Understanding the rational approximation of the exponential integrator (REXI)

Martin Schreiber (M.Schreiber@exeter.ac.uk)
Pedro S. Peixoto (pedrosp@ime.usp.br)
et. al.

August 9, 2015

!!!THIS IS A PRELIMINARY, NON PROOF-READ DOCUMENT!!!

This document serves as the basis for implementing the Rational approximation of the EXponential Integrator (REXI). Here, we purely focus on the linear part of the shallow-water equations (SWE) and show the different steps to approximate solving this linear part with an exponential integrator. This paper mainly summarises previous work on REXI.

1 Problem formulation

We use linearised shallow water equations (SWE) with respect to a rest state with mean water depth of H and defined for perturbations of height h (see [1]). The linear operator (L) may be written as

$$L(U) := \begin{pmatrix} 0 & H\delta_x & H\delta_y \\ g\delta_x & 0 & -f \\ g\delta_y & f & 0 \end{pmatrix} U$$

where $U := (h, u, v)^T$. Here, we neglect all non-linear terms and consider f constant (f-plane approximation).

The time evolution of the PDE, with the subscript t denoting the derivative in time, is given by

$$U_t = L(U).$$

It is further worth noting, that this system describes an oscillatory system (2D wave equation), hence the operator L is hyperbolic and has imaginary eigenvalues.

2 Exponential integrator

Linear initial value differential problems are well known to be solvable with exponential integrators for arbitrary time step sizes via

$$U(t) = e^{Lt}U(0).$$

see e.g. [5]. However, this is typically quite expensive to compute and analytic solutions only exist for some simplified system of equations, see e.g. [1] for f-plane shallow-water equations. These exponential integrators can be approximated with rational functions and this paper is on giving insight into this approximation.

3 Underlying idea of rational approximation

Terry et. al. [2] developed a rational approximation of the exponential integrator. First, we like to get more insight into it with a one-dimensional formulation before applying REXI to a rational approximation of a linear operator. Our main target is to find an approximation of an operator with a *complex exponential shape*, in our case e^{ix}, which (in one-dimension) is given as a function $f(x)$. We will end up in an approximation given by the following rational approximation:

$$e^{ix} \approx \sum_{n=-N}^{N} \frac{\beta_n}{ix - \alpha_n}$$

with complex coefficients α_n and β_n. We point out that the coefficients α_n will always have non zero real part, so no singularity occurs with the rational function.

3.1 Step A) Approximation of solution space

First, we assume that we can use Gaussian curves as basis functions for our approximation So first we find an approximation of one of our underlying Gaussian basis function

$$\psi_h(x) := (4\pi)^{-\frac{1}{2}} e^{-x^2/(4h^2)}$$

In this formulation, h can be interpreted as the horizontal "stretching" of the basis function. Note the similarities to the Gaussian distribution, but by dropping certain parts of the vertical scaling as it is required for probability distributions. We can now approximate our function $f(x)$ with a superposition of basis functions $\psi_h(x)$ by

$$f(x) \approx \sum_{m=-M}^{M} b_m \psi_h(x + mh)$$

with M controlling the interval of approximation (~size of "domain of interest") and h will be related to the accuracy of integration (~resolution in "domain of interest").

We choose h small enough so that the support of the Fourier transform of f is mainly localised within $[-1/(2h), 1/(2h)]$ (i.e. almost zero outside this interval). M is chosen such that the approximation will be adequate in the interval $|x| < Mh$.

To compute the coefficients b_m, we rewrite the previous equation in Fourier space with

$$\frac{\hat{f}(\xi)}{\hat{\psi}_h(\xi)} = \sum_{m=\infty}^{\infty} b_m e^{2\pi imh\xi},$$

where the $\hat{\cdot}$ symbols indicate the Fourier transforms of the respective functions. The b_m are now the Fourier coefficients of the series for the function $\frac{\hat{f}(\xi)}{\hat{\psi}_h(\xi)}$ and be calculated as [1],

$$b_m = h \int_{-\frac{1}{2h}}^{\frac{1}{2h}} e^{-2\pi imh\xi} \frac{\hat{f}(\xi)}{\hat{\psi}_h(\xi)} d\xi,$$

for $m \in \mathbb{Z}$.

