Electronic Transactions on Numerical Analysis.
Volume 60, pp. 136–168, 2024.

ISSN 1068–9613.
DOI: 10.1553/etna_vol60s136

ETNA
Kent State University and
Johann Radon Institute (RICAM)

CONVERGENCE ANALYSIS OF A KRYLOV SUBSPACE SPECTRAL METHOD FOR THE 1D WAVE EQUATION IN AN INHOMOGENEOUS MEDIUM*

BAILEY RESTER†, ANZHELIKA VASILYEVA†, AND JAMES V. LAMBERS†

Abstract. This paper presents a convergence analysis of a Krylov subspace spectral (KSS) method applied to an 1D wave equation in an inhomogeneous medium. It will be shown that for sufficiently regular initial data, this KSS method yields unconditional stability, spectral accuracy in space, and second-order accuracy in time in the case of constant wave speed and a bandlimited reaction term coefficient. Numerical experiments that corroborate the established theory are included along with an investigation of generalizations, such as to higher space dimensions and nonlinear PDEs, that features performance comparisons with other Krylov subspace-based time-stepping methods. This paper also includes the first stability analysis of a KSS method that does not assume a bandlimited reaction term coefficient.

Key words. spectral methods, wave equation, convergence analysis, variable coefficients

AMS subject classifications. 65M70, 65M12, 65F60

1. Introduction.

Consider the 1D wave equation in an inhomogeneous medium,

$$u_{tt} = (p(x)u_x)_x + q(x)u, \tag{1.1}$$

on a bounded domain with appropriate initial and boundary conditions. Analytical methods are not practical for this problem since the coefficients are not constant. For instance, applying separation of variables [8] would result in a spatial ODE that cannot be solved analytically, and therefore numerical methods are needed. However, standard time-stepping methods, such as Runge-Kutta methods or multistep methods, suffer from a lack of scalability. As the number of grid points increases, a smaller time step would be needed due to the CFL condition [16] for explicit methods or an increasingly ill-conditioned system must be solved for implicit methods. It follows that increasing the number of grid points significantly increases the computational expense. Therefore, a more practical numerical method for solving this kind of variable-coefficient PDE is desirable.

Krylov subspace spectral (KSS) methods are high-order accurate, explicit time-stepping methods that possess a stability characteristic of implicit methods [33]. By contrast with other time-stepping methods, KSS methods employ a componentwise approach in which each Fourier coefficient of the solution is computed using an approximation of the solution operator of the PDE that is tailored to that coefficient. This customization is based on techniques for approximating bilinear forms involving matrix functions by treating them as Riemann-Stieltjes integrals [9]. This componentwise approach allows KSS methods to circumvent difficulties caused by stiffness, and thus they scale effectively to higher spatial resolution [4].

A first-order KSS method applied to the heat equation with a constant leading coefficient was proven to be unconditionally stable [23, 24], as well as a second-order KSS method applied to the wave equation with a constant leading coefficient [22]. In all of these studies, lower-order coefficients of the spatial differential operator were assumed to be bandlimited. A first-order KSS method applied to the heat equation with a bandlimited leading coefficient is also unconditionally stable [33]. In this paper, we analyze the stability of a KSS method applied to the wave equation with bandlimited coefficients.

*Received January 20, 2023. Accepted March 8, 2024. Published online on April 3, 2024. Recommended by Maya Neycheva.

†School of Mathematics and Natural Sciences, The University of Southern Mississippi, 118 College Dr #5043, Hattiesburg, MS 39406, USA (resterbailey@yahoo.com).

The outline of the paper is as follows. Section 2 provides an overview of KSS methods, as applied to the wave equation. Section 3 presents a stability analysis of a second-order KSS method applied to the PDE (1.1) with bandlimited coefficients $p(x)$ and $q(x)$ and periodic boundary conditions. In that same section, a full convergence analysis in the case of $p(x) \equiv$ constant is carried out. Corroborating numerical experiments are given in Section 4, along with applications of the second-order KSS method to more general problems. This section also includes performance comparisons between the KSS method and other time-stepping methods, particularly those that also make use of Krylov subspaces. Upon demonstrating through numerical experiments that the assumptions on the coefficients of the PDE made in Section 3 are not necessary for convergence, an additional stability analysis is conducted in which $q(x)$ is not assumed to be bandlimited, a task that has not previously been performed for a KSS method. Conclusions and ideas for future work are given in Section 5.

2. Background.

Consider the second-order wave equation

$$u_{tt} + Lu = 0 \qquad \text{on } (0, 2\pi) \times (0, \infty), \tag{2.1}$$

$$u(x, 0) = f(x),\ u_t(x, 0) = g(x), \qquad 0 < x < 2\pi, \tag{2.2}$$

with periodic boundary conditions

$$u(0, t) = u(2\pi, t), \qquad t > 0. \tag{2.3}$$

The spatial differential operator L is defined by

$$Lu = -(p(x)u_x)_x + q(x)u, \tag{2.4}$$

where we assume $p(x) > 0$ and $q(x) \geq 0$ to guarantee that L is self-adjoint and positive definite.

A spectral representation of the operator L allows us to describe the solution operator, the propagator, as a function of L [15]. By introducing

$$\begin{aligned} f_{11}(\lambda) &= f_{22}(\lambda) = \cos(\lambda^{1/2}\Delta t), \\ f_{12}(\lambda) &= \lambda^{-1/2} \sin(\lambda^{1/2}\Delta t), \\ f_{21}(\lambda) &= -\lambda f_{12}(\lambda), \end{aligned} \tag{2.5}$$

we can describe the evolution of the solution by

$$\begin{bmatrix} u(x, t + \Delta t) \\ u_t(x, t + \Delta t) \end{bmatrix} = \begin{bmatrix} f_{11}(L) & f_{12}(L) \\ f_{21}(L) & f_{22}(L) \end{bmatrix} \begin{bmatrix} u(x, t) \\ u_t(x, t) \end{bmatrix}.$$

In view of the periodic boundary conditions, we can express the solution at time $t + \Delta t$ as a sum of Fourier series,

$$\begin{aligned} u(x, t + \Delta t) = \frac{1}{2\pi} \sum_{\omega=-\infty}^{\infty} e^{i\omega x} \langle e^{i\omega \cdot}, \cos(L^{1/2}\Delta t) u(\cdot, t) \rangle \\ + \frac{1}{2\pi} \sum_{\omega=-\infty}^{\infty} e^{i\omega x} \langle e^{i\omega \cdot}, L^{-1/2} \sin(L^{1/2}\Delta t) u_t(\cdot, t) \rangle, \end{aligned}$$

where $\langle \cdot, \cdot \rangle$ is the standard inner product of functions on $(0, 2\pi)$.

Upon spatial discretization, each Fourier coefficient in the above series is approximated by an expression of the form

$$\mathbf{u}^H f(A) \mathbf{v}, \tag{2.6}$$

where $\mathbf{u}$ and $\mathbf{v}$ are N-vectors, A is an $N \times N$ symmetric positive definite matrix, and f is either f_{11} or f_{12}. In [9], Golub and Meurant describe algorithms for approximating such bilinear forms involving matrix functions by treating them as Riemann-Stieltjes integrals that can be approximated through Gauss quadrature over an interval containing the eigenvalues of A.

In the case of $\mathbf{u} = \mathbf{v}$, the Gauss quadrature rule is constructed by applying the Lanczos algorithm to A with an initial vector $\mathbf{u}$. The Gauss quadrature nodes and weights are then obtained from the eigenvalues and eigenvectors of the tridiagonal matrix of the recursion coefficients produced by the Lanczos iteration. In the case of $\mathbf{u} \neq \mathbf{v}$, the unsymmetric Lanczos algorithm can be used instead, with initial vectors $\mathbf{u}$ and $\mathbf{v}$, but this may yield a quadrature rule that does not have real positive weights, which can be numerically unstable [1].

For this case, a block approach can be used instead [9]: the block Lanczos algorithm [10] is applied to A with an initial block $\begin{bmatrix}\mathbf{u} & \mathbf{v}\end{bmatrix}$. The iteration produces a block tridiagonal matrix with 2×2 blocks, and as before, its eigenvalues and eigenvectors yield Gauss quadrature nodes and (matrix-valued) weights.

We now describe how this approach is applied to the solution of the problem (2.1)–(2.3). Let $\mathbf{u}^n$ and $\mathbf{u}_t^n$ be the computed solution at time t_n and its time derivative, respectively, and let $\hat{\mathbf{e}}_\omega$ be a discretization of $\hat{e}_\omega(x) = e^{i\omega x}$. For each wave number $\omega = -N/2+1, \ldots, N/2$, we define

$$R_0 = \begin{bmatrix} \frac{1}{N}\hat{\mathbf{e}}_\omega & \mathbf{u}^n \end{bmatrix}, \qquad \tilde{R}_0 = \begin{bmatrix} \frac{1}{N}\hat{\mathbf{e}}_\omega & \mathbf{u}_t^n \end{bmatrix},$$

and then compute the QR factorizations

$$R_0 = X_1 B_0, \qquad \tilde{R}_0 = \tilde{X}_1 \tilde{B}_0.$$

The block Lanczos iteration yields $\mathcal{T}_K$ and $\tilde{\mathcal{T}}_K$ from X_1 and $\tilde{X}_1$. Then, the Fourier coefficients of the solution and its time derivative are approximated by

$$[\hat{\mathbf{u}}^{n+1}]_\omega = \left[B_0^H \cos[\mathcal{T}_K^{1/2}\Delta t]_{1:2,1:2} B_0\right]_{12} + \left[\tilde{B}_0^H (\tilde{\mathcal{T}}_K^{-1/2} \sin[\tilde{\mathcal{T}}_K^{1/2}\Delta t])_{1:2,1:2} \tilde{B}_0\right]_{12},$$

$$[\hat{\mathbf{u}}_t^{n+1}]_\omega = -\left[B_0^H (\mathcal{T}_K^{1/2} \sin[\mathcal{T}_K^{1/2}\Delta t])_{1:2,1:2} B_0\right]_{12} + \left[\tilde{B}_0^H \cos[\tilde{\mathcal{T}}_K^{1/2}\Delta t]_{1:2,1:2} \tilde{B}_0\right]_{12}.$$

Let $u(x, \Delta t)$ be the exact solution, and let $\tilde{u}(x, \Delta t)$ be the approximate solution. If K iterations of the block Lanczos method are performed, then [22], for $\omega = -N/2+1, \ldots, N/2$,

$$|\langle \hat{e}_\omega, u(\cdot, \Delta t) - \tilde{u}(\cdot, \Delta t)\rangle| = O(\Delta t^{4K}),$$

$$|\langle \hat{e}_\omega, u_t(\cdot, \Delta t) - \tilde{u}_t(\cdot, \Delta t)\rangle| = O(\Delta t^{4K-1}).$$

The high order of accuracy in time is due to the second derivative with respect to time in the PDE. In addition to their high-order accuracy in time, the following results about the stability of KSS methods have been proven for various problems:

- the heat equation $u_t = p u_{xx} + q(x)u$, where p is constant and $q(x)$ is bandlimited: a first-order KSS method is unconditionally stable [24],
- the wave equation $u_{tt} = p u_{xx} + q(x)u$, where p is constant and $q(x)$ is bandlimited: a second-order KSS method is unconditionally stable [22],
- the reaction-diffusion system of the form $\mathbf{v}_t = L\mathbf{v}$: a first-order KSS method with constant diffusion coefficient and bandlimited reaction term coefficient is unconditionally stable [33],

- the wave equation $u_{tt} = pu_{xx} + q(x)u$, where p is constant and $q(x)$ is bandlimited: a second-order block KSS method is unconditionally stable [22], and
- the heat equation $u_t = (p(x)u_x)_x + q(x)u$ where $p(x)$ and $q(x)$ are bandlimited: a first-order block KSS method is unconditionally stable [33].

KSS methods use a significantly different approach to computing matrix function-vector products of the form $\varphi(A)\mathbf{b}$ than Krylov subspace methods from the literature (see, for example, [19]). Such Krylov subspace methods approximate the function φ with either a polynomial or rational function. Depending on the function φ, the approximating function may need to be of high degree to ensure sufficient accuracy. When such methods are used to solve stiff systems of ODEs obtained from spatial discretization of PDEs, the degree can grow substantially when the time steps or the number of grid points increases.

This is demonstrated in [4], where it was also shown that, by contrast, KSS methods do not suffer from this loss of scalability. Each Fourier coefficient of the solution is obtained using its own frequency-dependent approximation that is of a low degree determined by the desired order of temporal accuracy. This is possible because each Fourier coefficient is equivalent to a Riemann-Stieltjes integral with a frequency-dependent measure that is nearly constant over most of the integration domain [4], and therefore the integral is determined primarily by the behavior of the integrand over only a small, frequency-dependent portion of this domain.

In Section 4 it will be demonstrated that this componentwise approach to time stepping provides an advantage over other time-stepping methods that apply the same approximation of the exponential to all components of the solution.

3. Convergence analysis.

We will now analyze convergence of a second-order KSS method with $K = 1$ for the IVP (2.1)–(2.4) under the assumptions that the Fourier coefficients $\hat{p}(\omega)$, $\hat{q}(\omega)$ of $p(x)$ and $q(x)$, respectively, satisfy $\hat{p}(\omega) = \hat{q}(\omega) = 0$ when $|\omega| > \omega_{\text{max}}$, for some threshold ω_{max}. That is, we assume that $p(x)$ and $q(x)$ are bandlimited.

We first carry out spatial discretization. We use a uniform grid with spacing $\Delta x = 2\pi/N$, where N is assumed to be even. Then, we let $\mathbf{x}_N$ be an N-vector of grid points

$$x_j = j\Delta x, \qquad j = 0, 1, 2, \dots, N-1.$$

Let ω_j be the corresponding wave numbers

$$\omega_j = j - N/2, \qquad j = 1, 2, \dots, N.$$

We then denote by D_N an $N \times N$ matrix that discretizes the second-derivative operator using the discrete Fourier transform:

$$D_N = F_N^{-1}\Lambda_N F_N,$$

where

$$[F_N]_{jk} = \frac{1}{N}e^{-i\omega_j x_k}, \qquad [\Lambda_N]_{jj} = -\omega_j^2.$$

We also let I_N denote the $N \times N$ identity matrix, whereas I is the identity operator on functions of x.

Let $u_1 = u$ and $u_2 = u_t$. We rewrite (2.1) as the first-order system

$$\begin{aligned} \frac{\partial u_1}{\partial t} &= u_2, \\ \frac{\partial u_2}{\partial t} &= -Lu_1, \end{aligned} \tag{3.1}$$

which, for convenience, we write as

$$\mathbf{v}_t = \tilde{L}\mathbf{v}, \qquad \mathbf{v} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}, \quad \tilde{L} = \begin{bmatrix} 0 & I \\ -L & 0 \end{bmatrix}. \tag{3.2}$$

Spatial discretization of (3.2) yields a system of ODEs

$$\mathbf{v}_N'(t) = \tilde{L}_N \mathbf{v}(t), \tag{3.3}$$

where

$$\mathbf{v}_N(t) = \begin{bmatrix} \mathbf{u}_{1,N}(t) \\ \mathbf{u}_{2,N}(t) \end{bmatrix}$$

is the spatial discretization of the vector field $\mathbf{v}$ and $\tilde{L}_N$ is a $2N \times 2N$ matrix which has the 2×2 block structure

$$\tilde{L}_N = \begin{bmatrix} 0 & I_N \\ -L_N & 0 \end{bmatrix}.$$

We define the exact solution operator of (3.1) as

$$S(t) = \exp[\tilde{L}t] = \begin{bmatrix} S_{11}(t) & S_{12}(t) \\ S_{12}(t) & S_{22}(t) \end{bmatrix} = \begin{bmatrix} R_0(t) & R_1(t) \\ -L\,R_1(t) & R_0(t) \end{bmatrix}, \tag{3.4}$$

where, as before, $R_1(t) = L^{-1/2}\sin(L^{1/2}t)$ and $R_0(t) = \cos(L^{1/2}t)$. Then we let

$$S_N(\Delta t) = \begin{bmatrix} S_{N,11}(\Delta t) & S_{N,12}(\Delta t) \\ S_{N,12}(\Delta t) & S_{N,22}(\Delta t) \end{bmatrix}, \tag{3.5}$$

where each $S_{N,ij}(\Delta t)$ is the approximation of $S_{ij}(\Delta t)$ by the KSS method.

A KSS method applied to (2.1) with $K = 1$ uses two block Gauss quadrature nodes for each Fourier coefficient. Using an approach described in [28], we estimate these nodes rather than using the block Lanczos iteration explicitly. This significantly improves the efficiency of KSS methods, but it will also simplify the convergence analysis to be carried out in this section. The quadrature nodes will be prescribed as follows:

$$l_{1,\omega} = 0, \quad l_{2,\omega} = \overline{p}\omega^2 + \overline{q}, \qquad \omega = -N/2+1, \ldots, N/2, \tag{3.6}$$

where $\overline{p}$ and $\overline{q}$ are the average values of $p(x)$ and $q(x)$, respectively, on $[0, 2\pi]$.

