A Unification and Investigation of Rational Approximation of Exponential Integration Methods

Martin Schreiber, Jed Brown

A GENERALIZED RATIONAL APPROXIMATION OF EXPONENTIAL INTEGRATION (REXI) FOR MASSIVELY PARALLEL TIME INTEGRATION

MARTIN SCHREIBER* AND JED BROWN†

Abstract. Solving partial differential equations (PDEs) is one of the most traditional tasks in scientific computing. In this work, we consider numerical solutions of initial value problems (IVPs) problems partly or entirely given by linear PDEs and how to compute solutions with a method we refer to as rational approximation of exponential integration (REXI). REXI replaces a typically sequential timestepping method with a sum of rational terms, leading to the possibility to parallelize over this sum. Hence, this method can potentially exploit additional degrees of parallelization for scaling problems limited in their spatial scalability to large-scale supercomputers.

The main contribution of this work lies in developing the "unified REXI" in which we show algebraic equivalence to other methods developed up to five decades ago. Such methods cover, e.g., diagonalization of the Butcher table for implicit Runge-Kutta methods, Cauchy-contour integration-based methods, and direct approximations. To our best knowledge, this is the first time of such a comparison and deep investigation of all these methods.

Finally, we will show the applicability of REXI to the nonlinear shallow-water equations on the rotating sphere, including HPC results. While previous REXI studies have focused on exposing more parallelism to enable faster time to solution, we also consider efficiency at prescribed accuracy and find that diagonalized Gauss Runge-Kutta methods (formulated as REXI) are compelling highly efficient methods.

Key words. Exponential integrators, rational approximation, parallel-in-time, Cauchy contour, Butcher table, diagonalization

AMS subject classifications.

1. Introduction.

Time integration of IVPs is one of the most traditional tasks in scientific computing, having seen two centuries of research. The IVPs we are interested in are given entirely or partly by linear autonomous PDEs, which are ubiquitous in applications ranging from daily weather forecasting [11] to full waveform inversion [43]. Integration of such systems is sequential in time using conventional methods such as explicit and diagonally implicit Runge-Kutta [29, 21]: Without special structure [20], the state at each stage is necessary to compute the next stage, either explicitly or implicitly. The time step size is typically limited by stability and/or accuracy requirements and the method is purely sequential in the time dimension.

With the desire to solve PDEs with ever-higher resolutions, the demands on high-performance computers (HPC) have increased. The steady and ongoing increase in

*Univ. Grenoble Alpes / Laboratoire Jean Kuntzmann / Inria, Grenoble, France (martin.schreiber@univ-grenoble-alpes.fr), Department of Informatics, Technical University of Munich, Germany (martin.schreiber@tum.de), https://www.martin-schreiber.info

†Department of Computer Science, University of Colorado, Boulder, USA (jed@jedbrown.org, https://jedbrown.org)

Submitted to the editors DATE.

Funding: This project has received funding from the Federal Ministry of Education and Research and the European High-Performance Computing Joint Undertaking (JU) under grant agreement No 955701. The JU receives support from the European Union's Horizon 2020 research and innovation programme and Belgium, France, Germany, Switzerland.
Martin Schreiber gratefully acknowledges KONWIHR funding as part of the project "Parallel in Time Integration with Rational Approximations targeting Weather and Climate Simulations".
Jed Brown acknowledges support from the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, applied mathematics program.

HPC performance is provided almost exclusively by increased parallelism; increasing resolution in space (spatial scalability) can be solved in the same amount of time per time step, but the wallclock time to simulate for a fixed physical duration increases due to the increasing number of time steps to satisfy the Courant-Friedrichs-Lewy (CFL) constraint [9] for transport phenomena. Consequently, refinement to increase accuracy on a transient physical problem is always a scaling challenge, and many applications are unable to increase spatial resolution without sacrificing external timelines such as IPCC assessment reports [6] or design/manufacturing timelines. Parallelism in the time dimension is seen as an opportunity to utilize greater parallelism to meet stringent simulation timelines. The Rational Approximation of Exponential Integration (REXI) family of methods, which we briefly explain next, are a promising candidate for hyperbolic PDEs. Consider a linear autonomous PDE given by $\frac{\partial \mathcal{U}(t)}{\partial t} = \mathcal{L}\mathcal{U}(t)$ with $\mathcal{U}(t)$ the current state and $\mathcal{L}$ a linear differential operator. Discretizing state variable and operator in space leads to

$$\frac{\partial U(t)}{\partial t} = LU(t) \tag{1.1}$$

with L the discrete linear operator and $U(t)$ the discrete state variables at time t. Solving such IVPs have been intensively studied over the last decades with various approaches, and one of the direct methods is the application of an exponential integration

$$U(t + \Delta t) = \exp(\Delta t L) U(t) \tag{1.2}$$

with the solution $U(t)$ at time t. We want to emphasize that no time discretization has been introduced and that the only approximations are related to space. The REXI method exploits the feature that $\exp(\Delta t L)$ only needs to be approximated within a spectrum related to time step size Δt and the spectrum of L. In the present work, we can express a variety of different time integration methods by what we refer to as the "unified REXI" formulation given by

$$U(t + \Delta t) \approx \gamma U(t) + \underbrace{\sum_{n=1}^{N} \beta_n (\Delta t L - \alpha_n)^{-1} U(t)}_{\text{Parallelization}} \tag{1.3}$$

with the time step size given by Δt, and (typically) complex valued REXI coefficients α_n, β_n and real-valued γ is a new generalization that we will use in the following sections. The remainder of this work investigates different ways to infer these REXI coefficients and their relation to a class of Runge-Kutta methods. Based on this, we will study their numerical properties in the linear and nonlinear context.

2. Related Work.

2.1. Exponential integration.

Exponential integration methods are formulated for nonlinear systems factored as

$$\frac{\partial U(t)}{\partial t} = LU(t) + N(U(t)), \tag{2.1}$$

where the linear part L is intended to capture the "fast" dynamics and N is the remaining nonlinear part. An exact ansatz for advancing this split equation over a

finite time interval is given by

$$U(t+\Delta t) = \exp(\Delta t L)U(t) + \int_0^{\Delta t} \exp((\Delta t - \tau)L)N(U(t+\tau))d\tau. \tag{2.2}$$

In this form, the linear parts are integrated precisely by an exponential function, hence overcoming potential stiffness challenges caused by the linear parts. Due to this advantageous property, the interest in these exponential integrators has steadily increased over the last decades (see, e.g., [25, 19]) where various approaches have been taken to approximate the integral of the nonlinearities. One of the most commonly known approximations of the integral is, e.g., given by (see [10])

$$U(t+\Delta t) = \varphi_0(\Delta t L)U(t) + \Delta t \varphi_1(\Delta t L)N(U(t)) \tag{2.3}$$

where we used the notations $\varphi_0(Z) = e^Z$ and $\varphi_1(Z) = \frac{e^Z - I}{Z}$. We skip further examples for discretized exponential integrator formulations and only like to point out the φ functions to be omnipresent in higher-order exponential integration methods, which is generally given, e.g., by

$$\varphi_{i+1}(Z) = (\varphi_i(Z) - \varphi_i(0))\, Z^{-1} \quad \text{for } i \geq 0. \tag{2.4}$$

An investigation of all different varieties of discretizations of exponential integrators incorporating the nonlinearities is beyond the scope of this work, and we continue on the linear parts. These linear parts can be either given by full linear PDEs or by time integrating only a part of linear PDEs where the underlying requirement of time integration results in problems of the form $U(t+\Delta t) = \varphi_0(\Delta t L)U(t) = \exp(\Delta t L)U(t)$. In contrast to state-of-the-art time integration methods, which are used in operational codes, exponential integrators for linear operators avoid any time-discretization errors. However, the computational complexity can be tremendous and triggered the development of various ways to tackle this challenge [25]. We will briefly summarize the ones recently researched, namely based on Krylov subspaces and REXI.

The exponential can be approximated using Krylov subspace solvers (see [26, 39, 40, 8]) where we see polynomial approximations (e.g., based on Chebyshev) as a subclass of them. The advantage of such methods is their simplicity – assuming the Krylov solver framework given – since only vector multiplications with the linear operator are required. However, the potential drawbacks of Krylov subspace solvers are their inherent property of sequential iterations over the Krylov subspace, hence not providing ways to exploit additional degrees of parallelization. An alternative is to use the REXI method, which will be discussed in the next section in further detail.

2.2. REXI. The particular way to evaluate the φ functions in the present work is strongly related to Padé approximations, which can be used as a first instance to approximate the φ functions. This approximation is most naturally related to how all Runge-Kutta formulations, e.g., based on the Butcher table, can be formulated for linear autonomous operators. However, higher-order polynomials in the denominator also make the development of solvers for such Padé approximations more challenging, and a partial fraction decomposition can be used. This well-known decomposition transforms a higher-order Padé approximation into a sum of lower-order terms, which can eventually be used to develop solvers parallelized over all terms. Although not being explained in the context of Padé approximations, REXI methods using the partial fraction decomposed form have been developed in different contexts. They

can be interpreted as a Padé approximation, which is why they are mentioned in this context.

In what follows, we will provide an overview of different methods, which can all be phrased in REXI form. One of the earliest REXI formulations for hyperbolic PDE time integrators is related to the Laplace transformation (cf. [23, 7]). Here, the PDE is transformed with the Laplace operator, where the backward transform is conducted with a Cauchy Contour integral. This transformation can be again related to an exponential integration scheme, namely to the Cauchy Contour method mentioned below, see also [41]. More recently, time integration based on Laplace transformations with a circle-based Cauchy contour integration have been more intensively studied in [28] with ODEs, in particular filtering properties. However, it needed a more extensive (community) effort to develop other, e.g., higher-order methods around them, as has been extensively the case for exponential integration methods. This lack of advancements with the REXI Laplace transform is also why we only concentrate on exponential integration-based formulations. We want to point out that the same approaches could also be taken from the Laplace transform perspective.

Another way to infer REXI coefficients originates from the REXI method based on Gaussian basis functions originally developed in [18], which only targets purely oscillatory problems (hence L has only imaginary or zero eigenvalues). This method also showed excellent properties regarding the wallclock-time vs. error for the linear shallow-water equations on the plane (see [33]) and on the rotating sphere [32, 34].

Although initially developed for analytical reasons, the Cauchy contour integration method can indeed be used for REXI time integration. As pointed out above, one of the first times this has been used as a REXI-like method was with the Laplace transformations. However, exponential formulations (see Eq. 2.2) provide a more direct and substantial established way to integrate in time. Here, the property of φ_i being an analytical function plays a fundamental role in the Cauchy contour integration method as well as the exponentially fast converging trapezoidal rule to approximate the contour [41]. This method has already been used in different works: The approximation of $\varphi_i(x)$ evaluations on scalar values has been used in various works to overcome singularities of $\varphi_{i>0}$ singularities at the origin, see, e.g., [5]. It has been used mainly for parabolic problems [35], also pointing out the potential of parallelization for the first time, as well as using a Carathéodory-Fejér method [31]. Regarding real applications, it was applied to nonlinear shallow-water equations on the rotating sphere [34], providing improved wallclock time-to-solution by using an enlarged and shifted contour to avoid numerical cancellation errors.

2.3. Parallel-in-time. Overcoming the wallclock time limitations of simulations, which cannot be accomplished by any further increase of parallelization in the spatial dimension, is the main focus of the parallel-in-time algorithms. Here, two different types of approaches exist: (a) minimally-invasive methods that take existing time integration methods and incorporate them into an iterative-in-time correction scheme (see, e.g., Parareal [22] and PFASST [24]); and (b) invasive methods that replace an existing time stepping with one that works entirely differently. Very often, one likes to use a combination of these approaches to enhance the convergence speed of the correction scheme in time. REXI is an invasive parallel-in-time algorithm (see [33]) since it requires efficient complex-valued solvers for each REXI term.