Since we are interested in approximating $f(x) = e^{ix}$, we can simplify the equation by using the response in frequency space $\hat{f}(\xi) = \delta(\xi - \frac{1}{2\pi})$, where here δ is the Dirac distribution, and

$$b_m = h\, e^{-imh} \hat{\psi}_h(\frac{1}{2\pi})^{-1}.$$

The Fourier transform of the Gaussian function is well known and given by

$$\hat{\psi}_h(\xi) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{4\pi}} e^{-\left(\frac{x}{2h}\right)^2} e^{-2\pi ix\xi} dx = he^{-(2h\pi\xi)^2}$$

where we used that $\int_{-\infty}^{\infty} e^{-\left(\frac{x}{2h}\right)^2} dx = h\sqrt{4\pi}$. For the case $\xi = \frac{1}{2\pi}$, we get

$$\hat{\psi}_h\left(\frac{1}{2\pi}\right) = h\, e^{-h^2}.$$

Finally, one can obtain the equation

$$b_m = h\, e^{-imh} \frac{1}{h\, e^{-h^2}} = e^{-imh} e^{h^2}$$

to compute the coefficients b_m for $f(x) = e^{ix}$.

[1] see [2], page 11

3.2 Step B) Approximation of basis function

The second step is the approximation of the basis function $\psi_h(x)$ itself with a rational approximation, see [6]. Our basis function is given by

$$\psi_h(x) := (4\pi)^{-\frac{1}{2}} e^{-x^2/(4h^2)}$$

and a close-to-optimal approximation of $\psi_1(x)$ with a sum of rational functions is given by

$$\psi_1(x) \approx Re\left(\sum_{l=-L}^{L} \frac{a_l}{ix + (\mu + i\,l)}\right)$$

with the μ and a_l given in [6], Table 1. We can generalise this approximation to arbitrary chosen h via

$$\psi_h(x) \approx Re\left(\sum_{l=-L}^{L} \frac{a_l}{i\frac{x}{h} + (\mu + i\,l)}\right)$$

3.3 Step C) Approximation of the approximation

We then combine the approximation (B) of the approximation (A), yielding the approximation for f given by

$$\tilde{f}(x) = \sum_{m=-M}^{M} b_m \psi_h(x + mh) = \sum_{m=-M}^{M} b_m Re\left(\sum_{l=-L}^{L} \frac{a_l}{i\frac{x+mh}{h} + (\mu + i\,l)}\right)$$

$$= \sum_{m=-M}^{M} b_m \sum_{l=-L}^{L} Re\left(\frac{ha_l}{ix + h(\mu + i(m+l))}\right).$$

We can simplify the summations assuming $n = m + l$ and inverting their order,

$$\begin{aligned}
\tilde{f}(x) &= \sum_{l=-L}^{L} \sum_{m=-M}^{M} b_m Re\left(\frac{ha_l}{ix + h(\mu + i(m+l))}\right) \\
&= \sum_{l=-L}^{L} \sum_{n=-M+l}^{M+l} b_{n-l} Re\left(\frac{ha_l}{ix + h(\mu + in)}\right) \\
&= \sum_{n=-N}^{N} \sum_{k=L_1}^{L_2} b_{n-k} Re\left(\frac{ha_k}{ix + h(\mu + in)}\right),
\end{aligned}$$

where $N = L + M$, $L_1 = max(-L, n - M)$ and $L_2 = min(L, n + M)$.

