A Unification and Investigation of Rational Approximation of Exponential Integration Methods

Martin Schreiber, Jed Brown

A Unification and Investigation of Rational Approximation of Exponential Integration Methods

Martin Schreiber[a,c,d,*], Jed Brown[b]

[a] *Univ. Grenoble Alpes / Laboratoire Jean Kuntzmann / CNRS/ Inria, Grenoble, France*
[b] *Department of Computer Science, University of Colorado, Boulder, USA, USA*
[c] *Inria AIRSEA team , 700 Avenue Centrale, Grenoble, 38058, France*
[d] *TUM School of Computation, Information and Technology, Boltzmannstrasse 3, Garching b. München, 85748, Germany*

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 of parallelizing 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.

We present the "unified REXI" method 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. In the present work, we target the hyperbolic problems considered to be a particularly challenging task. We provide for the first time a deep numerical investigation, discussion, and comparison of all these methods. In particular, we account for numerical problems and, if possible, workarounds for them. Finally, we demonstrate and compare the performance of REXI with off-the-shelf time integrators using the nonlinear shallow-water equations on the rotating sphere on a high-performance computing system.

While previous REXI studies have focused on exposing more parallelism to enable faster time to solution, we also consider computing resource efficiency at prescribed accuracy and find that diagonalized lower-order Gauss Runge-Kutta methods (formulated as REXI) are compelling highly efficient methods leading to a 64× reduction of the required computational resources compared to existing work.

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 hyperbolic PDEs, which are ubiquitous in applications ranging from daily weather forecasting [11] to full waveform inversion [46]. Integration of such systems is sequential in time using conventional methods such as explicit and diagonally implicit Runge-Kutta [33, 24]: Without special structure [23], 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 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

*Corresponding author

increasing number of time steps to satisfy the Courant-Friedrichs-Lewy (CFL) constraint [8] 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 [5] or design/manufacturing timelines. Parallelism in the time dimension [16] 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 linear 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 the state variables and operators in space leads to $\frac{dU(t)}{dt} = LU(t)$ with L the discrete linear operator as a matrix and $U(t)$ the discrete state variables at time t as a vector. 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

$$U(t + \Delta t) = \exp(\Delta t L) U(t) \tag{1}$$

which avoids any time discretization, hence, errors in this form. In this form, a rational approximation of $\exp(\Delta t L)$ can be used to only approximate a particular space related to time step size Δt and the spectrum of L which can be written as

$$U(t + \Delta t) \quad \approx \quad \gamma U(t) + \underbrace{\sum_{n=1}^{N} \beta_n (\Delta t L - \alpha_n)^{-1} U(t)}_{\text{Parallelization}} . \tag{2}$$

with (typically) complex valued REXI coefficients α_n, β_n and real-valued γ. This form provides a unified way to reformulate a variety of other time integration schemes which is the reason why we call it "unified REXI" formulation and use it in the remainder of our work.

2. Related Work

2.1. Exponential integration

Exponential integration methods are formulated for nonlinear systems written as

$$\frac{dU(t)}{dt} = LU(t) + N(U(t)), \tag{3}$$

where the linear part L is intended to capture the "fast" dynamics (limiting the time step size for stability reasons) and N is the remaining nonlinear part. An exact solution for advancing this split equation over a finite time interval is given by

$$U(t + \Delta t) = \exp(\Delta t L) U(t) + \int_0^{\Delta t} \exp((\Delta t - \tau) L) N(U(t + \tau)) d\tau. \tag{4}$$

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., [28, 22, 10, 12, 31]) 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 [9])

$$U(t + \Delta t) \approx \varphi_0(\Delta t L) U(t) + \Delta t \varphi_1(\Delta t L) N(U(t)) \tag{5}$$

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, related to the present work, point out the φ functions omnipresent in higher-order exponential integration methods, which are given, e.g., by

$$\varphi_{i+1}(Z) \quad = \quad (\varphi_i(Z) - \varphi_i(0)) \, Z^{-1} \quad \text{for } i \geq 0. \tag{6}$$

An investigation of all different varieties of discretizations of exponential integrators incorporating the nonlinearities is beyond the scope of this work, and we continue with an in-depth investigation of REXI approximations 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 $U(t+\Delta t) = \varphi_0(\Delta t L)U(t) = \exp(\Delta t L)U(t)$. However, the computational complexity of computing these terms can be tremendous and triggered the development of various ways to tackle this challenge [28]. REXI is one of such candidates which is deeply investigated in the present work.

2.2. Krylov subspaces

The exponential can be approximated using Krylov subspace solvers (see [29, 42, 43, 7]) 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. We would like to point out that REXI also requires solvers for linear systems of equations, which might be, again, based on Krylov subspace solvers.

2.3. Laplace transforms

One of the earliest REXI formulations for hyperbolic linear PDE time integrators is related to the Laplace transformation (cf. [26, 6]). 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 [45]. More recently, time integration based on Laplace transformations with a circle-based Cauchy contour integration have been more intensively studied in [32] with ODEs. 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. We point out that the same approaches developed from the exponential integration perspective, partly used in the present work, could also be taken from the Laplace transform perspective.

2.4. Parallel-in-time

Finally, exponential integration plays an important role in parallel-in-time methods, which seek to further reduce time to solution beyond the strong-scaling limit from spatial parallelism. 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 [25] and PFASST [27]); 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 can be seen as an invasive parallel-in-time algorithm (see [36]) since it requires efficient complex-valued solvers for each REXI term, with certain challenges as discussed later.

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 to obtain $\frac{\partial U(t)}{\partial t} = LU(t) = Q\Lambda Q^{-1}U(t)$. 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 (2) becomes

$$u_i(t+\Delta t) = \exp(\Delta t \lambda_i)u_i(t) \approx \gamma u_i(t) + \sum_{n=1}^{N} \beta_n (\Delta t \lambda_i - \alpha_n)^{-1} u_i(t). \tag{7}$$

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. From the PDE perspective, this can be directly related to hyperbolic and parabolic PDEs, respectively.

We note that it is possible to reduce the workload by a factor of approximately two for real-valued operators L when the poles α consist of complex conjugate pairs (see [26, 21]). Since this optimization does not change the relative performance of the methods we consider here, for simplicity, we do not apply it.

3.1. REXI-derived higher-order φ forms

Particular higher-order exponential time integrators such as (5) require evaluations of $\varphi_{i>0}$ as a basic building block in these higher-order integrators. REXI coefficients for these functions are so far computed with methods tailored to them, see [21, 35]. We briefly present an new 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

$$\begin{aligned}
\varphi_{i+1}(x) &= \frac{\varphi_i(x) - \varphi_i(0)}{x} && (8)\\
&\approx \frac{\gamma + \sum_n \frac{\beta_n}{x-\alpha_n} - \varphi_i(0)}{x} && (9)\\
&= \sum_n \left(\frac{\beta_n}{\alpha_n (x - \alpha_n)} \right) + \frac{1}{x} \left(\underbrace{\sum_n \left(\frac{\beta_n}{-\alpha_n} \right) + \gamma - \varphi_i(0)}_{=0} \right) = \sum_n \frac{\frac{\beta_n}{\alpha_n}}{x - \alpha_n}. && (10)
\end{aligned}$$

The cancellation of the terms is a consequence of the stationary modes requiring $\sum_n \left(\frac{\beta_n}{-\alpha_n} \right) + \gamma = \varphi_i(0)$. This approach provides for the first time a way to compute coefficients for $\varphi_{i>0}$ using a REXI method based on Butcher-table diagonalization (see Sec. 4.1).

3.2. Linear solvers for REXI terms

REXI replaces the original exponential term with independent systems of linear equations to be solved, hence, requires efficient linear solvers. Over the last decades, this efficiency aspect for particular α_n terms turned out to be a very challenging task for arbitrary grids. E.g., in the context of shallow-water equations, this results for some α_n 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., [13]). The present work is based on solvers which have been particularly tailored to this challenge which exploit spherical harmonics. Depending on which terms are involved, this leads to diagonal (only gravity modes) but also non-diagonal pentadiagonal (gravity modes and Coriolis effect) linear equations in spectral space which allows an efficient solution of each individual REXI term (see [35] for more information).

4. REXI methods for hyperbolic/oscillatory systems

These sections provide an overview and derivation of different methods and their translation to the unified REXI representation in Eq. (2). Since approximating diffusive problems is relatively straightforward, we focus on purely oscillatory problems with $\lambda \in i\,\mathbb{R}$. Here, we focus on methods which are the most promising candidates for hyperbolic problems and present them in characteristic form in Eq. (7) without loss of generality for them to hold in system form given by Eq. (2).

Throughout the numerical studies, we will use the error

$$e(z) = \left| \gamma + \sum_n \beta_n (z - \alpha_n)^{-1} - \exp(z) \right| \qquad (11)$$

to compute the deviation from $\varphi_0(z) = \exp(z)$ with $z = \lambda\Delta t$ denoting the point on the complex plane to evaluate. Targeting hyperbolic problems, the REXI methods we consider have complex-conjugate poles α, thus $e(z) = e(\bar{z})$ and so we only plot errors for $Im(z) \geq 0$.

The source code to reproduce all ODE-related plots is made publicly available under this url: https://doi.org/10.5281/zenodo.14917194 .

4.1. B-REXI: Butcher/Bickart

A Butcher table [2] provides a canonical representation of N-stage Runge-Kutta methods [33, 24] in terms of a matrix $A \in \mathbb{R}^{N\times N}$ and completion vector $\mathbf{b} \in \mathbb{R}^N$, with $\mathbf{c} = A\mathbf{1}$ determining the abscissa. Given these coefficients, 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

$$y_s = u_n + \Delta t \sum_{j=1}^{N} A_{sj} f(t + c_j\Delta t, y_j), \quad i = 1, ..., N \tag{12}$$

followed by evaluating the completion formula $u_{n+1} = u_n + \Delta t \sum_{j=1}^{N} b_j f(t + c_j\Delta t, y_j)$ with the time step size Δt, and the vector of stage solutions $\mathbf{y} = \{y_j\}_{j=1}^{N}$.

