IMA Journal of Numerical Analysis (2015) **35**, 1315–1341
doi:10.1093/imanum/dru031
Advance Access publication on July 31, 2014

Convergence analysis of some second-order parareal algorithms

Shu-Lin Wu
School of Science, Sichuan University of Science and Engineering, Zigong, Sichuan 643000, P. R. China
wushulin_ylp@163.com wushulin.hust@gmail.com

[Received on 23 July 2013; revised on 21 February 2014]

In the past 10 years, the 'parareal' (parallel-in-time) algorithm has attracted lots of attention thanks to its excellent performance in scientific computing. The parareal algorithm is iterative and is characterized by two propagators $\mathcal{G}$ and $\mathcal{F}$ which are associated with a coarse step size ΔT and a fine step size Δt, respectively, where $\Delta T = J\Delta t$ and $J \geqslant 2$ is an integer. When we apply this algorithm to large-scale systems of ordinary differential equations obtained by semidiscretizing partial differential equations, two questions arise naturally. (I) Is the error between the iterate and the target solution contractive at each iteration for any choice of the discretization parameters ΔT, J and Δx? (II) How small can the contraction factor be and can such a contraction factor be independent of the discretization parameters? For linear problems $\mathbf{u}' = A\mathbf{u} + g$ with symmetric negative-definite matrix A, when the implicit Euler method is used as both the $\mathcal{G}$- and $\mathcal{F}$-propagators, positive answers to these two questions were given by Mathew *et al.* (2010, *SIAM J. Sci. Comput.*, **32**, 1180–1200) and the contraction factor can be bounded by 0.298 for any choice of the discretization parameters. In this paper, for the case that the implicit Euler method is used as the $\mathcal{G}$-propagator, we provide a positive answer to (I) for three second-order $\mathcal{F}$-propagators: the trapezoidal method, the TR/BDF2 method and the two-stage diagonally implicit Runge–Kutta (2s-DIRK) method. For (II), we prove that the contraction factors can be bounded by 0.316 and $\frac{1}{3}$ for the 2s-DIRK method and the TR/BDF2 method (provided the parameter γ involved in TR/BDF2 satisfies $\gamma \in [0.043, 0.977]$), respectively, and both bounds are independent of the discretization parameters. For the trapezoidal method, we show that a uniform bound (less than 1) of the contraction factor does not exist. Numerical results are presented to validate the theoretical prediction.

Keywords: parareal algorithm; implicit Euler method; TR/BDF2 method; trapezoidal method; 2s-DIRK method.

1. Introduction

Large-scale linear systems of ordinary differential equations (ODEs)

$$\mathbf{u}' = A\mathbf{u} + \mathbf{g}, \quad A \in \mathbb{R}^{m\times m}, \tag{1.1}$$

arise frequently in the numerical approximation of complex partial differential equations (PDEs), such as the two-sided spatial fractional diffusions (Podlubny, 1999; Zhuang *et al.*, 2009), the Galerkin equations for random diffusion problems (Xiu & Karniadakis, 2002; Xiu & Shen, 2009) and time-dependent PDE-constrained optimization problems (Mathew *et al.*, 2010; Pearson *et al.*, 2012), etc. In these fields, the matrix A is usually symmetric and negative definite and the number of nonzero elements can be $\mathcal{O}(10^2) \sim \mathcal{O}(10^8)$. Efficient numerical approximation for these problems therefore 'requires' solving the large-scale linear systems of ODEs efficiently.

For long-time computation, it is meaningful to use a parallel strategy to reduce the total computational cost over the whole time interval. Here, we adopt the parareal algorithm invented by Lions *et al.* (2001) for this purpose. The parareal algorithm can be regarded as a variant of the multiple shooting method or the multigrid-in-time algorithm; see the work by Gander & Vandewalle (2007). The advantage of the parareal algorithm is that it allows the computation of the solution later in time, before having fully accurate approximations at earlier times, while the global accuracy of the iterative process after a few iterations is comparable with that given by a sequential numerical method used on a fine discretization in time. Nowadays, the parareal algorithm, as well as its variants proposed by Gander & Petcu (2008), Wu *et al.* (2009), Minion (2010) and Dai & Maday (2013), has been used in many fields by many authors, such as molecular-dynamics simulations by Baffico *et al.* (2002), morphological transformation simulations by He & He (2012), structural (fluid)-dynamics simulations by Cortial & Farhat (2008) and Farhat & Chandesris (2003), optimal control by Maday *et al.* (2007) and Mathew *et al.* (2010) and numerical approximation of the Hamiltonian by Dai *et al.* (2013) and Gander & Hairer (2014), turbulent plasma simulations by Reynolds-Barredo *et al.* (2012), numerical investigation of the acoustic–advection system by Ruprecht & Krause (2012), etc.

The parareal algorithm is defined by two time propagators, namely $\mathcal{F}$ and $\mathcal{G}$, which are associated with different step sizes. In general, there are three time steps in the parareal framework: the synchronization time step ΔT_s, the time step ΔT used by the coarse propagator to move forward by the time ΔT_s and the time step Δt used by the fine propagator to move forward by the time ΔT_s. Throughout this paper, we let $\Delta T = \Delta T_s$, i.e., the coarse propagator is always used for one time step. Moreover, we assume $\Delta T = J\Delta t$, where $J \geqslant 2$ is an integer. In the work by Gander & Vandewalle (2007), the authors presented a convergence analysis for the model problem

$$u' = \lambda u, \quad \lambda \in \mathbb{R}^-. \tag{1.2}$$

When the implicit Euler method is used as the coarse propagator, it was proved that the parareal algorithm is convergent if

$$\mathcal{R}_f^J\left(\frac{z}{J}\right) > \frac{1+z}{1-z}, \tag{1.3}$$

where $z = \lambda\Delta T < 0$ and $\mathcal{R}_f$ is the stability function of the fine propagator $\mathcal{F}$. For the symmetric negative-definite systems of ODEs (1.1), we need to prove that the parareal algorithm satisfies (1.3) for all $z < 0$ and because of the complexity of the stability function $\mathcal{R}_f$ of the $\mathcal{F}$-propagator, a rigorous proof is difficult. Such a proof is important, because it insures that the error between the iterate and the target solution (i.e., the converged solution) is contractive at each iteration. Naturally, it would be further desirable that the contraction factor is small and is insensitive to the change of mesh parameters: (1) a small contraction factor ensures that the error diminishes to zero rapidly; (2) the insensitivity ensures that refining the meshes to improve the accuracy of the converged solution does not cause deterioration of the convergence rate.

In a recent work by Mathew *et al.* (2010, see Lemma 4.3), the authors proved that for the parareal algorithm using the implicit Euler method as both the $\mathcal{G}$- and $\mathcal{F}$-propagators the error is contractive at each iteration and the contraction factor can be bounded from above by 0.298! The proof given by Mathew *et al.* (2010) depends on the simple form of the stability function of the implicit Euler method. For higher-order $\mathcal{F}$-propagators, such as the trapezoidal method (Hairer & Wanner, 2002), the TR/BDF2 method (Bank *et al.*, 1985a,b; Dharmaraja *et al.*, 2010) and the two-stage diagonally implicit Runge–Kutta (2s-DIRK) method, there are no results about the contraction factor so far. In this paper, we consider the parareal algorithm using the implicit Euler method as the $\mathcal{G}$-propagator and the aforementioned three time integrators as the $\mathcal{F}$-propagator. We denote these three algorithms by Parareal-TR, Parareal-TR/BDF2 and Parareal-2s-DIRK throughout this paper. For even integers $J \geqslant 2$, we prove that these three parareal algorithms satisfy (1.3) for all $z < 0$. Furthermore, we prove that for the Parareal-2s-DIRK algorithm the contraction factor can be bounded by 0.316—a quantity only slightly larger than 0.298. For the Parareal-TR/BDF2 algorithm, a free parameter $\gamma\,(\in (0,1))$ is involved and we prove that if $\gamma \in [0.043, 0.977]$, the contraction factor can be bounded by $\frac{1}{3}$. These two bounds are independent of J. However, for the algorithm Parareal-TR, the contraction factor cannot be uniformly bounded from above by a value strictly less than 1.

The remainder of this paper is organized as follows. In Section 2, we review the framework of the parareal algorithm. In Section 3, we prove that the algorithms Parareal-TR, Parareal-TR/BDF2 and Parareal-2s-DIRK satisfy (1.3) for all $z < 0$. In Section 4, we focus on deriving a sharp bound for the contraction factor for the three second-order parareal algorithms. Section 5 provides some numerical results to support our theoretical analysis. We finish this paper by presenting some conclusions in Section 6.

2. The parareal algorithm

For the system of ODEs

$$\begin{cases} u'(t) = f(t, u(t)), & t \in [0, T], \\ u(0) = u_0, & t = 0, \end{cases} \tag{2.1}$$

where the function $f\colon \mathbb{R} \times \mathbb{R}^m \to \mathbb{R}^m$ is Lipschitz continuous, the parareal algorithm can be described as follows. First, the whole time interval $[0, T]$ is divided into N large time-slices $[T_n, T_{n+1}]$, $n = 0, 1, \ldots, N-1$. We suppose that all the time-slices are of uniform size, i.e., $T_{n+1} - T_n = \Delta T = T/N$. Second, we divide each large time-slice $[T_n, T_{n+1}]$ into J $(\geqslant 2)$ small time-slices $[T_{n+(j\Delta T/J)}, T_{n+((j+1)\frac{\Delta T}{J})}]$, $j = 0, 1, \ldots, J-1$. Then, two numerical propagators $\mathcal{G}$ and $\mathcal{F}$ are assigned to the coarse and fine time grids, respectively, where $\mathcal{G}$ is usually a low-order and inexpensive numerical method and $\mathcal{F}$ is usually a higher-order method. We designate by the symbol $\ominus$ the time-sequential implementation, and by the symbol $\oplus$ the time-parallel implementation. Then, the parareal algorithm proposed by Lions *et al.* (2001) is given as in Algorithm 2.1.

Clearly, the argument $\tilde{u}_{n+1}$ can be written as $\tilde{u}_{n+1} = \mathcal{F}^J(T_n, u_n^k, \Delta t)$ and therefore Algorithm 2.1 can be written compactly as

$$u_{n+1}^{k+1} = \mathcal{G}(T_n, u_n^{k+1}, \Delta T) + \mathcal{F}^J(T_n, u_n^k, \Delta t) - \mathcal{G}(T_n, u_n^k, \Delta T), \tag{2.2}$$

where $\Delta t = \Delta T/J$ and $\mathcal{F}^J(T_n, u_n^k, \Delta t)$ denotes a value calculated by running J steps of the fine propagator $\mathcal{F}$ with initial value u_n^k and the fine step size Δt.

REMARK 2.2 At the $(k+1)$th iteration, the quantities $\{u_n^k\}_1^N$ are known from the kth iteration and hence N processors can be employed to compute the N quantities $\{\mathcal{F}^J(T_n, u_n^k, \Delta t)\}_{n=1}^N$ and $\{\mathcal{G}(T_n, u_n^k, \Delta T)\}_{n=1}^N$ simultaneously. Finally, upon convergence, the parareal algorithm (2.2) will generate a series of values $\{u_n\}_{n=1}^N$ which satisfy $u_{n+1} = \mathcal{F}^J(T_n, u_n, \Delta t)$. That is, the approximation at the coarse time point T_n will have achieved the accuracy of the $\mathcal{F}$-propagator.

Since the implicit Euler method is stable for all possible choices of step size ΔT (Hairer & Wanner, 2002) and in the context of the parareal algorithm the coarse propagator $\mathcal{G}$ is forced to take a large step

Algorithm 2.1 Parareal algorithm.

⊖ **Initialization**: Perform sequential computation $u_{n+1}^0 = \mathcal{G}(T_n, u_n^0, \Delta T)$ with $u_0^0 = u_0, n = 0, 1, \ldots, N-1$;

For $k = 0, 1, \ldots$

⊕ **Step 1** On each subinterval $[T_n, T_{n+1}]$, compute $\tilde{u}_{n+((j+1)/J)} = \mathcal{F}(T_{n+(j/J)}, \tilde{u}_{n+j/J}, \frac{\Delta T}{J})$ with initial value $\tilde{u}_n = u_n^k$, where $T_{n+j/J} = T_n + \frac{j\Delta T}{J}$ and $j = 0, 1, \ldots, J-1$;

⊖ **Step 2** Perform sequential corrections

$$u_{n+1}^{k+1} = \mathcal{G}(T_n, u_n^{k+1}, \Delta T) + \tilde{u}_{n+1} - \mathcal{G}(T_n, u_n^k, \Delta T),$$

where $u_0^{k+1} = u_0, n = 0, 1, \ldots, N-1$;

⊖ **Step 3** If $\{u_n^{k+1}\}_{n=1}^N$ satisfy stopping criteria, terminate the iteration; otherwise go to **Step 1**.

size, using the implicit Euler method as the coarse propagator is therefore reasonable. We can also consider some other implicit methods, such as multistage implicit Runge–Kutta methods (Radau, SDIRK, etc.), but they are all much more expensive than the implicit Euler method. For the fine propagator, we consider three time integrators and we describe them for problem (2.1):

- the trapezoidal method: $u_{n+1} = u_n + (\Delta t/2)[f(t_n, u_n) + f(t_{n+1}, u_{n+1})]$;
- the TR/BDF2 method ($\gamma \in (0, 1)$):

$$\begin{aligned} \frac{\tilde{u}_n - u_n}{\gamma \Delta t} &= \frac{1}{2}[f(t_n, u_n) + f(\tilde{t}_n, \tilde{u}_n)], \quad \tilde{t}_n = t_n + \gamma \Delta t, \\ \frac{\gamma(2-\gamma)u_{n+1} - \tilde{u}_n + (1-\gamma)^2 u_n}{\gamma \Delta t} &= (1-\gamma) f(t_{n+1}, u_{n+1}); \end{aligned} \tag{2.3}$$

- the 2s-DIRK method: $\tilde{u}_n = u_n + \theta \Delta t\, f(t_n + \theta \Delta t, \tilde{u}_n)$, $u_{n+1} = \hat{u}_n + \theta \Delta t\, f(t_{n+1}, u_{n+1})$, where $\theta = 1 - 1/\sqrt{2}$ and $\hat{u}_n = ((1-\theta)/\theta)\tilde{u}_n + (1 - (1-\theta)/\theta)u_n$.

