, like the heat equation. In particular, for an affine function of the form $F(X) = A(X) + B$, it holds $\ell = \frac{1}{2}\lambda_{\max}(A + A^T)$; see [20, Ch. I.10]. The quantity ε_r is called the *modeling error* and decreases when the rank r increases. For our problems of interest, this quantity is typically very small. Finally, the existence of L is only needed to guarantee the uniqueness of (1) but it will actually not appear in our analysis. We can therefore allow L to be large, as is the case for discretized parabolic PDEs.

Standard theory for perturbations of ODEs allows us to obtain the following error bound from the assumptions above:

Theorem 1 (Error of DLRA [23]) *Under Assumption 1, the DLRA verifies*

$$\left\| \psi_r^h(Y_0) - \phi^h(X_0) \right\| \le e^{\ell t} \|Y_0 - X_0\| + \varepsilon_r \int_0^t e^{\ell s} ds, \tag{6}$$

where ϕ^h is the flow of the original problem (1) *and ψ_r^h is the flow of its DLRA* (5) *for rank r.*

The solution of DLRA (5) is quasi-optimal with the best rank approximation. This can be seen already in Theorem 1 for short time intervals. Similar estimates exist for parabolic problems [5] and for longer time when there is a sufficiently large gap in the singular values and when their derivatives are bounded [25].

2.3 Contributions

In this paper, we propose a new algorithm, called *low-rank Parareal*. As far as we know, this is the first parallel-in-time integrator for low-rank approximations. We analyze the proposed algorithm when the function F in (1) is affine. To this end, we extend the analysis of the classical Parareal algorithm in [14] to a more general setup where the coarse problem is different from the fine problem. We can prove that the method converges for big steps (large h) on diffusive problems ($\ell < 0$). In numerical experiments, we confirm this behavior. In addition, the method also performs well empirically with a less strict condition on h and on a non-affine problem.

3 Low-rank Parareal

We now present our low-rank Parareal algorithm for solving (1). Since the cost of most discrete integrators for DLRA scales quadratically[1] with the approximation rank, we take the coarse time stepper as DLRA with a small rank q. Likewise, the fine time stepper is DLRA with a large rank r. We can even take $r = m$, which corresponds to computing the exact solution as the fine time stepper since $Y \in \mathbb{R}^{m \times m}$.

Definition 3 (Low-rank Parareal) Consider two ranks $q < r$. The low-rank Parareal algorithm iterates

(Initial value)	$Y_0^k = Y_0,$	(7)
(Initial approximation)	$Y_{n+1}^0 = \psi_q^h \circ \mathscr{T}_q(Y_n^0) + \mathscr{E}_n,$	(8)
(Iteration)	$Y_{n+1}^{k+1} = \psi_r^h \circ \mathscr{T}_r(Y_n^k) + \psi_q^h \circ \mathscr{T}_q(Y_n^{k+1}) - \psi_q^h \circ \mathscr{T}_q(Y_n^k),$	(9)

where $\psi_r^h(Z)$ is the solution of (5) at time h with initial value $Y_0 = Z$, and $\mathscr{T}_r$ is the orthogonal projection onto $\mathscr{M}_r$. The notations ψ_q^h and $\mathscr{T}_q$ are similar but apply to rank q. The matrices $\mathscr{E}_n$ are small perturbations such that $\text{rank}(Y_{n+1}^0) = r + 2q$ and can be chosen randomly.[2]

Observe that the rank of Y_n^k is at most $r + 2q$ for all n, k. The low-rank structure is therefore preserved over the iterations. The matrices $\mathscr{E}_n$ insure that each iteration has a rank between r and $r + 2q$. These matrices impact only the initial error but do not have any role in the convergence of the algorithm as is shown later in the analysis. An efficient implementation should store the low-rank matrices in a factorized manner. In this context, the truncated SVD can be efficiently performed. The DLRA flows ψ_r^h and ψ_q^h can only be computed for relatively small problems. For larger problems, a suitable DLRA integrator must be used; see Section 4 for implementation details.

---

[1] While the actual cost can be larger, at the very least the methods typically compute compact QR factorizations to normalize the approximations. This costs $O(mr^2 + r^3)$ flops in our setting.

[2] One could even take all initial Y_{n+1}^0 random, as is sometimes also done in standard Parareal. The important property is $\text{rank}(Y_{n+1}^0) = r + 2q$ so that the low-rank Parareal iterations (9) are performed on the correct manifold.

3.1 Convergence analysis

Let $X_n = X(t_n)$ be the solution of the full problem (1) at time t_n. Let Y_n^k be the corresponding low-rank Parareal solution at iteration k. We are interested in bounding the error of the algorithm,

$$E_n^k = X_n - Y_n^k, \tag{10}$$

for all relevant n and k. To this end, we make the following assumption:

Assumption 2 (Affine vector field) *The function F is affine linear and autonomous, that is,*

$$F(X) = A(X) + B$$

with $A\colon \mathbb{R}^{m \times m} \to \mathbb{R}^{m \times m}$ a linear operator and $B \in \mathbb{R}^{m \times m}$.

The following lemma gives us a recursion for the Frobenius norm of the error. This recursion will be fundamental in deriving our convergence bounds later on when we generalize the proof for standard Parareal from [14].

Lemma 1 (Iteration of the error) *Under the Assumptions 1 and 2, the error of low-rank Parareal verifies*

$$\left\| E_{n+1}^{k+1} \right\| \leq e^{\ell h} C_{r,q} \left\| E_n^k \right\| + e^{\ell h} C_q \left\| E_n^{k+1} \right\| + e^{\ell h} \max_{n \geq 0} \| X_n - \mathscr{T}_r(X_n) \| + (2\varepsilon_q + \varepsilon_r) \int_0^h e^{\ell(h-s)} ds. \tag{11}$$

The constants $\ell, \varepsilon_q, \varepsilon_r$ are defined in Assumption 1. Moreover, $C_{r,q}$ and C_q are the Lipschitz constants of $\mathscr{T}_r - \mathscr{T}_q$ and $\mathscr{T}_q$.

Proof Our proof is similar to the one in [24] where first the continuous version of the approximation error of DLRA is studied. Denote by $\phi^h(Z)$ the solution of (1) at time h with initial value $Y_0 = Z$. By definition, the discrete error is

$$E_{n+1}^{k+1} = \phi^h(X_n) - \psi_r^h \circ \mathscr{T}_r(Y_n^k) - \psi_q^h \circ \mathscr{T}_q(Y_n^{k+1}) + \psi_q^h \circ \mathscr{T}_q(Y_n^k).$$

We can interpret each term above as a flow from t_n to $t_{n+1} = t_n + h$. Denote these flows by $X(t), Z(t), W(t)$, and $V(t)$ with the initial values

$$X(t_n) = X_n, \ Z(t_n) = \mathscr{T}_r(Y_n^k), \ W(t_n) = \mathscr{T}_q(Y_n^{k+1}), \ V(t_n) = \mathscr{T}_q(Y_n^k). \tag{12}$$

Defining the continuous error as

$$E(t) = X(t) - Z(t) - W(t) + V(t),$$

we then get the identity $E_{n+1}^{k+1} = E(t_n + h)$.

We proceed by bounding $\|E(t)\|$. By definition of the flows above, we have (omitting the dependence on t in the notation)

$$
\begin{aligned}
\dot{E} &= F(X) - \mathscr{P}_Z F(Z) - \mathscr{P}_W F(W) + \mathscr{P}_V F(V) \\
&= F(X - Z - W + V) + F(Z) - \mathscr{P}_Z F(Z) - \mathscr{P}_W F(W) + F(W) - F(V) + \mathscr{P}_V F(V),
\end{aligned}
$$

where the last equality holds since the function F is affine. Using Assumption 1 and Cauchy–Schwarz, we compute

$$
\begin{aligned}
\frac{1}{2}\frac{d}{dt}\|E(t)\|^2 &= \left\langle E, \dot{E} \right\rangle \\
&= \langle E, F(E) \rangle + \langle E, F(Z) - \mathscr{P}_Z F(Z) \rangle + \langle E, F(W) - \mathscr{P}_W F(W) \rangle - \langle E, F(V) - \mathscr{P}_V F(V) \rangle \\
&\leq \ell \|E\|^2 + \varepsilon_r \|E\| + 2\varepsilon_q \|E\|.
\end{aligned}
$$

Since $\frac{d}{dt}\|E(t)\|^2 = 2\|E(t)\|\frac{d}{dt}\|E(t)\|$, we therefore obtain the differential inequality

$$\frac{d}{dt}\|E(t)\| \leq \ell \|E\| + \varepsilon_r + 2\varepsilon_q.$$

Grönwall's lemma allows us to conclude

$$\|E(t_n + h)\| \leq \|E(t_n)\| e^{\ell h} + (2\varepsilon_q + \varepsilon_r) \int_{t_n}^{t_n + h} e^{\ell(h-s)} ds. \tag{13}$$

