J Sci Comput (2018) 76:777–811
https://doi.org/10.1007/s10915-017-0639-9

A High Order Compact Time/Space Finite Difference Scheme for the Wave Equation with Variable Speed of Sound

Steven Britt[1] · **Eli Turkel**[1] · **Semyon Tsynkov**[2,3]

Received: 7 April 2017 / Revised: 21 December 2017 / Accepted: 28 December 2017 /
Published online: 17 January 2018

Abstract We consider fourth order accurate compact schemes, in both space and time, for the second order wave equation with a variable speed of sound. We demonstrate that usually this is much more efficient than lower order schemes despite being implicit and only conditionally stable. Fast time marching of the implicit scheme is accomplished by iterative methods such as conjugate gradient and multigrid. For conjugate gradient, an upper bound on the convergence rate of the iterations is obtained by eigenvalue analysis of the scheme. The implicit discretization technique is such that the spatial and temporal convergence orders can be adjusted independently of each other. In special cases, the spatial error dominates the problem, and then an unconditionally stable second order accurate scheme in time with fourth order accuracy in space is more efficient. Computations confirm the design convergence rate for the inhomogeneous, variable wave speed equation and also confirm the pollution effect for these time dependent problems.

Work supported by the US-Israel Binational Science Foundation (BSF) under Grant # 2014048 and by US Army Research Office (ARO) under Grant # W911NF-16-1-0115. S. Britt was supported by the Raymond and Beverly Sackler Post-Doctoral Scholarship at Tel Aviv University and a Fulbright Postdoctoral Scholarship funded by the US-Israel Educational Foundation.

✉ Steven Britt
darrellstevenbritt@gmail.com

Eli Turkel
turkel@post.tau.ac.il
http://math.tau.ac.il/∼turkel/

Semyon Tsynkov
tsynkov@math.ncsu.edu
http://www.math.ncsu.edu/∼stsynkov/

[1] School of Mathematical Sciences, Tel Aviv University, Ramat Aviv, 69978 Tel Aviv, Israel

[2] Department of Mathematics, North Carolina State University, Box 8205, Raleigh, NC 27695, USA

[3] Moscow Institute of Physics and Technology, Dolgoprudny, Russia 141700

Keywords Compact finite differences · Wave equation · Variable coefficients · High order accuracy

1 Introduction

The acoustic wave equation describes the propagation of waves in the atmosphere and can also be used as a simplification for modeling electromagnetic waves or elastic waves in a solid. The methods developed here for the acoustic wave equation may be extended to more complicated hyperbolic systems. In particular, the fourth order scheme in time is applicable to any second order equation in time (i.e., in which second time derivatives appear). If first order time derivatives also appear, then an extension of the method described here is straightforward. In the proposed scheme, the time and space portions are discretized separately. Hence, a fourth order approximation in space depends only on a similar fourth order discretization for the time harmonic version of the equation. The resulting spatial equation in this case is positive definite elliptic rather than non-positive as with the Helmholtz equation.

Even though the formulation is implicit, the overall scheme remains efficient due to the high order of accuracy and the use of rapidly converging iterative methods for time marching. Since the spatial equation at each time step is symmetric positive definite, conjugate gradient (CG) and multigrid (MG) can be used. Performance is improved by using a 2nd order accurate initial guess at each time step from the standard explicit scheme, and this further reduces the number of iterations per time step.

Kreiss and Oliger [28] showed for a simple convective equation that the error increases nonlinearly with time and that the rate depends on the order of the scheme. Bayliss et al. [6] and later Babushka and coworkers [4,18] demonstrated a similar effect for the Helmholtz equation: the error increases nonlinearly with the wave number, but this increase is slower for high order schemes. For example, the test solution $u = U(x, y) \cos \Omega t$ to the wave equation has wavenumber $k = \frac{\Omega}{c}$. The pollution effect dictates that, for a p^{th} order time accurate scheme, the quantity $k^{p+1}h^p$ should remain constant in order for the error to remain constant as the wavenumber k increases. This means that higher order methods can achieve the same error while using fewer grid nodes, making them more efficient in terms of both storage and CPU time than low order methods. This applies to both the temporal [28] and spatial errors [4,6,18]. Our results computationally verify the pollution effect for the wave equation and the increased efficiency of the higher order schemes. Moreover, the results demonstrate improved efficiency of the higher order method despite the need to invert an elliptic equation at every time step. We stress that the only way to achieve a compact high order scheme (i.e., higher than second order) is to use an implicit scheme.

In this work we only consider second order equations rather than a first order approach, e.g. [5,23]. The straightforward way to achieve higher spatial accuracy is to increase the stencil size, such as in [14,37,43]. This method can also be extended to treat boundaries and interfaces with either conforming or nonconforming grids using the summation-by-parts technique [27,34,41,42] (see also the reviews [19,39]). One can extend this to higher order accurate formulae in time [36]. The fourth order time scheme we develop is based directly on the PDE and results in a compact spatial scheme that involves only three time levels (see, e.g., [11,13,20,43]). As a result, no special difficulties arise at the initial time step beyond that of the usual wave equation.

Some early papers to consider high order methods, in space or time, for the acoustic wave equation were [3,13,16]. Alford considered a fourth order method in space while Ciment considered a compact fourth order in space and time scheme. Implicit portions were inserted in various operators and solved by ADI. The scheme of Dablain involved a very large stencil to achieve very high order in space. The scheme of Henshaw [25] solves the wave equation explicitly with 4th order in both time and space using only 3 time levels; however, the spatial scheme does not use a compact stencil and hence additional numerical boundary conditions are required to maintain high order accuracy. Several authors, e.g. [2,11,12,15,20,24,26], consider the ODE $\frac{\partial^2 u}{\partial t^2} + Au = 0$. They develop a scheme which is fourth order accurate in time by considering the θ scheme and replacing higher order time derivatives by powers of A or equivalently replace a fourth time derivative by fourth space derivatives. This has the disadvantage of either requiring large matrix vector multiplications and/or results in a non-compact scheme in space. The approach of Chabassier et al. is based on finding an optimal scheme for a given problem, whereas our focus is on simplicity and, in addition, utility for the method of difference potentials [9,35]. An ADI approach is considered in [1,13,32,33] while [17] includes a discussion of splitting errors.

Our scheme is compact in both space and time and so has no special difficulties with initial conditions or at boundaries, even for complicated boundary conditions (see, e.g., [9]). Compact schemes also result in matrices with lower bandwidths, which can be inverted more efficiently. Compact spatial schemes of 4th and 6th order for the Helmholtz equation have been developed [8,38,40]. Since the bandwidth is small and the spatial equation is positive definite, efficient iterative schemes such as multigrid or conjugate gradient are applicable. In the present work, we solve the resultant (modified) Helmholtz equation by both direct and iterative methods to compare their efficiency.

Furthermore, we have developed the scheme for the inhomogeneous and variable wavespeed equation, as in our previous work on the Helmholtz equation [8]. The speed of sound in the wave equation is a function of the media properties. In many applications the speed is not constant with respect to the density, pressure, or temperature of the material. We will assume that the material properties do not vary in time but only in space, i.e., that the speed of sound is independent of time. In some cases the Laplacian term can also have variable coefficients. This occurs primarily in electromagnetic problems. The method described herein also extends to this case as long as the Laplacian is in divergence form (i.e., div(grad u)). For general shaped domains, we suggest using difference potentials [9,35] rather than a grid transformation, and thus we keep the original Cartesian or polar form of the Laplacian.

The structure of the paper is as follows. In Sect. 2, we derive a parameterized scheme in time for the inhomogeneous, variable coefficient wave equation. The scheme is compact in time. We also provide a stability analysis which shows unconditional stability in the 2nd order case and provides a stability condition in the 4th order case. In Sect. 3, we specify a 4th order compact FD solver in space and address Dirichlet and Neumann boundary conditions. Section 4 discusses the use of CG and MG as solvers for time marching of the implicit scheme, and an eigenvalue analysis for general compact FD schemes for the modified Helmholtz equation is presented. This result is used to prove fast convergence of CG for the steady-state equation resulting from our implicit discretization of the wave equation using the FD scheme of Sect. 3. We present numerical results in Sect. 5 confirming the design rate of the scheme, demonstrating the pollution effect for solutions of increasing frequency, and comparing direct and iterative solvers for time marching of the implicit scheme. Concluding remarks and future directions are given in Sect. 6. Additional numerical investigations of iterative solvers for the modified Helmholtz equation are provided in “Appendix A”.

2 High Order Discretization of the Wave Equation in Time

2.1 Derivation of the Time-Marching Scheme

We consider the variable coefficient wave equation in 2D:

$$u_{tt} = c^2 \Delta u + F, \tag{1}$$

where $c = c(x, y)$ and $F = F(x, y, t)$ are given. The Laplacian term Δu can be extended to $\frac{\partial}{\partial x}\left(A(x, y)\frac{\partial u}{\partial x}\right) + \frac{\partial}{\partial y}\left(B(x, y)\frac{\partial u}{\partial y}\right)$ in a straightforward manner [8] . We wish to solve the interior initial-boundary value problem on a square domain of side length $2s$ centered at the origin. Consider the initial boundary value problem:

$$\begin{aligned} \ell_{x,y=\pm s}(u) &= \phi(x, y, t), \\ u(x, y, t = 0) &= u^0(x, y), \\ u_t(x, y, t = 0) &= \psi(x, y), \end{aligned} \tag{2}$$

where ℓ is either the identity (Dirichlet) or normal derivative (Neumann) operator along each edge.

We consider a semi-discrete approach by first discretizing the equation in time and later in space. Denote by h_t the uniform time step so that $t_n = nh_t$, and let δ_t^2 be the second central difference operator, $\delta_t^2 u^n = u^{n+1} - 2u^n + u^{n-1}$. The time derivative of (1) is approximated by

$$u_{tt}^n = \frac{1}{h_t^2}\delta_t^2 u^n - \frac{h_t^2}{12}u_{tttt}^n + \mathcal{O}\left(h_t^4\right).$$

Differentiating the wave equation (1), we replace the fourth time derivative as follows:

$$u_{tt}^n = \frac{1}{h_t^2}\delta_t^2 u^n - \frac{h_t^2}{12}\left(c^2 \Delta u_{tt}^n + F_{tt}^n\right) + \mathcal{O}\left(h_t^4\right). \tag{3}$$

We then replace Δu_{tt} and F_{tt} by 2nd order central differences in time, resulting in a fourth order approximation in time:

$$u_{tt}^n = \frac{1}{h_t^2}\delta_t^2 u^n - \frac{c^2}{12}\delta_t^2 \Delta u^n - \frac{1}{12}\delta_t^2 F^n + \mathcal{O}\left(h_t^4\right). \tag{4}$$

We consider the following approximation to Eq. (1) centered at the time t_n using a free parameter θ (sometimes referred to as the θ-scheme, see, e.g., [11,31]):

$$\frac{1}{h_t^2}\delta_t^2 u^n = c^2 \Delta(u^n) + \theta c^2 \delta_t^2 \Delta(u^n) + F^n + \theta \delta_t^2 F^n. \tag{5}$$

Observe from (4) that the choice $\theta = \frac{1}{12}$ results in a 4th order approximation of the wave equation in time. Rearranging (5) to gather the upper time level terms yields

$$\begin{aligned} \theta \Delta(u^{n+1}) - \frac{u^{n+1}}{c^2 h_t^2} = {} & 2\left(\theta \Delta(u^n) - \frac{u^n}{c^2 h_t^2}\right) - \left(\theta \Delta(u^{n-1}) - \frac{u^{n-1}}{c^2 h_t^2}\right) - \Delta(u^n) \\ & - \frac{1}{c^2}F^n - \frac{\theta}{c^2}\delta_t^2 F^n. \end{aligned} \tag{6}$$

From (6) we observe that, for any space discretization, the scheme (5) is

- explicit and second order in time if $\theta = 0$,

- implicit and equivalent to the second order Crank–Nicolson scheme if $\theta = \frac{1}{2}$,
- implicit and fourth order in time if $\theta = \frac{1}{12}$. Therefore the scheme must be implicit in order to have a fourth order accurate scheme in time which uses only three time levels.

We now consider the implementation of (5). The most straightforward idea is to solve (6) for u^{n+1}. We define $f^n := \Delta(u^n) - \frac{1}{\theta c^2 h_t^2} u^n$ and $\tilde{F}^n := -\frac{1}{\theta c^2} F^n - \frac{1}{c^2}\delta_t^2 F^n$. We then rewrite (6) as

$$\Delta(u^{n+1}) - \frac{1}{\theta c^2 h_t^2} u^{n+1} = 2 f^n - f^{n-1} - \frac{1}{\theta}\Delta(u^n) + \tilde{F}^n. \tag{7}$$

By the definition of f^n, we substitute $\Delta(u^n) = f^n + \frac{1}{\theta c^2 h_t^2} u^n$ into (7) and gather like terms to obtain

$$\Delta(u^{n+1}) - \frac{1}{\theta c^2 h_t^2} u^{n+1} = \left(2 - \frac{1}{\theta}\right) f^n - f^{n-1} - \frac{1}{\theta^2 c^2 h_t^2} u^n + \tilde{F}^n.$$

This suggests the recurrence

$$f^{n+1} = \left(2 - \frac{1}{\theta}\right) f^n - f^{n-1} - \frac{1}{\theta^2 c^2 h_t^2} u^n + \tilde{F}^n \tag{8}$$

for the right-hand side. We now solve the following spatial equation at each time step:

$$\Delta(u^{n+1}) - \frac{1}{\theta c^2 h_t^2} u^{n+1} = f^{n+1}. \tag{9}$$

Denote the spatial finite difference operator on the left-hand side of (9) by $\frac{1}{h_x^2} L_h$. Rather than approximating the terms such as $\Delta u^n - \frac{1}{\theta c^2 h_t^2} u^n$ on the right-hand side of (6) at each step by the matrix multiplication $f^n = \frac{1}{h_x^2} L_h u^n$ (which also requires storing two previous time levels of the solution), we only need to store two previous time levels of f^n and one previous time level of the solution and then perform the additions and subtractions of (8). Thus, the recurrence relation (8) efficiently computes the right-hand side at each step. Furthermore, computing the terms $f^n = \frac{1}{h_x^2} L_h u^n$ by matrix multiplication would require further consideration at the boundary since the 4th order compact FD schemes in space will require that an approximation of Δ be applied to the right-hand side. By computing f^{n+1} from the recurrence (8), we avoid approximating Δ^2, which is difficult to achieve compactly at the boundary nodes for boundary conditions other than Dirichlet. Even though f^n contains Δu^n in its definition, the use of the recursion formula (8) for the right-hand side circumvents this issue since f^0 and f^1 are known on the entire grid from the initial condition, so that f^{n+1} is known on the entire grid at each time step.

