REFACTORIZATION OF THE MIDPOINT RULE

JOHN BURKARDT * AND CATALIN TRENCHEA†

Key words. Backward Euler, midpoint rule, second-order, symplectic, Hamiltonian, energy conservation, A-stable and B-stable, blackbox / legacy code, partitioning algorithms, time adaptivity

Abstract. An alternative formulation of the midpoint method is employed to analyze its advantages as an implicit second-order absolutely stable timestepping method. Legacy codes originally using the backward Euler method can be upgraded to this method by inserting a single line of new code. We show that the midpoint method, and a theta-like generalization, are B-stable. We outline three estimates of local truncation error that allow adaptive time-stepping.

1. One line of code to change a Backward Euler code into to a second-order, unconditionally stable, conservative method. For the numerical approximation of a general evolution equation:

$$y'(t) = f(t, y(t)), \tag{1.1}$$

on the mesh points $\{t_n\}_{n\geq 0}$, and with the timestep τ_n, such that:

$$t_{n+1} = t_n + \tau_n, \quad t_{n+1/2} = t_n + \tfrac{1}{2}\tau_n,$$

we recall the classical midpoint quadrature rule:

$$\frac{y_{n+1} - y_n}{\tau_n} = f(t_{n+1/2}, y_{n+1/2}), \tag{1.2}$$

where $y_n \approx y(t_n)$. The method (1.2) is ubiquitously presented and used [8, 11, 15, 18, 19, 21, 26, 27, 29–31] in the apparently different form:

$$\frac{y_{n+1} - y_n}{\tau_n} = f\left(t_{n+1/2}, \frac{y_{n+1} + y_n}{2}\right). \tag{1.3}$$

The reason for the wide use of (1.3) instead of (1.2) (see e.g. [27, page 133]) is due to the natural question: '*but which value should we take for* $y_{n+1/2}$*?*'. The method (1.3) is an implicit second-order A-stable time-stepping method, and is the preferred method for solving evolutive conservative systems of partial differential equations (PDEs), along with the second order backward differentiation formula (BDF2) for dissipative PDEs.

From an algorithmic viewpoint, increasing the numerical accuracy of a complex legacy code, based on the first-order backward Euler (BE) method, to a second-order A-stable method, can be a difficult task. One straightforward solution would be to apply non-intrusive minimal modifications to the algorithm, i.e., by adding a few lines of code and post-processing the original BE solution into a 'filtered' higher-order solution. This is currently done in geophysical fluid dynamics, to improve the stability and accuracy of the solution to the leap-frog (explicit midpoint) method, by filtering it with Robert-Asselin or Robert-Asselin-Williams filters [3, 25, 33, 36–38, 40–42]. Recently, the BE solution was filtered into the solution to a second-order linear multistep method (LMM), similar to a BDF2 solution (see e.g., [24]), with a reduced discrete curvature and numerical dissipation.

Most LMMs [16, 27], when considered with variable steps, do not preserve the zero-stability or unconditional A-stability properties of the constant step versions. For example, the variable step version of the trapezoidal method (Crank-Nicolson) is unstable [16], [39, pp. 181-182]; similarly, BDF2 loses zero-stability and A-stability when used with a variable stepsize. The trapezoidal method, even in the constant step case, is *A*-stable but not *B*-stable [1]. Also, "it is not known which of the LMMs preserve quadratic invariants" [5].

An alternative non-intrusive modification to the BE method, with the goal of defining a family of second-order, variable step, unconditionally stable one-step methods, relies on the successful resolution of the above question regarding (1.2). This alternative is based on the fact that both the midpoint (1.2) and the trapezoidal methods can be viewed as a sequence of backward-Euler then forward-Euler methods, respectively a forward-Euler then a backward-Euler method, where the first computation is performed at the time $t_{n+1/2}$, see e.g., [26, page 223] and [17, page 57].

*Department of Mathematics, University of Pittsburgh, Pittsburgh, PA 15260, USA. Email: jvb25@pitt.edu.
†Department of Mathematics, University of Pittsburgh, Pittsburgh, PA 15260, USA. Email: trenchea@pitt.edu.

Consequently, consider implementing the midpoint rule (1.2) by solving a backward-Euler step at the half-integer time step $t_{n+1/2}$, followed by a forward-Euler step to t_{n+1}:

$$\frac{y_{n+1/2}-y_n}{\tau_n/2} = f(t_{n+1/2}, y_{n+1/2}), \tag{BE}$$

$$\frac{y_{n+1}-y_{n+1/2}}{\tau_n/2} = f(t_{n+1/2}, y_{n+1/2}). \tag{FE}$$

We point out that solving the equations (BE)-(FE) is equivalent to, and reduces to only solving (BE), and then applying a time-filter, as the (FE) step is equivalent to a linear extrapolation. Hence we evaluate $y_{n+1} = 2y_{n+1/2} - y_n$, and the equation (BE)-(FE) can be thought of as a single process designated as (BEFE):

$$\begin{cases} \dfrac{y_{n+1/2}-y_n}{\tau_n/2} = f(t_{n+1/2}, y_{n+1/2}), \\ y_{n+1} = 2y_{n+1/2} - y_n. \end{cases} \tag{BEFE}$$

Notice that the second step can also be written as:

$$y_{n+1/2} = \frac{y_{n+1}+y_n}{2},$$

and therefore both (1.2) *and* (1.3) *yield exactly the same numerical approximations*, i.e., (1.2) is a second-order accurate, unconditionally A-stable method. The second formulation (BEFE), while equivalent, makes it obvious how to bootstrap an existing Euler code to a second-order accurate, unconditionally energy stable, conservative, symplectic code. An important characteristic of (1.2) is the fact that it is a one-leg two-step method, which makes it easy to view it as a variable-step method, without losing the stability property. There are several options as to how to adapt the time-step τ_n (see e.g., [22, 23]), namely how to estimate the local truncation error.