3. Unified REXI formulation. We start directly with the REXI formulation which will provide a standard fundament for the different variants to infer REXI coefficients. Given a discrete linear operator L, we can use an eigendecomposition

$L = Q\Lambda Q^{-1}$ with the eigenvectors stored in the columns of Q and the eigenvalues placed correspondingly on the diagonal of Λ.

$$\frac{\partial U(t)}{\partial t} = LU(t) = Q\Lambda Q^{-1}U(t) \tag{3.1}$$

where Q and Λ are the matrices with the eigenvectors and eigenvalues on the diagonal, respectively. In terms of the characteristic variable $u = Q^{-1}U$ and due to diagonal-only Λ, we get independent equations of the form

$$\frac{\partial u_i(t)}{\partial t} = \lambda_i u_i(t)$$

with λ_i the individual Eigenvalues on the diagonal of Λ. In characteristic variables, the unified REXI formulation (1.3) becomes

$$u_i(t+\Delta t) \approx \gamma u_i(t) + \sum_{n=1}^{N} \beta_n (\Delta t \lambda_i - I\alpha_n)^{-1} u_i(t). \tag{3.2}$$

Since each component u_i is decoupled, we can freely drop the subscript. For the purpose of time integration, the linear operator L is completely described by its eigenvalues λ, where imaginary components $\Im(\lambda)$ represent oscillation and negative real values $\Re(\lambda) < 0$ describe a diffusive/damping behavior. Note that substituting $\lambda = 1, t = 0, \Delta t = x, u(0) = 1$ in (3.2) yields $\exp(x) = \gamma + \sum_n \beta_n (x - \alpha_n)^{-1}$, which provides intuition as a sum of rational functions.

3.1. Exploiting symmetry of coefficients. We note that it is possible to reduce the workload by a factor of two for real-valued operators L when the poles α consist of complex conjugate pairs (see, e.g., [23, 18]). This optimization does not change the relative performance of the methods we consider here, so for simplicity, we do not apply it.

3.2. REXI-derived higher-order φ forms. Particular higher-order exponential time integrators such as (2.3) require evaluations of higher-order $\varphi_{i|i>0}$. REXI coefficients for these functions are so far computed with methods tailored to them, see [18, 32]. We briefly present an new alternative way to compute them which is easily applicable. Given REXI coefficients for

$$\varphi_i(x) \approx \gamma + \sum_n \beta_n (x - \alpha_n)^{-1}$$

we can compute higher-order REXI approximations with

$$\varphi_{i+1}(x) = \frac{\varphi_i(x) - \varphi_i(0)}{x} = \frac{\gamma + \sum_n \frac{\beta_n}{x-\alpha_n} - \varphi_i(0)}{x}$$

$$= \sum_n \left(\frac{\beta_n}{\alpha_n (x - \alpha_n)} \right) + \frac{1}{x} \left(\underbrace{\sum_n \left(-\frac{\beta_n}{\alpha_n} \right) + \gamma - \varphi_i(0)}_{=0} \right) = \sum_n \frac{\frac{\beta_n}{\alpha_n}}{x - \alpha_n}.$$

The cancellation of the terms is a consequence of the stationary modes which require $\sum_n \left(-\frac{\beta_n}{\alpha_n} \right) + \gamma = \varphi_i(0)$. Note that this leads to different coefficients compared to tailored computations.

3.3. Linear solvers for REXI terms. Efficient solvers are required for each REXI term. Over the last decades, this efficiency aspect turned out to be a very challenging task. E.g., in the context of shallow-water equations, this results in the original Helmholtz problem (rather than a backward Euler time step) where it is known that no off-the-shelf solvers such as GMRES and multigrid methods work in a highly-scalable way (see, e.g., [12]). This is ongoing research, and in the present work, we are using solvers developed in spherical harmonics formulations, hence a solver tailored particularly to one PDE problem.

4. REXI methods. These sections cover various ways to infer REXI coefficients, which represent, from our point of view, the most interesting cases. The goal is not to show all methods in great detail but their fundamental properties.

Although we present methods in characteristic form (3.2), the proposed methods also hold in system form (1.3). In the following, we will use the error

$$e(z) = \left| \gamma + \sum_n \beta_n (z - \alpha_n)^{-1} - \exp(z) \right| \tag{4.1}$$

to compute the deviation from $\varphi_0(z) = \exp(z)$ with $z = \lambda \Delta t$ denoting the point on the complex plane to evaluate. Since approximating diffusive problems is relatively straightforward, we focus on purely oscillatory problems with $\lambda \in i\,\mathbb{R}$. The REXI methods we consider have complex-conjugate poles α, thus $e(z) = e(\bar{z})$ and so we only plot errors for $\Im[z] \geq 0$.

4.1. B-REXI: Butcher/Bickart. A Butcher table [2] provides a canonical representation of s-stage Runge-Kutta methods [29, 21] in terms of a matrix $A \in \mathbb{R}^{s \times s}$ and completion vector $b \in \mathbb{R}^s$, with $c = A\mathbf{1}$ determining the abscissa (which we will see are related to REXI poles and $\mathbf{1}$ is a column vector of ones). The coefficients are selected to achieve the desired order of accuracy and stability properties as well as solution procedure, such as explicit, diagonally implicit, and fully implicit.

For fully nonlinear and non-autonomous ODEs $\frac{\partial u}{\partial t} = f(t, u)$, a Runge-Kutta method in Butcher form requires solving a system of stage equations

$$y_s = u_n + \Delta t \sum_{j=1}^{S} A_{sj} f(t + c_j \Delta t, y_j), \quad i = 1, ..., S \tag{4.2}$$

and evaluating the completion formula $u_{n+1} = u_n + \Delta t \sum_{j=1}^{S} b_j f(t + c_j \Delta t, y_j)$. Here, Δt is the time step size, and $\mathbf{y} = \{y_j\}_{j=1}^S$ is the vector of stage solutions. For linear autonomous equations, we can choose characteristic variables, in which case $f(t, u) = \lambda u$, and the stage equations (4.2) reduce to $\mathbf{y} = \mathbf{1}u + \Delta t \lambda A \mathbf{y}$ and

$$u_{n+1} = \underbrace{\left[1 + \Delta t \lambda b^T (I - \Delta t \lambda A)^{-1} \mathbf{1}\right]}_{R(\Delta t \lambda)} u_n, \tag{4.3}$$

where we have identified the stability function $R(z) \approx \exp(z)$.

4.1.1. Derivation. We now show that unified REXI is algebraically equivalent to Runge-Kutta methods with a diagonal Butcher matrix A, starting with a decomposition inspired by the solution method developed independently by [3, 1]. Given an eigendecomposition $A = EDE^{-1}$ (which exists for the collocation methods we will consider [16]), we can rewrite (4.3) as

$$u_{n+1} = \left[1 + \Delta t \lambda b^T E (I - \Delta t \lambda D)^{-1} E^{-1} \mathbf{1}\right] u_n. \tag{4.4}$$

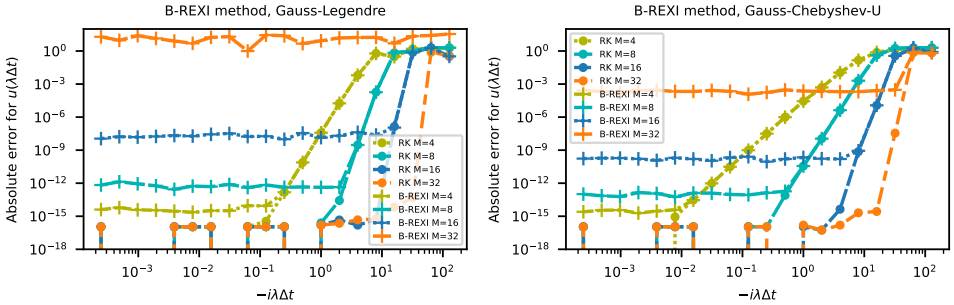

FIG. 4.1. *Error studies for the B-REXI method with (a) Gauss-Legendre and (b) Chebyshev quadrature points for the error given in Eq. (4.1). Each color refers to the same number of stages, marked to B-REXI or RK form. The non-diagonalized version provides significantly better results compared to the diagonalized version. In particular, results with B-REXI using 32 or more stages suffer from significant defects in the solution.*

With $W = \mathrm{diag}(E^{-1}\mathbf{1})^{-1}$, we may transform to

$$u_{n+1} = \left[1 + \Delta t\lambda \underbrace{b^T E W^{-1}}_{\tilde{b}^T}(I - \Delta t\lambda D)^{-1} \underbrace{W E^{-1}\mathbf{1}}_{\mathbf{1}}\right] u_n, \tag{4.5}$$

which is a diagonal Runge-Kutta method with A replaced by D and the original completion vector b replaced by $\tilde{b}$. Rewriting this to a REXI form leads to

$$\begin{aligned} u_{n+1} &= u_n + \Delta t\tilde{\boldsymbol{b}}^T \left(-(\Delta t D)^{-1}\right)\left(I + (\Delta t\lambda D - I)^{-1}\right)\mathbf{1}u_n \\ &= \underbrace{\left(1 - \tilde{\boldsymbol{b}}^T D^{-1}\mathbf{1}\right)}_{\gamma} u_n + \underbrace{\left(-\tilde{\boldsymbol{b}}^T D^{-2}\right)}_{\beta^T}\left(\Delta t\lambda - \underbrace{D^{-1}}_{\mathrm{diag}(\alpha)}\right)^{-1}\mathbf{1}u_n. \end{aligned} \tag{4.6}$$

Finally, we can write this in the unified REXI formulation (1.3) with

$$\gamma = 1 - \tilde{\boldsymbol{b}}^T D^{-1}\mathbf{1} \qquad \beta^T = -\tilde{\boldsymbol{b}}^T D^{-2} \qquad \alpha = \mathrm{diag}\left(D^{-1}\right). \tag{4.7}$$

We have derived a transformation from implicit RK method with nonzero eigenvalues to REXI form with the same stability function. Given a REXI method, one can construct an equivalent diagonal RK method (with complex coefficients) via $D = \mathrm{diag}(\alpha)^{-1}$ and $\tilde{\boldsymbol{b}}^T = -\beta^T D^2$. Note that a conventional Butcher table $A, \boldsymbol{b}^T$ is not uniquely determined by this procedure. We remark that standard techniques for analyzing Runge-Kutta methods can readily be applied to REXI methods. This includes barriers such as Theorem 4.3 of [20], which establishes that diagonal (parallel) RK methods can be no more than second order accurate for nonlinear problems.

4.1.2. Error studies. We choose the Gauss-Legendre and Chebyshev quadrature points for the error studies, with results given in Figure 4.1. We can observe that increasing the number of stages in the non-diagonalized version (using a dense Butcher table) always improves accuracy per stage. In contrast, B-REXI accuracy degrades when too many stages are used, becoming apparent beyond 8 stages. This effect is related to ill-conditioning that can be interpreted via the condition number of the eigenbasis E that effects diagonalization (4.5) or via the 1-norm of the completion vector $\tilde{\boldsymbol{b}}$, as shown in Figure 4.2. Note that completion vectors must sum to 1 so $\|\tilde{\boldsymbol{b}}\|_1 = 1$ is optimal (and indeed holds for the original completion vector $\boldsymbol{b}$);

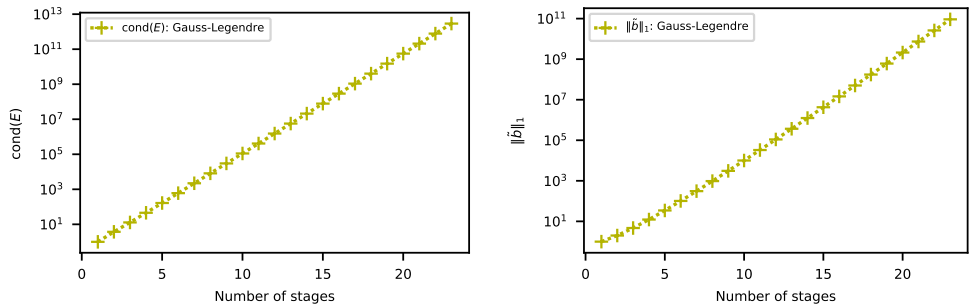

FIG. 4.2. *Condition number of the eigenbasis E for the B-REXI method on Gauss-Legendre collocation points (left) and 1-norm of completion vector $\tilde{\mathbf{b}}$ for the diagonalized method (right). The rounding errors incurred by the exponential growth precludes use of this approach for many stages.*

a large 1-norm indicates the existence of large positive and negative entries, leading to cancellation errors. Despite this downside, the numerical experiments of §6 will show that these B-REXI (diagonalized Gauss Runge-Kutta) methods with lower stage counts are remarkably efficient compared to the other (better-conditioned) families.