From (12), we get

$$E(t_n) = X_n - \mathscr{T}_r(Y_n^k) - \mathscr{T}_q(Y_n^{k+1}) + \mathscr{T}_q(Y_n^k).$$

4

Denoting $\mathscr{T}_r^{\perp} = I - \mathscr{T}_r$ and $\mathscr{T}_{r,q} = \mathscr{T}_r - \mathscr{T}_q$, we get after rearranging terms

$$E(t_n) = \mathscr{T}_r^{\perp}(X_n) + \mathscr{T}_r(X_n) - \mathscr{T}_q(X_n) - \mathscr{T}_r(Y_n^k) + \mathscr{T}_q(Y_n^k) + \mathscr{T}_q(X_n) - \mathscr{T}_q(Y_n^{k+1})$$
$$= \mathscr{T}_r^{\perp}(X_n) + \mathscr{T}_{r,q}(X_n) - \mathscr{T}_{r,q}(Y_n^k) + \mathscr{T}_q(X_n) - \mathscr{T}_q(Y_n^{k+1}).$$

Taking norms gives

$$\|E(t_n)\| \leq \left\| \mathscr{T}_r^{\perp}(X_n) \right\| + \left\| \mathscr{T}_{r,q}(X_n) - \mathscr{T}_{r,q}(Y_n^k) \right\| - \left\| \mathscr{T}_q(X_n) - \mathscr{T}_q(Y_n^{k+1}) \right\|$$
$$\leq \max_{n \geq 0} \left\| \mathscr{T}_r^{\perp}(X_n) \right\| + C_{r,q} \left\| E_n^k \right\| + C_q \left\| E_n^{k+1} \right\|, \tag{14}$$

where $C_{r,q}$ and C_q are the Lipschitz constants of $\mathscr{T}_{r,q}$ and $\mathscr{T}_q$ respectively. Combining inequalities (13) and (14) gives the statement of the lemma.

We now study the error recursion (11) in more detail. To this end, let us slightly rewrite it as

$$\left\| E_{n+1}^{k+1} \right\| \leq \alpha \left\| E_n^k \right\| + \beta \left\| E_n^{k+1} \right\| + \kappa, \quad \left\| E_n^0 \right\| \leq \gamma, \tag{15}$$

with the non-negative constants

$$\alpha = e^{\ell h} C_{r,q}, \quad \beta = e^{\ell h} C_q, \quad \gamma = \max_{n \geq 0} \left\| E_n^0 \right\|,$$
$$\kappa = e^{\ell h} \max_{n \geq 0} \|X_n - \mathscr{T}_r(X_n)\| + (2\varepsilon_q + \varepsilon_r) \int_0^h e^{\ell(h-s)} ds. \tag{16}$$

Our first result is a linear convergence bound, up to the DLRA approximation error. It is similar to the linear bound for standard Parareal.

Theorem 2 (Linear convergence) *Under the Assumptions 1 and 2, and if $\alpha + \beta < 1$, low-rank Parareal verifies for all $k \in \mathbb{N}$ the linear bound*

$$\max_{n \geq 0} \left\| E_n^k \right\| \leq \left(\frac{\alpha}{1-\beta} \right)^k \max_{n \geq 0} \left\| E_n^0 \right\| + \frac{\kappa}{1-\alpha-\beta}, \tag{17}$$

where α, β, κ are defined in (16).

Proof Define $e_\star^k = \max_{n \geq 0} \left\| E_n^k \right\|$. Taking the maximum for $n \geq 0$ of both sides of (15), we obtain

$$e_\star^{k+1} \leq \alpha e_\star^k + \beta e_\star^{k+1} + \kappa.$$

By assumption, $0 \leq \beta < 1$ and we can therefore obtain the recursion

$$e_\star^{k+1} \leq \frac{\alpha}{1-\beta} e_\star^k + \frac{\kappa}{1-\beta},$$

with solution

$$e_\star^k \leq \left(\frac{\alpha}{1-\beta} \right)^k e_\star^0 + \frac{\kappa}{1-\alpha-\beta} \left[1 - \left(\frac{\alpha}{1-\beta} \right)^k \right].$$

By assumption, we also have $0 \leq \frac{\alpha}{1-\beta} < 1$, which allows us to obtain the statement of the theorem.

Next, we present a more refined superlinear bound. To this end, we require the following technical lemma that solves the equality version of the double iteration (15). A similar result, but without the κ term and only as an upper bound, already appeared in [14, Thm. 1]. Our proof is therefore similar but more elaborate.

Lemma 2 *Let $\alpha, \beta, \gamma, \kappa \in \mathbb{R}$ be any non-negative constants such that $\alpha < 1$ and $\beta < 1$. Let e_n^k be a sequence depending on $n, k \in \mathbb{N}$ such that*

$$e_{n+1}^{k+1} = \alpha e_n^k + \beta e_n^{k+1} + \kappa, \qquad e_{n+1}^0 = \gamma, \qquad e_0^k = 0. \tag{18}$$

Then,

$$e_n^k = \kappa \sum_{j=0}^{k-1} \sum_{i=0}^{n-j-1} \binom{i+j}{i} \alpha^j \beta^i + \begin{cases} 0 & \text{if } n \leq k, \\ \gamma \alpha^k \sum_{i=0}^{n-k-1} \binom{i+k-1}{i} \beta^i & \text{if } n \geq k+1. \end{cases} \tag{19}$$

Proof The idea is to use the generating function $\rho_k(\xi) = \sum_{n=1}^{\infty} e_n^k \xi^n$ for $k \geq 1$. Multiplying (18) by ξ^{n+1} and summing over n, we obtain

$$\sum_{n=0}^{\infty} e_{n+1}^{k+1} \xi^{n+1} = \sum_{n=0}^{\infty} \alpha e_n^k \xi^{n+1} + \sum_{n=0}^{\infty} \beta e_n^{k+1} \xi^{n+1} + \sum_{n=0}^{\infty} \kappa \xi^{n+1}, \qquad \sum_{n=0}^{\infty} e_{n+1}^0 \xi^{n+1} = \sum_{n=0}^{\infty} \gamma \xi^{n+1}.$$

Since $e_0^k = 0$ for all k, this gives the relations

$$\rho_{k+1}(\xi) = \alpha \xi \rho_k(\xi) + \beta \xi \rho_{k+1}(\xi) + \kappa \frac{\xi}{1-\xi}, \qquad \rho_0(\xi) = \gamma \frac{\xi}{1-\xi}.$$

We can therefore obtain the linear recurrence

$$\rho_{k+1}(\xi) = a \rho_k(\xi) + b, \quad \text{where } a = \frac{\alpha \xi}{1-\beta \xi}, \, b = \frac{\kappa \xi}{(1-\xi)(1-\beta \xi)}.$$

Its solution satisfies

$$\rho_k(\xi) = \frac{\alpha^k \xi^k}{(1-\beta \xi)^k} \frac{\gamma \xi}{1-\xi} + \sum_{j=0}^{k-1} \frac{\alpha^j \xi^j}{(1-\beta \xi)^{j+1}} \frac{\kappa \xi}{1-\xi}.$$

It remains to compute the coefficients in the power series of the above formula since by definition of $\rho_k(\xi) = \sum_{n=1}^{\infty} e_n^k \xi^n$ they equal the unknowns e_n^k. The binomial series formula for $|z| < 1$,

$$\frac{1}{(1-z)^{k+1}} = \sum_{i=0}^{\infty} \binom{i+k}{i} z^i, \tag{20}$$

together with the Cauchy product gives

$$\frac{1}{(1-\beta \xi)^k} \frac{1}{1-\xi} = \sum_{i=0}^{\infty} \binom{i+k-1}{i} \beta^i \xi^i \cdot \sum_{i=0}^{\infty} \xi^i = \sum_{n=0}^{\infty} \left(\sum_{\ell=0}^{n} \binom{\ell+k-1}{\ell} \beta^\ell \right) \xi^n$$

$$\frac{1}{(1-\beta \xi)^{j+1}} \frac{1}{1-\xi} = \sum_{i=0}^{\infty} \binom{i+j}{i} \beta^i \xi^i \cdot \sum_{i=0}^{\infty} \xi^i = \sum_{n=0}^{\infty} \left(\sum_{\ell=0}^{n} \binom{\ell+j}{\ell} \beta^\ell \right) \xi^n.$$

Hence, the first term in $\rho_k(\xi)$ satisfies

$$\frac{\alpha^k \xi^k}{(1-\beta \xi)^k} \frac{\gamma \xi}{1-\xi} = \gamma \alpha^k \sum_{n=0}^{\infty} \left(\sum_{\ell=0}^{n} \binom{\ell+k-1}{\ell} \beta^\ell \right) \xi^{n+k+1} = \gamma \alpha^k \sum_{n=k+1}^{\infty} \left(\sum_{\ell=0}^{n-k-1} \binom{\ell+k-1}{\ell} \beta^\ell \right) \xi^n,$$

while the second term can be written as

$$\sum_{j=0}^{k-1} \frac{\alpha^j \xi^j}{(1-\beta \xi)^{j+1}} \frac{\kappa \xi}{1-\xi} = \kappa \sum_{j=0}^{k-1} \sum_{n=0}^{\infty} \sum_{\ell=0}^{n} \binom{\ell+j}{\ell} \alpha^j \beta^\ell \xi^{n+j+1} = \kappa \sum_{n=1}^{\infty} \sum_{j=0}^{k-1} \left(\sum_{\ell=0}^{n-1} \binom{\ell+j}{\ell} \alpha^j \beta^\ell \right) \xi^{n+j}.$$