This implementation (BEFE) of the midpoint method is consequential from the viewpoint of its potential applications for time-stepping methods of complex partial differential equations. The first advantage is the ease of non-intrusive implementation: it takes one line of code to transform a first-order dissipative method to a second-order accurate, energy conservative, stable method. (We recall that Dahlquist's barrier limits the accuracy of A-stable linear multistep methods to second-order.) Moreover, the midpoint rule is a symplectic method for general Hamiltonian systems, conserving all quadratic Hamiltonians [2, 5], unconditionally stable (A-stable and B-stable [1, 7]). Another important remark is that the constant in the local truncation error of (1.3), when seen in the implementation (BEFE), is $\frac{1}{24}$, instead of the usual $-\frac{1}{12}$. Thirdly, time-adaptivity can also be implemented with non-intrusive minimal algorithmic changes, mitigating the fact that the midpoint rule is not a Poisson map [26].

For coupled complex systems, like ocean-atmosphere, groundwater-surface water, fluid-structure interactions, or magnetohydrodynamics, the current trend is to employ partitioning methods of implicit-explicit type, which solve each equation separately by a legacy code, and transfer information between the subdomains and algorithms. This breaking of the monolithic approach routinely comes at the cost of stability. Most existing partitioned stable methods are only first-order accurate in time. The (BEFE) implementation opens the path of extending the current partitioning first-order stable methods to second-order accurate variable-step unconditionally stable methods, by manipulating the computed solution at $t_{n+1/2}$ in a stable manner. Recently, this approach has been applied to problems in fluid-structure interaction [6], magnetohydrodynamics and ocean-atmosphere modeling. Note also that the computed solution at $t_{n+1/2}$ allows further manipulation, such as modular spatial filtering, in order to improve the qualitative properties of the numerical simulations [34, 35].

2. Generalization to a θ-like method. We remark also that (BEFE) is a particular instance of the one-leg 'θ-like' method:

$$\frac{y_{n+1}-y_n}{\tau_n} = f(t_{n+\theta}, y_{n+\theta_n}), \tag{2.1}$$

implemented as:

$$\begin{cases} \dfrac{y_{n+\theta_n}-y_n}{\theta_n\tau_n} = f(t_{n+\theta_n}, y_{n+\theta_n}), \\ \dfrac{y_{n+1}-y_{n+\theta_n}}{(1-\theta_n)\tau_n} = f(t_{n+\theta_n}, y_{n+\theta_n}). \end{cases} \tag{2.2}$$

which can be rewritten as:

$$\begin{cases} \dfrac{y_{n+\theta_n}-y_n}{\theta_n\tau_n} = f(t_{n+\theta_n}, y_{n+\theta_n}), \\ y_{n+1} = \frac{1}{\theta_n} y_{n+\theta_n} - \left(\frac{1}{\theta_n}-1\right) y_n. \end{cases} \tag{2.3}$$

Notice that (2.1) is not the classical linear multistep θ method [20, page 182], but Cauchy's one-leg version (see e.g. [9, pp. 40], also [12–14]):

$$\frac{y_{n+1}-y_n}{\tau_n} = f\big(t_{n+\theta}, \theta_n y_{n+1} + (1-\theta_n) y_n\big), \tag{2.4}$$

since, as above, we have from the second part of (2.3) that $y_{n+\theta_n} = \theta_n y_{n+1} + (1-\theta_n) y_n$.

REMARK 1. *It was recently proved in [4], for the Navier-Stokes equations, that the solutions constructed using the one-leg method* (2.4) *(with* $\theta_n = \frac{1}{2} + \tau_n^{1-\varepsilon}$*) for the finite-difference time-discretization and the finite elements methods for the spatial-discretization, give rise to suitable weak solutions in the sense of Scheffer and Caffarelli-Kohn-Nirenberg.* In the following we mean stability in the sense of B-stability [7], which implies A-stability [12]. We say a method is B-stable if for any f satisfying the condition that $\langle f(u)-f(v), u-v\rangle \le 0$ for u, v in a Hilbert space, it holds that $\|y_{n+1}-z_{n+1}\| \le \|y_n - z_n\|$, where $\{y_n\}_{n\ge 0}, \{z_n\}_{n\ge 0}$ are two sequences of approximations computed with the method. Dahlquist introduced a similar criterion for certain types of multistep methods, G-stability (Dahlquist 1975, see e.g., [11] or [28, p.308]), which is equivalent to A-stability for constant step linear multistep methods.

