From semidiscrete to fully discrete: stability of Runge-Kutta schemes by the energy method. II

Eitan Tadmor*

Abstract

We study the stability of Runge-Kutta methods for the time integration of semidiscrete systems associated with time dependent PDEs. These semidiscrete systems amount to large systems of ODEs with the possibility that the matrices involved are far from being normal. The stability question of their Runge-Kutta methods, therefore, cannot be addressed by the familiar scalar arguments of eigenvalues lying in the corresponding region of absolute stability. Instead, we replace this scalar spectral analysis by the energy method, where stability of the fully discrete Runge-Kutta methods takes into account the full eigenstructure of the problem at hand.

We discuss two energy method approaches that guarantee the stability of fully-discrete Runge-Kutta methods for sufficiently small CFL condition, $\Delta t \leq \Delta t_0$. In the first approach, Runge-Kutta methods are shown to preserve stability for the subclass of coercive semidiscrete problems. A second approach treats the more general class of semibounded problems. It is shown that their time integration by third-order Runge-Kutta method is stable under a slightly more restrictive CFL condition.

We conclude by utilizing these two approaches to examine the stability of Runge-Kutta discretizations of semidiscrete advection-diffusion problems. Our study includes a detailed stability analysis for prototype examples of one-sided and centered finite differencing and pseudospectral Fourier and Jacobi-based methods.

*Department of Mathematics, UCLA, Los-Angeles CA 90095, USA. email: tadmor@math.ucla.edu. Research was supported in part by NSF grants DMS01-07428, DMS01-07917 and ONR grant N00014-91-J-1076.

Contents

1 Introduction. The semidiscrete method of lines

We are concerned with the stability of Runge-Kutta (RK) methods for the approximate solution of time-dependent problems

$$\partial_t u = L(x, t, \partial_x)u. \tag{1}$$

For simplicity of the presentation our discussion ignores the explicit time dependence, assuming $L(x, t, \partial_x) = L(x, \partial_x)$, and we note in passing that our results apply, mutatis mutandis, to the case of time-dependent coefficients. Consult for example, our discussion in §5.2 below.

As a first step in the discretization of (1), the linear differential operator L is replaced by an appropriate spatial discretization. In a typical scenario, such L's are replaced by finite dimensional approximations of the form $L_N = P_N L P_N$, where P_N is a projection to an N-dimensional computational space involving a small spatial scale, e.g., a small grid size of order $\Delta x \sim 1/N$. The resulting semidiscrete approximation of (1), called the *method of lines*, amounts to an $N \times N$ system of ODEs of the form

$$\frac{du_N}{dt} = L_N u_N, \tag{2}$$

and we are interested to know whether the numerical solution of (2) by Runge-Kutta (RK) methods preserves the stability of such approximants, *independently* of their increasing size with N.

The differential operator $L = L(x, \partial_x)$ is local[1]. The discrete L_N's, however, need not have any special local structure. Boundary conditions and other side constraints may destroy the local property of finite difference stencils and pseudospectral Fourier and Jacobi-based methods naturally lead to global stencils. In general, the L_N's need not be uniformly diagonalizable, with the possibility of ill-conditioning due to increasingly large condition numbers as N increases. These difficulties are outlined in our earlier companion paper [LevTad98, §2.3.2], and are demonstrated in the examples revisited in §5. The RK stability question for such L_N's, therefore, cannot be addressed by the familiar scalar arguments, based on the requirement that the eigenvalues of L_N lie inside the absolute stability region of the corresponding scalar RK method. In this paper, we replace this potentially misleading spectral analysis by the energy method, where stability is examined by taking into account the full eigenstructure of L_N.

The paper is organized as follows. In §2 we discuss different notions of stability for the semidiscrete method of lines (2). Specifically, we indicate the equivalence, upon proper re-norming, between the notions of stability, (3), and strong stability, (5). We assume that the latter holds by requiring that L_N is *semibounded* i.e., that there exists a symmterizer – a positive definite matrix $H = H_N > 0$ such that *for all* u_N's,

$$\langle L_N u_N, H_N u_N\rangle + \langle u_N, H_N L_N u_N\rangle \leq 2\gamma\langle u_N, H_N u_N\rangle.$$

The existence of such a symmetrizer is usually a part of the stability question, and in many physically relevant problems, it is induced by a well-posed underlying differential problem. The central issue in this paper is whether the strong stability of such semidiscrete systems is preserved by its RK time discretizations. That is, whether the H-weighted semiboundedness of L_N implies that for sufficiently small time-step Δt, the fully-discrete RK method remains (strongly) stable. The time-step restriction in this context amounts to the celebrated Courant-Friedrichs-Levy (CFL) condition, [RicMor67],[GusKreOli].

The stability question of the fully discrete RK schemes is treated in §3 and §4. In §3, we treat the important subclass of *coercive* L_N's, for which there exists a fixed $\eta > 0$ such that *for all* u_N's,

$$\langle L_N u_N, H_N u_N\rangle + \langle u_N, H_N L_N u_N\rangle \leq -\eta\langle L_N u_N, H_N L_N u_N\rangle.$$

The relevance of coercivity in this context was first pointed out in [LevTad98]. It guarantees the strong stability of the first-order accurate forward Euler time dis-

[1] $L(x, \partial_x)$ is local in the sense that the support of $Lu(\cdot)$ is contained in $\{\text{supp } u(\cdot)\}$, so that in particular, the value of $L(x, \partial_x)u(x,t)$ is dictated by the infinitesimal neighborhood (x,t).

cretization, $u_N(t+\Delta t) = u_N(t) + \Delta t L_N u_N(t)$, under the CFL time-step restriction $\Delta t \leq \eta$. In §3 we recall the systematic study in [GotShuTad99] where we convert this first-order stability result to higher, third- and forth-order RK discretizations of coercive (and possibly nonlinear) semidiscrete systems. The main coercivity result of this section is summarized in Proposition 1. In §4, we return to general semibounded problems, beyond the subclass of coercive L_N's. Theorem 2 summarizes the strong stability result of [Lev98] for the third-order RK method, under a (possibly smaller) CFL restriction, $\Delta t \leq 1/\|L_N\|$. The corresponding stability question for fourth-order RK discretizations of general semibounded problems remains open. We should note in passing a different approach to this question of preserving stability by RK methods, [KreWu93], which is based on the closely related (yet weaker) notion of *resolvent stability* outlined in the Appendix.

Equipped with these two stability criteria, Proposition 1 and Theorem 2, we conclude in §5, with a series of four examples. In §5.1, we consider one-sided differencing of scalar advection equation as a favorite prototype model for non-normal systems, whose stability study based on naive spectral analysis could be misleading. The so-called energy method approach outlined in sections 3 and 4 yields a sharp CFL time-step restriction by taking into account the full eigenstructure of the semidiscrete L_N's in question. In §5.2, we revisit the example of Cauchy problems governed by linear advection-diffusion systems, with spatial discretization using general finite difference centered stencils, [LevTad98, §4]. It is here that Theorem 2 offers an advantage over the coercivity argument summarized in Proposition 1. We prove stability under a CFL condition which is valid *uniformly* with respect to (w.r.t) the amount of diffusion and in particular, we recover a sharp CFL stability condition for the limiting case of pure advection. In §5.3, we turn to consider advection equations based on Fourier differencing. Fourier differencing lacks the local character of finite difference stencils, which in turns implies that spatial discretization based on Fourier differencing is *not* semibounded, [Tad87]. Instead, the open stability question of the corresponding semidiscrete Fourier method was answered in [GooHouTad94] using an intricate construction of a symmetrizer H_N. We use Theorem 2 to extend this H-weighted stability result to the fully-discrete third-order RK method. In §5.4, we discuss the stability question for pseudospectral Jacobi-based methods for mixed initial-boundary advection problems. Again, the derivation of a CFL stability condition depends on a nontrivial construction of an appropriate symmetrizer, carried out in [GotTad91]. We revisit the coercivity stability argument [LevTad98, §4.2], which is compared with the strong stability argument of Theorem 2.

2 Weighted L^2 stability and semiboundedness

We assume that the semidiscrete method (2) is *stable* in the sense that there exists a constant K (independent of t) such that for arbitrary initial data $u_N(0)$, the corresponding semidiscrete solution $u_N(t)$ has bounded growth relative to its initial size, i.e.,

$$|u_N(t)| \leq K|u_N(0)|, \qquad \forall u_N(0). \tag{3}$$

Stability can be equivalently expressed in terms of the solution operator, $u_N(t) = e^{L_N t}u_N(0)$, as

$$\|e^{L_N t}\| := \sup_{u_N(0)} \frac{|e^{L_N t}u_N(0)|}{|u_N(0)|} \leq K. \tag{4}$$

We note that in general situations, the discrete solution might grow exponentially in time, $u_N(t) \sim Ke^{\gamma t}u_N(0)$, and that in this case (3) applies to the renormalized solution $\tilde{u}_N(t) = e^{-\gamma t}u_N(t)$, i.e., replacing L_N with $L_N - \gamma I_N$. Such exponential growth should be admitted in order to entertain the presence of low-order terms. For example, if L_N is stable then it can be shown that $e^{(L_N+B_N)t}$ is upperbounded by $K\exp(\gamma t)$ with $\gamma = K\|B_N\|$. Thus, by allowing exponential growth in time, the notion of stability remains invariant under the addition of bounded perturbations, which is an essential property of any useful stability definition. This point was emphasized early on by the instructive counterexamples of H.-O. Kreiss, [RicMor67, §5.2]. The essence of the stability definition (3), therefore, is not the growth in time but the uniform bound w.r.t. N — the discrete solution should remain bounded as we refine the small spatial scale.