We can simplify this for linear autonomous equations by choosing characteristic variables, in which case $f(t, u) = \lambda u$, and the stage equations (12) reduce to $\mathbf{y} = \mathbf{1}u + \Delta t\lambda A\mathbf{y}$ and

$$u_{n+1} = \underbrace{\left[1 + \Delta t\lambda \mathbf{b}^T (I - \Delta t\lambda A)^{-1}\mathbf{1}\right]}_{R(\Delta t\lambda)} u_n, \tag{13}$$

where we have identified the stability function $R(z) \approx \exp(z)$ which should approximate the exponential function. Next, we relate this to the unified REXI form (see Eq. (7)).

4.1.1. Derivation

We now show that, for linear equations, 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 [18]), we can rewrite Eq. (13) as

$$u_{n+1} = \left[1 + \Delta t\lambda \mathbf{b}^T E(I - \Delta t\lambda D)^{-1} E^{-1}\mathbf{1}\right] u_n. \tag{14}$$

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

$$u_{n+1} = \left[1 + \Delta t\lambda \underbrace{\mathbf{b}^T EW^{-1}}_{\tilde{\mathbf{b}}^T} (I - \Delta t\lambda D)^{-1} \underbrace{WE^{-1}\mathbf{1}}_{\mathbf{1}}\right] u_n, \tag{15}$$

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

$$\begin{aligned} u_{n+1} &= u_n + \Delta t\tilde{\boldsymbol{b}}^T \left(-(\Delta tD)^{-1}\right) \left(I + (\Delta t\lambda D - I)^{-1}\right) \mathbf{1}u_n \\ &= \underbrace{\left(1 - \tilde{\boldsymbol{b}}^T D^{-1}\mathbf{1}\right)}_{\gamma} u_n + \underbrace{\left(-\tilde{\boldsymbol{b}}^T D^{-2}\right)}_{\beta^T} \left(\Delta t\lambda - \underbrace{D^{-1}}_{\mathrm{diag}(\alpha)}\right)^{-1} \mathbf{1}u_n. \end{aligned} \tag{16}$$

Finally, we can write this in the unified REXI formulation (2) with

$$\gamma = 1 - \tilde{\boldsymbol{b}}^T D^{-1}\mathbf{1} \qquad \beta^T = -\tilde{\boldsymbol{b}}^T D^{-2} \qquad \alpha = \mathrm{diag}\left(D^{-1}\right) \tag{17}$$

to which we will refer to as the B-REXI method. To summarize, 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 [23], which establishes that diagonal (parallel) RK methods can be no more than second order accurate for nonlinear problems.

We like to point out that our new method for computing higher-order $\varphi_{i>0}$ terms presented in Eq. (8) makes B-REXI also applicable to higher-order exponential integration methods (see, e.g., Eq. (5)) for the first time. Although the present paper investigates REXI for linear problems, [19, IV.8 (p. 121-122)] shows how to use this B-REXI method in the context of nonlinear problems with a block diagonalization.

4.1.2. Error studies

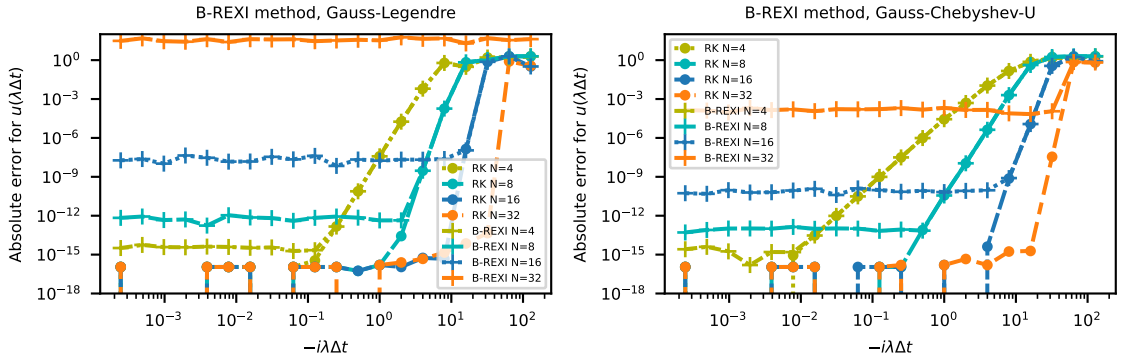

Figure 1: Error studies for the B-REXI method with (a) Gauss-Legendre and (b) Chebyshev quadrature points for the error given in Eq. (11). Each color refers to the same number of stages. Markers refer to B-REXI or RK form. The non-diagonalized version provides significantly better results compared to the diagonalized version. In particular, results with B-REXI using $N = 32$ or more stages suffer from significant defects in the solution.

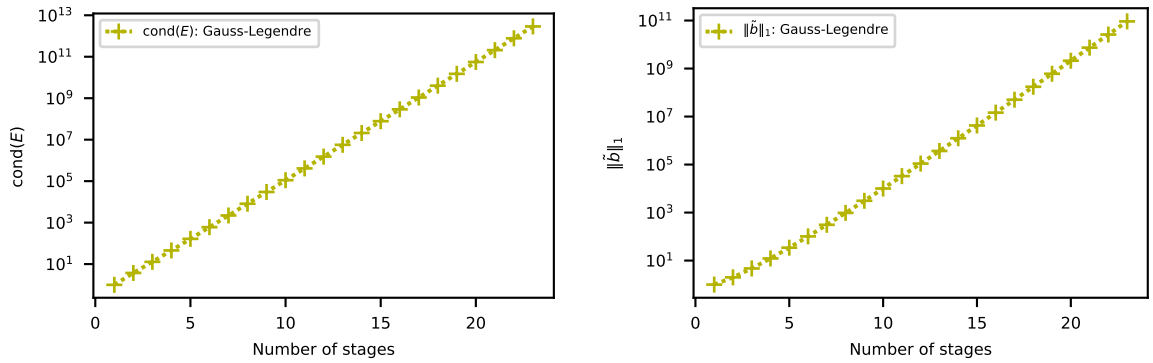

Figure 2: Condition number of the eigenbasis E for the B-REXI method on Gauss-Legendre collocation points (left) and 1-norm of completion vector $\tilde{\boldsymbol{b}}$ for the diagonalized method (right). The rounding errors incurred by the exponential growth precludes use of this approach for many stages.

We choose the Gauss-Legendre and Chebyshev quadrature points for the error studies based on Eq. (11), with results given in Figure 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 the diagonalization of Eq. (15) or via the 1-norm of the completion vector $\tilde{\boldsymbol{b}} = W^{-1}E^T\boldsymbol{b}$, as shown in Figure 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 $\boldsymbol{b}$); 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 methods with lower stage counts are remarkably efficient compared to the other (better-conditioned) families with higher stage counts.

4.1.3. Relation to Crank-Nicolson

We close this section by showing the relation between the B-REXI approximation with the Gauss-Legendre quadrature using just a simple quadrature point centered at the interval. This leads to the terms $\gamma = -1$, $\alpha = 2$ and $\beta = -4$ which yields the REXI approximation $\exp(x) \approx -1 + \frac{-4}{x-2} = \frac{1+\frac{1}{2}x}{1-\frac{1}{2}x}$ where the last equation shows the relation to the Crank-Nicolson formulation with a midpoint rule (the forward Euler is on the nominator and backward Euler on the denominator for $x = \Delta t L$ and $\Delta t = 0.5$). In other words, the REXI method allows time integrating with a Crank-Nicolson formulation using just a single REXI term. This will also account later for numerical results of B-REXI equivalent to the Crank-Nicolson method.

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 [21]. 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

$$\psi_h(x) = (4\pi)^{-\frac{1}{2}} e^{-x^2/(4h^2)} \approx Re\left(\sum_{k=-W}^{W} \frac{\omega_k}{i\frac{x}{h} + (\mu + ik)} \right). \tag{18}$$

which can be approximated up to numerical double precision (see [21] for coefficients) using $W = 11$, hence $L = 2W + 1 = 23$ terms in total. 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 $x \in [-Mh; Mh]$ in Fourier space, yielding $\exp(ix) \approx \sum_{k=-M}^{M} \nu_k \psi_h(x + kh)$. Both steps are then combined, resulting in the approximation $Re\,(\exp(ix)) \approx \sum_{n=-M-W}^{M+W} Re\left(\beta_n^{Re}(ix + \alpha_n)^{-1}\right)$ where we only showed the Re one. Combining the real and imaginary approximation then results in

$$\exp(ix) \quad \approx \quad \sum_{n=-M-W}^{M+W} \beta_n (ix - \alpha_n)^{-1} \tag{19}$$

which resembles the unified REXI formulation with $\gamma = 0$, and x related directly to the imaginary value. 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 [21]) or directly with the φ_0 REXI coefficients as derived in §3.1. It is important to note that this approximation was derived only for purely oscillatory functions and, hence, does not directly apply to problems with non-zero real eigenvalues components.

4.2.2. Error studies

We investigate the errors (see Eq. (11)) of the T-REXI method in Figure 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 [21]) 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.