PROPOSITION 2.1. *The midpoint method* (BE)-(FE)*, and the* θ*-method* (2.2) *for* $\frac{1}{2} \le \theta_n \le 1$*, are unconditionally-stable, and the following equality holds:*

$$\frac{1}{2}\|y_{n+1}\|^2 - \frac{1}{2}\|y_n\|^2 + \frac{2\theta_n - 1}{2}\|y_{n+1}-y_n\|^2 = \tau_n \langle f(t_{n+\theta_n}, y_{n+\theta_n}), y_{n+\theta_n}\rangle.$$

Proof. We prove the result only for (2.2), since the midpoint method is obtained by taking $\theta_n = 1/2$. First, for B-stability, we consider the equation (2.4) for $\{y_{n+1}\}$ and respectively $\{z_{n+1}\}$. Then subtract, take the inner product with $\tau_n(y_{n+\theta_n} - z_{n+\theta_n})$, use the Cauchy-Schwarz inequality and the definition to obtain

$$\begin{aligned} 0 &\ge \tau_n \langle f(y_{n+\theta_n}) - f(z_{n+\theta_n}), y_{n+\theta_n} - z_{n+\theta_n}\rangle \\ &= \langle y_{n+1} - z_{n+1}, \theta_n(y_{n+1}-z_{n+1}) + (1-\theta_n)(y_n - z_n)\rangle \\ &\ge \big(\theta_n \|y_{n+1}-z_{n+1}\| + (1-\theta_n)\|y_n - z_n\|\big)\big(\|y_{n+1}-z_{n+1}\| - \|y_n - z_n\|\big), \end{aligned}$$

which yields $\|y_{n+1}-z_{n+1}\| \le \|y_n - z_n\|$.
For the energy equality we proceed in a similar manner. Multiplying both equations in (2.2) by $\theta_n \tau_n y_{n+\theta_n}$ and $(1-\theta_n)\tau_n y_{n+\theta_n}$ respectively, and applying the polarization identity we obtain:

$$\begin{aligned} \frac{1}{2}\|y_{n+\theta_n}\|^2 - \frac{1}{2}\|y_n\|^2 + \frac{1}{2}\|y_{n+\theta_n} - y_n\|^2 &= \theta_n \tau_n f(t_{n+\theta_n}, y_{n+\theta_n}) y_{n+\theta_n}, \\ \frac{1}{2}\|y_{n+1}\|^2 - \frac{1}{2}\|y_{n+\theta_n}\|^2 - \frac{1}{2}\|y_{n+1} - y_{n+\theta_n}\|^2 &= (1-\theta_n)\tau_n f(t_{n+\theta_n}, y_{n+\theta_n}) y_{n+\theta_n}. \end{aligned}$$

Summation and the use of (2.2) completes the argument. □

3. Time-step adaptivity. We begin this section by a small observation: the local truncation error[1] of the midpoint method (BEFE) is:

$$T_{n+1} = \frac{1}{24}\tau_n^3 y'''(t_{n+1/2}) + \mathscr{O}(\tau_n^5). \tag{3.1}$$

The same formula holds for the '$\theta-$like' method (2.1), provided $\theta_n = \frac{1}{2} + \frac{1}{2}\tau_n^2$.

[1] The local truncation error holds provided the solution is smooth enough.

Therefore, we can adaptively adjust the time step τ_n by enforcing an estimate of the local truncation error (1.2), denoted $\widehat{T}_{n+1}$, to equal a tolerance, i.e., such that the $\|\widehat{T}_{n+1}\| \approx \texttt{tol}$ (see e.g. [23]). The time-step τ_n^{new} which imposes that $\hat{T}_{n+1}$ is sufficiently small is given by:

$$\tau_n^{new} = \kappa \tau_n \left| \frac{\texttt{tol}}{\|\widehat{T}_{n+1}\|} \right|^{\frac{1}{3}}, \tag{3.2}$$

where $\kappa = 1$. In our computations, we found that more conservative coefficient values of the (safety coefficient [27, p.168], [26, p.255]) $\kappa = 0.90 \div 0.95$ minimize the number of time step rejections in the adaptive algorithm, while increasing the number of time intervals[2].

There are numerous ways in which the time-step adaptivity can be implemented (see e.g. [23]), out of which we present three methods. The first choice is based on the estimation of the LTE using Taylor expansions. The other two options estimate the local truncation error by the difference between the numerical midpoint solution and a second-order, and respectively a third-order approximation, given by formulae similar to the explicit Adams-Bashforth 2 (AB2) and Adams-Bashforth 3 (AB3) methods. These two methods are related to the classical AB2 and AB3 (see e.g., [27, p. 398]), the difference being that they use the function values evaluated at half-times $f_{n+1/2}, f_{n-1/2}, f_{n-3/2}, f_{n-5/2}$.

Result: *Adaptive midpoint rule*
initialization: set tol, compute y_1 and τ_0 with a one step second-order accurate method, such that τ_0 is in the convergence range (see e.g., [10, page 367]);
compute y_2 and τ_1 with a second order accurate method, $t_2 = t_1 + \tau_1$;
$t^{\text{new}} = t_2$, $\tau^{\text{new}} = \tau_1$;
for $n \geq 2$ (i.e., t^{new}, τ^{new}, y_n, y_{n-1}, y_{n-2} are given);
while $t^{\text{new}} \leq T$ **do**
  $\tau_n \leftarrow \tau^{\text{new}}$;
  evaluate y_{n+1} with the midpoint rule (1.2);
  evaluate $\widehat{T}_{n+1}$ with (LTE-Taylor), (LTE-AB2) or (LTE-AB3);
  $\tau^{\text{new}} \leftarrow \kappa \tau_n \left| \texttt{tol} / \|\widehat{T}_{n+1}\| \right|^{\frac{1}{3}}$;
  if $\|\widehat{T}_{n+1}\| \leq \texttt{tol}$ **then**
    $t_{n+1} \leftarrow t_n + \tau^{\text{new}}$, $t^{\text{new}} \leftarrow t_{n+1}$, $n \leftarrow n+1$
  end
end