But this notion of stability is not sharp since $e^{L_N t} \sim I_N + \mathcal{O}(\|L_N\|t)$, and we therefore should expect a stability constant $K \sim 1$, at least for $t \sim 0$. This brings us to the notion of *strong stability*, requiring (3) to hold with $K = 1$,

$$|\!|\!| u_N(t) |\!|\!| \leq |\!|\!| u_N(0) |\!|\!|, \qquad \forall u_N(0), \tag{5}$$

where $|\!|\!| \cdot |\!|\!|$ is a possibly new norm. So far, we have not specified a particular norm to be used in connection with these notions of stability. Indeed, the choice of the specific vector norms, $|\cdot|$ or $|\!|\!| \cdot |\!|\!|$ is an essential part of the stability problem itself.

It follows that a semidiscrete system which is stable w.r.t. a given norm $|\cdot|$, is necessarily strongly stable w.r.t. another norm — the norm given by

$$|\!|\!| w |\!|\!| := \sup_{s>0} \frac{|e^{L_N s}w|}{|w|}.$$

Indeed, if our method is stable so that (3) holds, then the sup on the right defines a proper new vector norm $|\!|\!| \cdot |\!|\!|$ such that $|\!|\!| e^{L_N t}u_N(0) |\!|\!| \leq |\!|\!| u_N(0) |\!|\!|$; i.e., by *re-norming*, our semidiscrete method becomes strongly stable. This is a constructive procedure to convert a stable method into a strongly stable method by identifying an appropriate norm that induces the sharper strong stability estimate (5). Unfortunately, this procedure is too difficult to work with, since it lacks any geometrical structure, and instead we seek strong stability that is induced by the following property of semiboundedness.

The system (2) is *semibounded* if there exists a symmetric, positive-definite $H = H_N$ such that[2]

$$L_N^\top H_N + H_N L_N \leq 0. \tag{6}$$

[2]Here and below, $\{\cdot\}^\top$ denote the transpose relative to the Euclidian inner product $\langle\cdot,\cdot\rangle$, and we employ the usual order between hermitian matrices, namely, $H \leq J$ iff $\langle w, Hw\rangle \leq \langle w, Jw\rangle$ for all w's.

The symmetrizer H_N is assumed to be uniformly bounded, i.e., there exists a constant $c > 0$, independent of N such that

$$0 < \frac{1}{c} \leq H_N \leq c \tag{7}$$

The symmetrizer $H = H_N$ induces the H-weighted norm $|w|_H^2 := \langle w, Hw \rangle$. It follows that the semibounded method (6)-(7) is strongly stable with respect to the weighted norm $|\cdot|_H$,

$$\frac{d}{dt}|u_N|_H^2 = \langle \frac{d}{dt}u_N, u_N\rangle_H + \langle u_N, \frac{d}{dt}u_N\rangle_H = \langle L_N u_N, H_N u_N\rangle + \langle u_N, H_N L_N u_N\rangle =$$

$$= \langle u_N, \left[L_N^\top H_N + H_N L_N\right] u_N\rangle \leq 0.$$

Moreover, the uniform bound (7) implies the weighted, possibly N-dependent norm, $|w|_{H_N}$, is in fact equivalent to the usual L^2 Euclidean norm,via

$$0 < \frac{1}{c}|w|_{L^2} \leq |w|_{H_N} \leq c|w|_{L^2},$$

which in turn yields the L^2 stability estimate (3), $|u_N(t)|_{L^2} \leq K|u_N(0)|_{L^2}$, with $K = c^2$.

It turns out that most stable systems encountered in applications — including many physically-relevant problems governed by hyperbolic and parabolic systems of PDEs, are in fact semibounded w.r.t an appropriate H-weighted norm[3] The existence of a symmetrizer $H = H_N$ in the context of semidiscrete approximations is usually inferred from the well-posedness of the underlying PDE, and granted such (uniform) semiboundedness, (6)-(7), then strong stability follows. Moreover, the inverse implication, namely, strong stability $\Longrightarrow$ semiboundedness, holds for the important class of *families* of finite-dimensional matrices, $\{M\}_{M\in\mathcal{F}}$,

$$L_N = \sum_{M\in\mathcal{F}} \oplus M, \quad \dim(M) \leq Const. \tag{8}$$

This is the content of the H-condition in the celebrated Kreiss Matrix Theorem [RicMor67, §4.9].

It is therefore this notion of stability, namely, H-weighted semiboundedness, that will be used throughout this paper as the starting point for our study of stability preservation by the fully discrete Runge-Kutta methods.

We close by noting that weighted H-stability amounts to L_N being negative-definite with respect to the H-weighted product $\langle\cdot, H\cdot\rangle$, namely, $\Re e_H L_N := L_N^\top H_N + H_N L_N \leq 0$, and in particular, the spectrum of L_N lies in the left side of the plane, $\Re e\lambda(L_N) \leq 0$. In case of time growth, $|u_N(t)|_H \leq e^{\gamma t}|u_N(0)|_H$, then L_N is replaced by $L_N - \gamma I_N$, and the semiboundedness requirement (6) then reads

$$2\Re e_H L_N := L_N^\top H_N + L_N H_N \leq 2\gamma H_N, \tag{9}$$

with the corresponding spectrum, $\Re e\lambda(L_N) \leq \gamma$.

[3]In many such systems, the existence of a symmetrizer is intimately linked to the existence of an entropy function associated with the underlying nonlinear system.

3 Strong stability of coercive methods

To discretize (1) in time, we introduce a time-step Δt. Runge-Kutta (RK) approximations of the semidiscrete method (2) are based on the polynomial expansion

$$e^{L_N t} \sim \left[I + \Delta t L_N + \frac{1}{2}(\Delta t L_N)^2 + \frac{1}{6}(\Delta t L_N)^3 + \ldots\right]^{\frac{t}{\Delta t}}.$$

The corresponding k-stage fully-discrete RK method then reads

$$u_N(t^n + \Delta t) = \left[p_0 + p_1 \Delta t L_N + \frac{p_2}{2}(\Delta t L_N)^2 + \ldots + \frac{p_k}{k!}(\Delta t L_N)^k\right] u_N(t^n), \quad (10)$$

and it is s-order accurate if its first $s+1$ coefficients, $p_0 = p_1 = \ldots = p_s = 1$. We inquire whether the semidiscrete problem (1) carries over to its fully discrete RK approximation (10). The answer is negative already for the first-order RK method, the so-called forward Euler method,

$$u(t^{n+1}) = (I + \Delta t L_N) u(t^n). \quad (11)$$

The stability of (11) fails for arbitrary semibounded L's as shown, for example, by considering the purely imaginary spectrum associated with skew-symmetric differencing, consult (35) or (36) below.

To guarantee the strong stability of (11), measured by an appropriate H-weighted norm, it is necessary and sufficient for L_N to be (uniformly) *coercive*, in the sense that there exists a constant $\eta > 0$, independent of N, such that

$$L_N^\top H_N + H_N L_N \leq -\eta L_N^\top H L_N. \quad (12)$$

The geometric interpretation of the coercivity condition (12) requires in a generic case — say, for normal L_N's, that the eigenvalues of such L_N are contained in the left-plane circle, $|\lambda + \frac{1}{\eta}| \leq \frac{1}{\eta}$.

Restricting attention to coercive L_N's, the first-order forward Euler method (11) is strongly stable, $\|I + \Delta t L_N\|_H \leq 1$, for sufficiently small time-step $\Delta t \leq \eta$. The issue then is preserving this strong stability for higher order RK methods. The main result of [LevTad98] proves strong stability of third- and fourth-order RK discretizations of linear coercive problems. A systematic study of preserving stability from the first order Euler to higher order RK discretizations of linear (as well as *nonlinear*) problems, is presented in [GotShuTad99] which is also surveyed in another paper of this volume, [Shu02].

We conclude this section with two prototype examples of strong stability for the third- and fourth-order RK methods of coercive problems, [GotShuTad99, §3]. For the third-order three-stage case (k=s=3 in (10)), we write

$$\begin{aligned} P_3(\Delta t L_N) &:= I + \Delta t L_N + \frac{1}{2}(\Delta t L_N)^2 + \frac{1}{6}(\Delta t L_N)^3 \quad (13) \\ &\equiv \frac{1}{3} + \frac{1}{2}(I + \Delta t L_N) + \frac{1}{6}(I + \Delta t L_N)^3 \end{aligned}$$

Coercivity implies the strong stability of the first-order Euler method, $\|I + \Delta t L_N\| \le 1$, and we conclude that under the same time-step restriction, $\Delta t \le \eta$, this stability is preserved in the third-order case, $\|P_3(\Delta t L_N)\|_H \le \frac{1}{3} + \frac{1}{2} + \frac{1}{6} = 1$.

