

Solving Forward and Inverse Helmholtz Equations via Controllability Methods

Inauguraldissertation

zur

Erlangung der Würde eines Doktors der Philosophie

vorgelegt der

Philosophisch-Naturwissenschaftlichen Fakultät

der Universität Basel

von

Jet Hoe Tang

aus den Niederlanden

Basel, 2020

Originaldokument gespeichert auf dem Dokumentenserver der Universität

Basel

edoc.unibas.ch

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät
auf Antrag von

Prof. Dr. Marcus J. Grote
Prof. Dr. Martin Gander

Basel, den 17. September 2019

Prof. Dr. Martin Spiess
Dekan

Abstract

Waves are useful for probing an unknown medium by illuminating it with a source. To infer the characteristics of the medium from (boundary) measurements, for instance, one typically formulates inverse scattering problems in frequency domain as a PDE-constrained optimization problem. Finding the medium, where the simulated wave field matches the measured (real) wave field, the inverse problem requires the repeated solutions of forward (Helmholtz) problems. Typically, standard numerical methods, e.g. direct solvers or iterative methods, are used to solve the forward problem. However, large-scaled (or high-frequent) scattering problems are known being competitive in computation and storage for standard methods. Moreover, since the optimization problem is severely ill-posed and has a large number of local minima, the inverse problem requires additional regularization akin to minimizing the total variation. Finding a suitable regularization for the inverse problem is critical to tackle the ill-posedness and to reduce the computational cost and storage requirement.

In my thesis, we first apply standard methods to forward problems. Then, we consider the controllability method (CM) for solving the forward problem: it instead reformulates the problem in the time domain and seeks the time-harmonic solution of the corresponding wave equation. By iteratively reducing the mismatch between the solution at initial time and after one period with the conjugate gradient (CG) method, the CMCG method greatly speeds up the convergence to the time-harmonic asymptotic limit. Moreover, each conjugate gradient iteration solely relies on standard numerical algorithms, which are inherently parallel and robust against higher frequencies. Based on the original CM, introduced in 1994 by Bristeau et al., for sound-soft scattering problems, we extend the CMCG method to general boundary-value problems governed by the Helmholtz equation. Numerical results not only show the usefulness, robustness, and efficiency of the CMCG method for solving the forward problem, but also demonstrate remarkably accurate solutions.

Second, we formulate the PDE-constrained optimization problem governed by the inverse scattering problem to reconstruct the unknown medium. Instead of a grid-based discrete representation combined with standard Tikhonov-type regularization, the unknown medium is projected to a small finite-dimensional subspace, which is iteratively adapted using dynamic thresholding. The adaptive (spectral) space is governed by solving several Poisson-type eigenvalue problems. To tackle the ill-posedness that the Newton-type optimization method converges to a false local minimum, we combine the adaptive spectral inversion (ASI) method with the frequency stepping strategy. Numerical examples illustrate the usefulness of the ASI approach, which not only efficiently and remarkably reduces the dimension of the solution space, but also yields an accurate and robust method.

Acknowledgements

This thesis was written at the Department of Mathematics and Computer Science at the University of Basel and was partly supported by the Swiss National Science Foundation.

My sincere thanks go to my supervisor Prof. Dr. Marcus J. Grote for giving me the opportunity of doing my PhD and for his guidance, advice, and support.

I am very grateful to Prof. Dr. Frédéric Nataf and Dr. Pierre-Henri Tournier for their collaboration, support, helpful discussion, and meetings. I would like to thank Prof. Dr. Martin Gander for his agreeing to act as a co-referee in my thesis committee and Prof. Dr. Frédéric Nataf for acting as an additional expert during my PhD examination.

My sincere thanks also go to the colleagues Dr. Daniel H. Baffet, Yannik Gleichmann, Dr. Marie Graff-Kray, Dr. Michaela Mehlin, Simon Michel, and Dr. Uri Nahum.

Special thanks to my office mate Dr. Daniel H. Baffet for the helpful discussion and proofreading my thesis. Many thanks go to Yannik Gleichmann and Dr. Uri Nahum for proofreading this thesis.

Contents

| | | |
|-----------|---|------------|
| 1 | Introduction | 1 |
| 1.1 | Model problems | 1 |
| 1.2 | Controllability methods | 4 |
| 1.3 | Inverse scattering problems | 6 |
| 1.4 | Outline of the thesis | 9 |
| | | |
| I | Time-harmonic wave equations | 11 |
| 2 | Model problems: Helmholtz equation | 13 |
| 2.1 | Analytical background | 13 |
| 2.2 | Helmholtz equation | 15 |
| 2.3 | Existence and uniqueness | 16 |
| 2.4 | Finite element discretization | 17 |
| 2.5 | Standard numerical methods | 19 |
| 3 | Controllability methods for solving Helmholtz problems | 21 |
| 3.1 | Analytical background | 22 |
| 3.2 | Time-dependent wave equation | 23 |
| 3.3 | Time-periodic and time-harmonic solutions of the wave equation | 29 |
| 3.4 | Exact controllability methods | 35 |
| 3.5 | Frequency filtering procedure | 45 |
| 3.6 | Initial run-up phase | 50 |
| 3.7 | CMCG Algorithm and computational cost | 51 |
| 3.8 | Numerical experiments I: parameter study | 52 |
| 3.9 | Numerical experiments II: Helmholtz equations in physically bounded or unbounded domains | 61 |
| 3.10 | Numerical experiments III: HPC-CMCG method | 65 |
| 4 | More applications with the CMCG approach | 71 |
| 4.1 | Multiple time-harmonic sources | 71 |
| 4.2 | Helmholtz equation with perfectly matched layer | 79 |
| 4.3 | Helmholtz equation in first-order formulation | 85 |
| | | |
| II | Inversion of time-harmonic wave equations | 99 |
| 5 | Inverse Helmholtz problems | 101 |
| 5.1 | PDE-constrained optimization problem | 102 |
| 5.2 | Adaptive spectral decomposition | 106 |
| 5.3 | Adaptive spectral inversion | 118 |
| | | |
| | Conclusion | 129 |

List of Figures

| | | |
|------|--|----|
| 2.1 | Typical bounded computational domain Ω | 13 |
| 2.2 | (Overlapping) domain decomposition | 20 |
| 3.1 | Long-time behavior: initial values z_0 and z_1 of (3.5) | 24 |
| 3.2 | Long-time behavior: snapshots of $z(\cdot, t)$ in $\Omega = (0, L)$, $L = 4$, at time $t = 0, 2, 4, 6, 8$, and 12 | 24 |
| 3.3 | Nodes for linear ($r = 1$, left), quadratic ($r = 2$, center), and cubic ($r = 3$, right) polynomials on a triangle | 27 |
| 3.4 | Nodes for linear ($r = 1$, left), quadratic ($r = 2$, center), and cubic ($r = 3$, right) polynomials on a tetrahedron | 27 |
| 3.5 | Computational mesh with local refinement near the open wedge ob- stacle; “fine part” is indicated by darker (green) triangles. | 28 |
| 3.6 | One-dimensional Helmholtz equation with Neumann condition at $x =$ 0 and Sommerfeld condition at $x = 1$. Exact and numerical solution using the CMCG method with the cost functional J_1 ; (a) without com- patibility condition; (b) with compatibility condition (3.41). | 38 |
| 3.7 | Helmholtz equation in a bounded domain: numerical solutions of (2.9) in $\Omega = (0, 1)$ with Neumann conditions at $x = 0, 1$ using the CMCG method either with J_1 (a), J_2 (b), J_3 (c), or J_∞ (d). | 43 |
| 3.8 | Smooth transient kernel: (a) transient kernel $\theta_{tr}(t)$ from $t = 0$ to $t = t_{tr}$; (b) $\cos(\omega t)$; (c) $\cos(\omega t)\theta_{tr}(t)$ | 51 |
| 3.9 | Semi-discrete convergence: Comparison of the numerical solution u_h , obtained with the CMCG method, and u_h^* , obtained with a direct solver. | 54 |
| 3.10 | Plane wave solution: the relative error of the numerical solution in the unit square using the CMCG method for varying mesh size h | 54 |
| 3.11 | CMCG iterations vs. wave number: One-dimensional Helmholtz equation (2.9) with Dirichlet condition at $x = 0$ and Sommerfeld con- dition at $x = 1$. The number of CMCG iterations vs. the wave number k for a tolerance $Tol = 10^{-6}, 10^{-12}$: (a) with pollution error; (b) without pollution error. The relative L^2 - and H^1 -errors between the exact so- lution of (2.9) and the numerical solution, obtained with the CMCG method, for a tolerance 10^{-12} ; (c) with pollution error; (d) without pol- lution error. | 55 |
| 3.12 | Computational domain Ω with a convex square (a) or a nonconvex cavity (b) shaped obstacle | 56 |
| 3.13 | CG iterations and stopping criteria: relative CG residual $ u_h^{(\ell)} _{CG}$ in (3.76), Helmholtz residual $ u_h^{(\ell)} _H$ in (3.78), and periodicity mismatch $ u_h^{(\ell)} _J$ in (3.77) at the ℓ -th CG iteration. | 58 |

| | | |
|------|---|----|
| 3.14 | CMCG method vs. long-time wave equation solver: plane wave scattering from a convex (a) or a nonconvex obstacle (b). Comparison between the numerical solution, $u_h^{(\ell)}$, obtained with the CMCG method at the ℓ -th CG iteration and the approximate solution $w_h^{(\ell)}$, obtained via (3.80) from the solution of the wave equation at time $t = \ell T$ without controllability. | 58 |
| 3.15 | Initial run-up: plane wave scattering problems from (a) a convex or (b) a nonconvex obstacle: total number of forward and backward wave equations solved over one period $[0, T]$ until convergence. | 59 |
| 3.16 | Plane wave in a semi-opened wave guide: Comparison of the number of CG iterations, obtained with the CMCG method, with the width L of the computational domain $\Omega = (0, L) \times (0, 1)$ and time interval $(0, mT)$ | 60 |
| 3.17 | Sound-soft scattering: contour lines of the numerical solution with the CMCG method combined with explicit local time-stepping. | 62 |
| 3.18 | Sound-soft scattering: relative residual and relative error at each CG iteration using the CMCG method combined with a local time-stepping method with a time-step $\Delta\tau = \Delta t/q$ with $q = 1, \dots, 9$ (left); zoom on the relative residuals at each CG iteration for varying q (right). In the left frame the curves for different q essentially coincide at this scale. | 62 |
| 3.19 | Sound-hard scattering: contour lines of the total wave field obtained with the CMCG method and the compatibility condition (3.40)–(3.41). | 63 |
| 3.20 | Scattering from a penetrable inhomogeneous inclusion: the squared wave speed c^2 is constant inside and outside of the kite-shaped inclusion (a); contour lines of the numerical solution obtained with the CMCG method using the BGT-1 absorbing condition (3.50) and the functional J_1 (b). | 64 |
| 3.21 | physically bounded domain with an inhomogeneous medium | 65 |
| 3.22 | contour lines of the numerical solution | 65 |
| 3.23 | Closed waveguide: the inhomogeneity a is constant inside and outside of the ellipse-shaped inclusion (top); contour lines of the numerical solution of (2.9) using the CMCG method with the functional J_3 defined by (3.43) with $t_1 = \frac{T}{4}$, $t_2 = \frac{T}{2}$, and $t_3 = T$ (bottom). | 65 |
| 3.24 | Marmousi model: propagation velocity $1.5 \leq c(x) \leq 5.5$ [km/s] | 66 |
| 3.25 | Marmousi model: real part of the scattered field with $\omega = 2\pi\nu$, $\nu = 250$ Hz | 67 |
| 3.26 | Marmousi model: total CPU-time in seconds for varying number of cores. For each frequency ν , the FE-discretization and problem size remain fixed. | 67 |
| 3.27 | 3D-cavity: (a) front view of the opening with inner and outer radius; (b) longitudinal cross-section. | 68 |
| 3.28 | 3D-sound-soft scattering problem: total computational time in second by solving (2.9) with the CMCG method in parallel architecture by varying the number of cores for a fixed configuration – see Table 3.3 | 69 |
| 3.29 | 3D-cavity: total wave field (2.9) with $\nu = 6$ obtained with the CMCG method | 69 |
| 4.1 | One-dimensional superposition of Neumann problems: numerical superposition u_h of Helmholtz equations (4.13), obtained with the CMCG-method. | 75 |

| | | |
|------|---|-----|
| 4.2 | One-dimensional superposition of Neumann problems: numerical eigenmodes $u_h^{(\ell)}$, $\ell = 1, \dots, 3$, from the superposition u_h , obtained with the filtering procedure; top row: real part of each eigenmodes; bottom row: imaginary part of each eigenmodes. | 75 |
| 4.3 | Superposition from inhomogeneous inclusion: the inhomogeneous medium a is constant inside and outside of the kite-shaped inclusion (a); the illustration of a minimizer of J_1 (b)–(c), obtained with the CMCG method. | 76 |
| 4.4 | Scattering from inhomogeneous inclusion: (a)–(f) numerical solutions $u_h^{(j)}$ of (4.13), obtained from u_h with the filtering procedure ($\phi_j(t) = e^{i\omega_j t}$); top row: $\text{Re}\{u_h^{(j)}\}$, $j = 1, 2, 3$; bottom row: $\text{Im}\{u_h^{(j)}\}$, $j = 1, 2, 3$ | 77 |
| 4.5 | Scattering from inhomogeneous inclusion and point sources: minimizer $u_h^{(p)}$ of J_1 , obtained with the CMCG method for superposition with the source term $\tilde{f}^{(p)}$, $p = 1, 2, 3$, in (4.4) given by (4.16). | 78 |
| 4.6 | Scattering from an inhomogeneous inclusion and point sources: numerical solutions $u_h^{(j,\ell)}$ of (2.9) with $\omega = \omega_j$ and $f = f^{(\ell)}$, $j, \ell = 1, 2, 3$, obtained from $u_h^{(p)}$, $p = 1, 2, 3$, with the filtering procedure. | 78 |
| 4.7 | Computational domain Ω , surrounded by the PML layer Ω_{PML} | 80 |
| 4.8 | Cubic polynomial damping function ζ | 83 |
| 4.9 | Absorbing boundary condition vs. perfectly matched layer: numerical results of the scattering problem (2.9), obtained with the CMCG method combined with either J_1 ($\Gamma_S = \partial\Omega$) or J_{PML} ; (a) – (b) the relative L^2 - and H^1 -error of the numerical solution u_h at each CMCG iteration, (c) the relative CG residual; (d) the cost functional J_1 and J_{PML} | 84 |
| 4.10 | Sound-soft scattering problem with PML: (a) the inhomogeneous medium a is constant inside and outside of the kite-shaped inclusion; (b) contour lines of the numerical solution of (2.9) using the CMCG method with the functional J_{PML} defined by (4.19). | 85 |
| 4.11 | HDG semi-discrete convergence: comparison of the numerical solution u_h , obtained with the CMCG method, and u_h^* , obtained with a direct solver for the same fixed \mathcal{P}^2 -FE discretization (H^1 -conforming or HDG). | 94 |
| 4.12 | One-dimensional sound-soft scattering: the number of CG iterations vs. the wave number $k = 2^i$, $i = 0, \dots, 5$ | 96 |
| 4.13 | Physically bounded domain: comparison of the exact solution u of (4.29) with the numerical solution u_h , obtained with the CMCG method for the first-order formulation with HDG discretization, either with or without filtering procedure. | 97 |
| 5.1 | Scatterer from circular and star-shaped inclusion: (a) exact total profile u ; (b) reconstruction of u with TV-penalty term; (c) reconstruction of u without regularization term. | 106 |
| 5.2 | Illustration of the definitions (5.29), (5.32), and (5.36): (left) shapes A^k , B^m ; (right) background $(\bar{B}^1 \cup \bar{B}^2) \cap \bar{D}_\delta$, boundary \bar{S} , and region \bar{S}_δ | 109 |
| 5.3 | Illustration of definitions (5.34) and (5.35) with homogeneous background: (left) shapes A^k ; (right) approximated shapes A_δ^k of A^k and their neighborhood U_δ^k | 109 |
| 5.4 | Illustration of Minkowski content (5.42) of B_1 and $U_\delta = B_{1+\delta} \setminus \bar{B}_{1-\delta}$ | 110 |
| 5.5 | Illustration of assertion (i) in Lemma 8. | 114 |

| | | |
|------|--|-----|
| 5.6 | Adaptive spectral approximation of a piecewise constant medium: (a) exact total medium u_δ in FE space; (b) exact background u_δ^0 in FE space; (c) L^2 -orthogonal projection $P_8[u_\delta]$, given by (5.83), of u_δ in AS space with eight eigenfunctions. | 118 |
| 5.7 | Adaptive spectral approximation of a piecewise constant medium: (a) background φ_0 in (5.24) with $w = u_\delta$ and $\varepsilon = 10^{-8}$; (b)–(i) eigenfunctions $\varphi_1, \dots, \varphi_8$ of (5.26) to the first eight eigenvalues $\lambda_1, \dots, \lambda_8$ | 119 |
| 5.8 | Two circular inclusions: comparison of the reconstructions of the unknown medium u (a), either with the grid-based nodal basis (b) or with the AS basis, either without (c) or with (d) dynamic size reduction from Section 5.3.1. | 123 |
| 5.9 | Two circular inclusions: comparison of the misfit J or \widehat{J} (a), the relative L^2 -error (b), and the frequency (c) in each iteration (frequency step for nodal basis and ASI iteration for AS basis) of the inversion method in nodal basis and in AS basis; (d) comparison of the dimension of the AS space in each ASI iteration, obtained with the ASI method with and without dynamic size reduction, while the number of grid-based nodal basis is fixed ($ndofs = 80'401$). | 123 |
| 5.10 | Three circular inclusions: (a) total medium u ; (b) ASI reconstruction. | 124 |
| 5.11 | Three circular inclusions: (a)–(c) eigenfunctions in (5.26) with the exact medium $w = u$; (d)–(f) basis of the AS space $U^{(m)}$ in the last iteration $m = 24$, obtained with the ASI method combined with the space size reduction approach in Section 5.3.1. | 125 |
| 5.12 | Salt dome (Pluto) model: comparison of the exact profile u (a) with the AS approximation (b), obtained with the L^2 -orthogonal projection, and with $u^{(m)}$ (d), obtained with the ASI method after $m = 13$ ASI iterations starting with the initial guess $u^{(0)}$ (c). | 126 |
| 5.13 | Salt dome (Pluto) model: comparison of the misfit (a), the relative L^2 error (b), the number of basis functions K_m (c), and the frequency ν (d) in each ASI iteration m | 126 |

List of Tables

| | | |
|-----|--|----|
| 3.1 | Computational cost and storage of the CMCG method using the original functional J_1 combined with the compatibility condition, denoted by J_1^{cc} , and with the filtering procedure, denoted by $J_1^{fp,cc}$, or the alternative functionals J_m, J_∞ relative to that using J_1 . Crosses indicate that the CMCG method generally does not yield the correct time-harmonic solution. The top line indicates the number of periods $[0, T]$ for which the forward or backward wave equation (3.5) must be solved. | 53 |
| 3.2 | 2D-Marmousi model: \mathcal{P}^2 -FE with 15 points per wave length and $n_T = T/\Delta t = 390$ time steps. | 66 |
| 3.3 | 3D-cavity resonator: \mathcal{P}^1 -FE; as η increases, the ratio $hk^{3/2}$ remains constant to avoid pollution errors [7] | 68 |
| 4.1 | HDG superconvergence: numerical error $\ u - u_h\ $ vs. mesh size h , obtained with the CMCG method for the first-order formulation with \mathcal{P}^2 -HDG discretization for various boundary conditions, either with or without compatibility condition (CC), filtering procedure (FP), and post-processing (PP). | 95 |

List of Symbols

Notation for parameters

| | |
|-----------|--|
| CFL | Courant-Friedrichs-Lewy constant |
| Tol | tolerance |
| T | time period $T = (2\pi)/\omega$ |
| Γ | boundary $\Gamma \subset \partial\Omega$ |
| Ω | computational domain |
| λ | wave length $\lambda = (2\pi)/k$ |
| ν | ordinary frequency $\nu = 1/T$ |
| ω | angular frequency $\omega = 2\pi\nu$ |
| ρ | density |
| a | reciprocal density $a = 1/\rho$ |
| c | wave speed |
| d | spatial dimension $d = 1, 2, 3$ |
| k | wave number $k = \omega/c$ |
| $ndofs$ | number of degrees of freedoms |
| n | outer normal vector |
| $relres$ | relative residual |
| u | squared wave speed $u = c^2$ |

Notation for solutions

| | |
|--------------|---|
| \mathbf{p} | $= \nabla y$ |
| u | time-independent wave field (Part I) |
| v | time-dependent velocity or time-harmonic wave field |
| y^*, z | time-harmonic (Part I) or time-independent (Part II) adjoint wave field |
| y | time-harmonic (Part I) or time-independent (Part II) wave field |

Notation for sets and function spaces

| | |
|--------------------|--|
| C^r | function space of r -times continuously differentiable functions, $0 \leq r \leq +\infty$ |
| C_c^∞ | function space of smooth compactly supported functions |
| $H(\text{div})$ | $= \{\mathbf{q} \in (L^2)^d \mid \nabla \cdot \mathbf{q} \in L^2\}$ |
| H_0^1 | closure of C_c^∞ in H^1 |
| H_D^1 | $= \{w \in H^1 \mid w = 0 \text{ on } \Gamma_D\}$ |
| H_N^{div} | $= \{\mathbf{q} \in H(\text{div}) \mid \mathbf{q} \cdot \mathbf{n} = 0 \text{ on } \Gamma_N\}$ |
| $W^{r,p}$ | Sobolev space, i.e. $L^p = W^{0,p}$, $p \in [1, +\infty]$, and $H^1 = W^{1,2}$ |
| \mathcal{H}^s | s -dimensional Hausdorff measure, $s \leq d$ |
| \mathcal{L} | d -dimensional Lebesgue measure |
| \mathcal{P}^r | piecewise polynomial functions with degree $\leq r$ |
| \mathcal{T}_h | finite set of elements |
| \mathcal{V}_h | finite element function space |

Notation for methods and conditions

| | |
|-------|--|
| AS | Adaptive Spectral |
| ASI | Adaptive Spectral Inversion |
| BFGS | Broyden–Fletcher–Goldfarb–Shanno |
| BGT | Bayliss-Gunzburger-Turkel (boundary condition) |
| CC | Compatibility Condition |
| CFL | Courant-Friedrichs-Lewy (condition) |
| CG | Conjugate Gradient |
| CM | Controllability Method |
| DD | Domain Decomposition |
| DG | Discontinuous Galerkin (Finite Element) |
| FD | Finite Difference |
| FE | Finite Element |
| FP | (Frequency) Filtering Procedure |
| GenEO | Generalized Eigenproblems in the Overlap |
| GMRES | Generalized Minimal Residual |
| HDG | Hybridizable Discontinuous Galerkin (Finite Element) |
| HPDDM | High-Performance unified framework for Domain Decomposition Method |
| LF | Leap-Frog (or second order central finite difference) |
| LTS | Local Time-Stepping |
| LU | LU (decomposition) |
| ML | Mass-Lumping |
| PML | Perfectly Matched Layer |
| PP | (local) Post-Processing |
| QR | QR (decomposition) |
| RK | Runge-Kutta |
| SE | Spectral (Finite) Element |
| SVD | Singular Value Decomposition |
| TV | Total Variation (regularization) |
| WP | Wolfe-Powell (line search) |

CHAPTER 1

Introduction

In recent years, numerical simulations of wave phenomena has been successfully applied to both computational sciences and engineering applications such as seismic reflection used in geophysical exploration [129, 94, 120, 118, 27, 84], tomography in medical imaging [1, 113, 115], and in many other fields.

The goal of exploration of an unknown in a bounded medium is to determine physical properties and position of the unknowns, the scatterer, with less a priori knowledge. While waves propagate through the entire medium, in particular through the penetrable scatterer, one attempts to gather information from the scattered wave field measured near to the boundary. From the scatterer the speed of propagation varies in inhomogeneity. Once the inhomogeneity of the scatterer is known, the simulated wave field is identically equal to the measured data (or the observation) when there is no noise in the observation [4, 43, 77, 84, 132].

However, the scatterer is in general unknown. Hence, one is looking for the (exact) inhomogeneity of the scatterer so that the simulated wave field and the observation are identical, which describes the so-called inverse scattering problem.

In physics, the wave mentioned above can be classified into two main types of waves: mechanical and electromagnetic waves. Mechanical waves including acoustic waves such as the sound or water wave.

In this work we focus on both time-dependent and time-independent acoustic waves, which are described by the (acoustic) wave equation. Here the time-independent wave equation coincides with the Helmholtz equation resulted from Fourier transformation of the wave equation to the frequency domain. It is well-known that it is difficult to solve the Helmholtz equation in the high-frequency regime [55, 20, 61], which is still an active research area. In recent years, parallel methods have become more popular in computational sciences due to efficiency and rise of technology in massively parallel machines [104, 55, 20, 27, 21, 75].

1.1 Model problems

The time-independent scattering problems are modeled by the Helmholtz equation

$$-\nabla \cdot (a(x)\nabla u(x)) - k^2(x)u(x) = f(x). \quad (1.1)$$

It is well-known that (1.1) has a unique solution for certain choices of wave numbers $k = \omega/c$, wave frequency ω , wave speed c , $a(x) = 1/\rho(x)$, the density ρ , source term f , and boundary conditions; e.g. for homogeneous scattering problems ($c \equiv \text{const}$) with a Sommerfeld-type impedance boundary condition [95, 93]. Based on the unique

continuation principle [98, 136], Sauter et al. recently extended this result to the variable coefficient case [74, 125].

The discrete Helmholtz problems resulting from finite element (FE) [24, 122] or finite difference (FD) discretizations are equivalent to solving the linear system of equations,

$$\mathbf{A}_h \mathbf{u}_h = \mathbf{F}_h. \quad (1.2)$$

Here \mathbf{A}_h and \mathbf{F}_h result from the discretization with mesh size h of the partial differential equation and \mathbf{u}_h is the discrete solution of (1.1). The $N \times N$ system matrix \mathbf{A}_h is sparse, structured, complex symmetric, but not Hermitian, and with

$$N \sim h^{-d}, \quad d = 1, 2, 3, \quad (1.3)$$

where d is the spatial dimension.

For a reasonably accurate numerical solution one typically sets at least “10 grid points per wave length” [6, 7, 20, 110, 134], that is

$$kh \approx \frac{2\pi}{10}. \quad (1.4)$$

However, the discrete error $e_h = u - u_h$ of the FE solution u_h grows with k and (1.4) due to the numerical dispersion. In [14, 6], for the quasioptimality of the FE solution,

$$\begin{aligned} \|e_h\|_{H^1} &\leq Ch^r, \\ \|e_h\|_{L^2} &\leq Ch^{r+1}, \end{aligned}$$

with $C > 0$ independent of h , one requires

$$hk^2 \approx \text{const.}$$

Hence one needs a finer mesh and requires e.g. for the P^r -FE solution u_h

$$N \sim h^{-d} \sim k^d \frac{r+1}{r} \quad (1.5)$$

– see [6, 7, 110, 134]. Consequently, the restriction (1.5) in comparison to (1.3) with (1.4) makes the problem (1.2) harder to solve for $k \gg 1$ (see further Section 2.4).

Here we consider some typical class of numerical methods for solving (1.1): (i) direct solvers, (ii) classical splitting or multigrid methods, (iii) Krylov iterative methods

(i) Direct solver. A classical direct solver akin to the LU decomposition method for solving (1.2). Here one instead considers

$$\mathbf{A}_h \mathbf{u}_h = \mathbf{L}_h(\mathbf{U}_h \mathbf{u}_h) = \mathbf{F}_h, \quad (1.6)$$

where \mathbf{L}_h is lower triangular and \mathbf{U}_h upper triangular,

$$\begin{aligned} \mathbf{L}_h \mathbf{v}_h &= \mathbf{F}_h && \text{(forward substitution),} \\ \mathbf{U}_h \mathbf{u}_h &= \mathbf{v}_h && \text{(backward substitution).} \end{aligned}$$

The decomposition of the dense matrix \mathbf{A}_h , e.g., from boundary element or spectral discretization, requires a computational work of $\mathcal{O}(N^3)$ and the two-phase substitution steps requires a computational work of $\mathcal{O}(N^2)$ [41, 55, 61]. For the sparse matrix

\mathbf{A}_h , e.g., from finite element or finite difference discretizations on regular meshes, the elimination and decomposition require $\mathcal{O}(N^{3/2})$ on two-dimensional meshes and $\mathcal{O}(N^2)$ on three-dimensional meshes [61]. For scattering problems with m multiple source terms, (1.2) splits into m linear equations, where the system matrix \mathbf{A}_h remains the same for all source terms and the LU decomposition needs only be computed once.

Direct solvers based on Gauss elimination are robust with respect to the condition number of \mathbf{A}_h . However, for higher frequencies (1.2) leads to a large system of equations due to (1.5). As a consequence, (1.2) typically becomes too expensive in computational cost and storage requirement. Moreover, direct solvers typically are hard to parallelize on distributed memory architectures.

(ii) Classical splitting or multigrid methods. Instead of solving (1.2) directly, classical matrix splitting methods consider

$$\mathbf{A}_h \mathbf{u}_h = (\mathbf{M}_h + \mathbf{N}_h) \mathbf{u}_h = \mathbf{F}_h \quad (1.7)$$

with

$$\mathbf{A}_h = \mathbf{M}_h + \mathbf{N}_h,$$

where \mathbf{M}_h is regular, e.g. (damped) Jacobi, Gauss-Seidel (GS) or, successive over relaxation (SOR). One then solves the iterative problem

$$\mathbf{M}_h \mathbf{u}_h^{m+1} = \mathbf{F}_h - \mathbf{N}_h \mathbf{u}_h^m. \quad (1.8)$$

Nowadays, classical splitting iterations are only used as smoothers and as preconditioner in multigrid and Krylov iterative methods, respectively.

The damped Jacobi iteration effectively removes the error of the high-frequency modes (oscillation part). However, the error of the low-frequency modes (smooth part) converges slowly [55]. Therefore, the damped Jacobi iteration, the smoother, is then used in the smoothing process. Next, the smooth part can be solved efficiently on the coarse grid. In the coarse grid correction, each multigrid iteration first restricts the solution \mathbf{u}_h^m to a coarse grid (restriction step) and solves the linear system on the coarse grid. After that the solution is prolonged back to the finer grid via interpolation (prolongation step).

(iii) Krylov iterative methods. Here (the system matrix) \mathbf{A}_h in (1.2) is sparse and complex symmetric, however indefinite and not Hermitian. As a consequence, the classical conjugate gradient (CG) method cannot directly be applied to (1.2). Instead one typically uses iterative methods such as generalized minimal residual (GMRES) [124], biconjugate gradient stabilized (BiCGSTAB)[133], or quasi-minimal residual (QMR)[60] methods, because they require less computational storage than the direct solver. However, it is well-known that the convergence of these methods is very sensitive to the condition number of \mathbf{A}_h . Moreover, the convergence deteriorates the higher frequency, and thereby it is also very sensitive depending on the frequency [55]. Consequently, the problem (1.2) is hard for iterative methods. Therefore, one requires “good” and efficient preconditioners. Again, instead of (1.2), one considers the equivalent modification,

$$\mathbf{P}^{-1} \mathbf{A}_h \mathbf{u}_h = \mathbf{P}^{-1} \mathbf{F}_h, \quad (\text{left preconditioning}) \quad (1.9a)$$

$$\mathbf{A}_h \mathbf{P}^{-1} \mathbf{v}_h = \mathbf{F}_h, \quad \mathbf{P} \mathbf{u}_h = \mathbf{v}_h, \quad (\text{right preconditioning}) \quad (1.9b)$$

where \mathbf{P} is an invertible matrix.

In recent years, a growing number of increasingly sophisticated preconditioners has been proposed for the iterative solution of the Helmholtz equation: algebraic preconditioners derived from matrix splittings e.g. Jacobi or incomplete LU (ILU) decomposition; shifted Laplacian by adding a complex-valued damping zeroth term [54, 55, 107, 2, 63], which leads to modern multigrid methods [30, 31, 20]; domain decomposition methods based on Schwarz decomposition are first proposed by Miller in 1965 and then by e.g. [65, 62, 64, 75]; algebraic multilevel or sweeping preconditioners using approximate LDL^T factorization [53].

In particular, a rigorous theory providing rates of convergence for the Helmholtz problem with absorption has been developed in [75] and extended to the absorptive Maxwell equations in [21]. While some of those preconditioners may achieve a desirable frequency independent convergence behavior in special situations [53], that optimal behavior is often rapidly lost in the presence of strong heterogeneity. Moreover, they are typically tied to a particular discretization or fail to achieve optimal scaling on massively parallel architectures.

1.2 Controllability methods

In [25, 26], Glowinski et al. proposed an alternative approach to solve the Helmholtz equation in the time domain instead of in frequency domain, which corresponds to the time-dependent wave equation,

$$\frac{1}{c^2(x)} \frac{\partial^2 y(x, t)}{\partial t^2} - \nabla \cdot (a(x) \nabla y(x, t)) = \operatorname{Re}\{f(x) e^{-i\omega t}\} \quad (1.10)$$

with the initial conditions

$$y(x, 0) = y_0(x), \quad \frac{\partial y(x, 0)}{\partial t} = y_1(x). \quad (1.11)$$

It is well-known that (1.10) has a unique solution y , which continuously depends on the (unknown) initial value (y_0, y_1) [56]. For sound-soft scattering problems, if y is a time-periodic solution of (1.10) with the known period $T = (2\pi)/\omega$, then y corresponds to

$$y(x, t) = \operatorname{Re}\{u(x) e^{-i\omega t}\}, \quad (1.12)$$

where u is the unique solution of (1.1) [11]. It is clear that (1.12) immediately yields

$$u(x) = y(x, 0) + \frac{i}{\omega} \frac{\partial y(x, 0)}{\partial t} = y_0(x) + \frac{i}{\omega} y_1(x). \quad (1.13)$$

Following [11], for arbitrary initial value (y_0, y_1) , the asymptotic limit of the corresponding solution of the wave equation converges to the time-harmonic solution $\operatorname{Re}\{u(x) e^{-i\omega t}\}$ as t tends to $+\infty$. Hence, simply integrating (1.10) for a long time eventually yields the time-periodic solution y . Moreover, Mur suggested in [114] that convergence of the time-harmonically forced wave equation (1.10) to the time-harmonic asymptotic regime can be accelerated by pre-multiplying the time-harmonic sources in (1.10) with a smooth transient function. For scattering from non-convex obstacles with ray trapping, simply solving the time-harmonically forced wave equation over a very long time fails to reach the long-time asymptotic final time-harmonic state because of trapped modes.

In [25, 26], Glowinski et al. reformulated the problem using the exact controllability method (CM) [104] as a least-squares problem with the cost functional

$$J(y_0, y_1) = \frac{1}{2} \left\| \sqrt{a} \nabla(y(x, T) - y_0(x)) \right\|^2 + \frac{1}{2} \left\| \frac{1}{c(x)} (y_t(x, T) - y_1(x)) \right\|^2, \quad (1.14)$$

which penalizes the periodic misfit of the solution y of (1.10) with the initial value (y_0, y_1) after one time period $T = \frac{2\pi}{\omega}$. Clearly, J is non-negative and once J is zero, y is T -periodic and the solution of (1.1) is given by (1.13). Hence, (y_0, y_1) is a (global) minimizer of J . Since J is convex and quadratic, they proposed to use the conjugate gradient (CG) method to find a global minimizer of J , which requires the Fréchet derivative J' of J in each CG iteration. While J' only lies in the dual space $H^{-1} \times L^2$, it needs the Riesz representative in the correct space [108] to ensure that the solution remains sufficiently regular and in $H^1 \times L^2$, which coincides with a Poisson-like strongly elliptic problem. In contrast to simply integrating the wave equation for a long-time, the CMCG method always accelerates the convergence to the long-time asymptotic time-harmonic limit and yields a judicious solution of (1.1).

The CMCG method in [25, 26] relied on a piecewise linear finite element (FE) discretization in space and the second-order leapfrog scheme in time. Low-order FE discretizations, however, are notoriously prone to the pollution effects [7, 50]. Moreover, local mesh refinement imposes a severe CFL stability constraint on explicit time integration, as the maximal time-step is dictated by the smallest element in the mesh.

Recently, Heikkola et al. [91, 92] combined higher-order \mathcal{P}^3 spectral FE with the classical fourth-order Runge-Kutta (RK) method to mitigate the pollution effect. To avoid inverting the mass matrix, they used the mass-lumping based on the order-preserved Gauss-Legendre-Lobatto quadrature nodes. In contrast to [25, 26], they followed the discretize-then-optimize approach rather than the optimize-then-discretize approach. Moreover, they also used an algebraic multigrid preconditioner for solving the (strongly) elliptic problem to remove mesh dependence of the convergence at fixed frequency.

In scattering problems, one requires a boundary condition at infinity to ensure the well-posedness of the problem. In [25, 26, 91, 92], they considered scattered wave fields satisfying the first-order absorbing boundary condition, which is inaccurate at higher frequency due to numerical reflections. To reduce undesired numerical reflections, one uses higher-order absorbing boundary conditions on the artificial boundary [52, 12, 76]. Following [102], the CMCG method can immediately be applied to scattering problems satisfying these higher-order boundary conditions.

In [70, 68], Glowinski et al. derived an equivalent first-order (or mixed) formulation for sound-soft scattering problems, where the solution instead lies in $(L^2)^{d+1}$. Moreover, the gradient of the corresponding cost function J in (1.14) operates in $(L^2)^{d+1}$, and hence all CG iterates automatically lie in the correct solution space $(L^2)^{d+1}$. Therefore, the solution of a strongly elliptic problem is no longer needed. They proposed to use Ravier-Thomas (RT) elements for the discretization in space [97, 69]. Following [15, 69], there is mass-lumping for RT quadrilateral elements akin to mass-lumping for classical finite elements; however, so far there is no mass-lumping for RT triangular elements. Thus each time step needs solving a linear equation system, which is very costly in computational time and is not perfectly parallelizable.

So far the controllability method has always been applied to sound-soft scattering problems [25, 26, 91, 92], where both Dirichlet and impedance conditions are imposed at the boundary. However, for boundary value problems governed by the

Helmholtz equation with any other combination of boundary conditions, the original CMCG method will generally fail, because the minimizer of J is no longer unique. An alternative functional J_∞ proposed in [11] recovers the uniqueness of the minimizer in all those problem settings, but it requires storing the entire history of the solution of the wave equation, which may be prohibitive for large-scale problems.

1.3 Inverse scattering problems

In the inversion of the scattering problems the squared propagation speed u of the wave field is unknown. One typically considers two classes of numerical methods to determine the unknown u (or the scatterer) from measurements y^{obs} : qualitative and quantitative methods.

In general, qualitative methods [101, 29] require less a-priori knowledge about the scatterer u . These methods return a qualitative estimation of the location and the shape of the scatterer. In contrast, quantitative methods such as full waveform inversion are applied to recover more physical details of the scatterer [130, 115].

The time-dependent full waveform inversion (FWI) was originally formulated by Lailly [103] and Tarantola [129]. Later, Gauthier, Tarantola, and Virieux first applied the FWI method to the two-dimensional seismic problems [66]. In [119, 121], the FWI approach was proposed to solve frequency-domain FWI problems. This is typically formulated as a PDE-constrained nonlinear least-squares (NLS) problem [104, 130, 132, 32, 27],

$$\begin{aligned} \min_{u \in V} J[u], \quad J[u] &= \frac{1}{2} \|y - y^{\text{obs}}\|_{L^2(D)}^2, \\ \text{s.t. } y &= y(u) \text{ solves (1.1),} \end{aligned} \quad (1.15)$$

where V is a Banach space and either $D \subset \Omega$ or $D \subset \partial\Omega$. It minimizes the misfit between simulated y and measured y^{obs} data with respect to the L^2 -norm. Other norms for the misfit such as L^∞ -norm or L^1 -norm are discussed in [130]. However, the measured data in general is perturbed by (additive or multiplicative) noise, and thus the optimal control u_* in (1.15) and the exact scatterer u may not coincide. Moreover, the NLS problem is ill-posed. To obtain well-posedness, one adds a-priori information of various nature to the (original) NLS problem in the regularization procedure [32]. It is crucial to choose the optimization and regularization method to have the desired regularization property.

1.3.1 Optimization methods

To minimize $J[u]$, one typically uses a standard Newton-type iteration: classical (or full) Newton, truncated Newton, quasi-Newton such as Broyden-Fletcher-Goldfarb-Shanno (BFGS), or Gauss-Newton optimization methods [94, 117, 43, 113, 27, 77, 79, 115]. This leads to an update process

$$\mathbf{H}_\ell \mathbf{d}_\ell = \mathbf{g}_\ell, \quad (1.16)$$

where \mathbf{H}_ℓ is a symmetric positive definite matrix (e.g. an approximation of the Hessian matrix of J), \mathbf{d}_ℓ is the unknown new search direction, and \mathbf{g}_ℓ is a known vector (e.g. the gradient of J). The matrix \mathbf{H}_ℓ or its inversion \mathbf{H}_ℓ^{-1} are usually not computed due to lack of storage and the amount of computational work.

In [94], Hicks et al. proposed applying the full Newton method to the optimization problem (1.15). To compute the exact Hessian matrix, they presented a new

efficient strategy using a “backpropagation” approach, which is similar to that used to compute the derivative of J . However, it still involves a large number of repeated solutions of (1.1).

For the Gauss–Newton method, one typically uses the CG method to compute the approximated Hessian matrix \mathbf{H}_ℓ . Since each CG iteration also involves a large number of repeated solutions of linear systems [94, 115], one requires a suitable stopping criterion to reduce the number of CG iterations. In [51], Eisenstat et al. proposed to impose the Eisenstat-Walker stopping criterion for the CG method. Although it does not solve (1.16) exactly, it still returns a judicious solution.

In contrast to the full Newton or the Gauss–Newton method, the approximated (inverse) Hessian matrix in the quasi-Newton method follows immediately from the derivative of J , and hence no several repeated solutions of (1.1) are required. However, the approximated Hessian matrix may be dense, and thus one instead uses the limited memory BFGS (L-BFGS) method [106, 34].

1.3.2 Regularization strategies

Here we introduce some important classical regularization methods.

(i) Regularization with penalty term. It is well-known that the inverse problem is ill-posed, since the problem may have several local (or global) minima. Hence optimization methods may fail to converge to the exact solution u . Moreover, the perturbations in the observation may impede the convergence. To overcome the problem of getting stuck in a local minimum, the problem needs to be regularized, which is usually done by adding, for instance, a total variation [58, 16, 100, 3] or Tikhonov-type regularization term [131, 71] such as

$$R[u] = \int_{\Omega} |\nabla u(x)| \, dx \quad (\text{total variation}), \quad (1.17)$$

$$R[u] = \frac{1}{2} \int_{\Omega} |\nabla u(x)|^2 \, dx \quad (H^1\text{-penalty}), \quad (1.18)$$

$$R[u] = \frac{1}{2} \int_{\Omega} u^2(x) \, dx \quad (L^2\text{-penalty}). \quad (1.19)$$

Instead of minimizing J in (1.15), one minimizes $J[u] + \alpha R[u]$. Since the additional regularization term α involves additional parameters, an adaptive strategy is proposed to determine α [58, 86, 115].

(ii) Regularization by parameterization. Instead of minimizing the cost functional J in V , one seeks a minimizer

$$\beta = \underset{\widehat{\beta} \in W}{\operatorname{argmin}} J[\psi(\widehat{\beta})], \quad (1.20)$$

where ψ is the parameterization map with $\psi(W) \subset V$. The parameterization is performed to reduce the number of unknown parameters. The choice of parameterization map generally depends on the inverse problem [32].

(iii) Regularization by size reduction. To reduce the number of unknown parameters, one instead considers the candidate of u in $W \subset V$ and seeks

$$w = \underset{\widehat{w} \in W}{\operatorname{argmin}} J[\widehat{w}]. \quad (1.21)$$

One approach to define the subset W is to add new constraints from a-priori information to the control [32].

(iv) **Regularization by filtering or denoising.** Another regularization approach by filtering or denoising [72, 85] is proposed to remove noise and extract useful data from the inverse problem. One reduces the space of the admissible parameter with the low-rank approximation based on the singular value decomposition (SVD) of a matrix. For a large ill-posed matrix, however, one typically cannot determine the SVD due to computational cost and storage requirement [72, 85].

1.3.3 Adaptive spectral decomposition

In [43], de Buhan and Kray presented a new approach based on [44] for viscoelasticity to solve the inverse scattering problem for wave equations. They first considered the series

$$u = \sum_{\ell=1}^{\infty} \beta_{\ell} \varphi_{\ell}, \quad (1.22)$$

where φ_{ℓ} are eigenfunctions satisfying the eigenvalue problem

$$B[\varphi_{\ell}, v] := -(\nabla \cdot (\mu[u] \nabla \varphi_{\ell}), v) = \lambda_{\ell}(\varphi_{\ell}, v) \quad (1.23)$$

with

$$\mu = \frac{1}{\max\{|\nabla u|, \varepsilon\}}, \quad \varepsilon > 0. \quad (1.24)$$

Here $\{\lambda_{\ell}\}$ denotes the monotonically increasing sequence of eigenvalues. To ensure that μ is well-defined and the denominator does not vanish, they added a small $\varepsilon > 0$.

Now, to regularize the solution u in (1.22), they truncated the sum by a finite number K ,

$$u \approx \sum_{\ell=1}^K \beta_{\ell} \varphi_{\ell}. \quad (1.25)$$

It results

$$\min_{v \in V} \{J[v] + \alpha R[v]\}, \quad (1.26)$$

where $V = \text{span}\{\varphi_{\ell}\}_{\ell=1}^K$ is the corresponding adaptive spectral (AS) space (or adaptive eigenspace).

In [77, 79], Grote et al. applied the adaptive spectral decomposition to the inverse scattering problem in frequency domain. Instead of adding an additional penalty term αR , they proposed only to regularize the unknown parameter u in discretization ($\alpha = 0$). Then they combined the AS approach with the optimal control problem [104, 132] for the cost functional (1.15) with either the Gauss-Newton or truncated Newton method [94, 117, 84]. In addition, they proposed to use the frequency continuation strategy from [35, 10, 94] to prevent that the optimization procedure converges to a false local minimum. This approach requires solving a series of inverse scattering problems with growing frequencies, known as frequency stepping. The optimization procedure for each frequency starts with the initial guess from the solution of the previous lower frequency. By starting at the lowest frequency, it avoids converging to a false local minimum. By ending at the highest frequency, it allows to capture more detailed properties of the scatterer. While increasing the frequency, they imposed that the number K of eigenfunctions should grow proportionally with the frequency; hence $K \sim \omega$, e.g. $K = 4\omega$.

In [77], Grote et al. stated the connection of total variation (TV) R given by (1.17) to μ with $\varepsilon = 0$, namely

$$\langle R'[u], v \rangle = \left(\frac{\nabla u}{|\nabla u|}, \nabla v \right) = (\mu[u] \nabla u, \nabla v), \quad v \in H_0^1(\Omega). \quad (1.27)$$

Later, in [79, 115], Grote and Nahum extended the AS approach based on the total variation regularization to other regularizations based on the following penalty term: Gaussian penalty term, Lorentz penalty term, and the Tikhonov L^2 -penalty term.

The numerical examples showed that for profiles u consisting of a sum of m characteristic functions $\{\chi^\ell\}_{\ell=1}^m$, each eigenfunction φ_ℓ of (1.23) based on the TV regularization term, $1 \leq \ell \leq m$, remarkably matches the characteristic function χ^ℓ . As a consequence, the adaptive spectral space for u only requires m eigenfunctions; namely $K = m$ (independent of ω) and $V = \text{span}\{\varphi_\ell\}_{\ell=1}^m$.

1.4 Outline of the thesis

This thesis consists of two parts: forward problem (Part I) and inverse problem (Part II).

In the first part we introduce the model problem, the Helmholtz equation, more precisely the source term, the boundary conditions, the solution space, and its variational formulation. Moreover, we consider some theoretical preliminaries about the uniqueness and regularity of the solution. Then, we consider the discrete model problem with FE method and a list of some theorems about the convergence in space. Next, we reformulate the Helmholtz equation as an optimal control or least-squares problem with exact controllability. We first derive a family of time periodic solutions of the wave equation, which immediately shows that the CMCG method exactly solves the Helmholtz equation. However, the original CMCG method may fail for Helmholtz equations with general boundary conditions, since the solution may not be unique. We thus extend the original CMCG method to more general cases and present some approaches to recover the uniqueness [83].

In the presence of local mesh refinement, the CFL stability constraint on explicit time integration leads to a small time-step. To overcome the bottleneck from the CFL restriction due to local mesh refinement, we combine the CMCG method with high-order explicit local time-stepping (LTS) methods [78, 81].

It is known that the (first-order) absorbing boundary condition is inaccurate due to numerical reflections. We consider two approaches. We first replace the absorbing boundary condition by more accurate boundary condition such as Bayliss-Gunzburger-Turkel (BGT) condition [12, 76]. Secondly, we add a damping perfectly matched layer (PML) surrounding the propagation region of interest [19, 82, 99, 81, 8]. Here the solution in this layer, outside the physical domain, is damped exponentially, and as a consequence, no reflections return from the layer.

In addition, we illustrate that the CMCG approach also yields the superposition of waves of different modes when all source terms are superposition of different modes. Then we show that we obtain all modes separately after a simple frequency filtering procedure.

At the end of this part, we consider the CM based on the wave equation in first-order form. We again show the extension of the CMCG approach for solving arbitrary boundary value problems governed by the Helmholtz equation. In addition, we combine the CMCG method with the recent hybrid discontinuous Galerkin (HDG) method [37, 38], which not only inherits the block diagonal mass-matrix from

the discontinuous Galerkin discretization, but also the superconvergence in space after a simple local post-processing step.

We present a series of numerical examples and experiments which confirms the usefulness of the CMCG method. In addition, we also consider the CMCG method on a massively parallel architecture [80].

In the second part of my thesis, we introduce the inversion of the scattering problem and formulate this problem as a nonlinear optimal control problem. Instead of minimizing the cost functional with a penalty TV-like regularization term, we formulate the solution as a linear combination of the (adaptive) spectral basis, obtained from the eigenvalue problem (1.23). The truncation of the linear combination then leads to regularization of the solution. Here we describe the adaptive spectral (AS) decomposition and present some approximation theory of the AS space [9]. Since the accuracy of the solution and the performance of the regularization using the AS approach depend on the truncation, we introduce a dynamic thresholding strategy to truncate the sum (1.22) of the solution. Following [10, 77, 79], we use a frequency continuation strategy to reconstruct the solution correctly. As the AS space is updated adaptively for each frequency and the AS spaces may not span the entire space \mathbb{R}^N , the current solution may not lie in the new eigenspace after the update. To ensure that the solution is represented correctly in the new eigenspace, we apply the singular value decomposition (SVD) and modified Gram-Schmidt method to merge all AS spaces. Finally, we present a series of numerical experiments to show the efficiency of the adaptive eigenspace and the usefulness of this approach to solve inverse scattering problems.

Part I

Time-harmonic wave equations

CHAPTER 2

Model problems: Helmholtz equation

In this chapter we introduce the model problem, the Helmholtz equation, and consider the existence, uniqueness, and regularity of the solution, as well as the variational and discrete formulation.

We always assume that the computational domain $\Omega \subset \mathbb{R}^d$, $d \leq 3$, is bounded, connected, and open with a Lipschitz boundary $\Gamma = \Gamma_S \cup \Gamma_N \cup \Gamma_D$, which generally consists of physical and artificial boundaries, see Figure 2.1. Moreover, we impose a Dirichlet boundary conditions on Γ_D , Neumann boundary conditions on Γ_N , and impedance or Sommerfeld-like absorbing boundary conditions on Γ_S .

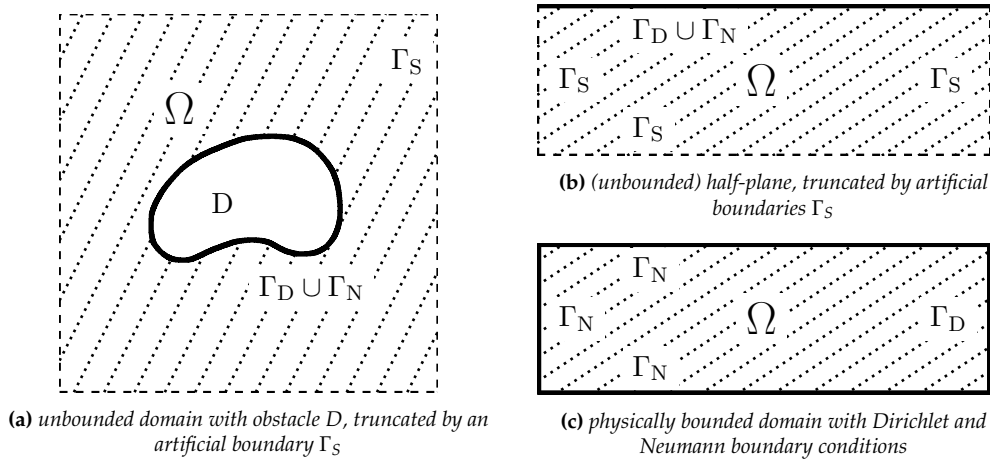


Figure 2.1: Typical bounded computational domain Ω

2.1 Analytical background

Here we summarize important definitions, notations, and theorems without any detailed proofs. More detailed explanations and proofs can be found in the literature [135, 57, 56, 116].

Definition 1. Let $f : \Omega \rightarrow \mathbb{K}$, $\mathbb{K} = \mathbb{R}$ or \mathbb{C} , be a (Lebesgue)-measurable function.

- The function f is α^{th} -weakly differentiable with the multiindex $\alpha = (\alpha_1, \dots, \alpha_d) \in \mathbb{N}_{\geq 0}^d$, if there exists a measurable function g such that

$$\int_{\Omega} f(x) D^{\alpha} \varphi(x) dx = (-1)^{|\alpha|} \int_{\Omega} g(x) \varphi(x) dx, \quad \forall \varphi \in C_c^{\infty}(\Omega), \quad (2.1)$$

where $|\alpha| = \alpha_1 + \dots + \alpha_d$ and

$$D^{\alpha} = \frac{\partial^{\alpha_1}}{\partial x_1^{\alpha_1}} \dots \frac{\partial^{\alpha_d}}{\partial x_d^{\alpha_d}}.$$

If g exists, then the derivative denoted by $D^{\alpha} f$ is unique up to a zero measure set.

- The function f is in $L^p(\Omega)$ for some $p \in [1, +\infty]$, if

$$\|f\|_{L^p(\Omega)} := \begin{cases} \left(\int_{\Omega} |f(x)|^p dx \right)^{\frac{1}{p}} < +\infty, & 1 \leq p < +\infty, \\ \text{ess sup}_{x \in \Omega} |f(x)| < +\infty, & p = +\infty. \end{cases} \quad (2.2)$$

- The function f is in the Sobolev space $W^{r,p}(\Omega)$, if $D^{\alpha} f$, $|\alpha| \leq r \in \mathbb{N}_0$, exists and

$$D^{\alpha} f \in L^p(\Omega). \quad (2.3)$$

The Sobolev space $W^{r,p}(\Omega)$ is a Banach space for $p \in [1, +\infty]$ equipped with the norm

$$\|f\|_{W^{r,p}(\Omega)} := \left(\sum_{|\alpha| \leq r} \|D^{\alpha} f\|_{L^p(\Omega)}^p \right)^{\frac{1}{p}}, \quad f \in W^{r,p}(\Omega). \quad (2.4)$$

For non-integer $r > 0$ and $p \in [1, +\infty]$, $W^{r,p}(\Omega)$ denotes the Sobolev-Slobodeckij space equipped with the norm

$$\|f\|_{W^{r,p}(\Omega)} := \left(\|f\|_{W^{[r],p}(\Omega)}^p + \sum_{|\alpha|=[r]} \int_{\Omega} \frac{|D^{\alpha} f(x) - D^{\alpha} f(y)|^p}{|x-y|^{d+(r-[r])p}} dx \right)^{\frac{1}{p}}, \quad (2.5)$$

see [135, Definition §3.1] and [116, §2.3.8].

We write $H^r(\Omega) = W^{r,2}(\Omega)$ for $p = 2$ and $r \geq 0$, where $H^r(\Omega)$ is a Hilbert space equipped with the (standard) inner product

$$(f, g)_{H^r(\Omega)} := \sum_{|\alpha| \leq r} \int_{\Omega} (D^{\alpha} f(x)) \overline{(D^{\alpha} g(x))} dx, \quad (2.6)$$

for $r \in \mathbb{N}_0$, $f, g \in H^r(\Omega)$, and

$$(f, g)_{H^r(\Omega)} := (f, g)_{H^{[r]}(\Omega)} + \sum_{|\alpha|=[r]} \int_{\Omega} \frac{(D^{\alpha} f(x) - D^{\alpha} f(y)) \overline{(D^{\alpha} g(x) - D^{\alpha} g(y))}}{|x-y|^{d+2(r-[r])}} dx \quad (2.7)$$

for $r \notin \mathbb{N}_0$, $f, g \in H^r(\Omega)$.

It is well-known that $C_c^{\infty}(\Omega)$ is dense in $W^{r,p}(\Omega)$ [56]. Let $W_0^{r,p}(\Omega)$ denote the closure of $C_c^{\infty}(\Omega)$ in $W^{r,p}(\Omega)$, in particular $H_0^r(\Omega)$ denotes the closure of $C_c^{\infty}(\Omega)$ in $H^r(\Omega)$. Then, the dual space of $H_0^r(\Omega)$ is denoted by $H^{-r}(\Omega)$; in particular the dual space $L^2(\Omega)$ is self-dual.

Next, for $f \in H^r(\Omega)$, $r \geq \frac{1}{2}$, we characterize the trace of f on the boundary $\Gamma = \partial\Omega$ by the trace operator T . The existence of T follows from the trace theorem in [135, 57, 56, 116]:

Theorem 1 (Trace extension). *There exists $T : H^r(\Omega) \rightarrow H^{r-\frac{1}{2}}(\Gamma)$, $\frac{1}{2} < r \leq \frac{3}{2}$, and a constant $C > 0$ such that*

$$(i) \quad Tf = f|_{\Gamma} \quad \text{if } f \in C^\infty(\overline{\Omega}), \quad \text{and}$$

$$(ii) \quad \|Tf\|_{H^{r-\frac{1}{2}}(\Gamma)} \leq C\|f\|_{H^1(\Omega)}.$$

Theorem §5.5.2 in [56] implies that

$$H_0^1(\Omega) = \{f \in H^1(\Omega) \mid f = 0 \text{ on } \partial\Omega\}.$$

Theorem 2 (Poincaré–Wirtinger inequality). *Let Ω be a bounded, connected, and open Lipschitz domain. Then there exists a constant $C > 0$, depending only on d and Ω , such that*

$$\left\| f - \frac{1}{\mathcal{L}(\Omega)} \int_{\Omega} f \, dx \right\|_{L^2(\Omega)} \leq C \|\nabla f\|_{L^2(\Omega)}, \quad f \in H^1(\Omega).$$

Theorem 3 (Poincaré inequality). *Let $\Omega \subset \mathbb{R}^d$, $d = 1, 2, 3$, be an open, bounded and connected domain, which satisfies the cone property. Then there exists $C > 0$ such that*

$$\|f\|_{L^2(\Omega)} \leq C \left(\|\nabla f\|_{L^2(\Omega)} + \frac{1}{\sqrt{\mathcal{H}^{d-1}(\Gamma)}} \left| \int_{\Gamma} f(x) \, ds \right| \right). \quad (2.8)$$

for any $u \in H^1(\Omega)$ and Lipschitz hypersurface $\Gamma \subset \partial\Omega$, $\mathcal{H}^{d-1}(\Gamma) > 0$.

The proof of Theorem 2 and 3 can be found in [42, 22].

Remark 1. (i) Here $\mathcal{H}^{d-1}(\Gamma)$ denotes the $(d-1)$ -dimensional Hausdorff measure of $\Gamma \subset \mathbb{R}^d$, and $\mathcal{L}(U)$ the d -dimensional Lebesgue measure of $U \subset \mathbb{R}^d$ [57, §2].

(ii) Theorem 3 immediately implies that for $\Gamma \subset \partial\Omega$, $\mathcal{H}^{d-1}(\Gamma) > 0$, there is $C > 0$ such that

$$\|f\|_{L^2(\Omega)} \leq C \|\nabla f\|_{L^2(\Omega)} \quad \forall f \in H^1(\Omega), \quad f|_{\Gamma} = 0.$$

2.2 Helmholtz equation

Let u denote the wave field satisfying the Helmholtz equation

$$-\nabla \cdot (a(x) \nabla u(x)) - k^2(x) u(x) = f(x), \quad x \in \Omega, \quad (2.9a)$$

$$a(x) \frac{\partial u(x)}{\partial n} - i\sqrt{a(x)}k(x)u(x) = g_S(x), \quad x \in \Gamma_S, \quad (2.9b)$$

$$a(x) \frac{\partial u(x)}{\partial n} = g_N(x), \quad x \in \Gamma_N, \quad (2.9c)$$

$$u(x) = g_D(x), \quad x \in \Gamma_D. \quad (2.9d)$$

Here $k(x) = \omega/c(x)$ denotes the wave number, $\omega > 0$ the frequency, $c \in L^\infty(\Omega)$ the propagation speed of the medium with $c(x) \geq c_{\min} > 0$, $\rho \in L^\infty(\Omega)$ the density with $\rho(x) \geq \rho_{\min} > 0$, and $a(x) = 1/\rho(x)$. Moreover, n denotes the unit outward normal and $f \in L^2(\Omega)$, $g_N \in L^2(\Gamma_N)$, $g_S \in L^2(\Gamma_S)$, and $g_D \in H^{1/2}(\Gamma_D)$ are known sources, which may vanish.

Next, we consider the variational formulation of (2.9): find a (weak-) solution $u \in H^1(\Omega)$ satisfying $u = g_D$ on Γ_D and

$$A[u, v] = F[v], \quad \forall v \in H_D^1, \quad (2.10)$$

where

$$A[u, v] = \int_{\Omega} a \nabla u \cdot \nabla \bar{v} \, dx - \int_{\Omega} k^2 u \bar{v} \, dx - i \int_{\Gamma_S} \sqrt{a} k u \bar{v} \, ds, \quad (2.11)$$

$$F[v] = \int_{\Omega} f \bar{v} \, dx + \int_{\Gamma_N} g_N \bar{v} \, ds + \int_{\Gamma_S} g_S \bar{v} \, ds, \quad (2.12)$$

$$H_D^1 = \{w \in H^1(\Omega) \mid w = 0 \text{ on } \Gamma_D\}. \quad (2.13)$$

Since $A[u, v]$ is in general not coercive, one cannot use the the Lax-Milgram Theorem to show that (2.10) is well-posed.

2.3 Existence and uniqueness

The solution of (2.9) in general may neither exist nor be unique in $H^1(\Omega)$ when the homogeneous Helmholtz equation with $f \equiv g_S \equiv g_N \equiv g_D \equiv 0$ has a non-trivial solution in $H^1(\Omega)$.

For space dimension $d = 1$, the well-posedness of (2.9) has been discussed in [14, 5, 49]. For $d = 2, 3$, $a \equiv 1$, c constant (e.g. $c = 1$), $k = \omega$, and $\Gamma = \Gamma_S$, (2.9) (or (2.10)) has a unique (weak-) solution for any $f \in H^{-1}(\Omega)$, $g_S \in H^{-\frac{1}{2}}(\Gamma)$ depending continuously on the data [109, Proposition §8.1.3]. Later Cummings et al. showed in [40] the stability for a star-shaped Ω with respect to a ball. In [93], Hetmaniuk showed that the Helmholtz equation (2.9) with $a \equiv 1$ and $g_D \equiv g_N \equiv 0$ has a unique weak solution $u \in H^1(\Omega)$ which satisfies the stability condition: there is a constant $C > 0$ independent of u , f , and g_S such that

$$k \|u\|_{L^2(\Omega)} + \|\nabla u\|_{L^2(\Omega)} \leq C (\|f\|_{L^2(\Omega)} + \|g_S\|_{L^2(\Gamma_S)}).$$

For $a \equiv 1$ and $k \in L^\infty(\Omega)$ with $0 < k_{\min} \leq k(x) \leq k_{\max}$, Dörfler et al. proved the uniqueness and existence of sound-soft scattering problems [48, Proposition §3.2]. Recently, Graham et al. extended the theorem in [73] to an exterior Dirichlet problem, where they assumed that a, c are Lipschitz. In [74], Sauter and Graham proved, using the unique continuation principle, the existence and uniqueness of (2.9) with variable coefficients $a, c \in L^\infty(\Omega)$, which are bounded away from zero, for $d \leq 2$; however, for $d \geq 3$, they required that a is in addition Lipschitz continuous.