The definition of f^n in the spatial equation (9) solved at each new time step implicitly contains Δu^n. Therefore, we consider an alternative form of the scheme (6) which has a simpler expression for the right-hand side. From (5), we have

$$\frac{1}{h_t^2 c^2}\delta_t^2 u^n = \Delta(u^n + \theta\delta_t^2 u^n) - \theta\tilde{F}^n. \tag{10}$$

We introduce a new variable

$$v^{n+1} = u^n + \theta\delta_t^2 u^n, \quad \theta \neq 0. \tag{11}$$

which implies $\delta_t^2 u^n = \frac{v^{n+1}-u^n}{\theta}$. Substituting (11) into (10) yields the spatial equation

$$\Delta v^{n+1} - \frac{1}{\theta c^2 h_t^2} v^{n+1} = -\frac{1}{\theta c^2 h_t^2} u^n + \theta \tilde{F}^n := f_v^{n+1}, \tag{12}$$

which is solved for v^{n+1} at each time step. The solution of the wave equation at the upper time level is then given by solving for u^{n+1} in (11):

$$u^{n+1} = 2u^n - u^{n-1} + \frac{v^{n+1} - u^n}{\theta}.$$

The right-hand side f_v^{n+1} of the spatial equation (12) at each step involves the solution itself at the prior time step but not its derivatives. The two versions are linearly equivalent, and computations with the two variants yielded almost identical results. There might be advantages to each scheme when considering difference potentials [35] for the wave equation to handle general geometries.

2.2 Stability Analysis

The stability analysis presented below is a generalization of the proof by Z. Li [30]. He considered a modified wave equation in one dimension using a second order spatial scheme. We consider the original wave equation with a variable wave speed in multiple dimensions. The following analysis is general provided that the spatial scheme is self-adjoint negative definite. For example, the wave equation can be extended to include a self-adjoint Laplacian with variable coefficients. Hence, we have added several improvements to the proof of Li.

We consider the stability analysis for a generalized wave equation

$$\frac{1}{c^2}\frac{\partial^2 u}{\partial t^2} = Lu. \tag{13}$$

Let $\frac{1}{h_x^2} L_h$ be the numerical approximation to L (in Sect. 2.1, $L \equiv \Delta$), where h_x is the uniform spatial step size and the operator L_h has the following properties:

1. L_h is negative definite, i.e., there exists a real inner product so that $(u, v) = (v, u)$ and $(-L_h u, u) \geq 0$. Further, we require that $0 < L_{\text{lower}} \|u\|^2 \leq (-L_h u, u) \leq L_{\text{upper}} \|u\|^2$.
2. L_h is self-adjoint, so that $(L_h u, v) = (u, L_h v)$. Hence there exists a symmetric or an anti-symmetric matrix M which satisfies $M^2 = L_h$. Thus $(L_h u, u) = (Mu, Mu)$, showing that $(-L_h u, u)$ is a norm. Note that in the case of a one-dimensional PDE L_h is a second derivative while M is a first derivative which is anti-symmetric.

The θ scheme is given by

$$\begin{aligned} \frac{u^{m+1} - 2u^m + u^{m-1}}{c^2 h_t^2} &= \frac{1}{h_x^2} L_h \left(\theta u^{m+1} + (1-2\theta) u^m + \theta u^{m-1} \right) \\ &= \frac{1}{h_x^2} L_h \left(\theta (u^{m+1} - u^m) - \theta (u^m - u^{m-1}) + u^m \right). \end{aligned} \tag{14}$$

For convenience, define $v^{m+1} = u^{m+1} - u^m$. Observe that $v^{m+1} - v^m = u^{m+1} - 2u^m + u^{m-1}$. Multiplying both sides of (14) by h_x^2 and taking the inner product with $u^{m+1} - u^{m-1} = v^{m+1} + v^m$, we obtain

$$\frac{h_x^2}{c^2 h_t^2}(v^{m+1} - v^m, v^{m+1} + v^m) = (L_h u^m + \theta L_h (v^{m+1} - v^m), v^{m+1} + v^m).$$

Define $\lambda(\overrightarrow{x}) = \frac{c(\overrightarrow{x})h_t}{h_x}$, which we shall refer to as the Courant–Friedrichs–Lewy (CFL) number. Simplifying, we obtain

$$\frac{1}{\lambda^2}(\|v^{m+1}\|^2 - \|v^m\|^2) = (L_h u^m, v^{m+1} + v^m) + \theta(L_h(v^{m+1} - v^m), v^{m+1} + v^m). \tag{15}$$

Since L_h is self-adjoint, the cross term in the last term of (15) is zero, so that (15) reduces to

$$\frac{1}{\lambda^2}(\|v^{m+1}\|^2 - \|v^m\|^2) = (L_h u^m, v^{m+1} + v^m) + \theta(L(v^{m+1}, v^{m+1}) - (L_h v^m, v^m)),$$

and rearranging yields

$$\frac{1}{\lambda^2}\|v^{m+1}\|^2 - \theta(Lv^{m+1}, v^{m+1}) = \frac{1}{\lambda^2}\|v^m\|^2 - \theta(L_h v^m, v^m) + (L_h u^m, v^{m+1} + v^m). \tag{16}$$

We rewrite the last term in (16) using the following identity $v^{m+1} + v^m = u^{m+1} - u^{m-1}$ and

$$\begin{aligned}&(L_h u^m, u^{m+1}) - (L_h u^m, u^{m-1})\\ &\quad = 1/4\left[(L_h v^m, v^m) - (L_h v^{m+1}, v^{m+1})\right.\\ &\qquad \left. -(L_h(u^m + u^{m-1}), u^m + u^{m-1}) + (L_h(u^m + u^{m+1}), u^m + u^{m+1})\right].\end{aligned} \tag{17}$$

This is verified by expanding the right hand side. Combining (16) with (17) we get

$$\begin{aligned}&\frac{1}{\lambda^2}\|v^{m+1}\|^2 + (1/4 - \theta)(L_h v^{m+1}, v^{m+1}) - 1/4(L_h(u^{m+1} + u^m), u^{m+1} + u^m)\\ &\quad = \frac{1}{\lambda^2}\|v^m\|^2 + (1/4 - \theta)(L_h v^m, v^m) - 1/4(L_h(u^m + u^{m-1}), u^m + u^{m-1}).\end{aligned} \tag{18}$$

Define

$$S_m = \frac{1}{\lambda^2}\|v^m\|^2 + (1/4 - \theta)(L_h v^m, v^m) - 1/4(L_h(u^m + u^{m-1}), u^m + u^{m-1}).$$

Then (18) is equivalent to

$$S_{m+1} = S_m,$$

meaning that the quantity S_m is constant through the calculation.

We now check the stability of the scheme in two cases:

1. $\theta \geq 1/4$
   Since L_h is negative definite, every term in S_m is positive. Hence, defining $\|u\|_E^2 = S_m$, we have energy conservation of $\|u\|_E$, and the scheme is unconditionally stable.
2. $0 \leq \theta < 1/4$
   We now use the assumption that $0 < L_{\text{lower}}\|u\|^2 \leq (-L_h u, u) \leq L_{\text{upper}}\|u\|^2$. We then get

$$\begin{aligned}&\left(\frac{1}{\lambda^2} - (1/4 - \theta)L_{\text{upper}}\right)\|v^m\|^2 + \frac{L_{\text{lower}}}{4}\|u^m + u^{m-1}\|^2 \leq S_m \leq\\ &\left(\frac{1}{\lambda^2} + (1/4 - \theta)L_{\text{lower}}\right)\|v^m\|^2 + \frac{L_{\text{upper}}}{4}\|u^m + u^{m-1}\|^2.\end{aligned}$$

Therefore, S_m is equivalent to the norm $\|u^m - u^{m-1}\|^2 + \|u^m + u^{m-1}\|^2$, which is equivalent to $\|u^m\|^2 + \|u^{m-1}\|^2$ if and only if $\frac{1}{\lambda^2} - (1/4 - \theta)L_{\text{upper}} \geq 0$. Thus, when $\theta < 1/4$ the scheme is stable provided that

$$\max_{\overrightarrow{x}} \lambda(\overrightarrow{x})^2 \leq \frac{1}{(1/4 - \theta)L_{\text{upper}}}. \tag{19}$$

For $\theta = 1/12$, this yields $\lambda^2 \leq \frac{6}{L_{\text{upper}}}$ which is 50% larger than for the explicit scheme $\theta = 0$. For the five point second order central difference stencil, $L_{\text{upper}} = 8$ and the stability condition is $\lambda \leq \sqrt{0.75}$. For a fourth order space approximation to the Helmholtz equation, the value of L_{upper} depends on the details of the scheme (to be discussed in the next section).

3 High Order Spatial Discretization

The two schemes (9) and (12) share the form of a modified Helmholtz equation

$$\Delta w - k^2 w = g, \tag{20}$$

with $k^2 = \frac{1}{\theta c^2 h_t^2}$. Since, $k^2 > 0$, Eq. (20) differs substantially from the conventional Helmholtz equation. The quantity k is not a physical wavenumber. It is rather a parameter of the discrete approximation that depends primarily on the time step h_t. We note that the wave equation (1) can be reduced to the conventional Helmholtz equation by a Fourier transform in time. In doing so, the resulting Helmholtz equation is parameterized by the dual Fourier variable ω, which is called the frequency, or, equivalently, by the wavenumber $k = \omega/c$. From the standpoint of solving Eq. (1) numerically though, it may only be practical to use the aforementioned reduction if it is known ahead of time that the original solution of Eq. (1) contains no more than a small number of discrete frequencies. Otherwise, in the case of a continuous spectrum or a broadband solution, it is efficient to integrate equation (1) directly in the time domain rather than replace it with a collection of Helmholtz equations in the frequency domain.

We solve the two-dimensional equation (20) by the compact finite difference scheme given in [38] on a Cartesian grid which is equally spaced in both directions with step size $h_x = h_y$. The scheme developed in [8] is for an equation with a variable coefficient Laplacian, and we note that for the equation with constant coefficients in the Laplacian term the schemes of [38] and [8] coincide.

Let u_s and u_c denote, respectively, the sums of the four side and corner points:

$$\begin{aligned} u_s &= u_{m+1,n} + u_{m-1,n} + u_{m,n+1} + u_{m,n-1} \\ u_c &= u_{m+1,n+1} + u_{m+1,n-1} + u_{m-1,n+1} + u_{m-1,n-1}. \end{aligned}$$

Let g_s and g_c denote the corresponding sums for the inhomogeneous term of (20). Then any compact scheme for the Helmholtz equation (20) may be written in the form

$$A_0 u_{m,n} + A_s u_s + A_c u_c = B_0 g_{m,n} + B_s g_s + B_c g_c, \tag{21}$$

where A_0, A_s, and A_c represent, respectively, the coefficients of the center, side, and corner nodes of the compact stencil acting on the solution u. Then the 4th order scheme described in [38] is given by

$$A_0 = -\frac{10}{3} + \frac{2}{3}k^2 h_x^2, \quad A_s = \frac{2}{3} + \frac{k^2 h_x^2}{12}, \quad A_c = \frac{1}{6}. \tag{22}$$

We define the stencil operating on the right-hand side of (20) by B_i corresponding to the A_i. Then

$$B_0 = \frac{2h_x^2}{3}, \quad B_s = \frac{h_x^2}{12}, \quad B_c = 0. \tag{23}$$

Note that $\frac{1}{h_x^2} L_h$ in the stability analysis of Sect. 2.2 is the approximation to the Laplace operator, which corresponds to the case $k = 0$ but still includes the right-hand side. Formula (22) can be extended to the more general Laplacian of the form $\operatorname{div}(A(x, y) \operatorname{grad} u)$ using the finite difference scheme in [8]. It can also be extended to polar coordinates with fourth order accuracy, see [7].

In Sect. 2.2 we derived the stability condition (19) in terms of the bound L_{upper}. We now evaluate the stability constant for the case of constant coefficients. For an implicit scheme, $L_h = P^{-1}Q$. Because we use central differences, both P and Q are symmetric. For constant coefficients, P and Q commute. Since P^{-1} is positive definite (see (22)) we can introduce a new norm $(u, v)_P = (P^{-1}u, v)$. Since Q is negative definite and self adjoint (see (23)), the stability arguments of Sect. 2.2 apply to the implicit scheme. For the fourth order implicit scheme it is easier to evaluate L_{upper} in the Fourier domain. For a periodic boundary problem, we apply the Fourier transform to these finite difference formulae. For a Dirichlet boundary condition, we use a sine transform. By Parseval's theorem, the stability condition of the Fourier transform in L_2 implies the same stability condition of the original finite difference formula. Recall from (13) and the ensuing stability analysis that $\frac{1}{h_x^2} L_h$ is the discrete approximation to the Laplacian (i.e., $k = 0$), and we denote the Fourier transform of $\frac{1}{h_x^2} L_h$ as $\mathbf{L}$, which is a scalar function of ξ and η. The negative of the Fourier transform of the left-hand side of Eq. (21) with coefficients (22–23) and $k = 0$ is then given by

$$-\mathbf{L} = \frac{\frac{10}{3} - \frac{4}{3}(\cos(\xi) + \cos(\eta)) - \frac{2}{3}\cos(\xi)\cos(\eta)}{\frac{2}{3} + \frac{1}{6}(\cos(\xi) + \cos(\eta))}. \tag{24}$$

A MATLAB search verifies that the maximum occurs at $(\xi, \eta) = (\pi, \pi)$. In physical space this represents the mode which oscillates between $+1$ and -1, $u = (-1)^{i+j}$. Therefore, (24) is bounded by $|\mathbf{L}| \leq \frac{\frac{10}{3} + \frac{8}{3} - \frac{2}{3}}{\frac{1}{3}} = 16$. Substituting this into (19), we arrive at the stability condition $\lambda^2 \leq \frac{3}{8}$, from which we may compute the CFL number, $\lambda = \frac{ch_t}{h_x}$. For example, choosing $c = 0.9$ gives the stability condition $\frac{h_t}{h_x} \leq \frac{\sqrt{.375}}{.9} \approx .67$.

Remark – $c(\overrightarrow{x})$ appears in the term of $\delta_t^2 u$. There are no assumptions on c except that it is not a function of time, which is standard.

- The only assumption on L_h is that it is symmetric negative definite. Hence, L_h can contain variable coefficients and also boundary conditions that satisfy this condition.
- The Fourier space was used only to calculate the upper bound L_{upper}. The assumption is that this upper bound is not sensitive to variable coefficients or boundary conditions. Hence, we use a periodic domain with constant coefficients. In practice, the stability bound is always used with a safety factor for such reasons.

3.1 Boundary Conditions

In the present work we consider both Dirichlet and Neumann boundary conditions. At each time step we solve a modified Helmholtz equation. For Dirichlet BCs, it is straightforward to impose the boundary condition within the Helmholtz solver.

The following Neumann BC procedure for compact schemes is introduced in [38]. We consider a Neumann condition on the right edge of the square domain, $\frac{\partial u}{\partial x}\big|_{x=x_N} = \phi(y,t)$. For simplicity, we assume that the wave equation (1) is homogeneous ($F \equiv 0$) with a constant wavenumber c. This assumption implies no loss of generality as one can always consider that the inhomogeneity and/or variable speed of sound are present only "well inside" the domain. On the other hand, allowing, for example, the wave equation (1) to have a non-zero right-hand side F all the way up to the boundary can lead only to insignificant modifications of the approach described below in this section.