To interpolate the functions $f_{ij}(\lambda)$ from (2.5) at the nodes $l_{1,\omega}$, $l_{2,\omega}$, we compute the slopes

$$M_{ij,\omega} = \frac{f_{ij}(l_{2,\omega}) - f_{ij}(l_{1,\omega})}{l_{2,\omega} - l_{1,\omega}}, \qquad \omega = -N/2+1, \ldots, N/2, \quad i,j = 1,2.$$

We then describe the computed solution at time t_{n+1} by

$$\begin{aligned} \mathbf{u}^{n+1} &= \mathbf{z}_{11} + \mathbf{z}_{12}, \\ \mathbf{u}_t^{n+1} &= \mathbf{z}_{21} + \mathbf{z}_{22}, \end{aligned} \tag{3.7}$$

where

$$\mathbf{z}_{i1} = S_{N,i1}(\Delta t)\mathbf{u}^n, \quad \mathbf{z}_{i2} = S_{N,i2}(\Delta t)\mathbf{u}_t^n, \qquad i = 1,2.$$

We also define $\tilde{p} = p - \bar{p}$, $\tilde{q} = q - \bar{q}$, and let P_N, Q_N, $\tilde{P}_N$, and $\tilde{Q}_N$ be diagonal matrices with the values of the coefficients $p(x)$, $q(x)$, $\tilde{p}(x)$, and $\tilde{q}(x)$, respectively, at the grid points on the main diagonal.

Let $I_\omega = \{k \in \mathbb{Z} | 0 < |k - \omega| \leq \omega_{\max}\}$. The discrete Fourier coefficients of $\mathbf{z}_{ij}$, $i, j = 1, 2$, are then given by

$$\begin{aligned}
\hat{\mathbf{z}}_{11}(\omega) &= S_{11}(l_{2,\omega})(\hat{\mathbf{e}}_\omega^H \mathbf{u}^n) + M_{11,\omega} \hat{\mathbf{e}}_\omega^H (L_N - l_{2,\omega} I)\mathbf{u}^n \\
&= S_{11}(l_{2,\omega})\hat{u}(\omega) - i\omega M_{11,\omega} \sum_{k \in I_\omega} \hat{p}(\omega - k) i(k) \hat{u}(k) \\
&\qquad + M_{11,\omega} \sum_{k \in I_\omega} \hat{q}(\omega - k)\hat{u}(k),
\end{aligned} \tag{3.8}$$

$$\begin{aligned}
\hat{\mathbf{z}}_{21}(\omega) &= S_{21}(l_{2,\omega})\hat{u}(\omega) - i\omega M_{21,\omega} \sum_{k \in I_\omega} \hat{p}(\omega - k) i(k) \hat{u}(k) \\
&\qquad + M_{21,\omega} \sum_{k \in I_\omega} \hat{q}(\omega - k)\hat{u}(k),
\end{aligned} \tag{3.9}$$

$$\begin{aligned}
\hat{\mathbf{z}}_{12}(\omega) &= S_{12}(l_{2,\omega})\hat{u}_t(\omega) - i\omega M_{12,\omega} \sum_{k \in I_\omega} \hat{p}(\omega - k) i(k) \hat{u}_t(k) \\
&\qquad + M_{12,\omega} \sum_{k \in I_\omega} \hat{q}(\omega - k)\hat{u}_t(k), \\
\hat{\mathbf{z}}_{22}(\omega) &= S_{22}(l_{2,\omega})\hat{u}_t(\omega) - i\omega M_{22,\omega} \sum_{k \in I_\omega} \hat{p}(\omega - k) i(k) \hat{u}_t(k) \\
&\qquad + M_{22,\omega} \sum_{k \in I_\omega} \hat{q}(\omega - k)\hat{u}_t(k),
\end{aligned}$$

where $i = \sqrt{-1}$ and $\omega = -N/2 + 1, \ldots, N/2$. To obtain these formulas, we used the simplification

$$\begin{aligned}
\hat{\mathbf{e}}_\omega^H (L_N - l_{2,\omega} I)\mathbf{u}^n &= \hat{\mathbf{e}}_\omega^H [-D_N P_N D_N + Q_N]\mathbf{u}^n - l_{2,\omega}(\hat{\mathbf{e}}_\omega^H \mathbf{u}^n) \\
&= -i\omega \hat{\mathbf{e}}_\omega^H P_N D_N \mathbf{u}^n + \hat{\mathbf{e}}_\omega^H Q_N \mathbf{u}^n - l_{2,\omega}(\hat{\mathbf{e}}_\omega^H \mathbf{u}^n) \\
&= (\bar{p}\omega^2 + \bar{q})(\hat{\mathbf{e}}_\omega^H \mathbf{u}^n) - i\omega \hat{\mathbf{e}}_\omega^H \tilde{P}_N D_N \mathbf{u}^n + \hat{\mathbf{e}}_\omega^H \tilde{Q}_N \mathbf{u}^n - l_{2,\omega}(\hat{\mathbf{e}}_\omega^H \mathbf{u}^n) \\
&= -i\omega \hat{\mathbf{e}}_\omega^H \tilde{P}_N D_N \mathbf{u}^n + \hat{\mathbf{e}}_\omega^H \tilde{Q}_N \mathbf{u}^n.
\end{aligned}$$

To bound the error, we need to establish an upper bound for a norm of the approximate solution operator $S_N(\Delta t)$. We elect to use the C-norm, defined by

$$\| (u, v) \|_C^2 = \langle u, Cu \rangle + \langle v, v \rangle,$$

where, as before, $\langle \cdot, \cdot \rangle$ is the standard inner product on $(0, 2\pi)$, and its discrete counterpart, the C_N-norm, defined by

$$\| (\mathbf{u}, \mathbf{v}) \|_{C_N}^2 = \mathbf{u}^T C_N \mathbf{u} + \|\mathbf{v}\|_2^2.$$

Here, the $N \times N$ matrix C_N discretizes the constant-coefficient differential operator $C = -\bar{p}\partial_{xx} + \bar{q}I$, where u and v are N-vectors. We choose to bound the C_N-norm of the solution operator for convenience because the operator C_N has a very simple expression in Fourier space due to its constant coefficients, which simplifies the analysis.

3.1. Stability. We wish to express $\|S_N(\Delta t)\|_{C_N}$ as the 2-norm of some matrix since that will be easier to bound. We define $\|S_N(\Delta t)\|_{C_N}$ by

$$\|S_N(\Delta t)\|_{C_N}^2 = \sup_{\mathbf{w}=(\mathbf{u},\mathbf{v})\neq\mathbf{0}} \frac{\|S_N(\Delta t)\mathbf{w}\|_{C_N}^2}{\|\mathbf{w}\|_{C_N}^2}.$$

Then

$$\|S_N(\Delta t)\|_{C_N}^2 = \sup_{(\mathbf{u},\mathbf{v})\neq\mathbf{0}} \frac{\tilde{\mathbf{u}}^T C_N \tilde{\mathbf{u}} + \|\tilde{\mathbf{v}}_2\|^2}{\mathbf{u}^T C_N \mathbf{u} + \|\mathbf{v}_2\|^2},$$

where

$$\begin{bmatrix} \tilde{\mathbf{u}} \\ \tilde{\mathbf{v}} \end{bmatrix} = S_N(\Delta t) \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix} = S_N(\Delta t)\mathbf{w}.$$

In matrix form, we have

$$\|S_N(\Delta t)\|_{C_N}^2 = \sup \frac{\mathbf{w}^T S_N(\Delta t)^T \tilde{C}_N S_N(\Delta t)\mathbf{w}}{\mathbf{w}^T \tilde{C}_N \mathbf{w}} = \sup \frac{\mathbf{w}^T S_N(\Delta t)^T \tilde{C}_N S_N(\Delta t)\mathbf{w}}{(\tilde{C}_N^{1/2}\mathbf{w})^T(\tilde{C}_N^{1/2}\mathbf{w})},$$

where $\tilde{C}_N = \begin{bmatrix} C_N & 0 \\ 0 & I \end{bmatrix}$. Let $\mathbf{z} = \tilde{C}_N^{1/2}\mathbf{w}$. Then,

$$\|S_N(\Delta t)\|_{C_N}^2 = \sup \frac{\mathbf{z}^T(\tilde{C}_N^{1/2} S_N(\Delta t)\tilde{C}_N^{-1/2})^T(\tilde{C}_N^{1/2} S_N(\Delta t)\tilde{C}_N^{-1/2})\mathbf{z}}{\mathbf{z}^T\mathbf{z}}.$$

Therefore,

$$\|S_N(\Delta t)\|_{C_N} = \|B\|_2 = \sqrt{\rho\left(B^T B\right)} \leq \sqrt{\|G\|_\infty},$$

where $B = \tilde{C}_N^{1/2} S_N(\Delta t)\tilde{C}_N^{-1/2}$, and

$$G = B^T B = \begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix}, \tag{3.10}$$

with

$$\begin{aligned} G_{11} &= C_N^{-1/2} S_{N,11}(\Delta t)^T C_N S_{N,11}(\Delta t) C_N^{-1/2} \\ &\qquad + C_N^{-1/2} S_{N,21}(\Delta t)^T S_{N,21}(\Delta t) C_N^{-1/2}, \\ G_{12} &= C_N^{-1/2} S_{N,11}(\Delta t)^T C_N S_{N,12}(\Delta t) + C_N^{-1/2} S_{N,21}(\Delta t)^T S_{N,22}(\Delta t), \\ G_{21} &= S_{N,12}(\Delta t)^T C_N S_{N,11}(\Delta t) C_N^{-1/2} + S_{N,22}(\Delta t)^T S_{N,21}(\Delta t) C_N^{-1/2}, \\ G_{22} &= S_{N,12}(\Delta t)^T C_N S_{N,12}(\Delta t) + S_{N,22}(\Delta t)^T S_{N,22}(\Delta t). \end{aligned}$$

To obtain a bound for the C_N-norm of the overall approximate solution operator $S_N(\Delta t)$, we will proceed by bounding $\|G\|_\infty$ through bounding $\|G_{ij}\|_\infty$, for $i, j = 1, 2$.

To bound the norm of each such block, we use expressions for the computed solution $\mathbf{u}^{n+1}$, $\mathbf{u}_t^{n+1}$ at time t_{n+1} in terms of $\mathbf{u}^n$ and $\mathbf{u}_t^n$. We begin with

$$\begin{aligned}\|(\mathbf{u}^{n+1},\mathbf{u}_t^{n+1})\|_{C_N}^2 &= (\mathbf{u}^{n+1})^T C_N(\mathbf{u}^{n+1}) + (\mathbf{u}_t^{n+1})^T(\mathbf{u}_t^{n+1}) \\ &= [\mathbf{u}^n]^T \bar{G}_{11}\mathbf{u}^n + [\mathbf{u}^n]^T\bar{G}_{12}\mathbf{u}_t^n + [\mathbf{u}_t^n]^T\bar{G}_{21}\mathbf{u}^n + [\mathbf{u}_t^n]^T\bar{G}_{22}\mathbf{u}_t^n,\end{aligned}$$

where

$$\begin{aligned}\bar{G}_{11} &= S_{N,11}(\Delta t)^T C_N S_{N,11}(\Delta t) + S_{N,21}(\Delta t)^T S_{N,21}(\Delta t),\\ \bar{G}_{12} &= S_{N,11}(\Delta t)^T C_N S_{N,12}(\Delta t) + S_{N,21}(\Delta t)^T S_{N,22}(\Delta t),\\ \bar{G}_{21} &= S_{N,12}(\Delta t)^T C_N S_{N,12}(\Delta t) + S_{N,22}(\Delta t)^T S_{N,21}(\Delta t),\\ \bar{G}_{22} &= S_{N,12}(\Delta t)^T C_N S_{N,12}(\Delta t) + S_{N,22}(\Delta t)^T S_{N,22}(\Delta t).\end{aligned}$$

We note that

$$G_{11} = C_N^{-1/2}\bar{G}_{11}C_N^{-1/2}, \qquad G_{12} = C_N^{-1/2}\bar{G}_{12}, \qquad G_{21} = \bar{G}_{21}C_N^{-1/2}, \qquad G_{22} = \bar{G}_{22}.$$

Therefore, we can proceed by bounding the entries of each $\bar{G}_{ij}$, for $i,j = 1,2$.

LEMMA 3.1. *Assume* $\hat{p}(\omega) = 0$ *and* $\hat{q}(\omega) = 0$ *for* $|\omega| > \omega_{\max}$. *Then the matrix* G_{11} *defined in* (3.10) *satisfies*

$$\begin{aligned}\|G_{11}\|_\infty \le 1 &+ C_{11,p}\|\tilde{p}\|_\infty \Delta t^2 N^2 + C_{11,q}\|\tilde{q}\|_\infty\Delta t^2 + C_{11,p^2}\|\tilde{p}\|_\infty^2 \Delta t^2 N^2 \\ &+ C_{11,pq}\|\tilde{p}\|_\infty\|\tilde{q}\|_\infty\Delta t^2 + C_{11,q^2}\|\tilde{q}\|_\infty^2\Delta t^2,\end{aligned}$$

where the constants $C_{11,p}$, $C_{11,q}$, C_{11,p^2}, $C_{11,pq}$, *and* C_{11,q^2} *are independent of* N *and* Δt.

Proof. Let $\hat{I}_N = \{-N/2+1,\dots,N/2\}$. From (3.7) we have

$$\begin{aligned}[\mathbf{u}^n]^T\bar{G}_{11}\mathbf{u}^n &= \mathbf{z}_{11}^T C_N \mathbf{z}_{11} + \mathbf{z}_{21}^T\mathbf{z}_{21}\\ &= \sum_{\omega\in\hat{I}_N} \overline{\hat{\mathbf{z}}_{11}(\omega)}\hat{\mathbf{z}}_{11}(\omega)(\bar{p}\omega^2+\bar{q}) + \sum_{\omega\in\hat{I}_N}\overline{\hat{\mathbf{z}}_{21}(\omega)}\hat{\mathbf{z}}_{21}(\omega)\\ &= \sum_{j\in\hat{I}_N}\sum_{k\in\hat{I}_N}\hat{u}(-j)\hat{u}(k)\left[\bar{A}+\bar{B}+\bar{C}+\bar{D}\right]_{jk},\end{aligned}$$

where, by (3.8) and (3.9), we have, for $j,k\in\hat{I}_N$,

$$\begin{aligned}\bar{A}_{jj} &= (S_{11}(l_{2,j}))^2(\bar{p}j^2+\bar{q}) + (S_{21}(l_{2,j}))^2\\ &= \cos^2\left(\sqrt{\bar{p}j^2+\bar{q}}\Delta t\right)(\bar{p}j^2+\bar{q}) + \left(-\left(\bar{p}j^2+\bar{q}\right)^{1/2}\sin\left(\sqrt{\bar{p}j^2+\bar{q}}\Delta t\right)\right)^2\\ &= \bar{p}j^2+\bar{q},\\ \bar{B}_{jk} &= -jkS_{11}(l_{2,k})M_{11,k}\hat{p}(k-j)(\bar{p}k^2+\bar{q}) + S_{11}(l_{2,k})M_{11,k}\hat{q}(k-j)(\bar{p}k^2+\bar{q})\\ &\qquad - jkS_{21}(l_{2,k})M_{21,k}\hat{p}(k-j) + S_{21}(l_{2,k})M_{21,k}\hat{q}(k-j), \qquad j\ne k,\\ \bar{C}_{jk} &= -jkS_{11}(l_{2,j})M_{11,j}\hat{p}(k-j)(\bar{p}j^2+\bar{q}) + S_{11}(l_{2,j})M_{11,j}\hat{q}(k-j)(\bar{p}j^2+\bar{q})\\ &\qquad - jkS_{21}(l_{2,j})M_{21,j}\hat{p}(k-j) + S_{21}(l_{2,j})M_{21,j}\hat{q}(k-j), \qquad j\ne k,\\ \bar{D}_{jk} &= -jk\sum_{\omega\in\hat{I}_N\setminus\{k,j\}}\omega^2\hat{p}(\omega-k)\hat{p}(j-\omega)\left(M_{11,\omega}^2(\bar{p}\omega^2+\bar{q}) + M_{21,\omega}^2\right)\\ &\qquad + (-k)\sum_{\omega\in\hat{I}_N\setminus\{k,j\}}\omega\hat{p}(\omega-k)\hat{q}(j-\omega)\left(M_{11,\omega}^2(\bar{p}\omega^2+\bar{q}) + M_{21,\omega}^2\right)\\ &\qquad + (-j)\sum_{\omega\in\hat{I}_N\setminus\{k,j\}}\omega\hat{q}(\omega-k)\hat{p}(j-\omega)\left(M_{11,\omega}^2(\bar{p}\omega^2+\bar{q}) + M_{21,\omega}^2\right)\end{aligned}$$

ETNA
Kent State University and
Johann Radon Institute (RICAM)

$$+ \sum_{\omega \in \hat{I}_N \setminus \{k,j\}} \hat{q}(\omega - k)\hat{q}(j-\omega) \left(M_{11,\omega}^2 (\overline{p}\omega^2 + \overline{q}) + M_{21,\omega}^2 \right),$$

with $\bar{A}_{jk} = 0$ for $j \neq k$, and $\bar{B}_{jj} = \bar{C}_{jj} = 0$ for $j \in \hat{I}_N$.