Now, we summarize the result in the following theorem (see [74, Theorem 2.1] and [125, Theorem 1]):

Theorem 4. *Let $\Omega \subset \mathbb{R}^d$, $d = 1, 2, 3$, be a bounded Lipschitz domain and $a, c \in L^\infty(\Omega)$ with $0 < a_{\min} \leq a(x) \leq a_{\max}$ and $0 < c_{\min} \leq c(x) \leq c_{\max}$. In addition, $a \in C^{0,1}(\bar{\Omega})$ for $d \geq 3$. Then, (2.9) has a unique weak solution in $H^1(\Omega)$.*

In this thesis we assume that the solution of (2.9) (or (2.10)) exists. Again, it is known that the solution of (2.9) is not unique if, and only if, the homogeneous Helmholtz equation with $f \equiv g_S \equiv g_D \equiv g_N \equiv 0$, which coincides with the eigenvalue

problem (2.14), has a non-trivial solution v :

$$L[v] := -\nabla \cdot (a(x) \nabla v(x)) = k^2(x) v(x) = \frac{\omega^2}{c^2(x)} v(x), \quad x \in \Omega, \quad (2.14a)$$

$$\sqrt{a(x)} \frac{\partial v(x)}{\partial n} = ik(x) v(x) = i \frac{\omega}{c(x)} v(x), \quad x \in \Gamma_S, \quad (2.14b)$$

$$\frac{\partial v(x)}{\partial n} = 0, \quad x \in \Gamma_N, \quad (2.14c)$$

$$v(x) = 0, \quad x \in \Gamma_D. \quad (2.14d)$$

Assume that a non-trivial solution v of (2.14) exists. Then ω^2 is called an eigenvalue of L and v a corresponding eigenfunction of L in $H_D^1(\Omega)$. The spectrum Σ_L of L denotes the set of all eigenvalues. The theory of Fredholm in [56, §6.2] implies that Σ_L is real, nonnegative, and at most countable. If $\Sigma_L = \{\omega_\ell^2\}_{\ell=0}^\infty$ is infinite, then ω_ℓ^2 tends to $+\infty$ as $\ell \rightarrow +\infty$.

Let us consider (2.14) with $\omega = 0$. Then (2.14) becomes a Laplacian-type elliptic equation,

$$\nabla \cdot (a(x) \nabla v(x)) = 0, \quad \text{in } \Omega, \quad (2.15a)$$

$$v(x) = 0, \quad x \in \Gamma_D, \quad (2.15b)$$

$$\frac{\partial v(x)}{\partial n} = 0, \quad x \in \Gamma_N \cup \Gamma_S. \quad (2.15c)$$

A function $v \in H^1(\Omega)$ satisfies (2.15) if, and only if, v is constant almost everywhere. That immediately follows from the fact that ∇v is identically zero in Ω , since

$$0 = B[v, v] = \int_{\Omega} a |\nabla u|^2 dx \geq a_{\min} \int_{\Omega} |\nabla u|^2 dx$$

and $a_{\min} > 0$. Moreover, v is also identically zero when $\mathcal{H}^{d-1}(\Gamma_D) > 0$.

Now, we summarize the result in the following proposition.

Proposition 1. *Let $v \in H_D^1$ be a (weak-) solution of (2.15). Then v is constant a.e. in Ω . Furthermore, if $\mathcal{H}^{d-1}(\Gamma_D) > 0$, v vanishes throughout Ω .*

2.4 Finite element discretization

Here we apply the standard H^1 -conforming finite element (FE) method on (2.10) based on the references [24, 122].

Let $\{\mathcal{T}_h\}_{h>0}$ be a family of shape-regular grids of the domain Ω with $h = \max_{K \in \mathcal{T}_h} h_K$, where h_K is the diameter of the element K in \mathcal{T}_h . Then we define

$$\mathcal{P}_h^r = \{v \in C^0(\overline{\Omega}) \mid v|_K \in \mathbb{P}^r, \forall K \in \mathcal{T}_h\},$$

where \mathbb{P}^r denotes the space of polynomials with degree lower than or equal to r . Next, we consider

$$\mathcal{V}_h = \text{Span}\{\varphi_j\}_{j=1}^{ndofs} \subset \mathcal{P}_h^r, \quad ndofs = \dim(\mathcal{V}_h) < +\infty.$$

For $g_D = 0$ on Γ_D or $\mathcal{H}^{d-1}(\Gamma_D) = 0$, we now look for a discrete solution $u_h \in \mathcal{V}_h$ which satisfies

$$A[u_h, v_h] = F[v_h], \quad \forall v_h \in \mathcal{V}_h \cap H_D^1. \quad (2.16)$$

Remark 2. In practice, for $g_D \neq 0$ and $\mathcal{H}^{d-1}(\Gamma_D) > 0$, we first find a lifting function $R_D \in H^1(\Omega)$ of boundary data [122, §3.3.3], with

$$R_D = g_D \quad \text{on } \Gamma_D.$$

Then instead of the solution u of (2.9), we consider the solution $u_0 = u - R_D$ of the “modified” Helmholtz equation

$$-\nabla \cdot (a \nabla u_0) - k^2 u_0 = f + \nabla \cdot (a \nabla R_D) + k^2 R_D, \quad x \in \Omega, \quad (2.17a)$$

$$a \frac{\partial u_0}{\partial n} - i\sqrt{a} k u_0 = g_S - a \frac{\partial R_D}{\partial n} + i\sqrt{a} k R_D, \quad x \in \Gamma_S, \quad (2.17b)$$

$$a \frac{\partial u_0}{\partial n} = g_N - a \frac{\partial R_D}{\partial n}, \quad x \in \Gamma_N, \quad (2.17c)$$

$$u_0 = 0, \quad x \in \Gamma_D. \quad (2.17d)$$

As a consequence, the solution u of (2.9) is given by

$$u = u_0 + R_D.$$

For simplicity, we shall henceforth assume that $g_D = 0$ on Γ_D when $\mathcal{H}^{d-1}(\Gamma_D) > 0$.

Next, we define

$$\begin{aligned} (\mathbf{M}_h)_{j,\ell} &= \int_{\Omega} \frac{1}{c^2(x)} \varphi_{\ell}(x) \varphi_j(x) dx, & (\mathbf{S}_h)_{j,\ell} &= \int_{\Gamma_S} \frac{\sqrt{a(x)}}{c(x)} \varphi_{\ell}(x) \varphi_j(x) ds, \\ (\mathbf{K}_h)_{j,\ell} &= \int_{\Omega} a(x) \nabla \varphi_{\ell}(x) \cdot \nabla \varphi_j(x) dx, & (\mathbf{F}_h)_j &= F[\varphi_j]. \end{aligned} \quad (2.18)$$

Then the discrete linear operator of $A[\cdot, \cdot]$ is given by

$$\mathbf{A}_h = -\omega^2 \mathbf{M}_h + \mathbf{K}_h - i\omega \mathbf{S}_h,$$

and (2.16) becomes a linear system

$$\mathbf{A}_h \mathbf{u}_h = \mathbf{F}_h, \quad (2.19)$$

where the solution $\mathbf{u}_h \in \mathcal{P}_h^r$ of (2.19) represents the coefficients of u_h with respect to the basis functions $\{\varphi_j\}$. From [6, 7, 111, 48], when the numerical error between u_h and u is bounded by the constant $C > 0$ so that the (quasi-) optimal estimate

$$\|u - u_h\|_{L^2(\Omega)} \leq Ch^{r+1} \quad (2.20)$$

holds true, the constant C depends continuously on the constant wave number k with

$$C \sim k^{r+2}, \quad (2.21)$$

in particular, for $r = 1$ in [14],

$$C \sim k^3.$$

In practice, for a reasonably accurate solution using \mathcal{P}^1 -FE one requires the rule-of-thumb [87, 20] of

$$kh \approx \frac{2\pi}{10} \quad (2.22)$$

coinciding with at least 10 grid points per wave length,

$$\lambda = \frac{2\pi}{k}.$$

However, the numerical error (2.20) with $hk \sim \text{const}$ increases with growing k due to (2.21). To avoid the so-called wave phase or pollution error arising from increasing k , one instead needs the (stricter) condition

$$k(hk)^{r+1} \sim \text{const} \quad (2.23)$$

for a large wave number k or low-order r [13, 6]. Moreover, to obtain quasi-optimal convergence, one requires

$$k(hk)^r \sim \text{const}, \quad (2.24)$$

see [6, 7].

2.5 Standard numerical methods

In compare to iterative methods, a classical direct solver akin to the LU decomposition (or factorization) method based on Gauss elimination is of advantage to solve linear system of non-Hermitian, ill-conditioned matrix. It decomposes the system matrix \mathbf{A}_h , resulting from (2.19), into a lower triangular \mathbf{L}_h and an upper triangular \mathbf{U}_h matrix [72, §3.2], namely

$$\mathbf{A}_h = \mathbf{L}_h \mathbf{U}_h. \quad (2.25)$$

One solves (2.19) in two substitution steps:

$$\begin{aligned} \mathbf{L}_h \mathbf{v}_h &= \mathbf{F}_h && \text{(forward substitution),} \\ \mathbf{U}_h \mathbf{u}_h &= \mathbf{v}_h && \text{(backward substitution).} \end{aligned}$$

For a large frequency k with (2.24), solving (2.19), resulting e.g. from \mathcal{P}^r -FE, with the direct solver typically becomes too expensive in both computational cost and storage requirement [55, 20, 61]:

- computational cost of $\mathcal{O}(N^{3/2})$ and storage requirement of $\mathcal{O}(N \log N)$ for $d = 2$
- computational cost of $\mathcal{O}(N^2)$ and storage requirement of $\mathcal{O}(N^{4/3})$ for $d = 3$

Moreover, classical direct solvers are hard to parallelize.

One instead uses e.g. iterative methods, which require less storage requirement, with an efficient preconditioner such as the “shifted-Laplacian”, “incomplete LU decomposition”, and “domain decomposition”.

The Schwarz based domain decomposition (DD) method gained popularity in the last decades as a preconditioner for solving elliptic problems. The alternating Schwarz procedure,

$$\left\{ \begin{array}{lll} -\Delta u_1^m & = & f \quad \text{in } \Omega_1, \\ u_1^m & = & g \quad \text{on } \partial\Omega_1 \setminus \Gamma, \\ u_1^m & = & u_2^{m-1} \quad \text{on } \Gamma, \end{array} \right. \quad \left\{ \begin{array}{lll} -\Delta u_2^m & = & f \quad \text{in } \Omega_2, \\ u_2^m & = & g \quad \text{on } \partial\Omega_2 \setminus \Gamma, \\ u_2^m & = & u_1^m \quad \text{on } \Gamma, \end{array} \right.$$

was originally proposed by Schwarz in 1870 [126] at the continuous level to prove the solvability of the Dirichlet equation. Here he decomposed the physical domain Ω into two subdomains Ω_1 and Ω_2 with overlaps – see Figure 2.2a.

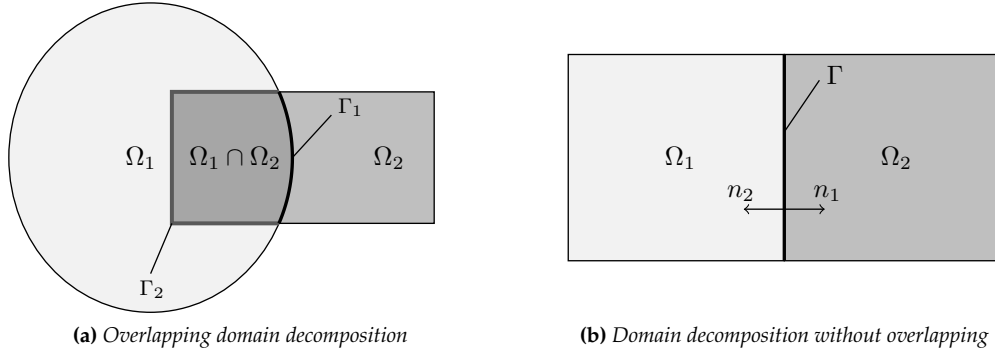


Figure 2.2: (Overlapping) domain decomposition

In 1965, Miller first proposed in [112] to use the alternating Schwarz iteration as a numerical method to solve Poisson-like problems. Later, in [105], Lions introduced a parallel Schwarz approach for parallel computing.

In the recent years, a growing number of domain decomposition based preconditioners has been proposed for the iterative solution of the Helmholtz equation [55, 47],

$$\left\{ \begin{array}{ll} -(\Delta + k^2)u_1^m &= f \quad \text{in } \Omega_1, \\ u_1^m &= g \quad \text{on } \partial\Omega_1 \setminus \Gamma_1, \\ u_1^m &= u_2^{m-1} \quad \text{on } \Gamma_1, \end{array} \right. \quad \left\{ \begin{array}{ll} -(\Delta + k^2)u_2^m &= f \quad \text{in } \Omega_2, \\ u_2^m &= g \quad \text{on } \partial\Omega_2 \setminus \Gamma_2, \\ u_2^m &= u_1^m \quad \text{on } \Gamma_2. \end{array} \right.$$

However, this DD method without overlaps and with the Dirichlet (or Neumann) transmission conditions at the interfaces Γ , shown Figure 2.2b, will not converge. Therefore, a new Robin transmission conditions were proposed for the DD method for solving the Helmholtz [17, 65]. Then the corresponding algorithm is given by

$$\left\{ \begin{array}{ll} -(\Delta + k^2)u_1^m &= f \quad \text{in } \Omega_1, \\ u_1^m &= g \quad \text{on } \partial\Omega_1 \cap \partial\Omega, \\ \left(\frac{\partial}{\partial n_1} + \alpha \right) (u_1^m) &= \left(\frac{\partial}{\partial n_1} + \alpha \right) (u_2^{m-1}) \quad \text{on } \Gamma, \end{array} \right.$$

and

$$\left\{ \begin{array}{ll} -(\Delta + k^2)u_2^m &= f \quad \text{in } \Omega_2, \\ u_2^m &= g \quad \text{on } \partial\Omega_2 \cap \partial\Omega, \\ \left(\frac{\partial}{\partial n_2} + \alpha \right) (u_2^m) &= \left(\frac{\partial}{\partial n_2} + \alpha \right) (u_1^{m-1}) \quad \text{on } \Gamma, \end{array} \right.$$

for a positive number α .

To achieve a convergence rate independent of the number of subdomains, one combines the DDM method with a coarse grid correction [64, 47].

In this thesis, we apply the direct solver based on the LU decomposition (2.25) to solve (2.19) when $N > 0$ is sufficiently small. Otherwise, we use the generalized minimal residual (GMRES) method [124], combined with the one- or two-level Optimized Restricted Additive Schwarz (ORAS) preconditioner [47] with overlap, to solve (2.19).

CHAPTER 3

Controllability methods for solving Helmholtz problems

Here we consider an alternative method using the controllability technique [104] for the Helmholtz problem. The approach is to solve the Helmholtz equation (2.9) in the time domain instead of the frequency domain, which indicates the (time-harmonic) wave equation (1.10). Then, the corresponding time-harmonic wave field y is given by

$$y(x, t) = \operatorname{Re}\{u(x) e^{-i\omega t}\}, \quad (3.1)$$

which directly yields the solution u of the Helmholtz equation, namely

$$u(x) = y(x, 0) + \frac{i}{\omega} \frac{\partial}{\partial t} y(x, 0) = y_0(x) + \frac{i}{\omega} y_1(x).$$

However, the initial values (y_0, y_1) of (1.10) are unknown. To find a time-harmonic solution, Bristeau et al. showed in [25, 26] that for sound-soft scattering problems one only requires the (unique) time-periodic solution of the wave equation with

$$y(x, T) = y(x, 0), \quad \frac{\partial}{\partial t} y(x, T) = \frac{\partial}{\partial t} y(x, 0), \quad x \in \Omega,$$

and the known period $T = 2\pi/\omega$. The uniqueness implies that y fulfills (3.1). To determine the unknown initial values, they formulated the scattering problem as a PDE-constrained least-squares problem.

First, we introduce some preliminaries about functions in the time-dependent (Sobolev) space and some references to the analytic background. Second, we consider the wave equation and some fundamental analysis for the CM approach such as the uniqueness of the solution. For solving the time integration of a wave equation, we derive the semi-discrete problem of the wave equation using a Galerkin approach and the finite element (FE) discretization. Then, we precisely describe some classical numerical methods for solving the semi-discrete time dependent wave equation. In the presence of local mesh refinement, we combined the CM with high-order explicit local time-stepping (LTS) methods [78, 81] to overcome the bottleneck due to an overly stringent CFL stability constraint. Third, we provide some analytical theory about time-periodic solutions of the wave equation, which directly yields the time-harmonic solution with additional conditions or post-processing. The analysis is based on the Fourier theory, which shows that a time-periodic solution is a superposition of various eigenmodes. We will discuss how to recover the solution from the superposition. Next, for solving the Helmholtz equation, we formulate the quadratic least-squares problem and finally apply the conjugate gradient (CG) method to the optimization problem. Following [11, 83, 80], the original CM approach, which was proposed for solving sound-soft scattering problems,

however, may fail for solving general boundary value problems governed by the Helmholtz equation. Therefore, we will discuss how to extend the CMCG method for Helmholtz problems in more general settings [83, 80]. To accelerate the convergence of the CG method, based on [114], we show how to find an accurate initial guess for the CG method. Finally, we show a series of numerical experiments, obtained with the CMCG method, to verify the usefulness and robustness of the method. To illustrate the accuracy and efficiency of both the original and the modified (or extended) CMCG method, we compare the original CMCG method with the extended CMCG method and with the direct solver. Then, we present series of numerical results to demonstrate the parallel performances of the HPC controllability method.

3.1 Analytical background

Here we introduce some analytical backgrounds based on §5.9.2 in [56], which are essential to describe Sobolev spaces involving time and to prove the existence and uniqueness of the solution. We first consider function spaces, which include functions mapping the time into Banach or even Hilbert spaces.

Definition 2. Let V be a Banach space equipped with the norm $\|\cdot\|_V$. Then

(i) the function space

$$C^r([0, T]; V), \quad r \geq 0, \quad (3.2)$$

consists of all r -times continuously differentiable functions $y : [0, T] \rightarrow V$ with

$$\|y\|_{C^r([0, T]; V)} := \max_{t \in [0, T]} \|y(\cdot, t)\|_V < +\infty,$$

(ii) and the (Sobolev) function space

$$L^p([0, T]; V), \quad p \geq 0, \quad (3.3)$$

consists of all measurable functions $y : [0, T] \rightarrow V$, with

$$\|y\|_{L^2([0, T]; V)} := \left(\int_0^T \|y(t)\|_V^2 dt \right)^{1/2} < +\infty.$$

Definition 3. Let $y, z \in L^p([0, T]; V)$. Then z is the weak derivative of y , written $y_t = z$ or $\frac{\partial}{\partial t} y = z$, provided

$$\int_0^T y(\cdot, t) \psi'(t) dt = - \int_0^T z(\cdot, t) \psi(t) dt, \quad \forall \psi \in C_c^1([0, T]).$$

Moreover, y belongs to the Sobolev space

$$H^1([0, T]; V), \quad (3.4)$$

if y_t exists and

$$\|y\|_{H^1([0, T]; V)} := \left(\|y\|_{L^2([0, T]; V)}^2 + \|y_t\|_{L^2([0, T]; V)}^2 \right)^{1/2} < +\infty.$$

Lemma 1. Let $y \in H^1([0, T]; H^1(\Omega)) \cap L^2([0, T]; L^2(\Omega))$ and $V \subset H^1(\Omega)$ be a subset. Then, for a fixed $\varphi \in V$, the mapping,

$$t \mapsto \int_{\Omega} y(x, t) \varphi(x) dx,$$

is absolutely continuous in time $t \in [0, T]$ after possibly being redefined on a set of measure zero. For $t = 0$ (or $t = T$) we take the left (or right) limit.

Proof. According to Theorem 2 (i)-(ii) from [56, §5.9.2] and the fundamental theorem of Lebesgue integral calculus (see [57, Theorem §1.30]), the mapping

$$t \mapsto y(\cdot, t)$$

is absolutely continuous with respect to the L^2 -norm after possibly being redefined on a set of measure zero. Let $\varphi \in V$ and $\{t_\ell\}_{\ell=1}^\infty \subset [0, T]$ with

$$t_\ell \rightarrow t \in [0, T], \quad \ell \rightarrow +\infty.$$

Then the continuity of the mapping $t \mapsto y(\cdot, t)$ and the Cauchy-Schwarz inequality yield

$$\left| \int_{\Omega} (y(x, t_\ell) - y(x, t)) \varphi(x) dx \right| \leq \|y(\cdot, t_\ell) - y(\cdot, t)\|_{L^2(\Omega)} \|\varphi\|_{L^2(\Omega)} \rightarrow 0, \quad \ell \rightarrow +\infty,$$

which shows the continuity of $(y(\cdot, t), \varphi)_{L^2(\Omega)}$ and proves the first assertion.

Following Theorem 2 in §5.9.2 in [56] by replacing $\|y(\cdot, t)\|^2$ with $(y(\cdot, t), \varphi)$, it holds

$$\begin{aligned} \int_{\Omega} y(x, t) \varphi(x) dx - \int_{\Omega} y(x, s) \varphi(x) dx &= \int_{\Omega} (y(x, t) - y(x, s)) \varphi(x) dx \\ &= \int_{\Omega} \int_s^t \frac{\partial y(x, \tau)}{\partial \tau} \varphi(x) d\tau dx \\ &= \int_s^t \frac{\partial}{\partial \tau} \left[\int_{\Omega} y(x, \tau) \varphi(x) dx \right] d\tau. \end{aligned}$$

This yields the absolute continuity and completes the lemma. □

3.2 Time-dependent wave equation

Let y denote the solution of the wave equation,

$$\frac{1}{c^2(x)} \frac{\partial^2 y(x, t)}{\partial t^2} - \nabla \cdot (a(x) \nabla y(x, t)) = \operatorname{Re}\{f(x) e^{-i\omega t}\}, \quad x \in \Omega, t > 0, \quad (3.5a)$$

$$a(x) \frac{\partial y(x, t)}{\partial n} + \frac{\sqrt{a(x)}}{c(x)} \frac{\partial y(x, t)}{\partial t} = \operatorname{Re}\{g_S(x) e^{-i\omega t}\}, \quad x \in \Gamma_S, t > 0, \quad (3.5b)$$

$$a(x) \frac{\partial y(x, t)}{\partial n} = \operatorname{Re}\{g_N(x) e^{-i\omega t}\}, \quad x \in \Gamma_N, t > 0, \quad (3.5c)$$

$$y(x, t) = \operatorname{Re}\{g_D(x) e^{-i\omega t}\}, \quad x \in \Gamma_D, t > 0, \quad (3.5d)$$

$$y(x, 0) = y_0, \quad x \in \Omega, \quad (3.5e)$$

$$y_t(x, 0) = y_1, \quad x \in \Omega, \quad (3.5f)$$

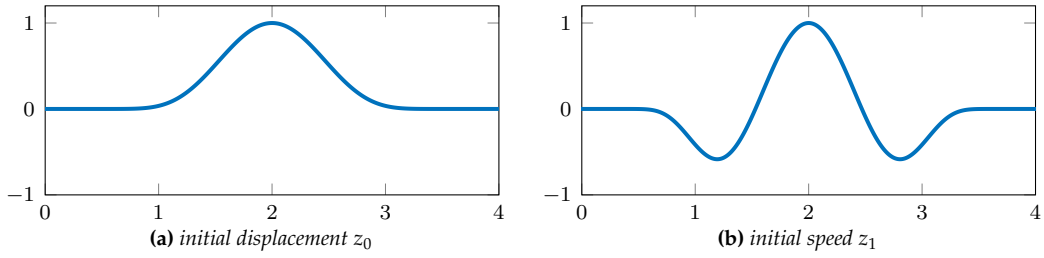


Figure 3.1: Long-time behavior: initial values z_0 and z_1 of (3.5)

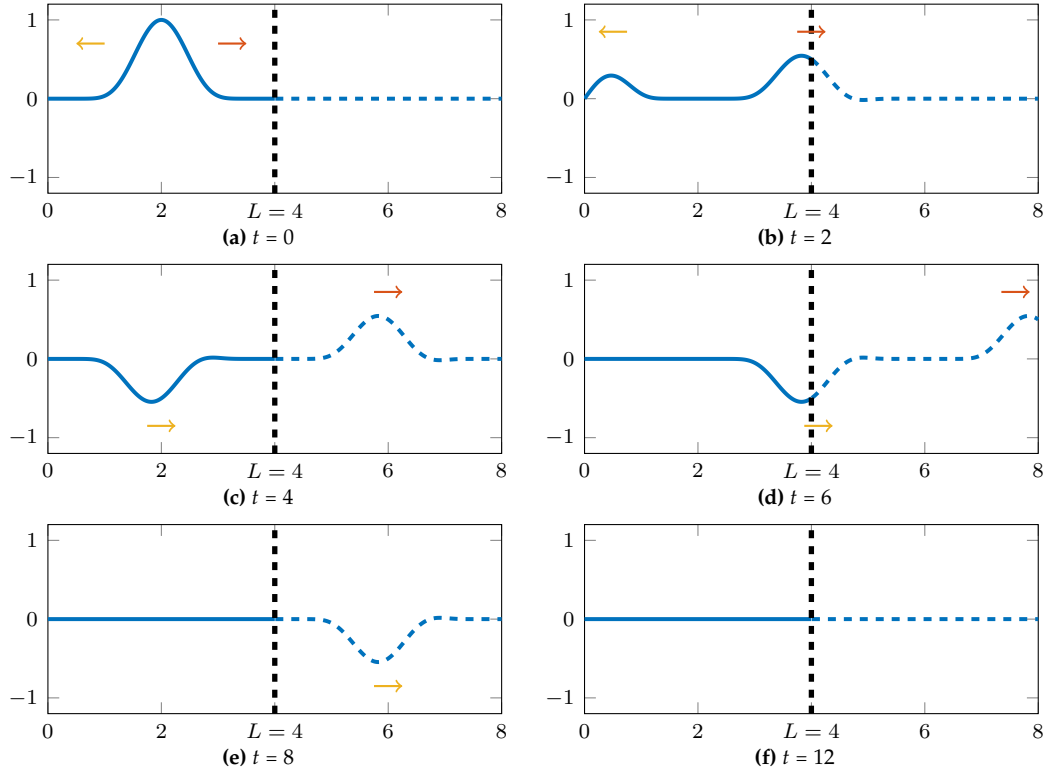


Figure 3.2: Long-time behavior: snapshots of $z(\cdot, t)$ in $\Omega = (0, L)$, $L = 4$, at time $t = 0, 2, 4, 6, 8$, and 12 .

for the (unknown) initial values (y_0, y_1) .

Suppose that y is time-harmonic, namely $y(x, t) = \text{Re}\{u(x)e^{-i\omega t}\}$. Then

$$y_0(x) = \text{Re}\{u(x)\}, \quad y_1(x) = \omega \text{Im}\{u(x)\} \quad (3.6)$$

which implies that

$$u(x) = y_0(x) + \frac{i}{\omega} y_1(x). \quad (3.7)$$

3.2.1 Asymptotic solution of the time-harmonic wave equation

Here we consider the second-order hyperbolic differential equation (3.5) in Ω , $\Gamma = \partial\Omega$ smooth, with $a \equiv c \equiv 1$ in the first-order formulation

$$\frac{\partial}{\partial t} Y(t) = F e^{-i\omega t} + B Y(t), \quad (3.8)$$

with the linear operator

$$B = \begin{pmatrix} 0 & \text{id} \\ \Delta & 0 \end{pmatrix}$$

and

$$Y(t) = (y, y_t)^\top, \quad F = (0, f)^\top, \quad Y(0) = Y_0 \in H^1(\Omega) \times L^2(\Omega).$$

Following [11, 96],

$$Y(x, t) = U(x) e^{-i\omega t} + Z(t)(U_0(x) - U(x))$$

solves (3.8), where

$$Z(t) = \exp(tB), \quad U(x) = (i\omega - B)^{-1}F(x).$$

Furthermore, $Z(t)$ tends to zero when t tends to infinity. Consequently, we obtain the asymptotic behavior

$$y(x, t) \sim u(x) e^{-i\omega t}, \quad t \rightarrow +\infty,$$

and hence

$$y(x, mT) + \frac{i}{\omega} y_t(x, mT) \sim u(x) e^{-i\omega mT} = u(x), \quad m \rightarrow +\infty,$$

where u solves (2.9). We write $y(x, t) = u(x) e^{-i\omega t} + z(x, t)$. Here z denotes a perturbation, which satisfies (3.5) with $f = g_S = g_N = g_D = 0$ and tends to zero when $t \rightarrow +\infty$.

Let us consider a one-dimensional solution z of (3.5) with $a = c = 1$, $f = 0$ in $\Omega = (0, L)$, $L > 0$, and

$$z(x, 0) = z_0(x), \quad z_t(x, 0) = z_1(x), \quad x \in \Omega.$$

Here we impose that $\Gamma_D = \{0\}$, $\Gamma_S = \{L\}$, $z_0, z_1 \in C^\infty(\mathbb{R})$ with $\text{supp}(z_0), \text{supp}(z_1) \subset (0, L)$. Following [56], the formula of D'Alembert with the extension

$$z_0(-x) = -z_0(x), \quad z_1(-x) = -z_1(x), \quad x \leq 0,$$

implies that

$$z(x, t) = \begin{cases} \frac{1}{2}(z_0(x+t) + z_0(x-t)) + \frac{1}{2} \int_{x-t}^{x+t} z_1(\tau) d\tau, & 0 \leq t \leq x, \\ \frac{1}{2}(z_0(x+t) - z_0(t-x)) + \frac{1}{2} \int_{-x+t}^{x+t} z_1(\tau) d\tau, & 0 \leq x \leq t. \end{cases}$$

Clearly, $z(x, t) = 0$ for $t \geq 2L$ and $x \in \Omega$.

To illustrate how the perturbation z vanishes identically, we consider the example in $\Omega = (0, L)$, $L = 4$, with

$$z_0(x) = \begin{cases} e^{10} e^{\frac{40}{x(x-4)}}, & x \in (0, 4), \\ 0, & \text{otherwise}, \end{cases} \quad z_1(x) = \begin{cases} -\frac{e^{10}}{5} \frac{\partial^2}{\partial x^2} \left(e^{\frac{40}{x(x-4)}} \right), & x \in (0, 4), \\ 0, & \text{otherwise}, \end{cases}$$

where z_0 is a "Gaussian"-like and z_1 a "Ricker wavelet"-like function – see Figure 3.1. Figure 3.2 monitors the solution z of (3.5) at the time step $t = 0, 2, 4, 6, 8$, and 12. The solution splits into two parts. One moves to the left and is reflected back at $x = 0$.

The other propagates to the right and leaves the medium at $x = L = 4$. As expected, the wave vanishes identically the latest at $t = 2L = 8$.

3.2.2 Law of dissipation of energy and uniqueness

Let $E : C^0([0, T]; H^1(\Omega)) \cap C^1([0, T]; L^2(\Omega)) \rightarrow \mathbb{R}$ denote the energy of the wave equation with

$$E[y_0, y_1](t) = \frac{1}{2} \int_{\Omega} a(x) |\nabla y_0(x, t)|^2 dx + \frac{1}{2} \int_{\Omega} \frac{1}{c^2(x)} y_1(x, t)^2 dx. \quad (3.9)$$

It is well-known that the energy of the solution of a homogeneous wave equation is either monotonically decreasing or constant in time, which is summarized in the following proposition.

Proposition 2. *Let y be the solution of the homogeneous wave equation (3.5) with $f \equiv g_S \equiv g_N \equiv g_D \equiv 0$. Then the energy function $E[y, y_t]$ in (3.9) is conservative in time t when $\mathcal{H}^{d-1}(\Gamma_S) = 0$. Otherwise, the energy $E[y, y_t]$ is dissipative when $\mathcal{H}^{d-1}(\Gamma_S) > 0$.*

Proof. Green's formula and (3.5) yield

$$\begin{aligned} \frac{\partial E[y, y_t](t)}{\partial t} &= \int_{\Omega} a(x) \nabla y(x, t) \cdot \nabla y_t(x, t) dx + \int_{\Omega} \frac{1}{c^2(x)} y_{tt}(x, t) y_t(x, t) dx \\ &= \int_{\Omega} a(x) \nabla y(x, t) \cdot \nabla y_t(x, t) dx + \int_{\Omega} y_t(x, t) (\nabla \cdot (a(x) \nabla y(x, t))) dx \\ &= - \int_{\Gamma_S} \frac{\sqrt{a(x)}}{c(x)} y_t(x, t)^2 ds \leq 0, \end{aligned}$$

which clearly shows the proposition. \square

Proposition 2 immediately yields the uniqueness of the solution of (3.5):

Lemma 2. *Assume that the solution $y \in C^0([0, T]; H^1(\Omega)) \cap C^1([0, T]; L^2(\Omega))$ of the wave equation (3.5) exists. Then y is unique (up to a zero measure).*

Proof. Let $y_1, y_2 \in C^0([0, T]; H^1(\Omega)) \cap C^1([0, T]; L^2(\Omega))$ be solutions of (3.5). Then $z := y_1 - y_2 \in C^0([0, T]; H^1(\Omega)) \cap C^1([0, T]; L^2(\Omega))$ solves the homogeneous wave equation, where f, g_D, g_S, g_N and the initial values z_0, z_1 are identically zero. Following Proposition 2, we obtain

$$0 \leq E[z, z_t](t) \leq E[z, z_t](0) = 0, \quad \forall t > 0.$$

This implies, together with the definition (3.9) and the fact that a, c are positive and bounded away from zero, that $\nabla z, z_t$ are identically zero in Ω ; hence z is constant almost everywhere in $\Omega \times [0, T]$. In particular, for $t = 0$, we obtain

$$z(x, t) = z(x, 0) = 0, \quad \forall x, t \in \Omega \times [0, T],$$

which proves the lemma. \square

3.2.3 Variational formulation and discretization of the wave equations

Here we seek $y(t) \in \{w \in H^1(\Omega) \mid w(x) = \operatorname{Re}\{g_D(x) e^{-i\omega t}\}, x \in \Gamma_D\}$, $t \in [0, T]$, such that

$$\tilde{A}[y, z; t] := \frac{\partial^2}{\partial t^2} \int_{\Omega} \frac{1}{c^2} y(t) z dx + \int_{\Omega} a \nabla y(t) \nabla z dx + \frac{\partial}{\partial t} \int_{\Gamma_S} \frac{\sqrt{a}}{c} y(t) z ds = \operatorname{Re}\{F e^{-i\omega t}\},$$

for any $z \in H_D^1$ in (2.13), where F from Equation (2.12). Recalling the definition of (2.18), the semi-discrete wave equation of (3.5) is given by

$$\mathbf{M}_h \frac{\partial^2}{\partial t^2} \mathbf{y}(t) + \mathbf{K}_h \mathbf{y}(t) + \mathbf{S}_h \frac{\partial}{\partial t} \mathbf{y}(t) = \text{Re}\{\mathbf{F}_h e^{-i\omega t}\}. \quad (3.10)$$

Mass-lumping approach

Usually we avoid inverting the mass-matrix \mathbf{M}_h in (3.10) at each time step via order preserving mass-lumping [122, §10] which, however, introduces an additional spatial discretization error. Here we use the bubble function technique described in [36, 39, 67] for \mathcal{P}^r -FE discretizations (see Figures 3.3 and 3.4). The lumped mass matrix \mathbf{M}_h is diagonal and positive definite.

Remark 3. *Alternatively, one can instead use the \mathcal{P}^r -spectral element (SE) method, combined with Gauss-Legendre-Lobatto (GLL) integration formula [122, §10 Spectral methods], for the spatial discretization. For nonconforming FE discretizations, one typically instead uses the discontinuous Galerkin (DG) finite element discretization [123].*

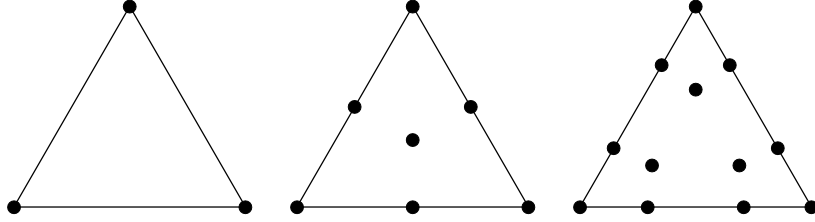


Figure 3.3: Nodes for linear ($r = 1$, left), quadratic ($r = 2$, center), and cubic ($r = 3$, right) polynomials on a triangle

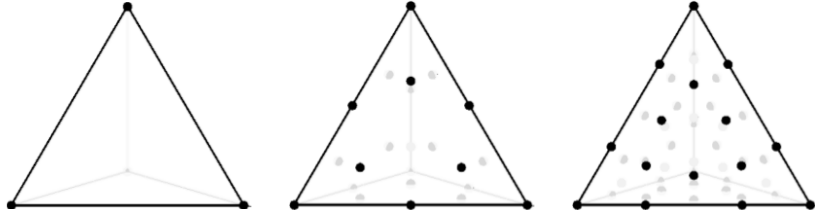


Figure 3.4: Nodes for linear ($r = 1$, left), quadratic ($r = 2$, center), and cubic ($r = 3$, right) polynomials on a tetrahedron

3.2.4 Numerical methods for time integration

Now, we list a series of numerical method for solving the time integration of (3.5).

Leap-Frog (LF) method

By using centered differences to approximate the first and second derivatives in (3.10), we obtain the Leap-Frog time integration,

$$\mathbf{M}_h \frac{\mathbf{y}^{m+1} - 2\mathbf{y}^m + \mathbf{y}^{m-1}}{\Delta t^2} + \mathbf{K}_h \mathbf{y}^m + \mathbf{S}_h \frac{\mathbf{y}^{m+1} - \mathbf{y}^{m-1}}{2\Delta t} = \text{Re}\{\mathbf{F}_h e^{-i\omega t_m}\} \quad (3.11)$$

for solving (3.10), where \mathbf{y}^m denotes the discrete approximation of $\mathbf{y}(t_m)$. The initial variables are given by

$$(\mathbf{y}^0)_j = y(x_j), \quad (\mathbf{y}_t^0)_j = y_t(x_j), \quad x_j \in \Omega_h, \quad 1 \leq j \leq \text{ndofs}.$$

Next, we determine \mathbf{y}^1 with the Taylor expansion and (3.10) with $t = 0$, namely

$$\mathbf{y}^1 = \mathbf{y}^0 + \Delta t \frac{\partial}{\partial t} \mathbf{y} \Big|_{t=0} + \frac{\Delta t^2}{2} \frac{\partial^2}{\partial t^2} \mathbf{y} \Big|_{t=0} = \mathbf{y}^0 + \Delta t \mathbf{y}_t^0 + \frac{\Delta t^2}{2} \mathbf{M}_h^{-1} (\text{Re}\{\mathbf{F}_h\} - \mathbf{K}_h \mathbf{y}^0 - \mathbf{S}_h \mathbf{y}_t^0).$$

Classical Runge-Kutta (RK4) method

Here we transform the second-order wave equation (3.10) into a first-order system of ordinary differential equations:

$$\frac{\partial}{\partial t} \mathbf{Y}_h(t) = \Phi(t, \mathbf{Y}_h(t)), \quad \mathbf{Y}_h = \begin{bmatrix} \mathbf{y} & \frac{\partial}{\partial t} \mathbf{y} \end{bmatrix}^\top, \quad (3.12)$$

$$\Phi(t, \mathbf{Y}_h) = \mathbf{B} \mathbf{Y}_h + \begin{pmatrix} \mathbf{0} \\ \text{Re}\{\mathbf{F}_h e^{-i\omega t}\} \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} \mathbf{0} & \mathbf{I} \\ -\mathbf{M}_h^{-1} \mathbf{K}_h & -\mathbf{M}_h^{-1} \mathbf{S}_h \end{pmatrix}. \quad (3.13)$$

By using the RK4 method for the numerical time integration of (3.12), we have:

$$\begin{aligned} \mathbf{K}_{1,m} &= \Phi(t_m, \mathbf{Y}_h^m) = \mathbf{B} \mathbf{Y}_h^m + \mathbf{R}_m \\ \mathbf{K}_{2,m} &= \Phi(t_m + \frac{\Delta t}{2}, \mathbf{Y}_h^m + \frac{\Delta t}{2} \mathbf{K}_{1,m}) = \mathbf{B}(\mathbf{Y}_h^m + \frac{\Delta t}{2} \mathbf{K}_{1,m}) + \mathbf{R}_{m+\frac{1}{2}} \\ \mathbf{K}_{3,m} &= \Phi(t_m + \frac{\Delta t}{2}, \mathbf{Y}_h^m + \frac{\Delta t}{2} \mathbf{K}_{2,m}) = \mathbf{B}(\mathbf{Y}_h^m + \frac{\Delta t}{2} \mathbf{K}_{2,m}) + \mathbf{R}_{m+\frac{1}{2}} \\ \mathbf{K}_{4,m} &= \Phi(t_m + \Delta t, \mathbf{Y}_h^m + \Delta t \mathbf{K}_{3,m}) = \mathbf{B}(\mathbf{Y}_h^m + \Delta t \mathbf{K}_{3,m}) + \mathbf{R}_{m+1} \\ \mathbf{Y}_h^{m+1} &= \mathbf{Y}_h^m + \frac{\Delta t}{6} (\mathbf{K}_{1,m} + 2\mathbf{K}_{2,m} + 2\mathbf{K}_{3,m} + \mathbf{K}_{4,m}) \end{aligned} \quad (3.14)$$

Local Time-Stepping based on explicit Runge-Kutta methods

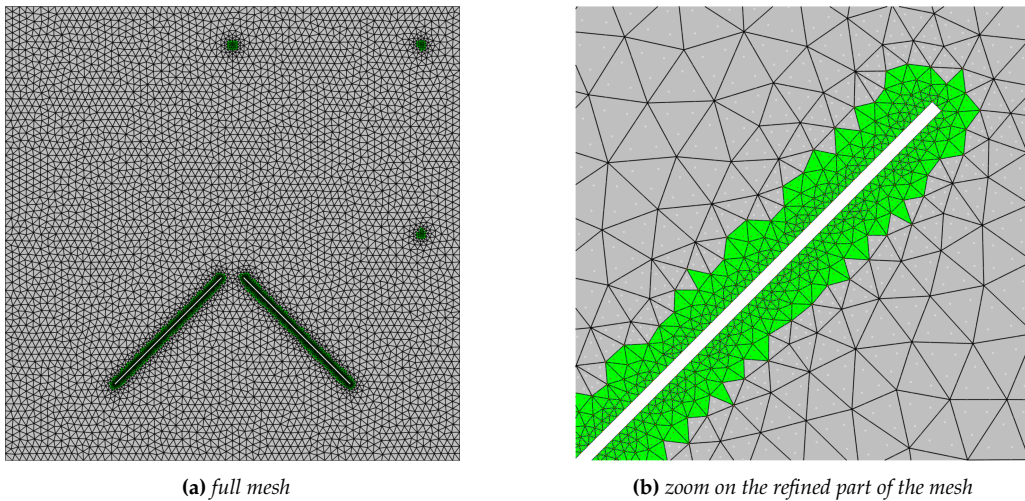


Figure 3.5: Computational mesh with local refinement near the open wedge obstacle; “fine part” is indicated by darker (green) triangles.

Here we consider the local time-stepping methods based on the classical Runge-Kutta scheme – see [78, 81]. The triangular mesh \mathcal{T}_h from the spatial domain Ω has

Algorithm 1: Local Time-Stepping based on RK4

Input: $\mathbf{Y}_0 := \mathbf{Y}_h^m$ initial values, q refinement ratio, Δt time-steps, $\Delta \tau = \frac{\Delta t}{q}$

$\mathbf{w}_0^m = \mathbf{B}[\mathbf{I} - \mathbf{P}]\mathbf{Y}_0 + [\mathbf{I} - \mathbf{P}]\mathbf{R}_m$

$\mathbf{w}_1^m = \mathbf{B}[\mathbf{I} - \mathbf{P}][\mathbf{B}\mathbf{Y}_0 + \mathbf{R}_m] + \frac{1}{\Delta t}[\mathbf{I} - \mathbf{P}][-3\mathbf{R}_m + 4\mathbf{R}_{m+\frac{1}{2}} - \mathbf{R}_{m+1}]$

$\mathbf{w}_2^m = \mathbf{B}[\mathbf{I} - \mathbf{P}][\mathbf{B}^2\mathbf{Y}_0 + \mathbf{B}\mathbf{R}_m + \frac{1}{\Delta t}[-3\mathbf{R}_m + 4\mathbf{R}_{m+\frac{1}{2}} - \mathbf{R}_{m+1}]] + \frac{4}{\Delta t^2}[\mathbf{I} - \mathbf{P}][\mathbf{R}_m - 2\mathbf{R}_{m+\frac{1}{2}} + \mathbf{R}_{m+1}]$

$\mathbf{w}_3^m = \mathbf{B}[\mathbf{I} - \mathbf{P}][\mathbf{B}^3\mathbf{Y}_0 + \mathbf{B}^2\mathbf{R}_m + \frac{1}{\Delta t}\mathbf{B}[-3\mathbf{R}_m + 4\mathbf{R}_{m+\frac{1}{2}} - \mathbf{R}_{m+1}] + \frac{4}{\Delta t^2}[\mathbf{R}_m - 2\mathbf{R}_{m+\frac{1}{2}} + \mathbf{R}_{m+1}]]$

for $\ell = 0, \dots, q-1$ **do**

$\mathbf{K}_{1, \frac{\ell+1}{q}} = \sum_{j=0}^3 \frac{(\ell\Delta\tau)^j}{j!} \mathbf{w}_j^m + \mathbf{B}\mathbf{P}\mathbf{Y}_{\frac{\ell}{q}} + \mathbf{P}\mathbf{R}_{m, \ell}$

$\mathbf{K}_{2, \frac{\ell+1}{q}} = \sum_{j=0}^3 \frac{((\ell+\frac{1}{2})\Delta\tau)^j}{j!} \mathbf{w}_j^m + \mathbf{B}\mathbf{P}[\tilde{\mathbf{Y}}_{\frac{\ell}{q}} + \frac{\Delta\tau}{2}\mathbf{K}_{1, \frac{\ell+1}{q}}] + \mathbf{P}\mathbf{R}_{m, \ell+\frac{1}{2}}$

$\mathbf{K}_{3, \frac{\ell+1}{q}} = \sum_{j=0}^3 \frac{((\ell+\frac{1}{2})\Delta\tau)^j}{j!} \mathbf{w}_j^m + \mathbf{B}\mathbf{P}[\tilde{\mathbf{Y}}_{\frac{\ell}{q}} + \frac{\Delta\tau}{2}\mathbf{K}_{2, \frac{\ell+1}{q}}] + \mathbf{P}\mathbf{R}_{m, \ell+\frac{1}{2}}$

$\mathbf{K}_{4, \frac{\ell+1}{q}} = \sum_{j=0}^3 \frac{((\ell+1)\Delta\tau)^j}{j!} \mathbf{w}_j^m + \mathbf{B}\mathbf{P}[\tilde{\mathbf{Y}}_{\frac{\ell}{q}} + \Delta\tau\mathbf{K}_{3, \frac{\ell+1}{q}}] + \mathbf{P}\mathbf{R}_{m, \ell+1}$

$\mathbf{Y}_{\frac{\ell+1}{q}} = \mathbf{Y}_{\frac{\ell}{q}} + \frac{\Delta\tau}{6}[\mathbf{K}_{1, \frac{\ell+1}{q}} + 2\mathbf{K}_{2, \frac{\ell+1}{q}} + 2\mathbf{K}_{3, \frac{\ell+1}{q}} + \mathbf{K}_{4, \frac{\ell+1}{q}}]$

return $\mathbf{Y}_h^{m+1} = \mathbf{Y}_1$

some local refinements. The idea of this method is to decompose the spatial domain into fine and coarse parts. We denote the diagonal \mathbf{P} as the refinement matrix, where $\mathbf{P}_{\ell\ell} \in \{0, 1\}$ and $\mathbf{P}_{\ell\ell}=1$ if, and only if, x_ℓ lies in the fine part (see Figure 3.5):

$$\mathbf{Z} = (\mathbf{I} - \mathbf{P})\mathbf{Z} + \mathbf{P}\mathbf{Z} = \mathbf{Y}_h^C + \mathbf{Y}_h^F,$$

where \mathbf{Y}_h^C and \mathbf{Y}_h^F are computed separately with different local time steps, $\Delta\tau$ and $\Delta t = q\Delta\tau$, and different techniques based on the Runge-Kutta scheme.

Let h^{coarse} be the smallest mesh size from all coarse elements and h^{fine} is the corresponding smallest mesh size from the fine part. The number of local steps is equal to the local refinement ratio

$$q \approx \frac{h^{\text{coarse}}}{h^{\text{fine}}}. \quad (3.15)$$

We impose $\Delta\tau = \Delta t/q$, the local time-steps. The LTS Algorithm 1 can be found in [78, Section 3.3].

3.3 Time-periodic and time-harmonic solutions of the wave equation

In the controllability approach, we seek an initial value of the wave equation (3.5) such that its solution is time-periodic with a known time period T . For certain situations, a time-periodic solution immediately yields the time-harmonic solution of (2.9). Therefore, we first present some definition and analytical proposition for periodic functions based on the Fourier theory.

Definition 4 (Time periodic function). Let $y \in C(\mathbb{R}; H)$ and H be a Hilbert space. We say y is T -time periodic (or just T -periodic) if

$$y(\cdot, T + t) = y(\cdot, t), \quad \forall t \in \mathbb{R}. \quad (3.16)$$

Now, let $y \in C^0([0, T]; H^1(\Omega)) \cap C^1([0, T]; L^2(\Omega))$ be a solution of (3.5) with

$$y(\cdot, T) = y(\cdot, 0) = y_0, \quad \frac{\partial y(\cdot, T)}{\partial t} = \frac{\partial y(\cdot, 0)}{\partial t} = y_1. \quad (3.17)$$

Then y is T -time periodic in sense that y satisfies (3.5) in $\Omega \times \mathbb{R}$ and (3.16). To show that, we consider the solution y of (3.5) in $[T, 2T]$ and write

$$\tilde{y}(\cdot, t) = y(\cdot, T + t), \quad t \in [0, T].$$

The definition and the condition (3.17) immediately imply that

$$\tilde{y}(\cdot, 0) = y(\cdot, T) = y(\cdot, 0), \quad \frac{\partial \tilde{y}(\cdot, 0)}{\partial t} = \frac{\partial y(\cdot, T)}{\partial t} = \frac{\partial y(\cdot, 0)}{\partial t}.$$

Furthermore, \tilde{y} solves the same wave equation (3.5) as y in $[0, T]$, since all source terms in (3.5) are also T -periodic. The uniqueness lemma 2 yields

$$y(\cdot, T + t) = \tilde{y}(\cdot, t) = y(\cdot, t), \quad \forall t \in [0, T].$$

Therefore, we conclude with a simple “bootstrap argument” that y is T -periodic as its initial value is equal to itself after one period. Now, we summarize this result in the following proposition.

Proposition 3. The solution $y \in C^0(\mathbb{R}; H^1(\Omega)) \cap C^1(\mathbb{R}; L^2(\Omega))$ of the wave equation (3.5) is T -periodic if, and only if, (3.17) is fulfilled.

3.3.1 Time-harmonic and time-periodic wave equation

Here we consider the relation between the time-harmonic solution $\operatorname{Re}\{u(x) e^{-i\omega t}\}$ and the time-periodic solution $y(x, t)$ of the wave equation (3.5). We will see that both solutions are indeed equivalent for sound-soft scattering problems, where we impose both Dirichlet and impedance boundary conditions in (2.9).

We first introduce some basic preliminaries of Fourier series for periodic functions in $C^0([0, T])$. Let

$$\oint_0^T v \, dt = \frac{1}{T} \int_0^T v \, dt \quad (3.18)$$

denote the mean of a function v over the interval $[0, T]$.

Theorem 5 (Fourier’s Theorem). Suppose $f \in C^0([0, T])$ is a T -periodic function. Then there is a complex sequence γ_ℓ such that

$$g_m \rightarrow f \text{ in } L^2\text{-norm, as } m \rightarrow +\infty,$$

where

$$g_m(t) = \sum_{|\ell| \leq m} c_\ell \exp(i\omega \ell t), \quad c_\ell = \oint_0^T f(t) \exp(-i\omega \ell t) \, dt. \quad (3.19)$$

Proof. See Theorem 3.2 in Chapter 4 [127]. □

Corollary 1. Let $f \in C^1([0, T])$. Suppose that f, f' are T -periodic and that g_m is the partial Fourier series of f given by (3.19) with the Fourier coefficients c_ℓ . Then

$$f'(t) = \lim_{m \rightarrow +\infty} g'_m(t), \quad t \in [0, T].$$

Proof. Let $\{d_\ell\} \subset \mathbb{C}$ be the Fourier coefficients of f' and

$$h_m(t) = \sum_{|\ell| \leq m} d_\ell \exp(i\omega \ell t), \quad m \geq 0.$$

Then

$$f'(t) = \lim_{m \rightarrow +\infty} h_m(t),$$

where

$$h_m(t) = \sum_{|\ell| \leq m} d_\ell \exp(i\omega \ell t), \quad d_\ell = \int_0^T f'(t) \exp(-i\omega \ell t) dt.$$

Integration by parts and the time-periodicity of y yield

$$d_\ell = \int_0^T f'(t) \exp(-i\omega \ell t) dt = i\omega \ell \int_0^T f(t) \exp(-i\omega \ell t) dt + \frac{f(T) - f(0)}{T} = i\omega \ell c_\ell.$$

Now, we have

$$h_m(t) = \sum_{|\ell| \leq m} d_\ell \exp(i\omega \ell t) = \sum_{|\ell| \leq m} i\omega \ell c_\ell \exp(i\omega \ell t) = g'_m(t)$$

which completes the proof. \square

Following [11], for sound-soft scattering problems, where we impose both Dirichlet and impedance boundary conditions in (2.9), the T -time periodic solution y of (3.5) immediately yields the unique solution u of (2.9); namely u is given by (3.7). In general, however, a T -time periodic of (3.5) neither is unique nor does it necessarily yield the (unique) time-harmonic solution, as shown in the following theorem.

Theorem 6. Let $u \in H^1(\Omega)$ be the unique solution of the Helmholtz equation (2.9) and $y \in C^0([0, T]; H^1(\Omega)) \cap C^1([0, T]; L^2(\Omega))$ a (real-valued) solution of the wave equation (3.5) with initial conditions $(y_0, y_1) \in H^1(\Omega) \times L^2(\Omega)$. If ∇y and y_t are time periodic with period $T = 2\pi/\omega$, then y admits the Fourier series expansion

$$(y(\cdot, t), \varphi) = (\operatorname{Re}\{u e^{-i\omega t}\}, \varphi) + (\lambda + \eta t, \varphi) + \sum_{|\ell| > 1} (\gamma_\ell e^{-i\omega \ell t}, \varphi) \quad (3.20)$$

for any $\varphi \in H_D^1$, where the constants $\lambda, \eta \in \mathbb{R}$ and the complex-valued (eigen-)functions $\gamma_\ell \in H_D^1$, $|\ell| \geq 2$, satisfy $\overline{\gamma_\ell} = \gamma_{-\ell}$ and

$$-\nabla \cdot (a(x) \nabla \gamma_\ell(x)) = (\ell k(x))^2 \gamma_\ell(x), \quad x \in \Omega, \quad (3.21a)$$

$$\sqrt{a(x)} \frac{\partial \gamma_\ell(x)}{\partial n} = -i\ell k(x) \gamma_\ell(x), \quad x \in \Gamma_S, \quad (3.21b)$$

$$\frac{\partial \gamma_\ell(x)}{\partial n} = 0, \quad x \in \Gamma_N, \quad (3.21c)$$

$$\gamma_\ell(x) = 0, \quad x \in \Gamma_D. \quad (3.21d)$$

Furthermore, if $\mathcal{H}^{d-1}(\Gamma_D \cup \Gamma_S) > 0$, then $\eta = 0$ and if $\mathcal{H}^{d-1}(\Gamma_D) > 0$, then $\lambda = \eta = 0$.

Here (\cdot, \cdot) denotes the standard inner product on $L^2(\Omega)$.

Remark 4. Following (3.20) in Theorem 6, it immediately follows for $t = 0$ that

$$(v, \varphi) = (u, \varphi) + (\lambda + \frac{i}{\omega}\eta, \varphi) + \sum_{|\ell|>1}^{\infty} (\alpha_{\ell} + i\ell\beta_{\ell}, \varphi) \quad \forall \varphi \in H_D^1, \quad (3.22)$$

where $\alpha_{\ell} = \operatorname{Re}\{\gamma_{\ell}\}$ and $\beta_{\ell} = \operatorname{Im}\{\gamma_{\ell}\}$ satisfy

$$-\nabla \cdot (a(x) \nabla \alpha_{\ell}(x)) = (\ell k(x))^2 \alpha_{\ell}(x), \quad x \in \Omega, \quad (3.23a)$$

$$\sqrt{a(x)} \frac{\partial \alpha_{\ell}(x)}{\partial n} = -\ell k(x) \beta_{\ell}(x), \quad x \in \Gamma_S, \quad (3.23b)$$

$$\frac{\partial \alpha_{\ell}(x)}{\partial n} = 0, \quad x \in \Gamma_N, \quad (3.23c)$$

$$\alpha_{\ell}(x) = 0, \quad x \in \Gamma_D, \quad (3.23d)$$

and

$$-\nabla \cdot (a(x) \nabla \beta_{\ell}(x)) = (\ell k(x))^2 \beta_{\ell}(x), \quad x \in \Omega, \quad (3.24a)$$

$$\sqrt{a(x)} \frac{\partial \beta_{\ell}(x)}{\partial n} = \ell k(x) \alpha_{\ell}(x), \quad x \in \Gamma_S, \quad (3.24b)$$

$$\frac{\partial \beta_{\ell}(x)}{\partial n} = 0, \quad x \in \Gamma_N, \quad (3.24c)$$

$$\beta_{\ell}(x) = 0, \quad x \in \Gamma_D. \quad (3.24d)$$

Here we obtain $u = y_0 + (i/\omega)y_1$ precisely when $\lambda = \eta = 0$ and all $\alpha_{\ell}, \beta_{\ell}, |\ell| > 1$, vanish identically.

Proof. For the proof, we distinguish between (i) $\mathcal{H}^{d-1}(\Gamma_D \cup \Gamma_S) > 0$ and (ii) $\Gamma_D \cup \Gamma_S = \emptyset$. (i) For $\mathcal{H}^{d-1}(\Gamma_D \cup \Gamma_S) > 0$, let z denote the difference between $y(x, t)$ and the time-harmonic solution:

$$z(x, t) = y(x, t) - \operatorname{Re}\{u(x) e^{-i\omega t}\}. \quad (3.25)$$

It is easy to verify that z satisfies

$$\frac{1}{c^2(x)} \frac{\partial^2 z(x, t)}{\partial t^2} - \nabla \cdot (a(x) \nabla z(x, t)) = 0, \quad (x, t) \in \Omega \times (0, T), \quad (3.26a)$$

$$\sqrt{a(x)} \frac{\partial z(x, t)}{\partial n} + \frac{1}{c(x)} \frac{\partial z(x, t)}{\partial t} = 0, \quad (x, t) \in \Gamma_S \times (0, T), \quad (3.26b)$$

$$\frac{\partial}{\partial n} z(x, t) = 0, \quad (x, t) \in \Gamma_N \times (0, T), \quad (3.26c)$$

$$z(x, t) = 0, \quad (x, t) \in \Gamma_D \times (0, T). \quad (3.26d)$$

Now, let $\eta(x) = z(x, T) - z(x, 0)$. Since both ∇z and z_t are T -periodic,

$$\nabla[z(\cdot, T) - z(\cdot, 0)] = 0$$

throughout Ω and thus $\eta(x) = \eta$ is constant a.e. in Ω .

If $\mathcal{H}^{d-1}(\Gamma_D) > 0$, η vanishes on Γ_D and hence $\eta = 0$ everywhere in Ω . Otherwise $\Gamma_D = \emptyset$ and thus $\mathcal{H}^{d-1}(\Gamma_S) > 0$. By integrating (3.26a) and using Green's formula together with the homogeneous boundary conditions on Γ_N and Γ_S , we obtain

$$0 = \int_0^T \int_{\Omega} \left(\frac{1}{c^2} z_{tt} - \nabla \cdot (a \nabla z) \right) dx dt = \int_{\Omega} \frac{1}{c^2} z_t \Big|_0^T dx + \int_{\Gamma_S} \frac{\sqrt{a}}{c} z \Big|_0^T ds = \eta \int_{\Gamma_S} \frac{\sqrt{a}}{c} ds,$$

where we have used (3.26b) and the T -periodicity of z_t . Since $a(x), c(x) > 0$ and $\mathcal{H}^{d-1}(\Gamma_S) > 0$, η must equal zero. Therefore, z is T -periodic and for any $\varphi \in H_D^1$, the function $(z(\cdot, t), \varphi) : [0, T] \rightarrow \mathbb{R}$ is also T -periodic and continuous, followed from Lemma 1. Thus, we can expand it in a Fourier series as

$$(z(\cdot, t), \varphi) = \sum_{\ell=-\infty}^{\infty} \tilde{\gamma}_\ell e^{i\omega\ell t}, \quad 0 \leq t \leq T, \quad (3.27)$$

with Fourier coefficients

$$\tilde{\gamma}_\ell = \int_0^T (z(\cdot, t), \varphi) e^{-i\omega\ell t} dt \in \mathbb{C}, \quad \ell \in \mathbb{Z}. \quad (3.28)$$

Now let

$$\gamma_\ell(x) = \int_0^T z(x, t) e^{-i\omega\ell t} dt \in H^1(\Omega), \quad \ell \in \mathbb{Z}. \quad (3.29)$$

Then $\tilde{\gamma}_\ell = (\gamma_\ell, \varphi)$.

Next, we verify that γ_ℓ solves the eigenvalue problem (3.21). First, we replace γ_ℓ on the left side of (3.21) by its definition (3.29), use (3.26a), and integrate by parts twice with respect to time the resulting expression to obtain

$$\begin{aligned} -\nabla \cdot (a(x) \nabla \gamma_\ell(x)) &= - \int_0^T \nabla \cdot (a(x) \nabla z(x, t)) e^{-i\omega\ell t} dt \\ &= -\frac{1}{c^2(x)} \int_0^T z_{tt}(x, t) e^{-i\omega\ell t} dt \\ &= (\ell k(x))^2 \int_0^T z(x, t) e^{-i\omega\ell t} dt \\ &\quad - \frac{i\ell k(x)}{c(x)T} z(x, t) e^{-i\omega\ell t} \Big|_0^T - \frac{1}{c^2(x)T} z_t(x, t) e^{-i\omega\ell t} \Big|_0^T, \quad x \in \Omega. \end{aligned}$$

Since both z and z_t are T -periodic, the last two terms vanish, which yields

$$-\nabla \cdot (a(x) \nabla \gamma_\ell(x)) = (\ell k(x))^2 \int_0^T z(x, t) e^{-i\omega\ell t} dt = (\ell k(x))^2 \gamma_\ell(x), \quad x \in \Omega,$$

by definition (3.29) of γ_ℓ .