3.1. Estimation of the local truncation error using Taylor expansions. In order to estimate the numerical value of $\widehat{T}_{n+1}$, we need to evaluate $y'''(t_n)$. We proceed by using Taylor expansions:

$$\begin{aligned}
y'(t_{n+1/2}) &= y'(t_n) + \frac{\tau_n}{2} y''(t_n) + \frac{\tau_n^2}{8} y'''(t_n) + \mathscr{O}(\tau_n^3), \\
y'(t_{n-1/2}) &= y'(t_n) - \frac{\tau_{n-1}}{2} y''(t_n) + \frac{\tau_{n-1}^2}{8} y'''(t_n) + \mathscr{O}(\tau_{n-1}^3), \\
y'(t_{n-3/2}) &= y'(t_n) - \frac{2\tau_{n-1} + \tau_{n-2}}{2} y''(t_n) + \frac{(2\tau_{n-1} + \tau_{n-2})^2}{8} y'''(t_n) + \mathscr{O}(\tau_{n-1}^3 + \tau_{n-2}^3),
\end{aligned}$$

which, eliminating $y'(t_n)$ and $y''(t_n)$, gives:

$$\frac{y'(t_{n+1/2}) - y'(t_{n-1/2})}{\tau_n + \tau_{n-1}} - \frac{y'(t_{n-1/2}) - y'(t_{n-3/2})}{\tau_{n-1} + \tau_{n-2}} = \frac{1}{8}(\tau_n + 2\tau_{n-1} + \tau_{n-2}) y'''(t_n) + \mathscr{O}(\tau_n^2 + \tau_{n-1}^2 + \tau_{n-2}^2).$$

Using the numerical method (1.2), the LTE (3.1) can finally be estimated in terms of the computed solutions:

$$\widehat{T}_{n+1} = \frac{\tau_n^3}{3(\tau_n + 2\tau_{n-1} + \tau_{n-2})} \left(\frac{f_{n+1/2} - f_{n-1/2}}{\tau_n + \tau_{n-1}} - \frac{f_{n-1/2} - f_{n-3/2}}{\tau_{n-1} + \tau_{n-2}} \right) \tag{LTE-Taylor}$$

[2] From (3.2) we see that if $\widehat{T}_{n+1} > \texttt{tol}$, then τ_n is decreased, and the algorithm repeats the midpoint rule step with a reduced time-step. Respectively, if $\widehat{T}_{n+1} \leq \texttt{tol}$, then τ_n is increased, and the computation moves to the next time interval, with the increased time step. The safety factor $\kappa < 1$ reduces the probability of the new time-steps being rejected in the [**if** $\|\widehat{T}_{n+1}\| \leq \texttt{tol}$] test in the Algorithm 1.

$$= \frac{\tau_n^3}{3(\tau_n+2\tau_{n-1}+\tau_{n-2})}\left(\frac{\frac{y_{n+1}-y_n}{\tau_n}-\frac{y_n-y_{n-1}}{\tau_{n-1}}}{\tau_n+\tau_{n-1}}-\frac{\frac{y_n-y_{n-1}}{\tau_{n-1}}-\frac{y_{n-1}-y_{n-2}}{\tau_{n-2}}}{\tau_{n-1}+\tau_{n-2}}\right)$$
$$= \frac{\tau_n^3}{3(\tau_n+2\tau_{n-1}+\tau_{n-2})}\Big(y_{n+1}\frac{1}{\tau_n(\tau_n+\tau_{n-1})}-y_n\frac{\tau_n+\tau_{n-1}+\tau_{n-2}}{\tau_n\tau_{n-1}(\tau_{n-1}+\tau_{n-2})}$$
$$+y_{n-1}\frac{\tau_n+\tau_{n-1}+\tau_{n-2}}{\tau_{n-1}\tau_{n-2}(\tau_n+\tau_{n-1})}-y_{n-2}\frac{1}{\tau_{n-2}(\tau_{n-1}+\tau_{n-2})}\Big).$$

3.2. Estimation of the local truncation error using a variable step AB-2 solution. Here we estimate the local truncation error at t_{n+1} by evaluating the difference between the $\mathscr{O}(\Delta t^2)$ midpoint-solution y_{n+1} and another second-order approximation, $\widetilde{y_{n+1}^{AB2}}$, obtained by a variable-step Adams-Bashforth 2-like method. Let $\Pi_1(t)$ be the polynomial interpolating $f(y(t))$ at nodes $\{t_{n-1/2}, t_{n-3/2}\}$ and values $\{f_{n-1/2}, f_{n-3/2}\}$, which by (BEFE) denote:

$$f_{n-1/2}=\frac{y_n-y_{n-1}}{\tau_{n-1}}, \qquad f_{n-3/2}=\frac{y_{n-1}-y_{n-2}}{\tau_{n-2}}.$$

Then the solution to the AB2-like with variable step is:

$$\widetilde{y_{n+1}^{AB2}} = y_n+\int_{t_n}^{t_{n+1}}\Pi_1(t)\,dt = y_n+f_{n-1/2}\frac{\tau_n(\tau_n+2\tau_{n-1}+\tau_{n-2})}{\tau_{n-1}+\tau_{n-2}}-f_{n-3/2}\frac{\tau_n(\tau_n+\tau_{n-1})}{\tau_{n-1}+\tau_{n-2}}$$
$$= y_n\frac{(\tau_n+\tau_{n-1})(\tau_n+\tau_{n-1}+\tau_{n-2})}{\tau_{n-1}(\tau_{n-1}+\tau_{n-2})}-y_{n-1}\frac{\tau_n(\tau_n+\tau_{n-1}+\tau_{n-2})}{\tau_{n-1}\tau_{n-2}}+y_{n-2}\frac{\tau_n(\tau_n+\tau_{n-1})}{\tau_{n-2}(\tau_{n-1}+\tau_{n-2})}, \qquad \text{(AB2-like)}$$

and its local truncation error (under the 'localization assumption', i.e. back values are exact, see e.g. [23, p.70], [32, p.56]) can be written:

$$\widetilde{T_{n+1}^{AB2}} = \tau_n^3 y'''(t_{n+1/2})\left(\frac{1}{24}+\frac{1}{8}\left(1+\frac{\tau_{n-1}}{\tau_n}\right)\left(1+2\frac{\tau_{n-1}}{\tau_n}+\frac{\tau_{n-2}}{\tau_n}\right)\right).$$

