SIAM J. SCI. COMPUT.
Vol. 41, No. 6, pp. A3476–A3497

PARAREAL CONVERGENCE FOR OSCILLATORY PDEs WITH FINITE TIME-SCALE SEPARATION*

ADAM G. PEDDLE†, TERRY HAUT‡, AND BETH WINGATE†

Abstract. In [*SIAM J. Sci. Comput.*, 36 (2014), pp. A693–A713] the authors present a new coarse propagator for the parareal method applied to oscillatory PDEs that exhibit time-scale separation and show, under certain regularity constraints, superlinear convergence which leads to significant parallel speedups over standard parareal methods. The error bound depends on the degree of time-scale separation, ϵ, and the coarse time step, ΔT, and relies on a bound that holds only in the limit of small ϵ. The main result of the present paper is a generalization of this error bound that also holds for finite values of ϵ, which can be important for applications in the absence of scale separation. The new error bound is found to depend on an additional parameter, η, the averaging window used in the nonlinear term of the coarse propagator. The new proof gives insight into how the parareal method can converge even for finite values of ϵ. It is also a significant technical advance over the proof presented in [*SIAM J. Sci. Comput.*, 36 (2014), pp. A693–A713]; it requires the introduction of a stiffness regulator function that allows us to control the oscillatory stiffness in the nonlinear term. The new convergence concepts developed in the new proof are confirmed using numerical simulations.

Key words. parareal, parallel in time, wave averaging, oscillatory stiffness, time-scale separation

AMS subject classifications. 65M12, 65D30, 65B99, 65L04, 68W10, 34E10

DOI. 10.1137/17M1131611

1. Introduction.

In this paper we are interested in the convergence of the asymptotic parallel-in-time (APinT) [14] parareal method [21, 22] for oscillatory systems of equations with the following form:

$$\frac{d\mathbf{u}}{dt} + \frac{1}{\varepsilon} L\mathbf{u} + \mathcal{N}(\mathbf{u}) = 0, \tag{1.1}$$

$$\mathbf{u}(t)|_{t=0} = \mathbf{u}^0. \tag{1.2}$$

Here, $\mathbf{u}$ is the vector of unknowns, L is a skew-Hermitian matrix with purely imaginary eigenvalues, and $\mathcal{N}(\cdot)$ is a nonlinear operator. The linear term induces temporal oscillations on an $\mathcal{O}(\varepsilon)$ time scale, which can require the use of prohibitively small time steps for standard numerical integrators if ε is small and if the temporal oscillations of $\mathbf{u}(t)$ are significant (e.g., $L\mathbf{u}(t)$ is not small). In addition, convergence of the parareal method for such highly oscillatory problems can require a time step that scales like $\mathcal{O}(\varepsilon)$ [12]. Our motivation for studying (1.1) comes from the development of efficient time-stepping schemes for solving spatially discretized PDEs that arise in geophysical fluid applications (see [19]), where it is important that the time step ΔT can be chosen on a time scale that is independent of ε. The analysis in the current paper assumes that PDEs can be written as a system of ODEs as in (1.1).

*Submitted to the journal's Methods and Algorithms for Scientific Computing section May 23, 2017; accepted for publication (in revised form) July 5, 2019; published electronically November 12, 2019.

https://doi.org/10.1137/17M1131611

Funding: This work was supported by the Engineering and Physical Sciences Research Council (EPSRC) through grant EP/R029628/1, the University of Exeter College of Engineering, Mathematics and Physical Sciences, Los Alamos National Laboratory LDRD funding, and the Lawrence Livermore National Laboratory.

†CEMPS, University of Exeter, Exeter EX4 4QF, UK (ap553@exeter.ac.uk, B.Wingate@exeter.ac.uk).

‡Lawrence Livermore National Laboratory, Livermore, CA 94550 (haut3@llnl.com).

In order to understand how the concepts in this paper differ from those proposed in [14], we will first sketch the mathematical ideas of the coarse propagator, pointing readers to sections in the paper where more details will be described later. We also assume the reader has a basic knowledge of the parareal method, but we provide a short section 2 where the basic concepts and notation are described.

The coarse solver studied in this paper is based on two key steps. The first step is to approximate (1.1) by using a coordinate transformation,

$$\mathbf{u}(t) \approx e^{-\frac{1}{\epsilon}Lt}\ \mathbf{v}(t), \tag{1.3}$$

$$\frac{d\mathbf{v}}{dt}(t) + e^{\frac{1}{\epsilon}Lt}\mathcal{N}(e^{-\frac{1}{\epsilon}Lt}\mathbf{v}(t)) = 0, \tag{1.4}$$

$$\mathbf{v}(t)|_{t=0} = \mathbf{u}^0, \tag{1.5}$$

which removes the stiff linear term. However, taking another time derivative of (1.4) shows that there are still some oscillations remaining. Therefore, to regularize (1.4) further, a time average is applied. This produces the ϵ-dependent equation that is used as an approximate coarse solver in the APinT parareal method,

$$\mathbf{u}(t) \approx e^{-\frac{1}{\epsilon}Lt}\ \overline{\mathbf{v}}(t), \tag{1.6}$$

$$\frac{d\overline{\mathbf{v}}}{dt}(t) + \frac{1}{\eta}\int_0^\eta\ e^{\frac{1}{\epsilon}Ls}\mathcal{N}(e^{-\frac{1}{\epsilon}Ls}\overline{\mathbf{v}}(t))ds = 0, \tag{1.7}$$

$$\overline{\mathbf{v}}(t)|_{t=0} = \mathbf{u}^0. \tag{1.8}$$

The averaging in (1.7) introduces a new parameter, η, whose effect on the convergence of the method must be included to understand the convergence of the APinT method. Also notice that the integral is over s, a parameter that appears only in the matrix exponentials. The approximation (1.6)–(1.7) is based on an analysis of PDEs [6, 18, 26] which provides an error bound in the limit of $\epsilon \to 0$ and will be discussed in more depth in section 3.1. The second step in constructing the coarse solver is the numerical approximation of (1.7) which introduces another key parameter, the coarse time step, ΔT.

To summarize, the key parameters of the error bound for APinT are (1) the degree of time-scale separation, ϵ, (2) the coarse time step, ΔT, and (3) the averaging window, η. We are now in a position to discuss the assumptions made in [14] and the significance of the approach taken in this paper to extend the range of applicability of the error bound. Readers who wish to understand more details before examining the main concepts can find a short discussion of the parareal method in section 2 and more details about how we treat the coarse solver mathematically in section 1.2, with the details of the proof finally given in section 4.

We now discuss one of the assumptions used in the error estimate for the convergence of APinT in [14]. In particular, the error bound is a function of ϵ and ΔT, with no dependency on the averaging window, η. This is because the authors used an important relationship between the time-scale separation, ϵ, and the averaging window, η, based on analysis in [6, 18, 26, 8] which holds only in the limit of $\epsilon \to 0$, and which will be discussed in more detail in section 3.1. For the purposes of the introduction we describe this relationship in words: in the limit as $\epsilon \to 0$ and $\eta \to \infty$, the error between the solution of (1.1) and its approximation (1.6)–(1.7) is $o(1)$. This approximation is valid as long as ϵ is small enough to be considered within the asymptotic limit.

However, developing error bounds for finite ϵ, which range from $\epsilon \ll 1$ to $\epsilon = \mathcal{O}(1)$, is important for physical applications. In fact, some physical phenomena, such

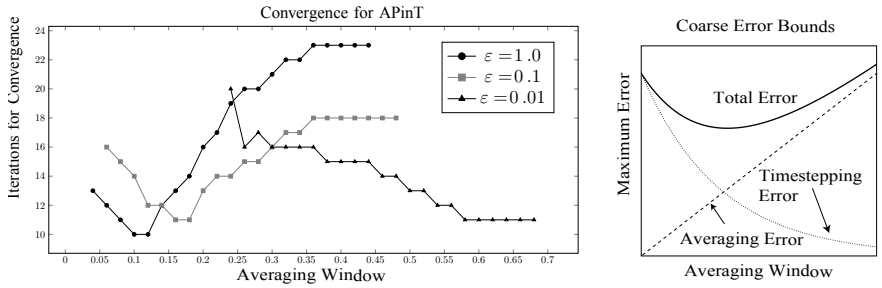

FIG. 1.1. Left: *The number of iterations required for convergence of the APinT method for the two-dimensional rotating shallow water equations across three values of time-scale separation, ε (cf. section* 1*). Note that towards the small-ε limit, shown by the line represented by triangles, the convergence improves with an increase in the size of the averaging window, as is consistent with the asymptotic theory. Note, however, that a clear minimum is visible outside of this limit, especially for $\varepsilon = 1$, which marks a departure from the asymptotic theory. This makes clear both that the convergence of the method depends on the degree of scale separation, ε, and that the width of the averaging window, shown here proportional to the coarse time step, may be chosen to control it. We shall rigorously explain this in section* 4. Right: *A graph of the error bounds found in section* 4. *The wide dash represents the error estimate due to averaging the nonlinearity. The dotted line represents the error from the time-stepping. The solid line represents their sum, which is the total error using the APinT coarse propagator. This can be compared to the graph on the left, which shows a minimum in the convergence of parareal using the APinT coarse propagator.*

as those occurring in numerical weather prediction, have finite frequencies inherent in the problem; e.g., Earth's rotation rate is finite. Because of this issue, some numerical weather prediction codes no longer use the simpler reduced equations derived from small nondimensional parameters, choosing to use the full set of equations even though they contain "fast" motions [7]. Furthermore, even if the value of ϵ is small enough to be considered to be within the asymptotic limit, it may not remain so as the simulation unfolds. This is because small nondimensional parameters are constructed using characteristic velocity, length, and time scales that change as solutions evolve. As an example of this behavior for a case relevant to the numerical simulations in section 5, see [24], where the authors discuss how regimes which start out with small Froude number, a small parameter related to oscillations due to the force of gravitation, grow to O(1) as the flow adjusts with time.

We now demonstrate the effect of η and ϵ on the convergence of parareal through numerical evidence for APinT solutions of the rotating shallow water equations on the left side of Figure 1.1. The figure's vertical axis shows the total number of iterations required for the APinT method to converge to a specific error tolerance for three different orders of magnitude of ϵ. When ϵ is small, the oscillations are fast, and when ϵ is large, the oscillations are slower. The horizontal axis represents a time-averaging interval, η. This averaging interval is an important part of regularizing the effects of oscillations on the nonlinearity and will be discussed in detail in section 3. For now η can be interpreted as the degree to which the oscillations have been smoothed. That is, when $\eta = 0$, there is no averaging and all the oscillations are present in the solution. When $\eta \to \infty$, this is the asymptotic case when $\epsilon \to 0$ and the nonlinear effects of oscillations on the solution have been averaged away. This asymptotic limiting behavior is demonstrated by the case when $\epsilon = .01$ (represented by ▲ on the graph). In this case, as η increases, the number of iterations not only decreases, but approaches a minimal plateau. Beyond this point, increasing the averaging window

has no effect on convergence because the effect of the oscillations has already been averaged out. For the other two cases, for $\epsilon = .1$ (■ on the graph) and $\epsilon = 1$ (• on the graph), it can be seen that as η increases, the number of parareal iterations to converge decreases until it reaches a minimum. Rather than reaching a plateau, as in the asymptotic case, we see that the number of iterations to converge increases. This indicates there is a value of η for which there is at least one minimum.