Putting everything together, we have

$$\sum_{n=0}^{\infty} e_n^k \xi^n = \gamma \alpha^k \sum_{n=k+1}^{\infty} \left(\sum_{\ell=0}^{n-k-1} \binom{\ell+k-1}{\ell} \beta^\ell \right) \xi^n + \kappa \sum_{m=1}^{\infty} \sum_{j=0}^{k-1} \left(\sum_{\ell=0}^{m-1} \binom{\ell+j}{\ell} \alpha^j \beta^\ell \right) \xi^{m+j}.$$

Finally, we can identify the coefficient e_n^k in front of ξ^n with those on the right-hand side. The coefficient for ξ^n in the first term is clearly nonzero only when $n \geq k+1$. In the second term, there is only one m for every j such that $m+j = n$. Substituting $m = n - j$ allows us to identify the coefficient of ξ^n.

Using the previous lemma, we can obtain a convergence bound that is superlinear in k.

Theorem 3 (Superlinear convergence) *Under the Assumptions 1 and 2, and if $\alpha + \beta < 1$, the error of low-rank Parareal satisfies for all $n, k \in \mathbb{N}$ the bound*

$$\left\| E_n^k \right\| \leq \frac{\alpha^k}{(k-1)!} \frac{\prod_{j=2}^{k}(n-j)}{1-\beta} \max_{n \geq 0} \left\| E_n^0 \right\| + \frac{\kappa}{1-\alpha-\beta}, \tag{21}$$

where α, β, κ are defined in (16).

Proof Define $e_n^k = \left\| E_n^k \right\|$. By Lemma 1, the terms e_n^k verify the relation described in Lemma 2 with $=$ replaced by $\leq$ in (18). Hence, the solution (19) from Lemma 2 will be an upper bound for e_n^k.

Since $0 \leq \alpha + \beta < 1$ and using the binomial series formula (20), we bound the first term in (19) as

$$
\kappa \sum_{j=0}^{k-1} \sum_{i=0}^{n-j-1} \binom{i+j}{i} \alpha^j \beta^i \leq \kappa \sum_{j=0}^{k-1} \sum_{i=0}^{\infty} \binom{i+j}{i} \alpha^j \beta^i = \kappa \sum_{j=0}^{k-1} \alpha^j \frac{1}{(1-\beta)^{j+1}}
$$

$$
\leq \frac{\kappa}{1-\beta} \sum_{j=0}^{\infty} \left(\frac{\alpha}{1-\beta} \right)^j = \frac{\kappa}{1-\beta} \frac{1}{1-\frac{\alpha}{1-\beta}} = \frac{\kappa}{1-\alpha-\beta}.
$$

For $0 \leq i \leq n-k-1$ and $n \geq k+1$, observe that

$$
\frac{(i+k-1)!}{i!} = \prod_{j=1}^{k} (i+j) \leq \prod_{j=1}^{k} (n-k-1+j) = \prod_{j=2}^{k} (n-j).
$$

Since $0 \leq \beta < 1$, we can therefore bound the second term as

$$
\gamma \, \alpha^k \sum_{i=0}^{n-k-1} \binom{i+k-1}{i} \beta^i = \gamma \, \alpha^k \sum_{i=0}^{n-k-1} \frac{(i+k-1)!}{i!(k-1)!} \beta^i \leq \gamma \, \frac{\alpha^k}{(k-1)!} \prod_{j=2}^{k} (n-j) \sum_{i=0}^{n-k-1} \beta^i
$$

$$
\leq \gamma \, \frac{\alpha^k}{(k-1)!} \frac{\prod_{j=2}^{k}(n-j)}{1-\beta}.
$$

The conclusion now follows by the definition of γ.

The proof above can be modified to obtain a simple linear bound that is similar but different to the one from Theorem 2:

Theorem 4 (Another linear convergence bound) *Under Assumptions 1 and 2, and if $\alpha + \beta < 1$, the error of low-rank Parareal satisfies for all $n, k \in \mathbb{N}$ the bound*

$$
\left\| E_n^k \right\| \leq \alpha^k (1+\beta)^{n-1} \max_{n \geq 0} \left\| E_n^0 \right\| + \frac{\kappa}{1-\alpha-\beta}, \tag{22}
$$

where α, β, κ are defined in (16).

Proof We repeat the proof for the superlinear bound but this time, the second term is bounded as

$$
\gamma \, \alpha^k \sum_{i=0}^{n-k-1} \binom{i+k-1}{i} \beta^i \leq \gamma \, \alpha^k \sum_{i=0}^{n-1} \binom{n-1}{i} \beta^i = \gamma \, \alpha^k \, (1+\beta)^{n-1}. \square
$$

Remark 1 In the proof above, yet another bound based on (20) is

$$
\gamma \, \alpha^k \sum_{i=0}^{n-k-1} \binom{i+k-1}{i} \beta^i \leq \gamma \, \alpha^k \sum_{i=0}^{\infty} \binom{i+k-1}{i} \beta^i = \gamma \, \alpha^k \, \frac{1}{(1-\beta)^k}.
$$

This time we recover the linear bound from Theorem 2.

3.2 Summary of the convergence bounds

In the previous section, we have proven four upper bounds for the error of low-rank Parareal. The first is directly obtained from Lemma 2. It is the tightest bound but its expression is too unwieldy for practical use. The other three bounds can be summarized as

$$
\left\| E_n^k \right\| \leq B_{n,k} \max_{n \geq 0} \left\| E_n^0 \right\| + \frac{\kappa}{1-\alpha-\beta}, \tag{23}
$$

with

$B_{n,k}$	rate of (23) in k
$\alpha^k(1-\beta)^{-k}$	linear
$\alpha^k(1+\beta)^{n-1}$	linear
$\alpha^k(1-\beta)^{-1}\frac{\Pi_{j=2}^k(n-j)}{(k-1)!}$	superlinear

Each of these practical bounds describes different phases of the convergence, and none is always better than the others. In Figure 1, we have plotted all four bounds for realistic values of α and β. We took $\kappa = 10^{-15} \approx \varepsilon_{\text{mach}}$ since it only determines the stagnation of the error and would interfere with judging the transient behavior of the convergence plot. Furthermore, the errors $e_n^0 = \gamma = 1$ at the start of the iteration $k = 0$ were chosen arbitrarily since they have little influence on the results.

The bounds above depend on $\alpha = e^{\ell h}C_{r,q}$ and $\beta = e^{\ell h}C_q$, where C_q and $C_{r,q}$ are the Lipschitz constants of $\mathscr{T}_q$ and $\mathscr{T}_{r,q}$ respectively; see (16). While it seems difficult to give a priori results on the size C_q and $C_{r,q}$, we can bound them up to first order in the theorem below. Note also that in the important case of $\ell < 0$, the constants α and β can be made as small as desired by taking h sufficiently large.

Theorem 5 (Lipschitz constants) *Let $A, \tilde{A} \in \mathbb{R}^{m \times n}$. Then*

$$\|\mathscr{T}_q(A) - \mathscr{T}_q(\tilde{A})\| \leq \frac{\sigma_q}{\sigma_q - \sigma_{q+1}}\|A - \tilde{A}\| + O(\|A - \tilde{A}\|^2), \tag{24}$$

where σ_q is the qth singular value of A. Moreover,

$$\|\mathscr{T}_{r,q}(A) - \mathscr{T}_{r,q}(\tilde{A})\| \leq \left(\frac{\sigma_q}{\sigma_q - \sigma_{q+1}} + \frac{\sigma_r}{\sigma_r - \sigma_{r+1}}\right)\|A - \tilde{A}\| + O(\|A - \tilde{A}\|^2). \tag{25}$$

Proof For the first inequality, we refer to [2, Theorem 2] and [10, Theorem 24]. The second inequality follows from the first by the triangle inequality,

$$\|\mathscr{T}_{r,q}(A) - \mathscr{T}_{r,q}(\tilde{A})\|_F \leq \|\mathscr{T}_q(A) - \mathscr{T}_q(\tilde{A})\| + \|\mathscr{T}_r(A) - \mathscr{T}_r(\tilde{A})\|. \square$$

In many applications with low-rank matrices, the singular values of the underlying matrix are rapidly decaying. In particular, when the singular values decay exponentially like $\sigma_k \approx e^{-ck}$ for some $c > 0$, we have

$$\frac{\sigma_q}{\sigma_q - \sigma_{q+1}} = \frac{1}{1 - \sigma_{q+1}/\sigma_q} \approx \frac{1}{1 - e^{-c}}. \tag{26}$$

This last quantity decreases quickly to 1 when c grows. Even for $c = 1$, it is less than 1.6. We therefore see that the constants in Theorem 5 are not too large in this case.

