SIAM J. NUMER. ANAL.
Vol. 49, No. 1, pp. 232–263

LINEAR STABILITY OF PARTITIONED RUNGE–KUTTA METHODS*

R. I. MCLACHLAN†, Y. SUN‡, AND P. S. P. TSE‡

Abstract. We study the linear stability of partitioned Runge–Kutta (PRK) methods applied to linear separable Hamiltonian ODEs and to the semidiscretization of certain Hamiltonian PDEs. We extend the work of Jay and Petzold [*Highly Oscillatory Systems and Periodic Stability*, Preprint 95-015, Army High Performance Computing Research Center, Stanford, CA, 1995] by presenting simplified expressions of the trace of the stability matrix, $\operatorname{tr} M_s$, for the Lobatto IIIA–IIIB family of symplectic PRK methods. By making the connection to Padé approximants and continued fractions, we study the asymptotic behavior of $\operatorname{tr} M_s(\omega)$ as a function of the frequency ω and stage number s.

Key words. asymptotic behavior, continued fractions, linear stability, Lobatto IIIA–IIIB methods, Padé approximants, trace of stability matrix

AMS subject classifications. 33C20, 34E05, 65L06, 65L07, 65L20

DOI. 10.1137/100787234

1. Introduction. Partitioned Runge–Kutta (PRK) methods have a checkered history. They were first introduced in the 1970s for the integration of certain stiff differential equations. This area did not develop, partly because of a lack of naturally partitioned stiff systems. There was renewed interest in the 1990s with the advent of symplectic integration of Hamiltonian systems, with their natural partitioning into position (q) and momentum (p) variables. In 1993, Sanz-Serna and Abia [27] and Sun [28] found conditions on the parameters for the s-stage PRK method

$$\begin{aligned} Q_i &= q_0 + h\sum_{j=1}^{s} a_{ij} f(Q_j, P_j), \\ P_i &= p_0 + h\sum_{j=1}^{s} \hat{a}_{ij} g(Q_j, P_j), \\ q_1 &= q_0 + h\sum_{j=1}^{s} b_j f(Q_j, P_j), \\ p_1 &= p_0 + h\sum_{j=1}^{s} \hat{b}_j g(Q_j, P_j) \end{aligned} \tag{1.1}$$

*Received by the editors March 2, 2010; accepted for publication (in revised form) November 8, 2010; published electronically February 8, 2011.
http://www.siam.org/journals/sinum/49-1/78723.html

†IFS, Massey University, Palmerston North, New Zealand (r.mclachlan@massey.ac.nz). This author's research was supported by the Marsden Fund of the Royal Society of New Zealand.

‡LSEC, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, P.O. Box 2719, Beijing 100190, China (sunyj@lsec.cc.ac.cn, ptse@lsec.cc.ac.cn). The second author's research was supported by the National Natural Science Foundation of China (60931002) and the National Basic Research Program of China (2010CB832702). The third author's research was supported by China Postdoctoral Science Foundation.

to be symplectic when applied to the canonical Hamiltonian system

$$\begin{aligned} \dot{q} &= f(q,p) = \frac{\partial H}{\partial p}, \\ \dot{p} &= g(q,p) = -\frac{\partial H}{\partial q}, \end{aligned} \tag{1.2}$$

where $q \in \mathbb{R}^n$, $p \in \mathbb{R}^n$, and the Hamiltonian $H\colon \mathbb{R}^{2n} \to \mathbb{R}$. Unlike symplectic Runge–Kutta (RK) methods, PRK methods can be explicit, but only when the Hamiltonian is separable, i.e., $H = T(p)+V(q)$, and in this case they reduce to composition methods. The study of these methods from the PRK and composition points of view proceeded in parallel. In 1995, Sun [29] constructed some families of implicit symplectic PRK methods, in particular, the family of most concern in the present paper, namely the *Lobatto IIIA–IIIB* family of methods based on Lobatto quadrature. In this family, the (a_{ij}, b_j) coefficients are those of the Lobatto IIIA RK method and the $(\hat{a}_{ij}, \hat{b}_j)$ coefficients are those of the Lobatto IIIB RK method, methods introduced in 1969 by Ehle [8]. (There is also a partnered family of symplectic PRK methods, with A and B coefficients swapped.) For brevity we will call the Lobatto IIIA–IIIB methods *Lobatto PRK* methods. The method with $s \geq 2$ stages has order $2s-2$. Its first ($s = 2$) member is known as the *generalized leapfrog method* and can be written in the form

$$\begin{aligned} P &= p_0 + \frac{1}{2} h g(q_0, P), \\ q_1 &= q_0 + \frac{1}{2} h \left(f(q_0, P) + f(q_1, P) \right), \\ p_1 &= P + \frac{1}{2} h g(q_1, P). \end{aligned} \tag{1.3}$$

(When H is separable, this reduces to the explicit, symplectic leapfrog method.) This is sometimes used for symplectic integration of nonseparable systems [15], often together with a composition to increase the order, because it requires solving one system of n algebraic equations rather than the $2n$ equations required by the midpoint rule. For some H the n equations may be particularly easy to solve. However, for high order symplectic integration of nonseparable systems, Gauss RK tends to be preferred over Lobatto PRK because of its optimal order ($2s$), A-stability, and very small error constants; when the system is not stiff, the nonlinear equations can be solved fairly easily and quickly [19].

In 1996 Jay [13] discovered that Lobatto PRK was suitable for the symplectic integration of Hamiltonian systems subject to holonomic constraints of the form $G(q) = 0$. Gauss RK suffers an order reduction (from $2s$ to s) because the constraints are not enforced at the endpoints, while Lobatto PRK is still superconvergent of order $2s-2$. In addition, the presence of the constraint usually forces one to use an implicit method anyway. However, in 2007 Jay [14] modified Gauss RK so that order $2s$ was retained for constrained systems, thus removing the apparent advantage of PRK methods here.

In 2003, Grimm and Scherer [9] generalized the W-transformation of Hairer, Nørsett, and Wanner [10] to PRK methods, obtaining, amongst other things, a construction of all high order symplectic PRK methods.

In 1995, Jay and Petzold in an unpublished report [12] studied the linear stability of Lobatto PRK and proved that none of this family of methods are P-stable—roughly, that they are not unconditionally stable when applied to the harmonic oscillator. They concluded that they were not suitable for highly oscillatory systems. For nonstiff systems, the significance of this result is not so clear. Consider comparing the midpoint

rule and the generalized leapfrog method (1.3). The former is P-stable, but as the high frequencies and their interactions with the low frequencies are not captured correctly anyway [2], the non-P-stability of (1.3) is not as bad as thought.

In 2000, Reich [23] suggested the use of Gauss RK methods for the spatial discretization of Hamiltonian PDEs, i.e., wave equations. Combined with symplectic time integration, a conservation law that is formally a discrete analogue of the multisymplectic conservation law of the PDE can be obtained. Furthermore, its behavior on linear systems has some interesting features: for example, the midpoint (box) method can qualitatively preserve the dispersion relation of any system of Hamiltonian PDEs for all time and space steps. Unfortunately, it leads to fully implicit systems of equations that may not have a solution. To avoid this, Ryland, McLachlan, and Frank [25] considered the use of partitioned symplectic PRK methods for spatial discretization, finding conditions on the PDE under which the Lobatto PRK methods lead to explicit spatial discretizations, central differences of second-order spatial derivatives being the lowest-order member, and the spatial analogue of the leapfrog method. In this application, we know of no alternative to the use of partitioned methods.

Both Gauss and Lobatto methods are variational and hence can be derived in the context of Galerkin finite element schemes with quadrature [18]; see also the discussion of the relationship between (especially Gauss) RK methods and Galerkin finite element methods in [5].

In the application to spatial semidiscretization, it is vital that the PRK method has a certain stability property that is different from that arising in time integration. However, we will show below that the response of the method on the harmonic oscillator contains all the information required to understand its stability and dispersion when used in spatial semidiscretization. Thus, a complete understanding of the linear stability of PRK methods, and Lobatto PRK methods, in particular, is required.

An outline of the paper is as follows. In section 2, we show that the linear stability analysis of partitioned methods is substantially harder than that of nonpartitioned methods, because no normal form allows one to reduce to low-dimensional systems. We focus on separable Hamiltonian systems, for which the harmonic oscillator is a normal form, but show that nonseparability can influence linear stability. Section 3 gives a general treatment of PRK methods applied to the harmonic oscillator and introduces the central object of our study, the *stability function* $\operatorname{tr} M(\omega)$ and the *stability region* $\{\omega \in \mathbb{R} \colon |\operatorname{tr} M(\omega)| \le 2\}$. In section 4 we prove that the (compositional) inverse of the stability function also determines the stability of PRK when used as a spatial discretization of certain Hamiltonian PDEs such as the nonlinear wave equation. In section 5 we calculate and discuss the stability function of Lobatto IIIA–IIIB for two to six stages; certain obvious patterns observed in these cases become conjectures that are proved later in the paper. Section 6 reviews an unpublished work of Jay and Petzold [12] that is used in section 7 to establish our key results, a complete description of the stability region for Lobatto IIIA–IIIB for any number of stages and an explicit expression for the stability function as a rational function of two explicit continued fractions. These continued fractions are related to the diagonal Padé approximants to the exponential function which allows us in section 8 to describe the asymptotic behavior of the stability function in various regimes.

2. Normal forms of partitioned linear systems. The stability analysis of linear methods like RK methods centers on the response $y_0 \mapsto y_1 = R(h\lambda)y_0$ of the method to the Dahlquist test equation $\dot{y} = \lambda y$, λ, $y \in \mathbb{C}$. $R \colon \mathbb{C} \to \mathbb{C}$ contains all the information about the behavior of the method on linear problems. This is

because, when applied to the linear system $\dot{x} = Ax$, $x \in \mathbb{R}^n$, the RK method yields $x_0 \mapsto x_1 = R(hA)x_0$. The matrix A can be put in its Jordan normal form, and, applied to a Jordan block J with eigenvalue λ, $R(J)$ is the Toeplitz matrix

$$\begin{pmatrix} R(\lambda) & \lambda R'(\lambda) & \frac{1}{2!}\lambda^2 R''(\lambda) & \dots \\ 0 & R(\lambda) & \lambda R'(\lambda) & \ddots \\ \vdots & \ddots & \ddots & \ddots \end{pmatrix} \tag{2.1}$$

from which information on stability and linear contractivity can be derived from $R(z)$ [11]. The crucial point that allows this approach to succeed is that the linear change of variables that puts A in its normal form commutes with the Runge–Kutta discretization, that is, RK is *linearly covariant* [20].

The situation is different with PRK methods because they are not covariant with respect to nonseparable linear transformations. The behavior of the method does not depend only on the eigenvalues (or Jordan normal form) of the system.

Example 2.1. The Hamiltonians $H_1 = \frac{1}{2}(p^2 - q^2)$ and $H_2 = pq$ both generate two-dimensional linear systems with eigenvalues ± 1. The generalized leapfrog method (1.3) leads to symplectic linear maps $y_0 \mapsto y_1 = M_i(h)y_0$ with

$$\begin{aligned} M_1(h) &= \begin{pmatrix} 1 + \frac{h^2}{2} & h \\ h + \frac{h^3}{4} & 1 + \frac{h^2}{2} \end{pmatrix}, \ \lambda = \frac{1}{2}(2 + h^2 \pm \sqrt{4 + h^2}); \\ M_2(h) &= \begin{pmatrix} \frac{2+h}{2-h} & 0 \\ 0 & \frac{2-h}{2+h} \end{pmatrix}, \ \lambda = \frac{2 \pm h}{2 \mp h}. \end{aligned} \tag{2.2}$$

On H_1, the method reduces to the leapfrog method and has the same stability properties as the differential equation (namely $0 < \lambda_1 < 1 < \lambda_2$, $\lambda_1 \to 0$, $\lambda_2 \to \infty$ as $h \to \infty$) for all h. On H_2, the method reduces to the midpoint rule and only has the correct stability properties for $0 < h < 2$. For $h = 2$ the method is undefined, and for $h > 2$ the eigenvalues have the wrong sign and the wrong limits as $h \to \infty$.

THEOREM 2.1. *Under invertible partitioned linear maps*

$$q \mapsto X_1 q, \quad p \mapsto X_2 p, \tag{2.3}$$

partitioned ODEs

$$\begin{pmatrix} \dot{q} \\ \dot{p} \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} q \\ p \end{pmatrix} \tag{2.4}$$

cannot in general be (i) *diagonalized,* (ii) *block diagonalized with* 2×2 *blocks,* (iii) *block diagonalized in the Hamiltonian case by partitioned symplectic maps, or* (iv) *block diagonalized in the Hamiltonian case by arbitrary partitioned linear maps.*

Proof. Cases (i)–(iii) are subsumed by case (iv), but it is instructive to consider them first. The transformed system is

$$\begin{pmatrix} \dot{q} \\ \dot{p} \end{pmatrix} = \begin{pmatrix} X_1^{-1} A X_1 & X_1^{-1} B X_2 \\ X_2^{-1} C X_1 & X_2^{-1} D X_2 \end{pmatrix} \begin{pmatrix} q \\ p \end{pmatrix}. \tag{2.5}$$

For (i), diagonalization requires $X_1^{-1} B X_2 = 0 \Rightarrow B = 0$. For (ii), A and D can generically be diagonalized; doing so determines X_1 and X_2 up to scalings and permutations

of the eigenvectors. This freedom is not enough to (generically) diagonalize B and C. For (iii), if the system is Hamiltonian, then $B = B^\top$, $C = C^\top$, and $D = -A^\top$; if the transformation is symplectic, then $X_2 = X_1^{-\top}$. We can still diagonalize A and D but not the symmetric matrices B and C. For (iv), allowing arbitrary invertible X_1, X_2 does not help, as the additional (scaling and permutation only) symmetry still does not help diagonalize B or C. ∎

In the two-dimensional case, the transformed system is

$$\begin{pmatrix}\dot q\\ \dot p\end{pmatrix} = \begin{pmatrix} a & b\frac{x_2}{x_1}\\ c\frac{x_1}{x_2} & d\end{pmatrix}\begin{pmatrix} q\\ p\end{pmatrix} \tag{2.6}$$

so that we have the normal forms

$$\begin{pmatrix} a & b\\ b & d\end{pmatrix},\quad \begin{pmatrix} a & b\\ -b & d\end{pmatrix},\quad \begin{pmatrix} a & 0\\ 1 & d\end{pmatrix} \tag{2.7}$$

in the cases $bc > 0$, $bc < 0$, and $b = 0$, respectively. In the case that the system is Hamiltonian, we also have $d = -a$, giving 2-parameter normal forms with eigenvalues $\pm\sqrt{a^2+b^2}$, $\pm\sqrt{a^2-b^2}$, and $\pm a$, respectively.