It is the goal of this paper to provide the understanding required to explain the minimum in Figure 1.1 found for finite values of ϵ by deriving a new error estimate for the coarse propagator. This result extends the work of [14], which was valid for the case when $\epsilon \to 0$. The proof differs significantly from that of [14] because the error bounds connecting the coarse time step ΔT, the time-averaging window η, and the parameter ϵ are valid even in the absence of scale separation (see sections 3 and 4).

The remainder of this introduction is separated into two subsections. In subsection 1.1 we will discuss related efforts in the parareal literature to solve problems of type (1.1). Then in subsection 1.2 we will sketch the key numerical concepts, unique to the APinT coarse propagator studied in this paper, that will be used to derive the new error estimate. The latter subsection will act as a guide to understanding the plan of the paper.

1.1. Related work. Examples of applications of the parareal algorithm being applied to parabolic PDEs include simulations of financial markets (i.e., the Black–Scholes equation for an American put [4]) and a nonlinear parabolic evolution equation via the finite element method [16]. Hyperbolic systems solved with parareal include simulation of molecular dynamics [3], fluid/structure interaction [9], solution of the Navier–Stokes equations [10], and reservoir modelling [13]. In all of these applications, the degree of oscillatory stiffness was not sufficient to impede convergence, but oscillatory stiffness is known (cf. [12, 1]) to be an important issue standing in the way of fast convergence rates for the parareal method.

There have been several modifications to the parareal method which apply to highly oscillatory systems and which assume that a system may be separated into fast and slow variables. In terms of ODEs, the authors of [20] have proposed a multiscale method for singularly perturbed ODEs where the fast dynamics are dissipative. In 2016, Ariel, Kim, and Tsai [1] proposed a method for highly oscillatory ODEs which is multiscale in nature but does not require explicit knowledge of the fast and slow variables. In 2014, Gander and Hairer [11] suggested parareal methods for Hamiltonian dynamics. Approaches using symplectic integrators with applications to molecular dynamics are presented in, for example, [2] and [5]. The method of [14], also for oscillatory PDEs, is motivated by the analysis of fast singular limits of nonlinear PDEs. It is the coarse propagator of [14] which we examine closely in this paper and discuss conceptually in the next section. Finally, we discuss the relationship of this work to the paper of Gander and Vandewalle [12], which studies the convergence of parareal and gives examples for two classic linear PDEs: linear advection, which has imaginary eigenvalues (oscillatory), and the heat equation, which has negative real eigenvalues (dissipative). Relevant to this paper is the case of linear advection, which is a much simpler PDE than the one studied in this paper, for which it is shown that even if the parareal method converges, it may not do so quickly. The key aspects of this paper that differentiate it from the linear advection case studied in [12] is that first, we focus on nonlinear problems, and second, we assume we can rotate the solution into the space of the oscillations through a coordinate transformation described in section 1.2. This coordinate transformation, through the matrix exponential, is a challenging

problem in its own right. An example of a parallelizable method of computing this operator in the context of the method proposed in this paper was discussed in [15].

1.2. Key concepts of the coarse propagator. The goal of this paper is to provide a new understanding of the convergence of the APinT parareal method for finite ϵ, depicted on the left side of Figure 1.1, through a new analysis of the coarse propagator. To that end, we sketch the key concepts required to identify the two main sources of error in the coarse propagator. At the end of this section, we will use these concepts to outline the plan of the paper.

First we introduce a coordinate transformation

$$\mathbf{u} = e^{-Lt/\epsilon}\mathbf{v}. \tag{1.9}$$

Then, defining

$$\mathcal{N}(s, \mathbf{x}) = e^{sL}\mathcal{N}\left(e^{-sL}\mathbf{x}\right), \tag{1.10}$$

equation (1.1) can be written as

$$\frac{d\mathbf{v}}{dt}(t) = \mathcal{N}\left(\frac{t}{\varepsilon}, \mathbf{v}(t)\right). \tag{1.11}$$

Comparing (1.9)–(1.11) with (1.1) shows that the mapping (1.9) removes the stiff linear term from the evolution equation. However, though the stiff linear term is gone, taking another derivative of (1.11) shows there will still be oscillations present. To make further progress, the second step in the construction of the coarse propagator is to introduce a "fast-wave-average" into the evolution equation described by (1.11),

$$\frac{d\overline{\mathbf{v}}}{dt}(t) = \overline{\mathcal{N}}_\eta\left(\frac{t}{\varepsilon}, \overline{\mathbf{v}}(t)\right). \tag{1.12}$$

Here the operator $\overline{\mathcal{N}}_\eta(t/\varepsilon, \overline{\mathbf{v}}(t))$ is defined by averaging over the fast variable t/ε with respect to a smooth bump function ρ:

$$\overline{\mathcal{N}}_\eta\left(\frac{t}{\varepsilon}, \overline{\mathbf{v}}(t)\right) = \frac{1}{\eta}\int_0^\eta \rho\left(\frac{t+s}{\eta}\right)\mathcal{N}\left(\frac{t+s}{\varepsilon}, \overline{\mathbf{v}}(t)\right) ds. \tag{1.13}$$

We next consider the main sources of error in approximating the exact solution (1.11) with the numerical approximation of its average, (1.12). As mentioned earlier, throughout this paper we will assume that the mapping (1.9) is exact and that we can discretize PDEs into the form (1.1) in order to use the notation and concepts of ODE analysis. Let $\mathbf{x}(t)$ represent the exact solution at time t, let $\overline{\mathbf{y}}^\eta(t)$ represent the averaged approximation to the exact equation based on a time-averaging interval of η, and let $\overline{\mathbf{y}}^\eta_{\Delta t}$ represent the time discretization of the average $\overline{\mathbf{y}}^\eta(t)$ with time step Δt. Then we can write, using the triangle inequality,

$$\begin{aligned} ||\mathbf{x}(t) - \overline{\mathbf{y}}^\eta_{\Delta t}|| &= ||\mathbf{x}(t) - \overline{\mathbf{y}}^\eta(t) + \overline{\mathbf{y}}^\eta(t) - \overline{\mathbf{y}}^\eta_{\Delta T}(t)|| \\ &\le \underbrace{||\mathbf{x}(t) - \overline{\mathbf{y}}^\eta(t)||}_{\text{averaging error}} + \underbrace{||\overline{\mathbf{y}}^\eta(t) - \overline{\mathbf{y}}^\eta_{\Delta t}(t)||}_{\text{time-stepping error}}. \end{aligned} \tag{1.14}$$

This error estimate indicates there are two main sources of error: the error due to the time-averaging and the error due to the numerical approximation of the time-averaged evolution equation.

To guide the reader on how the proof will proceed, we visually sketch the outcome of the error analysis of the coarse solver on the right side of Figure 1.1; this will be discussed in sections 3 and 4. The error due to time-averaging is sketched by the dotted line, while the error due to time-stepping the averaged equation is sketched by the dashed line. We can then examine the total error, the solid line, in which we find a "sweet spot" where the addition of the two sources of error reaches a minimum. The similarity between the parareal performance we observed in our example calculation and our error bounds can be seen by comparing the two sides of Figure 1.1.

In the next section we give a short overview of the parareal method to set the context of the work and define our notation. In section 3 we discuss the coarse propagator and introduce the details of the nonlinear averaging. It is in this section that we briefly introduce the concept of resonant sets for constructing error bounds in the average, nonlinear operator for finite ϵ. With that in mind, we proceed in section 4 to prove the error bounds and therefore the convergence of the APinT method for finite ϵ. Finally, in section 5 we give an example for the shallow water equations, where we discuss the concrete form of the coarse solution and the role played by ordered resonant sets. We also present numerical experiments for the one-dimensional rotating shallow water equations that demonstrate the concepts used in the error bound.

2. The parareal algorithm. In this paper we assume the reader has a basic understanding of the parareal method [21, 22]. We include this section to sketch the main concepts and define our notation. The basic approach of the parareal method is to take large time steps ΔT in serial using a coarse integrator of (1.1), and to iteratively refine the solutions in parallel using small time steps Δt and a more accurate integrator. This can result in significant speedup in real (wall-clock) time if the parareal iterations converge rapidly, and either the ratio $\Delta T/\Delta t$ of coarse and fine step sizes is large, or the cost of the coarse solver is much cheaper than that of the fine solver.

For all the discussions in this paper we assume that we are interested in solving (1.1) on the interval $t \in [0, 1]$. To set notation, let $\varphi_{t_{n-1},t_n}(\mathbf{u}_0)$ and $\overline{\varphi}_{t_{n-1},t_n}(\mathbf{u}_0)$ denote the solutions of the original ODE (1.11) and the time-averaged ODE (1.12) at time $t_n = n\Delta T$, respectively, starting from the initial condition $\mathbf{u}_0$ at $t_n = (n-1)\,\Delta T$. Similarly, let $\tilde{\varphi}_{t_{n-1},t_n}(\mathbf{u}_0)$ denote a numerical approximation to the solution at time $t_n = n\Delta T$ of the time-averaged ODE, starting from the initial condition $\mathbf{u}_0$ at $t_{n-1} = (n-1)\,\Delta T$, where the approximate solution is obtained via a one-step time-stepping method.

We divide the time domain into N finite subintervals, $[n\Delta T, (n+1)\Delta T]$, where $n = 0, \ldots, N-1$. The parareal algorithm begins with a coarse solve and then proceeds by computing approximations to the solution, $\mathbf{v}_n^k$, iteratively, as follows:

$$\mathbf{v}_n^k = \tilde{\varphi}_{t_{n-1},t_n}\left(\mathbf{v}_{n-1}^k\right) + \left(\varphi_{t_{n-1},t_n}\left(\mathbf{v}_{n-1}^{k-1}\right) - \tilde{\varphi}_{t_{n-1},t_n}\left(\mathbf{v}_{n-1}^{k-1}\right)\right). \tag{2.1}$$

Here, since the quantities $\mathbf{v}_{n-1}^{k-1}$ in the difference $\varphi_{\Delta T}(\mathbf{v}_{n-1}^{k-1}) - \tilde{\varphi}_{\Delta T}(\mathbf{v}_{n-1}^{k-1})$ are already computed at iteration k, the difference can be computed in parallel for all n. Since the computation of $\overline{\varphi}_{\Delta T}(\mathbf{v}_{n-1}^k)$ is cheap, the overall parallel computation cost will be cheap provided the iterates converge quickly.

3. The coarse propagator. In this section we complete the mathematical description of the coarse propagator that we introduced in section 1.2. These concepts are the underpinning ideas required to understand the error bounds of section 4.