We begin with a second order approximation of the boundary condition at the right edge using grid nodes along the line of ghost points $x_{N+1} = x_N + h_x$ which lie outside of the computational grid and will be eliminated from the final expressions:

$$\frac{u^n_{N+1,j} - u^n_{N-1,j}}{2h_x} = \frac{\partial u^n_{N,j}}{\partial x} + \frac{h_x^2}{6}\frac{\partial^3 u^n_{N,j}}{\partial x^3} + \mathcal{O}\left(h_x^4\right). \tag{25}$$

Solving for the ghost point $u^n_{N+1,j}$ yields

$$u^n_{N+1,j} = u^n_{N-1,j} + \underbrace{2h_x\frac{\partial u^n_{N,j}}{\partial x} + \frac{h_x^3}{3}\frac{\partial^3 u^n_{N,j}}{\partial x^3}}_{:=\zeta^n_{N,j}} + \mathcal{O}\left(h_x^5\right). \tag{26}$$

At each time step, a linear system is obtained from the implicit time discretization (9) and the compact spatial scheme (21), with the right-hand side f^{n+1} for each time step defined recursively by (8). The Neumann BC is enforced by substituting (26) into the discrete equation (21) at the right boundary $x = x_N$, and we divide by 2 to maintain symmetry of the system:

$$\begin{aligned}&\frac{1}{2}A_0 u^n_{N,j} + \frac{1}{2}A_s\left(2u^n_{N-1,j} + u^n_{N,j+1} + u^n_{N,j-1}\right) + A_c\left(u^n_{N-1,j+1} + u^n_{N-1,j-1}\right)\\ &\quad = \frac{1}{2}\left(B_0 f^{n+1}_{N,j} + B_s f^{n+1}_s\right) - \frac{1}{2}\left(A_s\zeta^n_{N,j} + A_c\left(\zeta^n_{N,j+1} + \zeta^n_{N,j-1}\right)\right).\end{aligned}$$

Note that the term f^{n+1}_s contains the term $f^{n+1}_{N+1,j}$ with values on the ghost line. We substitute (26) into (9) to obtain

$$\begin{aligned}f^{n+1}_{N+1,j} &= \Delta u^{n+1}_{N+1,j} - k^2 u^{n+1}_{N+1,j} = \Delta u^{n+1}_{N-1,j} - k^2 u^{n+1}_{N-1,j} + \Delta\zeta^{n+1}_{N,j} - k^2\zeta^{n+1}_{N,j}\\ &= f^{n+1}_{N-1,j} + \Delta\zeta^{n+1}_{N,j} - k^2\zeta^{n+1}_{N,j}.\end{aligned} \tag{27}$$

We now show how to evaluate the terms $\zeta^n_{N,j}$ and $\Delta\zeta^n_{N,j}$ at the right-hand boundary using equation-based substitution of the wave equation along with the Neumann boundary condition. For $\zeta^n_{N,j}$, we must show how to compute the first and third derivatives of $u^n_{N,j}$ with respect to x. The first derivative is immediate from the Neumann boundary condition, $\frac{\partial u^n_{N,j}}{\partial x} = \phi^n_j$. For the third derivative, we differentiate the wave equation (1) and solve for u_{xxx} to obtain

$$u_{xxx} = \frac{1}{c^2}u_{xtt} - u_{xyy}, \tag{28}$$

which at time t_n can be evaluated at the grid node (x_N, y_j) by substituting the y and t derivatives of the boundary condition:

$$\frac{\partial^3 u^n_{N,j}}{\partial x^3} = \frac{1}{c^2}\frac{\partial^2\phi^n_j}{\partial t^2} - \frac{\partial^2\phi^n_j}{\partial y^2}. \tag{29}$$

The terms ϕ_{tt} and ϕ_{yy} can be computed either analytically or by 2nd order central differences so that (25) is 4th order accurate. If, on the other hand, $F \neq 0$ in Eq. (1), then instead of (28) we will have $u_{xxx} = \frac{1}{c^2}u_{xtt} - u_{xyy} - F_x$, where F is known, and the corresponding expressions below, in particular, formulae (28′), will change accordingly.

To compute $\Delta\zeta^n_{N,j}$, we have

$$\Delta\zeta = 2h_x\left(\frac{\partial^3 u}{\partial x^3} + \frac{\partial^3 u}{\partial x \partial y^2}\right) + \frac{h_x^3}{3}\left(\frac{\partial^5 u}{\partial x^5} + \frac{\partial^5 u}{\partial x^3 \partial y^2}\right). \tag{30}$$

At the right boundary, $\frac{\partial^3 u}{\partial x^3}$ is evaluated by (29) and as before $\frac{\partial^3 u^n_{N,j}}{\partial x \partial y^2} = \frac{\partial^2 \phi^n_j}{\partial y^2}$. To obtain expressions for the higher order derivatives of (30) at the right boundary that contain only known quantities, we differentiate (28) twice with respect to x, y, and t to obtain the expressions

$$\begin{aligned} u_{xxxyy} &= \frac{1}{c^2}u_{xyytt} - u_{xyyyy}, \\ u_{xxxtt} &= \frac{1}{c^2}u_{xtttt} - u_{xyytt}, \\ u_{xxxxx} &= \frac{1}{c^2}u_{xxxtt} - u_{xxxyy} = \frac{1}{c^4}u_{xtttt} - \frac{2}{c^2}u_{xyytt} + u_{xyyyy}, \end{aligned} \tag{28′}$$

where the final expression for the fifth x derivative is obtained by substitution of the previous two formulas. At the right boundary of the square, we thus compute the h_x^3 terms of (30) by

$$\begin{aligned} \frac{\partial^5 u^n_{N,j}}{\partial x^3 y^2} &= \frac{1}{c^2}\frac{\partial^4 \phi^n_j}{\partial y^2 \partial t^2} - \frac{\partial^4 \phi^n_j}{\partial y^4}, \\ \frac{\partial^5 u^n_{N,j}}{\partial x^5} &= \frac{1}{c^4}\frac{\partial^4 \phi^n_j}{\partial t^4} - \frac{2}{c^2}\frac{\partial^4 \phi^n_j}{\partial y^2 \partial t^2} + \frac{\partial^4 \phi^n_j}{\partial y^4}, \end{aligned}$$

where all y and t derivatives of the function ϕ along the right boundary $x = s$ of the square are assumed to either be known analytically or with sufficient accuracy by finite differences.

In summary, the Eq. (26) enforce the Neumann boundary condition at the right boundary $x = x_N$ with fourth order accuracy, with the right-hand side computed using (27). The resulting system is symmetric positive definite. The same procedure can be used to treat a Neumann BC on other edges of the square. Examples of more general boundary conditions can be found in [8,9,35,40].

4 Iterative Solvers for the Modified Helmholtz Equation

We now present an analysis of conjugate gradient and multigrid solvers for the modified Helmholtz equation with specific emphasis on the case which arises in the implicit discretization of the wave equation. We begin with an eigenvalue analysis of general compact schemes for the modified Helmholtz equation. Using this analysis, we compute the condition number of the FD matrix for the scheme (22), which provides an upper bound on the convergence rate of the conjugate gradient method for the wave equation, see Sect. 4.1. For multigrid methods using a damped Jacobi smoother, the same eigenvalue analysis facilitates the computation of the optimal parameter for damping the high-frequency modes for the wave equation, which we derive in Sect. 4.2. Section 4.3 discusses choices of initial guess for iterative solvers for the wave equation using the previously computed numerical solutions from prior time steps.

Consider the following generalization of the Helmholtz equation (20) in negative definite form,

$$-\Delta w - Kw = -g, \tag{31}$$

which is the usual Helmholtz equation if $K = k^2 > 0$ or the modified Helmholtz equation if $K = -k^2 < 0$. For the purpose of analysis, we consider (31) in 2D on a square domain of side length π, $S = \{0 \le x, y \le \pi\}$, with homogeneous Dirichlet boundary conditions on the edges in both the x and y directions. Let A be a general Cartesian FD scheme on the compact 3×3 stencil with uniform grid spacing $h_x = h_y = \frac{\pi}{M}$, so that there are M nodes in each direction. In this paper, we are primarily interested in the special case where the modified Helmholtz equation results from the discretization of the wave equation by the θ method; however, the following analysis can be applied to more general cases as well.

The Fourier eigenfunctions on the interior satisfying the Dirichlet BC can be expressed as the product of sines in each direction:

$$\phi_{m,n}^{(k_1,k_2)} = \sin(mk_1h_x)\sin(nk_2h_x), \quad 1 \le m, n, k_1, k_2 \le M-1. \tag{32}$$

We seek the eigenvalues of A, $\lambda_{k_1,k_2}(A)$, using the Fourier modes (32) by solving $A\phi^{(k_1,k_2)} = \lambda_{k_1,k_2}(A)\phi^{(k_1,k_2)}$. We will make use of the trigonometric identity

$$\sin(\alpha+\beta) \pm \sin(\alpha-\beta) = 2\left(1 - 2\sin^2\beta/2\right)\sin\alpha \tag{33}$$

to simplify sums of the Fourier modes. First, observe that

$$\begin{aligned}\phi_{m+1,n}^{(k_1,k_2)} + \phi_{m-1,n}^{(k_1,k_2)} &= \sin((m+1)k_1h_x)\sin(nk_2h_x) + \sin((m-1)k_1h_x)\sin(nk_2h_x)\\ &= 2\left(1 - 2\sin^2\frac{k_1h_x}{2}\right)\phi^{(k_1,k_2)}.\end{aligned} \tag{34}$$

An analogous argument applies to the remaining terms of $\phi_s^{(k_1,k_2)}$, which added together with (34) yields

$$\phi_s^{(k_1,k_2)} = 4\left[1 - \left(\sin^2\frac{k_1h_x}{2} + \sin^2\frac{k_2h_x}{2}\right)\right]\phi^{(k_1,k_2)}.$$

Using the same identity, we simplify the sum of the corner points multiplied by A_c:

$$\phi_{m+1,n\pm1}^{(k_1,k_2)} + \phi_{m-1,n\pm1}^{(k_1,k_2)} = \left(2\left[1 - 2\sin^2\frac{k_1h_x}{2}\right]\sin(mk_1h_x)\right)\sin((n\pm1)k_2h_x). \tag{35}$$

The identity (33) can be applied a second time to the sum of the corner points in (35), which yields

$$\phi_c^{(k_1,k_2)} = 4\left[1 - 2\left(\sin^2\frac{k_1h_x}{2} + \sin^2\frac{k_2h_x}{2}\right) + 4\sin^2\frac{k_1h_x}{2}\sin^2\frac{k_2h_x}{2}\right]\phi^{(k_1,k_2)}. \tag{36}$$

Combining (34) and (36), it follows that the eigenvalues satisfying $A\phi^{(k_1,k_2)} = \lambda_{k_1,k_2}(A)$ $\phi^{(k_1,k_2)}$ are given by:

$$\begin{aligned}\lambda_{k_1,k_2}(A) = A_0 + 4(A_s + A_c) - 4(A_s + 2A_c)\left(\sin^2\frac{k_1h_x}{2} + \sin^2\frac{k_2h_x}{2}\right)\\ + 16A_c\sin^2\frac{k_1h_x}{2}\sin^2\frac{k_2h_x}{2}.\end{aligned} \tag{37}$$

4.1 Conjugate Gradient

For the Conjugate Gradient method with a Hermitian positive definite linear system $Ax = b$, it is well known that the error of the n^{th} iteration, e_n, is bounded by

$$\frac{\|e_n\|_A}{\|e_0\|_A} \le 2\left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^n, \tag{38}$$

where $\kappa = \frac{\lambda_{max}(A)}{\lambda_{min}(A)}$ is the condition number of A (see, for example, [22]).

Equation (37) represents the eigenvalues of a general compact scheme. Using the values of the coefficients A_0, A_s, and A_c of the fourth order scheme (22), we have

$$A_0 + 4(A_s + A_c) = \frac{10}{3} - \frac{2}{3}Kh_x^2 + 4\left[-\frac{2}{3} - \frac{1}{12}Kh_x^2 - \frac{1}{6}\right] = -Kh_x^2,$$

and

$$A_s + 2A_c = \left[-\frac{2}{3} - \frac{1}{12}Kh_x^2\right] + 2\left(-\frac{1}{6}\right) = -1 - \frac{1}{12}Kh_x^2.$$

Substituting into (37), the eigenvalues of A are given by

$$\lambda_{k_1,k_2}(A) = -Kh_x^2 + \left(4 + \frac{1}{3}Kh_x^2\right)\left(\sin^2\frac{k_1h_x}{2} + \sin^2\frac{k_2h_x}{2}\right) \\ - \frac{8}{3}\sin^2\frac{k_1h_x}{2}\sin^2\frac{k_2h_x}{2}.$$

For simplicity, assume that $Kh_x^2 \le -12$, so that the coefficients of all the sine terms are negative (note that this condition is satisfied for the steady-state equation of interest in this paper). Because sine is monotonic on the interval $[0, \frac{\pi}{2}]$ and all sine terms have negative coefficients, their sum (and, therefore, the eigenvalues themselves) will be monotonic with respect to k_1, k_2. We then conclude that the largest and smallest eigenvalues will result from the cases $k_1 = k_2 = 0$ and $k_1 = k_2 = M$, respectively. This results in

$$\lambda_{0,0}(A) = -Kh_x^2, \\ \lambda_{M,M}(A) = \frac{16}{3} - \frac{1}{3}Kh_x^2.$$

Note that the eigenvalues are all positive since we assume $Kh_x^2 \le -12$ and thus the smallest and largest eigenvalues are $\lambda_{\min}(A) = \lambda_{M,M}$ and $\lambda_{\max}(A) = \lambda_{0,0}$.

In the case where we want to solve the Helmholtz equation (31) for any fixed K, the quantity $|Kh_x^2|$ becomes smaller as the grid is refined (in contrast to the cases resulting from the wave equation above, for which we assumed $Kh_x^2 < -12$.) In particular, the eigenvalue $\lambda_{0,0}(A)$ tends to 0 while $\lambda_{M,M}(A)$ tends to $\frac{16}{3}$. Thus the condition number $\kappa(A) = \lambda_{\max}(A)/\lambda_{\min}(A)$ approaches infinity as the step size h_x becomes smaller, and the convergence factor of CG from (38) will tend to 1, indicating that convergence of CG iterations may become very slow as the grid is refined. Our numerical tests confirm this analysis, and rapid convergence was observed for the modified Helmholtz equation when $Kh_x^2 < -12$ with slower convergence as the grid is refined for the case when $Kh_x^2 \to 0$.

For the wave equation discretized by the θ method, one obtains the generalized Helmholtz equation (31) with $Kh_x^2 = -\frac{1}{\theta CFL^2}$.[1] Then the condition number of A is given by

$$\kappa = \frac{\lambda_{\max}(A)}{\lambda_{\min}(A)} = \frac{\frac{1}{\theta CFL^2}}{\frac{16}{3} + \frac{1}{3\theta CFL^2}} = \frac{3}{16\theta CFL^2 + 1}.$$

For the 4th order scheme in time, $\theta = \frac{1}{12}$ and the stability condition is $CFL^2 \leq \frac{3}{8}$. Taking $CFL^2 = \frac{3}{8}$ gives the condition number

$$\kappa(A) = \frac{3}{16 \cdot \frac{1}{12} \cdot \frac{3}{8} + 1} = \frac{3}{\frac{3}{2}} = 2.$$

From (38) the convergence factor $\frac{\sqrt{\kappa}+1}{\sqrt{\kappa}-1}$ of CG iterations is

$$\frac{\sqrt{2}-1}{\sqrt{2}+1} \approx 0.17.$$

For the 2nd order scheme in time, $\theta \geq \frac{1}{4}$. Even though the scheme is unconditionally stable, it is most efficient to take $CFL = h_x$. In that case, the inequality $Kh_x^2 = -\frac{1}{\theta CFL^2} < -12$ is satisfied for $h_x^2 < \frac{1}{12\theta}$ and the condition number is

$$\kappa(A) = \frac{3}{16\theta h_x^2 + 1} \rightarrow 3 \text{ as } h_x \rightarrow 0.$$

Thus, from (38) the convergence factor of CG iterations for the scheme with $\theta \geq \frac{1}{4}$ as the grid is refined using $CFL = h_x$ is

$$\frac{\sqrt{3}-1}{\sqrt{3}+1} \approx 0.27.$$

In both cases, the condition number is well behaved and fast convergence of CG iterations is guaranteed.