From Hairer & Wanner (2002), we know that the trapezoidal method and the 2s-DIRK method are second-order methods. From the work by Bank *et al.* (1985a,b), we know that the TR/BDF2 method is also a second-order method. The algorithms using these three time integrators as the fine propagator are, respectively, called Parareal-TR, Parareal-TR/BDF2 and Parareal-2s-DIRK throughout this paper. The algorithm using the implicit Euler method as both the $\mathcal{G}$- and $\mathcal{F}$-propagators is denoted by Parareal-Euler.

Remark 2.3 The TR/BDF2 method is less known compared with the other two methods and here we provide some background for this method. TR/BDF2 was originally invented by Bank *et al.* (1985a,b) and it is self-starting: from (2.3), we see that the trapezoidal method determines $\tilde{\mathbf{u}}_n$ from $\mathbf{u}_n$, and then $\mathbf{u}_{n+1}$ comes from BDF2. The computing time for the TR/BDF2 method for the linear problems (1.1) is dominated by the solution of large-scale linear equations like $\mathcal{A}\mathbf{x} = \mathbf{b}$, where the coefficient matrix $\mathcal{A}$

and the vector **b** are

$$\mathcal{A}_{\mathrm{TR}} = I - \frac{\gamma \Delta t}{2} A, \quad \mathbf{b}_{\mathrm{TR}} = \mathbf{u}_n + \frac{\gamma \Delta t}{2}(A\mathbf{u}_n + \mathbf{g}_n + \tilde{\mathbf{g}}_n), \quad \text{(trapezoidal)}$$
$$\mathcal{A}_{\mathrm{BDF2}} = (2-\gamma)I - (1-\gamma)\Delta t A, \quad \mathbf{b}_{\mathrm{BDF2}} = \gamma^{-1}[\tilde{\mathbf{u}}_n - (1-\gamma)^2\mathbf{u}_n] + (1-\gamma)\Delta t \mathbf{g}_{n+1}, \quad \text{(BDF2)}.$$

It would be desirable if these two coefficient matrices were equal or proportional with the same step size Δt. If this is true, then the matrices $\mathcal{A}_{\mathrm{TR}}$ and $\mathcal{A}_{\mathrm{BDF2}}$ can be factored into LU once and for all. Fortunately, this purpose can be realized by letting $\gamma = 2 - \sqrt{2}$, which leads to

$$\sqrt{2}\left[I - \frac{\left(2-\sqrt{2}\right)\Delta t}{2} A\right] = \sqrt{2} I - (\sqrt{2}-1)\Delta t A \quad \text{(i.e., } \sqrt{2}\mathcal{A}_{\mathrm{TR}} = \mathcal{A}_{\mathrm{BDF2}}).$$

Besides this merit, Hosea & Shampine (1996) proved that the choice $\gamma = 2-\sqrt{2}$ results in the minimum local truncation error. Recently, Dharmaraja *et al.* (2010) proved that the choice $\gamma = 2-\sqrt{2}$ also results in the largest stability region on the complex plane and they call it the 'magic' choice. With the magic choice $\gamma = 2-\sqrt{2}$, the TR/BDF2 method is called `ode23tb` among the MATLAB solvers for systems of ODEs.

3. Convergence analysis

In this section, we prove convergence of the three second-order parareal algorithms. For the linear systems of ODEs (1.1) with $A \in \mathbb{R}^{m\times m}$ a symmetric negative-definite matrix (therefore it can be diagonalized and all the eigenvalues are negative real numbers), the following result concerning the convergence of the parareal algorithm on bounded and sufficiently long time intervals holds for any choice of the fine propagators and can be obtained directly by the analysis given by Gander & Vandewalle (2007).

THEOREM 3.1 (General results deduced from Gander & Vandewalle, 2007) Let $\mathcal{F}$ be a time integrator with stability function $\mathcal{R}_f(z)$ and let $\sigma(A) = \{\lambda_1, \ldots, \lambda_m\}$ be the set of eigenvalues of the matrix A in (1.1). Then, the errors $\{\mathbf{e}_n^k\}_{n=1}^N$ of the parareal algorithm using the implicit Euler method as the coarse propagator satisfy

$$\begin{aligned} T < +\infty: \max_{1\leqslant n\leqslant N} \|V\mathbf{e}_n^k\|_\infty &\leqslant \frac{(\rho^s(\Delta T, J))^k}{k!} \prod_{k_0=1}^{k} (N-k_0) \max_{1\leqslant n\leqslant N} \|V\mathbf{e}_n^0\|_\infty, \\ T = +\infty: \sup_n \|V\mathbf{e}_n^k\|_\infty &\leqslant (\rho^l(\Delta T, J))^k \sup_n \|V\mathbf{e}_n^0\|_\infty, \end{aligned} \tag{3.1}$$

where $N = T/\Delta T$, $V \in \mathbb{R}^{m\times m}$ consists of the eigenvectors of A (i.e., $V^{-1}AV = \mathrm{diag}(\lambda_1, \ldots, \lambda_m)$), $k \geqslant 1$ is the iteration index, and the arguments $\rho^s(\Delta T, J)$ and $\rho^l(\Delta T, J)$ are defined by

$$\begin{aligned} \rho^s(\Delta T, J) &= \max_{\lambda\in\sigma(A)} \mathcal{K}^s(\lambda\Delta T, J) \quad \text{with } \mathcal{K}^s(z, J) = \left|\mathcal{R}_f^J(z/J) - \frac{1}{1-z}\right|, \\ \rho^l(\Delta T, J) &= \max_{\lambda\in\sigma(A)} \mathcal{K}^l(\lambda\Delta T, J) \quad \text{with } \mathcal{K}^l(z, J) = \frac{\mathcal{K}^s(z, J)}{1 - |1/(1-z)|}. \end{aligned} \tag{3.2}$$

One can see that the algorithm converges superlinearly on bounded time intervals, as the division by $k!$ in (3.1) shows. Moreover, Theorem 3.1 shows that the parareal algorithm terminates, with a converged solution for any ΔT on any bounded time interval, in at most $N-1$ iterations. However, one would like to stop the iteration process with a converged or sufficiently accurate solution well before $N-1$ iterations, since otherwise there is no gain from using the parareal algorithm. It is therefore interesting and important to prove

$$\rho^s(\Delta T, J) < 1, \quad \rho^l(\Delta T, J) < 1. \tag{3.3}$$

To this end, it is sufficient to prove

$$\mathcal{K}^s(z, J) < 1, \ \ \mathcal{K}^l(z, J) < 1, \quad \forall z < 0. \tag{3.4}$$

3.1 *The Parareal-TR/BDF2 method*

We first consider the case that the TR/BDF2 method is used as the $\mathcal{F}$-propagator. In this case, the stability function $\mathcal{R}_f$ is (Dharmaraja *et al.*, 2010)

$$\mathcal{R}_f(z) = \mathcal{R}_{\text{TR-BDF2}}(z,\gamma) = \frac{2\gamma - 4 - (2-2\gamma+\gamma^2)z}{\gamma(\gamma-1)z^2 + (2-\gamma^2)z + 2\gamma - 4}, \quad \gamma \in (0,1).$$

Since the TR/BDF2 method is L-stable (Dharmaraja *et al.*, 2010), we have

$$|\mathcal{R}_{\text{TR-BDF2}}(z,\gamma)| < 1 \ (\forall z \in \mathbb{C}^-) \quad \text{and} \quad \lim_{z\to-\infty} |\mathcal{R}_{\text{TR-BDF2}}(z,\gamma)| = 0. \tag{3.5}$$

The following auxiliary lemma is useful for our analysis.

LEMMA 3.2 Let $\gamma \in (0,1)$, $z \in (0,+\infty)$ and $\Phi(z,\gamma) = (2\gamma - 4 + (2-2\gamma+\gamma^2)z)/(\gamma(\gamma-1)z^2 - (2-\gamma^2)z + 2\gamma - 4)$. Then for any $\gamma \in (0,1)$, the function $\Phi(z,\gamma)$ has a unique positive root located at $z_0 = 2((2-\gamma)/(2-2\gamma+\gamma^2))$. Moreover, we have

(1) for any $\gamma \in (0,1)$, Φ has a unique local minimizer

$$z_+ = 2\frac{\gamma(1-\gamma)(2-\gamma) + (2-\gamma)\sqrt{\gamma(1-\gamma)(2-\gamma)}}{\gamma(1-\gamma)(2-2\gamma+\gamma^2)};$$

(2) the quantities z_+ and z_0 satisfy $z_+ > z_0 > 1$.

Proof. It is clear that $\Phi(z,\gamma)$ has a unique positive root, which is $z_0 = 2((2-\gamma)/(2-2\gamma+\gamma^2))$. Routine calculation yields

$$\partial_z\Phi(z,\gamma) = \frac{\gamma(1-\gamma)(2-2\gamma+\gamma^2)z^2 - 4\gamma(1-\gamma)(2-\gamma)z - 4(2-\gamma)^2}{[\gamma(\gamma-1)z^2 - (2-\gamma^2)z + 2\gamma - 4]^2}. \tag{3.6}$$

Hence, $\partial_z\Phi(z,\gamma)$ has two roots:

$$z_\pm = 2\frac{\gamma(1-\gamma)(2-\gamma) \pm \sqrt{[\gamma(1-\gamma)(2-\gamma)]^2 + \gamma(1-\gamma)(2-\gamma)^2(2-2\gamma+\gamma^2)}}{\gamma(1-\gamma)(2-2\gamma+\gamma^2)}.$$

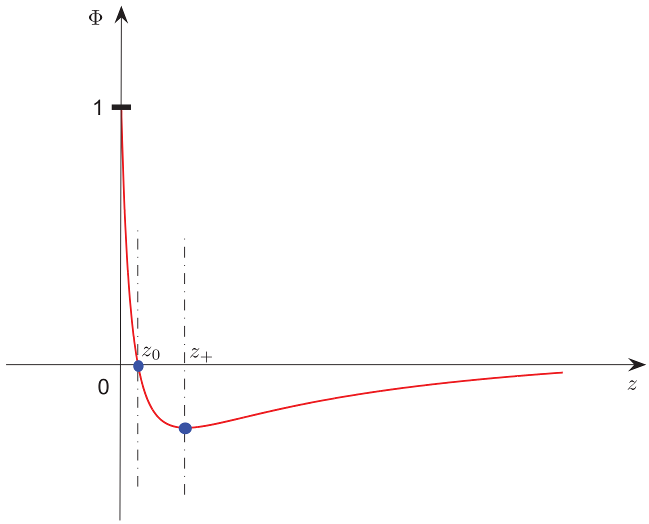

FIG. 1. The profile of Φ as a function of $z \in \mathbb{R}^+$.

The smaller one, z_-, can be excluded since $z_- < 0$ for $\gamma \in (0,1)$. Hence, for $z \in \mathbb{R}^+$ the function $\Phi(z,\gamma)$ has a unique local extreme located at $z = z_+$. Since $\Phi(0,\gamma) = 1$ and $|\Phi(z,\gamma)| < 1$ (this can be seen by noting that $\Phi(z,\gamma) = \mathcal{R}_{\text{TR-BDF2}}(-z,\gamma)$ and then using (3.5)), it can be straightforwardly deduced that z_+ is the unique local minimizer of Φ. For $\gamma \in (0,1)$, it is easy to get

$$z_+ = z_0 \left(\frac{\gamma(1-\gamma) + \sqrt{\gamma(1-\gamma)(2-\gamma)}}{\gamma(1-\gamma)} \right) > z_0.$$

Moreover, it holds that $z_0 - 1 = (2-\gamma^2)/(2-2\gamma+\gamma^2) > 0$, which gives $z_0 > 1$. □

From Lemma 3.2, we know that the function $\Phi(\gamma, z)$ is decreasing for $z \in (0, z_+)$ and is increasing for $z \geqslant z_+$. Typical behaviour of Φ is shown in Fig. 1.

THEOREM 3.3 (Parareal-TR/BDF2 on bounded time intervals) Let $J \geqslant 2$ be an even integer and let the $\mathcal{F}$-propagator of the parareal algorithm be the TR/BDF2 method. Then, the contraction factor $\mathcal{K}^s_{\text{TR-BDF2}}(z,\gamma,J)$ of the Parareal-TR/BDF2 algorithm on bounded time intervals satisfies

$$\mathcal{K}^s_{\text{TR-BDF2}}(z,\gamma,J) < 1 \quad \forall\, z < 0 \text{ and } \gamma \in (0,1), \tag{3.7}$$

where $\mathcal{K}^s_{\text{TR-BDF2}}(z,\gamma,J) = |\mathcal{R}^J_{\text{TR-BDF2}}(z/J,\gamma) - 1/(1-z)|$.

Proof. Since the TR/BDF2 method is used as the $\mathcal{F}$-propagator, we have

$$\mathcal{K}^s(z,J) := \mathcal{K}^s_{\text{TR-BDF2}}(z,\gamma,J) = \left| \Phi^J\left(-\frac{z}{J},\gamma\right) - \frac{1}{1-z} \right| \quad \forall\, z < 0. \tag{3.8}$$

Because $\mathcal{R}_{\text{TR-BDF2}}(z,\gamma) = \Phi(-z,\gamma)$, from (3.5) we know $\Phi^J(-z/J,\gamma) \in [0,1)$ (for all $(z,\lambda) \in \mathbb{R}^- \times (0,1)$), if J is an even integer. This, together with $1/(1-z) \in (0,1)$ (for all $z < 0$), implies (3.7). □

For odd J ($\geqslant 3$), we cannot prove (3.7) for all $\gamma \in (0,1)$. But it can be shown that (3.7) holds for the popular choices $\gamma = \frac{1}{2}$ and $\gamma = 2-\sqrt{2}$. Indeed, by using Lemma 3.2, we know that $\Phi^J(-z/J,\gamma) \in [0,1)$ for $z \in [-Jz_0, 0)$ and that $\Phi^J(-z/J,\gamma) \in [\Phi^J(Jz_+,\gamma), 0)$ for $z \leqslant (-\infty, -Jz_0)$. Hence, $\mathcal{K}^s_{\text{TR-BDF2}}(z,\gamma,J) < 1$ for $z \in [-Jz_0, 0)$. For $z < (-\infty, -Jz_0)$, it is easy to get

$$\max_{z \leqslant -Jz_0} \mathcal{K}^s_{\text{TR-BDF2}}(z,\gamma,J) = \max_{z \leqslant -Jz_0} \left| \Phi^J\left(-\frac{z}{J},\gamma\right) - \frac{1}{1-z} \right| < \frac{1}{1+Jz_0} - \Phi^J(z_+,\gamma)$$
$$\leqslant \frac{1}{1+3z_0} - \Phi^3(z_+,\gamma).$$

Then, for $\gamma = \frac{1}{2}$ and $\gamma = 2-\sqrt{2}$, it is easy to get $1/(1+3z_0) - \Phi^3(z_+,\gamma) \approx 0.04826655$ and 0.04775175, respectively.