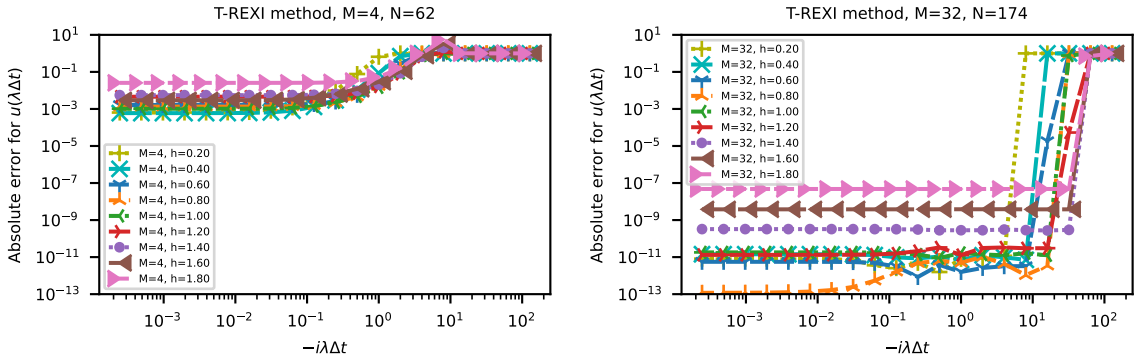

Figure 3: Error studies for the T-REXI method for different M and h values. M relates to the number of Gaussian functions via $2M+1$ to approximate an oscillation with $N = 2(2M+L)$ number of REXI terms. Left image: We can observe a very high error for a low number of Gaussian basis bumps, which cannot 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 $h \approx 0.8$.

4.3. CI-REXI: Cauchy Contour Integral method

Cauchy Contour Integral (CI) methods offer yet another way to infer the REXI coefficients (see e.g. [44, 4, 37]). We start with the general CI equation given by

$$g(x) = \frac{1}{2\pi i} \oint_\Gamma \frac{g(z)}{z-x} dz \tag{20}$$

where $g(x)$ is one of the analytic φ_i functions from Eq. (6), Γ the contour enclosing the eigenvalue λ for ODEs and all eigenvalues of L for PDEs. For PDEs, not enclosing particular eigenvalues can be also used for a filtering as discussed later.

We like to hint that this does not require explicit computations of the eigenvalues: E.g., for wave equations with spherical harmonics, the largest eigenvalues can be computed analytically. If this is not possible, an existing off-the-shelf time integration scheme with a known stability region can be used by relating the spectral radius of L to the maximum stable time step size and the known stability region.

4.3.1. REXI Derivation

Here, we focus on 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

$$g(x) = \frac{1}{2\pi i} \oint_0^1 \frac{g(\sigma(w))}{\sigma(w)-x} \sigma'(w) dw = \oint_0^1 \frac{i(2\pi)^{-1} g(\sigma(w))\sigma'(w)}{x-\sigma(w)} dw. \tag{21}$$

With the exponentially fast converging trapezoidal rule on periodic boundaries (see [45]) and N trapezoidal points in total, we obtain

$$g(x) \approx \frac{1}{N} \sum_{n=1}^{N} \frac{i(2\pi)^{-1} g(\sigma(w_n))\sigma'(w_n)}{x-\sigma(w_n)} \quad \text{with} \quad w_n = \frac{n}{N}. \tag{22}$$

Again, we can infer a unified REXI formulation (see Eq. (2)) by setting

$$\alpha_n = \sigma(w_n), \qquad \beta_n = \frac{i g(\sigma(w_n))\sigma'(w_n)}{N 2\pi}, \qquad \gamma = 0. \tag{23}$$

A study of all kinds of contour shapes (rectangle, polygonal shapes, etc.) is beyond the scope of this work. To investigate CI-REXI with at least one contour targeting oscillatory problems, we choose the ellipse

contour as the most natural and trivial one. It 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, leading to REXI coefficients

$$\alpha_n = \sigma(w_n), \qquad \gamma = 0, \tag{24}$$

$$\beta_n = \frac{i}{N} \exp(\sigma(w)) \left(-R_x \sin(iw2\pi) + iR_y \cos(iw2\pi)\right). \tag{25}$$

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. The ellipse will be used for numerical studies to show its superiority to another REXI method. But before that, we will use the circle contour to discuss highly relevant numerical properties in the next section.

4.3.2. Characterization and numerical issues

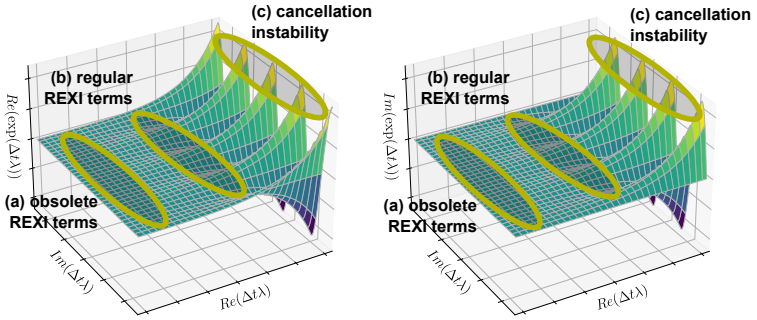

Figure 4: Complex plane for the real (left image) and imaginary (right image) value of $\exp(x)$. We highlight the different areas related to the different β characterizations.

In this section, we introduce for the first time a characterization of the REXI terms to which we refer to as the β characterization, with an overview in Figure 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 can identify three different cases: obsolete (rational terms to be truncated off), regular (rational terms which need to be computed), and cancellation-prone REXI terms (rational terms leading to numerical issues). We discuss these cases in the following.

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 $\lim_{\sigma \to -\infty} |\beta_n| = 0$. Once a particular β_n coefficient undershoots a threshold ϵ_β, the corresponding REXI term can be removed if $\beta_n < \bar{\epsilon}_\beta$ and $\bar{\epsilon}_\beta = \epsilon_\beta / N$. The last equation incorporates that a higher numerical resolution with an increase of N poles results in smaller values of the β weights. This is a crucial property for the CI-REXI method, as it allows for a significant reduction of the workload.

b) Regular REXI terms: This characterization refers to those REXI terms that should be incorporated in the approximation in the regular way. They are not small enough to be truncated off and not large enough to cause numerical issues (see next characterization).

c) Cancellation-prone REXI terms: These terms are related to the contour $Re(\sigma(x)) \to +\infty$. Approximating the exp function for larger positive real values leads to exponential increase of the magnitude of the β values (see Eq. (25)). In addition, different β values parallel to the imaginary axis oscillate between positive and negative values. Consequently, this results in severe cancellation errors in this region, hence, no contour should pass through this region.

Examples of different contours are provided in Figure 5. Each contour targets a particular problem. Starting with the left image, 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

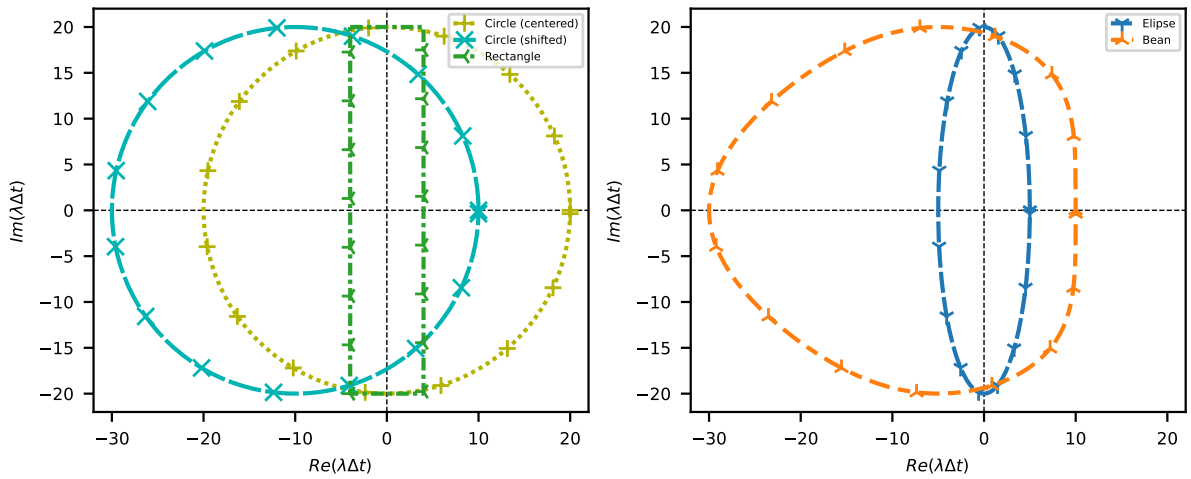

Figure 5: Selection of different contours which can be used with the CI-REXI method. Each contour results in different numerical properties, see text for more information.

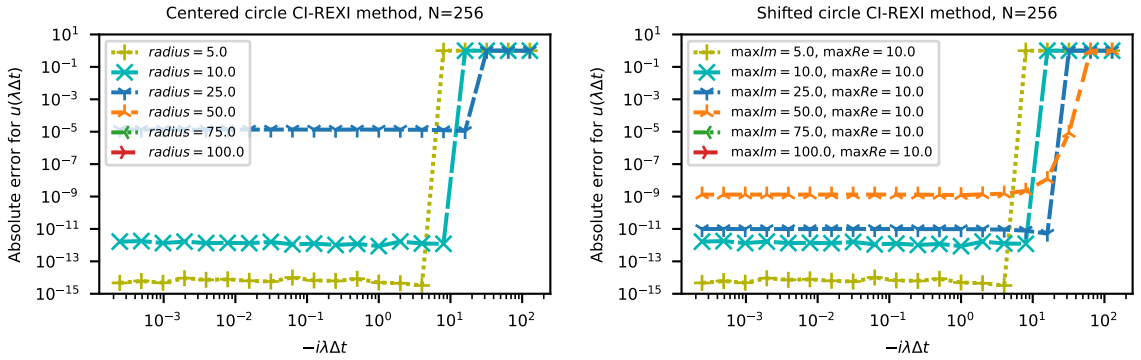

Figure 6: Error studies for the centered circle (left) and shifted circle (right) CI-REXI method. The centered circle suffers from cancellation effects for large radii, whereas the shifted circle limits these effects. In particular, for a larger imaginary spectrum to be approximated, adding more REXI terms leads to improved accuracy, which is not the case for the centered circle CI-REXI method.