This section is organized as follows. First, in section 3.1 we sketch the relevant results from the theory of fast singular limits to provide the motivation for the approx-

imation used for the coarse propagator. An additional motivation for examining this limiting case is to allow a description of what we expect in the limit of $\epsilon \to 0$. Following this, in section 3.2, we will complete the details of the mathematical description of the coarse propagator; in particular we shall describe the fast-wave-averaging of the nonlinear term.

3.1. Relation of the coarse propagator to fast singular limits in PDEs. The coarse propagator used in this paper has its roots in the mathematics of fast singular limits as applied to (1.1) in the limit as $\epsilon \to 0$ [6, 18, 26, 8]. The analysis begins by separating time into two scales, a slow time scale, t, and a fast time scale, τ. Then the time derivative and its corresponding asymptotic expansion are

$$\frac{\partial}{\partial t} \to \frac{\partial}{\partial t} + \frac{1}{\epsilon}\frac{\partial}{\partial \tau}, \qquad \mathbf{u} = \mathbf{u}^0(t,\tau) + \epsilon \mathbf{u}^1(t,\tau) + \mathrm{O}\left(\epsilon^2\right). \tag{3.1}$$

The leading order solution is

$$\mathbf{u}^0(t,\tau) = e^{-\tau L}\overline{\mathbf{u}}(t) + o(1), \tag{3.2}$$

with

$$\frac{\partial \overline{\mathbf{u}}(t)}{\partial t} + \lim_{\tau\to\infty} \frac{1}{\tau}\int_0^\tau e^{sL}\mathcal{N}(e^{-sL}\overline{\mathbf{u}}(t))\,\mathrm{d}s = 0, \tag{3.3}$$

$$\overline{\mathbf{u}}(t)|_{t=0} = \mathbf{u}^0. \tag{3.4}$$

Equation (3.3) provides the definition for the solution $\overline{\mathbf{u}}(t)$ as an average over the variable s of the nonlinear combination of the matrix exponential operators. One of the conclusions of this theory is that the leading order solution given in (3.2) contains both slow and fast components of the solution because the matrix exponential itself, comprising the eigenfrequencies and eigenfunctions of the skew-Hermitian linear operator, contains all the frequencies, including low frequencies and sometimes zero frequencies, of the system of equations.

3.2. Nonlinear averaging in the coarse propagator. In this section we consider an approximation, similar to the one made in the last section, for when ϵ is finite rather than infinite. This implies an average over a finite time interval, η, rather than an infinite one. From a numerical point of view, the finite time interval can be thought of as a technique for regularizing the nonlinear equation over the interval of the coarse time step.

We remind the reader that in section 1.2 we presented the two main sources of error in the numerical approximation of the APinT coarse propagator. The first step uses the matrix exponential of the skew-Hermitian linear operator to map the unknowns of (1.1), denoted as the vector $\mathbf{v}$, into a new coordinate vector $\mathbf{u}$, arriving at equations (1.11). Though it may be tempting to think of $\mathbf{v}$ as the fast coordinate and $\mathbf{u}$ as the slow coordinate, the mapping described by (1.9) will contain all the frequencies of the system, which will also include low-frequency modes which make important contributions to the solution over the interval of the coarse time step, as well as possible zero-frequency modes. It is therefore more advantageous to think of the matrix exponential as mapping the unknowns of the evolution equation into the space of the oscillations. The second step approximates the exact mapped solution described by (1.11) by a new equation that describes a smoothed, time-averaged solution, (1.12). The solution to (1.12) is the definition of the averaged solution $\overline{\mathbf{v}}$.

To understand the behavior of this averaging operator, we assume that $\mathcal{N}(s,\mathbf{x})$ can be written as a phase expansion in the form

$$\mathcal{N}(s,\mathbf{x}) = \sum_n e^{i\omega_n s}\mathcal{N}_n(\mathbf{x}), \quad \omega_n \in \mathbb{R}, \tag{3.5}$$

where $\omega_0 = 0$ and $\omega_n \neq 0$ for $n \neq 0$. The $i\omega_n s$ in (3.5) depend on the form of the nonlinear term. In geophysical fluid applications with quadratic nonlinearity,

$$\mathcal{N}(s,\mathbf{x}) = e^{sL}\mathcal{B}\left(e^{-sL}\mathbf{x}, e^{-sL}\mathbf{x}\right),$$

where $\mathcal{B}$ is a bilinear form and L is a skew-Hermitian matrix; the functional form (3.5) that arises from decomposing $\mathbf{x}$ in terms of eigenvectors $\mathbf{w}_n$ of L with corresponding eigenvalues $i\lambda_n$ is given by

$$\mathcal{N}(s,\mathbf{x}) = \sum_{n_3}\sum_{n_1,n_2} e^{i\left(\lambda_{n_3}-\lambda_{n_2}-\lambda_{n_1}\right)s} a_{n_1,n_2,n_3}\left(\mathbf{x}\cdot\mathbf{w}_{n_1}\right)\left(\mathbf{x}\cdot\mathbf{w}_{n_2}\right)\mathbf{w}_{n_3}, \tag{3.6}$$

where $a_{n_1,n_2,n_3} = \mathcal{B}(\mathbf{w}_{n_1},\mathbf{w}_{n_2})\cdot\mathbf{w}_{n_3}$. In this case, the frequencies of (3.5) are a combination of the frequencies of the linear operator L,

$$\omega_{n_3} = \lambda_{n_3} - \lambda_{n_2} - \lambda_{n_1}. \tag{3.7}$$

Examining (3.6), it is seen that three fast frequencies can be combined to produce a total frequency, ω_{n3}, that is small, leading to a low-frequency solution. Because it is the combination of fast frequencies that leads to low-frequency dynamics, we point out that numerical methods that cause errors in these frequencies could have a further effect of also causing errors in the low-frequency component of the nonlinear solution. A concrete example of (3.6) will be given in section 5. In addition, the function $\mathcal{N}_0(\mathbf{x})$ governing the slow time behavior in the decomposition (3.5) is given by all linear combinations of terms $a_{n_1,n_2,n_3}\left(\mathbf{x}\cdot\mathbf{w}_{n_1}\right)\left(\mathbf{x}\cdot\mathbf{w}_{n_2}\right)\mathbf{w}_{n_3}$ for which $\lambda_{n_3}-\lambda_{n_2}-\lambda_{n_1}=0$.

From (3.5),

$$\frac{d\overline{\mathbf{v}}}{dt}(t) = \sum_n e^{i\omega_n t/\varepsilon}\left(\int_0^1 \rho(s)\, e^{i\omega_n \eta s/\varepsilon} ds\right)\mathcal{N}_n\left(\overline{\mathbf{v}}(t)\right). \tag{3.8}$$

Choosing, e.g., $\rho(s) = \rho_0 \exp\left(-1/\left(s(1-s)\right)\right)$, it turns out (see [17]) that there are constants a_0 and a_1 for which

$$\left|\int_0^1 \rho(s)\, e^{i\omega_n \eta s/\varepsilon} ds\right| \lesssim a_0\left|e^{-a_1\sqrt{\omega_n \eta/\varepsilon}}\right|. \tag{3.9}$$

Note that when $\eta \gg \varepsilon/\min_{\omega_n\neq 0}|\omega_n|$,

$$\frac{d\overline{\mathbf{v}}}{dt}(t) \approx \mathcal{N}_0\left(\overline{\mathbf{v}}(t)\right),$$

and the above identity is exact in the limit of $\eta \to \infty$.

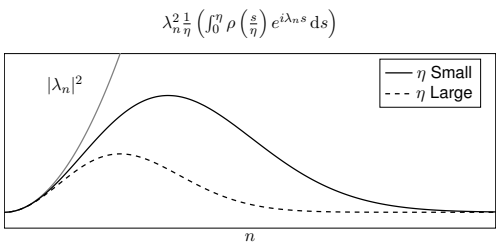

FIG. 4.1. *A schematic depiction of the time-smoothing properties of the time-averaging kernel* $\rho(\cdot)$.

4. Error bounds.

4.1. Basic ideas behind the error bounds and a scaling result. The basic idea behind the time-stepping error bounds is that the time-stepping error for an order p time-stepping method scales like

$$\|\overline{\mathbf{v}}(t_n) - \mathbf{v}_n\| \sim \Delta T^p \left\| \left(\frac{d}{dt} \right)^{p+1} \overline{\mathbf{v}}(t) \right\| \tag{4.1}$$

$$\sim \max_{\omega_0 \leq \omega} \left| \left(\frac{\omega}{\varepsilon} \right)^{p+1} \int_0^1 \rho(s) \, e^{i\omega\eta s/\varepsilon} ds \right| \tag{4.2}$$

$$\sim a_0 \max_{\omega_0 \leq \omega} \left(\frac{\omega}{\varepsilon} \right)^{p+1} e^{-a_1 \sqrt{\omega_n \eta/\varepsilon}}, \tag{4.3}$$

where $\mathbf{v}_n$ denotes the numerical approximation to $\overline{\mathbf{v}}(t_n)$ and we used the decomposition (3.5) and the bound (3.9). The qualitative behavior for the time-smoothing properties of

$$\max_{\omega_0 \leq \omega} \left| \left(\frac{\omega}{\varepsilon} \right)^{p+1} \int_0^1 \rho(s) \, e^{i\omega\eta s/\varepsilon} ds \right|$$

as a function of ω is displayed in Figure 4.1.

Also, it can be shown (see Corollary 4.1) that the difference between $\overline{\mathbf{v}}(t_n)$ and $\mathbf{v}(t_n)$ scales like

$$\|\mathbf{v}(t_n) - \overline{\mathbf{v}}(t_n)\| \sim C\eta\varepsilon. \tag{4.4}$$

Since $\|\mathbf{v}(t_n) - \mathbf{v}_n\| \leq \|\mathbf{v}(t_n) - \overline{\mathbf{v}}(t_n)\| + \|\overline{\mathbf{v}}(t_n) - \mathbf{v}_n\|$, combining (4.3) and (4.4) suggests that each parareal iteration attenuates the error by a factor of

$$\frac{C\eta\varepsilon}{\Delta T} + \Delta T^p a_0 \max_{\omega_0 \leq \omega} \left(\frac{\omega}{\varepsilon} \right)^p e^{-a_1 \sqrt{\omega_n \eta/\varepsilon}}, \tag{4.5}$$

which represents a balance between errors from the time homogenization and the time-stepping errors. This heuristic argument is formalized in Theorem 4.2.