To obtain an upper bound for $\|G_{11}\|_\infty$, we use the following bounds for $S_{ij}(l_{2,\omega})$ and $M_{ij,\omega}$, which are the coefficients in the linear approximations of the various components of the solution operator:

$$\begin{aligned}
|S_{11}(l_{2,\omega})| &\leq \left| \cos\left(l_{2,\omega}^{1/2} \Delta t \right) \right| \leq 1, \\
|S_{21}(l_{2,\omega})| &\leq \left| -l_{2,\omega}^{1/2} \sin\left(l_{2,\omega}^{1/2} \Delta t \right) \right| \leq \left| l_{2,\omega}^{1/2} \right| l_{2,\omega}^{1/2} \Delta t = l_{2,\omega} \Delta t, \\
|M_{11,\omega}| &\leq \frac{\Delta t^2}{2}, \\
|M_{11,\omega}| &\leq \frac{\Delta t}{(\overline{p}\omega^2 + \overline{q})^{1/2}}, \\
|M_{21,\omega}| &\leq \Delta t.
\end{aligned}$$

We have multiple bounds for M_{11} so that different terms will have the same order of magnitude in terms of N and Δt. Then, for $j \neq k$, we have

$$\begin{aligned}
\left| \bar{B}_{jk} \right| &\leq \left| jk\hat{p}(k-j)(\overline{p}k^2 + \overline{q}) \frac{\Delta t^2}{2} \right| + \left| \frac{\Delta t^2}{2} \hat{q}(k-j)(\overline{p}k^2 + \overline{q}) \right| \\
&\qquad + \left| jk\hat{p}(k-j)(\overline{p}k^2 + \overline{q})\Delta t^2 \right| + \left| \hat{q}(k-j)(\overline{p}k^2 + \overline{q})\Delta t^2 \right| \\
&\leq \frac{3}{2} \Delta t^2 (\overline{p}k^2 + \overline{q}) \left(|jk\hat{p}(k-j)| + |\hat{q}(k-j)| \right), \\
|\bar{C}_{jk}| &\leq \left| jk\hat{p}(k-j)(\overline{p}j^2 + \overline{q}) \frac{\Delta t^2}{2} \right| + \left| \hat{q}(k-j)(\overline{p}j^2 + \overline{q}) \frac{\Delta t^2}{2} \right| \\
&\qquad + \left| jk\hat{p}(k-j)(\overline{p}j^2 + \overline{q})\Delta t^2 \right| + \left| \hat{q}(k-j)(\overline{p}j^2 + \overline{q})\Delta t^2 \right| \\
&\leq \frac{3}{2} \Delta t^2 (\overline{p}j^2 + \overline{q}) \left(|jk\hat{p}(k-j)| + |\hat{q}(k-j)| \right), \\
|\bar{D}_{jk}| &\leq \Bigg| jk \sum_{\omega \in \hat{I}_N \setminus \{k,j\}} \omega^2 |\hat{p}(\omega - k)\hat{p}(j-\omega)| \left(2\Delta t^2 \right) \\
&\qquad + k \sum_{\omega \in \hat{I}_N \setminus \{k,j\}} \omega |\hat{p}(\omega - k)\hat{q}(j-\omega)| \left(2\Delta t^2 \right) \\
&\qquad + j \sum_{\omega \in \hat{I}_N \setminus \{k,j\}} \omega |\hat{q}(\omega - k)\hat{p}(j-\omega)| \left(2\Delta t^2 \right) \\
&\qquad + \sum_{\omega \in \hat{I}_N \setminus \{k,j\}} |\hat{q}(\omega - k)\hat{q}(j-\omega)| \left(2\Delta t^2 \right) \Bigg|.
\end{aligned}$$

From these bounds, we obtain

$$\begin{aligned}
&\|G_{11}\|_\infty \\
&\leq \max_{1 \leq j \leq N} \sum_{k \in \hat{I}_N \setminus j} (\bar{p}j^2 + \bar{q})^{-1/2} \left| \bar{A}_{jk} + \bar{B}_{jk} + \bar{C}_{jk} + \bar{D}_{jk} \right| (\bar{p}k^2 + \bar{q})^{-1/2}
\end{aligned}$$

$$\le \max_{1\le j\le N}\Bigg\{(\bar p j^2+\bar q)^{-1/2}(\bar p j^2+\bar q)^{-1/2}(\bar p j^2+\bar q)$$
$$+\frac{3}{2}\sum_{k\in\hat I_N\setminus j}\left|jk\hat p(k-j)(\bar p j^2+\bar q)^{-1/2}(\bar p k^2+\bar q)^{1/2}\Delta t^2\right|$$
$$+\frac{3}{2}\sum_{k\in\hat I_N\setminus j}\left|\hat q(k-j)(\bar p j^2+\bar q)^{-1/2}(\bar p k^2+\bar q)^{1/2}\Delta t^2\right|$$
$$+\frac{3}{2}\sum_{k\in\hat I_N\setminus j}\left|jk\hat p(k-j)(\bar p j^2+\bar q)^{1/2}(\bar p k^2+\bar q)^{-1/2}\Delta t^2\right|$$
$$+\frac{3}{2}\sum_{k\in\hat I_N\setminus j}\left|\hat q(k-j)(\bar p j^2+\bar q)^{1/2}(\bar p k^2+\bar q)^{-1/2}\Delta t^2\right|$$
$$+\sum_{k\in\hat I_N}\left|jk(\bar p j^2+\bar q)^{-1/2}(\bar p k^2+\bar q)^{-1/2}\sum_{\omega\in\hat I_N\setminus\{k,j\}}\omega^2|\hat p(\omega-k)\hat p(j-\omega)|\Delta t^2\right|$$
$$+\sum_{k\in\hat I_N}\left|k(\bar p j^2+\bar q)^{-1/2}(\bar p k^2+\bar q)^{-1/2}\sum_{\omega\in\hat I_N\setminus\{k,j\}}\omega|\hat p(\omega-k)\hat q(j-\omega)|\Delta t^2\right|$$
$$+\sum_{k\in\hat I_N}\left|j(\bar p j^2+\bar q)^{-1/2}(\bar p k^2+\bar q)^{-1/2}\sum_{\omega\in\hat I_N\setminus\{k,j\}}\omega|\hat q(\omega-k)\hat p(j-\omega)|\Delta t^2\right|$$
$$+\sum_{k\in\hat I_N}\left|(\bar p j^2+\bar q)^{-1/2}(\bar p k^2+\bar q)^{-1/2}\sum_{\omega\in\hat I_N\setminus\{k,j\}}|\hat q(\omega-k)\hat q(j-\omega)|\Delta t^2\right|\Bigg\}$$
$$\le \max_{1\le j\le N}\Bigg\{1+\frac{3}{2}j\Delta t^2\|\tilde p\|_\infty(\bar p j^2+\bar q)^{-1/2}\sum_{k\in I_j}\left|k(\bar p k^2+\bar q)^{1/2}\right|$$
$$+\frac{3}{2}\Delta t^2\|\tilde q\|_\infty(\bar p j^2+\bar q)^{-1/2}\sum_{k\in I_j}\left|(\bar p k^2+\bar q)^{1/2}\right| \tag{3.11a}$$
$$+\frac{3}{2}j\Delta t^2\|\tilde p\|_\infty(\bar p j^2+\bar q)^{1/2}\sum_{k\in I_j}\left|k(\bar p k^2+\bar q)^{-1/2}\right| \tag{3.11b}$$
$$+\frac{3}{2}\Delta t^2\|\tilde q\|_\infty(\bar p j^2+\bar q)^{1/2}\sum_{k\in I_j}\left|(\bar p k^2+\bar q)^{-1/2}\right| \tag{3.11c}$$
$$+j\Delta t^2\|\tilde p\|_\infty^2(\bar p j^2+\bar q)^{-1/2}\sum_{k\in\hat I_N}\left|k(\bar p k^2+\bar q)^{-1/2}\sum_{\omega\in I_j\cap I_k}\omega^2\right| \tag{3.11d}$$
$$+\Delta t^2\|\tilde p\|_\infty\|\tilde q\|_\infty(\bar p j^2+\bar q)^{-1/2}\sum_{k\in\hat I_N}\left|k(\bar p k^2+\bar q)^{-1/2}\sum_{\omega\in I_j\cap I_k}\omega\right| \tag{3.11e}$$
$$+j\Delta t^2\|\tilde p\|_\infty\|\tilde q\|_\infty(\bar p j^2+\bar q)^{-1/2}\sum_{k\in\hat I_N}\left|(\bar p k^2+\bar q)^{-1/2}\sum_{\omega\in I_j\cap I_k}\omega\right| \tag{3.11f}$$

ETNA
Kent State University and
Johann Radon Institute (RICAM)

$$+\Delta t^2\|\tilde{q}\|_\infty^2(\overline{p}j^2+\overline{q})^{-1/2}\sum_{k\in\hat{I}_N}\left|(\overline{p}k^2+\overline{q})^{-1/2}\sum_{\omega\in I_j\cap I_k}1\right|\Bigg\}. \tag{3.11g}$$

We can bound each of the summations in (3.11) as in the following examples.

- We first derive a bound for

$$(\overline{p}j^2+\overline{q})^{1/2}\sum_{k\in I_j}\left|(\overline{p}k^2+\overline{q})^{-1/2}\right|. \tag{3.12}$$

For $|j|>\omega_{\max}$, we have

$$\begin{aligned}(\overline{p}j^2+\overline{q})^{1/2}\sum_{k\in I_j}\left|(\overline{p}k^2+\overline{q})^{-1/2}\right| &\le (\overline{p}j^2+\overline{q})^{1/2}\sum_{k\in I_j}\frac{1}{(\overline{p}(|j|-\omega_{\max})^2+\overline{q})^{1/2}}\\ &\le (\overline{p}j^2+\overline{q})^{1/2}\frac{2\omega_{\max}}{(\overline{p}(|j|-\omega_{\max})^2+\overline{q})^{1/2}}.\end{aligned}$$

As $|j|\to\infty$, we obtain

$$\lim_{j\to\infty}(\overline{p}j^2+\overline{q})^{1/2}\sum_{k\in I_j}\left|(\overline{p}k^2+\overline{q})^{-1/2}\right|\le 2\omega_{\max}.$$

Therefore, the expression (3.12) can be bounded independently of N.

- Next, we derive a bound for

$$(\overline{p}j^2+\overline{q})^{-1/2}\sum_{k\in\hat{I}_N}\left|k(\overline{p}k^2+\overline{q})^{-1/2}\sum_{\omega\in I_j\cap I_k}\omega\right|. \tag{3.13}$$

We have

$$\begin{aligned}\sum_{k\in\hat{I}_N}\left|k(\overline{p}k^2+\overline{q})^{-1/2}\sum_{\omega\in I_j\cap I_k}\omega\right| &\le \sum_{k\in\hat{I}_N}\sum_{\omega\in I_j\cap I_k}\left|k(\overline{p}k^2+\overline{q})^{-1/2}\omega\right|\\ &\le \sum_{\omega\in I_j}|\omega|\sum_{k\in I_\omega}\left|k(\overline{p}k^2+\overline{q})^{-1/2}\right|\\ &\le \sum_{\omega\in I_j}|\omega|\sum_{\eta\in I_0}\left|(\omega+\eta)(\overline{p}(\omega+\eta)^2+\overline{q})^{-1/2}\right|.\end{aligned}$$

If $j>2\omega_{\max}$, then $\omega>\omega_{\max}$, and

$$\sum_{k\in\hat{I}_N}\left|k(\overline{p}k^2+\overline{q})^{-1/2}\sum_{\omega\in I_j\cap I_k}\omega\right|\le\sum_{\omega\in I_j}|j+\omega_{\max}|\sum_{\eta\in I_0}\left|\frac{j+2\omega_{\max}}{(\overline{p}(j-2\omega_{\max})^2+\overline{q})^{1/2}}\right|.$$

We then have

$$\lim_{j\to\infty}(\overline{p}j^2+\overline{q})^{-1/2}\sum_{k\in\hat{I}_N}\left|k(\overline{p}k^2+\overline{q})^{-1/2}\sum_{\omega\in I_j\cap I_k}\omega\right|\le\frac{1}{\overline{p}}.$$

ETNA
Kent State University and
Johann Radon Institute (RICAM)

We conclude that the expression (3.13) can also be bounded independently of N.

Using a similar approach to bound the remaining summations in (3.11), we obtain

$$\begin{aligned}\|G_{11}\|_\infty \leq 1 + C_{11,p}\|\tilde{p}\|_\infty \Delta t^2 N^2 + C_{11,q}\|\tilde{q}\|_\infty \Delta t^2 + C_{11,p^2}\|\tilde{p}\|_\infty^2 \Delta t^2 N^2 \\ + C_{11,pq}\|\tilde{p}\|_\infty \|\tilde{q}\|_\infty \Delta t^2 + C_{11,q^2}\|\tilde{q}\|_\infty^2 \Delta t^2,\end{aligned}$$

with constants $C_{11,p}$, $C_{11,q}$, C_{11,p^2}, $C_{11,pq}$, C_{11,q^2} that are independent of N and Δt, which completes the proof. ∎

Using the same approach, we find that the matrix G_{22} defined in (3.10) satisfies a bound of the same form as that of G_{11} with appropriate constant factors.

LEMMA 3.2. *Assume $\hat{p}(\omega) = 0$ and $\hat{q}(\omega) = 0$ for $|\omega| > \omega_{\max}$. Then the matrix G_{12} defined in (3.10) satisfies*

$$\begin{aligned}\|G_{12}\|_\infty \leq C_{12,p}\|\tilde{p}\|_\infty \Delta t N + C_{12,q}\|\tilde{q}\|_\infty \Delta t + C_{12,p^2}\|\tilde{p}\|_\infty^2 \Delta t N \\ + C_{12,pq}\|\tilde{p}\|_\infty \|\tilde{q}\|_\infty \Delta t + C_{12,q^2}\|\tilde{q}\|_\infty^2 \Delta t,\end{aligned}$$

where each constant is independent of N and Δt.

Proof. From (3.7), we have

$$\begin{aligned}[\mathbf{u}^n]^T \bar{G}_{12} \mathbf{u}_t^n &= \sum_{\omega \in \hat{I}_N} \overline{\hat{\mathbf{z}}_{11}(\omega)} \hat{\mathbf{z}}_{12}(\omega)(\bar{p}\omega^2 + \bar{q}) + \sum_{\omega \in \hat{I}_N} \overline{\hat{\mathbf{z}}_{21}(\omega)} \hat{\mathbf{z}}_{22}(\omega) \\ &= \sum_{j \in \hat{I}_N} \sum_{k \in \hat{I}_N} \hat{u}(-j)\hat{u}_t(k) \left[\bar{A} + \bar{B} + \bar{C} + \bar{D}\right]_{jk},\end{aligned}$$

where, for $j, k \in \hat{I}_N$,

$$\begin{aligned}\bar{A}_{jj} &= S_{11}(l_{2,j})S_{12}(l_{2,j})(\bar{p}j^2 + \bar{q}) + S_{21}(l_{2,j})S_{22}(l_{2,j}) \\ &= \cos\left((\bar{p}j^2 + \bar{q})^{1/2}\Delta t\right)(\bar{p}j^2 + q)^{-1/2} \sin\left(\sqrt{\bar{p}j^2 + q}\Delta t\right)(\bar{p}j^2 + \bar{q}) \\ &\quad + \left(-(\bar{p}j^2 + \bar{q})^{1/2} \sin\left((\bar{p}j^2 + \bar{q})^{1/2}\Delta t\right) \cos\left(\sqrt{\bar{p}j^2 + \bar{q}}\Delta t\right)\right) \\ &= \cos\left((\bar{p}j^2 + \bar{q})^{1/2}\Delta t\right)(\bar{p}j^2 + q)^{1/2} \sin\left(\sqrt{\bar{p}j^2 + q}\Delta t\right) \\ &\quad - (\bar{p}j^2 + \bar{q})^{1/2} \sin\left((\bar{p}j^2 + \bar{q})^{1/2}\Delta t\right) \cos\left(\sqrt{\bar{p}j^2 + \bar{q}}\Delta t\right) = 0, \\ \bar{B}_{jk} &= -jkS_{12}(l_{2,k})M_{11,k}\hat{p}(k-j)(\bar{p}k^2 + \bar{q}) + S_{12}(l_{2,k})M_{11,k}\hat{q}(k-j)(\bar{p}k^2 + \bar{q}) \\ &\quad - jkS_{22}(l_{2,k})M_{21,k}\hat{p}(k-j) + S_{22}(l_{2,k})M_{21,k}\hat{q}(k-j), \qquad j \neq k, \\ \bar{C}_{jk} &= -jkS_{11}(l_{2,j})M_{12,j}\hat{p}(k-j)(\bar{p}j^2 + \bar{q}) + S_{11}(l_{2,j})M_{12,j}\hat{q}(k-j)(\bar{p}j^2 + \bar{q}) \\ &\quad - jkS_{21}(l_{2,j})M_{22,j}\hat{p}(k-j) + S_{21}(l_{2,j})M_{22,j}\hat{q}(k-j), \qquad j \neq k, \\ \bar{D}_{jk} &= -jk \sum_{\omega \in \hat{I}_N \setminus \{k,j\}} \omega^2 \hat{p}(j-\omega)\hat{p}(\omega-k)\left(M_{11,\omega}M_{12,\omega}(\bar{p}\omega^2 + \bar{q}) + M_{21,\omega}M_{22,\omega}\right) \\ &\quad + (-j) \sum_{\omega \in \hat{I}_N \setminus \{k,j\}} \omega \hat{p}(j-\omega)\hat{q}(\omega-k)\left(M_{11,\omega}M_{12,\omega}(\bar{p}\omega^2 + \bar{q}) + M_{21,\omega}M_{22,\omega}\right) \\ &\quad + (-k) \sum_{\omega \in \hat{I}_N \setminus \{k,j\}} \omega \hat{q}(j-\omega)\hat{p}(\omega-k)\left(M_{11,\omega}M_{12,\omega}(\bar{p}\omega^2 + \bar{q}) + M_{21,\omega}M_{22,\omega}\right) \\ &\quad + \sum_{\omega \in \hat{I}_N \setminus \{k,j\}} \hat{q}(j-\omega)\hat{q}(\omega-k)\left(M_{11,\omega}M_{12,\omega}(\bar{p}\omega^2 + \bar{q}) + M_{21,\omega}M_{22,\omega}\right),\end{aligned}$$

with $\bar{A}_{jk} = 0$ for $j \neq k$, and $\bar{B}_{jj} = \bar{C}_{jj} = 0$ for $j \in \hat{I}_N$.