Next, we consider the parareal algorithm as a time-integration method on sufficiently long time intervals. In this case, we need to prove $\mathcal{K}^l(z,J) < 1$ for all $z < 0$.

THEOREM 3.4 (Parareal-TR/BDF2 on sufficiently long time intervals) With the same assumptions stated in Theorem 3.3, the contraction factor $\mathcal{K}^l_{\text{TR-BDF2}}(z,\gamma,J)$ of the Parareal-TR/BDF2 algorithm on sufficiently long time intervals satisfies

$$\mathcal{K}^l_{\text{TR-BDF2}}(z,\gamma,J) < 1 \quad \forall z < 0 \text{ and } \gamma \in (0,1), \tag{3.9}$$

where

$$\mathcal{K}^l_{\text{TR-BDF2}}(z,\gamma,J) = \frac{\mathcal{K}^s_{\text{TR-BDF2}}(z,\gamma,J)}{1-|1/(1-z)|}.$$

Proof. Note that, for $z \in (-\infty, 0)$, it holds that

$$\mathcal{K}^l_{\text{TR-BDF2}}(z,\gamma,J) = \frac{|\Phi^J(-z/J,\gamma) - 1/(1-z)|}{1-1/(1-z)} = \frac{|(1-z)\Phi^J(-z/J,\gamma) - 1|}{-z},$$

where the function Φ is defined by Lemma 3.2. Hence, proposition (3.9) is equivalent to

$$\frac{|(1+z)\Phi^J(z/J,\gamma) - 1|}{z} < 1 \quad \forall z \in (0,\infty) \text{ and } \gamma \in (0,1). \tag{3.10}$$

Moreover, for $z > 0$ it is easy to get

$$\frac{|(1+z)\Phi^J(z/J,\gamma) - 1|}{z} < 1 \Leftrightarrow 1 - z < (1+z)\Phi^J\left(\frac{z}{J},\gamma\right) < 1 + z \Leftrightarrow \Phi^J\left(\frac{z}{J},\gamma\right) > \frac{1-z}{1+z},$$

where we have used the fact that, for even integers J ($\geqslant 2$), it holds that $\Phi^J(z/J,\gamma) \in [0,1)$. The proof of (3.9) is therefore equivalent to proving

$$\Phi^J\left(\frac{z}{J},\gamma\right) > \frac{1-z}{1+z} \quad \forall z \in (0,\infty) \text{ and } \gamma \in (0,1). \tag{3.11}$$

Clearly, for even J we just need to prove (3.11) for $z \in (0,1]$ and the proof consists of two parts: (a) proving (3.11) for $J = 1$ and (b) proving $\Phi^{J_2}(z/J_2,\gamma) > \Phi^{J_1}(z/J_1,\gamma)$ for $z \in (0,1]$ and any integers J_1 and J_2 satisfying $J_2 > J_1 \geqslant 1$. Claim (b) will be proved later (see Lemma 3.5) and claim (a) now. From Lemma 3.2 (or Fig. 1), it is easy to see that, for $z \in (0,1]$, the denominator of $\Phi(z,\gamma)$ is negative, i.e.,

$\gamma(\gamma-1)z^2-(2-\gamma^2)z+2\gamma-4<0$. A routine calculation yields

$$\Phi(z,\gamma)-\frac{1-z}{1+z}=\frac{2(\gamma-2)+(\gamma^2-\gamma)z+(\gamma^2-\gamma)z^2}{\gamma(\gamma-1)z^2-(2-\gamma^2)z+2\gamma-4}\times\frac{z}{1+z}.$$

It is easy to verify $2(\gamma-2)+(\gamma^2-\gamma)z+(\gamma^2-\gamma)z^2<0$ (for all $z\in(0,1],\gamma\in(0,1)$); thus (3.11) holds for $J=1$. □

LEMMA 3.5 Suppose $z\in(0,z_0J^*]$ and $\gamma\in(0,1)$. Then it holds that

$$\Phi^{J_2}\left(\frac{z}{J_2},\gamma\right)>\Phi^{J_1}\left(\frac{z}{J_1},\gamma\right)\quad\forall J_2>J_1\geqslant J^*,\tag{3.12}$$

where $z_0=(4-2r)/(2-2r+r^2)$ is the unique positive root of $\Phi(z,\gamma)$.

Proof. For any $J>J^*$, a partial derivative of $\Phi^J(z/J,\gamma)$ with respect to J leads to

$$\partial_J\left[\Phi^J\left(\frac{z}{J},\gamma\right)\right]=\Phi^J\left(\frac{z}{J},\gamma\right)\left[\log\Phi\left(\frac{z}{J},\gamma\right)-\frac{z}{J}\frac{\partial_z\Phi(z/J,\gamma)}{\Phi(z/J,\gamma)}\right].\tag{3.13}$$

Let $s=z/J$, $r(s)=\Phi(s,\gamma)$ and $R(s)=\log r(s)-s(r'(s)/r(s))$. Clearly, $\partial_J[\Phi^J(z/J,\gamma)]=r^J(s)R(s)$. Since $z\in(0,z_0J]$ and $J>J^*$, we have $s\in(0,z_0)$ and then from Lemma 3.2 we know $r(s)=\Phi(z/J,\gamma)>0$. Hence, proving (3.12) is equivalent to proving $R(s)>0$ for all $s\in(0,z_0)$ and $\gamma\in(0,1)$. To this end, we note $R(s)=\log r(s)-s(\mathrm{d}\log r(s)/\mathrm{d}s)$, which gives

$$R'(s)=\frac{\mathrm{d}\log r(s)}{\mathrm{d}s}-\frac{\mathrm{d}\log r(s)}{\mathrm{d}s}-s\frac{\mathrm{d}^2\log r(s)}{\mathrm{d}s^2}=-s\frac{\mathrm{d}(r'(s)/r(s))}{\mathrm{d}s}.$$

Let $\mathcal{N}(s)=2\gamma-4+(2-2\gamma+\gamma^2)s,\mathcal{D}(s)=\gamma(\gamma-1)s^2-(2-\gamma^2)s+2\gamma-4$ and then a tedious (but routine) calculation yields

$$\frac{\mathrm{d}(r'(s)/r(s))}{\mathrm{d}s}=\frac{R_1(s)+R_2(s)}{[\mathcal{N}(s)\mathcal{D}(s)]^2},\tag{3.14a}$$

where $R_1(s)=-(2-2\gamma+\gamma^2)^2\mathcal{D}^2(s)$ and $R_2(s)=\mathcal{N}^2(s)[(\mathcal{D}'(s))^2-2\gamma(\gamma-1)\mathcal{D}(s)]$. Clearly,

$$\mathcal{N}(s)<0,\ R_1(s)<0,\quad\forall(s,\gamma)\in(0,z_0)\times(0,1).\tag{3.14b}$$

For the function $R_2(s)$, we have $R_2'(s)=2\mathcal{N}(s)\bar{R}_2(s)$, where

$$\bar{R}_2(s)=(2-2\gamma+\gamma^2)[(\mathcal{D}'(s))^2-2\gamma(\gamma-1)\mathcal{D}(s)]+\gamma(\gamma-1)\mathcal{D}'(s)\mathcal{N}(s).$$

In the sequel, we prove $\bar{R}_2(s)\geqslant0$. Routine calculation yields

$$\begin{aligned}\bar{R}_2'(s)=&(2-2\gamma+\gamma^2)[2\mathcal{D}'(s)\mathcal{D}''(s)-2\gamma(\gamma-1)\mathcal{D}'(s)]+\gamma(\gamma-1)[\mathcal{D}''(s)\mathcal{N}(s)+\mathcal{D}'(s)\mathcal{N}'(s)]\\=&3\gamma(\gamma-1)(2-2\gamma+\gamma^2)\mathcal{D}'(s)+2\gamma^2(\gamma-1)^2\mathcal{N}(s),\end{aligned}$$

and thus $\bar{R}_2'(0)=\gamma(\gamma-1)(2-2\gamma+\gamma^2)(5\gamma^2-2\gamma-6)>0$ for all $\gamma\in(0,1)$. Moreover, it holds that

$$\bar{R}_2''(s)=3\gamma(\gamma-1)(2-2\gamma+\gamma^2)\mathcal{D}''(s)+2\gamma^2(\gamma-1)^2\mathcal{N}'(s)=8\gamma^2(\gamma-1)^2(2-2\gamma+\gamma^2)>0,$$

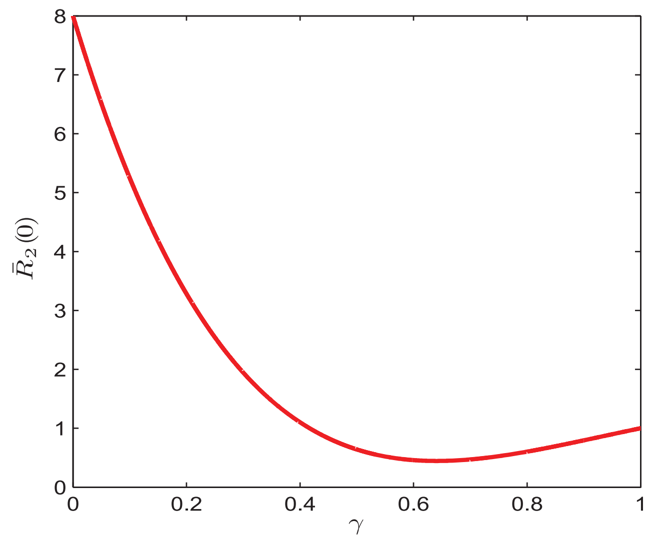

FIG. 2. Positivity of $\bar{R}_2(0)$ for $\gamma \in (0, 1)$.

and this, together with $\bar{R}'(0) > 0$, implies $\bar{R}_2'(s) > 0$. Note that $\bar{R}_2(0) = (2-\gamma^2)^2(2-2\gamma+\gamma^2) - 2\gamma(\gamma-1)(\gamma-2)(6-4\gamma+\gamma^2)$ is a sixth-order polynomial of γ and it can be shown numerically that $\bar{R}_2(0) > 0$ for all $\gamma \in (0, 1)$; see Fig. 2. Hence, by using $\bar{R}_2'(s) > 0$ we get $\bar{R}_2(s) > 0$ for all $(s, \lambda) \in (0, z_0) \times (0, 1)$. This, together with $R_2'(s) = 2\mathcal{N}(s)\bar{R}_2(s)$ and $\mathcal{N}(s) < 0$, gives

$$R_2(s) < 0 \quad \forall (s, \gamma) \in (0, z_0) \times (0, 1). \tag{3.14c}$$

Combining (3.14b) and (3.14c) leads to $\mathrm{d}(r'(s)/r(s))/\mathrm{d}s < 0$ in (3.14a), which further implies $R'(s) > 0$. Since $r(0) = \Phi(0, \gamma) = 1$, we get $R(0) = 0$. Now, it is clear that $R(s) > 0$ holds for all $(s, \gamma) \in (0, z_0) \times (0, 1)$ and this completes the proof. □

For odd J ($\geqslant 3$), it is, however, difficult to prove (3.9). This is because the function $\Phi(z/J, \gamma)$ will be negative for $z > Jz_0$ and therefore it is unclear whether (3.11) still holds for $z > Jz_0$.

3.2 *The parareal-TR algorithm*

For the Parareal-TR algorithm, the stability function of the $\mathcal{F}$-propagator is

$$\mathcal{R}_f(z) = \mathcal{R}_{\mathrm{TR}}(z) = \frac{2+z}{2-z}, \tag{3.15}$$

which can be regarded as a special case of the stability function of the TR/BDF2 method when $\gamma = 1$ or $\gamma = 0$, because $\mathcal{R}_{\text{TR-BDF2}}(z, 1) = \mathcal{R}_{\text{TR-BDF2}}(z, 0) = \mathcal{R}_{\mathrm{TR}}(z)$.

THEOREM 3.6 (Parareal-TR on bounded time intervals) Let $J \geqslant 2$ be an even integer and let the trapezoidal method be the $\mathcal{F}$-propagator. Then the contraction factor $\mathcal{K}^s_{\mathrm{TR}}(z, J)$ of the Parareal-TR algorithm on bounded time intervals satisfies

$$\mathcal{K}^s_{\mathrm{TR}}(z, J) < 1 \quad \forall z < 0, \tag{3.16}$$

where $\mathcal{K}^s_{\mathrm{TR}}(z, J) = |\mathcal{R}^J_{\mathrm{TR}}(z/J) - 1/(1-z)|$.

Proof. The proof is essentially the same as the proof for the Parareal-TR/BDF2 algorithm and we omit it. □

THEOREM 3.7 (Parareal-TR on sufficiently long time intervals) With the same assumptions as Theorem 3.6, the contraction factor $\mathcal{K}^l_{\mathrm{TR}}(z, J)$ of the Parareal-TR algorithm on sufficiently long time intervals satisfies

$$\mathcal{K}^l_{\mathrm{TR}}(z, J) < 1 \quad \forall z < 0, \tag{3.17}$$

where

$$\mathcal{K}^l_{\mathrm{TR}}(z, J) = \frac{\mathcal{K}^s_{\mathrm{TR}}(z, J)}{1 - |1/(1-z)|}.$$

Proof. The proof of this theorem is quite similar to that of Theorem 3.4. Particularly, for $z \in (0, 1]$ it can be proved that

(1) $(2-z)/(2+z) > (1-z)/(1+z)$ (i.e., (3.11) with $J = 1$ and $\gamma = 1$);

(2) $((2-z/J_2)/(2+z/J_2))^{J_2} > ((2-z/J_1)/(2+z/J_1))^{J_1}$ and for all $J_2 > J_1 \geqslant 1$ (i.e., Lemma 3.5 with $\gamma = 1$).