4.1.3. Relation to Crank-Nicolson. Since this will be relevant for the results section, we would like to show the relation between the B-REXI approximation with a single pole using the Gauss-Legendre quadrature using just a simple pole (centered at the interval). This will lead to the terms $\gamma = -1$, $\alpha = 2$ and $\beta = -4$ which yields the REXI approximation $\exp(x) \approx -1 + \frac{-4}{x-2} = \frac{1+\frac{1}{2}x}{1-\frac{1}{2}x}$ with the equation on the right hand side matching the Crank-Nicolson formulation. This REXI approximation with a single term resembles the Crank-Nicolson formulation with a midpoint rule (forward Euler on nominator and backward Euler on denominator for $x = \Delta t L$ and a half-time step size). This will explain that later numerical results with B-REXI match the Crank-Nicolson method. Using more REXI poles will result in even higher-order approximations.

4.2. T-REXI: Terry's Rational Approximation of the Exponential Integrator. The approach which we will refer to as T-REXI was introduced in [18]. Several steps are required to gain the α and β coefficients. Since these steps account for the computational workload and the properties, we briefly describe the derivation, including a discussion on the advantages and limitations of this method.

4.2.1. Derivation. The first step consists of an approximation of a Gaussian basis function $\psi_c(x)$ as follows:

$$\psi_h(x) = (4\pi)^{-\frac{1}{2}} e^{-x^2/(4h^2)} \approx Re\left(\sum_{k=-W}^{W} \frac{\omega_k}{i\frac{x}{h} + (\mu + ik)} \right) \tag{4.8}$$

Using $W = 11$, hence $L = 2W + 1 = 23$ terms in total, is sufficient for an accurate approximation up to numerical double precision (see [18]). The advantage of this representation is an efficient representation of the Gaussian basis function in Fourier space. The proxy with the Gaussian basis function allows for computing the coefficients ν_k for an approximation of an oscillatory function within an approximate range

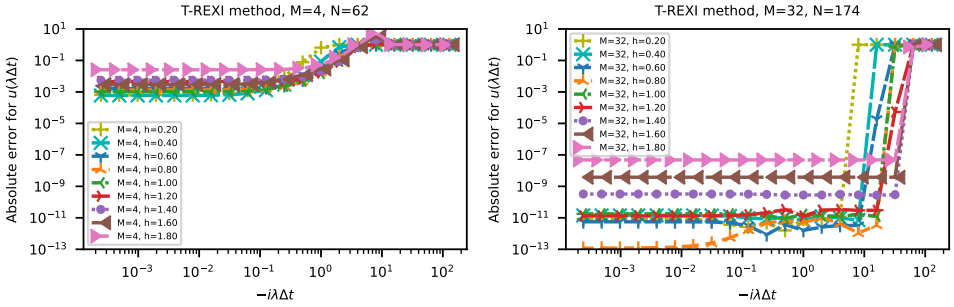

FIG. 4.3. *Error studies for the T-REXI method for different M and h values. M relates to the number of Gaussian functions via $2M+1$ to approximate an oscillation with $N = 2(2M+L)$ number of REXI terms. Left image: We can observe a very high error for a low number of Gaussian basis bumps, which cannot be improved by changing h. Right image: Using significantly more Gaussian functions leads to significant improvements. In particular, we observe that optimal values for h influence the quality of the approximation. An optimum can be observed for $h \approx 0.8$.*

$x \in [-Mh; Mh]$ in Fourier space, yielding

$$\exp(ix) \approx \sum_{k=-M}^{M} \nu_k \psi_h(x + kh). \tag{4.9}$$

Both steps are then combined, resulting in the approximation

$$Re\left(\exp(ix)\right) \approx \sum_{n=-M-W}^{M+W} Re\left(\beta_n^{Re}(ix + \alpha_n)^{-1}\right) \tag{4.10}$$

where we only showed the Re one. We combine Re and Im to the form

$$\exp(ix) \approx \sum_{n=-M-W}^{M+W} \beta_n(ix - \alpha_n)^{-1} \tag{4.11}$$

eventually leading to the REXI formulation with $\gamma = 0$, but x related directly to the imaginary value on the complex plane. So far, we only targeted the φ_0 function, and we like to point out that this method can also be used to approximate other φ_i terms (see [18]) or directly with the REXI coefficients (see §3.2). We want to emphasize that this approximation was derived only for purely oscillatory functions and, hence, does not include approximations with non-zero real eigenvalues components.

4.2.2. Error studies. We investigate the errors of the T-REXI method in Figure 4.3. On the left image, we can observe that we need a minimum number of Gaussian basis functions to approximate the oscillations. The right image shows exceptionally accurate results for $h \approx 0.8$ in the range $x \in [0; 10]$ and a rather large region of accuracy of about $e(x = 128) \leq 10^{-11}$. Other figures (not included) show that increasing M leads to a linear increase of the size of the region of high accuracy (see [18]) with an optimum value of $h \approx 0.8$. For the remainder of this work, we will use $h \approx 1.0$ as a compromise between accuracy and total workload.

4.3. CI-REXI: Cauchy Contour Integral method. Cauchy Contour Integral (CI) methods offer an alternative way to infer the REXI coefficients (see e.g. [42, 4, 34]). We start with the general CI equation given by

$$g(x) = \frac{1}{2\pi i} \oint_{\Gamma} \frac{g(z)}{z - x} dz \tag{4.12}$$

where $g(x)$ is one of the analytic φ_i functions (2.4), Γ the contour enclosing the eigenvalue λ for ODEs and all eigenvalues on the diagonal of Λ for PDEs.

4.3.1. REXI Derivation. Regarding the contour, we can use different approaches. In what follows, we used parametrized contours $\Gamma = \{\sigma(w)|w \in [0;1]\}$ with the contour function $\sigma(w) : \mathbb{R} \to \mathbb{C}$. Using integration by substitution and the contour function, we obtain

$$g(x) = \frac{1}{2\pi i} \oint_0^1 \frac{g(\sigma(w))}{\sigma(w) - x} \sigma'(w) dw = \oint_0^1 \frac{i(2\pi)^{-1} g(\sigma(w))\sigma'(w)}{x - \sigma(w)} dw. \tag{4.13}$$

Using the trapezoidal rule, which is exponentially fast converging on periodic boundaries (see [41]) with N trapezoidal points in total, we obtain

$$g(x) \approx \frac{1}{N} \sum_{n=1}^{N} \frac{i(2\pi)^{-1} g(\sigma(w_n))\sigma'(w_n)}{x - \sigma(w_n)} \quad \text{with} \quad w_n = \frac{n}{N}. \tag{4.14}$$

Again, we can infer a unified REXI formulation (1.3) by setting

$$\alpha_n = \sigma(w_n) \qquad \beta_n = \frac{ig(\sigma(w_n))\sigma'(w_n)}{N2\pi} \qquad \gamma = 0. \tag{4.15}$$

An ellipse contour is given by $\sigma(w) = R_x \cos(iw2\pi) + iR_y \sin(iw2\pi) - \mu$ with μ related to the center of the ellipse. This leads to the coefficients

$$\alpha_n = R_x \cos(iw2\pi) + iR_y \sin(iw2\pi) - \mu \tag{4.16}$$

$$\beta_n = \frac{i}{N} \exp(\sigma(w)) \left(-R_x \sin(iw2\pi) + iR_y \cos(iw2\pi)\right) \tag{4.17}$$

A study of all kinds of contour shapes (rectangle, bean, polygonal shapes, etc.) is beyond the scope of this work. In the next section, we will mainly focus on the circle to show interesting characterizations and use the ellipse for numerical studies to show its superiority to another REXI method. In the following, we will refer to the special case of a circle as CI-REXI and to the ellipse case as CI-EL-REXI.

4.3.2. Characterization and numerical issues. Next, we characterize the REXI terms, referred to as the β characterization, with an overview in Figure 4.4. We remind the reader that REXI approximates functions with a linear combination of rational basis functions. Depending on the placement of these functions (related to α_n) and the weighting of each basis (related to β_n), we have three different cases:

a) Obsolete REXI terms: Contours $Re(\sigma(x)) \to -\infty$ relating to areas of the contour in the distant negative real axis on the complex plane have exponentially fast decaying β coefficients, hence $\lim_{\sigma\to-\infty} \beta_n = \lim_{\sigma\to-\infty} \frac{i \exp(Re(\sigma(w_n)))\sigma'(w_n)}{N2\pi} = 0$. Once a particular β_n coefficient undershoots a threshold ϵ_β, the corresponding REXI term can be removed if $\beta_n < \overline{\epsilon}_\beta$ and $\overline{\epsilon}_\beta = \epsilon_\beta / N$. The last equation incorporates that a higher numerical resolution results in smaller values of the β weights.

b) Regular REXI terms: This characterization refers to those REXI terms that can be incorporated in the approximation in a useful way.

c) Cancellation-prone REXI terms: These terms are related to the contour $Re(\sigma(x)) \to +\infty$. Approximating the exp function in the far distance to the right of the origin leads to exponentially increasing the β values. An oscillatory function is also parallel to the imaginary axis, which is approximated. Both effects lead to

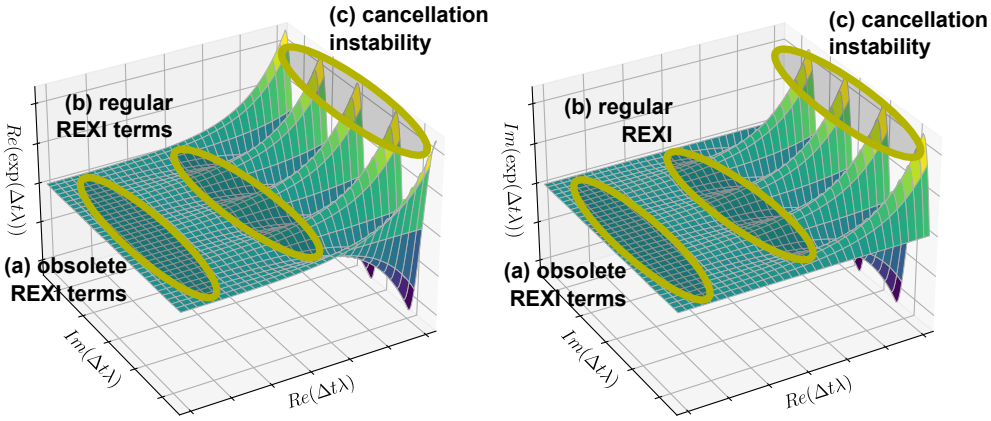

FIG. 4.4. *Complex plane for the real (left image) and imaginary (right image) value of* $\exp(x)$. *We highlight the different areas related to the different* β *characterizations.*

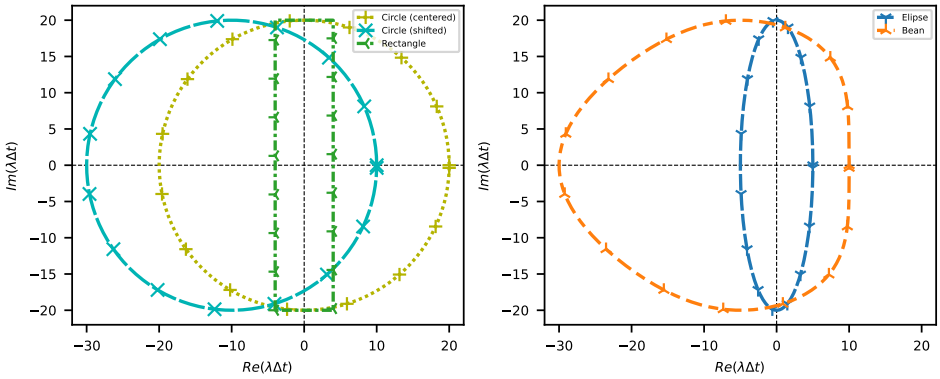

FIG. 4.5. *Selection of contours used with the CI-REXI method.*

very large positive and negative numbers, resulting in severe cancellation errors in this region. Consequently, this region should be avoided.