4.2 Multigrid

Classical iterative solvers such as Jacobi and Gauss–Seidel are well known to require large numbers of iterations to converge, with the number of iterations increasing as the grid is refined. Multigrid (MG) methods utilize computations on coarse grids to reduce the overall number of iterations required for iterative solvers. For the Poisson equation, it is well established that the number of V-cycles (that is, MG iterations) needed to achieve convergence using the full multigrid (FMG) algorithm is independent of the grid size, resulting in a very efficient solution method. We consider the multigrid method for solving the modified Helmholtz equation (31), both on its own and as the time marching step of implicit schemes for the wave equation.

Let the number of grid points in each coordinate direction be $N_{h_x} = 2^n + 1$. Coarsening the grid by a factor of 2 results in $N_{2h_x} = 2^{n-1} + 1$ grid points that coincide with the fine

[1] For the (4,4) scheme, $\theta = \frac{1}{12}$ and thus $Kh_x^2 = -\frac{1}{\theta CFL^2} < -12$ whenever $CFL < 1$, which is already guaranteed by the stability condition (see Sect. 3).

grid at every other node. We choose the injection operator to project from a fine grid to a coarse grid, which can be written as

$$v^{2h_x} = I_{h_x}^{2h_x} v^{h_x}.$$

To interpolate from a coarse grid to a fine one, we use full weighting, which we write in operator form as $v^{h_x} = I_{2h_x}^{h_x} v^{2h_x}$. The iterative method used in conjunction with MG is known as a "smoother" and performing iterations as "smoothing." The most common smoothers are damped Jacobi and Gauss–Seidel.

We introduce a damping factor ω. Then the damped Jacobi iteration matrix is $R_\omega = I - \omega D^{-1}A$, and therefore the asymptotic convergence rate of the residual is equal to the spectral radius $\rho(R_\omega) = 1 - \frac{\omega}{A_0}\rho(A)$. The eigenvalues are $\lambda_{k_1,k_2}(R_\omega) = 1 - \frac{\omega}{A_0}\lambda_{k_1,k_2}(A)$ with $\lambda_{k_1,k_2}(A)$ given by (37) for a general compact scheme.

The convergence of damped Jacobi requires $\rho(R_\omega) < 1$. For the purposes of multigrid, the damping parameter ω should be chosen so that the absolute value of the largest eigenvalue of the high order modes of R_ω is minimized. We proceed by enforcing the following condition on the eigenvalues of the highest Fourier modes in both x and y:

$$-\lambda_{\frac{M}{2},\frac{M}{2}}(R_\omega) = \lambda_{M,M}(R_\omega), \tag{39}$$

which is a necessary condition for the optimal high-mode damping parameter. For $k_1 = k_2 = \frac{M}{2}$, we have $\sin^2\frac{k_1 h_x}{2} = \sin^2\frac{k_2 h_x}{2} = \sin^2\frac{\pi}{4} = 1/2$, and similarly $k_1 = k_2 = M$ gives $\sin^2\frac{k_1 h_x}{2} = \sin^2\frac{k_2 h_x}{2} = \sin^2\frac{\pi}{2} = 1$. The eigenvalue associated with the middle modes (i.e., the left-hand side of condition (39)) reduces to

$$\begin{aligned}\lambda_{\frac{M}{2},\frac{M}{2}}(R_\omega) &= 1 - \frac{\omega}{A_0}\left[A_0 + 4(A_s + A_c) - 4(A_s + 2A_c)\left(\frac{1}{2} + \frac{1}{2}\right) + 16A_c \cdot \frac{1}{2}\cdot\frac{1}{2}\right]\\ &= 1 - \frac{\omega}{A_0}\left(A_0 + 4A_s + 4A_c - 4A_s - 8A_c + 4A_c\right)\\ &= 1 - \omega.\end{aligned} \tag{40}$$

Thus, for any compact scheme, the eigenvalue of A associated with the middle modes is simply A_0 and results in an eigenvalue of $1 - \omega$ for the damped Jacobi iteration matrix. The eigenvalue associated with the highest modes which appears on the right-hand side of (39) becomes

$$\begin{aligned}\lambda_{M,M}(R_\omega) &= 1 - \frac{\omega}{A_0}\left[A_0 + 4(A_s + A_c) - 4(A_s + 2A_c)(1+1) + 16A_c \cdot 1 \cdot 1\right]\\ &= 1 - \frac{\omega}{A_0}\left(A_0 + 4(A_c - A_s)\right).\end{aligned} \tag{41}$$

Therefore the eigenvalue of A for the highest mode is $\lambda_{M,M}(A) = A_0 + 4(A_c - A_s)$. We now solve (39) for ω using (40) and (41):

$$\begin{aligned}-\lambda_{\frac{M}{2},\frac{M}{2}}(R_\omega) = \omega - 1 = 1 - \frac{\omega}{A_0}\left(A_0 + 4(A_c - A_s)\right) = \lambda_{M,M}(R_\omega)\\ \omega\left[1 + \frac{A_0 + 4(A_c - A_s)}{A_0}\right] = 2\\ \omega = \frac{A_0}{A_0 + 2(A_c - A_s)},\end{aligned} \tag{42}$$

where we assume that $A_0 + 2(A_c - A_s) \neq 0$; otherwise, (39) is false. Using ω from (42), the eigenvalues of the iteration matrix R_ω are given by

$$\begin{aligned}
&\lambda_{k_1,k_2}(R_\omega)\\
&\quad = 1 - \left(\frac{A_0}{A_0 + 2(A_c - A_s)}\right)\frac{1}{A_0}\Bigg[A_0 + 4(A_s + A_c)\\
&\quad - 4(A_s + 2A_c)\left(\sin^2\frac{k_1 h_x}{2} + \sin^2\frac{k_2 h_x}{2}\right) + 16A_c \sin^2\frac{k_1 h_x}{2}\sin^2\frac{k_2 h_x}{2}\Bigg]\\
&\quad = \frac{-2(A_c + 3A_s) - 4(A_s + 2A_c)(\sin^2\frac{k_1 h_x}{2} + \sin^2\frac{k_2 h_x}{2}) + 16A_c \sin^2\frac{k_1 h_x}{2}\sin^2\frac{k_2 h_x}{2}}{A_0 + 2(A_c - A_s)}.
\end{aligned} \tag{43}$$

We now compute the largest eigenvalue, in absolute value, of the iteration matrix for the high-frequency modes. Using $h_x = \frac{\pi}{M}$, the arguments of the sines in (43) are $\frac{\pi}{4} \leq \frac{k_1\pi}{2M}, \frac{k_2\pi}{2M} \leq \frac{\pi}{2}$ for the high-frequency modes $\frac{M}{2} \leq k_1, k_2 \leq M - 1$. Since the sine function is monotonic on the interval $\left[\frac{\pi}{4}, \frac{\pi}{2}\right]$, it attains its maxima and minima only at the boundaries of this interval. We consider the following cases:

Case 1 If $m = n = \frac{M}{2}$, we have from (40) that

$$\left|\lambda_{\frac{M}{2},\frac{M}{2}}(R_\omega)\right| = \left|1 - \frac{A_0}{A_0 + 2(A_c - A_s)}\right| = \left|\frac{2(A_c - A_s)}{A_0 + 2(A_c - A_s)}\right|.$$

The case $m = n = M$ is identical due to condition (39).
Case 2 If $m = \frac{M}{2}, n = M$, then from (37) we have

$$\begin{aligned}
\left|\lambda_{\frac{M}{2},M}(R_\omega)\right| &= \left|1 - \frac{\omega}{A_0}\left[A_0 + 4(A_s + A_c) - 4(A_s + 2A_c)\left(\frac{1}{2} + 1\right) + 16A_c \cdot \frac{1}{2} \cdot 1\right]\right|\\
&= \left|\frac{2A_c}{A_0 + 2(A_c - A_s)}\right|.
\end{aligned}$$

The case $m = M, n = \frac{M}{2}$ is identical by symmetry of (43) with respect to k_1 and k_2. Therefore, the largest eigenvalue of R_ω for the modes $\frac{M}{2} \leq k_1, k_2 \leq M$ is

$$\max_{\frac{M}{2} \leq k_1, k_2 \leq M}\left\{\lambda_{k_1,k_2}(R_\omega)\right\} = \max\left\{\left|\frac{2(A_c - A_s)}{A_0 + 2(A_c - A_s)}\right|, \left|\frac{2A_c}{A_0 + 2(A_c - A_s)}\right|\right\}. \tag{44}$$

Note that we have already assumed the denominator of (44) is nonzero in order to find the parameter ω in (42). Using the values of the coefficients A_0, A_s, and A_c of the fourth order scheme (22), we have

$$A_0 + 4(A_s + A_c) = -Kh_x^2,$$

and

$$A_s + 2A_c = -1 - \frac{1}{12}Kh_x^2.$$

Substituting into (37), the eigenvalues of A are given by

$$\begin{aligned}
\lambda_{k_1,k_2}(A) = -Kh_x^2 + \left(4 + \frac{1}{3}Kh_x^2\right)\left(\sin^2\frac{k_1 h_x}{2} + \sin^2\frac{k_2 h_x}{2}\right)\\
- \frac{8}{3}\sin^2\frac{k_1 h_x}{2}\sin^2\frac{k_2 h_x}{2}.
\end{aligned}$$

For the compact 4th order scheme (22), the optimal damping parameter (42) is given by

$$\omega = \frac{A_0}{A_0 + 2(A_c - A_s)} = \frac{4Kh_x^2 - 20}{3Kh_x^2 - 26}. \tag{45}$$

If $Kh_x^2 \to 0$, which includes the Poisson equation $K = 0$, we have $\omega = \frac{10}{13} \approx 0.77$. This is near the value of the optimal damping parameter for the Poisson equation using the standard 5-point central difference stencil in 2D, which is $\omega = 0.8$. For $K \neq 0$, we require $Kh_x^2 \neq \frac{26}{3}$.

For the eigenvalues, we reduce the following expressions from the general form (43):

$$A_0 + 2(A_c - A_s) = \frac{1}{6}\left(26 - 3Kh_x^2\right)$$

and

$$A_c + 3A_s = -\frac{7}{6} - \frac{1}{12}Kh_x^2.$$

Substituting into (43), the eigenvalues of the iteration matrix R_ω are

$$\begin{aligned}\lambda_{k_1,k_2}(R_\omega) = \frac{1}{26 - 3Kh_x^2}&\left[-7 - \frac{1}{2}Kh_x^2 + \left(24 + 2Kh_x^2\right)\left(\sin^2\frac{k_1 h_x}{2} + \sin^2\frac{k_2 h_x}{2}\right)\right.\\ &\left.- \left(16 + Kh_x^2\right)\sin^2\frac{k_1 h_x}{2}\sin^2\frac{k_2 h_x}{2}\right].\end{aligned} \tag{46}$$

Then

$$\left|\lambda_{\frac{M}{2},\frac{M}{2}}(R_\omega)\right| = \left|\lambda_{M,M}(R_\omega)\right| = \left|\frac{2(A_c - A_s)}{A_0 + 2(A_c - A_s)}\right| = \left|\frac{-Kh_x^2 - 6}{3Kh_x^2 - 26}\right|, \tag{47}$$

and

$$\left|\lambda_{\frac{M}{2},M}(R_\omega)\right| = \left|\lambda_{M,\frac{M}{2}}(R_\omega)\right| = \left|\frac{2A_c}{A_0 + 2(A_c - A_s)}\right| = \left|\frac{2}{3Kh_x^2 - 26}\right|.$$

Observe that for the Laplace equation, $K = 0$, the maximum eigenvalue for the high modes is $\frac{3}{13} \approx 0.23$, which is an improvement over the five-point stencil, which has a largest high-mode eigenvalue of $\frac{1}{3}$ for $K = 0$ (this can be seen from the preceding analysis with $A_0 = 4$, $A_s = -1$, $A_c = 0$).

When solving the modified Helmholtz equation with a fixed $K < 0$, the quantity Kh_x^2 vanishes as the grid is refined. Then $\omega \to \frac{10}{13}$ and the largest high-mode eigenvalue approaches $\frac{3}{13} \approx 0.23$. When the modified Helmholtz equation results from the implicit time discretization of the wave equation, we have $Kh_x^2 = -\frac{h_x^2}{\theta c^2 h_t^2} = -\frac{1}{\theta CFL^2}$. The (2,4) scheme ($\theta \geq \frac{1}{4}$) is unconditionally stable. Nevertheless, it is most efficient to equalize the space and time errors by choosing $h_t \approx h_x^2$ or, equivalently, $CFL \approx h_x$. In this case, the quantity $Kh_x^2 = -\frac{1}{\theta h_x^2}$ approaches negative infinity as the grid is refined, and the absolute value of the largest high-mode eigenvalue of R_ω is given by (46) and approaches $\frac{1}{3}$.

The fourth order time scheme ($\theta = \frac{1}{12}$) is conditionally stable with the requirement that $CFL \leq \frac{\sqrt{.375}}{c}$. The stability condition for the high order scheme does not depend on the grid size h_x, and therefore the quantity $Kh_x^2 = \frac{1}{\theta CFL^2}$ will remain fixed as the grid is refined. For example, if we take a constant wave speed $c = 1$, then the largest CFL satisfying the stability condition is $CFL^2 = 0.375$, which gives $Kh_x^2 = -32$. The absolute value of the largest high-mode eigenvalue of R_ω with $\gamma = 0$ is given by (47), $|\lambda_{\max}| = \frac{13}{61} \approx 0.21$.

4.3 Initial Guess for Implicit Solution of the Wave Equation

When solving the steady-state equation (20) resulting from the implicit discretization of the wave equation, convergence of iterative methods can be improved by using the computed solution at previous time steps to calculate an initial guess. The formulation (9) already utilizes two backwards time levels. A first order backwards interpolation is given by $u^{n+1} \approx u^n + \mathcal{O}(h_t)$, and 2nd order is given by $u^{n+1} \approx 2u^n - u^{n-1} + \mathcal{O}\left(h_t^2\right)$. Numerical results show that using either of these approximations as an initial guess improves the convergence of both CG and MG.

Intuitively, the reduction in the number of iterations is due to the fact that we begin with a low-order initial guess and only the high order behavior is left to be resolved. From this insight, we observe that a better second-order initial guess may be obtained using the standard 2nd order explicit scheme,

$$x_0^{n+1} = 2u^n - u^{n-1} + h_t^2 \left(c^2 \Delta_2 u^n + F^n\right), \tag{48}$$

where Δ_2 is the 2nd order 5-point central difference operator. The initial guess (48) has the form of the 2nd order explicit scheme, but takes as its inputs the 4th order approximations u^n and u^{n-1} rather than 2nd order approximations. Hence, the accuracy of this initial guess will exceed that of the true 2nd order explicit scheme at time t_{n+1}. These initial guess choices are compared for the conjugate gradient method in Sect. 5.