We still need to verify that γ_ℓ also satisfies the boundary conditions (3.21b), (3.21c), and (3.21d). As $z(\cdot, t)$ vanishes on Γ_D , so does γ_ℓ since

$$\gamma_\ell(x) = \int_0^T z(x, t) e^{-i\omega\ell t} dt = 0, \quad x \in \Gamma_D.$$

To verify that γ_ℓ satisfies (3.21b) on Γ_S , we start from (3.29) and again use integration by parts in time, the periodicity of z , and (3.26b) to obtain

$$\begin{aligned} \sqrt{a(x)} \frac{\partial \gamma_\ell(x)}{\partial n} &= \int_0^T \sqrt{a(x)} \frac{\partial z(x, t)}{\partial n} e^{-i\omega\ell t} dt = - \int_0^T \frac{1}{c(x)} \frac{\partial z(x, t)}{\partial t} e^{-i\omega\ell t} dt \\ &= -i\ell k(x) \int_0^T z(x, t) e^{-i\omega\ell t} dt - \frac{1}{c(x)T} z(x, t) e^{-i\omega\ell t} \Big|_0^T \\ &= -i\ell k(x) \gamma_\ell(x), \quad x \in \Gamma_S, \end{aligned}$$

by definition (3.29) of γ_ℓ . Similarly, (3.21c) follows from the definition (3.29) and Equation (3.26c). Hence, γ_ℓ indeed satisfies (3.21) for $\ell \in \mathbb{Z}$.

For $\ell = -1$, (3.21) corresponds to (2.9) with $u = \gamma_{-1} = \overline{\gamma_1}$, homogeneous boundary conditions, and no sources. By uniqueness, γ_{-1} , together with its complex conjugate γ_1 , is therefore identically zero. By Proposition 1, γ_0 also vanishes identically.

Next, we use (3.27) in (3.25), together with (3.28) and (3.29), and rearrange terms to obtain

$$(y(\cdot, t), \varphi) = (\operatorname{Re}\{u e^{-i\omega t}\}, \varphi) + (\lambda, \varphi) + \sum_{|\ell| > 1} (\gamma_\ell, \varphi) e^{i\omega \ell t}, \quad \forall \varphi \in H_D^1,$$

which corresponds to (3.20) with $\eta = 0$.

(ii) For $\Gamma_D = \Gamma_S = \emptyset$, that is in a situation of pure Neumann boundary conditions, y is not T -periodic in general and hence

$$\eta(x) = \frac{1}{T} (y(x, T) - y(x, 0))$$

is no longer zero. However, $\nabla \eta(x) \equiv 0$ by assumption and thus η is constant a.e. in Ω . Now, let

$$z(x, t) = y(x, t) - \operatorname{Re}\{u(x) e^{-i\omega t}\} - \eta t.$$

Clearly, z satisfies

$$\begin{aligned} \frac{1}{c^2(x)} \frac{\partial^2 z(x, t)}{\partial t^2} - \nabla \cdot (a(x) \nabla z(x, t)) &= 0, & (x, t) \in \Omega \times [0, T], \\ \frac{\partial z(x, t)}{\partial n} &= 0, & (x, t) \in \Gamma_N \times [0, T], \end{aligned}$$

and is T -periodic by definition of η . Therefore, z again admits a Fourier series expansion with coefficients γ_ℓ which satisfies (3.21). A similar argument as in case (i) concludes the proof. \square

From Theorem 6, we conclude that the initial values y_0 and y_1 of a time-periodic solution are in general not unique whenever λ , η , or γ_ℓ is not identically zero. For sound-soft scattering problems where $\mathcal{H}^{d-1}(\Gamma_S) > 0$ and $\mathcal{H}^{d-1}(\Gamma_D) > 0$, however, the constants λ, η and the eigenfunctions all vanish. We summarize this result in the following corollary – see also [102].

Corollary 2. *Let $u \in H^1(\Omega)$ be the unique solution of (2.9) with $\mathcal{H}^{d-1}(\Gamma_S) > 0$ and $y \in C^0([0, T]; H^1(\Omega)) \cap C^1([0, T]; L^2(\Omega))$ be a solution of (3.5) with initial conditions $(y_0, y_1) \in H^1(\Omega) \times L^2(\Omega)$. If ∇y and y_t are time periodic with period $T = 2\pi/\omega$, then*

$$u = y_0 + \frac{i}{\omega} y_1 + \lambda. \quad (3.31)$$

Furthermore, if $\mathcal{H}^{d-1}(\Gamma_D) > 0$, then

$$u = y_0 + \frac{i}{\omega} y_1.$$

Proof. From Theorem 6 we know that u satisfies the Fourier expansion (3.22) with $\eta = 0$. Since each eigenvalue problem (3.21) corresponds to the Helmholtz equation (2.9) with frequency $\omega \ell$ and $f = g_S = g_D = g_N \equiv 0$, $\ell \neq 0$, which has only the trivial solution, we conclude that $\gamma_\ell = \alpha_\ell = \beta_\ell = 0$, $|\ell| > 1$. Therefore (3.31) holds.

In addition, Theorem 6 implies that $\lambda = 0$ when $\mathcal{H}^{d-1}(\Gamma_D) > 0$, which, together with (3.31), concludes the corollary. \square

3.4 Exact controllability methods

Following Theorem 6 in Section 3.3, the initial values (y_0, y_1) of the time-periodic solution y of (3.5) immediately yield the solution u of Helmholtz equation (2.9) given by (3.7). Now, we discuss how to reformulate (2.9) as a least-squares optimization problem [25, 26] to determine the initial values (y_0, y_1) . Here we consider four kinds of Helmholtz equations:

- (i) sound-soft scattering problems:
 $\mathcal{H}^{d-1}(\Gamma_S) > 0$, $\mathcal{H}^{d-1}(\Gamma_D) > 0$ and Γ_N may be empty
- (ii) sound-hard scattering problems:
 $\mathcal{H}^{d-1}(\Gamma_S) > 0$, $\mathcal{H}^{d-1}(\Gamma_N) > 0$ and $\Gamma_D = \emptyset$
- (iii) scattering from a penetrable inhomogeneous inclusion:
 $\Gamma_S = \partial\Omega$ and $\Gamma_D = \Gamma_N = \emptyset$
- (iv) Helmholtz equation in a physically bounded domain:
 $\Gamma_S = \emptyset$ and Γ_D or Γ_N may vanish

3.4.1 Sound-soft scattering problems

In [25, 26, 90, 91, 102], the authors only considered sound-soft scattering problems, where they imposed Sommerfeld-like impedance boundary conditions on Γ_S and Dirichlet boundary conditions on Γ_D . From Corollary 2, the solution y of (3.5) with the initial value (y_0, y_1) is time-harmonic if, and only if, y and (y_0, y_1) fulfill (3.17). To determine the unknown (control) variables y_0, y_1 , one considers the least-squares optimization problem

$$\min_{(y_0, y_1)} J(y_0, y_1)$$

with a penalty cost functional J . There are two different approaches, optimize-then-discretize [104, 132, 25, 26, 83, 80] and discretize-then-optimize [90, 91].

In the discretize-then-optimize approach, one first discretizes (2.9), e.g. using FE, and afterwards formulates the least-squares optimization problem in \mathbb{R}^{2n} . The penalty functional is given by $J_1 : \mathbb{R}^{2n} \rightarrow \mathbb{R}_{\geq 0}$,

$$J_1(\mathbf{y}_0, \mathbf{y}_1) = \frac{1}{2} \begin{pmatrix} \mathbf{y}(T) - \mathbf{y}_0 \\ \frac{\partial}{\partial t} \mathbf{y}(T) - \mathbf{y}_1 \end{pmatrix}^\top \begin{pmatrix} \mathbf{K}_h & \\ & \mathbf{M}_h \end{pmatrix} \begin{pmatrix} \mathbf{y}(T) - \mathbf{y}_0 \\ \frac{\partial}{\partial t} \mathbf{y}(T) - \mathbf{y}_1 \end{pmatrix}, \quad (3.32)$$

where $\mathbf{y}(t)$ is the semi-discrete solution of (3.5) with the (discrete) initial value $(\mathbf{y}_0, \mathbf{y}_1)$, the stiffness matrix \mathbf{K}_h and the mass matrix \mathbf{M}_h are defined in (2.18).

Here we use the optimize-then-discretize approach and consider the penalty functional $J_1 : H^1(\Omega) \times L^2(\Omega) \rightarrow \mathbb{R}_{\geq 0}$,

$$J_1(y_0, y_1) = \frac{1}{2} \int_{\Omega} a(x) |\nabla(y(x, T) - y_0(x))|^2 dx + \frac{1}{2} \int_{\Omega} \frac{1}{c^2(x)} (y_t(x, T) - y_1(x))^2 dx, \quad (3.33)$$

where $y = y[y_0, y_1]$ solves (3.5) with the initial value (y_0, y_1) . Obviously, J_1 is non-negative and convex. Since

$$y[\operatorname{Re} u, \omega \operatorname{Im} u](x, t) = \operatorname{Re}\{u(x) e^{-i\omega t}\},$$

we obtain

$$J_1(\operatorname{Re} u, \omega \operatorname{Im} u) = 0.$$

Following Corollary 2 and the definition (3.33), J_1 has a unique (global) minimizer $(y_0, y_1) = (\operatorname{Re} u, \omega \operatorname{Im} u)$, which yields the solution u of (2.9) through (3.7). Therefore, the functional J_1 is minimal at (y_0, y_1) if, and only if,

$$J_1(y_0, y_1) = 0,$$

and if, and only if, the gradient or Fréchet derivative J'_1 vanishes identically at (y_0, y_1) .

We derive the gradient with the adjoint method [104]. First, the chain rule yields

$$\begin{aligned} \langle J'_1(y_0, y_1), (\delta y_0, \delta y_1) \rangle &= \int_{\Omega} a(x) \nabla(y(x, T) - y_0(x)) \cdot \nabla(\delta y(x, T) - \delta y_0(x)) dx \\ &\quad + \int_{\Omega} \frac{1}{c^2(x)} (y_t(x, T) - y_1(x)) (\delta y_t(x, T) - \delta y_1(x)) dx, \end{aligned} \quad (3.34)$$

for a perturbation $(\delta y_0, \delta y_1) \in H_D^1(\Omega) \times L^2(\Omega)$, where δy solves the homogeneous wave equation (3.5) with $f = g_S = g_N = g_D = 0$ and $\langle \cdot, \cdot \rangle$ denotes the standard duality pairing in $((H_D^1 \times L^2(\Omega))', H_D^1 \times L^2(\Omega))$. Next, multiplication of (3.5) by an arbitrary function p and integration by parts over Ω and $[0, T]$ then imply that

$$\begin{aligned} 0 &= \int_0^T \int_{\Omega} \left[\frac{1}{c^2} \delta y_{tt} - \nabla \cdot (a \nabla \delta y) \right] p dx dt \\ &= \int_0^T \int_{\Omega} \left[\frac{1}{c^2} p_{tt} - \nabla \cdot (a \nabla p) \right] \delta y dx dt + \int_0^T \int_{\Gamma_S} \left[a \frac{\partial}{\partial n} p - \frac{\sqrt{a}}{c} p_t \right] \delta y ds dt \\ &\quad + \int_0^T \int_{\Gamma_N} a \frac{\partial}{\partial n} p \delta y ds dt + \int_{\Gamma_S} \frac{\sqrt{a}}{c} [p \delta y] \Big|_0^T ds + \int_{\Omega} \frac{1}{c^2} \left[p \delta y_t - p_t \delta y \right] \Big|_0^T dx. \end{aligned} \quad (3.35)$$

We further have

$$\begin{aligned} \langle J'_1(v), \delta v \rangle &= - \int_{\Omega} a(x) \nabla(y(x, T) - y_0(x)) \cdot \nabla \delta y_0(x) dx \\ &\quad - \int_{\Omega} \frac{1}{c^2(x)} (y_t(x, T) - y_1(x)) \delta y_1(x) dx \\ &\quad + \int_{\Omega} \frac{1}{c^2(x)} (p(x, 0) \delta y_1(x) - p_t(x, 0) \delta y_0(x)) dx + \int_{\Gamma_S} \frac{\sqrt{a(x)}}{c(x)} p(x, 0) \delta y_0(x) ds, \end{aligned} \quad (3.36)$$

where p solves the solution of the adjoint (backward) wave equation

$$\frac{1}{c^2(x)} \frac{\partial^2 p(x, t)}{\partial t^2} - \nabla \cdot (a(x) \nabla p(x, t)) = 0, \quad x \in \Omega, t > 0, \quad (3.37a)$$

$$\sqrt{a(x)} \frac{\partial p(x, t)}{\partial n} - \frac{1}{c(x)} \frac{\partial p(x, t)}{\partial t} = 0, \quad x \in \Gamma_S, t > 0 \quad (3.37b)$$

$$\frac{\partial p(x, t)}{\partial n} = 0, \quad x \in \Gamma_N, t > 0, \quad (3.37c)$$

$$p(x, t) = 0, \quad x \in \Gamma_D, t > 0, \quad (3.37d)$$

$$p(x, T) = p_0(x), \quad p_t(x, T) = p_1(x), \quad x \in \Omega, \quad (3.37e)$$

with the initial conditions

$$p_0(x) = y_t(x, T) - y_1(x), \quad x \in \Omega, \quad (3.37f)$$

$$\int_{\Omega} \frac{1}{c^2} p_1 w dx = \int_{\Gamma_S} \frac{\sqrt{a}}{c} p_0 w ds - \int_{\Omega} a \nabla(y(\cdot, T) - y_0) \cdot \nabla w dx, \quad w \in H_D^1. \quad (3.37g)$$

The derivation of (3.36) and (3.37) and the CMCG Algorithm can be found in [25, 26]. Hence in each CG iteration, we compute $J'_1(y_0, y_1)$ by solving the forward and backward (adjoint) wave equations (3.5) and (3.37) over one period $[0, T]$. If the time-step Δt until $T = 2\pi/\omega$ is chosen inversely proportional to the frequency ω , the number of time-steps over one period remains independent of ω . Moreover, each CG iteration requires an explicit (Riesz) representer $\tilde{g} = (\tilde{g}_0, \tilde{g}_1) \in H_D^1(\Omega) \times L^2(\Omega)$ of the gradient $g = (g_0, g_1) = J'_1(y_0, y_1)$, which is determined by solving the symmetric and coercive elliptic problem [26, 108]:

$$\begin{aligned} \int_{\Omega} a \nabla \tilde{g}_0 \cdot \nabla \varphi \, dx &= \int_{\Omega} g_0(x) \varphi(x) \, dx \\ &= \int_{\Omega} \left[a(x) \nabla (y_0(x) - y(x, T)) \cdot \nabla \varphi(x) - p_t(x, 0) \varphi(x) \right] dx \\ &\quad + \int_{\Gamma_S} \frac{\sqrt{a(x)}}{c(x)} p(x, 0) \varphi(x) \, ds, \quad \forall \varphi \in H_D^1, \end{aligned} \quad (3.38a)$$

$$\tilde{g}_1(x) = c^2(x) g_1(x) = y_1(x) - y_t(x, T) + p(x, 0), \quad x \in \Omega. \quad (3.38b)$$

3.4.2 Sound-hard scattering problems or scattering from inhomogeneous inclusion

Let u be the (unique) solution of the Helmholtz equation (2.9) with $\mathcal{H}^{d-1}(\Gamma_S) > 0$, but now $\Gamma_D = \emptyset$, as in a typical sound-hard scattering problem where Γ_N corresponds to the boundary of the obstacle and Γ_S to the exterior artificial boundary. Since $\mathcal{H}^{d-1}(\Gamma_S) > 0$, we again immediately conclude from Theorem 6 that η and γ_ℓ , together with α_ℓ and β_ℓ , identically vanish in (3.22). However, in contrast to the situation of sound-soft scattering from Section 3.4.1, λ is no longer necessarily zero. In fact, Corollary 2 implies that any (global) minimizer v of J_1 necessarily corresponds to an arbitrary constant shift of u , i.e. $v = u + \lambda$, $\lambda \in \mathbb{R}$. To remove the spurious constant shift and restore uniqueness, we propose three distinct approaches:

- (i) determine λ by enforcing a compatibility condition,
- (ii) replace J_1 by an alternative functional or
- (iii) use an absorbing boundary condition on Γ_S with a zeroth-order term.

(i) Compatibility condition.

From Corollary 2 we know that if $J_1(y_0, y_1) = 0$ there exists a constant $\lambda \in \mathbb{R}$ such that

$$v(x) = y_0(x) + \frac{i}{\omega} y_1(x) = u(x) + \lambda.$$

However, since $\mathcal{H}^{d-1}(\Gamma_S) > 0$, (2.9) always has a unique solution u . Hence, we shall now derive an additional condition from (2.9) to determine λ .

In the absence of Γ_D , by (2.10), we have

$$A[u, 1] = A[v - \lambda, 1] = A[v, 1] - \lambda A[1, 1] = F[1], \quad (3.39)$$

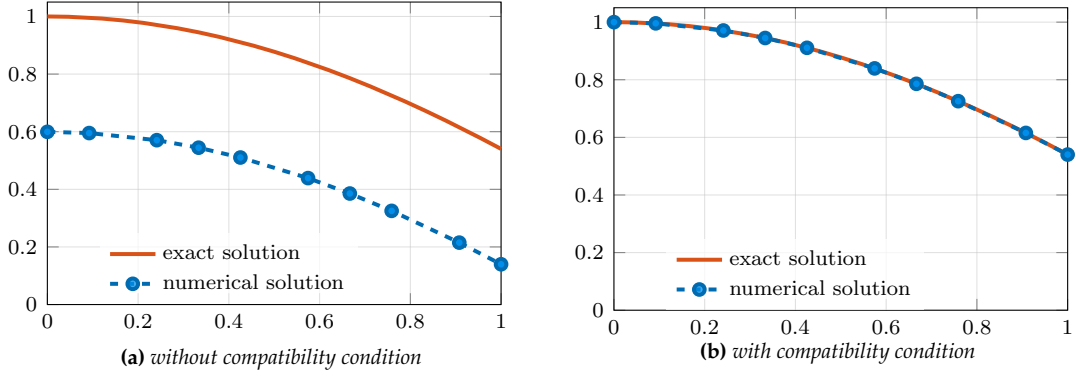


Figure 3.6: One-dimensional Helmholtz equation with Neumann condition at $x = 0$ and Sommerfeld condition at $x = 1$. Exact and numerical solution using the CMCG method with the cost functional J_1 ; (a) without compatibility condition; (b) with compatibility condition (3.41).

where A and F are given by (2.11) and (2.12), respectively. This yields after some algebra the compatibility condition

$$\lambda = \frac{A[v, 1] - F[1]}{A[1, 1]} = - \frac{\int_{\Omega} k^2 v \, dx + \int_{\Omega} f \, dx + \int_{\Gamma_S} [g_S + i\sqrt{a} kv] \, ds + \int_{\Gamma_N} g_N \, ds}{\int_{\Omega} k^2 \, dx + i \int_{\Gamma_S} \sqrt{a} k \, ds}. \quad (3.40)$$

Thus, once we have found a T -periodic solution $v = y_0 + (i/\omega) y_1$ of (3.5), we determine the correct (unique) time-harmonic solution u by applying the constant shift

$$u = v - \lambda, \quad (3.41)$$

with λ given by (3.40). We denote the combined use of J_1 with the compatibility condition (3.40)–(3.41) by J_1^{cc} .

To illustrate the usefulness of the compatibility condition (3.40)–(3.41), we now consider the following simple one-dimensional example. Let

$$u(x) = e^{ix}, \quad x \in \Omega = (0, 1),$$

be the exact solution of (2.9) with $\Gamma_N = \{0\}$, $\Gamma_S = \{1\}$, $\omega = a = c = 1$, and $f \equiv 0$. In Figure 3.6, we display the exact and the numerical solution obtained with the CMCG method using the functional J_1 in (3.33). The CMCG iteration converges to a minimizer v , which contains a spurious constant shift from u . By computing λ in (3.40) and subtracting it from v , we recover u , as shown in Figure 3.6b.

(ii) Alternative functionals.

Since the coerciveness of J_1 in (3.33) "depends in a subtle way on the geometry of Ω " [11], Bardos and Rauch proposed an alternative cost functional, J_{∞} , which is coercive even in situations with trapping rays. In [26], Bristeau et al. used the following real-valued version of J_{∞} :

$$\begin{aligned} J_{\infty}(y_0, y_1) = & \frac{1}{2} \int_0^T \int_{\Omega} a(x) |\nabla(y(x, t) - \operatorname{Re}\{v(x) e^{-i\omega t}\})|^2 \, dx \, dt \\ & + \frac{1}{2} \int_0^T \int_{\Omega} \frac{1}{c^2(x)} (y_t(x, t) - \omega \operatorname{Im}\{v(x) e^{-i\omega t}\})^2 \, dx \, dt, \end{aligned} \quad (3.42)$$

where y solves (3.5) and $v = y_0 + (i/\omega) y_1$. In contrast to J_1 , the functional J_∞ penalizes the departure from the desired time-harmonic behavior not only after one period at $t = T$, but in fact at all times. In [26, Section 7.6.2], the CMCG method with J_1 or J_∞ was applied to wave scattering from a perfectly conducting circular cavity with a crack and the numerical "solutions obtained with the two control approaches were the same".

When the cost functional J_1 is replaced by J_∞ in the CMCG Algorithm (Section 3.7), the Fréchet derivative of J_∞ requires the solution of modified forward and backward wave equations akin to (3.5) and (3.37). However, unlike in the original CMCG approach with J_1 , the backward wave equation now depends explicitly on the entire solution of the forward wave equation. To avoid storing the entire history of the forward wave equation, which may be prohibitive for large-scale problems, we consider instead the alternative cost functional

$$J_m(y_0, y_1) = \frac{1}{2} \sum_{j=1}^m \int_{\Omega} a(x) |\nabla(y(x, t_j) - \operatorname{Re}\{v(x) e^{-i\omega t_j}\})|^2 dx + \frac{1}{2} \sum_{j=1}^m \int_{\Omega} \frac{1}{c^2(x)} (y_t(x, t_j) - \omega \operatorname{Im}\{v(x) e^{-i\omega t_j}\})^2 dx, \quad (3.43)$$

where y solves (3.5) and $v = y_0 + (i/\omega) y_1$. Here $0 < t_1 < t_2 < \dots < t_m = T$ are $m \geq 1$ fixed time instants – we shall only use $m \leq 3$. Note that J_m coincides with J_1 for $m = 1$ and $t_1 = T$. In general, the uniqueness of a global minimizer of J_m depends on the specific boundary condition in (2.9) and on the choice of $\{t_j\}_{j=1}^m$. In particular, for $\Gamma_D = \emptyset$ and $\mathcal{H}^{d-1}(\Gamma_S) > 0$, the functional J_2 in (3.43) with $t_1 = T/4$ and $t_2 = T$ always has a unique global minimizer.

Theorem 7. *Let $u \in H^1(\Omega)$ be the unique solution of (2.9) with $\Gamma_D = \emptyset$ and $\mathcal{H}^{d-1}(\Gamma_S) > 0$. For $(y_0, y_1) \in H^1(\Omega) \times H^1(\Omega)$, the following assertions are equivalent:*

- (i) $u = y_0 + (i/\omega) y_1$.
- (ii) $J_2(y_0, y_1) = 0$ with J_2 as in (3.43) with $m = 2$, $t_1 = T/4$ and $t_2 = T$.

Proof. The implication (i) \Rightarrow (ii) obviously holds, since the solution y of (3.5) with initial values $(y_0, y_1) = (\operatorname{Re} u, \omega \operatorname{Im} u)$ is unique and equal to $\operatorname{Re}\{u(x) e^{-i\omega t}\}$.

To show that (ii) \Rightarrow (i), let $(y_0, y_1) \in H^1(\Omega) \times H^1(\Omega)$ be a minimizer of J_2 and $v = y_0 + (i/\omega) y_1$. Then, for the corresponding solution y of (3.5),

$$\nabla(y(x, T) - y(x, 0)) = 0, \quad y_t(x, T) - y_t(x, 0) = 0 \quad \text{in } \Omega$$

and thus both ∇y and y_t are T -periodic. From Corollary 2, we infer that

$$y(x, t) = \operatorname{Re}\{u(x) e^{-i\omega t}\} + \lambda$$

for some constant $\lambda \in \mathbb{R}$ and that $v = u + \lambda$. Thus

$$y(x, t) = \operatorname{Re}\{(v(x) - \lambda) e^{-i\omega t}\} + \lambda = \operatorname{Re}\{v(x) e^{-i\omega t}\} + \lambda (1 - \cos(\omega t)). \quad (3.44)$$

Since $J_2(y_0, y_1) = 0$, we also have

$$y_t\left(x, \frac{T}{4}\right) - \omega \operatorname{Im}\{v(x) e^{-i\omega \frac{T}{4}}\} = 0, \quad x \in \Omega. \quad (3.45)$$

We differentiate (3.44) in time, set $t = T/4$, and use (3.45) to obtain

$$0 = y_t\left(x, \frac{T}{4}\right) - \omega \operatorname{Im}\{v(x) e^{-i\omega \frac{T}{4}}\} - \omega \lambda \sin\left(\frac{\omega T}{4}\right) = -\omega \lambda \sin\left(\frac{\pi}{2}\right) = -\omega \lambda. \quad (3.46)$$

Since $\omega > 0$, (3.46) implies that $\lambda = 0$ and hence that $u = v$, which concludes the proof. \square

Remark 5. The choice of t_1 in Theorem 7 is not unique. In fact, Theorem 7 holds for any $t_1 \in (0, T)$ with $\sin(\omega t_1) \neq 0$, i.e. $t_1 \neq T/2$. The (somewhat arbitrary) value $t_1 = T/4$ lies between the first two critical values 0 and $T/2$.

To apply the CMCG method with the alternative functional J_m , we need to compute its Fréchet derivative. Let $v = (y_0, y_1) \in H^1(\Omega) \times H^1(\Omega)$, $\delta v = (\delta y_0, \delta y_1)$ be an arbitrary perturbation and define

$$\begin{aligned} \tilde{v}(x, t) &= \operatorname{Re} \left\{ (y_0(x) + \frac{i}{\omega} y_1(x)) e^{-i\omega t} \right\}, & x \in \Omega, t > 0, \\ \tilde{\delta v}(x, t) &= \operatorname{Re} \left\{ (\delta y_0(x) + \frac{i}{\omega} \delta y_1(x)) e^{-i\omega t} \right\}, & x \in \Omega, t > 0. \end{aligned}$$

Then, the Fréchet derivative of J_m is

$$\begin{aligned} \langle J'_m(v), \delta v \rangle &= - \sum_{\ell=1}^m \int_{\Omega} a(x) \nabla(y(x, t_{\ell}) - \tilde{v}(x, t_{\ell})) \cdot \nabla \tilde{\delta v}(x, t_{\ell}) dx \\ &\quad - \sum_{\ell=1}^m \int_{\Omega} \frac{1}{c^2(x)} (y_t(x, t_{\ell}) - \tilde{v}_t(x, t_{\ell})) \tilde{\delta v}_t(x, t_{\ell}) dx \\ &\quad + \int_{\Omega} \frac{1}{c^2(x)} (p(x, 0) \delta y_1(x) - p_t(x, 0) \delta y_0(x)) dx + \int_{\Gamma_S} \frac{\sqrt{a(x)}}{c(x)} p(x, 0) \delta y_0(x) ds. \end{aligned} \quad (3.47)$$

Here $p = \sum_{\ell=1}^m p^{\ell}$ and each p^{ℓ} solves the adjoint wave equation (3.37) with the initial conditions $p^{\ell}(x, t_{\ell}) = p_0^{\ell}(x)$ and $p_t^{\ell}(x, t_{\ell}) = p_1^{\ell}(x)$, $1 \leq \ell \leq m$, defined by

$$\begin{aligned} p_0^{\ell}(x) &= y_t(x, t_{\ell}) - \tilde{v}_t(x, t_{\ell}), & x \in \Omega, \\ \int_{\Omega} \frac{1}{c^2(x)} p_1^{\ell}(x) w(x) dx &= \int_{\Gamma_S} \frac{\sqrt{a(x)}}{c(x)} p_0^{\ell}(x) w(x) ds \\ &\quad - \int_{\Omega} a(x) \nabla(y(x, t_{\ell}) - \tilde{v}(x, t_{\ell})) \cdot \nabla w(x) dx, & \forall w \in H^1(\Omega). \end{aligned}$$

The CMCG method with the cost functional J_m in (3.47) requires the solution of a single forward wave equation until time T and m independent backward wave equations over $[0, t_{\ell}]$, $\ell = 1, \dots, m$. To ensure that both $y_0^{(k)}$ and $y_1^{(k)}$ remain in $H^1(\Omega)$ during the CMCG iteration, we now determine an explicit Riesz representer $\tilde{g}^{(k)} = (\tilde{g}_0^{(k)}, \tilde{g}_1^{(k)}) \in H^1(\Omega) \times H^1(\Omega)$ of $g^{(k)} = J'_m(v^{(k)})$ by solving an elliptic coercive problem similar to (3.38a) for both $\tilde{g}_0^{(k)}$ and $\tilde{g}_1^{(k)}$. For $m = 2$, $t_1 = T/4$ and $t_2 = T$, for instance, the CMCG Algorithm based on the functional J_2 always computes the correct minimizer at a modest 12.5% increase in the computational cost over using J_1 – see Section 3.7 for details.

Next, the functional J_{∞} given by (3.42) not only has a uniqueness solution for scattering problems, but also for Helmholtz equation with general boundary condition. We show that in the following theorem.

Theorem 8. Let $u \in H^1(\Omega)$ be the unique solution of the Helmholtz equation (2.9) and $y \in C^1([0, T]; H^1(\Omega))$ be the solution of the wave equation (3.5) with the initial values $y_0, y_1 \in H^1(\Omega)$. Suppose that $J_\infty(y_0, y_1) = 0$. Then

$$u = y_0 + \frac{i}{\omega} y_1.$$

Proof. Following $J_\infty(y_0, y_1) = 0$, we have

$$\begin{aligned} \frac{\partial}{\partial t} y(x, t) - \frac{\partial}{\partial t} \operatorname{Re}\{v(x) e^{-i\omega t}\} &= 0, \\ \nabla y(x, t) - \nabla \operatorname{Re}\{v(x) e^{-i\omega t}\} &= 0, \end{aligned}$$

with $v = y_0 + \frac{i}{\omega} y_1$, which implies that

$$y(x, t) - \operatorname{Re}\{v(x) e^{-i\omega t}\} \equiv \mu \in \mathbb{R}.$$

Since $y(x, 0) = y_0(x)$, we immediately have $\mu = 0$, which concludes the proof. \square

For the sake of completeness, we also list the derivative of $J_\infty(v)$

$$\begin{aligned} \langle J'_\infty(v), \delta v \rangle &= - \int_0^T \int_\Omega a(x) \nabla(y(x, t) - \tilde{v}(x, t)) \nabla \tilde{\delta v}(x, t) \, dx \, dt \\ &\quad - \int_0^T \int_\Omega \frac{1}{c^2(x)} (y_t(x, t) - \tilde{v}_t(x, t)) \tilde{\delta v}_t(x, t) \, dx \, dt \\ &\quad + \int_\Omega \frac{1}{c^2(x)} (p(x, 0) \delta y_1(x) - p_t(x, 0) \delta y_0(x)) \, dx + \int_{\Gamma_S} \frac{\sqrt{a(x)}}{c(x)} p(x, 0) \delta y_0(x) \, ds, \end{aligned} \quad (3.48)$$

where p is the solution of the adjoint wave equation (3.49)

$$\frac{1}{c^2} p_{tt} - \nabla \cdot (a \nabla p) = -\nabla \cdot (a \nabla (y - \tilde{v})) - \frac{1}{c^2} (y_{tt} - \tilde{v}_{tt}) \quad \text{in } \Omega \times (0, T), \quad (3.49a)$$

$$a \frac{\partial p}{\partial n} - \frac{\sqrt{a}}{c} p_t = a \frac{\partial (y - \tilde{v})}{\partial n} \quad \text{on } \Gamma_S \times (0, T), \quad (3.49b)$$

$$\frac{\partial p}{\partial n} = \frac{\partial (y - \tilde{v})}{\partial n} \quad \text{on } \Gamma_N \times (0, T), \quad (3.49c)$$

$$p = 0 \quad \text{on } \Gamma_D \times (0, T), \quad (3.49d)$$

$$p(x, T) = 0 \quad \text{for } x \in \Omega, \quad (3.49e)$$

$$p_t(x, T) = -(y_t(x, T) - \tilde{v}_t(x, T)) \quad \text{for } x \in \Omega. \quad (3.49f)$$

Remark 6. For the CMCG method with J_∞ , the adjoint state method in (3.49) requires the time-reversed solution $y(\cdot, T - t)$ of the forward problem (3.5), and hence we need to store the entire history of y . For a large problem, in particular three-dimensional problems, the storage requirement may be extensive. Using reverse time migration (RTM) with checkpoints, one can reduce the storage [128]. However, this involves additional time integration of (3.5) and leads to a trade-off between computational cost and storage requirement.

(iii) Absorbing boundary condition with zeroth-order term.

From Corollary 2, we know that the eigenvalue problems (3.21) with $\Gamma_D = \emptyset$, $\mathcal{H}^{d-1}(\Gamma_S) > 0$, and the Sommerfeld condition (3.21b) only have the trivial solution for $\ell \neq 0$. However, we may also replace the simple Sommerfeld-like impedance condition (2.9b),

with $a \equiv 1$, by the (more accurate) first-order Bayliss-Gunzburger-Turkel (BGT-1) condition [12, 76] in $d = 2, 3$ dimensions, for instance:

$$\frac{\partial u(x)}{\partial n} - ik(x)u(x) + \frac{1}{(4-d)r}u(x) = g_S(x), \quad x \in \Gamma_S. \quad (3.50a)$$

The corresponding time-harmonic wave field y solves (3.5) with (3.5b) replaced by

$$\frac{\partial y(x, t)}{\partial n} + \frac{1}{c(x)} \frac{\partial y(x, t)}{\partial t} + \frac{1}{(4-d)r}y(x, t) = \operatorname{Re}\{g_S(x) e^{-i\omega t}\}, \quad x \in \Gamma_S, t > 0, \quad (3.50b)$$

while Corollary 6 still holds with (3.21b) replaced by

$$\frac{\partial \gamma_\ell(x)}{\partial n} - i\ell k(x)\gamma_\ell(x) + \frac{1}{(4-d)r}\gamma_\ell(x) = 0, \quad x \in \Gamma_S. \quad (3.51)$$

Because of the strictly positive (or negative) definite imaginary part of the zeroth-order coefficient in (3.50b), the eigenvalue problems (3.21) still only have the trivial solution for $\ell \neq 0$. Moreover even for $\ell = 0$, the nonzero coefficient of the zeroth-order remaining term in (3.51) now implies that γ_0 is not only constant but in fact zero everywhere in Ω . In this case, the T -periodicity of y is sufficient to guarantee convergence to the (unique) solution of the Helmholtz equation (2.9).

Clearly, other absorbing boundary conditions [76] may also permit the use of J_1 without any modification. Exact nonreflecting boundary conditions based on integral formulations or DtN maps [76], for instance, could also be used but will generally lead to nonlocal space-time operators on the artificial boundary. Alternatively, perfectly matched layers can probably be used as is and will lead to local formulations without convolution integrals in time when combined with appropriate auxiliary functions – see Section 4.2.

3.4.3 Helmholtz equation in physically bounded domains

In Sections 3.4.1 and 3.4.2, we considered various scattering problems where we impose an impedance boundary condition (2.9b) on part of the boundary Γ_S . Here, we consider the Helmholtz equation (2.9) with $\Gamma_S = \emptyset$, i.e. with pure Dirichlet or Neumann boundary condition. Clearly, to guarantee the uniqueness of the solution u of (2.9), we now always assume that ω^2 is not an eigenvalue of the principal elliptic part. Still, the CMCG method with the cost functional J_1 from (3.33) in general will not yield the correct solution u , not even shifted by an arbitrary constant. Indeed, let $(y_0, y_1) \in H^1(\Omega) \times L^2(\Omega)$ be a minimizer of J_1 . According to Theorem 6, $v = y_0 + (i/\omega) y_1$ then satisfies (3.22) with γ_ℓ being the solution of (3.21). While for $\mathcal{H}^{d-1}(\Gamma_S) > 0$ all eigenfunction γ_ℓ , $|\ell| > 1$, necessarily vanish (Corollary 2), this is no longer the case when $\Gamma_S = \emptyset$, as $(\omega\ell)^2$ may be an eigenvalue of the principal part for some $|\ell| > 1$, even when ω^2 is not.

To illustrate this added ambiguity, we now consider the following simple one-dimensional example. Let f be chosen so that

$$u(x) = 4x(1-x), \quad x \in \Omega = (0, 1),$$

is the solution of (2.9) with $a = c = 1$, $\omega = \pi/4$, and Neumann boundary conditions $u'(0) = 4$ and $u'(1) = -4$. Since ω^2 does not lie in the spectrum $\Sigma = \{(\ell\pi)^2 : \ell \in \mathbb{Z}\}$ of (3.21), (2.9) is well-posed. However, $(\omega\ell)^2 \in \Sigma$ for any $\ell \in 4\mathbb{Z}$ with corresponding eigenfunction $\alpha_\ell(x) = \cos(\omega\ell x)$.

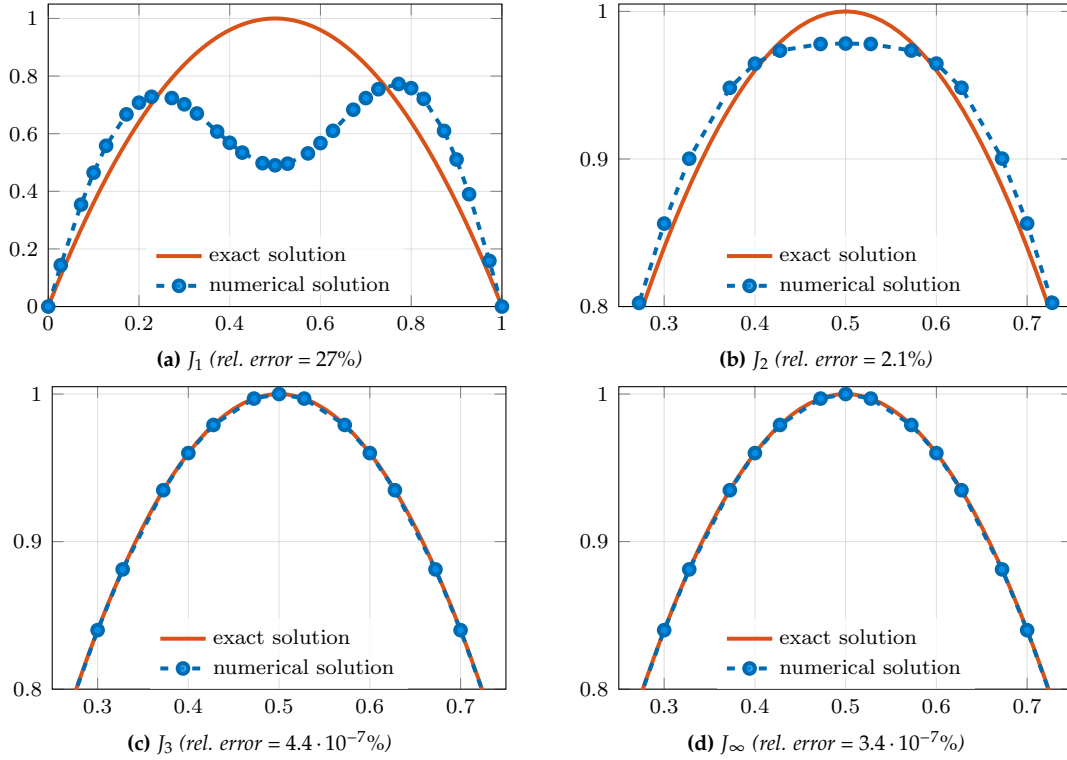


Figure 3.7: Helmholtz equation in a bounded domain: numerical solutions of (2.9) in $\Omega = (0, 1)$ with Neumann conditions at $x = 0, 1$ using the CMCG method either with J_1 (a), J_2 (b), J_3 (c), or J_∞ (d).

Now, let (y_0, y_1) be an arbitrary initial value and y be the corresponding solution of the wave equation (3.5). Then the solution \hat{y} of (3.5) with the initial value $(y_0 + \alpha_\ell, y_1)$ is $\hat{y}(x, t) = y(x, t) + \alpha_\ell(x) \cos(\omega \ell t)$. Since $T = 2\pi/\omega$,

$$\hat{y}(x, T) - (y_0(x) + \alpha_\ell(x)) = y(x, T) + \alpha_\ell(x) \cos(\omega \ell T) - (y_0(x) + \alpha_\ell(x)) = y(x, T) - y_0(x)$$

and

$$\hat{y}_t(x, T) - y_1(x) = y_t(x, T) - \omega \ell \alpha_\ell(x) \sin(\omega \ell T) - y_1(x) = y_t(x, T) - y_1(x)$$

for any $\ell \in 4\mathbb{Z}$. Thus

$$J_1(y_0 + \alpha_\ell, y_1) = J_1(y_0, y_1), \quad (y_0, y_1) \in H^1(\Omega) \times L^2(\Omega), \quad \ell \in 4\mathbb{Z},$$

and minimizers of J_1 are neither unique nor do they simply differ by a constant.

As shown in Figures 3.7a and 3.7b, the CMCG method using J_1 , or even J_2 , does not yield the correct minimizer u . However, the cost functional J_∞ in (3.42) remains strictly convex; hence, the CMCG method based on J_∞ converges to the correct minimizer u , as shown in Figure 3.7d. Similarly, the cost functional J_3 in (3.43) with $t_1 = T/4$, $t_2 = T/2$, and $t_3 = T$ always yields u , as illustrated in Figure 3.7c and shown in the following theorem.

Theorem 9. Let $u \in H^1(\Omega)$ be the unique solution of the Helmholtz equation (2.9) and $(y_0, y_1) \in H^1(\Omega) \times H^1(\Omega)$. The following assertions are equivalent:

- (i) $u = y_0 + (i/\omega) y_1$.
- (ii) $J_3(y_0, y_1) = 0$ with $t_1 = T/4$, $t_2 = T/2$, $t_3 = T$, and $m = 3$ in (3.43).

Proof. As in the proof of Theorem 7, the implication (i) \Rightarrow (ii) immediately holds.

Now, let $(y_0, y_1) \in H^1(\Omega) \times H^1(\Omega)$ be a minimizer of J_3 and y be the solution of (3.5). Then, $\nabla(y(x, T) - y(x, 0))$ and $y_t(x, T) - y_t(x, 0)$ vanish identically in Ω . Next, we define the “misfit function”

$$r(x, t) = y(x, t) - \operatorname{Re}\{v(x) e^{-i\omega t}\} \quad (3.52)$$

and use (3.20) and (3.22) from Theorem 6 to replace v and y in (3.52). Let γ_ℓ be the Fourier coefficients from Theorem 6, $\alpha_\ell = \operatorname{Re}\{\gamma_\ell\}$, and $\beta_\ell = \operatorname{Im}\{\gamma_\ell\}$. Then this yields

$$\begin{aligned} (r(\cdot, t), \varphi) &= (y(\cdot, t) - \operatorname{Re}\{u e^{-i\omega t}\}, \varphi) - (\operatorname{Re}\{v e^{-i\omega t}\} - \operatorname{Re}\{u e^{-i\omega t}\}, \varphi) \\ &= (\lambda(1 - \cos(\omega t)) + \eta(t - \omega^{-1} \sin(\omega t)), \varphi) \\ &\quad + \sum_{\ell=2}^{\infty} \left\{ (\alpha_\ell(\cos(\omega \ell t) - \cos(\omega t)) + \beta_\ell(\sin(\omega \ell t) - \ell \sin(\omega t)), \varphi) \right\}. \end{aligned} \quad (3.53)$$

Differentiating once with respect to time, we obtain

$$\begin{aligned} (r_t(\cdot, t), \varphi) &= (\omega \lambda \sin(\omega t) + \eta(1 - \cos(\omega t)), \varphi) \\ &\quad + \omega \sum_{\ell=2}^{\infty} \left\{ (-\alpha_\ell(\ell \sin(\omega \ell t) - \sin(\omega t)) + \ell \beta_\ell(\cos(\omega \ell t) - \cos(\omega t)), \varphi) \right\}. \end{aligned} \quad (3.54)$$

Since $\nabla r(\cdot, t_j) = 0$ and $r_t(\cdot, t_j) = 0$ for $t_j \in \{T/4, T/2\}$, $r(\cdot, t_j) \equiv \mu_j$ is constant a.e. in Ω , $j = 1, 2$, while for $t = t_2 = T/2 = \pi/\omega$, (3.53) and (3.54) reduce to

$$\begin{aligned} (\mu_2, \varphi) &= (2\lambda + \eta \frac{T}{2}, \varphi) + \sum_{\ell=2}^{\infty} ((-1)^\ell + 1)(\alpha_\ell, \varphi), \\ 0 &= (2\eta, \varphi) + \omega \sum_{\ell=2}^{\infty} \ell((-1)^\ell + 1)(\beta_\ell, \varphi). \end{aligned}$$

Thus,

$$(2\lambda + \eta \frac{T}{2} - \mu_2, \varphi) + 2 \sum_{\substack{\ell=2 \\ \ell \text{ even}}}^{\infty} (\alpha_\ell, \varphi) = (2\eta, \varphi) + 2\omega \sum_{\substack{\ell=2 \\ \ell \text{ even}}}^{\infty} \ell(\beta_\ell, \varphi) = 0$$

for any test function $\varphi \in H_D^1(\Omega)$.

If $\mathcal{H}^{d-1}(\Gamma_D) > 0$, $\mu_1 = \mu_2 = 0$ and $\lambda = \eta = 0$ by Theorem 6. Otherwise if $\Gamma_D = \emptyset$, $2\lambda + \eta T/2 - \mu_2$ and 2η both solve (3.21) with $\ell = 0$ and hence are eigenfunctions with zero eigenvalue unless they vanish identically. Similarly, α_ℓ and β_ℓ are either eigenfunctions of (3.23) and (3.24) for the eigenvalue $(\omega \ell)^2$ or identically zero. By orthogonality, $\eta = 0$ and the eigenfunctions $\alpha_\ell = 0$ and $\beta_\ell = 0$ for any even $\ell \geq 2$.

Now, we set $t = t_1 = T/4 = \pi/(2\omega)$ in (3.53) and (3.54) to obtain

$$\begin{aligned} (\mu_1, \varphi) &= (\lambda, \varphi) - \sum_{\substack{\ell=3 \\ \ell \text{ odd}}}^{\infty} ((-1)^\ell + \ell)(\beta_\ell, \varphi) = (\lambda, \varphi) - \sum_{\substack{\ell=3 \\ \ell \text{ odd}}}^{\infty} (\ell - 1)(\beta_\ell, \varphi), \\ 0 &= (\omega \lambda, \varphi) + \omega \sum_{\substack{\ell=3 \\ \ell \text{ odd}}}^{\infty} (\ell(-1)^\ell + 1)(\alpha_\ell, \varphi) = (\omega \lambda, \varphi) - \omega \sum_{\substack{\ell=3 \\ \ell \text{ odd}}}^{\infty} (\ell - 1)(\alpha_\ell, \varphi). \end{aligned}$$

If α_ℓ and β_ℓ do not identically vanish, they are eigenfunctions of (3.23) and (3.24), respectively. Since $\ell - 1 > 0$ for any odd number $\ell \geq 3$, the orthogonality of the eigenfunctions implies $\alpha_\ell = 0$ and $\beta_\ell = 0$ for any odd $\ell \geq 3$ and $\lambda = 0$. Hence, $\alpha_\ell = \beta_\ell = 0$ for ℓ and $\lambda = \eta = 0$, which concludes the proof. \square

3.5 Frequency filtering procedure

From Theorem 6 we conclude that a minimizer of J_1 generally yields a time-dependent solution y of (3.5), which contains a constant shift determined by λ , a linearly growing part determined by η , and higher frequency harmonics determined by γ_ℓ , all superimposed on the desired time-harmonic field u with fundamental frequency ω . Those spurious modes can be eliminated by replacing J_1 with an alternative energy functional at a small extra computational cost [83]. Instead we now propose an alternative approach via filtering which removes all spurious modes without requiring a modified energy functional. However, for Neumann problems $\Gamma_N = \partial\Omega$, the spurious constant η in the Fourier expansion (3.22) is still present in any minimizer of J_1 . To determine η , we again apply the compatibility condition (3.39).

3.5.1 Filtering procedure

Let $y(x, t)$ be the time-dependent solution of (3.5) that corresponds to a minimizer (y_0, y_1) of J_1 . Next, we define $\widehat{y} \in \{w \in H^1(\Omega) \mid w = g_D \text{ on } \Gamma_D\}$ as

$$\widehat{y}(x) = \oint_0^T \left[y(x, t) + \frac{i}{\omega} y_t(x, t) \right] e^{i\omega t} dt. \quad (3.55)$$

To extract $u(x)$ from $y(x, t)$, we now take advantage of the mutual orthogonality of different time harmonics $\exp(i\omega \ell t)$ in $L^2(0, T)$. Hence, we multiply (3.20) with $e^{i\omega t}$ and integrate in time over $(0, T)$ to obtain

$$\begin{aligned} \widehat{y}(x) &= \oint_0^T \left(\operatorname{Re}\{u e^{-i\omega t}\} + \lambda + \eta t + i \operatorname{Im}\{u e^{-i\omega t}\} + \frac{i\eta}{\omega} \right) e^{i\omega t} dt \\ &= \oint_0^T u e^{-i\omega t} e^{i\omega t} dt - \frac{i\eta}{\omega} = u - \frac{i\eta}{\omega}. \end{aligned} \quad (3.56)$$

This yields

$$u(x) = \widehat{y}(x) + \frac{i\eta}{\omega}, \quad x \in \Omega \quad (3.57)$$

where λ and all γ_ℓ have vanished but the constant η is still undetermined.

If $\mathcal{H}^{d-1}(\Gamma_S) > 0$ or $\mathcal{H}^{d-1}(\Gamma_D) > 0$, Theorem 6 implies that $\eta = 0$ and thus $u(x) = \widehat{y}(x)$. Otherwise in the pure Neumann case ($\Gamma_N = \partial\Omega$), we determine η by integrating (3.57), multiplied by $k^2(x)$, over Ω and using the compatibility condition

$$- \int_\Omega k^2(x) u(x) dx = \int_\Omega f(x) dx + \int_{\partial\Omega} g_N(x) ds \quad (3.58)$$

from (2.9a). This immediately yields the remaining constant

$$\frac{i\eta}{\omega} = - \frac{1}{\|k\|_{L^2(\Omega)}^2} \left(\int_\Omega f(x) dx + \int_{\partial\Omega} g_N(x) ds + \int_\Omega k^2(x) \widehat{y}(x) dx \right). \quad (3.59)$$

We summarize the above derivation in the following proposition.

Proposition 4. *Let $u \in H^1(\Omega)$ be the unique solution of (2.9) and y the time dependent solution of (3.5) corresponding to a minimizer $(y_0, y_1) \in H^1(\Omega) \times L^2(\Omega)$ of J , i.e. $J(y_0, y_1) = 0$. Then u is given by (3.57) with $\eta = 0$ if $\mathcal{H}^{d-1}(\Gamma_S) > 0$ or $\mathcal{H}^{d-1}(\Gamma_D) > 0$, and with η given by (3.59) when $\Gamma_N = \partial\Omega$.*

Not only does the above filtering approach allow us to use the original cost functional J_1 , it also involves a negligible computational effort or storage amount, as the time integral for \hat{y} can be calculated cumulatively via numerical quadrature during the solution of the wave equation (3.5).

For an even number of steps n_T , for instance, the Simpson's quadrature rule yields

$$\hat{y}(x) \approx \frac{1}{3n_T} \sum_{m=1}^{n_T/2} [\tilde{v}(x, t_{2m-2}) + 4\tilde{v}(x, t_{2m-1}) + \tilde{v}(x, t_{2m})],$$

where $\tilde{v}(x, t) = (y(x, t) + \frac{i}{\omega} y_t(x, t)) e^{i\omega t}$ and $t_m = m\Delta t$, $m = 0, \dots, n_T$.

We denote the use of the original functional J_1 , combined with the filtering procedure and the compatibility condition, by $J_1^{f,p,cc}$.

3.5.2 Error estimate

We consider the accuracy and rate of convergence of the numerical solution u_h to the Helmholtz problem, obtained with the CMCG method and the frequency filtering procedure. The rate of convergence of the CMCG method may not only depend on the mesh size h but also on the time step Δt and the tolerance Tol in the CG method – see Section 3.7.

Let y be the solution of the wave equation with the initial values y_0, y_1 and \hat{y} given in (3.55). Then we first show that the error of $\|u - \hat{y}\|$ in L^2 -norm depends on the departure from $y(0), y_t(0)$ after one period at $t = T$. Second, we show that u_h converges to the numerical solution u_h^* of (2.19), obtained by the direct solver, for the same spatial discretization when Δt and Tol both tend to zero. As a consequence, for sufficiently small Δt and Tol , we obtain the same rate of convergence of u_h as the rate of the direct solver when h tends to zero.

Theorem 10. Assume $\Omega \subset \mathbb{R}^d$, $d = 1, 2, 3$, be an open, bounded and connected domain, with a Lipschitz boundary $\partial\Omega$. Let $u \in H^1(\Omega)$ be the unique complex-valued solution of (2.9) and $y \in C^0(0, T; H^1(\Omega)) \cap C^1(0, T; L^2(\Omega))$ the solution of (3.5) with the initial value $(y_0, y_1) \in H^1(\Omega) \times L^2(\Omega)$ and $T = \frac{2\pi}{\omega}$. Then v given by

$$v(x) = \int_0^T (y(x, t) + \frac{i}{\omega} y_t(x, t)) e^{i\omega t} dt \in H^1(\Omega) \quad (3.60)$$

satisfies the boundary conditions (2.9c), (2.9d) and there are constants $C, C_S > 0$ such that

$$\left\| a \frac{\partial}{\partial n} v - ik\sqrt{a}v - g_S \right\|_{H^{-\frac{1}{2}}(\Gamma_S)} \leq C_S (\|y(\cdot, T) - y_0\|_{H^1(\Omega)} + \|y_t(\cdot, T) - y_1\|_{H^{-\frac{1}{2}}(\Gamma_S)}), \quad (3.61)$$

and

$$\|u - v\|_{L^2(\Omega)} \leq C (\|y(\cdot, T) - y_0\|_{H^1(\Omega)} + \|y_t(\cdot, T) - y_1\|_{L^2(\Omega)}). \quad (3.62)$$

In particular, if y, y_t both are T -periodic then v satisfies the absorbing boundary condition (2.9) and $u = v$ a.e. in Ω .

Remark 7. (i) In the presence of absorbing boundary conditions ($\mathcal{H}^{d-1}(\Gamma_S) > 0$), the solution $v(x)$, resulted from the frequency filtering procedure given in (3.60), may not satisfy the desired condition (2.9b) when the solution $y(x, t)$ of the wave equation is not exactly T -time periodic. However, Theorem 10 provides the continuous dependence of the departure of v from (2.9b) on the departure of y from periodicity. For $y_t \in H^1(\Omega)$,

we have the estimate

$$\left\| a \frac{\partial}{\partial n} v - ik\sqrt{a}v - g_S \right\|_{H^{-\frac{1}{2}}(\Gamma_S)} \leq C_S (\|y(\cdot, T) - y_0\|_{H^1(\Omega)} + \|y_t(\cdot, T) - y_1\|_{H^1(\Omega)}).$$

(ii) For then solution v in (3.60), where the corresponding time-dependent solution y is only “closed” to T -time periodic, the error of $u - v$ converges uniformly to zero when the departure of y from periodicity tends to zero.

Proof of Theorem 10. (i) We first show that v given by (3.60) lies in $H^1(\Omega)$ and satisfies the boundary conditions (3.5c) and (3.5d).

By the definition (3.60) of v and integration by parts in time, we have

$$v(x) = \int_0^T \left(y(x, t) + \frac{i}{\omega} y_t(x, t) \right) e^{i\omega t} dt = 2 \int_0^T y(x, t) e^{i\omega t} dt + \frac{i}{T\omega} y(x, t) \Big|_0^T.$$

The regularity of $y \in C^1(0, T; H^1(\Omega))$ immediately implies that $v \in H^1(\Omega)$.

Next, the boundary condition (3.5d) yields

$$v = \int_0^T (y(\cdot, t) + \frac{i}{\omega} y_t(\cdot, t)) e^{i\omega t} dt = \int_0^T (\operatorname{Re}\{g_D e^{-i\omega t}\} + i \operatorname{Im}\{g_D e^{-i\omega t}\}) e^{i\omega t} dt = g_D$$

on the boundary Γ_D . Similarly, boundary condition (3.5c) yields

$$a \frac{\partial}{\partial n} v = a \int_0^T (\operatorname{Re}\{g_N e^{-i\omega t}\} + i \operatorname{Im}\{g_N e^{-i\omega t}\}) e^{i\omega t} dt = g_N$$

on Γ_N , which proves the first assertion.

(ii) Second, we show that v satisfies (3.61). The definition (3.60) yields

$$\begin{aligned} a \frac{\partial}{\partial n} v - ik\sqrt{a}v &= \int_0^T \left(a \frac{\partial}{\partial n} (y(\cdot, t) + \frac{i}{\omega} y_t(\cdot, t)) - ik\sqrt{a} (y(\cdot, t) + \frac{i}{\omega} y_t(\cdot, t)) \right) e^{i\omega t} dt \\ &= \int_0^T \left(a \frac{\partial}{\partial n} y(\cdot, t) + \frac{\sqrt{a}}{c} y_t(\cdot, t) + \frac{i}{\omega} \left(a \frac{\partial}{\partial n} y_t(\cdot, t) - \sqrt{a} \frac{\omega^2}{c} y(x, t) \right) \right) e^{i\omega t} dt. \end{aligned}$$

We apply twice integration by parts over $[0, T]$ on the last term to obtain

$$\begin{aligned} a \frac{\partial}{\partial n} v - ik\sqrt{a}v &= \int_0^T \left(a \frac{\partial}{\partial n} y(\cdot, t) + \frac{\sqrt{a}}{c} y_t(\cdot, t) + \frac{i}{\omega} \frac{\partial}{\partial t} \left(a \frac{\partial}{\partial n} y(\cdot, t) + \frac{\sqrt{a}}{c} y_t(\cdot, t) \right) \right) e^{i\omega t} dt \\ &\quad - \frac{\sqrt{a}}{c} y(\cdot, t) \Big|_{t=0}^{t=T} - \frac{i\sqrt{a}}{\omega c} y_t(\cdot, t) \Big|_{t=0}^{t=T} \\ &= g_S - \frac{\sqrt{a}}{c} y(\cdot, t) \Big|_{t=0}^{t=T} - \frac{i\sqrt{a}}{\omega c} y_t(\cdot, t) \Big|_{t=0}^{t=T}, \end{aligned}$$

by the boundary condition (3.5b). This implies, together with the triangle inequality and the trace theorem [56, §5.5], that there is a constant $C_{tr} > 0$ such that

$$\left\| a \frac{\partial}{\partial n} v - ik\sqrt{a}v - g_S \right\|_{H^{-\frac{1}{2}}(\Gamma_S)} \leq C_S (\|y(\cdot, T) - y_0\|_{H^1(\Omega)} + \|y_t(\cdot, T) - y_1\|_{H^{-\frac{1}{2}}(\Gamma_S)})$$

with

$$C_S = \|\sqrt{a}\|_{L^\infty(\Omega)} \left(\frac{C_{tr}}{c_{\min}} + \frac{1}{\omega c_{\min}} \right),$$

which proves the first assertion (3.61).

(iii) Next, we define

$$z(x, t) = \operatorname{Re}\{u(x) e^{-i\omega t}\} - y(x, t), \quad (3.63)$$

which solves the wave equation (3.5) with $f \equiv g_D \equiv g_N \equiv g_S \equiv 0$ and

$$z(x, 0) = \operatorname{Re}\{u(x)\} - y_0(x), \quad z_t(x, 0) = \omega \operatorname{Im}\{u(x)\} - y_1(x).$$

By the identity

$$\begin{aligned} u(x) &= \int_0^T u(x) e^{-i\omega t} e^{i\omega t} dt = \int_0^T \left(\operatorname{Re}\{u(x) e^{-i\omega t}\} + i \operatorname{Im}\{u(x) e^{-i\omega t}\} \right) e^{i\omega t} dt \\ &= \int_0^T \left(\operatorname{Re}\{u(x) e^{-i\omega t}\} + \frac{i}{\omega} \operatorname{Re}\{-i\omega u(x) e^{-i\omega t}\} \right) e^{i\omega t} dt \end{aligned}$$

and the definitions (3.63) and (3.60), we obtain

$$u(x) - v(x) = \int_0^T \left(z(x, t) + \frac{i}{\omega} z_t(x, t) \right) e^{i\omega t} dt.$$

Since $e^{\pm i\omega t}$ is T -periodic, integration by parts over $[0, T]$ on z_t and the definition of z yield

$$u(x) - v(x) = 2 \int_0^T z(x, t) e^{i\omega t} dt + \frac{i}{\omega} y(x, t) \Big|_{t=0}^{t=T}. \quad (3.64)$$

Now, we define

$$\psi(x) = \int_0^T z(x, t) e^{i\omega t} dt. \quad (3.65)$$

Then for $\varphi \in H_D^1(\Omega)$, integration by parts in time yields

$$\begin{aligned} (-k^2 \psi - \nabla \cdot (a \nabla \psi), \varphi) &= \int_0^T \int_\Omega (-k^2(x) z(x, t) - \nabla \cdot (a(x) \nabla z(x, t))) e^{i\omega t} \bar{\varphi}(x) dt dx \\ &= \int_0^T \int_\Omega \left[\frac{1}{c^2(x)} z_{tt}(x, t) - \nabla \cdot (a(x) \nabla z(x, t)) \right] e^{i\omega t} \bar{\varphi}(x) dt dx \\ &\quad - \frac{1}{T} \int_\Omega \frac{1}{c^2(x)} z_t(x, t) \bar{\varphi}(x) \Big|_0^T dx + \frac{i\omega}{T} \int_\Omega \frac{1}{c^2(x)} z(x, t) \bar{\varphi}(x) \Big|_0^T dx. \end{aligned}$$

Since z solves the homogeneous wave equation (3.5), we have

$$(-k^2 \psi - \nabla \cdot (a \nabla \psi), \varphi) = \int_\Omega f_\psi(x) \bar{\varphi}(x) dx$$

with

$$f_\psi(x) = \frac{i\omega}{c^2(x)T} y(x, t) \Big|_{t=0}^{t=T} - \frac{1}{c^2(x)T} y_t(x, t) \Big|_{t=0}^{t=T} \in L^2(\Omega). \quad (3.66)$$

Moreover, the definition (3.65) of ψ and the boundary condition (3.5b) follow

$$\int_{\Gamma_S} \left(a \frac{\partial \psi}{\partial n} - ik\sqrt{a}\psi \right) \bar{\varphi} ds = \int_{\Gamma_S} \int_0^T \left(-\frac{\sqrt{a(x)}}{c(x)} z_t(x, t) - ik(x)\sqrt{a(x)} z(x, t) \right) e^{i\omega t} \bar{\varphi}(x) dt ds.$$

By integration by parts, we get

$$\int_{\Gamma_S} \left(a(x) \frac{\partial}{\partial n} \phi(x) - ik(x)\sqrt{a(x)} \phi(x) \right) \bar{\varphi}(x) ds = \int_{\Gamma_S} g_\phi(x) \bar{\varphi}(x) ds$$

with

$$g_\psi(x) = -\frac{\sqrt{a}}{cT} (z(x, T) - z(x, 0)) = -\frac{\sqrt{a}}{cT} (y(x, T) - y(x, 0)) \in L^2(\Gamma_S). \quad (3.67)$$

Since z satisfies the boundary conditions (3.5c) and (3.5d) with $g_D \equiv g_N \equiv 0$, by the definition (3.65) we get

$$\psi(x) = 0, \quad x \in \Gamma_D, \quad \frac{\partial}{\partial n} \psi(x) = 0, \quad x \in \Gamma_N.$$

It results that ψ solves the Helmholtz equation with $f \equiv f_\psi$, $g_S \equiv g_\psi$, and $g_D \equiv g_N \equiv 0$. Moreover, the definitions (3.66) and (3.67) yield

$$\|f_\psi\|_{L^2(\Omega)} \leq \frac{\omega}{c_{\min}^2 T} \|y(\cdot, T) - y(\cdot, 0)\|_{L^2(\Omega)} + \frac{1}{c_{\min}^2 T} \|y_t(\cdot, T) - y_t(\cdot, 0)\|_{L^2(\Omega)} \quad (3.68)$$

and, together with the constant $C_{tr} > 0$ from the trace operator,

$$\|g_\psi\|_{L^2(\Gamma_S)} \leq \frac{C_{tr} \|\sqrt{a}\|_{L^\infty(\Omega)}}{c_{\min} T} \|y(\cdot, T) - y(\cdot, 0)\|_{H^1(\Omega)}. \quad (3.69)$$

Under the assumption that the Helmholtz equation has a unique solution, by the boundedness of the inverse [56, §6.2], there is a constant $C_H > 0$ independent of ϕ , f_ϕ , and g_ϕ such that

$$\|\psi\|_{L^2(\Omega)} \leq C_H (\|f_\psi\|_{L^2(\Omega)} + \|g_\psi\|_{L^2(\Gamma_S)}).$$

This, together with (3.64), (3.68), and (3.69), results

$$\|u - v\|_{L^2(\Omega)} \leq C (\|y(\cdot, T) - y(\cdot, 0)\|_{H^1(\Omega)} + \|y_t(\cdot, T) - y_t(\cdot, 0)\|_{L^2(\Omega)})$$

with

$$C = \frac{2C_H \omega}{c_{\min}^2 T} + \frac{C_H}{c_{\min}^2 T} + \frac{2C_H C_{tr} \|\sqrt{a}\|_{L^\infty(\Omega)}}{c_{\min} T} + \frac{1}{\omega},$$

which concludes the assertion (3.62). \square

Following Theorem 10, we have the following estimate for the semi-discrete y_h and fully-discrete solution $y_{h,\Delta t}$ of (3.5).