The first proposition can be easily verified. For the second one, we need to return to Lemma 3.5. For $\gamma = 1$, it is clear that $R_1(s) = -(s+2)^2$ and $R_2(s) = (s-2)^2$. Hence, $R_1(s) + R_2(s) < 0$ for $s \in (0, 1]$ and from (3.14a) we know $\mathrm{d}(r'(s)/r(s))/\mathrm{d}s < 0$ for all $s \in (0, 1]$. Therefore, we get $R'(s) > 0$ for $s \in (0, 1]$, which completes the proof. □

3.3 *The Parareal-2s-DIRK algorithm*

It remains to consider the case that the 2s-DIRK method is used as the $\mathcal{F}$-propagator. The stability function of the 2s-DIRK method is (Hairer & Wanner, 2002)

$$\mathcal{R}_{\text{2s-DIRK}}(z) = \frac{1 + z(1-2\theta)}{(1-z\theta)^2}, \quad \theta = 1 - \frac{1}{\sqrt{2}}.$$

Since the 2s-DIRK method is L-stable (Hairer & Wanner, 2002), we have

$$|\mathcal{R}_{\text{2s-DIRK}}(z)| < 1 \ (\forall z \in \mathbb{C}^-) \quad \text{and} \quad \lim_{z \to -\infty} |\mathcal{R}_{\text{2s-DIRK}}(z)| = 0. \tag{3.18}$$

LEMMA 3.8 For $z \in (0, +\infty)$ and $\theta = 1 - 1/\sqrt{2}$, the function $\Psi(z) = (1 - z(1-2\theta))/(1+z\theta)^2$ has a unique positive root located at $\tilde{z}_0 = 1/(1-2\theta)$ and a unique local minimizer $\tilde{z}_+ = 1/(\theta - 2\theta^2)$.

Proof. It is clear that $\Psi(z, \theta)$ has a unique positive root, which is $\tilde{z}_0 = 2((2-\theta)/(2-2\theta+\theta^2))$. Routine calculation yields

$$\Psi'(z) = \frac{z(\theta - 2\theta^2) - 1}{(1+z\theta)^3}. \tag{3.19}$$

Hence, $\Psi'(z)$ has a unique root $\tilde{z}_+ = 1/(\theta - 2\theta^2)$, which is the local minimizer of Ψ. □

From Lemma 3.8, we know that the function $\Psi(z)$ is decreasing for $z \in (0, \tilde{z}_+)$ and is increasing for $z \geqslant \tilde{z}_+$.

Theorem 3.9 (Parareal-2s-DIRK on bounded time intervals) Let $J \geqslant 2$ be an even integer and let the $\mathcal{F}$-propagator of the parareal algorithm be the 2s-DIRK method. Then, the contraction factor $\mathcal{K}^s_{\text{2s-DIRK}}(z, J)$ of the Parareal-2s-DIRK algorithm on bounded time intervals satisfies

$$\mathcal{K}^s_{\text{2s-DIRK}}(z, J) < 1 \quad \forall z < 0, \tag{3.20}$$

where $\mathcal{K}^s_{\text{2s-DIRK}}(z, J) = |\mathcal{R}^J_{\text{2s-DIRK}}(z/J) - 1/(1-z)|$.

Proof. Since $\mathcal{R}_{\text{2s-DIRK}}(z) = \Psi(-z)$, we have

$$\mathcal{K}^s_{\text{2s-DIRK}}(z, J) = \left| \Psi^J\left(-\frac{z}{J}\right) - \frac{1}{1-z} \right| \quad \forall z < 0. \tag{3.21}$$

From (3.18), we know that, for even integers J $(\geqslant 2)$, it holds that $\Psi^J(-z/J) \in [0, 1)$ (for all $z < 0$). This, together with $1/(1-z) \in (0, 1)$ (for all $z < 0$), implies (3.20). □

Theorem 3.10 (Parareal-2s-DIRK on sufficiently long time intervals) With the same assumptions stated in Theorem 3.9, the contraction factor $\mathcal{K}^l_{\text{2s-DIRK}}(z, J)$ of the Parareal-2s-DIRK algorithm on sufficiently long time intervals satisfies

$$\mathcal{K}^l_{\text{2s-DIRK}}(z, J) < 1 \quad \forall z < 0, \tag{3.22}$$

where

$$\mathcal{K}^l_{\text{2s-DIRK}}(z, J) = \frac{\mathcal{K}^s_{\text{2s-DIRK}}(z, J)}{1 - |1/(1-z)|}.$$

Proof. It holds that

$$\mathcal{K}^l_{\text{2s-DIRK}}(z, J) = \frac{|\Psi^J(-z/J) - 1/(1-z)|}{1 - 1/(1-z)} = \frac{|(1-z)\Psi^J(-z/J) - 1|}{-z} \quad \forall z < 0,$$

where the function Ψ is defined by Lemma 3.8. Hence, proposition (3.22) is equivalent to

$$\frac{|(1+z)\Psi^J(z/J) - 1|}{z} < 1 \quad \forall z \in (0, \infty). \tag{3.23}$$

For $z > 0$, we have

$$\frac{|(1+z)\Psi^J(z/J) - 1|}{z} < 1 \Leftrightarrow 1 - z < (1+z)\Psi^J\left(\frac{z}{J}\right) < 1 + z \Leftrightarrow \Psi^J\left(\frac{z}{J}\right) > \frac{1-z}{1+z},$$

where we have used $\Psi^J(z/J) \in [0, 1)$ for even integers J $(\geqslant 2)$. The proof of (3.22) is therefore equivalent to

$$\Psi^J\left(\frac{z}{J}\right) > \frac{1-z}{1+z} \quad \forall z \in (0, \infty). \tag{3.24}$$

Clearly, for even J we just need to prove (3.24) for $z \in (0, 1]$ and this procedure is similar to the one for the Parareal-TR/BDF2 method. To be precise, the proof is also divided into two parts: (a) proving (3.24) for $J = 1$ and (b) proving $\Psi^{J_2}(z/J_2) > \Psi^{J_1}(z/J_1)$ for $z \in (0, 1]$ and any integers J_1 and J_2 satisfying

$J_2 > J_1 \geqslant 1$. Claim (b) will be proved later (see Lemma 3.11) and claim (a) now. A routine calculation yields

$$\Psi(z) - \frac{1-z}{1+z} = \frac{(1-\theta z)^2 + (6\theta - 1 - \theta^2)z}{(1+\theta z)^3}.$$

For $\theta = 1 - 1/\sqrt{2}$, it is easy to verify $6\theta - 1 - \theta^2 > 0$. Hence, we know that (3.24) holds for $J = 1$. $\square$

LEMMA 3.11 Suppose $z \in (0, \tilde{z}_0 J^*]$ and then it holds that

$$\Psi^{J_2}\left(\frac{z}{J_2}\right) > \Psi^{J_1}\left(\frac{z}{J_1}\right) \quad \forall\, J_2 > J_1 \geqslant J^*, \tag{3.25}$$

where $\tilde{z}_0 = 1/(1-2\theta)$ is the unique positive root of $\Psi(z)$.

Proof. For any $J > J^*$, a derivative of $\Psi^J(z/J)$ with respect to J gives

$$\partial_J\left[\Psi^J\left(\frac{z}{J}\right)\right] = \Psi^J\left(\frac{z}{J}\right)\left[\log\Psi\left(\frac{z}{J}\right) - \frac{z}{J}\frac{\Psi'(z/J)}{\Psi(z/J)}\right]. \tag{3.26}$$

Let $s = z/J$, $r(s) = \Psi(s)$ and $R(s) = \log r(s) - s(r'(s)/r(s))$. Clearly, $\partial_J[\Psi^J(z/J)] = r^J(s)R(s)$. Since $z \in (0, \tilde{z}_0 J]$ and $J > J^*$, we have $s \in (0, \tilde{z}_0)$ and then from Lemma 3.8 we know $r(s) = \Psi(z/J) > 0$. Hence, proving (3.25) is equivalent to proving $R(s) > 0$ for all $s \in (0, \tilde{z}_0)$. To this end, we note $R(s) = \log r(s) - s(\mathrm{d}\log r(s)/\mathrm{d}s)$ and this gives

$$R'(s) = \frac{\mathrm{d}\log r(s)}{\mathrm{d}s} - \frac{\mathrm{d}\log r(s)}{\mathrm{d}s} - s\frac{\mathrm{d}^2\log r(s)}{\mathrm{d}s^2} = -s\frac{\mathrm{d}(r'(s)/r(s))}{\mathrm{d}s}.$$

A tedious (but routine) calculation yields

$$\frac{\mathrm{d}(r'(s)/r(s))}{\mathrm{d}s} = \frac{z^2(\theta - 2\theta^2)^2 - 2z(\theta - 2\theta^2) + 4\theta - 2\theta^2 - 1}{[(1+z\theta)(1-z(1-2\theta))]^2}. \tag{3.27}$$

For $\theta = 1 - 1/\sqrt{2}$ and $z \in (0, \tilde{z}_0)$, it is easy to verify $z^2(\theta - 2\theta^2)^2 - 2z(\theta - 2\theta^2) + 4\theta - 2\theta^2 - 1 < 0$. This implies $\mathrm{d}(r'(s)/r(s))/\mathrm{d}s < 0$, which further implies $R'(s) > 0$. Since $r(0) = \Psi(0) = 1$, we get $R(0) = 0$. Now, it is clear that $R(s) > 0$ holds for all $s \in (0, \tilde{z}_0)$ and this completes the proof. $\square$

Similarly to the analysis for the Parareal-TR/BDF2 algorithm, it is difficult to prove (3.22) for odd integers J ($\geqslant 3$) because in this case the function $\Psi(z/J)$ will be negative for $z > J\tilde{z}_0$ and therefore it is unclear whether (3.24) still holds for $z > J\tilde{z}_0$.

4. Sharp bound for the contraction factor on sufficiently long time intervals

We have proved global contraction $\mathcal{K}^l(z, J) < 1$ over the whole negative z-axis. However, this just guarantees convergence of the parareal algorithm and so far we know nothing about the convergence rate. To make a qualitative evaluation of the algorithm, it is necessary to prove that $\mathcal{K}^l(z, J)$ can be bounded from above by a constant $\varrho \in (0, 1)$, i.e., $\mathcal{K}^l(z, J) \leqslant \varrho$ for all $z < 0$. If this is true for some choice of the coarse and the fine propagators, the corresponding parareal algorithm will be robust with respect to the mesh parameters, Δx, ΔT and J, etc.

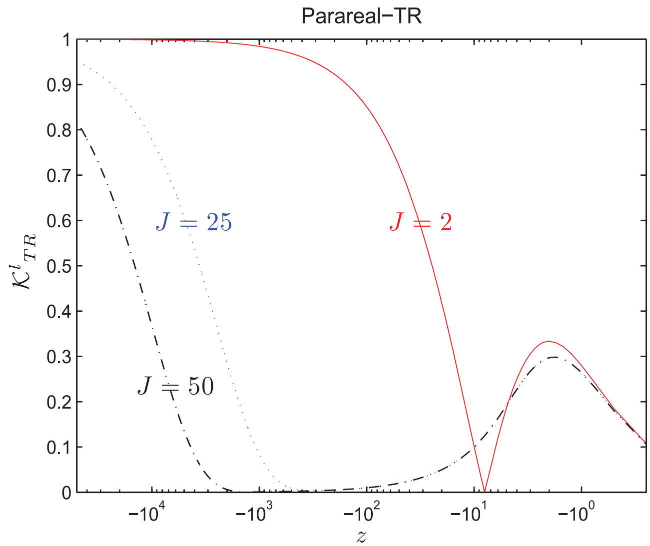

FIG. 3. The contraction factor $\mathcal{K}^l_{\mathrm{TR}}(z, J)$ as a function of z.

We begin our analysis from numerical illustrations. In Fig. 3, we have plotted $\mathcal{K}^l_{\mathrm{TR}}(z, J)$ as a function of z for three choices of J, from which one can deduce that the Parareal-TR algorithm cannot behave robustly with respect to refinement of meshes because $\mathcal{K}^l_{\mathrm{TR}}(z, J) \to 1$ as $z \to -\infty$. In fact, since $\lim_{z\to-\infty} \mathcal{R}_{\mathrm{IM}}(z) = 0$ and $\lim_{z\to-\infty} \mathcal{R}_{\mathrm{TR}}(z) = -1$, it is easy to get $\lim_{z\to-\infty} \mathcal{K}^l_{\mathrm{TR}}(z, J) = 1$ for all $J \geqslant 2$.