A similar argument applies to the fourth-order four-stage case (k=s=4), where we use the factorization,

$$P_4(\Delta t L_N) := I + \Delta t L_N + \frac{1}{2}(\Delta t L_N)^2 + \frac{1}{6}(\Delta t L_N)^3 + \frac{1}{24}(\Delta t L_N)^4 \qquad (14)$$

$$\equiv \frac{3}{8} + \frac{1}{3}(I + \Delta t L_N) + \frac{1}{4}(I + \Delta t L_N)^2 + \frac{1}{24}(I + \Delta t L_N)^4$$

Coercivity implies $\|I + \Delta t L_N\|_H \le 1$ for sufficiently small time-step, $\Delta t \le \eta$, and we conclude that stability is preserved in the forth-order case, $\|P_4(\Delta t L_N)\|_H \le \frac{3}{8} + \frac{1}{3} + \frac{1}{4} + \frac{1}{24} = 1$. We summarize by stating

Proposition 1. *Consider the general coercive methods of lines (2),(12), with coercivity constant η. Their third- and fourth-order RK time discretizations are strongly stable, $|u_N(t)|_H \le |u_N(0)|_H$, for sufficiently small time-step, $\Delta t \le \eta$.*

The strong stability preserving s-stage, s-order accurate RK methods for general *nonlinear* problems, up to s=8, are listed in [GotShuTad99, Table3.1]. The general s-order linear case was communicated to us by H. Liu.

4 Beyond coercivity – strong stability of RK methods

It is well known that high-order RK methods are more 'faithful' approximations of the ODE system (2) than the first-order forward Euler method, and in this context we seek to remove the restriction of coercivity, (12), that is tied to the first-order RK method. Thus, we seek strong stability for higher-order RK approximations of semidiscrete problems governed by general semibounded L_N's. The next result, communicated in [Lev98], shows that (weighted) strong stability is preserved for the third-order RK scheme, thus removing the previous restriction to the subclass of coercive problems.

Theorem 2. *[Levermore] Consider the semidiscrete H-stable ODE system (2),(6). Its third order accurate RK approximation,*

$$u_N(t^{n+1}) = P_3(\Delta t L_N)u_N(t^n), \quad P_3(\Delta t L_N) := I + \Delta t L_N + \frac{1}{2}(\Delta t L_N)^2 + \frac{1}{6}(\Delta t L_N)^3,$$

is strongly stable, $|u_N(t^n)|_H \le |u_N(0)|_H$, under CFL time-step restriction

$$\Delta t \|L_N\|_H \le 1. \qquad (15)$$

Proof. We seek strong stability under the H-weighted norm

$$\|P_3(\Delta t)\|_H^2 = \sup_{u \ne 0} \frac{\langle P_3^\top(\Delta t L_N) H P_3(\Delta t L_N)u, u\rangle}{\langle u, Hu\rangle} \le 1, \qquad (16)$$

yielding $|u_N(t^{n+1})|_H \leq |u_N(t^n)|_H \leq \cdots |u_N(0)|_H$, which in view of (7) implies the L^2-stability statement, $|u_N(t^n)|_{L^2} \leq c^2 |u_N(0)|_{L^2}$.

We begin by setting $P_3 \equiv P_3(\Delta t L_N) = I + \mathcal{L}R(\mathcal{L})$ with $R = R(\mathcal{L}) := I + \mathcal{L}/2 + \mathcal{L}^2/6$ expressed in terms of $\mathcal{L} = \mathcal{L}_N := \Delta t L_N$. We compute

$$P_3^\top H P_3 - H = (I + R^\top \mathcal{L}^\top) H (I + \mathcal{L}R) - H = R^\top \mathcal{L}^\top H \mathcal{L} R + R^\top \mathcal{L}^\top H + H \mathcal{L} R.$$

Inserting $I = R - \mathcal{L}/2 - \mathcal{L}^2/6$ into the right and left of the last two terms, we find

$$\begin{aligned} P_3^\top H P_3 - H &= (\frac{1}{2} + \frac{1}{2}) R^\top \mathcal{L}^\top H \mathcal{L} R \\ &\quad + R^\top \mathcal{L}^\top H (R - \frac{1}{2}\mathcal{L} - \frac{1}{6}\mathcal{L}^2) + (R - \frac{1}{2}\mathcal{L} - \frac{1}{6}\mathcal{L}^2)^\top H \mathcal{L} R \\ &= R^\top (\mathcal{L}^\top H + H\mathcal{L}) R \\ &\quad + R^\top \mathcal{L}^\top H \left(\frac{1}{2}\mathcal{L}R - \frac{1}{2}\mathcal{L} - \frac{1}{6}\mathcal{L}^2 \right) + \left(\frac{1}{2}\mathcal{L}R - \frac{1}{2}\mathcal{L} - \frac{1}{6}\mathcal{L}^2 \right)^\top H \mathcal{L} R. \end{aligned}$$

The first term on the right is H-negative, as is $\mathcal{L}$ by assumption, and hence

$$P_3 H P_3 - H \leq R^\top \mathcal{L}^\top H \left(\frac{1}{2}\mathcal{L}R - \frac{1}{2}\mathcal{L} - \frac{1}{6}\mathcal{L}^2 \right) + \left(\frac{1}{2}\mathcal{L}R - \frac{1}{2}\mathcal{L} - \frac{1}{6}\mathcal{L}^2 \right)^\top H \mathcal{L} R \quad (17)$$

Next, we treat the two expressions on the right of (17). First, note that

$$\frac{1}{2}\mathcal{L}R - \frac{1}{2}\mathcal{L} - \frac{1}{6}\mathcal{L}^2 = \frac{1}{12}(\mathcal{L}^2 + \mathcal{L}^3) = \frac{1}{12}\mathcal{L}^2 (I + \mathcal{L}).$$

Second, we decompose the term $\mathcal{L}R$ into

$$\mathcal{L}R = \mathcal{L} + \mathcal{L}^2 - \frac{1}{2}(I - \mathcal{L}/3)\mathcal{L}^2.$$

With this, the RHS of (17) amounts to

$$\begin{aligned} P_3^\top H P_3 - H &\leq \frac{1}{12}(\mathcal{L} + \mathcal{L}^2)^\top H (\mathcal{L}^2 + \mathcal{L}^3) + \frac{1}{12}(\mathcal{L}^2 + \mathcal{L}^3)^\top H (\mathcal{L} + \mathcal{L}^2) \\ &\quad - \frac{1}{24}(I - \mathcal{L}/3)^\top (\mathcal{L}^2)^\top H \mathcal{L}^2 (I + \mathcal{L}) - \frac{1}{24}(I + \mathcal{L})^\top (\mathcal{L}^2)^\top H \mathcal{L}^2 (I - \mathcal{L}/3). \end{aligned}$$

Again, the sum of the first two terms is H-negative, as is $\mathcal{L}$ by assumption, and

$$\begin{aligned} &\frac{1}{12}(\mathcal{L} + \mathcal{L}^2)^\top H (\mathcal{L}^2 + \mathcal{L}^3) + \frac{1}{12}(\mathcal{L}^2 + \mathcal{L}^3)^\top H (\mathcal{L} + \mathcal{L}^2) \\ &\qquad = \frac{1}{12}(\mathcal{L} + \mathcal{L}^2)^\top (\mathcal{L}^\top H + H\mathcal{L})(\mathcal{L} + \mathcal{L}^2) \leq 0. \end{aligned}$$

Hence, we are left with

$$P_3^\top H P_3 - H \leq -\frac{1}{24}\Re e (I - \mathcal{L}/3)^\top \mathcal{H} (I + \mathcal{L}), \quad (18)$$

where $\mathcal{H} := (\mathcal{L}^2)^\top H\mathcal{L}^2 > 0$. One checks that the last upperbound is nonpositive,

$$-\Re e\langle (I-\mathcal{L}/3)u,\ (I+\mathcal{L})u\rangle_{\mathcal{H}} = -|u|^2_{\mathcal{H}} - \frac{2}{3}\Re e\langle \mathcal{L}u, u\rangle_{\mathcal{H}} + \frac{1}{3}|\mathcal{L}u|^2_{\mathcal{H}}$$
$$\equiv \frac{1}{3}|(I-\mathcal{L})u|^2_{\mathcal{H}} - \frac{4}{3}|u|^2_{\mathcal{H}} \le 0,$$

provided

$$\|I-\mathcal{L}\|_{\mathcal{H}} \le 2. \tag{19}$$

This leads to the sufficient CFL stability condition, $\|\mathcal{L}\|_{\mathcal{H}} \le 1$, which in turn is guaranteed by the CFL restriction (15)

$$\|\mathcal{L}\|_{\mathcal{H}} = \sup_{u\neq 0}\frac{|\mathcal{L}u|_{\mathcal{H}}}{|u|_{\mathcal{H}}} = \sup_{u\neq 0}\frac{|\mathcal{L}^3u|_H}{|\mathcal{L}^2u|_H} \le \Delta t\cdot\|L_N\|_H < 1.$$

□

Remarks.

1. *On the time-step restriction.* I. How does the stability restriction stated in Theorem 2 compare with the previous stability criterion for coercive problems? According to Proposition 1, the stability of the third-order RK method is guaranteed for coercive L_N's under the time-step restriction

$$\Delta t \le \eta := 2\inf_{u\neq 0}\frac{|\Re e\langle L_Nu,u\rangle_H|}{|L_Nu|^2_H} \tag{20}$$

An upper bound of the expression on right-hand-side leads to *twice* the time-step restriction (15) of Theorem 2,

$$\Delta t \le 2\inf_{u\neq 0}\frac{|\Re e\langle L_Nu,u\rangle_H|}{|L_Nu|^2_H} \le$$
$$\le 2\inf_{u\neq 0}\frac{|L_Nu|_H|u|_H}{|L_Nu|^2_H} = 2\frac{1}{\sup_{u\neq 0}(|L_Nu|_H/|u|_H)} = 2\frac{1}{\|L_N\|}.$$

This shows that the strong stability statement of Theorem 2 may require a more restrictive – up to half the time-step restriction based on the coercivity argued in Proposition 1. In the generic case of normal L_N's, the last theorem requires the eigenvalues of such L_N to lie in the strip $\Delta t\cdot\Re e\lambda \in [-1,0)$, which is only half the width of the circle $|\Delta t\cdot\lambda+1| \le 1$ we met earlier, in the context of coercive problems. Nevertheless, the gain of Theorem 2 is due to its applicability to a larger class of semibounded problems (in agreement with the fact of having an *infinite* strip of strong stability along the imaginary x-axis). The examples in §5 will demonstrate these points.