Corollary 3. *Let $u \in H^1(\Omega)$ be the unique complex-valued solution of (2.9) and $y \in C^0([0, T]; H^1(\Omega)) \cap C^1([0, T]; L^2(\Omega))$ be the real-valued solution of (3.5) with initial value $(y_0, y_1) \in H^1(\Omega) \times L^2(\Omega)$ and $T = \frac{2\pi}{\omega}$. Furthermore, let $u_h \in V_h \subset H^1(\Omega)$ be the Galerkin approximation of u , $y_h \in C^0([0, T]; V_h) \cap C^1([0, T]; W_h)$, $W_h \subset L^2(\Omega)$, be the semi-discrete approximation of y ,*

and $y_{h,\Delta t}$ be the full-discrete approximation of y , where

$$v(x) = \int_0^T (y(x,t) + \frac{i}{\omega} y_t(x,t)) e^{i\omega t} dt, \quad (3.70a)$$

$$v_h(x) = \int_0^T (y_h(x,t) + \frac{i}{\omega} (y_h)_t(x,t)) e^{i\omega t} dt, \quad (3.70b)$$

$$v_{h,\Delta t}(x) = \int_0^T (y_{h,\Delta t}(x,t) + \frac{i}{\omega} (y_{h,\Delta t})_t(x,t)) e^{i\omega t} dt. \quad (3.70c)$$

Suppose that there are constants r, s, C_h , and C_t independent of y, h , and Δt such that

$$\max\{\|y_h - y\|, \|(y_h)_t - y_t\|\} \leq C_h h^{r+1}, \quad \|y_{h,\Delta t} - y_h\| \leq C_t \Delta t^s. \quad (3.71)$$

Then there is a constant $C > 0$ such that

$$\|u - v_{h,\Delta t}\|_{L^2(\Omega)} \leq C(h^{r+1} + \Delta t^s + \|y(\cdot, T) - y(\cdot, 0)\|_{H^1(\Omega)} + \|y_t(\cdot, T) - y_t(\cdot, 0)\|_{L^2(\Omega)}).$$

Proof. The estimation follows directly from the assumptions (3.71), the definition (3.70), and Theorem 10, together with

$$\|u - v_{h,\Delta t}\| \leq \|u - v\| + \|v - v_{h,\Delta t}\| \leq \|u - v\| + \|v - v_h\| + \|v_h - v_{h,\Delta t}\|.$$

Remark 8. Note that $\|v - v_h\|$ is the error in spatial discretization, $\|v_h - v_{h,\Delta t}\|$ the error in the numerical time integration and $\|u - v\|$ the departure from the solution of the wave equation $y(\cdot, 0), y_t(\cdot, 0)$ after one period at $t = T$. Following Theorem 10, the error $\|u - v\|$ tends to zero when the misfit $\|y(\cdot, T) - y_0\|_{H^1}$ and $\|y_t(\cdot, T) - y_1\|_{L^2}$ tend to zero. Furthermore, v_h and $v_{h,\Delta t}$ converge to v and v_h , respectively, when $h, \Delta t \rightarrow 0$. □

3.6 Initial run-up phase

In [114], Mur suggested that convergence of the time-harmonically forced wave equation (3.5) to the time-harmonic asymptotic regime can be accelerated by pre-multiplying the time-harmonic sources in (3.5) with the smooth transient function θ_{tr} from zero to one,

$$\theta_{tr}(t) = \begin{cases} \left(2 - \sin\left(\frac{t}{t_{tr}} \frac{\pi}{2}\right)\right) \sin\left(\frac{t}{t_{tr}} \frac{\pi}{2}\right), & 0 \leq t \leq t_{tr}, \\ 1, & t \geq t_{tr}, \end{cases} \quad (3.72)$$

active during the initial time interval $[0, t_{tr}]$, $t_{tr} = mT$, and $m \in \mathbb{N}_0$ – see also [26]. Figure 3.8 monitors the function $\theta_{tr}(t)$, $t \in [0, t_{tr}]$.

For a given initial value $(\widehat{y}_0, \widehat{y}_1)$ we first solve (3.5) over $t \in [0, t_{tr}]$ with $\text{Re}\{f(x)e^{-i\omega t}\}$ replaced by $\theta(t) \text{Re}\{f(x)e^{-i\omega t}\}$, and $\text{Re}\{g_S(x)e^{-i\omega t}\}$, $\text{Re}\{g_N(x)e^{-i\omega t}\}$, and $\text{Re}\{g_D(x)e^{-i\omega t}\}$, respectively. For a real source term, e.g. $f(x) \in \mathbb{R}$, the time-harmonic solution is given by

$$\text{Re}\{f(x)e^{-i\omega t}\} = f(x) \cos(\omega t).$$

Instead of $\cos(\omega t)$ (Figure 3.8b), the source f is multiplied with $\cos(\omega t)\theta_{tr}(t)$ (Figure 3.8c).

Let \widehat{y} be the corresponding solution of (3.5). Then we set

$$y_0(x) = \widehat{y}(x, t_{tr}), \quad y_1(x) = \widehat{y}_t(x, t_{tr}). \quad (3.73)$$

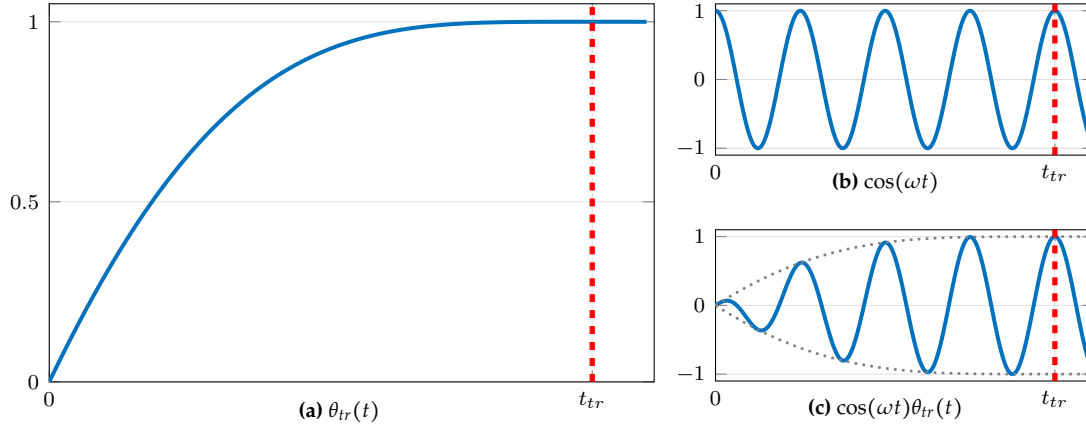


Figure 3.8: Smooth transient kernel: (a) transient kernel $\theta_{tr}(t)$ from $t = 0$ to $t = t_{tr}$; (b) $\cos(\omega t)$; (c) $\cos(\omega t)\theta_{tr}(t)$.

Here we always set $\widehat{y}_0 = \widehat{y}_1 = 0$. Moreover, the size of the initial time interval t_{tr} is chosen so that

$$t_{tr} = mT, \quad m \in \mathbb{N}_0, \quad T = (2\pi)/\omega,$$

and the wave travels at least once across the entire computational domain Ω during the time interval $[0, t_{tr}]$; hence, we set

$$m = \left\lceil \frac{\sqrt{\text{width}^2 + \text{height}^2 + \text{depth}^2}}{c_{\min} T} \right\rceil.$$

3.7 CMCG Algorithm and computational cost

We list the full CMCG Algorithm 2 – see [25, 26, 83, 80].

Every iteration of the CMCG Algorithm requires the solution of one strongly elliptic problem (3.38a) together with the solution of one forward and one backward wave equation over one period $[0, T]$ and $m - 1$ backward wave equations for $m = 2, 3$. For J_2 or J_3 , however, the additional backward (or adjoint) wave equations are not solved over the entire period $[0, T]$ but only over the shorter time interval $[0, t_1]$ for J_2 , or $[0, t_1]$ and $[0, t_2]$ for J_3 . Hence if we neglect the (much smaller) computational cost of the single elliptic problem, the relative increased cost of using either J_2 or J_3 over that of using J_1 is proportional precisely to the additional number of time-steps required for those (one or two) additional shorter time intervals $[0, t_j]$. The computational cost of each single time-step is always merely proportional to the number of degrees of freedom in the spatial FE discretization, when explicit time integration is used to solve the (forward and backward) wave equations.

In Table 3.1, we estimate the computational effort and storage of the CMCG method using different penalty functionals relative to that using the original functional J_1 . Here $n_T = T/\Delta t$ denotes the number of time steps in the solution of the forward wave equation over one period $[0, T]$. The functional J_m is slightly more expensive for $m = 2, 3$, because each time instant t_j , $1 \leq j \leq m$, needs the solution of an adjoint wave equation, but only over $[0, t_j]$. Therefore, the functional J_2 with $t_1 = T/4$ and $t_2 = T$ requires n_T time-steps to obtain the forward solution (3.5) at $t = T$ and $(n_T + n_T/4)$ time-steps for the two adjoint wave equations (3.37), which yields a total of $2.25 \times n_T$ time-steps. Hence, the use of the functional J_2 leads to a

Algorithm 2: CMCG Algorithm

-
- 1 **Input:** Initial guess $v^{(0)} = (y_0^{(0)}, y_1^{(0)})$, tolerance Tol , cost functional J (e.g. $J = J_1$)
 - 2 **Output:** u_h
 - 3 Solve the forward and the backward wave equations (3.5) and (3.37) to determine the gradient of J , $g^{(0)} = J'(v^{(0)})$, defined by (3.34).
 - 4 Solve the coercive elliptic problem (3.38a) with $g = g^{(0)}$ to determine the new search direction $\tilde{g}^{(0)}$.
 - 5 Set $r^{(0)} = d^{(0)} = \tilde{g}^{(0)}$, $relres^{(0)} = 1$, and $\ell = 0$.
 - 6 **while** $relres^{(\ell)} > Tol$ **do**
 - 7 Solve the wave equation (3.5) with $f = g_D = g_S = g_N = 0$ and the initial values $d^{(\ell)} = (d_0^{(\ell)}, d_1^{(\ell)})$ and the backward wave equation (3.37).
 - 8 Compute the gradient $g^{(\ell)} = J'(d^{(\ell)})$ defined by (3.34).
 - 9 Solve the coercive elliptic problem (3.38a) with $g = g^{(\ell)}$ to get $\tilde{g}^{(\ell)}$.
 - 10
$$\alpha_\ell = \frac{\|\sqrt{a}\nabla r_0^{(\ell)}\|_{L^2(\Omega)}^2 + \|(1/c)r_1^{(\ell)}\|_{L^2(\Omega)}^2}{(a\nabla\tilde{g}_0^{(\ell)}, \nabla d_0^{(\ell)})_{L^2(\Omega)} + ((1/c^2)\tilde{g}_1^{(\ell)}, d_1^{(\ell)})_{L^2(\Omega)}}.$$
 - 11 $v^{(\ell+1)} = v^{(\ell)} - \alpha_\ell d^{(\ell)}$
 - 12 $r^{(\ell+1)} = r^{(\ell)} - \alpha_\ell \tilde{g}^{(\ell)}$
 - 13
$$\beta_\ell = \frac{\|\sqrt{a}\nabla r_0^{(\ell+1)}\|_{L^2(\Omega)}^2 + \|(1/c)r_1^{(\ell+1)}\|_{L^2(\Omega)}^2}{\|\sqrt{a}\nabla r_0^{(\ell)}\|_{L^2(\Omega)}^2 + \|(1/c)r_1^{(\ell)}\|_{L^2(\Omega)}^2}$$
 - 14 $d^{(\ell+1)} = r^{(\ell+1)} + \beta_\ell d^{(\ell)}$
 - 15
$$relres^{(\ell+1)} = \sqrt{\frac{\|\sqrt{a}\nabla r_0^{(\ell+1)}\|_{L^2(\Omega)}^2 + \|(1/c)r_1^{(\ell+1)}\|_{L^2(\Omega)}^2}{\|\sqrt{a}\nabla r_0^{(0)}\|_{L^2(\Omega)}^2 + \|(1/c)r_1^{(0)}\|_{L^2(\Omega)}^2}}$$
 - 16 $\ell \leftarrow \ell + 1$
 - 17 **return** $u_h = y_0^{(\ell)} + \frac{i}{\omega}y_1^{(\ell)}$, $v^{(\ell)} = (y_0^{(\ell)}, y_1^{(\ell)})$
-

$(2.25 - 2)/2 = 12.5\%$ relative increase in computational cost over J_1 . Similarly, the functional J_3 with $t_1 = T/4$, $t_2 = T/2$, and $t_3 = T$ requires n_T time-steps for the forward solution and $(n_T/4 + n_T/2 + n_T)$ time-steps for the adjoint solutions, which leads to $2.75 \times n_T$ time-steps in total.

In contrast, the functional J_∞ only needs the solution of a single forward and a single backward wave equation over $[0, T]$, which yields the same computational effort as the functional J_1 . However, as the adjoint wave equation then involves the entire history of the forward solution, it becomes necessary to store the forward solution at all time steps $t_j = j\Delta t$, $1 \leq j \leq n_T$.

3.8 Numerical experiments I: parameter study

Here we present a series of numerical examples that illustrate the accuracy and convergence behavior of the CMCG method. First, we verify that the numerical solution u_h of (2.9), obtained with the CMCG method, converges to the numerical solution u_h^* obtained with a direct solver for the same spatial FE discretization as the time step $\Delta t \rightarrow 0$ in the numerical integration of (3.5). Second, we consider a plane wave exact

| | J_1 | J_1^{cc} | J_2 | J_3 | J_∞ | $J_1^{fp,cc}$ |
|-----------------------------|-------|------------|-------|-------|------------|---------------|
| number of periods $[0, T]$ | 2 | 2 | 2.25 | 2.75 | 2 | 2 |
| relative computational time | 1 | 1 | 1.125 | 1.375 | 1 | 1 |
| relative memory requirement | 1 | 1 | 2 | 3 | n_T | 1 |
| sound-soft scattering | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| sound-hard scattering | × | ✓ | ✓ | ✓ | ✓ | ✓ |
| physically bounded domain | × | × | × | ✓ | ✓ | ✓ |

Table 3.1: Computational cost and storage of the CMCG method using the original functional J_1 combined with the compatibility condition, denoted by J_1^{cc} , and with the filtering procedure, denoted by $J_1^{fp,cc}$, or the alternative functionals J_m , J_∞ relative to that using J_1 . Crosses indicate that the CMCG method generally does not yield the correct time-harmonic solution. The top line indicates the number of periods $[0, T]$ for which the forward or backward wave equation (3.5) must be solved.

solution and verify that the CMCG method achieves the expected rates of convergence. Next, we consider the one-dimensional exact solution and progressively increase the wave number to study how the number of CG iterations in the CMCG Algorithm depends on the wave number, while refining the mesh to keep the number of mesh points per wave length fixed. Fourth, we evaluate different stopping criteria for the CG iteration in the CMCG Algorithm from Section 3.7. We also compare the CMCG Algorithm to a long-time solution of the wave equation without controllability (“do-nothing” approach) to demonstrate its effectiveness, in particular for non-convex obstacles. Moreover, we show how an initial run-up yields a judicious initial guess (y_0, y_1) for the CG iteration thereby further accelerating convergence. Finally, we study how the number of CG iterations in the CMCG Algorithm depends on the width of the computational domain $\Omega = (0, L) \times (0, 1)$ and the time interval $(0, mT)$, $T = \frac{2\pi}{\omega}$, while increasing the width L and the number of time periods m .

3.8.1 Semi-discrete convergence

First, we consider a 1D example to show for a fixed FE-mesh that the numerical solution u_h , obtained with the CMCG method, converges to the numerical solution u_h^* , obtained with a direct solver from the same \mathcal{P}^3 -FE discretization of the Helmholtz equation (2.9), as $\Delta t \rightarrow 0$.

Now, let u be the exact solution of (2.9) in $\Omega = (0, 1)$ with $\omega = k = 8\pi$, $a = c = 1$ and $f = 0$:

$$u(x) = -\exp(ikx), \quad \text{with} \quad u(0) = -1, \quad u'(1) - ik u(1) = 0.$$

For the time integration of (3.5), we use the standard explicit fourth-order Runge-Kutta (RK4) method. Usually we avoid inverting the mass-matrix at each time step via order preserving mass-lumping which, however, introduces an additional spatial discretization error. Here to ensure a consistent comparison, we compute both u_h and u_h^* without mass-lumping. For the CG iteration, we choose $y_0^{(0)} \equiv 0$, $y_1^{(0)} \equiv 0$ and fix the tolerance to $Tol = 10^{-14}$ to ensure convergence to machine precision accuracy.

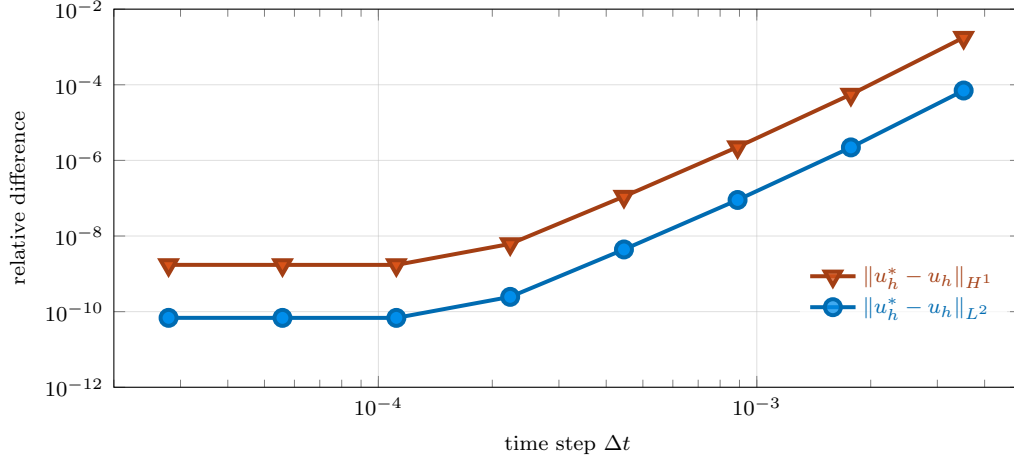


Figure 3.9: Semi-discrete convergence: Comparison of the numerical solution u_h , obtained with the CMCG method, and u_h^* , obtained with a direct solver.

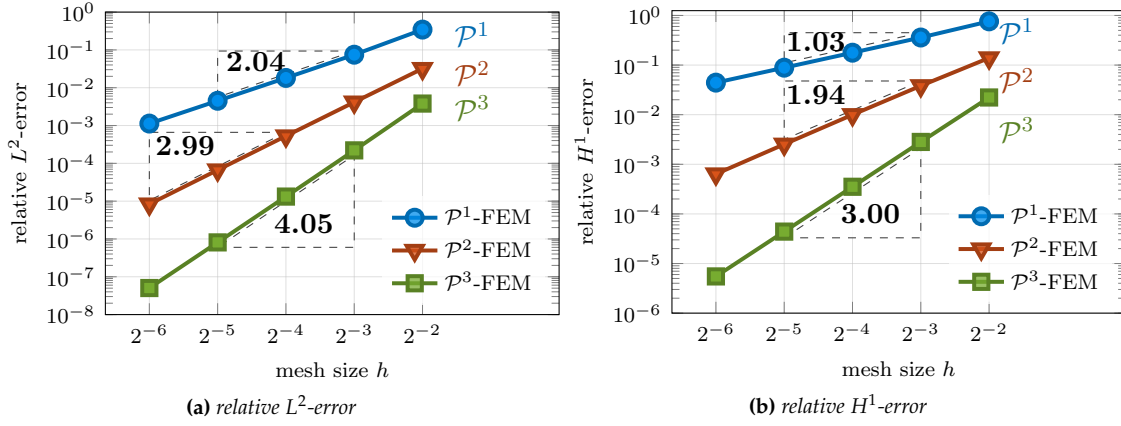


Figure 3.10: Plane wave solution: the relative error of the numerical solution in the unit square using the CMCG method for varying mesh size h .

In Figure 3.9, we monitor both the L^2 - and H^1 -difference between the numerical solution u_h^* , obtained with a direct solver, and u_h , obtained with the CMCG method. As expected, for increasingly smaller Δt and a fixed stringent tolerance in the CG iteration, the numerical solution u_h of the CMCG method always converges to the discrete solution u_h^* of (2.9) for the same FE discretization.

3.8.2 Plane Wave Solution

We verify that the controllability method with the functional J_1 from (3.33) indeed achieves the same rate of convergence as the direct solution of the Helmholtz equation. Here, we consider a plane wave solution $u(x, y) = u^{in}(x, y)$, where

$$u^{in}(x, y) = \exp(ik(x \cos(\theta) + y \sin(\theta))) \quad (3.74)$$

with the angle of incidence $\theta = 135^\circ$ and wave number $k = \omega = 2\pi$ ($c \equiv 1$). Hence we consider (2.9) in the unit square $\Omega = (0, 1) \times (0, 1)$ and set the boundary conditions accordingly with $\Gamma_D = [0, 1] \times \{0, 1\}$ and $\Gamma_S = \{0, 1\} \times (0, 1)$. We use \mathcal{P}^r -FE with (order preserving) mass-lumping for the spatial discretization with $r = 1, \dots, 3$ and the classical fourth-order Runge-Kutta (RK4) method for the time integration of (3.5). The tolerance in the CMCG Algorithm (Section 3.7) is set to $Tol = 10^{-11}$.

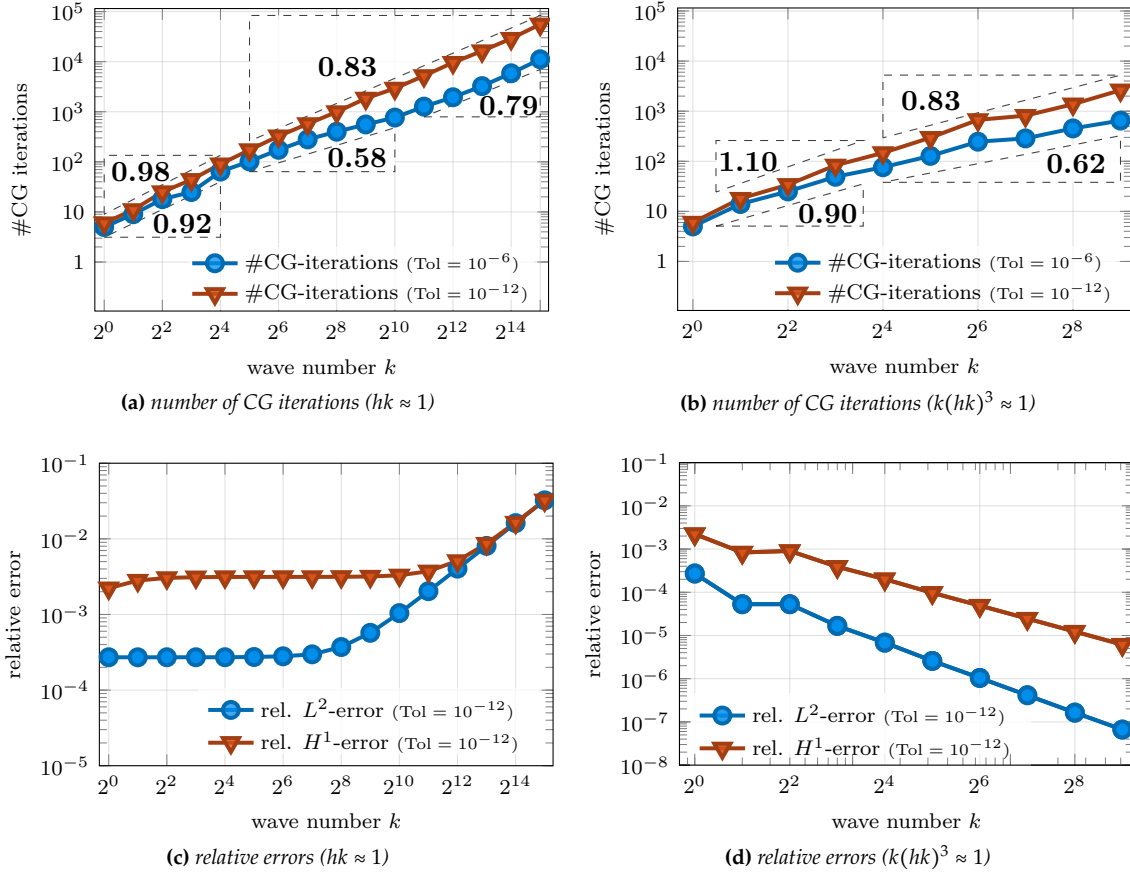


Figure 3.11: CMCG iterations vs. wave number: One-dimensional Helmholtz equation (2.9) with Dirichlet condition at $x = 0$ and Sommerfeld condition at $x = 1$. The number of CMCG iterations vs. the wave number k for a tolerance $\text{Tol} = 10^{-6}, 10^{-12}$: (a) with pollution error; (b) without pollution error. The relative L^2 - and H^1 -errors between the exact solution of (2.9) and the numerical solution, obtained with the CMCG method, for a tolerance 10^{-12} ; (c) with pollution error; (d) without pollution error.

In Figure 3.10, we observe that the relative error of the numerical solution, obtained with the CMCG method for a sequence of meshes with mesh size $h = 2^{-\ell}$, $\ell = 2, \dots, 6$, converges with the (expected) optimal rate as h^{r+1} (or h^r) with respect to the L^2 -norm (or H^1 -norm) – see Section 2.4. Clearly, as the mesh size decreases we also reduce the time-step Δt for numerical stability.

3.8.3 Number of CG iterations in the CMCG method for various frequency

Even when the geometry and the problem's set-up remain fixed, classical iterative methods require increasingly many iterations for the discrete Helmholtz problem at higher wave number, $k = \omega/c$ [7, 55, 50]. To study how the number of CG iterations in the CMCG Algorithm depends on k (or ω), we again consider the one-dimensional exact solution,

$$u(x) = -\exp(ikx),$$

(2.9) with $a \equiv c \equiv 1$ and $f \equiv 0$ together with $u(0) = -1$ at $x = 0$ and an (exact) absorbing boundary condition at $x = 1$. For the numerical solution of (3.5), we use \mathcal{P}^3 -FE in space and the RK4 method in time.

Now, we progressively increase the wave number as $k = \omega = 2^i$, $i = 0, \dots, 15$, while refining the mesh to keep the number of mesh points per wave length fixed,

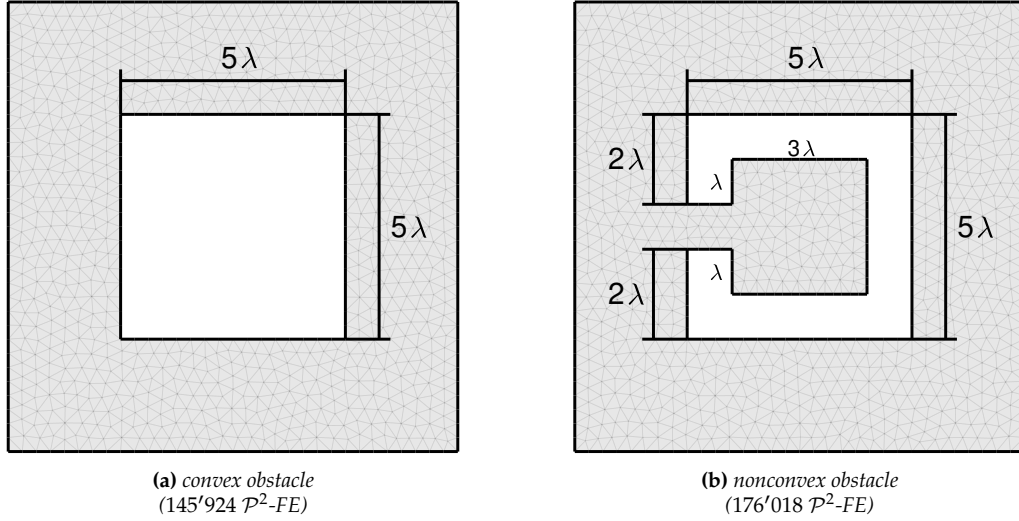


Figure 3.12: Computational domain Ω with a convex square (a) or a nonconvex cavity (b) shaped obstacle

that is $kh \approx 1$. In Figure 3.11a, we show the number of CMCG iterations required to reach a fixed tolerance of $Tol = 10^{-6}, 10^{-12}$ in Step 6 in Algorithm 2. We observe that the number of CG iterations vs. k essentially increases linearly with an approximate slope of $0.58 - 0.92$ for $Tol = 10^{-6}$ and $0.83 - 0.98$ for $Tol = 10^{-12}$.

When the number of mesh points per wave length remains fixed as k increases, the accuracy of the numerical solution generally suffers from the pollution effect [7, 50]. Indeed, as shown in Figure 3.11c the error in the CMCG solution increases with frequency if $hk \approx 1$. To avoid the pollution error, we thus repeat the same numerical experiment but let the mesh size h decrease slightly faster as $k(kh)^3 \approx 1$ while increasing k – see [7]. Now as shown Figure 3.11d both the L^2 - and H^1 -error remain bounded as the frequency increases. Again in Figure 3.11b, we observe a linear increase in the number of CMCG iterations vs. k with an approximate slope of $0.62 - 0.90$ for $Tol = 10^{-6}$ and $0.83 - 1.10$ for $Tol = 10^{-12}$. Hence, the number of CG iterations in the CMCG method generally increases with the wave number k both with and without control of the pollution error.

3.8.4 CG iteration and initial run-up

Next, we compare different stopping criteria for the CG iteration in the CMCG Algorithm 2. We then illustrate how the CMCG method greatly accelerates the convergence of a solution of the wave equation to its long-time asymptotic limit, in particular for nonconvex obstacles. Finally, we show how an initial run-up yields a judicious initial guess for the CG iteration, which further accelerates the convergence of the CMCG Algorithm.

Hence, we consider a two-dimensional sound-soft scattering problem (2.9) with $a = c = 1$, $k = \omega = 2\pi$, $f = g_D = g_N = 0$, and $g_S = -(\frac{\partial}{\partial n} - ik)u^{in}$ in a bounded square domain $\Omega = (0, 10\lambda) \times (0, 10\lambda)$, $\lambda = 1$, either with a convex obstacle or a semi-open square shaped cavity. On the boundary Γ_D of the obstacle, we impose a homogeneous Dirichlet condition and on the exterior boundary Γ_S a Sommerfeld-like absorbing condition on the total wave field. The incident plane wave u^{in} given by (3.74) impinges with the angle $\theta = 135^\circ$ upon the obstacle.

CG iteration and stopping criteria

In Algorithm 2 in Section 3.7, the CMCG method terminates at the ℓ -th iteration and returns

$$u_h^{(\ell)} = y_0^{(\ell)} + (i/\omega)y_1^{(\ell)} \quad (3.75)$$

when the *relative CG-residual* in Step 15,

$$|u_h^{(\ell)}|_{\text{CG}} := \sqrt{\frac{\|\sqrt{a} \nabla r_0^{(\ell+1)}\|_{L^2(\Omega)}^2 + \|(1/c) r_1^{(\ell+1)}\|_{L^2(\Omega)}^2}{\|\sqrt{a} \nabla r_0^{(0)}\|_{L^2(\Omega)}^2 + \|(1/c) r_1^{(0)}\|_{L^2(\Omega)}^2}}, \quad (3.76)$$

is less than the tolerance *Tol*. Indeed, a small CG-residual indicates that the gradient of J_1 is sufficiently small at $(y_0^{(\ell)}, y_1^{(\ell)})$ and thus that a minimum has been reached.

Since the cost functional J_1 also vanishes at the minimum, we can use J_1 itself, instead of its gradient, to monitor convergence of the CG iteration via the *relative periodicity misfit*,

$$|u_h^{(\ell)}|_J := \frac{\sqrt{J_1(y_0^{(\ell)}, y_1^{(\ell)})}}{\|f\|_{L^2(\Omega)} + \|g_S\|_{L^2(\Gamma_S)}}. \quad (3.77)$$

In fact, the convergence criterion (3.77) is typically used in long-time simulations of the wave equation without controllability (“do-nothing” approach) to determine the current misfit from periodicity in the energy norm.

Alternatively, we may also directly compute the current *relative Helmholtz residual* from (2.9):

$$|u_h^{(\ell)}|_H := \frac{\|\mathbf{A}_h \mathbf{u}_h^{(\ell)} - \mathbf{F}_h\|_2}{\|\mathbf{F}_h\|_2}, \quad (3.78)$$

where \mathbf{A}_h and \mathbf{F}_h result from a FE discretization in Section 2.4 of (2.9) without mass-lumping, $\mathbf{u}_h^{(\ell)}$ corresponds to the discrete vector of FE coefficients of $u_h^{(\ell)}$, and $\|\cdot\|_2$ denotes the discrete Euclidean norm.

In Figure 3.13, we monitor $|u_h^{(\ell)}|_{\text{CG}}$, $|u_h^{(\ell)}|_J$, and $|u_h^{(\ell)}|_H$, defined in (3.76)–(3.78) for the CMCG solution $u_h^{(\ell)}$ at the ℓ -th CG iteration. Whether for a convex (Figure 3.12a) or a nonconvex (Figure 3.12b) obstacle, both the CG-residual $|u_h^{(\ell)}|_{\text{CG}}$ and the periodicity misfit $|u_h^{(\ell)}|_J$ rapidly converge to zero. In contrast, the Helmholtz residual $|u_h^{(\ell)}|_H$ stagnates beyond the first hundred CG iterations, as the mass-matrix that appears in \mathbf{A}_h in (3.78) is discretized here without mass-lumping. That additional discretization error together with the numerical error in the time integration of (3.5) both prevent the discrete Helmholtz residual $|u_h^{(\ell)}|_H$ from converging to zero; hence, (3.78) is generally not a reliable stopping criterion for the CMCG method, unless the spatial FE discretizations used in (2.9) and (2.19) are identical.

CMCG method vs. long-time wave equation solver

In general, the solution $w(x, t)$ of the time-harmonically forced wave equation (3.5) converges asymptotically to the time-harmonic solution [96, 11]

$$w(x, t) \sim \text{Re}\{u(x) e^{-i\omega t}\} \quad \text{as } t \rightarrow +\infty, \quad (3.79)$$

where u is the (unique) solution of the Helmholtz equation (2.9) – see Section 3.2.1. Thus, with a wave equation solver at hand, one can in principle compute u from

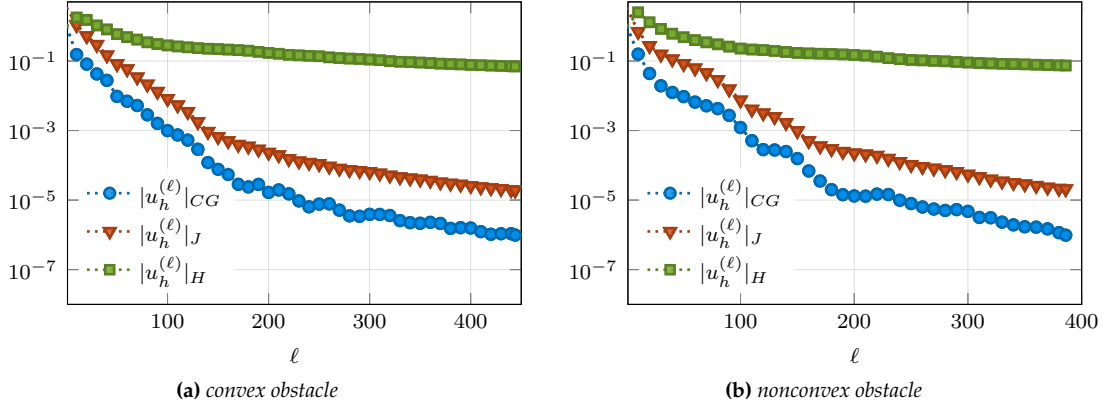


Figure 3.13: CG iterations and stopping criteria: relative CG residual $|u_h^{(\ell)}|_{CG}$ in (3.76), Helmholtz residual $|u_h^{(\ell)}|_H$ in (3.78), and periodicity mismatch $|u_h^{(\ell)}|_J$ in (3.77) at the ℓ -th CG iteration.

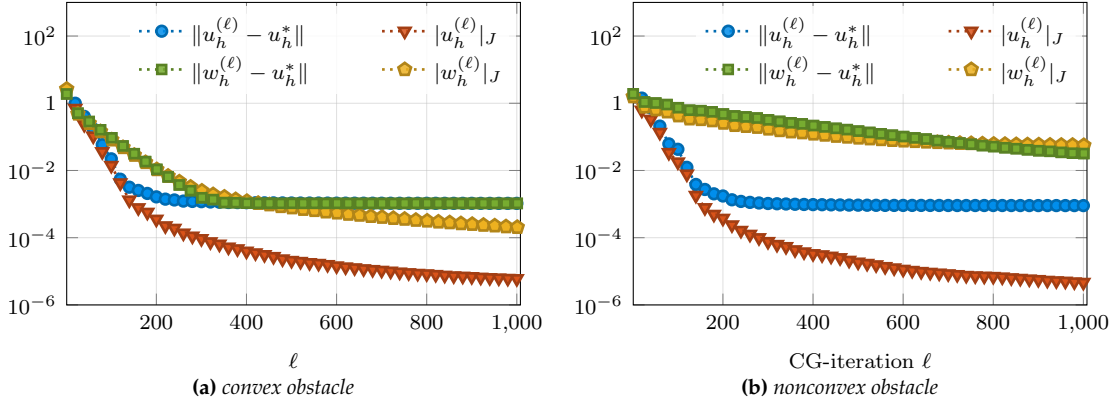


Figure 3.14: CMCG method vs. long-time wave equation solver: plane wave scattering from a convex (a) or a nonconvex obstacle (b). Comparison between the numerical solution, $u_h^{(\ell)}$, obtained with the CMCG method at the ℓ -th CG iteration and the approximate solution $w_h^{(\ell)}$, obtained via (3.80) from the solution of the wave equation at time $t = \ell T$ without controllability.

w by solving (3.5) without controllability until a quasi-periodic regime is reached. Given the current value of $w(\cdot, t)$ at time $t = \ell T$, $\ell \geq 1$, one can extract it from the complex-valued approximate solution of (2.9),

$$w_h^{(\ell)} := w(\cdot, \ell T) + \frac{i}{\omega} \frac{\partial}{\partial t} w(\cdot, \ell T), \quad \ell \geq 1, \quad T = (2\pi)/\omega, \quad (3.80)$$

which converges to u as $\ell \rightarrow +\infty$. This “do-nothing” approach only requires the time integration of (3.5) without controllability or CG iteration, but it may converge arbitrarily slowly for nonconvex obstacles due to trapped modes [26, 83].

In Figure 3.14, we monitor the periodicity misfit of $|u_h^{(\ell)}|_J$ and $|w_h^{(\ell)}|_J$, where $u_h^{(\ell)}$ is the CMCG solution at the ℓ -th CG iteration and $w_h^{(\ell)}$ is given by (3.80). In addition, we also compare both numerical solutions with the direct solution u_h^* of the linear system (2.19), resulting from the same underlying FE discretization, yet without mass-lumping.

We observe that the asymptotic solution $w_h^{(\ell)}$ and the CMCG solution $u_h^{(\ell)}$ indeed both converge to the time-harmonic solution u_h^* , until the additional errors caused by mass-lumping and the time discretization dominate the total error – see Section

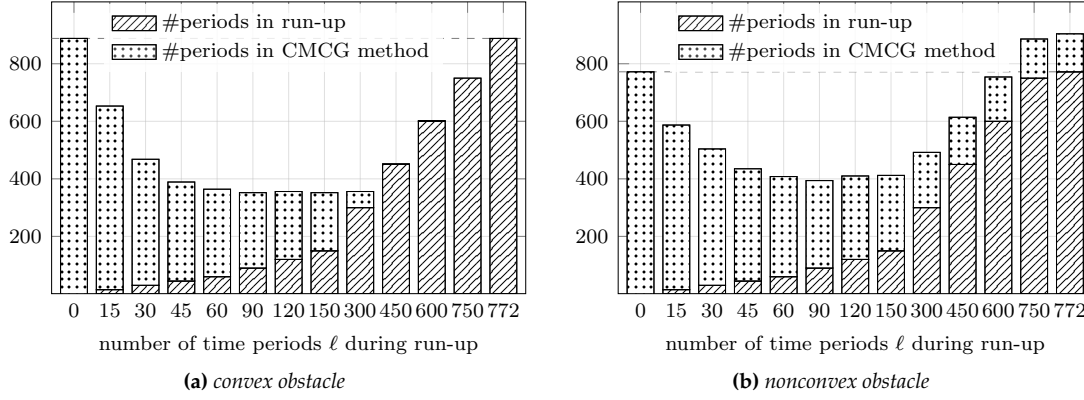


Figure 3.15: Initial run-up: plane wave scattering problems from (a) a convex or (b) a nonconvex obstacle: total number of forward and backward wave equations solved over one period $[0, T]$ until convergence.

3.8.1. For the convex obstacle, the number of CG iterations required by $u_h^{(\ell)}$ is only half the number of time periods needed for $w_h^{(\ell)}$ to reach the same level of accuracy. Since each CG iteration requires not only the solution of a forward and backward wave equation but also of the elliptic problem (3.38a), simply computing a long-time solution of the time-harmonically forced wave equation (3.5) without controllability in fact proves cheaper here than the CMCG Algorithm.

For a nonconvex obstacle, however, the long-time numerical solution of the time-dependent wave equation $w_h^{(\ell)}$ converges extremely slowly and fails to reach the asymptotic time-harmonic regime even after 1000 periods. In contrast, the convergence of the CMCG solution $u_h^{(\ell)}$ remains remarkably insensitive to the non-convexity of the obstacle.

Initial run-up

In [114], Mur suggested that convergence of the time-harmonically forced wave equation (3.5) to the time-harmonic asymptotic regime can be accelerated by pre-multiplying the time-harmonic sources in (3.5) with the smooth transient function θ_{tr} given in (3.72).

Again, we consider plane wave scattering either from a convex or nonconvex obstacle – see Figure 3.12. Now, we first solve the wave equation (3.5) with the modified source terms and zero initial conditions until time $t = \ell T$, $\ell \geq 1$, which yields the time-dependent solution y_{tr} . After that initial run-up phase, we then apply the CMCG Algorithm using the initial guess

$$y_0^{(0)}(x) = y_{tr}(x, \ell T), \quad y_1^{(0)}(x) = \frac{\partial}{\partial t} y_{tr}(x, \ell T), \quad x \in \Omega.$$

To estimate the total computational effort, we count the total number of time periods for which the (forward or backward) wave equation is solved: ℓ during initial run-up and $2 \times \#iter_{CG}$ during the CG iteration. In Figure 3.15 we display the total number $2 \times \#iter_{CG} + \ell$ of time periods needed until convergence with $Tol = 10^{-6}$, as we vary the number of periods ℓ in the initial run-up.

For a convex obstacle, the CMCG Algorithm without any initial run-up requires 888 time periods. However, as above, convergence can also be achieved at a comparable computational effort simply by solving the wave equation, here with the

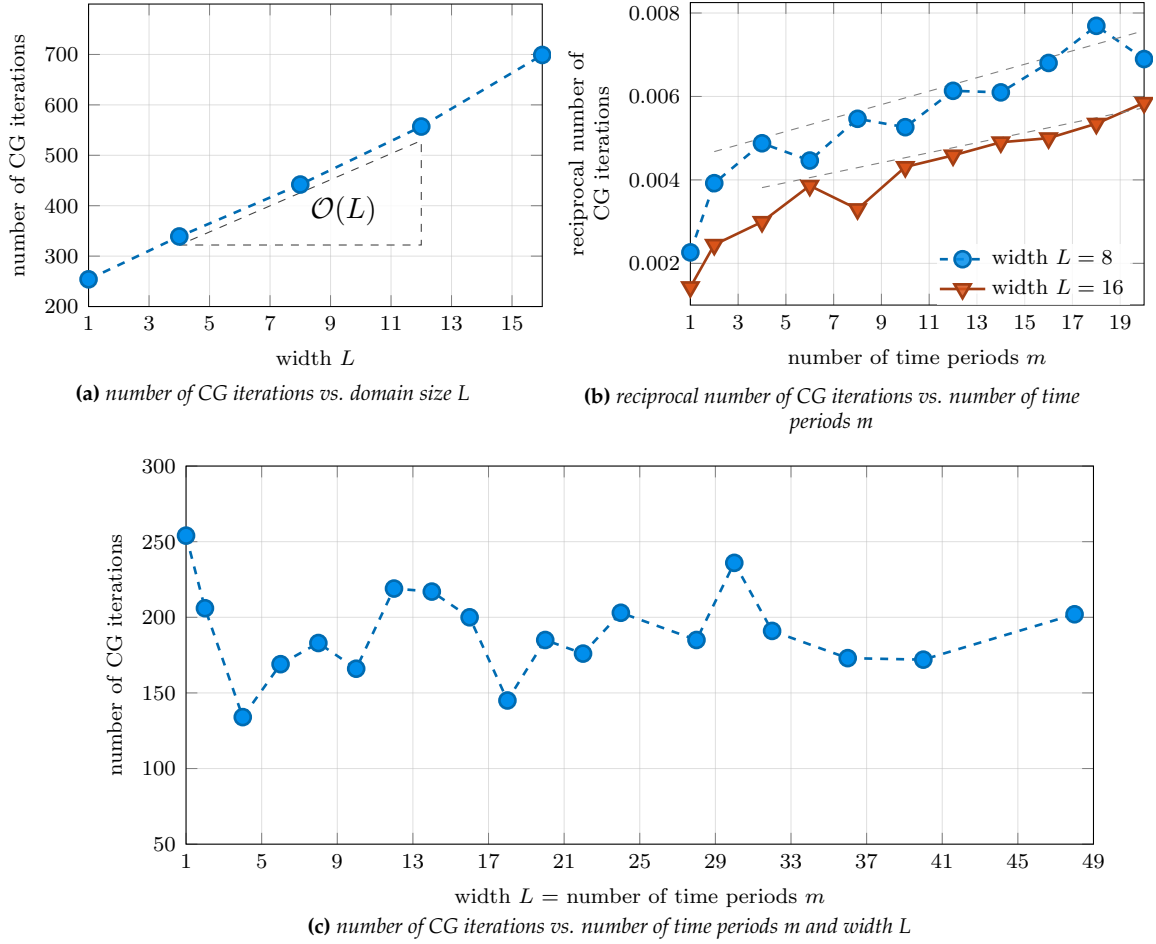


Figure 3.16: Plane wave in a semi-opened wave guide: Comparison of the number of CG iterations, obtained with the CMCG method, with the width L of the computational domain $\Omega = (0, L) \times (0, 1)$ and time interval $(0, mT)$.

source terms pre-multiplied by θ_{tr} in (3.72). Still, the minimal computational cost is achieved when both the initial run-up and the CMCG Algorithm are combined.

For the nonconvex obstacle, however, simply solving the time-harmonically forced wave equation over a very long time, let be it with or without $\theta_{tr}(t)$ smoothing, fails to reach the long-time asymptotic final time-harmonic state. Regardless of the length of the initial run-up, convergence indeed cannot be achieved here (within 1000 time periods) without controllability because of trapped modes. Nevertheless, the initial run-up always speeds up the convergence of the CMCG method by providing a judicious initial guess for the CG iteration.

3.8.5 Computational domain and time interval

Here we consider the plane wave solution $u = -u^{in}$, given by (3.74) with $\theta = 30^\circ$, of the scattering problem (2.9) with $k = \omega = 2\pi$, $a = c = 1$, and the exact source terms in a semi-open waveguide $\Omega = (0, L) \times (0, 1)$, $L > 0$. We impose a physical boundary condition on $\Gamma_D = \{x_1 = 0\} \cup \{x_1 = L\} \cup \{x_2 = 0\}$ and an impedance boundary condition on $\Gamma_S = \{x_2 = L\}$. For the numerical solution of (3.5), we use \mathcal{P}^3 -FE in space and the RK4 method in time.

Here we apply the CMCG method for a fixed CG tolerance $Tol = 10^{-6}$ to find an initial value (y_0, y_1) such that the corresponding solution y of (3.5) is mT -time

periodic with $T = 2\pi/\omega$ and $m \geq 1$. We compare the number of CG iterations with the width of the computational domain L and the number of time periods m .

Note that Theorem 6 provides constants $\lambda, \eta \in \mathbb{R}$ and a sequence of $\gamma_\ell \in H^1(\Omega)$ solving the eigenvalue problem (3.21) with $(\omega\ell)/m$ instead of $\omega\ell$:

$$(y(\cdot, t), \varphi) = (\operatorname{Re}\{u e^{-i\frac{\omega}{m}t}\}, \varphi) + (\lambda + \eta t, \varphi) + \sum_{|\ell|>1}^{\infty} (\gamma_\ell, \varphi) e^{i\frac{\omega}{m}\ell t}, \quad \varphi \in H_D^1(\Omega). \quad (3.81)$$

Since $\mathcal{H}^{d-1}(\Gamma_D), \mathcal{H}^{d-1}(\Gamma_S) > 0$, the eigenvalue problem has only the trivial solution such that λ, η , and $\gamma_\ell, |\ell| > 1$, all vanish precisely. As a result, we have $y(x, t) = \operatorname{Re}\{u(x) e^{-i\frac{\omega}{m}t}\}$. Moreover, Equation (3.81) and its first derivative in time with $t = 0$ yield the desired time-harmonics solution (3.7).

We observe that for a fixed time interval $(0, T)$ ($m = 1$) in the time integration of (3.5), the number of CG iterations increases linearly to the width L – see Figure 3.16a. On the other hand, in Figure 3.16b for a fixed width $L = 8$ or $L = 16$, the number of CG iterations decreases when the time interval $(0, \ell T)$ grows large. Next, we vary both the computational domain size $\Omega = (0, L) \times (0, 1)$ and time interval $(0, mT)$, where $L = m$ range from 1 to 48. Figure 3.16c monitors the number of CG iterations, which lies between 134 and 254. For instance, in comparison of $L = m = 2$ and $L = m = 48$, the CMCG method requires the number of CG iterations of 206 and 202, respectively, which does not have significant differences.

We conclude that keeping the ratio between the width of the domain L and the number of the time periods m constant, the number of CG iterations may neither increase nor decrease significantly. Since each CG iteration requires the solution of an elliptic problem (3.38a) and the time integration of (3.5) is inherently parallel, it is recommended to enlarge the time interval for a large computational domain when an efficient parallel solver for the time-dependent wave equation is available.

3.9 Numerical experiments II: Helmholtz equations in physically bounded or unbounded domains

Here we present a series of numerical experiments to illustrate the usefulness of controllability methods (CMCG) for the solution of the Helmholtz equation (2.9) in various typical configurations. First, we apply the CMCG method with the original energy functional J_1 from (3.33) to a typical sound-soft scattering problem, as in [25, 26]. To overcome the bottleneck in the time integration of (3.5) due to local mesh refinement, we replace standard time marching by Runge-Kutta based explicit local time-stepping methods [78, 81]. Second, we apply the CMCG method to a sound-hard scattering problem, where we restore uniqueness by imposing the compatibility condition from Section 3.4.2. Next, we also apply the CMCG method to scattering from an inhomogeneous inclusion with a first-order absorbing boundary condition [12] to demonstrate that uniqueness is then inherently guaranteed (see Section 3.4.2). Finally, we show how the CMCG method can also be efficiently applied in a physically bounded domain by using the functional J_3 in (3.43).

3.9.1 Sound-Soft Scattering

Again, we apply the CMCG method to a typical sound-soft scattering problem from the open wedge shown in Figure 3.5 in Section 3.2.4. Hence we consider (2.9) with $\omega = 36\pi, a = c = 1, \Gamma_D$ the boundary of the wedge ($g_D = 0$) and $\Gamma_N = \emptyset$. The

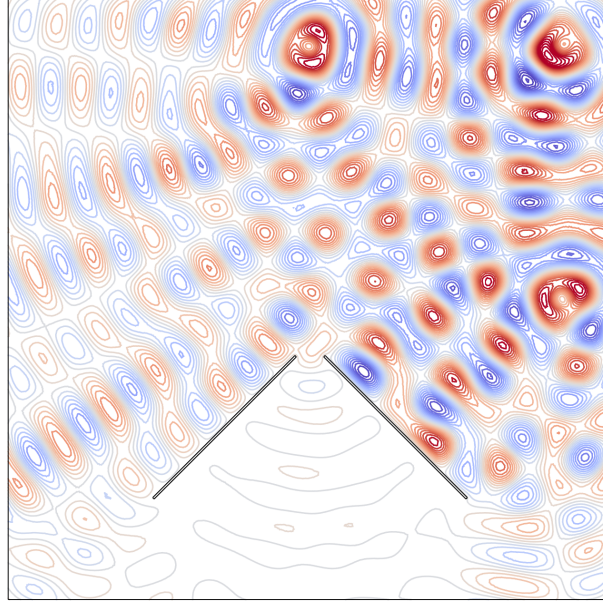


Figure 3.17: Sound-soft scattering: contour lines of the numerical solution with the CMCG method combined with explicit local time-stepping.

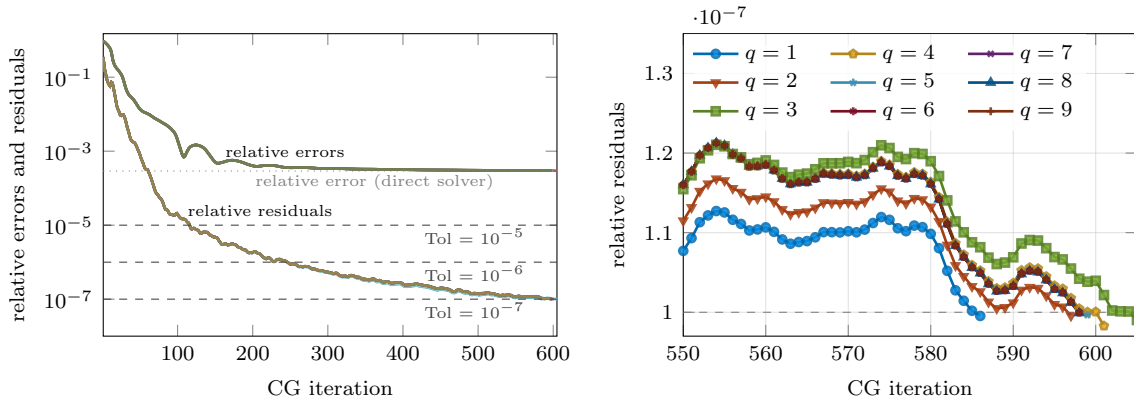


Figure 3.18: Sound-soft scattering: relative residual and relative error at each CG iteration using the CMCG method combined with a local time-stepping method with a time-step $\Delta\tau = \Delta t/q$ with $q = 1, \dots, 9$ (left); zoom on the relative residuals at each CG iteration for varying q (right). In the left frame the curves for different q essentially coincide at this scale.

unbounded exterior is truncated by a square artificial boundary Γ_S , where we impose the absorbing boundary condition (2.9b) with $g_S = 0$. The initial conditions are set to zero, while the incident wave originates from three point sources located at $(0.75, 0.4)$, $(0.5, 0.65)$, and $(0.75, 0.65)$ in the top right corner of Ω .

For spatial discretization, we use \mathcal{P}^3 -FE with mass-lumping and the mesh shown in Figure 3.5 with 79'917 nodes and 12'116 triangles. To overcome the bottleneck from the CFL-restriction on the time step Δt due to local mesh refinement near the wedge, we opt for local time-stepping (LTS) methods based on the classical fourth-order RK method [78]. Hence, we split the mesh into “fine” and “coarse” elements – see Figure 3.5 – and use small local time steps of size $\Delta\tau = \Delta t/q$ but only in the “fine” part. Here, q denotes the refinement ratio between the smallest mesh size h_{\min}^{coarse} in the “coarse” part and h_{\min}^{fine} in the “fine” part, given by (3.15) with $h = h_{\min}$. Figure 3.17 displays the contour lines of the numerical solution of (2.9) obtained by the CMCG Algorithm (Section 3.7) with the functional J_1 from (3.33) and a tolerance $\text{Tol} = 10^{-7}$.

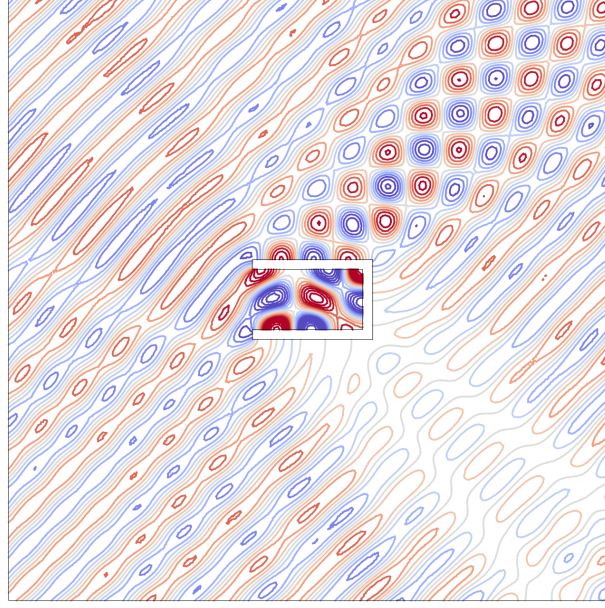


Figure 3.19: Sound-hard scattering: contour lines of the total wave field obtained with the CMCG method and the compatibility condition (3.40)–(3.41).

The numerical solution obtained with the CMCG method differs by only 0.15% from that obtained through direct solution of the Helmholtz equation (2.9). Both solutions differ from a reference solution on a finer mesh by less than 3%.

Now, we combine the CMCG method with local time-stepping using different mesh refinement ratios q . For a fixed mesh, shown in Figure 3.5, we vary the partition into “fine” and “coarse” elements such that $q = 1, \dots, 9$ in (3.15). Regardless of the tolerance $Tol = 10^{-5}, 10^{-6}, 10^{-7}$ in the CMCG Algorithm, we observe in Figure 3.18 that the relative residuals and the numbers of iterations remain identical independently of q . In summary, the CMCG method yields a comparable accuracy to the direct solution of the Helmholtz equation while the convergence of the CMCG method remains unaffected by the local time-stepping strategy.

3.9.2 Sound-hard scattering

Here we apply the CMCG method to the typical situation of an incident plane wave scattered from a sound-hard cavity, as shown in Figure 3.19. Hence we consider (2.9) with $\Gamma_D = \emptyset$, Γ_N the boundary of the obstacle ($g_N = 0$) and Γ_S the exterior square artificial boundary. Since $\Gamma_D = \emptyset$, the original cost functional J_1 in (3.33) does not have a unique minimizer. To remove the spurious constant shift and thus obtain the correct (unique) solution, shown in Figure 3.19, we post-process the solution obtained with J_1 by applying the compatibility condition (3.40)–(3.41) derived in Section 3.4.2 (i). Here, the angle of the incident wave (3.74) is $\phi = 315^\circ$ with $a = c = 1$ and $\omega = 8\pi$ whereas the mesh in $\Omega = (0, 3) \times (0, 3)$ consists of 8'146 \mathcal{P}^3 -FE with 53'456 nodes. For the sake of comparison, we compute a reference solution by solving the Helmholtz equation directly on a finer mesh. The CMCG method yields a relative error of 0.27% with a tolerance $Tol = 10^{-6}$, comparable to that in the direct solution of (2.9) on the same mesh with a relative error of 0.29%.

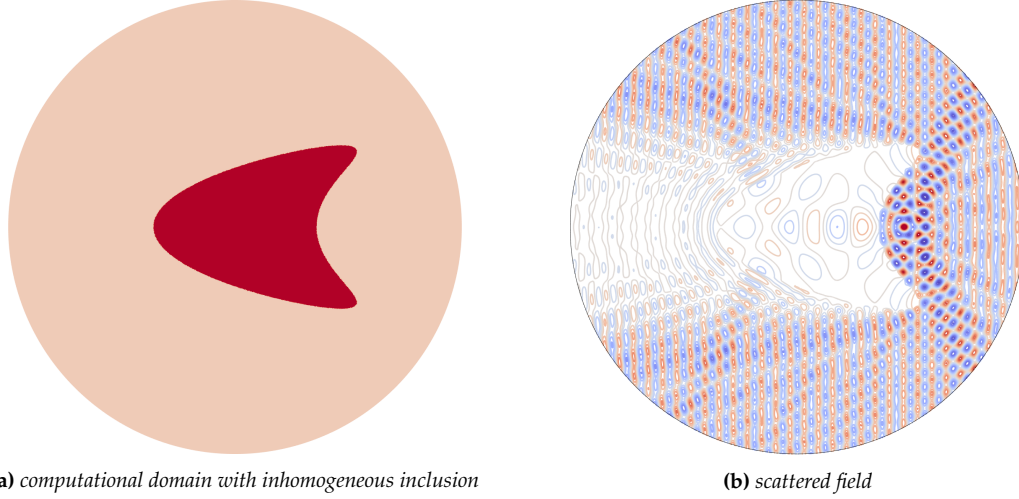


Figure 3.20: Scattering from a penetrable inhomogeneous inclusion: the squared wave speed c^2 is constant inside and outside of the kite-shaped inclusion (a); contour lines of the numerical solution obtained with the CMCG method using the BGT-1 absorbing condition (3.50) and the functional J_1 (b).

3.9.3 Scattering from Inhomogeneous Inclusion

We consider scattering from a penetrable inhomogeneous inclusion with a circular artificial boundary Γ_S ; hence, we consider (2.9) with $\Gamma_D = \Gamma_N = \emptyset$. Instead of applying the compatibility condition (3.39), we enforce uniqueness by replacing the (homogeneous) Sommerfeld-like condition (2.9b) and (3.5b) on Γ_S by the more accurate first-order (homogeneous) Bayliss-Gunzburger-Turkel (BGT-1) absorbing condition (3.50a) and (3.50b), respectively – see Section 3.4.2 (iii).

Again, we use \mathcal{P}^3 -FEM with 667'225 nodes and 102'504 elements and apply the RK4 method for the time integration of (3.5). The frequency $\omega = 32\pi$, $a \equiv 1$, and $c^2(x) = 1 + 8 \cdot H(x)$, where $H(x)$ is the indicator function of the kite-shaped inclusion – see Figure 3.20a. In Figure 3.20b, we display the scattered field for an incident plane wave (3.74) with $\phi = 180^\circ$ obtained by the CMCG method using the original functional J_1 from (3.33). Both the numerical solution obtained with the CMCG method and that obtained by solving (2.9) directly yield a 2.1% relative error with respect to a reference solution computed on a finer mesh. Thus, the BGT-1 condition, like any other absorbing boundary condition with a positive (or negative) definite zeroth-order term, permits the use of the CMCG approach with the original cost functional J_1 even when $\Gamma_D = \Gamma_N = \emptyset$.