Example 2.2. Consider the two-dimensional Hamiltonian system

$$\begin{pmatrix}\dot q\\ \dot p\end{pmatrix} = \begin{pmatrix} \mu & \omega\\ -\omega & -\mu\end{pmatrix}\begin{pmatrix} q\\ p\end{pmatrix} \tag{2.8}$$

with eigenvalues $\pm i\sqrt{\omega^2-\mu^2}$. It is stable when $|\mu| \le |\omega|$. The generalized leapfrog method (1.3) generates the map $(q_0, p_0)^\top \mapsto M(h\omega, h\mu)(q_0, p_0)^\top$ with

$$M(\omega,\mu) = \frac{1}{4-\mu^2}\begin{pmatrix} 4+\mu^2-2\omega^2+4\mu & 4\omega\\ \omega(-4-\mu^2+\omega^2) & 4+\mu^2-2\omega^2-4\mu\end{pmatrix} \tag{2.9}$$

and is stable (i.e., $|\operatorname{tr} M| \le 2$; see below) when $|h\omega| \le 2$ and $|\mu| \le |\omega|$. Thus, the nonseparability (μ) does not influence the numerical stability.

However, a longer calculation (not shown) for the 3-stage, fourth-order Lobatto PRK method gives that it is stable when $|h\omega| \le \sqrt{24}$, $|\omega| \le |\mu|$, and $\frac{h^2\omega^2}{12+h^2\mu^2} \notin (\frac{2}{3}, 1)$. Thus in this case the nonseparability *does* influence the numerical stability.

We do not know of a simple normal form for the general or Hamiltonian $2n$-dimensional case. However, in the separable Hamiltonian case ($A = D = 0$, B and C symmetric), taking $X_1 = BX_2$ gives I in the upper right corner and $X_2^{-1}CBX_2$ in the lower right corner of the matrix in (2.6); if B is invertible, then CB is similar to the symmetric matrix $B^{1/2}CB^{1/2}$, hence diagonalizable, giving the normal form

$$\begin{pmatrix}\dot q\\ \dot p\end{pmatrix} = \begin{pmatrix} 0 & I\\ \Lambda & 0\end{pmatrix}\begin{pmatrix} q\\ p\end{pmatrix}, \tag{2.10}$$

where Λ is diagonal and real. This gives a block-diagonalization into n two-dimensional systems $\dot q = p$, $\dot p = \lambda q$, each a harmonic oscillator (if $\lambda < 0$), Hamiltonian saddle (if $\lambda > 0$), or degenerate (if $\lambda = 0$). In this case we can therefore determine the behavior of the PRK method on the $2n$-dimensional linear system by considering only two-dimensional test problems.

We will henceforth restrict our attention to the harmonic oscillator case, taking the test equation in the form

$$\begin{pmatrix}\dot q\\ \dot p\end{pmatrix} = \begin{pmatrix} 0 & \omega\\ -\omega & 0\end{pmatrix}\begin{pmatrix} q\\ p\end{pmatrix} \tag{2.11}$$

with $q, p \in \mathbb{R}$. While this is, in fact, the standard test equation for symplectic integrators, we wish to emphasize that this test equation is not sufficient for understanding the linear response of an integrator even in the two-dimensional case, as Examples 2.1 and 2.2 show. Getting any kind of result in the general case, i.e., for (2.4), looks difficult.

3. Stability functions and stability regions for PRK methods. Applying a PRK method to (2.11) gives a two-dimensional linear map $y_0 \mapsto y_1 := M(h\omega)y_0$. The eigenvalues of M are $\frac{1}{2}\operatorname{tr} M \pm ((\frac{1}{2}\operatorname{tr} M)^2 - \det M)^{1/2}$. If the method is symplectic, then $\det M = 1$, and the eigenvalues lie on the unit circle iff $|\operatorname{tr} M| \le 2$. This is the stability criterion for two-dimensional symplectic maps. If $|\operatorname{tr} M| \le 2$, the eigenvalues are $\mathrm{e}^{\pm \mathrm{i}\theta}$, where $\cos\theta = \frac{1}{2}\operatorname{tr} M$; thus, as $\operatorname{tr} M$ decreases from 2 to -2, say, the eigenvalues move around the unit circle from 1 to -1.

DEFINITION 3.1. *The* stability function *of a symplectic PRK method (for the test problem* (2.11)*) is* $\operatorname{tr} M(\omega)$. *The* stability region *of a symplectic PRK method is*

$$\{\omega \in \mathbb{R} \colon |\operatorname{tr} M(\omega)| \le 2\}. \tag{3.1}$$

The method is P-stable *if its stability region is* $\mathbb{R}$.[1]

In case the PRK method is an RK method with stability function $R(z)$, we have $\operatorname{tr} M(h\omega) = 2\operatorname{Re} R(\mathrm{i}h\omega)$, so we can use $\operatorname{tr} M$ in place of R for RK methods. Symplectic RK methods always have $R(z)R(-z) = 1$; hence they have $|R(\mathrm{i}\omega)| = 1$ and are P-stable.

The exact solution of (2.11) is

$$\begin{pmatrix} q(t) \\ p(t) \end{pmatrix} = M_{\text{exact}}(\omega t) \begin{pmatrix} q_0 \\ p_0 \end{pmatrix},$$

where

$$M_{\text{exact}}(\omega t) = \begin{pmatrix} \cos(\omega t) & \sin(\omega t) \\ -\sin(\omega t) & \cos(\omega t) \end{pmatrix}. \tag{3.2}$$

Therefore

$$\operatorname{tr}(M_{\text{exact}}(\omega)) = 2\cos(\omega).$$

Let I_s be the $s \times s$ identity matrix, let $\mathbf{1}_s$ be the vector $(1, \dots, 1)^\top \in \mathbb{R}^n$, let A, $\hat{A}$ be the matrices of the PRK coefficients, and let b, $\hat{b}$ be the vectors of the PRK weights. The application of a PRK method to (2.11) yields the linear system

$$\begin{pmatrix} I_s & -\omega A \\ \omega\hat{A} & I_s \end{pmatrix} \begin{pmatrix} Q \\ P \end{pmatrix} = \begin{pmatrix} \mathbf{1}_s q_0 \\ \mathbf{1}_s p_0 \end{pmatrix}, \tag{3.3}$$

$$\begin{pmatrix} q_1 \\ p_1 \end{pmatrix} = \begin{pmatrix} q_0 \\ p_0 \end{pmatrix} + h\omega \begin{pmatrix} 0 & b^\top \\ -\hat{b}^\top & 0 \end{pmatrix} \begin{pmatrix} Q \\ P \end{pmatrix}. \tag{3.4}$$

Hence we get

$$\begin{pmatrix} q_1 \\ p_1 \end{pmatrix} = M(h\omega) \begin{pmatrix} q_0 \\ p_0 \end{pmatrix} \tag{3.5}$$

[1] If $\operatorname{tr} M(\omega) = \pm 2$, then the map has a double eigenvalue at ± 1 and may be algebraically unstable with solutions that are $\mathcal{O}(t)$, whereas the solutions of the ODE may be either $\mathcal{O}(1)$ or $\mathcal{O}(t)$. We still call the method stable in this case.

with

$$M(\omega) = I_2 + \omega \begin{pmatrix} 0 & b^\top \\ -\hat{b}^\top & 0 \end{pmatrix} \begin{pmatrix} I_s & -\omega A \\ \omega \hat{A} & I_s \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{1}_s & 0 \\ 0 & \mathbf{1}_s \end{pmatrix}. \tag{3.6}$$

This already allows the stability functions to be calculated for any specific PRK method.

4. Stability function determines dispersion relation in multisymplectic integration. We have given one interpretation of the stability function $\operatorname{tr} M(h\omega)$; namely, in the time integration of (2.11), a PRK method is

$$\begin{aligned} &|\operatorname{tr} M(\omega)| \le 2 : \text{stable, with eigenvalues } \mathrm{e}^{\pm \mathrm{i}\theta}, \cos\theta = \frac{1}{2}\operatorname{tr} M, \\ &|\operatorname{tr} M(\omega)| > 2 : \text{unstable.} \end{aligned} \tag{4.1}$$

Some examples of $\operatorname{tr} M(\omega)$ are given in Figure 5.1 for Lobatto PRK. In this interpretation, the practical stability limit of the method is

$$\omega^* := \inf\{\omega : |\operatorname{tr} M(\omega)| > 2\}. \tag{4.2}$$

In other words, where the stability region consists of several intervals, only the smallest one is relevant.

We now give another interpretation of the stability function in which all of the intervals are relevant.

Multisymplecticity is the extension of symplecticity to Hamiltonian PDEs. We consider here PDEs that can be written in the multi-Hamiltonian form [6]

$$\mathsf{K} z_t + \mathsf{L} z_x = \nabla_z S(z), \tag{4.3}$$

where $z(t,x) \in \mathbb{R}^n$, K and L are skew-symmetric matrices, and $S(z)$ is a smooth function. Along solutions $z(t,x)$ to a PDE of this form, the multisymplectic conservation law $(\mathsf{K}\mathrm{d}z \wedge \mathrm{d}z)_t + (\mathsf{L}\mathrm{d}z \wedge \mathrm{d}z)_x = 0$ holds, where $\mathsf{K}\mathrm{d}z_t + \mathsf{L}\mathrm{d}z_x = D_{zz}S(z)\mathrm{d}z$. By analogy with Hamiltonian ODEs, multisymplectic integrators are those for which a discrete analogue of the multisymplectic conservation law holds [6]. (In contrast to the case of symplectic integrators, in multisymplectic integrators, the discrete conservation law depends on the method.)

If RK methods (or PRK methods with suitable partitionings) are applied to the time *and* space derivatives in (4.3), one obtains a system of discrete equations that formally satisfies a discrete multisymplectic conservation law [26]. An example is the box scheme that arises from applying the midpoint rule in both space and time. In [3] it was demonstrated that the box scheme gave very smooth solutions to the Korteweg–de Vries equation at large time and space steps, and this was linked to the fact that the box scheme preserves (in a certain sense) the dispersion relation of any multi-Hamiltonian PDE for all time and space steps. Specifically, the linear multi-Hamiltonian PDE

$$\mathsf{K} z_t + \mathsf{L} z_x = \mathsf{S} z \tag{4.4}$$

has a periodic solution $z = \mathrm{e}^{\mathrm{i}kx + \mathrm{i}\omega t} y$ if the dispersion relation

$$\det(\mathrm{i}\omega \mathsf{K} + \mathrm{i}k\mathsf{L} - \mathsf{S}) = 0 \tag{4.5}$$

holds. The box scheme with $z(t,x) \approx Z(n\Delta t, m\Delta x)$ has a periodic solution $Z = \mathrm{e}^{\mathrm{i}Km\Delta x + \mathrm{i}\Omega n\Delta t}Y$ iff (4.5) holds with

$$\mathrm{e}^{\mathrm{i}K\Delta x} = R(\mathrm{i}k\Delta x), \quad \mathrm{e}^{\mathrm{i}\Omega\Delta t} = R(\mathrm{i}\omega\Delta t), \tag{4.6}$$

where R is the stability function of the midpoint rule. In this way the frequency space $\mathbb{R}^2$ of the PDE is mapped diffeomorphically into the discrete frequency space $(-\pi,\pi)^2$ via the phase of R on the imaginary axis, i.e., by the response of the midpoint rule to the harmonic oscillator. It is in this sense that the entire dispersion relation is qualitatively preserved, including the number of branches and the sign of the group velocity.

Let z_m be the node (grid point) variables, and let $Z_{m,1},\dots,Z_{m,s}$ be the stage variables at grid point m of a (P)RK method. Applying the method as a spatial semidiscretization yields a differential algebraic equation (DAE) in (z_m, Z_{mj}). If the z_m variables can be locally algebraically eliminated to determine $\frac{\partial}{\partial t}Z_{mj}$ as explicit local functions of Z, then we call the semidiscretization *explicit*. An advantage is that well-defined ODEs are then obtained regardless of boundary conditions; implicit discretizations (like the box scheme) may not yield well-defined ODEs. Theorem 4.1 of [26] gives sufficient conditions on K, L, $S(z)$, and the partitioning for a PRK method that satisfies

$$a_{1j} = 0,\ a_{rj} = b_j,\ \hat{a}_{jr} = 0,\ \hat{a}_{j1} = b_1, \quad 1 \le j \le s \tag{4.7}$$

and

$$\det C \neq 0, \quad C_{i-1,j-1} = \sum_{k,l} a_{ik}(b_l - \delta_{kl})\hat{a}_{lj}, \quad 2 \le i,j \le s-1 \tag{4.8}$$

to generate explicit semidiscretizations. (Lobatto IIIA–IIIB satisfies (4.7),(4.8).) An example is the nonlinear wave equation $u_{tt} = u_{xx} - V'(u)$, for which 2-stage Lobatto yields the explicit ODEs

$$\frac{\partial^2}{\partial t^2}U_{m,1} = (\Delta x)^{-2}(U_{m-1,1} - 2U_{m,1} + U_{m+1,1}) - V'(U_{m,1}) \tag{4.9}$$

and $U_{m,2} = U_{m+1,1}$, while 3-stage Lobatto yields the explicit ODEs

(4.10)

$$\frac{\partial^2}{\partial t^2}U_{m,1} = (\Delta x)^{-2}(-U_{m-1,1} + 8U_{m-1,2} - 14U_{m,1} + 8U_{m,2} - U_{m+1,1}) - V'(U_{m,1}),$$

$$\frac{\partial^2}{\partial t^2}U_{m,2} = (\Delta x)^{-2}(4U_{m,1} - 8U_{m,2} + 4U_{m+1,1}) - V'(U_{m,2}),$$

and $U_{m,3} = U_{m+1,1}$. In general, because of (4.7), s-stage Lobatto leads to $s-1$ independent ODEs per grid point.

Thus, the question arises as to the dispersion relation of PDEs that are (semi-)discretized in this way. Of course, we cannot expect unconditional preservation as achieved by the box scheme via (4.6). However, we do have the following result, that the dispersion relation, stability, and stiffness of the discretization is completely determined by $\operatorname{tr} M(\omega)$ and, in particular, by *(all of)* its stability intervals.

THEOREM 4.1. *Consider a linear multi-Hamiltonian PDE* (4.4) *satisfying the conditions of Theorem* 4.1 *of* [26] *such that all solutions with periodic initial data are periodic, i.e., all solutions of the dispersion relation* (4.5) *are real for any fixed real value of the spatial wavenumber* k.

(i) *The explicit ODEs obtained from an s-stage PRK method satisfying* (4.7), (4.8) *have periodic solutions of the form $Z = \mathrm{e}^{\mathrm{i}Km\Delta x + \mathrm{i}\omega t}Y$ iff the continuous dispersion relation* (4.5) *holds together with the mapping of continuous to discrete frequencies given by*

$$\cos(K\Delta x) = \frac{1}{2}\operatorname{tr} M(k\Delta x). \tag{4.11}$$

(ii) *The linear ODEs have purely imaginary eigenvalues—and hence the semidiscretization is stable—iff the stability function satisfies the condition that*

$$\frac{1}{2}\operatorname{tr} M(k\Delta x) = a \tag{4.12}$$

has precisely $s-1$ solutions, counting multiplicity, for each value of a in $[-1,1]$.