For brevity, we denote the error coefficient in the right hand side, which depends on timestep ratios, by:

$$\mathscr{R}_n=\frac{1}{24}+\frac{1}{8}\left(1+\frac{\tau_{n-1}}{\tau_n}\right)\left(1+2\frac{\tau_{n-1}}{\tau_n}+\frac{\tau_{n-2}}{\tau_n}\right).$$

Then, from (3.1) and the expression above, we obtain the following approximation of the local truncation error of the midpoint rule (BEFE):

$$\widehat{T}_{n+1}=(y_{n+1}^{\text{midpoint}}-\widetilde{y_{n+1}^{AB2}})\frac{1}{1-1/(24\mathscr{R}_n)}, \qquad \text{(LTE-AB2)}$$

where $y_{n+1}^{\text{midpoint}}$ denotes the midpoint solution from (BEFE), and $\widetilde{y_{n+1}^{AB2}}$ is given in (AB2-like).

3.3. Estimation of the local truncation error using a variable step AB-3 solution. We choose to estimate the local truncation error at t_{n+1} by evaluating the difference between the $\mathscr{O}(\Delta t^2)$ midpoint-solution y_{n+1} and a third-order approximation, u_{n+1}, obtained by the variable-step Adams-Bashforth 3 method (see e.g., [27, p. 398]). We denote:

$$f_{n-1/2}=\frac{y_n-y_{n-1}}{\tau_{n-1}}, \qquad f_{n-3/2}=\frac{y_{n-1}-y_{n-2}}{\tau_{n-2}}, \qquad f_{n-5/2}=\frac{y_{n-2}-y_{n-3}}{\tau_{n-3}},$$

and let $\Pi_2(t)$ be the polynomial interpolating $f(y(t))$ at nodes $\{t_{n-1/2}, t_{n-3/2}, t_{n-5/2}\}$ and values $\{f_{n-1/2}, f_{n-3/2}, f_{n-5/2}\}$:

$$\Pi_2(t)=f_{n-1/2}+\frac{f_{n-1/2}-f_{n-3/2}}{t_{n-1/2}-t_{n-3/2}}(t-t_{n-1/2})+\frac{\frac{f_{n-1/2}-f_{n-3/2}}{t_{n-1/2}-t_{n-3/2}}-\frac{f_{n-3/2}-f_{n-5/2}}{t_{n-3/2}-t_{n-5/2}}}{t_{n-1/2}-t_{n-5/2}}(t-t_{n-1/2})(t-t_{n-3/2}).$$

Hence:

$$u_{n+1}\approx y_n+\int_{t_n}^{t_{n+1}}\Pi_2(t)\,dt = y_n+\tau_n\left[f_{n-1/2}+\frac{f_{n-1/2}-f_{n-3/2}}{t_{n-1/2}-t_{n-3/2}}\frac{\tau_n+\tau_{n-1}}{2}\right.$$

$$+\frac{\frac{f_{n-1/2}-f_{n-3/2}}{t_{n-1/2}-t_{n-3/2}}-\frac{f_{n-3/2}-f_{n-5/2}}{t_{n-3/2}-t_{n-5/2}}}{t_{n-1/2}-t_{n-5/2}}\cdot\left(\frac{1}{3}\tau_n^2+\frac{1}{2}\tau_{n-1}^2+\frac{3}{4}\tau_n\tau_{n-1}+\frac{1}{4}\tau_n\tau_{n-2}+\frac{1}{4}\tau_{n-1}\tau_{n-2}\right)\Bigg],$$

and therefore the local truncation error can be approximated by:

$$\widehat{T}_{n+1}=\tau_n\Bigg[f_{n-1/2}+\frac{f_{n-1/2}-f_{n-3/2}}{t_{n-1/2}-t_{n-3/2}}\frac{\tau_n+\tau_{n-1}}{2} \tag{LTE-AB3}$$
$$+\frac{\frac{f_{n-1/2}-f_{n-3/2}}{t_{n-1/2}-t_{n-3/2}}-\frac{f_{n-3/2}-f_{n-5/2}}{t_{n-3/2}-t_{n-5/2}}}{t_{n-1/2}-t_{n-5/2}}\cdot\left(\frac{1}{3}\tau_n^2+\frac{1}{2}\tau_{n-1}^2+\frac{3}{4}\tau_n\tau_{n-1}+\frac{1}{4}\tau_n\tau_{n-2}+\frac{1}{4}\tau_{n-1}\tau_{n-2}\right)\Bigg].$$

3.4. The conservation property. One of the most characteristic features of an ODE solver is its accuracy. An ODE problem is often thought of simply as a need to approximate the value $y(T)$ to high precision with minimal effort. Often, however, the variables controlled by an ODE inherently obey implicit laws as well. In a sense, the variables have both a landscape or manifold of allowed behaviors, and a dynamic that defines their progression in time from one behavior to another. As a simple example, in the common SIR model of infection, there is generally an implicit conservation law that $S+I+R=$ constant, which could be made explicit by summing the three equations. An ODE solver applied to an SIR model may produce solution values that are accurate, in the sense that the computed S(T), I(T), and R(T) are close to the true values, while not exactly satisfying the conservation law. Thus, a small, perhaps fractional number of people have been created by the limited accuracy of the solver. Often, such deviations are not noticed, or regarded as an unavoidable error. But a conservation law can represent a physical law (conservation of energy or mass) or a logical law (conservation of the total population). Thus aside from accuracy, another important feature of an ODE solver can be the ability to preserve a conserved quantity exactly. This is especially true in physics, astronomy, and even climate simulation, where researchers modeling the flow of a glacier over a century expect the initial and final masses to agree until the last few decimals. Correspondingly, the mathematical field of geometric integration [26] has arisen, concentrating on the geometric structures imposed by satisfying conservation laws.