2. *On the time-step restriction.* II. Theorem 2 raises the question whether its time-step restriction (15) is sharp. An optimal time-step restriction for the third order RK approximation of general normal operators can be derived by the usual

scalar stability argument, requiring that the eigenvalues of $P_3(\Delta t L)$ be contained in the region of absolute stability, so that $\|P_3(\Delta t \lambda(L_N))\| \leq 1$. For skew-symmetric L_N's, for example, this yields the time-step restriction $\Delta t \leq \sqrt{3}/\|L_N\|$. Indeed, Theorem 2 is sharp enough to include these optimal cases. For normal L_N's, for example, one arrives at the stability condition (19) which now reads

$$\|I - \mathcal{L}\|_{\mathcal{H}} = \max_{\lambda} |1 - \Delta t \lambda(L_N)| \leq 2,$$

yielding

$$\Delta t \times \max_{\lambda = \lambda(L_N)} \sqrt{|\lambda|^2 - 2\Re e\lambda} \leq \sqrt{3}, \quad L_N \text{ normal.} \tag{21}$$

In particular, for skew-symmetric L_N's we have, $\Re e\lambda(L_N) = 0$, $\max \lambda(L_N) = \|L_N\|$, and (21) recovers the sharp CFL condition, $\Delta t \leq \sqrt{3}/\|L_N\|$. We conjecture that (15) is optimal for general semibounded, non-diagonalizable systems. We note in passing that the even additional factor of $\sqrt{3}$ does not fully compensate for the twice larger time-step restriction in the coercive case.

3. *On preserving (strong) stability by the fourth-order RK method.* The absolute stability regions of the generic first- and second-order RK methods do not contain any interval along the imaginary x-axis, and hence they cannot preserve stability of arbitrary semibounded problems. Their stability fails, for example, for skew-symmetric L_N's with increasing spectrum along the x-axis. As argued in [LevTad98], one therefore needs to restrict attention to the subclass of coercive L_N's for preserving the stability of these first- and second-order RK methods. Theorem 2 addresses the issue of preserving strong stability for general semibounded L_N's, starting with the third-order three-stage RK method, and the same question is being raised for the ubiquitous fourth-order four-stage RK method. Namely, we ask whether for sufficiently small time-step, $\Delta t \|L_N\|_H \leq \Delta t_0$, there holds

$$\|P_4(\Delta t L_N)\|_H \leq 1, \quad P_4(\Delta t L_N) := \sum_{j=0}^{4} \frac{1}{j!} (\Delta t L_N)^j. \tag{22}$$

Our preliminary studies indicate the following *saturation* result. Namely, that preserving semidiscrete stability by the fourth-order RK methods requires more than the four stages encoded in $P_4(\Delta t L_N)$. Instead, we conjecture that one can preserve the semidiscrete H-weighted stability (6), by allowing additional stages with higher order terms, so that (22) is replaced by

$$\| \sum_{j=0}^{4} \frac{1}{j!} (\Delta t L_N)^j + p_5(\Delta t L_N)^5 + \ldots \|_H \leq 1. \tag{23}$$

4. *On the generalized resolvent stability of the fully discrete RK methods.* We conclude this section by recalling the closely related stability result of Kreiss & Wu, [KreWu93]. They consider class of locally stable[4] RK methods, including the generic

[4]Local stability of an RK method is understood in the sense that its region of absolute stability contains a semi-circle centered at the origin (and in particular, therefore, it contains an interval along the imaginary x-axis), consult [KreSch92].

k-stage s-order methods, k=s=$3,4$. The stability result [KreWu93, Theorem 3.2] states that such RK methods preserve the generalized stability of semidiscrete problems. Here, *generalized* stability is interpreted in the sense of appropriate resolvent estimates that replace the closely related notions of semidiscrete semiboundedness (6), and fully-discrete power-boundedness. Details are provided in the Appendix.

5 Examples

5.1 One sided differencing of advection equations

Our first example deals with the advection equation

$$u_t(x,t) = a u_x(x,t), \quad -1 \le x \le 1, \quad a = Const > 0 \tag{24}$$

augmented with zero boundary condition $u(x,t)_{|x=1} = 0$. We introduce a spatial equidistant gridpoints, $x_j = -1+j\Delta x$, $\Delta x := 2/N$, and we use one-sided differences for spatial differencing to obtain (here and below we let $u(\cdot,t)$ denote both the exact solution and its semidiscrete approximation, which can be distinguished by the context)

$$\frac{d}{dt}u(x_j,t) = a\frac{u(x_{j+1},t)-u(x_j,t)}{\Delta x}, \quad j = 0,1,\ldots,N-1 \tag{25}$$

augmented with zero boundary conditions $u(x_N,t) = 0$. Thus, with $u_N(t) := (u(x_0,t), u(x_1,t),\ldots,u(x_{N-1},t))^\top$ the method of lines (25) amounts to the $N \times N$ semidiscrete system

$$\frac{d}{dt}u_N = L_N u_N, \quad L_N = \frac{a}{\Delta x}\begin{bmatrix} -1 & 1 & & & & \\ & -1 & 1 & & & \\ & & & \ddots & & \\ & & & \ddots & 1 & \\ & & & & -1 \end{bmatrix}, \tag{26}$$

and we turn to consider the stability of the third-order RK time discretization

$$u_N(t^{n+1}) = \Big[I + \Delta t L_N + \frac{(\Delta t L_N)^2}{2} + \frac{(\Delta t L_N)^3}{6}\Big]u_N(t^n). \tag{27}$$

System (26) serves as a favorite prototype example demonstrating that a scalar stability argument based on naive eigenvalues analysis of non-normal matrices can be misleading; among a host of references we mention one of the firsts due to Godunov and Ryabenkii, [GodRya63], and more recent ones in [GusKreOli], [Tre96], and the detailed discussion in [Ise96, §14.1],... Indeed, verifying that the eigenvalues of $\Delta t L_N$, lie in the absolute stability region of the (scalar) third-order RK method, $|P_3(-a\frac{\Delta t}{\Delta x})| \le 1$, leads to the wrong CFL condition $\frac{\Delta t}{\Delta x}|a| \le 2.5$. This scalar argument fails to guarantee stability since it does not capture the power-growth of the increasingly larger Jordan blocks of the type encountered in (26), consult

[LevTad98, §2.3.2]. Instead, stability of this initial-boundary value problem can be verified by the normal mode analysis of [GusKreSun72], and to this end we proceed as follows. Expressed in terms of the one-sided divided difference operator, $D_+w := (w(\cdot+\Delta x)-w(\cdot))/\Delta x$, the RK method (27) consists of an interior difference scheme

$$u(x_j,t^{n+1}) = \Big[I+a\Delta t D_+ + \frac{(a\Delta t)^2}{2}D_+^2 + \frac{(a\Delta t)^3}{6}D_+^3\Big]u(x_j,t^n), \quad j=0,1,\ldots,N-3, \tag{28}$$

augmented by the prescribed boundary data,

$$u(x_j,t^{n+1}) = 0, \quad j = N-2, N-1, N. \tag{29}$$

A necessary von Neumann stability condition for the interior scheme (28) requires

$$\sup_{\xi} |P_3\Big(a\frac{\Delta t}{\Delta x}(e^{i\xi\Delta x}-1)\Big)| \le 1, \tag{30}$$

which leads to the CFL time-step limitation $\frac{\Delta t}{\Delta x}|a| < 1.25$. The stability of this interior scheme combined with the *translatory* boundary conditions (29) follows from the general stability results for translatory boundary conditions of [GolTad81], based on the normal mode analysis of [GusKreSun72].

We now turn to examine the same stability question of (27) using the two general approaches outlined in Proposition 1 and Theorem 2. We start by noting that L_N is negative if and only if $a > 0$,

$$2\Re e L_N := L_N^\top + L_N = \frac{a}{\Delta x}\begin{bmatrix} -2 & 1 & & & \\ 1 & -2 & 1 & & \\ & & \ddots & & \\ & & \ddots & 1 \\ & & & 1 & -2 \end{bmatrix} \le 0, \quad a > 0.$$

To check the coercivity of L_N we compute

$$L_N^\top L_N = \frac{a^2}{(\Delta x)^2}\begin{bmatrix} 2 & 1 & & & & \\ -1 & 2 & -1 & & & \\ & -1 & 2 & -1 & & \\ & & & \ddots & & \\ & & & \ddots & & \\ & & & & 2 & -1 \\ & & & & -1 & 2 \end{bmatrix}.$$

It follows that the coercivity condition (12) with $H = I$ holds so that $L_N^\top + L_N \le -\eta L_N^\top L_N$, provided $\eta a^2/(\Delta x)^2 \le a/\Delta x$, i.e., with coercivity constant $\eta = \Delta x/a$. We conclude

Corollary 3. *Consider the advection equation (24) which is integrated by one-sided spatial differencing (25) and the third-order RK time discretization (27). The resulting fully-discrete scheme is strongly stable,* $|u_N(t)| \leq |u_N(0)|$*, under the CFL condition*

$$\frac{\Delta t}{\Delta x}|a| \leq 1. \tag{31}$$

Remarks.

