CrossMark

Convergence of Summation-by-Parts Finite Difference Methods for the Wave Equation

Siyang Wang[1] · Gunilla Kreiss[1]

Abstract When using a finite difference method to solve a time dependent partial differential equation, the truncation error is often larger at a few grid points near a boundary or grid interface than in the interior. In computations, the observed convergence rate is often higher than the order of the large truncation error. In this paper, we develop techniques for analyzing this phenomenon, and particularly consider the second order wave equation. The equation is discretized by a finite difference operator satisfying a summation by parts property, and the boundary and grid interface conditions are imposed weakly by the simultaneous approximation term method. It is well-known that if the semi-discretized wave equation satisfies the determinant condition, that is the boundary system in Laplace space is nonsingular for all $Re(s) \geq 0$, two orders are gained from the large truncation error localized at a few grid points. By performing a normal mode analysis, we show that many common discretizations do not satisfy the determinant condition at $s = 0$. We then carefully analyze the error equation to determine the gain in the convergence rate. The result shows that stability does not automatically imply a gain of two orders in the convergence rate. The precise gain can be lower than, equal to or higher than two orders, depending on the boundary condition and numerical boundary treatment. The accuracy analysis is verified by numerical experiments, and very good agreement is obtained.

✉ Siyang Wang
   siyang.wang@it.uu.se

[1] Department of Information Technology, Uppsala University, Box 337, 751 05 Uppsala, Sweden

1 Introduction

In many physical problems arising in for example acoustics, seismology, and electromagnetism, the governing equations can be formulated as systems of second order time dependent hyperbolic partial differential equations (PDE). For wave propagation problems with smooth solutions, it has been shown that high order accurate time marching methods as well as high order spatial discretizations solve these problems more efficiently than low order methods [7,9,10]. Examples of high order accurate methods to discretize the wave equation include the discontinuous Galerkin method [5] and the spectral method [23]. In this paper, we in particular consider high order finite difference methods.

High order finite difference methods have been widely used for solving wave propagation problems. One major difficulty with high order spatial discretizations is the numerical treatment of boundary conditions and grid interface conditions. To achieve both stability and high accuracy, one candidate is the summation-by-parts simultaneous–approximation–term (SBP–SAT) finite difference method [4,22]. SBP finite difference operators were originally constructed in [11] for first derivatives. In this paper, we use the SBP finite difference operators in [15] to approximate second derivatives. Boundary conditions and grid interface conditions are imposed weakly by the SAT technique [2,3,13,14]. In the SBP–SAT framework, the energy method is often used to derive an energy estimate that guarantees stability.

A commonly used measure for accuracy is convergence rate, typically in L_2 norm. The convergence rate indicates how fast the error in the numerical solution approaches zero as the mesh size goes to zero. The error in the numerical solution is caused by truncation errors. The truncation error near a boundary is often larger than in the interior of the computational domain, but the large boundary truncation error is localized at a fixed number of grid points. As a consequence, its effect on the convergence rate may be weakened. This is often called *gain in convergence*. A rule of thumb says m orders are gained in convergence for a PDE with m^{th} order spatial derivative, and such a gain is termed as *optimal*. The analysis of gain in convergence for different PDEs has been a long–standing research topic.

It is well–known that by directly applying the energy method to the error equation, 1/2 order is gained in the convergence rate compared with the largest truncation error. A noticeable exception is that in [1], it is proven that 3/2 orders are gained in the convergence rate for parabolic problems by a careful derivation of the energy estimate. However, numerical experiments in [1] show a gain of two orders in convergence. This indicates that although an energy estimate gives an upper bound of the error, it is often not sharp.

A more powerful tool for stability and accuracy analysis is normal mode analysis, which is used for example in [6] to prove that under reasonable conditions one order is gained in the convergence rate for first order hyperbolic systems. The idea is based on Laplace transforming the error equation in time, which leads to a system of linear equations referred to as the boundary system. The optimal gain straightforwardly follows if the boundary system is nonsingular for all $\text{Re}(s) \geq 0$, where s is the dual variable of time in Laplace space. For such cases we use the same terminology as in [8, pp. 292] and say that the boundary system satisfies the determinant condition.

In [21] the concept of pointwise stability is put forward as a condition implying the determinant condition, and therefore leading to the optimal gain. In [22] a sufficient condition of pointwise stability for an initial–boundary–value problem with a first derivative in time is discussed. However, we find that for the wave equation, the determinant condition does not follow from pointwise stability as defined in [21], and such an example is presented in "Appendix 2".

In fact, there are many examples of discretized problems that violate the determinant condition at points along the imaginary axis, even though the discretization is stable by an energy estimate. In [17], the Schrödinger equation with a grid interface is considered and is shown to be of this type. Analysis in Laplace space is performed and yields sharper error estimates than the $1/2$ order gain obtained by applying the energy method to the error equation in physical space.

In this paper, we consider the wave equation in the second order form. We find that also for this problem, the determinant condition is violated in many interesting settings even though there is an energy estimate. In particular, we consider problems with Dirichlet boundary conditions, Neumann boundary conditions and a problem with a grid interface, which are referred to as the Dirichlet problem, the Neumann problem and the interface problem, respectively. If the determinant condition is satisfied, we get the optimal convergence rate. If the determinant condition is not satisfied on the imaginary axis, the results in [8] are not valid. We then carefully derive an estimate for the solution of the boundary system to see how much is gained in the convergence rate. In addition, we have found by a careful analysis of the boundary system that in certain cases the gain in convergence is higher than the optimal gain, which we refer to as *super–convergence*.

The main contributions in this paper are: for the second order wave equation (1) an energy estimate does not imply an optimal convergence rate; (2) the determinant condition is not necessary for an optimal convergence rate; (3) if there is an energy estimate but the determinant condition is not satisfied, there can be an optimal gain of order 2, or a non–optimal gain of order 1, or only order $1/2$; (4) in certain cases it is possible to obtain a super–convergence, i.e. a gain of 2.5 orders.

The rest of the paper is organized as follows. We start in Sect. 2 with the SBP–SAT method applied to the one dimensional wave equation with Dirichlet boundary conditions. We apply the energy method and normal mode analysis, and derive results that correspond to the second, fourth and sixth order schemes. We then discuss the Neumann problem in Sect. 3. In Sect. 4, we consider the one dimensional wave equation on a grid with a grid interface. In Sect. 5, we perform numerical experiments. The computational convergence results support the accuracy analysis. In the end, we conclude the work in Sect. 6.

2 The One Dimensional Wave Equation with Dirichlet Doundary Conditions

To describe the properties of the equation and numerical methods, we need the following definitions. Let $w_1(x)$ and $w_2(x)$ be real–valued functions in $L_2[0, \infty)$. An inner product is defined as $(w_1, w_2) = \int_0^\infty w_1 w_2 dx$ with a corresponding norm $\|w_1\|^2 = (w_1, w_1)$, and w_1 is said to be in L_2 if $\|w_1\| < \infty$.

The second order wave equation on a half–line in one space dimension takes the form

$$U_{tt} = U_{xx} + F, \ x \in [0, \infty), \ 0 \le t \le t_f,$$
$$U(x, 0) = G(x), \ U_t(x, 0) = \bar{G}(x). \tag{1}$$

We consider the equation up to some finite time t_f. For boundary conditions, we use either the Dirichlet boundary condition:

$$U(0, t) = M(t), \tag{2}$$

or the Neumann boundary condition

$$U_x(0, t) = M(t). \tag{3}$$

The forcing function F, the initial data G, $\bar{G}$ and the boundary data M are compatible smooth functions with compact support. As a consequence, the true solution is smooth. For the problem (1) in a bounded domain, there is one boundary condition at each boundary. For the half line problem in consideration, the right boundary condition is substituted by requiring that the L_2 norm of the solution is bounded, i.e. $\|U(\cdot, t)\| < \infty$. If the corresponding half line and Cauchy problems are well–posed, then the original initial–boundary–value problem is well–posed [8, pp. 256].

The problem (1) with (2) or (3) is well–posed if the solution can be bounded in terms of the data. Multiplying Eq. (1) by U_t and integrating in space, with the homogeneous Dirichlet or Neumann boundary condition, it follows from the integration–by–parts principle

$$\frac{d}{dt}\|\|U\|\| \leq \|F\|, \tag{4}$$

where $\|\|U\|\|^2 = \|U_t\|^2 + \|U_x\|^2$ is the continuous energy, a semi–norm of the solution U. The standard energy estimate follows from an integration in time of (4),

$$
\begin{aligned}
\|\|U\|\| &\leq \sqrt{\|G_x\|^2 + \|\bar{G}\|^2} + \int_0^t \|F(\cdot, z)\| dz \\
&\leq \sqrt{\|G_x\|^2 + \|\bar{G}\|^2} + t \max_{0 \leq z \leq t} \|F(\cdot, z)\|
\end{aligned}
\tag{5}
$$

Therefore, problem (1) is well–posed with either (2) or (3).

Remark 1 We are interested in the solution U, but the standard energy analysis above only shows that a semi–norm of U is bounded. To include U itself in the estimate, we use the relation

$$\frac{d}{dt}\|U\|^2 = 2\|U\|\frac{d}{dt}\|U\|$$

and

$$\frac{d}{dt}\|U\|^2 = \int_0^\infty \frac{\partial}{\partial t} U^2 dx = \int_0^\infty 2UU_t dx \leq 2\|U\|\|U_t\|$$

to obtain

$$\frac{d}{dt}\|U\| \leq \|U_t\| \leq \|\|U\|\|.$$

An integration in time, together with (5), gives

$$
\begin{aligned}
\|U\| &\leq \|G\| + t\sqrt{\|G_x\|^2 + \|\bar{G}\|^2} + \int_0^t \left(\int_0^y \|F(\cdot, z)\| dz \right) dy \\
&\leq \|G\| + t\sqrt{\|G_x\|^2 + \|\bar{G}\|^2} + t^2 \max_{0 \leq z \leq t} \|F(\cdot, z)\|
\end{aligned}
\tag{6}
$$

Therefore, $\|U\|$ is bounded by the data in any bounded time interval $t \in [0, t_f]$, and is in the space of L_2.

To discretize the equation in space, we introduce an equidistant grid $x_i = ih$, $i = 0, 1, 2, \cdots$, and a grid function $u_i(t) \approx U(x_i, t)$. Furthermore, let $u(t) = [u_0(t), u_1(t), \cdots]^T$ where ^{T} denotes the transpose. We also define an inner product and norm for the grid functions a and b in $\mathcal{R}$ as $(a, b)_H = a^T H b$ and $\|a\|_H^2 = a^T H a$, respectively, where H is a symmetric

Table 1 α_{2p} values

$2p$	2	4	6	8	10
α_{2p}	0.4	0.2508560249	0.1878715026	0.0015782259	0.0351202265

positive definite operator on the space of grid functions. The norm $\| \cdot \|_H$ is equivalent to the standard discrete L$_2$ norm, and they are the same if $H = hI$ with an identity operator I.

An SBP operator has the standard central finite difference stencil in the interior and a special one–sided stencil near the boundary. The boundary closure is chosen such that the SBP operator mimics integration–by–parts via its associated inner product. The SBP operator approximating the second derivative $D \approx \partial^2 / \partial x^2$ has a structure

$$D = H^{-1}(-A + BS),\tag{7}$$

where H is diagonal and positive definite, A is symmetric positive semi-definite and $B = \text{diag}(-1, 0, 0, \cdots)$. S is a one sided approximation of the first derivative at the boundary.

In this paper, we use the diagonal norm SBP operators constructed in [15]. Although these operators are termed as $2p^{th}$ order accurate, the truncation error of D is $\mathcal{O}(h^{2p})$ in the interior but only $\mathcal{O}(h^p)$ near the boundary. The truncation error of S is $\mathcal{O}(h^{p+1})$. Such operators are constructed for $p = 1, 2, 3, 4$ in [15], and $p = 5$ in [12].