3.9.4 Closed wave guide

Finally, we consider the Helmholtz equation (2.9) in a physically bounded domain $\Omega = (-5, 1) \times (-0.5, 0.5)$ without any impedance boundary condition, $\Gamma_S = \emptyset$. At the right vertical entry boundary $x = 1$, the wave field satisfies the Dirichlet boundary condition (2.9d) with $g \equiv 1$; elsewhere, it satisfies a homogeneous Neumann boundary condition. Here, we have $\omega = 25$, $c = 1$, and $a(x) = 1 + H(x)$, where H is the indicator function of the ellipse-shaped inclusion shown in Figure 3.21.

Again, we use \mathcal{P}^3 -FEM with 30'518 nodes and 4'622 triangles for the spatial discretization and the RK4 method for the time integration of (3.5). Here, the solution, shown in Figure 3.22, obtained using the CMCG method and the cost functional J_3 in (3.43) with a relative error of 5.7% is slightly more accurate than that obtained by



Figure 3.21: *physically bounded domain with an inhomogeneous medium*

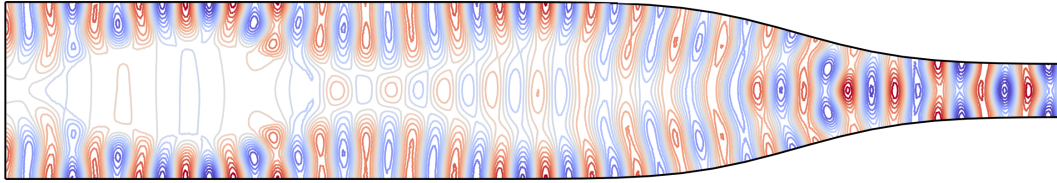


Figure 3.22: *contour lines of the numerical solution*

Figure 3.23: *Closed waveguide: the inhomogeneity a is constant inside and outside of the ellipse-shaped inclusion (top); contour lines of the numerical solution of (2.9) using the CMCG method with the functional J_3 defined by (3.43) with $t_1 = \frac{T}{4}$, $t_2 = \frac{T}{2}$, and $t_3 = T$ (bottom).*

the direct solution of the linear system with a relative error of 6.3%, probably due to the high condition number of the discrete Helmholtz problem.

3.10 Numerical experiments III: HPC-CMCG method

The CMCG method leads to inherently parallel non-intrusive algorithms, as long as an efficient parallel solver for the time-dependent wave equation is available. Here we demonstrate that the CMCG approach, which requires the solution of the elliptic problem (3.38a) at each CG iteration, nonetheless achieves strong scalability on a massively parallel architecture.

The CMCG Algorithm from Section 3.7 is implemented within FreeFem++ [88], an open source finite element software written in C++. FreeFem++ defines a high-level Domain Specific Language (DSL) and natively supports distributed parallelism with Message Passing Interface (MPI). The parallel implementation of the CMCG method relies on the spatial decomposition of the computational domain Ω into multiple subdomains, each assigned to a single computing core. Local finite element spaces are then defined on the local meshes of the subdomains, effectively distributing the global set of degrees of freedom across the available cores.

The bulk of the computational work for solving the forward and backward wave equations in Steps 3 and 8 of the CMCG Algorithm simply consists in performing a sparse matrix-vector product at each time step, which is easily parallelized in this domain decomposition framework: it amounts to performing local matrix-vector products in parallel on the local set of degrees of freedom corresponding to each subdomain, followed by local exchange of shared values between neighboring subdomains.

While the explicit time integration of the wave equation, e.g. RK4 and LF method, is trivially parallelized thanks to mass-lumping in Section 3.2.3, achieving good parallel scalability for the elliptic problem in Steps 4 and 8 of the CMCG Algorithm is more difficult. Here we use domain decomposition (DD) methods [47], which are well-known to produce robust and scalable parallel preconditioners for the iterative solution of large scale partial differential equations. We use the parallel DD

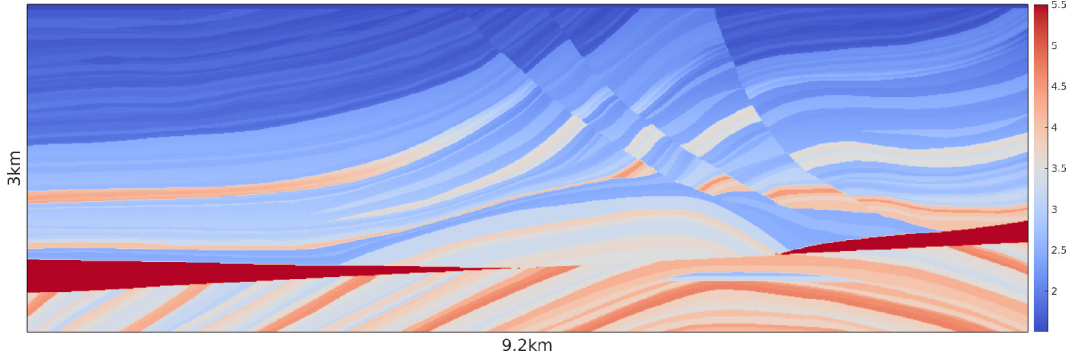


Figure 3.24: *Marmousi model*: propagation velocity $1.5 \leq c(x) \leq 5.5$ [km/s]

| Frequency ν [Hz] | Wave number $k = \omega/c = 2\pi\nu/c$ | #Unknowns $ndof$ | #Nodes 24 cores per node |
|-------------------------|---|---------------------|-----------------------------|
| 10 | 11 – 42 | 1'658'443 | 1–8 |
| 20 | 22 – 84 | 6'628'881 | 1–16 |
| 40 | 45 – 168 | 26'505'761 | 8–64 |
| 60 | 68 – 252 | 59'630'641 | 16–128 |
| 80 | 91 – 336 | 106'003'521 | 16–128 |
| 160 | 182 – 671 | 423'975'041 | 64–256 |
| 250 | 285 – 1048 | 1'035'241'009 | 128–512 |

Table 3.2: *2D-Marmousi model*: \mathcal{P}^2 -FE with 15 points per wave length and $n_T = T/\Delta t = 390$ time steps.

library HPDDM [89], which implements efficiently various Schwarz and substructuring methods in C++11 with MPI and Open Multi-Processing (OpenMP) for parallelism and is interfaced with FreeFem++.

The elliptic problem (3.38a) in the CMCG algorithm is solved by HPDDM using a two-level overlapping Schwarz DD preconditioner, where the coarse space is built using Generalized Eigenproblems in the Overlap (GenEO) [46]. The GenEO approach has proved effective in producing highly scalable preconditioners for solving various elliptic problems [21, 46].

All computations were performed on the supercomputer OCCIGEN at CINES, France¹, with 50'544 (*Intel XEON Haswell*) cores.

3.10.1 2D Marmousi model

First, we consider the two-dimensional geophysical Marmousi model [23] of (2.9) in $\Omega = (0, 9.2) \times (0, 3)$ with $a \equiv 1$, $k(x) = \omega/c(x)$, $\omega = 2\pi\nu$, the wave frequency ν , and

$$f(x) = \exp(-2000 |x - x^*|^2), \quad x^* = (6, -3/16).$$

The velocity profile $c(x)$ is shown in Figure 3.24. We apply a homogeneous Dirichlet condition at the top ($\Gamma_D = \{x_2 = 0\}$) and absorbing boundary conditions on the lateral and lower boundaries ($\Gamma_S = \partial\Omega \setminus \Gamma_D$). For the spatial discretization, we use a \mathcal{P}^2 -FE method with (order preserving) mass-lumping and at least 15 points per wave length λ . We apply the LF method for the time integration of (3.5). To speed-up the convergence of the CMCG method, we use an initial run-up from Section 3.6 with the smooth transient kernel θ_{tr} in (3.72), so that the wave travels at least once across

¹<https://www.cines.fr/calcul/materiels/occigen/>

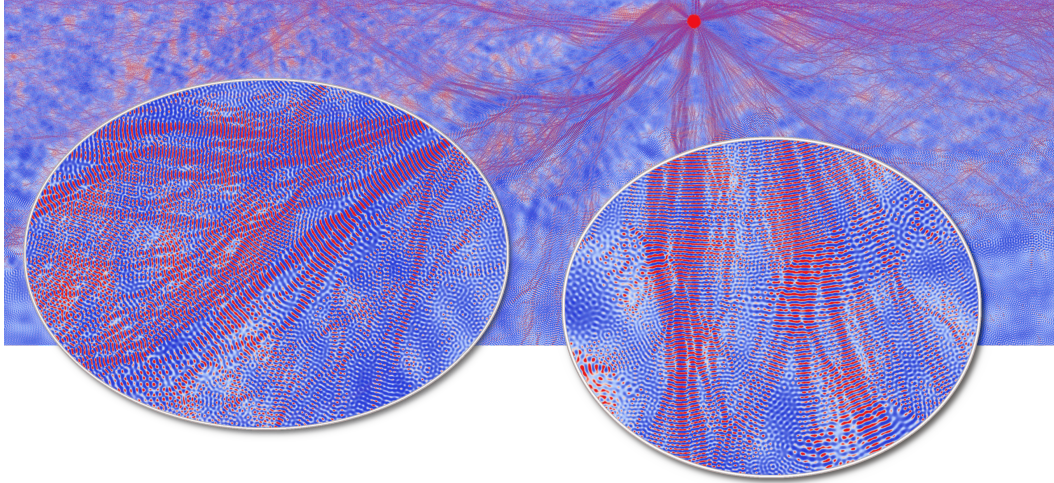


Figure 3.25: *Marmousi model*: real part of the scattered field with $\omega = 2\pi\nu$, $\nu = 250$ Hz

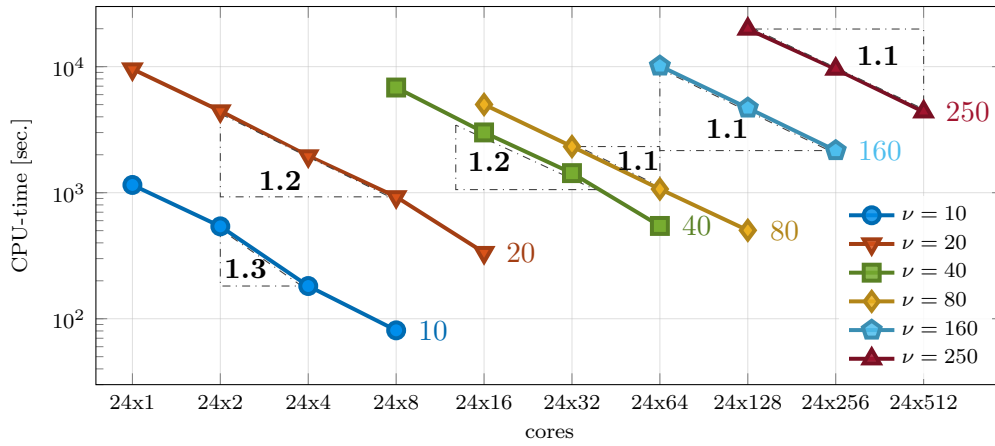


Figure 3.26: *Marmousi model*: total CPU-time in seconds for varying number of cores. For each frequency ν , the FE-discretization and problem size remain fixed.

the entire computational domain during the time interval $[0, t_{tr}]$; hence, we set

$$m = \left\lceil \frac{\sqrt{9.2^2 + 3^2}}{T_{c_{\min}}} \right\rceil, \quad t_{tr} = mT, \quad T = (2\pi)/\omega.$$

For any particular frequency ν , we apply the CMCG method for fixed parameters and FE-mesh while increasing the number of (CPU) cores. Figure 3.25 displays the real part of the wave field with $\nu = 250$ Hz. In Figure 3.26, we observe linear speed up (strong scaling) at every frequency with increasing number of cores. In fact, the speed up is even slightly better than linear due to cache effects, but also because the cost of the direct solver used on each subdomain decreases superlinearly with the decreasing size of subdomains as the number of cores increases. As the frequency ν increases, both the period $T = 1/\nu$ and the time-step Δt decrease, so that the number of time steps per CG iteration remains constant. Since the number of CG iterations does not grow here with increasing ν , the bulk of the computational work in the CMCG Algorithm in fact shifts to the run-up phase. For $\nu = 10$ Hz, for instance, the CMCG Algorithm stops after 273 CG iterations, while 74% of the total computational time is spent in the time integration of (3.5), 16% in the elliptic solver (DDM) and 10% in the initial run-up. In contrast, for $\nu = 250$ Hz, the CMCG Algorithm already stops

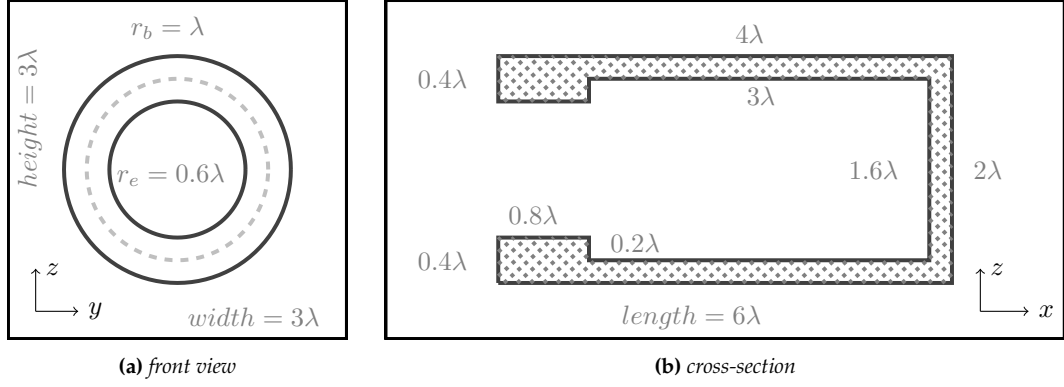


Figure 3.27: 3D-cavity: (a) front view of the opening with inner and outer radius; (b) longitudinal cross-section.

| Frequency $\nu = 2\pi\omega$ | #Unknowns $ndof$ | #Tetrahedra | CG iterations | #Nodes 24 cores per node |
|---------------------------------|---------------------|-------------|---------------|-----------------------------|
| 2 | $8.17 \cdot 10^5$ | 5'051'049 | 239 | 1–8 |
| 3 | $5.22 \cdot 10^6$ | 31'190'000 | 440 | 2–32 |
| 4 | $1.9 \cdot 10^7$ | 114'391'112 | 607 | 32–96 |
| 6 | $1.18 \cdot 10^8$ | 703'590'464 | 578 | 64–128 |

Table 3.3: 3D-cavity resonator: \mathcal{P}^1 -FE; as η increases, the ratio $hk^{3/2}$ remains constant to avoid pollution errors [7]

after 5 CG iterations, while 99% of the total computational time is spent in the initial run-up and 1% in the CG iteration. By modifying the run-up time t_{tr} , one could arbitrarily shift the relative computational cost between run-up and CG iterations and thus further optimize for a minimal total execution time.

3.10.2 3D cavity

Here we solve a sound-soft scattering problem (2.9) with $a \equiv c \equiv 1$, $k = \omega = 2\pi\nu$, $f \equiv g_D \equiv g_N \equiv 0$, and

$$g_S = -(\partial_n - ik)u^{in}, \quad u^{in}(\mathbf{x}) = \exp(ik\mathbf{x}^\top \mathbf{d}), \quad \mathbf{d} = (1/2, 0, \sqrt{3}/2)^\top,$$

in $\Omega = (0, 6\lambda) \times (0, 3\lambda) \times (0, 3\lambda)$, $\lambda = 1$, from a cavity – see Figure 3.27. We impose a homogeneous Dirichlet boundary condition on the obstacle and a Sommerfeld-like absorbing boundary condition on the exterior boundary. We discretize (3.5) with \mathcal{P}^1 -FE in space and the second-order LF method for time integration. To control the pollution error and obtain an accurate numerical solution, we set $hk^{3/2} \sim \text{const}$, as we increase the frequency ν . Figure 3.29 shows the total wave field of (2.9) with $\nu = 6$ inside the cavity. For fixed parameters and mesh size, we now solve (2.9) at frequency $\nu = 2, 3, 4, 6$ with the CMCG method using an increasing number of cores – see Table 3.3. Again, we observe in Figure 3.28 (better than) linear (strong) scaling with increasing number of cores.

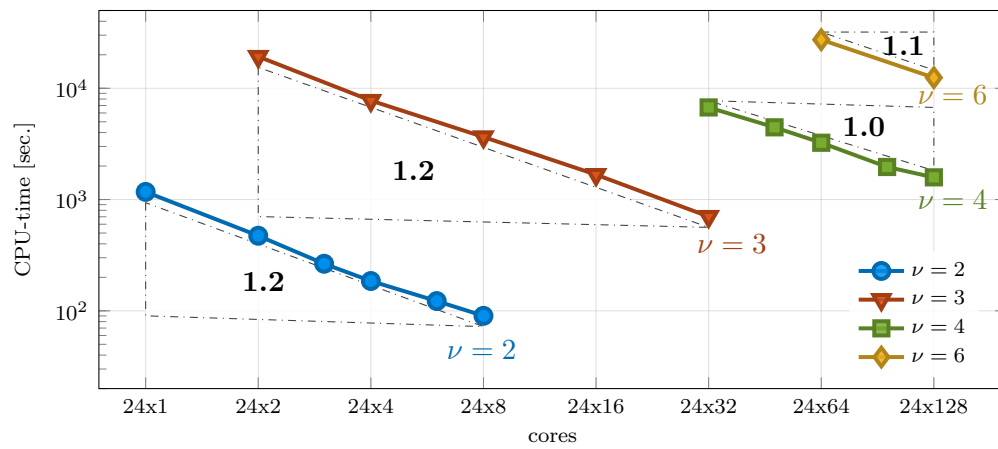


Figure 3.28: 3D-sound-soft scattering problem: total computational time in second by solving (2.9) with the CMCG method in parallel architecture by varying the number of cores for a fixed configuration – see Table 3.3

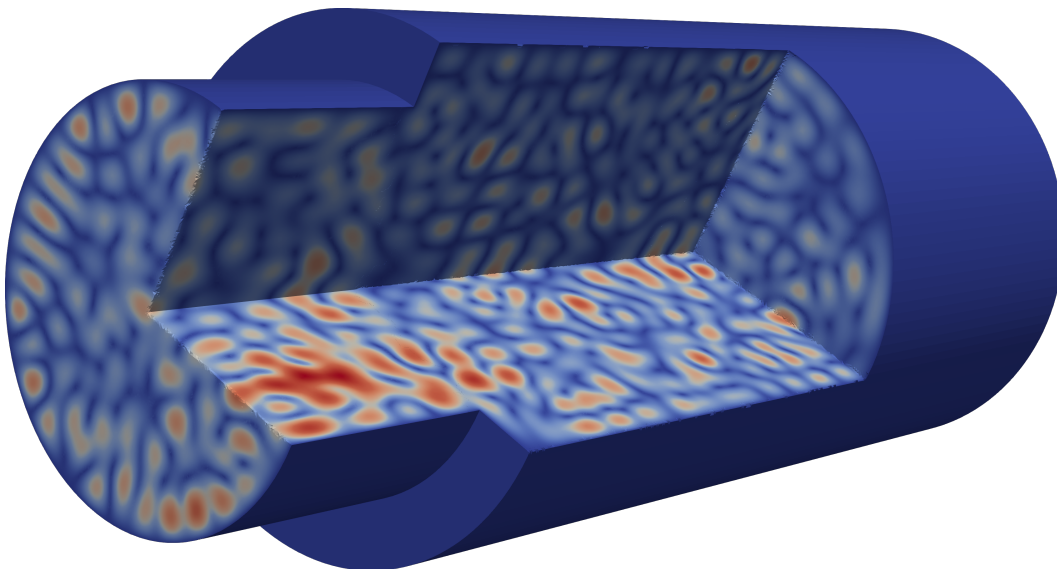


Figure 3.29: 3D-cavity: total wave field (2.9) with $\nu = 6$ obtained with the CMCG method

CHAPTER 4

More applications with the CMCG approach

In this chapter we extend the controllability method (CM), combined with the conjugate gradient (CG) method, to further applications.

In Section 4.1, we first show how to use the CMCG method to compute in one shot the superposition of several time-harmonic (single frequency) wave fields. The filtering procedure from Section 3.5 permits to extract the time-harmonic solution at the desired frequency. Next, in Section 4.2, we formulate the CMCG method for unbounded scattering problems in a computational domain surrounded by a perfectly matched layer (PML). This formulation completely replaces the absorbing boundary condition on Γ_S by the PML approach [18, 82, 81, 8]. Finally, in Section 4.3, we write the CMCG Algorithm from Section 3.7 for the first-order formulation [70, 68, 80]. Since the derivative of the functional J_1 in (3.33), which is needed during the CG update, here operates only in $(L^2(\Omega))^d \times L^2(\Omega)$, which is self-dual. This implies that the Riesz representative at every CG iteration, obtained by solving the strongly elliptic problem (3.38a), is no longer needed. As a consequence, the CMCG becomes inherently parallel.

4.1 Multiple time-harmonic sources

4.1.1 Controllability method for superposition of time-harmonic waves

Let $\omega > 0$ and $0 < \omega_1 < \dots < \omega_m < +\infty$ be frequencies with

$$\mathcal{J} = \left\{ \frac{\omega_j}{\omega} \mid j = 1, \dots, m \right\} \subset \mathbb{N}. \quad (4.1)$$

Then, we seek the solutions $u^{(j)}$ of the (2.9) with $\omega = \omega_j$, $f = f^{(j)}$, $g_D = g_D^{(j)}$, $g_N = g_N^{(j)}$ and $g_S = g_S^{(j)}$, $j = 1, \dots, m$.

Now, instead of solving the Helmholtz equations directly in the frequency domain, we again reformulate (2.9) in the time domain. Hence,

$$y(x, t) = \sum_{j=1}^m \operatorname{Re}\{u^{(j)}(x) e^{-i\omega_j t}\}, \quad (4.2)$$

satisfies the (real-valued) time-dependent wave equation (3.5) with the initial values

$$y_0 = \sum_{j=1}^m \operatorname{Re}\{u^{(j)}(x)\}, \quad y_1 = \omega \sum_{j=1}^m \operatorname{Im}\{u^{(j)}(x)\}, \quad (4.3)$$

and the source terms

$$\begin{aligned}\tilde{f}(x, t) &= \sum_{j=1}^m \operatorname{Re}\{f^{(j)}(x) e^{-i\omega_j t}\}, & \tilde{g}_S(x, t) &= \sum_{j=1}^m \operatorname{Re}\{g_S^{(j)}(x) e^{-i\omega_j t}\}, \\ \tilde{g}_N(x, t) &= \sum_{j=1}^m \operatorname{Re}\{g_N^{(j)}(x) e^{-i\omega_j t}\}, & \tilde{g}_D(x, t) &= \sum_{j=1}^m \operatorname{Re}\{g_D^{(j)}(x) e^{-i\omega_j t}\}.\end{aligned}\quad (4.4)$$

By Lemma 2, the wave equation (3.5), with the source terms given by (4.4) and the initial values $y_0 \in H^1(\Omega)$ and $y_1 \in L^2(\Omega)$ given by (4.3), is well-posed and has a unique solution $y \in C^0(0, T; H^1(\Omega)) \cap C^1(0, T; L^2(\Omega))$. Moreover, for sound-soft scattering problems, the time-periodic solution y of (3.5) is also unique and satisfies (4.2). For general Helmholtz equation, however, a time-periodic solution is not unique, which is formulated in the following theorem.

Theorem 11. *Let $y \in C^0([0, T]; H^1(\Omega)) \cap C^1([0, T]; L^2(\Omega))$ be a (real-valued) solution of the wave equation (3.5) with initial conditions $(y_0, y_1) \in H^1(\Omega) \times L^2(\Omega)$. If ∇y and y_t are time periodic with period $T = 2\pi/\omega$, then y admits the Fourier series expansion*

$$(y(\cdot, t), \varphi) = \sum_{\ell=1}^m (\operatorname{Re}\{u^{(\ell)}(x) e^{-i\omega_\ell t}\}, \varphi) + (\lambda + \eta t, \varphi) + \sum_{|\ell| \notin J} (\gamma_\ell, \varphi) e^{i\omega_\ell t} \quad (4.5)$$

for any $\varphi \in H_D^1$ defined in (2.13), where the constants $\eta, \lambda \in \mathbb{R}$, and the complex-valued functions $\gamma_\ell \in H^1$, $|\ell| \geq 1$, solve (2.9) with the frequency ω_ℓ and source terms

$$\tilde{f}^{(\ell)} = \int_0^T \tilde{f}(x, t) e^{-i\omega_\ell t} dt, \quad \tilde{g}_S^{(\ell)} = \int_0^T \tilde{g}_S(x, t) e^{-i\omega_\ell t} dt, \quad (4.6)$$

$$\tilde{g}_N^{(\ell)} = \int_0^T \tilde{g}_N(x, t) e^{-i\omega_\ell t} dt, \quad \tilde{g}_D^{(\ell)} = \int_0^T \tilde{g}_D(x, t) e^{-i\omega_\ell t} dt, \quad (4.7)$$

instead of ω , f , g_S , g_N , and g_D , respectively. Here we recall the notation f_0^T given by (3.18).

If $\mathcal{H}^{d-1}(\Gamma_S) > 0$ then $\eta = 0$, and if $\mathcal{H}^{d-1}(\Gamma_D) > 0$ then $\lambda = \eta = 0$.

Remark 9. (i) For $\omega_1 = \omega$ and $m = 1$, Theorem 11 coincides with Theorem 6 from Section 3.3.1 with $\omega = \omega_1$, $f = f^{(1)}$, $g_S = g_S^{(1)}$, $g_N = g_N^{(1)}$, and $g_D = g_D^{(1)}$.

(ii) The mutual orthogonality of different time harmonics $e^{-i\omega_\ell t}$ in $L^2(0, T)$ yields

$$\tilde{f}^{(\ell)} = \begin{cases} \int_0^T \operatorname{Re}\{f^{(j)} e^{-i\omega_\ell t}\} e^{-i\omega_\ell t} dt = \frac{\overline{f^{(j)}}}{2}, & \ell = \frac{\omega_j}{\omega} \in \mathcal{J}, \\ \int_0^T \operatorname{Re}\{f^{(j)} e^{i\omega_\ell t}\} e^{-i\omega_\ell t} dt = \frac{f^{(j)}}{2}, & -\ell = \frac{\omega_j}{\omega} \in \mathcal{J}, \\ 0, & |\ell| \notin \mathcal{J}. \end{cases} \quad (4.8)$$

Proof of Theorem 11. It can be proved in an analogous way as Theorem 6 by replacing $\operatorname{Re}\{f(x) e^{-i\omega t}\}$, $\operatorname{Re}\{g_S(x) e^{-i\omega t}\}$, $\operatorname{Re}\{g_N(x) e^{-i\omega t}\}$, and $\operatorname{Re}\{g_D(x) e^{-i\omega t}\}$ by $\tilde{f}^{(\ell)}(x, t)$, $\tilde{g}_S^{(\ell)}(x, t)$, $\tilde{g}_N^{(\ell)}(x, t)$, and $\tilde{g}_D^{(\ell)}(x, t)$ in (4.4), respectively. \square

For scattering problems ($\mathcal{H}^{d-1}(\Gamma_S) > 0$), the homogeneous Helmholtz equation (2.9) with $f \equiv g_S \equiv g_N \equiv g_D \equiv 0$ has only the trivial solution, which yields

$$y = \lambda + \sum_{\ell \in \mathcal{J}} (\gamma_\ell e^{i\omega_\ell t} + \gamma_{-\ell} e^{-i\omega_\ell t}), \quad 0 \leq t \leq T. \quad (4.9)$$

In particular, for sound-soft scattering ($\mathcal{H}^{d-1}(\Gamma_D), \mathcal{H}^{d-1}(\Gamma_S) > 0$),

$$y = \sum_{\ell \in \mathcal{J}} (\gamma_\ell e^{i\omega_\ell t} + \gamma_{-\ell} e^{-i\omega_\ell t}), \quad 0 \leq t \leq T. \quad (4.10)$$

Similar to the frequency filtering procedure for a single time-harmonic wave in Section 3.5, we propose to use the filtering procedure to extract all desired time-harmonic waves.

4.1.2 Frequency filtering procedure

We recall the frequency filtering procedure in Section 3.5. Let y be a T -time periodic solution of the wave equation (3.5) with the initial value (y_0, y_1) , where $T = (2\pi)/\omega$. By Theorem 11, y admits the Fourier series expansion

$$y(\cdot, t) = \sum_{\ell=1}^m \operatorname{Re}\{u^{(\ell)}(x) e^{-i\omega_\ell t}\} + \lambda + \eta t + \sum_{|\ell| \notin \mathcal{J}} \gamma_\ell e^{i\omega_\ell t}.$$

Now, we define

$$\widehat{y}_j(x) = \int_0^T \left(y(x, t) + \frac{i}{\omega_j} y_t(x, t) \right) e^{i\omega_j t} dt, \quad j = 1, \dots, m, \quad (4.11)$$

in an analogous way with that in Section 3.5. Then, we obtain

$$\widehat{y}_j = \int_0^T \left[\sum_{\ell=1}^m \left(\operatorname{Re}\{u^{(\ell)} e^{-i\omega_\ell t}\} + \frac{\omega_\ell}{\omega_j} \operatorname{Im}\{u^{(\ell)} e^{-i\omega_\ell t}\} \right) + \eta t + \sum_{|\ell| \notin \mathcal{J}} \gamma_\ell(x) e^{i\omega_\ell t} \right] e^{-i\omega_j t} dt.$$

The mutual orthogonality of different time harmonics $e^{-i\omega_\ell t}$ in $L^2([0, T])$ yields

$$\widehat{y}_j = u^{(j)} - \frac{i\eta}{\omega_j} \quad (4.12)$$

where $u^{(j)}$ solves the (desired) Helmholtz equation

$$-\nabla \cdot (a(x) \nabla u^{(j)}(x)) - \frac{\omega_j^2}{c^2(x)} u^{(j)}(x) = f^{(j)}(x), \quad x \in \Omega, \quad (4.13a)$$

$$a(x) \frac{\partial u^{(j)}(x)}{\partial n} - i\omega_j \frac{\sqrt{a(x)}}{c(x)} u^{(j)}(x) = g_S^{(j)}(x), \quad x \in \Gamma_S, \quad (4.13b)$$

$$a(x) \frac{\partial u^{(j)}(x)}{\partial n} = g_N^{(j)}(x), \quad x \in \Gamma_N, \quad (4.13c)$$

$$u^{(j)}(x) = g_D^{(j)}(x), \quad x \in \Gamma_D. \quad (4.13d)$$

The constant η immediately vanishes when $\mathcal{H}^{d-1}(\Gamma_D \cup \Gamma_S) > 0$. Otherwise, for $\Gamma_D = \Gamma_S = \emptyset$, and $\Gamma_N = \partial\Omega$, (2.9) corresponds to a pure Neumann problem. We determine η using the compatibility condition directly derived from the Helmholtz equation. By integration of (4.13a) and using Green's formula, together with (4.12) and the boundary condition (4.13c), we obtain

$$-\frac{i\eta}{\omega_j} = \frac{1}{\|\frac{\omega_j}{c}\|^2} \left(\int_\Omega f^{(j)}(x) dx + \int_{\Gamma_N} g_N^{(j)} ds + \int_\Omega \left(\frac{\omega_j}{c(x)} \right)^2 \widehat{y}^{(j)}(x) dx \right). \quad (4.14)$$

4.1.3 CMCG algorithm and computational cost

To determine the unknown initial values y_0, y_1 in (4.3), we minimize the functional J_1 in (3.33). Here $y(x, t)$ is the solution of the wave equation (3.5) with instead the source terms given by (4.4).

Since the formulation of the CMCG method in CMCG Algorithm 2 only requires time-periodic sources, we can immediately apply the algorithm by replacing the source terms in (3.5) with those given by (4.4) to find the minimizer (y_0, y_1) . To recover the desired time-harmonics, we apply the filtering procedure in (4.12). In the absence of Γ_S and Γ_D , we impose the compatibility condition in (4.14) to restore the uniqueness.

Next, the computational cost of the CMCG Algorithm for computing the superposition of m time-harmonics is the same as the computational cost for solving single frequency (2.9). Second, the filtering procedure only requires the integration of the time-dependent solution $y(x, t)$ in time, which is computed on the fly using quadrature formula (e.g. Simpson's rule), and hence it does not need to store the entire history of $y(x, t)$. Since the solution $y(x, t)$ is already computed in the CMCG algorithm, the filtering procedure has almost no additional computational cost. Finally, the compatibility condition only requires simple matrix-vector multiplications, which can be computed easily.

As a consequence, the CMCG method solves m Helmholtz equations in one shot with the same computational cost as the cost for one Helmholtz equation. However, here we require that the final time $T = \frac{2\pi}{\omega_1}$. This implies that we require a larger number of time steps $n_T = n_{T_1} := T/\Delta t$ than the number of time steps n_{T_j} ,

$$n_{T_j} = \frac{2\pi}{\omega_j} < \frac{2\pi}{\omega_1} = n_T, \quad j = 2, \dots, m,$$

for the CMCG method for one frequency ω_j . Moreover, to avoid pollution errors, we require that the mesh size h is sufficiently small with respect to the highest frequency ω_m . Hence, we may need to solve the CMCG method on a finer mesh with a larger number of time steps than the CMCG method for single frequency.

4.1.4 Numerical examples

Here we present series of numerical examples to show the usefulness of the CMCG method and verify the accuracy and correctness of numerical solutions, obtained with the filtering procedure for the desired harmonic wave field.

First, we illustrate the spurious constant shift in the one-dimensional superposition of the several solutions of the Neumann problem (2.9), $\Gamma_N = \partial\Omega$, obtained with the CMCG method. To recover all desired eigenmodes, we apply the filtering procedure. In addition, to restore the uniqueness of the numerical solution, we impose the compatibility condition (4.14). Second, we apply the CMCG method to solve in one shot a two-dimensional scattering problem from inhomogeneous inclusion (4.13) from three plane waves with different frequencies ω_ℓ , $\ell = 1, 2, 3$. Again, we use the filtering procedure to extract the desired eigenmodes. Finally, we solve scattering problems of (2.9) with the source term $f = f^{(\ell)}$ and frequency $\omega = \omega_m$, $m, \ell = 1, 2, 3$, with the CMCG method. The nine wave fields $u_{\ell,m}$ are determined by the filtering procedure.

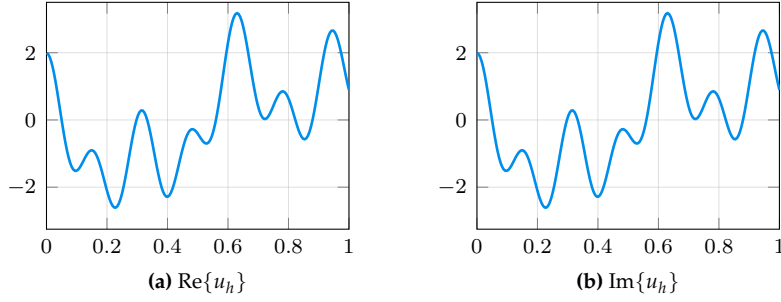


Figure 4.1: One-dimensional superposition of Neumann problems: numerical superposition u_h of Helmholtz equations (4.13), obtained with the CMCG-method.

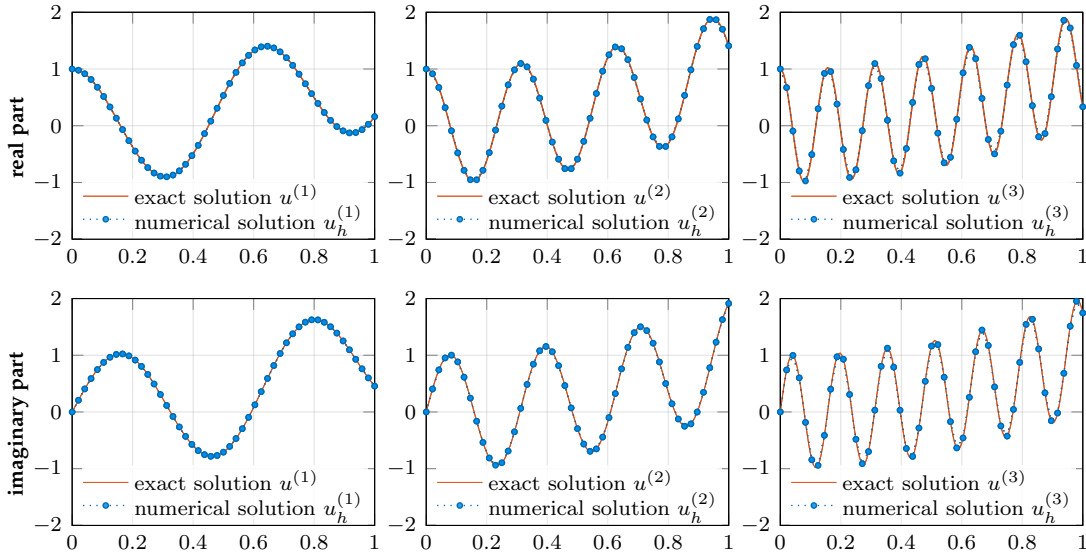


Figure 4.2: One-dimensional superposition of Neumann problems: numerical eigenmodes $u_h^{(\ell)}$, $\ell = 1, \dots, 3$, from the superposition u_h , obtained with the filtering procedure; top row: real part of each eigenmodes; bottom row: imaginary part of each eigenmodes.

Neumann problems

First, we consider the Neumann problem (4.13) in $\Omega = (0, 1)$, $\Gamma = \Gamma_N$, with

$$a = c = 1, \quad \omega = 5, \quad \omega_j = 2^j \omega, \quad j = 1, \dots, m = 3.$$

Let the source terms $f^{(j)}$ and $g_N^{(j)}$ be chosen so that

$$u^{(j)} = e^{i\omega_j x} + (1+i)x^2, \quad j = 1, \dots, m,$$

solves (4.13). Since ω_j is not in the spectrum Σ_L of the linear operator from Section 2.3, (4.13) has a unique solution for $j = 1, \dots, m$. The corresponding time-dependent source terms \tilde{f} and \tilde{g}_N in (3.5) are given by (4.4).

Now, we use the RK4 method for solving the time integration of (3.5) with the order preserving mass-lumping and with the time step $\Delta t = 0.15h$, $h = 2^{-6}$. Then the CG iteration stops when $Tol = 10^{-10}$ in Step 6 in the CMCG Algorithm is reached. We apply the filtering procedure to extract the desired eigenmode $u_h^{(j)}$, $j = 1, 2, 3$, from the superposition u_h shown in Figure 4.1, obtained with the CMCG method for the superposition of time-harmonic waves in Section 4.1.3. To remove the spurious

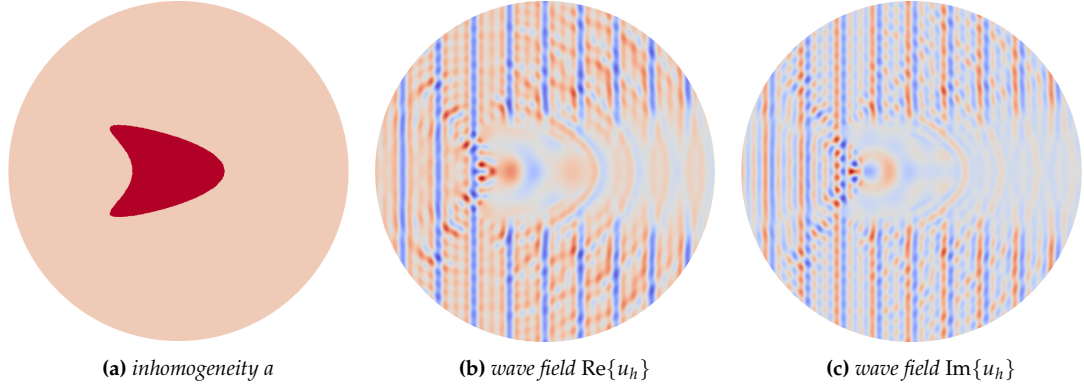


Figure 4.3: Superposition from inhomogeneous inclusion: the inhomogeneous medium a is constant inside and outside of the kite-shaped inclusion (a); the illustration of a minimizer of J_1 (b)–(c), obtained with the CMCG method.

constant shift from the solution $u_h^{(j)}$, we impose the compatibility condition (4.14). Figure 4.2 illustrates both the real and the imaginary part of each eigenmode $u_h^{(j)}$, $j = 1, 2, 3$. The CMCG method, combined with the filtering procedure and the compatibility condition, yields the relative L^2 -errors $\|u_h^{(\ell)} - u^{(\ell)}\|$ of $8.1 \cdot 10^{-6}$, $6.6 \cdot 10^{-5}$, and $5.6 \cdot 10^{-4}$, $\ell = 1, 2, 3$. Here the L^2 -error increases due to the pollution error.

Scattering from inhomogeneous inclusion: plane waves

Here we consider the scattering from a penetrable inhomogeneous inclusion (2.9) in the unit disk $\Omega = \{x \in \mathbb{R}^2 \mid \|x\| \leq 1\}$ with $\Gamma_S = \partial\Omega$, series of frequencies $\omega_j = j\omega$, $\omega = 15$, $j = 1, \dots, m = 3$, $c = 1$, and $a(x) = 1 + 5 \cdot H(x) \in [1, 6]$, where $H(x)$ is the indicator function of the kite-shaped inclusion – see Figure 4.3a. Furthermore,

$$f^{(j)}(x) \equiv 0, \quad g_S^{(j)}(x) = -\frac{\partial}{\partial n} u_j^{\text{in}}(x) + i\omega_j u_j^{\text{in}}(x),$$

where $u_j^{\text{in}}(x) = e^{i\omega_j x_1}$ is the incident plane wave (3.74) with $\phi = 0^\circ$. We use \mathcal{P}^3 -FE for the spatial discretization with order preserving mass-lumping and the classical fourth-order Runge-Kutta (RK4) method for the time integration of (3.5). In addition, the tolerance in the CMCG algorithm is set to $\text{Tol} = 10^{-8}$.

In Figures 4.3b – 4.3c, we display the superposition u_h of scattering wave fields governed from an incident plane wave (3.74), obtained by the CMCG method. In Figures 4.4a – 4.4f, we monitor both the real and imaginary part of the numerical solutions $u_h^{(j)}$, $j = 1, 2, 3$.

To compare the accuracy of $u_h^{(j)}$, obtained from u_h with the filtering procedure in (4.12), we compute the (reference) FE Galerkin solution $u_h^{*,(j)}$, $j = 1, 2, 3$, corresponding to the direct solution of the linear system. The system,

$$\mathbf{A}_h^{(j)} \mathbf{u}_h^{*,(j)} = \mathbf{F}_h^{(j)}, \quad (4.15)$$

results from the same standard H^1 -conforming \mathcal{P}^3 -FE discretization of the Helmholtz equation (4.13) with $\omega = \omega_j$, $f = f^{(j)}$ and $g_S^{(j)}$. Here we obtain a relative difference

$$\|u_h^{*,(j)} - u_h^{(j)}\|_{L^2(\Omega)}, \quad j = 1, 2, 3,$$

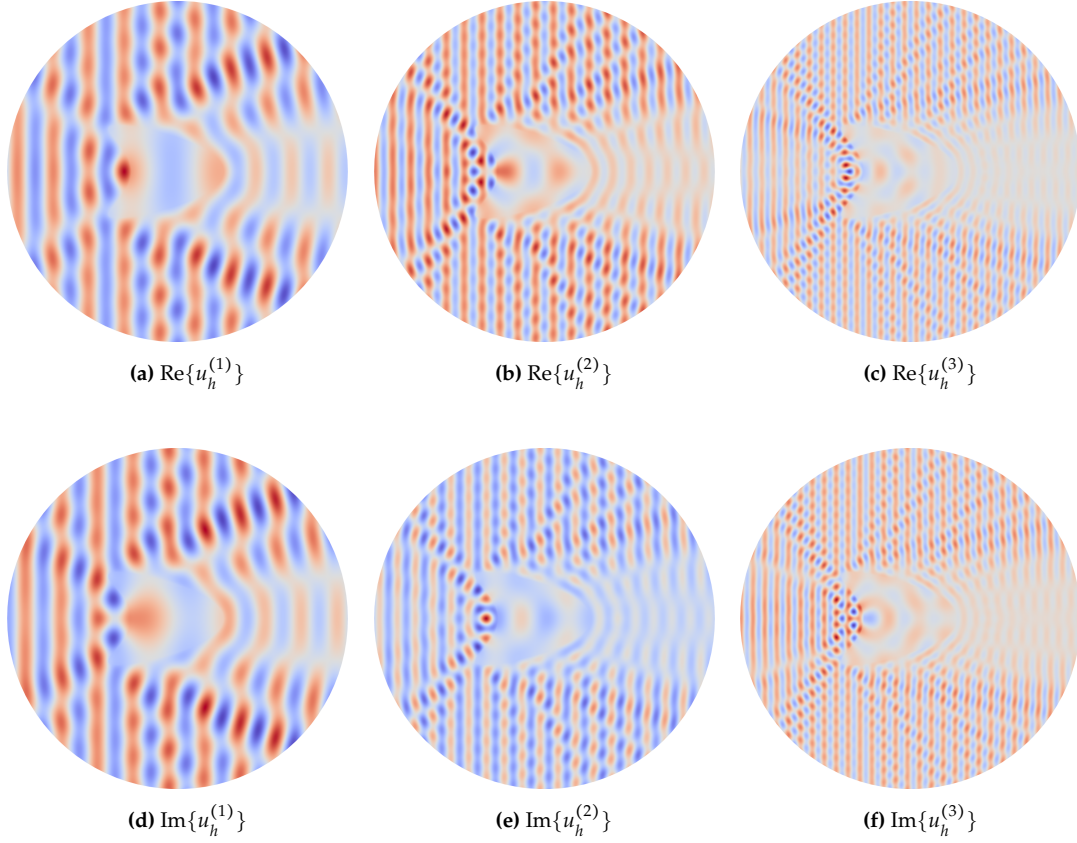


Figure 4.4: Scattering from inhomogeneous inclusion: (a)-(f) numerical solutions $u_h^{(j)}$ of (4.13), obtained from u_h with the filtering procedure $(\phi_j(t) = e^{i\omega_j t})$; top row: $\text{Re}\{u_h^{(j)}\}$, $j = 1, 2, 3$; bottom row: $\text{Im}\{u_h^{(j)}\}$, $j = 1, 2, 3$.

of $3.8 \cdot 10^{-6}$, $5.9 \cdot 10^{-6}$, and $7.3 \cdot 10^{-4}$.

Scattering from inhomogeneous inclusion and point sources

Finally, we consider scattering problems from a kite-shaped inhomogeneous inclusion and from different Gaussian point sources. Let $u^{(j,\ell)}$ denote the solution of (2.9) in $\Omega = (0, 1) \times (0, 1)$ with the frequency

$$\omega_j = j\omega, \quad \omega = 20, \quad j = 1, \dots, m = 3,$$

and the source term

$$f^{(\ell)}(x) = 5000 e^{-5000(x-x_\ell^*)^2}, \quad \ell = 1, \dots, m = 3,$$

with

$$x_1^* = (0.05 \ 0.95)^\top, \quad x_2^* = (0.5 \ 0.95)^\top, \quad x_3^* = (0.95 \ 0.95)^\top.$$

Here we impose the homogeneous absorbing boundary condition (2.9b) with $g_S^{(\ell)} = 0$, $\ell = 1, 2, 3$, on the boundary $\Gamma_S = \partial\Omega$. Furthermore, $c = 1$ and $a(x) = 1 + 4 H(x)$, where $H(x)$ is the indicator function of the star-shaped inclusion.

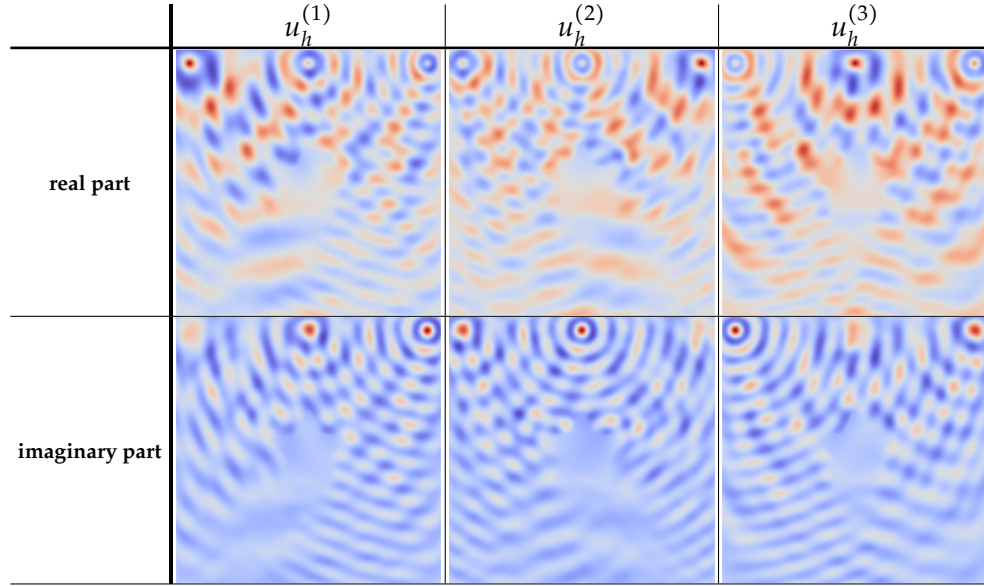


Figure 4.5: Scattering from inhomogeneous inclusion and point sources: minimizer $u_h^{(p)}$ of J_1 , obtained with the CMCG method for superposition with the source term $\tilde{f}^{(p)}$, $p = 1, 2, 3$, in (4.4) given by (4.16).

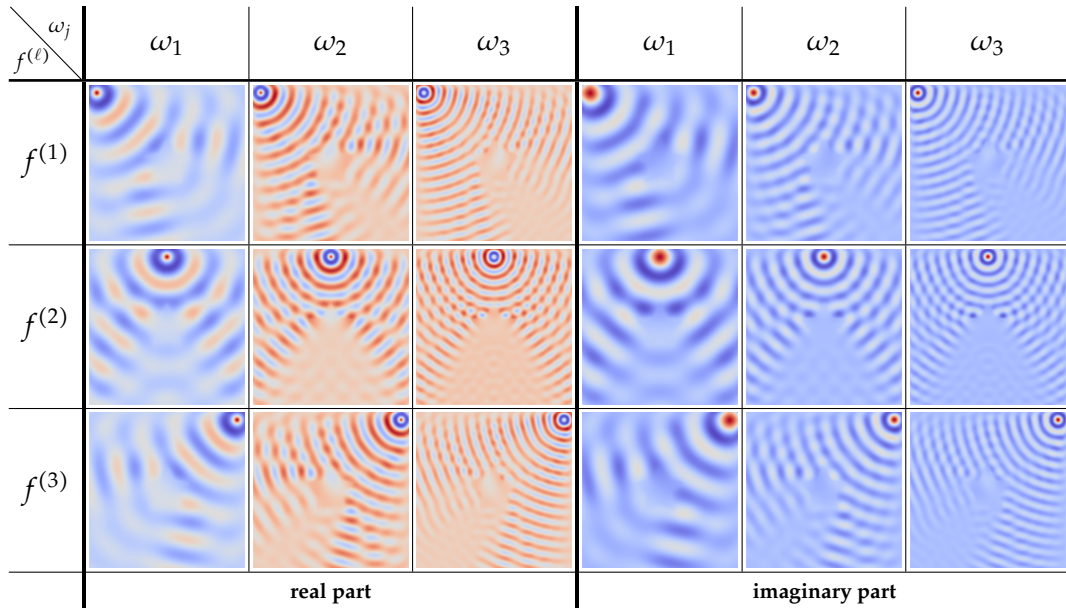


Figure 4.6: Scattering from an inhomogeneous inclusion and point sources: numerical solutions $u_h^{(j, \ell)}$ of (2.9) with $\omega = \omega_j$ and $f = f^{(\ell)}$, $j, \ell = 1, 2, 3$, obtained from $u_h^{(p)}$, $p = 1, 2, 3$, with the filtering procedure.

First, we apply the CMCG method for the superposition with the time-harmonic source term $\tilde{f}^{(p)}(x, t)$, $p = 1, 2, 3$,

$$\tilde{f}^{(1)}(x, t) = \text{Re}\{f^{(1)}(x) e^{-i\omega_1 t} + f^{(2)}(x) e^{-i\omega_2 t} + f^{(3)}(x) e^{-i\omega_3 t}\}, \quad (4.16a)$$

$$\tilde{f}^{(2)}(x, t) = \text{Re}\{f^{(3)}(x) e^{-i\omega_1 t} + f^{(1)}(x) e^{-i\omega_2 t} + f^{(2)}(x) e^{-i\omega_3 t}\}, \quad (4.16b)$$

$$\tilde{f}^{(3)}(x, t) = \text{Re}\{f^{(2)}(x) e^{-i\omega_1 t} + f^{(3)}(x) e^{-i\omega_2 t} + f^{(1)}(x) e^{-i\omega_3 t}\}, \quad (4.16c)$$

to compute the numerical superposition $u_h^{(p)}$ shown in Figure 4.5. Then, we apply the filtering procedure in (4.12) to the solution $u_h^{(p)}$, $p = 1, \dots, 3$, which yields $(u_h^{(1,1)}, u_h^{(2,2)}, u_h^{(3,3)})$ from $u_h^{(1)}$, $(u_h^{(1,3)}, u_h^{(2,1)}, u_h^{(3,2)})$ from $u_h^{(2)}$, and $(u_h^{(1,2)}, u_h^{(2,3)}, u_h^{(3,1)})$ from $u_h^{(3)}$ – see Figure 4.6.

Again, we use \mathcal{P}^3 -FE discretization with order preserving mass-lumping (105'841 unknowns) and the classical fourth-order Runge-Kutta (RK4) method for the time integration of (3.5). The CG iteration stops when the tolerance $Tol = 10^{-8}$ is reached.

To compare the accuracy of $u_h^{(j,\ell)}$ with $u^{(j,\ell)}$, we compute the reference solution $u_h^{*,(j,\ell)}$ of (4.15), $j, \ell = 1, 2, 3$, with the direct solver. We observe that the relative L^2 -difference $\|u_h^{*,(j,\ell)} - u_h^{(j,\ell)}\|$ lies in $[1.8 \cdot 10^{-6}, 2.5 \cdot 10^{-6}]$ for $j = 1$, in $[1.9 \cdot 10^{-5}, 2.8 \cdot 10^{-5}]$ for $j = 2$, and in $[1.2 \cdot 10^{-4}, 4.4 \cdot 10^{-4}]$ for $j = 3$.

4.2 Helmholtz equation with perfectly matched layer

The perfectly matched layer (PML) was first formulated by Bérenger in [18] in order to solve time-dependent Maxwell's equations in unbounded domains; it absorbs the electromagnetic waves in the absorbing layer without spurious reflection. Since then the PML has been applied to both time-dependent and time-harmonic wave equations. Later this approach has been proven not only to be robust and flexible, but also very accurate.

Here we apply the PML approach for solving unbounded Helmholtz problems. Following [82, 81, 8], we use the PML formulation for the second-order wave equation and write the CMCG method for the “inherited” PML formulation for (2.9).

First, we recall the PML formulation for the wave equation from [82]. Next, instead of using J_1 in (3.33) for the CMCG method for PML formulation, we introduce a new cost functional J_{PML} . It basically extends the integration over Ω in J_1 to the integration over the domain Ω_∞ , which includes the PML layer. Finally, we show some numerical examples, which first shows the accuracy of the numerical solution and second illustrates the usefulness of the CMCG method, combined with the PML approach.

4.2.1 Perfectly matched layer

Here we assume that the source f is compactly supported in $\Omega \subset \mathbb{R}^2$ and is identically zero outside Ω . We extend Ω to a bounded open set

$$\Omega_\infty = \overline{\Omega} \cup \Omega_{PML}, \quad \Omega \cap \Omega_{PML} = \emptyset,$$

with a Lipschitz boundary $\Gamma = \partial\Omega_\infty$ – see Figure 4.7.

We let $\zeta_1, \zeta_2 \geq 0$ denote the damping function with

$$\begin{cases} \zeta_i > 0 & \text{in } \Omega_{PML}, \\ \zeta_i = 0 & \text{in } \overline{\Omega}, \end{cases} \quad i = 1, 2. \quad (4.17)$$

In addition, we define

$$\Gamma_1 = -\begin{pmatrix} \zeta_1 & 0 \\ 0 & \zeta_2 \end{pmatrix}, \quad \Gamma_2 = \begin{pmatrix} \zeta_2 - \zeta_1 & 0 \\ 0 & \zeta_1 - \zeta_2 \end{pmatrix}.$$

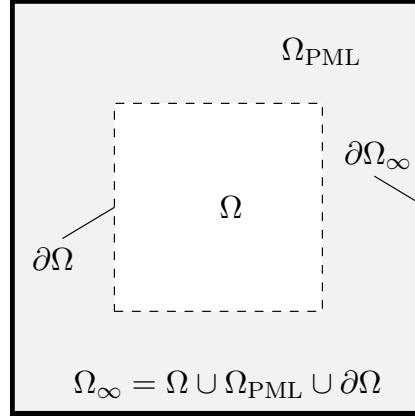


Figure 4.7: Computational domain Ω , surrounded by the PML layer Ω_{PML} .

Let (y, ϕ) be the solution of the PML modified wave equation [82, 8],

$$\frac{1}{c^2} \frac{\partial^2 y}{\partial t^2} - \nabla \cdot (a \nabla y) = \text{Re}\{f e^{-i\omega t}\} \quad \text{in } \Omega \times (0, T) \quad (4.18a)$$

and

$$\frac{1}{c^2} \frac{\partial^2 y}{\partial t^2} - \nabla \cdot (a \nabla y) + (\zeta_1 + \zeta_2) \frac{\partial y}{\partial t} + \zeta_1 \zeta_2 y - \nabla \cdot \phi = 0 \quad \text{in } \Omega_{PML} \times (0, T), \quad (4.18b)$$

$$\frac{\partial}{\partial t} \phi - \Gamma_1 \phi - a \Gamma_2 \nabla y = 0 \quad \text{in } \Omega_{PML} \times (0, T), \quad (4.18c)$$

$$y = 0 \quad \text{on } \Gamma \times (0, T), \quad (4.18d)$$

$$\phi = 0 \quad \text{in } \Omega_{PML} \times (0, T) \quad (4.18e)$$

with the initial conditions

$$y(\cdot, 0) = y_0, \quad \frac{\partial y(\cdot, 0)}{\partial t} = y_1 \quad \text{in } \Omega_\infty, \quad (4.18f)$$

$$\phi(\cdot, 0) = 0 \quad \text{in } \Omega_{PML}. \quad (4.18g)$$

Here the auxiliary function ϕ vanishes identically in $\overline{\Omega}$. Again, we are looking for the unknown initial values y_0, y_1 such that the corresponding wave solution is time-periodic.

We note that a three-dimensional PML formulation can be found in [82].

4.2.2 Controllability method

To determine the initial values y_0, y_1 , we formulate a PDE-constrained least-squares problem. First, we define the cost functional

$$J_{PML}(y_0, y_1) = \frac{1}{2} \int_{\Omega_\infty} a(x) |\nabla y(x, T) - \nabla y_0(x)|^2 dx + \frac{1}{2} \int_{\Omega_\infty} \frac{1}{c^2(x)} (y_t(x, T) - y_1(x))^2 dx. \quad (4.19)$$

To optimize J_{PML} , we require the Fréchet derivative of J_{PML} , which is given by

$$\begin{aligned} \langle J'_{PML}(y_0, y_1), (\delta y_0, \delta y_1) \rangle &= \int_{\Omega_\infty} a(x) \nabla(y(x, T) - y(x, 0)) \cdot \nabla(\delta y(x, T) - \delta y_0(x)) dx \\ &\quad + \int_{\Omega_\infty} \frac{1}{c^2(x)} (y_t(x, T) - y_1(x)) (\delta y_t(x, T) - \delta y_1(x)) dx \end{aligned} \quad (4.20)$$

for the perturbation (y_0, y_1) . Here δy_0 and δy_1 solve the homogeneous wave equation (4.18) with $f = g_S = g_N = g_D = 0$ and the initial values $(\delta y_0, \delta y_1, \delta \phi_0) = 0$. Then, the integration of (4.18a) and (4.18b) over $\Omega_\infty \times (0, T)$ yields

$$\begin{aligned} 0 &= \int_0^T \int_{\Omega_\infty} \left[\frac{1}{c^2} \delta y_{tt} - \nabla \cdot (a \nabla \delta y) \right] p dx dt \\ &\quad + \int_0^T \int_{\Omega_{PML}} \left[(\zeta_1 + \zeta_2) \delta y_t + \zeta_1 \zeta_2 \delta y - \nabla \cdot \delta \phi \right] p dx dt \\ &= \int_0^T \int_{\Omega_\infty} \left[\frac{1}{c^2} p_{tt} - \nabla \cdot (a \nabla p) \right] \delta y dx dt + \int_0^T \int_{\Omega_{PML}} \left[-(\zeta_1 + \zeta_2) p_t + \zeta_1 \zeta_2 p \right] \delta y dx dt \\ &\quad + \int_0^T \int_{\partial \Omega_\infty} a \left[-(\nabla \delta y \cdot n) p + (\nabla p \cdot n) \delta y \right] ds dt + \int_0^T \int_{\Omega_{PML}} \delta \phi \cdot \nabla p dx dt \\ &\quad - \int_0^T \int_{\partial \Omega_{PML}} (\delta \phi \cdot n) p ds dt + \int_{\Omega_\infty} \frac{1}{c^2} \left[p \delta y_t - p_t \delta y \right] \Big|_0^T dx + \int_{\Omega_{PML}} (\zeta_1 + \zeta_2) p \delta y \Big|_0^T dx. \end{aligned} \quad (4.21)$$

From the integration of (4.18c), multiplied with ψ , $\phi = \delta \phi$, and integration by parts, together with the fact that Γ_1 and Γ_2 are diagonal and $\delta y = 0$ on Γ , we obtain

$$\begin{aligned} 0 &= \int_0^T \int_{\Omega_{PML}} \frac{1}{a} \Gamma_2^{-1} \left[\delta \phi_t - \Gamma_1 \delta \phi - a \Gamma_2 (\nabla \delta y) \right] \cdot \psi dx dt \\ &= \int_0^T \int_{\Omega_{PML}} \frac{1}{a} \Gamma_2^{-1} \left[-\psi_t - \Gamma_1 \psi \right] \cdot \delta \phi + \delta y \nabla \cdot \psi dx dt + \int_{\Omega_{PML}} \frac{1}{a} (\Gamma_2^{-1} \psi) \cdot \delta \phi \Big|_0^T dx. \end{aligned} \quad (4.22)$$

Equations (4.21)–(4.22) and integration by parts, together with $p = 0$ on $\partial \Omega_\infty$ and $\psi = 0$ in $\bar{\Omega}$, yield

$$\begin{aligned} 0 &= \int_0^T \int_{\Omega} \left[\frac{1}{c^2} p_{tt} - \nabla \cdot (a \nabla p) \right] \delta y dx dt \\ &\quad + \int_{\Omega_\infty} \frac{1}{c^2} \left[p \delta y_t - p_t \delta y \right] \Big|_0^T dx + \int_{\Omega_{PML}} (\zeta_1 + \zeta_2) p \delta y \Big|_0^T dx - \int_{\Omega_{PML}} \frac{1}{a} (\Gamma_2^{-1} \psi) \cdot \delta \phi \Big|_0^T dx \\ &\quad + \int_0^T \int_{\Omega_{PML}} \left[\frac{1}{c^2} p_{tt} - \nabla \cdot (a \nabla p) - (\zeta_1 + \zeta_2) p_t + \zeta_1 \zeta_2 p - \nabla \cdot \psi \right] \delta y dx dt \\ &\quad + \int_0^T \int_{\Omega_{PML}} \frac{1}{a} \Gamma_2^{-1} \left[\psi_t + \Gamma_1 \psi + a \Gamma_2 \nabla p \right] \cdot \delta \phi dx dt. \end{aligned} \quad (4.23)$$

Let $(p, \boldsymbol{\psi})$ be the solution of the PML modified adjoint wave equation

$$\frac{1}{c^2} \frac{\partial^2 p}{\partial t^2} - \nabla \cdot (a \nabla p) = 0 \quad \text{in } \Omega \times (0, T), \quad (4.24a)$$

$$\frac{1}{c^2} \frac{\partial^2 p}{\partial t^2} - \nabla \cdot (a \nabla p) - (\zeta_1 + \zeta_2) \frac{\partial p}{\partial t} + \zeta_1 \zeta_2 p - \nabla \cdot \boldsymbol{\psi} = 0 \quad \text{in } \Omega_{PML} \times (0, T), \quad (4.24b)$$

$$\frac{\partial}{\partial t} \boldsymbol{\psi} + \Gamma_1 \boldsymbol{\psi} + a \Gamma_2 \nabla p = 0 \quad \text{in } \Omega_{PML} \times (0, T), \quad (4.24c)$$

$$p = 0 \quad \text{on } \Gamma \times (0, T), \quad (4.24d)$$

$$\boldsymbol{\psi} = 0 \quad \text{in } \overline{\Omega} \times (0, T) \quad (4.24e)$$

with the initial conditions

$$p(x, T) = y_t(x, T) - y_1(x), \quad x \in \Omega_\infty, \quad (4.24f)$$

$$\int_{\Omega_\infty} \frac{1}{c^2} p_t(\cdot, T) \varphi \, dx = \int_{\Omega_{PML}} (\zeta_1 + \zeta_2) p(\cdot, T) \varphi \, dx - \int_{\Omega_\infty} a \nabla (y(\cdot, T) - y_0) \cdot \nabla \varphi \, dx \quad (4.24g)$$

for all $\varphi \in H_0^1(\Omega_\infty)$ and $\boldsymbol{\psi}(\cdot, T) = 0$ in Ω_{PML} . Equations (4.23) and (4.24) then yield

$$\begin{aligned} & \int_{\Omega_\infty} \frac{1}{c^2(x)} \left[p(x, 0) \delta y_1(x) - p_t(x, 0) \delta y_0(x) \right] dx + \int_{\Omega_{PML}} (\zeta_1 + \zeta_2) p(x, 0) \delta y_0(x) \, dx \\ &= \int_{\Omega_\infty} \frac{1}{c^2(x)} \left[p(x, T) \delta y_t(x, T) - p_t(x, T) \delta y(x, T) \right] dx \\ &+ \int_{\Omega_{PML}} (\zeta_1(x) + \zeta_2(x)) p(x, T) \delta y(x, T) \, dx \\ &= \int_{\Omega_\infty} \frac{1}{c^2(x)} (y_t(x, T) - y_1(x)) \delta y_t(x, T) \, dx \\ &+ \int_{\Omega_\infty} a(x) \nabla (y(x, T) - y_0(x)) \cdot \nabla \delta y(x, T) \, dx. \end{aligned} \quad (4.25)$$

Combined with (4.20) we have

$$\begin{aligned} \langle J'_{PML}(y_0, y_1), (\delta y_0, \delta y_1) \rangle &= - \int_{\Omega_\infty} a(x) \nabla (y(x, T) - y_0(x)) \cdot \nabla \delta y_0(x) \, dx \\ &- \int_{\Omega_\infty} \frac{1}{c^2} p_t(x, 0) \delta y_0(x) \, dx \\ &+ \int_{\Omega_{PML}} (\zeta_1(x) + \zeta_2(x)) p(x, 0) \delta y_0(x) \, dx \\ &+ \int_{\Omega_\infty} \frac{1}{c^2(x)} (p(x, 0) - p(x, T)) \delta y_1(x) \, dx. \end{aligned} \quad (4.26)$$

Similar to (3.38) in Section 3.4, we need the corresponding (Riesz) representer of (4.26).