For the other two parareal algorithms studied in this paper, we have

$$\begin{aligned}
\lim_{z\to0-} \mathcal{K}^l_{\text{TR-BDF2}}(z, \gamma, J) &= \lim_{z\to0+} \frac{|(1+z)\Phi^J(z/J, \gamma) - 1|}{z} \\
&= \lim_{z\to0+} \left| \Phi^J\left(\frac{z}{J}, \gamma\right) + (1+z)\Phi^{J-1}\left(\frac{z}{J}, \gamma\right) \partial_z \Phi\left(\frac{z}{J}, \gamma\right) \right| = 0, \\
\lim_{z\to0-} \mathcal{K}^l_{\text{2s-DIRK}}(z, J) &= \lim_{z\to0+} \frac{|(1+z)\Psi^J(z/J) - 1|}{z} \\
&= \lim_{z\to0+} \left| \Psi^J\left(\frac{z}{J}\right) + (1+z)\Psi^{J-1}\left(\frac{z}{J}\right) \partial_z \Psi\left(\frac{z}{J}\right) \right| = 0,
\end{aligned} \tag{4.1}$$

where we have used $\partial_z \Phi(z/J, \gamma)|_{z=0} = -1$ (respectively, $\partial_z \Psi(z/J)|_{z=0} = -1$) and this can be obtained by letting $z = 0$ in (3.6) (respectively, (3.19)). The above calculation implies that refining ΔT and/or pushing the matrix A to be nearly singular (i.e., the eigenvalue(s) of A close to zero) will not cause deterioration of the convergence rate of the parareal algorithm. Besides this, we also have

$$\lim_{z\to-\infty} \mathcal{K}^l_{\text{TR-BDF2}}(z, \gamma, J) = 0 \ (\forall\, \gamma \in (0, 1)), \quad \lim_{z\to-\infty} \mathcal{K}^l_{\text{2s-DIRK}}(z, J) = 0, \tag{4.2}$$

and this implies that refining the spatial mesh Δx, to obtain a more accurate space approximation, will not cause deterioration of the convergence rate of the parareal algorithm, either. In Fig. 4, we have plotted the curves of $\mathcal{K}^l_{\text{TR-BDF2}}(z, \gamma, J)$ (with $\gamma = 2 - \sqrt{2}$) and $\mathcal{K}^l_{\text{2s-DIRK}}(z, J)$ for several choices of J,

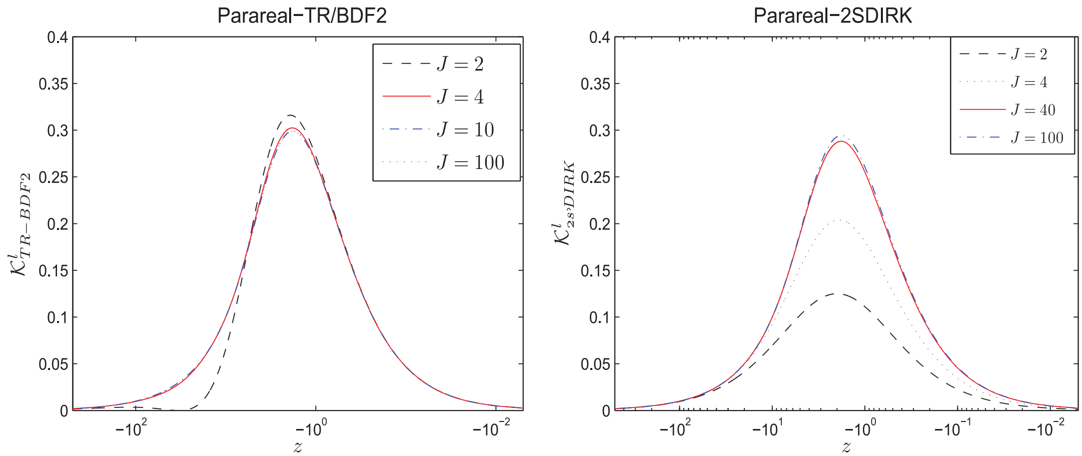

FIG. 4. The contraction factors $\mathcal{K}^l_{\text{TR-BDF2}}(z,\gamma,J)$ (with $\gamma = 2-\sqrt{2}$) and $\mathcal{K}^l_{\text{2s-DIRK}}(z,J)$ as functions of z.

where we see the same property implied by (4.1) and (4.2). Moreover, we see that, for given J, both $\mathcal{K}^l_{\text{TR-BDF2}}(z,\gamma,J)$ and $\mathcal{K}^l_{\text{2s-DIRK}}(z,J)$ 'have a local maximum'. Hence, the properties (4.1) and (4.2) are not sufficient to evaluate the parareal algorithm and we need to know whether the contraction factor can be uniformly bounded from above by a value (strictly less than 1).

THEOREM 4.1 (Parareal-TR/BDF2) Let z_0 and z_+ be the arguments given by Lemma 3.2. Then, for any even integer J ($\geqslant 2$), it holds that

$$\max_{z\leqslant 0} \mathcal{K}^l_{\text{TR-BDF2}}(z,\gamma,J) \leqslant \varrho^*, \tag{4.3}$$

where ϱ^* is the minimal value of ϱ $(\in [\max\{1/2z_0, \Phi^2(z_+,\gamma)\}, (2-\gamma)/(2-\gamma^2)))$ such that $\mathcal{P}_4(z^*(\varrho),\varrho) \geqslant 0$. Here, for each ϱ, the function $\mathcal{P}_4(z,\varrho)$ is a fourth-order polynomial of z and $z^*(\varrho)$ is the unique positive root of $\partial_z \mathcal{P}_4(z,\varrho)$:

$$\begin{aligned}\mathcal{P}_4(z,\varrho) = {} & \varrho(\gamma-\gamma^2)^2 z^4 + (\gamma-\gamma^2)[4\varrho(2-\gamma^2)+\gamma^2-\gamma]z^3 + 4[R(\gamma)\varrho + S(\gamma)]z^2 \\ & - 32(2-\gamma)(2-\gamma-\varrho(2-\gamma^2))z + 64\varrho(2-\gamma)^2,\end{aligned} \tag{4.4}$$

where $R(\gamma) = \gamma^4 + 4\gamma^3 - 16\gamma^2 + 8\gamma + 4$ and $S(\gamma) = -3\gamma^3 + 10\gamma^2 - 10\gamma + 4$.

Proof. For any $\varrho > 0$, we have

$$\mathcal{K}^l_{\text{TR-BDF2}}(z,\gamma,J) \leqslant \varrho \ (\forall\, z<0) \Leftrightarrow \frac{1-\varrho z}{1+z} \leqslant \Phi^J\left(\frac{z}{J},\gamma\right) \leqslant \frac{1+\varrho z}{1+z} \ (\forall\, z>0). \tag{4.5}$$

This implies that a sharp bound of the contraction factor is the minimal value of ϱ such that the two inequalities $(1-\varrho z)/(1+z) \leqslant \Phi^J(z/J,\gamma)$ and $\Phi^J(z/J,\gamma) \leqslant (1+\varrho z)/(1+z)$ hold. Our analysis is therefore divided into two steps: in Step A, we prove the first inequality for $\varrho \in [\varrho^*, (2-\gamma)/(2-\gamma^2))$ with $\varrho^* \geqslant 1/2z_0$ and in Step B, we prove the second one for $\varrho \geqslant \Phi^2(z_+,\gamma)$.

Step A. *The proof of* $(1-\varrho z)/(1+z) \leqslant \Phi^J(z/J,\gamma)$ *for* $\varrho \in [\varrho^*, (2-\gamma)/(2-\gamma^2))$ *and* $\varrho^* \geqslant 1/2z_0$. Clearly, for even integers J ($\geqslant 2$) we just need to consider $z \in (0, 1/\varrho)$. Since $1/\varrho \leqslant 2z_0$, by using Lemma 3.5 we know that $\Phi^J(z/J,\gamma) \geqslant \Phi^2(z/2,\gamma)$ holds for all $J \geqslant 2$ and $z \in (0,1/\varrho)$. Hence, it is sufficient to prove $(1-\varrho z)/(1+z) \leqslant \Phi^2(z/2,\gamma)$ for $z \in (0,1/\varrho)$ and a tedious but very routine calculation yields that this is equivalent to $\mathcal{P}_4(z,\varrho) \geqslant 0$, i.e.,

$$\frac{1-\varrho z}{1+z} \leqslant \Phi^J\left(\frac{z}{J},\gamma\right) \Leftarrow \frac{1-\varrho z}{1+z} \leqslant \Phi^2\left(\frac{z}{2},\gamma\right) \Leftrightarrow \mathcal{P}_4(z,\varrho) \geqslant 0 \quad \forall z \in (0,1/\varrho).$$

For a given ϱ, we need to know the minimum of $\mathcal{P}_4(z,\varrho)$ over the interval $(0,1/\varrho)$. To this end, we note that, for any $\gamma \in (0,1)$, it holds that $R(\gamma) \geqslant 0$, $S(\gamma) \geqslant 0$ and $2(2-\gamma^2)/z_0 + \gamma^2 - \gamma > 0$ (these can be verified easily). This implies that the cubic polynomial $\mathcal{P}_3(z,\varrho) := \partial_z \mathcal{P}_4(z,\varrho)$ is a monotonic increasing function, since all the coefficients of the quadratic polynomial $\mathcal{P}_2(z,\varrho) := \partial_{zz}\mathcal{P}_4(z,\varrho)$ are non-negative. We next claim that, for given $\varrho \in [\varrho^*, (2-\gamma)/(2-\gamma^2))$, the function $\mathcal{P}_4(z,\varrho)$ has a unique minimizer located at $z = z^*(\varrho)$, where $z^*(\varrho)$ is the unique root of $\partial_z\mathcal{P}_4(z,\varrho)$. Note that, for $\varrho < (2-\gamma)/(2-\gamma^2)$, it holds that $\partial_z \mathcal{P}_4(0,\varrho) = -32(2-\gamma)(2-\gamma-\varrho(2-\gamma^2)) < 0$ and

$$\partial_z\mathcal{P}_4(1/\varrho,\varrho) \geqslant \frac{8}{\varrho}\tilde{P}(\varrho,\gamma), \tag{4.6}$$

where $\tilde{P}(\varrho,\gamma) = 4\varrho^2(2-\gamma)(2-\gamma^2) + (R(\gamma) - 4(2-\gamma)^2)\varrho + S(\gamma)$. For any $\gamma \in (0,1)$, the function $\tilde{P}(\varrho,\gamma)$ is a quadratic polynomial of ϱ and its symmetry axis is $\tilde{\varrho}(\gamma) := (4(2-\gamma)^2 - R(\gamma))/(2(2-\gamma)(2-\gamma^2))$, which corresponds to the global minimizer of $\tilde{P}(\varrho,\gamma)$. By regarding $\tilde{P}(\tilde{\varrho}(\gamma),\gamma)$ as a function of γ, we have plotted its profile in Fig. 5 and it is clear that $\tilde{P}(\tilde{\varrho}(\gamma),\gamma) > 0$ for all $\gamma \in (0,1)$. Hence, from (4.6) we have $\partial_z\mathcal{P}_4(1/\varrho,\varrho) > 0$ and this together with $\partial_z\mathcal{P}_4(0,\varrho) < 0$ implies that the function $\partial_z\mathcal{P}_4(z,\varrho)$ has a unique root $z^*(\varrho) \in (0,1/\varrho)$, which is the global minimizer of $\mathcal{P}_4(z,\varrho)$. Since $\partial_\varrho\mathcal{P}_4(z,\varrho) > 0$ and $\partial_z\mathcal{P}_4(z^*(\varrho),\varrho) = 0$, we have

$$\frac{\mathrm{d}\mathcal{P}_4(z^*(\varrho),\varrho)}{\varrho} = \left.\frac{\partial\mathcal{P}_4(z,\varrho)}{\partial\varrho}\right|_{z=z^*(\varrho)} + \left.\frac{\partial\mathcal{P}_4(z,\varrho)}{\partial z}\right|_{z=z^*(\varrho)} = \left.\frac{\partial\mathcal{P}_4(z,\varrho)}{\partial\varrho}\right|_{z=z^*(\varrho)} > 0,$$

and this implies $\min_{z\in(0,1/\varrho)}\mathcal{P}_4(z,\varrho) \geqslant 0$ $(\Rightarrow (1-\varrho z)/(1+z) \leqslant \Phi^J(z/J,\gamma))$, provided $(2-\gamma)/(2-\gamma^2) > \varrho \geqslant \varrho^*$ and $\varrho^* \geqslant 1/2z_0$.

Step B. *The proof of* $\Phi^J(z/J,\gamma) \leqslant (1+\varrho z)/(1+z)$ *for* $\varrho \geqslant \Phi^2(z_+,\gamma)$. We prove this inequality on two intervals: $z \in (0, Jz_0)$ and $z \in [Jz_0, +\infty)$. For the case $z \in (0,Jz_0)$, we know from Lemma 3.5 that

$$\Phi^J\left(\frac{z}{J},\gamma\right) \leqslant \lim_{J\to+\infty}\Phi^J\left(\frac{z}{J},\gamma\right) = \lim_{J\to+\infty}\left(1-\frac{z}{J}\mathcal{I}(J)\right)^J \quad \forall z \in (0,Jz_0), \tag{4.7}$$

where $\mathcal{I}(J) = ((\gamma^2-\gamma)(z/J) + 2\gamma - 4)/((\gamma^2-\gamma)(z/J)^2 - (2-\gamma^2)(z/J) + 2\gamma - 4)$ and it satisfies $\mathcal{I}(J) > 1$ for $J \gg 1$ and $\lim_{J\to+\infty}\mathcal{I}(J) = 1$. Hence, we have

$$\Phi^J\left(\frac{z}{J},\gamma\right) \leqslant e^{-z} \quad \forall z \in (0,Jz_0). \tag{4.8}$$

Since $e^{-z} < 1/(1+z)$ (for all $z > 0$), we have $\Phi^J(z/J,\gamma) < 1/(1+z) < (1+\varrho z)/(1+z)$ for all $\varrho > 0$ and $z > 0$. It remains to consider the case $z \geqslant Jz_0$. From Lemma 3.2, we know that, for even integers J ($\geqslant 2$), the function $\Phi^J(z/J,\gamma)$ has a unique local maximum $\Phi^J(z_+,\gamma)$—which is also the global

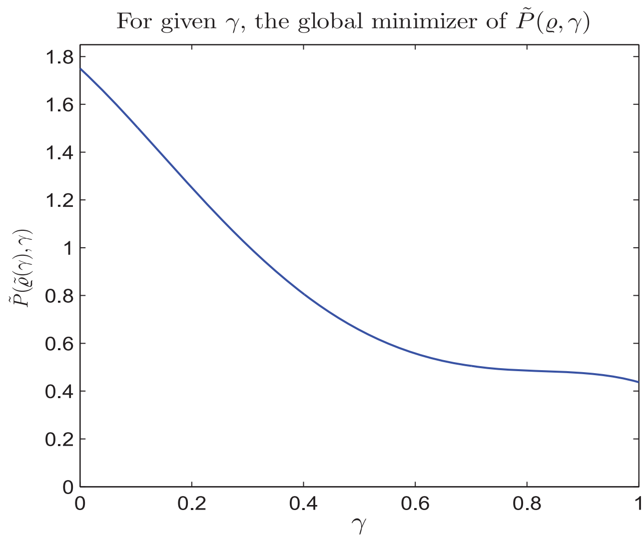

FIG. 5. Positivity of the global minimizer $\tilde{P}(\tilde{\varrho}(\gamma),\gamma)$ of $\tilde{P}(\varrho,\gamma)$.