1. The CFL stability condition offered in the last corollary is more restrictive than the von Neumann condition of $\frac{\Delta t}{\Delta x}|a| \leq 1.25$, associated with the normal mode stability analysis of (27). We note, however, that normal mode analysis leads to stability in the sense of satisfying the resolvent stability estimate (65), which is somewhat weaker than the strong stability asserted in Corollary 3.

2. As indicated earlier, the coercivity argument allows for a CFL stability condition (31) that is twice as large as the one stated in Theorem 2, $\Delta t \leq 1/\|L_N\| = \Delta x/2a$. The advantage of the latter, however, in treating the larger class of arbitrary semibounded problems is demonstrated in our next example.

5.2 Local centered differencing of advection-diffusion equations

We consider the one-dimensional system of advection-diffusion equation with variable coefficients,

$$u_t = A(x,t)u_x + (Q(x,t)u_x)_x, \quad -1 \leq x \leq 1, \tag{32}$$

subject to given initial conditions, $u(x,0) = u_0(x)$ and periodic boundary conditions, $u(-1,t) = u(1,t)$. Here, the advective part is driven by a symmetric $A(\cdot,t) \in C^1$ and diffusion is governed by positive definite viscosity, $0 < q_0 \leq Q(x,t) \leq Q_0$, and we note the possibility of time dependent coefficients in this case.

Discretization in space employs centered divided differences, expressed in terms of the translation operator $Tw := w(\cdot + \Delta x)$,

$$D_+ = \frac{1}{\Delta x}\sum_{k\geq 0}\alpha_k T^k, \quad D_- := -D_+^\top = \frac{-1}{\Delta x}\sum_{k\geq 0}\alpha_k T^{-k}, \quad D_0 := \frac{1}{2}(D_+ + D_-). \tag{33}$$

The difference operators are assumed to be *local* in the sense of having a bounded stencil, consult [Tad87, §2], $|\alpha|_1 := \sum_{k>0} k|\alpha_k| \leq Const.$, so that in particular,

$$|D_\pm u| \leq |\alpha|\frac{|u|}{\Delta x}, \qquad |\alpha| := \sum_{k>0}|\alpha_k| \leq \sum_{k>0}k|\alpha_k| \leq Const. \tag{34}$$

There is a variety of such local spatial discretizations and we mention below three prototype examples. A second order centered finite-difference stencil, corresponding to $\alpha_{\pm 1} = \pm 1$ and augmented with periodic boundary conditions, is represented by the $N \times N$ circulant differentiation matrix of the form, $D_2 = \{D_{jk} = \frac{1}{\Delta x}\alpha_{(j-k)[modN]}\}$,

$$D_2 = \frac{1}{2\Delta x}\begin{bmatrix} 0 & 1 & \cdots & 0 & -1 \\ -1 & 0 & & & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & & & 0 & 1 \\ 1 & 0 & \cdots & -1 & 0 \end{bmatrix}. \tag{35}$$

A fourth order periodic stencil, corresponding to $(\alpha_{\pm 1}, \alpha_{\pm 2}) = (\pm 4/3, \mp 1/6)$ reads

$$D_4 = \frac{1}{12\Delta x}\begin{bmatrix} 0 & 8 & -1 & \cdots & -8 & 1 \\ -8 & 0 & 8 & -1 & \ldots & -8 \\ 1 & -8 & 0 & 8 & -1 & \cdots \\ \vdots & & & \ddots & & \vdots \\ 8 & & 1 & -8 & 0 & 8 \\ -1 & 8 & \cdots & 1 & -8 & 0 \end{bmatrix}. \tag{36}$$

Local stencils need not be finite in order to satisfy the locality property (34). For example, the fourth order finite element discretization of ∂_x can be realized by the local $N \times N$ circulant matrix

$$D_{\text{fem}} = \begin{bmatrix} 4 & 1 & \cdots & 1 & 1 \\ 1 & 4 & & & 1 \\ \vdots & & \ddots & & \vdots \\ 1 & & & 4 & 1 \\ 1 & 1 & \cdots & 1 & 4 \end{bmatrix}^{-1} \cdot \frac{3}{\Delta x}\begin{bmatrix} 0 & 1 & \cdots & 0 & -1 \\ -1 & 0 & & & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & & & 0 & 1 \\ 1 & 0 & \cdots & -1 & 0 \end{bmatrix}.$$

Augmented with periodic boundary conditions, the corresponding centered semidiscrete method of lines for $u_N(t) := (u(x_1, t), \ldots, u(x_N, t))$ reads, $\partial_t u_N = L_N u_N$, where $L = L_N(t)$ is the $N \times N$ matrix associated with the centered differencing

$$(Lu(t))_j = A(x_j, t)(D_0 u)_j + D_-(Q(x_j, t)D_+ u)_j, \tag{37}$$

and we turn to study the stability of its RK time discretization.

We begin with the constant coefficient case, $A(x,t) \equiv A$. In this case, the centered differencing AD_0 is skew-symmetric,

$$(AD_0)^\top + AD_0 = 0, \tag{38}$$

and summation by parts then yields, $\langle (L^\top + L)u, u\rangle \le -2q_0 |D_+ u|^2$. Together with the straightforward upper bound

$$|L_N u| \le \|A\| \cdot |D_+ u| + Q_0 \frac{|\alpha|}{\Delta x} |D_+ u|, \tag{39}$$

it follows that the corresponding difference operator, L_N, is coercive with coercivity constant η given by,

$$L_N^\top + L_N \le -\eta L_N^\top L_N, \quad \eta = \frac{2q_0}{(\|A\| + Q_0|\alpha|/\Delta x)^2}, \tag{40}$$

consult [LevTad98, §4]. By Proposition 1, coercivity implies the strong stability of the third- and fourth-order RK time discretizations of (37), $\|P_s(\Delta t L)\| \leq 1, \ s = 3, 4$, under the time-step restriction

$$\Delta t \leq \frac{2q_0}{(\|A\| + |\alpha| Q_0/\Delta x)^2}. \tag{41}$$

We observe that the presence of numerical dissipation, $q_0 > 0$, is necessary in order for this coercivity argument to hold. This should be compared with the strong stability statement in Theorem 2. The upper bounds in (34),(39) yield $\|L_N\| \leq |\alpha|(\|A\| + Q_0|\alpha|/\Delta x)/\Delta x$, which implies strong stability under the improved time-step restriction

$$\Delta t \leq \frac{\Delta x}{|\alpha|(\|A\| + Q_0|\alpha|/\Delta x)}. \tag{42}$$

Indeed, comparing with the time-step restriction in (41) we arrive at

$$\frac{2q_0}{(\|A\| + |\alpha| Q_0/\Delta x)^2} \Big/ \frac{\Delta x}{|\alpha|(\|A\| + Q_0|\alpha|/\Delta x)} \leq \frac{2q_0}{Q_0} < 2.$$

Thus, as before, we find that coercivity offers a time-step that may be up to twice larger than the one offered by strong stability stated in Theorem 2. The main advantage of (42), however, is its *independence* of the viscosity amplitude, q_0. In the particular limiting case of pure advection, $Q \equiv 0$, and we end up with the usual advection stability restriction

$$\frac{\Delta t}{\Delta x}\|A\| \cdot |\alpha| \leq 1. \tag{43}$$

A constant coefficient von Neumann stability analysis leads to the similar time-step restriction (though with a smaller $|\alpha| = \sup_\xi |\sum \alpha_k e^{ik\xi}|$) — an enjoyable sharpness.

The above approach is based on direct energy bounds, which enable us to extend this stability result to the variable coefficients case, $A = A(x, t)$. The extension can be worked out along the lines of [Tad87]. In this case, $A(\cdot, t)D_0$ is no longer skew-symmetric. Instead, since the difference operators $D_\pm$ are assumed to be local (34), the skew-symmetry (38) is replaced by the following commutator bound, consult [Tad87, condition L, §2]

$$\left|\Big(A(\cdot,t)D_0\Big)^\top + A(\cdot,t)D_0 u\Big)_j = D_0(A(x_j,t)u) - A(x_j,t)D_0 u\right| \leq \gamma\Delta x, \quad \gamma \sim |\alpha|_1 \cdot |A(\cdot,t)|_{C^1},$$

and a straightforward perturbation argument yields $\|P_3(\Delta t L_N(t^k))\| \leq 1 + \gamma\Delta x$, with a constant γ *independent* of N and time. Stability follows by successive application of $P_3(\Delta t L_N)$ under the hyperbolic time-step restriction, $\Delta t \cong \Delta x$,

$$|u_N(t = t^n)| \leq \Pi_{k=0}^n \|P_3(\Delta t L(t^k)\| \cdot |u_N(0)| \leq (1 + \gamma\Delta x)^{t^n/\Delta t} \leq exp(\gamma t^n) \cdot |u_N(0)|.$$

In the purely diffusive case, $A \equiv 0$, we are led to the usual parabolic time step restriction $\Delta t \cong (\Delta x)^2$, which preserves strong stability. We summarize by stating

Corollary 4. *Consider the discrete approximation of the advection-diffusion system of equations (32), based on a spatial centered differencing (37) using a general local stencil (33), (34). Its third-order RK time discretization,* $u_N(t^{n+1}) = P_3(\Delta t L_N)u_N(t^n)$ *is strongly stable,* $|u_N(t)| \le exp(\gamma t)|u_N(0)|$, *under the time-step restriction*

$$\frac{\Delta t}{\Delta x}|\alpha| \le \frac{1}{\sup_x \|A(\cdot,t)\| + Q_0|\alpha|/\Delta x}.$$

We conclude by pointing out that our stability arguments apply in the current example of time-dependent coefficients since the symmetrizer in this case is the identity matrix, $H_N = I$. Similar arguments apply as long as the symmetrizer, $H_N = H_N(t)$, remains smoothly dependent on t.