Given the general form (4.5) for the error attenuation, the goal is to choose a time-averaging window $\eta = \eta(\varepsilon, \Delta T)$ that yields fast convergence, where the coarse time step ΔT is only constrained by the slow time scales (if $\varepsilon \ll 1$). To show that such a choice for η exists, write $\eta(\varepsilon, \Delta T) = \varepsilon^s \Delta T^2 / \omega_{\min}$, where $0 < s < 1$. Looking at

$$\frac{\eta(\varepsilon, \Delta T)\, \varepsilon}{\Delta T} + (\Delta T \, |\omega_{\max}|)^{p+1} \left(\frac{1}{\varepsilon} \right)^{p+1} \left| e^{-C_0 \sqrt{\Delta T/\varepsilon^{1-s}}} \right|,$$

we need the quantity

$$\varepsilon^{1+s}\Delta T/\omega_{\min} + |\omega_{\max}| \left(\Delta T\,|\omega_{\max}|\right)^{p+1} \left(\frac{1}{\varepsilon}\right)^{p+1} \left|e^{-C_0\sqrt{\Delta T/\varepsilon^{1-s}}}\right| \tag{4.6}$$

to be small. Notice that, since $1-s>0$, the second term in (4.6),

$$\left(\Delta T\,|\omega_{\max}|\right)^{p+1} \left(\frac{1}{\varepsilon}\right)^{p+1} \left|e^{-C_0\sqrt{\Delta T/\varepsilon^{1-s}}}\right|,$$

goes to zero superexponentially fast as $\varepsilon \to 0$, and is bounded for all values of ε given a fixed coarse time step ΔT. Also, since $s>0$, the time window η over which the averaging is performed goes to zero as $\varepsilon \to 0$.

To connect the form of the error attenuation (4.6) to the discretization of geophysically relevant PDEs (e.g., the shallow water equations), note that the size of $\omega_{\min}$ and $\omega_{\max}$ will scale with the grid spacing N^{-1} in some way, e.g., $\omega_{\min} \sim N^{-1}$ and $|\omega_{\max}| \sim N$. This requires choosing a coarse time step $\Delta T \lesssim N^{-1}$ in order for the error (4.6) to be small. Importantly, this choice for the coarse time step ΔT is independent of ε, as desired.

4.2. Statement and proof of error bounds. We now derive error bounds for the parareal iteration. Our error bounds rely on the following result, which follows from Proposition A.1 in the appendix and the change of variables $\tau = t/\varepsilon$. We note that Proposition A.1 is a slight modification of the proof of Lemma 3.2.9 in [25]. For notational simplicity, throughout the proof we let C denote a generic constant that implicitly depends on the constants λ, M, and C_0 in Corollary 4.1. The key point is that the constant C is not expected to grow as $\varepsilon \to 0$.

COROLLARY 4.1. *Consider*

$$\frac{d\mathbf{v}}{dt}(t) = \mathcal{N}\left(\frac{t}{\varepsilon}, \mathbf{v}(t)\right), \quad 0 \le t \le 1,$$

and its averaged version

$$\frac{d\overline{\mathbf{v}}}{dt}(t) = \overline{\mathcal{N}}_\eta\left(\frac{t}{\varepsilon}, \overline{\mathbf{v}}(t)\right), \quad 0 \le t \le 1,$$

where

$$\overline{\mathcal{N}}_\eta\left(\frac{t}{\varepsilon}, \overline{\mathbf{v}}(t)\right) = \frac{1}{\eta}\int_0^\eta \rho\left(\frac{t+s}{\eta}\right)\mathcal{N}\left(\frac{t+s}{\varepsilon}, \overline{\mathbf{v}}(t)\right) ds.$$

Assume that there are constants λ and M such that

$$\max_{0\le t,t_1,t_2\le 1} \left\|\mathcal{N}\left(\frac{\tau_0}{\varepsilon}, \mathbf{v}(\tau_1)\right) - \mathcal{N}\left(\frac{\tau_0}{\varepsilon}, \mathbf{v}(\tau_2)\right)\right\| \le \lambda \|\mathbf{v}(\tau_1) - \mathbf{v}(\tau_2)\|$$

and

$$M = \max_{0\le\tau\le 1} \left\|\mathcal{N}\left(\frac{\tau}{\varepsilon}, \mathbf{v}(\tau)\right)\right\| < \infty, \quad C_0 = \int_0^1 \rho(s)\, s ds. \tag{4.7}$$

Then

$$\|\mathbf{v}(t) - \overline{\mathbf{v}}(t)\| = C_0(1+\lambda) M\eta\varepsilon e^{\varepsilon\lambda t}, \quad 0 \le t \le 1.$$

Proof. Use the change of variables $\tau = t/\varepsilon$ and Proposition A.1. ∎

Before stating the hypothesis on the propagation operators $\varphi_{t_{n-1},t_n}(\mathbf{v}_0)$, $\overline{\varphi}_{t_{n-1},t_n}(\mathbf{v}_0)$, and $\tilde{\varphi}_{t_{n-1},t_n}(\mathbf{v}_0)$, we consider forward Euler as a motivating example (recall that the propagation operators are defined in section 2). For forward Euler, the local truncation error from time-stepping on the time-averaged equation (1.12) satisfies

$$\overline{\varphi}_{t_{n-1},t_n}(\mathbf{v}) - \tilde{\varphi}_{t_{n-1},t_n}(\mathbf{v}) = \frac{1}{2}\left[\frac{d^2 v_j}{dt^2}\left(\tau_j^n\right)\right]_j \Delta T^2, \quad t_{n-1} \leq \tau_j^n \leq t_n.$$

Now, defining

$$\kappa(\varepsilon,\eta,\omega) = \int_0^1 \rho(s)\, e^{i\omega\eta s/\varepsilon} ds$$

and

$$\mathcal{N}^{(1)}_{m_1,m_2}(\overline{\mathbf{v}}(t)) = J\mathcal{N}_{m_1}(\overline{\mathbf{v}}(t))\,\mathcal{N}_{m_2}(\overline{\mathbf{v}}(t)),$$

a direct calculation from (3.8) yields

$$\left(\frac{d}{dt}\right)^2 \mathbf{v}(t) = \sum_m e^{i\omega_m t/\varepsilon}\left(\frac{i\omega_m}{\varepsilon}\right)\kappa(\varepsilon,\eta,\omega_m)\,\mathcal{N}_m(\overline{\mathbf{v}}(t)) + \sum_{m_1,m_2} e^{i\left(\omega_{m_1}+\omega_{m_2}\right)t/\varepsilon}\kappa(\varepsilon,\eta,\omega_{m_1})\,\kappa(\varepsilon,\eta,\omega_{m_2})\,\mathcal{N}^{(1)}_{m_1,m_2}(\overline{\mathbf{v}}(t)).$$

Using the fact that $|\kappa(\varepsilon,\eta,\omega_{m_1})| \leq 1$, there are smooth functions $\mathcal{M}_0(\mathbf{v},\varepsilon,\eta,\Delta T)$ and $\mathcal{M}_0(\mathbf{v})$ such that

$$\overline{\varphi}_{t_{n-1},t_n}(\mathbf{v}) - \tilde{\varphi}_{t_{n-1},t_n}(\mathbf{v}) = \mathcal{M}_0(\mathbf{v};\varepsilon,\eta,\Delta T)$$

and

$$\|\mathcal{M}_0(\mathbf{v},\varepsilon,\eta,\Delta T)\| \leq C\Delta T^2 \max_{\omega_0\leq\omega}\left\{\left|\frac{\omega}{\varepsilon}\right|\kappa(\varepsilon,\eta,\omega)\right\}\|\mathcal{M}_0(\mathbf{v})\|.$$

More generally, for a general order p one-step time-stepping scheme, a calculation analogous to the one above shows that there are smooth functions $\mathcal{M}_0(\mathbf{v},\varepsilon,\eta,\Delta T)$ and $\mathcal{M}_0(\mathbf{v})$ such that the local truncation error from time-stepping on (1.12) satisfies

$$\overline{\varphi}_{t_{n-1},t_n}(\mathbf{v}) - \tilde{\varphi}_{t_{n-1},t_n}(\mathbf{v}) = \mathcal{M}_0(\mathbf{v},\varepsilon,\eta,\Delta T), \tag{4.8}$$

where

$$\|\mathcal{M}_0(\mathbf{v};\varepsilon,\eta)\| \leq C\Delta T^{p+1} \max_{\omega_0\leq\omega}\left\{\left|\frac{\omega}{\varepsilon}\right|^p\kappa(\varepsilon,\eta,\omega)\right\}\|\mathcal{M}_0(\mathbf{v})\|. \tag{4.9}$$

We also note that, from Corollary 4.1, there are smooth functions $\mathcal{M}_1(\mathbf{v},\varepsilon,\eta,\Delta T)$ and $\mathcal{M}_1(\mathbf{v})$ such that

$$\varphi_{t_{n-1},t_n}(\mathbf{v}) - \overline{\varphi}_{t_{n-1},t_n}(\mathbf{v}) \leq \mathcal{M}_1(\mathbf{v},\varepsilon,\eta,\Delta T) \tag{4.10}$$

and

$$\|\mathcal{M}_1(\mathbf{v},\varepsilon,\eta,\Delta T)\| \leq C\eta\varepsilon\|\mathcal{M}_1(\mathbf{v})\|. \tag{4.11}$$

We introduce the main assumptions on the propagation operators $\varphi_{t_{n-1},t_n}(\mathbf{v}_0)$, $\overline{\varphi}_{t_{n-1},t_n}(\mathbf{v}_0)$, and $\tilde{\varphi}_{t_{n-1},t_n}(\mathbf{v}_0)$. To do so, it will be convenient to introduce the operators $\mathcal{E}_{t_{n-1},t_n}(\cdot)$ and $\tilde{\mathcal{E}}_{t_{n-1},t_n}(\cdot)$:

$$\mathcal{E}_{t_{n-1},t_n}(\mathbf{v}) = \varphi_{t_{n-1},t_n}(\mathbf{v}) - \overline{\varphi}_{t_{n-1},t_n}(\mathbf{v})$$

and

$$\tilde{\mathcal{E}}_{t_{n-1},t_n}(\mathbf{v}) = \overline{\varphi}_{t_{n-1},t_n}(\mathbf{v}) - \tilde{\varphi}_{t_{n-1},t_n}(\mathbf{v}).$$

Using (4.8), (4.9), (4.10), and (4.11), we see that

(4.12)
$$\left\|\tilde{\mathcal{E}}_{t_{n-1},t_n}(\mathbf{v}_1) - \tilde{\mathcal{E}}_{t_{n-1},t_n}(\mathbf{v}_2)\right\| \leq C\Delta T^{p+1} \max_{\omega_0 \leq \omega} \left\{ \left|\frac{\omega}{\varepsilon}\right|^p \kappa(\varepsilon,\eta,\omega) \right\} \|\mathcal{M}_0(\mathbf{v}_1) - \mathcal{M}_0(\mathbf{v}_2)\|$$

and

$$\left\|\mathcal{E}_{t_{n-1},t_n}(\mathbf{v}_1) - \mathcal{E}_{t_{n-1},t_n}(\mathbf{v}_2)\right\| \leq C\eta\varepsilon \|\mathcal{M}_1(\mathbf{v}_1) - \mathcal{M}_1(\mathbf{v}_2)\|. \tag{4.13}$$

Finally, we assume that the propagation operator $\tilde{\varphi}_{t_{n-1},t_n}(\cdot)$ associated with performing the time-stepping method on the time-averaged equation satisfies

$$\left\|\tilde{\varphi}_{t_{n-1},t_n}(\mathbf{v}_1) - \tilde{\varphi}_{t_{n-1},t_n}(\mathbf{v}_2)\right\| \leq (1 + C\Delta T)\|\mathbf{v}_1 - \mathbf{v}_2\|. \tag{4.14}$$

From the decomposition (3.5), we expect the above bound to hold for all standard time-stepping methods, and with a constant C that doesn't scale with ε.