An SBP operator itself does not impose any boundary condition. It is important that boundary conditions are imposed in a way such that an energy estimate can be derived to ensure stability. Such numerical boundary treatments for the wave equation include the projection method [16] and the ghost point approach [19], which impose boundary conditions strongly. In this paper, we instead use a weak enforcement technique, the SAT method [3], since it is easy to derive an energy estimate even in higher dimensions. The semi–discretization of (1) and (2) with the homogeneous boundary data reads:

$$u_{tt} = Du + P_D u + f,\tag{8}$$

where $P_D = -H^{-1}S^T E_0 - \tau h^{-1}H^{-1}E_0$, $e_0 = [1, 0, 0, \cdots]^T$, $E_0 = e_0 e_0^T$ and $f(t)$ is the restriction of $F(x, t)$ to the grid. On the right hand side, the first term Du is an SBP approximation of U_{xx}, while the second term $P_D u$ imposes weakly the boundary condition $U(0, t) = 0$. It acts as a penalty term dragging the numerical solution at the boundary towards zero so that the boundary condition is simultaneously approximated, but the boundary condition is in general not satisfied exactly. The penalty parameter τ is to be determined so that an energy estimate of the discrete solution exists, which ensures stability.

2.1 Stability

In [2,13,14], it is shown that the operator A in (7) can be written as

$$A = h\alpha_{2p}(BS)^T BS + \tilde{A},\tag{9}$$

where $\alpha_{2p} > 0$ is a constant independent of h and $\tilde{A}$ is symmetric positive semi-definite. This is often called the borrowing trick. The values of α_{2p} are computed for $p = 1, 2$ and 3 in [13]. We use the same technique to compute α_{2p} for $p = 4$ and 5, and list the results in Table 1.

To derive an energy estimate, we introduce the energy

$$\|u\|_{D,h}^2 = \|u_t\|_H^2 + \|u\|_{\tilde{A}}^2 + \left(\sqrt{h\alpha_{2p}}(BSu)_0 - \frac{u_0}{\sqrt{h\alpha_{2p}}}\right)^2 + h^{-1}(\tau - \frac{1}{\alpha_{2p}})u_0^2, \quad (10)$$

where $\tau \geq 1/\alpha_{2p}$. Multiplying Eq. (8) by $u_t^T H$ from the left, it follows from (7) and (9)

$$\frac{d}{dt}(\|u\|_{D,h}^2) = 2u_t^T Hf.$$

The discrete energy estimate

$$\|u(t)\|_{D,h} \leq \|u(0)\|_{D,h} + \int_0^t \|f(z)\|_H dz$$
$$\leq \|u(0)\|_{D,h} + t \max_{0 \leq z \leq t} \|f(z)\|_H$$

is an analogue of the continuous energy estimate (5). We conclude that the semi–discretization (8) is energy stable if $\tau \geq 1/\alpha_{2p}$. The term $\|u\|_H$ can be included in the energy estimate in the same way as in Remark 1.

2.2 Accuracy Analysis by the Energy Method

The error equation for the pointwise error $\zeta_j = U(x_j, t) - u_j(t)$ is

$$\zeta_{tt} = D\zeta + P_D\zeta + T^{2p}, \quad (11)$$

where $\|\zeta(t)\|_h < \infty$, T^{2p} is the truncation error and $2p$ is the accuracy order of the SBP operator D. The solution has compact support during the entire time interval in consideration, so does the truncation error. Therefore, $T_j^{2p} = 0$, $j \geq J$ for some $J \sim \mathcal{O}(h^{-1})$. The first m components of T^{2p} are of order $\mathcal{O}(h^p)$, and all the other nonzero components are of order $\mathcal{O}(h^{2p})$. For $2p = 2, 4, 6, 8, 10$, the corresponding values of m are 1, 4, 6, 8, 11. In the analysis, we only consider the leading term of the truncation error. We introduce the interior truncation error $T^{2p,I}$, and the boundary truncation error $T^{2p,B}$ such that $T^{2p} = h^{2p}T^{2p,I} + h^p T^{2p,B}$, where $T^{2p,I}$ and $T^{2p,B}$ are independent of h, but depend on the derivatives of U. Since the number of grid points with the large truncation error $\mathcal{O}(h^p)$ is finite and independent of h, we have

$$\|T^{2p}\|_h^2 = h(h^{4p}\underbrace{\sum_{i=m}^{\infty}|T_i^{2p,I}|^2}_{\mathcal{O}(h^{-1})} + h^{2p}\underbrace{\sum_{i=0}^{m-1}|T_i^{2p,B}|^2}_{\mathcal{O}(1)}).$$

For $2p = 2, 4, 6, 8, 10$ and small h, the first term is much smaller than the second one. Thus, we have $\|T^{2p}\|_h \leq Kh^{p+1/2}$. Here and in the rest of the paper, we use the capital letter K in the estimate to denote some constant independent of the grid spacing. The constant K can be made precise by Taylor expansions, but we do not distinguish them from one estimate to another for a sake of simplified notations.

By applying the energy method to the error equation (11), we obtain for the discrete energy defined by (10)

$$\|\zeta\|_{D,h} \leq Kh^{p+1/2}. \quad (12)$$

This means that by the energy method we can only prove a gain in accuracy of 1/2 order. Therefore, the convergence rate is at least $p + 1/2$ if the numerical scheme is stable, that is if the penalty parameter $\tau \geq 1/\alpha_{2p}$.

2.3 Normal Mode Analysis for the Boundary Truncation Error

To derive a sharper estimate, we partition the pointwise error into two parts, the interior error ϵ^I and the boundary error ϵ, such that $\zeta = \epsilon^I + \epsilon$. ϵ^I is the error due to the interior truncation error, and ϵ is the error due to the boundary truncation error. ϵ^I can be estimated by the energy method, yielding

$$\left\|\left\|\epsilon^I\right\|\right\|_{D,h} \leq Kh^{2p}. \tag{13}$$

As mentioned above, $\|\epsilon^I\|_H$ can be bounded similarly.

We perform a normal mode analysis to analyze ϵ for the second, fourth and sixth order SBP-SAT schemes. The boundary error equation is

$$\epsilon_{tt} = D\epsilon + P_D\epsilon + h^p T^{2p,B}, \tag{14}$$

where $\|\epsilon\|_h < \infty$. In the analysis of ϵ, we take the Laplace transform in time of (14),

$$s^2\hat{\epsilon} = D\hat{\epsilon} + P_D\hat{\epsilon} + h^p \hat{T}^{2p,B} \tag{15}$$

for $\mathrm{Re}(s) > 0$, where $\hat{}$ denotes Laplace transform. After multiplying by h^2 on both sides of (15), we obtain with the notation $\tilde{s} = sh$

$$\tilde{s}^2\hat{\epsilon} = h^2 D\hat{\epsilon} + h^2 P_D\hat{\epsilon} + h^{p+2} \hat{T}^{2p,B}. \tag{16}$$

Note that the coefficients in the first two terms in the right hand side of (16) are h-independent, and that $\hat{T}^{2p,B}$ is nonzero only at a few points near the boundary.

There are essentially two steps in the normal mode analysis. In the first step, by a detailed analysis of the error equation (16), we derive an estimate of the solution to (16) in terms of the forcing. We use Parseval's relation in the second step to derive an estimate for the error in the physical space, which involves an inverse Laplace transform. As will be seen later, the integration is performed along the vertical line $\mathrm{Re}(\tilde{s}) = \eta h$ with $\eta > 0$ a fixed constant independent of h. It is therefore important to derive a sharp error estimate in Laplace space when $\mathrm{Re}(\tilde{s})$ goes to zero. The final estimate in the physical space is in the form

$$\sqrt{\int_0^{t_f} \|\epsilon(\cdot, t)\|_h^2 dt} \leq Kh^q, \tag{17}$$

where K depends only on η, the final time t_f and the derivatives of the true solution U.

Convergence rate is an asymptotic property, and we need the solution to be smooth and well resolved. By the smoothness assumption $|\hat{T}^{2p}(\eta + i\xi)|$ decreases fast for large $|\xi|$, making contributions from $|\xi| > \bar{K}$ insignificant for some constant $\bar{K}$. In the analysis we will only consider $s = \eta + i\xi$, with η positive and $|\xi| < \bar{K}$. Correspondingly we only consider $|\tilde{s}| \ll 1$, $\mathrm{Re}(\tilde{s}) = \mathcal{O}(h)$. In particular it suffices to consider $\tilde{s} = 0$ when checking the determinant condition. We formalize this discussion by making the following assumption.

Assumption 1 There is a constant $\bar{K} < \infty$ such that contributions from $s = \eta + i\xi$, $|\xi| > \bar{K}$ to the estimate (17) do not influence the rate q.

We summarize the convergence result for the Dirichlet problem in Theorem 1.

Table 2 Theoretical convergence result for the one dimensional wave equation with the Dirichlet boundary condition with $\tau > 1/\alpha_{2p}$

$2p$	$q_B = p + \text{Gain}$	$q = \min(q_B, 2p)$
2	3.5=1+2.5	2
4	4 =2+2	4
6	5.5=3+2.5	5.5

$2p$: interior order. q_B : order of the boundary error. q: convergence rate

Theorem 1 *Consider the second, fourth and sixth order stable SBP-SAT approximations (8) of the second order wave equation (1) with the Dirichlet boundary condition. With Assumption 1, the rates q in (17) depend on τ:*

1. *If $\tau = 1/\alpha_{2p}$, then $q = p + 1/2$, i.e. the gain in convergence is only half an order.*
2. *If $\tau > 1/\alpha_{2p}$, then the interior and boundary truncation errors lead to errors $\mathcal{O}(h^{2p})$ and $\mathcal{O}(h^{q_B})$ in the solution, respectively. The values q_B and the overall rates $q = \min(2p, q_B)$ are listed in Table 2.*

Remark 2 The penalty parameter τ has an effect on the convergence rate for a stable scheme. More precisely, with the same SBP operator, the convergence rate is always higher with $\tau > 1/\alpha_{2p}$ than with $\tau = 1/\alpha_{2p}$. We therefore should always choose $\tau > 1/\alpha_{2p}$ in computations, but bearing in mind that a moderate value of τ is appropriate because a very large τ has a negative impact on the CFL condition.

The convergence rates are optimal for the second and fourth order schemes, while super–convergence is obtained with the sixth order scheme. We also comment that the overall convergence rate for the second order scheme is dominated by the interior truncation error.

After some preliminaries we will prove Theorem 1 for $2p = 2$, 4 and 6 in Sects. 2.3.2, 2.3.3 and 2.3.4, respectively.