(iii) *(Only) the part of the continuous dispersion relation corresponding to values of k in the stability region is captured by the semidiscretization. Gaps in the stability region lead to spurious jumps and critical points in the dispersion relation. If, in the continuous dispersion relation, $\omega \to \infty$ as $k \to \infty$, then for sufficiently small Δx the largest eigenvalue of the ODEs (that is, the stiffness) is determined by $k^*\Delta x$, where*

$$k^*\Delta x := \sup\{k\Delta x : |\operatorname{tr} M(k\Delta x)| \le 2\}. \tag{4.13}$$

Proof.

(i) For such a PDE, the first-order space derivatives may be eliminated to write the PDE as second order in space, i.e., $u_{xx} = f(u, u_t)$, where f is linear. The time-dependence of u is assumed to be periodic, i.e., $u = \mathrm{e}^{\mathrm{i}\omega t}\tilde{u}$, giving $\tilde{u}_{xx} = \tilde{F}(\omega)\tilde{u}$, which has periodic solutions proportional to $\mathrm{e}^{\mathrm{i}kx}$ for each (k,ω) satisfying (4.5). That is, the equation now takes the form of the harmonic oscillator test equation with wavenumber k. This ODE with independent variable x and parameter ω is now discretized by the PRK method with step size Δx. If $k\Delta x$ is in the stability region, the response is periodic and at grid point x_m is proportional to $\mathrm{e}^{\mathrm{i}Km\Delta x}$, where (4.11) holds. Y is given by the values of the stage variables. If $k\Delta x$ is not in the stability region, the semidiscretization does not have a periodic solution.

(ii) The analogy in (i) with time integration yields the solution at the grid points. However, the dependent variables in the semidiscretization are the stage values $U_{m,j}$, of which (when (4.7) holds) there are $s-1$ independent values per grid point. To be stable, the solution of the ODEs must be periodic for any initial values of the $U_{m,j}$. Grouping the stage values in each cell into a vector in $\mathbb{R}^{s-1}$ and taking a $\mathbb{C}^{s-1}$-valued Fourier transform with respect to m, for each wavenumber in the discrete frequency domain, $K\Delta x \in [-\pi,\pi]$, the ODEs must support $s-1$ periodic solutions. Substituting this range of K values into (4.11) gives the result.

(iii) This is mostly a corollary of (i). For the spurious critical points, write the continuous dispersion relation as $P(\omega,k) = 0$ so that the discrete dispersion relation is $P(\omega, k(K)) = 0$. Differentiating with respect to K gives

$$\frac{\partial P}{\partial \omega}\frac{\partial \omega}{\partial K} + \frac{\partial P}{\partial k}\frac{\partial k}{\partial K} = 0. \tag{4.14}$$

On the other hand, differentiating (4.11) with respect to K gives

$$-\sin(K\Delta x) = \frac{1}{2}(\operatorname{tr} M)'(k\Delta x)\frac{\partial k}{\partial K}. \tag{4.15}$$

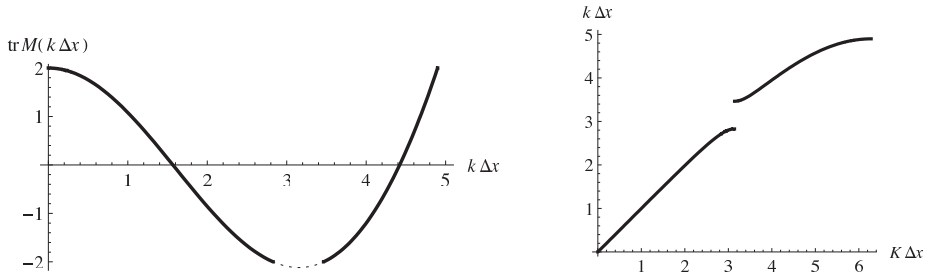

FIG. 4.1. *The stability function* $\frac{1}{2}\operatorname{tr} M_3(k\Delta x)$ *for* 3*-stage Lobatto IIIA–IIIB (left), drawn as a solid line in the stability region* $[0, 2\sqrt{2}] \cup [2\sqrt{3}, 2\sqrt{6}]$*, and the effective frequency* $k\Delta x$ *appearing in the discrete dispersion relation* (4.5), (4.12) *(right). Endpoints of the stability region generate spurious jumps and critical points in the discrete dispersion relation.*

Therefore, at a value of k such that $|\operatorname{tr} M(k\Delta x)| = 2$, we get $|\cos(K\Delta x)| = 1$ and $\sin(K\Delta x) = 0$. Therefore, if $(\operatorname{tr} M)'(k\Delta x) \neq 0$ and $\frac{\partial P}{\partial \omega} \neq 0$ (violating the latter would lead to a genuine critical point in $k(\omega)$), then $\frac{\partial k}{\partial K} = 0$ and hence $\frac{\partial \omega}{\partial K} = 0$, a spurious critical point in the dispersion relation for $\omega(K)$. ∎

One can summarize Theorem 4.1 by saying that the behavior of PRK methods is determined in time integration by the stability function and in space integration by its (compositional) inverse.

Note that if the PDE does not satisfy the required separability assumptions to lead to a PDE that is second order in space, then PRK may still be applied, but it will not lead to separable ODEs, and the stability function associated with the separable test problem (2.11) will not determine the stability or dispersion of the discretization.

Some of the assumptions of the theorem can be relaxed. For general PRK methods that do not satisfy (4.7), the condition in Theorem 4.1 is modified to require s (instead of $s-1$) solutions. For RK methods, one also needs s solutions in general, but the condition on separability of the PDE can be dropped [25]. This condition is in fact quite stringent. Gauss RK satisfies it, and hence is stable in this sense, and we shall see that Lobatto PRK satisfies it (Corollary 7.5). From Table 5.2 one can check this for $2 \leq s \leq 6$. Most other symplectic integrators that we have checked do not satisfy it and hence are not useful in multisymplectic integration (see the examples in Figure 4.2). (Such compositions were proposed in [7]. They are in fact unstable and cannot be used. Also, the order of a composition method when used in time discretization is not, as claimed in [7], the same as the order it achieves in space discretization, because the stage values and not the node values carry the information. At best the stage order can be attained.) In particular, we make the following conjecture.

CONJECTURE 4.1. *No composition method of order higher than* 2*, where the base method is either the midpoint rule or the leapfrog method, leads to stable semidiscretizations in the sense of Theorem* 4.1*.*

Because of the known problems with Gauss RK, this conjecture further focuses attention on PRK methods satisfying (4.7) and Lobatto PRK in particular.

The 1–1 correspondence between discrete and continuous frequencies established by (4.11) indicates that the $s-2$ high-frequency solutions (e.g., the right-hand side of Figure 4.1) do correspond to physical waves and are not numerical artifacts. To make a more precise statement requires comparing continuous and discrete eigenfunctions, not just eigenvalues; only in the simplest case ($s = 2$, which reduces to central differences

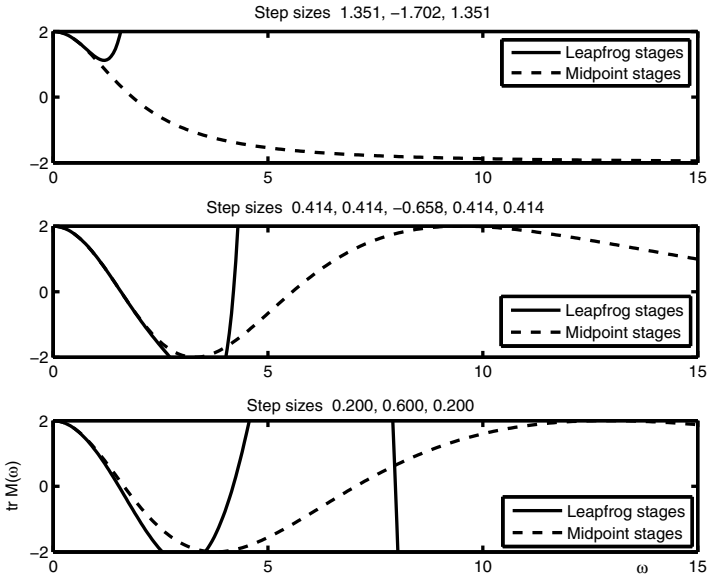

FIG. 4.2. *Response of three different compositions of leapfrog and the midpoint rule to the harmonic oscillator. Top: step sizes* $(\alpha, 1-2\alpha, \alpha)$ *for* $\alpha = 1/(2-2^{1/3})$*, a 3-stage, fourth-order composition; middle: step sizes* $(\beta, \beta, 1-4\beta, \beta, \beta)$ *for* $\beta = 1/(4-4^{1/3})$*, a much more accurate 5-stage, fourth-order composition; bottom: step sizes* $(0.2, 0.6, 0.2)$*, a 3-stage, second-order composition. When used as spatial semidiscretizations, the fourth-order methods are unstable and the second-order method is stable (see Theorem* 4.1*).*

on a uniform grid) does the restriction of the continuous eigenfunctions to the grid coincide with the discrete eigenfunctions. A start in this direction is made in [24], in which it is shown that the highest-frequency ($k \to \infty$) eigenvector for Gauss RK is a sawtooth on the nonuniform Gauss points. We plan a more detailed study of the eigenvectors in the future.

5. Stability regions for Lobatto PRK methods of fixed order. The parameters for s-stage Lobatto IIIA–IIIB methods are determined by [11]

$$
\begin{aligned}
B(r) &: \sum_{i=1}^{s} b_i c_i^{k-1} = \frac{1}{k}, \quad 1 \le k \le r, \\
C(r) &: \sum_{j=1}^{s} a_{ij} c_j^{k-1} = \frac{1}{k} c_i^k, \quad 1 \le i \le s,\ 1 \le k \le r, \\
D(r) &: \sum_{i=1}^{s} b_i c_i^{k-1} \hat{a}_{ij} = \frac{1}{k} b_j (1 - c_j^k), \quad 1 \le j \le s,\ 1 \le k \le r,
\end{aligned}
\tag{5.1}
$$

and the c_i are the zeros of

$$
\frac{\mathrm{d}^{s-2}}{\mathrm{d}x^{s-2}} \left(x^{s-1} (x-1)^{s-1} \right). \tag{5.2}
$$

Plugging into (3.6) allows one to calculate $\operatorname{tr} M_s(\omega)$ for any specific value of s in terms of $2s \times 2s$ determinants (although calculating the c_i requires solving cubics for $8 \le s \le 11$ and quintics or higher for $s \ge 12$). In Table 5.1 we give a list of the first

TABLE 5.1
The trace of the stability matrix M_s for the Lobatto IIIA–IIIB method applied to the separable Hamiltonian ODE (2.11). *The stage values are* $s = 2, 3, 4, 5, 6$.

s	$\operatorname{tr} M_s(\omega)$
2	$2 - \omega^2$
3	$\dfrac{48 - 22\omega^2 + \omega^4}{24 + \omega^2}$
4	$\dfrac{3600 - 1680\omega^2 + 92\omega^4 - \omega^6}{1800 + 60\omega^2 + \omega^4}$
5	$\dfrac{564480 - 267120\omega^2 + 16176\omega^4 - 260\omega^6 + \omega^8}{282240 + 7560\omega^2 + 108\omega^4 + \omega^6}$
6	$\dfrac{152409600 - 72817920\omega^2 + 2698280\omega^4 - 90384\omega^6 + 590\omega^8 - \omega^{10}}{76204800 + 1693440\omega^2 + 20160\omega^4 + 168\omega^6 + \omega^8}$

TABLE 5.2
The numerical stability intervals (i.e., $\{\omega : |\operatorname{tr} M_s(\omega)| \le 2\}$*) for the Lobatto IIIA–IIIB method when applied to the separable Hamiltonian ODE* (2.11). *The stage values are* $s = 2, 3, 4, 5, 6$.

s	Numerical stability intervals
2	$[0, 2]$
3	$[0, 2.82842], [3.4641, 4.89897]$
4	$[0, 3.11272], [3.16228, 5.47723], [7.74597, 8.62038]$
5	$[0, 3.14045], [3.14247, 6.05405], [6.48074, 8.25455], [13.04319, 13.54062]$
6	$[0, 3.14156], [3.14162, 6.25301], [6.30594, 8.84147], [10.10600, 11.35341], [19.49962, 19.79795]$

five rational functions $\operatorname{tr} M_s(\omega)$, and Table 5.2 displays the numerically calculated stability regions. We also plot $\operatorname{tr} M_s$ (see Figure 5.1).

From these finite-s results, we make the following observations that one can check directly to be true for $2 \le s \le 5$ and that we will later show to be true for all $s \ge 2$.

- $\operatorname{tr} M_s(\omega)$ is an even rational function of degree $2s - 2$ over $2s - 4$.
- The $2s - 2$ zeros of $\operatorname{tr} M_s(\omega)$ are all real.
- None of the $2s - 4$ poles of $\operatorname{tr} M_s(\omega)$ are real.
- The function $\operatorname{tr} M_s(\omega)$ crosses the boundaries of the stability region ± 2 exactly $2s - 2$ times, with precisely one critical point in each unstable region.

Furthermore, based on Figure 5.2 we conjecture (and later prove) that

$$\operatorname{tr} M_s(\omega) \to 2\cos(\omega) \quad \text{as } s \to \infty \text{ for all fixed } \omega \in \mathbb{R},$$

which is exactly what we expected since $2\cos(\omega)$ is precisely the trace of the stability matrix $M_{\text{exact}}(\omega)$ from the exact solution (3.2) of the linear Hamiltonian ODE. Figure 5.2 also suggests the stronger conjecture, that there exists an α^* such that

$$\operatorname{tr} M_s(\alpha s) \to 2\cos(\alpha s) \quad \text{as } s \to \infty \text{ for all fixed } \alpha \text{ satisfying } |\alpha| < \alpha^*. \tag{5.3}$$

Beyond the evidence from small values of s, another motivation comes from what is known about the stability of Gauss RK. For these, $R(z)$ is the $[s/s]$ Padé approximant to e^z. These functions have been studied very intensively since they were introduced by Hermite in 1873 and developed in Padé's thesis of 1892 (they are implicit in

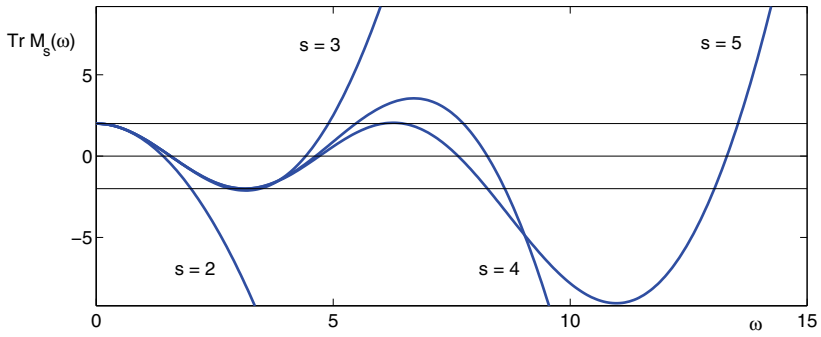

FIG. 5.1. *Plots of* $\operatorname{tr} M_s(\omega)$ *against* ω *for the Lobatto IIIA–IIIB method with stage values* $s = 2, 3, 4, 5$ *when applied to the separable Hamiltonian ODE* (2.11). *The intervals of* ω, *where the trace is between* ± 2, *is the stability region.*

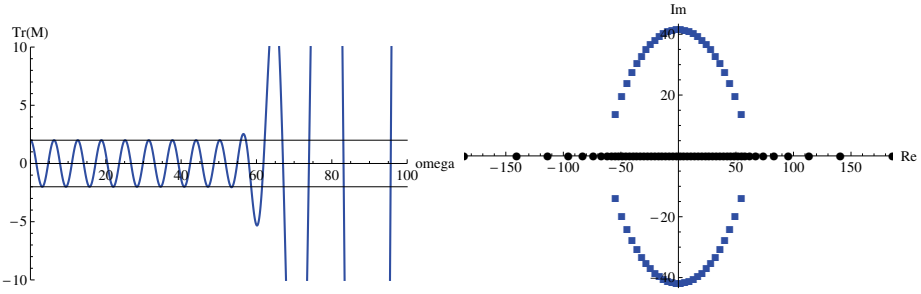

FIG. 5.2. *Plots of* $\operatorname{tr} M_s(\omega)$ *against* ω *(left), and the locations of the poles (boxes) and zeros (dots) of* $\operatorname{tr} M_s(\omega)$ *on the complex plane (right) for the Lobatto IIIA–IIIB method with stage value* $s = 30$ *applied to the separable Hamiltonian ODE* (2.11).