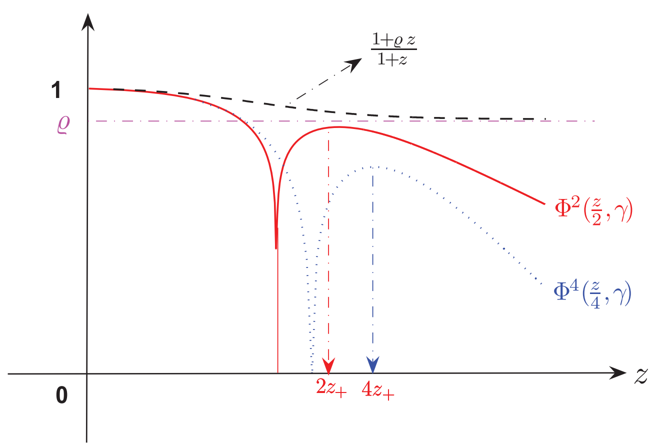

FIG. 6. An illustration of the properties $(1+\varrho z)/(1+z) \geqslant \varrho$ and $\lim_{z\to\infty}((1+\varrho z)/(1+z)) = \varrho$, and $\Phi^J(z_+,\gamma)$ decreases as J increases.

maximum for $z \in [Jz_0, +\infty)$, located at $z = Jz_+$. Hence, to prove $\Phi^J(z/J,\gamma) \leqslant (1+\varrho z)/(1+z)$ for $z \geqslant Jz_0$, it is sufficient to prove

$$\Phi^J(z_+,\gamma) \leqslant \frac{1+\varrho z}{1+z} \quad \forall z \in (0, Jz_0). \tag{4.9}$$

Since $(1+\varrho z)/(1+z) \geqslant \varrho$, $\lim_{z\to\infty}((1+\varrho z)/(1+z)) = \varrho$ and $\Phi^J(z_+,\gamma)$ decreases as J increases (see Fig. 6 for illustration), it is sufficient to require $\varrho \geqslant \Phi^2(z_+,\gamma)$ to ensure (4.9). This, together with (4.8), completes the proof of Step B. Now, combining Steps A and B finishes the whole proof. □

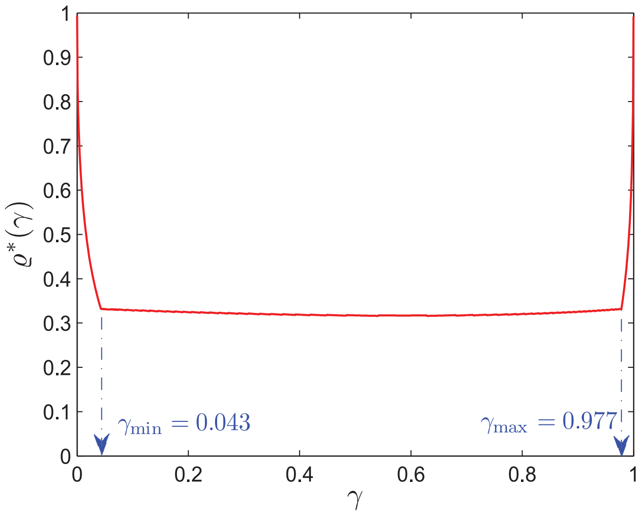

FIG. 7. The bound of the contraction factor $\mathcal{K}^l_{\text{TR-BDF2}}$ (i.e., $\varrho^*(\gamma)$) as a function of γ, a free parameter involved in the TR/BDF2 method.

REMARK 4.2 The bound ϱ^* of the contraction factor $\mathcal{K}^l_{\text{TR/BDF2}}$ only depends on the time integration parameter γ. For a given $\gamma \in (0,1)$, it is very easy to get ϱ^* numerically:

$$\varrho^*(\gamma) := \min \varrho, \qquad \text{subject to} \begin{cases} \max\left\{\dfrac{1}{2z_0}, \Phi^2(z_+,\gamma)\right\} \leqslant \varrho < \dfrac{2-\gamma}{2-\gamma^2}, \\ \mathcal{P}_4(z^*(\varrho),\varrho) \geqslant 0, \end{cases} \tag{4.10}$$

where $z^*(\varrho)$ and $\mathcal{P}_4$ are defined by Theorem 4.1. In Fig. 7, we have plotted the profile of $\varrho^*(\gamma)$ and it can be found that (1) for $0.043 \leqslant \gamma \leqslant 0.977$, it holds that $\varrho^*(\gamma) \simeq \frac{1}{3}$; (2) for $\gamma \to 0$ (or $\gamma \to 1$), the bound ϱ^* rapidly approaches 1. An explanation for the second observation is that, for $\gamma \to 0$ or $\gamma \to 1$, the stability function of the TR/BDF2 method is reduced to that of the trapezoidal method, and the contraction factor $\mathcal{K}^l_{\text{TR}}$ approaches 1 when $z \to -\infty$ (see the curves plotted in Fig. 3).

THEOREM 4.3 (Parareal-2s-DIRK) Let $\tilde{z}_0$ and $\tilde{z}_+$ be the arguments given by Lemma 3.8. Then for any even integer J ($\geqslant 2$), it holds that

$$\max_{z \leqslant 0} \mathcal{K}^l_{\text{2s-DIRK}}(z,J) \leqslant 0.316. \tag{4.11}$$

Proof. Similarly to the analysis for the Parareal-TR/BDF2 algorithm, for any $\varrho > 0$ we have

$$\mathcal{K}^l_{\text{2s-DIRK}}(z,J) \leqslant \varrho \ (\forall\, z<0) \Leftrightarrow \frac{1-\varrho z}{1+z} \leqslant \Psi^J\left(\frac{z}{J}\right) \leqslant \frac{1+\varrho z}{1+z} \ (\forall\, z>0), \tag{4.12}$$

where Ψ is defined in Lemma 3.8. From (4.12), we focus in the sequel on proving $(1-\varrho z)/(1+z) \leqslant \Psi^J(z/J)$ and $\Psi^J(z/J) \leqslant (1+\varrho z)/(1+z)$ for $\varrho \geqslant 0.316$ and this is divided into two steps.

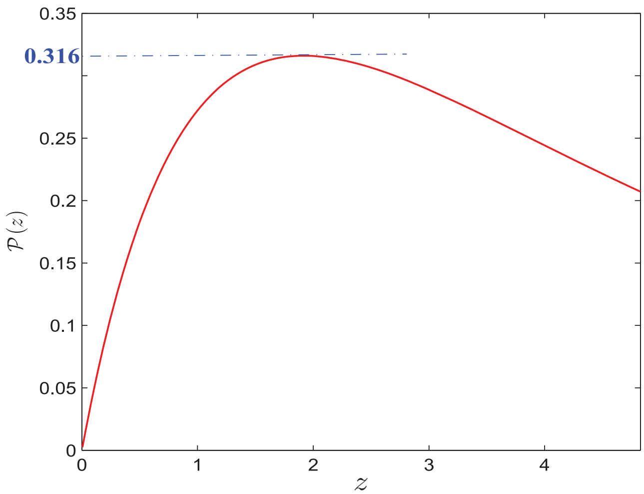

FIG. 8. Global maximum of $\mathcal{P}(z)$ is 0.316.

Step I. *The proof of* $(1-\varrho z)/(1+z) \leqslant \Psi^J(z/J)$ *with* $\varrho \geqslant 1/2\tilde{z}_0$. Clearly, for even integers J ($\geqslant$ 2) we just need to consider $z \in (0, 1/\varrho)$. Since $1/\varrho \leqslant 2\tilde{z}_0$, by using Lemma 3.11 we have $\Psi^J(z/J) \geqslant \Psi^2(z/2)$ for all $J \geqslant 2$ and for all $z \in (0, 1/\varrho)$. Hence, it is sufficient to prove $(1-\varrho z)/(1+z) \leqslant \Psi^2(z/2)$ for $z \in (0, 1/\varrho)$ and after a tedious but very routine calculation it can be shown that this is equivalent to $\varrho \geqslant \mathcal{P}(z)$, i.e.,

$$\frac{1-\varrho z}{1+z} \leqslant \Psi^J\left(\frac{z}{J}\right) \Leftarrow \frac{1-\varrho z}{1+z} \leqslant \Psi^2\left(\frac{z}{2}\right) \Leftrightarrow \varrho \geqslant \mathcal{P}(z) \quad \forall\, z \in (0, 1/\varrho), \tag{4.13}$$

where $\mathcal{P}(z) = (1-(1+z)\Psi^2(z/2))/z$. For $\theta = 1 - 1/\sqrt{2}$ and $z \in (0, 2\tilde{z}_0)$, we have plotted the profile of $\mathcal{P}(z)$ in Fig. 8 and it is clear to see that the global maximum of $\mathcal{P}(z)$ for $z \in (0, 2\tilde{z}_0)$ is 0.316. Hence, a first bound of ϱ, which ensures the inequality $(1-\varrho z)/(1+z) \leqslant \Psi^J(z/J)$ (for all $J \geqslant 2$), is

$$\varrho \geqslant \max\left\{\frac{1}{2\tilde{z}_0}, \max_{z\in(0,1/(2\tilde{z}_0))} \mathcal{P}(z)\right\} = 0.316. \tag{4.14}$$

Step II. *The proof of* $\Psi^J(z/J) \leqslant (1+\varrho z)/(1+z)$ *for* $\varrho \geqslant \Psi^2(\tilde{z}_+)$. We prove this inequality on two intervals: $z \in (0, J\tilde{z}_0)$ and $z \in [J\tilde{z}_0, +\infty)$. For $z \in (0, J\tilde{z}_0)$, we know from Lemma 3.11 that

$$\Psi^J\left(\frac{z}{J}\right) \leqslant \lim_{J\to+\infty} \Psi^J\left(\frac{z}{J}\right) = \lim_{J\to+\infty}\left(1 - \frac{z}{J}\mathcal{I}(J)\right)^J \quad \forall\, z \in (0, J\tilde{z}_0), \tag{4.15}$$

where $\mathcal{I}(J) = (1+(z/J)\theta^2)/(1+(z/J)\theta)^2$ and it satisfies $\lim_{J\to+\infty}\mathcal{I}(J) = 1$. Hence, we have

$$\Psi^J\left(\frac{z}{J}\right) \leqslant e^{-z} \quad \forall\, z \in (0, J\tilde{z}_0). \tag{4.16}$$

Since $e^{-z} < 1/(1+z)$ (for all $z > 0$), we have $\Psi^J(z/J) < 1/(1+z) < (1+\varrho z)/(1+z)$ for all $\varrho > 0$ and $z > 0$. It remains to consider the case $z \geqslant J\tilde{z}_0$. From Lemma 3.8, for even integers J ($\geqslant 2$) we know that the function $\Psi^J(z/J)$ has a unique local maximum $\Psi^J(\tilde{z}_+)$—which is also the global maximum for $z \in [J\tilde{z}_0, +\infty)$, located at $z = J\tilde{z}_+$. Hence, to prove $\Psi^J(z/J) \leqslant (1+\varrho z)/(1+z)$ for $z \geqslant J\tilde{z}_0$, it is sufficient to prove

$$\Psi^J(\tilde{z}_+) \leqslant \frac{1+\varrho z}{1+z} \quad \forall z > J\tilde{z}_0. \tag{4.17}$$

Since $(1+\varrho z)/(1+z) \geqslant \varrho$, $\lim_{z\to\infty}((1+\varrho z)/(1+z)) = \varrho$ and $\Psi^J(\tilde{z}_+)$ decreases as J increases (these properties are the same as that of the Parareal-TR/BDF2 algorithm and see Fig. 6 for illustration), it is sufficient to require $\varrho \geqslant \Psi^2(\tilde{z}_+)$ to ensure (4.17). For $\theta = 1 - 1/\sqrt{2}$, it is easy to get $\Psi^2(\tilde{z}_+) = 0.042893218813453$. This, together with (4.14), gives the lower bound of ϱ, $\varrho_{\min} = 0.316$, which ensures the two inequalities in the right-hand side of (4.12). □

REMARK 4.4 (Efficiency of the parareal algorithms) Assume that the computational cost of the implicit Euler method (i.e., the coarse propagator $\mathcal{G}$) is $\mathbb{T}_{\text{unit}}$ for one time step. Then, the computational cost of the fine propagator $\mathcal{F}$ can be denoted by $c \times \mathbb{T}_{\text{unit}}$ for one time step, where $c = 1$ if the implicit Euler and trapezoidal methods are used as the fine propagator, and $c = 2$ if the 2s-DIRK and TR/BDF2 methods are used as the fine propagator. With this notation, the total computational cost of the sequential computing by using the fine propagator $\mathcal{F}$ is

$$\mathbb{T}_{\text{total}} = N \times J \times c \times \mathbb{T}_{\text{unit}}. \tag{4.18}$$

Moreover, the computational cost for a single iteration of the parareal algorithm can be expressed as

$$\underbrace{N \times \mathbb{T}_{\text{unit}}}_{\text{sequential correction by } \mathcal{G}} + \underbrace{J \times c \times \mathbb{T}_{\text{unit}} + \mathbb{T}_{\text{unit}}}_{\text{parallel computing by } \mathcal{F} \text{ and } \mathcal{G}}. \tag{4.19}$$

The number of iterations required to achieve the given tolerance ϵ is $k \approx \log\epsilon/\log\rho^l(\Delta T, J)$, where $\rho^l(\Delta T, J)$ denotes the upper bound of the contraction factor of the parareal algorithm. Hence, the total computational cost of the parareal algorithm to achieve the desired tolerance is

$$\frac{\log\epsilon}{\log\rho^l(\Delta T, J)}(N + Jc + 1)\mathbb{T}_{\text{unit}}. \tag{4.20}$$

The parallelism efficiency of the parareal algorithm can therefore be expressed as

$$\mathbb{E}(N, \Delta T, J) := \frac{(\log\epsilon/\log\rho^l(\Delta T, J))(N + Jc + 1)\mathbb{T}_{\text{unit}}}{\mathbb{T}_{\text{total}}} = \frac{\log\epsilon}{\log\rho^l(\Delta T, J)}\left(\frac{1}{Jc} + \frac{Jc+1}{NJc}\right). \tag{4.21}$$

For the Parareal-Euler, Parareal-TR/BDF2 and Parareal-2s-DIRK algorithms (with $\gamma \in [0.043, 0.977]$), the argument ρ^l can be bounded by a value around 0.3, independent of J and ΔT; hence, the parallelism efficiency of these three parareal algorithms can be bounded by

$$\mathbb{E}(N, \Delta T, J) \leqslant \frac{-\log\epsilon}{1.2}\left(\frac{1}{Jc} + \frac{Jc+1}{NJc}\right) \leqslant \left(\frac{1}{Jc} + \frac{Jc+1}{NJc}\right)\log\frac{1}{\epsilon}. \tag{4.22}$$

This implies that increasing J, which improves the accuracy of the converged solution, or increasing the number of paralleled coarse time-slices, does not cause deterioration of the parallelism efficiency of the

Parareal-Euler, Parareal-TR/BDF2 and Parareal-2s-DIRK algorithms. The property of the parallelism efficiency of the Parareal-TR algorithm is, however, not clear at the moment, because $\rho^l \approx 1$ when the minimal eigenvalue of the coefficient matrix A is sufficiently small.