One of the remarkable features of the midpoint method is that it can preserve conserved quantities that are implicit in a system of ODE's, as long as that quantity can be expressed in terms of a polynomial in the state variables, of quadratic degree or less. Thus, if the midpoint method is applied to an SIR model, the total population stays exactly the same, from beginning to end. The midpoint method conserves the energy of many simple mechanical systems, such as the pendulum. The conservation law for a common predator-prey model involves logarithms, and so will not be conserved exactly. However, even here, the midpoint method can stay much closer to the original conservation value than many other methods.

As a simple example of a conservation law, we can consider an ODE that describes a closed path along the surface of a unit sphere, as mentioned by Hairer [27].

$$\begin{aligned}\frac{dx}{dt}&=(\frac{1}{c}-\frac{1}{b})zy\\ \frac{dy}{dt}&=(\frac{1}{a}-\frac{1}{c})xz\\ \frac{dz}{dt}&=(\frac{1}{b}-\frac{1}{a})yx\end{aligned}$$

with $a=1.6, b=1, c=\frac{2}{3}$. A conserved quantity is, of course, $x^2+y^2+z^2=1$

Here, it is obvious that there must be both a landscape (the surface) and a dynamic (how to move) that are wrapped together into the ODE system. An ODE solver that approximates both landscape and dynamic will construct a path that immediately leaves the surface of the sphere and gradually drifts away. This deviation can be reduced but never eliminated. As our test, we start at time $t=0.0$ with the initial condition $(x_0,y_0,z_0)=(\cos(0.9),0.0,\sin(0.9))$. Figure [3.1] compares the conservation plots for solutions using 1,000 Euler steps versus 100 midpoint steps and 100 RK4 steps: in integrating to $T=50$. While the Euler method has struggled, the higher precision midpoint and RK4 methods seem to have done an excellent job. MATLAB's suite of ODE solvers, best typified by `ode45()`, which include adaptive stepping, also showed no obvious conservation difficulty.

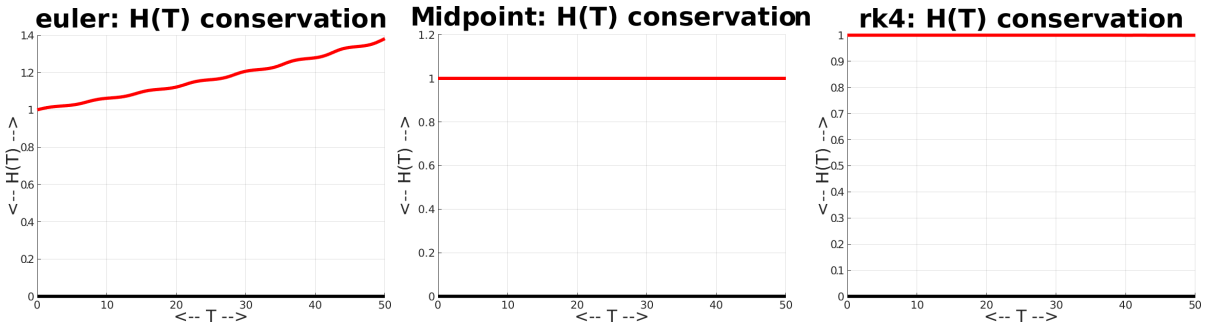

FIG. 3.1. *Conservation over* $[0,50]$ *for Euler, Midpoint and RK4.*

As a stronger test, then, we extend the problem interval to $0 \leq t \leq 10,000$. For the fixed step Midpoint and RK4 codes, we use 20,000 equal steps, while for the adaptive MATLAB solvers `ode15s()`, `ode23()`, `ode23s`, `ode45()` use their default settings. The Midpoint method has a perfect conservation history, while the RK4 results