Remark 2 In the analysis, a sufficiently large gap in the singular values is required at both the coarse rank and the fine rank. In our experiments, we observed that such a gap is indeed required at the coarse rank, but not at the fine rank. It suggests that the bound (25) can therefore probably be improved.

4 Numerical experiments

We now show numerical experiments for our low-rank Parareal algorithm. We implemented the algorithm in Python 3.10 and all computations were performed on a MacBook Pro with a M1 processor and 16GB of RAM. The complete code is available at GitHub so that all the experiments can be reproduced. The DLRA steps are solved by the second-order projector-splitting integrator from [29]. Since the problems considered are stiff, we used sufficiently many substeps of this integrator so that the coarse and fine solvers within low-rank Parareal can be considered exact.
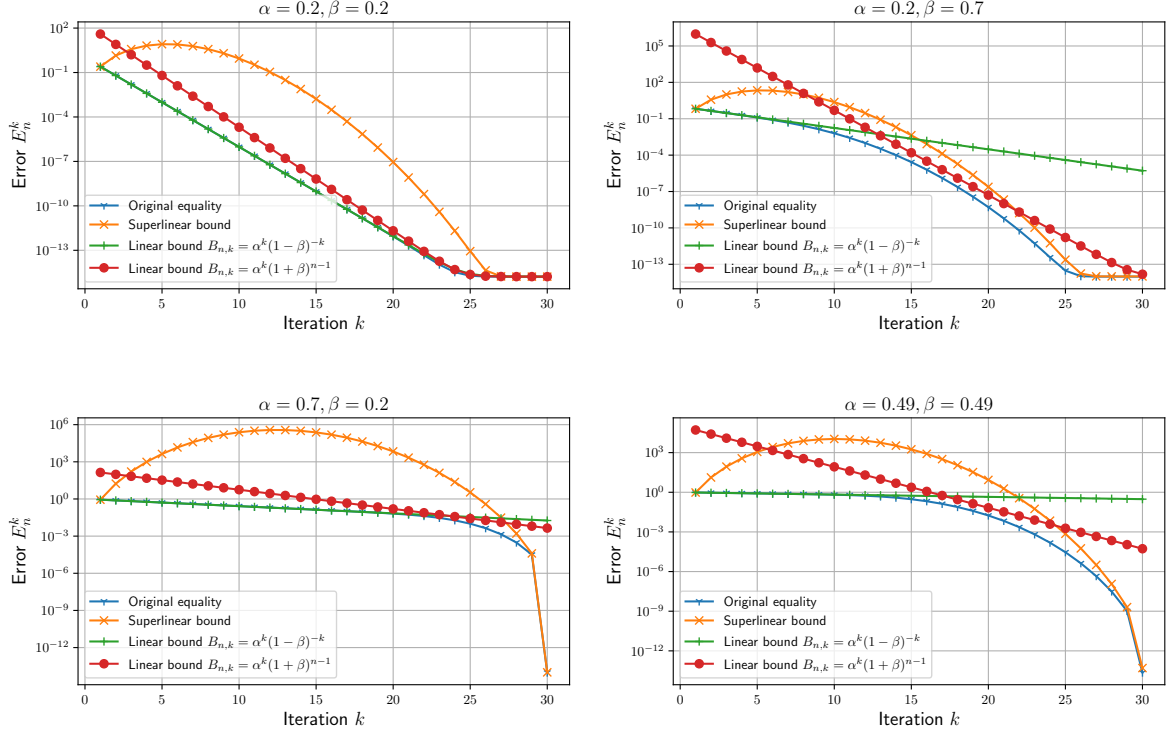
Fig. 1: Bounds derived for several values of α and β. In all panels $n = 30$, $\gamma = 1$, and $\kappa = 10^{-15}$.

4.1 Lyapunov equation

Consider the differential Lyapunov equation,

$$\dot{X}(t) = AX(t) + X(t)A + CC^T, \quad X(0) = X_0, \tag{27}$$

where $A \in \mathbb{R}^{m \times m}$ is a symmetric matrix, and $C \in \mathbb{R}^{m \times k}$ is a tall matrix for some $k \le m$. This initial value problem admits a unique solution for $t \in [0,T]$ for any $T > 0$. The most typical example of (27) is the heat equation on a square with separable source term. Other applications can be found in [31].

Assumption 3 *The matrix $A \in \mathbb{R}^{m \times m}$ is symmetric and strictly negative definite.*

Under Assumption 3, the one-sided Lipschitz constant ℓ for (27) is strictly negative. Indeed, the linear Lyapunov operator $\mathscr{A}(X) = AX + XA$ has the symmetric matrix representation $A \otimes I + I \otimes A$ with eigenvalues $\lambda_i(A) + \lambda_j(A)$ for $1 \le i, j \le m$; see [19, Ch. 12.3]. As in [20, Ch. I.10], we therefore get immediately that $\ell = 2 \max_i \lambda_i(A) < 0$. Moreover, since $\mathscr{A}$ is invertible, we can write the closed-form solution of (27) as

$$X(t) = e^{t\mathscr{A}}(X_0) + \mathscr{A}^{-1}(e^{t\mathscr{A}}(CC^T) - CC^T), \tag{28}$$

which can be easily verified by differentiation using properties of the matrix exponential $e^{t\mathscr{A}}(Z) = e^{tA}Ze^{tA}$.

The following result shows that the solution of (27) can be well approximated by low rank. It is the analogue to a similar result for the algebraic Lyapunov equation $\mathscr{A}(X) = CC^T$. The latter result is well known, but we did not find a proof for the former in the literature.

Lemma 3 (Low-rank approximability of Lyapunov ODE) *Let $\sigma_i(X_0)$ be the ith singular value of X_0 and likewise for $\sigma_i(CC^T)$. Under Assumption 3, the solution $X(t)$ of (27) has an approximation*

$$Y(t) \text{ of rank at most } r_0 + 2r\rho$$

for any $0 \le r_0, r, \rho \le m$ with error

$$\|X(t) - Y(t)\|_2 \le e^{\ell t} \sigma_{r_0+1}(X_0) + \left(\frac{e^{t\ell} - 1}{\ell} \right) \left(4 \exp\left(\frac{-\pi^2 \rho}{\log(4\kappa_A)} \right) \|CC^T\|_2 + \sigma_{r+1}(CC^T) \right),$$

where $\kappa_A = \|A\|_2 \|A^{-1}\|_2$ and $\ell = 2 \max_i \lambda_i(A)$.

9

Proof The aim is to approximate the following two terms that make up the closed-form solution $X(t)$ in (28):

$$X_1(t) = e^{t\mathscr{A}}(X_0), \quad X_2(t) = \mathscr{A}^{-1}(e^{t\mathscr{A}}(CC^T) - CC^T).$$

The first term $X_1(t)$ can be treated directly. By the truncated SVD, the initial value satisfies

$$X_0 = Y_0 + E_0 \text{ where } \text{rank}(Y_0) = r_0 \text{ and } \|E_0\|_2 = \sigma_{r_0+1}(X_0).$$

By Assumption 3, the operator $\mathscr{A}$ is full rank. We therefore obtain

$$X_1(t) = e^{t\mathscr{A}}(X_0) = e^{At}Y_0e^{At} + e^{At}E_0e^{At} = Y_1(t) + E_1(t), \tag{29}$$

where $\text{rank}(Y_1(t)) = \text{rank}(Y_0) = r_0$ and $\|E_1(t)\|_2 \le e^{\ell t}\sigma_{r_0+1}(X_0)$ since $\ell = 2\max_i \lambda_i(A)$.