We will define the poles as

$$\alpha_n := h(\mu + in). \tag{1}$$

For the real part of $f(x)$ we have that

$$\begin{aligned}
Re(\tilde{f}(x)) &= \sum_{n=-N}^{N} \sum_{k=L_1}^{L_2} Re(b_{n-k}) Re\left(\frac{ha_k}{ix+\alpha_n}\right) \\
&= Re\left(\sum_{n=-N}^{N} \sum_{k=L_1}^{L_2} Re(b_{n-k}) \frac{ha_k}{ix+\alpha_n}\right) \\
&= Re\left(\sum_{n=-N}^{N} \frac{1}{ix+\alpha_n} h \sum_{k=L_1}^{L_2} Re(b_{n-k}) a_k\right).
\end{aligned}$$

By setting the real residue as

$$\beta_n^{Re} := h \sum_{k=L_1}^{L_2} Re(b_{n-k}) a_k, \tag{2}$$

we have that

$$Re(\tilde{f}(x)) = Re\left(\sum_{n=-N}^{N} \frac{\beta_n^{Re}}{ix+\alpha_n}\right).$$

For the imaginary part of $f(x)$ we analogously have that

$$Im(\tilde{f}(x)) = Re\left(\sum_{n=-N}^{N} \frac{\beta_n^{Im}}{ix+\alpha_n}\right),$$

where

$$\beta_n^{Im} := h \sum_{k=L_1}^{L_2} Im(b_{n-k}) a_k.$$

This finally leads us to the REXI approximation

$$e^{ix} \approx \sum_{n=-N}^{N} Re\left(\frac{\beta_n^{Re}}{ix+\alpha_n}\right) + i\, Re\left(\frac{\beta_n^{Im}}{ix+\alpha_n}\right)$$

for the complex-valued function e^{ix}.

4 Matrix exponential

Finally, we like to apply REXI to a formulation such as

$$U(t) := e^{tL} U(0).$$

To see the relationship between the approximation of e^{ix} with e^{tL} we assume that L is skew hermitian and therefore has only purely imaginary eigenvalues, and maybe decomposed as $\Sigma\Lambda\Sigma^H$, yielding

$$e^{tL} = \sum_{k=0}^{\infty} \frac{tL}{k!} = \Sigma \left(\sum_{k=0}^{\infty} \frac{t\Lambda}{k!} \right) \Sigma^H = \Sigma e^{t\Lambda} \Sigma^H,$$

where we used the orthonormality of Σ to remove it from the summation, and

$$e^{t\Lambda} = \begin{pmatrix} \cdots & & \\ & e^{i\lambda_n t} & \\ & & \cdots \end{pmatrix}$$

where we hare explicitly detached the imaginary unit from the eigenvalues, therefore λ_n are assumed real. Since $e^{t\Lambda}$ is diagonal, it can be eigenvalue-wise approximated in the same way as in e^{ix}.

Some important points about the choice of M and h have to be made at this point. We know that e^{ix} is accurately approximated with REXI for the interval $|x| < hM$, where h is chosen small enough to obtain a good approximation in step (A), and M will define the interval size and number of approximation points. In the matrix case, M has to be chosen so that $hM > t\bar{\lambda}$, where $\bar{\lambda} = \max_n |\lambda_n|$, in order to capture all wavelengths of L. In other other words, hM need to be set to capture the fastest wave. Note that if this is used as a time stepping method, with time step $t = \tau$, then, the larger the timestep, the larger M will be. Exact evaluations of the choices for h and M may be done based on equation (3.6) of [2].

5 Filtering

The method described in the previous section is well defined for skew hermitian L. If L is not skew hermitian, the real eigenvalues might cause the REXI to have absolute values larger than 1, which can lead to instabilities if used as time stepping method.

To ensure that the REXI is bounded by unit, a filtering process is proposed in [2]. REXI is prone to exceed unit in the neighbourhood of $|t\lambda| \approx hM$, therefore in the highest frequencies. The idea is to construct a rational function $S(ix)$ that is approximately 1 in a smaller interval $|t\lambda| < hM_0$, with $M_0 < M$, and decays very fast to zero outside this interval. Then we multiply this filters function to the original REXI, which will lead to a unit bounded REXI.

Further details of how $S(ix)$ is computed will be added later.