Euler's continued fraction expansion for e given in 1748 (*Introductio in Analysin Infinitorum*, Book 1, Chapter 18). We have $\lim_{s\to\infty} R_s(z) = \mathrm{e}^z$ for all z, and asymptotic expansions of the error $\mathrm{e}^z - R(z)$ are known that are uniformly valid in z. $R_s(\alpha s)$ also converges for all α, but it converges to $\mathrm{e}^{\alpha s}$ only for α inside a certain lens-shaped region. From this one can conclude that $\mathrm{Re} R_s(\mathrm{i}\alpha s) \to \cos(\alpha s)$ for all $|\alpha| < 2$. The poles and zeros of the approximant cluster onto the boundary of the lens-shaped region with known density [4]. Although the approximations in Table 5.1 are *not* Padé approximants, we began our work with the conviction that the Lobatto methods are so naturally defined that there should be some simple interpretation of $\operatorname{tr} M_s(\omega)$ as an approximation to $2\cos\omega$.

6. Known results on stability of Lobatto PRK methods of general order. We review the results of Jay and Petzold [12]. Their key theorem is that *no* member of the Lobatto PRK family is P-stable. They show that the stability function is a rational function whose numerator has degree $2s-2$ and whose denominator has degree $\le 2s-4$, so that $\operatorname{tr} M_s(\infty) = \infty$. First, they use the identity

$$v^\top N^{-1} w = \frac{\det(N + wv^\top)}{\det(N)} - 1 \tag{6.1}$$

valid for any invertible matrix N and any vectors v and w to show that

$$M_s(\omega) = \frac{1}{q(\omega)} \begin{pmatrix} p_{11}(\omega) & p_{12}(\omega) \\ p_{21}(\omega) & p_{22}(\omega) \end{pmatrix} \tag{6.2}$$

with

$$p_{11}(\omega) := \det \begin{pmatrix} I_s & -\omega(A - \mathbf{1}_s b^\top) \\ \omega \hat{A} & I_s \end{pmatrix}, \quad p_{12}(\omega) := \det \begin{pmatrix} I_s & -\omega A \\ \omega \hat{A} & I_s + \omega \mathbf{1}_s b^\top \end{pmatrix} - q, \tag{6.3}$$

$$p_{21}(\omega) := \det \begin{pmatrix} I_s - \omega \mathbf{1}_s \hat{b}^\top & -\omega A \\ \omega \hat{A} & I_s \end{pmatrix} - q, \quad p_{22}(\omega) := \det \begin{pmatrix} I_s & -\omega A \\ \omega(\hat{A} - \mathbf{1}_s \hat{b}^\top) & I_s \end{pmatrix}, \tag{6.4}$$

and

$$q(\omega) := \det \begin{pmatrix} I_s & -\omega A \\ \omega \hat{A} & I_s \end{pmatrix}. \tag{6.5}$$

Second, they use the W-transformation $X := W^\top BAW$, $\widehat{X} := W^\top B\hat{A}W$, where $B = \operatorname{diag}(b_1, \ldots, b_s)$ and $W_{ij} = P_{j-1}(c_i)$ with $P_j(x)$ the jth shifted Legendre polynomial and c_i the nodes of Lobatto quadrature. The W-transformation is extensively used in RK theory; see, e.g., [11]. (Note that in [9], a slightly different, "generalized" W-transformation is used to study PRK methods.)

Thus far, their treatment applies to arbitrary PRK methods. Now, we specialize to Lobatto IIIA–IIIB methods. Let

$$(\widehat{X}_0)_{s-1} := \begin{pmatrix} 1/2 & -\xi_1 & & \\ \xi_1 & 0 & \ddots & \\ & \ddots & \ddots & -\xi_{s-2} \\ & & \xi_{s-2} & 0 \end{pmatrix}, \quad \xi_k := \frac{1}{2\sqrt{4k^2 - 1}}, \tag{6.6}$$

and

$$\beta := (0, 0, \ldots, -\xi_{s-1} u)^\top, \quad u := \sum_{i=1}^{s} b_i P_{s-1}^2(c_i). \tag{6.7}$$

Then it is known [29] that the matrices X, $\widehat{X}$ for Lobatto PRK are given by

$$X := (X_0 \mid 0_s), \quad \widehat{X} := (\widehat{X}_0^\top \mid 0_s)^\top, \tag{6.8}$$

where

$$X_0 := ((\widehat{X}_0)_{s-1}^\top \mid -\beta)^\top, \quad \widehat{X}_0 := ((\widehat{X}_0)_{s-1} \mid \beta). \tag{6.9}$$

Using these, Jay and Petzold [12] show that

$$p_{11}(\omega) = \frac{1}{u} \det \begin{pmatrix} I_{s-1} & \omega \widehat{X}_0 \\ \omega \widehat{X}_0^\top & D_s \end{pmatrix}, \quad p_{22}(\omega) = \frac{1}{u} \det \begin{pmatrix} I_{s-1} & -\omega X_0^\top \\ -\omega X_0 & D_s \end{pmatrix}, \tag{6.10}$$

where

$$D_s := \operatorname{diag}(1, 1, \ldots, u). \tag{6.11}$$

7. Determination of the trace of $M_s(\omega)$. First, we need the value of u in (6.7) for methods based on Lobatto quadrature.

PROPOSITION 7.1.

$$u := \sum_{i=1}^{s} b_i P_{s-1}^2(c_i) = \frac{2s-1}{s-1}. \tag{7.1}$$

The proof is in Appendix A. Now, let

$$Y_{s-1} = \begin{pmatrix} 0 & -\xi_1 & & \\ \xi_1 & 0 & \ddots & \\ & \ddots & 0 & -\xi_{s-2} \\ & & \xi_{s-2} & 0 \end{pmatrix}. \tag{7.2}$$

This is the "regular" part of the matrices appearing in $\operatorname{tr} M_s(\omega)$. Further, we know that $\det(I_{s-1} + zY_{s-1})$ is directly related to the continued fraction expansion of the stability function of high-order RK methods [11, Theorem 5.18]. Therefore, we try to write the stability function in terms of Y_{s-1}.

PROPOSITION 7.2. *Let*

$$U := \frac{1}{u}\beta\beta^\top \tag{7.3}$$

be the $(s-1)\times(s-1)$ *matrix whose only nonzero entry is* $U_{s-1,s-1} = u\xi_{s-1}^2$, *and let*

$$B_s(\omega) := \det\left(I_{s-1} + \omega^2 Y_{s-1}^2 - \omega^2 U\right). \tag{7.4}$$

Then

$$\operatorname{tr} M_s(\omega) = \frac{2\left(1 - \dfrac{\omega^2}{4}\dfrac{B_{s-1}(\omega)}{B_s(\omega)}\right)}{1 + \dfrac{\omega^2}{4}\dfrac{B_{s-1}(\omega)}{B_s(\omega)}}. \tag{7.5}$$

Proof. Applying $\det\left(\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right) = \det(D)\det(A - BD^{-1}C)$ to (6.10) gives

$$p_{11}(\omega) = \det\left(I_{s-1} - \omega^2(\widehat{X}_0)_{s-1}(\widehat{X}_0)_{s-1}^\top - \omega^2 U\right), \tag{7.6}$$

$$p_{22}(\omega) = \det\left(I_{s-1} - \omega^2(\widehat{X}_0)_{s-1}^\top(\widehat{X}_0)_{s-1} - \omega^2 U\right). \tag{7.7}$$

We decompose $(\widehat{X}_0)_{s-1}$ as

$$(\widehat{X}_0)_{s-1} = \frac{1}{2}e_1 e_1^\top + Y_{s-1} \tag{7.8}$$

with $e_1 = (1, 0, \ldots, 0)^\top$, giving

$$\begin{aligned} p_{11}(\omega) &= \det\left(I_{s-1} - \omega^2 U - \omega^2\left(\frac{1}{2}e_1e_1^\top + Y_{s-1}\right)\left(\frac{1}{2}e_1e_1^\top - Y_{s-1}\right)\right) \\ &= \det\left(I_{s-1} + \omega^2 Y_{s-1}^2 - \omega^2 U - \omega^2 Z\right), \end{aligned} \tag{7.9}$$

where Z is the $(s-1)\times(s-1)$ matrix for which all entries are zero except for the top left 2×2 block, which is

$$\begin{pmatrix} \frac{1}{4} & -\frac{1}{2}\xi_1 \\ -\frac{1}{2}\xi_1 & 0 \end{pmatrix}. \tag{7.10}$$

Expanding the determinant in (7.9) along the first row, we get

$$\begin{aligned} p_{11}(\omega) &= \det\left(I_{s-1} + \omega^2 Y_{s-1}^2 - \omega^2 U\right) \\ &\quad - \frac{\omega^2}{4}\det\left(I_{s-2} + \omega^2 Y_{s-2}^2 - \omega^2 U_{s-2}\right), \end{aligned} \tag{7.11}$$

where Y_k (resp., U_k) is the $k\times k$ matrix given by the last k rows and columns of Y_{s-1} (resp., U). Similarly, $p_{22}(\omega)$ can be reformulated to obtain

$$p_{11}(\omega) = p_{22}(\omega). \tag{7.12}$$

Substituting the W-transformation into the expression (6.5) for $q(\omega)$ gives

$$q(\omega) = \det\left(I_{s-1} + \omega^2 (\widehat{X}_0)_{s-1}^2 - \omega^2 U\right),$$

and now using (7.8) and proceeding as for $p_{11}(\omega)$ above, we get

$$q(\omega) = \det\left(I_{s-1} + \omega^2 Y_{s-1}^2 - \omega^2 U\right) + \frac{\omega^2}{4}\det\left(I_{s-2} + \omega^2 Y_{s-2}^2 - \omega^2 U_{s-2}\right). \tag{7.13}$$

Now (6.2), (7.11), (7.12), and (7.13) give the result. ∎

The trace of stability matrix $M_s(\omega)$ is a rational function, and we have the following proposition.

PROPOSITION 7.3. *The roots of* $\operatorname{tr} M_s(\omega)$ *are real.*

Proof. In this proof only, let $W = (\widehat{X}_0)_{s-1}(\widehat{X}_0)_{s-1}^\top + U$. From (7.6) we have

$$p_{11}(\omega) = \det(I_{s-1} - \omega^2 W).$$

Let V_k be the $k\times k$ matrix $V_k := (\widehat{X}_0)_k(\widehat{X}_0)_k^\top$, and let W_k be the $k\times k$ matrix given by the first k rows and columns of W. We have

$$\det(W) = \det(V_{s-1}) + \xi_{s-1}^2 u \det(W_{s-2}). \tag{7.14}$$

Recursively expanding the determinant, we have

$$\begin{aligned} \det(W_{s-2}) &= \det(V_{s-2}) + \xi_{s-2}^2 \det(W_{s-3}), \ldots, \\ \det(W_1) &= \frac{1}{4} + \xi_1^2. \end{aligned} \tag{7.15}$$

Now $\det(V_k) \geq 0$, (7.14), and (7.15) give $\det(W_i) > 0$ $(i = 1, \ldots, s-2)$ and $\det(W) > 0$. Sylvester's criterion implies that matrix W is positive definite; therefore the roots of $p_{11}(\omega)$, namely plus and minus the reciprocals of the square roots of the eigenvalues of W, are real. ∎

Note that the proof implies that the degree of the numerator of $\operatorname{tr} M_s(\omega)$ is exactly $2s-2$, as was already proved in [12].

From the point of view of stability, the following result is more important.

PROPOSITION 7.4. *All zeros of the rational functions* $\operatorname{tr} M_s(\omega) \pm 2$ *are real.*

Proof. From (6.2) and (7.12), we know

$$\operatorname{tr} M_s(\omega) \pm 2 = \frac{2(p_{11}(\omega) \pm q(\omega))}{q(\omega)}, \tag{7.16}$$

where

$$p_{11}(\omega) + q(\omega) = 2 \det \left(I_{s-1} - \omega^2 (Y_{s-1}^\top Y_{s-1} + U) \right), \tag{7.17}$$

$$p_{11}(\omega) - q(\omega) = -\frac{\omega^2}{2} \det \left(I_{s-2} - \omega^2 (Y_{s-2}^\top Y_{s-2} + U_{s-2}) \right) \tag{7.18}$$

are derived by using (7.11) and (7.13). Let $R = Y_{s-1}^\top Y_{s-1} + U$, let R_k (resp., $\widehat{Y}_k$) be the $k \times k$ matrix given by the first k rows and columns of R (resp., Y_{s-1}), and let $Q_k = \widehat{Y}_k^\top \widehat{Y}_k$. Recursively expanding the determinant gives

$$\begin{aligned} \det(R) &= \det(Q_{s-1}) + \xi_{s-1}^2 u \det(R_{s-2}), \\ \det(R_{s-2}) &= \det(Q_{s-2}) + \xi_{s-2}^2 \det(R_{s-3}), \ldots, \\ \det(R_1) &= \xi_1^2. \end{aligned}$$

As in the proof of Proposition 7.3, we discover R is positive definite. A similar calculation shows that $Y_{s-2}^\top Y_{s-2} + U_{s-2}$ is also positive definite, proving the result. ∎

COROLLARY 7.5. $\operatorname{tr} M_s(\omega) - 2$ *and* $\operatorname{tr} M_s(\omega) + 2$ *have precisely* $2s - 2$ *roots. The stability region (in* $\omega \geq 0$*) consists of fewer than the maximum possible number* $s - 1$ *intervals, only if* $\operatorname{tr} M_s(\omega) \pm 2$ *has multiple roots.* $\operatorname{tr} M_s(\omega)$ *has precisely* $2s - 1$ *turning points, none in the interior of the stability region. For each* a *in* $[-1, 1]$*, the equation* $\frac{1}{2} \operatorname{tr} M_s(\omega) = a$ *has precisely* $s - 1$ *solutions. Lobatto PRK provides stable semidiscretizations in the sense of Theorem* 4.1*.*

Proof. $\frac{1}{2} \operatorname{tr} M_s(\omega) - a$ is a rational function of numerator degree $2s - 2$ with no poles on the real axis. For $a = \pm 1$ we know it has $2s - 2$ real zeros. By continuity, it must have exactly this many also when $|a| < 1$. A turning point in $|a| < 1$ would give more zeros, a contradiction. The stability function is even (being a function of ω^2); hence for $a = 1$ and $a = -1$ (values that determine the boundaries of the stability region) there are $s - 1$ solutions in $\omega \geq 0$. This is exactly the stability condition required in Theorem 4.1. ∎

The proof of Proposition 7.3 gives an expression for $\operatorname{tr} M_s(\omega)$ in terms of determinants of 5-diagonal matrices. These could be expanded using a 5-term recurrence relation to get some information for general s. However, knowing that determinants of tridiagonal matrices are much simpler, being related to classical continued fractions, we now express the determinants of 5-diagonal matrices as a product of determinants of tridiagonal matrices, each with a simple structure.