We can finally state the main theoretical result of the paper.

THEOREM 4.2. *Let $\mathbf{v}(t_n)$ denote the solution to the ODE (1.11) at time $t_n = n\Delta T$, and let $\mathbf{v}_n^k$ denote the approximation to $\mathbf{v}(t_n)$ obtained after k parareal iterations, where the coarse parareal integrator corresponds to applying an order p time-stepping method to the time-averaged equation (1.12). Then, defining*

$$\kappa_p(\varepsilon,\eta) = \max_{\omega_0 \leq \omega} \left(\frac{\omega}{\varepsilon}\right)^p \int_0^1 \rho(s)\, e^{i\omega\eta s/\varepsilon} ds,$$

the error $\mathbf{v}(t_n) - \mathbf{v}_n^k$ after the kth parareal iteration satisfies

$$\left\|\mathbf{v}(t_n) - \mathbf{v}_n^k\right\| \leq \frac{C^{k+1}}{(k+1)!} e^C \left(\varepsilon\eta + \kappa_p(\varepsilon,\eta)\Delta T^p\right) \left(\frac{\eta\varepsilon}{\Delta T} + \kappa_p(\varepsilon,\eta)\Delta T^p\right)^k, \tag{4.15}$$

where the constant C implicitly depends on λ, M, and C_0 in Corollary 4.1.

Proof. The proof follows along the same lines as Theorem 1 in [11].

First consider when $k = 0$. Since the local truncation error satisfies

$$\left\|\overline{\varphi}_{t_{n-1},t_n}(\mathbf{v}_0) - \tilde{\varphi}_{t_{n-1},t_n}(\mathbf{v}_0)\right\| \leq C\Delta T^{p+1}\kappa_p(\varepsilon,\eta)\|\mathbf{v}_0\|,$$

the standard proof for one-step time-stepping methods shows that

$$\left\|\overline{\mathbf{v}}(t_n) - \mathbf{v}_n^0\right\| \leq C\kappa_p(\varepsilon,\eta)\Delta T^p \|\mathbf{v}_0\|.$$

Also, from Proposition A.1,

$$\|\mathbf{v}(t_n) - \overline{\mathbf{v}}(t_n)\| \leq C\varepsilon\eta.$$

Therefore,

$$\begin{aligned}\left\|\mathbf{v}(t_n) - \mathbf{v}_n^0\right\| &\leq \|\mathbf{v}(t_n) - \overline{\mathbf{v}}(t_n)\| + \left\|\overline{\mathbf{v}}(t_n) - \mathbf{v}_n^0\right\| \\ &\leq C\varepsilon\eta + C\kappa_p(\varepsilon,\eta)\Delta T^p.\end{aligned}$$

Now consider some $k \geq 1$. Recall that the parareal scheme can be written as

$$\mathbf{v}_n^k = \tilde{\varphi}_{t_{n-1},t_n}\left(\mathbf{v}_{n-1}^k\right) + \left(\varphi_{t_{n-1},t_n}\left(\mathbf{v}_{n-1}^{k-1}\right) - \tilde{\varphi}_{t_{n-1},t_n}\left(\mathbf{v}_{n-1}^{k-1}\right)\right).$$

We can rewrite this as

$$\begin{aligned}\left\|\mathbf{v}\left(t_n\right) - \mathbf{v}_n^k\right\| = &\left(\tilde{\varphi}_{t_{n-1},t_n}\left(\mathbf{v}\left(t_{n-1}\right)\right) - \tilde{\varphi}_{t_{n-1},t_n}\left(\mathbf{v}_{n-1}^k\right)\right)\\ &+ \left(\mathcal{E}_{t_{n-1},t_n}\left(\mathbf{v}\left(t_{n-1}\right)\right) - \mathcal{E}_{t_{n-1},t_n}\left(\mathbf{v}_{n-1}^{k-1}\right)\right)\\ &+ \left(\tilde{\mathcal{E}}_{t_{n-1},t_n}\left(\mathbf{v}\left(t_{n-1}\right)\right) - \tilde{\mathcal{E}}_{t_{n-1},t_n}\left(\mathbf{v}_{n-1}^{k-1}\right)\right).\end{aligned}$$

From the bounds (4.12), (4.13), and (4.14) and the assumed smoothness of the operators $\mathcal{M}_0(\cdot)$ and $\mathcal{M}_1(\cdot)$,

$$\left\|\tilde{\varphi}_{t_{n-1},t_n}\left(\mathbf{v}\left(t_{n-1}\right)\right) - \tilde{\varphi}_{t_{n-1},t_n}\left(\mathbf{v}_{n-1}^k\right)\right\| \leq \left(1 + C\Delta T\right)\left\|\mathbf{v}\left(T_{n-1}\right) - \mathbf{v}_{n-1}^k\right\|,$$

$$\left\|\mathcal{E}_{t_{n-1},t_n}\left(\mathbf{v}\left(t_{n-1}\right)\right) - \mathcal{E}_{t_{n-1},t_n}\left(\mathbf{v}_{n-1}^{k-1}\right)\right\| \leq C\eta\varepsilon\left\|\mathbf{v}\left(T_{n-1}\right) - \mathbf{v}_{n-1}^{k-1}\right\|,$$

and

$$\left\|\tilde{\mathcal{E}}_{t_{n-1},t_n}\left(\mathbf{v}\left(t_{n-1}\right)\right) - \tilde{\mathcal{E}}_{t_{n-1},t_n}\left(\mathbf{v}_{n-1}^{k-1}\right)\right\| \leq C\kappa_p\left(\varepsilon,\eta\right)\Delta T^{p+1}\left\|\mathbf{v}\left(t_{n-1}\right) - \mathbf{v}_{n-1}^{k-1}\right\|.$$

Therefore,

$$\begin{aligned}\left\|\mathbf{v}\left(T_n\right) - \mathbf{v}_n^k\right\| \leq &\left(1 + C\Delta T\right)\left\|\mathbf{v}\left(T_{n-1}\right) - \mathbf{v}_{n-1}^k\right\|\\ &+ C\left(\eta\varepsilon + \kappa_p\left(\varepsilon,\eta\right)\Delta T^{p+1}\right)\left\|\mathbf{v}\left(t_{n-1}\right) - \mathbf{v}_{n-1}^{k-1}\right\|.\end{aligned}$$

Finally, following the same steps as in Theorem 1 in [11],

$$\begin{aligned}\left\|\mathbf{v}\left(t_n\right) - \mathbf{v}_n^k\right\| &\leq C\left(\varepsilon\eta + \kappa_p\left(\varepsilon,\eta\right)\Delta T^p\right)\frac{\left(C\left(\eta\varepsilon + \kappa_p\left(\varepsilon,\eta\right)\Delta T^{p+1}\right)\right)^k}{(k+1)!}\left(1 + C\Delta T\right)^{n-k-1} n^k\\ &\leq \frac{C^k}{(k+1)!}e^{C\left(t_n - t_{k+1}\right)}\left(\varepsilon\eta + \kappa_p\left(\varepsilon,\eta\right)\Delta T^p\right)\left(\Delta T n\right)^k\left(\eta\varepsilon/\Delta T + \kappa_p\left(\varepsilon,\eta\right)\Delta T^p\right)^k\\ &\leq \frac{C^{k+1}}{(k+1)!}e^C\left(\varepsilon\eta + \kappa_p\left(\varepsilon,\eta\right)\Delta T^p\right)\left(\frac{\eta\varepsilon}{\Delta T} + \kappa_p\left(\varepsilon,\eta\right)\Delta T^p\right)^k,\end{aligned}$$

where we used the fact that $n\Delta T \leq 1$ and $(t_n - t_{k+1}) \leq 1$ in the last step. ∎

5. The one-dimensional rotating shallow water equations. In this section we present an example of the method applied to the one-dimensional rotating shallow water equations [14], which has quadratic nonlinearity. We first discuss the spectral decomposition, including concrete definitions of the resonant sets, then show numerical verification of the bound proved in Corollary 4.1 and Theorem 4.2 of section 4. Let (1.1) exist in the one-dimensional 2π-periodic domain, and let the vector of unknowns be

$$\mathbf{u}(t,x) = \left(v_1(t,x), v_2(t,x), h(t,x)\right)^T. \tag{5.1}$$

We then write the linear and nonlinear operators in (1.1) as

$$L = \begin{pmatrix} 0 & -1 & F^{-1/2}\partial_x \\ 1 & 0 & 0 \\ F^{-1/2}\partial_x & 0 & 0 \end{pmatrix}, \quad \mathcal{N}(\mathbf{u}) = \begin{pmatrix} v_1(v_1)_x \\ v_1(v_2)_x \\ (hv_1)_x \end{pmatrix} \tag{5.2}$$

for some constant $F \in \mathbb{R}$. The corresponding purely imaginary eigenvalues are

$$\lambda_k^\alpha = \alpha\lambda_0, \quad \lambda_0 = \sqrt{1 + F^{-1}k^2}, \quad \alpha = -1, 0, +1, \tag{5.3}$$

where $\mathbf{k} = (k_1, k_2)$ is the wave number, α indicates whether the mode is slow ($\alpha = 0$) or fast ($\alpha = \pm 1$), and $\mathbf{r}_\mathbf{k}^\alpha$ are the eigenvectors,

$$\begin{matrix} \mathbf{r}_k^{-1} \end{matrix} \quad \begin{matrix} \mathbf{r}_k^{0} \end{matrix} \quad \begin{matrix} \mathbf{r}_k^{+1} \end{matrix} \qquad \begin{matrix} \mathbf{r}_0^{-1} \end{matrix} \quad \begin{matrix} \mathbf{r}_0^{0} \end{matrix} \quad \begin{matrix} \mathbf{r}_0^{+1} \end{matrix}$$

$$\begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{i}{\sqrt{2}\lambda_0} \\ \frac{F^{-1/2}k}{\sqrt{2}\lambda_0} \end{pmatrix} \quad \begin{pmatrix} 0 \\ \frac{F^{-1/2}ik}{\lambda_0} \\ \frac{1}{\lambda_0} \end{pmatrix} \quad \begin{pmatrix} -\frac{1}{\sqrt{2}} \\ -\frac{i}{\sqrt{2}\lambda_0} \\ \frac{F^{-1/2}k}{\sqrt{2}\lambda_0} \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} \frac{i}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{pmatrix} \quad \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \quad \begin{pmatrix} \frac{-i}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{pmatrix}. \tag{5.4}$$