4.2.3 Numerical experiments

In this section we consider two scattering problems (2.9) in an unbounded domain. To avoid or reduce the reflection from the artificial boundary, we apply the perfectly matched layer (PML) with the layer size L .

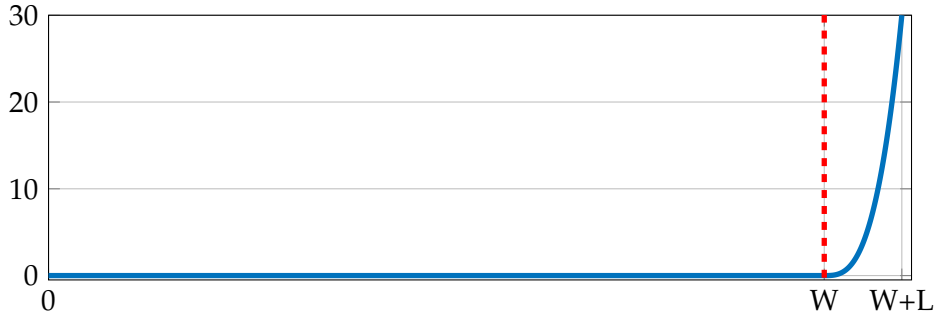


Figure 4.8: Cubic polynomial damping function ζ

In both examples, the damping function $\zeta(z)$ in (4.17) in the layer $W \leq |z| \leq W + L$ is given by the cubic polynomial

$$\zeta(z) = \zeta_0 \left(\frac{|z| - W}{L} \right)^3, \quad (4.27)$$

where ζ_0 is typically set to 20 or 30 – see Figure 4.8.

We first consider the accuracy of the numerical solution of (2.9) in $[-W, W] \times [-W, W]$, obtained with the CMCG method combined with J_{PML} . Furthermore, we compare the numerical solution with that, obtained with the original CMCG method with J_1 . Second, we consider the scattering problem (2.9) with a partial physically bounded domain.

Helmholtz equation in an unbounded domain

First, we consider the numerical solution u_h of the two-dimensional scattering problem (2.9) in $(-W, W) \times (-H, H)$, $W = H = 4$, with $a = c = 1$, $\omega = k = 4\pi$, and the exact solution u given by

$$u(x) = \begin{cases} e^{\frac{1}{16(|x|^2 - W^2)}} \cos\left(\frac{\pi|x|^2}{2}\right), & |x| < W, \\ 0, & |x| \geq W. \end{cases} \quad (4.28)$$

Let f be chosen such that (4.28) solves (2.9).

We apply the CMCG method for solving (2.9), with Ω surrounded by PML with the width of $L = 0.2$, combined with the cost functional J_{PML} . Here the time integration of (4.18), with \mathcal{P}^3 -FE discretization (92'233 nodes and 14'112 elements), is solved by the RK4 method. For comparison, we also compute the solution of (2.9) with $\Gamma_S = \partial\Omega$ and the exact g_S in (2.9b), obtained with the CMCG method with the original cost functional J_1^{CC} combined with the compatibility condition (4.14).

Now, Figures 4.9a and 4.9b monitor the relative errors of both methods. Here we observe that the both relative L^2 - and H^1 -errors decrease and saturate at some point. We finally obtain a numerical solution, obtained with the original approach, with a relative L^2 - and H^1 -error of 0.0045 and 0.0212, respectively. Next, the CMCG method with PML approach yields a slightly better result than the original approach with a relative L^2 - and H^1 -error of 0.0040 and 0.0210, respectively.

Figures 4.9c and 4.9d monitor the relative CG residual and the cost functional J_{PML} from the CMCG algorithm at each CG iteration. We observe that both CMCG methods combined with J_1^{CC} or with J_{PML} converge to the given tolerance $Tol = 10^{-6}$. The CMCG method with J_1 terminates after 144 CG iterations while the CMCG

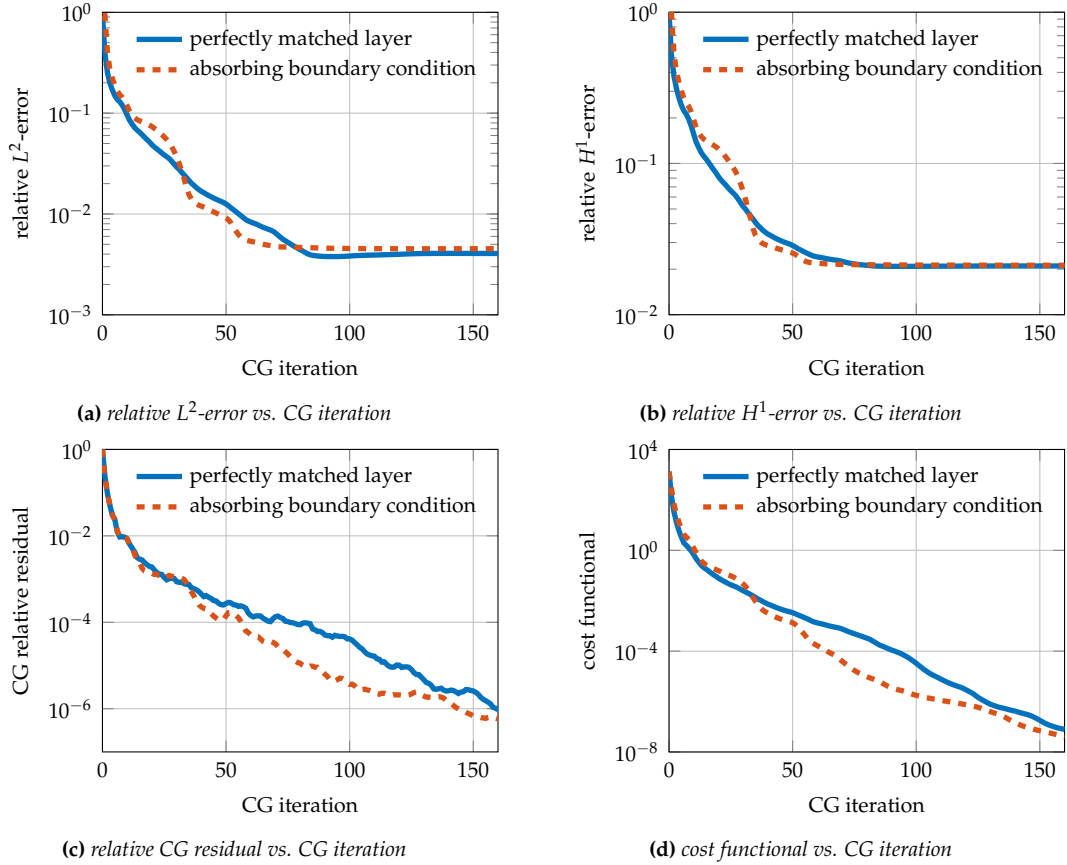


Figure 4.9: Absorbing boundary condition vs. perfectly matched layer: numerical results of the scattering problem (2.9), obtained with the CMCG method combined with either J_1 ($\Gamma_S = \partial\Omega$) or J_{PML} ; (a) – (b) the relative L^2 - and H^1 -error of the numerical solution u_h at each CMCG iteration, (c) the relative CG residual; (d) the cost functional J_1 and J_{PML} .

method with J_{PML} stops after 160 CG iterations. We therefore conclude that both methods terminate in comparable number of CG iterations with comparable accuracy.

Sound-soft scattering from inhomogeneous inclusion

Here we consider the scattering problem (2.9) with $\omega = 20\pi$, $c = 1$, and $a(x) = 1 + H(x)$, where $H(x)$ is an indicator function of the kite-shaped inclusion – see Figure 4.10a. We impose a homogeneous boundary condition on the arc boundary $\Gamma_D = \{x \mid x_2 \geq 0, |x| = 2\}$. The wave goes out through the boundary $\Gamma_{PML} = \{x_1 = -2\} \cup \{x_1 = 2\} \cup \{x_2 = -1\}$. The source term f ,

$$f(x) = -10^3 e^{-800(x_1^2 + (x_2 - 1)^2)}, \quad x \in \Omega_\infty,$$

lies inside the domain Ω at $(0, 1)$.

We use \mathcal{P}^3 -FEM with 165'880 nodes and 25'422 elements for the spatial discretization and the RK4 method for the time integration of (3.5). Furthermore, we add a PML layer with a PML width of $L = 0.5$.

We observe the contour lines of the numerical solution shown in Figure 4.10b, obtained with the CMCG method with (4.19). Here the outgoing wave is indeed damped in PML such that the wave near to boundary vanishes identically.

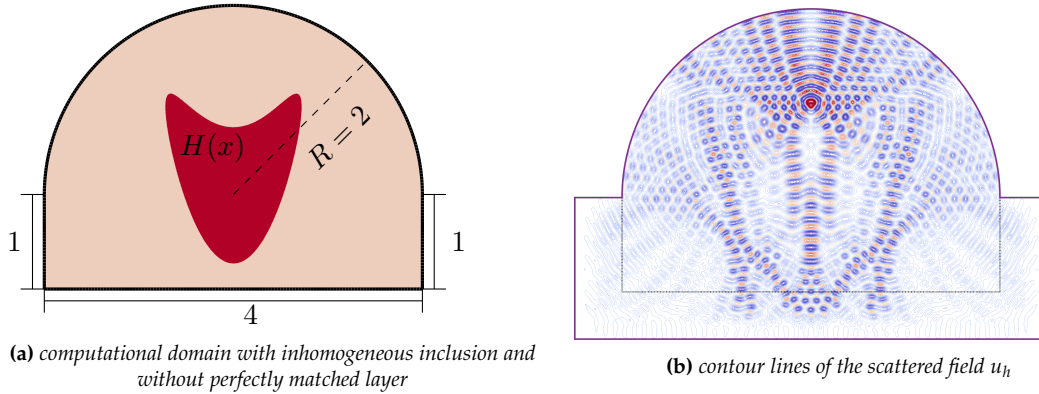


Figure 4.10: Sound-soft scattering problem with PML: (a) the inhomogeneous medium a is constant inside and outside of the kite-shaped inclusion; (b) contour lines of the numerical solution of (2.9) using the CMCG method with the functional J_{PML} defined by (4.19).

In summary, the PML approach applied to the CMCG method leads to a comparably accurate and probably parallelizable method for scattering problems in frequency domain.

4.3 Helmholtz equation in first-order formulation

The CMCG Algorithm from Section 3.7 iterates on the initial value $(y_0, y_1) \in H^1(\Omega) \times L^2(\Omega)$ of the second-order wave equation (3.5) until its solution is T -time periodic. However, the gradient of the cost functional $J_1(y_0, y_1)$, which is needed during the CG update, only lies in the dual space $H^{-1}(\Omega) \times L^2(\Omega)$. To ensure that the solution remains sufficiently regular in $H^1(\Omega) \times L^2(\Omega)$, the corresponding Riesz representative is computed at every CG iteration by solving the strongly elliptic problem (3.38a). In [70], Glowinski et al. derived an equivalent first-order formulation for sound-soft scattering problems, where the solution instead lies in $(L^2(\Omega))^{d+1}$, which is self-dual or is isomorph to its dual space. As a consequence, all CG iterates automatically lie in the correct solution space $(L^2(\Omega))^{d+1}$, while the solution of (3.38a) is no longer needed.

4.3.1 Model problem

Following [70, 33], we let $z = -i\omega u$, $\mathbf{r} = \nabla u$ and rewrite the Helmholtz equation (2.9), with $a \equiv 1$, equivalently as

$$-\nabla \cdot \mathbf{r}(x) - i \frac{\omega}{c^2(x)} z(x) = f(x), \quad x \in \Omega, \quad (4.29a)$$

$$-i\omega \mathbf{r}(x) = \nabla z(x), \quad x \in \Omega, \quad (4.29b)$$

$$\mathbf{r}(x) \cdot \mathbf{n} + \frac{1}{c(x)} z(x) = g_S(x), \quad x \in \Gamma_S, \quad (4.29c)$$

$$\mathbf{r}(x) \cdot \mathbf{n} = g_N(x) \quad x \in \Gamma_N, \quad (4.29d)$$

$$z(x) = -i\omega g_D(x), \quad x \in \Gamma_D. \quad (4.29e)$$

Here $\nabla \cdot \mathbf{r} = \text{div}(\mathbf{r})$ denotes the divergence of \mathbf{r} .

We always assume for any particular choice of ω , c , f , and boundary conditions (4.29c)-(4.29e) on $\Gamma = \Gamma_D \cup \Gamma_S \cup \Gamma_N$ that (4.29) has a unique solution $(\mathbf{r}, z) \in H(\text{div}; \Omega) \times$

$L^2(\Omega)$ with $\mathbf{r} \cdot \mathbf{n} = g_N$ on Γ_N , where

$$H(\operatorname{div}; \Omega) = \{\mathbf{p} \in (L^2(\Omega))^d \mid \nabla \cdot \mathbf{p} \in L^2(\Omega)\}. \quad (4.30)$$

Remark 10. For the unique solution $u \in H^1(\Omega)$ of (2.9), $\widehat{\mathbf{r}} = \nabla u \in (L^2(\Omega))^d$ and $\widehat{z} = -i\omega u \in L^2(\Omega)$ clearly satisfy (4.29) by applying $\varphi \in C_c^\infty(\Omega)$ in its variational formulation:

$$(\nabla \cdot \widehat{\mathbf{r}}, \varphi) = (f + i\frac{\omega}{c^2} \widehat{z}, \varphi), \quad \forall \varphi \in C_c^\infty(\Omega).$$

We conclude that $\nabla \cdot \widehat{\mathbf{r}} = f + i\omega c^{-2} \widehat{z} \in L^2(\Omega)$ a.e. in Ω , and hence $\widehat{\mathbf{r}} \in H(\operatorname{div}; \Omega)$. By the uniqueness, it indeed follows that $\mathbf{r} = \widehat{\mathbf{r}} = \nabla u$ and $z = \widehat{z} = -i\omega u$ a.e. in Ω .

Let $\mathbf{p} = \nabla y$ and $v = y_t$. Then we rewrite the time-dependent wave equation (3.5) in first-order form:

$$\frac{1}{c^2(x)} \frac{\partial}{\partial t} v(x, t) - \nabla \cdot \mathbf{p}(x, t) = \operatorname{Re}\{f(x) e^{-i\omega t}\}, \quad x \in \Omega, t > 0, \quad (4.31a)$$

$$\frac{\partial}{\partial t} \mathbf{p}(x, t) = \nabla v(x, t), \quad x \in \Omega, t > 0, \quad (4.31b)$$

$$\mathbf{p}(x, t) \cdot \mathbf{n} + \frac{1}{c(x)} v(x, t) = \operatorname{Re}\{g_S(x) e^{-i\omega t}\}, \quad x \in \Gamma_S, t > 0, \quad (4.31c)$$

$$\mathbf{p}(x, t) \cdot \mathbf{n} = \operatorname{Re}\{g_N(x) e^{-i\omega t}\}, \quad x \in \Gamma_N, t > 0, \quad (4.31d)$$

$$v(x, t) = \operatorname{Re}\{-i\omega g_D(x) e^{-i\omega t}\}, \quad x \in \Gamma_D, t > 0 \quad (4.31e)$$

with the initial conditions

$$\mathbf{p}(x, 0) = \mathbf{p}_0(x) \in \mathbb{R}^d, \quad v(x, 0) = v_0(x) \in \mathbb{R}, \quad x \in \Omega. \quad (4.31f)$$

The (mixed) variational formulation is to find $\mathbf{p}(t) \in H(\operatorname{div}; \Omega)$ and $v(t) \in L^2(\Omega)$ for a fixed $t > 0$ such that

$$\int_{\Omega} \left(\frac{1}{c^2(x)} \frac{\partial}{\partial t} v(x, t) - \nabla \cdot \mathbf{p}(x, t) \right) w(x) dx = \operatorname{Re} \left\{ \int_{\Omega} (f(x) w(x)) e^{-i\omega t} dx \right\}, \quad (4.32a)$$

$$\begin{aligned} & \int_{\Omega} (\mathbf{p}_t(x, t) \mathbf{q}(x) + v(x, t) \nabla \cdot \mathbf{q}(x)) dx + \int_{\Gamma_S} c(x) (\mathbf{p}(x, t) \cdot \mathbf{n}) (\mathbf{q}(x) \cdot \mathbf{n}) ds \\ &= \operatorname{Re} \left\{ \left(\int_{\Gamma_S} c(x) g_S(x) \mathbf{q}(x) \cdot \mathbf{n} ds - i\omega \int_{\Gamma_D} g_D(x) \mathbf{q}(x) \cdot \mathbf{n} ds \right) e^{-i\omega t} \right\}, \end{aligned} \quad (4.32b)$$

for any $(\mathbf{q}, w) \in H_N^{\operatorname{div}} \times L^2(\Omega)$, where

$$H_N^{\operatorname{div}} = \{\mathbf{q} \in H(\operatorname{div}; \Omega) \mid \mathbf{q} \cdot \mathbf{n} = 0 \text{ on } \Gamma_N\}. \quad (4.33)$$

The solution (\mathbf{p}, v) of (4.31) exists for $(\mathbf{p}_0, v_0) \in H(\operatorname{div}; \Omega) \times L^2(\Omega)$ and lies in the function space \mathcal{Q} [15, 97, 56],

$$\mathcal{Q} = C^0([0, T]; H(\operatorname{div}; \Omega) \times L^2(\Omega)) \cap C^1([0, T]; (L^2(\Omega))^{d+1}). \quad (4.34)$$

Here the solution of (4.31) is also unique. Since the second-order formulation (3.5) yields a solution of (4.31), it establishes the relation between both formulations, which we formulate in the following lemma.

Lemma 3. Suppose the first-order formulation (4.31) has a solution $(\mathbf{p}, v) \in \mathcal{Q}$. Then (\mathbf{p}, v) is unique.

Furthermore, if y solves the second-order formulation (3.5) with $\nabla y(\cdot, 0) = \mathbf{p}(\cdot, 0)$ and $y_t(\cdot, 0) = v(\cdot, 0)$, then $\nabla y = \mathbf{p}$ and $y_t = v$ a.e. in Ω .

Proof. (i) We show the uniqueness by the energy dissipation law

$$E[\mathbf{p}, v](t) := \frac{1}{2} \int_{\Omega} |\mathbf{p}(x, t)|^2 dx + \frac{1}{2} \int_{\Omega} \frac{1}{c^2(x)} |v(x, t)|^2 dx \leq E[\mathbf{p}, v](0).$$

Let (\mathbf{p}_1, v_1) and (\mathbf{p}_2, v_2) be solutions of (4.31) in $H(\operatorname{div}; \Omega) \times L^2(\Omega)$. Then $(\mathbf{q} = \mathbf{p}_1 - \mathbf{p}_2, w = v_1 - v_2)$ solves (4.31) with $f \equiv g_S \equiv g_D \equiv g_N \equiv 0$, $\mathbf{q}(\cdot, 0) = 0$, and $w(\cdot, 0) = 0$ in Ω . To show that the energy $E[\mathbf{q}, w]$ is monotonically decreasing in $t \geq 0$, we consider the derivative of the energy,

$$\frac{\partial}{\partial t} E[\mathbf{q}, w](t) = \int_{\Omega} \mathbf{q}_t(x, t) \cdot \mathbf{q}(x, t) dx + \int_{\Omega} \frac{1}{c^2(x)} w_t(x, t) w(x, t) dx.$$

Equations (4.31a) and (4.31b) follow

$$\frac{\partial}{\partial t} E[\mathbf{q}, w](t) = \int_{\Omega} \nabla w(x, t) \cdot \mathbf{q}(x, t) dx + \int_{\Omega} \frac{1}{c^2(x)} \nabla \cdot (\mathbf{q}(x, t)) w(x, t) dx.$$

Green's formula, together with the boundary conditions (4.31c) – (4.31e), yields

$$\frac{\partial}{\partial t} E[\mathbf{q}, w](t) = \int_{\partial\Omega} w(x, t) \mathbf{q}(x, t) \cdot \mathbf{n} ds = - \int_{\Gamma_S} c(x) w(x, t)^2 ds \leq 0.$$

Hence $E[\mathbf{q}, w](t)$ is monotonically decreasing and

$$0 \leq E[\mathbf{q}, w](t) \leq E[\mathbf{q}, w](0) = 0,$$

which implies that $\mathbf{q} \equiv 0$ and $w \equiv 0$.

(ii) Since $(\nabla y, y_t) \in \mathcal{Q}$ satisfies (4.31a) – (4.31e), the uniqueness (i) implies that $\mathbf{p} = \nabla y$ and $v = y_t$ a.e. in Ω . \square

For sound-soft scattering problems ($\mathcal{H}^{d-1}(\Gamma_D), \mathcal{H}^{d-1}(\Gamma_S) > 0$), the time-periodic solution y of (4.31) is unique, and therefore, corresponds to the (unique) time-harmonic solution $\operatorname{Re}\{u(x) e^{-i\omega t}\}$ of (4.31). For general boundary value problems governed by the Helmholtz equation, however, a time-periodic solution of (4.31) is not necessarily unique, as shown in the following theorem.

Theorem 12. Let $u \in H^1(\Omega)$ be the unique solution of (2.9) and $(\mathbf{p}, v) \in \mathcal{Q}$ be a real-valued solution of (4.31) with initial values $(\mathbf{p}_0, v_0) \in H(\operatorname{div}; \Omega) \times L^2(\Omega)$. If \mathbf{p} and v are time periodic with period $T = 2\pi/\omega$, then \mathbf{p} and v admit the Fourier series representation

$$\mathbf{p}(\cdot, t) = \operatorname{Re}\{\nabla u e^{-i\omega t}\} + \lambda + \sum_{|\ell| \geq 1} \gamma_{\ell}^{\mathbf{p}} e^{-i\omega \ell t}, \quad (4.35a)$$

$$v(\cdot, t) = \omega \operatorname{Im}\{u e^{-i\omega t}\} + \eta + \sum_{|\ell| \geq 1} \gamma_{\ell}^v e^{-i\omega \ell t}, \quad (4.35b)$$

where the constant $\eta \in \mathbb{R}$, $\lambda \in H_N^{\operatorname{div}}$ with

$$\int_{\Omega} \lambda \cdot \nabla \varphi dx = 0, \quad \forall \varphi \in H^1(\Omega), \varphi|_{\Gamma_D} \equiv 0, \quad (4.36)$$

and the complex-valued eigenfunctions $\gamma_\ell^p \in H_N^{\text{div}}$, $\gamma_\ell^v \in L^2(\Omega)$, $|\ell| > 1$, satisfy

$$-c^2(x)\nabla \cdot \gamma_\ell^p(x) + i\omega\ell\gamma_\ell^v(x) = 0, \quad x \in \Omega, \quad (4.37a)$$

$$i\omega\ell\gamma_\ell^p(x) = \nabla\gamma_\ell^v(x), \quad x \in \Omega, \quad (4.37b)$$

$$c(x)\gamma_\ell^p(x) \cdot \mathbf{n} + \gamma_\ell^v(x) = 0, \quad x \in \Gamma_S, \quad (4.37c)$$

$$\gamma_\ell^p(x) \cdot \mathbf{n} = 0, \quad x \in \Gamma_N, \quad (4.37d)$$

$$\gamma_\ell^v(x) = 0, \quad x \in \Gamma_D. \quad (4.37e)$$

Furthermore, if $\mathcal{H}^{d-1}(\Gamma_S \cup \Gamma_D) > 0$, then $\eta = 0$.

Proof. Let

$$\mathbf{q}(\cdot, t) = \mathbf{p}(\cdot, t) - \text{Re}\{\nabla u e^{-i\omega t}\}, \quad w(\cdot, t) = v(\cdot, t) - \omega \text{Im} u e^{-i\omega t}.$$

Then \mathbf{q} and w satisfy (4.31) with $f \equiv g_D \equiv g_S \equiv g_N \equiv 0$ and initial values

$$\mathbf{q}(x, 0) = \mathbf{p}_0(x) - \text{Re}\{\nabla u(x)\}, \quad w(x, 0) = v_0(x) - \omega \text{Im}\{u(x)\}, \quad x \in \Omega.$$

Since \mathbf{p} and v are T -periodic, so are \mathbf{q} and w . Moreover, the mappings

$$t \mapsto (\mathbf{q}(\cdot, t), \boldsymbol{\psi}), \quad t \mapsto (w(\cdot, t), \varphi)$$

are T -periodic and continuous for any $(\boldsymbol{\psi}, \varphi) \in H_N^{\text{div}} \times L^2(\Omega)$ [56]. Hence, they admit the Fourier series representation,

$$(\mathbf{q}(\cdot, t), \boldsymbol{\psi}) = \sum_{\ell=-\infty}^{\infty} \widehat{\gamma}_\ell^p e^{i\omega\ell t}, \quad (w(\cdot, t), \varphi) = \sum_{\ell=-\infty}^{\infty} \widehat{\gamma}_\ell^v e^{i\omega\ell t},$$

where $\widehat{\gamma}_\ell^p \in \mathbb{C}^d$, $\widehat{\gamma}_\ell^v \in \mathbb{C}$. Next, we define

$$\gamma_\ell^p(x) = \oint_0^T \mathbf{q}(x, t) e^{-i\omega\ell t} dt, \quad \gamma_\ell^v(x) = \oint_0^T w(x, t) e^{-i\omega\ell t} dt, \quad (4.38)$$

which implies that

$$\widehat{\gamma}_\ell^p = (\gamma_\ell^p, \boldsymbol{\psi}), \quad \widehat{\gamma}_\ell^v = (\gamma_\ell^v, \varphi).$$

We shall now show that γ_ℓ^p and γ_ℓ^v satisfy (4.37) for all $|\ell| \geq 1$. First, integration by parts, (4.31a) – (4.31b), and the periodicity of \mathbf{q} and w imply that

$$\begin{aligned} \gamma_\ell^v(x) &= \oint_0^T w_t(x, t) \frac{e^{-i\omega\ell t}}{i\omega\ell} dt - \frac{w(x, t) e^{-i\omega\ell t}}{i\omega\ell T} \Big|_0^T = \oint_0^T c^2(x) \nabla \cdot \mathbf{q}(x, t) \frac{e^{-i\omega\ell t}}{i\omega\ell} dt, \\ \gamma_\ell^p(x) &= \oint_0^T \mathbf{q}_t(x, t) \frac{e^{-i\omega\ell t}}{i\omega\ell} dt - \frac{\mathbf{q}(x, t) e^{-i\omega\ell t}}{i\omega\ell T} \Big|_0^T = \oint_0^T \nabla w(x, t) \frac{e^{-i\omega\ell t}}{i\omega\ell} dt. \end{aligned}$$

Together with the definition (4.38) of γ_ℓ^p and γ_ℓ^v , we thus immediately obtain

$$i\omega\ell\gamma_\ell^v - c^2\nabla \cdot \gamma_\ell^p = 0, \quad i\omega\ell\gamma_\ell^p = \nabla\gamma_\ell^v \quad \text{in } \Omega.$$

Since $w(x, t) = 0$ for $x \in \Gamma_D$, we infer from (4.38) that

$$\int_{\Gamma_D} \gamma_\ell^v(x) \varphi(x) ds = \oint_0^T \int_{\Gamma_D} w(x, t) \varphi(x) ds e^{-i\omega\ell t} dt = 0, \quad \varphi \in L^2(\Gamma_D),$$

and hence γ_ℓ^v satisfies (4.37e). Similarly, (4.37c), (4.37d) follow from the fact that \mathbf{q} and w satisfy (4.31c), (4.31d) with $g_N \equiv g_S \equiv 0$. Hence $\gamma_\ell^p, \gamma_\ell^v$ satisfy (4.37) for all $|\ell| \geq 1$. In fact for $\ell = 1$, (4.37) corresponds to (2.9) in first-order formulation with $\gamma_1^p = \nabla \bar{u}$, $\gamma_1^v = i\omega \bar{u}$, homogeneous boundary conditions and no sources. By uniqueness, γ_1^p and γ_1^v , together with their complex conjugates, are therefore identically zero.

Next, we consider γ_0^p, γ_0^v . Again, since \mathbf{q} and w satisfy (4.31a)-(4.31e) with $f = 0$ and homogeneous boundary conditions, we obtain from (4.38) with $\ell = 0$ and the periodicity of \mathbf{q} and w

$$\int_{\Omega} (\nabla \cdot \gamma_0^p) \varphi \, dx = \int_0^T \int_{\Omega} \frac{1}{c^2} w_t \varphi \, dx dt = 0, \quad \forall \varphi \in L^2(\Omega), \quad (4.39)$$

$$\begin{aligned} \int_{\Omega} \gamma_0^v \nabla \cdot \boldsymbol{\psi} \, dx &= \int_0^T \int_{\Omega} \mathbf{q}_t \cdot \boldsymbol{\psi} \, dx dt - \int_0^T \int_{\Gamma_S} w \boldsymbol{\psi} \cdot \mathbf{n} \, ds dt \\ &= \int_0^T \int_{\Gamma_S} c \mathbf{q} \cdot \mathbf{n} \boldsymbol{\psi} \cdot \mathbf{n} \, ds dt = \int_{\Gamma_S} c \gamma_0^p \cdot \mathbf{n} \boldsymbol{\psi} \cdot \mathbf{n} \, ds, \quad \forall \boldsymbol{\psi} \in H_N^{\text{div}}. \end{aligned} \quad (4.40)$$

In particular, (4.39)-(4.40) implies with $\varphi = \gamma_0^v$ and $\boldsymbol{\psi} = \gamma_0^p$ that

$$\int_{\Gamma_S} c |\gamma_0^p \cdot \mathbf{n}|^2 \, ds = 0,$$

and hence, $\gamma_0^p \cdot \mathbf{n} = 0$ on Γ_S , since $c > 0$. Moreover, Green's formula, together with (4.39) and the homogeneous boundary conditions, implies that

$$\int_{\Omega} \gamma_0^p \cdot \nabla \varphi \, dx = - \int_{\Omega} (\nabla \cdot \gamma_0^p) \varphi \, dx + \int_{\partial\Omega} \gamma_0^p \cdot \mathbf{n} \varphi \, ds = 0, \quad \forall \varphi \in H_D^1,$$

and therefore $\lambda = \gamma_0^p$ satisfies (4.36).

To show that γ_0^v is constant, we now let $\varphi \in C_c^\infty(\Omega)$ and $\boldsymbol{\psi} = \mathbf{e}_j \varphi \in H(\text{div}; \Omega)$, $j = 1, \dots, d$, where \mathbf{e}_j is the j -th unit basis vector of \mathbb{R}^d . Integration of (4.31b) over $[0, T]$, the definition (4.38) with $\ell = 0$ and the periodicity of \mathbf{q} then yield

$$0 = \int_0^T \int_{\Omega} \mathbf{q}_t \cdot \boldsymbol{\psi} \, dx dt = - \int_0^T \int_{\Omega} w \nabla \cdot \boldsymbol{\psi} \, dx dt = - \int_{\Omega} \gamma_0^v \frac{\partial \varphi}{\partial x_j} \, dx. \quad (4.41)$$

From (4.41), we conclude that $\partial_{x_j} \gamma_0^v = 0$, $j = 1, \dots, d$, which implies

$$\gamma_0^v(x) \equiv \eta, \quad \gamma_0^v \in H^1(\Omega).$$

Since γ_0^v satisfies (4.31e) with $\ell = 0$, $\eta = \gamma_0^v = 0$, if $\mathcal{H}^{d-1}(\Gamma_D) > 0$. Similarly, if $\mathcal{H}^{d-1}(\Gamma_S) > 0$, (4.31c), together with $\gamma_0^p \cdot \mathbf{n} = 0$ on Γ_S , yields

$$0 = \int_0^T \int_{\Omega} [c(x) \mathbf{q}(x, t) \cdot \mathbf{n} + w(x, t)] \, dt = c(x) \gamma_0^p(x) \cdot \mathbf{n} + \gamma_0^v(x) = \eta, \quad x \in \Gamma_S.$$

Thus, $\eta = 0$ when $\mathcal{H}^{d-1}(\Gamma_D \cup \Gamma_S) > 0$, which completes the proof. \square

For sound-soft scattering problems, where $\mathcal{H}^{d-1}(\Gamma_D) > 0$ and $\mathcal{H}^{d-1}(\Gamma_S) > 0$, $\eta = 0$ and all eigenfunctions $\gamma_\ell^p, \gamma_\ell^v$, $|\ell| > 1$ of (4.37) trivially vanish in (4.35). Therefore, (4.35)-(4.36) in Theorem 12 with $t = 0$ imply that

$$\begin{aligned} (\mathbf{p}_0, \nabla \varphi) &= (\text{Re}\{\nabla u\}, \nabla \varphi), \quad \varphi \in H_D^1(\Omega), \\ v_0 &= \omega \text{Im}\{u\}. \end{aligned}$$

From the real part of (2.9) we then conclude that

$$u = -\frac{1}{k^2} (\operatorname{Re}\{f\} + \nabla \cdot \mathbf{p}_0) + \frac{i}{\omega} v_0. \quad (4.42)$$

4.3.2 Fundamental Frequency Filter

When the CMCG method is applied to the first-order formulation (4.31), any time-periodic solution of (4.31) in (4.35) generally consists of spurious perturbations η , λ and eigenfunctions γ_ℓ^p and γ_ℓ^v superimposed on the desired (unique) solution u of (2.9). To extract u from (\mathbf{p}_0, v_0) , we apply a filtering approach, similar to that in Section 3.5, and thereby restore uniqueness. Again, we multiply the Fourier series representation in (4.35) of v by $e^{i\omega t}$ and integrate over $(0, T)$. Since η and λ are independent of time, while $e^{i\omega t}$ is orthogonal to $e^{i\omega \ell t}$, $|\ell| > 1$, all spurious modes vanish and the resulting expression simplifies to:

$$\frac{2}{T} \int_0^T v(\cdot, t) e^{i\omega t} dt = \frac{2}{T} \int_0^T \operatorname{Re}\{-i\omega u e^{-i\omega t}\} e^{i\omega t} dt = -i\omega u,$$

which immediately yields

$$u(x) = \frac{2i}{T\omega} \int_0^T v(\cdot, t) e^{i\omega t} dt. \quad (4.43)$$

We summarize this result in the following proposition.

Proposition 5. *Let $u \in H^1(\Omega)$ be the unique solution of (2.9) and $(\mathbf{p}, v) \in \mathcal{Q}$ be a T -time periodic solution of (4.31). Then u is given by (4.43).*

4.3.3 Exact controllability method

Here we introduce the corresponding penalty functional $\widehat{J}: H(\operatorname{div}; \Omega) \times L^2(\Omega) \rightarrow \mathbb{R}_{\geq 0}$ to the penalty functional J_1 from (3.33) with

$$\widehat{J}(\mathbf{p}_0, v_0) = \frac{1}{2} \int_\Omega |\mathbf{p}(x, T) - \mathbf{p}_0(x)|^2 dx + \frac{1}{2} \int_\Omega \frac{1}{c^2(x)} (v(x, T) - v_0(x))^2 dx. \quad (4.44)$$

Here again, the CG method requires the Fréchet derivative $\widehat{J}': H(\operatorname{div}; \Omega) \times L^2(\Omega) \rightarrow (H(\operatorname{div}; \Omega))' \times L^2(\Omega)$,

$$\begin{aligned} \langle \widehat{J}'(\mathbf{p}_0, v_0), (\delta \mathbf{p}_0, \delta v_0) \rangle &= \int_\Omega (\mathbf{p}(x, T) - \mathbf{p}_0(x)) \cdot (\delta \mathbf{p}(x, T) - \delta \mathbf{p}_0(x)) dx \\ &\quad + \int_\Omega \frac{1}{c^2(x)} (v(x, T) - v_0(x)) (\delta v(x, T) - \delta v_0(x)) dx \end{aligned} \quad (4.45)$$

for a perturbation $(\delta \mathbf{p}_0, \delta v_0) \in H(\operatorname{div}; \Omega) \times L^2(\Omega)$. Here $\delta \mathbf{p}$ and δv solve the homogeneous wave equation (4.31) (or (4.32)) with $f = g_S = g_N = g_D = 0$, and initial values $(\delta \mathbf{p}_0, \delta v_0)$. Then Equations (4.32a), with $w = v^* \in L^2(\Omega)$, and (4.32b), with $\mathbf{q} = \mathbf{p}^* \in H_N^{\operatorname{div}}(\Omega)$, and integration by parts in time yield

$$\begin{aligned} 0 &= \int_0^T \int_\Omega \left[\frac{1}{c^2} \delta v_t - \nabla \cdot \delta \mathbf{p} \right] v^* dx dt \\ &= \int_\Omega \frac{1}{c^2} \delta v v^* \Big|_0^T dx - \int_0^T \int_\Omega \left[\frac{1}{c^2} \delta v v_t^* + (\nabla \cdot \delta \mathbf{p}) v^* \right] dx dt, \end{aligned} \quad (4.46)$$

$$\begin{aligned}
0 &= \int_0^T \left[\int_{\Omega} (\delta \mathbf{p}_t \cdot \mathbf{p}^* + \delta v \nabla \cdot \mathbf{p}^*) dx - \int_{\Gamma_S} c (\delta \mathbf{p} \cdot \mathbf{n})(\mathbf{p}^* \cdot \mathbf{n}) ds \right] dt \\
&= \int_{\Omega} \delta \mathbf{p} \cdot \mathbf{p}^* \Big|_0^T dx - \int_0^T \left[\int_{\Omega} (\delta \mathbf{p} \cdot \mathbf{p}_t^* - \delta v \nabla \cdot \mathbf{p}^*) dx + \int_{\Gamma_S} c (\delta \mathbf{p} \cdot \mathbf{n})(\mathbf{p}^* \cdot \mathbf{n}) ds \right] dt.
\end{aligned} \tag{4.47}$$

Furthermore, integration by parts and Equations (4.46) and (4.47) imply that

$$\begin{aligned}
\int_{\Omega} \left[\frac{1}{c^2} \delta v v^* - \delta \mathbf{p} \cdot \mathbf{p}^* \right] \Big|_0^T dx &= \int_0^T \int_{\Omega} \left[\frac{1}{c^2} v_t^* + \nabla \cdot \mathbf{p}^* \right] \delta v dx dt \\
&\quad + \int_0^T \int_{\Omega} [v^* (\nabla \cdot \delta \mathbf{p}) - (\mathbf{p}_t^* \cdot \delta \mathbf{p})] dx dt \\
&\quad - \int_0^T \int_{\Gamma_S} c (\delta \mathbf{p} \cdot \mathbf{n})(\mathbf{p}^* \cdot \mathbf{n}) ds dt \\
&= \int_0^T \int_{\Omega} \left[\frac{1}{c^2} v_t^* + \nabla \cdot \mathbf{p}^* \right] \delta v dx dt \\
&\quad - \int_0^T \int_{\Omega} [\mathbf{p}_t^* + \nabla v^*] \cdot \delta \mathbf{p} dx dt \\
&\quad + \int_0^T \left[\int_{\Gamma_D} v^* (\delta \mathbf{p} \cdot \mathbf{n}) ds + \int_{\Gamma_S} (v^* - c \delta \mathbf{p} \cdot \mathbf{n})(\mathbf{p}^* \cdot \mathbf{n}) ds \right] dt.
\end{aligned} \tag{4.48}$$

Let $(\mathbf{p}^*, v^*) \in Q$ be the solution of the adjoint wave equation in first-order form,

$$\frac{1}{c^2(x)} \frac{\partial v^*(x, t)}{\partial t} + \nabla \cdot \mathbf{p}^*(x, t) = 0, \quad x \in \Omega, t > 0, \tag{4.49a}$$

$$\frac{\partial \mathbf{p}^*(x, t)}{\partial t} + \nabla v^*(x, t) = 0, \quad x \in \Omega, t > 0, \tag{4.49b}$$

$$c(x) \mathbf{p}^*(x, t) \cdot \mathbf{n} + v(x, t) = 0, \quad x \in \Gamma_S, t > 0, \tag{4.49c}$$

$$\mathbf{p}^*(x, t) \cdot \mathbf{n} = 0, \quad x \in \Gamma_N, t > 0, \tag{4.49d}$$

$$v^*(x, t) = 0, \quad x \in \Gamma_D, t > 0, \tag{4.49e}$$

with the initial conditions

$$\mathbf{p}^*(x, T) = \mathbf{p}(x, T) - \mathbf{p}_0(x), \quad x \in \Omega, \tag{4.49f}$$

$$v^*(x, T) = -(v(x, T) - v_0(x)), \quad x \in \Omega. \tag{4.49g}$$

Then, Equation (4.48) yields

$$\begin{aligned}
&\int_{\Omega} \left[\delta \mathbf{p}_0(x) \cdot \mathbf{p}^*(x, 0) - \frac{1}{c^2(x)} \delta v_0(x) v^*(x, 0) \right] dx \\
&= \int_{\Omega} \left[\delta \mathbf{p}(x, T) \cdot \mathbf{p}^*(x, T) - \frac{1}{c^2(x)} \delta v(x, T) v^*(x, T) \right] dx \\
&= \int_{\Omega} \left[\delta \mathbf{p}(x, T) \cdot (\mathbf{p}(x, T) - \mathbf{p}_0(x)) + \frac{1}{c^2(x)} \delta v(x, T) (v(x, T) - v_0(x)) \right] dx,
\end{aligned}$$

and, together with (4.45), we obtain

$$\begin{aligned} \langle \widehat{J}'(\mathbf{p}_0, v_0), (\delta \mathbf{p}_0, \delta v_0) \rangle &= \int_{\Omega} (\mathbf{p}^*(x, 0) - \mathbf{p}^*(x, T)) \cdot \delta \mathbf{p}_0(x) dx \\ &\quad - \int_{\Omega} \frac{1}{c^2(x)} (v^*(x, 0) - v^*(x, T)) \delta v_0(x) dx. \end{aligned} \quad (4.50)$$

In contrast to J'_1 in (3.34), for $(g_0, g_1) = \widehat{J}'(\mathbf{p}_0, v_0)$, we have

$$g_0(x) = \mathbf{p}^*(x, 0) - \mathbf{p}^*(x, T) \in H_N^{\text{div}}, \quad (4.51)$$

$$g_1(x) = -\frac{1}{c^2(x)} (v^*(x, 0) - v^*(x, T)) \in L^2(\Omega), \quad (4.52)$$

and hence no Riesz representative is needed.

4.3.4 Hybridizable discontinuous Galerkin discretization

In [70], Glowinski et al. used classical (mixed) Raviart-Thomas (RT) finite elements to discretize (4.31). Since no mass-lumping is available for RT elements on triangles or on tetrahedra [15], each time-step then requires the inversion of the mass-matrix. To avoid that extra computational cost, which also impedes parallelization, we instead consider the recent hybrid discontinuous Galerkin (HDG) finite element (FE) method [37, 38] to discretize (2.9) in first-order formulation and (4.31). Then, the mass-matrix is block-diagonal, with (small and constant) block size equal to the number of degrees of freedoms per element, so that the time-stepping scheme becomes truly explicit and inherently parallel.

Let \mathcal{T}_h denote a regular triangulation of Ω_h , \mathcal{E}_h the set of all faces and \mathcal{P}^r the space of polynomials of degree r . In addition, we define

$$\mathbf{P}_h = \{\mathbf{r} \in (L^2(\Omega))^d : \mathbf{r}|_K \in (\mathcal{P}^r(K))^d, \forall K \in \mathcal{T}_h\}, \quad (4.53)$$

$$V_h = \{w \in L^2(\Omega) : w|_K \in \mathcal{P}^r(K), \forall K \in \mathcal{T}_h\}, \quad (4.54)$$

$$M_h = \{\mu \in L^2(\mathcal{E}_h) : \mu|_F \in \mathcal{P}^r(F), \forall F \in \mathcal{E}_h\}. \quad (4.55)$$

Next, we consider the semi-discrete wave equation (4.32),

$$\frac{\partial}{\partial t} \left(\frac{1}{c^2} v_h, w \right)_{\mathcal{T}_h} = (f, w)_{\mathcal{T}_h} - (\nabla \cdot \mathbf{p}_h, w)_{\mathcal{T}_h} + \langle \widehat{\mathbf{p}}_h \cdot \mathbf{n}, w \rangle_{\partial \mathcal{T}_h}, \quad (4.56a)$$

$$\frac{\partial}{\partial t} (p_h, \mathbf{r})_{\mathcal{T}_h} = - (v_h, \nabla \cdot \mathbf{r})_{\mathcal{T}_h} + \langle \widehat{v}_h, \mathbf{r} \cdot \mathbf{n} \rangle_{\partial \mathcal{T}_h}, \quad (4.56b)$$

$$\langle \widehat{\mathbf{p}}_h \cdot \mathbf{n}, \mu \rangle_{\partial \mathcal{T}_h \setminus \partial \Omega_h} = 0, \quad (4.56c)$$

with the boundary conditions

$$\langle \widehat{\mathbf{p}}_h \cdot \mathbf{n} + \frac{1}{c} \widehat{v}_h, \mu \rangle_{\partial \Omega_h \cap \Gamma_S} = \langle \text{Re}\{g_S \exp(-i\omega t)\}, \mu \rangle_{\partial \Omega_h \cap \Gamma_S}, \quad (4.56d)$$

$$\langle \widehat{\mathbf{p}}_h \cdot \mathbf{n}, \mu \rangle_{\partial \Omega_h \cap \Gamma_N} = \langle \text{Re}\{g_N \exp(-i\omega t)\}, \mu \rangle_{\partial \Omega_h \cap \Gamma_N}, \quad (4.56e)$$

$$\langle \widehat{v}_h, \mu \rangle_{\partial \Omega_h \cap \Gamma_D} = \left\langle \frac{\partial}{\partial t} \text{Re}\{g_D \exp(-i\omega t)\}, \mu \right\rangle_{\partial \Omega_h \cap \Gamma_D}, \quad (4.56f)$$

for some $(\mathbf{r}, w, \mu) \in \mathbf{P}_h \times V_h \times M_h$ and $t \in (0, T)$. Here $(\cdot, \cdot)_K$ denotes the standard inner product of $L^2(K)$, $K \in \mathcal{T}_h$, and $\langle \cdot, \cdot \rangle_F$ the standard scalar product of $L^2(F)$, $F \in \mathcal{E}_h$, and

the numerical flux is given by

$$\widehat{\mathbf{p}}_h \cdot \mathbf{n} = \mathbf{p}_h \cdot \mathbf{n} - \tau(v_h - \widehat{v}_h) \text{ on } \partial\mathcal{T}_h \quad (4.57)$$

with the numerical flux parameter τ . Following [38], the HDG Galerkin FE formulation reads: find $(\mathbf{p}_h, v_h, \widehat{v}_h) \in \mathbf{P}_h \times V_h \times M_h$ such that

$$\left(\frac{\partial \mathbf{p}_h}{\partial t}, \mathbf{r}\right)_K = -(v_h, \nabla \cdot \mathbf{r})_K + \langle \widehat{v}_h, \mathbf{r} \cdot \mathbf{n} \rangle_{\partial K}, \quad (4.58a)$$

$$\left(\frac{1}{c^2} \frac{\partial v_h}{\partial t}, w\right)_K = (f, w)_K - (\mathbf{p}_h, \nabla w)_K + \langle \widehat{\mathbf{p}}_h \cdot \mathbf{n}, w \rangle_{\partial K}, \quad (4.58b)$$

$$\langle \widehat{\mathbf{p}}_h \cdot \mathbf{n} + \frac{1}{c} \widehat{v}_h, \mu \rangle_{\partial K \cap \Gamma_S} = \langle \text{Re}\{g_S \exp(-i\omega t)\}, \mu \rangle_{\partial K \cap \Gamma_S}, \quad (4.58c)$$

$$\langle \widehat{\mathbf{p}}_h \cdot \mathbf{n}, \mu \rangle_{\partial K \cap \Gamma_N} = \langle \text{Re}\{g_N \exp(-i\omega t)\}, \mu \rangle_{\partial K \cap \Gamma_N}, \quad (4.58d)$$

$$\langle \widehat{v}_h, \mu \rangle_{\partial K \cap \Gamma_D} = \langle \omega \text{Im}\{g_D \exp(-i\omega t)\}, \mu \rangle_{\partial K \cap \Gamma_D}, \quad (4.58e)$$

for all $(\mathbf{r}, w, \mu) \in \mathbf{P}_h \times V_h \times M_h$, $K \in \mathcal{T}_h$ and $t \in [0, T]$, where

$$\widehat{\mathbf{p}}_h \cdot \mathbf{n} = \mathbf{p}_h \cdot \mathbf{n} - \tau(v_h - \widehat{v}_h) \text{ on } \partial K, \quad (4.58f)$$

with the stabilization function τ in [38]. In addition, it satisfies the initial conditions

$$\mathbf{p}_h(x, 0) = \mathbf{p}_0(x), \quad v_h(x, 0) = v_0(x), \quad x \in \Omega. \quad (4.58g)$$

For the first-order system (4.58) of ordinary differential equations, we use standard explicit Runge-Kutta (RK) methods.

4.3.5 Post-processing and superconvergence

For a FE discretization with piecewise polynomials of degree r , we usually expect convergence as $\mathcal{O}(h^{r+1})$ with respect to the L^2 -norm. For the above HDG discretization, however, an extra power in h can be achieved by applying a cheap local post-processing step [38, §2.8]. The same (super-) convergence in space of order $r+2$ using only P^r -FE can be achieved with the CMCG method by applying the local post-processing step to the numerical solutions $(\mathbf{p}_h^{n_T}, v_h^{n_T})$ of (4.31) at the final time $T = n_T \Delta t$.

Let $(\mathbf{p}_h^m, v_h^m, \widehat{v}_h^m)$ denotes the fully discrete solution of (4.58) at $t_m = m\Delta t$. First, we compute the new (more accurate) approximation $\mathbf{p}_h^{n_T,*}$ of $\mathbf{p}(\cdot, T)$ by solving the local problem

$$(\mathbf{p}_h^{n_T,*}, \boldsymbol{\psi})_{L^2(K)} = -(v_h^{n_T}, \nabla \cdot \boldsymbol{\psi})_{L^2(K)} + \langle \widehat{v}_h^{n_T}, \boldsymbol{\psi} \cdot \mathbf{n} \rangle_{\partial K}, \quad \forall \boldsymbol{\psi} \in \mathbf{P}_h$$

on each $K \in \mathcal{T}_h$. Then, we calculate the additional approximations $y_h^{n_T,*}$ of $y(\cdot, T) = \text{Re}\{u(x)\}$ given by (4.42), $v_h^{n_T,*}$ of $v(\cdot, T)$ in $\mathcal{P}^{r+1}(K)$, which satisfies

$$\begin{aligned} (\nabla y_h^{n_T,*}, \nabla \varphi)_{L^2(K)} &= (\mathbf{p}_h^{n_T}, \nabla \varphi)_{L^2(K)}, & \forall \varphi \in \mathcal{P}^{r+1}(K), \\ (y_h^{n_T,*}, 1)_{L^2(K)} &= (y_h^{n_T}, 1)_{L^2(K)}, \\ (\nabla v_h^{n_T,*}, \nabla \varphi)_{L^2(K)} &= (\mathbf{p}_h^{n_T,*}, \nabla \varphi)_{L^2(K)}, & \forall \varphi \in \mathcal{P}^{r+1}(K), \\ (v_h^{n_T,*}, 1)_{L^2(K)} &= (v_h^{n_T}, 1)_{L^2(K)}, \end{aligned}$$

for any element $K \in \mathcal{T}_h$. The new approximate solution u is then given by (4.43) with \mathbf{p} and v replaced by $\mathbf{p}_h^{n_T,*}$ and $v_h^{n_T,*}$.

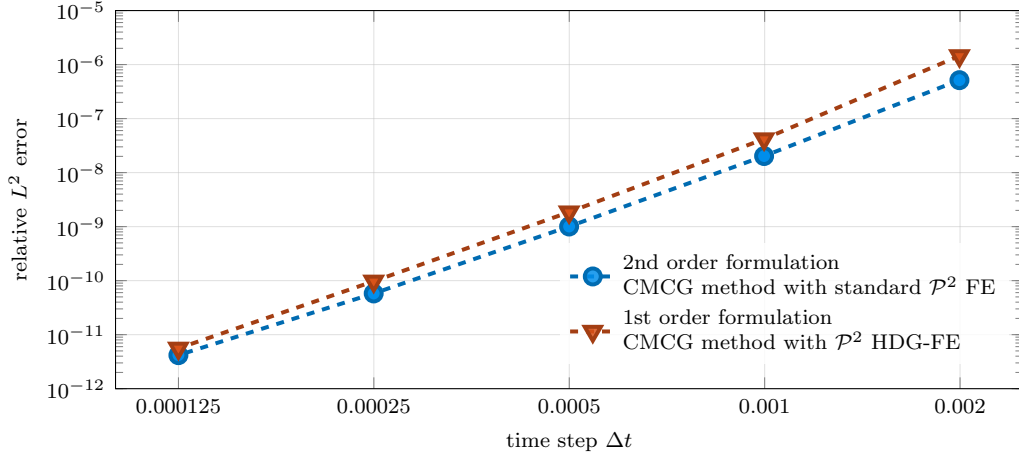


Figure 4.11: HDG semi-discrete convergence: comparison of the numerical solution u_h , obtained with the CMCG method, and u_h^* , obtained with a direct solver for the same fixed \mathcal{P}^2 -FE discretization (H^1 -conforming or HDG).

4.3.6 Numerical experiments

Here we present a series of numerical examples that illustrate the usefulness, accuracy and convergence behavior of the CMCG method for the first-order formulation in various typical configurations. First, we verify that the numerical solution u_h of (4.29), obtained with the CMCG method, converges to the numerical solution u_h^* , obtained with a direct solver for the same spatial HDG-FE discretization, as the time step $\Delta t \rightarrow 0$ in the numerical integration of (4.31). Second, we verify that the CMCG method for the first-order formulation with HDG-FE discretization achieves the expected rates of super convergence after a local post-processing when the minimizer of the cost functional \hat{J} in (4.44) is unique. Third, we study how the number of CG iterations in the CMCG Algorithm depends on the wave number k . Finally, we apply the CMCG method to a Helmholtz equation in a physically bounded domain, where we restore the uniqueness using the filtering procedure from Section 4.3.2.

Semi-discrete convergence

First, we consider a simple 1D example to show for a fixed FE-mesh that the numerical solution u_h , obtained with the CMCG method, converges to the numerical solution u_h^* of (2.9) in first-order formulation, obtained with a direct solver, as $\Delta t \rightarrow 0$. Hence we consider the following solution u of (2.9) in $\Omega = (0, 1)$ with $\omega = k = 6\pi$, $a = c = 1$, and $f \equiv 0$:

$$u(x) = \exp(ikx), \quad \text{with} \quad u(0) = 1, \quad u'(1) - ik u(1) = 0.$$

We use the standard explicit fourth order Runge-Kutta (RK4) method for the time integration of (4.31) in the CMCG Algorithm. For the CG iteration, we always choose $y_0^{(0)} \equiv 0$, $y_1^{(0)} \equiv 0$ and fix the tolerance to $Tol = 10^{-14}$ to ensure convergence to machine precision accuracy.

In Figure 4.11, we monitor the difference between the numerical solution u_h^* of (4.29), obtained with a direct solver, and u_h , obtained with the CMCG method using the first-order formulation. As expected, for increasingly smaller Δt and a fixed stringent tolerance in the CG iteration, the numerical solution of the CMCG method always converges to the discrete solution of the Helmholtz equation. For the sake

| Dirichlet-Sommerfeld | $h = 2^{-1}$ | $h = 2^{-2}$ | $h = 2^{-3}$ | $h = 2^{-4}$ | $h = 2^{-5}$ | $h = 2^{-6}$ | rate |
|-----------------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|------|
| HDG CMCG | $4.2 \cdot 10^{-2}$ | $5.7 \cdot 10^{-3}$ | $7.6 \cdot 10^{-4}$ | $9.8 \cdot 10^{-5}$ | $1.2 \cdot 10^{-5}$ | $1.6 \cdot 10^{-6}$ | 2.94 |
| with (FP) and (CC) | $4.2 \cdot 10^{-2}$ | $5.7 \cdot 10^{-3}$ | $7.6 \cdot 10^{-4}$ | $9.8 \cdot 10^{-5}$ | $1.2 \cdot 10^{-5}$ | $1.6 \cdot 10^{-6}$ | 2.94 |
| with (FP), (CC), and (PP) | $1.3 \cdot 10^{-2}$ | $6.9 \cdot 10^{-4}$ | $3.9 \cdot 10^{-5}$ | $2.3 \cdot 10^{-6}$ | $1.4 \cdot 10^{-7}$ | $9.0 \cdot 10^{-9}$ | 4.10 |
| Neumann-Sommerfeld | $h = 2^{-1}$ | $h = 2^{-2}$ | $h = 2^{-3}$ | $h = 2^{-4}$ | $h = 2^{-5}$ | $h = 2^{-6}$ | rate |
| HDG CMCG | $4.2 \cdot 10^{-2}$ | $5.7 \cdot 10^{-3}$ | $7.6 \cdot 10^{-4}$ | $9.8 \cdot 10^{-5}$ | $1.2 \cdot 10^{-5}$ | $1.6 \cdot 10^{-6}$ | 2.94 |
| with (FP) and (CC) | $4.2 \cdot 10^{-2}$ | $5.7 \cdot 10^{-3}$ | $7.6 \cdot 10^{-4}$ | $9.8 \cdot 10^{-5}$ | $1.2 \cdot 10^{-5}$ | $1.6 \cdot 10^{-6}$ | 2.94 |
| with (FP), (CC), and (PP) | $1.3 \cdot 10^{-2}$ | $6.9 \cdot 10^{-4}$ | $3.9 \cdot 10^{-5}$ | $2.3 \cdot 10^{-6}$ | $1.4 \cdot 10^{-7}$ | $9.0 \cdot 10^{-9}$ | 4.10 |
| Pure Sommerfeld | $h = 2^{-1}$ | $h = 2^{-2}$ | $h = 2^{-3}$ | $h = 2^{-4}$ | $h = 2^{-5}$ | $h = 2^{-6}$ | rate |
| HDG CMCG | $4.2 \cdot 10^{-2}$ | $5.7 \cdot 10^{-3}$ | $7.6 \cdot 10^{-4}$ | $9.8 \cdot 10^{-5}$ | $1.2 \cdot 10^{-5}$ | $1.6 \cdot 10^{-6}$ | 2.94 |
| with (FP) and (CC) | $4.3 \cdot 10^{-2}$ | $6.5 \cdot 10^{-3}$ | $8.9 \cdot 10^{-4}$ | $1.1 \cdot 10^{-4}$ | $1.4 \cdot 10^{-5}$ | $1.8 \cdot 10^{-6}$ | 2.91 |
| with (FP), (CC), and (PP) | $1.3 \cdot 10^{-2}$ | $6.9 \cdot 10^{-4}$ | $3.9 \cdot 10^{-5}$ | $2.3 \cdot 10^{-6}$ | $1.4 \cdot 10^{-7}$ | $9.0 \cdot 10^{-9}$ | 4.10 |
| Pure Dirichlet | $h = 2^{-1}$ | $h = 2^{-2}$ | $h = 2^{-3}$ | $h = 2^{-4}$ | $h = 2^{-5}$ | $h = 2^{-6}$ | rate |
| HDG CMCG | $2.2 \cdot 10^{-1}$ | $1.1 \cdot 10^{-1}$ | $5.5 \cdot 10^{-2}$ | $2.7 \cdot 10^{-2}$ | $1.4 \cdot 10^{-2}$ | $6.8 \cdot 10^{-3}$ | 1.01 |
| with (FP) and (CC) | $4.5 \cdot 10^{-2}$ | $5.8 \cdot 10^{-3}$ | $7.6 \cdot 10^{-4}$ | $9.8 \cdot 10^{-5}$ | $1.2 \cdot 10^{-5}$ | $1.6 \cdot 10^{-6}$ | 2.97 |
| with (FP), (CC), and (PP) | $1.3 \cdot 10^{-2}$ | $6.8 \cdot 10^{-4}$ | $3.8 \cdot 10^{-5}$ | $2.3 \cdot 10^{-6}$ | $1.4 \cdot 10^{-7}$ | $9.0 \cdot 10^{-9}$ | 4.09 |
| Pure Neumann | $h = 2^{-1}$ | $h = 2^{-2}$ | $h = 2^{-3}$ | $h = 2^{-4}$ | $h = 2^{-5}$ | $h = 2^{-6}$ | rate |
| HDG CMCG | $3.9 \cdot 10^{-2}$ | $1.0 \cdot 10^{-2}$ | $9.5 \cdot 10^{-3}$ | $9.5 \cdot 10^{-3}$ | $9.6 \cdot 10^{-3}$ | $9.6 \cdot 10^{-3}$ | 0.41 |
| with (FP) and (CC) | $3.9 \cdot 10^{-2}$ | $5.7 \cdot 10^{-3}$ | $7.6 \cdot 10^{-4}$ | $9.8 \cdot 10^{-5}$ | $1.2 \cdot 10^{-5}$ | $1.6 \cdot 10^{-6}$ | 2.92 |
| with (FP), (CC), and (PP) | $1.4 \cdot 10^{-2}$ | $6.9 \cdot 10^{-4}$ | $3.9 \cdot 10^{-5}$ | $2.3 \cdot 10^{-6}$ | $1.4 \cdot 10^{-7}$ | $9.0 \cdot 10^{-9}$ | 4.11 |
| Dirichlet-Neumann | $h = 2^{-1}$ | $h = 2^{-2}$ | $h = 2^{-3}$ | $h = 2^{-4}$ | $h = 2^{-5}$ | $h = 2^{-6}$ | rate |
| HDG CMCG | $4.4 \cdot 10^{-2}$ | $5.8 \cdot 10^{-3}$ | $7.6 \cdot 10^{-4}$ | $9.8 \cdot 10^{-5}$ | $1.9 \cdot 10^{-2}$ | $1.9 \cdot 10^{-2}$ | 0.24 |
| with (FP) and (CC) | $4.4 \cdot 10^{-2}$ | $5.8 \cdot 10^{-3}$ | $7.6 \cdot 10^{-4}$ | $9.8 \cdot 10^{-5}$ | $1.2 \cdot 10^{-5}$ | $1.6 \cdot 10^{-6}$ | 2.95 |
| with (FP), (CC), and (PP) | $1.9 \cdot 10^{-2}$ | $8.3 \cdot 10^{-4}$ | $4.1 \cdot 10^{-5}$ | $2.4 \cdot 10^{-6}$ | $1.5 \cdot 10^{-7}$ | $9.0 \cdot 10^{-9}$ | 4.20 |

Table 4.1: HDG superconvergence: numerical error $\|u - u_h\|$ vs. mesh size h , obtained with the CMCG method for the first-order formulation with \mathcal{P}^2 -HDG discretization for various boundary conditions, either with or without compatibility condition (CC), filtering procedure (FP), and post-processing (PP).

of comparison, we also compute the numerical solution, obtained with the original CMCG method applied to the second-order formulation using (continuous) \mathcal{P}^2 -FEM with mass-lumping – see Figure 4.11.