5. Numerical results

In this section, we perform several numerical experiments to test the three second-order parareal algorithms studied in this paper. To this end, we consider the two-sided fractional diffusion equations (FDEs)

$$\begin{cases} \dfrac{\partial u(x,t)}{\partial t} = d(x)\left(\dfrac{\partial^\alpha u(x,t)}{\partial_+ x^\alpha} + \dfrac{\partial^\alpha u(x,t)}{\partial_- x^\alpha}\right) + tx\sin(t), & (x,t) \in (0,1)\times(0,30), \\ u(0,t) = u(1,t) = 0, & t \in (0,30), \\ u(x,0) = 0, & x \in [0,1], \end{cases} \tag{5.1}$$

where $d(x) = 2 + x + \cos(3\pi x)$, $\alpha \in (1,2)$ is the fractional order, $\partial^\alpha u(x,t)/\partial_+ x^\alpha$ (respectively, $\partial^\alpha u(x,t)/\partial_- x^\alpha$) denotes the left-sided (respectively, right-sided) fractional derivative and is defined in the Grünwald–Letnikov sense (Podlubny, 1999):

$$\frac{\partial^\alpha u(x,t)}{\partial_+ x^\alpha} = \frac{1}{\Gamma(2-\alpha)}\frac{\partial^2}{\partial x^2}\int_0^x \frac{u(s,t)}{(x-s)^{\alpha-1}}\,\mathrm{d}s, \quad \frac{\partial^\alpha u(x,t)}{\partial_- x^\alpha} = \frac{1}{\Gamma(2-\alpha)}\frac{\partial^2}{\partial x^2}\int_x^1 \frac{u(s,t)}{(s-x)^{\alpha-1}}\,\mathrm{d}s, \tag{5.2}$$

where Γ is the Gamma function. There is a critical difference between the behaviour of an integer-order derivative and a fractional one, which is that the latter is a nonlocal operator, and because of this, solving fractional PDEs is much more time consuming.

To solve (5.1) and (5.2) numerically, we first use the so-called shifted Grünwald approximations to discretize the two fractional derivatives:

$$\frac{\partial^\alpha u(x_m,t)}{\partial_+ x^\alpha} \approx \frac{1}{(\Delta x)^\alpha}\sum_{j=0}^{m+1} g_j^{(\alpha)} u(x_{m-(j-1)},t), \quad \frac{\partial^\alpha u(x,t)}{\partial_- x^\alpha} \approx \frac{1}{(\Delta x)^\alpha}\sum_{j=0}^{M-m+1} g_j^{(\alpha)} u(x_{m+(j-1)},t), \tag{5.3a}$$

where M is an integer, $\Delta x = L/M$ denotes the spatial discretization parameter, $x_m = m\Delta x$ for $m = 0,1,\ldots,M$ and $g_j^{(\alpha)}$ is defined by $g_j^{(\alpha)} = (-1)^j\binom{\alpha}{j}$, and can be evaluated recursively as

$$g_0^{(\alpha)} = 1,\ g_j^{(\alpha)} = \left(1 - \frac{\alpha+1}{j}\right) g_{j-1}^{(\alpha)}, \quad j = 1,2,\ldots. \tag{5.3b}$$

Note that, for $\alpha = 2$, (5.3a) is reduced to the centered finite difference scheme, which is a classical approximation of $\partial_{xx}u$. After space discretization, we arrive at the following system of ODEs:

$$\frac{\mathrm{d}u_m(t)}{\mathrm{d}t} = \frac{d_m}{(\Delta x)^\alpha}\left(\sum_{j=0}^{m+1} g_j^{(\alpha)} u_{m-j+1}(t) + \sum_{j=0}^{M-m+1} g_j^{(\alpha)} u_{m-j+1}(t)\right) + s_m(t), \tag{5.4}$$

where $d_m = d(x_m)$, $s_m(t) = tx_m\sin(t)$, $u_m(t) \approx u(x_m,t)$, $x_m = m\Delta x$ and $m = 0,1,\ldots,M$. Define

$$D = \mathrm{diag}(d_1,d_2,\ldots,d_{M-1}), \quad \mathbf{u}(t) = (u_1(t),u_2(t),\ldots,u_{M-1}(t))^{\mathrm{T}}, \quad \mathbf{s}(t) = (s_1(t),s_2(t),\ldots,s_{M-1}(t))^{\mathrm{T}}, \tag{5.5}$$

and then the system of ODEs (5.4) can be rewritten as

$$\begin{cases} \dfrac{\mathrm{d}\mathbf{u}(t)}{\mathrm{d}t} = A\mathbf{u}(t) + \mathbf{g}(t), & t > 0, \\ \mathbf{u}(0) = 0, & t = 0, \end{cases} \tag{5.6}$$

where

$$A = \frac{1}{(\Delta x)^{\alpha}}(DG_{\alpha} + (DG_{\alpha})^{\mathrm{T}}), \quad G_{\alpha} = \begin{pmatrix} g_1^{(\alpha)} & g_0^{(\alpha)} & 0 & \cdots & 0 & 0 \\ g_2^{(\alpha)} & g_1^{(\alpha)} & g_0^{(\alpha)} & 0 & \cdots & 0 \\ \vdots & g_2^{(\alpha)} & g_1^{(\alpha)} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ g_{M-2}^{(\alpha)} & \ddots & \ddots & \ddots & g_1^{(\alpha)} & g_0^{(\alpha)} \\ g_{M-1}^{(\alpha)} & g_{M-2}^{(\alpha)} & \cdots & \cdots & g_2^{(\alpha)} & g_1^{(\alpha)} \end{pmatrix}. \tag{5.7}$$

Clearly, the coefficient matrix A is symmetric. Since $d(x) \geqslant d_{\min} > 0$, from the work by Podlubny (1999), Meerschaert *et al.* (2006) and Meerschaert & Tadjeran (2006) we know that (1) the matrix A is negative definite; (2) numerical schemes based on (5.6) and (5.7) and implicit time integrators are unconditionally stable.

In our numerical experiments, we use several fractional orders α for FDEs (5.1). The initial iterate for the parareal algorithm is chosen randomly and we stop the iteration process when the error between the iterate and the target solution is less than $1.0e \times 10^{-8}$, i.e.,

$$\max_{N \geqslant n \geqslant 1} \|\mathbf{u}_n^k - \mathbf{u}_n\|_{\infty} \leqslant 1.0e \times 10^{-8}. \tag{5.8}$$

EXAMPLE 5.1 In this first set of numerical results, we compare the convergence rates of the four parareal algorithms: Parareal-Euler, Parareal-TR/BDF2 (with magic parameter $\gamma = 2 - \sqrt{2}$), Parareal-TR and Parareal-2s-DIRK. By choosing $\Delta x = 0.01$, $J = 20$ and $\Delta T = 0.1$ for the discretization parameters, we have plotted in Fig. 9 the measured error at each iteration, under two choices of the fractional order: $\alpha = 1.25$ (left) and $\alpha = 1.8$ (right). We see that with a small fractional order α the convergence rates of the four parareal algorithms are equal, while for large α the performance of the Parareal-TR algorithm is worse compared with the other three algorithms. This observation can be explained by examining the distribution of eigenvalues of the coefficient matrix A: $\lambda \in [-5.434e \times 10^3, -2.4921]$ for $\alpha = 1.25$ and $\lambda \in [-1.002 \times 10^5, -24.0589]$ for $\alpha = 1.8$. In Fig. 10, we have plotted the contraction factor $\mathcal{K}_{\mathrm{TR}}^{l}$ with $J = 20$ and from the left panel we can see that, for $\alpha = 1.25$, the contraction factor can be bounded by 0.3 for $\Delta T\lambda \in [-5.434e \times 10^2, -0.24921]$, which implies that the Parareal-TR algorithm has the same convergence rate as the other three algorithms. However, for $\alpha = 1.8$ we see from the right panel that the contraction factor can reach 0.85 and therefore in this case the convergence rate of the Parareal-TR algorithm is worse.

We next consider the asymptotic dependence of the four parareal algorithms on the mesh parameters, Δx, ΔT and J. Since the contraction factors of Parareal-Euler, Parareal-TR/BDF2 and Parareal-2s-DIRK can be bounded uniformly by a value around 0.3, we can imagine that these three algorithms are insensitive to the change of mesh parameters. We illustrate this theoretical prediction on the left-hand side of Fig. 11 by plotting the number of iterations needed to satisfy the termination criterion (5.8). We see that the Parareal-Euler, Parareal-TR/BDF2 and Parareal-2s-DIRK algorithms are really

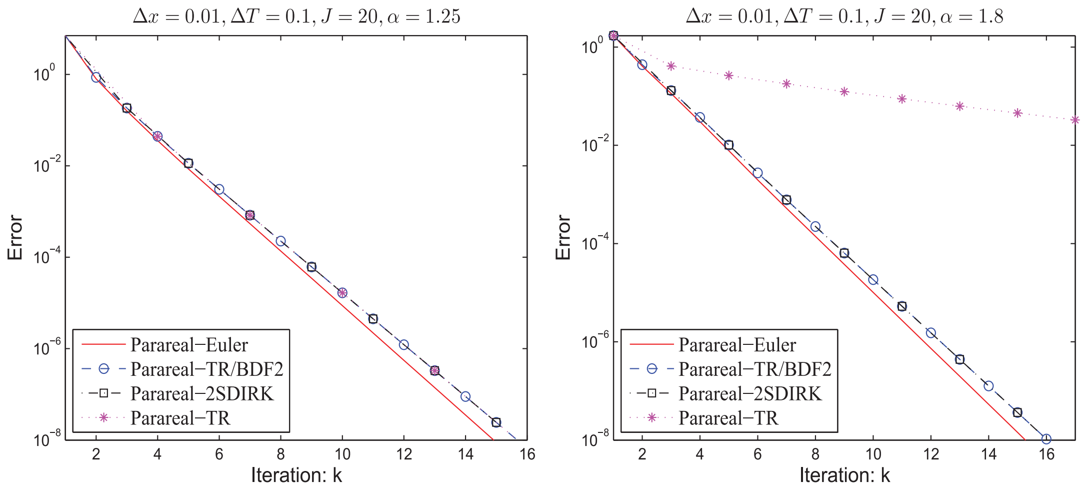

FIG. 9. Measured error of the four parareal algorithms at each iteration: $\alpha = 1.25$ (left) and $\alpha = 1.8$ (right).

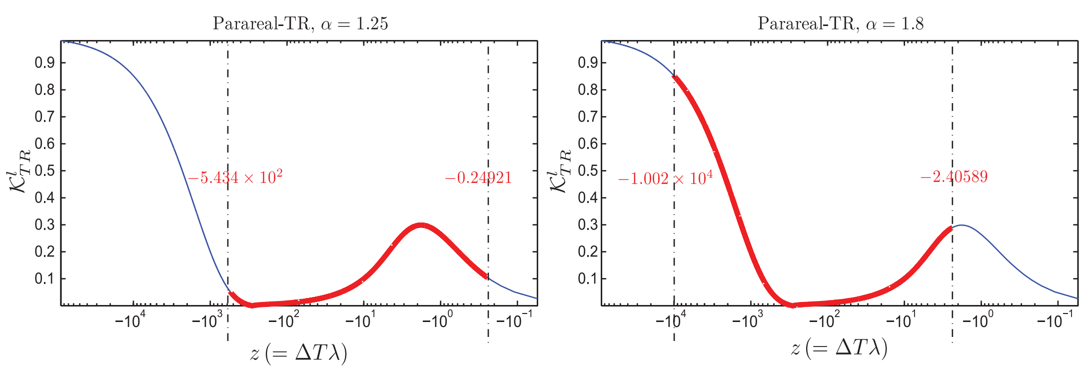

FIG. 10. Contraction factor $\mathcal{K}^l_{\mathrm{TR}}$ of the Parareal-TR algorithm when $J = 20$. Left: $\alpha = 1.25$; right: $\alpha = 1.8$.

insensitive to the change of mesh parameters. For the Parareal-TR algorithm, from the subfigures on the left we see that (1) the converge rate deteriorates rapidly as we refine Δx, or increase (respectively, decrease) the argument ΔT (respectively, J); (2) for Δx, J large (or ΔT small), it converges as fast as the other three algorithms. The explanation for this observation is the same as we have described for Fig. 10: for Δx, J small (or ΔT large) we have $\Delta T\lambda_{\min} \to -\infty$ and therefore $\mathcal{K}^l_{\mathrm{TR}}(z, J) \to 1$, while otherwise the contraction factor $\mathcal{K}^l_{\mathrm{TR}}(z, J)$ can be bounded from above by a constant around 0.3 for $z = \Delta T\lambda \in [\Delta T\lambda_{\min}, \Delta T\lambda_{\max}]$, where $\lambda_{\min}$ (respectively, $\lambda_{\max}$) is the minimal (respectively, maximal) eigenvalue of the matrix A.