The left three eigenfunctions correspond to the case when $k \neq 0$, and the right three eigenfunctions are for the case when $k = 0$. We can then write the concrete form of the fast-wave-average equation (1.12) as

$$\frac{d\sigma_\mathbf{k}^\alpha(t)}{dt} + \frac{1}{\eta}\int_0^\eta \sum_{\mathbf{k}\in\mathbb{Z}^2}\sum_{\alpha=-1}^{1}\left(\sum_{\mathbf{k}_1,\mathbf{k}_2,\alpha_1,\alpha_2\in\mathcal{S}_{\mathbf{k},\alpha}^{\epsilon_\beta}} \sigma_{\mathbf{k}_1}^{\alpha_1}(t)\sigma_{\mathbf{k}_2}^{\alpha_2}(t) \times C_{\mathbf{k},\mathbf{k}_1,\mathbf{k}_2}^{\alpha,\alpha_1\alpha_2} e^{i(\mathbf{k}\cdot\mathbf{x}) - i\omega_{\mathbf{k},\mathbf{k}_1,\mathbf{k}_2}^{\alpha,\alpha_1,\alpha_2}s/\varepsilon}\right)\mathbf{r}_\mathbf{k}^\alpha ds = 0, \tag{5.5}$$

where $\omega_{\mathbf{k},\mathbf{k}1,\mathbf{k}2}^{\alpha,\alpha_1,\alpha_2} = \lambda_{\mathbf{k}_1}^{\alpha_1} + \lambda_{\mathbf{k}_2}^{\alpha_2} - \lambda_\mathbf{k}^\alpha$ and $C_{\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}}^{\alpha_1,\alpha_2,\alpha}$ is an interaction coefficient that can be found in [23], i.e.,

$$C_{\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}}^{\alpha_1,\alpha_2,\alpha} = \frac{i}{2}\Big[(\mathbf{k_2}\cdot\mathbf{v}_{\mathbf{H}k_1}^{\alpha_1})\langle\mathbf{r}_{k_2}^{\alpha_2},\mathbf{r}_k^\alpha\rangle + (\mathbf{k_1}\cdot\mathbf{v}_{\mathbf{H}k_2}^{\alpha_2})\langle\mathbf{r}_{k_1}^{\alpha_1},\mathbf{r}_k^\alpha\rangle + (\mathbf{k_1}\cdot\mathbf{v}_{\mathbf{H}k_1}^{\alpha_1})\mathbf{r}_{k_2}^{\alpha_2}[3]\mathbf{r}_k^\alpha[3] + (\mathbf{k_2}\cdot\mathbf{v}_{\mathbf{H}k_2}^{\alpha_2})\mathbf{r}_{k_1}^{\alpha_1}[3]\mathbf{r}_k^\alpha[3]\Big], \tag{5.6}$$

where we've used the notation $\mathbf{k} = (k, 0)$ is a wave vector with only one component and $\mathbf{v}_{\mathbf{H}k}^\alpha = (\mathbf{r}_k^\alpha[1], \mathbf{r}_k^\alpha[2])$. The ordered resonant sets are, for this special case of quadratic nonlinearity, the following:

$$\mathcal{S}_{\mathbf{k},\alpha}^{\epsilon_\beta} = \left\{(\mathbf{k}_1,\mathbf{k}_2,\alpha_1,\alpha_2) : \mathbf{k} = \mathbf{k}_1 + \mathbf{k}_2, \quad \epsilon_{\beta-1} < \frac{1}{\varepsilon}|\lambda_\mathbf{k}^\alpha - \lambda_{\mathbf{k}_1}^{\alpha_1} + \lambda_{\mathbf{k}_2}^{\alpha_2}| \leq \epsilon_\beta\right\}, \tag{5.7}$$

where $\epsilon_0 = 0$ by definition. Because the ordered set (5.7) involves linear combinations of three wave vectors and three frequencies of the linear operator, it is also used to indicate ordered "triad" interactions. Compare the exponential part of the phase expansion in (3.6) with (5.5), where the role of the resonant sets is more concretely defined. Specifically, when the integral in (5.5) over s is close to ϵ, the oscillatory term makes low-frequency contributions to the solution because the phase is nearly constant. Beyond the low-frequency contributions, the oscillations cancel themselves out, making almost no contribution because they have been averaged away. Therefore, the extent to which the ordered resonant sets are retained and rejected by the averaging procedure is fundamental for accuracy of the method.

In Figure 5.1 we show numerical solutions to illustrate the coarse solver's solution relative to the fine solver's solution. The numerical simulations are for the one-dimensional shallow water equations described in this section with an initial condition

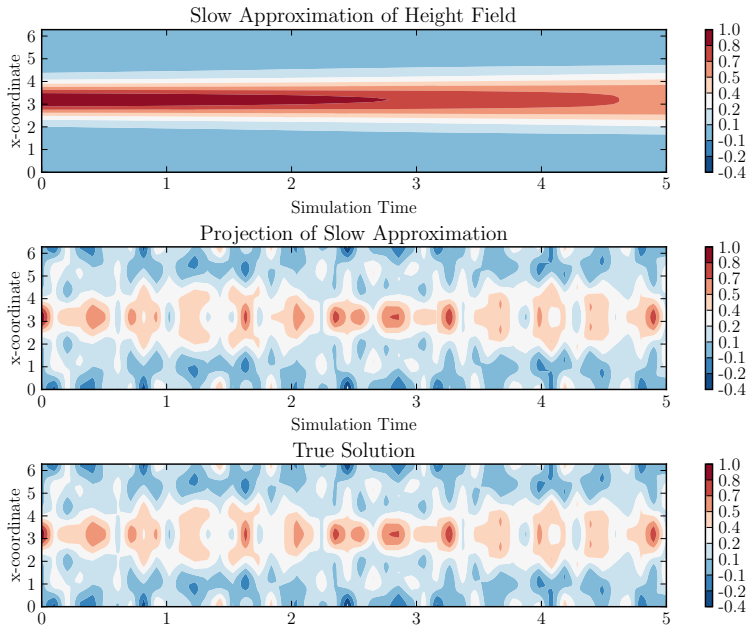

FIG. 5.1. *A comparison of the solution derived from the averaging method with the "true" solution. All three plots show spatio-temporal oscillations in the height field of the one-dimensional rotating shallow water equations with the time coordinate on the horizontal axis and the spatial coordinate on the vertical axis. The top plot shows the fast-wave-averaged approximation of the height field (the third component of* $\overline{\mathbf{u}}$*). The time-stepping is performed over this slower quantity with decreased oscillatory stiffness and therefore an increase in time step. The middle plot is the transformation of this quantity back into normal space by application of the matrix exponential,* $e^{-t/\varepsilon L}$*. This is the coarse solution which is used after the first coarse solve. The quality of this when compared to the solution computed with the fine solver shown in the last plot (to which the APinT algorithm converges) is what allows rapid convergence of parareal. In this example,* $\varepsilon = 0.01$ *and the averaging window* $\eta = 1.0$*.*

of a Gaussian height field. A second-order Strang splitting method was used for both the coarse and fine solves. The figure shows an amplitude versus time plot of the low-frequency solution seen by the coarse propagator (upper panel) and its relation to that of the fine solution, which contains all the fast oscillations (bottom panel) for the stiff case where $\varepsilon = 0.01$. The middle panel shows the coarse solution of (1.12) transformed into the original coordinates by the matrix exponential $e^{-t/\varepsilon L}$. Comparing the middle panel and the bottom panel shows numerical evidence that the coarse solver solution is close to the full solution (bottom panel). It also demonstrates that the averaged solution (top panel) lacks the rapid oscillations of the fine solution and thus allows large time steps to be taken by the coarse solver. We refer the reader to the figure's caption for more details.

We next present convergence results for three different values of ϵ as an illustrative example of Theorem 4.2. Figure 5.2 shows the norm of the coarse error, i.e., $\|\mathbf{x}(t) - \overline{\mathbf{y}}^{\eta}_{\mathbf{\Delta T}}(t)\|_2$, computed relative to the fine time step versus the width of the averaging window, η, where $0 \leq t \leq 1$, $\delta t = 2\text{e}{-4}$, and the spatial resolution is $N = 64$. For the smallest ε we can identify the qualities we expect of the asymptotic solution: when we choose $\eta \to \infty$ as $\varepsilon \to 0$, the error between the full solution and the coarse (slow) solution reaches a minimum, and no further increases in accuracy are available.

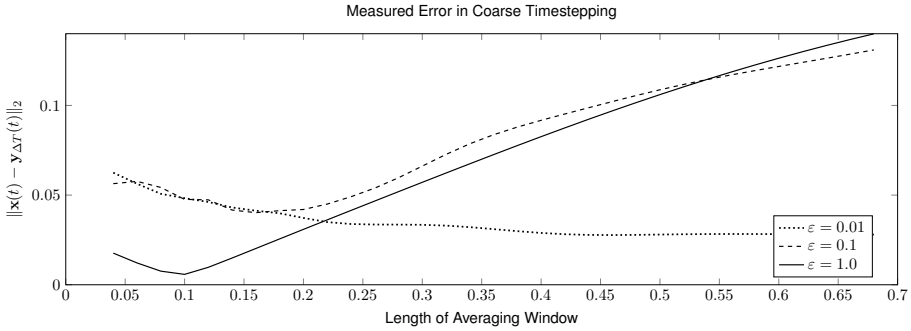

FIG. 5.2. *Computed coarse error for $\Delta T = 0.1$. This is a numerical estimate of the error corresponding to $\|\mathbf{x}(t) - \overline{\mathbf{y}}^{\eta}{}_{\Delta T}(t)\|$, computed by brute-force comparison of the averaged coarse solution to a finely computed reference solution. Note the clear existence of an optimal averaging window for the case where $\varepsilon = 1.0$, and the tendency towards the asymptotic theory, i.e., the error becoming inversely proportional to the averaging window length, η, as $\varepsilon \to 0$.*

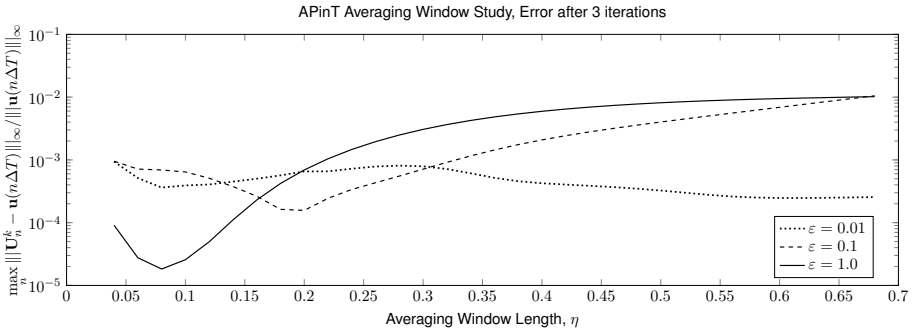

FIG. 5.3. *Iterative error in APinT with one-dimensional rotating shallow water equations after three iterations for $\Delta T = 0.1$. Whereas Figure 5.2 showed the measured total (i.e., time-stepping plus averaging) error in the coarse time-stepping, this figure shows the iterative error for a full parareal solve of the rotating shallow water equations after three iterations for the same computational conditions. The behavior with respect to variation of the averaging window, and particularly the location of the optimal window length, is well predicted by the brute-force computation of the coarse time-stepping error.*

However, for larger ε such as the two cases shown, there is a clear optimal size for the averaging window; i.e., a minimum is present predicted by Theorem 4.2. This demonstrates that the behavior in this example is consistent with the error bounds derived in this paper. Figure 5.3 shows the iterative error in the APinT method after three iterations for the same parameters as in Figure 5.2. Comparing the two figures shows the direct computation of the coarse time-stepping error provides good qualitative agreement with direct computation. This is also in agreement with the estimates in Theorem 4.2.