To obtain an upper bound for $\|G_{12}\|_\infty$, we use the following bounds for $S_{ij}(l_{2,\omega})$ and $M_{ij,\omega}$:

$$
\begin{aligned}
|S_{12}(l_{2,\omega})| &\leq \left| l_{2,\omega}^{-1/2} \sin(l_{2,\omega}^{1/2}\Delta t) \right| \leq \left| l_{2,\omega}^{-1/2} \right| l_{2,\omega}^{1/2}\Delta t = \Delta t, \\
|M_{11,\omega}| &= |M_{22,\omega}| \leq \frac{2}{\overline{p}\omega^2 + \overline{q}}, \\
|M_{12,\omega}| &\leq \frac{\Delta t}{\overline{p}\omega^2 + \overline{q}}.
\end{aligned}
$$

Then we have

$$
\begin{aligned}
|\bar{B}_{jk}| &\leq 3\left(|jk\hat{p}(k-j)\Delta t| + |\hat{q}(k-j)\Delta t|\right), \\
|\bar{C}_{jk}| &\leq 3\left(|jk\hat{p}(k-j)\Delta t| + |\hat{q}(k-j)\Delta t|\right),
\end{aligned}
$$

and

$$
\begin{aligned}
|\bar{D}_{jk}| \leq \Bigg| & jk \sum_{\omega \in \hat{I}_N \setminus \{k,j\}} \omega^2 |\hat{p}(j-\omega)\hat{p}(\omega-k)| \frac{4\Delta t}{\overline{p}\omega^2 + \overline{q}} \\
& + j \sum_{\omega \in \hat{I}_N \setminus \{k,j\}} \omega |\hat{p}(j-\omega)\hat{q}(\omega-k)| \frac{4\Delta t}{\overline{p}\omega^2 + \overline{q}} \\
& + k \sum_{\omega \in \hat{I}_N \setminus \{k,j\}} \omega |\hat{q}(j-\omega)\hat{p}(\omega-k)| \frac{4\Delta t}{\overline{p}\omega^2 + \overline{q}} \\
& + \sum_{\omega \in \hat{I}_N \setminus \{k,j\}} |\hat{q}(j-\omega)\hat{q}(\omega-k)| \frac{4\Delta t}{\overline{p}\omega^2 + \overline{q}} \Bigg|.
\end{aligned}
$$

From these bounds, we obtain

$$
\begin{aligned}
&\|G_{12}\|_\infty \\
&\quad \leq \max_{1 \leq j \leq N} \sum_{k \in \hat{I}_N \setminus j} \left| \left(\bar{A}_{jk} + \bar{B}_{jk} + \bar{C}_{jk} + \bar{D}_{jk} \right) (\bar{p}j^2 + \bar{q})^{-1/2} \right| \\
&\quad \leq \max_{1 \leq j \leq N} \Bigg\{ 6 \sum_{k \in \hat{I}_N \setminus j} \left| jk\hat{p}(k-j)(\bar{p}j^2 + \bar{q})^{-1/2}\Delta t \right| \\
&\qquad + 6 \sum_{k \in \hat{I}_N \setminus j} \left| \hat{q}(k-j)(\bar{p}j^2 + \bar{q})^{-1/2}\Delta t \right| \\
&\qquad + \sum_{k \in \hat{I}_N \setminus j} \left| jk \sum_{\omega \in \hat{I}_N \setminus \{k,j\}} \omega^2 |\hat{p}(j-\omega)\hat{p}(\omega-k)| (\bar{p}j^2 + \bar{q})^{-1/2} \frac{4\Delta t}{\overline{p}\omega^2 + \overline{q}} \right| \\
&\qquad + \sum_{k \in \hat{I}_N \setminus j} \left| j \sum_{\omega \in \hat{I}_N \setminus \{k,j\}} \omega |\hat{p}(j-\omega)\hat{q}(\omega-k)| (\bar{p}j^2 + \bar{q})^{-1/2} \frac{4\Delta t}{\overline{p}\omega^2 + \overline{q}} \right| \\
&\qquad + \sum_{k \in \hat{I}_N \setminus j} \left| k \sum_{\omega \in \hat{I}_N \setminus \{k,j\}} \omega |\hat{q}(j-\omega)\hat{p}(\omega-k)| (\bar{p}j^2 + \bar{q})^{-1/2} \frac{4\Delta t}{\overline{p}\omega^2 + \overline{q}} \right|
\end{aligned}
$$

ETNA
Kent State University and
Johann Radon Institute (RICAM)

$$
\begin{aligned}
&\quad + \sum_{k\in\hat{I}_N\setminus j}\left|\sum_{\omega\in\hat{I}_N\setminus\{k,j\}}|\hat{q}(j-\omega)\hat{q}(\omega-k)|(\bar{p}j^2+\bar{q})^{-1/2}\frac{4\Delta t}{\bar{p}\omega^2+\bar{q}}\right|\Bigg\}\\
&\le \max_{1\le j\le N}\Bigg\{6j\|\tilde{p}\|_\infty(\bar{p}j^2+\bar{q})^{-1/2}\Delta t\sum_{k\in I_j}|k|\\
&\quad + 6\|\tilde{q}\|_\infty(\bar{p}j^2+\bar{q})^{-1/2}\Delta t(2\omega_{max})\\
&\quad + j\|\tilde{p}\|_\infty^2(\bar{p}j^2+\bar{q})^{-1/2}\Delta t\frac{4}{\bar{p}}\sum_{k\in\hat{I}_N}\sum_{\omega\in I_j\cap I_k}|k|\\
&\quad + \|\tilde{p}\|_\infty\|\tilde{q}\|_\infty(\bar{p}j^2+\bar{q})^{-1/2}\Delta t\frac{4}{\bar{p}}\sum_{\omega\in I_j}\left(2j\omega_{max}+\sum_{k\in I_\omega}|k|\right)\\
&\quad + \|\tilde{q}\|_\infty^2(\bar{p}j^2+\bar{q})^{-1/2}\Delta t\frac{4}{\bar{q}}(4\omega_{max}^2)\Bigg\}.
\end{aligned}
$$

These summations can be bounded as in the proof of Lemma 3.1. As an example, we obtain a bound for

$$(\bar{p}j^2+\bar{q})^{-1/2}\sum_{k\in I_j}|k|\,. \tag{3.14}$$

If $|j|>\omega_{\max}$, then

$$\sum_{k\in I_j}|k| = \sum_{\eta\in I_0}|j-\eta| = \sum_{\eta=1}^{\omega_{\max}}|j-\eta|+|j+\eta| = \sum_{\eta=1}^{\omega_{\max}}2|j| = 2|j|\omega_{\max}.$$

It follows that

$$\lim_{j\to\infty}(\bar{p}j^2+\bar{q})^{-1/2}\sum_{k\in I_j}|k| = \lim_{j\to\infty}(\bar{p}j^2+\bar{q})^{-1/2}2|j|\omega_{\max} = \frac{2\omega_{\max}}{\bar{p}}.$$

That is, the expression (3.14) is bounded independently of N, whereas it would be $O(N)$ if $p(x)$ was not bandlimited.

Proceeding in a similar manner for the remaining summations, we conclude that there exist constants $C_{12,p}$, $C_{12,q}$, C_{12,p^2}, $C_{12,pq}$, and C_{12,q^2} such that

$$
\begin{aligned}
\|G_{12}\|_\infty \le C_{12,p}\|\tilde{p}\|_\infty\Delta tN + C_{12,q}\|\tilde{q}\|_\infty\Delta t + C_{12,p^2}\|\tilde{p}\|_\infty^2\Delta tN\\
+ C_{12,pq}\|\tilde{p}\|_\infty\|\tilde{q}\|_\infty\Delta t + C_{12,q^2}\|\tilde{q}\|_\infty^2\Delta t,
\end{aligned}
$$

which completes the proof. □

Using the same approach, it can be shown that the matrix G_{21} in (3.10) satisfies a bound of the same form as that of G_{12}.

We now prove a result that gives us reason to believe that the second-order KSS method applied to (2.1)–(2.3) may be unstable.

THEOREM 3.3. *Assume $\hat{p}(\omega) = 0$ and $\hat{q}(\omega) = 0$ for $|\omega| > \omega_{\max}$. Then the solution operator $S_N(\Delta t)$ satisfies*

$$\|S_N(\Delta t)\|_{C_N} \leq 1 + (C_p\|\tilde{p}\|_\infty N + C_q\|\tilde{q}\|_\infty)\Delta t, \tag{3.15}$$

where the constants C_p and C_q are independent of N and Δt.

Proof. From

$$\left\| \begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix} \right\|_\infty \leq \max\left\{\|G_{11}\|_\infty + \|G_{12}\|_\infty, \|G_{21}\|_\infty + \|G_{22}\|_\infty\right\},$$

we have, for some $k \in \{1, 2\}$,

$$\|S_N(\Delta t)\|_{C_N} = \|B\|_2 \leq \sqrt{\|G\|_\infty} \leq \sqrt{\|G_{k1}\|_\infty + \|G_{k2}\|_\infty}.$$

From Lemma 3.1 and Lemma 3.2, we have

$$\begin{aligned} \|G\|_\infty \leq 1 &+ \Delta t N \left(C_{k2,p}\|\tilde{p}\|_\infty + C_{k2,p^2}\|\tilde{p}\|_\infty^2\right) + \Delta t^2 N^2 \left(C_{k1,p}\|\tilde{p}\|_\infty + C_{k1,p^2}\|\tilde{p}\|_\infty^2\right) \\ &+ \Delta t \left(C_{k2,q}\|\tilde{q}\|_\infty + C_{k2,pq}\|\tilde{p}\|_\infty\|\tilde{q}\|_\infty + C_{k2,q^2}\|\tilde{q}\|_\infty^2\right) \\ &+ \Delta t^2 \left(C_{k1,q}\|\tilde{q}\|_\infty + C_{k1,pq}\|\tilde{p}\|_\infty\|\tilde{q}\|_\infty + C_{k1,q^2}\|\tilde{q}\|_\infty^2\right). \end{aligned}$$

Let

$$\begin{aligned} R_1 &= N\left(C_{k2,p}\|\tilde{p}\|_\infty + C_{k2,p^2}\|\tilde{p}\|_\infty^2\right) + C_{k2,q}\|\tilde{q}\|_\infty + C_{k2,pq}\|\tilde{p}\|_\infty\|\tilde{q}\|_\infty + C_{k2,q^2}\|\tilde{q}\|_\infty^2, \\ R_2 &= N^2\left(C_{k1,p}\|\tilde{p}\|_\infty + C_{k1,p^2}\|\tilde{p}\|_\infty^2\right) + C_{k1,q}\|\tilde{q}\|_\infty + C_{k1,pq}\|\tilde{p}\|_\infty\|\tilde{q}\|_\infty + C_{k1,q^2}\|\tilde{q}\|_\infty^2, \end{aligned}$$

and $R = \max\{R_2^{1/2}, R_1/2\}$. We then have

$$\|S_N(\Delta t)\|_{C_N} \leq \sqrt{\|G\|_\infty} \leq 1 + R\Delta t \leq 1 + (C_p\|\tilde{p}\|_\infty N + C_q\|\tilde{q}\|_\infty)\Delta t,$$

from which the result follows. □

While Theorem 3.3 does not prove that the bound in (3.15) is sharp, numerical experiments indicate that it actually is. In the case of $p(x) \equiv$ constant, we obtain a more favorable stability result.

COROLLARY 3.4. *Assume the leading coefficient $p(x)$ is constant. Then, under the assumptions of Theorem 3.3,*

$$\|S_N(\Delta t)\|_{C_N} \leq 1 + C_q\|\tilde{q}\|_\infty\Delta t.$$

Proof. Because the leading coefficient $p(x)$ is constant, we have $\tilde{p}(x) = p(x) - \bar{p} \equiv 0$. The result follows immediately from the last line of the proof of Theorem 3.3. □

Therefore, a second-order KSS method applied to (2.1)–(2.4), under the assumptions that $p(x)$ is constant and $q(x)$ is bandlimited, is unconditionally stable.

3.2. Consistency. For the remainder of this convergence analysis, we assume that the coefficient $p(x)$ from (2.4) is constant since stability has been proved only for this case.

Before we obtain an estimate of the local truncation error, we introduce additional notation. We first define the restriction operator

$$\mathcal{R}_N\mathbf{f}(x) = \mathbf{f}(\mathbf{x}_N)$$

ETNA
Kent State University and
Johann Radon Institute (RICAM)

and the interpolation operator

$$\mathcal{T}_N\mathbf{g} = \mathcal{T}_N\begin{bmatrix}\mathbf{g}_1\\ \mathbf{g}_2\end{bmatrix} = \begin{bmatrix}\sum_{\omega=-N/2+1}^{N/2} e^{i\omega x}\tilde{g}_1(\omega)\\ \\ \sum_{\omega=-N/2+1}^{N/2} e^{i\omega x}\tilde{g}_2(\omega)\end{bmatrix},$$

where, for $i = 1, 2$,

$$\tilde{g}_i(\omega) = \frac{1}{N}\sum_{j=1}^{N} e^{-i\omega x_j}[\mathbf{g}_i]_j\,.$$

Then, the operator $\mathcal{I}_N = \mathcal{T}_N\mathcal{R}_N$ on $L^2([0, 2\pi])\times L^2([0, 2\pi])$ computes the Fourier interpolant of each component function. By contrast, if we define

$$\hat{\mathcal{R}}_N\mathbf{f}(x) = \hat{\mathcal{R}}_N\begin{bmatrix}f_1(x)\\ f_2(x)\end{bmatrix} = \begin{bmatrix}\sum_{\omega=-N/2+1}^{N/2} e^{i\omega\mathbf{x}_N}\hat{f}_1(\omega)\\ \\ \sum_{\omega=-N/2+1}^{N/2} e^{i\omega\mathbf{x}_N}\hat{f}_2(\omega)\end{bmatrix},$$

where, for $i = 1, 2$,

$$\hat{f}_i(\omega) = \frac{1}{2\pi}\int_0^{2\pi} e^{-i\omega x} f_i(x)\,dx,$$

then the operator $\mathcal{P}_N = \mathcal{T}_N\hat{\mathcal{R}}_N$ on $L^2([0, 2\pi]) \times L^2([0, 2\pi])$ is the orthogonal projection operator onto $\operatorname{span}\{e^{i\omega x}\}_{\omega=-N/2+1}^{N/2}$. Finally, the continuous approximate solution operator $\tilde{S}_N(\Delta t) : L^2([0, 2\pi]) \to L^2([0, 2\pi])$ is defined by

$$\tilde{S}_N(\Delta t) = \mathcal{T}_N S_N(\Delta t)\mathcal{R}_N.$$

THEOREM 3.5. *Let* $\mathbf{f} \in H_p^{m+1}([0, 2\pi]) \times H_p^m([0, 2\pi])$ *for* $m \geq 4$. *Then, for the problem* (2.1)–(2.4) *on the domain* $(0, 2\pi) \times (0, T)$*, with* $p(x)$ *constant and* $q(x)$ *bandlimited, the two-node block KSS method with prescribed nodes* (3.6) *is consistent. That is, the local truncation error satisfies*

$$\frac{1}{\Delta t}\left\|\mathcal{I}_N\exp[\tilde{L}\Delta t]\mathbf{f} - \tilde{S}_N(\Delta t)\mathbf{f}\right\|_C \leq C_1\Delta x^{m-1} + C_2\Delta t^2,$$

where the constants C_1 *and* C_2 *are independent of* Δx *and* Δt.