2.3.1 Solution to the Error Equation

To begin with, we note that sufficiently far away from the boundary, the penalty terms and the boundary truncation error in (16) are not present. The coefficients in D correspond to standard central finite difference schemes:

$$2p = 2: \quad s^2\hat{\epsilon}_j = D_+ D_- \hat{\epsilon}_j, \quad j = 3, 4, 5, \cdots, \tag{18a}$$

$$2p = 4: \quad s^2\hat{\epsilon}_j = (D_+ D_- - \frac{1}{12}(D_+ D_-)^2)\hat{\epsilon}_j, \quad j = 4, 5, 6, \cdots, \tag{18b}$$

$$2p = 6: \quad s^2\hat{\epsilon}_j = (D_+ D_- - \frac{h^2}{12}(D_+ D_-)^2 + \frac{h^4}{90}(D_+ D_-)^3)\hat{\epsilon}_j, \quad j = 5, 6, 7, \cdots, \tag{18c}$$

where

$$D_+ \hat{\epsilon}_j = \frac{\hat{\epsilon}_{j+1} - \hat{\epsilon}_j}{h} \quad \text{and} \quad D_- \hat{\epsilon}_j = \frac{\hat{\epsilon}_j - \hat{\epsilon}_{j-1}}{h}.$$

The corresponding characteristic equations are

$$2p = 2: \quad \kappa^2 - (2 + \tilde{s}^2)\kappa + 1 = 0, \tag{19a}$$

$$2p = 4: \quad \kappa^4 - 16\kappa^3 + (30 + 12\tilde{s}^2)\kappa^2 - 16\kappa + 1 = 0, \tag{19b}$$

$$2p = 6: \quad 2\kappa^6 - 27\kappa^5 + 270\kappa^4 - (180\tilde{s}^2 + 490)\kappa^3 + 270\kappa^2 - 27\kappa + 2 = 0. \tag{19c}$$

It is easy to verify by the von Neumann analysis that the interior numerical scheme is stable when applied to the corresponding periodic problem. From Lemma 12.1.3 in [8, pp. 379], it is straightforward to prove that there is no root of (19) with $|\kappa| = 1$ for $\text{Re}(s) > 0$. We call a root κ an admissible root if $|\kappa| < 1$. Since the problem in consideration is a half line problem, any κ with $|\kappa| > 1$ is not admissible because it results in an unbounded solution. We will need the following specifics for the roots.

Lemma 1 *For $2p = 2, 4, 6$, the number of admissible roots of (19) satisfying $|\kappa| < 1$ for $\text{Re}(\tilde{s}) > 0$ is 1,2,3, respectively. In the vicinity of $\tilde{s} = 0$, they are given by*

$$2p = 2: \quad \kappa_1 = 1 - \tilde{s} + \mathcal{O}(\tilde{s}^2), \tag{20a}$$

$$2p = 4: \quad \kappa_1 = 1 - \tilde{s} + \mathcal{O}(\tilde{s}^2), \; \kappa_2 = 7 - 4\sqrt{3} + \mathcal{O}(\tilde{s}^2), \tag{20b}$$

$$2p = 6: \quad \kappa_1 = 1 - \tilde{s} + \mathcal{O}(\tilde{s}^2), \; \kappa_2 = 0.0519 - 0.0801i + \mathcal{O}(\tilde{s}^2),$$

$$\kappa_3 = 0.0519 + 0.0801i + \mathcal{O}(\tilde{s}^2). \tag{20c}$$

Proof $2p = 2$: Eq. (19a) has two roots: $\kappa_{1,2} = 1 + \frac{1}{2}\tilde{s}^2 \pm \frac{1}{2}\sqrt{\tilde{s}^4 + 4\tilde{s}^2}$. We find by Taylor expansion at $\tilde{s} = 0$ that $\kappa_1 = 1 - \tilde{s} + \mathcal{O}(\tilde{s}^2)$ and $\kappa_2 = 1 + \tilde{s} + \mathcal{O}(\tilde{s}^2)$. Thus, in the vicinity of $\tilde{s} = 0$ the admissible root is $\kappa_1 = 1 - \tilde{s} + \mathcal{O}(\tilde{s}^2)$.

$2p = 4$: Eq. (19b) has four roots:

$$\kappa_1 = 4 - \sqrt{9 - 3\tilde{s}^2} - \sqrt{24 - 3\tilde{s}^2 - 8\sqrt{9 - 3\tilde{s}^2}},$$

$$\kappa_2 = \sqrt{9 - 3\tilde{s}^2} - \sqrt{8\sqrt{9 - 3\tilde{s}^2} - 3\tilde{s}^2 + 24} + 4,$$

$$\kappa_3 = \sqrt{8\sqrt{9 - 3\tilde{s}^2} - 3\tilde{s}^2 + 24} + \sqrt{9 - 3\tilde{s}^2} + 4,$$

$$\kappa_4 = \sqrt{24 - 3\tilde{s}^2 - 8\sqrt{9 - 3\tilde{s}^2}} - \sqrt{9 - 3\tilde{s}^2} + 4.$$

We find by Taylor expansion at $\tilde{s} = 0$ that

$$\kappa_1 = 1 - \tilde{s} + \mathcal{O}(\tilde{s}^2), \qquad \kappa_2 = 7 - 4\sqrt{3} + \mathcal{O}(\tilde{s}^2),$$
$$\kappa_3 = 7 + 4\sqrt{3} + \mathcal{O}(\tilde{s}^2), \qquad \kappa_4 = 1 + \tilde{s} + \mathcal{O}(\tilde{s}^2).$$

Thus, the admissible roots are $\kappa_1 = 1 - \tilde{s} + \mathcal{O}(\tilde{s}^2)$ and $\kappa_2 = 7 - 4\sqrt{3} + \mathcal{O}(\tilde{s}^2)$.

$2p = 6$: There is no closed form of the solution to a sixth order equation like (19c). However, (19c) with $\tilde{s} = 0$ can be factorized to a second order polynomial multiplied by a fourth order polynomial, thus an analytical solution can be obtained. We then perform a perturbation analysis to the six roots, and find analytical expressions for their dependence on $\tilde{s}$ in a neighbourhood of $\tilde{s} = 0$. The three admissible roots are given by (20c). We show the numerical values with four digits since the analytical expressions are very lengthy. □

The pointwise error away from the boundary can be expressed as

$$2p = 2: \quad \hat{\epsilon}_j = \sigma_1 \kappa^{j-2}, \ j = 2, 3, 4, \cdots, \tag{21a}$$

$$2p = 4: \quad \hat{\epsilon}_j = \sigma_1 \kappa_1^{j-2} + \sigma_2 \kappa_2^{j-2}, \ j = 2, 3, 4, \cdots, \tag{21b}$$

$$2p = 6: \quad \hat{\epsilon}_j = \sigma_1 \kappa_1^{j-3} + \sigma_2 \kappa_2^{j-3} + \sigma_3 \kappa_3^{j-3}, \ j = 3, 4, 5, \cdots, \tag{21c}$$

where the coefficients σ are determined by the boundary closure. The error equation (16) corresponding to a few grid points near the boundary also determines the pointwise errors that are not determined by 21. The general form for the L_2 norm of $\hat{\epsilon}$ can be written as

$$
\begin{aligned}
\|\hat{\epsilon}\|_h^2 &= h \sum_{i=0}^{d} |\hat{\epsilon}_i|^2 + h \sum_{m=1}^{p} \sum_{j=d+1}^{\infty} |\sigma_m \kappa_m^{j-d-1}|^2, \\
&= h \sum_{i=0}^{d} |\hat{\epsilon}_i|^2 + h \sum_{m=1}^{p} \frac{|\sigma_m|^2}{1 - |\kappa_m|^2}
\end{aligned}
\tag{22}
$$

where $d = 1, 1, 2$ for $2p = 2, 4, 6$, respectively. Note that by Lemma 1 we have that in the second term the factor $\frac{1}{1-|\kappa_1|^2}$ cannot be bounded independent of h when $\tilde{s} = \mathcal{O}(h)$.

Lemma 2 *Consider $Re(\tilde{s}) = \eta h > 0$, where η is a constant independent of h. For $2p = 2, 4, 6$ the admissible root $\kappa_1(\tilde{s})$ in 20 satisfies*

$$\frac{1}{1 - |\kappa_1|^2} \leq \frac{1}{2\eta h}$$

to the leading order.

Proof Let $\tilde{s} = x + iy$ where x, y are real numbers. Then $|x| \geq \eta h$.

$$
\begin{aligned}
1 - |\kappa_1(\tilde{s})|^2 &= 1 - |1 - \tilde{s} + \mathcal{O}(\tilde{s}^2)|^2 \\
&\geq 1 - |1 - \tilde{s}|^2 + \mathcal{O}(|\tilde{s}|^2) \\
&= 1 - |1 - x - iy|^2 + \mathcal{O}(|\tilde{s}|^2) \\
&= 2x - x^2 - y^2 + \mathcal{O}(|\tilde{s}|^2) \\
&= 2\mathrm{Re}(\tilde{s}) + \mathcal{O}(|\tilde{s}|^2).
\end{aligned}
$$

The desired estimate follows.   □

By Lemma 2, (22) becomes

$$\|\hat{\epsilon}\|_h^2 \leq h \sum_{i=0}^{d} |\hat{\epsilon}_i|^2 + \frac{K}{\eta} |\sigma_1|^2 + Kh \sum_{m=2}^{p} |\sigma_m|^2. \tag{23}$$

In the following sections, we analyze the error equation (16) corresponding to the grid points near the boundary, and derive bounds for $\hat{\epsilon}_i$ and σ_m in (23). The final estimate in the physical space of the form (17) follows by Parseval's relation. To keep a balance between clarity and paper length, we give a very detailed analysis for the second order scheme, and only show the main steps of the proof for the fourth and sixth order schemes.

2.3.2 Proof of Theorem 1 for the Second Order Scheme

The first three rows of (16) are affected by the boundary closure. They are

$$
\begin{aligned}
\tilde{s}^2 \hat{\epsilon}_0 &= \hat{\epsilon}_0 - 2\hat{\epsilon}_1 + \hat{\epsilon}_2 + 3\hat{\epsilon}_0 - 2\tau\hat{\epsilon}_0 + h^3 \hat{T}_0^{2,B}, \\
\tilde{s}^2 \hat{\epsilon}_1 &= \hat{\epsilon}_0 - 2\hat{\epsilon}_1 + \hat{\epsilon}_2 - 2\hat{\epsilon}_0, \\
\tilde{s}^2 \hat{\epsilon}_2 &= \hat{\epsilon}_1 - 2\hat{\epsilon}_2 + \hat{\epsilon}_3 - \frac{1}{2}\hat{\epsilon}_0.
\end{aligned}
\tag{24}
$$

By Taylor expansion, it is straightforward to derive $\hat{T}_0^{2,B} = -\hat{U}_{xxx}(0, s)$ to the leading order. We obtain the boundary system by rewriting (24) in a matrix–vector multiplication form

$$
\underbrace{\begin{bmatrix}
-1 & \tilde{s}^2 - 4 + 2\tau & 2 \\
-1 & 1 & \tilde{s}^2 + 2 \\
\kappa_1 - 2 - \tilde{s}^2 & \frac{1}{2} & 1
\end{bmatrix}}_{C_{2D}(\tilde{s},\tau)}
\underbrace{\begin{bmatrix}
\sigma_1 \\
\hat{\epsilon}_0 \\
\hat{\epsilon}_1
\end{bmatrix}}_{\Sigma_{2D}}
= h^3 \underbrace{\begin{bmatrix}
-1 \\
0 \\
0
\end{bmatrix} \hat{U}_{xxx}(0, s)}_{\hat{T}_v^{2,B}}.
\tag{25}
$$

If the determinant condition is satisfied, i.e. C_{2D} is nonsingular for $\mathrm{Re}(\tilde{s}) \geq 0$, then $|\Sigma_{2D}|$ is bounded by a constant multiplying h^3 and a gain of two orders from the boundary truncation error follows. In other words, we obtain the optimal convergence rate if the determinant condition is satisfied. Below we demonstrate that by analyzing the components of Σ_{2D}, it is in certain cases possible to obtain an even higher order gain from the boundary truncation error, which is referred to as super–convergence.