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. Also a rectangle could be used to avoid this problem. The right image shows the ellipse contour which targets to approximate the spectrum on or close to the imaginary axis and the bean-shaped 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 which provided the best results for oscillatory problems.

4.3.3. Error studies

The first study is based on a circle centered at the origin, with studies for different radii. Results are given in Figure 6 with plots based on a fixed number of $N = 256$ REXI terms and the error given by Eq. (11). On the left handed image, 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 terms do not significantly improve the results (not shown here).

For numerical investigation of the shifted circle, rather than providing the coefficients explicitly, they are given implicitly by $maxIm$ and $maxRe$. In this way, $maxIm$ refers to the contour passing through the maximum approximation range for purely oscillatory systems and $maxRe$ refers to a parameter for avoiding

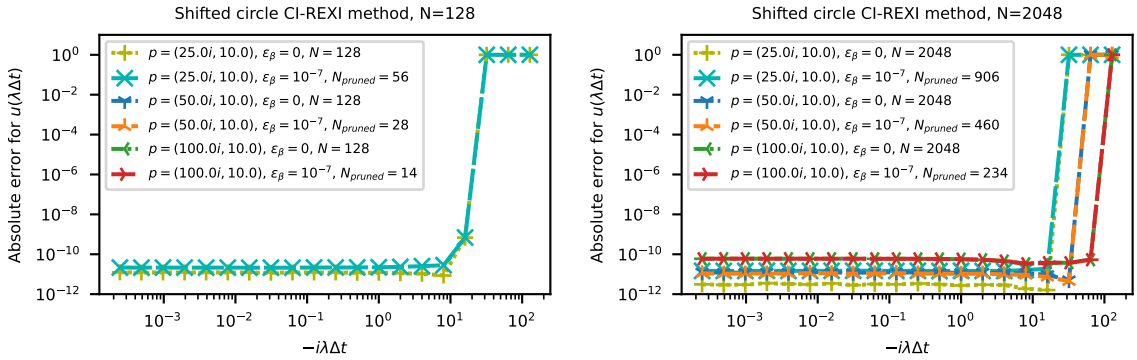

Figure 7: Error studies for the shifted circle with $N = 128$ REXI terms (left) and $N = 2048$ REXI terms (right) with different ϵ_β pruning values. We can observe significant reductions in the number of required REXI terms.

the region of cancellation effects (characterization (c)). Then, on the complex plane, the shifted circle is symmetric around the real axis and passes through the first point given by $(0, maxIm)$ and the second point given by $(maxRe, 0)$. On the right handed image, using such an shifted and enlarged circle, we observe to gain improved results that overcome previously discussed cancellation errors (see characterization (c) from above).

So far, we only investigated the error itself but neglected the total workload with results presented in Figure 7 where the parameters of the shifted circle are abbreviated with $p = (maxIm \cdot i, maxRe)$. Pruning β with ϵ_β (exploiting characterization (a)), we can reduce some workload significantly as depicted in Figure 7 for larger radii. For a moderate number of REXI terms $N = 128$ (left image, first two results), we observe that pruning results in a similar accuracy, hence, hardly impacting the results. Larger radii suffer from inaccuracies of the used quadrature, with errors outside the plotting range. For a larger number of REXI terms $N = 2048$ (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 from, e.g., $N = 2048$ terms to $N = 234$ terms.

5. Stability, normalization & filtering

So far, we have only investigated 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 briefly investigate such properties in this section for Dahlquists's ODE $\frac{du(t)}{dt} = \lambda u(t)$.

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 8): The stability reflects the A-stability of the Gauss-Legendre quadrature nodes on the entire left half plane. In particular, stability is given for the entire imaginary axis, a known property of this choice of quadrature points.

T-REXI (right image in Figure 8): 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 [21]).

CI-REXI: Finally, we look at the CI-REXI method based on Cauchy contour integral methods in Figure 9. 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

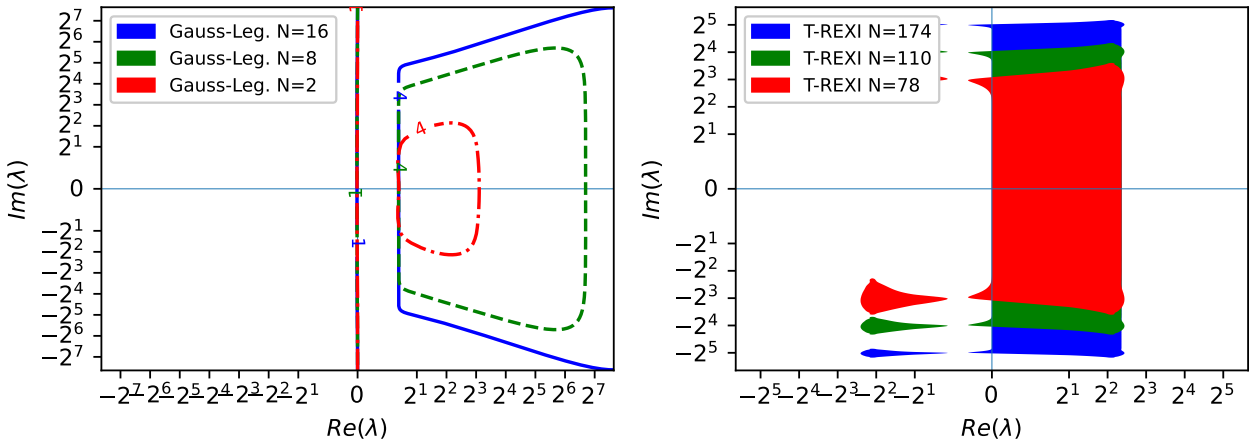

Figure 8: Areas in which the stability function $\langle R(z)\rangle \leq 1$ are stable, otherwise the method is unstable. For B-REXI (left): The stability region is exactly the left half plane for Gauss-Legendre methods. In addition, we plotted the contour for 4 (unstable). T-REXI (right): Filled areas refer to unstable regions. They in particular occur also on the left half plane at the boundaries of the approximation range.

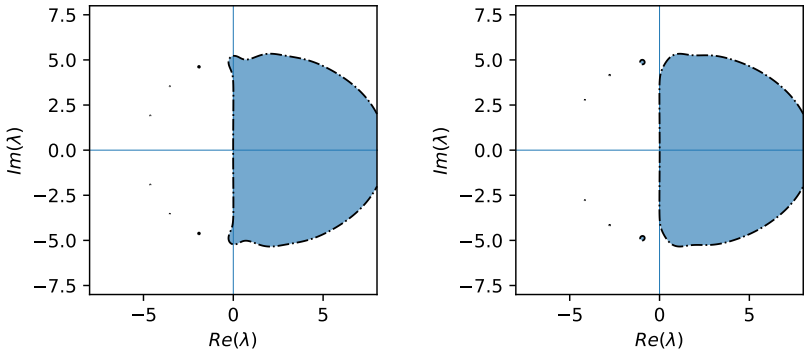

Figure 9: Stability plots for CI-REXI. The left image depicts the discrete contour points chosen so that one α pole lies on the imaginary axis, leading to instabilities if eigenvalues are in the proximity of this pole. The right image depicts a half-shifted variant of it which effectively avoids this instability issue, leading to a filtering in this range of spectrum.

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 can use the CI-REXI method as a filter for high frequencies by avoiding to place poles near the eigenvalues of the linear operator.

5.2. Normalization

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 T-REXI methods over multiple time steps. We will use Dahlquist's equation 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 modes which are nearby stationary balance. Such modes often play an important role for PDEs, e.g., for geostrophic balance in atmospheric simulations, and not preserving them might lead to spurious/parasitic modes.

An investigation of the absolute ODE errors at $t = 100$ is given in Figure 10. The left image shows that the T-REXI method suffers from significant defects in it. We account for these errors by numerical issues of coefficients for approximation of the Gaussian stored in the source-code which can lead to such round-off errors which consecutively accumulate in each time step. A normalization can be used to overcome this problem where stationary modes require $s = 1$ for $s = \sum_n \frac{\beta_n}{\alpha_n}$ and we can ensure this by simply rescaling

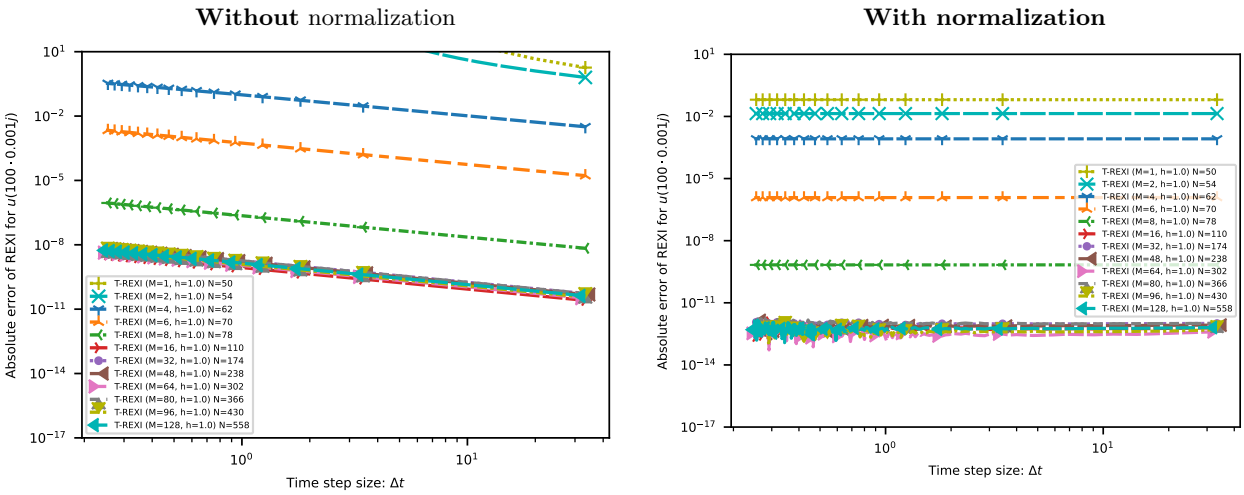

Figure 10: Error studies for different T-REXI methods of time step size Δt vs. absolute error at $u(t = 100)$. The left column shows errors without normalization, and the right column shows errors with normalization for near-stationary modes. Significant improvements can be observed for the T-REXI method. See the text for a detailed explanation of the results.