5 Numerical Results

The following computations are for the two dimensional wave equation. We solve the wave equation (1) on a square domain of side length $2s$ centered at the origin. The initial data u^0 and $u_t^0 = \psi$ are given (see (2)). One then obtains an approximation to u^1 by combining the Taylor expansion with an equation-based procedure as follows:

$$\begin{aligned} u^1 &= u^0 + h_t u_t^0 + \frac{h_t^2}{2} u_{tt}^0 + \frac{h_t^3}{6} u_{ttt}^0 + \frac{h_t^4}{24} u_{tttt}^0 + \mathcal{O}\left(h_t^5\right), \\ &= u^0 + h_t u_t^0 + \frac{h_t^2}{2}\left(c^2 \Delta u^0 + F^0\right) + \frac{h_t^3}{6}(c^2 \Delta u_t^0 + F_t^0) \\ &\quad + \frac{h_t^4}{24}\left(c^2 \Delta\left(c^2 \Delta u^0 + F^0\right) + F_{tt}^0\right) + \mathcal{O}\left(h_t^5\right). \end{aligned} \tag{49}$$

If u^0 and u_t^0 are given by explicit formulae, then the expressions of (49) can be computed exactly to give the desired approximation. Otherwise, (49) can be approximated by difference formulae. The initial conditions for a scheme of order p should be accurate, in time, to order $p+1$ [29].

We consider test solutions of the form

$$u(x, y, t) = V(x, y)\Phi(t). \tag{50}$$

Then

$$\begin{aligned} u_{tt} &= V(x, y)\Phi''(t) \\ \Delta u &= \Delta V(x, y)\Phi(t). \end{aligned}$$

Substituting into the wave equation, the corresponding right-hand side $F(x, y, t)$ for the test solution is

$$F(x, y, t) = \Phi''(t)V(x, y) - c^2(x, y)\Delta V(x, y)\Phi(t). \tag{51}$$

The general initial procedure (49) reduces to

$$\begin{aligned} u^0 &= V(x, y)\Phi(0), \\ u^1 &= V(x, y)\left(\Phi(0) + h_t\Phi'(0) + \frac{h_t^2}{2}\Phi''(0) + \frac{h_t^3}{6}\Phi'''(0) + \frac{h_t^4}{24}\Phi''''(0)\right) + \mathcal{O}\left(h_t^5\right), \end{aligned} \tag{52}$$

where (52) is truncated to $\mathcal{O}\left(h_t^3\right)$ accuracy for the 2nd order method in time and $\mathcal{O}\left(h_t^5\right)$ for the 4th order method in time. The Dirichlet boundary condition at time t_n for this case is given by

$$\begin{aligned} u(x, \pm s, t_n) &= V(x, \pm s)\Phi(t_n), \\ u(\pm s, y, t_n) &= V(\pm s, y)\Phi(t_n). \end{aligned}$$

We also consider examples in which a Neumann condition is specified at the right-edge of the square with Dirichlet is considered elsewhere,

$$\begin{aligned} u(-s, y, t_n) &= V(-s, y)\Phi(t_n), \\ u(x, \pm s, t_n) &= V(x, \pm s)\Phi(t_n), \\ \frac{\partial u}{\partial x}(s, y, t_n) &= V_x(s, y)\Phi(t_n), \end{aligned}$$

where the Neumann condition is enforced by Eq. (26).

The design accuracy of the scheme (9) in time is determined by the parameter θ, so the scheme is 4th order accurate in time when $\theta = \frac{1}{12}$ and 2nd order in time otherwise. The scheme is unconditionally stable and 2nd order in time for $\theta \geq \frac{1}{4}$. Tests which are 2nd order in time will use only the particular cases $\theta = \frac{1}{2}$, which reduces to the Crank–Nicolson scheme, and $\theta = \frac{1}{4}$, which is the smallest θ that allows unconditional stability. We represent the time and space order of the scheme as an ordered pair so that, for example, a (2,4)-order scheme represents a 2nd order scheme in time and 4th order in space. We refer to this simply as a (2,4) scheme. The 4th order FD scheme in space is described in Sect. 3. For the (2,2)-order simulations of the following sections, a simple 2nd order FD scheme in space is used by approximating the Laplacian by central differences on a five point stencil.

The numerical results in Sects. 5.1, 5.2 and 5.3 are obtained using a direct LU solver, while the subsequent numerical sections investigate the efficiency of conjugate gradient and multigrid as solvers. Additionally, a comparison is made to the standard explicit scheme, which is a (2,2) scheme with $\theta = 0$, as well as a (4,4) explicit scheme described in Sect. 5.8. For the (2,2) explicit scheme, we also use the five point central difference to approximate the Laplacian in space, while the (4,4) explicit scheme uses a 4th order non-compact 9-point star stencil at the interior nodes and 2nd order central differences at the boundaries (it is described in more detail in Sect. 5.8). In the ensuing discussion, we distinguish between the temporal error e_t and the spatial error e_x.

The simulations are implemented in MATLAB using the built-in sparse LU command `lu` as a direct solver as well as the built-in conjugate gradient implementation `pcg`. All of the multigrid routines, including the Jacobi and Gauss–Seidel smoothers, were written in MATLAB. All computations were performed on a Mac Pro with 64 GB of RAM and a 12-core Intel Xeon Processor E5-v2 at 2.7 GHz.

5.1 Constant Coefficient Examples

Let $s = \frac{\pi}{2}$, $V(x, y) = \cos ax \cos by$, and $\Phi(t) = \cos \Omega t$, with $\Omega = c\sqrt{a^2 + b^2}$ and $c(x, y) = c_0$ constant. Substitution into the wave equation (1) shows that this is a solution to the homogeneous equation, $F(x, y, t) = 0$. The test solution $u = \cos ax \cos by \cos \Omega t$ is zero along the boundary $x = \pm\frac{\pi}{2}$, $y = \pm\frac{\pi}{2}$. The initial condition u^0 is given, and we approximate the first time step u^1 by a Taylor expansion (52). For the 2nd order scheme in time (i.e., $\theta \neq \frac{1}{12}$), it is sufficient to take $\mathcal{O}\left(h_t^3\right)$ accuracy in (52): $u^1 = 1 - \frac{\Omega^2 h_t^2}{2}$. In order to maintain 4th order accuracy in time, we require $\mathcal{O}\left(h_t^5\right)$ accuracy in the expansion (52), which yields $u^1 = 1 - \frac{\Omega^2 h_t^2}{2} + \frac{\Omega^4 h_t^4}{24}$.

In Tables 1 and 2, we take $a = b = 1$ and $c = 0.9$, so that $\Omega = 0.9\sqrt{2}$. For the (4,4)-order scheme, it was shown in Sect. 3 that the stability condition for these parameters is $\frac{h_t}{h_x} \leq 0.67$, and we choose the CFL number λ so that h_t is below this threshold. The Dirichlet boundary conditions in this case simplify to $u(t) = 0$ at all the edges of the square, and the Neumann condition at the right edge is given by $\frac{\partial u}{\partial x}(\pi/2, y, t) = V_x(\pi/2, y, t)\Phi(t) = -\alpha \cos \beta y \cos \Omega t$. The final time is chosen to be $t_F = 2$.

The results of Table 1 clearly demonstrate the design convergence rate of each scheme. The (2,2) and (2,4) schemes each achieved better errors when using the smaller value of $\theta = \frac{1}{4}$. While the (2,4)-order scheme does not appear to have an advantage over the (2,2)-order scheme in Table 1, this is because we have only considered a fixed CFL of 0.9. In Table 2 we see that the (2,4)-order scheme can achieve much smaller errors on a fixed grid by reducing the CFL number. Since $e_t = \mathcal{O}\left(h_t^2\right)$ and $e_x = \mathcal{O}\left(h_x^4\right)$, we expect this behavior up to the point that $h_t \approx h_x^2$. Convergence tables based on lowering the CFL for the (2,2)-order and (4,4)-order schemes are therefore omitted since the time error e_t will quickly become smaller than the space error e_x, making the convergence rate in time unobservable in this manner.

We see that the errors for the (2,4) scheme are only slightly better than the (2,2) scheme for moderate CFL numbers. However, Table 2 demonstrates that the error can be dramatically improved by taking smaller CFL numbers with the (2,4) scheme. It has been found that one should reduce the CFL so that the temporal error matches the spatial error for a (2,4) scheme (see [21]). The results in terms of CPU time for this test problem are shown in Table 3.

Observe that the error of the (4,4)-order scheme on an 8×8 grid from Table 1 is $1.03 \cdot 10^{-5}$, which is already smaller than the (2,2)-order and (2,4)-order scheme on a 128×128 grid. We wish to directly compare the relative costs associated with the (2,2)-order, (2,4)-order, and (4,4)-order schemes. In Table 3, we consider an example using a test solution of the same form as the above examples but with $a = 2$ and $b = 5$, so that $\Omega = 0.9\sqrt{29} \approx 4.85$. Using the result of the (4,4)-order scheme on a 32×32 grid as the standard, we seek the number of grid nodes for which the error of the (2,2)-order and (2,4)-order schemes will be approximately the same. We then compare the CPU times for each method. Note that for a fixed number of grid nodes the implicit methods will have approximately the same cost per time step; therefore, we expect the higher order implicit schemes to exhibit better efficiency than the (2,2) scheme. We also include the (2,2) explicit scheme (i.e., $\theta = 0$) in the comparison. Further comparison of the (4,4) implicit scheme with explicit schemes is conducted in Sect. 5.8. We demonstrate the impact of taking a smaller CFL number (i.e., $h_t \approx h_x^2$) for the (2,4)-order scheme. The findings are summarized in Table 3.

We observe from Table 3 that the (4,4)-order scheme is by far the most efficient, followed by the (2,4)-order scheme with small CFL. Further reducing the CFL for the (2,4) scheme

Table 1 Design convergence is observed for the homogeneous, constant coefficient wave equation

Grid	Dirichlet		Neumann	
	Error	Convergence rate	Error	Convergence rate
(2,2)-Order, $\theta = 1/2$, CFL = 0.9				
8	$2.13 \cdot 10^{-2}$	–	$5.82 \cdot 10^{-2}$	–
16	$6.16 \cdot 10^{-3}$	1.79	$3.01 \cdot 10^{-3}$	1.83
32	$1.64 \cdot 10^{-3}$	1.91	$5.96 \cdot 10^{-3}$	1.80
64	$4.19 \cdot 10^{-4}$	1.97	$1.58 \cdot 10^{-3}$	1.92
128	$1.05 \cdot 10^{-4}$	1.99	$4.05 \cdot 10^{-4}$	1.96
(2,2)-Order, $\theta = 1/4$, CFL = 0.9				
8	$1.07 \cdot 10^{-2}$	–	$9.52 \cdot 10^{-3}$	–
16	$3.01 \cdot 10^{-3}$	1.83	$7.43 \cdot 10^{-3}$	0.36
32	$7.89 \cdot 10^{-4}$	1.93	$2.50 \cdot 10^{-3}$	1.57
64	$2.01 \cdot 10^{-4}$	1.97	$7.02 \cdot 10^{-4}$	1.83
128	$5.04 \cdot 10^{-5}$	1.99	$1.85 \cdot 10^{-4}$	1.92
(2,4)-Order, $\theta = 1/2$, CFL = 0.9				
8	$1.89 \cdot 10^{-2}$	–	$5.65 \cdot 10^{-2}$	–
16	$5.40 \cdot 10^{-3}$	1.81	$1.92 \cdot 10^{-2}$	1.56
32	$1.42 \cdot 10^{-3}$	1.92	$5.37 \cdot 10^{-3}$	1.84
64	$3.64 \cdot 10^{-4}$	1.97	$1.43 \cdot 10^{-3}$	1.93
128	$9.14 \cdot 10^{-5}$	1.99	$3.62 \cdot 10^{-4}$	1.96
(2,4)-Order, $\theta = 1/4$, CFL = 0.9				
8	$7.97 \cdot 10^{-3}$	–	$6.91 \cdot 10^{-3}$	–
16	$2.21 \cdot 10^{-3}$	1.85	$5.68 \cdot 10^{-3}$	0.28
32	$5.73 \cdot 10^{-4}$	1.95	$1.91 \cdot 10^{-3}$	1.62
64	$1.45 \cdot 10^{-4}$	1.98	$5.36 \cdot 10^{-4}$	1.83
128	$3.65 \cdot 10^{-5}$	1.99	$1.41 \cdot 10^{-4}$	1.93
(4,4)-Order, $\theta = 1/12$, CFL = 0.6				
8	$1.03 \cdot 10^{-5}$	–	$9.49 \cdot 10^{-5}$	–
16	$6.77 \cdot 10^{-7}$	3.92	$6.10 \cdot 10^{-6}$	3.96
32	$4.32 \cdot 10^{-8}$	3.97	$3.80 \cdot 10^{-7}$	4.00
64	$2.72 \cdot 10^{-9}$	3.99	$2.44 \cdot 10^{-8}$	3.96
128	$1.64 \cdot 10^{-10}$	4.05	$1.54 \cdot 10^{-9}$	3.99

The test solution is $u = \cos x \cos y \cos 0.9\sqrt{2}t$. The constant wave speed is $c = 0.9$ and the final time is $t_F = 2$

did not result in any gain in accuracy since this is the point at which the spatial and temporal errors are roughly the same order, $e_t \approx e_x$. The largest possible CFL for stability was the most efficient for the conditionally stable schemes, as we observed that lowering the CFL for the (2,2) explicit scheme and (4,4) implicit scheme did not improve the error but increased the computational cost. Lowering the CFL for the (2,2) implicit scheme did not improve the accuracy.

Table 2 (2,4)-Order convergence in time on a fixed 64 × 64 grid

CFL	Dirichlet		Neumann	
	Error	Convergence rate	Error	Convergence rate
(2,4)-Order, $\theta = 1/2$				
3.6	$5.40 \cdot 10^{-3}$	–	$1.91 \cdot 10^{-2}$	–
1.8	$1.42 \cdot 10^{-3}$	1.92	$5.37 \cdot 10^{-3}$	1.83
0.9	$3.64 \cdot 10^{-4}$	1.97	$1.41 \cdot 10^{-3}$	1.92
0.45	$9.14 \cdot 10^{-5}$	1.99	$3.62 \cdot 10^{-4}$	1.96
0.225	$2.29 \cdot 10^{-5}$	2.00	$9.15 \cdot 10^{-5}$	1.98
0.1125	$5.73 \cdot 10^{-6}$	2.00	$2.30 \cdot 10^{-5}$	1.99
(2,4)-Order, $\theta = 1/4$				
3.6	$5.40 \cdot 10^{-3}$	–	$5.69 \cdot 10^{-3}$	–
1.8	$1.42 \cdot 10^{-3}$	1.92	$1.91 \cdot 10^{-3}$	1.57
0.9	$3.64 \cdot 10^{-4}$	1.97	$5.36 \cdot 10^{-4}$	1.83
0.45	$9.14 \cdot 10^{-5}$	1.99	$1.41 \cdot 10^{-4}$	1.93
0.225	$2.29 \cdot 10^{-5}$	2.00	$3.61 \cdot 10^{-5}$	1.97
0.1125	$5.73 \cdot 10^{-6}$	2.00	$9.13 \cdot 10^{-6}$	1.98

The test solution is $u = \cos x \cos y \cos 0.9\sqrt{2}t$, which results in a homogeneous wave equation. The constant wave speed is $c = 0.9$ and the final time is $t_F = 2$

Table 3 Comparison of the running time required to achieve a similar error with different schemes

Grid	Error	CFL	# Time steps	CPU time (s)
(4,4)-Order, $\theta = 1/12$				
32	$3.85 \cdot 10^{-4}$	0.68	30	0.014
(2,2)-Order, $\theta = 0$ (explicit)				
256	$3.71 \cdot 10^{-4}$	0.79	207	0.94
(2,2)-Order, $\theta = 1/4$				
512	$4.19 \cdot 10^{-4}$	0.5	652	33.44
480	$3.53 \cdot 10^{-4}$	0.25	1222	56.87
(2,4)-Order, $\theta = 1/4$				
320	$3.72 \cdot 10^{-4}$	0.5	407	10.71
49	$3.64 \cdot 10^{-4}$	0.064	487	0.24

For the (2,2) explicit scheme and (4,4) implicit scheme, the CFL is chosen to be the largest value that still allows for stability. The test solution is $u = \cos x \cos y \cos 0.9\sqrt{2}t$ with a constant wave speed $c = 0.9$ and final time $t_F = 2$

5.2 Variable Speed of Sound Examples

We next consider test problems where the speed of sound $c(x, y)$ is variable. At each time step, the modified Helmholtz equation (20) now has the variable parameter

$$k^2(x, y) = \frac{1}{\theta c^2(x, y) h_t^2}, \qquad c(x, y) \neq 0.$$

All tests were performed on a square of side length 2 (i.e., $s = 1$) centered at the origin. Because the (4,4)-order scheme is only conditionally stable, the largest allowable CFL will depend on the maximum value of $c(x, y)$ on the domain. As a result of the chosen test solution, the wave equation becomes inhomogeneous with the right-hand side given by (51).