In Lemma 2, we use $\mathrm{Re}(\tilde{s}) = \eta h$ to estimate $1/(1 - |\kappa_1|^2)$, with η a constant independent of h. When analyzing the boundary system, we need to use the same $\tilde{s} = \eta h$ because that is where the inverse Laplace transform is performed later. We write the solution to (25) as $\Sigma_{2D}(\tilde{s}, \tau) = h^3 C_{2D}^{-1} \hat{T}_v^{2,B}$ for $\mathrm{Re}(\tilde{s}) = \eta h$. Since η is a constant and h is arbitrarily small, it is important to check how $\Sigma_{2D}(\tilde{s}, \tau)$ behaves as $\tilde{s}$ approaches zero. We start by setting $\tilde{s} = 0$, and find that when $\tau \neq 2.5$ the determinant condition is satisfied, i.e. $C_{2D}(0, \tau)$ is nonsingular; but the determinant condition is not satisfied when $\tau = 2.5$. The stability condition $\tau \geq 1/\alpha_2 = 2.5$ motivates us to consider these two cases separately.

In the case $\tau > 2.5$, the determinant condition is satisfied and we can expect at least an optimal gain of 2 orders in convergence. A perturbation analysis of Σ_{2D} at $\tilde{s} = 0$ shows that

$$
\begin{aligned}
\sigma_1 &= \left[-\frac{\tilde{s}^2}{2(2\tau - 5)} + \mathcal{O}(\tilde{s}^3) \right] \hat{U}_{xxx}(0, s) h^3 = \left[-\frac{\eta^2}{2(2\tau - 5)} + \mathcal{O}(\tilde{s}) \right] \hat{U}_{xxx}(0, s) h^5, \\
\hat{\epsilon}_0 &= \left[-\frac{1}{2\tau - 5} + \mathcal{O}(\tilde{s}^2) \right] \hat{U}_{xxx}(0, s) h^3, \quad \hat{\epsilon}_1 = \left[\frac{1}{2\tau - 5} + \mathcal{O}(\tilde{s}^2) \right] \hat{U}_{xxx}(0, s) h^3,
\end{aligned}
$$

which together with (23) leads to

$$
\|\hat{\epsilon}\|_h^2 \leq h \sum_{i=0}^{d} |\hat{\epsilon}_i|^2 + \frac{K}{\eta} |\sigma_1|^2 \leq K h^7 |\hat{U}_{xxx}(0, s)|^2.
\tag{26}
$$

Remember that $|\tilde{s}| \ll 1$ corresponds to $|s| < \bar{K}$ for some constant $\bar{K}$ and $h \ll 1$. By Assumption 1 and Parseval's relation, we have

$$\int_0^\infty e^{-2\eta t} \|\epsilon\|_h^2 dt = \frac{1}{2\pi} \int_{-\infty}^\infty \|\hat{\epsilon}(\eta + i\xi)\|_h^2 d\xi$$

$$\leq \frac{Kh^7}{2\pi} \int_{-\infty}^\infty |\hat{U}_{xxx}(0, \eta + i\xi)|^2 d\xi$$

$$= Kh^7 \int_0^\infty e^{-2\eta t} |U_{xxx}(0, t)|^2 dt.$$

We have derived the estimate for $\hat{\epsilon}$ in the vicinity of $\tilde{s} = 0$, but Assumption 1 allows us to use this estimate when integrating ξ from $-\infty$ to ∞. By arguing that future cannot affect past [8, pp. 294], we can replace the upper limit of the integrals on both sides by a finite time t_f. Since

$$\int_0^{t_f} e^{-2\eta t_f} \|\epsilon\|_h^2 dt \leq \int_0^{t_f} e^{-2\eta t} \|\epsilon\|_h^2 dt,$$

$$\int_0^{t_f} e^{-2\eta t} |U_{xxx}(0, t)|^2 dt \leq \int_0^{t_f} |U_{xxx}(0, t)|^2 dt,$$

we obtain

$$\sqrt{\int_0^{t_f} \|\epsilon\|_h^2 dt} \leq h^{3.5} \sqrt{Ke^{2\eta t_f} \int_0^{t_f} |U_{xxx}(0, t)|^2 dt}.$$

Thus, the boundary error is $\mathcal{O}(h^{3.5})$, which is 2.5 orders higher than the boundary truncation error. In practical computations, this super–convergence is not seen because the dominating source of error is the interior error $\mathcal{O}(h^2)$ given by (13). We conclude that the overall convergence rate is 2, and this proves that $q_B = 3.5$ and $q = 2$ in Theorem 1.

Remark 3 By only checking the determinant condition, we can prove the optimal convergence rate, i.e. a gain of two orders from the boundary truncation error. The above analysis demonstrates that it is important to analyze the components of the solution to the boundary system. Super–convergence is obtained when σ_1 is much smaller than the other components of the solution to the boundary system when h is sufficiently small. In this case, the error in the solution is larger at a few grid points near the boundary than in the interior.

In the case $\tau = 2.5$, the last two components of Σ_{2D} are infinite when $\tilde{s} = 0$ since $C_{2D}(0, \tau)$ is singular and the determinant condition is not satisfied. We again perform a perturbation analysis to the solution Σ_{2D} and obtain

$$\sigma_1 = \left[-\frac{1}{3} + \mathcal{O}(\tilde{s}) \right] \hat{U}_{xxx}(0, s)h^3,$$

$$\hat{\epsilon}_0 = \left[-\frac{2}{3}\tilde{s}^{-2} + \mathcal{O}(1) \right] \hat{U}_{xxx}(0, s)h^3 = \left[-\frac{2}{3}\eta^{-2} + \mathcal{O}(h^2) \right] \hat{U}_{xxx}(0, s)h,$$

$$\hat{\epsilon}_0 = \left[\frac{1}{3}\tilde{s}^{-2} + \mathcal{O}(1) \right] \hat{U}_{xxx}(0, s)h^3 = \left[\frac{1}{3}\eta^{-2} + \mathcal{O}(h^2) \right] \hat{U}_{xxx}(0, s)h.$$

This, in the same way as the case $\tau > 2.5$, leads to a gain of 0.5 order from the boundary truncation error and the overall convergence rate $p + 1/2 = 1.5$, which is the same as predicted by the energy estimate.

2.3.3 Proof of Theorem 1 for the Fourth Order Scheme

The first four rows of (16) are affected by the penalty terms, which can be written as the boundary system

$$C_{4D}(\tilde{s}, \tau) \Sigma_{4D} = h^4 \hat{T}_v^{4,B}. \tag{27}$$

As in the second order case in Sect. 2.3.2, we again consider $\text{Re}(\tilde{s}) = \eta h$ and write the solution to the boundary system (27) as $\Sigma_{4D} = h^4 C_{4D}^{-1}(\tilde{s}, \tau) \hat{T}_v^{4,B}$ where $\Sigma_{4D} = [\sigma_1, \sigma_2, \hat{\epsilon}_0, \hat{\epsilon}_1]^T$, and

$$\hat{T}_v^{4,B} = [\frac{11}{12}, -\frac{1}{12}, \frac{5}{516}, \frac{11}{588}]^T \hat{U}_{xxxx}(0, s). \tag{28}$$

The matrix $C_{4D}(0, \tau)$ is presented in "Appendix 1".

In the case $\tau > 1/\alpha_4$, $C_{4D}(0, \tau)$ is nonsingular and the determinant condition is satisfied. All the four components of Σ_{4D} are order $\mathcal{O}(h^4)$ with σ_1 independent of τ. This gives the estimate to the leading order

$$\|\hat{\epsilon}\|_h^2 \le \frac{K}{\eta} |\hat{U}_{xxxx}(0, s)|^2 h^8.$$

By Parseval's relation and by arguing that future cannot affect past, we obtain

$$\sqrt{\int_0^{t_f} \|\epsilon\|_h^2 dt} \le h^4 \sqrt{\frac{K e^{2\eta t_f}}{\eta} \int_0^{t_f} |U_{xxxx}(0, t)|^2 dt}.$$

Thus, the boundary error is $\mathcal{O}(h^4)$. In this case, the interior error is also $\mathcal{O}(h^4)$ given by (13). Therefore, the convergence rate is 4. This proves that $q_B = q = 4$ in Theorem 1.

In the case $\tau = 1/\alpha_4$, $C_{4D}(0, \tau_4)$ is singular and the determinant condition is not satisfied. By the energy estimate (12), the convergence rate is $p + 1/2 = 2.5$.

2.3.4 Proof of Theorem 1 for the Sixth Order Scheme

Similar to the previous two cases, we consider the six–by–six boundary system and analyze its solution $\Sigma_{6D} = [\sigma_1, \sigma_2, \sigma_3, \hat{\epsilon}_0, \hat{\epsilon}_1, \hat{\epsilon}_2]^T$:

$$C_{6D}(\tilde{s}, \tau) \Sigma_{6D} = h^5 \hat{T}_v^{6,B}, \tag{29}$$

where

$$\hat{T}_v^{6,B} = [-\frac{157525}{163788}, \frac{3869}{17580}, -\frac{30409}{54220}, \frac{65723}{321540}, -\frac{80821}{472620}, \frac{9015}{175204}]^T \hat{U}_{xxxxx}(0, s). \tag{30}$$

The matrix $C_{6D}(0, \tau)$ is presented in "Appendix 1". It is singular if $\tau = 1/\alpha_6$, and nonsingular otherwise.

When $\tau > 1/\alpha_6$, we write the solution to (29) as $\Sigma_{6D}(\tilde{s}, \tau) = h^5 C_{6D}^{-1} \hat{T}_v^{6,B}$. A calculation of $\Sigma_{6D}(\tilde{s}, \tau)$ at $\tilde{s} = 0$ shows that $\sigma_1 = 0$ and the other five components $\mathcal{O}(h^5)$, and a further perturbation analysis gives $\sigma_1 \sim \mathcal{O}(h^7)$ when $\tilde{s} = \mathcal{O}(h)$. As a consequence, from (23) we obtain that the boundary error is $\mathcal{O}(h^{5.5})$. In this case, the interior error is $\mathcal{O}(h^6)$ given by (13). Thus, the convergence rate is 5.5, which is a super–convergence and is half order higher than the optimal convergence. This proves that $q_B = q = 5.5$ in Theorem 1.

When $\tau = 1/\alpha_6$, the convergence rate $p + 1/2 = 3.5$ is given by the energy estimate (12).

3 The One Dimensional Wave Equation with Neumann Boundary Conditions

We consider Eq. (1) with the Neumann boundary condition (3), and use the same assumption of the data as for the Dirichlet problem. As will be seen later, the determinant condition is not satisfied. Our analysis gives sharp error estimates for such problems as well.

3.1 Stability

To discretize the equation in space, we use the grid and grid functions introduced for the Dirichlet problem. The semi–discretized equation of the Neumann problem is

$$u_{tt} = Du + H^{-1}E_0 Su + f. \tag{31}$$

On the right hand side of (31), the first term approximates U_{xx} and the second term imposes weakly the boundary condition $U_x(0, t) = 0$. Since we consider a half line problem, the SBP operator D is in the form $D = H^{-1}(-A - E_0 S)$. As a consequence, the semi–discretization (31) can be written as

$$u_{tt} = -H^{-1}Au + f. \tag{32}$$

Multiplying Equation (32) by $u_t^T H$ from the left, we obtain

$$\frac{d}{dt}\left(\|u_t\|_H^2 + \|u\|_A^2\right) = 2u_t^T Hf.$$

The discrete energy $\|\|u\|\|_{N,h}^2 = \|u_t\|_H^2 + \|u\|_A^2$ is bounded as

$$\begin{aligned}\|\|u(t)\|\|_{N,h} &\leq \|\|u(0)\|\|_{N,h} + \int_0^t \|f(z)\|_H dz \\ &\leq \|\|u(0)\|\|_{N,h} + t \max_{0 \leq z \leq t} \|f(z)\|_H\end{aligned} \tag{33}$$

Here again, $\|u\|_H$ can be included in the energy estimate as in Remark 1.