Proof. We split the local truncation error into two components:

$$\begin{aligned}E_1(\Delta t, \Delta x) &= \mathcal{I}_N\exp(\tilde{L}\Delta t)\mathbf{f}(x) - \mathcal{T}_N\exp(\tilde{L}_N\Delta t)\mathcal{R}_N\mathbf{f}(x)\\ E_2(\Delta t, \Delta x) &= \mathcal{T}_N\exp(\tilde{L}_N\Delta t)\mathcal{R}_N\mathbf{f}(x) - \tilde{S}_N(\Delta t)\mathbf{f}(x)\\ &= \mathcal{T}_N[\exp(\tilde{L}_N\Delta t) - S_N(\Delta t)]\mathcal{R}_N\mathbf{f}(x).\end{aligned}$$

First, we bound $E_1(\Delta t, \Delta x)$. Because of the regularity of $\mathbf{f}$, we have

$$\|\mathbf{f} - \mathbf{f}_N\|_C \leq C_0\Delta x^m$$

for some constant C_0 (see [17, Theorem 2.16]). Next, we note that the exact solution $\mathbf{v}(x, t) = \exp[\tilde{L}t]\mathbf{f}(x)$ has the spectral decomposition

$$\mathbf{v}(x, t) = \sum_{k=1}^{\infty} e^{\mu_k t}\mathbf{v}_k(x)\langle\mathbf{v}_k, \mathbf{f}\rangle,$$

ETNA
Kent State University and
Johann Radon Institute (RICAM)

where $\{\mu_k\}_{k=1}^{\infty}$ are the purely imaginary eigenvalues of $\tilde{L}$ and $\{\mathbf{v}_k(x)\}_{k=1}^{\infty}$ are the corresponding orthonormal eigenfunctions, each of which belongs to $C_p^{\infty}[0, 2\pi]$. Using this spectral decomposition, it can be shown using an approach similar to that in [7, Section 7.2, Theorem 6] for other hyperbolic PDEs that if $\mathbf{f} \in H_p^{m+1}([0, 2\pi]) \times H_p^m([0, 2\pi])$, then $\mathbf{v}(x, t) \in L^2(0, T, H_p^{m+1}([0, 2\pi]) \times H_p^m([0, 2\pi]))$. That is, the regularity of $\mathbf{f}(x)$ is preserved in $\mathbf{v}(x, \Delta t)$ for each fixed $\Delta t > 0$. Therefore, there exists a constant C_T such that

$$\|(I - \mathcal{I}_N)\mathbf{v}(\cdot, \Delta t)\|_C \leq C_T \Delta x^m, \qquad 0 < \Delta t \leq T. \tag{3.16}$$

Using an approach based on [2] and applied in [33], we write $E_1(\Delta t, \Delta x)$ as

$$\begin{aligned} \mathbf{e}_N(x, t) &= \mathcal{I}_N \mathbf{v}(x, t) - \mathcal{T}_N \exp(\tilde{L}_N t)\mathcal{R}_N \mathbf{f}(x) \\ &= \mathcal{I}_N \mathbf{v}(x, t) - \exp(\mathcal{I}_N \tilde{L} t)\mathcal{I}_N \mathbf{f}(x)\,. \end{aligned}$$

Then, $\mathbf{e}_N(x, t)$ solves the IVP

$$\frac{\partial}{\partial t}\mathbf{e}_N = \mathcal{I}_N \tilde{L}\mathbf{e}_N + \mathcal{I}_N \tilde{L}(I - \mathcal{I}_N)\mathbf{v}, \qquad \mathbf{e}_N(x, 0) = \mathbf{0},$$

and therefore

$$\mathbf{e}_N(x, \Delta t) = \int_0^{\Delta t} e^{\mathcal{I}_N \tilde{L}(\Delta t - \tau)} \mathcal{I}_N \tilde{L}(I - \mathcal{I}_N)\mathbf{v}(x, \tau)\, d\tau.$$

From (3.16) and applying [7, Section 7.2, Theorem 6], it follows that

$$\begin{aligned} \|E_1(\Delta t, \Delta x)\|_C &= \|\mathbf{e}_N(x, \Delta t)\|_C \\ &\leq \Delta t \max_{0 \leq \tau \leq \Delta t} \left\| \mathcal{I}_N \tilde{L} e^{\mathcal{I}_N \tilde{L}(\Delta t - \tau)}(I - \mathcal{I}_N)\mathbf{v}(\cdot, \tau) \right\|_C \\ &\leq C_1 \Delta t \Delta x^{m-1}, \end{aligned}$$

where the constant C_1 is independent of Δx and Δt. Here we note that because the coefficients of L are assumed to be constant or bandlimited, $E_1(\Delta t, \Delta x)$ does not include an aliasing error.

Now, we examine $E_2(\Delta t, \Delta x)$. If we let

$$\mathcal{R}_N \mathbf{f} = \begin{bmatrix} \mathbf{f}_{N,1} \\ \mathbf{f}_{N,2} \end{bmatrix}, \qquad E_2(\Delta t, \Delta x) = \mathcal{T}_N \begin{bmatrix} \mathbf{e}_{N,1} \\ \mathbf{e}_{N,2} \end{bmatrix},$$

then we have

$$\begin{aligned} \mathbf{e}_{N,1} &= [\cos(L_N^{1/2}\Delta t) - S_{N,11}(\Delta t)]\mathbf{f}_{N,1} + [L_N^{-1/2}\sin(L_N^{1/2}\Delta t) - S_{N,12}(\Delta t)]\mathbf{f}_{N,2}, \\ \mathbf{e}_{N,2} &= [-L_N^{1/2}\sin(L_N^{1/2}\Delta t) - S_{N,21}(\Delta t)]\mathbf{f}_{N,1} + [\cos(L_N^{1/2}\Delta t) - S_{N,22}(\Delta t)]\mathbf{f}_{N,2}. \end{aligned}$$

We have, by Parseval's identity,

$$\|E_2(\Delta t, \Delta x)\|_C^2 = 2\pi \sum_{\omega=-N/2+1}^{N/2} (\overline{p}\omega^2 + \overline{q}) \left| \frac{1}{N}\hat{\mathbf{e}}_\omega^H \mathbf{e}_{N,1} \right|^2 + \left| \frac{1}{N}\hat{\mathbf{e}}_\omega^H \mathbf{e}_{N,2} \right|^2.$$

For each $\omega = -N/2 + 1, \ldots, N/2$ and $i, j = 1, 2$, we use the polynomial interpolation error in $S_{N,ij}$ to obtain

ETNA
Kent State University and
Johann Radon Institute (RICAM)

$$
\begin{aligned}
\hat{\mathbf{e}}_\omega^H \mathbf{e}_{N,1} &= \frac{1}{2}\,\frac{\partial^2}{\partial\lambda^2}\left[\cos(\lambda^{1/2}\Delta t)\right]\Big|_{\lambda=\xi_{\omega,11}} \hat{\mathbf{e}}_\omega^H (L_N - l_{1,\omega}I)(L_N - l_{2,\omega}I)\mathbf{f}_{N,1} \\
&\quad + \frac{1}{2}\,\frac{\partial^2}{\partial\lambda^2}\left[\lambda^{-1/2}\sin(\lambda^{1/2}\Delta t)\right]\Big|_{\lambda=\xi_{\omega,12}} \hat{\mathbf{e}}_\omega^H (L_N - l_{1,\omega}I)(L_N - l_{2,\omega}I)\mathbf{f}_{N,2}, \\
\hat{\mathbf{e}}_\omega^H \mathbf{e}_{N,2}, &= \frac{1}{2}\,\frac{\partial^2}{\partial\lambda^2}\left[-\lambda^{1/2}\sin(\lambda^{1/2}\Delta t)\right]\Big|_{\lambda=\xi_{\omega,21}} \hat{\mathbf{e}}_\omega^H (L_N - l_{1,\omega}I)(L_N - l_{2,\omega}I)\mathbf{f}_{N,1} \\
&\quad + \frac{1}{2}\,\frac{\partial^2}{\partial\lambda^2}\left[\cos(\lambda^{1/2}\Delta t)\right]\Big|_{\lambda=\xi_{\omega,22}} \hat{\mathbf{e}}_\omega^H (L_N - l_{1,\omega}I)(L_N - l_{2,\omega}I)\mathbf{f}_{N,2},
\end{aligned}
$$

where $\xi_{\omega,ij} \in [l_{1,\omega}, l_{2,\omega}]$, for $i,j = 1,2$. In view of the regularity of $\mathbf{f}(x)$ and the fact that L_N is a discretization of a second-order differential operator with bandlimited coefficients and a constant leading coefficient, we have

$$
\left|\frac{1}{N}\hat{\mathbf{e}}_0^H (L_N - l_{1,0}I)(L_N - l_{2,0}I)\mathbf{f}_{N,j}\right| = \left|\frac{1}{N}(L_N\tilde{\mathbf{q}}_N)^T \mathbf{f}_{N,j}\right| \le \|L_N\tilde{\mathbf{q}}_N\|_\infty \|\mathbf{f}_{N,j}\|_\infty.
$$

It follows that there exist constants C_{ij}, for $i,j = 1,2$, independent of N and Δt, such that

$$
\begin{aligned}
\bar{q}^{1/2}\left|\frac{1}{N}\hat{\mathbf{e}}_0^H \mathbf{e}_{N,1}\right| &\le \Delta t^4 C_{11} + \Delta t^5 C_{12}, \\
\left|\frac{1}{N}\hat{\mathbf{e}}_0^H \mathbf{e}_{N,2}\right| &\le \Delta t^3 C_{21} + \Delta t^4 C_{22},
\end{aligned}
$$

and for $\omega = -N/2+1, \ldots, -1, 1, \ldots, N/2$, by a Taylor expansion of the sines and cosines in $\hat{\mathbf{e}}_\omega^H \mathbf{e}_{N,1}$ and $\hat{\mathbf{e}}_\omega^H \mathbf{e}_{N,2}$, we have

$$
\begin{aligned}
(\bar{p}\omega^2 + \bar{q})^{1/2}\left|\frac{1}{N}\hat{\mathbf{e}}_\omega^H \mathbf{e}_{N,1}\right| &\le \Delta t^4 \frac{C_{11}}{|\omega|^{m-2}} + \Delta t^5 \frac{C_{12}}{|\omega|^{m-3}}, \\
\left|\frac{1}{N}\hat{\mathbf{e}}_\omega^H \mathbf{e}_{N,2}\right| &\le \Delta t^3 \frac{C_{21}}{|\omega|^{m-1}} + \Delta t^4 \frac{C_{22}}{|\omega|^{m-2}}.
\end{aligned}
$$

It is important to note that because the leading coefficient $p(x)$ of L is constant, we have $(L_N - l_{2,\omega}I)\hat{\mathbf{e}}_\omega = \tilde{\mathbf{q}}_N \circ \hat{\mathbf{e}}_\omega$, where $\circ$ denotes componentwise multiplication. Therefore, this expression is bounded independently of ω.

Finally, we obtain

$$
\begin{aligned}
&\|E_2(\Delta t, \Delta x)\|_C \\
&\quad \le \frac{\Delta t^3}{\sqrt{2\pi}}\Bigg[C_{11}^2\Delta t^2\left(1 + 2\sum_{\omega=1}^{\infty}\omega^{4-2m}\right) + 2C_{11}C_{12}\Delta t^3\left(1 + 2\sum_{\omega=1}^{\infty}\omega^{5-2m}\right) \\
&\qquad + C_{12}^2\Delta t^4\left(1 + 2\sum_{\omega=1}^{\infty}\omega^{6-2m}\right) + C_{21}^2\left(1 + 2\sum_{\omega=1}^{\infty}\omega^{2-2m}\right) \\
&\qquad + 2C_{21}C_{22}\Delta t\left(1 + 2\sum_{\omega=1}^{\infty}\omega^{3-2m}\right) + C_{22}^2\Delta t^2\left(1 + 2\sum_{\omega=1}^{\infty}\omega^{4-2m}\right)\Bigg]^{1/2} \\
&\quad \le C_2\Delta t^3.
\end{aligned}
$$

Since $m \ge 4$, it follows that all of the summations over ω converge to a sum that can be bounded independently of N. We conclude that the constant C_2 is independent of Δx and Δt. □

ETNA
Kent State University and
Johann Radon Institute (RICAM)

3.3. Convergence. Now we can prove that the second-order KSS method converges for the problem (2.1)–(2.4) in the case of $p(x)$ being constant and $q(x)$ bandlimited. We say that a method is convergent of order (m,n) if there exist constants C_t and C_x, independent of the time step Δt and the grid spacing $\Delta x = 2\pi/N$, such that

$$\|\mathbf{u}(\cdot,t) - \mathbf{u}_N(\cdot,t)\|_C \leq C_t\Delta t^m + C_x\Delta x^n, \qquad 0 \leq t \leq T,$$

where $\mathbf{u}(x,t)$ is the exact solution and $\mathbf{u}_N(x,t)$ is the approximate solution computed using an N-point grid.

THEOREM 3.6. *Under the assumptions of Theorem 3.5, the two-node block KSS method with prescribed nodes* (3.6) *is convergent of order* $(2, m-1)$.

Proof. We recall that the operator $S(\Delta t)$ from (3.4) is the exact solution operator for the problem (2.1)–(2.4). For any nonnegative integer n and fixed grid size N, we define

$$E_n = \|\mathcal{I}_N S(\Delta t)^n \mathbf{f} - \tilde{S}_N(\Delta t)^n \mathcal{I}_N \mathbf{f}\|_C.$$

Then, by Theorem 3.5 and Corollary 3.4, there exist constants C_1, C_2, and C_q such that

$$\begin{aligned}
E_{n+1} &= \|\mathcal{I}_N S(\Delta t)^{n+1}\mathbf{f} - \tilde{S}_N(\Delta t)^{n+1}\mathcal{I}_N\mathbf{f}\|_C \\
&= \|\mathcal{I}_N S(\Delta t)S(\Delta t)^n\mathbf{f} - \tilde{S}_N(\Delta t)\tilde{S}_N(\Delta t)^n\mathcal{I}_N\mathbf{f}\|_C \\
&= \|\mathcal{I}_N S(\Delta t)S(\Delta t)^n\mathbf{f} - \tilde{S}_N(\Delta t)S(\Delta t)^n\mathbf{f} \\
&\qquad + \tilde{S}_N(\Delta t)S(\Delta t)^n\mathbf{f} - \tilde{S}_N(\Delta t)\tilde{S}_N(\Delta t)^n\mathcal{I}_N\mathbf{f}\|_C \\
&\leq \|\mathcal{I}_N S(\Delta t)S(\Delta t)^n\mathbf{f} - \tilde{S}_N(\Delta t)S(\Delta t)^n\mathbf{f}\|_C \\
&\qquad + \|\tilde{S}_N(\Delta t)[\mathcal{I}_N S(\Delta t)^n\mathbf{f} - \tilde{S}_N(\Delta t)^n\mathcal{I}_N\mathbf{f}]\|_C \\
&\leq \|\mathcal{I}_N S(\Delta t)S(\Delta t)^n\mathbf{f} - \tilde{S}_N(\Delta t)S(\Delta t)^n\mathbf{f}\|_C \\
&\qquad + \|\mathcal{T}_N S_N(\Delta t)\mathcal{R}_N[\mathcal{I}_N S(\Delta t)^n\mathbf{f} - \tilde{S}_N(\Delta t)^n\mathcal{I}_N\mathbf{f}]\|_C \\
&\leq \|\mathcal{I}_N S(\Delta t)S(\Delta t)^n\mathbf{f} - \tilde{S}_N(\Delta t)S(\Delta t)^n\mathbf{f}\|_C \\
&\qquad + \|\mathcal{T}_N S_N(\Delta t)\hat{\mathcal{R}}_N[\mathcal{I}_N S(\Delta t)^n\mathbf{f} - \tilde{S}_N(\Delta t)^n\mathcal{I}_N\mathbf{f}]\|_C \\
&\leq \|\mathcal{I}_N S(\Delta t)\mathbf{u}(\cdot,t_n) - \tilde{S}_N(\Delta t)\mathcal{I}_N\mathbf{u}(\cdot,t_n)\|_C + \|S_N(\Delta t)\|_{C_N} E_n \\
&\leq C_1\Delta t^3 + C_2\Delta t\Delta x^{m-1} + (1 + C_q\|\tilde{q}\|_\infty\Delta t)E_n.
\end{aligned}$$

It follows that

$$E_n \leq \frac{e^{C_q\|\tilde{q}\|_\infty T} - 1}{1 + C_q\|\tilde{q}\|_\infty\Delta t - 1}(C_1\Delta t^3 + C_2\Delta t\Delta x^{m-1}) \leq \tilde{C}_1\Delta t^2 + \tilde{C}_2\Delta x^{m-1}$$

for some constants $\tilde{C}_1$ and $\tilde{C}_2$ that depend only on T. We conclude that

$$\begin{aligned}
\|\mathbf{u}(\cdot,t_n) - \mathbf{u}_N(\cdot,t_n)\|_C &\leq \|\mathcal{I}_N\mathbf{u}(\cdot,t_n) - \mathbf{u}_N(\cdot,t_n)\|_C + \|(I - \mathcal{I}_N)\mathbf{u}(\cdot,t_n)\|_C \\
&\leq \tilde{C}_1\Delta t^2 + \tilde{C}_2\Delta x^{m-1} + \tilde{C}_3\Delta x^m. \qquad \square
\end{aligned}$$

4. Numerical experiments. We now perform some numerical experiments to corroborate the theory presented in Section 3. For each test case, the relative error was estimated using the ℓ_2-norm, in comparison to a reference solution computed by the MATLAB ODE solver `ode15s` [32], with absolute and relative tolerances set to 10^{-12}.

4.1. Constant leading coefficient. We consider the initial value problem

$$u_{tt} + Lu = 0, \qquad 0 < x < 2\pi, \quad t > 0, \tag{4.1}$$

where L is of the form (2.4) with

$$p(x) = 1, \tag{4.2}$$

$$q(x) = 1 + \frac{1}{2}\sin x + \frac{1}{4}\cos 2x + \frac{1}{8}\sin 3x. \tag{4.3}$$

The initial conditions are

$$u(x,0) = \begin{cases} 1 - \frac{2}{\pi}|x-\pi| & \pi/2 \le x \le 3\pi/2, \\ 0 & 0 \le x < \pi/2. \quad 3\pi/2 < x < 2\pi, \end{cases} \tag{4.4}$$

$$u_t(x,0) = 0, \quad 0 < x < 2\pi, \tag{4.5}$$

and we impose periodic boundary conditions. We note that the initial data belongs to $H_p^{m+1}([0,2\pi]) \times H_p^m([0,2\pi])$ for $m = 1$, which is not sufficiently regular to satisfy the assumptions of Theorem 3.5.