β_n so that $\beta_n^{\text{new}} = \frac{\beta_n}{s}$. The results in the right image depict a significant drop of errors from 10^{-8} to about 10^{-13} with 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.

B-REXI is not prone to this problem (based on a Taylor expansion) and the CI-REXI method provides sufficient double-precision accuracy of stationary modes by using already just a few REXI terms which is the reason why we skip investigating them here.

5.3. Filtering

This section briefly remarks on the filtering capabilities [26, 21] of the different REXI methods when applied to oscillatory/hyperbolic problems. A filter for differential equations reduces the amplitude of $\varphi_i(x)$ if $|x|$ becomes sufficiently large and damps out the corresponding modes stronger the larger $|x|$ becomes. We recap x to be linearly related to the spectrum of the linear operator and larger values (typically) relate to faster oscillations and higher spatial frequencies. A reduction of the amplitude of such fast oscillations is indeed a desirable property to filter out (setting them to or close to zero) the so-called "fast modes" for $|Im(x)|$ exceeding a given threshold related to the fast modes. Such filtering of fast frequencies can be found in, e.g., atmosphere [26] and ocean [38] simulations.

The **CI-REXI** method approximates the $\exp(x)$ function with exponentially fast convergence [45] as long as the contour encloses the x values. Consequently, highly accurate dispersion properties can be obtained in this case. Orthogonal to this, the CI-REXI method can also be used as a filter for high frequencies by placing the contour that some points of the spectrum (the fast modes) are exterior of the contour and avoiding contour quadrature points in the proximity of the spectrum, see right image in Fig. 9. Then, the CI-REXI approximation of $\exp(x)$ would lead to a damping of these modes outside the contour. The more distant they are from the contour, the stronger they are filtered due to the rational approximation of $\lim_{|x|\to\infty} \beta(x-\alpha)^{-1} = 0$. The damping effect is observed consistently in the results, but a full mathematical explanation remains an open question.

A direct application of the **T-REXI** would lead to instabilities at the boundaries of the approximation range, see right image in Fig. 8. A remedy to this was to add an additional filter proposed in [21] which, however, adds additional computational costs.

For the **B-REXI** method, we differentiate between two cases with quadrature nodes leading to $|\gamma| = 1$ (e.g., Gauss-Legendre) and $|\gamma| < 1$ (e.g., Gauss-Lobatto) for Eq. (7). In the first case $|\gamma| = 1$ we observe a

preservation of amplitude for $x \to \pm i\mathbb{R}$ due to vanishing rational terms and there is no filtering. In contrast, the $|\gamma| < 1$ case can lead to a reduction of the amplitude for $x \to \pm i\mathbb{R}$. As a closing remark, we did not observe any collocation methods leading to $|\gamma| > 1$.

6. Comparison of REXI methods

This section aims to provide guidance about which REXI method is best and will be explored in different ways. A 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.

The PDE studies are conducted with the SWEET software publicly accessible under the tag `https://gitlab.inria.fr/sweet/sweet/-/tree/paper_unification_rexi_methods`. Python code for the computations of REXI coefficients can be found in the subfolder `mule_local/python/rexi`.

6.1. ODE

We use Dahlquist's equation $u_t = \lambda u$ using $\lambda = 1i$ and simulation results at $t = 100$ with $u(0) = (1+i)/\sqrt{2}$ as an initial condition. We compare various REXI methods in Figure 11 with the total numbers of REXI coefficients given by N. The **B-REXI** method (left upper image) performs extremely well for small step sizes where only a few REXI terms are required. For larger time step sizes of $\Delta t \approx 10$, using 16 REXI terms 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 terms if only small time step sizes should be taken due to the rational approximation of the Gaussian. This improves once larger step sizes are taken.

In addition, we also investigated the **CI-EL-REXI** (right bottom image) method, which is a natural choice for purely oscillatory problems. Here, the parameters of the ellipse extension in the horizontal and vertical is given by R_x and R_y, respectively. 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 and also provides the filtering property discussed above. Consequently, for this purely hyperbolic problems and large time step sizes, it can be considered superior to the other methods.

6.2. PDE example

In this final section, we will investigate different REXI methods with the shallow-water equations (SWE) on the rotating sphere. Since including the T-REXI method would not provide any beneficial insight, since the CI-REXI method is computationally cheaper and provides additional benefits, see previous section, we skip this method in the following studies for sake of brevity.

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 inviscid SWE are given by

$$\frac{\partial}{\partial t}\begin{pmatrix} \Phi \\ \vec{V} \end{pmatrix} = \underbrace{\begin{pmatrix} -\overline{\Phi}\nabla\cdot\vec{V} \\ -\nabla\Phi \end{pmatrix}}_{L_g\text{: linear gravity}} + \underbrace{\begin{pmatrix} 0 \\ -f\vec{k}\times\vec{V} \end{pmatrix}}_{L_c\text{: linear Coriolis}} + \underbrace{\begin{pmatrix} -\nabla\cdot\left(\Phi'\vec{V}\right) \\ -\vec{V}\cdot\nabla\vec{V} \end{pmatrix}}_{N\text{: nonlinear term}} \tag{26}$$

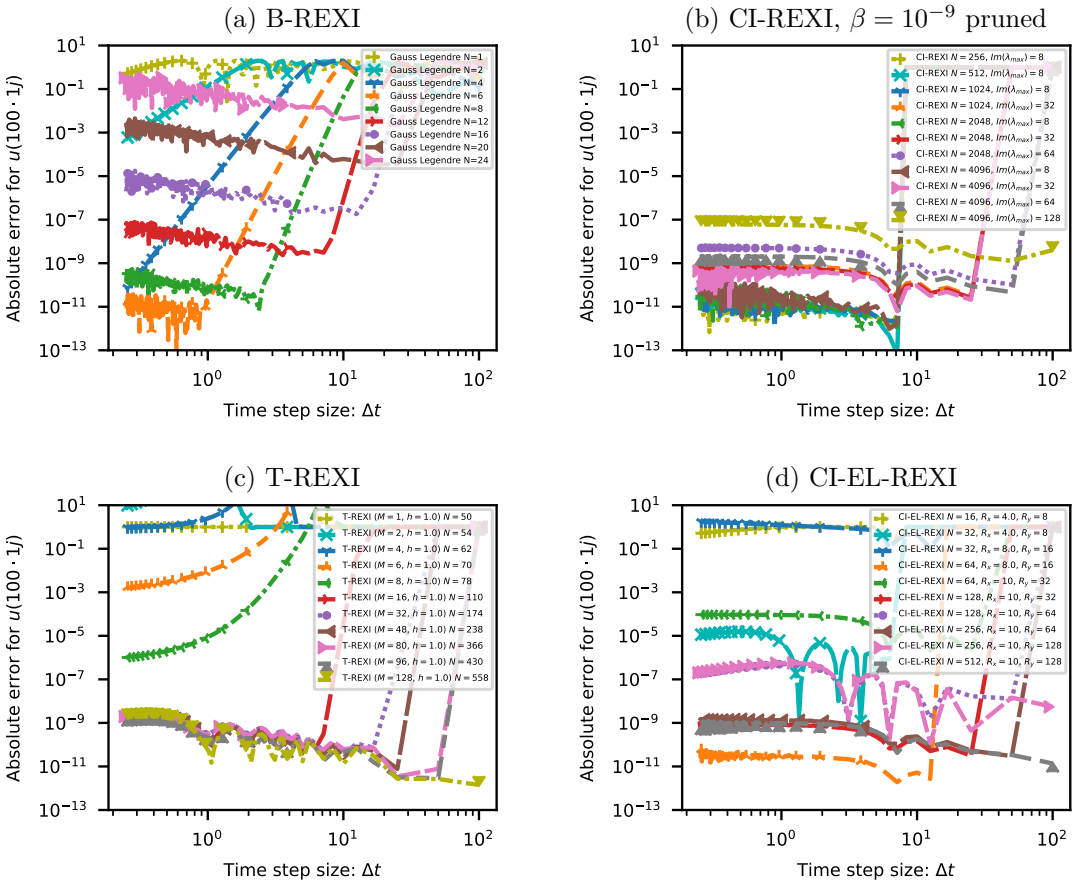

Figure 11: Error studies with oscillatory Dahlquist equation for different REXI methods with time step size Δt vs. absolute error at $u(t = 100)$. B-REXI is suitable only for smaller timestep sizes. CI-REXI can be tuned to allow also very large time step sizes. T-REXI requires many REXI terms for small 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 REXI terms for similar accuracy.

with the horizontal velocity $\vec{V}$ on the longitude/latitude field, the 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 $\overline{h} = 10^5 m$, the perturbed geopotential Φ' given by $\Phi' = \Phi - \overline{\Phi}$, and the Coriolis effect $f = 2\Omega \sin(\phi)$ with latitude ϕ and angular rate of rotation Ω.

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 [35, 17] for a detailed description of the spherical harmonics. In particular, we work with the vorticity-divergence formulation in spectral space to avoid spurious modes if one would convert the velocity to spectral space. The standard $\frac{2}{3}$ rule [30] is used for anti-aliasing to evaluate bi-non-linearities.