Let

$$\overline{Y}_s = \begin{pmatrix} 0 & -\xi_1 & & & \\ \xi_1 & 0 & \ddots & & \\ & \ddots & 0 & -\xi_{s-2} & \\ & & \xi_{s-2} & 0 & -\xi_{s-1}\sqrt{u} \\ & & & \xi_{s-1}\sqrt{u} & 0 \end{pmatrix}. \tag{7.19}$$

Using the notation of Proposition 7.2, we have the following proposition.

PROPOSITION 7.6. $B_s(\omega) = \det(I_{s-1} + \mathrm{i}\omega Y_{s-1}) \det(I_s + \mathrm{i}\omega \overline{Y}_s)$.

Proof. With $I_{s-1} + \omega^2 Y_{s-1}^2 = (I_{s-1} + \mathrm{i}\omega Y_{s-1})(I_{s-1} - \mathrm{i}\omega Y_{s-1})$ and $\det(I_{s-1} + \mathrm{i}\omega Y_{s-1}) = \det(I_{s-1} - \mathrm{i}\omega Y_{s-1})$, we derive

$$\begin{aligned} B_s(\omega) &= \det\left(I_{s-1} + \omega^2 Y_{s-1}^2 - \omega^2 U\right) \\ &= \det(I_{s-1} + \omega^2 Y_{s-1}^2) \det\left(I_{s-1} - \omega^2 U (I_{s-1} + \omega^2 Y_{s-1}^2)^{-1}\right) \\ &= \det(I_{s-1} + \mathrm{i}\omega Y_{s-1}) \det\left(I_{s-1} - \mathrm{i}\omega Y_{s-1} - \omega^2 U (I_{s-1} + \mathrm{i}\omega Y_{s-1})^{-1}\right). \end{aligned} \tag{7.20}$$

Letting $F = I_{s-1} + \mathrm{i}\omega Y_{s-1}$, and expanding (7.20) along the last row, we have

$$\begin{aligned} \det\left(I_{s-1} - \mathrm{i}\omega Y_{s-1} - \omega^2 U F^{-1}\right) &= f_{s-1,s-1} - \mathrm{i}\omega \xi_{s-2} f_{s-2,s-1} \\ &\quad - \frac{\omega^2 \xi_{s-1}^2 u}{\det(F)} \left(f_{s-1,s-1}^2 + f_{s-2,s-1}^2 + \cdots + f_{1,s-1}^2\right), \end{aligned} \tag{7.21}$$

where $f_{i,j}$, $i, j = 1, \ldots, s-1$, are the elements of $\mathrm{adj}(F)$, the adjugate matrix of F.[2] Let F_k be the $k \times k$ matrix given by the first k rows and columns of F. Recursively expanding the $f_{i,s-1}$ gives

$$\begin{aligned} f_{s-1,s-1} &= \det(F_{s-2}), \\ f_{s-2,s-1} &= -\mathrm{i}\omega \xi_{s-2} \det(F_{s-3}), \ldots, \\ f_{1,s-1} &= (-\mathrm{i})^{s-2} \omega^{s-2} \xi_1, \ldots, \xi_{s-2}; \end{aligned} \tag{7.22}$$

therefore

$$\begin{aligned} &f_{s-1,s-1}^2 + f_{s-2,s-1}^2 + \cdots + f_{1,s-1}^2 \\ &= \det(F_{s-2})^2 - \omega^2 \xi_{s-2}^2 \det(F_{s-3})^2 + \cdots + \omega^{2s-4} (-1)^{s-2} \xi_1^2, \ldots, \xi_{s-2}^2. \end{aligned} \tag{7.23}$$

We calculate the first two terms of (7.23) and have

$$\begin{aligned} &\det(F_{s-2})^2 - \omega^2 \xi_{s-2}^2 \det(F_{s-3})^2 \\ &= \left(\det(F_{s-2}) + \omega^2 \xi_{s-2}^2 \det(F_{s-3})\right) \left(\det(F_{s-2}) - \det(F_{s-3})\right) \\ &\quad + (1 - \omega^2 \xi_{s-2}^2) \det(F_{s-2}) \det(F_{s-3}). \end{aligned} \tag{7.24}$$

For $\det(F_i)$ $(2 \le i \le s-2)$ and $F_{s-1} = F$, we have the following recursion:

$$\det(F_{i+1}) = \det(F_i) - \omega^2 \xi_i^2 \det(F_{i-1}). \tag{7.25}$$

By using (7.25), it follows from (7.24) that

$$\begin{aligned} &\det(F_{s-2})^2 - \omega^2 \xi_{s-2}^2 \det(F_{s-3})^2 \\ &= \det(F_{s-2}) \det(F_{s-1}) - \omega^4 \xi_{s-2}^2 \xi_{s-3}^2 \det(F_{s-3}) \det(F_{s-4}). \end{aligned} \tag{7.26}$$

Substituting (7.22), (7.23), and (7.26) into (7.21), recursively, we obtain

$$\begin{aligned} &\det\left(I_{s-1} - \mathrm{i}\omega Y_{s-1} - \omega^2 U F^{-1}\right) \\ &= \det(F_{s-2}) - \omega^2 \xi_{s-2}^2 \det(F_{s-3}) - \omega^2 \xi_{s-1}^2 u \det(F_{s-2}) \\ &= \det(F_{s-1}) - \omega^2 \xi_{s-1}^2 u \det(F_{s-2}) \\ &= \det(I + \mathrm{i}\omega \overline{Y}_s), \end{aligned} \tag{7.27}$$

[2] $\mathrm{adj}(F) = \det(F) F^{-1}$ is the transpose of the cofactors (signed minors) of F, i.e., $(-1)^{i+j} f_{i,j}$ is the determinant of the matrix given by F with row j and column i deleted.

where the last equality holds by determinant expansion along the last row and column. Combining (7.20) and the above equality gives the result. ∎

In light of Proposition 7.2 (especially (7.5)) and Proposition 7.6, it makes sense to focus on $(1-\frac{1}{2}\operatorname{tr} M_s(\omega))/(1+\frac{1}{2}\operatorname{tr} M_s(\omega)) = (\omega^2 B_{s-1}(\omega))/(4B_s(\omega))$. Note that $(1-\cos\omega)/(1+\cos\omega) = \tan^2\frac{1}{2}\omega$. In order to relate our approximation to standard continued fractions we evaluate at 2ω instead of ω.

DEFINITION 7.7. *Let*

$$G_s(\omega) := \frac{1-\frac{1}{2}\operatorname{tr} M_s(2\omega)}{1+\frac{1}{2}\operatorname{tr} M_s(2\omega)} \quad (\approx \tan^2\omega). \tag{7.28}$$

From Proposition 7.6, we have

$$\begin{aligned} G_s(\omega) &= \omega^2 \frac{B_{s-1}(2\omega)}{B_s(2\omega)} \\ &= \omega^2 \frac{\det(I_{s-2}+2\mathrm{i}\omega Y_{s-2})}{\det(I_{s-1}+2\mathrm{i}\omega Y_{s-1})} \frac{\det(I_{s-1}+2\mathrm{i}\omega \overline{Y}_{s-1})}{\det(I_s+2\mathrm{i}\omega \overline{Y}_s)}, \end{aligned} \tag{7.29}$$

where $\overline{Y}_{s-1}$ is the $(s-1)\times(s-1)$ matrix given by the last $s-1$ rows and columns of $\overline{Y}_s$.

THEOREM 7.8. *For the s-stage Lobatto IIIA–IIIB method, $G_s(\omega)$ has the expression*

$$G_s(\omega) = C_s(\omega)\widehat{C}_s(\omega), \tag{7.30}$$

where $C_s(\omega) = \omega\frac{\det(I_{s-2}+2\mathrm{i}\omega Y_{s-2})}{\det(I_{s-1}+2\mathrm{i}\omega Y_{s-1})}$ and $\widehat{C}_s(\omega) = \omega\frac{\det(I_{s-1}+2\mathrm{i}\omega \overline{Y}_{s-1})}{\det(I_s+2\mathrm{i}\omega \overline{Y}_s)}$ are two rational functions which have the finite continued fraction expansions

$$C_s(\omega) = \frac{\omega|}{|1} - \frac{\omega^2|}{|3} - \frac{\omega^2|}{|5} - \cdots \frac{\omega^2|}{|2s-3}, \tag{7.31}$$

$$\widehat{C}_s(\omega) = \frac{\omega|}{|1} - \frac{\omega^2|}{|3} - \frac{\omega^2|}{|5} - \cdots \frac{\omega^2|}{|2s-3} - \frac{\omega^2|}{|s-1}. \tag{7.32}$$

Proof. From Theorem 5.18 of [11], we have

$$\omega\frac{\det(I_{s-2}+2\mathrm{i}\omega Y_{s-2})}{\det(I_{s-1}+2\mathrm{i}\omega Y_{s-1})} = \frac{\omega|}{|1} - \frac{4\omega^2\xi_1^2|}{|1} - \cdots - \frac{4\omega^2\xi_{s-2}^2|}{|1}. \tag{7.33}$$

Recall that $4\xi_k^2 = 1/((2k-1)(2k+1))$. Multiplying the second numerator and denominator by $1/4\xi_1^2 = 3$, the third numerator and denominator by $4\xi_1^2/4\xi_2^2 = 5$, and the kth numerator and denominator by $2k-1$ yields the equivalent continued fraction representation (7.31).

For $\widehat{C}_s(\omega)$, we have the similar continued fraction with the final numerator adjusted by $4\omega^2 u\xi_{s-1}^2(2s-3) = \omega^2/(s-1)$. In the calculation, $u = (2s-1)/(s-1)$ has been used by Proposition 7.1; therefore the final numerator is ω^2 and the final denominator is $s-1$, giving (7.32). ∎

Such continued fractions are Padé approximants of the functions with the same Maclaurin series, typically giving a "staircase" sequence of approximants formed from the diagonal and sub- or superdiagonal elements of the Padé table. In this case, because $\tan\omega$ is an odd function,[3] $C_s(\omega)$ is the Padé approximant to $\tan\omega$ of type

[3] "If functions are even or odd, they are degenerate in a rather trivial way, and there is no purpose in making a great issue of this" [4, sect. 4.2].

$[s-1/s-1]$, which is actually of degree $[s-2/s-1]$ if s is odd, and of degree $[s-1/s-2]$ if s is even. The continued fraction of $\tan\omega$ is $\tan(\omega) = \frac{\omega|}{|1} - \frac{\omega^2|}{|3} - \frac{\omega^2|}{|5} - \frac{\omega^2|}{|7} - \cdots$, whose partial sums converge to $\tan\omega$ for all $\omega \neq (2k+1)\pi/2$, $k \in \mathbb{Z}$ [4, eq. (4.6.2)].

Further, the continued fractions in Theorem 7.8 can be expressed in closed form as follows.

THEOREM 7.9. *Let*

$$[n/n]_{\mathrm{e}^z}(z) = \frac{A_n(z)}{A_n(-z)} = \mathrm{e}^z - V_n(z) \tag{7.34}$$

be the diagonal Padé approximants to e^z *whose numerators are given explicitly by* [4, eq. (1.2.12)]

$$A_n(z) := {}_1F_1(-n,-2n,z) = \sum_{k=0}^{n} \frac{(-n)_k}{(-2n)_k k!} z^k = \sum_{k=0}^{n} \frac{n!(2n-k)!}{(n-k)!(2n)!k!} z^k, \tag{7.35}$$

and let

$$\begin{aligned} P_n(\omega) &:= A_n(2\mathrm{i}\omega) - A_n(-2\mathrm{i}\omega), \\ Q_n(\omega) &:= \mathrm{i}(A_n(2\mathrm{i}\omega) + A_n(-2\mathrm{i}\omega)). \end{aligned} \tag{7.36}$$

We have

$$C_s(\omega) = \frac{P_{s-1}(\omega)}{Q_{s-1}(\omega)}, \tag{7.37}$$

$$\widehat{C}_s(\omega) = \frac{(2s-1)P_s(\omega) - sP_{s-1}(\omega)}{(2s-1)Q_s(\omega) - sQ_{s-1}(\omega)}, \tag{7.38}$$

and

$$\operatorname{tr} M_s(2\omega) = 2\frac{1 - C_s(\omega)\widehat{C}_s(\omega)}{1 + C_s(\omega)\widehat{C}_s(\omega)}. \tag{7.39}$$

Proof. The homographic invariance of value transformations [4, Theorem 1.5.3], states that taking diagonal Padé approximants commutes with taking linear fractional transformations of function values $f(z) \mapsto (a+bf(z))/(c+df(z))$, provided $c+df(0) \neq 0$. The homographic invariance under argument transformations [4, Theorem 1.5.2] takes that diagonal Padé approximants commute with taking origin-preserving linear fractional transformations of arguments, i.e., $z \mapsto \alpha z/(1+\beta z)$. Using $\tan\omega = \frac{\mathrm{e}^{2\mathrm{i}\omega}-1}{\mathrm{i}(\mathrm{e}^{2\mathrm{i}\omega}+1)}$ and both types of transformations with $\alpha = 2\mathrm{i}$, $\beta = 0$, $a = -1$, $b = 1$, and $c = d = \mathrm{i}$ gives (7.37).