6 REXI on linear operators, "our little wild dog"

We want to evaluate $e^{\tau L}U(0)$ with REXI, where τ will be a time step size and $U(0)$ the initial condition for this time step. We will assume τ a-priori fixed, which implies that the coefficients in REXI will not change and may be pre-computed.

Although L has imaginary eigenvalues, we wish to evaluate the $e^{\tau L}U(0)$, which is real valued, therefore, we will use the real approximation of e^{ix}

$$exp(\tau L) \approx Re\left(\sum_{n=-N}^{N} \beta_n (\tau L - \alpha_n)^{-1}\right), \tag{3}$$

where β_n is given by equation (2) and α_n by equation (1). These coefficient may be pre-computed if L and τ are fixed.

Note, that for debugging purpose, their *imaginary values have to cancel out.* (PP: are you sure?)

Note an important property (see Sec. 3.3 in [2]). There's an anti-symmetry in the α_i coefficients, which avoids computing half of the inverses,

$$\overline{(L-\alpha)^{-1}U(0)} = (L-\overline{\alpha})^{-1}U(0).$$

6.1 Computing inverse of $(L-\alpha)^{-1}$

For computing the inverse, arbitrary solvers can be used. However we like to note, that α is a complex number. Hence, requiring solvers with support for solving in complex space. As an example, we consider a specialization on the shallow-water equations given above with

$$L(U(t)) := \begin{pmatrix} & H\delta_x & H\delta_y \\ g\delta_x & & -f \\ g\delta_y & f & \end{pmatrix} U(t)$$

$$U_t(t) := L(U(t)).$$

6.2 Handling τ in REXI

We recall the formulation of the solution as an exponential integrator

$$U(t) = e^{tL}U(0)$$

which formally allows us to join the integration in time given by t with the L operator in case of such a formulation.

We reformulate the REXI approximation scheme given by

$$(\tau L - \alpha)^{-1}U(\tau) = U(0)$$

and by factoring τ out, yielding

$$(L - \frac{\alpha}{\tau})^{-1}U(\tau)\tau^{-1} = U(0)$$

So instead of solving for $U(\tau)$, we are solving for $U^\tau(\tau) := U(\tau)\tau^{-1}$ as well as $\alpha^\tau := \frac{\alpha}{\tau}$.

To summarize, we have to solve the system of equations given by

$$(L-\alpha^\tau)^{-1}U^\tau(\tau) = U(0) \tag{4}$$

with $U(0)$ the initial conditions. For sake of simplicity, we stick to the formulation without the τ notation.

6.3 Solving as an elliptic problem

We wish to solve the differential problem for each time step

$$(L-\alpha)U = U_0$$

so that $U = (L-\alpha)^{-1}U_0$. Expanding the equations with the definition of L brings us to

$$\begin{aligned} -fv - g\eta_x - \alpha u &= u_0 & (5)\\ fu - g\eta_y - \alpha v &= v_0 & (6)\\ H(u_x+v_y) - \alpha\eta &= \eta_0 & (7) \end{aligned}$$

Let f be constant (f-plane approximation), $\delta := u_x + v_y$ be the wind divergence, $\zeta := v_x - u_y$ be the wind (relative) vorticity and $\Delta\eta := \eta_{xx} + \eta_{yy}$ the Laplacian of the fluid depth. We will re-write the problem in a divergence-vorticity formulation by taking 2 steps. First, sum the ∂_x of equation (5) and the ∂_y of equation (6), yielding

$$-f\zeta - g\Delta\eta - \alpha\delta = \delta_0, \tag{8}$$

then subtract the ∂_y of equation (5) from the ∂_x of equation (6), yielding

$$f\delta - \alpha\zeta = -\zeta_0. \tag{9}$$

Using equation (8) in equation (9) gives us

$$\left(1+\frac{\alpha^2}{f^2}\right)\delta = -\frac{\alpha}{f^2}(g\Delta\eta + \delta_0) - \frac{1}{f}\zeta_0,$$

$$\delta = -\frac{1}{f^2+\alpha^2}\left(\alpha g\Delta\eta + \alpha\delta_0 + f\zeta_0\right).$$