6.2.2. REXI solvers

Finding solvers for the systems of linear equations for arbitrary grids still remains a particular challenge for the application of REXI, see discussion in Sec. 3.2. Using spherical harmonics, we can find highly efficient

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

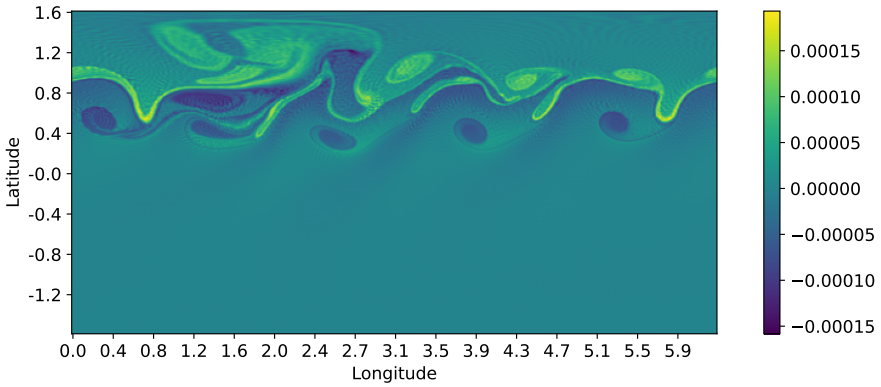

Figure 12: Vorticity field of barotropic instability benchmark after 8 days of inviscid shallow-water equations. We see the development of various large and small-scale vortices.

solvers in this space for integrating the term L_g with direct exponential integration (due to a diagonal linear system in spectral space) and also of $L_g + L_f$ for REXI and implicit Euler time integrators (with a penta-diagonal linear system in spectral space).

Regarding the direct exponential integration, an analytical solution can be found using the vorticity-divergence form, see also [41]. A diagonalization of the linear operator leads to the matrix of eigenvectors Q and eigenvalues $diag(\Lambda)$ for each individual spherical harmonics mode as

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

with $D = -\nabla^2$ and $G = -\overline{\Phi}$. We can then use $U(t + \Delta t) = Q \exp(\Delta t \Lambda) Q^{-1} U(t)$ for exponential time integration of the L_g term. From an algebraic perspective, this method matches the method in [20], where we have developed a more elegant and short derivation that does not require Laplace transforms.

We emphasize here that an exponential integration of the full linear terms $L = L_g + L_c$ (related to the Hough modes [47]) is not possible in this way with spherical harmonics. 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 [35]. The present work is based on the numerical reformulation of an implicit time stepper [40] of the form $(I - \Delta t L)^{-1}$ which has been transformed to solve a REXI term by using a complex-valued time step size.

6.2.3. Benchmark

Our benchmark is based on the barotropic instability test case (see [15]). This benchmark is initialized with a geostropically balanced initial condition, which is perturbed by a small Gaussian-shaped bell (see reference for detailed initial conditions). We time integrate this system for 8 days with results in Figure 12.

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 [39]. 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 1 where we use X and Y as representatives for either term in the PDE

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 1: Overview of time integration methods. Note that they can be composed together.

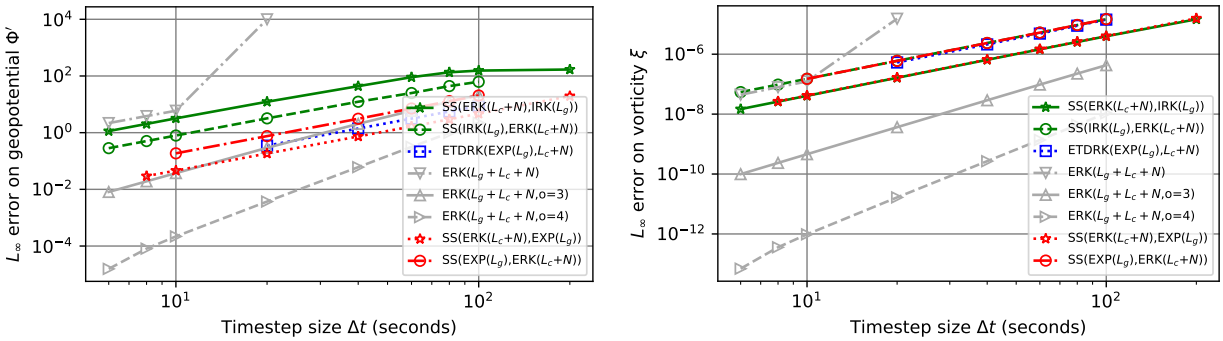

Figure 13: Studies **without REXI methods** (but using direct exponentiation) on all prognostic variables with error vs. time step size for the barotropic instability benchmark. We also include 2nd, 3rd, and 4th order Runge-Kutta based methods with gray lines.

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 Skylake Xeon 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 made use of the SHTNS library [34] which is based on FFTW [14]. We precomputed transformation plans and reused them for all studies to ensure the utilization of the same plans over all studies. Each batch job is set to timeout after 1 hour.

6.2.6. Performance comparison for splitting L_g and $L_c + N$

We start with an **error comparison of non-REXI methods** in Figure 13 which we will use as a baseline for further comparisons with REXI-based methods. Since conclusions to be drawn for the divergence field plots are similar to the conclusions of the geopotential, we only provide plots for the geopotential and vorticity. 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 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* variable is the Strang-split $SS(ERK(L_c + N), EXP(L_g))$, which we account for by the accurate treatment of the geopotential variable with the exponential. Since the *vorticity* field is not treated exponentially (time tendency for this in L_g is null), there is also no benefit visible in

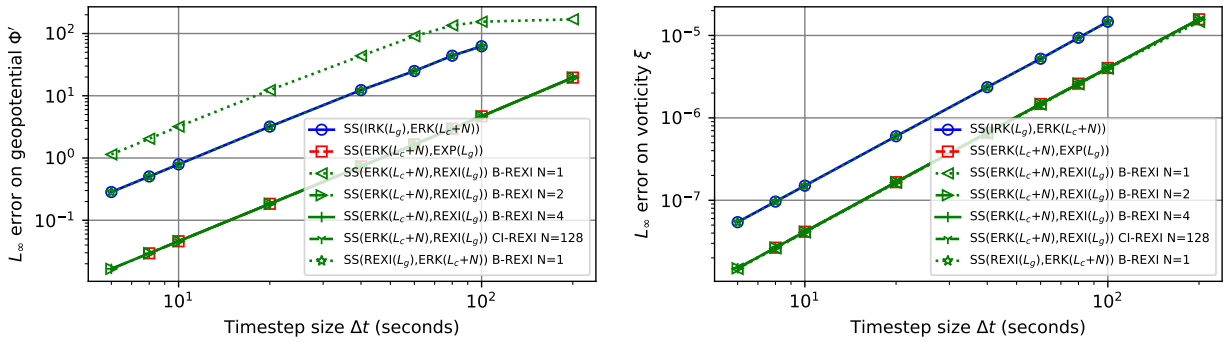

Figure 14: Studies **with REXI methods** on all prognostic variables with error vs. time step size for the barotropic instability benchmark.

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. However, it is still ranked 2nd for the geopotential and ranked lower for the other variables. We account for that by the way a 2nd order accurate Strang-splitting is performed: It 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 error studies** by comparing them with the best Strang-split exponential and implicit methods from the previous results in Figure 14. 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. 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. We attribute that to the Strang-splitting errors being larger than the errors from the approximation of the linear waves when using $N = 2$ poles, thus further resolution of the linear waves by using $N = 4$ poles does not lead to further reduction of this error norm. 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 [37].

We close this section by **HPC REXI studies** in Figure 15. For sake of better overview, we only plotted the most interesting candidates, skipping ETDRK which is worse than B-REXI methods, the explicit RK order 3 and 4 methods which 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. The CI-REXI method requires $N = 128$ terms and is consequently impacted by 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 $L = L_g + L_c$ and N

Next, we investigate the performance of REXI methods using a splitting into the linear term $L = L_g + L_c$ and the nonlinear term N where no direct computation of $\exp(\Delta t L)$ is possible.

Error plots are given in Figure 16 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

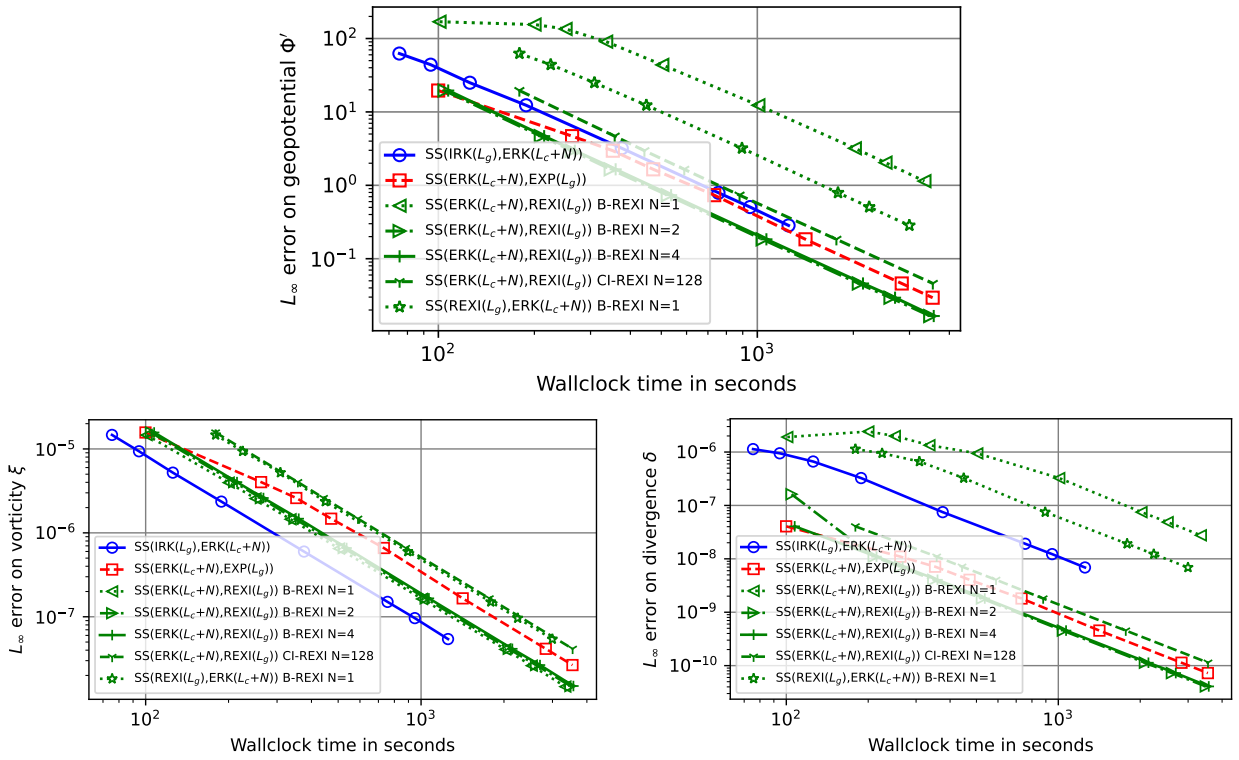

Figure 15: Studies **including REXI methods** with wallclock time vs. time step size for the barotropic instability benchmark.