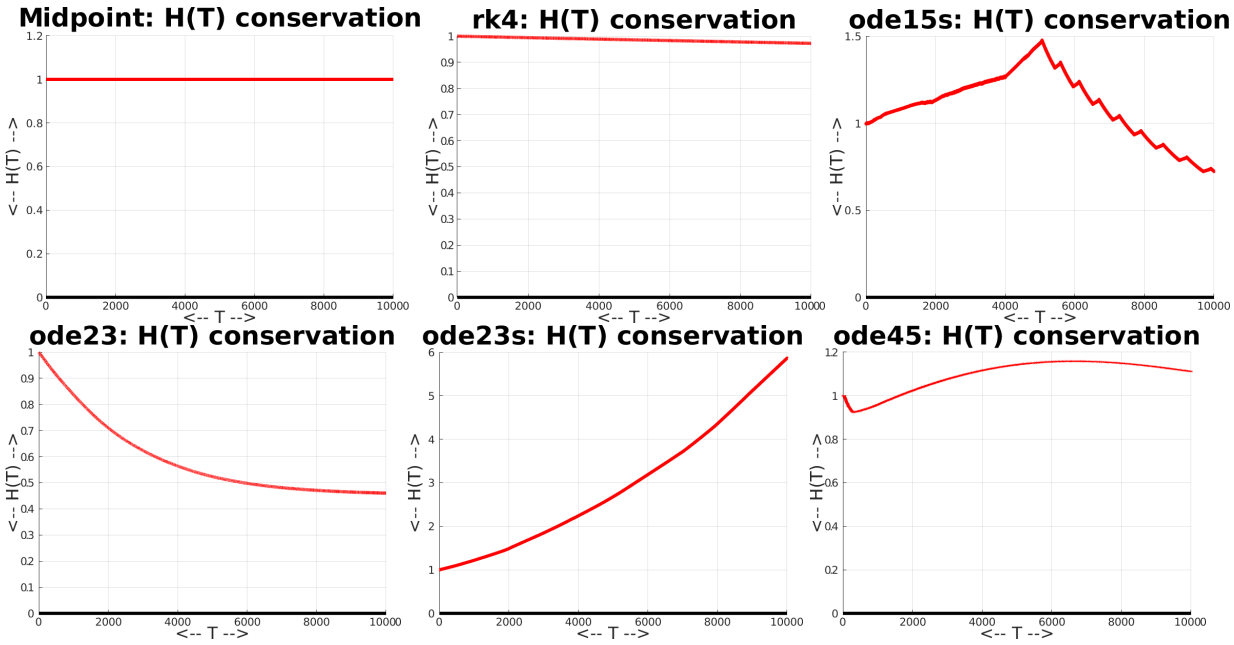

FIG. 3.2. *Conservation over* $[0,100000]$ *for Midpoint, RK4 and MATLAB solvers.*

are just beginning to show deterioration. On the other hand, each of the MATLAB solvers shows a serious failure in terms of conservation.

For this demonstration, a fixed stepsize version of the Midpoint method was used. It would be worthwhile to repeat the calculations using an adaptive version of the algorithm.

The conservation property of the Midpoint method applies to situations where the conserved quantity is a polynomial of at most quadratic degree in the variables. This is a common property of many systems from areas such as mechanics, electrodynamics, and astrophysics. However, the familiar Lotka-Volterra equations have a conserved quantity that involves logarithms of the state variables, and so is an example in which the Midpoint method cannot guarantee conservation.

While it seems natural to concentrate on accuracy or precision in an ODE solver, the example of the motion on a sphere may suggest why conservation is sometimes also a vital property. For a long term calculation, an accurate method may give us a solution that is "very close" to the true answer, but which unrealistically has left the surface of the sphere. A conservative solver may produce a solution that is further from the true solution, but which remains

on the sphere. It depends on the user's interest to decide which of these two solutions is actually furthest from "the truth".

REFERENCES

[1] G. AKRIVIS, *Numerical Methods for initial value problems*, Lecture Notes from a course taught at BCAM, Basque Center for Applied Mathematics, Bilbao, Basque Country, Spain, 2012.

[2] U. M. ASCHER AND S. REICH, *The midpoint scheme and variants for Hamiltonian systems: advantages and pitfalls*, SIAM J. Sci. Comput., 21 (1999), pp. 1045–1065.

[3] R. ASSELIN, *Frequency filter for time integrations*, Mon. Wea. Rev., 100 (1972), pp. 487–490.

[4] L. C. BERSELLI, S. FAGIOLI, AND S. SPIRITO, *Suitable weak solutions of the Navier-Stokes equations constructed by a space-time numerical discretization*, J. Math. Pures Appl. (9), 125 (2019), pp. 189–208.

[5] P. B. BOCHEV AND C. SCOVEL, *On quadratic invariants and symplectic structure*, BIT, 34 (1994), pp. 337–345.

[6] M. BUKAC AND C. TRENCHEA, *Boundary update via resolvent for fluid-structure interaction*, tech. rep., University of Pittsburgh, 2018.

[7] J. C. BUTCHER, *A stability property of implicit Runge-Kutta methods*, BIT Numerical Mathematics, 15 (1975), pp. 358–361.

[8] ———, *Numerical methods for ordinary differential equations*, John Wiley & Sons, Ltd., Chichester, third ed., 2016. With a foreword by J. M. Sanz-Serna.

[9] A.-L. CAUCHY, *Équations différentielles ordinaires*, Éditions Études Vivantes, Ltée., Ville Saint-Laurent, QC; Johnson Reprint Corp., New York, 1981. Cours inédit. Fragment. [Unpublished course. Fragment], With a preface by Jean Dieudonné, With an introduction by Christian Gilain.

[10] S. D. CONTE AND C. DE BOOR, *Elementary numerical analysis*, vol. 78 of Classics in Applied Mathematics, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2018.

[11] G. G. DAHLQUIST, *On stability and error analysis for stiff non-linear problems, Part I*, Dept. of Comp. Sci. Roy. Inst. of Technology, Report TRITA-NA-7508 (1975).

[12] ———, *Error analysis for a class of methods for stiff non-linear initial value problems*, in Numerical Analysis, G. Watson, ed., vol. 506 of Lecture Notes in Mathematics, Springer Berlin Heidelberg, 1976, pp. 60–72.

[13] ———, *G-stability is equivalent to A-stability*, BIT, 18 (1978), pp. 384–401.

[14] ———, *On one-leg multistep methods*, SIAM J. Numer. Anal., 20 (1983), pp. 1130–1138.