Finally, substituting δ in equation (7), that reads $H\delta - \alpha\eta = \eta_0$, results in

$$-\frac{H}{f^2+\alpha^2}\left(\alpha g\Delta\eta + \alpha\delta_0 + f\zeta_0\right) - \alpha\eta = \eta_0,$$

which may be simplified into the elliptic Helmholtz equation by multiplying by $-\frac{f^2+\alpha^2}{H\alpha g}$

$$\Delta\eta + \kappa^2\eta = r_0 \tag{10}$$

where

$$\kappa^2 = \frac{f^2 + \alpha^2}{Hg}$$

and

$$r_0 = \eta_0 - \frac{1}{g}\delta_0 - \frac{f}{\alpha g}\zeta_0.$$

Once η is calculated, we need to retrieve the velocities by solving the 2×2 system formed by equations (5) and (6), which gives

$$u = \frac{f}{f^2 + \alpha^2}\left[g(\eta_y - \frac{\alpha}{f}\eta_x) + v_0 - \frac{\alpha}{f}u_0\right] \tag{11}$$

and

$$v = \frac{-\alpha}{f^2 + \alpha^2}\left[g(\eta_y + \frac{f}{\alpha}\eta_x) + v_0 - \frac{f}{\alpha}u_0\right]. \tag{12}$$

6.4 Interpretation of τ

We like to close this section with a brief discussion of τ by having a look on the REXI reformulation

$$(L - \frac{\alpha}{\tau})^{-1}U(\tau)\tau^{-1} = U(0)$$

We see, that for an increasing τ, hence an integration in time over a larger time period, the poles given by α are getting closer. This can possibly lead to a loss in accuracy for the data sampled by the outer poles α_{-N} and α_N. Therefore, the number N of poles is expected to scale linearly with the size of the coarse time step,

$$|N| \propto \tau.$$

Indeed, we saw in section 4 that for larger τ, M needs to be larger!

7 Bringing everything together

Using the spectral methods (e.g. in SWEET), we can directly solve the Helmholtz problem for the height in Eq. (10) and then solver for the velocity in Eqs. (11,12). Note that the Helmholtz problem is in complex space, as α is complex. this is straightforward with spectral methods. for finite difference/element methods, the problems needs to be split into its real and imaginary parts.

Then, the problem is reduced to computing the REXI as given in Eq. (3). We like to note again, that the α_n and β_n are independent of the system L to solve, and the number of coefficients only depends on the accuracy and the resolution.

8 Notes on HPC

- The terms in REXI to solve are all independent. Hence, for latency avoiding, the communication can be interleaved with computations.
- The iterative solvers are memory bound. Instead of computing $c := a * b$ for the stencil operations, we could compute $\vec{c} := a\vec{b}$ with a one coefficient in the stencil. This allows vectorization over c and b on accelerator cards with strided memory access.
- It is unknown which method is more efficient to solve the system of equations:
  - iterative solvers have low memory access,
  - inverting the system and storing it as a sparse matrix allows fast direct solving but can yield more memory access operations.
- Splitting the solver into real and complex number would store them consecutively in memory. This has a potential to avoid non-strided memory access and using the same SIMD operations (Just a rough idea, TODO: check if this is really the case).

9 Acknowledgements

Thanks to Terry for the feedback & discussions!

References

[1] Formulations of the shallow-water equations, M. Schreiber

[2] High-order time-parallel approximation of evolution operators, T. Haut et. al.

[3] An asymptotic parallel-in-time method for highly oscillatory PDEs, T. Haut et. al.

[4] An invariant theory of the linearized shallow water equations with rotation and its application to a sphere and a plane, N. Paldor et. al.

[5] Nineteen Dubious Ways to Compute the Exponential of a Matrix, Twenty-Five Years Later, Cleve Moler and Charles Van Loan, SIAM review

[6] Near optimal rational approximations of large data sets, Damle, A., Beylkin, G., Haut, T. S. & Monzon