5.3 Global differencing — Fourier method for advection equations

We now consider the pseudospectral Fourier method as an example for *global* spatial differencing, whose stability necessitates the use of proper symmetrizer H_N. Fourier differencing starts with a given N-vector of gridvalues $(w_1, \ldots, w_N)$ prescribed at the N equidistant gridpoints $x_j = -\pi + j\Delta x,\ j =, 1, \ldots, N,\ \Delta x = 2\pi/N$. Let $\Psi_N(x) = \mathcal{I}_N w$ denote the trigonometric interpolant that is uniquely determined[5] by the prescribed data $\Psi_N(x_j) = w_j,\ j = 1, 2, \ldots, N$. The vector of discrete derivatives, $w_j^{'}$ is then computed by exact differentiation of this trigonometric interpolant

$$w_j^{'} := \frac{d}{dx}(\mathcal{I}_N w)(x)_{|x=x_j}.$$

We are interested in the RK integration of the semidiscrete Fourier method for the 2π-periodic scalar advection equation (which, for simplicity of the presentation, is taken here in its conservative form),

$$\frac{d}{dt}u(x_j,t) = (a(x_j)u(x_j,t))^{'} \tag{44}$$

The linear transformation of Fourier differencing, $(w_1, \ldots, w_N)^\top \mapsto (w_1^{'}, \ldots, w_N^{'})^\top$ for odd N's $N = 2n+1$, is realized by *global*, skew-symmetric $N \times N$ differentiation matrix $w_N^{'} = D_F w_N$, given by

$$(D_F)_{jk} = \frac{1}{\Delta x}\alpha_{(j-k)[\mathrm{mod} N]}, \quad \alpha_j = \frac{(-1)^j \Delta x}{2\sin(\frac{j\Delta x}{2})}.$$

The differencing stencil in this case is not local, $|\alpha|_1 = \sum k|\alpha_k| \sim N$, which in turns implies that the corresponding spatial operator

$$L_N = D_F \begin{bmatrix} a(x_1) & & \\ & \ddots & \\ & & a(x_N) \end{bmatrix},$$

[5] We note that Ψ_N is in fact a 2π periodic trigonometric polynomial of degree $n = [N/2]$.

is *not* L_2-semibounded, $\|L_N^\top + L_N\| \sim N$, [Tad87]. Instead, the stability of (44) was derived in [GooHouTad94] by construction of a nontrivial symmetrizer, $H = H_N$ such that (9) holds,

$$L_N^\top H_N + H_N L_N \leq H_N. \tag{45}$$

To study the stability of the fully discrete RK method, we therefore recall this construction of H_N for the special prototype case, $a(x) = \sin x$, [GooHouTad94, §2],

$$\frac{d}{dt}u_N(t) = D_F A_N u_N(t), \quad A_N = \begin{bmatrix} \sin(x_1) & & \\ & \ddots & \\ & & \sin(x_N) \end{bmatrix}. \tag{46}$$

To this end, we decompose the DFT of $u_N := (u(x_1,t),\ldots,u(x_N,t))^\top$ into its real and imaginary parts,

$$\hat{u}_N =: \hat{\mathbf{r}}_N + i\hat{\mathbf{j}}_N, \quad \hat{u}_N = F_N u_N, \quad F_N = \{(F)_{jk}\} = \frac{1}{N}e^{-ikj\Delta x}$$

The key observation in [GooHouTad94] is that the differences, $\rho_k^- := \hat{\mathbf{r}}_k - \hat{\mathbf{r}}_{k+1}$ and the sums $\rho_k^+ := \hat{\mathbf{j}}_k + \hat{\mathbf{j}}_{k+1}$ are governed by well-behaved skew-symmetric systems

$$\frac{d}{dt}\rho^\pm(t) = \frac{1}{2}(\pm I + \mathcal{S}_N)\rho^\pm(t), \tag{47}$$

$$\mathcal{S}_N = -\mathcal{S}_N^\top := \begin{bmatrix} 0 & N-1 & 0 & \ldots \\ 1-N & 0 & \ddots & 0 \\ 0 & \ddots & \ddots & 1 \\ \vdots & 0 & -1 & 0 \end{bmatrix} \oplus \begin{bmatrix} 0 & -1 & 0 & \ldots \\ 1 & 0 & \ddots & 0 \\ 0 & \ddots & \ddots & 1-N \\ \vdots & 0 & N-1 & 0 \end{bmatrix},$$

with a unitary solution operator $U(t) := e^{\mathcal{S}_N t}$. Expressed in terms of the $n \times n$ Jordan blocks, $\mathcal{J}_\pm = \begin{bmatrix} 1 & \pm 1 & \ldots & 0 \\ 0 & 1 & \ddots & \vdots \\ \vdots & & \ddots & \pm 1 \\ 0 & \ldots & 0 & 1 \end{bmatrix}$, the solution of $(47)_+$ reads

$$\rho^+(t) = e^{t/2}U(t)\rho^+(0), \quad \rho^+ := \left[\mathcal{J}_+ \oplus 1 \oplus \mathcal{J}_+^\top\right]\hat{\mathbf{j}},$$

with a similar expression for $\rho^- := \left[\mathcal{J}_- \oplus 1 \ominus \mathcal{J}_-^\top\right]\hat{\mathbf{r}}$. Returning to the original variables, we conclude

$$|\tilde{u}_N(t)|_H \leq e^{t/2}|\tilde{u}_N|_H, \quad \tilde{u}_N := \Re e\hat{u}_N \oplus \Im m\hat{u}_N$$

with a symmetrizer $H_N = H_N^- \oplus H_N^+$ given by

$$H_N^\pm := \mathcal{J}_\pm^\top \mathcal{J}_\pm \oplus 1 \oplus \mathcal{J}_\pm \mathcal{J}_\pm^\top = \tag{48}$$

$$= \left[\begin{array}{ccccc|c|ccccc}
1 & \pm1 & & & & & & & & & \\
\pm1 & 2 & \pm1 & & & & & & & & \\
 & \ddots & \ddots & \ddots & & & & & & & \\
 & & \ddots & \ddots & \pm1 & & & & & & \\
 & & & \pm1 & 2 & & & & & & \\ \hline
 & & & & & 1 & & & & & \\ \hline
 & & & & & & 2 & \pm1 & & & \\
 & & & & & & \pm1 & 2 & \pm1 & & \\
 & & & & & & & \ddots & \ddots & \ddots & \\
 & & & & & & & & \ddots & \ddots & \pm1 \\
 & & & & & & & & & \pm1 & 1
\end{array}\right].$$

This can be interpreted in the usual language of semiboundedness (45), where the real and imaginary decomposition of $\widehat{L}_N := F_N L_N F_N^*$, satisfy

$$\mathbf{R}_N^\top H_N^- + H_N^- \mathbf{R}_N = -H_N, \quad \mathbf{J}_N^\top H_N^+ + H_N^+ \mathbf{J}_N = H_N \quad F_N L_N F_N^* =: \mathbf{R}_N + i\mathbf{J}_N. \quad (49)$$

We should point out that the symmetrizer H_N does *not* satisfy the uniform L^2 equivalence bound (7), rather, $N^{-2} \le H_N \le 4$, yielding the weaker L^2 stability estimate

$$|u_N(t)|_{L^2} \le Const.N|u_N(0)|_{L^2}.$$

The notion of such a weak algebraic stability for general finite difference methods is rejected, due to potentially unacceptable exponential growth by low order perturbations[6], consult [RicMor67, §5.2]. Nevertheless, it is a truly remarkable phenomena that no such exponential loss can occur in the present context of Fourier method, since this algebraic stability estimate improves with the increased resolution of initial data with higher Sobolev-regularity, [GooHouTad94, Corollary 3.2]

$$|u_N(t)|_{L^2} \le Const.N^{(1-s)_+}|u_N(0)|_{W^s}, \quad |u_N|_{W^s} := \Big(\sum_k (1+|k|)^{2s}|\hat{u}_k|^2\Big)^{1/2}.$$

We now turn to consider the stability of RK time discretization for the semidiscrete Fourier problem (46). The linear growth of H_N^{-1} together with the precise semibounded statement (49) imply that $\widehat{L}_N$, and hence L_N, is not coercive, and we therefore appeal to the strong stability Theorem 2. To this end, we summarize the reformulation of (46) in the dual Fourier space as

$$\frac{d}{dt}\begin{bmatrix} \hat{\mathbf{r}} \\ \hat{\mathbf{j}} \end{bmatrix} = \widetilde{L}_N \begin{bmatrix} \hat{\mathbf{r}} \\ \hat{\mathbf{j}} \end{bmatrix}, \quad \widetilde{L}_N = \begin{bmatrix} \mathbf{R}_N & -\mathbf{J}_N \\ \mathbf{J}_N & \mathbf{R}_N \end{bmatrix}.$$