Expanding the determinants by the last row and column gives

$$\det(I_s + 2\mathrm{i}\omega\overline{Y}_s) = \det(I_{s-1} + 2\mathrm{i}\omega Y_{s-1}) - 4\omega^2\xi_{s-1}^2 u \det(I_{s-2} + 2\mathrm{i}\omega\widehat{Y}_{s-2}). \tag{7.40}$$

Similarly,

$$\det(I_s + 2\mathrm{i}\omega Y_s) = \det(I_{s-1} + 2\mathrm{i}\omega Y_{s-1}) - 4\omega^2\xi_{s-1}^2 \det(I_{s-2} + 2\mathrm{i}\omega\widehat{Y}_{s-2}), \tag{7.41}$$

where $\widehat{Y}_{s-2}$ is the $(s-2)\times(s-2)$ matrix given by the first $s-2$ rows and columns of Y_{s-1}. Eliminating $\det(I_{s-2} + 2\mathrm{i}\omega\widehat{Y}_{s-2})$ between these equations gives

$$\det(I_s + 2\mathrm{i}\omega\overline{Y}_s) = u\det(I_s + 2\mathrm{i}\omega Y_s) + (1-u)\det(I_{s-1} + 2\mathrm{i}\omega Y_{s-1}), \tag{7.42}$$

and similarly, we have

$$\det(I_{s-1} + 2\mathrm{i}\omega\overline{Y}_{s-1}) = u\det(I_{s-1} + 2\mathrm{i}\omega Y_{s-1}) + (1-u)\det(I_{s-2} + 2\mathrm{i}\omega Y_{s-2}), \tag{7.43}$$

where $u = (2s-1)/(s-1)$, and $\overline{Y}_{s-1}$ is the $(s-1)\times(s-1)$ matrix given by the last rows and columns of $\overline{Y}_s$. It is known from (7.35) that

$$A_{n+1}(z) = A_n(z) + \frac{z^2}{4(2n-1)(2n+1)}A_{n-1}(z), \tag{7.44}$$

and then from (7.36) and (7.44) we know $Q_n(\omega)$ satisfies the recurrence

$$Q_{n+1}(\omega) = Q_n(\omega) - \frac{\omega^2}{(2n-1)(2n+1)}Q_{n-1}(\omega) \tag{7.45}$$

with initial values $Q_1(\omega) = 2\mathrm{i}$, $Q_2(\omega) = 2\mathrm{i}(1-\omega^2/3)$. From (7.41) and (7.45), as $\xi_{s-1} = 1/(2\sqrt{(2s-3)(2s-1)})$, we know $\det(I_{s-1} + 2\mathrm{i}\omega Y_{s-1})$ satisfies the same recurrence as $Q_{s-1}(\omega)$ (7.45) with initial values $\det(I_1 + 2\mathrm{i}\omega\widehat{Y}_1) = 1$, $\det(I_2 + 2\mathrm{i}\omega\widehat{Y}_2) = 1-\omega^2/3$. Therefore,

$$\det(I_{s-1} + 2\mathrm{i}\omega Y_{s-1}) = \frac{1}{2\mathrm{i}}Q_{s-1}(\omega), \tag{7.46}$$

and similarly,

$$\det(I_{s-2} + 2\mathrm{i}\omega Y_{s-2}) = \frac{1}{2\mathrm{i}\omega}P_{s-1}(\omega). \tag{7.47}$$

Multiplying (7.42) and (7.43) by $s-1$ gives (7.38). Equation (7.39) follows from (7.28) and (7.30) and is just included for completeness. ∎

Recall $q(\omega)$ is the denominator of the trace of stability matrix, providing the information on poles of $\operatorname{tr} M_s(\omega)$. For the expression of $q(\omega)$ we have the following proposition.

PROPOSITION 7.10. *Let*

$$F_{l,s} := 2\mathrm{i}\frac{s!(2s-l)!2^l}{2s!(s-l)!l!}.$$

Then $q(\omega)$ can be expanded as

$$q(\omega) = \sum_{m=0}^{s-2}\left(-\frac{1}{4}\right)^{m+1} q_m\omega^{2m}, \tag{7.48}$$

where

$$q_m = \sum_{i=0}^{2m}(-1)^i F_{2m-i,s-1}\left(uF_{i,s} + (1-u)F_{i,s-1}\right) \tag{7.49}$$

when $0 \le m \le [(s-1)/2]$;

$$\begin{aligned} q_m = &\sum_{i=s}^{2m-s+1}(-1)^i F_{2m-i,s-1}\left(uF_{i,s} + (1-u)F_{i,s-1}\right) \\ &+(-1)^{s-1}(1-u)F_{2m-s+1,s-1}F_{s-1,s-1} \end{aligned} \tag{7.50}$$

when $[(s+1)/2] \le m \le s-2$.

Proof. By Proposition 7.6, and using (7.46), (7.47), $q(\omega)$ in (7.13) can be rewritten as

$$q(\omega) = -\frac{1}{4}Q_{s-1}(\omega/2)\Big(uQ_s(\omega/2) + (1-u)Q_{s-1}(\omega/2)\Big) - \frac{1}{4}P_{s-1}(\omega/2)\Big(uP_s(\omega/2) + (1-u)P_{s-1}(\omega/2)\Big). \tag{7.51}$$

It follows from (7.35) that $Q_s(\omega)$ and $P_s(\omega)$ can be expanded as, respectively,

$$Q_s(\omega) = \sum_{m=0}^{[s/2]} (-1)^l F_{2l,s}\omega^{2m}, \tag{7.52}$$

$$P_s(\omega) = \sum_{m=1}^{[(s+1)/2]} (-1)^{l+1}\frac{(2s+1-2l)l}{s+1-2l}F_{2l,s}\omega^{2m-1}, \tag{7.53}$$

where $F_{l,s}$ is as defined in the proposition. Substituting (7.52) and (7.53) into (7.51) gives the result. ∎

In [12], Jay and Petzold proved that the degree of $q(\omega) \le 2s-4$.

PROPOSITION 7.11. *The degree of $q(\omega)$ is equal to $2s-4$, and the coefficient in $q(\omega)$ of ω^{2s-4} is*

$$\frac{8s}{4^s(2s-3)!!^2}.$$

Proof. Using (7.48) and (7.50) gives the result. ∎

8. Asymptotic behavior of the trace of $M_s(\omega)$ and the stability region. We now study the asymptotic behavior $\operatorname{tr} M_s(\omega)$ in four different regimes.

8.1. Asymptotics for s fixed and $\omega \to 0$.

PROPOSITION 8.1. *For fixed s and $\omega \to 0$, we have*

$$\cos\omega - \frac{1}{2}\operatorname{tr} M_s(\omega) = \frac{1}{2s-2}e_{s-1}\omega^{2s} + \mathcal{O}(\omega^{2s+2}), \tag{8.1}$$

where

$$e_s := \frac{s!^2}{(2s)!(2s+1)!}. \tag{8.2}$$

Proof. For fixed n and $z \to 0$, we have [11]

$$\mathrm{e}^z - [n/n]_{\mathrm{e}^z}(z) = (-1)^n e_n z^{2n+1} + \mathcal{O}(z^{2n+2}). \tag{8.3}$$

Substituting into the above expression for C, $\widehat{C}$ gives

$$\tan\omega - C_s(\omega) = 4^{s-1}e_{s-1}\omega^{2s-1} + \mathcal{O}(\omega^{2s}) \tag{8.4}$$

and

$$\tan\omega - \widehat{C}_s(\omega) = -\frac{s}{s-1}4^{s-1}e_{s-1}\omega^{2s-1} + \mathcal{O}(\omega^{2s}); \tag{8.5}$$

note that the error in $\widehat{C}$ is opposite in sign and slightly larger in magnitude to that in C, so that together they nearly cancel. Combining these errors gives the error estimate (8.1). ∎

It is striking that the leading error coefficient is actually smaller (by a factor $\frac{1}{2s-2}$) than that of the Gauss RK method of the same order.

8.2. Asymptotics for s fixed and $\omega \to \infty$. The representation (7.39) is perfect for determining the boundary of the stability region. If $C_s(\omega) = 0$ or $\widehat{C}_s(\omega) = 0$, then $\operatorname{tr} M_s(2\omega) = 2$, while if $C_s(\omega)$ or $\widehat{C}_s(\omega)$ has a pole at $\omega = \omega^*$, then $\operatorname{tr} M_s(2\omega^*) = -2$. On the other hand, (7.39) is not so suitable for determining the zeros and poles of $\operatorname{tr} M_s(\omega)$ itself, as these involve the product $C_s(\omega)\widehat{C}_s(\omega)$.

First, we consider the part of the boundary of the stability region determined by zeros and poles of $C_s(\omega)$. From (7.36) and (7.37), $C_s(\omega)$ has a zero (resp., pole) if $[n/n]_{\mathrm{e}^z}(2\mathrm{i}\omega) = 1$ (resp., -1), where $n = s - 1$. Recall that as ω increases from 0 to ∞, the argument of this Padé approximant increases from 0 to $n\pi$. Thus, apart from the trivial zero at $\omega = 0$, this Padé approximant takes on the value ± 1 $n - 1$ times. The asymptotic behavior of $[n/n]_{\mathrm{e}^z}$ as $n \to \infty$ for fixed z and for large z can be used to give precise asymptotics of the boundary of the stability region.

We let $n = s - 1$, let $m = n(n+1) = s(s-1)$, and let the stability boundaries be $\omega_1, \omega_2, \ldots, \omega_{2s-3}$.

Expanding in a Taylor series about $z = \infty$ for fixed n using (7.35), (7.36), and (7.37) gives

$$\log[n/n]_{\mathrm{e}^z}(z) = n\mathrm{i}\pi + \frac{2m}{z} - \frac{2m(m-6)}{3z^3} + \mathcal{O}(z^{-5}). \tag{8.6}$$

The series has a finite radius of convergence that is $\mathcal{O}(n)$ as $n \to \infty$. Reverting the series gives that $\log[n/n]_{\mathrm{e}^z}(z) = w$ at

$$z = \frac{2m}{w - n\mathrm{i}\pi} + \left(\frac{1}{m} - \frac{1}{6}\right)(w - n\mathrm{i}\pi) + \frac{m^2 - 48m + 90}{45m^3}(w - n\mathrm{i}\pi)^3 + \mathcal{O}\left((w - n\mathrm{i}\pi)^5\right). \tag{8.7}$$

Evaluating at $w = (n-1)\mathrm{i}\pi$ gives the last stability boundary arising from C at

$$\omega_{2s-4} = \frac{2m}{\pi} + \left(\frac{1}{6} - \frac{1}{m}\right)\pi + \frac{m^2 - 48m + 90}{45m^3}\pi^3 + \mathcal{O}(m^{-1}). \tag{8.8}$$

A similar approach for the stability boundaries arising from $\widehat{C}$ gives a stability boundary at

$$\omega_{2s-3} = \frac{2m}{\pi} + \left(\frac{1}{6} + \frac{1}{m}\right)\pi + \frac{m^2 + 57m - 180}{45m^3}\pi^3 + \mathcal{O}(m^{-1}). \tag{8.9}$$

At these values of ω (or z), the kth term in the Taylor series (8.6) is $\mathcal{O}(n^{-k})$, which justifies the use of the series. Similarly, evaluating (8.7) at $w = (n-k)\mathrm{i}\pi$ gives the good estimate $\omega_{2s-2-k} \approx \frac{2m}{k\pi} + \frac{k\pi}{6}$ of the kth from the last stability boundary for fixed k.

This approach is very simple, using only Taylor series, but it does not give the coefficient of m^{-1}. (The series (8.9) converges up to the pole nearest ∞, and we know that as $n \to \infty$ this is near $z = 2\mathrm{i}n$.) However, numerically, including the first two $\mathcal{O}(m^{-1})$ terms (as above) does give a very precise estimate of the stability boundary. For example, the leading order estimate $2m/\pi + \pi/6$ gives 19.622 at $s = 6$, while (8.9) gives 19.789, and the actual stability limit is 19.798.

8.3. Asymptotics for $s \to \infty$, ω fixed. For fixed z and $n \to \infty$, we have the estimate [17, eq. II.14.2.13]

$$V_n(z) = (-1)^n e_n z^{2n+1} \mathrm{e}^{z + \frac{z^2}{8n+4}} \left(1 + \mathcal{O}(n^{-3})\right). \tag{8.10}$$

A Taylor expansion of $C(\omega/2)$ at $\omega = \pi$ gives that it has a pole (and hence $\operatorname{tr} M_s(\omega)$ has a stability boundary) at

$$\omega_2 = \pi + e_n \pi^{2n+1} \mathrm{e}^{\frac{-\pi^2}{8n+4}} \left(1 + \mathcal{O}(n^{-3})\right). \tag{8.11}$$

This estimate is already quite accurate for small n. For $n = 3$ (hence $s = 4$) it gives a stability boundary at $\omega \approx \pi + 0.0211$; the actual value is $\pi + 0.0207$. As $n \to \infty$ the dominant behavior is

$$\omega_2 \sim \pi + \frac{1}{\mathrm{e}} \left(\frac{\mathrm{e}\pi}{4n}\right)^{2n+1}. \tag{8.12}$$

The boundary of the kth stability region (for fixed k as $s \to \infty$) can be estimated similarly; its distance from $k\pi$ increases according to the order of the method, i.e., as k^{2n+1}.

A similar approach yields an estimate of the boundary of the stability region due to $\widehat{C}(\omega)$. However, because of the form of (7.38), we need the asymptotic expansion of $A_n(z)$ itself rather than just that of the Padé approximants $A_n(z)/A_n(-z)$. Equation (7.34) determines $A_n(z)$ up to multiplication by an even function of z. On the other hand, for z fixed and $n \to \infty$, we have [16, eq. I.4.8.16-19]

$$\mathrm{e}^{-z/2} A_n(z) \sim \sum_{k=0}^{\infty} (-1)^k d_k(z) (n + \tfrac{1}{2})^{-k}, \tag{8.13}$$

where $d_0(z) = 1$ and

$$8 d_{k+1}(z) = -4z d_k'(z) + \int_0^z t d_k(t)\, \mathrm{d}t, \tag{8.14}$$

i.e., $d_1(z) = \frac{z^2}{16}$, $d_2(z) = \frac{z^4}{512} - \frac{z^2}{16}$. The $d_k(z)$ are all even and provide the overall behavior of $\mathrm{e}^{z/2} A_n(z)$, but this approximation alone substituted in (7.38) simply yields the approximation $\widehat{C}_s(\omega) \sim \tan\omega$, which does not provide an error estimate. Therefore, we combine (7.34), (8.10), and (8.13) to get

$$\mathrm{e}^z - \frac{A_n(z) \mathrm{e}^{z/2}}{\sum_{k=0}^{2} d_k(z) (n + \frac{1}{2})^{-k}} \sim V_n(z) \left(1 + \mathcal{O}(n^{-3})\right), \tag{8.15}$$

where the $V_n(z)$ provides the exponentially small error terms that are subdominant to any finite term in the expansion (8.13). Now substituting (8.15), (8.10) into (7.36), (7.38) and proceeding as for (8.11) gives a stability boundary arising from $\widehat{C}_s(\omega)$ at

$$\omega_1 = \pi - (1 + \tfrac{1}{n}) e_n \pi^{2n+1} \mathrm{e}^{\frac{-\pi^2}{8n+4}} \left(1 + \mathcal{O}(n^{-3})\right). \tag{8.16}$$

The unstable region (ω_1, ω_2) is nearly centered on π.