Next, we focus on the second term $X_2(t)$. Like above, the source term can be decomposed as

$$CC^T = D + F \text{ where } \text{rank}(D) = r \text{ and } \|F\|_2 = \sigma_{r+1}(CC^T).$$

By linearity of the Lyapunov operator, we therefore obtain

$$X_2(t) = \mathscr{A}^{-1}(e^{t\mathscr{A}}D - D) + \mathscr{A}^{-1}(e^{t\mathscr{A}}F - F). \tag{30}$$

Denote $M = e^{t\mathscr{A}}D - D$. By definition of the Lyapunov operator $\mathscr{A}$, we have

$$S = \mathscr{A}^{-1}(M) \iff AS + SA = M.$$

As studied in [34] and then improved in [1], the singular values of the solution S are bounded as

$$\frac{\sigma_{\text{rank}(M)\rho+1}(S)}{\sigma_1(S)} \le 4\exp\left(\frac{-\pi^2\rho}{\log(4\kappa_A)}\right), \tag{31}$$

where $\kappa_A = \|A\|_2\|A^{-1}\|_2$ and $0 \le \rho \le m$. Since $\text{rank}(M) \le 2\text{rank}(D) = 2r$ by assumption on D, the bound (31) then implies that

$$S = Y_2(t) + \delta S(t),$$

where $\text{rank}(Y_2(t)) \le 2r\rho$ and

$$\|\delta S(t)\|_2 \le 4\exp\left(\frac{-\pi^2\rho}{\log(4\kappa_A)}\right)\left\|\mathscr{A}^{-1}(e^{t\mathscr{A}}D - D)\right\|_2 \le 4\exp\left(\frac{-\pi^2\rho}{\log(4\kappa_A)}\right)\frac{e^{t\ell}-1}{\ell}\|D\|_2,$$

where the last inequality holds by properties of the logarithmic norm μ of $\mathscr{A}$ which equals ℓ; see [37, Proposition 2.2]. Moreover, we can bound the last term in (30) as

$$\|E_2(t)\|_2 = \|\mathscr{A}^{-1}(e^{t\mathscr{A}}F - F)\|_2 \le \frac{e^{t\ell}-1}{\ell}\|F\|_2.$$

Putting (29) and (30) together, we obtained

$$X(t) = Y(t) + E(t), \quad Y(t) = Y_1(t) + Y_2(t), \quad E(t) = E_1(t) + \delta S(t) + E_2(t),$$

which proves the statement of the lemma.

The lemma shows that if X_0 and CC^T have good low-rank approximations, then the solution $X(t)$ of the differential Lyapunov equation has comparable low-rank approximations as well on $[0, T]$. Since $\ell < 0$, we can even take $T \to \infty$ and recover essentially the low-rank approximability of the Lyapunov equation $X(\infty) = \mathscr{A}^{-1}(CC^T)$. This is clearly visible when X_0 and CC^T are exactly of low rank, which we state as a simple corollary for convenience.

Corollary 1 *Under Assumption 3 and assuming that* $\text{rank}(X_0) = r_0$, $\text{rank}(CC^T) = r$, *the solution* $X(t)$ *of (28) has an approximation*

$$Y(t) \text{ of rank at most } r_0 + 2r\rho$$

for any $0 \le \rho \le m$ *with error*

$$\|X(t) - Y(t)\|_2 \le 4\frac{e^{t\ell}-1}{\ell}\exp\left(\frac{-\pi^2\rho}{\log(4\kappa_A)}\right)\|CC^T\|_2.$$
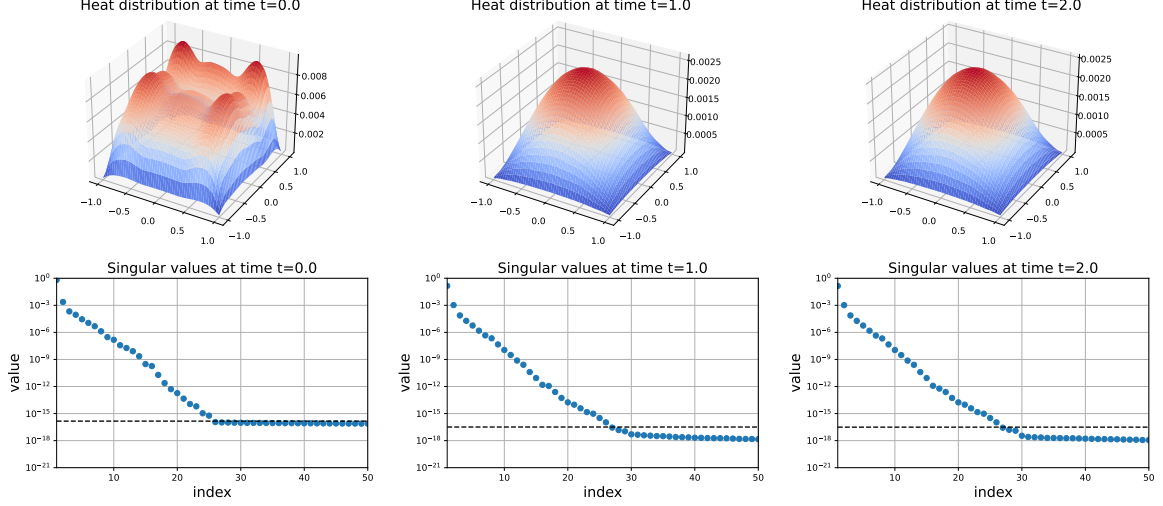
Fig. 2: Solution over time of the Lyapunov ODE (27) for the heat equation. Note the change of scale between $t = 0.0$ and $t = 1.0$.

The corollary clearly shows that the approximation error decreases exponentially when the approximation rank increases linearly via ρ. Furthermore, we see that the condition number of the matrix A has only a mild influence due to $\log(\kappa_A)$.

Remark 3 Corollary 1 can be compared to a similar result in [26]. In that work, the authors solve (27) with exact low-rank $X_0 = ZZ^T$ and CC^T using a Krylov subspace method. More specifically, with U_k an orthonormal matrix that spans the block Krylov space $K_k(A, [C\ Z])$, the projected Lyapunov equation

$$\dot{Y}_k = (U_k^T A U_k) Y_k + Y_k (U_k^T A U_k) + U_k^T C C^T U_k, \quad Y_k(0) = U_k^T X_0 U_k,$$

is used to define the approximation $X_k(t) = U_k Y_k(t) U_k^T$. The approximation error of $X_k(t)$ is studied in [26, Theorem 4.2]. Since $\text{rank}(X_k(t)) \le k(\text{rank}(Z) + \text{rank}(C))$, we therefore also get a result on the low-rank approximability of (27). This bound is, however, worse than ours since it does not give zero error for $t = 0$ and $k = 1$, for example. On the other hand, it is a bound for a discrete method whereas our Lemma 3 and Corollary 1 are statements about the exact solution.

We now apply the low-rank Parareal algorithm to the differential Lyapunov equation (27). Let $A = \Delta_{dx}$ be the $n \times n$ discrete Laplacian with zero Dirichlet boundary conditions obtained by standard centered differences on $[-1, 1]$. The Lyapunov equation is therefore a model for the 2D heat equation on $\Omega = [-1, 1]^2$. In the experiments, we used $n = 100$ spatial points and the time interval $[0, T] = [0, 2]$. The matrix C for the source is generated randomly with singular values $\sigma_i = 10^{-5(i-1)}$ where $i = 1, 2, \ldots$ so that its numerical rank is 4. In order to have a realistic initial value, X_0 is obtained as the exact solution at time $t = 0.01$ of the same ODE but with a random initial value $\tilde{X}_0$ with singular values $\sigma_i = 10^{-(i-1)}$.

Figure 2 is a 3D plot of the solution over time on Ω with its corresponding singular values. As we can see, the solution becomes almost stationary at $t = 1.0$. In addition, it stays low-rank over time in agreement to Lemma 3. Moreover, the singular values suggest to take the fine rank $r = 16$ for an error of the fine solver of order 10^{-12}.

The convergence of the error of the low-rank Parareal algorithm is shown in Figure 3. The algorithm converges linearly from the coarse rank solution to the fine rank solution. Figure 3a suggests that the coarse rank does not influence the convergence rate and it only reduces the initial error. This is consistent with our analysis. Indeed, since the singular values are exponentially decaying, the singular gap is approximately constant; see (26). Hence, the constants α and β from (16) that determine the convergence rate do not depend on the coarse rank q; as is shown up to first order in Theorem 5. Figure 3b shows that, similarly, the convergence rate does not depend on the fine rank either, although it limits the final error.