3.2 Accuracy

With the pointwise error defined as before, the error equation is

$$\zeta_{tt} = -H^{-1}A\zeta + T^{2p,N}, \tag{34}$$

where $\|\zeta\|_h < \infty$, $T^{2p,N}$ is the truncation error and $2p$ is the accuracy order of the SBP operator. In the same way as for the Dirichlet problem, by applying the energy method to the error equation (34), we obtain an estimate

$$\|\|\zeta\|\|_{N,h} \leq Kh^{p+1/2},$$

This means that the convergence rate of (32) is at least $p + 1/2$. In the following, we use normal mode analysis to derive a sharper bound of the error, which agrees well with the results from numerical experiments.

As in the Dirichlet problem, we partition the error ζ into two parts, the error ϵ^I due to the interior truncation error and the error ϵ due to the boundary truncation error. ϵ^I can be estimated by the energy method, yielding $\|\|\epsilon^I\|\|_{N,h} \leq Kh^{2p}$. The boundary error equation is

$$\epsilon_{tt} = -H^{-1}A\epsilon + h^p T^{2p,N,B}, \tag{35}$$

where $h^p T^{2p,N,B}$ is the boundary truncation error. $T^{2p,N,B}$ depends on the derivatives of the true solution U, but not h. Moreover, only the first m elements of $T^{2p,N,B}$ are nonzero, where m depends on p but not h.

We Laplace transform Equation (35) in time. With the notation $\tilde{s} = sh$, we obtain

$$\tilde{s}^2 \hat{\epsilon} = -h^2 H^{-1} A \hat{\epsilon} + h^{p+2} \hat{T}^{2p,N,B}. \tag{36}$$

Since the operator A in (36) is only symmetric positive semi–definite, we expect that the determinant condition is not satisfied for this problem. The characteristic equations of the interior error equations are the same as for the Dirichlet problem (19), and Lemma 1 and 2 are also applicable to the Neumann problem. Below we analyze the second, fourth and sixth order accurate cases.

3.2.1 Second Order Accurate Scheme

Only the first row of the error equation is affected by the boundary closure. In this case, the boundary system is the scalar equation

$$C_{2N}(\tilde{s})\sigma_1 = h^3 \hat{T}_0^{2,N,B}, \tag{37}$$

with

$$C_{2N}(\tilde{s}) = 2\tilde{s} + \mathcal{O}(\tilde{s}^2) \text{ and } \hat{T}_0^{2,N,B} = -\frac{2}{3}\hat{U}_{xxx}(0,s)$$

and the error is in the form (21a) with j starting at 0. Clearly the determinant condition is violated at $\tilde{s} = 0$, but we straightforwardly obtain

$$|\sigma_1| \leq \frac{K}{\eta}|\hat{U}_{xxx}(0,s)|h^2,$$

for $\tilde{s} = \mathcal{O}(h)$. In the same way as for the Dirichlet problem, an estimate in physical space of the type (17) follows with $q = 2$. Since the interior error is also $\mathcal{O}(h^2)$, the convergence rate is 2.

Remark 4 In this case, the gain from the boundary truncation error is only one order, and does not follow the $p + 2$ optimal gain. The interior error $\mathcal{O}(h^2)$ restricts the overall convergence rate to second order. Therefore, a gain of more than one order can normally not be observed. In the numerical experiments, we design a stable numerical scheme to verify that the gain in convergence is indeed only one order.

3.2.2 Fourth and Sixth Order Accurate Schemes

We follow the above approach to derive the four by four boundary system

$$C_{4N}(\tilde{s})\Sigma_{4N} = h^4 \hat{T}_u^{4,N,B},$$

for the fourth order scheme, and the sixth by sixth boundary system

$$C_{6N}(\tilde{s})\Sigma_{6N} = h^5 \hat{T}_u^{6,N,B},$$

for the sixth order scheme. The truncation errors $\hat{T}_u^{4,N,B}$, $\hat{T}_u^{6,N,B}$ are the same as the Dirichlet case (28), (30) except the coefficients of the first component are 43/204 and -53845/163788, respectively. The matrices $C_{4N}(0)$ and $C_{6N}(0)$ presented in "Appendix 1" are singular, indicating that the determinant condition is violated. We consider $\tilde{s} = \eta h$ with a constant η

Table 3 Theoretical convergence result for the one dimensional wave equation with the Neumann boundary condition

$2p$	$q_B = p + \text{Gain}$	$q = \min(2p, q_B)$
2	2 =1+1	2
4	4.5=2+2.5	4
6	5.5=3+2.5	5.5

$2p$: interior order. q_B: order of the boundary error. q: convergence rate

independent of h, and analyze the solution to the boundary system. For the fourth order case, σ_1 in (23) is $\sigma_1 \sim \mathcal{O}(\tilde{s}h^4) \sim \mathcal{O}(\eta h^5)$, and the other components are bounded by $\mathcal{O}(h^4)$. For the sixth order case $\sigma_1, \hat{\epsilon}_0 \sim \mathcal{O}(\tilde{s}h^5) \sim \mathcal{O}(\eta h^6)$, and the other components are bounded by $\mathcal{O}(h^5)$. Both cases lead to a gain of 2.5 orders from the boundary truncation error. We summarize the results on the Neumann problem in the following theorem.

Theorem 2 *Consider the second, fourth and sixth order stable SBP-SAT approximations (31) of the second order wave equation (1) with the Neumann boundary condition. With Assumption 1, the boundary truncation error lead to error $\mathcal{O}(h^{q_B})$ in the solution, where q_B and the overall rates $q = \min(2p, q_B)$ are listed in Table 3.*

4 The One Dimensional Wave Equation with a Grid Interface

In this section, we consider another example that does not satisfy the determinant condition. It is the Cauchy problem for the second order wave equation in one space dimension

$$U_{tt} = U_{xx} + F, \ x \in (-\infty, \infty), \ t \geq 0, \ \|U(\cdot, t)\| < \infty$$
$$U(x, 0) = G(x), \ U_t(x, 0) = \bar{G}(x), \tag{38}$$

where the forcing function F, the initial data G and $\bar{G}$ are compatible smooth functions with compact support. It is straightforward to derive the energy estimate of the form (5) for Eq. (38).

We solve the equation on a grid with a grid interface at $x = 0$. With the assumption that the true solution is smooth, it is natural to impose two interface conditions at $x = 0$: continuity of the solution and continuity of the first derivative of the solution.

We introduce the grid on the left $x_{-j} = -jh_L$, $j = 0, 1, 2, \cdots$, and the grid on the right $x_j = jh_R$, $j = 0, 1, 2, \cdots$, with the grid size h_L and h_R, respectively. The grid functions are $u_j^L(t) \approx U(x_{-j}, t)$ and $u_j^R(t) \approx U(x_j, t)$. Denote

$$e_{0L} = [\cdots, 0, 0, 1]^T, \quad e_{0R} = [1, 0, 0, \cdots]^T,$$
$$u^L = [\cdots, u_{-1}^L, u_0^L]^T, \quad u^R = [u_0^R, u_1^R, \cdots]^T.$$

Both $u_0^L = e_{0L}^T u^L$ and $u_0^R = e_{0R}^T u^R$ approximate $U(0, t)$. Both $(S_L u^L)_0 = e_{0L}^T S_L u^L$ and $(S_R u^R)_0 = e_{0R}^T S_R u^R$ approximate $U_x(0, t)$. To simplify notation, we define $\{u\} = u_0^L - u_0^R$ and $\{Su\} = (S_L u^L)_0 - (S_R u^R)_0$. The semi–discretized equation reads

$$u_{tt}^L = D_L u^L + P_{IL}\{u\} + f_L, \ u_{tt}^R = D_R u^R + P_{IR}\{u\} + f_R, \tag{39}$$

Table 4 Theoretical convergence result for the one dimensional wave equation with a grid interface when $\tau > \tilde{\tau}_{2p}$ in (40)

$2p$	$q_B^{(h_L=h_R)} = p + \text{Gain}$	$q_B^{(h_L \neq h_R)} = p + \text{Gain}$	$q = \min(2p, q_B)$
2	3.5=1+2.5	2=1+1	2
4	4.5=2+2.5	4=2+2	4

$2p$: interior order. $q_B = p + \text{Gain}$: order of the boundary error. q : convergence rate

where f_L and f_R are the restriction of $F(x, t)$ on the grid. The penalty terms $P_{IL}\{u\}$ and $P_{IR}\{u\}$ impose the interface conditions weakly and have the form

$$P_{IL}\{u\} = \frac{1}{2} H_L^{-1} S_L^T e_{0L}\{u\} - \frac{1}{2} H_L^{-1} e_{0L}\{Su\} - \tau H_L^{-1} e_{0L}\{u\},$$

$$P_{IR}\{u\} = \frac{1}{2} H_R^{-1} S_R^T e_{0R}\{u\} - \frac{1}{2} H_R^{-1} e_{0R}\{Su\} + \tau H_R^{-1} e_{0R}\{u\}.$$

The penalty parameters τ is chosen so that the semi–discretization is stable. By the energy method, the numerical scheme is stable if

$$\tau \geq \frac{h_L + h_R}{4\alpha_{2p} h_L h_R} := \tilde{\tau}_{2p}. \tag{40}$$

The stability proof is found in [13, Lemma 2.4, pp. 215]. If $h_L = h_R := h$, then $\tilde{\tau}_{2p} = 1/(2\alpha_{2p}h)$.

By the energy estimate, the convergence rate is at least $p+1/2$ if the semi–discretization is stable. In order to derive a sharper estimate, we follow the same approach as in the previous sections. The interior truncation error results in an error $\mathcal{O}(h^{2p})$ in the solution. In the remaining part of this section we will only consider the effect of the interface truncation error. Denote

$$\epsilon_{-j}^L(t) = U(x_{-j}, t) - u_{-j}^L(t), \quad \epsilon_j^R(t) = U(x_j, t) - u_j^R(t), \quad j = 0, 1, 2, \cdots.$$

The error equation in Laplace space reads

$$\tilde{s}_L^2 \hat{\epsilon}^L = h_L^2 D_L \hat{\epsilon}^L + h_L^2 P_{IL}\{\hat{\epsilon}\} + h_L^{p+2} \hat{T}^{2p,L,B}, \tag{41a}$$

$$\tilde{s}_R^2 \hat{\epsilon}^R = h_R^2 D_R \hat{\epsilon}^R + h_R^2 P_{IR}\{\hat{\epsilon}\} + h_R^{p+2} \hat{T}^{2p,R,B}, \tag{41b}$$

where $\tilde{s}_L = sh_L, \tilde{s}_R = sh_R$. Here the interface truncation errors $\hat{T}^{2p,L/R,B}$ are nonzero at only a few grid points near the interface. The errors are estimated for the second and fourth order schemes by the normal mode analysis, with the results summarized in the following theorem.

Theorem 3 *For the second and fourth order stable SBP-SAT approximation (39) of the wave equation with a grid interface, the numerical solution converges to the true solution at rate q. When $\tau = \tilde{\tau}_{2p}$ in (40), the rate $q = p + 1/2$ is given by the energy estimate. When $\tau > \tilde{\tau}_{2p}$, the rate q is listed in Table 4.*

Remark 5 In practical computations, we should always choose $\tau > \tilde{\tau}_{2p}$ so that the optimal convergence rate is obtained. Note that in this case, the gain in convergence depends on whether the grid spacing h_L is equal to h_R as shown in Table 4. The overall rate $q = \min(2p, q_B)$, however, is not affected by the relation of h_L and h_R.

We only need to analyze the values q_B when $\tau > \tilde{\tau}_{2p}$. We use the notation $h = h_L = rh_R$ where r is a fixed mesh size ratio. The analysis follows in the same way as before. We construct the characteristic equation on each side and solve them to obtain the general solution, which are given by (19) and (20), respectively. The boundary system is formulated by substituting the general solution to the error equation. The accuracy order is determined by how the solution of this system behaves with respect to h when $\tilde{s} = \mathcal{O}(h)$.