The results are presented in Table 4.1. As predicted by Theorem 3.5 and Corollary 3.4, we observe second-order accuracy in time, in spite of the low regularity of the initial data, even when the CFL limit is exceeded by using the same time step as the spatial resolution increases.

TABLE 4.1
Relative errors in the solution of (4.1)–(4.5) *with periodic boundary conditions on the domain* $(0, 2\pi) \times (0, 10)$*, using the second-order KSS method described in Section* 3 *with* N *grid points and time step* Δt*.*

Δt	$N = 256$	$N = 512$	$N = 1024$	$N = 2048$
$\pi/128$	1.38e-04	1.33e-04	1.30e-04	1.29e-04
$\pi/256$	3.32e-05	3.24e-05	3.25e-05	3.20e-05
$\pi/512$	8.04e-06	8.49e-06	8.27e-06	8.08e-06

4.2. Variable leading coefficient. We now solve the problem (4.1), with

$$p(x) = 1 - \frac{1}{2}\sin x + \frac{1}{4}\cos 2x \tag{4.6}$$

and initial conditions

$$u(x,0) = e^{-(x-\pi)^2}, \quad u_t(x,0) = 0, \qquad 0 < x < 2\pi. \tag{4.7}$$

The results are illustrated in Figures 4.1 and 4.2. We see that when the CFL number is larger than one, the method is unstable as high-frequency components quickly become the dominant terms of the solution, and their amplitudes grow without bound. On the other hand, when the CFL number is less than one, the solution is well-behaved.

The results of the experiments illustrated in Figures 4.1 and 4.2 are summarized in Table 4.2. As Δt is further decreased, the error in the second-order KSS method decreases as $O(\Delta t^2)$. The problem is also solved with `ode15s`, with its `MaxStep` and `InitialStep` parameters set to the value of each time step Δt used with KSS, in order to examine the behavior of the error as the maximum time step approaches zero. It is worth noting that `ode15s` employs adaptive time stepping while this implementation of KSS does not; adaptive time stepping for KSS methods was investigated in [5]. We observe that regardless of the maximum time step, `ode15s` produces a solution that is slightly more accurate than that of KSS, but KSS is significantly more efficient as long as the (fixed) time step is chosen sufficiently small.

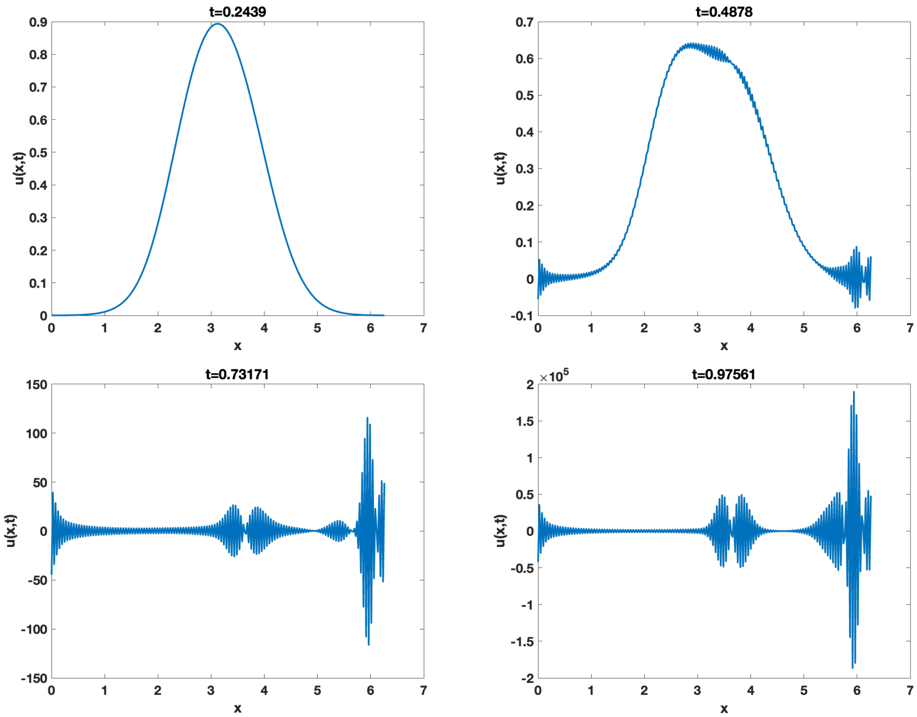

FIG. 4.1. *Solutions of* (4.1), (4.3), (4.6), (4.7) *on the domain* $(0, 2\pi) \times (0, 1)$ *using the second-order KSS method described in Section* 3 *with* $N = 256$ *grid points and time step CFL number* ≈ 1.74.

TABLE 4.2
Errors and timing for the examples from Figures 4.1 *and* 4.2.

		KSS		`ode15s`	
N	Δt	error	time	error	time
256	$\pi/128$	–	–	1.530e-05	1.284
	$\pi/256$	8.610e-05	0.006	1.282e-05	1.102
	$\pi/512$	1.941e-05	0.012	1.431e-05	1.099

4.3. Generalizations. In this paper, we have limited our analysis to the wave equation in one space dimension with periodic boundary conditions and spatial differentiation performed using the FFT. We now consider some variations of this problem, to investigate whether our conclusions may apply more broadly.

4.3.1. Finite differencing in space. We solve the problem (4.1), (4.2), (4.3), (4.7) on the domain $(0, 2\pi) \times (0, 10)$ with periodic boundary conditions and using centered differencing in space. Because of the change of spatial discretization, we modify the interpolation points from (3.6) by prescribing

$$l_{2,\omega} = \overline{p}\frac{2 - 2\cos(\omega\Delta x)}{\Delta x^2} + \overline{q}, \qquad \omega = -N/2 + 1, \ldots, N/2. \tag{4.8}$$

The results are shown in Table 4.3. It can be seen that the same unconditional stability that was established for spectral differentiation applies in the case of finite differencing, as an accurate solution is obtained even when the CFL number is as large as eight.

ETNA
Kent State University and
Johann Radon Institute (RICAM)

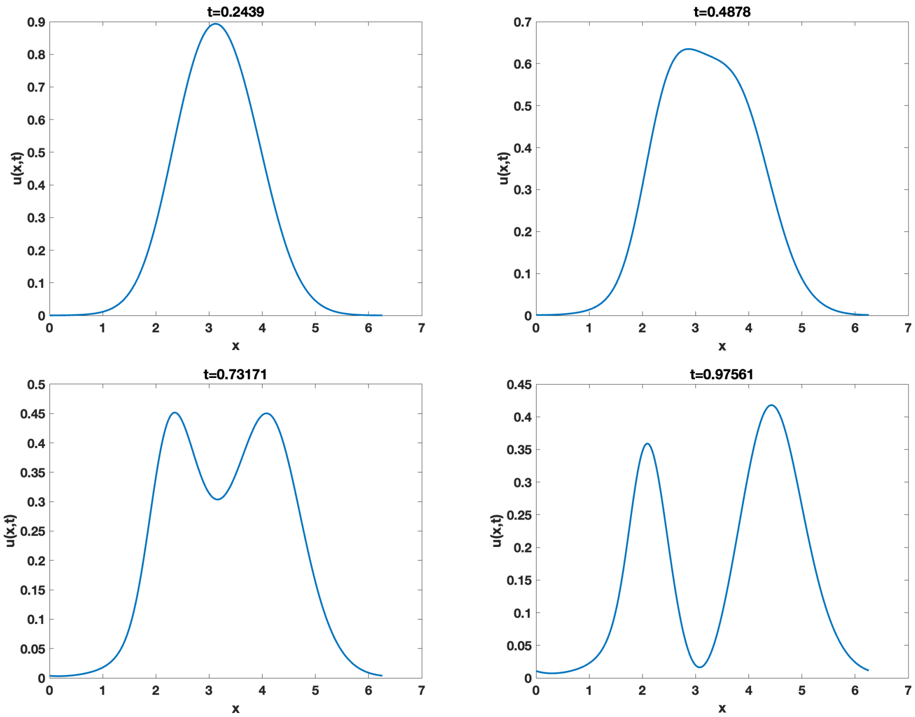

FIG. 4.2. *Solutions of* (4.1)*,* (4.3)*,* (4.6)*,* (4.7) *on the domain* $(0, 2\pi) \times (0, 1)$ *using the second-order KSS method described in Section 3 with* $N = 256$ *grid points and time step CFL number* ≈ 0.87*.*

TABLE 4.3
Relative errors in the solution of (4.1)*,* (4.2)*,* (4.3)*,* (4.7) *with periodic boundary conditions on the domain* $(0, 2\pi) \times (0, 10)$ *using the second-order KSS method described in Section 3 with* N *grid points, time step* Δt*, central differencing in space, and interpolation points* (4.8)*.*

Δt	$N = 256$	$N = 512$	$N = 1024$	$N = 2048$
$\pi/128$	1.00e-04	1.00e-04	1.00e-04	1.00e-04
$\pi/256$	2.47e-05	2.48e-05	2.48e-05	2.47e-05
$\pi/512$	6.15e-06	6.16e-06	6.15e-06	6.15e-06

4.3.2. Other boundary conditions. We repeat the problem from Section 4.3.1, except now with homogeneous Dirichlet boundary conditions. The interpolation points from (3.6) are modified as follows:

$$l_{2,\omega} = \overline{p}\frac{2 - 2\cos(\omega\Delta x/2)}{\Delta x^2} + \overline{q}, \qquad \omega = 0, 1, 2, \ldots, N-1. \tag{4.9}$$

As in the case of periodic boundary conditions, unconditional stability is indicated by the results, presented in Table 4.4.

4.3.3. Higher space dimension. We solve a two-dimensional wave equation

$$u_{tt} + Lu = 0, \qquad 0 < x, y < 2\pi, \quad 0 < t < 10, \tag{4.10}$$

where the differential operator L is defined by

$$Lu = -\Delta u + q(x, y)u, \tag{4.11}$$

ETNA
Kent State University and
Johann Radon Institute (RICAM)

TABLE 4.4
Relative errors in the solution of (4.1), (4.2), (4.3), (4.7) *with homogeneous Dirichlet boundary conditions on the domain* $(0, 2\pi) \times (0, 10)$ *using the second-order KSS method described in Section* 3 *with* N *grid points, time step* Δt*, central differencing in space, and interpolation points* (4.9).

Δt	$N = 256$	$N = 512$	$N = 1024$	$N = 2048$
$\pi/128$	1.53e-04	1.53e-04	1.53e-04	1.53e-04
$\pi/256$	3.85e-05	3.85e-05	3.85e-05	3.86e-05
$\pi/512$	9.67e-06	9.67e-06	9.67e-06	9.68e-06

with

$$q(x, y) = 1 + \frac{1}{2} \sin x \cos y + \frac{1}{4} \cos 2y + \frac{1}{8} \sin 3x. \tag{4.12}$$

Our initial conditions are

$$u(x, y, 0) = e^{-(x-\pi)^2+(y-\pi)^2}, \quad u_t(x, y, 0) = 0, \qquad 0 < x, y < 2\pi, \tag{4.13}$$

and we impose periodic boundary conditions. For the spatial discretization, we use an $N \times N$ grid with spacing $\Delta x = \Delta y = 2\pi/N$ and centered differencing in space. This leads to the choice of interpolation points

$$l_{1,\omega_1,\omega_2} = 0, \qquad l_{2,\omega_1,\omega_2} = \frac{1}{\Delta x^2}[4 - 2\cos(\omega_1 \Delta x) - 2\cos(\omega_2 \Delta y)] + \overline{q}, \tag{4.14}$$

for $\omega_1, \omega_2 = -N/2 + 1, \ldots, N/2$, where $\overline{q}$ is the average value of $q(x, y)$ on $(0, 2\pi)^2$. The results, shown in Table 4.5, indicate that unconditional stability holds again as the CFL number exceeds one without loss of accuracy or stability.

TABLE 4.5
Relative errors in the solution of (4.10), (4.12), (4.13) *with periodic boundary conditions on the domain* $(0, 2\pi)^2 \times (0, 10)$ *using the second-order KSS method described in Section* 3 *with* N *grid points per dimension, time step* Δt*, central differencing in space, and interpolation points* (4.14).

Δt	$N = 16$	$N = 32$	$N = 64$	$N = 128$
$\pi/8$	3.83e-02	4.14e-02	3.95e-02	3.84e-02
$\pi/16$	8.69e-03	8.80e-03	8.30e-03	8.06e-03
$\pi/32$	2.01e-03	1.95e-03	1.84e-03	1.78e-03

4.4. Non-bandlimited coefficients. We now carry out further investigation of the performance of KSS on problems beyond those considered in the convergence analysis from Section 3.

We consider the problem (4.1), (4.2), (4.4), (4.5) with periodic boundary conditions, first with

$$q(x) = \begin{cases} 1 + \frac{1}{2}\sin x & 0 \le x < \pi, \\ 1 - \frac{1}{2}\sin 2x & \pi \le x < 2\pi, \end{cases} \tag{4.15}$$

which is constructed to be continuous but not differentiable at $x = \pi$, and then with

$$q(x) = \begin{cases} 3/2 & 0 \le x < \pi, \\ 1/2 & \pi \le x < 2\pi. \end{cases} \tag{4.16}$$

ETNA
Kent State University and
Johann Radon Institute (RICAM)

TABLE 4.6
Relative errors in the solution of (4.1), (4.2), (4.4), (4.5), (4.15), with periodic boundary conditions on the domain $(0, 2\pi) \times (0, 1)$ using the second-order KSS method described in Section 3 with N grid points, time step Δt, and central differencing in space.

Δt	$N = 256$	$N = 512$	$N = 1024$	$N = 2048$
$\pi/128$	5.331e-05	5.452e-05	5.342e-05	5.220e-05
$\pi/256$	1.297e-05	1.336e-05	1.393e-05	1.381e-05
$\pi/512$	3.219e-06	3.396e-06	3.597e-06	3.531e-06

TABLE 4.7
Relative errors in the solution of (4.1), (4.2), (4.4), (4.5), (4.16) with periodic boundary conditions on the domain $(0, 2\pi) \times (0, 1)$ using the second-order KSS method described in Section 3 with N grid points, time step Δt, and central differencing in space.

Δt	$N = 256$	$N = 512$	$N = 1024$	$N = 2048$
$\pi/128$	5.891e-05	6.005e-05	5.917e-05	5.783e-05
$\pi/256$	1.417e-05	1.464e-05	1.533e-05	1.516e-05
$\pi/512$	3.574e-06	3.758e-06	3.924e-06	3.870e-06

As shown in Tables 4.6 and 4.7, the KSS method maintains stability and second-order accuracy in time. Based on these numerical results, we seek to strengthen the result of Corollary 3.4 by weakening the assumption about the regularity of $q(x)$.

THEOREM 4.1. *Assume $p(x) \equiv$ constant, $q(x)$ is 2π-periodic, and $q''(x)$ is piecewise C^1. Then the solution operator $S_N(\Delta t)$ satisfies*

$$\|S_N(\Delta t)\|_{C_N} \leq 1 + C_q \|\tilde{q}\|_\infty \Delta t, \tag{4.17}$$

where the constant C_q is independent of N and Δt.

Proof. The proof begins as in that of Theorem 3.3 and its supporting lemmas. We then have

$$\begin{aligned}
&\|G_{11}\|_\infty \\
&\quad \leq \max_{j \in \hat{I}_N} \left\{ 1 + \frac{3}{2} \sum_{k \in \hat{I}_N \setminus j} \left| \hat{q}(k-j)(\bar{p}j^2 + \bar{q})^{-1/2}(\bar{p}k^2 + \bar{q})^{1/2} \Delta t^2 \right| \right. \\
&\qquad + \frac{3}{2} \sum_{k \in \hat{I}_N \setminus j} \left| \hat{q}(k-j)(\bar{p}j^2 + \bar{q})^{1/2}(\bar{p}k^2 + \bar{q})^{-1/2} \Delta t^2 \right| \\
&\qquad \left. + \sum_{k \in \hat{I}_N} \left| (\bar{p}j^2 + \bar{q})^{-1/2}(\bar{p}k^2 + \bar{q})^{-1/2} \sum_{\omega \in \hat{I}_N \setminus \{k,j\}} |\hat{q}(\omega - k)\hat{q}(j - \omega)| \Delta t^2 \right| \right\}.
\end{aligned} \tag{4.18}$$

By the assumptions on $q(x)$, it follows from [16, Theorem A.1.3] that there exists a constant C_q such that

$$|\hat{q}(\omega)| \leq \frac{C_q}{|\omega|^3 + 1}.$$

Therefore, if we define

$$C_0 = \sup_{\omega \in \mathbb{Z} \setminus \{0\}} \frac{|\hat{q}(\omega)(|\omega|^3 + 1)|}{\|\tilde{q}\|_\infty},$$

it follows that for $\omega \neq 0$,

$$|\hat{q}(\omega)| \leq \frac{C_0\|\tilde{q}\|_\infty}{|\omega|^3+1}. \tag{4.19}$$

To bound the first summation in (4.18), we consider

$$\begin{aligned}\sum_{k\in\hat{I}_N\backslash j}\left|\hat{q}(k-j)\sqrt{\frac{\bar{p}k^2+\bar{q}}{\bar{p}j^2+\bar{q}}}\right| &\leq \frac{C_0\|\tilde{q}\|_\infty}{\sqrt{\bar{p}j^2+\bar{q}}}\sum_{k\in\hat{I}_N\backslash j}\frac{\sqrt{\bar{p}k^2+\bar{q}}}{|j-k|^3+1}\\ &\leq \frac{C_0\|\tilde{q}\|_\infty}{\sqrt{\bar{p}j^2+\bar{q}}}\sum_{k\in\hat{I}_N\backslash j}\frac{\sqrt{\bar{p}}|k|}{|j-k|^3+1}+\frac{\sqrt{\bar{q}}}{|j-k|^3+1}.\end{aligned} \tag{4.20}$$

From

$$\begin{aligned}\sum_{k\in\hat{I}_N\backslash j}\frac{|k|}{|j-k|^3+1} &\leq \max\{2,|j|\}+\sum_{k\in\hat{I}_N,|k-j|>1}\frac{|k|}{|j-k|^3+1}\\ &\leq \max\{2,|j|\}+2\sum_{u=2}^{N}\frac{u}{u^3+1}+\frac{|j|}{u^3+1}\\ &\leq \max\{2,|j|\}+2\int_1^\infty u^{-2}+|j|u^{-3}\,du\\ &\leq \max\{2,|j|\}+2\left(1+\frac{|j|}{2}\right),\end{aligned}$$

we can conclude that the expression from (4.20) is bounded independently of N.