Table 4 Design convergence of the (4,4) scheme for the inhomogeneous wave equation with variable wave speed $c^2(x, y) = \frac{x^2}{4} + 1$

Grid	Error	Convergence rate
8	$3.34 \cdot 10^{-2}$	–
16	$2.23 \cdot 10^{-3}$	3.90
32	$1.42 \cdot 10^{-4}$	3.97
64	$8.98 \cdot 10^{-6}$	3.98
128	$5.65 \cdot 10^{-7}$	3.99

The test solution is $u = \cos 5x \cos 3y \cos \Omega_0 t$ with CFL=0.5, $\Omega_0 = 5.85$, $t_F = 4$

Table 5 Design convergence of the (4,4) scheme for the inhomogeneous wave equation with variable wave speed $c^2(x, y) = \frac{x^2}{4} + 1$

Grid	Error	Convergence rate
8	$6.41 \cdot 10^{-3}$	–
16	$4.29 \cdot 10^{-4}$	3.90
32	$2.76 \cdot 10^{-5}$	3.96
64	$1.79 \cdot 10^{-6}$	3.95
128	$1.13 \cdot 10^{-7}$	3.98

The decaying test solution is $u = \cos 5x \cos 5ye^{-\Omega_0 t}$ with $CFL = 0.5$, $\Omega_0 = 5.85$, and $t_F = 2$

Since the wave speed is variable, we define $\Omega_0 := \bar{c}\sqrt{a^2 + b^2}$ where $\bar{c}$ is the average value of $c(x, y)$ on the domain.

For the first example, we take a test solution (50) with $V(x, y) = \cos(ax)\cos(by)$, where $a = 5$ and $b = 2$. Let $c^2(x, y) = \frac{x^2}{4} + 1$. The maximum of $c^2(x, y)$ occurs along the left and right boundaries and is equal to 1.25, so that the stability condition is approximately $\frac{h_t}{h_x} \leq 0.54$. We specify the time component of (50) to be $\Phi(t) = \cos(\Omega_0 t)$. The results are summarized in Table 4.

We next specify a test solution which is decaying in time. Let Φ in (50) be given by $\Phi(t) = e^{-\Omega_0 t}$. Let $V(x, y)$ and $c(x, y)$ be given as in the previous example, with the final time $t_F = 2$. The results are summarized in Table 5.

Next, we consider the Yukawa potential for nuclear particle forces,

$$c^2(r) = g^2 \frac{e^{-\sigma m r}}{r},$$

where $r = \sqrt{x^2 + y^2}$ is the polar radius, m represents the mass of a particle, and g and σ are a scaling constants. We choose $g^2 = 1$ to be the amplitude of the field and $\sigma m = 10$. The field goes to infinity at the origin $r = 0$. For the computations, we set the value at $r = 0$ to match the value at $r = h_r = h_x\sqrt{2}$, so that

$$c^2(0) \approx c^2(h_r) = c^2(h_x\sqrt{2}).$$

Table 6 The test solution for the Yukawa potential is $u = \cos 10x \cos 10y \cos 10\sqrt{2}t$ with variable wave speed $c^2(r) = \frac{e^{-10r}}{r}$

Grid	CFL	# Time steps	(4,4) Scheme		(2,4) scheme	
			Error	Conv. rate	Error	Conv. rate
8	0.634	13	$5.33 \cdot 10^{-2}$	–	$9.09 \cdot 10^{-1}$	–
16	0.131	123	$9.90 \cdot 10^{-4}$	5.75	$4.77 \cdot 10^{-3}$	7.57
32	$4.21 \cdot 10^{-2}$	761	$5.01 \cdot 10^{-5}$	4.30	$1.12 \cdot 10^{-4}$	5.41
64	$1.69 \cdot 10^{-2}$	3792	$2.97 \cdot 10^{-6}$	4.08	$4.36 \cdot 10^{-6}$	4.68
128	$7.56 \cdot 10^{-3}$	16,940	$1.81 \cdot 10^{-7}$	4.04	$2.19 \cdot 10^{-7}$	4.31

The wave equation with this test solution and wave speed is inhomogeneous. The CFL constraint results from the physics since c becomes larger at grid nodes closer to the origin as the grid is refined. The final time is $t_F = 1$

Recall from Sect. 3 the stability requirement of the (4,4)-order scheme is

$$\frac{h_t}{h_x} \leq \frac{\sqrt{0.375}}{\max c^2(r)} = \frac{\sqrt{0.375}h_r}{e^{-10h_r}} = \frac{\sqrt{0.75}h_x}{e^{-10\sqrt{2}h_x}}. \tag{53}$$

The CFL values for each grid determined by (53) are recorded in Table 6. This constraint implies that the (4,4)-order scheme will require increasingly small time steps as the grid is refined. We compare the results to the (2,4)-order scheme with $\theta = 1/4$, which we expect to perform similarly to the (4,4) scheme in this case since $h_t \propto h_x^2$ in the CFL condition (53). The test solution (50) is given by $V(x, y) = \cos(ax)\cos(by)$ with $a = b = 10$ and $\Phi(t) = \cos(\Omega_1 t)$ with $\Omega_1 = \sqrt{a^2 + b^2} = 10\sqrt{2}$. The final time for Table 6 is $t_F = 1$. For the Yukawa potential, the (4,4)-order scheme slightly outperforms the (2,4)-order scheme but with a distinct advantage on coarse grids. Their computation times are similar since the same CFL is used. Even though the (2,4)-order scheme is unconditionally stable, taking a larger CFL increases the error and is less efficient [21]. Note that our choice of replacing $c^2(0)$ by the value at $c^2(h_r)$ introduces a discontinuity in the first derivative of the wavenumber. In Table 6, no loss of convergence is observed. This is because we use a test solution which is known to be singularity free. In the general case, a smooth polynomial interpolation should be used instead.

Tables 4, 5 and 6 demonstrate that the design convergence rate of the (4,4) scheme is achieved for a variety of inhomogeneous and variable coefficient problems. The example of the Yukawa potential (see Table 6) shows that in some cases the (2,4)-order scheme may be as efficient as the (4,4)-order scheme due to the CFL constraint. We next construct examples for which the (2,4)-order scheme is computationally more efficient. This may happen, for example, if the frequency of the solution in time is substantially lower than that of the frequency in space. Consider the test solution given by $u = V(x, y)\Phi(t)$ with $V(x, y) = \cos(20x)\cos(20y)$, $\Phi(t) = \cos(\sqrt{2}t)$, and constant wavenumber $c(x, y) = 1$. This results in an inhomogeneous wave equation with right-hand side F given by (51). The advantage of taking a larger time step will be more pronounced over longer time intervals, and so we take a final time of $t_F = 10$ to reflect this in the following comparison.

The results of Table 7 show that the use of the (2,4)-order scheme with a larger CFL yields similar overall errors at a significantly lower computational cost. The CFL numbers for the (2,4) scheme were chosen by trial-and-error up to the point at which the overall error started

Table 7 Comparison of (4,4)-order and (2,4)-order schemes for the test solution $u = \cos 20x \cos 20y \cos \sqrt{2}t$, which has higher frequency in space than in time

The equation is inhomogeneous with constant wave speed $c = 1$, and the final time is $t_F = 10$

Grid	Error	CFL	# Time steps	CPU time (s)
(4,4)-Order, $\theta = 1/12$				
64	$9.27 \cdot 10^{-4}$	0.60	534	0.464
128	$5.86 \cdot 10^{-5}$	0.60	1067	4.265
256	$3.68 \cdot 10^{-6}$	0.60	2134	42.61
(2,4)-Order, $\theta = 1/4$				
64	$8.96 \cdot 10^{-4}$	0.60	534	0.470
64	$8.94 \cdot 10^{-4}$	2	160	0.157
64	$9.23 \cdot 10^{-4}$	2.5	128	0.115
128	$5.52 \cdot 10^{-5}$	0.60	1067	4.074
128	$3.21 \cdot 10^{-5}$	10	64	0.221
128	$6.55 \cdot 10^{-5}$	16	40	0.144
128	$9.99 \cdot 10^{-5}$	17.5	37	0.149
256	$3.61 \cdot 10^{-6}$	0.60	2134	37.73
256	$5.66 \cdot 10^{-6}$	10	128	2.223
256	$8.22 \cdot 10^{-6}$	11	117	2.003

to grow (i.e., when $e_t \approx e_x$). We observe that the largest CFL that maintains the same overall error of the (2,4) scheme is much greater for finer grids in this case.

A similar situation arises from the use of nonuniform meshes (which we do not consider here), in which case the time step is determined by the finest spatial mesh. We can simulate such a case by taking a variable propagation speed $c(x, y)$ which sharply increases from one side of the medium to the other. Let

$$c^2(x, y) = 1 + 3e^{20(y-1)}.$$

Define the test solution $u = V(x, y)\Phi(t)$ by $V(x, y) = \cos x \cos y$ and $\Phi(t) = \cos(\sqrt{2}t)$ (i.e., $a = b = 1$). This results in an inhomogeneous wave equation with right-hand side F given by (51). Note, that in this case the maximum of $c^2(x, y)$ is 4 at the top of edge of the square, so the stability condition for the (4,4) scheme is approximately $\frac{h_t}{h_x} \leq 0.15$. The final time for the simulations is $t_F = 2$. In Table 8, we compare the efficiency of the (2,4)-order and (4,4)-order scheme for this problem. In contrast to Table 7, we observe that the largest CFL that maintains the overall error becomes smaller as the spatial grid becomes finer.

5.3 Pollution Effect

We consider the homogeneous wave equation (1) and assume $u(x, y, t) = e^{i\Omega t} V(x, y)$ (more generally, we consider the Fourier transform of u in time). Then V solves $\Delta V + \frac{\Omega^2}{c^2} V = 0$. We increase the frequency Ω on successive grids by a factor $2^{4/5} \approx 1.741$ (this is the predicted growth of the dispersion error for a 4th order scheme when doubling the number of grid points [4,6]). We do so for the same variable-coefficient test solution used in Table 6, and we increase the parameter Ω_0 by a factor of 1.741 on each grid. We set $a = b$ so that $\Omega_0 = \bar{c}\sqrt{a^2 + a^2} = \bar{c}\sqrt{2}a$. Therefore, the desired increase in the frequency of the test solution is obtained by increasing a by a factor of 1.741 as the grid is refined by a factor of 2. We begin in this case with $a = 1$ on the coarsest grid. On the square of side length 2

Table 8 Comparison of (4,4)-order and (2,4)-order scheme for a sharply variable wavespeed $c^2(x, y) = 1 + 3e^{20(y-1)}$

Grid	Error	CFL	# Time steps	CPU time (s)
(4,4)-Order, $\theta = 1/12$				
64	$6.40 \cdot 10^{-4}$	0.20	279	0.267
128	$3.40 \cdot 10^{-5}$	0.20	640	2.331
256	$2.49 \cdot 10^{-6}$	0.20	1280	23.79
(2,4)-Order, $\theta = 1/4$				
64	$6.35 \cdot 10^{-4}$	0.20	279	0.273
64	$6.20 \cdot 10^{-4}$	0.9	72	0.061
64	$7.94 \cdot 10^{-4}$	1.5	43	0.044
128	$3.89 \cdot 10^{-5}$	0.20	640	2.342
128	$4.61 \cdot 10^{-5}$	0.8	160	0.575
256	$2.24 \cdot 10^{-6}$	0.20	1280	22.29
256	$2.84 \cdot 10^{-6}$	0.40	640	11.33
256	$4.50 \cdot 10^{-6}$	0.50	512	8.905

The inhomogeneous test solution is $u = \cos x \cos y \cos \sqrt{2}t$ with final time $t_F = 2$. A larger CFL number results in similar accuracy but at a lower computational cost for the (2,4) scheme

Table 9 The inhomogeneous wave equation with variable wave speed $c^2(x, y) = \frac{x^2}{4} + 1$ is solved for the test solution $u = \cos x \cos y \cos \Omega_0 t$ with $CFL = 0.4$ and final time $t_F = 2$

Grid	Error	Ω_0
8	$2.46 \cdot 10^{-5}$	1.56
16	$2.71 \cdot 10^{-5}$	2.72
32	$1.05 \cdot 10^{-5}$	4.73
64	$1.76 \cdot 10^{-5}$	8.23
128	$1.68 \cdot 10^{-5}$	14.33
256	$1.69 \cdot 10^{-5}$	24.96
512	$1.67 \cdot 10^{-5}$	43.46
1024	$1.14 \cdot 10^{-5}$	75.66

As the frequency Ω_0 of the solution increases by a factor of $2^{4/5}$ on each successive grid, the pollution quantity $k^{p+1}h^p$ remains constant

centered at the origin with $c^2(x, y) = \frac{x^2}{4} + 1$, we have $\bar{c} = 1.0885$. The results using the (4,4) implicit scheme are summarized in Table 9.

Since the error stays relatively constant in Table 9 as the wavenumber $k^2 = \frac{\Omega^2}{c^2}$ of the associated Helmholtz equation and number of grid nodes are varied according to the predicted growth of the dispersion error, the pollution effect is confirmed.

5.4 Observations

The scheme exhibits the design rate of convergence of 2nd order in time when $\theta = \frac{1}{2}$ or $\theta = \frac{1}{4}$ and 4th order in time when $\theta = \frac{1}{12}$. $\theta = \frac{1}{4}$ (the smallest possible θ for unconditional stability) yielded errors that were about half of the magnitude of the Crank–Nicolson scheme ($\theta = \frac{1}{2}$).

Fourth order convergence of the scheme when $\theta = \frac{1}{12}$ for the inhomogeneous equation with a variable speed of sound was confirmed. As evidenced by Table 3, the high order (4,4) scheme is typically more efficient than the (2,2)-order, and (2,4)-order implicit schemes despite the CFL restriction and we have confirmed that the computational efficiency of the (2,4) scheme is greatly improved using a CFL which is near the spatial step size h_x. Using a direct LU solver, the (4,4) scheme was also shown to be more efficient than the (2,2)-order explicit scheme. In Sects. 5.5, 5.6 and 5.7 we investigate the use of conjugate gradient and multigrid solvers, and in Sect. 5.8, we provide an additional comparison of the (4,4) implicit scheme with the (2,2) explicit scheme as well as a higher order (4,4) explicit scheme.