In Figure 4a, we investigate the convergence for several sizes n. Even though the problem is stiff, the convergence does not seem influenced by the size of the problem. Figure 4b shows the error of the algorithm applied to the problem with several step sizes. According to our analysis, the convergence is faster when the stepsize h is large; see (16).
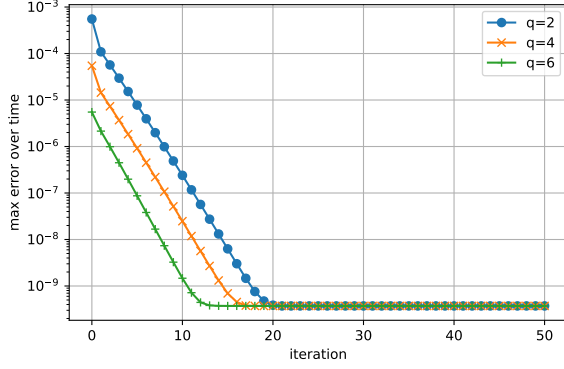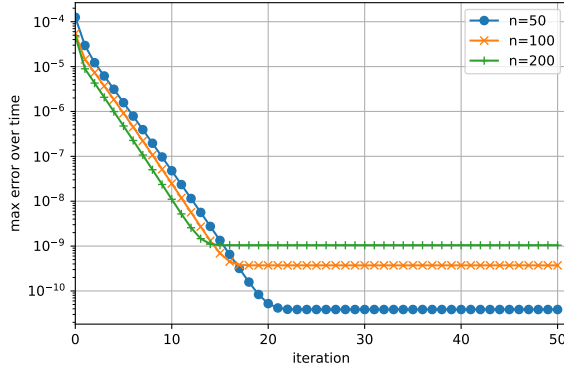
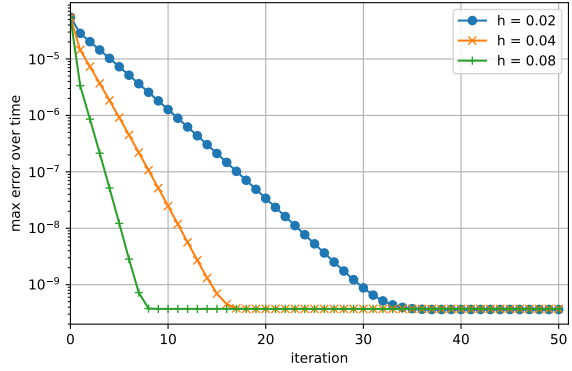(a) Several coarse ranks q with fine rank $r = 16$.
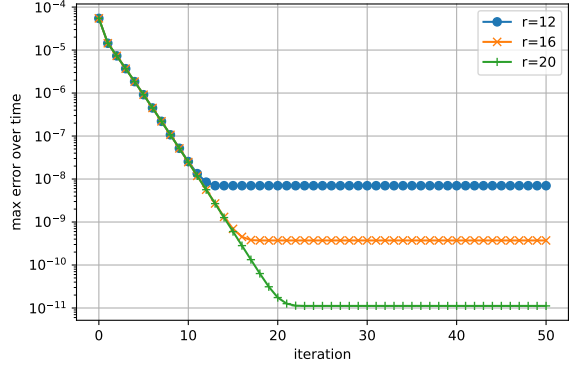
(b) Several fine ranks r with coarse rank $q = 4$.

Fig. 3: Convergence of the error of low-rank Parareal for the Lyapunov ODE (27) with $n = 100$ and $T = 2.0$. Influence of the coarse and fine ranks.

(a) Several sizes n with final time $T = 2.0$.

(b) Several stepsizes h with size $n = 100$.

Fig. 4: Convergence of the error of low-rank Parareal for the Lyapunov ODE (27) with coarse rank $q = 4$ and fine rank $r = 16$. Influence of size and final time.

4.2 Parametric cookie problem

We now solve a simplified version of the parametric cookie problem from [27]. Consider the ODE

$$\dot{Y} = -A_0 Y - A_1 Y C_1 + \mathbf{b}\mathbf{1}^T, \quad Y(0) = Y_0, \tag{32}$$

where the sparse matrices $A_0, A_1 \in \mathbb{R}^{1580 \times 1580}$, $\mathbf{b} \in \mathbb{R}^{1580}$, and $C_1 = \mathrm{diag}(c_1^1, c_1^2, \ldots, c_1^p)$ are given in [27]. The aim of this problem is to solve a heat problem simultaneously with several heat coefficients, denoted by $c_1^1, \ldots, c_1^p$.

In our experiments, we used $p = 101$ parameters with $c_1^1 = 0, c_1^2 = 1, \ldots, c_1^{101} = 100$. The initial value X_0 is obtained after computing the exact solution of (32) at time $t = 0.01$ with the zero matrix as initial value. The time interval is $[0, T] = [0, 0.1]$.

The singular values of the reference solution are shown in Figure 5. The stationary solution has good low-rank approximations, as was proved in [27, Thm. 2.4]. The singular value decay suggests that a fine rank $r = 16$ leads to full numerical accuracy.

In Figure 6, we applied the low-rank Parareal algorithm with several coarse ranks q and fine ranks r. Like for the Lyapunov equation, it seems that the convergence rate does not depend on the coarse rank q. In agreement to our analysis (see Figure 1), the convergence is linear in the first iterations and superlinear in the last iterations. In addition, the convergence is not influenced by the fine rank r.
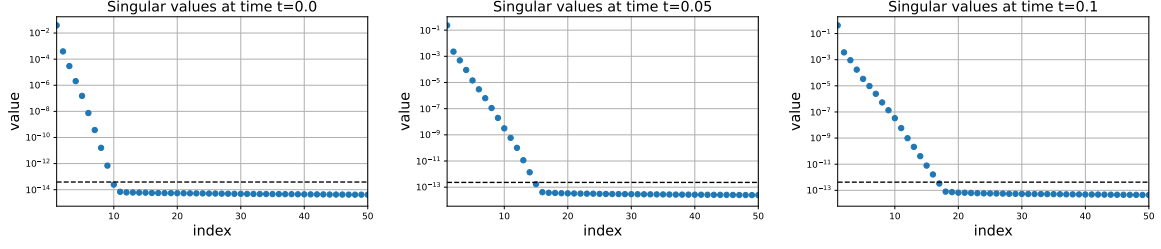
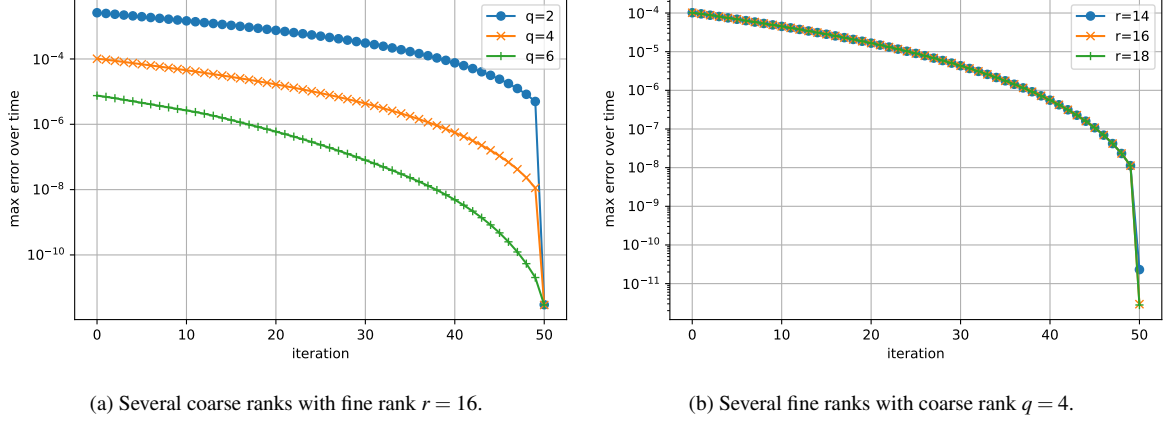Fig. 5: Singular values of the solution over time of the parametric cookie problem (32).

(a) Several coarse ranks with fine rank $r = 16$.

(b) Several fine ranks with coarse rank $q = 4$.

Fig. 6: Convergence of the error of low-rank Parareal for the parametric cookie problem (32). Influence of the coarse and fine ranks.

4.3 Riccati equation

The Riccati differential equation is given by

$$\dot{X}(t) = A^T X(t) + X(t)A + C^T C - X(t)SX(t), \quad X(0) = X_0, \tag{33}$$

where $X \in \mathbb{R}^{m \times m}$, $A \in \mathbb{R}^{m \times m}$, $C \in \mathbb{R}^{k \times m}$, and $S \in \mathbb{R}^{m \times m}$. We note that this is no longer an ODE with an affine vector field and hence our theoretical results do not apply here. As already studied in [33], we take $S = I$ and A is the spatial discretization of the diffusion operator

$$\mathscr{D} = \partial_x(\alpha(x)\partial_x(\cdot)) - \lambda I$$

on the spatial domain $\Omega = [0,1]$. Furthermore, we take $\alpha(x) = 2 + 2\cos(2\pi x)$ and $\lambda = 1$. The discretization is done by the finite volume method, as described in [15]. The tall matrix $C \in \mathbb{R}^{k \times m}$ is obtained from k independent vectors $\{1, e_1, \ldots, e_{(k-1)/2}, f_1, \ldots, f_{(k-1)/2}\}$, where

$$e_i(x) = \sqrt{2}\cos(2\pi kx) \quad \text{and} \quad f_i(x) = \sqrt{2}\sin(2\pi kx), \quad i = 1, \ldots, (q-1)/2, \tag{34}$$

are evaluated at the grid points $\{x_j\}_{j=1}^m$ with $x_j = \frac{j}{m+1}$. The time interval is $[0,T] = [0,0.1]$.

As for the other problems, the singular values of the solution (shown in Figure 7) indicate that we can expect good low-rank approximations on $[0,T]$. We choose the fine rank $r = 18$. The convergence of low-rank Parareal is shown in Figure 8. Unlike the previous problems, the coarse rank q has a more pronounced influence on the behavior of the convergence. While our theoretical results do not hold for this nonlinear problem, we still see that low-rank Parareal converges linearly when the coarse rank q is sufficiently large ($q = 6$, $q = 8$). The convergence is slower (but still superlinear) when $q = 4$. This could be due to the non-constant gaps in the singular values. The influence of the fine rank r is more like for the linear problems.