Next, we show that the second summation from (4.18),

$$\sum_{k\in\hat{I}_N\backslash j}\left|\hat{q}(k-j)(\bar{p}j^2+\bar{q})^{1/2}(\bar{p}k^2+\bar{q})^{-1/2}\Delta t^2\right|, \tag{4.21}$$

can also be bounded independently of N. Applying (4.19), we focus on

$$\begin{aligned}\sum_{k\in\hat{I}_N\backslash j}\left|\hat{q}(k-j)(\bar{p}k^2+\bar{q})^{-1/2}\right| &\leq \sum_{k\in\hat{I}_N\backslash j}\frac{(\bar{p}k^2+\bar{q})^{-1/2}}{|j-k|^3+1}\\ &\leq \frac{1}{(|j|^3+1)\bar{q}}+\frac{1}{\sqrt{\bar{p}}}\sum_{k\in\hat{I}_N\backslash\{0,j\}}\frac{1}{|j-k|^3|k|}.\end{aligned} \tag{4.22}$$

If $j=0$, then we have

$$\sum_{k\in\hat{I}_N\backslash 0}\frac{1}{|k|^3|k|} \leq 2\left(\int_1^\infty \frac{1}{k^4}\,dk+1\right) \leq \frac{8}{3}.$$

If $j>0$, then we bound the final summation in (4.22) as follows:

$$\begin{aligned}\sum_{k=-\frac{N}{2}+1}^{-1}\frac{1}{(j-k)^3(-k)} &\leq \frac{1}{(j+1)^3}+\int_{-\infty}^{-1}\frac{1}{(j-k)^3(-k)}\,dk\\ &\leq \frac{1}{(j+1)^3}-\frac{3j+2}{2j^2(j+1)^2}+\frac{\ln|j+1|}{j^3},\end{aligned}$$

ETNA
Kent State University and
Johann Radon Institute (RICAM)

$$\sum_{k=1}^{j-1} \frac{1}{(j-k)^3 k} \leq \frac{1}{j-1} + \frac{1}{(j-1)^3} + \int_1^{j-1} \frac{1}{(j-k)^3 k}\, dk$$
$$\leq \frac{1}{j-1} + \frac{1}{(j-1)^3} + \frac{j^3-6j+4}{2j^2(j-1)^2} + \frac{2\ln|j-1|}{j^3},$$
$$\sum_{k=j+1}^{N/2} \frac{1}{(k-j)^3 k} \leq \frac{1}{(j+1)} + \int_{j+1}^{\infty} \frac{1}{(k-j)^3 k}\, dk$$
$$\leq \frac{1}{(j+1)} + \frac{\ln|j+1|}{j^3} - \frac{1}{j^2} + \frac{1}{2j}.$$

As all of these portions of (4.22) are $O(j^{-1})$, we find that (4.21) is bounded independently of N.

Finally, we consider the third summation from (4.18),

$$\sum_{k\in \hat{I}_N} \left| (\bar{p}j^2+\bar{q})^{-1/2}(\bar{p}k^2+\bar{q})^{-1/2} \sum_{\omega\in \hat{I}_N\setminus\{k,j\}} |\hat{q}(\omega-k)\hat{q}(j-\omega)|\Delta t^2 \right|. \tag{4.23}$$

Applying (4.19) to the sum over ω, we then focus on bounding

$$\sum_{\omega\in \hat{I}_N\setminus\{k,j\}} \frac{1}{|\omega-k|^3|j-\omega|^3}. \tag{4.24}$$

Let $z \equiv k-j > 1$, and let $m = \lfloor (j+k)/2 \rfloor$. We then have

$$\sum_{\omega\in \hat{I}_N\setminus\{k,j\}} \frac{1}{|\omega-k|^3|j-\omega|^3}$$
$$\leq \frac{2}{|j-1-k|^3} + \frac{2}{|j+1-k|^3} + \int_{-\infty}^{j-1} \frac{1}{[(k-\omega)(j-\omega)]^3}\, d\omega$$
$$+ \int_{j+1}^{m} \frac{1}{[(\omega-k)(j-\omega)]^3}\, d\omega + \int_{m+1}^{k-1} \frac{1}{[(\omega-k)(j-\omega)]^3}\, d\omega$$
$$+ \int_{k+1}^{\infty} \frac{1}{[(\omega-k)(\omega-j)]^3}\, d\omega$$
$$\leq \frac{2}{|z+1|^3} + \frac{2}{|z-1|^3} + \frac{2z^2 - \frac{2z^2}{(z-1)^2} - \frac{12z}{z-1} + 24\ln|z-1|}{2z^5}$$
$$+ \frac{\frac{z^2}{(\frac{-z}{2}-1)^2} - \frac{z^2}{(\frac{z}{2}-1)^2} - \frac{6z}{(\frac{-z}{2}-1)} - \frac{6z}{(\frac{z}{2}-1)} - 12\ln|\frac{z}{2}+1| + 12\ln|\frac{-z}{2}+1|}{2z^5}$$
$$\leq \tilde{C} z^{-3}$$

for some constant $\tilde{C}$. By symmetry, the case of $z < 1$ is identical, and by direct evaluation, the terms corresponding to $|z| \leq 1$ are bounded independently of j. Summing the bounds for (4.24) over all $z \in \mathbb{Z}$, we conclude that (4.23) is bounded independently of N and is of the order $O(\Delta t^2)$.

In summary, there exist constants $C_{11,q}$ and C_{11,q^2}, independent of N and Δt, such that

$$\|G_{11}\|_\infty \leq 1 + C_{11,q}\|\tilde{q}\|_\infty \Delta t^2 + C_{11,q^2}\|\tilde{q}\|_\infty^2 \Delta t^2.$$

ETNA
Kent State University and
Johann Radon Institute (RICAM)

Using the same approach, we find that the matrix G_{22} defined in (3.10), with the assumption that $p(x)$ is constant, satisfies a bound of the same form as that of G_{11} with appropriate constant factors.

Proceeding as in the proof of Lemma 3.2, we have

$$\|G_{12}\|_\infty \leq \max_{1\leq j\leq N} \left\{ 6 \sum_{k\in \hat{I}_N\setminus j} \left| \hat{q}(k-j)(\bar{p}j^2+\bar{q})^{-1/2}\Delta t \right| \right.$$
$$\left. + \sum_{k\in\hat{I}_N\setminus j} \left| \sum_{\omega\in\hat{I}_N\setminus\{k,j\}} |\hat{q}(j-\omega)\hat{q}(\omega-k)|(\bar{p}j^2+\bar{q})^{-1/2} \frac{4\Delta t}{\bar{p}\omega^2+\bar{q}} \right| \right\}.$$

These summations can be bounded using the same approach as in the case of G_{11}, and therefore there exist constants $C_{12,q}$ and C_{12,q^2} independent of N and Δt such that

$$\|G_{12}\|_\infty \leq C_{12,q}\|\tilde{q}\|_\infty \Delta t + C_{12,q^2}\|\tilde{q}\|_\infty^2 \Delta t.$$

The same approach can also be used to show that the matrix G_{21} in (3.10), under the assumption that $p(x)$ is constant, satisfies a bound of the same form as that of G_{12}. Proceeding as in the proof of Theorem 3.3 yields (4.17). ◻

We now present numerical evidence that the conclusion of Theorem 4.1 holds under an even weaker assumption about the regularity of $q(x)$. Table 4.8 shows that for the differential operator (2.4), with $p(x)$ constant and $q(x)$ piecewise constant, $\|S_N(\Delta t)\|_{C_N}$ appears to be bounded independently of N. Unfortunately, such a bound cannot be proved using the same approach as in the proof of Theorem 4.1 as the upper bound established is not sufficiently sharp.

TABLE 4.8
$\|S_N(\Delta t)\|_{C_N}$ for various values of N and Δt, where $S_N(\Delta t)$, as defined in (3.5), is the approximate solution operator for the KSS method described in Section 3 for the problem (4.1), (4.2), (4.4), (4.5), (4.16).

Δt	$N = 256$	$N = 512$	$N = 1024$
1	1.272444	1.272439	1.272438
0.1	1.025053	1.025047	1.025045
0.01	1.002502	1.002502	1.002501
0.001	1.000250	1.000250	1.000250

4.5. Comparison with Krylov solvers. We will now compare the performance of our KSS method with an implicit time-stepping method in which a Krylov subspace method is used to solve systems of linear equations. We consider the problem (4.1), (4.2), (4.4), (4.5), (4.15) with periodic boundary conditions.

After spatial discretization, we solve the system of ODEs (3.3) using the trapezoidal rule for time stepping, which requires solving the systems of linear equations

$$\left(I_{2N} - \frac{\Delta t}{2}\tilde{L}_N\right)\mathbf{v}_N^{n+1} = \left(I_{2N} + \frac{\Delta t}{2}\tilde{L}_N\right)\mathbf{v}_N^n, \qquad n = 0,1,2,\ldots \tag{4.25}$$

To solve each system, we use GMRES with ILU(0) preconditioning [31]. The results are presented in Table 4.9. We observe that the trapezoidal rule is not nearly as accurate as KSS, even when the system (4.25) is solved to very high accuracy. Furthermore, the accuracy

ETNA
Kent State University and
Johann Radon Institute (RICAM)

deteriorates as the grid size increases, and second-order accuracy is not maintained. This is due to the fact that the initial data, and therefore the solution, is not smooth; with smooth initial data, the trapezoidal rule is more accurate and does not lose accuracy as N increases, though it is still not as accurate as KSS.

Finally, the number of iterates needed for convergence by GMRES, though reduced to some extent by the preconditioning, still increases with both N and Δt (whether the initial data is smooth or not), while the number of FFTs or matrix-vector multiplications required by KSS are not influenced by these parameters. Similar results were obtained when using BiCGSTAB in place of GMRES, except that, on average, even more iterations were required for convergence.

TABLE 4.9
Relative errors, execution times in seconds, and average iteration counts in the solution of (4.1), (4.2), (4.4), (4.5), (4.15) *with periodic boundary conditions on the domain* $(0, 2\pi) \times (0, 1)$ *using the second-order KSS method described in Section 3 (labeled "KSS") and the trapezoidal rule with GMRES and ILU preconditioning (labeled "GMRES"). Both methods use* N *grid points, time step* Δt*, and central differencing in space.*

		KSS		GMRES		
N	Δt	error	time	error	time	iter.
256	$\pi/64$	2.313e-04	0.001	4.685e-02	0.009	8
	$\pi/128$	5.891e-05	0.001	2.091e-02	0.013	5
	$\pi/256$	1.417e-05	0.002	1.080e-02	0.024	4
	$\pi/512$	3.574e-06	0.004	2.979e-03	0.044	3
	$\pi/1024$	8.958e-07	0.008	7.153e-04	0.087	3
	$\pi/2048$	2.242e-07	0.015	1.844e-04	0.176	3
512	$\pi/64$	2.203e-04	0.001	3.714e-02	0.022	13
	$\pi/128$	6.005e-05	0.002	2.574e-02	0.032	9
	$\pi/256$	1.464e-05	0.004	1.210e-02	0.050	6
	$\pi/512$	3.758e-06	0.007	4.789e-03	0.084	4
	$\pi/1024$	9.518e-07	0.015	1.921e-03	0.145	3
	$\pi/2048$	2.393e-07	0.029	5.675e-04	0.292	3
1024	$\pi/64$	2.112e-04	0.001	4.279e-02	0.073	22
	$\pi/128$	5.917e-05	0.002	2.155e-02	0.112	16
	$\pi/256$	1.533e-05	0.004	1.508e-02	0.177	10
	$\pi/512$	3.924e-06	0.007	7.994e-03	0.318	7
	$\pi/1024$	9.764e-07	0.015	3.316e-03	0.585	5
	$\pi/2048$	2.428e-07	0.031	1.304e-03	1.134	4

4.6. Comparison with exponential integrators. Next, we apply our KSS method to a nonlinear problem and compare its performance to that of exponential integrators that employ Krylov subspace methods to evaluate matrix function-vector products. We consider the Klein-Gordon equation [13]

$$u_{tt} = u_{xx} - u^3, \qquad 0 < x < 2\pi, \quad t > 0, \tag{4.26}$$

with initial conditions (4.4), (4.5) and periodic boundary conditions. The second-order KSS method described in Section 3 is compared with the following methods:

- A Gautschi-type method presented in [12, 18], in which matrix function-vector products are computed by applying Lanczos iteration, as described in [19]. This method will be referred to as "Gautschi-Krylov".

- The exponential Euler method [27], with matrix function-vector products computed using an adaptive Krylov iteration from [26]. This method will be referred to as "adaptive Krylov".

The results are shown in Table 4.10. For the Gautschi-Krylov and adaptive Krylov methods, the iteration counts reported in the table refer to the average number of matrix-vector multiplications performed in the course of approximating matrix function-vector products. For all three methods, the following computational expense is incurred during each time step:

- For KSS, three matrix-vector multiplications, three FFTs, and two inverse FFTs, in the course of approximating four matrix function-vector products with an $N \times N$ matrix.
- For Gautschi-Krylov, two matrix function-vector products, each involving, on average, the number of matrix-vector multiplications reported in Table 4.10 with an $N \times N$ matrix.
- For adaptive Krylov, one matrix function-vector product, involving, on average, the number of matrix-vector multiplications reported in Table 4.10 with a $2N \times 2N$ matrix.

As can be seen in the table, the number of matrix function-vector products required by the Gautschi-Krylov and adaptive Krylov methods increases with N and Δt. The accuracy of the KSS and adaptive Krylov method is comparable, while Gautschi-Krylov is somewhat more accurate than both, but KSS is significantly faster than both.

TABLE 4.10

Relative errors, execution times in seconds, and iteration counts in the solution of (4.4), (4.5), (4.26) with periodic boundary conditions on the domain $(0, 2\pi) \times (0, 1)$ using the second-order KSS method described in Section 3 (labeled "KSS"), the Gautschi-type method from [12, 18] (labeled "Gautschi-Krylov"), and the exponential Euler method [27] with adaptive Krylov iteration [26] (labeled "Adaptive Krylov"). All methods use N grid points, time step Δt, and central differencing in space.

		KSS		Gautschi-Krylov			Adaptive Krylov		
N	Δt	error	time	error	time	iter.	error	time	iter.
256	$\pi/64$	5.441e-04	0.001	1.796e-04	0.013	8	4.967e-04	0.037	16
	$\pi/128$	1.353e-04	0.001	4.309e-05	0.015	6	1.315e-04	0.049	11
	$\pi/256$	3.328e-05	0.002	1.036e-05	0.024	5	3.304e-05	0.084	8
	$\pi/512$	8.430e-06	0.004	2.533e-06	0.033	4	8.384e-06	0.161	7
	$\pi/1024$	2.105e-06	0.008	6.346e-07	0.055	4	2.099e-06	0.277	6
	$\pi/2048$	5.258e-07	0.016	1.588e-07	0.095	3	5.248e-07	0.542	5
512	$\pi/64$	5.221e-04	0.001	1.616e-04	0.025	12	4.919e-04	0.059	21
	$\pi/128$	1.376e-04	0.002	4.434e-05	0.028	9	1.307e-04	0.074	15
	$\pi/256$	3.351e-05	0.003	1.068e-05	0.036	7	3.285e-05	0.109	11
	$\pi/512$	8.375e-06	0.006	2.600e-06	0.059	6	8.331e-06	0.191	8
	$\pi/1024$	2.090e-06	0.012	6.228e-07	0.090	5	2.085e-06	0.360	7
	$\pi/2048$	5.223e-07	0.025	1.575e-07	0.148	4	5.216e-07	0.645	6
1024	$\pi/64$	5.158e-04	0.002	1.614e-04	0.110	18	4.893e-04	0.099	35
	$\pi/128$	1.348e-04	0.004	4.103e-05	0.073	12	1.299e-04	0.137	22
	$\pi/256$	3.358e-05	0.008	1.066e-05	0.075	9	3.267e-05	0.192	15
	$\pi/512$	8.423e-06	0.015	2.663e-06	0.089	7	8.293e-06	0.299	11
	$\pi/1024$	2.087e-06	0.031	6.429e-07	0.138	6	2.076e-06	0.502	9
	$\pi/2048$	5.201e-07	0.063	1.512e-07	0.228	5	5.191e-07	0.866	7

ETNA
Kent State University and
Johann Radon Institute (RICAM)

Next, we consider another Klein-Gordon equation,

$$u_{tt} = (p(x)u_x)_x - q(x)u - u^3, \qquad 0 < x < 2\pi, \quad t > 0, \tag{4.27}$$

with $p(x)$ from (4.6), $q(x)$ from (4.3), initial conditions (4.4), (4.5), and periodic boundary conditions. The results are shown in Table 4.11. We see that the KSS method cannot produce an accurate solution when $\Delta t > \Delta x$; the method is unstable in this case due to $p(x)$ varying with x. For Δt sufficiently small, KSS exhibits second-order accuracy, and an accuracy comparable to that of the adaptive Krylov method. The Gautschi-Krylov method is the most accurate method of the three, but KSS, when stable, is able to deliver higher accuracy in less time.