6. Conclusion. We have investigated the convergence of a parareal method using the APinT coarse solver, which provides a technique by which oscillatory-stiff equations may be solved with the parareal method. The convergence of this method is due to the averaging applied to the nonlinear terms of the coarse solution, which filters the fast waves and mitigates the oscillatory stiffness present in many of the equations of mathematical physics. This averaging must be performed over the en-

tirety of the nonlinear operator due to the role the resonances play in the oscillatory stiffness of the system.

By describing the error of the coarse solver in terms of the interplay between the average over the rapid oscillations and the time-stepping, we show that the method converges for finite scale separation, significantly extending the domain of applicability for this method.

We have shown here that this method is convergent across a wide range of scale separation, which is an improvement on the prior result [14] which held only in the small-ε limit.

Appendix A. Proof of the error bounds for the time-averaged equation. We now prove the following proposition (see also Lemma 3.2.9 in [25]).

PROPOSITION A.1. *Consider*

$$\frac{d\mathbf{v}}{d\tau}(\tau) = \varepsilon \mathcal{N}(\tau, \mathbf{v}(\tau)), \quad 0 \le \tau \le \frac{1}{\varepsilon},$$

and its averaged version

$$\frac{d\overline{\mathbf{v}}}{dt}(t) = \varepsilon \overline{\mathcal{N}}_\eta(\tau, \overline{\mathbf{v}}(\tau)), \quad 0 \le \tau \le \frac{1}{\varepsilon},$$

where

$$\overline{\mathcal{N}}_\eta(\tau, \overline{\mathbf{v}}(\tau)) = \frac{1}{\eta}\int_0^\eta \rho\left(\frac{s}{\eta}\right) \mathcal{N}(\tau + s, \overline{\mathbf{v}}(\tau))\, ds. \tag{A.1}$$

Assume there is a constant λ such that

$$\max_{0\le t,t_1,t_2\le\varepsilon^{-1}} \|\mathcal{N}(\tau_0, \mathbf{v}(\tau_1)) - \mathcal{N}(\tau_0, \mathbf{v}(\tau_2))\| \le \lambda \|\mathbf{v}(\tau_1) - \mathbf{v}(\tau_2)\|, \tag{A.2}$$

and define

$$M = \max_{0\le\tau\le\varepsilon^{-1}} \|\mathcal{N}(\tau, \mathbf{v}(\tau))\|, \quad C_0 = \int_0^1 \rho(s)\, s ds. \tag{A.3}$$

Then

$$\|\mathbf{v}(\tau) - \overline{\mathbf{v}}(\tau)\| = C_0(1+\lambda) M \eta \varepsilon e^{\varepsilon\lambda\tau}, \quad 0 \le \tau \le \frac{1}{\varepsilon}.$$

We now prove Proposition A.1. We first need the following lemma.

LEMMA A.2. *Suppose that $\boldsymbol{\phi}(t)$ is Lipschitz-continuous with Lipschitz constant λ:*

$$\|\boldsymbol{\phi}(t) - \boldsymbol{\phi}(s+t)\| \le \lambda s \quad \text{for all } s, t \ge 0.$$

Then

$$\left\|\boldsymbol{\phi}(t) - \boldsymbol{\phi}_\eta(t)\right\| \le C_0 \lambda \eta,$$

where

$$C_0 = \int_0^1 \rho(s)\, s ds. \tag{A.4}$$

Proof. Using that

$$\frac{1}{\eta}\int_0^\eta \rho\left(\frac{s}{\eta}\right) ds = \int_0^1 \rho(s)\, ds = 1,$$

we have that

$$\begin{aligned}\left\|\boldsymbol{\phi}(t)-\boldsymbol{\phi}_\eta(t)\right\| &= \left\|\boldsymbol{\phi}(t)-\frac{1}{\eta}\int_0^\eta \rho\left(\frac{s}{\eta}\right)\boldsymbol{\phi}(s+t)\,ds\right\| \\ &= \frac{1}{\eta}\int_0^\eta \rho\left(\frac{s}{\eta}\right)\left\|\boldsymbol{\phi}(t)-\boldsymbol{\phi}(s+t)\right\| ds \\ &\le \frac{1}{\eta}\int_0^\eta \rho\left(\frac{s}{\eta}\right) s\lambda ds \\ &= \eta\lambda\int_0^1 \rho(s)\, s ds. \end{aligned}$$

∎

LEMMA A.3. *Consider*

$$\frac{d\mathbf{v}}{dt}(t)=\varepsilon\mathcal{N}(t,\mathbf{v}(t)),\ \ 0\le t\le\varepsilon^{-1},$$

with $\mathcal{N}$ continuous in each argument. Define

$$\boldsymbol{\phi}(t)=\int_0^t \mathcal{N}(\tau,\mathbf{v}(\tau))\,d\tau, \tag{A.5}$$

$$\mathcal{N}_\eta(t,\mathbf{x})=\frac{1}{\eta}\int_0^\eta \rho\left(\frac{s}{\eta}\right)\mathcal{N}(t+s,\mathbf{x})\,ds,$$

and

$$\boldsymbol{\phi}_\eta(t)=\frac{1}{\eta}\int_0^\eta \rho\left(\frac{s}{\eta}\right)\boldsymbol{\phi}(s+t)\,ds.$$

Then

$$\left\|\boldsymbol{\phi}_\eta(t)-\int_0^t \mathcal{N}_\eta(\tau,\mathbf{v}(\tau))\,d\tau\right\|\le C_0(1+\lambda)M\eta.$$

Here $\mathcal{N}_\eta(\tau,\mathbf{v}(\tau))$ is defined via (A.1)*, C_0 is defined by* (A.4)*, and λ and M are defined by* (A.2) *and* (A.3)*.*

Proof. We calculate that

$$\begin{aligned}\boldsymbol{\phi}_\eta(t) &= \frac{1}{\eta}\int_0^\eta \rho\left(\frac{s}{\eta}\right)\boldsymbol{\phi}(s+t)\,ds \\ &= \frac{1}{\eta}\int_0^\eta \rho\left(\frac{s}{\eta}\right)\left(\int_0^{t+s}\mathcal{N}(\tau,\mathbf{v}(\tau))\,d\tau\right)ds \\ &= \frac{1}{\eta}\int_0^\eta \rho\left(\frac{s}{\eta}\right)\left(\int_s^{t+s}\mathcal{N}(\tau,\mathbf{v}(\tau))\,d\tau\right)ds + R_1 \\ &= \frac{1}{\eta}\int_0^\eta \rho\left(\frac{s}{\eta}\right)\left(\int_0^{t}\mathcal{N}(\tau+s,\mathbf{v}(\tau+s))\,d\tau\right)ds + R_1 \\ &= \frac{1}{\eta}\int_0^\eta \rho\left(\frac{s}{\eta}\right)\left(\int_0^{t}\mathcal{N}(\tau+s,\mathbf{v}(\tau))\,d\tau\right)ds + R_1 + R_2 \\ &= \int_0^t\left(\frac{1}{\eta}\int_0^\eta \rho\left(\frac{s}{\eta}\right)\mathcal{N}(\tau+s,\mathbf{v}(\tau))\,ds\right)d\tau + R_1 + R_2 \\ &= \int_0^t \mathcal{N}_\eta(\tau,\mathbf{v}(\tau))\,d\tau + R_1 + R_2. \end{aligned}$$

Here the remainder term R_1,

$$R_1 = \frac{1}{\eta}\int_0^\eta \rho\left(\frac{s}{\eta}\right)\left(\int_0^s \mathcal{N}\left(\tau, \mathbf{v}\left(\tau\right)\right) d\tau\right) ds,$$

is bounded as follows:

$$\begin{aligned}
\|R_1\| &= \frac{1}{\eta}\int_0^\eta \rho\left(\frac{s}{\eta}\right)\int_0^s \|\mathcal{N}\left(\tau, \mathbf{v}\left(\tau\right)\right)\| d\tau ds \\
&\leq \frac{1}{\eta}\int_0^\eta \rho\left(\frac{s}{\eta}\right)\int_0^s M d\tau ds \\
&= M\frac{1}{\eta}\int_0^\eta \rho\left(\frac{s}{\eta}\right) s ds \\
&= M\eta\int_0^1 \rho\left(s\right) s ds \\
&= C_0 M\eta.
\end{aligned}$$

Similarly, the remainder term R_2,

$$R_2\left(t\right) = \frac{1}{\eta}\int_0^\eta \rho\left(\frac{s}{\eta}\right)\int_0^t \left(\mathcal{N}\left(\tau+s, \mathbf{v}\left(\tau+s\right)\right) - \mathcal{N}\left(\tau+s, \mathbf{v}\left(\tau\right)\right)\right) d\tau ds,$$

is bounded by

$$\begin{aligned}
\|R_2\| &= \frac{1}{\eta}\int_0^\eta \rho\left(\frac{s}{\eta}\right)\int_0^t \|\mathcal{N}\left(\tau+s, \mathbf{v}\left(\tau+s\right)\right) - \mathcal{N}\left(\tau+s, \mathbf{v}\left(\tau\right)\right)\| d\tau ds \\
&\leq \frac{1}{\eta}\lambda\int_0^\eta \rho\left(\frac{s}{\eta}\right)\int_0^t \|\mathbf{v}\left(\tau+s\right) - \mathbf{v}\left(\tau\right)\| d\tau ds \\
&= \frac{1}{\eta}\lambda\int_0^\eta \rho\left(\frac{s}{\eta}\right)\int_0^t \left\|\int_\tau^{s+\tau} \frac{d\mathbf{v}}{d\sigma}\left(\sigma\right) d\sigma\right\| d\tau ds \\
&= \frac{1}{\eta}\lambda\int_0^\eta \rho\left(\frac{s}{\eta}\right)\int_0^t \left\|\int_\tau^{s+\tau} \varepsilon\mathcal{N}\left(\sigma, \mathbf{v}\left(\sigma\right)\right) d\sigma\right\| d\tau ds \\
&\leq \frac{1}{\eta}\varepsilon\lambda\int_0^\eta \rho\left(\frac{s}{\eta}\right)\int_0^t \int_\tau^{s+\tau} \|\mathcal{N}\left(\sigma, \mathbf{v}\left(\sigma\right)\right)\| d\sigma d\tau ds \\
&\leq \frac{1}{\eta}\varepsilon\lambda M\int_0^\eta \rho\left(\frac{s}{\eta}\right)\int_0^t \left(\left(\int_\tau^{s+\tau} d\sigma\right) d\tau\right) ds \\
&= \frac{1}{\eta}\varepsilon\lambda M\int_0^\eta \rho\left(\frac{s}{\eta}\right) s\left(\int_0^t d\tau\right) ds \\
&= \frac{1}{\eta}\varepsilon\lambda M t\int_0^\eta \rho\left(\frac{s}{\eta}\right) s ds \\
&= \left(C_0\lambda M\eta\right)\left(\varepsilon t\right) \\
&\leq C_0\lambda M\eta.
\end{aligned}$$

The second inequality uses the Lipschitz bound (A.2), the fourth inequality uses the bound (A.3), the last equality uses the definition (A.4) for C_0, and the last inequality uses that $0 \leq t \leq \varepsilon^{-1}$. ∎

We now prove Proposition A.1.