4.4 Rank-adaptive algorithm

Since the approximation rank of the solution is usually not known a priori, it is more convenient for the user to supply an approximation tolerance than an approximation rank. Even though the rank can change to satisfy the tolerance during
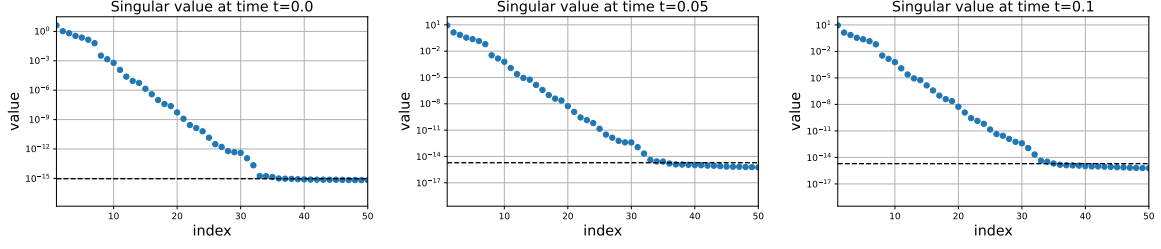
13

Fig. 7: Singular values of the solution over time of the Riccati ODE (33).

(a) Several coarse ranks with fine rank $r = 18$.

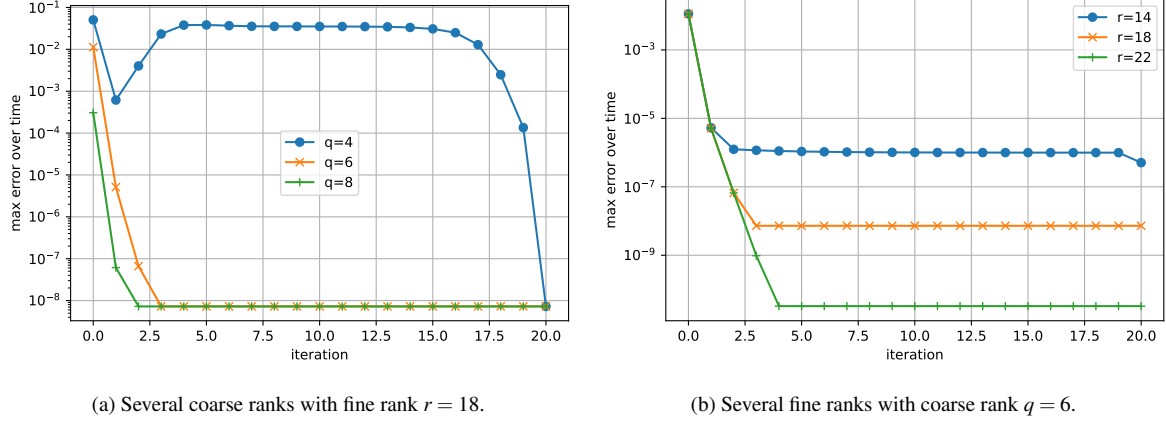(b) Several fine ranks with coarse rank $q = 6$.

Fig. 8: Convergence of the error of low-rank Parareal for the Riccati problem (33). Influence of the coarse and fine ranks.

the truncation steps, Algorithm 3 can be easily reformulated for such a rank adaptive setting. The key idea is to fix the coarse rank to keep the cost of the coarse solver low, while the fine rank is determined by a fine tolerance.

Definition 4 (Adaptive low-rank Parareal) Consider a small fixed rank q and a fine tolerance τ. The adaptive low-rank Parareal algorithm iterates

(Initial value)	$Y_0^k = Y_0,$	(35)
(Initial approximation)	$Y_{n+1}^0 = \psi_q^h \circ \mathscr{T}_q(Y_n^0) + \mathscr{E}_n,$	(36)
(Iteration)	$Y_{n+1}^{k+1} = \psi_{\mathrm{rank}(\mathscr{T}_\tau(Y_n^k))}^h \circ \mathscr{T}_\tau(Y_n^k) + \psi_q^h \circ \mathscr{T}_q(Y_n^{k+1}) - \psi_q^h \circ \mathscr{T}_q(Y_n^k),$	(37)

where the notation is similar to that of the previous Def. 3, except for $\mathscr{T}_\tau$ which represents the rank-adaptive truncation. In particular, $\mathscr{T}_\tau(Y)$ is the best rank q approximation of Y so that the $(q+1)$th singular value of Y equals the tolerance τ. The matrices $\mathscr{E}_n$ are small perturbations, randomly generated such that $\mathrm{rank}(Y_{n+1}^0) = \mathrm{rank}(Y_0)$ and its smallest singular value is larger than the fine tolerance τ.

Figure 9 shows the numerical behavior of this rank-adaptive algorithm. As we can see, the algorithm behaves as desired. Figure 9a shows the algorithm applied with several tolerances and is comparable to Figure 3b with several fine ranks. Figure 9b shows the rank of the solution over time. Already after two iterations, the rank is reduced to almost the numerical rank of the exact solution and the rank does not change much for the rest of the iterations.

5 Conclusion and future work

We proposed the first parallel-in-time algorithm for integrating a dynamical low-rank approximation (DLRA) of a matrix evolution equation. The algorithm follows the traditional Parareal scheme but it uses DLRA with a low rank as coarse integrator, whereas the fine integrator is DLRA with a higher rank. Taking into account the modeling error of DLRA, we presented an analysis of the algorithm and showed linear convergence as well as superlinear convergence under common assumptions and for affine linear vector fields, up to the modeling error.

In our numerical experiments, the algorithm behaved well on diffusive problems, which is similar to the original Parareal algorithm. Due to the significant difference in computational cost for the fine and coarse integrators, it is

(a) Several tolerances with coarse rank $q = 4$.

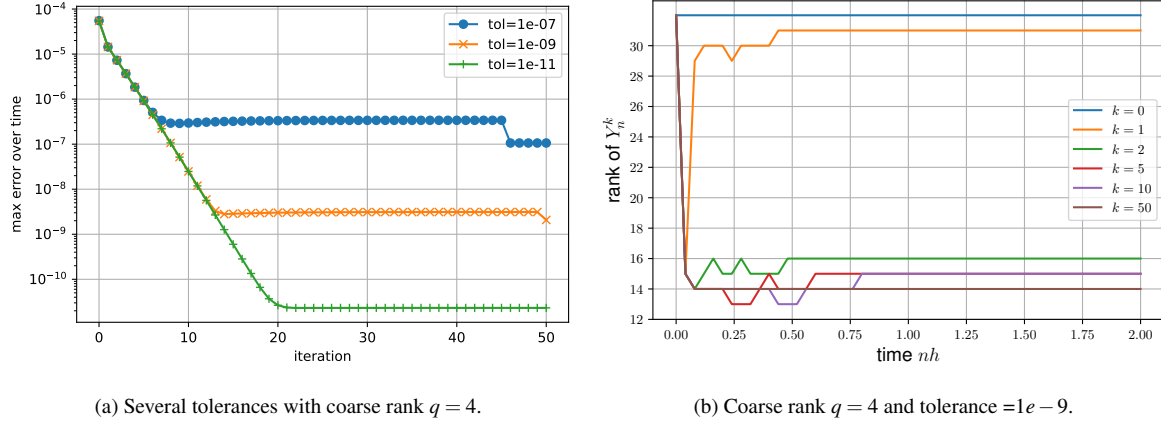(b) Coarse rank $q = 4$ and tolerance $= 1e - 9$.

Fig. 9: Adaptive low-rank Parareal. On the left, the algorithm is applied with several tolerances. On the right, the rank of the solution over time is shown for several iterations.

reasonable to expect good speed-up in actual parallel implementations. A proper parallel implementation to verify this claim is a natural future work. It may however be more appropriate to first generalize more efficient parallel-in-time algorithms, like Schwarz waveform relaxation and multigrid methods [13], to DLRA.

Since DLRA can also be used to obtain low-rank tensor approximations [30], another future work is to extend low-rank Parareal to tensor DLRA. Finally, our theoretical analysis assumes that the ODE has an affine vector field. Since this assumption was only needed in one step of the proof of Lemma 1, it might be possible that it can be relaxed to include certain non-linear vector fields.

Declarations

Conflicts of interest The authors have no conflicts of interest to declare.