From the preceeding asymptotics we now get that $C_s(\omega) \to \tan\omega$ and $\widehat{C}_s(\omega) \to \tan\omega$ for all $\omega \neq (2k+1)\pi$ and $1/C_s(\omega) \to \cot\omega$, $1/\widehat{C}_s(\omega) \to \cot\omega$ for all $\omega \neq 2k\pi$; combining, we have the following theorem.

THEOREM 8.2.

$$\operatorname{tr} M_s(\omega) \to 2\cos\omega \tag{8.17}$$

for all fixed ω as $s \to \infty$.

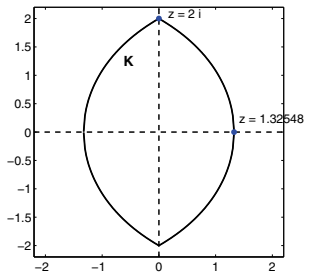

FIG. 8.1. *The poles and zeros of $R_n(\nu z)$ cluster along* $\mathbf{K}$*, and the relative error of this approximation to* $\mathrm{e}^{\nu z}$ *tends to zero only inside* $\mathbf{K}$.

8.4. Uniform asymptotics for $s \to \infty$. Let $R_n(z) := [n/n]_{\mathrm{e}^z}(z)$ be the diagonal Padé approximants to the exponential function. A great deal is known about these functions. As we have seen, they converge to e^z for all fixed z. However, if z is not fixed but is $\mathcal{O}(n)$, then they have a different asymptotic behavior. In particular, the zeros and poles of $R_n(nz)$ cluster along the boundary of an eye-shaped region (Figure 8.1) that we call $\mathbf{K}$ following Olver [21] who (as far as we have been able to determine) was the first to identify it. (It appears in his Figure 3, scaled by $\frac{i}{2}$; see also Figure 8.3 of [22].) The boundary of the eye $\mathbf{K}$ is given by $\partial\mathbf{K} := \zeta^{-1}\left(\left[\frac{-\mathrm{i}\pi}{2}, \frac{\mathrm{i}\pi}{2}\right]\right)$ together with its reflection in the imaginary axis, where

$$\zeta(z) = \sqrt{1 + \frac{z^2}{4}} + \ln \frac{z}{2\left(1 + \sqrt{1 + \frac{z^2}{4}}\right)}, \tag{8.18}$$

with branches chosen so that ζ is real when z is real and ζ is continuous in $|\arg z| \le \frac{\pi}{2}$. $\partial\mathbf{K}$ intersects the imaginary axis at $z = 2\mathrm{i}$, although $R_n(nz)$ has no zeros or poles actually on the imaginary axis. Despite an extensive literature on the behavior of $R_n(z)$ and associated special functions, we could not find in the literature a precise statement of its actual asymptotic behavior. We therefore develop it here.

PROPOSITION 8.3. *Let* $\mathbf{D}$ *be the quadrant* $0 \le \arg z \le \frac{\pi}{2}$. *We have the asymptotic relation*

$$R_n(\nu z) \sim \begin{cases} \mathrm{e}^{\nu z}, & z \text{ inside } \mathbf{K}, \\ (-1)^n \mathrm{e}^{\nu(z - 2\zeta)}, & z \text{ outside } \mathbf{K} \end{cases} \tag{8.19}$$

uniformly in z for any $z \in \mathbf{D}$, where $\nu = n + \frac{1}{2}$. For z in other quadrants, the asymptotic behavior of R_n can be obtained using $R_n(\overline{z}) = \overline{R_n(z)}$ and $R_n(-z) = 1/R_n(z)$.

The proof is in Appendix B.

For large z, we have

$$z - 2\zeta(z) \sim \frac{2}{z} - \frac{2}{3z^3} + \frac{4}{5z^5} - \cdots. \tag{8.20}$$

From Theorem 7.9, the stability boundaries are given by the poles and zeros of $C_s(\omega)$ and $\widehat{C}_s(\omega)$. From (7.36) and (7.37), the stability boundaries due to $C_s(\omega)$ are the points where $R_{n-1}(\mathrm{i}\omega) = \pm 1$. The asymptotic behavior of $R_n(z)$ found above hence explains the observed splitting of the stability domains into two sets, one on

which $\operatorname{tr} M_s(\omega) \approx 2\cos\omega$ with stability boundaries near integer multiples of π, and one in which the stability boundaries grow more rapidly.

In particular, there is a stability boundary (asymptotically) at $k\pi$ for $k = 1, \ldots, k^*$ and at $\nu g^{-1}(-\mathrm{i}k\pi/\nu)/\mathrm{i}$, $k = 1, \ldots, n-1-k^*$, where $g(z) = z - 2\zeta(z)$ and $k^* = \lfloor 2\nu/\pi \rfloor$ is determined by z lying inside or outside $\mathbf{K}$.[4]

The stability boundaries corresponding to $\widehat{C}$ can be found as follows. First, note from (7.36) we need only the behavior of A_n on the imaginary axis. Basic trigonometry gives us that

$$C_s(\omega) = \tan\theta_s(\omega), \quad \theta_s(\omega) := \arg A_n(2\mathrm{i}\omega), \tag{8.21}$$

and

$$\widehat{C}_s(\omega) = \tan\hat{\theta}_s(\omega), \quad \hat{\theta}_s(\omega) := \arg\left(uA_{n+1}(2\mathrm{i}\omega) + (1-u)A_n(2\mathrm{i}\omega)\right), \tag{8.22}$$

where $u = (2s-1)/(s-1)$. A Taylor expansion of $A_{n+1}(\nu z)/A_n(\nu z)$ for large n using (B.1), (B.5) gives

$$uA_{n+1}(\nu z) + (1-u)A_n(\nu z) = \varphi(n,z)A_n(\nu z), \tag{8.23}$$

where

$$\varphi(n,z) \sim \sqrt{1 + \frac{z^2}{4}} + \mathcal{O}\left(n^{-1}\right) \tag{8.24}$$

for z on the positive imaginary axis bounded away from 2i. Thus, as $n \to \infty$,

$$\begin{aligned} \hat{\theta}_s(\nu\omega) &= \theta_s(\nu\omega) + \arg\varphi(n,z) \\ &\sim \theta_s(\nu\omega) + \begin{cases} 0, & 0 \le \omega < 1, \\ \frac{\pi}{2}, & \omega > 1. \end{cases} \end{aligned} \tag{8.25}$$

The stability boundaries due to $\widehat{C}_s$ are located where $\hat{\theta}_s = k\pi/2$, $k \in \mathbb{Z}$, and so coincide (to this order of approximation) with those due to C_s.

Substituting (8.21), (8.22) into (7.39) gives

$$\operatorname{tr} M_s(2\omega) = \frac{\cos(\theta_s(\omega) + \hat{\theta}_s(\omega))}{\cos(\theta_s(\omega) - \hat{\theta}_s(\omega))}, \tag{8.26}$$

which, together with the asymptotic expansions above, yields our final desired result.

THEOREM 8.4. *For all $0 \le \omega < 2$ we have*

$$\operatorname{tr} M_s(\nu\omega) \to 2\cos(\nu\omega) \tag{8.27}$$

as $s \to \infty$, where $\nu = s + \frac{1}{2}$.

[4]For $n = 5$ (i.e., $s = 6$; see Table 5.2) this gives stability boundaries at π, 2π, 3π (here we pass $\mathbf{K}$), and 19.8062 (not quite as accurate as the simpler approximations of section 8.2). For $n = 10$, it gives stability boundaries at $\pi, 2\pi, \ldots, 6\pi \approx 18.9$ (here we pass $\mathbf{K}$), 25.18, 36.20, and 70.72, while the exact boundaries corresponding to C are at $(1+8.25\times10^{-16})\pi$, $(2+1.21\times10^{-9})\pi$, $(3+3.24\times10^{-6})\pi$, 4.00055π, 5.017π, 6.176π, 24.88, 36.03, and 70.53.

Appendix A. Proof of Proposition 7.1.

THEOREM A.1. *Let*

$$P_{s-1}(x) = \frac{\sqrt{2s-1}}{(s-1)!}\frac{\mathrm{d}}{\mathrm{d}x}\frac{\mathrm{d}^{s-2}}{\mathrm{d}x^{s-2}}\left(x^{s-1}(x-1)^{s-1}\right) \tag{A.1}$$

be the shifted Legendre polynomial of degree $s-1$. Let $c_1,\dots,c_s$ be the roots of the Lobatto polynomial

$$P_{\text{Lobatto}}(x) = \frac{\mathrm{d}^{s-2}}{\mathrm{d}x^{s-2}}\left(x^{s-1}(x-1)^{s-1}\right), \tag{A.2}$$

and let $b_1,\dots,b_s$ be the weights such that the following simplifying assumption is satisfied:

$$B(2s-2): \qquad \sum_{i=1}^{s} b_i c_i^{q-1} = \frac{1}{q} = \int_0^1 x^{q-1}\mathrm{d}x, \qquad 1 \le q \le 2s-2. \tag{A.3}$$

Then we have

$$u := \sum_{i=1}^{s} b_i P_{s-1}^2(c_i) = \frac{2s-1}{s-1}. \tag{A.4}$$

Proof. With the roots $c_1,\dots,c_s$, the Lobatto polynomial $P_{\text{Lobatto}}(x)$ can be reformulated as

$$P_{\text{Lobatto}}(x) = \frac{(2s-2)!}{s!}(x-c_1)\cdots(x-c_s). \tag{A.5}$$

Let V be the Vandermonde matrix

$$V = \begin{pmatrix} 1 & c_1 & \cdots & c_1^{s-1} \\ 1 & c_2 & \cdots & c_2^{s-1} \\ \cdots & \cdots & \cdots & \cdots \\ 1 & c_s & \cdots & c_s^{s-1} \end{pmatrix}, \tag{A.6}$$

and let $\rho_1,\dots,\rho_s$ be the coefficients of the polynomial $(x-c_1)\cdots(x-c_s)$. From (A.5) we have

$$\begin{aligned} P_{\text{Lobatto}}(x) &= \frac{(2s-2)!}{s!}\frac{(-1)^s}{\det(V)}\det\begin{pmatrix} 1 & x & \cdots & x^s \\ 1 & c_1 & \cdots & c_1^s \\ \cdots & \cdots & \cdots & \cdots \\ 1 & c_s & \cdots & c_s^s \end{pmatrix} \\ &= \frac{(2s-2)!}{s!}(x^s + \rho_1 x^{s-1} + \cdots + \rho_{s-1}x + \rho_s). \end{aligned}$$

For $s-1 \le q \le 2s-2$, condition $B(2s-2)$ implies that

$$\begin{pmatrix} b_1 & b_2 & \cdots & b_s \end{pmatrix}\begin{pmatrix} c_1^{s-2} & \cdots & \cdots & c_1^{2s-3} \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ c_s^{s-2} & \cdots & \cdots & c_s^{2s-3} \end{pmatrix} = \begin{pmatrix} \frac{1}{s-1} & \frac{1}{s} & \cdots & \frac{1}{2s-2} \end{pmatrix}; \tag{A.7}$$

therefore,

$$\begin{aligned}
\sum_{i=1}^{s} b_i c_i^{2s-2} &= \begin{pmatrix} b_1 & b_2 & \cdots & b_s \end{pmatrix} \begin{pmatrix} c_1^{2s-2} \\ \cdots \\ \\ c_s^{2s-2} \end{pmatrix} \\
&= \begin{pmatrix} \frac{1}{s-1} & \frac{1}{s} & \cdots & \frac{1}{2s-2} \end{pmatrix} \begin{pmatrix} c_1^{s-2} & \cdots & \cdots & c_1^{2s-3} \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ c_s^{s-2} & \cdots & \cdots & c_s^{2s-3} \end{pmatrix}^{-1} \begin{pmatrix} c_1^{2s-2} \\ \cdots \\ \\ c_s^{2s-2} \end{pmatrix} \\
&= -\frac{\rho_s}{s-1} - \cdots - \frac{\rho_1}{2s-2} \\
&= -\int_0^1 x^{s-2} \left(\rho_s + \cdots + \rho_1 x^{s-1}\right) \mathrm{d}x \\
&= -\int_0^1 x^{s-2} \left(\frac{s!}{(2s-2)!} P_{\mathrm{Lobatto}}(x) - x^s\right) \mathrm{d}x.
\end{aligned}$$

Furthermore, using integration by parts, we obtain

$$\begin{aligned}
\sum_{i=1}^{s} b_i c_i^{2s-2} &= \frac{1}{2s-1} - \frac{(-1)^{s-2} s!(s-2)!}{(2s-2)!} \int_0^1 x^{s-1}(x-1)^{s-1} \mathrm{d}x \\
&= \frac{1}{2s-1} + \frac{s!(s-2)!(s-1)!^2}{(2s-2)!^2(2s-1)}.
\end{aligned} \tag{A.8}$$

Let $a = \frac{\sqrt{2s-1}(2s-2)!}{(s-1)!^2}$ be the coefficient of highest order of $P_{s-1}(x)$. We rewrite $P_{s-1}(x)$ as

$$P_{s-1}(x) = P_{s-1}(x) - ax^{s-1} + ax^{s-1}, \tag{A.9}$$

and then

(A.10)

$$\sum_{i=1}^{s} b_i P_{s-1}^2(c_i) = \sum_{i=1}^{s} b_i \left(P_{s-1}(c_i) - ac_i^{s-1}\right) \left(P_{s-1}(c_i) + ac_i^{s-1}\right) + a^2 \sum_{i=1}^{s} b_i c_i^{2s-2}.$$

Noticing that $P_{s-1}(x) - ax^{s-1}$ is a polynomial of degree $s-2$, and (A.3) holds for polynomials of degree up to $2s-3$, it follows from (A.8) and (A.10) that

$$\begin{aligned}
\sum_{i=1}^{s} b_i P_{s-1}^2(c_i) &= \int_0^1 \left(P_{s-1}(x) - ax^{s-1}\right) \left(P_{s-1}(x) + ax^{s-1}\right) \mathrm{d}x + \frac{a^2}{2s-1} + \frac{s}{s-1} \\
&= \int_0^1 P_{s-1}^2(x) \mathrm{d}x + \frac{s}{s-1} \\
&= \frac{2s-1}{s-1}. \qquad \square
\end{aligned}$$

Appendix B. Asymptotic expansion of the diagonal Padé approximants to the exponential function.