Proof. First note that

$$\int_0^t \mathcal{N}\left(\tau, \mathbf{v}\left(\tau\right)\right) d\tau = \int_0^t \mathcal{N}_\eta\left(\tau, \mathbf{v}\left(\tau\right)\right) d\tau + E_0,$$

where

$$\left\|E_0\right\| \leq C_0\left(1+\lambda\right) M\eta. \tag{A.6}$$

The bound (A.6) follows from the definition (A.5) for $\boldsymbol{\phi}\left(t\right)$ and the inequality

$$\begin{aligned}
&\left\|\int_0^t \mathcal{N}\left(\tau, \mathbf{v}\left(\tau\right)\right) d\tau - \int_0^t \mathcal{N}_\eta\left(\tau, \mathbf{v}\left(\tau\right)\right) d\tau\right\| \\
&\quad = \left\|\boldsymbol{\phi}\left(t\right) - \int_0^t \mathcal{N}_\eta\left(\tau, \mathbf{v}\left(\tau\right)\right) d\tau\right\| \\
&\quad \leq \left\|\boldsymbol{\phi}\left(t\right) - \boldsymbol{\phi}_\eta\left(t\right)\right\| + \left\|\boldsymbol{\phi}_\eta\left(t\right) - \int_0^t \mathcal{N}_\eta\left(\tau, \mathbf{v}\left(\tau\right)\right) d\tau\right\|.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\mathbf{v}\left(t\right) &= \mathbf{v}\left(0\right) + \varepsilon \int_0^t \mathcal{N}\left(\tau, \mathbf{v}\left(\tau\right)\right) d\tau \\
&= \mathbf{v}\left(0\right) + \varepsilon \int_0^t \mathcal{N}_\eta\left(\tau, \mathbf{v}\left(\tau\right)\right) d\tau + \varepsilon E_0.
\end{aligned} \tag{A.7}$$

Using (A.7) and

$$\overline{\mathbf{v}}\left(t\right) = \mathbf{v}\left(0\right) + \varepsilon \int_0^t \mathcal{N}_\eta\left(\tau, \overline{\mathbf{v}}\left(t\right)\right) d\tau,$$

we have the bound

$$\begin{aligned}
\left\|\mathbf{v}\left(t\right) - \overline{\mathbf{v}}\left(t\right)\right\| &\leq \varepsilon \int_0^t \left\|\mathcal{N}_\eta\left(\tau, \mathbf{v}\left(\tau\right)\right) - \mathcal{N}_\eta\left(\tau, \overline{\mathbf{v}}\left(t\right)\right)\right\| d\tau + \left\|\varepsilon E_0\right\| \\
&\leq \varepsilon\lambda \int_0^t \left\|\mathbf{v}\left(\tau\right) - \overline{\mathbf{v}}\left(t\right)\right\| d\tau + C_0\left(1+\lambda\right) M\eta\varepsilon,
\end{aligned}$$

where we used the bound (A.6) for $\left\|\varepsilon E_0\right\|$. Finally, by Gronwall's inequality,

$$\left\|\mathbf{v}\left(t\right) - \overline{\mathbf{v}}\left(t\right)\right\| \leq C_0\left(1+\lambda\right) M\eta\varepsilon e^{\varepsilon\lambda\tau}.$$

∎

Acknowledgment. We would like to thank the two anonymous reviewers whose careful reading and relevant remarks contributed to improving this manuscript.

REFERENCES

[1] G. Ariel, S. J. Kim, and R. Tsai, *Parareal multiscale methods for highly oscillatory dynamical systems*, SIAM J. Sci. Comput., 38 (2016), pp. A3540–A3564, https://doi.org/10.1137/15M1011044.

[2] C. Audouze, M. Massot, and S. Volz, *Symplectic Multi-Time Step Parareal Algorithms Applied to Molecular Dynamics*, 2009, https://hal.archives-ouvertes.fr/hal-00358459.

[3] L. Baffico, S. Bernard, Y. Maday, G. Turinici, and G. Zérah, *Parallel-in-time molecular-dynamics simulations*, Phys. Rev. E, 66 (2002), 057701, https://doi.org/10.1103/PhysRevE.66.057701.

[4] G. Bal and Y. Maday, *A "parareal" time discretization for non-linear PDE's with application to the pricing of an American put*, in Recent Developments in Domain Decomposition Methods, L. F. Pavarino and A. Toselli, eds., Lect. Notes Comput. Sci. Eng. 23, Springer, Berlin, Heidelberg, 2002, pp. 189–202, https://doi.org/10.1007/978-3-642-56118-4_12.

[5] G. Bal and Q. Wu, *Symplectic parareal*, in Domain Decomposition Methods in Science and Engineering XVII, U. Langer, M. Discacciati, D. E. Keyes, O. B. Widlund, and W. Zulehner, eds., Lect. Notes Comput. Sci. Eng. 60, Springer, Berlin, Heidelberg, 2008, pp. 401–408.

[6] N. N. Bogoliubov and Y. A. Mitropolsky, *Asymptotic Methods in the Theory of Non-linear Oscillations*, Gordon and Breach, New York, 1961.

[7] T. Davies, A. Staniforth, N. Wood, and J. Thuburn, *Validity of anelastic and other equation sets as inferred from normal-mode analysis*, Q. J. R. Meteorol. Soc., 129 (2003), pp. 2761–2775, https://doi.org/10.1256/qj.02.1951.

[8] P. F. Embid and A. J. Majda, *Averaging over fast gravity waves for geophysical flows with arbitrary potential vorticity*, Comm. Partial Differential Equations, 21 (1996), pp. 619–658.

[9] C. Farhat and M. Chandesris, *Time-decomposed parallel time-integrators: Theory and feasibility studies for fluid, structure, and fluid-structure applications*, Internat. J. Numer. Methods Engrg., 58 (2003), pp. 1397–1434, https://doi.org/10.1002/nme.860.

[10] P. F. Fischer, F. Hecht, and Y. Maday, *A parareal in time semi-implicit approximation of the Navier-Stokes equations*, in Domain Decomposition Methods in Science and Engineering, T. J. Barth, M. Griebel, D. E. Keyes, R. M. Nieminen, D. Roose, T. Schlick, R. Kornhuber, R. Hoppe, J. Priaux, O. Pironneau, O. Widlund, and J. Xu, eds., Lect. Notes Comput. Sci. Eng. 40, Springer, Berlin, Heidelberg, 2005, pp. 433–440, https://doi.org/10.1007/3-540-26825-1_44.

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

[12] M. J. Gander and S. Vandewalle, *Analysis of the parareal time-parallel time-integration method*, SIAM J. Sci. Comput, 29 (2007), pp. 556–578, https://doi.org/10.1137/05064607X.

[13] I. Garrido, M. S. Espedal, and G. E. Fladmark, *A convergent algorithm for time parallelization applied to reservoir simulation*, in Domain Decomposition Methods in Science and Engineering, T. J. Barth, M. Griebel, D. E. Keyes, R. M. Nieminen, D. Roose, T. Schlick, R. Kornhuber, R. Hoppe, J. Périaux, O. Pironneau, O. Widlund, and J. Xu, eds., Lect. Notes Comput. Sci. Eng. 40, Springer, Berlin, Heidelberg, 2005, pp. 469–476, https://doi.org/10.1007/3-540-26825-1_48.

[14] T. Haut and B. Wingate, *An asymptotic parallel-in-time method for highly oscillatory PDEs*, SIAM J. Sci. Comput., 36 (2014), pp. A693–A713, https://doi.org/10.1137/130914577.

[15] T. S. Haut, T. Babb, P. G. Martinsson, and B. A. Wingate, *A high-order time-parallel scheme for solving wave propagation problems via the direct construction of an approximate time-evolution operator*, IMA J. Numer. Anal., 36 (2016), pp. 688–716.

[16] L. He, *The reduced basis technique as a coarse solver for parareal in time simulations*, J. Comput. Math., 28 (2010), pp. 676–692.

[17] S. G. Johnson, *Saddle-point integration of C_∞ "bump" functions*, preprint, https://arxiv.org/abs/1508.04376, 2015.

[18] S. Klainerman and A. Majda, *Singular limits of quasilinear hyperbolic systems with large parameters and the incompressible limit of compressible fluids*, Comm. Pure Appl. Math., 34 (1981), pp. 481–524.

[19] B. N. Lawrence, M. Rezny, R. Budich, P. Bauer, J. Behrens, M. Carter, W. Deconinck, R. Ford, C. Maynard, S. Mullerworth, C. Osuna, A. Porter, K. Serradell, S. Valcke, N. Wedi, and S. Wilson, *Crossing the chasm: How to develop weather and climate models for next generation computers?*, Geosci. Model Dev., 11 (2018), pp. 1799–1821, https://doi.org/10.5194/gmd-11-1799-2018.

[20] F. Legoll, T. Lelièvre, and G. Samaey, *A micro-macro parareal algorithm: Application to singularly perturbed ordinary differential equations*, SIAM J. Sci. Comput., 35 (2013), pp. A1951–A1986, https://doi.org/10.1137/120872681.

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

[22] Y. Maday and G. Turinici, *Parallel in time algorithms for quantum control: Parareal time discretization scheme*, Int. J. Quantum Chem., 93 (2003), pp. 223–228.

[23] A. Majda, *Introduction to PDEs and Waves for the Atmosphere and Ocean*, Courant Lect. Notes 9, New York University, Courant Institute of Mathematical Sciences, New York, American Mathematical Society, Providence, RI, 2003.
[24] L. M. Polvani, J. C. McWilliams, M. A. Spall, and R. Ford, *The coherent structures of shallow-water turbulence: Deformation-radius effects, cyclone/anticyclone asymmetry and gravity-wave generation*, Chaos, 4 (1994), pp. 177–186, https://doi.org/10.1063/1.166002.
[25] J. A. Sanders, F. Verhulst, and J. Murdock, *Averaging Methods in Nonlinear Dynamical Systems*, Appl. Math. Sci. 59, 2nd ed., Springer, New York, 2007.
[26] S. Schochet, *Fast singular limits of hyperbolic PDEs*, J. Differential Equations, 114 (1994), pp. 476–512.