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* ξ 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, we account for this by the errors in the nonlinear parts dominating the overall errors. Again, since the divergence δ shows results very similar to the geopotential, we skip them here for sake of brevity. Overall, the ETDRK methods show a poorer performance than the more straightforward SS approach.

Finally, we investigate the wallclock time vs. errors with **HPC studies** given in Figure 17. We can observe that fully explicit ERK methods provide excellent results due to their computational efficiency. 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 REXI methods. Again, the ETDRK method shows no real benefits. Investigating the *vorticity* ξ 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. Again, ETDRK are the worst performing methods requiring the most computational time. In investigation of the divergence δ shows results similar to the geopotential, hence, we skip it for sake of brevity.

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

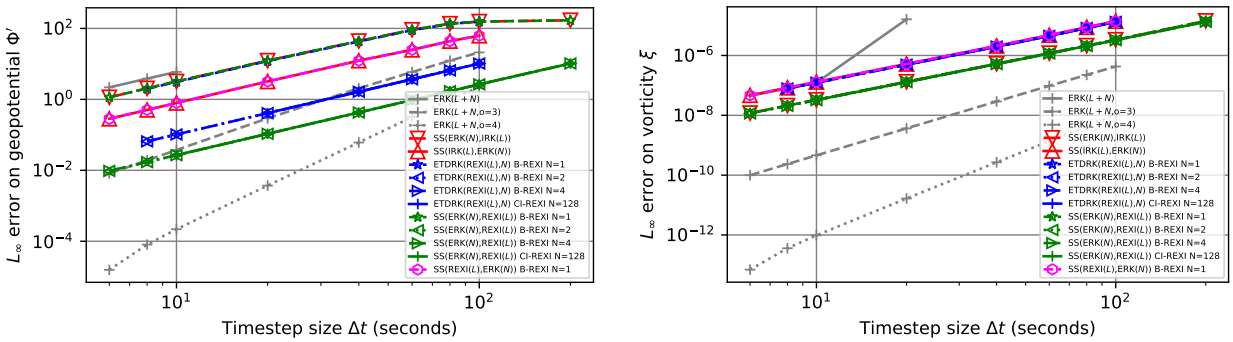

Figure 16: Error studies **using non-REXI methods** (using direct exponentiation) on all prognostic variables with error vs. time step size for the barotropic instability benchmark. (ETDRK data points are missing due to 1h timeouts of the job.)

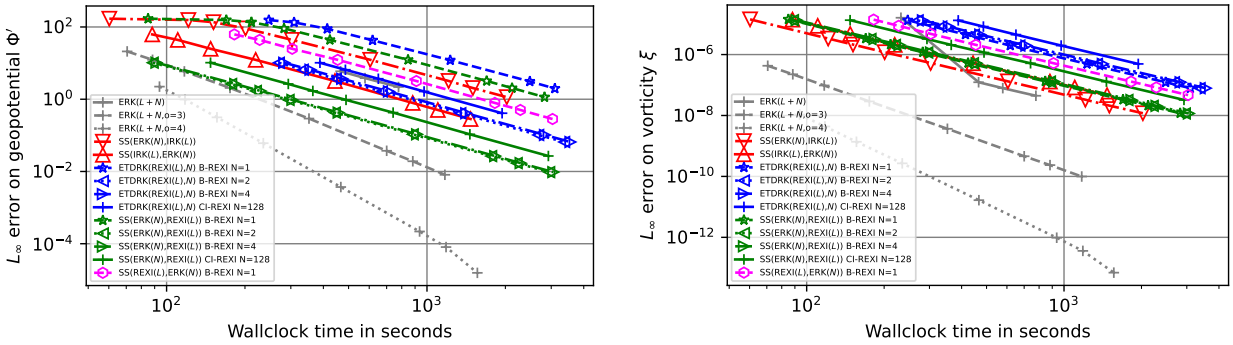

Figure 17: Studies **using non-REXI methods** (using direct exponentiation) on all prognostic variables with error vs. time step size for the barotropic instability benchmark. (ETDRK data points are missing due to 1h timeouts of the job.)

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 nonlinear parts dominate the overall errors.

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 applied on linear operators turning out to be computationally rather challenging.

This paper investigated different ways to approximate φ functions with Rational Approximations of Exponential Integration (REXI) with a unified REXI formulation. This formulation enabled us to express Butcher/Bickard-based B-REXI, Cauchy Contour integration-based CI-REXI, and Terry's T-REXI method within the same mathematical framework. For the application to higher-order exponential integration methods, we also derived an elegant way to compute higher-order φ functions based on REXI coefficients for lower-order φ opening the path to also use B-REXI methods for higher-order exponential integration formulations.

For the first time, an in-depth comparison of the approximation quality of each REXI method has been conducted including an explanation of numerical issues. We used linear ODEs where we studied and discussed properties of stability, convergence, and also the filtering capabilities. In addition, 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 $N = 2$ terms, leading to a significant reduction of computational effort compared to former REXI methods in this context using $N = 128$ terms. Consequently, regarding demands on computational resources, B-REXI showed a reduction of a factor of 64 compared to previous work leading to the same factor of significant savings of computational resources.

This work raises a particular future research question, motivated by the excellent results with the B-REXI method relating higher-order Runge-Kutta integration methods as approximations of exponential integration: When are Runge-Kutta-based methods (higher-order implicit Runge-Kutta methods, a subcycling with explicit Runge-Kutta methods, spectral deferred correction methods, etc.) sufficiently good approximations of φ_i terms and how do they compare from an HPC perspective to alternative methods?

Acknowledgements

Both authors like to thank Pedro S. Peixoto for pointing out the potential relation of exponential integration methods to Laplace transforms and Peter Lynch's work in this context. Martin Schreiber is grateful to NCAR for providing financial support and a very inspiring office space with a splendid view to the flatirons, which strongly supported this work. Both authors thank Matthew Normile for preliminary work as well as Finn Capelle and Raphael Schilling who indirectly contributed to this work with the REXInsight software. Both authors thank Ernst Hairer for valuable feedback on a preprint version. We also like to thank the two anonymous reviewers for their valuable feedback and suggestions that helped to improve the readability and understandability of the manuscript.

The authors gratefully acknowledge the Gauss Centre for SC e.V. (www.gauss-centre.eu) for funding this project by providing computing time on the GCS Supercomputer SUPERMUC-NG at Leibniz Supercomputing Centre (www.lrz.de).

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.

References

[1] Theodore A. Bickart. An Efficient Solution Process for Implicit Runge-Kutta Methods. *SIAM Journal on Numerical Analysis*, 14(6):1022–1027, 1977.

[2] John C. Butcher. Implicit Runge-Kutta Processes. *AMS*, 18(85):50–64, 1964.

[3] John C. Butcher. On the implementation of implicit Runge-Kutta methods. *BIT*, 16(3):237–240, 1976.

[4] Tommaso Buvoli. A Class of Exponential Integrators Based on Spectral Deferred Correction. pages 1–22, 2015. _eprint: 1504.05543.