On the right of Fig. 11, we show the efficiency factor $\mathbb{E}(N, \Delta T, J)$ (defined by (4.21)) for the four parareal algorithms, where the argument $\rho^l(\Delta T, J)$ is defined by the maximum of the contraction factor $\mathcal{K}^l$ over the interval $[\Delta T\lambda_{\min}, \Delta T\lambda_{\max}]$, i.e., $\rho^l(\Delta T, J) = \max_{z\in[\Delta T\lambda_{\min},\, \Delta T\lambda_{\max}]} \mathcal{K}^l(z, J)$. We see that the behaviour of $\mathbb{E}(N, \Delta T, J)$, for each parareal algorithm, is similar to that of the iteration number shown

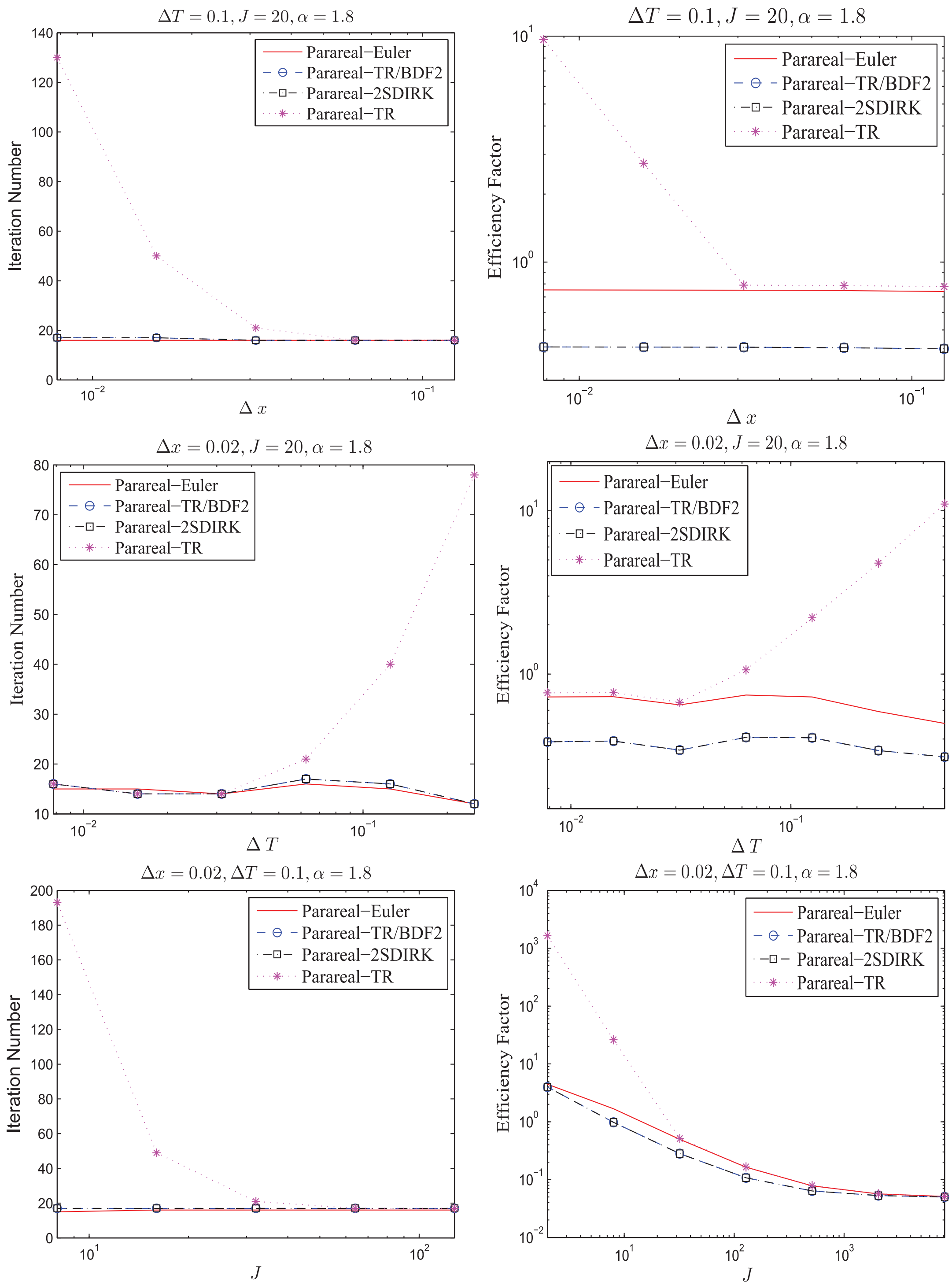

FIG. 11. Asymptotic behaviour of the four parareal algorithms with respect to Δx (top row), ΔT (middle row) and the multiplication ratio J (bottom row). Left: iteration number; right: efficiency factor $\mathbb{E}(N, \Delta T, J)$ defined by (4.21).

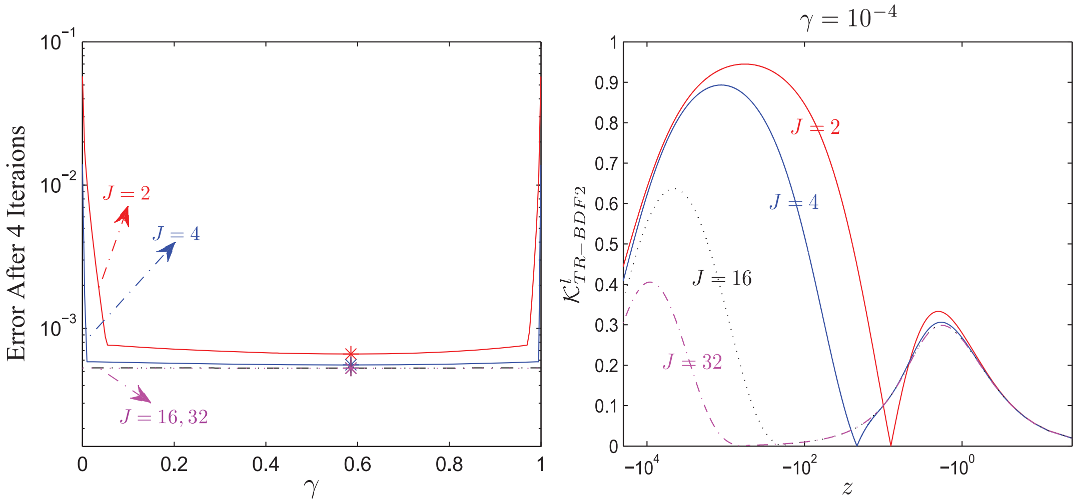

FIG. 12. Left: the maximal error $\max_{N\geqslant n\geqslant 1}\|\mathbf{u}_n^k - \mathbf{u}_n\|_\infty$ obtained by running the Parareal-TR/BDF2 algorithm for four iterations and various choices of the time integration parameter γ, with the magic choice $\gamma = 2-\sqrt{2}$ indicated by a star. Right: the profile of the contraction factor $\mathcal{K}^l_{\text{TR-BDF2}}$ with $\gamma = 10^{-4}$ and four different choices of J.

on the left. In particular, for the Parareal-Euler, Parareal-TR/BDF2 and Parareal-2s-DIRK algorithms, the parallelism efficiency is independent of Δx and does not deteriorate as we increase the multiplication ratio J (or refine the coarse step size ΔT). Moreover, we see that the parallelism efficiency of the Parareal-TR/BDF2 and Parareal-2s-DIRK algorithms is better than that of the Parareal-Euler algorithm; this is because the TR/BDF2 and 2s-DIRK methods are twice as expensive as the implicit Euler method.

EXAMPLE 5.2 For the TR/BDF2 method, a free parameter $\gamma \in (0,1)$ is involved and this parameter is therefore involved in the Parareal-TR/BDF2 algorithm. We remarked in Remark 4.2 that the influence of this parameter is negligible when $\gamma \in [0.043, 0.977]$. Here, we present numerical results to support this theoretical prediction. Let $\Delta x = \frac{1}{40}$ and $\Delta T = \frac{1}{4}$. Then, in the left panel of Fig. 12, we plot the maximal error $\max_{N\geqslant n\geqslant 1}\|\mathbf{u}_n^k - \mathbf{u}_n\|_\infty$ after running the parareal algorithm for four iterations, using various values for the free parameter γ; the *magic* choice $\gamma = 2-\sqrt{2}$ is indicated by a star. Two pieces of information can be obtained from this panel: (a) the influence of γ is slight for J large; (b) for J small (for example $J=2,4$), the influence of γ coincides with the information given by Fig. 7. An explanation for this is illustrated in Fig. 12 on the right: for J small, the contraction factor $\mathcal{K}^l_{\text{TR-BDF2}}$ rapidly approaches a value near 1 as z varies from 0 to $-\infty$, while for J large, $\mathcal{K}^l_{\text{TR-BDF2}}$ maintains a small value over a large interval, say $\mathcal{K}^l_{\text{TR-BDF2}}(z,\gamma,J) \leqslant 0.35$ for $z = \Delta T\lambda \in [-2000, 0]$. Hence, it is natural to see in a numerical computation that the influence of γ on the convergence rate is negligible when a large multiplication ratio J is used.

6. Conclusion

We have studied the convergence properties of three second-order parareal algorithms for large-scale symmetric negative-definite systems of ODEs. For an even multiplication ratio J, we prove that the error between the iterate and the target solution is contractive at each iteration. Moreover, we prove that the contraction factors of the Parareal-2s-DIRK algorithm and the Parareal-TR/BDF2 algorithm (with parameter γ satisfying $\gamma \in [0.043, 0.977]$) can be bounded by 0.316 and $\frac{1}{3}$, respectively. The bounds are independent of the discretization parameters (J, ΔT and Δx). For the Parareal-TR algorithm, a sharp bound (less than 1) for the contraction factor does not exist.

Acknowledgements

The authors are very grateful to the anonymous referees for their careful reading of a preliminary version of the manuscript and their valuable suggestions and comments, which greatly improved the quality of this paper.

Funding

This work was supported by the NSF of China (11301362, 11371157, 91130003) and the project of the Key Laboratory of Cambridge and Non-destructive Inspection of Sichuan Institutes of Higher Education (2013QZY01).

References

Baffico, L., Bernard, S., Maday, Y., Turinici, G. & Zérah, G. (2002) Parallel-in-time molecular-dynamics simulations. *Phys. Rev. E*, **66**, 057701.

Bank, R. E., Coughran Jr. W. M., Fichtner, W., Grosse, E. H., Rose, D. J. & Smith, R. K. (1985a) Transient simulation of silicon devices and circuits. *IEEE Trans. Comput. Aided Des.*, **4**, 436–451.

Bank, R. E., Coughran Jr. W. M., Fichtner, W., Grosse, E. H., Rose, D. J. & Smith, R. K. (1985b) Transient simulation of silicon devices and circuits. *IEEE Trans. Electron. Devices*, **32**, 1992–2007.

Cortial, J. & Farhat C. (2008) A time-parallel implicit method for accelerating the solution of nonlinear structural dynamics problems. *Int. J. Numer. Meth. Eng.*, **77**, 451–470.

Dai, X., Le Bris, C., Legoll, F. & Maday Y. (2013) Symmetric parareal algorithms for Hamiltonian systems. *M2AN Math. Model Numer. Anal.*, **47**, 717–742.

Dai, X. & Maday Y. (2013) Stable parareal in time method for first- and second-order hyperbolic systems. *SIAM J. Sci. Comput.*, **35**, A52–A78.

Dharmaraja, S., Wang, Y. & Strang, G. (2010) Optimal stability for trapezoidal-backward difference split-steps. *IMA J. Numer. Anal.*, **30**, 141–148.

Farhat, C. & Chandesris, M. (2003) Time-decomposed parallel time-integrators: theory and feasibility studies for fluid, structure, and fluid-structure applications. *Int. J. Numer. Meth. Eng.*, **58**, 1397–1434.

Gander, M. J. & Hairer, E. (2014) Analysis for parareal algorithms applied to Hamiltonian differential equations. *J. Comput. Appl. Math.*, **259**, 2–13.

Gander, M. & Petcu, M. (2008) Analysis of a Krylov subspace enhanced parareal algorithm for linear problems. *ESAIM Proc.*, **25**, 114–129.

Gander, M. J. & Vandewalle S. (2007) Analysis of the parareal time-parallel time-integration method. *SIAM J. Sci. Comput.*, **29**, 556–578.

Hairer, E. & Wanner, G. (2002) *Solving Ordinary Differential Equations II*, 2nd revised ed., Springer Series in Computational Mathematics. New York: Springer, 2002.

He, L. & He, M. (2012) Parareal in time simulation of morphological transformation in cubic alloys with spatially dependent composition. *Commun. Comput. Phys.*, **11**, 1697–1717.

Hosea, M. E. & Shampine L. F. (1996) Analysis and implementation of TR-BDF2. *Appl. Numer. Math.*, **30**, 21–37.

Lions, J.-L., Maday, Y. & Turinici, G. (2001) A 'parareal' in time discretization of PDE's. *C. R. Acad. Sci. Paris Sér. I Math.*, **332**, 661–668.

Maday, Y., Salomon, J. & Turinici, G. (2007) Monotonic parareal control for quantum systems. *SIAM J. Numer. Anal.*, **45**, 2468–2482.

Mathew, T. R., Sarkis, M. & Schaerer, C. E. (2010) Analysis of block parareal preconditioners for parabolic optimal control problems. *SIAM J. Sci. Comput.*, **32**, 1180–1200.

Meerschaert, M. M., Scheffler, H. P. & Tadjeran, C. (2006) Finite difference methods for two-dimensional fractional dispersion equation. *J. Comput. Phys.*, **211**, 249–261.

Meerschaert, M. M. & Tadjeran, C. (2006) Finite difference approximations for two-sided space-fractional partial differential equations. *Appl. Numer. Math.*, **56**, 80–90.

Minion, M. L. (2010) A hybrid parareal spectral deferred corrections method. *Appl. Math. Comput. Sci.*, **5**, 265–301.

Pearson, J. W., Stoll, M. & Wathen, A. J. (2012) Regularization-robust preconditioners for time-depdendent PDE-constrained optimization problems. *SIAM J. Matrix Anal. Appl.*, **33**, 1126–1152.

Podlubny, I. (1999) *Fractional Differential Equations*. New York: Academic Press, 1999.

Reynolds-Barredo, J. M., Newman, D. E., Sanchez, R., Samaddar, D. Berry, L. A. & Elwasif, W. R.(2012) Mechanisms for the convergence of time-parallelized, parareal turbulent plasma simulations. *J. Comput. Phys.*, **231**, 7851–7867.

Ruprecht, D. & Krause, R. (2012) Explicit parallel-in-time integration of a linear acoustic-advection system. *Comput. Fluids*, **59**, 72–83.

Wu, S. L., Shi, B. C. & Huang, C. M. (2009) Parareal-Richardson algorithm for solving nonlinear ODEs and PDEs. *Commun. Comput. Phys.*, **6**, 883–902.

Xiu, D. & Karniadakis, G. E. (2002) The Wiener–Askey polynomial chaos for stochastic differential equations. *SIAM J. Sci. Comput.*, **24**, 619–644.

Xiu, D. & Shen, J. (2009) Efficient stochastic Galerkin methods for random diffusion equations. *J. Comput. Phys.*, **228**, 266–281.

Zhuang, P., Liu, F., Anh, V. & Turner, I. (2009) Numerical methods for the variable-order fractional advection-diffusion equation with a nonlinear source term. *SIAM J. Numer. Anal.*, **47**, 1760–1781.