4.1 Proof of Theorem 3 for the Second Order Scheme

The last three rows of (41a) and the first three rows of (41b) are affected by the penalty terms, which lead to the six by six boundary system

$$C_{2I}(\tilde{s}, \tau)\Sigma = h^3 \hat{T}_u^{2,B},$$

where

$$\Sigma = [\sigma_{1L}, \sigma_{1R}, \hat{\epsilon}_0^L, \hat{\epsilon}_{-1}^L, \hat{\epsilon}_0^R, \hat{\epsilon}_1^R]^T,$$
$$\hat{T}_u^{2,B} = [\frac{2}{3} + \frac{1}{3r^2}, 0, 0, -\frac{2}{3r^3} - \frac{1}{3r}, 0, 0]^T \hat{U}_{xxx}(0, s).$$

The determinant condition is not satisfied because $C_{2I}(0, \tau)$ shown in "Appendix 1" is singular. We find that when $r = 1$,

$$\sigma_{1L}, \sigma_{1R} \sim \mathcal{O}(\tilde{s}^2 h^3) \sim \mathcal{O}(h^5) \text{ and } \hat{\epsilon}_0^L, \hat{\epsilon}_{-1}^L, \hat{\epsilon}_0^R, \hat{\epsilon}_1^R \sim \mathcal{O}(h^3),$$

which leads to a boundary error $\mathcal{O}(h^{3.5})$, i.e. a gain of 2.5 orders in convergence and $q_B^{(h_L=h_R)} = 3.5$, in the same way as in Sect. 2.3.2. When $r \neq 1$, all components of $\Sigma \sim \mathcal{O}(\tilde{s}^{-1} h^3) \sim \mathcal{O}(\tilde{h}^2)$. In this case, the gain in convergence is only 1 order, i.e. $q_B^{(h_L \neq h_R)} = 2$. In both cases, the overall convergence rate is $q = 2$.

4.2 Proof of Theorem 3 for the Fourth Order Scheme

In this case, the eight by eight boundary system takes the form

$$C_{4I}(\tilde{s}, \tau)\Sigma = h^4 \hat{T}_u^{4,B},$$

where

$$\Sigma = [\sigma_{1L}, \sigma_{2L}, \sigma_{1R}, \sigma_{2R}, \hat{\epsilon}_0^L, \hat{\epsilon}_{-1}^L, \hat{\epsilon}_0^R, \hat{\epsilon}_1^R]^T,$$
$$\hat{T}_u^{4,B} = [\frac{115}{204} - \frac{6}{17r^3}, -\frac{1}{12}, \frac{5}{516}, \frac{11}{588}, \frac{115}{204r^4} - \frac{6}{17r},$$
$$-\frac{1}{12r^4}, \frac{5}{516r^4}, \frac{11}{588r^4}]^T \hat{U}_{xxxx}(0, s).$$

The boundary system is also singular at $\tilde{s} = 0$, and the determinant condition is not satisfied. The matrix $C_{4I}(0, \tau)$ is presented in "Appendix 1". When $r = 1$,

$$\sigma_{1L}, \sigma_{1R} \sim \mathcal{O}(\tilde{s}h^4) \sim \mathcal{O}(h^5) \text{ and } \sigma_{2L}, \sigma_{2R}, \hat{\epsilon}_0^L, \hat{\epsilon}_{-1}^L, \hat{\epsilon}_0^R, \hat{\epsilon}_1^R \sim \mathcal{O}(h^4),$$

which leads to a boundary error $\mathcal{O}(h^{4.5})$, i.e. a gain of 2.5 orders in convergence and $q_B^{(h_L=h_R)} = 4.5$. When $r \neq 1$, all components of $\Sigma \sim \mathcal{O}(h^4)$. In this case, we gain two orders in convergence, i.e. $q_B^{(h_L \neq h_R)} = 4$. In both cases, the overall convergence rate is $q = 4$.
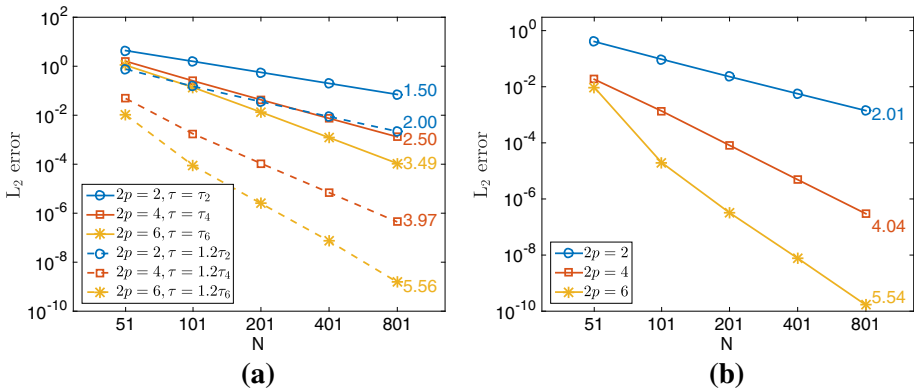
Fig. 1 A convergence test for the one dimensional wave equation with **a** Dirichlet and **b** Neumann boundary conditions

5 Numerical Experiments

In this section, we perform numerical experiments to verify the accuracy analysis. We investigate how the accuracy of the numerical solution is affected by the large truncation error localized near boundaries and grid interfaces. In the analysis, we use the half line problem. However, in the numerical experiments, we use a bounded domain and impose boundary conditions on all boundaries weakly. For the integration in time, we write the semi–discretized equation as a first order system in time and use the classical fourth order Runge–Kutta method. The step size Δt in time is chosen so that the temporal error has a negligible impact on the spatial convergence result. The value of Δt is given in each numerical experiment.

The L_2 error at a given time point are computed as the norm of the difference between the exact solution u^{ex} and the numerical solution u^h according to

$$\|u^h - u^{ex}\|_{L_2} = \sqrt{h^d (u^h - u^{ex})^T (u^h - u^{ex})},$$

where d is the spatial dimension. The convergence rate is computed by

$$q = \log\left(\frac{\|u^h - u^{ex}\|}{\|u^{2h} - u^{ex}\|}\right) \bigg/ \log\left(\frac{1}{2}\right). \tag{42}$$

5.1 The One Dimensional Wave Equation

We choose the analytical solution

$$u = \cos(10\pi x + 1)\cos(10\pi t + 2), \quad 0 \le x \le 1, \ 0 \le t \le 2, \tag{43}$$

and use it to impose the Dirichlet boundary conditions at $x = 0$ and $x = 1$. The step size in time $\Delta t = 0.1h$ is much smaller than the step size restricted by the stability condition. We use it because then the temporal error has a negligible impact on the spatial convergence rate. The analytical solution and its derivatives do not vanish at the boundaries at the final time $t = 2$. The errors in discrete L_2 norm are plotted versus the number of grid points in Fig. 1(a), and the convergence rates in L_2 norm are shown in the same figure.

With a $2p^{th}$ ($2p = 2, 4, 6$) order accurate method, the convergence rate is $p + 1/2$ in L_2 norm if the penalty parameter τ equals its limit. With a larger penalty parameter (increased by 20%), the accuracy of the numerical solution is improved significantly. The observed
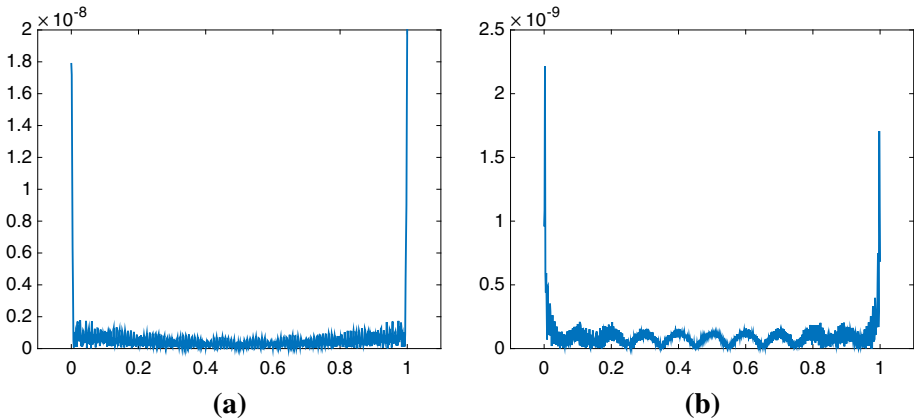
Fig. 2 Error plot for the. **a** Dirichlet and. **b** Neumann problem with the sixth order accurate scheme

convergence rates by (42) for the second, fourth and sixth order accurate schemes are 2.00, 3.97 and 5.56, which are very close to the theoretical rates 2, 4 and 5.5 in Table 2, Theorem 1. The super–convergence of the sixth order accurate scheme is also observed when solving the Schrödinger equation [17,18].

To test the Neumann problem, we again use (43) as the analytical solution and impose the Neumann boundary condition at two boundaries $x = 0$ and $x = 1$. In Fig. 1b we show the L_2 errors and the convergence rates in L_2 norm, which corroborate the accuracy analysis very well: for the second, fourth and sixth order accurate schemes, the observed convergence rates 2.01, 4.04 and 5.54 are quite close to the theoretical values 2, 4 and 5.5.

For the sixth order accurate scheme with $\tau > 1/\alpha_6$, the components of the solution to the boundary system are of different magnitude: $\sigma_1 \sim \mathcal{O}(h^7)$ while the other components are of order $\mathcal{O}(h^5)$. Therefore, the boundary truncation error only leads to $\mathcal{O}(h^7)$ interior error. The large error $\mathcal{O}(h^5)$ in the solution is located only at a few points near the boundary, which can be seen in Fig. 2a. The same is true for the Neumann problem with the sixth order accurate scheme, and the error plot in Fig. 2b also shows that the error in the solution is localized.

Next, we test how the accuracy and convergence are affected by the large truncation error localized near a grid interface. We choose the same analytical solution (43). The grid interface is located at $x = 0.5$ in the computational domain $\Omega = [0, 1]$, and the grid size ratio is 2:1. The step size in time is $\Delta t = 0.1h$, where h is the smaller grid size. Note that the analytical solution does not vanish at the grid interface. We use the SAT technique to impose the outer boundary condition weakly and choose the boundary penalty parameters strictly larger than their limits. The interface conditions are also imposed by the SAT technique. We use different interface penalty parameters to see how they affect the accuracy and convergence. The results are shown in Fig. 3a.

According to the accuracy analysis in Sect. 4, for second and fourth order accurate methods with $\tau = \tau_{2p}$, the expected convergence rates are 1.5 and 2.5 in L_2 norm. This is clearly observed in Fig. 3a. With a larger penalty parameter, a much better convergence result is obtained. For the second and fourth order methods, we get the optimal (second and fourth, respectively) order of convergence in L_2 norm. For the sixth order accurate scheme, the convergence rate is about 3.4 with $\tau = \tau_6$, which is in line with the $p + 1/2$ convergence
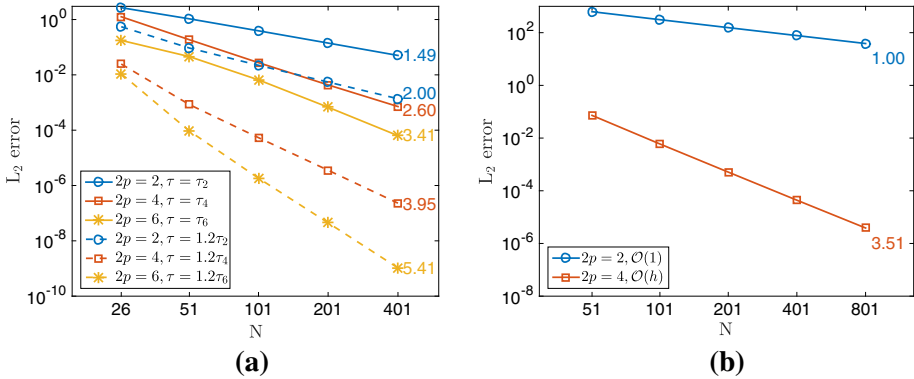
Fig. 3 Convergence tests for. **a** a grid interface problem. **b** a Neumann problem with schemes with added truncation errors

rate. With a larger penalty parameter, the convergence rate is about 5.4. Here we again observe a super–convergence phenomenon.