Examples of different contours are provided in Figure 4.5. Each contour targets a particular problem. The circle can be used for the approximation of a small spectral radius $\lambda\Delta t < 10$. Once requiring a larger approximation along the imaginary axis, the radius cannot be enlarged without sacrificing accuracy due to cancellation errors in β_n, see (c) above. This can be avoided by enlarging the radius and choosing the value μ, hence shifting the circle, to exclude a contour across areas with $Re(x) > 10$, which leads to the shifted circle. Other contours are, e.g., given by the ellipse or rectangle targeting the approximation of a spectrum on or close to the imaginary axis and the bean contour targeting an approximation of diffusive and oscillatory problems. Studies about these contours are beyond the scope of this work and we will focus on the (shifted) circle and ellipse throughout the remainder of this paper.

4.3.3. Error studies. We conduct error studies using the shifted circle CI-REXI method. The first study is based on a circle centered at the origin, with studies for different radii. The second is for a circle which is shifted to overcome problems related to the cancellation effects (see (c) above).

Results are given in Figure 4.6 with plots based on a fixed number of $N = 256$

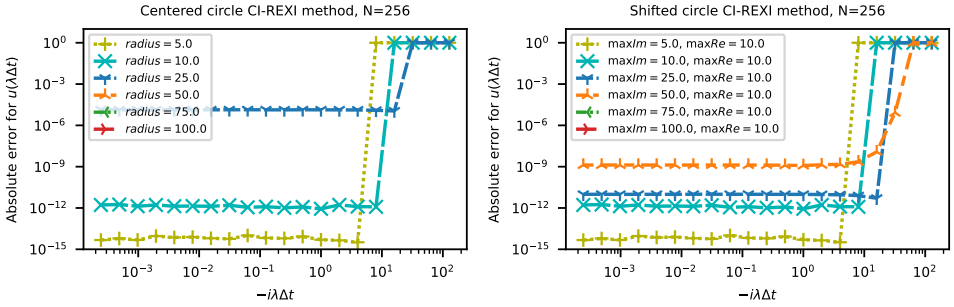

FIG. 4.6. *Error studies for the centered circle (left) and shifted circle (right) CI-REXI method. The centered circle suffers from cancellation effects for large radii, whereas the shifted circle limits these effects. In particular, for a larger imaginary spectrum to be approximated, adding more REXI poles leads to improved accuracy, which is not the case for the centered circle CI-REXI method.*

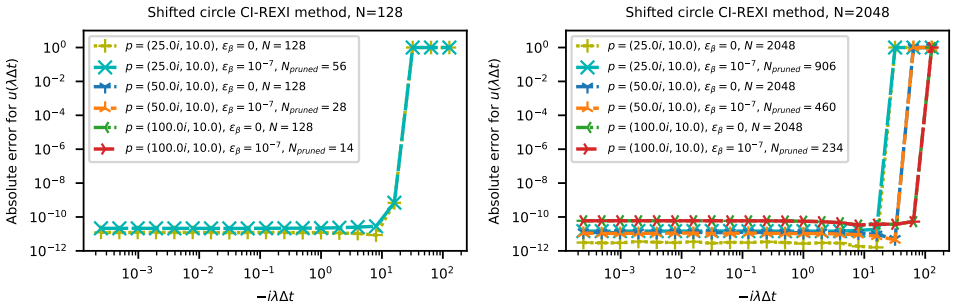

FIG. 4.7. *Error studies for the shifted circle with $N = 128$ REXI poles (left) and $N = 2048$ REXI poles (right) with different ϵ_β pruning values. We can observe significant reductions in the number of required REXI poles.*

REXI poles. We can observe that the errors significantly increase for the centered circle once the radius exceeds a certain threshold. In particular, errors for a larger radius – including a larger spectrum on the imaginary axis – are outside the plotting range. The results for using a higher number of REXI poles do not significantly improve the results. Using a shifted and enlarged circle, we can gain improved results that overcome cancellation errors.

So far, we only investigated the error itself but neglected the total workload. Pruning β with ϵ_β (exploiting characterization (a)), we can reduce some workload significantly as depicted in Figure 4.7 for larger radii. For a moderate number of REXI poles $N = 128$ (left image), we observe a pruning close to the accuracy of REXI itself, hardly impacting the results. In contrast, larger radii already suffer from inaccuracies of the used quadrature, with errors outside the plotting range. For a larger number of REXI poles $N = 2048$ (right image), we observe very robust pruning, hardly affecting the accuracy of the REXI approximation quality but leading to a significant reduction of the workload.

5. Stability, normalization & filtering. So far, we have only studied errors in approximating the φ_0 function with REXI methods. However, once we use REXI methods for time integrating differential equations, additional properties such as stability and convergence are assumed to be relevant. We will investigate these properties in this section for the ODE $\frac{du(t)}{dt} = \lambda u(t)$.

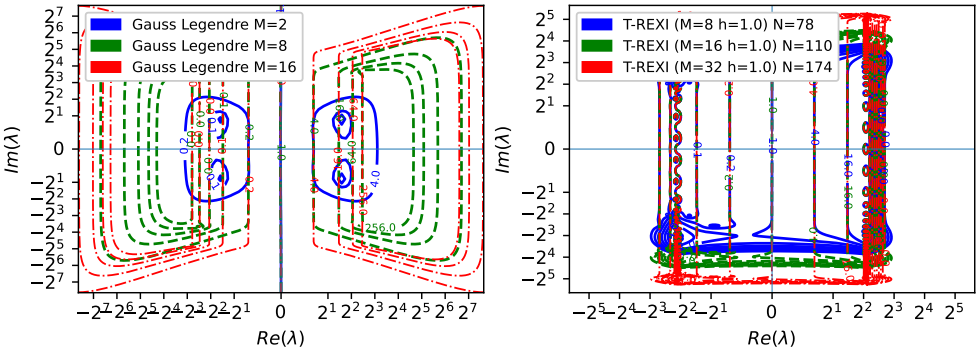

FIG. 5.1. *Stability plots. For B-REXI (left): we observe an excellent stability behavior known for collocation methods. T-REXI (right): We observe instabilities at the imaginary axis for the boundaries of the approximation range.*

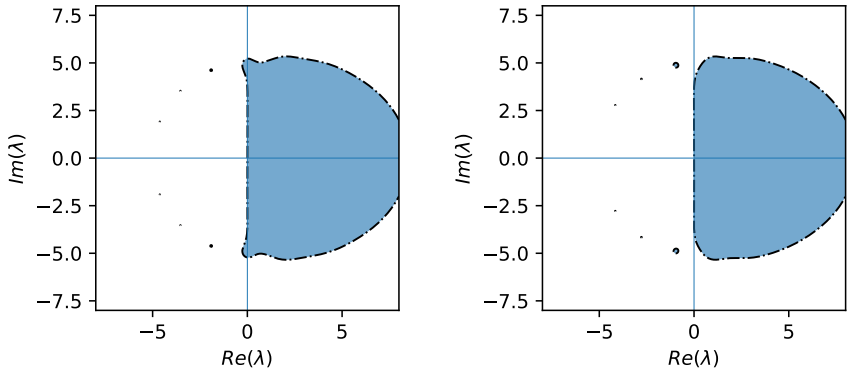

FIG. 5.2. *Stability plots for CI-REXI. The left image depicts the discrete contour points chosen so that one α pole lies on the imaginary axis. This leads to instabilities. The right image depicts a half-shifted variant of it.*

5.1. Stability. The stability plots are generated based on the stability function $R(\lambda)$, which is defined by the execution of a single-time step $u(t+\Delta t) = R(\Delta t\lambda)u(t)$. We will plot the amplification factor $|R(\lambda)|$ of the solution $u(t)$ over a time step $\Delta t = 1$.

B-REXI (left image in Figure 5.1): The stability reflects the A-stability of these methods on the entire left half plane. In particular, stability is given for the entire imaginary axis, a known property of collocation methods.

T-REXI (right image in Figure 5.1): We can observe that T-REXI provides excellent stability for purely imaginary values. However, we can observe instabilities on the imaginary axis once we reach the boundaries of the approximation range. This can be avoided by an additional T-REXI filter, which could be applied to obtain stability also outside the approximation range (see [18]).

CI-REXI: Finally, we look at the CI-REXI method based on Cauchy contour integral methods in Figure 5.2. The left image shows an unstable region along the imaginary axis. This is caused by an α pole directly placed on the imaginary axis. We can avoid this by choosing the support points of the trapezoidal rule differently. The right image shows a solution to this by shifting them by a half interval, effectively avoiding this instability, and CI-REXI becomes unconditionally stable for oscillatory systems. To summarize, if using the CI-REXI method, one should avoid placing poles

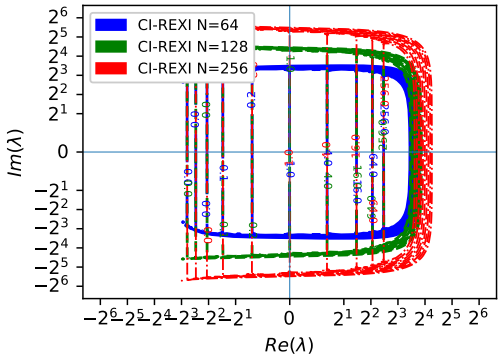

FIG. 5.3. *Contour lines for stability plots for CI-REXI with half-shifted intervals. We can observe an excellent stability region over this entire range.*

near the eigenvalues of the linear operator.

A contour plot comparing CI methods with an increasing number of poles and approximation range is provided in Figure 5.3. We adopted the contour of the circle to pass through the points $\pm 10i$, $\pm 20i$, and $\pm 40i$ on the imaginary axis for increasing the number of poles while keeping the contour never exceeding 10 on the real axis.

5.2. Normalization. This section concerns particular problems for stationary or nearby modes requiring special treatment with the T-REXI method. So far, we only assessed errors for a single time step, and this section will investigate the accuracy and conservation properties of stationary modes concerning REXI methods. We will use Dahlquist's equation (5) with $\lambda = 10^{-3}i$, which is time-integrated until $t = 100$ using different REXI methods. The particular choice of this low frequency is related to almost stationary modes of PDEs. Such modes play an important role, e.g., for geostrophic balance in atmospheric simulations, and not preserving them might lead to spurious/parasitic modes.

An investigation of the results at the absolute ODE errors at $t = 100$ is given in Figure 5.4. The left column shows REXI methods as they have been computed with the methods from before. We can observe that the CI-REXI method (top left image) has REXI coefficients preserving the stationary modes. However, the T-REXI suffers from significant defects in it. A normalization can be used to overcome this problem where stationary modes require $\sum_n \frac{\beta_n}{x-\alpha_n} = s = 1$ and we can ensure this by simply rescaling β_n so that $\beta_n^{\text{new}} = \frac{\beta_n}{s}$.

The results for this are given in the right column, where we observe relatively small improvements for the CI-REXI method (right top image). However, for the T-REXI method, the errors significantly drop from 10^{-8} to about 10^{-13} once applying this normalization. We also do not see any accuracy degradation for very large time step sizes. Hence, this normalization can be used without impacting the accuracy of other choices of λ, and we will use it throughout the remainder of this work.

We close with two side comments: First, studies for purely stationary modes ($\lambda = 0$, not shown here) showed that the errors are increasing for smaller time step sizes, but only due to round-off errors. Overall, these results still lie within the range of numerical precision; hence, we skipped them here. Second, we skipped the B-REXI method since it is not prone to this problem for a number of REXI terms usable as solvers.