Convergence and superconvergence

To illustrate the accuracy and verify the expected convergence rates for the HDG-FE discretizations in the CMCG method, we consider the one-dimensional solution

$$u(x) = -\exp(ikx), \quad x \in \Omega = (0, 1),$$

of (2.9) with $a = c = 1$ and $k = \omega = 5\pi/4$. For various examples, we impose that u satisfies the different combination of boundary conditions (2.9b) – (2.9d) related to the configurations in Section 3.4. Table 4.1 shows the relative L^2 -error $\|u - u_h\|$, obtained with the CMCG method for the first-order formulation (4.29), by using \mathcal{P}^2 -HDG ($r = 2$) discretization on a sequence of increasingly finer meshes $h = 2^{-i}$, $i = 1, \dots, 6$. Clearly as we refine the mesh, we always reduce the time-step in the RK4 method to satisfy the CFL stability condition. The CG iteration stops once the tolerance $Tol = 10^{-12}$ is reached. In addition, we compare the numerical solution with or without imposing compatibility condition (CC), filtering procedure (FP), and post-processing (PP).

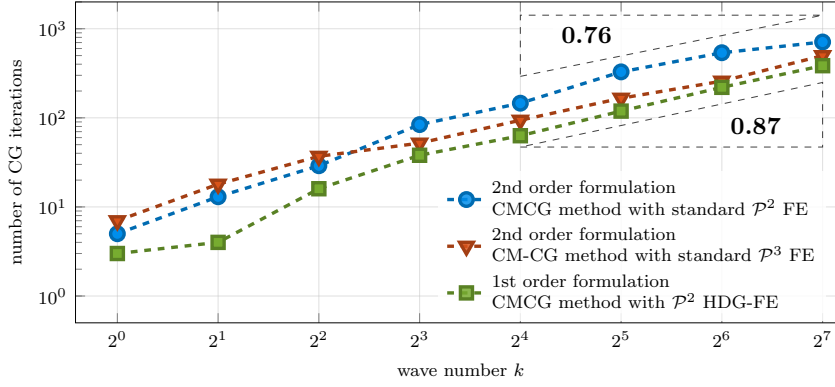


Figure 4.12: One-dimensional sound-soft scattering: the number of CG iterations vs. the wave number $k = 2^i$, $i = 0, \dots, 5$

We observe that the CMCG method for scattering problems ($\mathcal{H}^{d-1}(\Gamma_S) > 0$) yields a solution with the expected optimal convergence of order $r + 1$ with polynomials of degree r . But it fails for Helmholtz equation in physically bounded domains $\Gamma_S = \emptyset$ without filtering procedure and imposing the compatibility condition, since the minimizer of the cost functional is in general not unique. Next, we observe that the CMCG method even achieves a superconvergence of order $r + 2$, once local post-processing is applied to the solution in the final CG iteration.

Wave guide with varying frequency

Here we compare the CMCG method for both formulations for solving the Helmholtz equation (2.9) with the exact solution

$$u(x) = -e^{ikx}, \quad x \in \Omega = (0, 1),$$

exact source terms, and $a = c = 1$. We impose an impedance boundary condition on $\Gamma_S = \{1\}$ and a Dirichlet boundary condition on $\Gamma_D = \{0\}$. We use \mathcal{P}^2 -HDG-FE for first-order formulation and H^1 -conforming \mathcal{P}^2 -FE and \mathcal{P}^3 -FE for second-order formulation. Moreover, we apply the RK4 method for solving the time integration of (4.31) and (3.5), respectively.

Now, we progressively increase the wave number as $k = \omega = 2^i$, $i = 0, \dots, 7$, where the corresponding mesh size h decreases with $hk^{\frac{3}{2}} \sim \text{const}$ – see Section 2.4. Figure 4.12 monitors the number of CMCG iterations required to reach a fixed tolerance of 10^{-10} in Step 6 in Algorithm 2 for both formulations. We observe that the number of CG iterations vs. k essentially increases linearly with an approximate slope of between 0.76 and 0.87.

Hence, the number of CG iterations in the CMCG method generally increases with the wave number k both with and without control of the pollution error.

Physically bounded domain

In the absence of Dirichlet or impedance boundary conditions, the first-order formulation does not yield the correct minimizer of \widehat{J} . As a simple remedy, we proposed in Section 4.3.2 a filtering procedure which removes the unwanted spurious modes. To illustrate the effectiveness of the filtering procedure, we now consider the exact solution of (2.9),

$$u(x) = 16x^2(x-1)^2, \quad x \in \Omega = (0, 1), \quad (4.60)$$

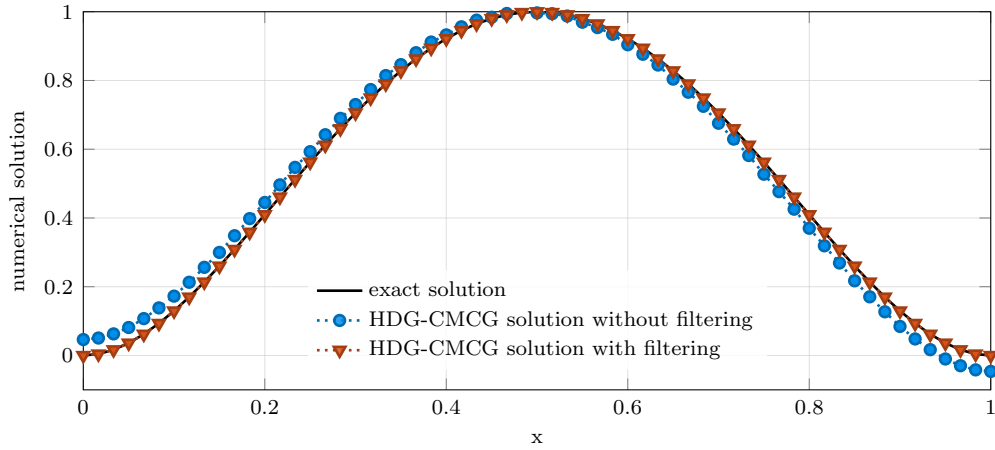


Figure 4.13: Physically bounded domain: comparison of the exact solution u of (4.29) with the numerical solution u_h , obtained with the CMCG method for the first-order formulation with HDG discretization, either with or without filtering procedure.

with homogeneous Neumann boundary conditions, $k = \omega = \pi/4$, and $a = c = 1$. Note that k^2 is not an eigenvalue of (3.21), and therefore (2.9) is well-posed. However, as $(4k)^2 = \pi^2$ indeed corresponds to the first eigenvalue of the negative Laplacian, the CMCG method in general will not yield the unique solution – see Theorems 6 and 12. Indeed as shown in Figure 4.13, the original CMCG method [70, 68] does not yield the exact solution u of (2.9). However, when the CMCG method is combined with the filtering approach proposed in Section 4.3.2, the desired numerical solution indeed coincides with u even for a pure Neumann problem.

Now, we summarize that the uniqueness of the Helmholtz equation using the CMCG method for the first-order formulation can be recovered by filtering procedure and imposing a compatibility condition.

Part II

Inversion of time-harmonic wave equations

CHAPTER 5

Inverse Helmholtz problems

To explore the (unknown) inhomogeneous inclusion, one typically formulates the (inverse) problem as an optimal control problem, where the control u is the squared wave speed. This leads to an ill-posed PDE-constrained optimization problem, which requires the repeated (state) solution y , depending on the control, of the scattering problem (2.9) from Part I. Here we seek after a minimizer u_* , where the corresponding (simulated) wave field y of (2.9) with $a = u_*$ coincides with the measured data y^{obs} . Since the measured data in general contains perturbations, the exact solution u may neither be found nor be unique, which makes the problem harder to solve. Moreover, to capture more detailed properties from the inhomogeneous inclusion, one requires high frequent solutions of (2.9), which again requires a fine mesh and hence leads to a large number of unknowns. This deteriorates the solvability of the optimization problem due to the growing number of local minima. Therefore, one typically requires adding a regularization (or penalty) term into the cost functional, which e.g. penalizes the total variation (TV) of the control [100, 28, 79]. However, such regularization itself involves additional parameters, so that determining appropriate parameter values for the reconstruction can be difficult.

Next, the representation of the control in the grid-based nodal basis, resulting e.g. either from the finite difference (FD) or finite element (FE) discretization, typically leads to a prohibitively large number of unknown variables. Therefore, we instead use the regularization by discretization and size reduction based on the adaptive spectral (AS) space approach in [77, 79], which not only significantly reduces the number of unknowns, but also is very robust and provides an accurate result.

In Section 5.1, we first formulate the inversion problem as a least-squares problem via an optimal control technique. Second, we derive an optimization formulation with penalty term for the regularization and subsequently apply a (standard) Newton-like optimization method on the least-squares problem. The numerical result shows the accuracy of the reconstructed medium and the usefulness of the regularization and illustrates the accuracy of the AS approximation of the medium. Next, in Section 5.2, we consider the regularization of the control u in parameterization based on the spectral decomposition of an (adaptive) linear elliptic operator, which again depends on u . In addition, we analyse the corresponding adaptive spectral (AS) space [9]. Here, a numerical example verifies the main result for a piecewise constant media. In Section 5.3, we apply the AS space approach on the optimal control problem. To reduce the dimension of the AS space efficiently and significantly, we introduce a dynamic size reduction approach based on low-rank approximation using the singular value decomposition (SVD) [72, 85]. Then, we perform a series of numerical experiments to illustrate the accuracy, convergence behavior of the adaptive spectral inversion (ASI), and the size reduction of the AS space.

5.1 PDE-constrained optimization problem

Let y denote the solution of the scattering problem (2.9),

$$-\nabla \cdot (u \nabla y) - \omega^2 y = f \quad \text{in } \Omega, \quad (5.1a)$$

$$\frac{\partial y}{\partial n} - i \frac{\omega}{\sqrt{u}} y = 0 \quad \text{on } \Gamma_S, \quad (5.1b)$$

$$\frac{\partial y}{\partial n} = 0 \quad \text{on } \Gamma_N, \quad (5.1c)$$

$$y = 0 \quad \text{on } \Gamma_D, \quad (5.1d)$$

in a bounded connected polygonal domain $\Omega \subset \mathbb{R}^d$, $d \leq 3$, with a Lipschitz boundary $\partial\Omega = \Gamma_S \cup \Gamma_N \cup \Gamma_D$, where $u(x) \geq u_0 > 0$ is unknown. In contrast to (2.9), the variable coefficients a and c^2 in (2.9) are replaced by the squared wave speed u .

Now, the goal is to explore the unknown (exact) inhomogeneous medium $u \in L^\infty(\Omega)$ from the measured data $y^{\text{obs}} \in L^2(\Gamma)$ on the partial boundary $\Gamma \subset \partial\Omega$ such that the solution y of (5.1) corresponds to y^{obs} . Since the measured data usually contains either additive or multiplicative Gaussian noise, even the solution of (5.1) with the true profile u may not exactly coincide with y^{obs} .

To determine an approximation of u , the inverse problem is formulated as a PDE-constrained optimization problem,

$$\min_{w \in V} J[w] \quad (5.2)$$

for some V , e.g. $V = L^\infty(\Omega)$. The misfit (cost) functional $J : V \rightarrow \mathbb{R}_{\geq 0}$ is

$$J[w] = \frac{1}{2} \sum_{\ell=1}^{N_s} \|y_\ell - y_\ell^{\text{obs}}\|_{L^2(\Gamma)}^2, \quad (5.3)$$

where $y_\ell = y_\ell[w]$ is the solution of (5.1) with $f = f_\ell$ and $g = g_\ell$, $\ell = 1, \dots, N_s$. It basically “penalizes” the misfit between y_ℓ and the observed data y_ℓ^{obs} on the boundary. Here we have N_s observations from N_s different source terms.

It is well-known that the inverse problem is in general severely ill-posed. Therefore, we require an additional regularization term such as the total variation (TV)[58],

$$TV_\varepsilon[w] = \int_\Omega \sqrt{|\nabla w|^2 + \varepsilon^2} \, dx, \quad \varepsilon > 0. \quad (5.4)$$

We note that the parameter $\varepsilon > 0$ is added to the TV regularization term, since ∇w may vanish such that the derivative of TV_ε with $\varepsilon = 0$ may not exist. It is clear that one can instead use, for instance, the regularization with Tikhonov, Gaussian, or Lorentzian penalty term [79, 115]. Now, we instead consider

$$\widehat{J}[w] = J[w] + \alpha R[w] \quad (5.5)$$

with the regularization parameter $\alpha \geq 0$ and $R = TV_\varepsilon$, $\varepsilon > 0$. As a consequence, we solve the PDE-constrained optimization problem,

$$\min_{w \in V} \widehat{J}[w] = \min_{w \in V} \{J[w] + \alpha R[w]\}. \quad (5.6)$$

The choice of α depends on the model problem and one usually requires that $\alpha R[w]$ is slightly smaller than $J[w]$. Once the parameter is too large, we only minimize the

total variation of w instead of $J[w]$. Many strategies such as the L -curve criterion [86] are proposed to choose the parameter adaptively. In [115], Nahum proposed

$$\alpha = \frac{J[u]}{R[u]}. \quad (5.7)$$

Based on [58] we determine α as follows: for a (local) minimizer u of $\widehat{J}[w]$ it holds

$$\langle \widehat{J}'[u], \delta u \rangle = \langle J'[u], \delta u \rangle + \alpha \langle R'[u], \delta u \rangle = 0, \quad \forall \delta u,$$

where $\langle \cdot, \cdot \rangle$ denotes the corresponding dual pairing and the derivatives of J' and R' are given in Section 5.1.1. For $\delta u = u$, we obtain

$$\alpha = -\frac{\langle J'[u], u \rangle}{\langle R'[u], u \rangle}. \quad (5.8)$$

To ensure that the total variation regularization $R = TV_\varepsilon$ in (5.4) is well-defined, we instead consider the approximation u_δ of u_* in $V := W^{1,\infty}(\Omega)$.

5.1.1 Newton-type optimization method

Similar to Section 3.4, we require the (Fréchet) derivative of the cost functional \widehat{J} in (5.5) for the optimization. Here we follow the optimize-then-discretize approach.

Let δu be a perturbation. Then we obtain

$$\langle \widehat{J}'[u], \delta u \rangle = \langle J'[u], \delta u \rangle + \alpha \langle TV'_\varepsilon[u], \delta u \rangle \quad (5.9)$$

with

$$\langle J'[u], \delta u \rangle = \sum_{\ell=1}^{N_s} \operatorname{Re} \left\{ \int_{\Gamma} (y_\ell - y_\ell^{\text{obs}}) \overline{\delta y_\ell} ds \right\} \quad (5.10)$$

and

$$\langle TV'_\varepsilon[u], \delta u \rangle = \left\langle -\nabla \cdot \left(\frac{\nabla u}{\sqrt{|\nabla u|^2 + \varepsilon^2}} \right), \delta u \right\rangle = \int_{\Omega} \frac{\nabla u \cdot \nabla \delta u}{\sqrt{|\nabla u|^2 + \varepsilon^2}} dx, \quad (5.11)$$

where the derivative $\delta y_\ell := \frac{\partial}{\partial u} y_\ell$ solves (5.1) with $u = \delta u$ and $f_\ell = g_\ell = 0$.

Now, we consider the linear operator $\mathcal{A}[u] : H^1(\Omega) \rightarrow (H^1(\Omega))'$,

$$\mathcal{A}[u] y = -\nabla \cdot (u \nabla y) - \omega^2 y. \quad (5.12)$$

By Equation (5.1a), we obtain $\mathcal{A}[u] y_\ell = f_\ell$ in Ω , and furthermore, we have

$$\frac{\partial}{\partial u} (\mathcal{A}[u] y_\ell) = \left(\frac{\partial}{\partial u} \mathcal{A}[u] \right) y_\ell + \mathcal{A}[u] \delta y_\ell = \frac{\partial}{\partial u} f_\ell = 0. \quad (5.13)$$

Let $z_\ell \in H^1(\Omega)$ be the solution to the adjoint scattering problem,

$$-\nabla \cdot (u \nabla z_\ell) - \omega^2 z_\ell = f_\ell^* := \begin{cases} y_\ell - y_\ell^{\text{obs}} & \text{on } \Gamma, \\ 0 & \text{in } \overline{\Omega} \setminus \Gamma, \end{cases} \quad (5.14a)$$

$$\frac{\partial z_\ell}{\partial n} + i \frac{\omega}{\sqrt{u}} z_\ell = 0 \quad \text{on } \Gamma_S, \quad (5.14b)$$

$$\frac{\partial z_\ell}{\partial n} = 0 \quad \text{on } \Gamma_N, \quad z_\ell = 0 \quad \text{on } \Gamma_D. \quad (5.14c)$$

By (5.14a) and (5.12), we again obtain

$$\mathcal{A}^*[u] z_\ell = f_\ell^* \quad \text{in } \Omega.$$

Next, we integrate (5.13) multiplied with z_ℓ over Ω . Integration by parts and the definition (5.12), together with the boundary conditions, yield

$$\begin{aligned} \int_{\Omega} (\mathcal{A}[u] \delta y_\ell) \bar{z}_\ell dx &= - \int_{\Omega} \nabla \cdot (u \nabla \delta y_\ell) \bar{z}_\ell dx - \omega^2 \int_{\Omega} \delta y \bar{z}_\ell dx \\ &= - \int_{\Omega} \nabla \cdot (u \nabla \bar{z}_\ell) \delta y_\ell dx - \omega^2 \int_{\Omega} \bar{z}_\ell \delta y dx \\ &= \int_{\Omega} (\mathcal{A}^*[u] \bar{z}_\ell) \delta y_\ell dx = \int_{\Gamma} (\overline{y_\ell - y_\ell^{\text{obs}}}) \delta y_\ell ds. \end{aligned} \quad (5.15)$$

Furthermore, Equation (5.13) implies that

$$\int_{\Omega} (\mathcal{A}[u] \delta y_\ell) \bar{z} dx = - \int_{\Omega} \left(\left(\frac{\partial}{\partial u} \mathcal{A}[u] \right) y_\ell \right) \bar{z}_\ell dx = - \int_{\Omega} \delta u \nabla y_\ell \cdot \nabla \bar{z}_\ell dx. \quad (5.16)$$

As a result, together with Equations (5.10), (5.15), and (5.16), we finally obtain

$$\langle J'[u], \delta u \rangle = - \text{Re} \left\{ \int_{\Omega} \sum_{\ell=1}^{N_s} \delta u \nabla y_\ell \cdot \nabla \bar{z}_\ell dx \right\}. \quad (5.17)$$

Here we consider u and $y = y[u]$ in (5.3) in the finite dimensional subspaces $\mathcal{V}_h = \{\psi_j\}_{j=1}^K \subset V$ and $W_h = \{\varphi_j\}_{j=1}^N \subset H^1(\Omega)$, respectively. Furthermore, $\beta \in \mathbb{R}^K$ denotes the discrete representative solution of u in \mathcal{V}_h with

$$u \approx \sum_{j=1}^K \beta^j \psi_j.$$

We then minimize the discrete cost function $\widehat{J}: \mathbb{R}^K \rightarrow \mathbb{R}_{\geq 0}$ by using its discrete gradient $\widehat{J}': \mathbb{R}^K \rightarrow \mathbb{R}^K$.

Now, we apply a Newton-type based gradient method with the iteration

$$\mathbf{H}_\ell \mathbf{d}_\ell = -\widehat{J}'[\beta_\ell], \quad (5.18a)$$

$$\beta_{\ell+1} = \beta_\ell + \rho_\ell \mathbf{d}_\ell, \quad (5.18b)$$

for some $\rho_\ell \in \mathbb{R}$ and symmetric positive definite matrix \mathbf{H}_ℓ to optimize the nonlinear least-squares problem (5.2), where \mathbf{d}_ℓ denotes the search direction. To determine the line step ρ_ℓ , we use the Wolfe-Powell or Armijo line search methods [117].

For $\mathbf{H}_\ell = \text{Hess}(\widehat{J})[\beta_\ell]$, the Hessian matrix of \widehat{J} , (5.18) corresponds to the classical Newton iteration. It is well-known that the classical Newton method has a quadratic rate of convergence when the iterative solution is near to a minimizer. However, the Newton method may fail when the iterative solution is close to a saddle point, since the Hessian matrix of the function becomes indefinite. Moreover, it may be expensive to determine the Hessian matrix of \widehat{J} exactly. Therefore, we use the BFGS method [117], a quasi-Newton method with rank-2 update. Since, for a large system, it still requires to store the (dense) Hessian matrix, and thus, we instead use the limited memory BFGS (L-BFGS) method [106, 34].

It is clear that other nonlinear optimization methods such as the nonlinear CG method (Fletcher-Reeves, Polak-Ribière, or Hestenes-Stiefel), inexact Newton method, or Gauss-Newton method, can also be applied to (5.2) [117].

To prevent that the optimization procedure converges to a false local minimum, we apply the frequency continuation strategy from [35, 10, 94]. Starting at the lowest frequency ν_1 , a fixed $\alpha_1 > 0$ in (5.5), and the initial guess $u^{(0)}$, we seek a local minimizer $u^{(1)}$ with the BFGS method. Next, we update the new regularization parameter α_m , $m >$, given by (5.8), namely,

$$\alpha_m = \theta_m \cdot \left| \frac{\langle J'[u^{(m)}], u^{(m)} \rangle}{\langle R'[u^{(m)}], u^{(m)} \rangle} \right|, \quad 0 < \theta_m \leq 1, \quad m \geq 1. \quad (5.19)$$

Then, we increase the frequency and rerun again the BFGS method to determine the new solution (or minimizer). Note that one can e.g. fix $\theta_m = \frac{1}{2}$ in (5.19), $m \geq 1$.

Here we list the algorithm of the (standard) inversion method using a grid-based nodal basis.

Algorithm 3: Standard inversion method with grid-based nodal basis

Input: initial guess $u^{(0)}$, regularization parameter α_1

Output: $u^{(m)}$

```

1 for  $m = 1, 2, \dots$  do
2   Find a (local) minimizer  $u^{(m)}$  of  $J + \alpha_m R$  in the FE space starting at  $u^{(m-1)}$ .
3   if  $\|u^{(m)} - u^{(m-1)}\| < Tol$  then
4     | Return  $u^{(m)}$  and stop.
5   Update the new regularization parameter  $\alpha_{m+1}$  by (5.19).
6    $\nu_{m+1} = \nu_m + \Delta\nu$ 
```

5.1.2 Numerical examples

Here we consider an example with the unknown piecewise constant profile

$$u(x) = 2 + \chi_S(x) + \chi_B(x),$$

shown in Figure 5.1a, where χ_S and χ_B denote the characteristic function of a disk B and a star S , respectively, with $S \subset\subset B$. Now, we apply the \mathcal{P}^1 -FE for the inverse problem (5.6) and \mathcal{P}^3 -FE for the Helmholtz equation (5.1). Next, we minimize the misfit \hat{J} in (5.5) between the synthetic data y and the observed data y^{obs} , where both are obtained by solving the Helmholtz equation with the corresponding squared wave speed. Here the observed data, recorded on the boundary $\Gamma = \partial\Omega$ from the eight Gaussian sources located at $(0.05, 0.05)$, $(0.5, 0.05)$, $(0.95, 0.05)$, $(0.05, 0.5)$, $(0.95, 0.5)$, $(0.05, 0.95)$, $(0.5, 0.95)$, and $(0.95, 0.95)$, contains 20% white noise. To avoid any inverse crime, the mesh for the observation is independent of the synthetic data and is about 30% finer than the mesh for the synthetic data.

Starting at the lowest frequency $\nu_1 = 4$ and the regularization parameter $\alpha_1 = 5 \cdot 10^{-5}$, we apply Algorithm 3 with the TV-penalty and $\varepsilon = 10^{-6}$ to reconstruct the unknown medium u . The frequency increases in each iteration by $\Delta\nu = 2$ up to $\nu_{20} = 42$.

Figure 5.1b shows the reconstruction $u^{(20)}$ of u with a relative L^2 -error of 3.7%. Here we observe that the inversion method captures precisely the shapes. Next, we repeat the same example without the TV-penalty term. It yields a larger relative L^2 -error of 16.89%. However, the reconstruction here shown in Figure 5.1c recovers the star-type shape better and sharper at the corners than the reconstruction with a penalty term.

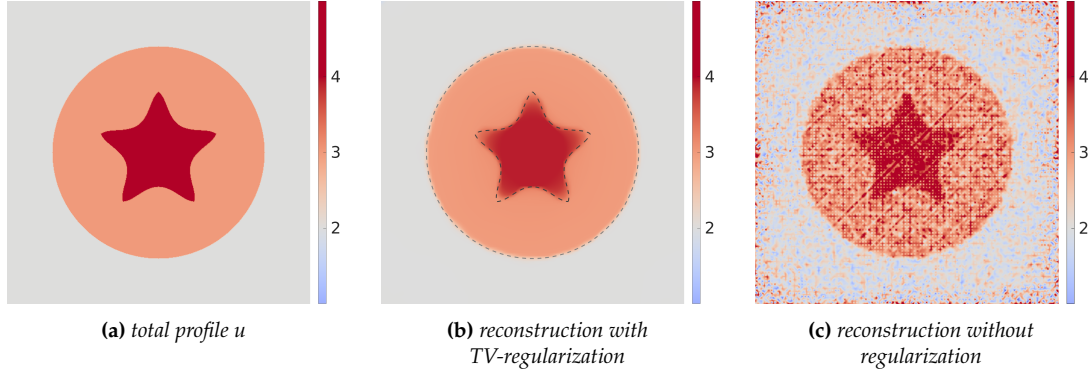


Figure 5.1: Scatterer from circular and star-shaped inclusion: (a) exact total profile u ; (b) reconstruction of u with TV-penalty term; (c) reconstruction of u without regularization term.

5.2 Adaptive spectral decomposition

It is well-known that the optimization problem of minimizing (5.3) in $L^\infty(\Omega)$ is, in general, severally ill-posed. To obtain a well-posed problem, we add a Tikhonov-type regularization term $R[u]$ into the misfit functional $J[u]$ in (5.3). However, this approach does not address the costs of solving an optimization problem in a higher dimensional space, as the dimension of set of discrete media is determined by the number of nodal variables resulting from the FD or FE discretization. An alternative approach for obtaining a well-posed problem in lower dimensional space, with keeping the media reasonably accurate, is proposed in [77, 79]. Instead of adding a penalty term in the original misfit functional J , the control u is regularized by parametrization in a finite-dimensional (spectral) space spanned by the eigenfunctions of the linear elliptic operator $L_\varepsilon[w]$, $\varepsilon > 0$. The linear elliptic operator

$$L_\varepsilon[w] = -\nabla \cdot (\mu_\varepsilon[w] \nabla), \quad \mu_\varepsilon[w] = \frac{1}{\sqrt{|\nabla w|^2 + \varepsilon^2}} \quad \text{or} \quad \mu_\varepsilon[w] = \frac{1}{\max\{|\nabla w|, \varepsilon\}},$$

with $w = u$, however, again depends on the control. Therefore, the spectral space requires an (initial) guess or approximation \hat{u} of u . For given $w = \hat{u}$, the (adaptive) spectral (AS) space consists of eigenfunctions φ_ℓ of the eigenvalue problem

$$L_\varepsilon[w](\varphi_\ell) = \lambda_\ell \varphi_\ell, \quad 0 < \lambda_1 \leq \lambda_2 \leq \dots$$

Then, the spectral expansion of u in AS space is truncated by a finite number K , namely

$$u \approx \sum_{\ell=1}^K \beta_\ell \varphi_\ell, \quad \beta_\ell \in \mathbb{R}.$$

For a piecewise constant medium u , numerical examples in [43, 77, 79] showed that the AS approach efficiently reduces the number of degrees of freedoms and still yields remarkably accurate approximations.

Note that for a piecewise constant function u , the operator $L_\varepsilon[u]$ is not well-defined. Therefore, we instead consider a regular approximation u_δ of u , e.g., a projection of u on an H^1 -conforming FE space.

Here we do the analysis of the AS space for piecewise constant media in a more general setting [9].

5.2.1 Adaptive spectral space

For the analysis, we assume that for a given function w , $\mu_\varepsilon[w]$, $\varepsilon > 0$, is given by

$$\mu_\varepsilon[w](x) = \hat{\mu}_\varepsilon(|\nabla w(x)|), \quad x \in \Omega, \quad (5.20)$$

where $\hat{\mu}_\varepsilon : [0, \infty) \rightarrow \mathbb{R}$ is a nonincreasing function that satisfies

$$\hat{\mu}_\varepsilon(0) = \varepsilon^{-1}, \quad 0 < \hat{\mu}_\varepsilon(\tau), \quad \tau \hat{\mu}_\varepsilon(\tau) \leq 1, \quad \tau \geq 0. \quad (5.21)$$

In particular, the above yields for w with $\nabla w \in L^\infty(\Omega)$,

$$0 < \hat{\mu}_\varepsilon(\|\nabla w\|_{L^\infty(\Omega)}) \leq \hat{\mu}_\varepsilon(|\nabla w|) = \mu_\varepsilon[w]. \quad (5.22)$$

In practice, we choose for $\hat{\mu}_\varepsilon$

$$\hat{\mu}_\varepsilon(\tau) = (\tau^q + \varepsilon^q)^{-\frac{1}{q}}, \quad q \in [1, \infty). \quad (5.23)$$

We note that μ_ε in (5.20) with $\hat{\mu}_\varepsilon$ given by (5.23), with $q = 2$, coincides with the derivative of TV_ε in (5.11) – see [79]. Furthermore, μ_ε in (5.20) with

$$\hat{\mu}_\varepsilon(\tau) = \max(\tau, \varepsilon)^{-1}$$

was originally used in [43, 77].

To include FE approximations in our analysis, we consider the eigenvalues and eigenfunctions of the operator $L_\varepsilon[u_\delta]$ in closed subspaces \mathcal{V}^δ and \mathcal{V}_0^δ of $H^1(\Omega)$ and $H_0^1(\Omega)$, respectively. More precisely, \mathcal{V}_0^δ is the subspace of \mathcal{V}^δ of functions with zero trace on $\partial\Omega$ – see the trace theorem in Section 2.1.

Let $\varphi_0 \in \mathcal{V}^\delta$ denote the solution of the strongly elliptic boundary value problem

$$B[\varphi_0, v] = 0, \quad \forall v \in \mathcal{V}_0^\delta, \quad (5.24a)$$

$$\varphi_0 = g, \quad \text{on } \partial\Omega, \quad (5.24b)$$

where $g = u_\delta$ on $\partial\Omega$ and $B[\cdot, \cdot]$ is the bilinear form given by

$$B[v, w] = \langle \mu_\varepsilon[u_\delta] \nabla v, \nabla w \rangle. \quad (5.25)$$

We say that $\lambda \in \mathbb{R}_{>0}$ is an eigenvalue of $L_\varepsilon[u_\delta]$ in \mathcal{V}_0^δ , if there exists an eigenfunction $0 \neq \varphi \in \mathcal{V}_0^\delta$ such that

$$B[\varphi, v] = \lambda \langle \varphi, v \rangle, \quad \forall v \in \mathcal{V}_0^\delta. \quad (5.26)$$

Following the spectral theorem [56, §6], the sequence of eigenvalues $\{\lambda_k\}$ of $L_\varepsilon[u_\delta]$ in \mathcal{V}_0^δ is strictly positive, real-valued, and nondecreasing. Moreover, the sequence of corresponding eigenfunctions $\{\varphi_k\}$ is an orthonormal basis with respect to the L^2 inner product of \mathcal{V}_0^δ . Clearly, the eigenvalues λ_k and eigenfunctions φ_k depend on δ and ε . To simplify notation, we write λ_k for $\lambda_k[\varepsilon, \delta]$, etc., when it is clear from the context.

5.2.2 Notation

In [77, 79], Grote et al. proposed to construct affine spaces of candidate functions adaptively in an iterative process. The procedure relies on the decomposition

$$u = u^0 + \tilde{u} \text{ in } \Omega, \quad u^0 \in \mathcal{V}^\delta, \quad \tilde{u} \in \mathcal{V}_0^\delta. \quad (5.27)$$

We suppose that the background u^0 of the medium has the form

$$u^0 = \sum_{m=1}^M \hat{u}^{0,m} \chi^{0,m}, \quad (5.28)$$

where $\chi^{0,m}$ denotes the characteristic function of a set

$$B^m = \tilde{B}^m \cap \overline{\Omega} \quad (5.29)$$

for some mutually disjoint connected open sets $\tilde{B}^1, \dots, \tilde{B}^M$ such that

$$\overline{\Omega} \subset \overline{\bigcup_{m=1}^M \tilde{B}^m}, \quad \mathcal{H}^{d-1}(\tilde{B}^m \cap \partial\Omega) > 0. \quad (5.30)$$

We further suppose that \tilde{u} is given by

$$\tilde{u} = \sum_{k=1}^K \hat{u}^k \chi^k, \quad \hat{u}^k \neq 0, \quad k = 1, \dots, K, \quad (5.31)$$

where χ^k is the characteristic function of a connected open set $A^k \subset \subset \Omega \setminus S$, where

$$S = \left(\bigcup_{m=1}^M \partial B^m \right) \setminus \partial\Omega, \quad (5.32)$$

and the sets A^1, \dots, A^K are open, connected, and with mutually disjoint boundaries.

We recall that for piecewise constant function u , we instead consider a regular approximation $u_\delta \in W^{1,\infty}(\Omega)$ of u . Suppose that $u_\delta = u_\delta^0 + \tilde{u}_\delta$, where

$$u_\delta^0 = \sum_{m=1}^M \hat{u}^{0,m} \chi_\delta^{0,m} \in \mathcal{V}^\delta, \quad \tilde{u}_\delta = \sum_{k=1}^K \hat{u}^k \chi_\delta^k \in \mathcal{V}_0^\delta. \quad (5.33)$$

The approximation $\{\chi_\delta^{0,m}\}_{\delta>0}$ of $\chi^{0,m}$ satisfies $\lim_{\delta \searrow 0} \chi_\delta^{0,m} = \chi^{0,m}$ in $L^2(\Omega)$ and $\chi_\delta^{0,m} \in \mathcal{V}^\delta$, for all δ . Similarly, for each $k = 1, \dots, K$, the approximation $\{\chi_\delta^k\}_{\delta>0}$ of χ^k satisfies $\lim_{\delta \searrow 0} \chi_\delta^k = \chi^k$ in $L^2(\Omega)$ and $\chi_\delta^k \in \mathcal{V}_0^\delta$, for all δ .

For $\delta > 0$, let

$$U_\delta = \bigcup_{k=1}^K U_\delta^k, \quad U_\delta^k = \{x \in \Omega \mid \text{dist}(x, (\partial A^k) \cap \Omega) < \delta\}, \quad (5.34)$$

$$A_\delta = \bigcup_{k=1}^K A_\delta^k, \quad A_\delta^k = (\Omega \setminus \overline{U_\delta^k}) \cap A^k, \quad (5.35)$$

$$S_\delta = \{x \in \Omega \mid \text{dist}(x, S) < \delta\}, \quad (5.36)$$

$$D_\delta = \Omega \setminus (\overline{U_\delta \cup S_\delta}). \quad (5.37)$$

The region U_δ^k basically surrounds the boundary of each shape A^k . For our analysis, U_δ^k includes the “discontinuous jump” of the characteristic function χ^k and S_δ the “discontinuous jump” in the background. Moreover, u is constant a.e. in D_δ and \tilde{u} vanishes identically in $D_\delta \setminus \overline{A_\delta}$. Figures 5.2 and 5.3 illustrate the definition of (5.29), (5.32), and (5.34)–(5.37).

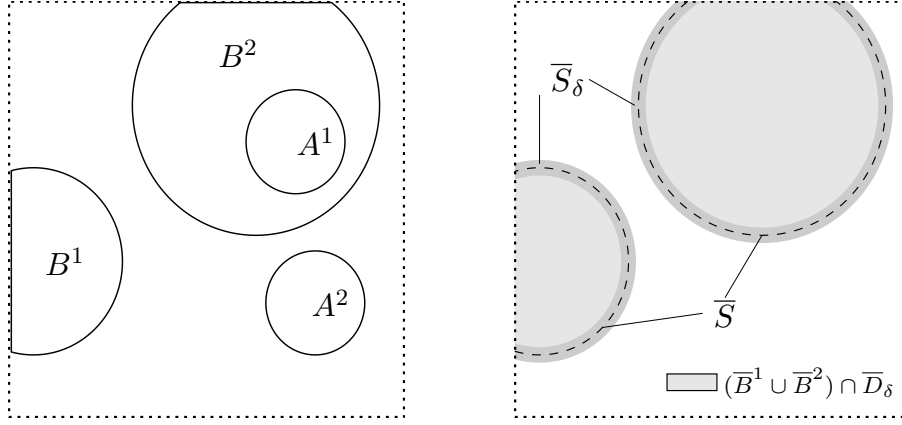


Figure 5.2: Illustration of the definitions (5.29), (5.32), and (5.36): (left) shapes A^k, B^m ; (right) background $(\bar{B}^1 \cup \bar{B}^2) \cap \bar{D}_\delta$, boundary \bar{S} , and region \bar{S}_δ .

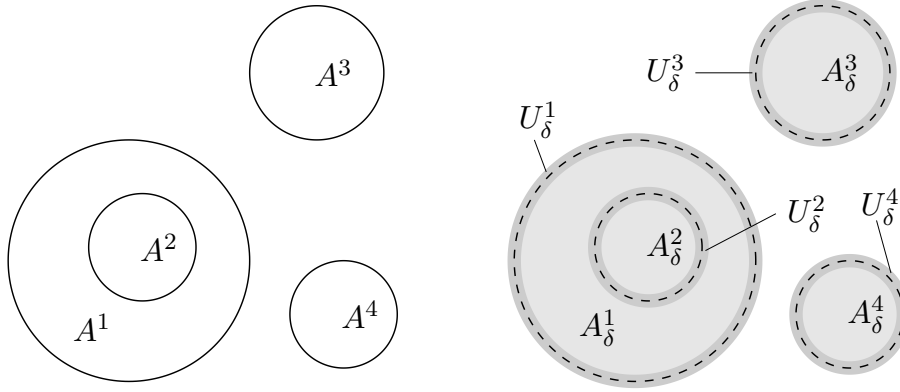


Figure 5.3: Illustration of definitions (5.34) and (5.35) with homogeneous background: (left) shapes A^k ; (right) approximated shapes A_δ^k of A^k and their neighborhood U_δ^k .

5.2.3 Statement of main results and discussion

The main analytic result, given by Theorem 13, provides estimates for the approximation φ_0 of the background u^0 and for the eigenvalues λ_k and eigenfunctions φ_k of $L_\varepsilon[u_\delta]$, $k = 1, \dots, K$; namely, there are $C_0, C > 0$ independent of $\delta, \varepsilon > 0$ such that

$$\|\nabla \varphi_0\|_{L^2(D_\delta)}^2 \leq C_0 \varepsilon, \quad (5.38a)$$

$$\|\nabla \varphi_\ell\|_{L^2(D_\delta)}^2 \leq C \varepsilon. \quad (5.38b)$$

Clearly, letting ε tends to zero, it follows for a fixed $\delta > 0$ that both the background φ_0 and the eigenfunction φ_ℓ are constant a.e. in D_δ . From the main theorem follows Corollary 4, which provides similar estimates for finite element formulations and for u_δ obtained as the convolution of u and the standard mollifier [57, 56].

Estimate (5.38a) for the background φ_0 in Theorem 13 directly follows from Lemma 5. However, the estimate (5.38b) for the eigenfunctions φ_k relies on Proposition 6 and Lemma 7, which provide the estimate of the eigenvalue λ_k and eigenfunction φ_k , $k = 1, \dots, K$, with

$$\lambda_k \leq \|\nabla \chi_k\|_{L^2(U_\delta)}^2, \quad \|\nabla \varphi_k\|_{L^2(D_\delta)}^2 \leq \lambda_k \varepsilon.$$

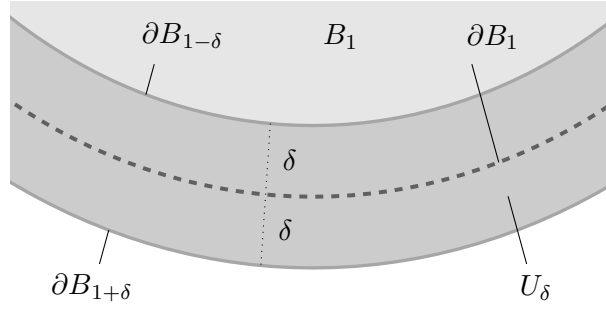


Figure 5.4: Illustration of Minkowski content (5.42) of B_1 and $U_\delta = B_{1+\delta} \setminus \overline{B_{1-\delta}}$.

Here we require for our estimation that the approximations $\{\chi_\delta^k\}_\delta$ and $\{\chi_\delta^{0,m}\}_\delta$ of each characteristic function $\{\chi^k\}$ and $\{\chi^{0,m}\}$, respectively, satisfy the estimates

$$\|\nabla \chi_\delta^{0,m}\|_{L^1(\Omega)} \leq C_0, \quad m = 1, \dots, M, \quad (5.39a)$$

$$\|\nabla \chi_\delta^k\|_{L^1(\Omega)} \leq C, \quad k = 1, \dots, K, \quad (5.39b)$$

for all δ sufficiently small and $C_0, C > 0$ independent of ε and δ .

Note that since χ_δ converges to a function χ with jump discontinuities, the gradients of χ_δ need not be bounded (in any norm) with respect to $\delta > 0$, for δ in a neighborhood of zero. Thus, whether estimate (5.39) is satisfied or not depends on properties of the method by which χ_δ is obtained, as well as on properties of the set A . The following lemma provides sufficient conditions for (5.39).

Lemma 4. *Let $A \subset \mathbb{R}^d$ be a bounded Lipschitz domain and $U_\delta = \{x \in \mathbb{R}^d \mid \text{dist}(x, \partial A) < \delta\}$ with $\delta \in (0, \eta]$, for $\eta > 0$. The following assertions hold.*

- (i) *There exists a constant $C > 0$ such that $\delta^{-1} \mathcal{L}(U_\delta) < C$, for every $\delta \in (0, \eta]$.*
- (ii) *If $\{g_\delta\}_{\delta \in (0, \eta]}$ is a family of functions such that $\text{supp}(g_\delta) \subset \subset \overline{U_\delta}$, for all $\delta \in (0, \eta]$ and some $p \in [1, \infty]$,*

$$\delta^{1-1/p} \|g_\delta\|_{L^p(\mathbb{R}^d)} \leq C_1 \quad (5.40)$$

(with the usual convention $1/\infty := 0$), then there exists a constant $C > 0$, such that for every $\delta \in (0, \eta]$,

$$\|g_\delta\|_{L^1(\mathbb{R}^d)} \leq C. \quad (5.41)$$

Remark 11. (i) *Note that in particular, for $p = \infty$, (5.40) reduces to $\delta |g_\delta| \leq C_1$ a.e. in \mathbb{R}^d for all $\delta \in (0, \eta]$. Also note that for $p = 1$, the conclusion of the lemma is trivial.*

- (ii) *For the proof of the estimate $\delta^{-1} \mathcal{L}(U_\delta) < C$ in Lemma (i), we need the theorem of the d -dimensional Minkowski content (see [59, Theorem 3.2.39] and [127, §4.1]), namely*

$$\lim_{\delta \searrow 0} \frac{\mathcal{L}(U_\delta)}{2\delta} = \mathcal{H}^{d-1}(\partial A). \quad (5.42)$$

To illustrate the Minkowski content (see Figure 5.4), we consider the unit disk $A = B_1$, where $B_r = B_r(0) = \{x \in \mathbb{R}^2 \mid \|x\| < r\}$. Then, by definition, we have $U_\delta = B_{1+\delta} \setminus \overline{B_{1-\delta}}$, $\delta > 0$, and

$$\frac{\mathcal{L}(U_\delta)}{2\delta} = \frac{(1+\delta)^2 \pi - (1-\delta)^2 \pi}{2\delta} = 2\pi = \mathcal{H}^1(\partial A), \quad \forall \delta > 0.$$

Proof of Lemma 4. (i) We first show that $\psi : [0, \eta] \rightarrow \mathbb{R}$, given by

$$\psi(\delta) = \begin{cases} \delta^{-1} \mathcal{L}(U_\delta) & \delta \neq 0, \\ 2\mathcal{H}^{d-1}(\partial A) & \delta = 0, \end{cases} \quad (5.43)$$

is bounded or, more precisely, ψ is continuous in $[0, \eta]$. Since A is a bounded Lipschitz domain, its boundary ∂A is $(d-1)$ -rectifiable. In particular, there exists a Lipschitz mapping from a bounded subset of \mathbb{R}^{d-1} onto ∂A . Hence the $(d-1)$ -dimensional Minkowski content and the $(d-1)$ -dimensional Hausdorff measure of ∂A coincide. Thus, we have $\lim_{\delta \searrow 0} \psi(\delta) = \psi(0)$, i.e., ψ is continuous at $\delta = 0$. Since $\rho(x) = \text{dist}(x, \partial A)$ is Lipschitz and satisfies $|\nabla \rho(x)| = 1$ a.e. $x \in U_\delta$ [45, §6], we have

$$\mathcal{L}(U_\delta) = \int_0^\delta \mathcal{H}^{d-1}(\rho^{-1}\{t\}) dt, \quad (5.44)$$

by the coarea formula in [57, §3.4]. Combining this with (5.42), we conclude that the function $\delta \mapsto \mathcal{L}(U_\delta)$ is absolutely continuous in $[0, \eta]$. It follows that ψ is continuous in $(0, \eta]$. Since it is also continuous at $\delta = 0$, we have that it is continuous in the entire closed interval $[0, \eta]$. Therefore, ψ is bounded in $[0, \eta]$, which proves (i).

(ii) The Hölder inequality, together with $\text{supp}(g_\delta) \subset \subset \overline{U_\delta}$, implies that

$$\|g_\delta\|_{L^1(\mathbb{R}^d)} = \int_{U_\delta} |g_\delta(x)| dx \leq \left(\mathcal{L}(U_\delta) \right)^{1-1/p} \|g_\delta\|_{L^p(\mathbb{R}^d)}. \quad (5.45)$$

Next, Equation (5.40) yields

$$\|g_\delta\|_{L^1(\mathbb{R}^d)} \leq C_1 \left(\delta^{-1} \mathcal{L}(U_\delta) \right)^{1-1/p}. \quad (5.46)$$

Thus the conclusion follows from assertion (i) of this lemma. \square

In the following results we suppose $\eta > 0$ is such that the sets A_η^k are nonempty, there holds

$$\overline{S_\eta} \cap \overline{U_\eta} = \emptyset, \quad \overline{U_\eta^k} \cap \overline{U_\eta^j} = \emptyset, \quad \forall k \neq j, \quad (5.47)$$

and for each connected component E_η of $D_\eta \setminus A_\eta$, there holds $\mathcal{H}^{d-1}(\partial E_\eta \cap \partial \Omega) > 0$. Since the boundaries of A^1, \dots, A^K are mutually disjoint and $A^k \subset \subset \Omega \setminus S$, $k = 1, \dots, K$, such an $\eta > 0$ exists. We further suppose for each $m = 1, \dots, M$,

$$\nabla \chi_\delta^{0,m} \in L^\infty(\Omega), \quad \text{supp}(\nabla \chi_\delta^{0,m}) \subset \subset \overline{S_\delta}, \quad \forall \delta \in (0, \eta], \quad (5.48)$$

and for each $k = 1, \dots, K$,

$$\nabla \chi_\delta^k \in L^\infty(\Omega), \quad \text{supp}(\nabla \chi_\delta^k) \subset \subset \overline{U_\delta^k}, \quad \forall \delta \in (0, \eta]. \quad (5.49)$$

Following these assumptions and (5.21), we obtain

$$\mu_\varepsilon[u_\delta] = \hat{\mu}_\varepsilon(0) = \varepsilon^{-1} \quad (5.50)$$

a.e. in D_δ for each $\delta \in (0, \eta]$, $\nabla u_\delta \in L^\infty(\Omega)$, and thus, for all $\varepsilon > 0$, (5.22) with $w = u_\delta$ is satisfied. In particular, the operator $L_\varepsilon[u_\delta]$ is uniformly elliptic in Ω .

Note that $\nabla \chi_\delta^k$ fulfills (5.49), where χ_δ^k is an (H^1 -conforming) FE-approximation of χ^k with

$$\chi_\delta^k = \begin{cases} 1, & \text{in } A_\delta^k, \\ 0, & \text{in } \Omega \setminus (A_\delta^k \cup U_\delta^k). \end{cases}$$

For the approximation φ_0 of the background u^0 , we have the following:

Lemma 5. *For every $\varepsilon > 0$ and $\delta \in (0, \eta]$ there holds*

$$\|\nabla \varphi_0\|_{L^2(D_\delta)}^2 \leq \varepsilon \|\nabla u_\delta^0\|_{L^1(\Omega)}. \quad (5.51)$$

Here we need the following general Lemma.

Lemma 6. *Let V be a Hilbert space and $B : V \times V \rightarrow \mathbb{R}$ a symmetric, coercive and continuous bilinear form.*

(i) *If $F : V \rightarrow \mathbb{R}$ is a continuous linear functional, then there exists a unique $u \in V$ such that*

$$B[u, v] = F[v] \quad \forall v \in V. \quad (5.52)$$

Furthermore, u is the unique minimizer in V of the functional G given by

$$G(v) = \frac{1}{2}B[v, v] - F[v]. \quad (5.53)$$

(ii) *Suppose that V is a subspace of \tilde{V} , $B : \tilde{V} \times \tilde{V} \rightarrow \mathbb{R}$ is a continuous extension of B , and $w \in \tilde{V}$. Then there exists a unique $u \in w + V$ such that*

$$B[u, v] = 0 \quad \forall v \in V. \quad (5.54)$$

Moreover, u is the unique minimizer of the functional I , given by $I(v) = B[v, v]$, in $w + V$, namely

$$B[u, u] \leq B[w + v, w + v] \quad \forall v \in V. \quad (5.55)$$

Proof. (i) According to the Theorem of Lax-Milgram there exists a unique solution $u \in V$ of (5.52) [56, §6.2]. Now, let $G : V \rightarrow \mathbb{R}$ defined by (5.53). Then for $u \neq v \in V$ it holds

$$\begin{aligned} G(v) &= \frac{1}{2}B[v, v] - F[v] = \frac{1}{2}B[v, v] - B[u, v] \\ &= \frac{1}{2}(B[v, v] - 2B[u, v] + B[u, u]) - \frac{1}{2}B[u, u] \\ &= \frac{1}{2}B[v - u, v - u] - \frac{1}{2}B[u, u] \\ &> -\frac{1}{2}B[u, u] = \frac{1}{2}B[u, u] - B[u, u] = \frac{1}{2}B[u, u] - F[u] = G(u). \end{aligned}$$

(ii) For $u \in w + V$, let $\hat{u} = u - w$. Then $\hat{u} \in V$ and u satisfies (5.54) if, and only if,

$$B[\hat{u}, v] = B[u - w, v] = B[u, v] - B[w, v] = -B[w, v], \quad \forall v \in V. \quad (5.56)$$

Since $\hat{F}[v] = -B[w, v]$ is continuous for a fixed w , there exists a unique $\hat{u} \in V$ satisfying (5.56). Moreover, (ii) implies that, for any $v \in V$, it holds

$$\frac{1}{2}B[\hat{u}, \hat{u}] + B[w, \hat{u}] = \frac{1}{2}B[\hat{u}, \hat{u}] - \hat{F}[\hat{u}] \leq \frac{1}{2}B[v, v] - \hat{F}[v] = \frac{1}{2}B[v, v] + B[w, v]. \quad (5.57)$$

Let $u = \hat{u} + w$. Then we have $u \in w + V$ and thus we recover (5.55) from (5.57), namely

$$\begin{aligned} B[u, u] &= B[\hat{u}, \hat{u}] - B[w, w] + 2B[u, w] \\ &\leq B[v, v] + 2B[w, v] - 2B[w, \hat{u}] - B[w, w] + 2B[u, w] \\ &= B[w + v, w + v] + 2B[w, u - \hat{u} - w] = B[w + v, w + v]. \end{aligned}$$

□

Proof of Lemma 5. Following Lemma 6, the solution φ_0 of (5.24) is the unique minimizer of the functional $I(v) = B[v, v]$ in $u_\delta + \mathcal{V}_0^\delta$, i.e.,

$$B[\varphi_0, \varphi_0] \leq B[u_\delta + v, u_\delta + v], \quad \forall v \in \mathcal{V}_0^\delta,$$

or equivalently

$$B[\varphi_0, \varphi_0] \leq B[u_\delta^0 + v, u_\delta^0 + v], \quad \forall v \in \mathcal{V}_0^\delta. \quad (5.58)$$

By the last estimate with $v = 0$ and the equality $u_\delta = u_\delta^0$, which holds in S_δ , we obtain

$$B[\varphi_0, \varphi_0] \leq B[u_\delta^0, u_\delta^0] = \int_{S_\delta} \mu_\varepsilon[u_\delta^0] |\nabla u_\delta^0|^2 dx \leq \|\nabla u_\delta^0\|_{L^1(S_\delta)}, \quad (5.59)$$

by (5.21) with $\tau = |\nabla u_\delta|$, and hence the conclusion. □

The estimates of the eigenfunctions are due to the following simple proposition:

Proposition 6. *For every $j \in \mathbb{N}$ there holds*

$$\|\nabla \varphi_j\|_{L^2(D_\delta)}^2 \leq \lambda_j \varepsilon, \quad \lambda_j = \lambda_j(\varepsilon, \delta). \quad (5.60)$$

Proof. The proposition follows from (5.26) and the definition of $B[\cdot, \cdot]$. □

Thus, to show that an eigenfunction φ_k is “almost” piecewise constant, we require an estimate of the eigenvalue λ_k of $L_\varepsilon[u_\delta]$ in \mathcal{V}_0^δ .

Lemma 7. *There exists a constant C , independent of u^1, \dots, u^K , such that for every $\varepsilon > 0$, $\delta \in (0, \eta]$, and $k = 1, \dots, K$, there holds*

$$\lambda_k \leq \frac{C \|\tau\|_2}{\min_j |\hat{u}^j|}, \quad \tau_j = \|\nabla \chi_\delta^j\|_{L^1(\Omega)}, \quad j = 1, \dots, K. \quad (5.61)$$

The proof of Lemma 7 requires the following result.

Lemma 8. *If the sets A_η^k , $k = 1, \dots, K$, are nonempty and (5.47) is satisfied for some $\eta > 0$, then the following assertions hold true:*

(i) *There exists $j = 1, \dots, K$ such that*

$$\mathcal{L}(\mathcal{A}_\eta^j \setminus \bigcup_{j \neq k=1}^K \overline{\mathcal{A}_\eta^k}) > 0, \quad \mathcal{A}_\eta^k = A_\eta^k \cap D_\eta. \quad (5.62)$$

(ii) *The restrictions of the functions χ^k , $k = 1, \dots, K$, to D_η are linearly independent.*

(iii) *If for every $k = 1, \dots, K$, (5.49) is satisfied, then there exists a constant $C > 0$ such that for every $\delta \in [0, \eta]$ and $\widehat{\psi} = (\psi^k)_{k=1}^K \in \mathbb{R}^K$, the following estimates are satisfied:*

$$\|\psi_\delta\|_{L^2(\Omega)}^2 \geq \|\psi\|_{L^2(D_\eta)}^2 \geq C \|\widehat{\psi}\|_2^2, \quad (5.63)$$

where $\|\cdot\|_2$ denotes the Euclidean norm in \mathbb{R}^K , $\psi = \sum_k \psi^k \chi^k$ and $\psi_\delta = \sum_k \psi^k \chi_\delta^k$.

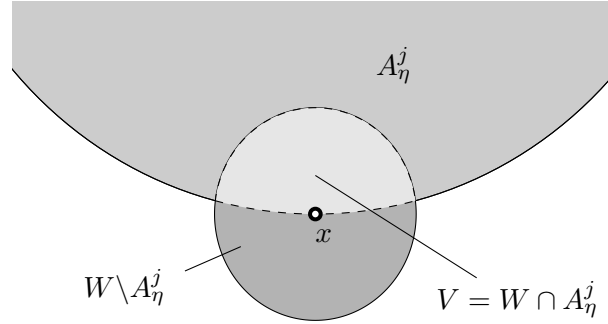


Figure 5.5: Illustration of assertion (i) in Lemma 8.

Proof of Lemma 8. (i) Let $x \in \partial A_\eta$. Then, by the definition of ∂A_η , for each neighborhood W of x , it holds $W \cap A_\eta \neq \emptyset$ and $W \cap (\Omega \setminus A_\eta) \neq \emptyset$. Since

$$W \cap A_\eta = \bigcup_{k=1}^K (W \cap A_\eta^k), \quad W \cap (\Omega \setminus A_\eta) = \bigcap_{k=1}^K (W \cap (\Omega \setminus A_\eta^k)), \quad (5.64)$$

we obtain $x \in W \cap (\overline{\Omega \setminus A_\eta^j})$, $\forall j$, and there is $j = 1, \dots, K$ with $x \in W \cap \overline{A_\eta^j}$. As a consequence, we have $x \in \partial A_\eta^j$. Note that $\partial A_\eta^j \subset \subset \partial U_\eta^j$. As the sets $\overline{U_\eta^k}$ with $k = 1, \dots, K$ are mutually disjoint, for each $k \neq j$, there holds $x \notin \partial A_\eta^k$. Similarly, we have $x \notin \partial S_\eta$. Hence, there exists an (sufficiently small) open set W such that $x \in W$, and $W \cap (\overline{A_\eta^k \cup U_\eta^k}) = \emptyset$, for all $k \neq j$, and $W \cap \overline{S_\eta} = \emptyset$. However, since $x \in \partial A_\eta^j$, the intersection $W \cap A_\eta^j$ is nonempty. Let $V = W \cap A_\eta^j$ shown Figure 5.5. By the above, we have $V \cap \overline{U_\eta^k} = \emptyset$, for all $k = 1, \dots, K$ and $V \cap \overline{S_\eta} = \emptyset$. Consequently, V is a nonempty open set satisfying $V \subset A_\eta^j \cap D_\eta = \mathcal{A}_\eta^j$ and $V \cap \overline{\mathcal{A}_\eta^k} \subset V \cap \overline{A_\eta^k} = \emptyset$, for $k \neq j$. Thus, we have $\mathcal{L}(V) > 0$ which proves (i).

(ii) Suppose

$$\sum_{k=1}^K \psi^k \chi^k = 0 \quad \text{a.e. in } D_\eta. \quad (5.65)$$

From (i) there is j satisfying (5.62). Without loss of generality, we assume that $j = K$. However, by (5.65) this can be true, only if $\psi^K = 0$. The above argument may be repeated by induction. Since there is a finite number of functions χ^k , the procedure stops only when there is only one function left; let it be χ^1 . Then (5.65) reduces to $\psi^1 \chi^1 = 0$. However, since $A_\eta^1 \cap D_\eta$ is open and nonempty (by the same argument), this implies that $\psi^1 = 0$, and we have that the restrictions of χ^k to D_η are linearly independent.