Tables 4, 5 and 6 demonstrate the design convergence rate of the scheme even for a variable speed of sound and inhomogeneous problems. Note that even though we do not treat coordinate transformations in this work, the stability analysis of Sect. 2.2 extends to the case of a self-adjoint Laplacian term and also polar coordinates. In the case of the Yukawa potential of Table 6, the CFL restriction of the (4,4) scheme resulted in similar errors for the (2,4) and (4,4) schemes but with no difference in computational cost. Tables 7 and 8 represent special cases, such as when the solution exhibits substantially more variation in space than in time or in the case of a nonuniform mesh, for which the (2,4) scheme is more efficient than the (4,4) scheme. This is because the overall error is dominated by the spatial error in these cases, even for a very large CFL number.

We remark that the parameter k of the modified Helmholtz equation (20) is inversely proportional to the time step h_t. Table 2 confirms that the convergence rate of the scheme is unaffected by the growth of the parameter k in (20) on a fixed grid as the CFL is lowered. Additionally, note that k becomes larger as the spatial grid becomes finer, and no loss of convergence is observed. On the other hand, we see in Table 9 that when the wavenumber of the associated Helmholtz equation (i.e., obtained by the Fourier transform for a particular solution) grows, the pollution effect is observed.

5.5 Conjugate Gradient Solver

The analysis of Sect. 4.1 indicates that CG will converge rapidly for the modified Helmholtz equation when Kh_x^2 does not tend to zero, as is the case for the implicit time marching of the wave equation. The parameters to be evaluated are the residual tolerance for terminating the CG iterations and the effect of the initial guess x_0 on the number of iterations needed. Assuming that no error estimate is available, we take the residual tolerance for terminating CG iterations to be 10^{-10} in all cases. Preconditioning is not likely to be beneficial for these problems given that the number of CG iterations in all test cases is already small. Preconditioning by an incomplete Cholesky factorization was found to reduce efficiency.

The choice of initial guess at time level t_{n+1} is denoted x_0^{n+1}. In Tables 10 and 11 we consider the four cases discussed in Sect. 4.3: $x_0^{n+1} = 0$, $x_0^{n+1} = u^n$, $x_0^{n+1} = 2u^n - u^{n-1}$,

Table 10 The required number of CG iterations for the (4,4) scheme with $c = 1$ and $CFL = 0.6$

x_0^{n+1}	Average # iterations	Largest # iterations	CPU time (s)
0	8.00	8	62.8
u^n	7.07	10	57.6
$2u^n - u^{n-1}$	6.78	10	55.7
$u_{explicit}^{n+1}$	4.80	8	44.8

The grid size is 256 × 256 with a final time of $t_f = 12$, so that there are 1630 time steps. The homogeneous test solution is $u = \sin 15x \cos 3y \cos \sqrt{234}t$

Table 11 The required number of CG iterations for the (2,4) scheme with $c = 1$ and $CFL = h_x$

x_0^{n+1}	Average # iterations	Largest # iterations	CPU time (s)
0	10	10	122
u^n	7.58	11	100
$2u^n - u^{n-1}$	5.96	7	85.3
$u^{n+1}_{explicit}$	3.99	6	68.8

The grid size is 128 × 128 with a final time of $t_f = 8$, so that there are 13,280 time steps. The homogeneous test solution is $u = \sin 15x \cos 3y \cos \sqrt{234}t$

and $x_0^{n+1} = u^{n+1}_{explicit}$, where $u^{n+1}_{explicit}$ is computed from the solutions u^n and u^{n-1} at prior time levels by the explicit formula (48). The tables display the largest number of CG iterations across all time steps as well as the average number of iterations per time step. The CPU times for the entire wave equation calculation are presented using MATLAB's built-in `pcg` command.

Tables 10 and 11 demonstrate that the number of iterations at each time step is low, even with a zero initial guess. However, a more accurate initial guess substantially reduces the number of iterations per time step and the overall efficiency of the scheme.

5.6 MG Solver

When solving the wave equation by multigrid, the initial guess is provided by the explicit formula (48) (see Sect. 4.3 and also Sect. 5.5). The domain and test solution are the same as in Sect. 5.5. When using FMG for the modified Helmholtz equation, we found it advantageous to use very coarse grids, with the coarsest being 2 × 2. However, for the wave equation, only one coarsening of the original grid was used, and in our testing the use of coarser grids beyond this made no improvement to the residual after each V-cycle.

Jacobi and Gauss–Seidel are tested as smoothers. Regarding damped Jacobi, the damping parameter obtained for general compact schemes in Sect. 4.2 resulted in no advantage over $\omega = 1$ (i.e., Jacobi without damping), and is thus excluded from the tests. When using Gauss–Seidel, alternating forward and backward sweeps are performed. For each smoother, we have experimentally chosen the number of V-cycles, ν_0, the number of smoothing steps before moving to a coarser grid, ν_1, and the number of smoothing steps after returning from a coarse grid, ν_2. The multigrid implementation, including the Jacobi and Gauss–Seidel smoothers, is written in MATLAB.

Figure 1 shows that the number of V-cycles needed in order for the residual to converge remains very small as the grid step size decreases. Convergence of the residual guarantees that the discretization error of the FD scheme has been reached, although it may be achieved sooner.

5.7 Comparison of Solvers

We now carry out a comparison of direct and iterative solvers. We take a test solution of the form $u(x, y, t) = V(x, y)\phi(t)$ with

$$V(x, y) = \begin{cases} (x^2 + y^2)^3(1 - (x^2 + y^2)^3) \sin\left(50 \arctan \frac{y}{x}\right), & \sqrt{x^2 + y^2} \le 1 \\ 0, & \sqrt{x^2 + y^2} > 1 \end{cases}, \tag{54}$$

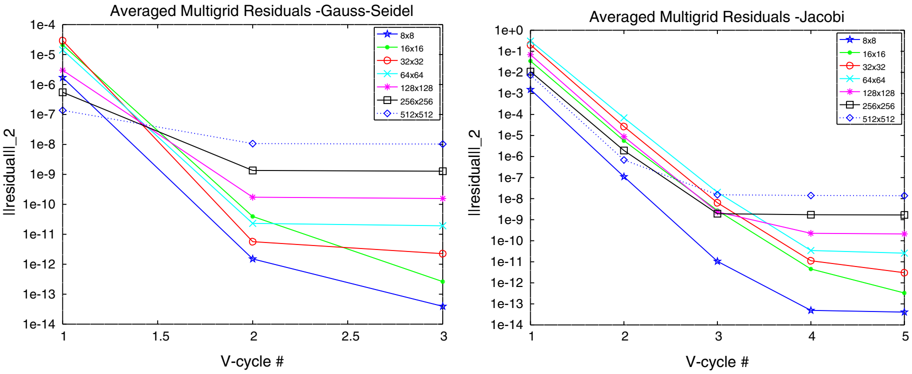

Fig. 1 Average residuals per V-cycle on each grid using a Gauss–Seidel (left) or Jacobi (right) smoother with $\nu_1 = \nu_2 = 4$. As the grid is refined, the residual converges in fewer V-cycles

Table 12 Total running times for direct LU and iterative CG and MG solvers for an inhomogeneous test solution $u = V(x, y) \cos 4t$ with $V(x, y)$ given in (54), $CFL = 0.6$, constant wave speed $c = 1$, and final time $t_F = 1$

Grid	Error	LU time	CG time	MG time (s)
64	$5.32 \cdot 10^{-2}$	0.052	0.18	0.098
128	$3.13 \cdot 10^{-3}$	0.28	0.77	0.47
256	$7.64 \cdot 10^{-5}$	3.07	6.70	4.01
512	$4.25 \cdot 10^{-6}$	31.0	43.6	34.6
1024	$2.57 \cdot 10^{-7}$	288	286	303
2048	$1.60 \cdot 10^{-8}$	2926	2139	2840

As the grid is refined, the iterative solvers become increasingly efficient compared to LU

and $\phi(t) = \cos 4t$ on a square domain of side length $2s = \pi$ centered at the origin. The test solution solves the inhomogeneous wave equation (1) with the right-hand side F given by (51). For Table 12 we solve the problem with Dirichlet boundary conditions, and the test solution is zero at the boundary. The final time is $t_F = 1$. For both CG and MG, the initial guess at each time step is given by the explicit formula (48). The tolerance for CG was 10^{-10}. For MG, a Jacobi smoother is used with $\nu_0 = 2$ V-cycles and $\nu_1 = \nu_2 = 2$ pre- and post-sweeps.

CG was the most efficient for fine grids, and this was due in part to an observed decrease in the average number of CG iterations per time step needed to meet the fixed residual tolerance of 10^{-10} as the grid was refined (for example, the average number of CG iterations per time step for the 64×64 grid was 12.91, but for the 512×512 grid it dropped to only 9.02). It is well known that direct solvers not only have large memory requirements, but also their computational complexity scales poorly as the dimension increases. Moreover, the Jacobi iterations used in MG are well suited to massively parallel computations, while the implementation we have used is serial. The fact that only two grids are needed in the multigrid V-cycle is also advantageous for parallel processing since this results in fewer idle cores throughout the algorithm. Finally, we point out that MATLAB's built-in `pcg` routine is highly optimized and pre-compiled while our multigrid implementation is written entirely in MATLAB, which is an interpreted language. Taking these factors into account, the results of Table 12 show that CG and MG will be very efficient solvers for 3D problems, and that MG may be particularly efficient using parallel processing.

5.8 Comparison of Implicit and Explicit Schemes

In this test, we compare the implicit (4,4) scheme with the standard (2,2) explicit scheme as well as a (4,4) explicit scheme. The test solution is given by $V(x, y) = \cos 7x \cos 7y$ and $\Phi(t) = \cos 7\sqrt{2}t$, resulting in a homogeneous wave equation, with Dirichlet boundary conditions and final time $t_F = 3$. The wave speed used is $c = 1$.

We construct an explicit scheme which is 4th order in both time and space as follows. The step of substituting a central difference in time for Δu^n_{tt} in Eq. (3) results in an implicit equation; therefore, to obtain an explicit 4th order scheme we instead use a second order backwards difference formula:

$$u^n_{tt} = \frac{1}{h_t^2}\left(2u^n - 5u^{n-1} + 4u^{n-2} - u^{n-3}\right) + \mathcal{O}\left(h_t^2\right).$$

Substituting into the wave equation (1) and rearranging for the upper time level, we obtain the 4th order accurate in time explicit formula:

$$u^{n+1} = 2u^n - u^{n-1} + \frac{h_t^2 c^2}{12}\Delta\left(14u^n - 5u^{n-1} + 4u^{n-2} - u^{n-3}\right) + h_t^2 F^n - \frac{h_t^4}{12}F^n_{tt} \quad (55)$$

To obtain 4th order accuracy in space, the Laplacian Δ is approximated with 4th order accuracy on the interior nodes by the 9-point star (non-compact) stencil. The 5-point (2nd order) central difference stencil is used at the boundary nodes. The (4,4) explicit scheme (55) requires additional backwards time levels, which are supplied at the initial time steps by Taylor expansions similar to (49). The CFL condition for the (4,4) explicit scheme was experimentally determined to be the same as that of the standard (2,2) explicit scheme for this test problem.

The solver used for the implicit scheme was CG with a tolerance of 10^{-12} and the initial guess $x^n_0 = 2u^n - u^{n-1}$ at each time step. For the explicit schemes, a CFL number of $\frac{2}{3}\sqrt{3/8} \approx 0.41$ is the largest that allows for stability, while the stability requirement for the implicit scheme allows a CFL of $\sqrt{3/8} \approx 0.61$ to be used. As noted in Sect. 2.2, the CFL restriction is 50% larger for the implicit scheme than the explicit schemes. The results are presented in Table 13. We observe graphically in Fig. 2 that the (4,4) implicit scheme is similar in efficiency to the (4,4) explicit scheme and even appears to be more efficient than the high order explicit scheme as the grid is refined. Both high order schemes are far more efficient than the (2,2) explicit scheme. We stress that the (4,4) explicit scheme is not compact.

6 Discussion

We have constructed an approximation to the variable coefficient two-dimensional wave equation which is fourth order in space and time and also compact in both space and time for Dirichlet and Neumann boundary conditions. Due to the compactness, high order global accuracy is achieved without any special treatment of the initial and boundary conditions; however, a compact high order scheme must be implicit. Even though the time discretization is implicit, a bounded CFL condition is required to achieve stability for the (4,4)-order scheme. A comparison with a (2,2)-order explicit scheme demonstrated the significant gain

Table 13 Comparison of the running times and errors of the (4,4) implicit scheme ($\theta = 1/12$) with the (2,2) explicit scheme and a (4,4) explicit scheme

Grid	(2,2) Explicit, $CFL = 0.41$		(4,4) Explicit, $CFL = 0.41$		(4,4) Implicit, $CFL = 0.61$	
	Error	Time (s)	Error	Time (s)	Error	Time (s)
16	$5.54 \cdot 10^{-1}$	$2.51 \cdot 10^{-3}$	1.17	$4.53 \cdot 10^{-3}$	$1.97 \cdot 10^{-1}$	$1.32 \cdot 10^{-2}$
32	$1.54 \cdot 10^{-1}$	$7.14 \cdot 10^{-3}$	$1.48 \cdot 10^{-1}$	$1.27 \cdot 10^{-2}$	$1.63 \cdot 10^{-2}$	$3.74 \cdot 10^{-2}$
64	$3.95 \cdot 10^{-2}$	$2.92 \cdot 10^{-2}$	$7.84 \cdot 10^{-3}$	$5.98 \cdot 10^{-2}$	$1.04 \cdot 10^{-3}$	0.18
128	$1.00 \cdot 10^{-2}$	0.14	$4.31 \cdot 10^{-4}$	0.26	$6.57 \cdot 10^{-5}$	0.74
256	$2.49 \cdot 10^{-3}$	0.67	$2.91 \cdot 10^{-5}$	1.34	$4.07 \cdot 10^{-6}$	6.58
512	$6.30 \cdot 10^{-4}$	4.99	$2.35 \cdot 10^{-6}$	10.3	$2.58 \cdot 10^{-7}$	52.9
1024	$1.58 \cdot 10^{-4}$	51.8	$2.22 \cdot 10^{-7}$	112	$2.36 \cdot 10^{-8}$	398

The wave equation is homogeneous with constant wave speed $c = 1$ and test solution $u = \cos 7x \cos 7y \cos 7\sqrt{2}t$. The final time is $t_F = 3$

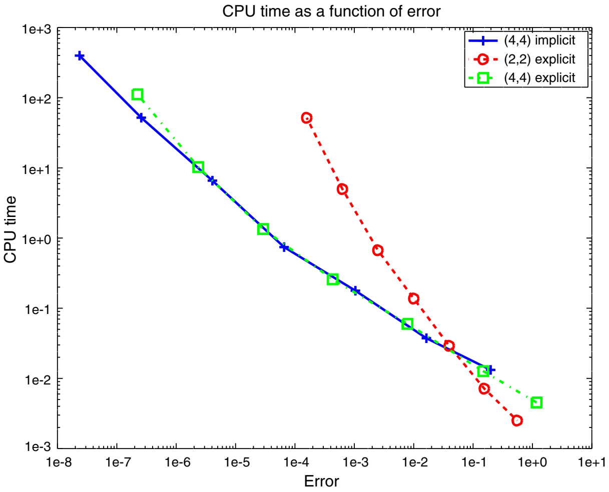

Fig. 2 The (4,4) implicit scheme solved by CG is similar in efficiency to a (4,4) explicit scheme up to an error of about 10^{-6} and becomes more efficient than the explicit scheme beyond that point

in efficiency of the high order scheme despite being implicit. Additionally, the (4,4) implicit scheme was shown to be similar in efficiency to a (4,4) explicit scheme. Comparisons with implicit second order time schemes, which are unconditionally stable, show a great gain in efficiency even though a large time step cannot be used for the (4,4)-order scheme. There are special cases for which the (2,4) scheme may be more efficient than the (4,4) scheme. An example occurs when a coarse resolution in time relative to the space resolution is sufficient or when nonuniform wave speeds or grids determine the CFL number.