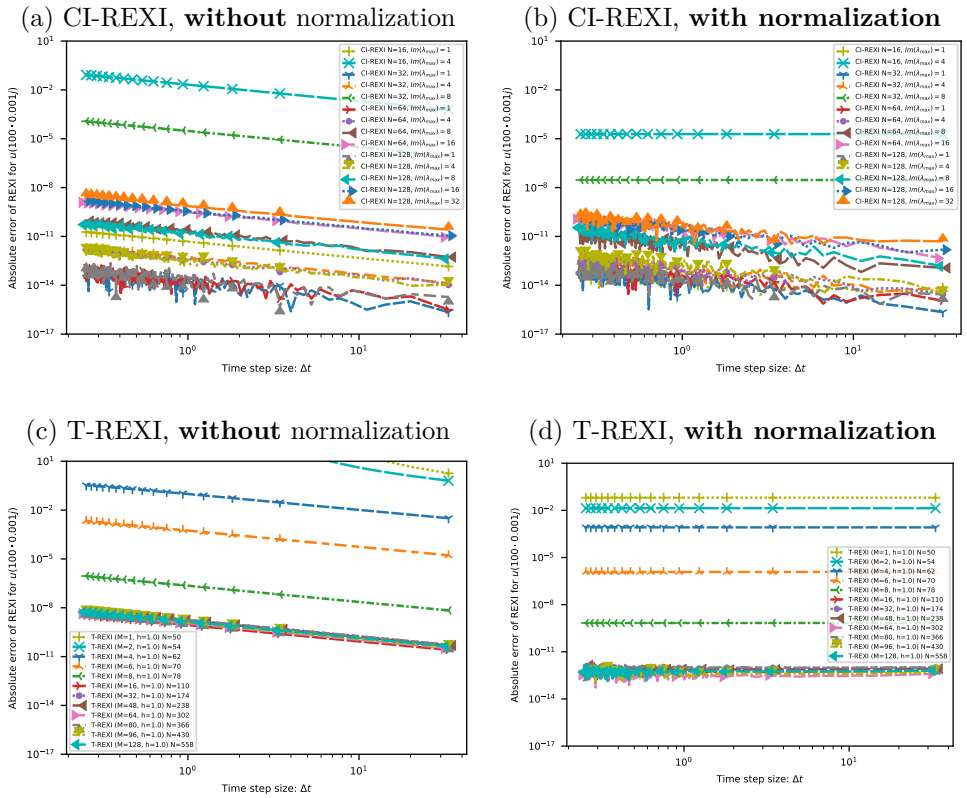

FIG. 5.4. *Error studies for different REXI methods of time step size Δt vs. absolute error at $u(t = 100)$. The left column shows errors without normalization, and the right column shows errors with normalization for near-stationary modes. As we can observe in the first row, the normalization for near-stationary modes with the CI-REXI method does not lead to any significant improvements. In contrast, significant improvements can be observed for the T-REXI method. See the text for a detailed explanation of the results.*

5.3. Filtering. This brief section points out the filtering capabilities of the different REXI methods. We define a filter to apply a reduction of the amplitude of $\varphi_i(x)$ for a particular set of eigenvalues. It is, in particular, desirable to filter out (setting them close to zero) the so-called "fast modes" for x starting at a threshold and to have a smooth transition of the change in amplitude towards filtering out modes. Since diffusive problems already have a reduction of amplitude given naturally by their mathematical properties, we will again solely focus on purely oscillatory problems, with results also applicable to a mix of oscillatory/diffusive problems.

Using the **B-REXI** method, we can observe that the stability contour follows exactly the imaginary axis. Hence, there is no filtering at all. For the **T-REXI** method, we skip a discussion of filtering due to the inherent instability at the boundaries of the approximation range and point out to an additional filter proposed in [18]. The **CI-REXI** method has a natural filtering. This is due to the property that points outside the contour are rapidly approaching 0 as a property of the Cauchy contour integral.

6. Comparison of REXI methods. This section aims to provide guidance about which REXI method is best, and we will explore this in different ways. A

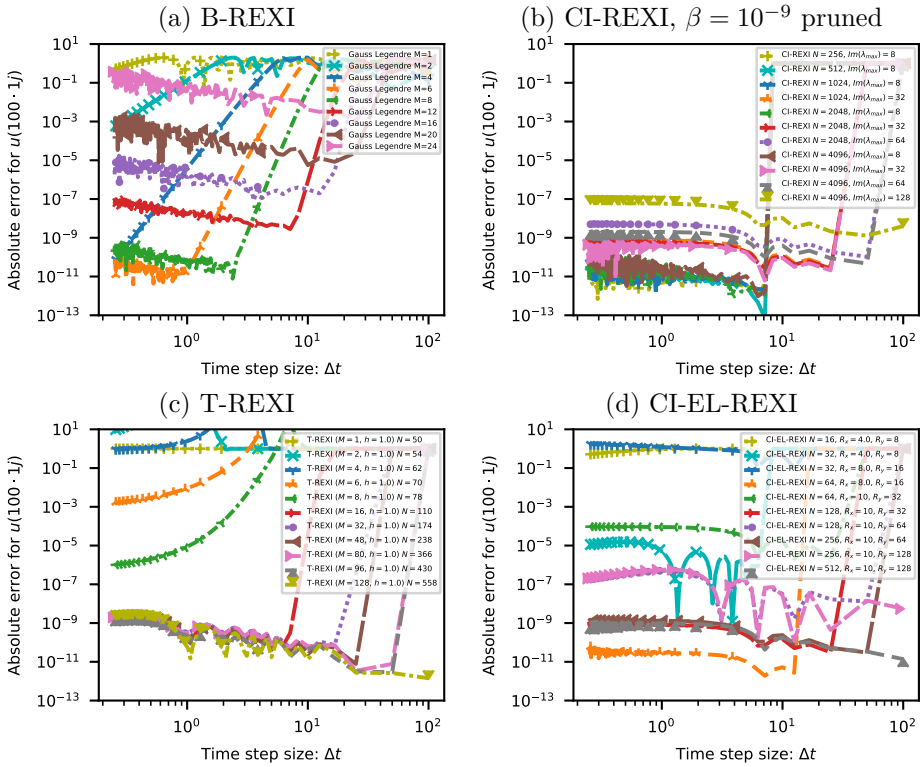

FIG. 6.1. *Error studies for different REXI methods of time step size Δt vs. absolute error at $u(t = 100)$. B-REXI (left top) is suitable only for smaller timestep sizes. CI-REXI can be tuned to allow also very large time step sizes. T-REXI requires many poles for small time step sizes and allows also very large time step sizes. CI-EL-REXI allows also very large time step sizes and in addition requires the least number of poles for similar accuracy.*

full exploration of all parameter combinations is obviously not possible. Hence, we focused on the ones that were most rational to us based on far more experiments than shown here. We first continue with concrete examples using a linear oscillatory ODE based on the Dahlquist equation followed by a PDE with the nonlinear shallow-water equations on the rotating sphere to gain insight into numerical properties once we apply this to more realistic test cases.

Based on the eigendecomposition, we classify linear operators as oscillatory, diffusive, or both. A purely oscillatory system [10] requires imaginary-only eigenvalues ($i\lambda \in \mathbb{R}$) whereas a diffusive behavior is based on negative real eigenvalues ($\lambda \in \mathbb{R}$ and $\lambda < 0$). Since oscillatory/hyperbolic systems belong to the most challenging problems for REXI methods, we will solely focus on them.

6.1. ODE. We investigate the ODE systems again with Dahlquist's equation (5) using $\lambda = 1$ and the simulation results at $t = 100$. We use $u(0) = (1 + i)/\sqrt{2}$ as an initial condition. We compare various REXI methods in Figure 6.1. The total numbers of REXI coefficients are given by N.

The **B-REXI** method (left upper image) performs extremely well for small step sizes where only a few poles are required. For larger time step sizes of $\Delta t \approx 10$, using 16 poles is sufficient to gain single precision accuracy.

The **CI-REXI** method (right top image) is tuned with a contour never exceeding

a real value of 10 and to include the points on the imaginary axis given by $Im(\lambda_{max})$. The CI-REXI method clearly outperforms the B-REXI method for medium-sized time step sizes and also allows taking very large time step sizes.

The **T-REXI** method (left bottom image) requires a significant number of REXI poles if only small time step sizes should be taken. This improves once larger step sizes are taken since the initial overheads of the large number of poles (due to the rational approximation of the Gaussian) have less relative impact on the number of REXI coefficients.

In addition, we also investigated the **CI-EL-REXI** (right bottom image) method, which is a natural choice for purely oscillatory problems. We chose the semi-major axis of the ellipse along the real axis in an empirical way and never exceeding 10 to avoid numerical issues. This method *outperforms both CI-REXI and T-REXI almost everywhere* regarding accuracy and number of terms required to solve it.

6.2. PDE example. In this final section, we will investigate different REXI methods with the shallow-water equations (SWE) on the rotating sphere. We decided not to investigate many different PDEs, but to go into depth of exponential integration for a single one which is of purely hyperbolic nature. We chose the SWE since they are frequently used to assess the quality and performance of discretizations in time and space concerning horizontal aspects of the full Euler equations solving the fluid dynamics equations related to the atmosphere. In velocity form, the nonlinear SWE are given by

$$\frac{\partial}{\partial t}\begin{pmatrix}\Phi\\ \vec{V}\end{pmatrix} = \underbrace{\begin{pmatrix}-\overline{\Phi}\nabla\cdot\vec{V}\\ -\nabla\Phi\end{pmatrix}}_{L_g\text{: linear gravity}} + \underbrace{\begin{pmatrix}0\\ -f\vec{k}\times\vec{V}\end{pmatrix}}_{L_c\text{: linear Coriolis}} + \underbrace{\begin{pmatrix}-\nabla\cdot\left(\Phi'\vec{V}\right)\\ -\vec{V}\cdot\nabla\vec{V}\end{pmatrix}}_{N\text{: nonlinear term}} \tag{6.1}$$

with the horizontal velocity $\vec{V}$ on the longitude/latitude field, geopotential $\Phi = g \cdot h$ with height h, average geopotential $\overline{\Phi} = g \cdot \overline{h}$ with average height $\overline{h}$, a linearization around a state $\overline{h} = 10^5 m$, Coriolis effect $f = 2\Omega\sin(\phi)$ with latitude ϕ and angular rate of rotation Ω. We like to emphasize that no (hyper)viscosity is used in this PDE to avoid a simplification of the problem due to diffusive effects.

We use this PDE due to its particularly interesting features: The linear gravity term L_g is the stiffest one and can be solved with exponential integrators either analytically or with REXI. We want to point out that a comparison of some methods has already been under investigation in former work [34] but solely with the CI-REXI method and the geopotential field, which has also been identified to be the best Strang-split method. Anyhow, this study also lacked comparisons with other variables, particularly other REXI methods, which will lead to new revelations, as presented in the following sections. Since including the T-REXI method would not provide any beneficial insight, since the CI-REXI method is computationally much cheaper and provides additional benefits, we skip this method in the following studies.

6.2.1. Spatial discretization. We solve these equations using the SWEET software[1] which utilizes spherical harmonics (SH) to solve these equations. Such a global spectral basis leads to a substantial reduction of spatial errors (besides a lack of nonlinear interactions at the limit of resolution), hence allowing us to put the focus on time integration methods. We like to refer to [32, 15] for a detailed description of the spherical harmonics. In particular, we work with the vorticity-divergence formulation

[1] https://sweet.gitlabpages.inria.fr/sweet-www/

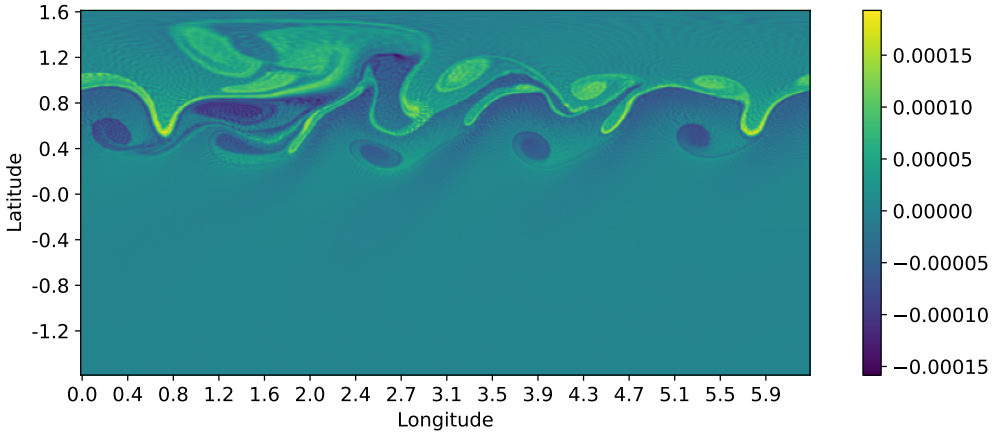

FIG. 6.2. *Vorticity field of barotropic instability benchmark after 8 days of inviscid shallow-water equations. We see the development of various large and small-scale vortices.*

in spectral space to avoid spurious modes if one would convert the velocity to spectral space. The standard $\frac{2}{3}$ rule [27] is used for anti-aliasing to evaluate bi-non-linearities.