Proof of Proposition 8.3. The proof is based on the asymptotic expansions of Olver [21] of Bessel functions of large argument, the connection being that

$$A_n(z) = {}_1F_1(-n,-2n,z) = \frac{n!}{(2n)!\sqrt{\pi}} z^{n+\frac{1}{2}} \mathrm{e}^{z/2} K_{n+\frac{1}{2}}(z/2) \tag{B.1}$$

and thus, setting $\nu = n + \frac{1}{2}$ and taking $-\pi < \arg z \le \pi$,

$$R_n(\nu z) = \mathrm{e}^{\mathrm{i}\nu\pi} \mathrm{e}^{\nu z} \frac{K_\nu(\nu z/2)}{K_\nu(-\nu z/2)}. \tag{B.2}$$

We obtain the asymptotic behavior of $R_n(\nu z)$ as $n \to \infty$ for any fixed z lying in the quadrant $\mathbf{D} := 0 \le \arg z \le \frac{\pi}{2}$. We divide $\mathbf{D}$ into six regions and consider each in turn. First, we give Olver's result. Let

$$\zeta = \sqrt{1 + \frac{z^2}{4}} + \ln \frac{z}{2\Big(1 + \sqrt{1 + \frac{z^2}{4}}\Big)}, \tag{B.3}$$

which maps the half plane $\Re z > 0$ conformally onto the union of the half plane $\Re\zeta > 0$ and the half strip $|\Im\zeta| < \frac{\pi}{2}$, $\Re\zeta \le 0$. (The interior of $\mathbf{K}$ maps into $\Re\zeta < 0$, while the exterior of $\mathbf{K}$ maps into $\Re\zeta > 0$.) Then [21, eq. (2.13)–(2.14)]

$$I_\nu(\nu z/2) \sim \frac{\mathrm{e}^{\nu\zeta}}{\sqrt{2\pi\nu}(1 + \frac{z^2}{4})^{1/4}} \tag{B.4}$$

and

$$K_\nu(\nu z/2) \sim \sqrt{\frac{\pi}{2\nu}} \frac{\mathrm{e}^{-\nu\zeta}}{(1 + \frac{z^2}{4})^{1/4}} \tag{B.5}$$

as $\nu \to \infty$, uniformly with respect to z in $|\arg z| \le \frac{\pi}{2} - \varepsilon$. (In fact, a complete asymptotic expansion in inverse powers of ν is provided.) In [22, p. 380], it is shown by analytic continuation that (B.5) also holds in $|\arg z| \le \frac{\pi}{2}$, provided z is bounded away from $\pm 2\mathrm{i}$.

Region 1. $\Re z > 0$, z inside $\mathbf{K}$. We use the identity

$$K_\nu(-z) = \mathrm{e}^{\mathrm{i}\nu\pi} K_\nu(z) + \mathrm{i}\pi I_\nu(z), \quad \nu \notin \mathbb{Z}. \tag{B.6}$$

In this region, $\Re\zeta < 0$ so $I_\nu(\nu z/2)$ is subdominant to $K_\nu(\nu z/2)$ and we have

$$R_n(\nu z) \sim \mathrm{e}^{\nu z}, \tag{B.7}$$

with the full asymptotic expansion proceeding in inverse powers of ν.

Region 2. $\Re z > 0$, z outside $\mathbf{K}$. We proceed as for Region 1, but now $\Re\zeta > 0$ is $K_\nu(\nu z/2)$ so subdominant to $I_\nu(\nu z/2)$ and we use (B.4) to get

$$\begin{aligned} R_n(\nu z) &= \mathrm{e}^{\mathrm{i}\nu\pi} \mathrm{e}^{\nu z} \frac{K_\nu(\nu z/2)}{K_\nu(-\nu z/2)} \\ &\sim \mathrm{e}^{\mathrm{i}\nu\pi} \mathrm{e}^{\nu z} \mathrm{e}^{-2\nu\zeta}/\mathrm{i} \\ &= (-1)^n \mathrm{e}^{\nu(z-2\zeta)}. \end{aligned} \tag{B.8}$$

Region 3. $\Re z = 0$, z inside $\mathbf{K}$. Here $-z = \overline{z}$, so we have

$$\begin{aligned} R_n(\nu z) &= \mathrm{e}^{\mathrm{i}\nu\pi}\mathrm{e}^{\nu z}\frac{K_\nu(\nu z/2)}{K_\nu(-\nu z/2)} \\ &= \mathrm{e}^{\mathrm{i}\nu\pi}\mathrm{e}^{\nu z}\mathrm{e}^{2\mathrm{i}\arg K_\nu(\nu z/2)} \\ &\sim \mathrm{e}^{\mathrm{i}\nu\pi}\mathrm{e}^{\nu z}\mathrm{e}^{2\mathrm{i}(-\nu)\frac{\pi}{2}} \\ &= \mathrm{e}^{\nu z}. \end{aligned} \tag{B.9}$$

Region 4. $\Re z = 0$, z outside $\mathbf{K}$. We proceed as for Region 3, but now $\arg(1 + z^2/4)^{1/4} = \frac{\pi}{4}$ and $\Re\zeta(z) = 0$ so $\arg \mathrm{e}^{-\nu\zeta} = -\nu\Im\zeta$; combining,

$$\begin{aligned} R_n(\nu z) &\sim \mathrm{e}^{\mathrm{i}\nu\pi}\mathrm{e}^{\nu z}\mathrm{e}^{-2\nu\zeta}\mathrm{e}^{-2\mathrm{i}\frac{\pi}{4}} \\ &= (-1)^n\mathrm{e}^{\nu(z-2\zeta)}. \end{aligned} \tag{B.10}$$

Region 5. $\Re z > 0$, $z \in \partial\mathbf{K}$. Here $\zeta \in [0, \frac{\pi}{2}\mathrm{i})$, so neither $K_\nu(\nu z/2)$ nor $I_\nu(\nu z/2)$ is dominant; we must retain them both. Combining (B.2), (B.4), (B.5), and (B.6) gives

$$R_n(\nu z) \sim \frac{e^{\nu(z+\zeta)}}{\mathrm{e}^{\nu\zeta} + (-1)^n\mathrm{e}^{-\nu\zeta}}. \tag{B.11}$$

Region 6. $z = 2\mathrm{i}$. Here we use

$$K_\nu(\mathrm{i}\nu) = -\frac{\mathrm{i}\pi}{2}\mathrm{e}^{-\frac{1}{2}\nu\mathrm{i}\pi}(J_\nu(\nu) - \mathrm{i}Y_\nu(\nu)) \tag{B.12}$$

together with equations (9.1.3), (9.3.31), and (9.3.32) of [1] to get

$$K_\nu(\mathrm{i}\nu) \sim -\frac{\mathrm{i}\pi}{2}\mathrm{e}^{-\frac{1}{2}\nu\mathrm{i}\pi}2^{1/3}3^{-2/3}\Gamma(2/3)^{-1}(1 + \sqrt{3}\mathrm{i})\nu^{-1/3}, \tag{B.13}$$

which, together with (B.2), gives

$$R_n(\nu z) \sim \mathrm{e}^{\nu z - \mathrm{i}\frac{\pi}{3}}. \tag{B.14}$$

(With a little more work, equations (9.3.35), (9.3.36) of [1] give expansions uniformly valid as $z \to 2\mathrm{i}$.)

Combining Regions 1–4 now proves the proposition. ∎

Note that (B.11) also provides asymptotic estimates of the poles (and hence zeros) of R_n, namely, they are asymptotically located at νz_j, where

$$\nu\zeta_j = \left(j + \frac{1}{2}\right)\pi, \quad j = 1, \ldots, n/2, \tag{B.15}$$

for n even, and at

$$\nu\zeta_j = j\pi, \quad j = 0, \ldots, (n-1)/2, \tag{B.16}$$

for n odd. These estimates are amazingly good. Even at $n = 1$, they estimate that the single pole is at $\nu\zeta^{-1}(0) = \frac{3}{2} \times 1.3254868386983634 \approx 1.988$, whereas the actual pole is at $z = 2$ (recall that $R_1(z) = \frac{1+\frac{z}{2}}{1-\frac{z}{2}}$)—an error of 0.6%. Note also that at $\zeta = 0$, $z = 1.3254868386983634$, $R_n(\nu z) \sim \frac{1}{2}\mathrm{e}^{\nu z}$ for n even. The convergence of the approximants to the asymptotic limit is shown in Figure B.1, where the change in behavior as $\partial\mathbf{K}$ is crossed is immediately apparent. The transition region, of width $\mathcal{O}(\nu^{-2/3})$ in the neighborhood of $z = 2\mathrm{i}$, can also be seen.

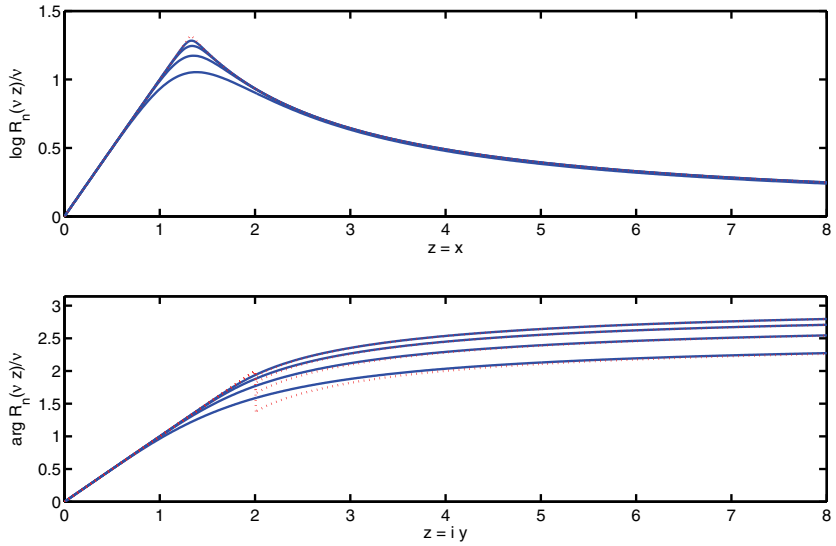

FIG. B.1. *Convergence of diagonal Padé approximants to* e^z. $\log R_n(\nu z)/\nu$ *(solid lines) is shown for* $n = 2$, 4, 8, *and* 16, *along the positive real axis (top) and imaginary axis (bottom), along with the asymptotic behavior given in Proposition* 8.3 *(dotted lines).*

Acknowledgment. We would like to thank Laurent Jay for useful discussions and for pointing out to us the report [12].

REFERENCES

[1] M. ABRAMOWITZ AND I. A. STEGUN, EDS., *Handbook of Mathematical Functions*, Dover Publications, New York, 1965.
[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] U. M. ASCHER AND R. I. MCLACHLAN, *Multisymplectic box schemes and the Korteweg-de Vries equation*, Workshop on Innovative Time Integrators for PDEs, Appl. Numer. Math., 48 (2004), pp. 255–269.
[4] G. A. BAKER, JR. AND P. GRAVES-MORRIS, *Padé Approximants*, 2nd ed., Encyclopedia Math. Appl., 59, Cambridge University Press, Cambridge, UK, 1996.
[5] C. L. BOTTASSO, *A new look at finite elements in time: A variational interpretation of Runge–Kutta methods*, Appl. Numer. Math., 25 (1997), pp. 355–368.
[6] T. J. BRIDGES AND S. REICH, *Multi-symplectic integrators: Numerical schemes for Hamiltonian PDEs that conserve symplecticity*, Phys. Lett. A, 284 (2001), pp. 184–193.
[7] J. CHEN AND M. QIN, *Multisymplectic composition integrators of high order*, J. Comput. Math., 21 (2003), pp. 647–656.
[8] B. L. EHLE, *On Padé Approximations to the Exponential Function and A-stable Methods for the Numerical Solution of Initial Value Problems*, Research Report CSRR 2010, University of Waterloo, Waterloo, Ontario, Canada, 1969.
[9] V. GRIMM AND R. SCHERER, *A generalized W-transformation for constructing symplectic partitioned Runge-Kutta methods*, BIT, 43 (2003), pp. 57–66.
[10] E. HAIRER, S. P. NØRSETT, AND G. WANNER, *Solving Ordinary Differential Equations.* I, *Nonstiff Problems,* 2nd ed., Springer Ser. Comput. Math. 8, Springer-Verlag, Berlin, 1993.
[11] E. HAIRER AND G. WANNER, *Solving Ordinary Differential Equations.* II, *Stiff and Differential-Algebraic Problems*, 2nd ed., Springer Ser. Comput. Math. 14, Springer-Verlag, Berlin, 1996.

[12] L. O. Jay and L. R. Petzold, *Highly Oscillatory Systems and Periodic Stability*, Preprint 95-015, Army High Performance Computing Research Center, Stanford, CA, 1995.
[13] L. O. Jay, *Symplectic partitioned Runge-Kutta methods for constrained Hamiltonian systems*, SIAM J. Numer. Anal., 33 (1996), pp. 368–387.
[14] L. O. Jay, *Specialized partitioned additive Runge-Kutta methods for systems of overdetermined DAEs with holonomic constraints*, SIAM J. Numer. Anal., 45 (2007), pp. 1814–1842.
[15] B. Leimkuhler and S. Reich, *Simulating Hamiltonian Dynamics*, Cambridge Monogr. Appl. Comput. Math. 14, Cambridge University Press, Cambridge, UK, 2004.
[16] Y. L. Luke, *The Special Functions and Their Approximations,* I, Mathematics in Science and Engineering 53, Academic Press, New York, 1969.
[17] Y. L. Luke, *The Special Functions and Their Approximations,* II, Mathematics in Science and Engineering 53, Academic Press, New York, 1969.
[18] J. E. Marsden and M. West, *Discrete mechanics and variational integrators*, Acta Numer., 10 (2001), pp. 357–514.
[19] R. I. McLachlan, *A new implementation of symplectic Runge–Kutta methods*, SIAM J. Sci. Comput., 29 (2007), pp. 1637–1649.
[20] R. I. McLachlan, G. R. W. Quispel, and G. S. Turner, *Numerical integrators that preserve symmetries and preserving symmetries*, SIAM J. Numer. Anal., 35 (1998), pp. 586–599.
[21] F. W. J. Olver, *The asymptotic expansion of Bessel functions of large order*, Philos. Trans. Roy. Soc. Ser. A, 247 (1954), pp. 328–368.
[22] F. W. J. Olver, *Asymptotics and Special Functions*, Academic Press, New York, 1974.
[23] S. Reich, *Multi-symplectic Runge-Kutta collocation methods for Hamiltonian wave equations*, J. Comput. Phys., 157 (2000), pp. 473–499.
[24] B. N. Ryland, *Multisymplectic Integration*, Ph.D. thesis, Massey University, Palmerston North, New Zealand.
[25] B. N. Ryland, R. I. McLachlan, and J. Frank, *On the multisymplecticity of partitioned Runge–Kutta and splitting methods*, Int. J. Comput. Math., 84 (2007), pp. 847–869.
[26] B. N. Ryland and R. I. McLachlan, *On multisymplecticity of partitioned Runge-Kutta methods*, SIAM J. Sci. Comput., 30 (2008), pp. 1318–1340.
[27] J. M. Sanz-Serna and L. Abia, *Order conditions for canonical Runge–Kutta schemes*, SIAM J. Numer. Anal., 28 (1991), pp. 1081–1096.
[28] G. Sun, *Symplectic partitioned Runge–Kutta methods*, J. Comput. Math., 11 (1993), pp. 365–372.
[29] G. Sun, *Construction of high order symplectic PRK methods*, J. Comput. Math., 13 (1995), pp. 40–50.