[5] Katherine Calvin, Dipak Dasgupta, Gerhard Krinner, Aditi Mukherji, Peter W. Thorne, Christopher Trisos, José Romero, Paulina Aldunce, Ko Barrett, Gabriel Blanco, William W.L. Cheung, Sarah Connors, Fatima Denton, Aïda Diongue-Niang, David Dodman, Matthias Garschagen, Oliver Geden, Bronwyn Hayward, Christopher Jones, Frank Jotzo, Thelma Krug, Rodel Lasco, Yune-Yi Lee, Valérie Masson-Delmotte, Malte Meinshausen, Katja Mintenbeck, Abdalah Mokssit, Friederike E.L. Otto, Minal Pathak, Anna Pirani, Elvira Poloczanska, Hans-Otto Pörtner, Aromar Revi, Debra C. Roberts, Joyashree Roy, Alex C. Ruane, Jim Skea, Priyadarshi R. Shukla, Raphael Slade, Aimée Slangen, Youba Sokona, Anna A. Sörensson, Melinda Tignor, Detlef Van Vuuren, Yi-Ming Wei, Harald Winkler, Panmao Zhai, Zinta Zommers, Jean-Charles Hourcade, Francis X. Johnson, Shonali Pachauri, Nicholas P. Simpson, Chandni Singh, Adelle Thomas, Edmond Totin, Paola Arias, Mercedes Bustamante, Ismail Elgizouli, Gregory Flato, Mark Howden, Carlos Méndez-Vallejo, Joy Jacqueline Pereira, Ramón Pichs-Madruga, Steven K. Rose, Yamina Saheb, Roberto Sánchez Rodríguez, Diana Ürge Vorsatz, Cunde Xiao, Noureddine Yassaa, Andrés Alegría, Kyle Armour, Birgit Bednar-Friedl, Kornelis Blok, Guéladio Cissé, Frank Dentener, Siri Eriksen, Erich Fischer, Gregory Garner, Céline Guivarch, Marjolijn Haasnoot, Gerrit Hansen, Mathias Hauser, Ed Hawkins, Tim Hermans, Robert Kopp, Noëmie Leprince-Ringuet, Jared Lewis, Debora Ley, Chloé Ludden, Leila Niamir, Zebedee Nicholls, Shreya Some, Sophie Szopa, Blair Trewin, Kaj-Ivar Van Der Wijst, Gundula Winter, Maximilian Witting, Arlene Birt, Meeyoung Ha, José Romero, Jinmi Kim, Erik F. Haites, Yonghun Jung, Robert Stavins, Arlene Birt, Meeyoung Ha, Dan Jezreel A. Orendain, Lance Ignon, Semin Park, Youngin Park, Andy Reisinger, Diego Cammaramo, Andreas Fischlin, Jan S. Fuglestvedt, Gerrit Hansen, Chloé Ludden, Valérie Masson-Delmotte, J.B. Robin Matthews, Katja Mintenbeck, Anna Pirani, Elvira Poloczanska, Noëmie Leprince-Ringuet, and Clotilde Péan. IPCC, 2023: Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. IPCC, Geneva, Switzerland. Technical report, Intergovernmental Panel on Climate Change (IPCC), July 2023. Edition: First.

[6] Colm Clancy and Peter Lynch. Laplace transform integration of the shallow-water equations. Part I: Eulerian formulation and Kelvin waves. *Quarterly Journal of the Royal Meteorological Society*, 137(656):792–799, 2011.
[7] Colm Clancy and Janusz A. Pudykiewicz. On the use of exponential time integration methods in atmospheric models. *Tellus, Series A: Dynamic Meteorology and Oceanography*, 65, 2013.
[8] R. Courant, H. Lewy, and K. Friedrichs. Über die partiellen Differenzengleichungen der mathematischen Physik. *Mathematische Annalen*, 1932.
[9] S. M. Cox and P. C. Matthews. Exponential time differencing for stiff systems. *Journal of Computational Physics*, 176(2):430–455, 2002. ISBN: 0021-9991.
[10] Nicolas Crouseilles, Lukas Einkemmer, and Josselin Massot. Exponential methods for solving hyperbolic problems with application to collisionless kinetic equations. *Journal of Computational Physics*, 420:109688, November 2020.
[11] ECMWF. The Strength of a Common Goal: A Roadmap To 2025. 2016.
[12] Lukas Einkemmer, Mayya Tokman, and John Loffeld. On the performance of exponential integrators for problems in magnetohydrodynamics. *Journal of Computational Physics*, 330:550–565, February 2017.
[13] O G Ernst and M J Gander. *Why it is Difficult to Solve Helmholtz Problems with Classical Iterative Methods*, volume 83. 2012. ISSN: 1098-6596 _eprint: arXiv:1011.1669v3.
[14] Matteo Frigo. A Fast Fourier Transform Compiler. 1999.
[15] Joseph Galewsky, Richard K. Scott, and Lorenzo M. Polvani. An initial-value problem for testing numerical models of the global shallow-water equations. *Tellus, Series A: Dynamic Meteorology and Oceanography*, 56(5):429–440, 2004. ISBN: 9781450324748.
[16] Martin J. Gander. 50 Years of Time Parallel Time Integration. pages 69–113, 2015. ISBN: 978-3-319-23321-5.
[17] JJ Hack and R Jakob. *Description of a global shallow water model based on the spectral transform method*, volume NCAR/TN-34. 1992. Publication Title: NCAR Technical Note.
[18] Ernst Hairer, S.P. Norsett, and Gerhard Wanner. *Solving ordinary differential equations I: Nonstiff problems*. 1987.
[19] Ernst Hairer and Gerhard Wanner. *Solving ordinary differential equations II: Stiff and differential-algebraic problems*, volume 14. 1991. ISSN: isbn:3540604529 Publication Title: SpringerVerlag.
[20] Eoghan Harney and Peter Lynch. Laplace transform integration of a baroclinic model. *Quarterly Journal of the Royal Meteorological Society*, (January):347–355, 2019.
[21] 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 Journal of Numerical Analysis*, 36(2):688–716, 2016. _eprint: 1402.5168.
[22] Marlis Hochbruck and Alexander Ostermann. *Exponential integrators*, volume 19. 2010. ISSN: 09624929 Publication Title: Acta Numerica.
[23] K. R. Jackson and S. P. Nørsett. The Potential for Parallelism in Runge-Kutta Methods. Part 1: RK Formulas in Standard Form. *SIAM Journal on Numerical Analysis*, 32(1):49–82, 1995.
[24] Wilhelm Kutta. Beitrag zur näherungsweisen integration totaler Differentialgleichungen. *Z. Math. Phys.*, (46):435–453, 1901.
[25] Jacques Louis Lions, Yvon Maday, and Gabriel Turinici. Résolution d'EDP par un schéma en temps "pararéel". *Comptes Rendus de l'Academie des Sciences - Series I: Mathematics*, 332(7):661–668, 2001. ISBN: 0764-4442.
[26] 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.
[27] Michael Minion, Robert Speck, Matthias Bolten, Matthew Emmett, and Daniel Ruprecht. Interweaving PFASST and Parallel Multigrid. pages 1–20, 2014. ISBN: 0001410105 _eprint: 1407.6486.
[28] Cleve Moler and Charles Van Loan. Nineteen Dubious Ways to Compute the Exponential of a Matrix, Twenty-Five Years Later. *SIAM Review*, 45(1):3–49, 2003. ISBN: 00361445 _eprint: arXiv:1011.1669v3.
[29] Jitse Niesen and Will M Wright. A Krylov subspace algorithm for evaluating the phi-functions appearing in exponential integrators. *ACM Transactions on Mathematical Software*, 38(November):20, 2012. ISBN: 0003-021X _eprint: 0907.4631.
[30] Steven A. Orszag. On the Elimination of Aliasing in Finite-Difference Schemes by Filtering High-Wavenumber Components. page 1, 1971. 2/3 rule.
[31] Fernando V. Ravelo, Martin Schreiber, and Pedro S. Peixoto. High-order exponential integration for seismic wave modeling. *Computational Geosciences*, 28(6):1349–1369, December 2024.
[32] Ana Cecilia Rojas Mendoza and Pedro da Silva Peixoto. Numerical solution of ordinary diferential equations using Laplace transform integration, 2020.
[33] Carl Runge. Über die numerische Auflösung von Differentialgleichungen. *Mathematische Annalen*, 46, 1895.
[34] Nathanaël Schaeffer. Efficient spherical harmonic transforms aimed at pseudospectral numerical simulations. *Geochemistry, Geophysics, Geosystems*, 14(3):751–758, 2013. _eprint: 1202.6522.
[35] Martin Schreiber and Richard Loft. A parallel time integrator for solving the linearized shallow water equations on the rotating sphere. *Numerical Linear Algebra with Applications*, 26(2), 2018.
[36] Martin Schreiber, Pedro S. Peixoto, Terry Haut, and Beth Wingate. Beyond spatial scalability limitations with a massively parallel method for linear oscillatory problems. *International Journal of High Performance Computing Applications*, 32(6):913–933, 2017.
[37] Martin Schreiber, Nathanaël Schaeffer, and Richard Loft. Exponential integrators with parallel-in-time rational approximations for the shallow-water equations on the rotating sphere. *Parallel Computing*, 85:56–65, 2019. Publisher: Elsevier B.V.
[38] Alexander F. Shchepetkin and James C. McWilliams. The regional oceanic modeling system (ROMS): a split-explicit, free-surface, topography-following-coordinate oceanic model. *Ocean Modelling*, 9(4):347–404, January 2005.

[39] Strang, Gilbert. On the Construction and Comparison of Difference Schemes. *SIAM Journal on Numerical Analysis*, 5(3):506–517, 1968.
[40] Clive Temperton. Treatment of the Coriolis Terms in Semi-Lagrangian Spectral Models. (March 1994), 1995.
[41] J. Thuburn, T.D. Ringler, W.C. Skamarock, and J.B. Klemp. Numerical representation of geostrophic modes on arbitrarily structured C-grids. *Journal of Computational Physics*, 228(22):8321–8335, December 2009.
[42] M. Tokman. Efficient integration of large stiff systems of ODEs with exponential propagation iterative (EPI) methods. *Journal of Computational Physics*, 213(2):748–776, April 2006.
[43] M. Tokman. A new class of exponential propagation iterative methods of Runge-Kutta type (EPIRK). *Journal of Computational Physics*, 230(24):8762–8778, 2011. Publisher: Elsevier Inc.
[44] L. N. Trefethen, J. A. C. Weideman, and T. Schmelzer. Talbot quadratures and rational approximations. *BIT Numerical Mathematics*, 46(3):653–670, September 2006.
[45] Lloyd N Trefethen and J A C Weideman. The Exponentially Convergent Trapezoidal Rule. 56(3):385–458, 2014.
[46] J. Virieux, A. Asnaashari, R. Brossier, L. Métivier, A. Ribodetti, and W. Zhou. *An introduction to full waveform inversion*. 2014. Publication Title: Encyclopedia of Exploration Geophysics.
[47] Houjun Wang, John P. Boyd, and Rashid A. Akmaev. On computation of Hough functions. *Geoscientific Model Development*, 9(4):1477–1488, 2016. ISBN: 1471239314.