References

1. Beckermann, B., Townsend, A.: On the Singular Values of Matrices with Displacement Structure. SIAM J. Matrix Anal. & Appl. **38**(4), 1227–1248 (2017). DOI 10.1137/16M1096426. URL https://epubs.siam.org/doi/10.1137/16M1096426
2. Breiding, P., Vannieuwenhoven, N.: Sensitivity of low-rank matrix recovery. arXiv:2103.00531 [cs, math] (2021). URL http://arxiv.org/abs/2103.00531. ArXiv: 2103.00531
3. Ceruti, G., Kusch, J., Lubich, C.: A rank-adaptive robust integrator for dynamical low-rank approximation. BIT Numerical Mathematics (2022). DOI 10.1007/s10543-021-00907-7
4. Ceruti, G., Lubich, C.: An unconventional robust integrator for dynamical low-rank approximation. BIT Numerical Mathematics **62**(1), 23–44 (2022). DOI 10.1007/s10543-021-00873-0
5. Conte, D.: Dynamical low-rank approximation to the solution of parabolic differential equations. Applied Numerical Mathematics **156**, 377–384 (2020). DOI 10.1016/j.apnum.2020.05.011. URL https://linkinghub.elsevier.com/retrieve/pii/S0168927420301550
6. Dektor, A., Rodgers, A., Venturi, D.: Rank-adaptive tensor methods for high-dimensional nonlinear PDEs. arXiv:2012.05962 [physics] (2021). URL http://arxiv.org/abs/2012.05962. ArXiv: 2012.05962
7. Einkemmer, L., Lubich, C.: A Quasi-Conservative Dynamical Low-Rank Algorithm for the Vlasov Equation. SIAM Journal on Scientific Computing **41**(5), B1061–B1081 (2019). DOI 10.1137/18M1218686
8. Emmett, M., Minion, M.: Toward an efficient parallel in time method for partial differential equations. Communications in Applied Mathematics and Computational Science **7**(1), 105–132 (2012)
9. Feppon, F., Lermusiaux, P.F.J.: Dynamically Orthogonal Numerical Schemes for Efficient Stochastic Advection and Lagrangian Transport. SIAM Rev. **60**(3), 595–625 (2018). DOI 10.1137/16M1109394. URL https://epubs.siam.org/doi/10.1137/16M1109394
10. Feppon, F., Lermusiaux, P.F.J.: A Geometric Approach to Dynamical Model Order Reduction. SIAM J. Matrix Anal. & Appl. **39**(1), 510–538 (2018). DOI 10.1137/16M1095202. URL https://epubs.siam.org/doi/10.1137/16M1095202
11. Friedhoff, S., Falgout, R.D., Kolev, T.V., MacLachlan, S., Schroder, J.B.: A multigrid-in-time algorithm for solving evolution equations in parallel. Tech. rep., Lawrence Livermore National Lab.(LLNL), Livermore, CA (United States) (2012)
12. Gander, M.J.: 50 years of time parallel time integration. In: Multiple Shooting and Time Domain Decomposition Methods, pp. 69–113. Springer (2015)

13. Gander, M.J.: Time Parallel Time Integration. In: Time Parallel Time Integration, p. 90. University of Geneva (2018)
14. Gander, M.J., Hairer, E.: Nonlinear Convergence Analysis for the Parareal Algorithm. In: T.J. Barth, M. Griebel, D.E. Keyes, R.M. Nieminen, D. Roose, T. Schlick, U. Langer, M. Discacciati, D.E. Keyes, O.B. Widlund, W. Zulehner (eds.) Domain Decomposition Methods in Science and Engineering XVII, vol. 60, pp. 45–56. Springer Berlin Heidelberg, Berlin, Heidelberg (2008). DOI 10.1007/978-3-540-75199-1_4. URL http://link.springer.com/10.1007/978-3-540-75199-1_4. Series Title: Lecture Notes in Computational Science and Engineering
15. Gander, M.J., Kwok, F.: Numerical Analysis of Partial Differential Equations Using Maple and MATLAB. Society for Industrial and Applied Mathematics, Philadelphia, PA (2018). DOI 10.1137/1.9781611975314. URL https://epubs.siam.org/doi/book/10.1137/1.9781611975314
16. Gander, M.J., Lunet, T., Ruprecht, D., Speck, R.: A unified analysis framework for iterative parallel-in-time algorithms. arXiv preprint arXiv:2203.16069 (2022)
17. Gander, M.J., Neumuller, M.: Analysis of a new space-time parallel multigrid algorithm for parabolic problems. SIAM Journal on Scientific Computing **38**(4), A2173–A2208 (2016)
18. Gander, M.J., Vandewalle, S.: Analysis of the Parareal Time-Parallel Time-Integration Method. SIAM J. Sci. Comput. **29**(2), 556–578 (2007). DOI 10.1137/05064607X. URL http://epubs.siam.org/doi/10.1137/05064607X
19. Golub, G.H., Van Loan, C.F.: Matrix computations, fourth edition edn. Johns Hopkins studies in the mathematical sciences. The Johns Hopkins University Press, Baltimore (2013). OCLC: ocn824733531
20. Hairer, E., Nørsett, S.P., Wanner, G.: Solving Ordinary Differential Equations I, *Springer Series in Computational Mathematics*, vol. 8. Springer Berlin Heidelberg, Berlin, Heidelberg (1987). DOI 10.1007/978-3-662-12607-3. URL http://link.springer.com/10.1007/978-3-662-12607-3
21. Hofer, C., Langer, U., Neumüller, M., Schneckenleitner, R.: Parallel and robust preconditioning for space-time isogeometric analysis of parabolic evolution problems. SIAM Journal on Scientific Computing **41**(3), A1793–A1821 (2019)
22. Khoromskij, B.N., Oseledets, I.V., Schneider, R.: Efficient time-stepping scheme for dynamics on TT-manifolds. Preprint (2012). URL https://www.mis.mpg.de/preprints/2012/preprint2012_24.pdf
23. 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)
24. Kieri, E., Vandereycken, B.: Projection Methods for Dynamical Low-Rank Approximation of High-Dimensional Problems. Computational Methods in Applied Mathematics **19**(1), 73–92 (2019). DOI 10.1515/cmam-2018-0029. URL https://www.degruyter.com/view/journals/cmam/19/1/article-p73.xml
25. Koch, O., Lubich, C.: Dynamical Low-Rank Approximation. SIAM J. Matrix Anal. & Appl. **29**(2), 434–454 (2007). DOI 10.1137/050639703. URL http://epubs.siam.org/doi/10.1137/050639703
26. Koskela, A., Mena, H.: Analysis of Krylov subspace approximation to large-scale differential Riccati equations. etna **52**, 431–454 (2020). DOI 10.1553/etna_vol52s431. URL https://hw.oeaw.ac.at?arp=0x003bd61b
27. Kressner, D., Tobler, C.: Low-Rank Tensor Krylov Subspace Methods for Parametrized Linear Systems. SIAM J. Matrix Anal. & Appl. **32**(4), 1288–1316 (2011). DOI 10.1137/100799010. URL http://epubs.siam.org/doi/10.1137/100799010
28. Lions, J.L., Maday, Y., Turinici, G.: Résolution d'EDP par un schéma en temps pararéel. Comptes Rendus de l'Académie des Sciences - Series I - Mathematics **332**(7), 661–668 (2001). DOI 10.1016/S0764-4442(00)01793-6. URL https://linkinghub.elsevier.com/retrieve/pii/S0764444200017936
29. Lubich, C., Oseledets, I.V.: A projector-splitting integrator for dynamical low-rank approximation. BIT Numerical Mathematics **54**(1), 171–188 (2014). DOI 10.1007/s10543-013-0454-0. URL http://link.springer.com/10.1007/s10543-013-0454-0
30. Lubich, C., Vandereycken, B., Walach, H.: Time integration of rank-constrained Tucker tensors. SIAM Journal on Numerical Analysis **56**(3), 1273–1290 (2018). DOI 10.1137/17M1146889
31. Mena, H., Ostermann, A., Pfurtscheller, L.M., Piazzola, C.: Numerical low-rank approximation of matrix differential equations. Journal of Computational and Applied Mathematics **340**, 602–614 (2018). DOI 10.1016/j.cam.2018.01.035
32. Ong, B.W., Schroder, J.B.: Applications of time parallelization. Computing and Visualization in Science **23**(1), 1–15 (2020)
33. 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). DOI 10.1137/18M1177901. URL https://epubs.siam.org/doi/10.1137/18M1177901
34. Penzl, T.: Eigenvalue decay bounds for solutions of Lyapunov equations: the symmetric case. Systems & Control Letters **40**(2), 139–144 (2000). DOI 10.1016/S0167-6911(00)00010-4. URL https://linkinghub.elsevier.com/retrieve/pii/S0167691100000104
35. Rodgers, A., Dektor, A., Venturi, D.: Adaptive integration of nonlinear evolution equations on tensor manifolds. arXiv:2008.00155 [physics] (2021). URL http://arxiv.org/abs/2008.00155. ArXiv: 2008.00155
36. Speck, R., Ruprecht, D., Krause, R., Emmett, M., Minion, M., Winkel, M., Gibbon, P.: A massively space-time parallel N-body solver. In: SC'12: Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis, pp. 1–11. IEEE (2012)
37. Söderlind, G.: The logarithmic norm. History and modern theory. BIT Numerical Mathematics **46**(3), 631–652 (2006). DOI 10.1007/s10543-006-0069-9