5.2 Non–Optimal Convergence

We have shown several cases where the gain in convergence does not follow the optimal $p + 2$ rule. For example, the sixth order accurate SBP–SAT scheme has a super–convergence property, which is both proved theoretically and observed in the numerical experiments. There are also cases where the accuracy analysis predicts a lower or higher order gain. Examples include the second and fourth order schemes for the Neumann problem, where analysis indicates a gain of 1 and 2.5 orders, respectively. In these cases, the interior error also plays an important role in the overall convergence rate. It is therefore unclear from the computation what the precise gain is in computations. We design the following two experiments to confirm that our accuracy analysis indeed gives sharp error estimates.

We use (43) as the analytical solution and impose the Neumann boundary condition at two boundaries $x = 0$ and $x = 1$. With this analytical solution, the equation does not have a forcing term, and $f = 0$ in (32). We modify the numerical scheme (32) to

$$u_{tt} = -H^{-1}Au + \mathcal{F}. \tag{44}$$

In the first experiment, we use the second order accurate scheme and choose the first component of $\mathcal{F}$ to be $(10\pi)^3$ and all the other components zero. We use the number $(10\pi)^3$ because the truncation error involves the third derivative of the true solution with the factor $(10\pi)^3$. The boundary truncation error of (44) has the same structure as in the original scheme, but increases to $\mathcal{O}(1)$, i.e. order 0. Note that energy stability still holds for (44). The L_2 error versus the number of grid points is plotted in Fig. 3b. Clearly, the convergence rate is 1, and is a gain of 1 order from the boundary truncation error. This corroborates the accuracy analysis very well.

In the second experiment, we use the fourth order accurate scheme and let the first four components of $\mathcal{F}$ be $h(10\pi)^4[43/204, -1/12, 5/516, 11/588]^T$, and the other components zero. The boundary truncation error of (44) has the same structure as in the original scheme, but increases to $\mathcal{O}(h)$. We plot the L_2 error versus the number of grid points in Fig. 3b, and observe a 3.5 orders convergence rate, i.e. a gain of 2.5 orders from the boundary truncation error. This again demonstrates that the error estimate is sharp.

6 Conclusion

For the second order wave equation a stable numerical scheme does not automatically satisfy the determinant condition, nor does it imply an optimal gain in convergence rate. We have considered stable SBP–SAT finite difference schemes for the Dirichlet, Neumann and interface problems. In all these cases the boundary/interface truncation error is larger than the interior truncation error. We find that the schemes satisfy the determinant condition only for the Dirichlet problem when the penalty parameter is greater than its limit value, and a gain of two orders in convergence follows by the standard analysis. By a careful analysis of the solution to the boundary system, we prove that the gain in convergence for some schemes is in fact 2.5 orders, i.e. half an order higher than the optimal gain.

For all the other cases the determinant condition is not satisfied, but there is nonetheless a gain in convergence of 0.5, 1, 2, or even 2.5 orders. In those cases, only a detailed analysis of the boundary system reveals the precise gain.

We have performed numerical experiments by using both the standard SBP–SAT scheme, and a modified version to confirm that our accuracy analysis gives sharp error estimates.

In this paper, we have performed accuracy analysis for problems in one space dimension. The technique and results can be straightforwardly generalized to two space dimensions if the boundary condition is periodic in one of the spatial dimensions. In a coming paper, we will show how to perform accuracy analysis for problems in two space dimensions when the boundary conditions in both dimensions are non–periodic. We will in particular consider a corner problem when the large truncation error is located on a few grid points at a conner in two space dimensions.

Appendix 1

The boundary systems for the high order schemes for the Dirichlet problem are $C_{4D}(0, \tau)$ in Sect. 2.3.3

$$C_{4D}(0, \tau) = \begin{bmatrix} -3 & 3 - 4\sqrt{3} & \frac{-122+48\tau}{17} & 5 \\ -1 & -1 & \frac{85}{59} & 2 \\ \frac{55}{43} & \frac{85+12\sqrt{3}}{43} & -\frac{68}{43} & -\frac{59}{43} \\ -\frac{1}{49} & -\frac{37+3\sqrt{3}}{49} & \frac{17}{49} & 0 \end{bmatrix},$$

and $C_{6D}(0, \tau)$ in Sect. 2.3.4

$$C_{6D}(0, \tau) = \begin{bmatrix} 2.9794 & 3.3686 + 0.0230i & 3.3686 - 0.0230i & 3.1651\tau - 9.3821 & 8.0245 & -8.2157 \\ 0.5705 & 1.0025 + 0.0469i & 1.0025 - 0.0469i & 1.8235 & 2.3504 & -1.8675 \\ -4.6151 & -6.7899 - 0.2266i & -6.7899 + 0.2266i & -4.1363 & -4.1376 & 8.1084 \\ 2.5660 & 4.5927 + 0.1966i & 4.5927 - 0.1966i & 0.8615 & 1.159 & -3.5117 \\ -1.2510 & -3.1883 - 0.2835i & -3.1883 + 0.2835i & -0.0951 & -0.9157 & 1.9876 \\ 0.1996 & 0.3281 + 0.1029i & 0.3281 - 0.1029i & -0.0400 & 0.1875 & -0.3471 \end{bmatrix}.$$

In Sect. 3.2.2, the boundary systems for the fourth and sixth order schemes for the Neumann problem are analyzed. We have for the fourth order scheme

$$
C_{4N}(0) = \begin{bmatrix}
\frac{5}{17} & \frac{11-4\sqrt{3}}{17} & \frac{54}{17} & -\frac{59}{17} \\
-1 & -1 & -1 & 2 \\
\frac{55}{43} & \frac{85+12\sqrt{3}}{43} & \frac{4}{43} & -\frac{59}{43} \\
-\frac{1}{49} & -\frac{37+8\sqrt{3}}{49} & \frac{1}{49} & 0
\end{bmatrix},
$$

and for sixth order scheme

$$
C_{6N}(0) = \begin{bmatrix}
-0.4494 & -0.8104 - 0.0403i & -0.8104 + 0.0403i & 3.8057 & -4.6357 & 1.2795 \\
0.5705 & 1.0025 + 0.0469i & 1.0025 - 0.0469i & -1.0534 & 2.3504 & -1.8675 \\
-4.6151 & -6.7899 - 0.2266i & -6.7899 + 0.2266i & 0.6442 & -4.1376 & 8.1084 \\
2.5660 & 4.5927 + 0.1966i & 4.5927 - 0.1966i & -0.2134 & 1.1591 & -3.5117 \\
-1.2510 & -3.1883 - 0.2835i & -3.1883 + 0.2835i & 0.1791 & -0.9157 & 1.9876 \\
0.1996 & 0.3281 + 0.1029i & 0.3281 - 0.1029i & -0.0400 & 0.1875 & -0.3471
\end{bmatrix}.
$$

The boundary systems for the interface problem are

$$
C_{2I}(0,\tau) = \begin{bmatrix}
-\frac{1}{2} & \frac{r}{2} & -1+2\tau h & 0 & \frac{3}{2}+\frac{3r}{2}-2\tau h & -2r \\
-1 & 0 & 0 & 2 & -1 & 0 \\
1 & 0 & -\frac{1}{4} & -1 & \frac{1}{4} & 0 \\
\frac{1}{2r} & -\frac{1}{2} & \frac{3}{2}+\frac{3}{2r}-\frac{2\tau h}{r} & -\frac{2}{r} & -1+\frac{2\tau h}{r} & 0 \\
0 & -1 & -1 & 0 & 0 & 2 \\
0 & 1 & \frac{1}{4} & 0 & -\frac{1}{4} & -1
\end{bmatrix},
$$

and

$$
C_{4I}(0,\tau) = \begin{bmatrix}
-\frac{23}{17} & \frac{31-36\sqrt{3}}{17} & \frac{28r}{17} & \frac{4r(8\sqrt{3}-5)}{17} & \frac{48h\tau-34}{17} & \frac{13}{17} & \frac{44r-48h\tau+44}{17} & -\frac{72r}{17} \\
-1 & -1 & 0 & 0 & \frac{13}{59} & 2 & -\frac{72}{59} & 0 \\
\frac{55}{43} & \frac{12\sqrt{3}+85}{43} & 0 & 0 & -\frac{32}{43} & -\frac{59}{43} & \frac{36}{43} & 0 \\
-\frac{1}{49} & -\frac{8\sqrt{3}+37}{49} & 0 & 0 & \frac{9}{49} & 0 & -\frac{8}{49} & 0 \\
\frac{28}{17r} & \frac{4(8\sqrt{3}-5)}{17r} & -\frac{23}{17} & \frac{31-36\sqrt{3}}{17} & \frac{44r-48h\tau+44}{17r} & -\frac{72}{17r} & \frac{48h\tau-34r}{17r} & \frac{13}{17} \\
0 & 0 & -1 & -1 & -\frac{72}{59} & 0 & \frac{13}{59} & 2 \\
0 & 0 & \frac{55}{43} & \frac{12\sqrt{3}+85}{43} & \frac{36}{43} & 0 & -\frac{32}{43} & -\frac{59}{43} \\
0 & 0 & -\frac{1}{49} & -\frac{8\sqrt{3}+37}{49} & -\frac{8}{49} & 0 & \frac{9}{49} & 0
\end{bmatrix}.
$$

Appendix 2

Consider the one dimensional wave equation (1) with the homogeneous Neumann boundary condition (3). To begin with, we will derive estimates for the continuous problem of U and U_t in maximum norm in terms of data. Then follows the corresponding derivation for the semi–discrete case.

The standard energy estimate (5) gives

$$
\|U_x\| \leq \sqrt{\|G_x\|^2 + \|\bar{G}\|^2} + t \max_{0 \leq \tau \leq t} \|F(\cdot,\tau)\|. \tag{45}
$$

Together with the estimate (6) and the Sobolev inequality [8, Lemma 8.3.1], we obtain that U is bounded by the data in maximum norm

$$\|U\|_\infty \leq \|U\| + \|U_x\|$$
$$\leq \|G\| + (1+t)\sqrt{\|G_x\|^2 + \|\bar{G}\|^2} + (t+t^2)\max_{0\leq\tau\leq t}\|F(\cdot,\tau)\| \qquad (46)$$

where $\|U(x,t)\|_\infty := \sup\{|U(x,t)| : x \geq 0\}$ for any fixed t.

Next, let $V = U_t$. Then V satisfies

$$V_{tt} = V_{xx} + F_t,$$

with initial condition

$$V(x,0) = \bar{G}(x),$$
$$V_t(x,0) = U_{tt}(x,0) = U_{xx}(x,0) + F(x,0) = G_{xx}(x) + F(x,0),$$

and homogeneous Neumann boundary condition

$$V_x(0,t) = 0.$$

The equation in V is in the same form as the equation in U, but with a different forcing and initial data. In the same way, V is bounded by the data in maximum norm

$$\|V\|_\infty \leq \|\bar{G}\| + (1+t)\sqrt{\|\bar{G}_x\|^2 + \|G_{xx} + F(\cdot,0)\|^2} + (t+t^2)\max_{0\leq\tau\leq t}\|F_\tau(\cdot,\tau)\|. \quad (47)$$

Before analyzing the semi–discrete case, we recall the definition of pointwise stability, Definition 2.2, in [21].