6.2.2. Time stepping solvers. We can find highly efficient solvers in spherical harmonics space for direct exponential integration (without numerical approximations), REXI, and implicit Euler time integrators.

Regarding the direct exponential integration, we can straightforwardly find a direct solution using the vorticity divergence form, see also [38]. Due to orthogonality, each mode can be separately written as

$$\begin{bmatrix} \partial_t \Phi' \\ \partial_t \delta \end{bmatrix} = \begin{bmatrix} & -\overline{\Phi} \\ -\nabla^2 & \end{bmatrix} \begin{bmatrix} \Phi' \\ \delta \end{bmatrix} = \begin{bmatrix} & G \\ D & \end{bmatrix} \begin{bmatrix} \Phi' \\ \delta \end{bmatrix}$$

with the famous identity $\nabla^2 = -n(n+1)$ for this harmonic. Using $D = -\nabla^2$ and $G = -\overline{\Phi}$ for convenience, we find the eigenvectors Q and eigenvalues $diag(\Lambda)$

$$Q = \begin{bmatrix} -\sqrt{\frac{G}{D}} & +\sqrt{\frac{G}{D}} \\ 1 & 1 \end{bmatrix} \qquad Q^{-1} = \begin{bmatrix} \frac{1}{2}\sqrt{\frac{D}{G}} & \frac{1}{2} \\ -\frac{1}{2}\sqrt{\frac{D}{G}} & \frac{1}{2} \end{bmatrix} \qquad \Lambda = \begin{bmatrix} -\sqrt{DG} & \\ & \sqrt{DG} \end{bmatrix}.$$

We can then use $U(t+\Delta t) = Q \exp(\Delta t \Lambda) Q^{-1}$. From an algebraic perspective, this method matches the method in [17], which uses a rather cumbersome derivation using Laplace transforms, whereas our derivation is more elegant and short. This method is also used for investigating errors

For the exponential integration of the full linear terms $L = L_g + L_c$, this would relate to the Hough modes [44] and no direct exponential solution has been derived yet. Hence, it requires evaluations of the form $(\Delta t L - \alpha)^{-1}$ with complex-valued α. The first time this was solved for REXI using spherical harmonics was based on a method requiring transformations to grid space [32]. The present work is based on an implicit time stepper [37] of the form $(I - \Delta t L)^{-1}$ which has been transformed to solve a REXI term by simply using a complex-valued time step size. We also used it for the implicit time integration of L as it has been originally suggested.

6.2.3. Benchmark. Our benchmark is based on the barotropic instability test case (see [14]). This benchmark is initialized with a geostropically balanced initial

Short notation	Description
$ERK(X, o = N)$	Explicit Runge-Kutta with order N
$IRK(X)$	Backward Euler using 2nd order Crank-Nicolson
$SS(X, Y)$	2nd order Strang-splitting as explained in the text with $F_1 = X$ and $F_2 = Y$
$EXP(X)$	Direct exponential integration on X
$REXI(X)$	A particular REXI method on X
$ETDRK(X, Y)$	2nd order ETDRK method with X being exponentially integrated and Y treated as the nonlinearity
$X + Y$	Time tendencies of terms X and Y are added

TABLE 6.1
Overview of time integration methods. Note that they can be composed together.

condition, which is perturbed by a small Gaussian bump (see reference for detailed initial conditions). We time integrate this system for 8 days with results in Figure 6.2.

6.2.4. Time integration. Regarding the particular Runge-Kutta (RK) based time integrators, we used 2nd order midpoint, 3rd order Heun, and classical 4th order RK. The reference solution to compute the errors is based on the 4th order RK with a time step size of $\Delta t = 5$.

Besides the methods already introduced, our investigation also includes the 2nd order Strang splitting (SS) method [36]. With SS, a PDE given by two terms $\frac{d}{dt}U = F_1(U) + F_2(U)$ can be integrated with 2nd order accuracy if a 2nd order accurate time integrator $R^{\Delta t}_{F_i}$ is provided for time step size Δt by $U(t + \Delta t) = R^{\frac{1}{2}\Delta t}_{F_1} \circ R^{\Delta t}_{F_2} \circ R^{\frac{1}{2}\Delta t}_{F_1}$. We use a function-like notation to refer to the particular time integration methods. An overview of this is given in Table 6.1 where we use X and Y as representatives for either term in the PDE such as L_g, L_c, and N or to refer to another time integrator. In the latter case, e.g., ERK, EXP, $REXI$, and IRK can both be used in the Strang-Splitting SS as arguments.

6.2.5. Hardware, parallelization & batch configuration. All results have been computed on the Thin Nodes of SUPERMUC-NG. Each node is equipped with two Intel SkylakeXeon Platinum 8174, resulting in two NUMA domains. For the spatial parallelization, we use solely OpenMP on one NUMA domain, resulting in a spatial scaling of up to 24 cores. Scalability for REXI is then based on MPI first by utilizing the 2nd NUMA domain, then other compute nodes. We gratefully acknowledge the usage of the SHTNS library [30] which is based on FFTW [13]. We precomputed transformation plans and reused them for all studies to ensure the utilization of the same ones over all studies. Each batch job is set to timeout after 1 hour, which follows the idea that the simulations should be finished within a specific time frame.

6.2.6. Performance comparison for splitting L_g and $L_c + N$. We start with a comparison of standard methods in Figure 6.3 which we will use as a baseline for further comparisons with REXI-based methods. Plots are given for the three prognostic variables, which we define here as the variables required as input to one time step, since results differ for all of them.

First, the higher-order 3rd- and 4th-order RK method can outperform other lower-order methods for smaller time step sizes depending on the variable under study. This

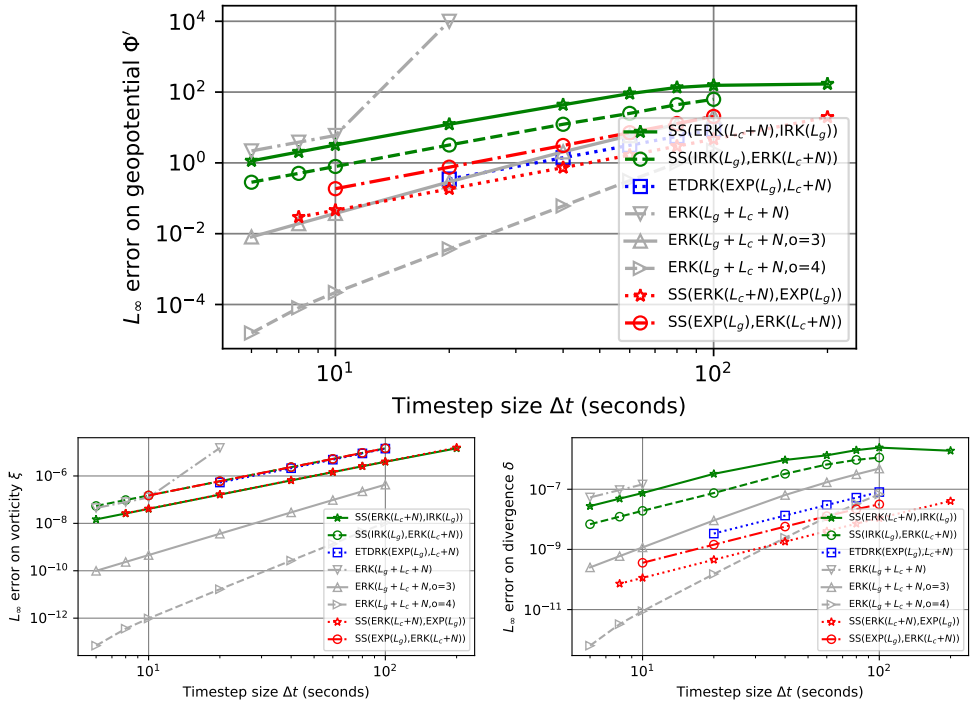

FIG. 6.3. *Studies **without REXI methods** (but using direct exponentiation) on all prognostic variables with error vs. time step size for the barotropic instability benchmark. We also include 2nd, 3rd, and 4th order Runge-Kutta based methods with gray lines.*

is a known phenomenon for higher-order time integration methods, and we wanted to include it to also see its max. stable time step size. We are primarily interested in very large time step sizes while still having a moderately small error.

The best method concerning the geopotential and the divergence variable is the Strang-split $SS(ERK(L_c+N), EXP(L_g))$, which we account for by the more accurate treatment with the exponential treatment of both variables. Since the vorticity field is not treated exponentially (time tendency for this in L_g is null), there's also no benefit visible in the comparison of the vorticity field.

The ETDRK method itself – although assumed to be an excellent off-the-shelf method – does not provide the overall best results compared to the rather straightforward Strang splitting. We can observe it to be the 2nd best for the geopotential and even lower ranked for the other variables. We account for that by the way a 2nd order accurate Strang-splitting is performed. This can be interpreted as a subcycling of time steps by executing two half-time steps for one of the terms (the time step size limiting one).

Next, we will continue with REXI studies by comparing them with the best Strang-split exponential and implicit methods from the previous results in Figure 6.4. Overall, we can observe a 2nd order convergence even if using only a single pole for the B-REXI methods.

Matching results for *SS(REXI,ERK) B-REXI N=1* and *SS(IRK,ERK)* are observed which is explained in §4.1.3: This particular B-REXI method resembles exactly the Crank-Nicolson method but uses one complex-valued pole to solve the system of equations.

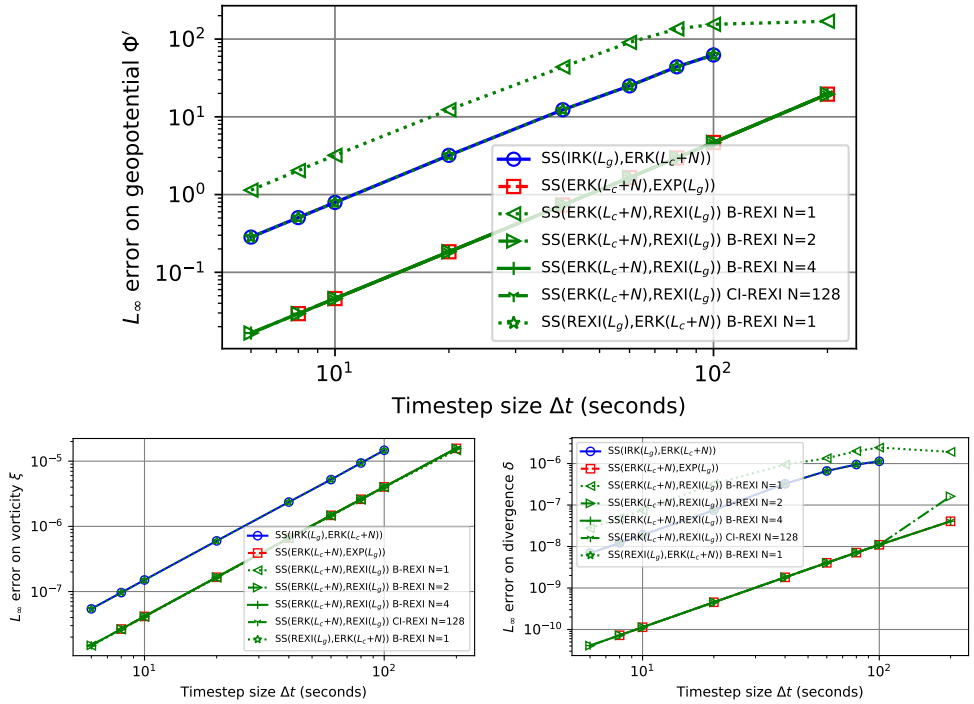

FIG. 6.4. *Studies **with REXI methods** on all prognostic variables with error vs. time step size for the barotropic instability benchmark.*

The other B-REXI methods outperform all alternatives except for the direct exponential integration EXP. We see one particularly interesting and highly important effect: The B-REXI method does not provide any further advantages using more than $N = 2$ poles. Even using $N = 4$ poles, the results are not further improved. A particularly important point is the comparison of the CI-REXI method with B-REXI, where absolutely no benefits are visible for $N = 128$ poles using CI-REXI compared to $N = 2$ poles using B-REXI. This clearly indicates that significant computational savings of a factor of 64 can be accomplished in this case compared to the former work.