TABLE 4.11

Relative errors, execution times in seconds, and iteration counts in the solution of (4.4), (4.5), (4.27) *with periodic boundary conditions on the domain* $(0, 2\pi) \times (0, 1)$ *using the second-order KSS method described in Section 3 (labeled "KSS"), the Gautschi-type method from [12, 18] (labeled "Gautschi-Krylov"), and the exponential Euler method [27] with adaptive Krylov iteration [26] (labeled "Adaptive Krylov"). All methods use* N *grid points, time step* Δt*, and central differencing in space.*

		KSS		Gautschi-Krylov			Adaptive Krylov		
N	Δt	error	time	error	time	iter.	error	time	iter.
256	$\pi/64$	–	–	4.011e-04	0.008	7	1.076e-03	0.025	11
	$\pi/128$	2.887e-04	0.001	1.079e-04	0.010	5	2.817e-04	0.042	8
	$\pi/256$	7.117e-05	0.003	2.738e-05	0.014	4	7.031e-05	0.071	6
	$\pi/512$	1.788e-05	0.005	7.010e-06	0.025	4	1.778e-05	0.137	5
	$\pi/1024$	4.455e-06	0.011	1.755e-06	0.043	3	4.442e-06	0.224	4
	$\pi/2048$	1.112e-06	0.021	4.391e-07	0.087	3	1.110e-06	0.352	3
512	$\pi/64$	–	–	4.010e-04	0.019	10	1.076e-03	0.051	20
	$\pi/128$	–	–	1.079e-04	0.019	7	2.816e-04	0.055	11
	$\pi/256$	7.113e-05	0.003	2.740e-05	0.029	6	7.027e-05	0.093	8
	$\pi/512$	1.788e-05	0.007	6.978e-06	0.041	4	1.777e-05	0.158	6
	$\pi/1024$	4.453e-06	0.014	1.755e-06	0.068	4	4.440e-06	0.308	5
	$\pi/2048$	1.111e-06	0.028	4.390e-07	0.113	3	1.110e-06	0.513	4
1024	$\pi/64$	–	–	4.010e-04	0.089	15	1.075e-03	0.100	55
	$\pi/128$	–	–	1.079e-04	0.055	11	2.815e-04	0.120	19
	$\pi/256$	–	–	2.737e-05	0.051	8	7.026e-05	0.167	12
	$\pi/512$	1.787e-05	0.012	6.981e-06	0.068	6	1.776e-05	0.215	8
	$\pi/1024$	4.452e-06	0.025	1.751e-06	0.104	5	4.439e-06	0.361	6
	$\pi/2048$	1.111e-06	0.052	4.389e-07	0.160	4	1.109e-06	0.699	5

Finally, we compare the Gautschi-Krylov method to a variation of the Gautschi-Krylov method in which any matrix function-vector products are computed using KSS; this variation is referred to as "Gautschi-KSS". As can be seen in Table 4.12, this variation combines the higher stability of Gautschi-Krylov with the efficiency and scalability of KSS. The Gautschi-KSS method is significantly faster than KSS alone (and therefore has an even greater advantage in terms of efficiency over Gautschi-Krylov) and is not unstable for $\Delta t > \Delta x$. For larger time steps, Gautschi-KSS does not always exhibit second-order accuracy; this is due to the lack of smoothness in the solution.

4.7. Discussion. This focus of this paper has been 1D wave propagation problems featuring initial data and time-independent coefficients with varying degrees of smoothness for

TABLE 4.12
Relative errors, execution times in seconds, and iteration counts in the solution of (4.4), (4.5), (4.27) with periodic boundary conditions on the domain $(0, 2\pi) \times (0, 1)$ using the Gautschi-type method from [12, 18] with matrix function-vector products computed as in [19] (labeled "Gautschi-Krylov") and the Gautschi-type method with matrix function-vector products computed via KSS (labeled "Gautschi-KSS"). All methods use N grid points, time step Δt, and central differencing in space.

		Gautschi-Krylov			Gautschi-KSS	
N	Δt	error	time	iter.	error	time
256	$\pi/64$	1.796e-04	0.013	8	1.933e-03	0.000
	$\pi/128$	4.309e-05	0.015	6	9.822e-05	0.001
	$\pi/256$	1.036e-05	0.024	5	2.158e-05	0.001
	$\pi/512$	2.533e-06	0.033	4	5.267e-06	0.003
	$\pi/1024$	6.346e-07	0.055	4	1.304e-06	0.005
	$\pi/2048$	1.588e-07	0.095	3	3.254e-07	0.011
512	$\pi/64$	1.616e-04	0.025	12	1.306e-03	0.001
	$\pi/128$	4.434e-05	0.028	9	4.793e-04	0.001
	$\pi/256$	1.068e-05	0.036	7	2.320e-05	0.002
	$\pi/512$	2.600e-06	0.059	6	5.339e-06	0.004
	$\pi/1024$	6.228e-07	0.090	5	1.300e-06	0.009
	$\pi/2048$	1.575e-07	0.148	4	3.228e-07	0.017
1024	$\pi/64$	1.614e-04	0.110	18	1.237e-03	0.001
	$\pi/128$	4.103e-05	0.073	12	3.384e-04	0.002
	$\pi/256$	1.066e-05	0.075	9	1.169e-04	0.004
	$\pi/512$	2.663e-06	0.089	7	5.598e-06	0.007
	$\pi/1024$	6.429e-07	0.138	6	1.314e-06	0.015
	$\pi/2048$	1.512e-07	0.228	5	3.238e-07	0.029

the purpose of a convergence analysis. By contrast, in [30], KSS methods have been applied to problems for modeling acoustic singular surfaces, such as shock waves and acceleration waves. In that work it is demonstrated that for such problems conventional approaches to time stepping—a Fourier spectral method in the case of a linear PDE with constant coefficients or an exponential integrator in the nonlinear case—can fail to produce a solution that is even qualitatively correct, whereas a KSS method is successful.

KSS methods are also not limited to IBVPs with time-independent data as the problems featured in [30] include time-dependent boundary conditions. More generally, features such as time-dependent coefficients can be handled by a combination of KSS methods with exponential integrators, which are not mutually exclusive as demonstrated in this section with the introduction of "Gautschi-KSS". A combination of KSS methods with exponential Rosenbrock methods for the application to larger-scale problems is forthcoming in [6], while a multistep formulation that can be applied "as is" to linear or nonlinear PDEs with or without time-dependent coefficients is soon to be introduced in [29].

5. Conclusion.

We have established an upper bound for the approximate solution operator of a second-order KSS method applied to the 1D wave equation with bandlimited coefficients and periodic boundary conditions. Unfortunately, the bound is not independent of the grid size, which indicates that the unconditional stability proved for the heat equation for the same kind of spatial differential operator in [33] does not extend to the wave equation. Numerical experiments support this assertion, while also suggesting that conditional stability may still hold. Unlike the CFL condition, which relates the spatial grid mesh and time step to the *magnitude* of the wave speed, a stability condition for a KSS method would likely relate the spatial grid mesh and time step to some measure of the *variation* in the wave speed.

We have also proved that the same KSS method applied to the wave equation with periodic boundary conditions is convergent with spectral accuracy in space and second-order accuracy in time, as well as unconditionally stable in the case of a constant wave speed and a bandlimited reaction term coefficient. This is the first result proving unconditional stability for a KSS method for the wave equation that approximates the solution operator of the PDE using prescribed interpolation points, as opposed to nodes of Gauss quadrature rules. Numerical experiments suggest that this unconditional stability also holds for related problems.

Furthermore, it has been demonstrated through numerical experiments and then proved that the assumption of a bandlimited reaction term coefficient is not necessary for unconditional stability. The proof of this result is the first stability analysis of a KSS method that does not require the coefficients of the spatial differential operator of the PDE to be either constant or bandlimited. Future work will extend this theory to other problems to which KSS methods have been applied. Finally, it has been shown that KSS methods can be effective for nonlinear wave equations with an advantage in efficiency and scalability over other time-stepping methods that use Krylov subspace iterations and that it is worthwhile to combine these approaches. Ongoing work involves combination of higher-order KSS methods and exponential integrators [20, 21] to improve on the combination presented in [4].

KSS methods, as presented in this paper and in [28], generalize the advantage of the Fourier spectral method for constant-coefficient linear PDEs—the ability to compute Fourier coefficients independently and with large time steps—to their variable-coefficient counterparts. Although the discrete Fourier transform has served as an essential ingredient in KSS methods in these works, it is important to note that KSS methods and the DFT are not inextricably linked. While the focus of this paper is mostly on problems in one space dimension with periodic boundary conditions, the main idea behind KSS methods—componentwise time stepping—can be employed effectively with any orthonormal basis (for example, a basis of modified sines as used in [25]) for which transformation between physical space and frequency space can be carried out efficiently. This allows for the development of KSS-like methods that use, for example, bases of Chebyshev polynomials or wavelets. For problems on non-rectangular domains, a combination with fictitious domain methods, such as the Fourier continuation approach of [3], is worthy of investigation. Another direction for future work is the addition of local time stepping [14], except in frequency space rather than physical space, to handle the case of variable wave speed by using smaller time steps for low-frequency components that are affected the most by such heterogeneity.

Acknowledgments. The authors wish to thank the anonymous referees for their helpful feedback, which led to substantial improvement of the manuscript.

REFERENCES

[1] K. E. ATKINSON, *An Introduction to Numerical Analysis*, 2nd ed., Wiley, New York, 1989.
[2] C. BARDOS AND E. TADMOR, *Stability and spectral convergence of Fourier method for nonlinear problems: on the shortcomings of the 2/3 de-aliasing method*, Numer. Math., 129 (2014), pp. 749–782.
[3] O. P. BRUNO AND P. JAGABANDHU, *Two-dimensional Fourier continuation and applications*, SIAM J. Sci. Comput., 44 (2022), pp. A964–A992.
[4] A. CIBOTARICA, J. V. LAMBERS, AND E. M. PALCHAK, *Solution of nonlinear time-dependent PDEs through componentwise approximation of matrix functions*, J. Comput. Phys., 321 (2016), pp. 1120–1143.
[5] H. DOZIER, *Enhancement of Krylov Subspace Spectral Methods Through the Use of the Residual*, PhD. Thesis, University of Southern Mississippi, Hattiesburg, 2019.
[6] C. DRUM, *Component-wise Rosenbrock methods for stiff systems of nonlinear ODEs*, in preparation.
[7] L. C. EVANS, *Partial Differential Equations*, American Mathematical Society, Providence, 1998.

[8] S. J. FARLOW, *Partial Differential Equations for Scientists and Engineers*, Dover, New York, 1993.

[9] G. H. GOLUB AND G. MEURANT, *Matrices, moments and quadrature*, in Numerical Analysis 1993, Proceedings of the 15th Dundee Conference, D. F. Griffiths and G. A. Watson, eds., Pitman Res. Notes Math. Ser., 303, Longman Scientific, Harlow, 1994, pp. 105–156.

[10] G. H. GOLUB AND R. UNDERWOOD, *The block Lanczos method for computing eigenvalues*, in Mathematical Software III., J. Rice, ed., Publication of the Mathematics Research Center, University of Wisconsin, 39, Academic Press, New York, 1977, pp. 361–377.

[11] J. GOODMAN, T. HOU, AND E. TADMOR, *On the stability of the unsmoothed Fourier method for hyperbolic equations*, Numer. Math., 67 (1994), pp. 93–129.

[12] V. GRIMM, *On the use of the Gautschi-type exponential integrator for wave equations*, in Numerical Mathematics and Advanced Applications, A. B. de Castro, D. Gómez, P. Quintela, and P. Salgado, eds., Springer, Berlin, 2006, pp. 557–563.

[13] W. GREINER, *Relativistic Quantum Mechanics. Wave Equations*, 3rd ed., Springer, Berlin, 2000.

[14] M. J. GROTE AND T. MITKOVA, *Explicit local time-stepping methods for time-dependent wave propagation*, in Direct and Inverse Problems in Wave Propagation and Applications, I. Graham, U. Langer, J. Melenk, and M. Sini, eds., Radon Ser. Comput. Appl. Math., 14, De Gruyter, Berlin, 2013, pp. 187–218.

[15] P. GUIDOTTI, J. V. LAMBERS, AND K. SOLNA, *Analysis of the 1D wave equation in inhomogeneous media*, Numer. Funct. Anal. Optim., 27 (2006), pp. 25–55.

[16] B. GUSTAFSSON, H.-O. KREISS, AND J. OLIGER, *Time Dependent Problems and Difference Methods*, Wiley, New York, 1995.

[17] J. S. HESTHAVEN, S. GOTTLIEB, AND D. GOTTLIEB, *Spectral Methods for Time-Dependent Problems*, Cambridge University Press, Cambridge, 2007.

[18] M. HOCHBRUCK AND C. LUBICH, *A Gautschi-type method for oscillatory second-order differential equations*, Numer. Math., 83 (1999), pp. 403–426.

[19] M. HOCHBRUCK, M., C. LUBICH AND H. SELHOFER, *Exponential integrators for large systems of differential equations*, SIAM J. Sci. Comput., 19 (1998), pp. 1552–1574.

[20] M. HOCHBRUCK AND A. OSTERMANN, *Explicit exponential Runge-Kutta methods for semilinear parabolic problems*, SIAM J. Numer. Anal., 43 (2005), pp. 1069–1090.

[21] ———, *Exponential integrators of Rosenbrock-type*, Tech. Report, Oberwolfach Reports 3, 2006, pp. 1107–1110.

[22] J. V. LAMBERS, *An explicit, stable, high-order spectral method for the wave equation based on the block Gaussian quadrature*, IAENG Int. J. Appl. Math., 38 (2008), pp. 233–248.

[23] ———, *Derivation of high-order spectral methods for time-dependent PDE using modified moments*, Electron. Trans. Numer. Anal., 28 (2007), pp. 114–135.
https://etna.ricam.oeaw.ac.at/vol.28.2007-2008/pp114-135.dir/pp114-135.pdf

[24] ———, *Enhancement of Krylov subspace spectral methods by block Lanczos iteration*, Electron. Trans. Numer. Anal., 31 (2008), pp. 86–109.
https://etna.ricam.oeaw.ac.at/vol.31.2008/pp86-109.dir/pp86-109.pdf

[25] J. V. LAMBERS AND P. M. JORDAN, *On the application of a Krylov subspace spectral method to poroacoustic shocks in inhomogeneous gases*, Numer. Methods Partial Differential Equations, 37 (2021), pp. 2955–2972.

[26] J. NIESEN AND W. M. WRIGHT, *Algorithm 919: a Krylov subspace algorithm for evaluating the φ-functions appearing in exponential integrators*, ACM Trans. Math. Software, 38 (2012), Art. 22, 19 pages.

[27] D. PHAN AND A. OSTERMANN, *Exponential integrators for second-order in time partial differential equations*, J. Sci. Comput., 93 (2022), Art. 58, 26 pages.

[28] E. M. PALCHAK, A. CIBOTARICA, AND J. V. LAMBERS, *Solution of time-dependent PDE through rapid estimation of block Gaussian quadrature nodes*, Linear Algebra Appl., 468 (2015), pp. 233–259.

[29] B. RESTER, *Spectral multistep methods for time-dependent PDEs I: explicit methods*, in preparation.

[30] B. RESTER, J. V. LAMBERS, AND P. M. JORDAN, *Acoustic singular surfaces in an exponential class of inhomogeneous gases: a new numerical approach based on Krylov subspace spectral methodologies*, Int. J. Non-Linear Mech., 156 (2023), Art. 104506, 19 pages.

[31] Y. SAAD, *Iterative Methods for Sparse Linear Systems*, PWS, Boston, 1996.

[32] L. F. SHAMPINE AND M. W. REICHELT, *The MATLAB ODE Suite*, SIAM J. Sci. Comput., 18 (1997), pp. 1–22.

[33] S. SHEIKHOLESLAMI, J. V. LAMBERS, AND C. WALKER, *Convergence analysis of Krylov subspace spectral methods for reaction-diffusion equations*, J. Sci. Comput., 78 (2019), pp. 1768–1789.