[15] G. G. DAHLQUIST AND Å. BJÖRCK, *Numerical methods*, Dover Publications, Inc., Mineola, NY, 2003. Translated from the Swedish by Ned Anderson, Reprint of the 1974 English translation.

[16] G. G. DAHLQUIST, W. LINIGER, AND O. NEVANLINNA, *Stability of two-step methods for variable integration steps*, SIAM J. Numer. Anal., 20 (1983), pp. 1071–1085.

[17] D. R. DURRAN, *Numerical methods for fluid dynamics*, vol. 32 of Texts in Applied Mathematics, Springer, New York, second ed., 2010. With applications to geophysics.

[18] W. E, *Principles of multiscale modeling*, Cambridge University Press, Cambridge, 2011.

[19] C. W. GEAR, *Numerical initial value problems in ordinary differential equations*, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1971.

[20] V. GIRAULT AND P.-A. RAVIART, *Finite element approximation of the Navier-Stokes equations*, vol. 749 of Lecture Notes in Mathematics, Springer-Verlag, Berlin, 1979.

[21] E. GODLEWSKI AND P.-A. RAVIART, *Numerical approximation of hyperbolic systems of conservation laws*, vol. 118 of Applied Mathematical Sciences, Springer-Verlag, New York, 1996.

[22] P. GRESHO, R. SANI, AND M. ENGELMAN, *Incompressible flow and the finite element method: advection-diffusion and isothermal laminar flow*, Incompressible Flow & the Finite Element Method, Wiley, 1998.

[23] D. F. GRIFFITHS AND D. J. HIGHAM, *Numerical methods for ordinary differential equations*, Springer Undergraduate Mathematics Series, Springer-Verlag London, Ltd., London, 2010. Initial value problems.

[24] A. GUZEL AND W. LAYTON, *Time filters increase accuracy of the fully implicit method*, BIT, 58 (2018), pp. 301–315.

[25] A. GUZEL AND C. TRENCHEA, *The Williams step increases the stability and accuracy of the hoRA time filter*, Appl. Numer. Math., 131 (2018), pp. 158–173.

[26] E. HAIRER, C. LUBICH, AND G. WANNER, *Geometric numerical integration*, vol. 31 of Springer Series in Computational Mathematics, Springer, Heidelberg, 2010. Structure-preserving algorithms for ordinary differential equations, Reprint of the second (2006) edition.

[27] E. HAIRER, S. P. NØRSETT, AND G. WANNER, *Solving ordinary differential equations. I*, vol. 8 of Springer Series in Computational Mathematics, Springer-Verlag, Berlin, second ed., 1993. Nonstiff problems.

[28] E. HAIRER AND G. WANNER, *Solving ordinary differential equations. II*, vol. 14 of Springer Series in Computational Mathematics, Springer-Verlag, Berlin, 2010. Stiff and differential-algebraic problems, Second revised edition.

[29] W. HUNDSDORFER AND J. VERWER, *Numerical solution of time-dependent advection-diffusion-reaction equations*, vol. 33 of Springer Series in Computational Mathematics, Springer-Verlag, Berlin, 2003.

[30] A. ISERLES, *A first course in the numerical analysis of differential equations*, Cambridge Texts in Applied Mathematics, Cambridge University Press, Cambridge, 1996.

[31] J. D. LAMBERT, *Computational methods in ordinary differential equations*, John Wiley & Sons, London-New York-Sydney, 1973. Introductory Mathematics for Scientists and Engineers.

[32] ———, *Numerical methods for ordinary differential systems*, John Wiley & Sons, Ltd., Chichester, 1991. The initial value problem.

[33] W. LAYTON, Y. LI, AND C. TRENCHEA, *Recent developments in IMEX methods with time filters for systems of evolution equations*, J. Comput. Appl. Math., 299 (2016), pp. 50–67.

[34] W. LAYTON, N. MAYS, M. NEDA, AND C. TRENCHEA, *Numerical analysis of modular regularization methods for the BDF2 time discretization of the Navier-Stokes equations*, ESAIM Math. Model. Numer. Anal., 48 (2014), pp. 765–793.

[35] W. LAYTON, L. G. REBHOLZ, AND C. TRENCHEA, *Modular nonlinear filter stabilization of methods for higher Reynolds numbers flow*,

J. Math. Fluid Mech., 14 (2012), pp. 325–354.
[36] Y. LI AND C. TRENCHEA, *A higher-order Robert–Asselin type time filter*, J. Comput. Phys., 259 (2014), pp. 23–32.
[37] ———, *Analysis of time filters used with the leapfrog scheme*, in Proceedings of the VI Conference on Computational Methods for Coupled Problems in Science and Engineering, Venice, Italy, May 2015, pp. 1261–1272.
[38] A. J. ROBERT, *The integration of a spectral model of the atmosphere by the implicit method*, Proc. WMO-IUGG Symp. on NWP, Tokyo, Japan Meteorological Agency, (1969), pp. 19–24.
[39] H. J. STETTER, *Analysis of discretization methods for ordinary differential equations*, Springer-Verlag, New York-Heidelberg, 1973. Springer Tracts in Natural Philosophy, Vol. 23.
[40] P. D. WILLIAMS, *A proposed modification to the Robert–Asselin time filter*, Mon. Wea. Rev., 137 (2009), pp. 2538–2546.
[41] ———, *The RAW filter: An improvement to the Robert–Asselin filter in semi-implicit integrations*, Mon. Wea. Rev., 139 (2011), pp. 1996–2007.
[42] ———, *Achieving seventh-order amplitude accuracy in leapfrog integrations*, Mon. Wea. Rev., 141 (2013), pp. 3037–3051.