A straightforward computation based on (47) then yields

$$\|\widetilde{L}_N\|_{H_N} = \sup \frac{|\rho^-(t) \oplus \rho^+(t)|_{L^2}}{|\rho^-(0) \oplus \rho^+(0)|_{L^2}} \le \frac{1+\|\mathcal{S}\|}{2},$$

[6] Exponential growth with N, not just with time!.

and with $\|\mathcal{S}_N\| \leq 2N-1$, we arrive at

Corollary 5. *Consider the fully-discrete Fourier approximation of the 2π-periodic advection equation $u_t = (\sin xu)_x$, using third-order RK time discretization,*

$$u_N(t^n + \Delta t) = P_3\Big(\Delta t D_F A_N\Big) u_N(t^n), \quad A_N = \begin{bmatrix} \sin(x_1) & & \\ & \ddots & \\ & & \sin(x_N) \end{bmatrix}. \quad (50)$$

The fully discrete scheme (50) is stable,

$$|u_N(t^n)|_{L^2} \leq Const.N^{(1-s)_+}|u_N(0)|_{W^s}, \quad |u_N|_{W^s} := \Big(\sum_k (1+|k|)^{2s}|\hat{u}_k|^2\Big)^{1/2},$$

under the CFL condition

$$\Delta t < \frac{1}{N}. \quad (51)$$

5.4 CFL condition for pseudospectral Jacobi-based methods

In this example, we analyze the stability of the high-order RK Chebyshev pseudospectral approximations of initial-boundary advection problem (24) with variable coefficients,

$$u_t = a(x)u_x, \quad 0 < a(x) < a_\infty, \quad -1 \leq x \leq 1, \quad u(x=1,t) = 0. \quad (52)$$

We focus our attention on the pseudospectral Chebyshev (semi-)discretization of (52), as a prototype for the general family of pseudospectral Jacobi-based methods discussed in [GotTad91]. To this end, we let $u_N(t) := (u(x_1,t),\ldots,u(x_N,t))$ denote the vector of computed values at the N Chebyshev collocation points $x_j := \cos(\frac{j\pi}{N+1})$, $j = N, N-1, \ldots, 1$. If $u_N(x,t)$ denotes the corresponding N-degree Chebyshev interpolant, based on these N interior points and augmented by the prescribed boundary condition $u_N(x=1,t) = 0$, we then set $u_j^{'}(t) = \partial_x u_N(x,t)|_{x=x_j}$. Being linear, this results in an $N \times N$ matrix representation, $u_N^{'} = D_T u_N$, with D_T denoting the so-called Chebyshev differentiation matrix. The corresponding semidiscrete Chebyshev method of lines for (52) then reads

$$\frac{du_N}{dt} = L_N u_N, \qquad L_N := A_N D_T, \quad A_N := \begin{bmatrix} a(x_1) & & \\ & \ddots & \\ & & a(x_N) \end{bmatrix}. \quad (53)$$

The pseudospectral Chebyshev approximation (53) is a primary example for the intricate stability issue associated with its fully discrete Runge-Kutta scheme. The Chebyshev differencing is based on a global stencil that leads to ill-conditioned differentiation matrix D_T.

To address the stability question, the computed discrete solution, $u_N(t)$ is first realized by its N-degree interpolant $u_N(x,t)$, which is governed by

$$\frac{\partial u_N}{\partial t}(x,t) - a\frac{\partial u_N}{\partial x}(x,t) = \tau(t)T'_{N+1}(x). \tag{54}$$

Here, T_k is the k-degree Chebyshev polynomial, and $\tau(t)$ is a free Lagrange multiplier which is dictated by the prescribed boundary conditions, $u_N(1,t)=0$.

The coercivity of (54) was verified in [GotTad91] using two essential inequalities that are inspired by the proper *weighted* stability of the advection equation (52). To motivate our choice of a weighted norm, we observe that (52) is well-posed w.r.t. the H-weighted norm, $\|u\|_H^2 := \int_{-1}^1 \sqrt{\frac{1+x}{1-x}}u^2(x)\frac{dx}{a(x)}$,

$$\frac{d}{dt}\|u(\cdot,t)\|_H^2 \leq -\int_{-1}^1 \frac{u^2(x,t)}{(1+x)^{1/2}(1-x)^{3/2}}dx \leq 0. \tag{55}$$

We note in passing that the initial-boundary problem (52) is not well-posed w.r.t. the usual Chebyshev weight, $\omega(x) := (1-x^2)^{-1/2}$, but (55) reveals that the problem is well-posed with the closely related $H(x) := (1+x)\omega(x)/a(x)$. Consequently, we consider the discrete norm utilizing the corresponding Chebyshev-Lobatto weights ω_j, $\omega_j = \frac{\pi}{N(1-x_j^2)}$,

$$\langle u_N, v_N\rangle_H := \sum_j (1+x_j)\omega_j u_N(x_j)v_N(x_j), \quad H_N = \begin{bmatrix} (1+x_1)\frac{\omega_1}{a(x_1)} & & \\ & \ddots & \\ & & (1+x_N)\frac{\omega_N}{a(x_N)} \end{bmatrix}$$

Equipped with these notations, we recall [GotTad91, Lemma 3.5 with $\alpha=\beta=-1/2$]

#1. $\Re e\langle u_N, L_N u_N\rangle_H \leq -\frac{1}{2}.\left\|\frac{u_N(x,t)}{1-x}\right\|^2_{(1-x)\omega(x)}$,

#2. $|L_N u_N|_H^2 \leq 2\max|a|^2(N+1)^2\left\|\frac{u_N(x,t)}{1-x}\right\|^2_{(1-x)\omega(x)}$.

The two inequalities for the constant coefficients case, $a(\cdot)\equiv a>0$, can be found in [GotTad91, eq. (3.37) and eq. (3.39)]. Similar estimates hold with variable coefficients, $a(\cdot)>0$, consult [GotTad91, eq. (6.18)] and respectively [GotTad91, eq. (6.19)].

Combining these two inequalities, we find that L is coercive with coercivity constant $\eta \simeq N^{-2}$, and Proposition 1 implies the stability of the fully discrete third- and forth-order RK methods under the CFL condition, consult [GotTad91, Theorem 4.2],

$$\Delta t \leq Const.\frac{1}{N^2 a_\infty}. \tag{56}$$

We shall now revisit the same stability question using Theorem 2. The spatial Chebyshev differencing of (52), L_N, is uniquely determined by setting $(L_N u_N)_j =$

$a(x_j)u'_N(x_j)$ augmented with zero boundary condition, $(L_N u_N)_0 = u_N(x_0) = 0$ at $x_0 = 1$. Since the N-degree polynomial interpolant $u_N(x,t)$ must vanish at $x=1$, it admits the factorization, $u_N(x,t) = (1-x)p_N(x)$, and a straightforward computation then yields,

$$\begin{aligned}
|L_N u_N|_H^2 = \sum_{j=1}^{N}(1+x_j)\omega_j a(x_j)(u'_N(x_j,t))^2 \leq \\
\leq 2a_\infty \sum_j (1+x_j)\omega_j (1-x_j)^2 (p'_N(x_j))^2 \\
+2\sum_j (1+x_j)\omega_j a(x_j) p_N^2(x_j) =: I_1 + I_2.
\end{aligned} \tag{57}$$

Application of the inverse inequality [GotTad91, Lemma 2.1], $\|p'_N\|_{(1-x^2)\mathrm{w}(x)} \leq N^2 \|p_N\|_{\mathrm{w}(x)}$, with $\mathrm{w}(x) = (1-x)\omega(x)$, implies that the first term on the right is bounded from above by

$$I_1 \leq 2N^2 a_\infty \sum_j (1-x_j)\omega_j p_N^2(x_j) \leq 2N^2 a_\infty \max_j \frac{a(x_j)}{1-x_j^2} |u_N|_H^2.$$

The second term on the right of (57) does not exceed

$$I_2 \leq 2\sum_j (1+x_j)\omega_j a(x_j) \frac{u_N^2(x_j)}{(1-x_j)^2} \leq 2a_\infty \max_j \frac{a(x_j)}{(1-x_j)^2} |u_N|_H^2,$$

and since $(1-x_j)^{-2} \leq N(1-x_j^2)^{-1}$, we find that the same upper bound of I_1 applies to I_2. We conclude

$$\|L_N\|_H = \sup \frac{|L_N u_N|_H}{|u_N|_H} \leq 2N\sqrt{a_\infty} \max_j \sqrt{\frac{a(x_j)}{1-x_j^2}},$$

arriving at the following.

Corollary 6. *Consider the fully-discrete Chebyshev approximation of the advection equation (52) with the third-order RK time discretization,*

$$u_N(t^n + \Delta t) = P_3\Big(\Delta t A_N D_T\Big) u_N(t^n), \quad A_N = \begin{bmatrix} a(x_1) & & \\ & \ddots & \\ & & a(x_N) \end{bmatrix}.$$

This fully discrete scheme is stable,

$$|u_N(t)|_H \leq |u_N(0)|_H, \quad |u_N|_H := \Big(\sum_j (1+x_j)\omega_j \frac{1}{a(x_j)} u_N^2(x_j)\Big)^{1/2},$$

under the CFL condition

$$\Delta t \leq \frac{1}{2N\sqrt{a_\infty}} \times \min_j \sqrt{\frac{1-x_j^2}{a(x_j)}}. \tag{58}$$

The CFL condition (58) implies the familiar time-step restriction for Chebyshev method of order $\Delta t \leq Const.N^{-2}$, the same we met earlier in (56). This is more restrictive, by an extra factor of $\mathcal{O}(N)$, than the usual advective CFL condition, e.g., (43), (51). Corollary 6 provides a more precise CFL bound, however, which reveals that the extra factor of $\mathcal{O}(N)$ in this strict CFL condition is due to the normalized speed, $\max_j \sqrt{a(x_j)/(1-x_j^2)} \leq N\sqrt{a_\infty}/\pi$. In particular, it suggests that by a change of variables which 'slows down' the transport velocity near the boundaries so that $\sqrt{a(x_j)/(1-x_j^2)}$ remains bounded, we could recover the improved CFL restriction $\Delta t \leq Const.N^{-1}$.