The implicitness of the schemes requires the solution of a positive definite elliptic system in space at each step. In this paper, this system was solved directly by LU, as well as iteratively by conjugate gradient and multigrid. Iterative solvers performed comparably with the direct solver in two dimensions for the scale of problems tested, and it was shown that the number of iterations needed for convergence at each time step was small in all cases.

Variable coefficient problems were easily treated. Our closely related work includes the application of the finite difference scheme tested in this paper to the method of difference potentials as an efficient means of solving the wave equation on nonconforming boundary shapes with high order accuracy [10].

The main objective of the current work is to present a detailed analysis of the high order accurate compact scheme for the wave equation, together with a set of supporting numerical simulations, in the case of two space dimensions. Accordingly, its most natural extension, which at the same time is very important from the standpoint of both analysis and applications, would be to build a similar scheme in the case of three space dimensions.

The considerations of Sect. 2.1 that lead to the time-marching schemes (5) and (9) will basically remain unchanged when moving from 2D to 3D, except that the Laplacian Δ will need to be taken in 3D. The stability argument of Sect. 2.2 will stay unaffected as well, except that the boundaries of the spectrum L_{lower} and L_{upper} will depend on the specific discrete approximation L_h of the operator L in 3D (in the simplest constant-coefficient case, $L \equiv \Delta$). The high order accurate spatial discretization of both the differential operator and the boundary conditions in 3D will also be similar to the corresponding 2D constructs of Sect. 3, except that the spatial stencil in 3D will contain 27 (rather than 9) nodes, and the expression (24) for the Fourier symbol will be modified accordingly. The discretization of the 3D modified Helmholtz equation with a variable speed of sound can be based on that presented in [40].

The most significant changes between 2D and 3D will be in the area of solvers. The direct LU solver will no longer provide a feasible approach because of memory and CPU requirements. In the case of constant coefficients, a very efficient solver can be built based on the FFT. However, it will not generalize to variable coefficients. In the case of variable coefficients in 3D, iterative solvers will basically remain the only viable option. Both conjugate gradients and multigrid in 3D will perform similarly to 2D (see Sects. 4, 5.5, and 5.6, as well as "Appendix A"). These schemes perform well for the 3D Laplace equation. Therefore, it is expected that they will converge even faster for the 3D modified Helmholtz equation, which is even more positive definite. Hence, they will be dramatically more efficient than direct methods (such as LU) for the solution of the spatial equation. The actual performance of iterative solvers in 3D needs to be carefully studied both analytically and experimentally. The 3D extension is currently underway, and the corresponding results will be reported in a future publication.

A Numerical Study of Iterative Methods for the Modified Helmholtz Equation

In the course of our study of iterative methods for the wave equation, some observations were made regarding the use of MG and CG for solving the modified Helmholtz equation.

It is well known that the number of V-cycles needed for the Poisson equation when using the second-order central difference stencil grows with the grid size when using a Jacobi smoother without damping (i.e., $\omega = 1$) and it is therefore advantageous to seek a damping parameter for the high frequencies. In Fig. 3, we confirm this classical result for the 2nd order central difference scheme with $\omega = 1$; however the compact scheme converges with $\omega = 1$ in a small number of V-cycles. The test solution is $u = \sin 5x \sin 3y$ on a square domain of side length $s = \pi$ centered at the origin and Dirichlet boundary conditions on all edges. Each V-cycle uses $\nu_1 = \nu_2 = 4$ pre- and post-sweeps of the Jacobi smoother. For the second order

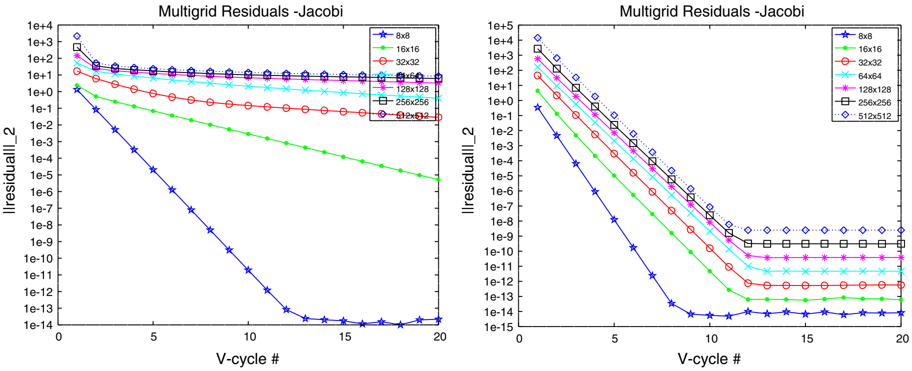

Fig. 3 For the 2nd order central difference scheme (left), Jacobi without damping leads to divergence for the Poisson equation, which is well known. The 4th order compact scheme (right) exhibits convergence within a small number of V-cycles even without damping for the Poisson equation

Table 14 The number of CG iterations increases for the modified Helmholtz equation with a fixed parameter $K = -50$ while the number of MG V-cycles remains constant

Grid	Error	CG iterations	MG V-cycles
16	$1.22 \cdot 10^{-2}$	17	7
32	$1.17 \cdot 10^{-3}$	36	10
64	$7.95 \cdot 10^{-5}$	74	13
128	$5.12 \cdot 10^{-6}$	149	14
256	$3.22 \cdot 10^{-7}$	294	13
512	$2.02 \cdot 10^{-8}$	576	12

The inhomogeneous test solution is $u = \sin 15x \sin 13y$. For MG, a Jacobi smoother ($\omega = 1$) was used with $\nu_1 = \nu_2 = 4$ pre- and post-smoothing iterations. The residual tolerance for CG was 10^{-10} on all grids

central difference scheme, a Jacobi smoother with $\omega = 1$ was also divergent for the modified Helmholtz equation with $K < 0$ but converged rapidly with $\omega = 4/5$, the classical optimal value for the Poisson equation in 2D. By contrast, use of the optimal damping parameter ω^* (see (45), Sect. 4.2) conferred no advantage for the 4th order compact scheme either for the Poisson or modified Helmholtz equation with $K < 0$.

In Sect. 4.1, our analysis showed that the error bound (38) for conjugate gradient was only well behaved when Kh_x^2 does not tend to zero as the grid is refined. The wave equation resulted in favorable cases in which the modified Helmholtz equation satisfied $Kh_x^2 = \frac{1}{\theta CFL^2}$, and this quantity is constant for the (4,4) scheme and actually increasing with the grid size for the (2,4) scheme with $CFL = h_x$. In the following example, we solve the modified Helmholtz equation with a fixed parameter $K = -50$ using the test solution $u = \sin 15x \sin 13y$ on a square of side length 2 centered at the origin and Dirichlet BCs. The residual tolerance for terminating CG iterations is 10^{-10} in Table 14, while the number of MG V-cycles is the point at which the residual converges.

Table 14 shows that the number of CG iterations doubles as the grid is refined by a factor of 2 for the case when K is fixed while the number of MG V-cycles remains small. This indicates that MG will in general be more efficient than CG when solving the modified Helmholtz equation. We ran a set of computations for various constant values of $K = k^2$ and found that the error was a function of kh, indicating that there is no pollution effect for the modified Helmholtz equation.

References

1. Abdulkadir, Y.A.: Comparison of finite difference schemes for the wave equation based on dispersion. J. Appl. Math. Phys. **3**, 1544–1562 (2015)
2. Agut, C., Diaz, J., Ezziani, A.: High-order discretizations for the wave equation based on the modified equation technique. In: 10ème Congrès Français d'Acoustique, Lyon, France (2010)
3. Alford, R.M., Kelley, K.R., Boore, D.M.: Accuracy of finite-difference modeling of the acoustic wave equation. Geophysics **39**(6), 834–842 (1974)
4. Babushka, I.M., Sauter, S.A.: Is the pollution effect of the FEM avoidable for the Helmholtz equation considering high wave numbers? SIAM J. Numer. Anal. **34**(6), 2392–2423 (1997)
5. Barucq, H., Calandra, H., Diaz, J., Ventimiglia, F.: High-order time discretization of the wave equation by Nabla-P scheme. ESAIM Proc. EDP Sci. **45**, 67–74 (2014)
6. Bayliss, A., Goldstein, C.I., Turkel, E.: On accuracy conditions for the numerical computation of waves. J. Comput. Phys. **59**, 396–404 (1985)
7. Britt, S., Tsynkov, S.V., Turkel, E.: A compact fourth order scheme for the Helmholtz equation in polar coordinates. J. Sci. Comput. **45**, 26–47 (2010)
8. Britt, S., Tsynkov, S.V., Turkel, E.: Numerical simulation of time-harmonic waves in inhomogeneous media using compact high order schemes. Commun. Comput. Phys. **9**, 520–541 (2011)
9. Britt, S., Tsynkov, S.V., Turkel, E.: A high order numerical method for the Helmholtz equation with non-standard boundary conditions. SIAM J. Sci. Comput. **35**, A2255–A2292 (2013)
10. Britt, S., Tsynkov, S.V., Turkel, E.: Numerical solution of the wave equation with variable wave speed on nonconforming domains by high-order difference potentials. J. Comput. Phys. **354**, 26–42 (2018)
11. Chabassier, J., Imperiale, S.: Introduction and study of fourth order theta schemes for linear wave equations. J. Comput. Appl. Math. **245**, 194–212 (2013)
12. Chabassier, J., Imperiale, S.: Fourth-order energy-preserving locally implicit time discretization for linear wave equations. Int. J. Numer. Methods Eng. **106**, 593–622 (2016)
13. Ciment, M., Leventhal, S.H.: Higher order compact implicit schemes for the wave equation. Math. Comput. **29**, 985–994 (1975)
14. Cohen, G.C.: Higher Order Numerical Methods for Transient Wave Equations. Springer, New York (2002)
15. Cohen, G.C., Joly, P.: Construction analysis of fourth-order finite difference schemes for the acoustic wave equation in nonhomogeneous media. SIAM J. Numer. Anal. **33**(4), 1266–1302 (1996)
16. Dablain, M.A.: The application of high order differencing to the scalar wave equation. Geophysics **51**(1), 54–66 (1986)
17. Das, S., Liaob, W., Guptac, A.: An efficient fourth-order low dispersive finite difference scheme for a 2-D acoustic wave equation. J. Comput. Appl. Math. **258**(1), 151–167 (2014)
18. Deraemaeker, A., Babushka, I., Bouillard, P.: Dispersion and pollution of the FEM solution for the Helmholtz equation in one, two and three dimensions. Int. J. Numer. Methods Eng. **46**(4), 471–499 (1999)
19. 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)
20. Gilbert, J., Joly, P.: Higher order time stepping for second order hyperbolic problems and optimal CFL conditions. Partial Differ. Equ. **16**, 67–93 (2008)
21. Gottlieb, D., Turkel, E.: Dissipative two-four methods for time dependent problems. Math. Comput. **30**, 703–723 (1976)
22. Greenbaum, A.: Iterative Methods for Solving Linear systems. SIAM, Philadelphia (1997)
23. Gustafsson, B., Mossberg, E.: Time compact high order difference methods for wave propagation. SIAM J. Sci. Comput. **26**, 259–271 (2004)
24. Hamilton, B., Bilbao, S.: Fourth order and optimized finite difference scheme for the 2-D wave equation. In: Proceedings of 16th International Conference on Digital Audio Effects (DAFx-13), Maynooth, Ireland, September 2–6 (2013)
25. Henshaw, W.: A high-order accurate parallel solver for Maxwell's equations on overlapping grids. SIAM J. Sci. Comput. **28**(5), 1730–1765 (2006)
26. Joly, P., Rogriguez, J.: Optimized higher order time discretization of second order hyperbolic problems: construction and numerical study. J. Comput. Appl. Math. **234**(6), 1953–1961 (2010)

27. Kozdon, J.E., Wilcox, L.C.: Stable coupling of non-conforming high-order finite difference methods. SIAM J. Sci. Comput. **38**(2), A923–A952 (2016)
28. Kreiss, H.-O., Oliger, J.: Comparison of accurate methods for the integration of hyperbolic equations. Tellus **24**(3), 199–215 (1972)
29. Lambert, J.D.: Computational Methods in Ordinary Differential Equations. Wiley, New York (1973)
30. Li, Z.: http://tinyurl.com/z5x7log or http://www.math.pku.edu.cn
31. Liang, H., Liu, M.Z., Lv, W.: Stability of theta-schemes in the numerical solution of a partial differential equation with piecewise continuous arguments. Appl. Math. Lett. **23**(2), 198–206 (2010)
32. Liao, W.Y.: On the dispersion, stability and accuracy of a compact higher-order finite difference scheme for 3D acoustic wave equation. J. Comput. Appl. Math. **270**, 571–583 (2013)
33. Liao, W., Yong, P., Dastour, H., Huang, J.: Efficient and accurate numerical simulation of acoustic wave propagation in a 2D heterogeneous media. Appl. Math. Comput. **321**, 385–400 (2018)
34. Mattsson, K., Ham, F., Iaccarino, G.: Stable boundary treatment for the wave equation on second-order form. J. Sci. Comput. **41**(3), 366–383 (2009)
35. Medvinsky, M., Tsynkov, S., Turkel, E.: The method of difference potentials for the Helmholtz equation using compact high order schemes. J. Sci. Comput. **53**(1), 150–193 (2012)
36. Nordström, J., Lundquist, T.: Summation-by-parts in time. J. Comput. Phys. **251**, 487–499 (2013)
37. Shubin, G.R., Bell, J.B.: The stability of numerical boundary treatments for compact high-order finite-difference schemes. J. Comput. Phys. **108**, 272–295 (1993)
38. Singer, I., Turkel, E.: High-order finite difference methods for the Helmholtz equation. Comput. Methods Appl. Mech. Eng. **163**(1–4), 343–358 (1998)
39. Svärd, M., Nordström, J.: Review of summation-by-parts schemes for initial-boundary-value problems. J. Comput. Phys. **268**, 17–38 (2014)
40. Turkel, E., Gordon, D., Gordon, R., Tsynkov, S.: Compact 2D and 3D sixth order schemes for the Helmholtz equation with variable wave number. J. Comput. Phys. **232**(1), 272–287 (2013)
41. Virta, K., Mattsson, K.: Acoustic wave propagation in complicated geometries and heterogenous media. J. Sci. Comput. **61**, 90–118 (2014)
42. Wang, S., Kreiss, G.: Convergence of summation-by-parts finite difference methods for the wave equation. J. Sci. Comput. **71**(1), 219–245 (2017)
43. Zeumi, A.: Fourth order symmetric finite difference schemes for the acoustic wave equation. BIT Numer. Math. **45**(3), 627–651 (2005)