We close this section by HPC studies in Figure 6.5. For sake of better overview, we only plotted the most promising candidates (ETDRK is worse than B-REXI methods, the explicit RK order 3 and 4 methods are better for larger wallclock times (smaller time steps), but unstable otherwise).

We start by comparing the performance of the direct exponential method EXP with the REXI method, where we would expect that the direct method is faster, which is not the case. We account for that by the direct method to be computationally more intensive (square root, exponential, etc., see §6.2.2) in order to solve for this term, whereas the B-REXI methods only require to evaluate two or 4 rational approximations. For the CI-REXI method, which requires $N = 128$ terms this is again different due to the higher MPI overheads resulting in a lower performance than the others.

Although the Strang-splitting method with the implicit term is computationally quite efficient to evaluate, its overall wallclock time performance is not optimal.

6.2.7. Performance comparison for splitting into L and N. Next, we investigate the performance of REXI methods using a splitting into the linear term

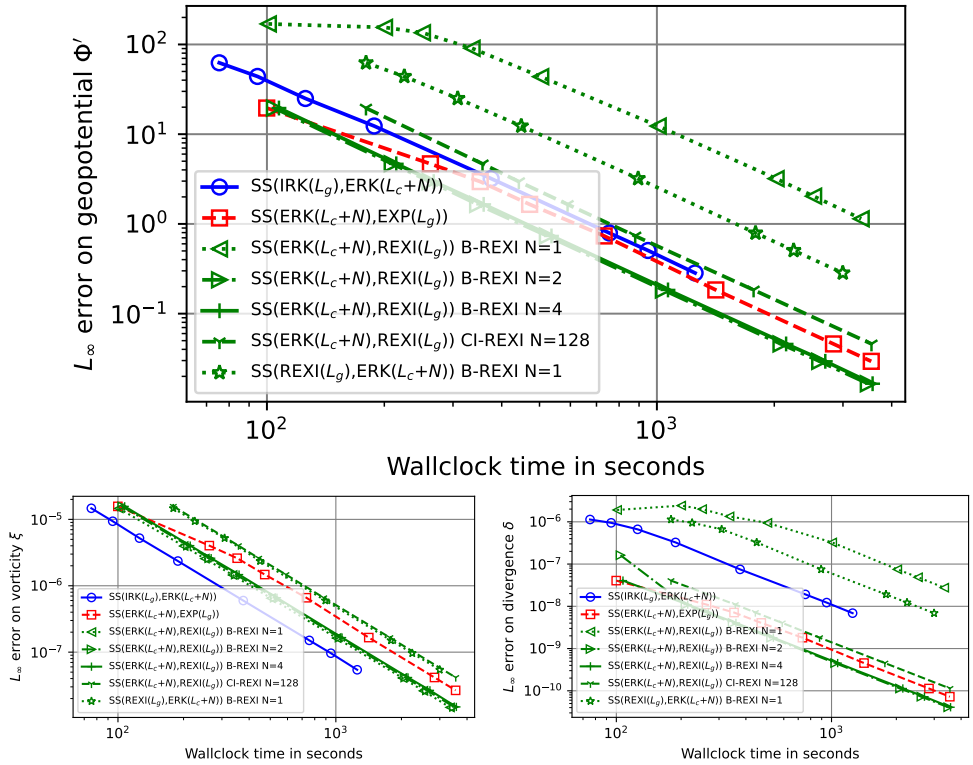

FIG. 6.5. *Studies **including REXI methods** with wallclock time vs. time step size for the barotropic instability benchmark.*

$L = L_g + L_c$ and the nonlinear term N. This leads to the situation that no direct computation of $\exp(\Delta t L)$ is possible, as previously explained. Plots are given in Figure 6.6 where some data points of ETDRK are missing due to the 1h time out of the job (see discussion before).

For the geopotential Φ', we can observe significant improvements in terms of accuracy. In particular, we can take very large time step sizes and still observe a convergence, whereas the 2nd order IRK-like methods already stagnate. With respect to the ETDRK scheme, its performance is worse compared to the best (straightforward) Strang-split methods.

For the vorticity η we can observe that the ETDRK method does not lead to any improvement. The best methods are the Strang-split IRK-based ones and some REXI-based methods. Hence, we do not see any improvement in the accuracy of the vorticity field by using exponential integration methods. This is kind of surprising at first glimpse since we expected a better treatment of the vorticity due to the exponential integration of the Coriolis effect. However, the errors in the nonlinear parts dominate the overall errors. Hence, this does not provide any better results.

The divergence δ study shows REXI methods to be the best ones. Again, the accuracy cannot be improved by using more than $N = 2$ poles. Everything beyond that would be an additional computational burden. The ETDRK methods again show a poorer performance than the more straightforward approach.

Finally, we investigate the wallclock time vs. errors with results given in Figure 6.7. We can observe that fully explicit ERK methods actually provide excellent results due

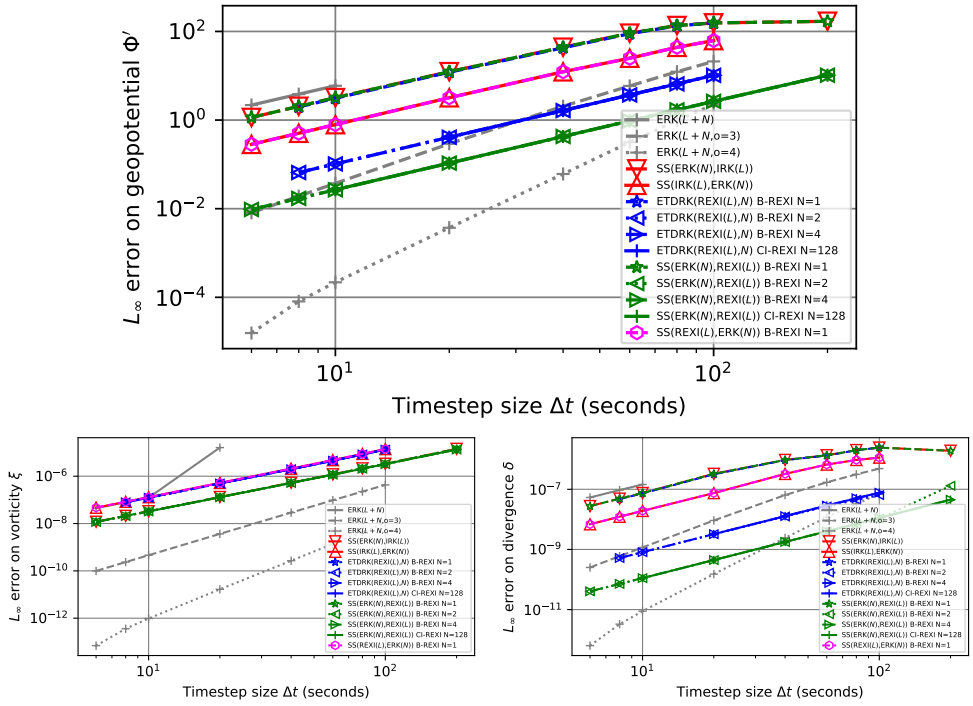

FIG. 6.6. *Studies **using non-REXI methods** (using direct exponentiation) on all prognostic variables with error vs. time step size for the barotropic instability benchmark. (ETDRK data points are missing due to 1h timeouts of the job.)*

to their computationally efficient way. In particular, the classical 4th-order accurate ERK method provides excellent results across all prognostic variables.

A closer look at the geopotential Φ' errors shows that the B-REXI-based methods with N=2 poles are to be preferred compared to all other methods. Again, the ETDRK method shows no real benefits.

Investigating the vorticity η leads to a different interpretation: Now, the implicit Strang-split method provides the best results which can be easily explained by the situation that the exponential treatment of the L_c term did not lead to any beneficial results already in the error vs. time step size plots and additional computational time is required here. Finally, ETDRK are literally the worst in here, not paying off at all.

The errors on the divergence δ show similar results compared to the geopotential, which is why we skip a detailed discussion here.

6.2.8. Summary of PDE results. The CI-REXI method with $N = 128$ poles is not beneficial at all compared to B-REXI with $N = 2$ poles. Using only $N = 2$ poles with the B-REXI method already provides the best results, and no improvement can be gained by adding more poles. This is actually quite surprising, with expectations of exponential integration methods to always provide significantly better results. However, using such a higher-order approximation seems to provide sufficient accuracy so that the errors from the splitting approach dominate the overall errors.

We would like to emphasize that all the statements are specific to the SWE on the rotating sphere PDE and should not be generalized.

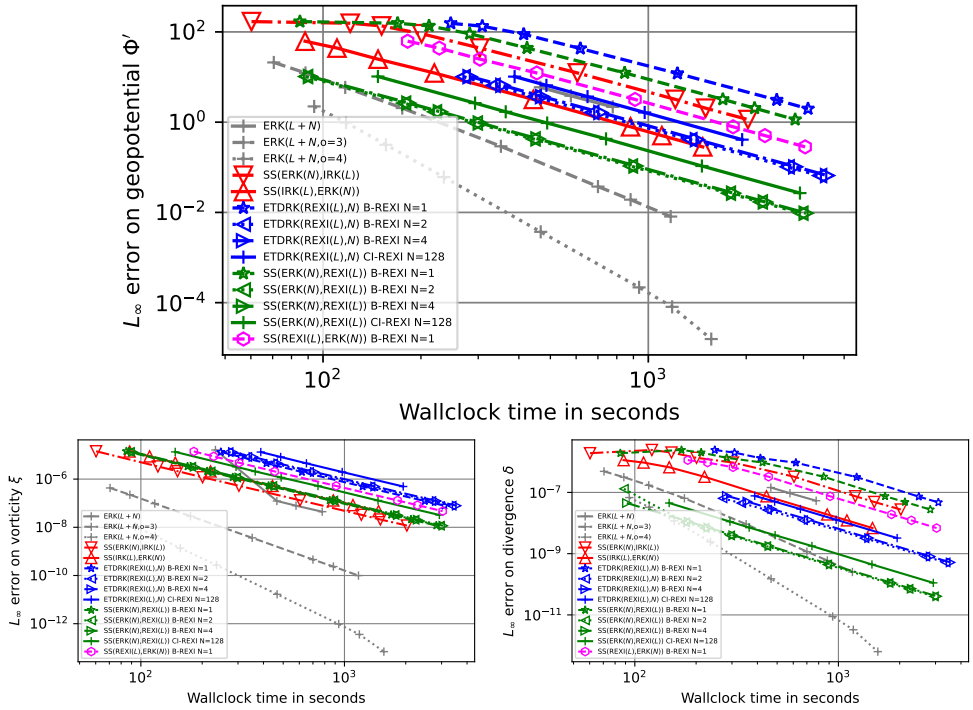

FIG. 6.7. *Studies **using non-REXI methods** (using direct exponentiation) on all prognostic variables with error vs. time step size for the barotropic instability benchmark. (ETDRK data points are missing due to 1h timeouts of the job.)*

7. Summary and Conclusions. Exponential integration methods are considered to be a way to integrate with high efficiency. As part of that, φ functions need to be solved, which turn out to be computationally rather challenging.