6 Appendix. On the resolvent stability condition

The notion of H-weighted stability (6) of L_N's guarantees stability of the semidiscrete (2) with respect to initial perturbations. That is, an initial perturbation, say of size $\mathcal{O}(\delta)$, is amplified by no more than $K\delta$ later in time. This notion is intimately related to yet another notion of stability – stability with respect to inhomogeneous perturbations. This is realized by a resolvent stability condition which we now explore. Here, we are led to investigate the stability of our algorithm in the presence of an inhomogeneous term, and to this end we consider the semidiscrete problem

$$\frac{du_N}{dt} = L_N u_N + F_N, \tag{59}$$

assuming, without loss of generality, zero initial values, $u_N(0) = 0$, (for otherwise, we can subtract the non vanishing initial data which instead can be added to the inhomogeneous term).

To analyze the stability of (59), we multiplying (59) by an exponential weight $e^{-\sigma t}$, $\sigma > 0$, and Fourier transform in time (setting $u_N \equiv 0$ for $t < 0$),

$$e^{-\sigma t}u_N(t) = \frac{1}{\sqrt{2\pi}}\int_{\xi=-\infty}^{\infty} \hat{u}_N(\xi)e^{\imath\xi t}d\xi, \qquad \hat{u}_N(\xi) := \frac{1}{\sqrt{2\pi}}\int_t u_N(t)e^{-\sigma t}e^{-\imath\xi t}dt.$$

The semidiscrete (2) reads

$$(\sigma + i\xi)\hat{u}_N(\xi) = L_N\hat{u}_N(\xi) + \hat{F}_N(\xi).$$

Abbreviating $s := \sigma + \imath\xi$, we arrive at the so-called resolvent-equation

$$\hat{u}_N(s) = (sI - L_N)^{-1}\hat{F}_N(s),$$

where $\hat{u}_N(s)$ stands for the Fourier-Laplace transform $\hat{u}_N(s) := \frac{1}{\sqrt{2\pi}}\int e^{-st}u_N(t)dt$.

By Parseval, we have

$$\|e^{-\sigma t}u_N\| = \|\hat{u}_N(s)\| \leq \|(sI-L_N)^{-1}\|\cdot\|\hat{F}_N(s)\| = \|(sI-L_N)^{-1}\|\cdot\|e^{-\sigma t}F_N\|. \tag{60}$$

Thus, the question of stability with respect to the inhomogeneous term F, boils down to the boundedness of the resolvent, $\|(sI-L_N)^{-1}\|$. Clearly, if the semidiscrete problem associated with L_N is stable (4), then by bounding its Laplace transform from above, we arrive at the resolvent stability estimate,

$$\|(sI - L_N)^{-1}\| \leq \int_t |e^{-st}| \cdot \|e^{L_N t}\| dt \leq \frac{K}{\Re es}. \tag{61}$$

Finally, we can utilize (60) coupled with Parseval, to translate (61) back into the "physical space". Namely, if we let L^2_σ denote the weighted L^2-norm $\|w\|_{L^2_\sigma} := \int_{t=0}^{\infty} e^{-\sigma t}|w(t)|^2 dt < \infty$, then the resulting stability of our inhomogeneous algorithm (59), states that

$$\|u_N\|_{L^2_\sigma} \leq \frac{K}{\sigma}\|F_N\|_{L^2_\sigma}, \quad \forall F_N \in L^2_\sigma,\ \sigma > 0. \tag{62}$$

Thus, the notion of stability in (4) implies the resolvent stability in (62). In particular, H-weighted stability implies the strict resolvent bound (61) with $K = 1$, i.e., $(sI - L_N)^{-1}\|_{H_N} \leq (\Re es)^{-1}$, which in turn yields the apriori estimate

$$\|u_N\|_{H^2_\sigma} \leq \frac{1}{\sigma}\|F_N\|_{H^2_\sigma}, \quad \|F_N\|_{H^2_\sigma} := \int_{t=0}^{\infty} e^{-\sigma t}|F_N(t)|^2_H dt < \infty,\ \forall \sigma > 0. \tag{63}$$

Similarly, the notion of strong stability of the inhomogeneous fully discrete method,

$$u_N(t^n + \Delta t) = P_k(\Delta t L_N)u_N(t^n) + F_N(t^n),$$

implies the resolvent condition,

$$\|(zI - L_N)^{-1}\| \leq \frac{K}{|z| - 1},\ \forall |z| > 1, \tag{64}$$

and the analogous stability estimate

$$\|u_N(t)\|_{H^2_\sigma} \leq \frac{K}{\sigma \Delta t}\|F_N(t)\|_{H^2_\sigma}, \tag{65}$$
$$\|F_N(t)\|_{H^2_\sigma} := \sum_{n=0}^{\infty} e^{-n\sigma\Delta t}|F_N(t^n)|^2_H \Delta t < \infty,\ \forall \sigma > 0.$$

The converse of these implications is a more intricate issue, with the important example of families of finite-dimensional matrices, (8), covered by the Kreiss matrix theorem, [RicMor67, §4.9]. Moreover, the Hille-Yoshida theory [Yos68] implies that strict resolvent estimate, (63) with $K = 1$, implies the semiboundedness for *arbitrary* L_N's.

The result of Kreiss and Wu deals with the preservation of *resolvent* stability by a class of RK methods, including the third- and fourth order methods (13) and (14).

Theorem 7. *[KreWu93, Theorem 3.2]. Assume the semidiscrete (2) is resolvent stable in the sense of satisfying (61) and consider its time integration by locally stable RK method. Then the corresponding inhomogeneous fully discrete system is resolvent stable so that its solution, $u_N(t)$, satisfies (65).*

Acknowledgments

I am grateful to Hailiang Liu and Donald Estep for reading the manuscript and making several helpful comments.

Bibliography

[GodRya63] S.K. Godunov & V. S. Ryabenkii, Special criteria of stability of boundary-value problems for non-self-adjoint difference equations, Uspekhi Mat. Nauk, 18 (1963), pp. 3-.

[GolTad81] M. Goldberg & E. Tadmor, Scheme-independent stability criteria for difference approximations of hyperbolic initial-boundary value problems. II, Math. of Comp. 36 (1981), 603-626.

[GooHouTad94] J. Goodman, T. Hou & E. Tadmor, On the stability of the unsmoothed Fourier method for hyperbolic equations, Numerische Mathematik 67(1) (1994), 93-129.

[GotShuTad99] S. Gottlieb, C-W. Shu & E. Tadmor, Strong stability-preserving high-order time discretization methods, SIAM Rev. 43 (2001), pp. 89-112.

[GotTad91] D. Gottlieb & E. Tadmor, The CFL condition for spectral approximations to hyperbolic initial-boundary value problems, Math. Comp. 56 (1991), pp. 565-588.

[GusKreSun72] B. Gustafsson, H.-O. Kreiss & A. Sundström, Stability theory of difference approximations for mixed initial boundary value problems. II, Math. Comp. 26 (1972), pp. 649-686.

[GusKreOli] B. Gustafsson, H.-O. Kreiss & J. Oliger, *Time Dependent Problems and Difference Methods*, Wiley-Interscience, 1995.

[Ise96] A. Iserles, *A First Course in the Numerical Analysis of Differential Equations*, Cambridge Texts in Appl. Math., 1996.

[KreSch92] H.-O. Kreiss & G. Scherer, Method of lines for hyperbolic differential equations, SINUM 29 (1992), 640-646.

[KreWu93] H.-O. Kreiss & Wu, On the stability of difference approximations for the initial boundary value problem, Appl. Numer. Math. 12 (1993), pp. 213-227.

[Lev98] D. Levermore, 1998, private communication.

[LevTad98] D. Levy & E. Tadmor, From semidiscrete to fully discrete: stability of Runge-Kutta schemes by the energy method, SIAM Rev. 40 (1998) 40-73.

[RicMor67] R. Richtmyer & B. Morton, *Difference Methods for Initial-Value Problems*, 2nd ed., John Wiley, New York, 1967.

[Shu02] C.-W. Shu, Strong stability-preserving high-order time discretization methods, *Collected Lectures on the Preservation of Stability under Discretization*, SIAM 2002, this volume.

[Tad87] E. Tadmor, Stability analysis of finite-difference, pseudospectral and Fourier-Galerkin approximations for time-dependent problems, SIAM Rev. 29 (1987), pp. 525-555.

[Tre96] N. Trefethen, *Finite Difference and Spectral Methods for Ordinary and Partial Differential Equations*, http://web.comlab.ox.ac.uk/oucl/work/nick.trefethen/pdetext.html, 1996.

[Yos68] Yoshida K., *Functional Analysis*, Springer-Verlag, New-York, 1968.