Definition 2.2 in [21]: The approximation, v, is strongly pointwise stable if, for all $h \leq h_0$, the estimate

$$\|v(t)\|_\infty^2 \leq K(t)(\|f\|^2 + \max_{0\leq\tau\leq t}\|F(\tau)\|^2 + \max_{0\leq\tau\leq t}g(\tau)^2) \qquad (48)$$

holds. Here $K(t)$ is a bounded function in any finite time interval and does not depend on the data. ($\|\cdot\|$ denotes some norm.) The approximation is pointwise stable if (48) holds with $g(t) = 0$.

In the above definition, $v(t)$ is the semi–discrete solution, f is the initial data, F is the forcing in the equation and g is the boundary data. We also recall Theorem 2.13 in [21].

Theorem 2.13 in [21]:If v and v_t are pointwise stable discrete solutions to (24'), then with $r = 2p - 2$ the global order of accuracy is $2p$.

In the above theorem, (24') is the SBP–SAT discretization of the wave equation, v is the semi–discrete solution, v_t is the time derivative of v, and r? is the order of boundary truncation error. This theorem says that for the wave equation if the semi–discrete solution and its time derivative are pointwise bounded, then two orders are gained in convergence.

In the following, we derive the pointwise estimates for the semi–discrete solution and its time derivative. The discrete energy estimate (33) gives

$$\|u\|_A \leq \sqrt{\|\bar{g}\|_H^2 + \|g\|_A^2} + t\max_{0\leq z\leq t}\|f(z)\|_H, \qquad (49)$$

where g and $\bar{g}$ are the restrictions of $G(x)$ and $\bar{G}(x)$ to the grid. By the Cauchy–Schwarz inequality, we have

$$2\|u\|_H\frac{d}{dt}\|u\|_H = \frac{d}{dt}\|u\|_H^2 \leq 2\|u\|_H\|u_t\|_H,$$

which leads to

$$\frac{d}{dt}\|u\|_H \le \|u_t\|_H \le \sqrt{\|\bar{g}\|_H^2 + \|g\|_A^2} + t \max_{0 \le z \le t} \|f(z)\|_H.$$

An integration in time yields

$$\|u\|_H \le \|g\|_H + t\sqrt{\|\bar{g}\|_H^2 + \|g\|_A^2} + t^2 \max_{0 \le z \le t} \|f(z)\|_H. \tag{50}$$

In the right hand side of (49) and (50), the H−norm causes no problem because it is equivalent to the standard discrete L$_2$ norm. The term $\|g\|_A^2$ is bounded by the following lemma.

Lemma 3 *Consider the initial data $G(x)$ in* (1) *with homogeneous Neumann boundary condition. Let g be the restriction of $G(x)$ to the grid. Then we have 1) $G_x(0) = 0$; 2)With the operator A in* (7), $\|g\|_A^2$ *is bounded by $G(x)$ and its derivatives up to order $2p + 2$.*

Proof 1) It follows from compatibility between initial and boundary data. We differentiate the initial condition $U(x, 0) = G(x)$ and obtain $U_x(x, 0) = G_x(x)$. It is compatible with the homogeneous Neumann boundary condition $U_x(0, t) = 0$ at $t = 0$, which implies $G_x(0) = 0$.

2) We have

$$\|g\|_A^2 = -g^T E_0 S g - g^T H D g$$
$$\le |G(0)||(Sg)_0| + \|g\|_H \|Dg\|_H, \tag{51}$$

where

$$(Sg)_0 = G_x(0) + T_S = T_S, \quad Dg = g_{xx} + T_D,$$

g_{xx} is the restriction of $G_{xx}(x)$ to the grid, T_S and T_D are the truncation errors of the operators S and D. More precisely, by Taylor expansion

$$T_S = h^{p+1} \sum_{i=1}^{k_1} a_i \frac{\partial^{p+2}}{\partial x^{p+2}} G(b_i),$$

where h is the grid spacing, p is the order of the boundary truncation error, k_1 is the stencil width of the operator S, $b_i \in [0, ih]$, and a_i depends on the precise form of S and the details of Taylor expansions. Clearly, $|T_S|$ is bounded by $\max_{0 \le x \le k_1 h} |\frac{\partial^{p+2}}{\partial x^{p+2}} G(x)|$. As a consequence, $|(Sg)_0| \le |T_S|$ is also bounded by the same data.

In a similar way, by using Taylor expansions, T_D can be expressed as $h^p \sum_{i=1}^{k_2} c_i \frac{\partial^{p+2}}{\partial x^{p+2}} G(d_i)$ and $h^{2p} \sum_{i=1}^{k_3} e_i \frac{\partial^{2p+2}}{\partial x^{2p+2}} G(f_i)$ at boundary and interior points, respectively, where k_2, k_3 depend on the stencil width of D. Therefore, $\|g\|_A^2$ is bounded by $|G(x)|$ and its derivatives up to order $2p + 2$. □

The estimates (49), (50), and Lemma 3 give that both $\|u\|_H$ and $\|u\|_A$ are bounded in terms of data. Then by Lemma 3.2 in [21], an analogue of the discrete Sobolve inequality, the numerical solution u is pointwise stable according to Definition 2.2 in [21].

Next, we show that u_t is also pointwise bounded. Let $v = u_t$. Then v satisfies

$$v_{tt} = -H^{-1}Av + f_t,$$
$$v(0) = \bar{g}, \quad v_t(0) = -H^{-1}Ag + f(0)$$

Similarly as before, we have

$$\|v\|_A \leq \sqrt{\|-H^{-1}Ag + f(0)\|_H^2 + \|\bar{g}\|_A^2} + t \max_{0 \leq z \leq t} \|f_z(z)\|_H, \tag{52}$$

and

$$\|v\|_H \leq \|\bar{g}\|_H + t\sqrt{\|-H^{-1}Ag + f(0)\|_H^2 + \|\bar{g}\|_A^2} + t^2 \max_{0 \leq z \leq t} \|f_z(z)\|_H. \tag{53}$$

The term $\|\bar{g}\|_A^2$ can be bounded by using Lemma 3. For the term $\|-H^{-1}Ag\|_H$, we note

$$-H^{-1}Ag = Dg + H^{-1}E_0Sg.$$

$\|Dg\|_H$ also appears in the term $\|g\|_A$ in Lemma 3 and we use the same bound here. Only the first component of $H^{-1}E_0Sg$ is nonzero and is

$$|(H^{-1}E_0Sg)_0| = Kh^{-1}|(G_x(0) + T_S)| = Kh^{-1}|T_S| \leq Kh^p \sum_{i=1}^{k_1} \left| a_i \frac{\partial^{p+2}}{\partial x^{p+2}} G(b_i) \right|,$$

where K depends on the particular form of S but not h. As a consequence, $\|-H^{-1}Ag\|_H$ is bounded by $G(x)$ and its derivatives up to order $2p + 2$ in maximum norm. We can then again use Lemma 3.2 in [21] to derive a pointwise bound on u_t.

In conclusion, with an SBP–SAT finite difference scheme for the wave equation with Neumann boundary condition, the numerical solutions u and u_t are pointwise bounded according to Definition 2.2 in [21]. However, the determinant condition is not satisfied as shown in Sect. 3, nor is the $p + 2$ optimal gain obtained. In particular, for $p = 1$ the gain is only 1 order.

References

1. Abarbanel, S., Ditkowski, A., Gustafsson, B.: On error bounds of finite difference approximations to partial differential equations—temporal behaviour and rate of convergence. J. Sci. Comput. **15**, 79–116 (2000)
2. Appelö, D., Kreiss, G.: Application of a perfectly matched layer to the nonlinear wave equation. Wave Motion **44**, 531–548 (2007)
3. Carpenter, M.H., Gottlieb, D., Abarbanel, S.: Time-stable boundary conditions for finite-difference schemes solving hyperbolic systems: methodology and application to high-order compact schemes. J. Comput. Phys. **111**, 220–236 (1994)
4. Fernández, D.C.D.R., Hicken, J.E., Zingg, D.W.: Review of summation-by-parts operators with simultaneous approximation terms for the numerical solution of partial differential equations. Comput. Fluids **95**, 171–196 (2014)
5. Grote, M.J., Schneebeli, A., Schötzau, D.: Discontinuous Galerkin finite element method for the wave equation. SIAM J. Numer. Anal. **44**, 2408–2431 (2006)
6. Gustafsson, B.: The convergence rate for difference approximations to mixed initial boundary value problems. Math. Comp. **29**, 396–406 (1975)
7. Gustafsson, B.: High Order Difference Methods for Time Dependent PDE. Springer, Berlin (2008)
8. Gustafsson, B., Kreiss, H.O., Oliger, J.: Time-Dependent Problems and Difference Methods. Wiley, New Jersey (2013)
9. Hagstrom, T., Hagstrom, G.: Grid stabilization of high-order one-sided differencing II: second-order wave equations. J. Comput. Phys. **231**, 7907–7931 (2012)
10. Kreiss, H.O., Oliger, J.: Comparison of accurate methods for the integration of hyperbolic equations. Tellus **24**, 199–215 (1972)
11. Kreiss, H.O., Scherer, G.: Finite element and finite difference methods for hyperbolic partial differential equations, Mathematical aspects of finite elements in partial differential equations. Symp. Proc. **33**, 195–212 (1974)

12. Mattsson, K., Almquist, M.: A solution to the stability issues with block norm summation by parts operators. J. Comput. Phys. **253**, 418–442 (2013)
13. Mattsson, K., Ham, F., Iaccarino, G.: Stable and accurate wave-propagation in discontinuous media. J. Comput. Phys. **227**, 8753–8767 (2008)
14. Mattsson, K., Ham, F., Iaccarino, G.: Stable boundary treatment for the wave equation on second-order form. J. Sci. Comput. **41**, 366–383 (2009)
15. Mattsson, K., Nordström, J.: Summation by parts operators for finite difference approximations of second derivatives. J. Comput. Phys. **199**, 503–540 (2004)
16. Mattsson, K., Nordström, J.: High order finite difference methods for wave propagation in discontinuous media. J. Comput. Phys. **220**, 249–269 (2006)
17. Nissen, A., Kreiss, G., Gerritsen, M.: Stability at nonconforming grid interfaces for a high order discretization of the Schrödinger equation. J. Sci. Comput. **53**, 528–551 (2012)
18. Nissen, A., Kreiss, G., Gerritsen, M.: High order stable finite difference methods for the Schrödinger equation. J. Sci. Comput. **55**, 173–199 (2013)
19. Petersson, N.A., Sjögreen, B.: Stable grid refinement and singular source discretiztion for seismic wave simulations. Comm. Comput. Phys. **8**, 1074–1110 (2010)
20. Svärd, M.: A note on L^∞ bounds and convergence rates of summation-by-parts schemes. BIT. Numer. Math. **54**, 823–830 (2014)
21. Svärd, M., Nordström, J.: On the order of accuracy for difference approximations of initial-boundary value problems. J. Comput. Phys. **218**, 333–352 (2006)
22. Svärd, M., Nordström, J.: Review of summation-by-parts schemes for initial-boundary-value problems. J. Comput. Phys. **268**, 17–38 (2014)
23. Taylor, N.W., Kidder, L.E., Teukolsky, S.A.: Spectral methods for the wave equation in second-order form. Phys. Rev. D **82**, 024037 (2010)
24. Virta, K., Mattsson, K.: Acoustic wave propagation in complicated geometries and heterogeneous media. J. Sci. Comput. **61**, 90–118 (2014)