This paper investigated different ways to approximate φ functions with rational approximations of exponential integration (REXI). The coefficients of REXI methods can be derived in many ways and we introduced a generalized REXI approach, finally allowing to express many different methods in this way. We showed this for the Butcher/Bickard-based REXI, Cauchy Contour integration based REXI and T(erry)-REXI method. All methods have been introduced in a way making its capabilities and limitations easily graspable. With respect to physical properties, the T-REXI method requires special treatment for (quasi-)stationary modes and became obsolete with CI-REXI. In addition, we derived an elegant way to compute higher-order φ functions based on REXI coefficients for lower-order φ.

An in-depth investigation of the approximation quality of each REXI method has been conducted including an explanation of numerical issues for all of the methods. Next, we put it into the context of time integration methods. We first used linear ODEs where we studies and discussed properties of stability, convergence and also the filtering capabilities. Second, we performed in-depth studies using the nonlinear shallow-water equations on the rotating sphere. Surprisingly, the best REXI method turned out B-REXI with only $N = 2$ poles, leading to a significant reduction of computational effort compared to former REXI methods in this context using $N = 128$ poles. Consequently, regarding demands on computational resources, B-REXI showed a reduction of a factor of 64 compared to previous work. This also means that a

683 higher-order implicit Runge-Kutta method is competitive to traditional exponential
684 integration methods for this PDE.

685 **Acknowledgements.** Both authors like to thank Pedro S. Peixoto for pointing
686 out the potential relation of exponential integration methods to Laplace transforms
687 and Peter Lynch's work in this context. Martin Schreiber is grateful to NCAR for
688 providing financial support and a very inspiring office space with a splendid view to the
689 flatirons, which strongly supported this work. Both authors thank Matthew Normile
690 for preliminary work as well as Finn Capelle and Raphael Schilling who indirectly
691 contributed to this work with the REXInsight software.
692 The authors gratefully acknowledge the Gauss Centre for SC e.V. (www.gauss-
693 centre.eu) for funding this project by providing computing time on the GCS Super-
694 computer SUPERMUC-NG at Leibniz Supercomputing Centre (www.lrz.de).

695 REFERENCES

696 [1] T. A. Bickart, *An Efficient Solution Process for Implicit Runge-Kutta Methods*, SIAM Journal
697 on Numerical Analysis, 14 (1977), pp. 1022–1027, https://doi.org/10.1137/0714069.
698 [2] J. C. Butcher, *Implicit Runge-Kutta Processes*, AMS, 18 (1964), pp. 50–64.
699 [3] J. C. Butcher, *On the implementation of implicit Runge-Kutta methods*, BIT, 16 (1976),
700 pp. 237–240, https://doi.org/10.1007/BF01932265.
701 [4] T. Buvoli, *A Class of Exponential Integrators Based on Spectral Deferred Correction*, (2015),
702 pp. 1–22, http://arxiv.org/abs/1504.05543.
703 [5] T. Buvoli, *A class of exponential integrators based on spectral deferred correction*, SIAM Jour-
704 nal on Scientific Computing, 42 (2020), pp. A1–A27, https://doi.org/10.1137/19M1256166.
705 [6] K. e. a. Calvin, *IPCC, 2023: Climate Change 2023: Synthesis Report. Contribution of Work-
706 ing Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on
707 Climate Change. IPCC, Geneva, Switzerland.*, tech. report, Intergovernmental Panel on
708 Climate Change (IPCC), July 2023, https://doi.org/10.59327/IPCC/AR6-9789291691647,
709 https://www.ipcc.ch/report/ar6/syr/ (accessed 2023-11-09). Edition: First.
710 [7] C. Clancy and P. Lynch, *Laplace transform integration of the shallow-water equations. Part
711 I: Eulerian formulation and Kelvin waves*, Quarterly Journal of the Royal Met. Society,
712 137 (2011), pp. 792–799, https://doi.org/10.1002/qj.793.
713 [8] C. Clancy and J. A. Pudykiewicz, *On the use of exponential time integration methods in
714 atmospheric models*, Tellus, Series A: Dynamic Meteorology and Oceanography, 65 (2013),
715 https://doi.org/10.3402/tellusa.v65i0.20898.
716 [9] R. Courant, H. Lewy, and K. Friedrichs, *Über die partiellen Differenzengleichungen der
717 mathematischen Physik*, Mathematische Annalen, (1932).
718 [10] S. M. Cox and P. C. Matthews, *Exponential time differencing for stiff systems*, Journal of
719 Computational Physics, 176 (2002), pp. 430–455, https://doi.org/10.1006/jcph.2002.6995.
720 [11] ECMWF, *The Strength of a Common Goal: A Roadmap To 2025*, (2016), https://www.ecmwf.
721 int/sites/default/files/ECMWF_Roadmap_to_2025.pdf.
722 [12] O. G. Ernst and M. J. Gander, *Why it is Difficult to Solve Helmholtz Problems with Classical
723 Iterative Methods*, vol. 83, 2012, https://doi.org/10.1007/978-3-642-22061-6.
724 [13] M. Frigo, *A Fast Fourier Transform Compiler*, 1999.
725 [14] J. Galewsky, R. K. Scott, and L. M. Polvani, *An initial-value problem for testing numerical
726 models of the global shallow-water equations*, Tellus, Series A: Dynamic Meteorology and
727 Oceanography, 56 (2004), pp. 429–440, https://doi.org/10.1111/j.1600-0870.2004.00071.x.
728 [15] J. Hack and R. Jakob, *Description of a global shallow water model based on the spectral
729 transform method*, vol. NCAR/TN-34, 1992. Publication Title: NCAR Technical Note.
730 [16] E. Hairer, S. Norsett, and G. Wanner, *Solving ordinary differential equations I: Nonstiff
731 problems*, 1987.
732 [17] E. Harney and P. Lynch, *Laplace transform integration of a baroclinic model*, Quarterly
733 Journal of the Royal Met. Society, (2019), pp. 347–355, https://doi.org/10.1002/qj.3435.
734 [18] T. S. Haut, T. Babb, P. G. Martinsson, and B. A. Wingate, *A high-order time-parallel
735 scheme for solving wave propagation problems via the direct construction of an approximate
736 time-evolution operator*, IMA Journal of Numerical Analysis, 36 (2016), pp. 688–716, https:
737 //doi.org/10.1093/imanum/drv021.
738 [19] M. Hochbruck and A. Ostermann, *Exponential integrators*, vol. 19, 2010, https://doi.org/

10.1017/S0962492910000048. ISSN: 09624929 Publication Title: Acta Numerica.
[20] K. R. Jackson and S. P. Nørsett, *The potential for parallelism in RK methods. Part 1: RK formulas in standard form*, SIAM Journal on Numerical Analysis, 32 (1995), pp. 49–82.
[21] W. Kutta, *Beitrag zur näherungsweisen integration totaler Differentialgleichungen*, Z. Math. Phys., (1901), pp. 435–453.
[22] J. L. Lions, Y. Maday, and G. Turinici, *Résolution d'EDP par un schéma en temps "pararéel"*, Comptes Rendus de l'Academie des Sciences - Series I: Mathematics, 332 (2001), pp. 661–668, https://doi.org/10.1016/S0764-4442(00)01793-6. ISBN: 0764-4442.
[23] P. Lynch, *Filtering integration schemes based on the Laplace and Z transforms*, 1991. ISSN: 00270644 Issue: 3 Pages: 653–666 Publication Title: Monthly Weather Review Volume: 119.
[24] M. Minion, R. Speck, M. Bolten, M. Emmett, and D. Ruprecht, *Interweaving PFASST and Parallel Multigrid*, (2014), pp. 1–20, https://doi.org/10.1137/14097536X.
[25] C. Moler and C. Van Loan, *Nineteen Dubious Ways to Compute the Exponential of a Matrix, Twenty-Five Years Later*, SIAM Review, 45 (2003), pp. 3–49, https://doi.org/10.1137/S00361445024180.
[26] J. Niesen and W. M. Wright, *A Krylov subspace algorithm for evaluating the phi-functions appearing in exponential integrators*, ACM Transactions on Mathematical Software, 38 (2012), p. 20, https://doi.org/10.1145/2168773.2168781.
[27] S. A. Orszag, *On the Elimination of Aliasing in Finite-Difference Schemes by Filtering High-Wavenumber Components*, (1971), p. 1. 2/3 rule.
[28] A. C. Rojas Mendoza and P. d. S. Peixoto, *Numerical Solution of Ordinary Differential equations using Laplace transform integration*, master's thesis, Univ. de São Paulo, 2020.
[29] C. Runge, *Über die numerische Auflösung von Differentialgleichungen*, Mathematische Annalen, 46 (1895).
[30] N. Schaeffer, *Efficient spherical harmonic transforms aimed at pseudospectral numerical simulations*, Geochemistry, Geophysics, Geosystems, 14 (2013), pp. 751–758.
[31] T. Schmelzer and L. N. Trefethen, *Evaluating matrix functions for exponential integrators via carathéodory-fejér approximation and contour integrals*, Electronic Transactions on Numerical Analysis, 29 (2007), pp. 1–18, https://doi.org/10.1007/s00586-009-1106-6.
[32] M. Schreiber and R. Loft, *A parallel time integrator for solving the linearized shallow water equations on the rotating sphere*, Numerical Linear Algebra with Applications, 26 (2018).
[33] M. Schreiber, P. S. Peixoto, T. Haut, and B. Wingate, *Beyond spatial scalability limitations with a massively parallel method for linear oscillatory problems*, International Journal of High Performance Computing Applications, 32 (2017), pp. 913–933.
[34] M. Schreiber, N. Schaeffer, and R. Loft, *Exponential integrators with parallel-in-time rat. approx. for the shallow-water equations on the rotating sphere*, Parallel Computing, 85 (2019), pp. 56–65, https://doi.org/10.1016/j.parco.2019.01.005. Publisher: Elsevier B.V.
[35] D. Sheen, I. H. Sloan, and V. Thomée, *A parallel method for time-discretization of parabolic problems based on contour integral representation and quadrature*, Mathematics of Computation, 69 (1999), pp. 177–196.
[36] Strang, Gilbert, *On the Construction and Comparison of Difference Schemes*, SIAM Journal on Numerical Analysis, 5 (1968), pp. 506–517, https://doi.org/10.1137/0705041.
[37] C. Temperton, *Treatment of the Coriolis Terms in Semi-Lagrangian Spectral Models*, (1995).
[38] J. Thuburn, T. Ringler, W. Skamarock, and J. Klemp, *Numerical representation of geostrophic modes on arbitrarily structured C-grids*, Journal of Computational Physics, 228 (2009), pp. 8321–8335, https://doi.org/10.1016/j.jcp.2009.08.006.
[39] M. Tokman, *Efficient integration of large stiff systems of ODEs with exponential propagation iterative (EPI) methods*, Journal of Computational Physics, 213 (2006), pp. 748–776.
[40] M. Tokman, *A new class of exponential propagation iterative methods of Runge-Kutta type (EPIRK)*, Journal of Computational Physics, 230 (2011), pp. 8762–8778, https://doi.org/10.1016/j.jcp.2011.08.023. Publisher: Elsevier Inc.
[41] L. N. Trefethen and J. A. C. Weideman, *The Exponentially Convergent Trapezoidal Rule*, 56 (2014), pp. 385–458.
[42] L. N. Trefethen, J. A. C. Weideman, and T. Schmelzer, *Talbot quadratures and rational approximations*, BIT Numerical Mathematics, 46 (2006), pp. 653–670.
[43] J. Virieux, A. Asnaashari, R. Brossier, L. Métivier, A. Ribodetti, and W. Zhou, *An introduction to full waveform inversion*, 2014, https://doi.org/10.1190/1.9781560803027.entry6. Publication Title: Encyclopedia of Exploration Geophysics.
[44] H. Wang, J. P. Boyd, and R. A. Akmaev, *On computation of Hough functions*, Geoscientific Model Development, 9 (2016), pp. 1477–1488, https://doi.org/10.5194/gmd-9-1477-2016. ISBN: 1471239314.