(iii) Let $\widehat{\psi} = (\psi^k) \in \mathbb{R}^K$, $\delta \in [0, \eta]$, $\psi = \sum_k \psi^k \chi^k$, and $\psi_\delta = \sum_k \psi^k \chi_\delta^k$, where $\chi_0^k = \chi^k$. As each χ_δ^k coincides with χ^k in $D_\delta = \Omega \setminus U_\delta$, we have

$$\|\psi_\delta\|_{L^2(\Omega)}^2 \geq \int_{D_\delta} \left| \sum_{k=1}^K \psi^k \chi_\delta^k \right|^2 = \int_{D_\delta} \left| \sum_{k=1}^K \psi^k \chi^k \right|^2 = \|\psi\|_{L^2(D_\delta)}^2 \geq \|\psi\|_{L^2(D_\eta)}^2, \quad (5.66)$$

for $\delta < \eta$, since $D_\eta \subset \subset D_\delta$. Expanding the term on the right hand side yields

$$\|\psi\|_{L^2(D_\eta)}^2 = \sum_{k,j=1}^K \psi^k \psi^j \int_{D_\eta} \chi^k \chi^j = \sum_{k,j=1}^K M_{kj} \psi^k \psi^j = \|\widehat{\psi}\|_{\mathbf{M}}^2, \quad (5.67)$$

where $\|v\|_{\mathbf{M}}^2 = v^\top \mathbf{M} v$, $v \in \mathbb{R}^K$, and

$$M_{kj} = \int_{D_\eta} \chi^k \chi^j = \mathcal{L}(A^k \cap A^j \cap D_\eta) > 0.$$

Hence $\|\psi\|_{L^2(D_\eta)}^2$ is a quadratic form in the vector $\widehat{\psi}$, represented by the symmetric matrix $\mathbf{M} = (M_{kj})$, and is positive for any $\widehat{\psi} \neq 0$, by assertion (ii) of this lemma. It follows that \mathbf{M} is symmetric positive definite, and thus $\|\cdot\|_{\mathbf{M}}$ is a norm in \mathbb{R}^K . As a consequence, it yields (5.63), together with (5.66) and (5.67), and hence the conclusion. \square

Proof of Lemma 7. It follows from spectral theory for symmetric elliptic operators and the Courant minimax principle [56, §6.5 – 6.6] that for each n ,

$$\lambda_n = \min_{0 \neq v \in \Phi_{n-1}^\perp} \frac{B[v, v]}{\|v\|^2}, \quad (5.68)$$

where $\Phi_{n-1} = \text{span}\{\varphi_j\}_{j=1}^{n-1}$, and W^\perp denotes the orthogonal complement of W in \mathcal{V}_0^δ with respect to the $L^2(\Omega)$ inner product. For

$$\psi_\delta = \sum_{k=1}^K \psi^k \chi_\delta^k, \quad \psi^k \in \mathbb{R}, \quad (5.69)$$

there holds

$$\langle \varphi_j, \psi_\delta \rangle = \sum_{k=1}^K \psi^k \langle \varphi_j, \chi_\delta^k \rangle. \quad (5.70)$$

Let $n \leq K$ and $\widehat{\psi} = (\psi^k)_{k=1}^K \in \mathbb{R}^K$ with $\|\widehat{\psi}\|_2 = 1$ such that ψ_δ given by (5.69) satisfies

$$\langle \varphi_j, \psi_\delta \rangle = 0, \quad j = 1, \dots, n-1. \quad (5.71)$$

One can find such a vector of coefficients $\widehat{\psi}$, since $n \leq K$ and hence the homogeneous linear system (5.71) has more variables than equations. By Lemma 8 there exists a constant $C_1 > 0$ independent of $\|\widehat{\psi}\|_2 = 1$, $\delta \in (0, \eta]$, or ε such that $\|\psi_\delta\|^2 \geq C_1$. Consequently, $0 \neq \psi_\delta \in \Phi_{n-1}^\perp$, and thus we recover

$$\lambda_n \leq \frac{B[\psi_\delta, \psi_\delta]}{\|\psi_\delta\|^2} \leq C_1^{-1} B[\psi_\delta, \psi_\delta]. \quad (5.72)$$

We write

$$B[\psi_\delta, \psi_\delta] = \int_{U_\delta} \mu_\varepsilon[u_\delta] |\nabla \psi_\delta|^2 dx. \quad (5.73)$$

Next, we show

$$|\nabla \psi_\delta(x)| < \frac{|\nabla u_\delta(x)|}{\min_j |u^j|}, \quad \text{a.e. } x \in U_\delta. \quad (5.74)$$

Let $m = \min_j |u^j|$. Since the sets $U_\delta^1, \dots, U_\delta^K$ are mutually disjoint, each $x \in U_\delta$ lies in precisely one U_δ^k , $k = 1, \dots, K$, and hence at most one of the gradients $\nabla \chi_\delta^k$ is nonzero. Therefore, in U_δ we have

$$|\nabla \psi_\delta| = \left| \sum_{k=1}^K \psi^k \nabla \chi_\delta^k \right| = \sum_{k=1}^K |\psi^k| |\nabla \chi_\delta^k| \leq \frac{1}{m} \sum_{k=1}^K |u^k| |\nabla \chi_\delta^k| = \frac{1}{m} \left| \sum_{k=1}^K u^k \nabla \chi_\delta^k \right|.$$

Since $S_\delta \cap U_\delta = \emptyset$, we have $\nabla u_\delta^0 = 0$ a.e. in U_δ and hence (5.74) is satisfied. Thus we obtain

$$B[\psi_\delta, \psi_\delta] \leq \frac{1}{m} \int_{U_\delta} |\nabla \psi_\delta| dx = \frac{1}{m} \|\nabla \psi_\delta\|_{L^1(\Omega)}. \quad (5.75)$$

It follows, together with Cauchy-Schwarz inequality, that

$$B[\psi_\delta, \psi_\delta] \leq \frac{1}{\min_j |u^j|} \sum_{k=1}^K |\psi^k| \tau_k \leq \frac{\|\tau\|_2}{\min_j |u^j|}, \quad (5.76)$$

which, combined with (5.72), yields the conclusion. \square

We summarize the results in the following theorem:

Theorem 13. *If A^1, \dots, A^K and B^1, \dots, B^M have Lipschitz boundaries and there exists a constant C such that for every $\delta \in (0, \eta]$, each of the functions $\chi_\delta = \chi_\delta^{0,m}$, $m = 1, \dots, M$, and $\chi_\delta = \chi_\delta^k$, $k = 1, \dots, K$, satisfies*

$$\delta \|\nabla \chi_\delta\|_{L^\infty(\Omega)} \leq C. \quad (5.77)$$

Then there exist constants C_0, C_1 , independent of the coefficients $\hat{u}^{0,1}, \dots, \hat{u}^{0,M}$ and $\hat{u}^1, \dots, \hat{u}^K$ such that for every $\varepsilon > 0$, and $\delta \in (0, \eta]$ the following estimates hold

$$\|\nabla \varphi_0\|_{L^2(D_\delta)}^2 \leq C_0 \max_m |\hat{u}^{0,m}| \varepsilon, \quad (5.78)$$

$$\lambda_k \leq \frac{C_1}{\min_j |\hat{u}^j|}, \quad \|\nabla \varphi_k\|_{L^2(D_\delta)}^2 \leq \frac{C_1}{\min_j |\hat{u}^j|} \varepsilon, \quad k = 1, \dots, K. \quad (5.79)$$

Moreover, for each Lipschitz domain $V \subset\subset D_\eta \setminus A_\eta$ with $\mathcal{H}^{d-1}(\partial V \cap \partial \Omega) > 0$, there exists a constant C_2 , independent of the coefficients $\hat{u}^{0,1}, \dots, \hat{u}^{0,M}$, and $\hat{u}^1, \dots, \hat{u}^K$, such that for every $\varepsilon > 0$, and $\delta \in (0, \eta]$ the following estimates are satisfied

$$\|u^0 - \varphi_0\|_{L^2(V)}^2 \leq C_2 \max_m |\hat{u}^{0,m}| \varepsilon, \quad (5.80)$$

$$\|\varphi_k\|_{L^2(V)}^2 \leq \frac{C_2}{\min_j |\hat{u}^j|} \varepsilon, \quad k = 1, \dots, K. \quad (5.81)$$

Proof. Estimates (5.78) and (5.79) follow easily from Lemma 4 with $p = \infty$, Lemmata 5 and 7, and Proposition 6. Estimates (5.80) and (5.81) follow from the Poincaré inequality in Section 2.1 and estimates (5.78) and (5.79), respectively, since $\mathcal{H}^{d-1}(\partial V \cap \partial \Omega) > 0$. \square

Remark 12. *For a fixed $\delta > 0$ and a connected open Lipschitz domain $V \subset\subset D_\delta$, the estimate (5.79) from Theorem 13 and Poincaré inequality in Section 2.1 provide constants $C_V, C > 0$, independent of δ , such that*

$$\left\| \varphi_k - \frac{1}{\mathcal{L}(V)} \int_V \varphi_k dx \right\|_{L^2(V)}^2 \leq C_V \|\nabla \varphi_k\|_{L^2(V)}^2 \leq C \varepsilon, \quad \forall \varepsilon > 0.$$

Essentially, estimates (5.78) and (5.79) imply that φ_0 and the first K eigenfunctions φ_k of $L_\varepsilon[u_\delta]$ are “almost” constant in each connected component of D_δ . In particular, φ_0 is almost constant in each connected component of $D_\delta \setminus A_\delta$. Since the Hausdorff measure of $\partial \Omega \cap \partial B^m$ is positive, for each $m = 1, \dots, M$, and φ_0 coincides with u_δ^0 on $\partial \Omega$, we have that φ_0 approximates u_δ^0 well in $D_\delta \setminus A_\delta$. Similarly, for $k = 1, \dots, K$, φ_k is almost constant in each connected component of $D_\delta \setminus A_\delta$, vanishes on $\partial \Omega$, and is, therefore, small in $D_\delta \setminus A_\delta$.

Consequently, u_δ can be well approximated in $\varphi_0 + \Phi_K$, for example, by its L^2 best approximation in $\varphi_0 + \Phi_K$,

$$\varphi_0 + P_K[u_\delta - \varphi_0], \quad (5.82)$$

where P_K is given by

$$\langle v - P_K[v], \varphi \rangle = 0, \quad \forall (v, \varphi) \in \mathcal{V}^\delta \times \Phi_K. \quad (5.83)$$

From Theorem 13 we may deduce estimates (5.78)–(5.81) for specific methods of approximation.

Corollary 4. *Let $u \in L^\infty(\Omega)$. Suppose that A^1, \dots, A^K and B^1, \dots, B^M have Lipschitz boundaries. Estimates (5.78)–(5.81) hold true in each of the following cases:*

- (i) *For each $\delta \in (0, \eta]$, u_δ is the convolution of u with the standard mollifier (e.g., [57, 56]), and $\mathcal{V}^\delta = H^1(\Omega)$.*
- (ii) *For each $\delta \in (0, \eta]$, u_δ is the FE projection of u into an H^1 -conforming FE space V_δ associated with a mesh \mathcal{T}_δ , where the family of meshes $\{\mathcal{T}_\delta\}_{\delta \in (0, \eta]}$ is regular and quasi-uniform (see, e.g., [24, 122]), and $\mathcal{V}^\delta = V_\delta$ or $\mathcal{V}^\delta = H^1(\Omega)$.*

Proof. (i) Here we verify that the convolution, $u_\delta = \rho_\delta * u$, $\delta > 0$, of $u \in W^{1,\infty}(\Omega)$, satisfies (5.77), where $\rho_\delta(x) = \delta^{-d} \rho(x/\delta)$ and $\rho \in C_c^\infty(\mathbb{R}^d)$ is the standard mollifier,

$$\rho(x) = \begin{cases} C \exp\left(\frac{1}{|x|^2-1}\right), & |x| < 1, \\ 0, & |x| \geq 1, \end{cases}, \quad C = \left(\int_{\{|x|<1\}} \exp\left(\frac{1}{|x|^2-1}\right) dx \right)^{-1}.$$

By definitions and integration by substitution, we have

$$\begin{aligned} \delta \left| \frac{\partial}{\partial x_j} u_\delta(x) \right| &\leq \delta \int_{\mathbb{R}^d} |u(y)| \left| \frac{\partial}{\partial x_j} \rho_\delta(y-x) \right| dy \\ &\leq \|u\|_{L^\infty(\Omega)} \int_{B_\delta(0)} \frac{1}{\delta^d} \left| \left(\frac{\partial}{\partial x_j} \rho \right) \left(\frac{y}{\delta} \right) \right| dy \leq \|u\|_{L^\infty(\Omega)} \int_{B_1(0)} \left| \frac{\partial}{\partial x_j} \rho(\hat{y}) \right| d\hat{y}, \end{aligned}$$

which yields

$$\delta \left| \nabla u_\delta(x) \right| \leq \|u\|_{L^\infty(\Omega)} \|\nabla \rho\|_{L^1(B_1(0))}, \quad x \in \Omega,$$

and hence, together with Theorem 13, the first assertion.

(ii) A proof for two-dimensional computational domains can be found in [24, §6.6] and in [122, §4.5]. Since the proof for two-dimensional domains is quite general, one can reuse and extend the proof for three-dimensional computational domains. \square

5.2.4 Numerical example

Here we consider the numerical example that shows the usefulness of the adaptive spectral (AS) space to verify the property shown in Corollary 4. We start with a piecewise constant function u given by

$$u = u^0 + \tilde{u}, \quad u^0(x) = \sum_{m=1}^M \hat{u}^{0,m} \chi^{0,m}(x), \quad \tilde{u}(x) = \sum_{k=1}^K \hat{u}^k \chi^k(x),$$

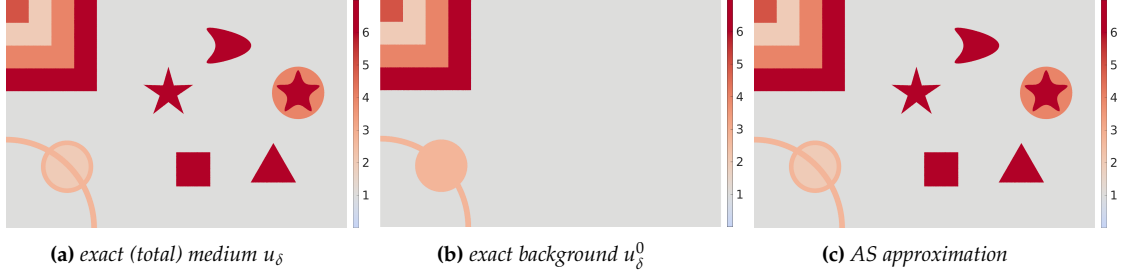


Figure 5.6: Adaptive spectral approximation of a piecewise constant medium: (a) exact total medium u_δ in FE space; (b) exact background u_δ^0 in FE space; (c) L^2 -orthogonal projection $P_8[u_\delta]$, given by (5.83), of u_δ in AS space with eight eigenfunctions.

for some coefficients $\hat{u}^{0,m}$ and \hat{u}^k , $m = 1, \dots, M$, $k = 1, \dots, K$, with $M = 5$ and $K = 8$. Here $\chi^{0,m}$ and χ^k denote the characteristic function of the shape B^m and A^k , respectively, where B^m touches the boundary $\partial\Omega$ in sense of $\mathcal{H}^{d-1}(\bar{B}^m \cap \partial\Omega) > 0$ and $A^k \subset\subset \Omega$.

First, we approximate u with \mathcal{P}^1 -FE in a regular and quasi-uniform triangular mesh denoted by $u_\delta \in W^{1,\infty}(\Omega)$, which is displayed in Figure 5.6a. Next, we compute the spectral space V_N in (5.26) with $w = u_\delta$, $N = 8$, and $\varepsilon = 10^{-8}$, and project u_δ into $v_N = P_N[u_\delta] \in V_N$, where P_N is given by (5.83). The reconstruction v_N of u_δ matches remarkably well with a relative L^2 -error of 0.15% in V_N – see Figure 5.6c.

In Figure 5.7a, we monitor the numerical background φ_0 of (5.24). Moreover, Figures 5.7b – 5.7i illustrate the eigenfunctions $\varphi_1, \dots, \varphi_8$ of (5.26) with the (smallest) eigenvalues $\lambda_1 \leq \dots \leq \lambda_8$. Here we observe that φ_0 is “almost” identical to the (exact) background u^0 . In addition, each eigenfunction φ_k accurately matches a characteristic function χ^k , $k = 1, \dots, 8$, except for φ_1 and φ_6 . Since the corresponding shapes A^1 and A^6 (after reordering) are overlapped to each other, φ_1 and φ_6 both instead are a linear combination of χ^1 and χ^6 . Moreover, the background precisely vanishes in every eigenfunction φ_k . Consequently, each eigenfunction corresponds very closely to a linear combination of the characteristic functions. This fulfills the property of the AS space shown in Corollary 4.

5.3 Adaptive spectral inversion

Here we present the adaptive spectral inversion (ASI) method and some numerical examples that illustrate the usefulness of the ASI method. First, we consider the inverse problem in Section 5.1 formulated as a PDE-constrained optimization problem to determine the unknown control u and provide a detailed description of the ASI algorithm. We then apply the adaptive spectral (AS) decomposition approach from Section 5.2 to regularize the ill-posed problem, where the solution u is projected to the control space consisting of a finite number of spectral basis functions. To have a small number of basis functions, we propose to use a dynamic size reduction of the control space.

Next, we apply the ASI method, combined with the space size reduction approach, to solve a two-dimensional inverse scattering problem. Finally, we show the numerical result of the scattering problem from the salt dome model.

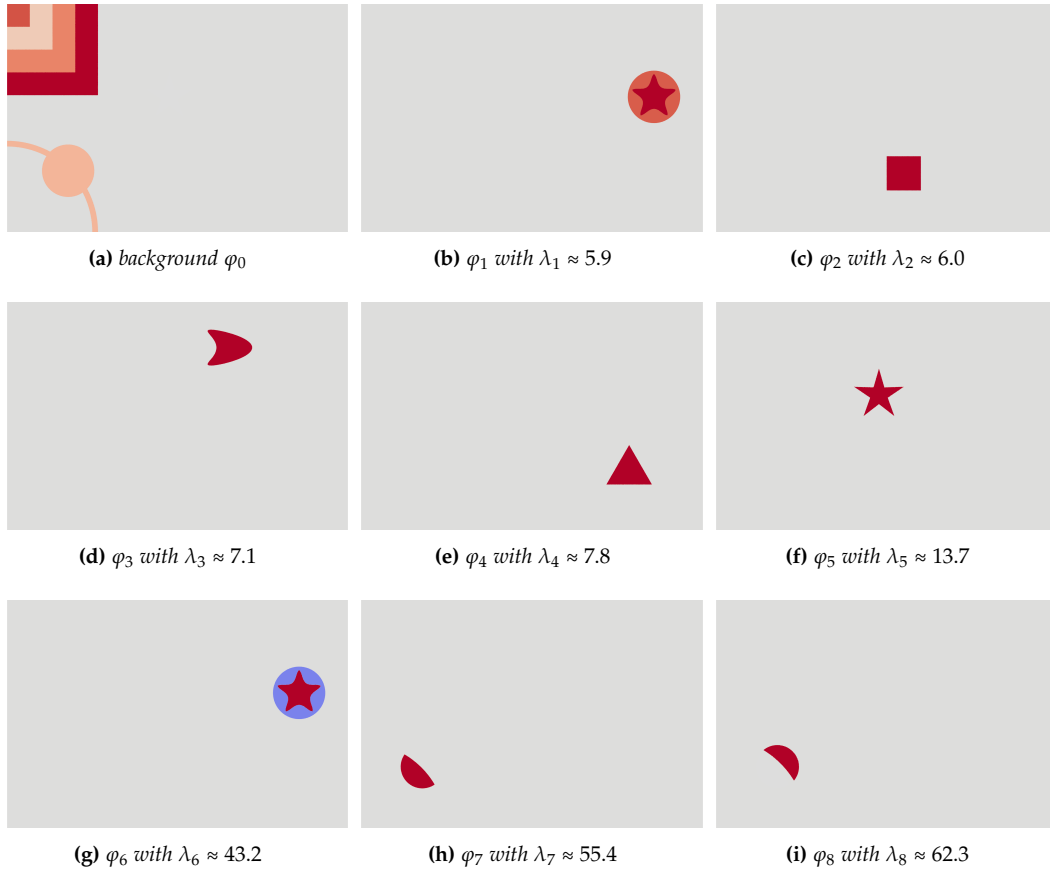


Figure 5.7: Adaptive spectral approximation of a piecewise constant medium: (a) background φ_0 in (5.24) with $w = u_\delta$ and $\varepsilon = 10^{-8}$; (b)–(i) eigenfunctions $\varphi_1, \dots, \varphi_8$ of (5.26) to the first eight eigenvalues $\lambda_1, \dots, \lambda_8$.

5.3.1 Adaptive spectral space with size reduction

Let us recall the PDE-constrained optimization problem,

$$u_* = \underset{v}{\operatorname{argmin}} J[v], \quad (5.84)$$

where J is defined in (5.3). Here we consider the representative of

$$u_* \approx \varphi_0 + \sum_{\ell=1}^K \beta_\ell \psi_\ell, \quad \beta_\ell \in \mathbb{R},$$

in the adaptive spectral (AS) space spanned by the eigenfunctions ψ_ℓ of (5.26), $\ell = 1, \dots, K$, and the background φ_0 given in (5.24). Since the exact medium u (or u_δ) is unknown, we start with an initial guess $u^{(0)}$ and compute

$$U^{(0)} = \varphi_0^{(0)} + \Phi^{(0)}, \quad \Phi^{(0)} = \operatorname{span}\{\psi_1^{(0)}, \dots, \psi_{K_0}^{(0)}\}, \quad K_0 = \dim(\Phi^{(0)}), \quad (5.85)$$

defined in (5.24) and (5.26) with $w = u^{(0)}$. We then instead minimize the discrete optimization problem (5.84),

$$\beta_*^{(0)} = \underset{\beta \in \mathbb{R}^{K_0}}{\operatorname{argmin}} J \left[\varphi_0^{(0)} + \sum_{\ell=1}^{K_0} \beta_\ell \psi_\ell^{(0)} \right],$$

and set

$$u_*^{(0)}(x) = \varphi_0^{(0)}(x) + \sum_{\ell=1}^{K_0} \beta_{*,\ell}^{(0)} \psi_\ell^{(0)}(x) \in U^{(0)}.$$

Now, the challenge is to update $U^{(m)}$, $m \geq 1$, adaptively while keeping the property of the previous AS space and in addition reducing the number of basis functions without losing the accuracy of the solution up to a given tolerance.

Let us consider the current space

$$U^{(m-1)} = \varphi_0^{(m-1)} + \Psi^{(m-1)}, \quad \Psi^{(m-1)} = \text{span}\{\psi_1^{(m-1)}, \dots, \psi_{K_{m-1}}^{(m-1)}\},$$

and compute the new AS space,

$$V^{(m)} = \varphi_0^{(m)} + \Phi^{(m)}, \quad \Phi^{(m)} = \text{span}\{\varphi_1^{(m)}, \dots, \varphi_{\tilde{K}_m}^{(m)}\}, \quad (5.86)$$

where \tilde{K}_m is the dimension of the new AS space, $\varphi_0^{(m)}$ solves (5.24), and $\varphi_\ell^{(m)}$ satisfies (5.26), $\ell = 1, \dots, \tilde{K}_m$, with $w = u_*^{(m-1)}$.

Following [77, 79], the new basis $U^{(m)} = V^{(m)}$ is formed and the old one $U^{(m-1)}$ is completely replaced by $V^{(m)}$. Instead of replacing the entire basis at each iteration, we now propose to pass relevant information in $U^{(m-1)}$ to $U^{(m)}$. Therefore, we first merge $U^{(m-1)}$ with $V^{(m)}$ and reduce then the merged space, including the previous background $\varphi_0^{(m)} - \varphi_0^{(m-1)}$,

$$\text{span}\{\varphi_0^{(m)} - \varphi_0^{(m-1)}, \psi_1^{(m-1)}, \dots, \psi_{K_{m-1}}^{(m-1)}, \varphi_1^{(m)}, \dots, \varphi_{\tilde{K}_m}^{(m)}\} \quad (5.87)$$

by using the singular value decomposition (SVD); namely, we replace the SVD of the matrix, consisting of all discrete vectors of (5.87), by its “trimmed down” version of the SVD – see [72, §2.5.4, §12.2] and [85].

Let $\psi_1^{(m)}, \dots, \psi_N^{(m)}$, for some N , denote an orthonormal basis with respect to the L^2 inner product of the resulted thin SVD space. Then, we consider the (orthonormal) projection

$$P[u^{(m-1)}] = \varphi_0^{(m)} + \sum_{\ell=1}^N \beta_\ell^{(m)} \psi_\ell^{(m)}, \quad \beta_\ell^{(m)} \in \mathbb{R}, \quad (5.88)$$

of $u_*^{(m-1)}$ into $\varphi_0^{(m)} + \text{span}\{\psi_1^{(m)}, \dots, \psi_N^{(m)}\}$, where P is given by (5.83).

To reduce the space dimension further, we remove all basis functions, where the corresponding coefficients β_ℓ in the solution are sufficiently small like

$$|\beta_\ell^{(m)}| \leq \text{ToI}_\beta, \quad \text{ToI}_\beta > 0.$$

Finally, we set after rearranging, e.g. $|\beta_1^{(m)}| \geq \dots \geq |\beta_N^{(m)}|$,

$$u^{(m)} = \varphi_0^{(m)} + \sum_{\ell=1}^{K_m} \beta_\ell^{(m)} \psi_\ell^{(m)}, \quad K_m \leq N,$$

and define

$$U^{(m)} = \varphi_0^{(m)} + \text{span}\{\psi_1^{(m)}, \dots, \psi_{K_m}^{(m)}\}. \quad (5.89)$$

Now, we list the complete ASI algorithm:

Algorithm 4: ASI algorithm

Input: initial guess $u^{(0)}$ of the medium u , initial space $U^{(0)}$ and frequency ν_1
Output: reconstruction $u^{(m)}$ of the medium

```

1 while  $m = 1, 2, \dots$  do
2   Find a (local) minimizer  $u_*^{(m)}$  of  $J$  in (5.84) starting with  $u^{(m-1)}$  in  $U^{(m-1)}$ .
3   if  $\|u_*^{(m)} - u_*^{(m-1)}\| < \text{ToI}_{ASI}$  then
4     Return  $u^{(m)} = u_*^{(m)}$  and stop.
5   Compute the new affine space  $V^{(m)}$  in (5.86).
6   Merge  $V^{(m)}$  with  $U^{(m-1)}$  given in (5.87) and
   trim the merged space with the SVD in (5.91).
7   Find an orthonormal basis  $\psi_1^{(m)}, \dots, \psi_K^{(m)}$  of the trimmed space.
8   Truncate the affine space with (5.89) to obtain the new reduced space  $U^{(m)}$ .
9   Determine the least-squares  $u^{(m)}$  of  $u_*^{(m)}$  in  $U^{(m)}$ .
10  if  $\|u^{(m)} - u^{(m-1)}\| < \varepsilon_\nu$  then
11    Increase the frequency,  $\nu_{m+1} = \nu_m + \Delta\nu$ .
12  else
13    Set  $\nu_{m+1} = \nu_m$ .
```

Remark 13. (i) The dimension of the new AS space in Step 5 given by \tilde{K}_m can be fixed for $m \geq 0$, or \tilde{K}_m increases linearly to the current frequency ν as in [77, 79].

(ii) To trim the space given by a matrix $\mathbf{B} \in \mathbb{R}^{q \times p}$, $q \geq p$, and $\text{rank}(\mathbf{B}) \geq 1$, e.g. resulting from the discrete basis function of (5.87), needed in Step 6, we compute its singular value decomposition

$$\mathbf{B} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top, \quad \mathbf{\Sigma} = \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_p \\ & & & 0 \end{pmatrix} \in \mathbb{R}^{q \times p}, \quad \sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_p > 0,$$

where $\mathbf{U} = (\mathbf{U}_1 \ \dots \ \mathbf{U}_q) \in \mathbb{R}^{q \times q}$ and $\mathbf{V} = (\mathbf{V}_1 \ \dots \ \mathbf{V}_p) \in \mathbb{R}^{p \times p}$ are orthogonal matrices. Next, we replace sufficiently small singular value σ_i ,

$$\frac{\sigma_i}{\sigma_1} \leq \varepsilon_\Sigma, \quad \varepsilon_\Sigma > 0, \quad (5.90)$$

by zero. Let $r \geq 1$ be the largest number where (5.90) is not fulfilled. Then we define the “trimmed” matrix $\tilde{\mathbf{B}}$ with $\text{rank}(\tilde{\mathbf{B}}) = r$ as

$$\tilde{\mathbf{B}} = \tilde{\mathbf{U}}\tilde{\mathbf{\Sigma}}\tilde{\mathbf{V}}^\top \in \mathbb{R}^{p \times r}, \quad (5.91)$$

where

$$\tilde{\mathbf{\Sigma}} = \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{pmatrix}, \quad \tilde{\mathbf{U}} = (\mathbf{U}_1 \ \dots \ \mathbf{U}_r), \quad \tilde{\mathbf{V}} = (\mathbf{V}_1 \ \dots \ \mathbf{V}_r). \quad (5.92)$$

The resulted matrix $\tilde{\mathbf{B}}$ is called the low-rank approximation of \mathbf{B} .

- (iii) The basis of the low-rank approximation of (5.87) gathered from the SVD method in Step 6, is in general not orthonormal with respect to the L^2 inner product. To ensure that the resulting basis functions in (5.91) are mutually orthonormal, we apply the QR decomposition or the (modified) Gram-Schmidt method [72, §5.2] to find an orthonormal basis $\psi_1^{(m)}, \dots, \psi_K^{(m)}$ of the trimmed space with respect to the standard inner product.
- (iv) In Step 8, one can also reduce the sum in the solution by finding the smallest subset $I \subset \{1, \dots, N\}$ such that

$$\left\| P[u^{(m-1)} - \varphi_0^{(m-1)}] - \sum_{\ell \in I} \beta_\ell^{(m)} \psi_\ell^{(m)} \right\| \leq \text{Tol}_\beta, \quad \text{Tol}_\beta > 0.$$

By Parseval's identity and under the assumption that $|\beta_1^{(m)}| \geq \dots \geq |\beta_N^{(m)}|$, one only needs to find the smallest $K_m \leq N$ such that

$$\left\| P[u^{(m-1)} - \varphi_0^{(m-1)}] - \sum_{\ell=1}^{K_m} \beta_\ell^{(m)} \psi_\ell^{(m)} \right\| = \left(\sum_{\ell=K_m+1}^N |\beta_\ell^{(m)}|^2 \right)^{\frac{1}{2}} \leq \text{Tol}_\beta$$

and sets instead

$$u^{(m)} = \varphi_0^{(m)} + \sum_{\ell=1}^{K_m} \beta_\ell^{(m)} \psi_\ell^{(m)}, \quad U^{(m)} = \varphi_0^{(m)} + \text{span}\{\psi_1^{(m)}, \dots, \psi_{K_m}^{(m)}\}.$$

5.3.2 Numerical examples

Here we present some numerical examples to illustrate the accuracy and the usefulness of the ASI method, as well as the behavior of the dimension of the adaptive spectral space.

First, we solve a PDE-constrained optimization problem (5.84) in nodal basis (Algorithm 3) and in ASI basis (Algorithm 4) to reconstruct the unknown piecewise constant medium u (or u_δ) with two obstacles. Then, we consider the accuracy and convergence of both methods. In addition, we show how the complexity of the degree of freedoms using the ASI method combined with the size reduction approach decreases efficiently. Next, we repeat the same example with a third obstacle and consider again the accuracy of the reconstruction, obtained with the ASI method, and the complexity of the degree of freedoms in the ASI approach. Finally, we apply the ASI method to a more realistic Salt-dome model from geosciences.

Numerical experiment 1: circular-shaped inclusions

We consider

$$u(x) = 2 + \frac{4}{3}\chi^1(x) + \chi^2(x), \quad x \in \Omega = (0, 1) \times (0, 1),$$

where χ^k denotes the characteristic function of the disk A^k , $k = 1, 2$, with the radius $r_1 = 0.115$ and $r_2 = 0.125$. The exact medium u now is approximated with the piecewise linear function in \mathcal{P}^1 -FE – see Figure 5.8a.

We minimize the misfit J , defined in (5.3), between the synthetic data y and the observed data y^{obs} , where both are obtained by solving the Helmholtz equation with $\Gamma_S = \partial\Omega$ and the squared wave speed u . The eight Gaussian sources originate at $(0.05, 0.05)$, $(0.5, 0.05)$, $(0.95, 0.05)$, $(0.05, 0.5)$, $(0.95, 0.5)$, $(0.05, 0.95)$, $(0.5, 0.95)$, and $(0.95, 0.95)$, and the data, which contains 20% white noise, is measured on the whole

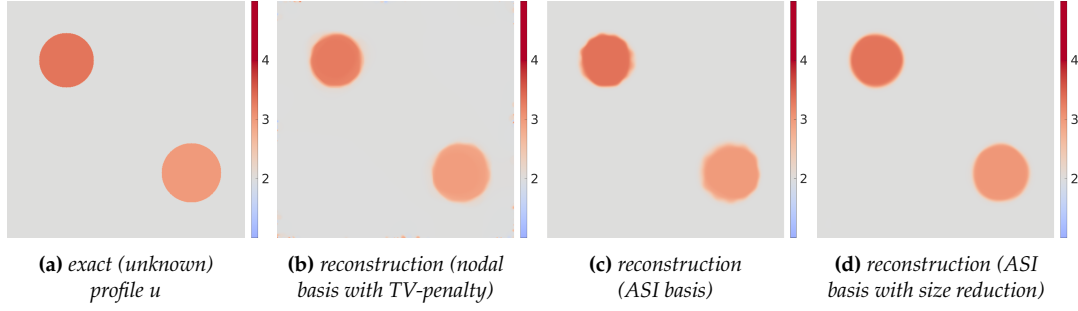


Figure 5.8: Two circular inclusions: comparison of the reconstructions of the unknown medium u (a), either with the grid-based nodal basis (b) or with the AS basis, either without (c) or with (d) dynamic size reduction from Section 5.3.1.

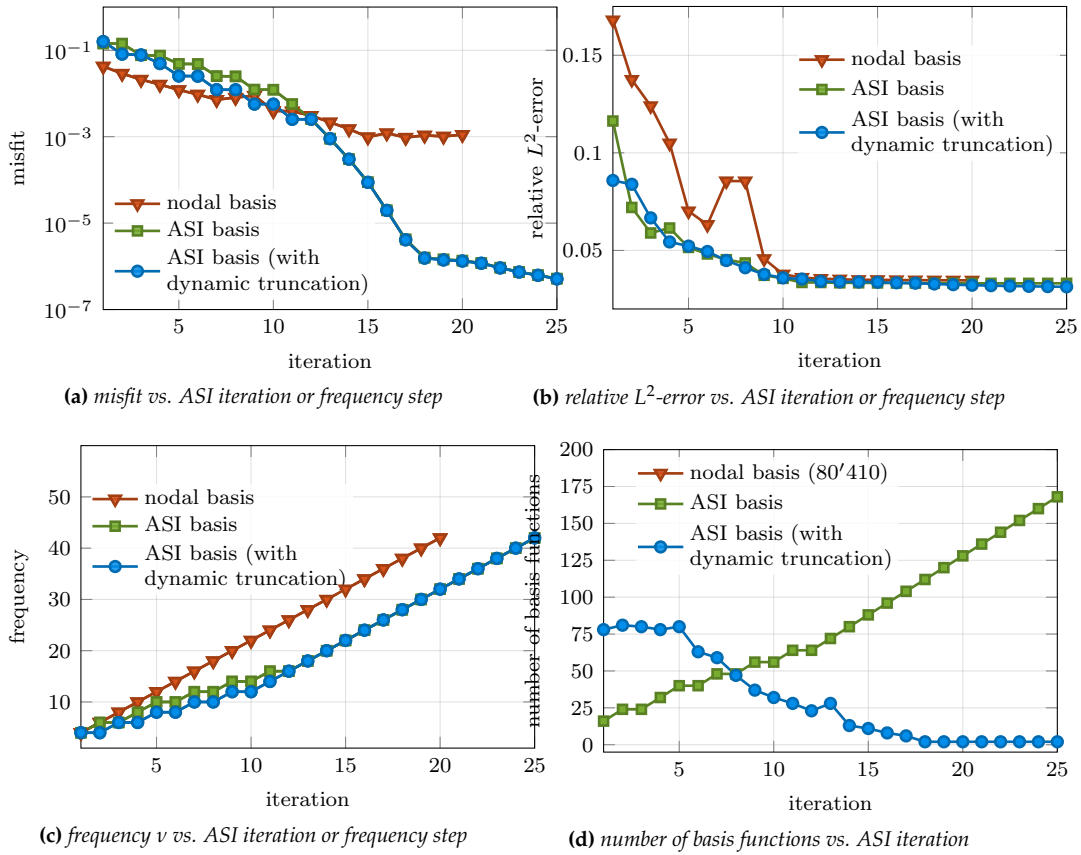


Figure 5.9: Two circular inclusions: comparison of the misfit J or \hat{J} (a), the relative L^2 -error (b), and the frequency (c) in each iteration (frequency step for nodal basis and in AS basis) of the inversion method in nodal basis and in AS basis; (d) comparison of the dimension of the AS space in each ASI iteration, obtained with the ASI method with and without dynamic size reduction, while the number of grid-based nodal basis is fixed ($\text{ndofs} = 80'401$).

boundary $\Gamma = \partial\Omega$. We use the \mathcal{P}^3 -FE for the spatial discretization. To avoid any inverse crime, the mesh for the observation is independent of the synthetic data and is about 30% finer than the mesh for the synthetic data.

Now, we apply the ASI method to the inverse problem (5.84). It starts at the lowest frequency $\nu = 4$, $\nu = 2\pi\omega$, and the adaptive spectral (AS) space $U^{(0)}$ in (5.85) with $\mu \equiv 1$. Then, we increase the frequency up to $\nu = 42$ with the frequency step $\Delta\nu = 2$ when the criterion for increment in Step 10 in the ASI algorithm 4 is fulfilled. Figures 5.9c and 5.9d monitor the frequency ν_m and the number of basis functions at

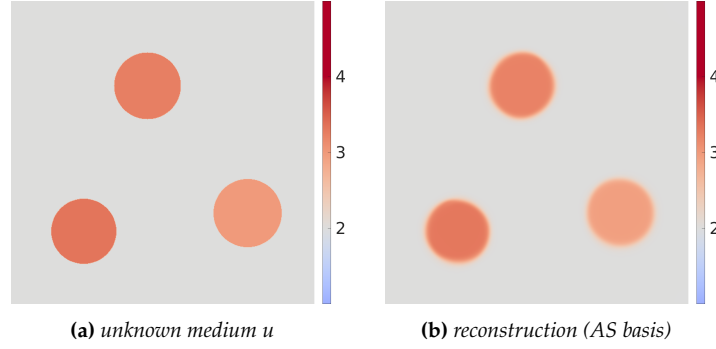


Figure 5.10: Three circular inclusions: (a) total medium u ; (b) ASI reconstruction.

each ASI iteration m .

First, we apply the ASI method from [77, 79] without the dynamic size reduction of the AS space. Starting with the spectral space $U^{(0)}$ of the (negative) Laplacian operator with $\dim(U^{(0)}) = 16$, the ASI method adapts the AS space $U^{(m)}$ given in (5.26) with $w = u^{(m)}$ and $\dim(U^{(m)}) = 4\nu_m$, $m = 1, \dots, 25$. The ASI method yields in the last iteration an AS space $U^{(m)}$, $m = 25$, with $\dim(U^{(m)}) = 168$ and a reconstruction shown in Figure 5.8c with a relative L^2 -error of 3.3%.

Second, we apply the ASI method with the dynamic space size reduction strategy starting with $\dim(U^{(0)}) = 50$. The dimension of the AS spaces $U^{(m)}$, $m = 1, \dots, 25$, resulting from the ASI method combined with size reduction approach in (5.89), vary from 3 to 89. The AS space $U^{(25)}$ in the last iteration only consists of two basis functions. Moreover, the ASI method even yields in the last step a reconstruction illustrated in Figure 5.8d with a lower relative L^2 -error of 3.1%.

Again, we repeat the same example and reconstruct the profile with the grid-based nodal basis with the TV-regularization. This yields a comparable accurate, relative L^2 -error of 3.5%. However, the control consists of 80'401 degrees of freedoms. In Figure 5.9b, we observe that the relative L^2 -error of all approaches reaches the same level of accuracy. In addition, Figure 5.9a monitors the misfit in each iteration m and shows that the misfit from ASI methods decreases monotonically and rapidly. In contrast, the misfit $\hat{J} = J + \alpha R$ in (5.5), obtained with the grid-based nodal approach, converges slowly and is stagnated at 0.00112 in the last iteration $m = 20$, where $J(u^{(m)}) = 0.00016$ and $\alpha_m R(u^{(m)}) = 0.00096$.

Next, we consider the scattering problem from three circular inhomogeneous inclusions,

$$u(x) = 2 + \frac{4}{3}\chi^1(x) + \chi^2(x) + \frac{5}{4}\chi^3(x), \quad x \in \Omega = (0, 1) \times (0, 1),$$

with the radii $r_1 = 0.1075$, $r_2 = 0.1125$, and $r_3 = 0.110$. We apply the ASI method, with the same settings as in the case of two disks, to reconstruct the unknown medium u . Figure 5.10a illustrates the exact profile. We impose 20% white noise on the observed data. The ASI method again starts with $\dim(U^{(0)}) = 50$ and yields here an AS space $U^{(m)}$, with $\dim(U^{(m)}) = 3$, in the last iteration $m = 24$. The corresponding reconstruction with a relative L^2 -error of 4% is shown in Figure 5.10b. Furthermore, Figure 5.11 displays the basis vectors $\psi_1^{(m)}$, $\psi_2^{(m)}$, and $\psi_3^{(m)}$ of $U^{(m)}$, which remarkably match the eigenfunctions φ_1 , φ_2 , and φ_3 in (5.26) with $w = u$.

To illustrate the robustness with respect to the observed data with perturbations, we compare the reconstruction of u , obtained from observations, containing white

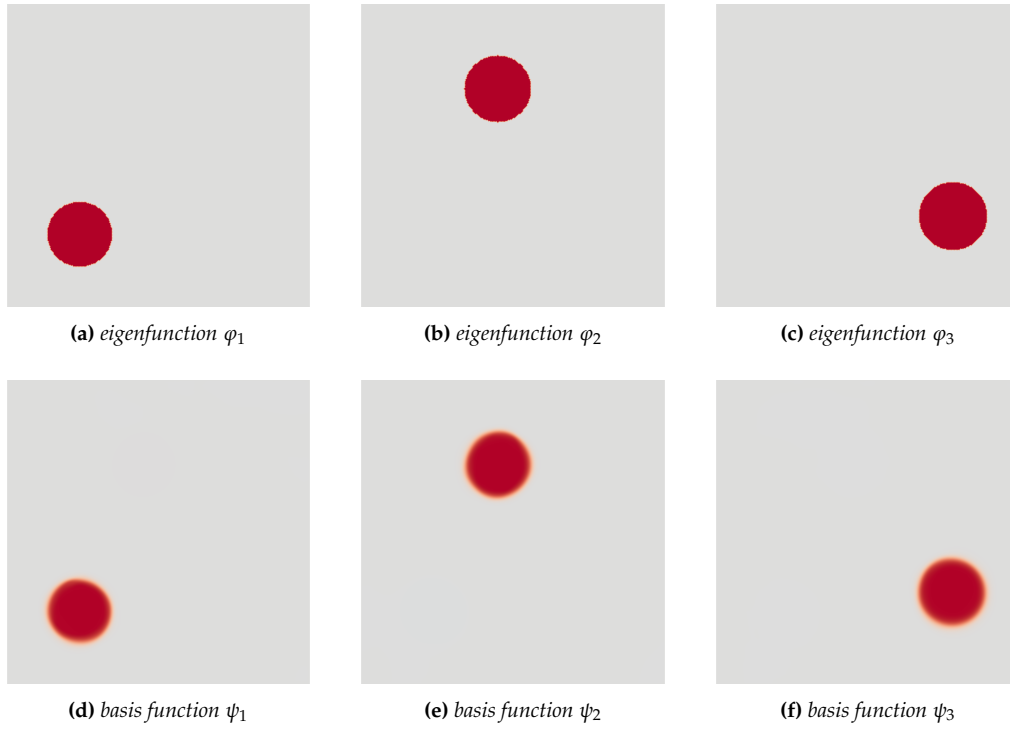


Figure 5.11: Three circular inclusions: (a)–(c) eigenfunctions in (5.26) with the exact medium $w = u$; (d)–(f) basis of the AS space $U^{(m)}$ in the last iteration $m = 24$, obtained with the ASI method combined with the space size reduction approach in Section 5.3.1.

noise of level 0%, 10%, 20%, or 40%, with the ASI method. Here we observe that the ASI method yields relative L^2 -errors between 3.9% and 4.1%.

We conclude that the adaptive spectral inversion (ASI), with the dynamic space size reduction approach, not only reduces the degree of unknowns significantly but also yields an accurate and robust method.

Numerical experiment 2: Salt-Pluto model

Let us consider the two-dimensional Pluto 1.5 model from geosciences generated by the “subsalt multiples attenuation and reduction technology” (SMAART). That is (2.9) in $\Omega = (0, 24.4) \times (-9, 0)$ [km] with the squared velocity profile $u(x)$ shown in Figure 5.12a. Here we impose that the solution y of (2.9) satisfies the absorbing boundary condition (2.9b) on the lateral and lower boundaries and the homogeneous (physical) Neumann boundary condition $\frac{\partial}{\partial n}y = 0$ on the surface ($y = 0$). The $N_s = 91$ source terms are located at the top, 50 meters beneath the surface ($y = 0.05$), distributed in about every 270 meters. Next, the data, which contains 20% white noise, is measured on the surface $\Gamma = \{y = 0\} \subset \partial\Omega$.

For the discretization of the synthetic data, we use a \mathcal{P}^3 -FE method with at least 20 points per wave length. To avoid any inverse crime, the mesh for the observation again is finer than the mesh for the synthetic data. Next, we apply \mathcal{P}^1 -FE with 54'572 vertices and 108'400 elements to present the control u_δ .

First, we show the approximation of u_δ in the AS space, by solving (5.26) with $w = u_\delta$ in μ_ϵ — see Figure 5.12b. Then, starting at $\nu = 0.5$ [Hz], with the initial guess $u^{(0)} = u_0^{(0)}$ shown in Figure 5.12c, and the eigenspace $U^{(0)}$ spanned by spectral basis of the negative Laplacian operator with $\dim(U^{(0)}) = 150$, we solve the inverse problem (5.84) by using the ASI method with increasing frequency ν up to 3 [Hz]. The ASI

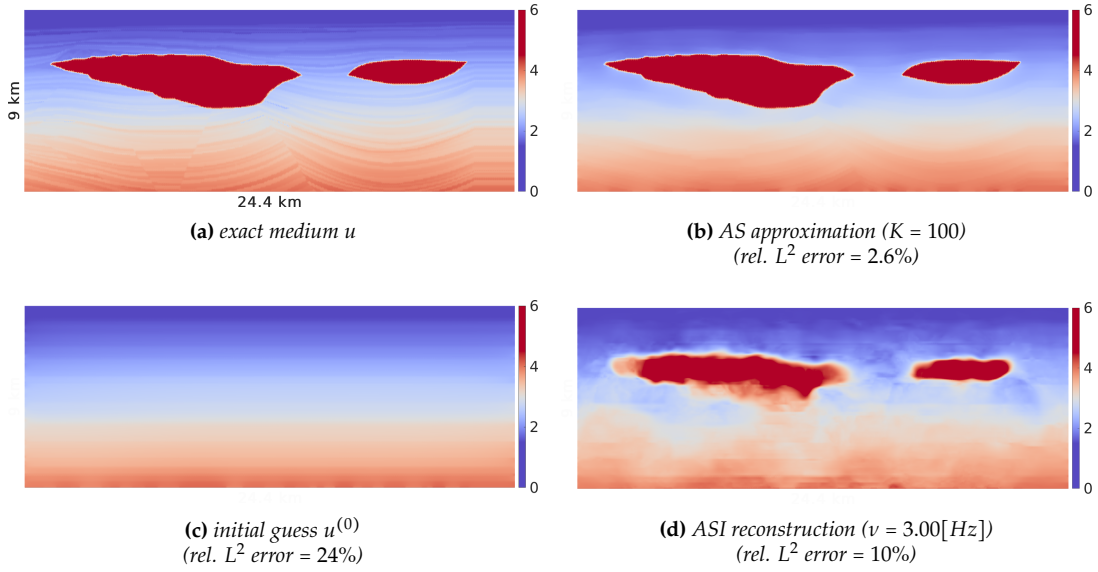


Figure 5.12: Salt dome (Pluto) model: comparison of the exact profile u (a) with the AS approximation (b), obtained with the L^2 -orthogonal projection, and with $u^{(m)}$ (d), obtained with the ASI method after $m = 13$ ASI iterations starting with the initial guess $u^{(0)}$ (c).

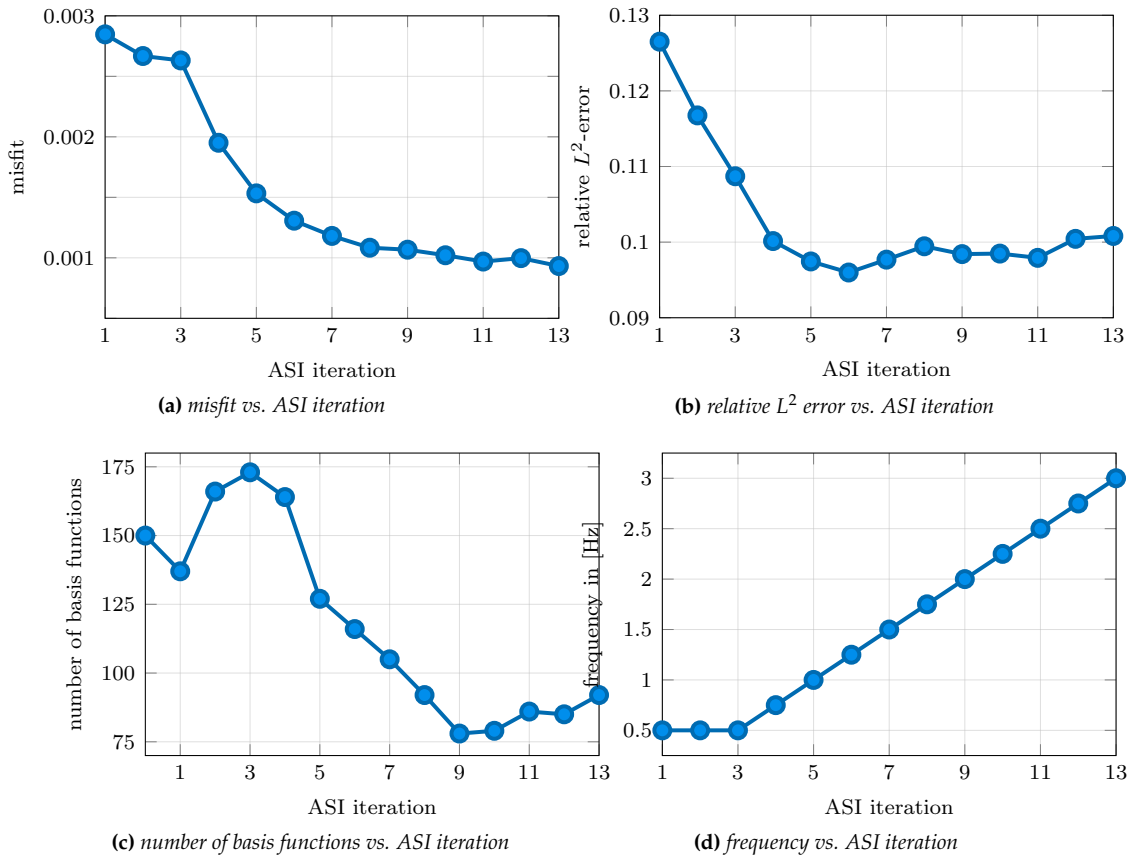


Figure 5.13: Salt dome (Pluto) model: comparison of the misfit (a), the relative L^2 error (b), the number of basis functions K_m (c), and the frequency ν (d) in each ASI iteration m .

method yields the reconstruction of u_δ in the last iteration, shown in Figure 5.12d, with an L^2 -error of 10%. In addition, Figure 5.13b monitors the history of the L^2 -error in each ASI iteration. We observe that it precisely captures the location and size of

the obstacles (salt-dome). However, the shape of the obstacle slightly differs from the exact shape, in particular the bottom of the large obstacle probably due to lack of measured data.

Figure 5.13c monitors the dimension of the $U^{(m)}$ in each ASI iteration m . Here we observe that the number of basis functions increases at the beginning from 150 to 173 and decreases and is stagnated at about 87. The frequency increases in each ASI iteration by $\Delta\nu = 0.25$ [Hz], shown in Figure 5.13d, when the criterion,

$$\|u_h^{(m)} - u_h^{(m-1)}\| \leq Tol_{ASI} = 2.5\%$$

is fulfilled.

In Figure 5.13a, we observe the convergence behavior of the misfit function J , which is monotonically decreasing.

Conclusion

Starting from inverse scattering problems in frequency domain for probing the unknown medium, we have seen that solving these problems requires the repeated multiple solutions of the time-harmonic wave equation or the (forward) Helmholtz equation. To gain more detailed properties of the medium, one requires the high-frequency solutions to the forward problem, which is well-known to be difficult to solve.

First, based on [25, 26], we have shown how to apply the controllability method (CM) combined with the conjugate gradient (CG) method as an alternative method for solving the Helmholtz equation. Akin to a shooting method, the CMCG method determines the time-harmonic solution of the corresponding wave equation in the time domain by iteratively reducing the departure from periodicity. Each CG iteration then requires the numerical solution of a forward and a backward wave equation over one period, together with the solution of a coercive elliptic problem independent of the frequency. Thanks to the well-known parallel efficiency of explicit methods combined with the excellent scalability of two-level domain decomposition preconditioners for coercive elliptic problems up to thousands of cores, the CMCG method is inherently parallel with strong scalability. Including a transient initial run-up to determine a judicious initial guess significantly accelerates the CG iteration. In fact, for scattering from convex obstacles, simply solving the time-harmonically forced wave equation over a long-time without any controllability can provide an even simpler, highly parallel Helmholtz solver. For nonconvex obstacles, however, solving the wave equation without any controllability is not a viable option, as the long time asymptotic convergence to the time-harmonic regime is simply too slow due to trapped modes. In all cases, the CMCG Algorithm combined with the initial run-up leads to the smallest time-to-solution. In the presence of local mesh refinement, high-order explicit local time-stepping (LTS) methods [78, 81] overcome the bottleneck due to an overly stringent CFL stability constraint.

However, the original CMCG method is based on the minimization of the cost functional J_1 in (3.33), which only guarantees convergence to the (unique) solution of the Helmholtz equation for sound-soft scattering problems, when both Dirichlet and impedance boundary conditions are imposed. For any other combination of boundary conditions, the minimization of J_1 generally does not lead to the correct time-harmonic solution. To extend the original CMCG method to boundary value problems governed by the Helmholtz equation with arbitrary combinations of Dirichlet, Neumann, or impedance boundary conditions, we have proposed to use alternative cost functionals or additional conditions and procedures to restore the uniqueness.

For scattering problems from sound-hard obstacles or inclusions, the compatibility condition eliminates the spurious constant shift in the minimizer and thus even permits the use of the original cost functional J_1 . Alternatively, (more accurate) absorbing boundary conditions that include a zeroth-order term also permit

the use of J_1 without modification. Exact nonreflecting boundary conditions based on integral formulations or DtN maps [76] could also be used but will generally lead to nonlocal space-time operators on the artificial boundary. Perfectly matched layers, however, lead to local formulations without convolution integrals in time when combined with appropriate auxiliary functions [99, 82, 8]. Next, for Helmholtz equation in physically bounded domains, the CMCG solution, obtained with the original cost functional, will generally contain higher order spurious eigenmodes. We thus have proposed a simple filtering procedure to remove them. In contrast to J_1 , the alternative cost functional J_3 in (3.43) always yields the true solution at little extra computational cost and storage requirement.

Based on the first-order formulation of the wave equation, the CMCG method uses a recent hybridized discontinuous Galerkin (HDG) discretization, which not only automatically yields a block-diagonal mass-matrix but also completely avoids solving the coercive elliptic problem that occurred in the CG method. Hence, it is trivially parallelized and even leads to superconvergence after a local post-processing step.

The CMCG approach developed here for the Helmholtz equation immediately generalizes to other time-harmonic vector wave equations from electromagnetics or elasticity [90].

Second, the inverse Helmholtz problems used to explore the unknown inhomogeneous medium is formulated as a nonlinear PDE-constrained optimal control problem for the misfit functional $J[u]$. Since it is well-known that minimizing J is severally ill-posed, a penalty term $\alpha R[u]$ based on the total variation is added to the optimization. The parameter α originating from the regularization is determined dynamically from a guess of u . This leads not only to a robust method, but is also flexible and accurate.

However, the representation of the control in the grid-based nodal basis, resulting e.g. either from the finite difference (FD) or finite element (FE) discretization, typically leads to a prohibitively large number of unknown variables. Therefore, we have shown how to instead use the adaptive spectral (AS) decomposition as regularization [43, 77, 79, 9]. It does not only significantly reduce the number of unknowns, but also yields a very robust and accurate method.

For piecewise constant medium, e.g. a sum of K linearly independent characteristics, our analytical theory has shown that the AS space only requires K eigenfunctions to accurately approximate the medium, which is finally verified by numerical examples.

In [77, 79], Grote et al. proposed to use the frequency stepping strategy to prevent that the optimization method converges to a false minimum. Moreover, for each frequency, the new AS space $U^{(m)}$, with the dimension increasing proportional to the frequency, completely replaces the previous AS space $U^{(m-1)}$. Based on that, the new merging approach guarantees that the information of the previous (affine) space $U^{(m-1)}$ of candidate functions passes to the new space $U^{(m)}$. Moreover, its dimension is dynamically controlled by the size reduction strategy, which is based on the low-rank approximation. The numerical experiments showed that the modified ASI approach not only is efficient, robust, and dynamically reduces the number of unknowns, but is also remarkably accurate.

Bibliography

- [1] G. S. Abdoulaev, A. H. Hielscher, and K. Ren. Optical tomography as a PDE-constrained optimization problem. *Inverse Problems*, 21(5):1507–1530, 2005.
- [2] T. Airaksinen and S. Mönkölä. Comparison between the shifted-Laplacian preconditioning and the controllability methods for computational acoustics. *Journal of Computational and Applied Mathematics*, 234(6):1796–1802, 2010.
- [3] A. Aravkin, E. Esser, L. Guasch, F. Herrmann, and T. van Leeuwen. Total variation regularization strategies in full-waveform inversion. *SIAM Journal on Imaging Sciences*, 11(1):376–406, 2018.
- [4] U. M. Ascher, E. Haber, and D. Oldenburg. On optimization techniques for solving nonlinear inverse problems. *Inverse Problems*, 16(5):1263–1280, 2000.
- [5] A. K. Aziz, R. B. Kellogg, and A. B. Stephens. A two point boundary value problem with a rapidly oscillating solution. *Numerische Mathematik*, 53:107–121, 1988.
- [6] I. Babuška and F. Ihlenburg. Finite element solution of the Helmholtz equation with high wave number Part I: the h-version of the FEM. *Computers and Mathematics with Applications*, 30(9):9–37, 1995.
- [7] I. M. Babuška and S. A. Sauter. Is the pollution effect of the FEM avoidable for the Helmholtz equation considering high wave numbers. *SIAM Journal on Numerical Analysis*, 34(6):2392–2423, 1997.
- [8] D. H. Baffet, M. J. Grote, S. Imperiale, and M. Kachanovska. Energy decay and stability of a perfectly matched layer for the wave equation. Technical Report No. 2018/13, University of Basel, Department of Mathematics and Computer Science, 2018.
- [9] D. H. Baffet, M. J. Grote, and J. H. Tang. Adaptive spectral decomposition for inverse scattering. Technical Report (in preparation), University of Basel, Department of Mathematics and Computer Science, 2020.
- [10] G. Bao and J. Lin. Imaging of local surface displacement on an infinite ground plane: the multiple frequency case. *SIAM Journal on Applied Mathematics*, 71(5):1733–1752, 2011.
- [11] C. Bardos and J. Rauch. Variational algorithms for Helmholtz equation using time evolution and artificial boundaries. *Asymptotic Analysis*, 9(2):101–117, 1994.
- [12] A. Bayliss, M. Gunzburger, and E. Turkel. Boundary conditions for the numerical solution of elliptic equations in exterior region. *SIAM Journal on Applied Mathematics*, 42(2):430–451, 1982.

- [13] A. Bayliss, M. Gunzburger, and E. Turkel. The numerical solution of the Helmholtz equation for wave propagation problems in underwater acoustics. *Computers and Mathematics with Applications*, 11(7–8):655–665, 1985.
- [14] A. Bayliss, M. Gunzburger, and E. Turkel. On accuracy conditions for the numerical computation of waves. *Journal of Computational Physics*, 59(3):396–404, 1985.
- [15] E. Bécache, P. Joly, and C. Tsogka. An analysis of new mixed finite elements for the approximation of wave propagation problems. *SIAM Journal on Numerical Analysis*, 37(4):1053–1084, 2000.
- [16] G. Bellettini, V. Caselles, and M. Novaga. The total variation flow in R^N . *Journal of Differential Equations*, 184(2):475–525, 2002.
- [17] J.-D. Benamou and B. Desprès. A domain decomposition method for the Helmholtz equation and related optimal control problems. *Journal of Computational Physics*, 136(1):68–82, 1997.
- [18] J.-P. Bérenger. A perfectly matched layer for the absorption of electromagnetic waves. *Journal of Computational Physics*, 114(2):185–200, 1994.
- [19] A. Bermúdez, L. Hervella-Nieto, A. Prieto, and R. Rodríguez. A family of exact bounded PML for the Helmholtz problem. *ECCOMAS 2004 - European Congress on Computational Methods in Applied Sciences and Engineering*, 2004.
- [20] M. Bollhöfer, M. J. Grote, and O. Schenk. Algebraic multilevel preconditioner for the Helmholtz equation in heterogeneous media. *SIAM Journal on Scientific Computing*, 31(5):3781–3805, 2009.
- [21] M. Bonazzoli, V. Dolean, I. G. Graham, E. A. Spence, and P.-H. Tournier. A two-level domain-decomposition preconditioner for the time-harmonic Maxwell’s equations. *Lecture Notes in Computational Science and Engineering*, 2018.
- [22] A. Boulkhémair and A. Chakib. On the uniform Poincaré inequality. *Communications in Partial Differential Equations*, 32(9):1439–1447, 2007.
- [23] A. Bourgeois, M. Bourget, P. Lailly, M. Poulet, P. Ricarte, and R. Versteeg. Marmousi, model and data. 1990.
- [24] D. Braess. *Finite Elemente*. Springer, 3 edition, 2003.
- [25] M.-O. Bristeau, R. Glowinski, and J. Périaux. On the numerical solution of the Helmholtz equation at large wave numbers using exact controllability methods. Application to scattering. In A. Quarteroni, J. Periaux, Y. A. Kuznetsov, and O. B. Widlund, editors, *Domain Decomposition Methods in Science and Engineering*, volume 157, pages 399–419. Contemporary Mathematics American Mathematical Society, 1994.
- [26] M.-O. Bristeau, R. Glowinski, and J. Périaux. Controllability methods for the computation of time-periodic solutions; Application to scattering. *Journal of Computational Physics*, 147(2):265–292, 1998.
- [27] R. Brossier, L. Métivier, S. Operto, and J. Virieux. Full waveform inversion and the truncated Newton method. *SIAM Journal on Scientific Computing*, 35(2):B401–B437, 2013.

- [28] M. Burger, G. Gilboa, and M. Moeller. Nonlinear spectral analysis via one-homogeneous functionals: overview and future prospects. *Journal of Mathematical Imaging and Vision*, 56(2):300–319, 2016.
- [29] F. Cakoni and D. Colton. *Qualitative Methods in Inverse Scattering Theory*. Springer, 1 edition, 2006.
- [30] H. Calandra, S. Gratton, X. Pinel, and X. Vasseur. An improved two-grid preconditioner for the solution of three-dimensional Helmholtz problems in heterogeneous media. *Numerical Linear Algebra with Applications*, 20(4):663–688, 2013.
- [31] H. Calandra, S. Gratton, and X. Vasseur. A geometric multigrid preconditioner for the solution of the Helmholtz equation in three-dimensional heterogeneous media on massively parallel computers. In D. Lahaye, J. Tang, and K. Vuik, editors, *Modern Solvers for Helmholtz Problems*, pages 141–155. Springer International Publishing, 2017.
- [32] G. Chavent. *Nonlinear Least Squares for Inverse Problems*. Springer, 2010.
- [33] H. Chen, P. Lu, and X. Xu. A hybridizable discontinuous Galerkin method for the Helmholtz equation with high wave number. *SIAM Journal on Numerical Analysis*, 51(4):2166–2188, 2013.
- [34] W. Chen, Z. Wang, and J. Zhou. Large-scale L-BFGS using MapReduce. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 27*, pages 1332–1340. Curran Associates, Inc., 2014.
- [35] Y. Chen. Inverse scattering via Heisenberg’s uncertainty principle. *Inverse Problems*, 13(2):253–282, 1997.
- [36] M. Chin-Joe-Kong, W. Mulder, and M. van Veldhuizen. Higher-order triangular and tetrahedral finite elements with mass lumping for solving the wave equation. *Journal of Engineering Mathematics*, 35(4):405–426, 1999.
- [37] B. Cockburn, N. Nguyen, and J. Peraire. An implicit high-order hybridizable discontinuous Galerkin method for linear convection–diffusion equations. *Journal of Computational Physics*, 228(9):3232–3254, 2009.
- [38] B. Cockburn, N. Nguyen, J. Peraire, and M. Stanglmeier. An explicit hybridizable discontinuous Galerkin method for the acoustic wave equation. *Computer Methods in Applied Mechanics and Engineering*, 300:748–769, 2016.
- [39] G. Cohen, P. Joly, J. E. Roberts, and N. Tordjman. Higher order triangular finite elements with mass lumping for the wave equation. *SIAM Journal on Numerical Analysis*, 38(6):2047–2078, 2001.
- [40] P. Cummings and X. Feng. Sharp regularity coefficient estimates for complex-valued acoustic and elastic Helmholtz equations. *Mathematical Models and Methods in Applied Sciences*, 16(1):139–160, 2006.
- [41] W. Dahmen and A. Reusken. *Numerik für Ingenieure und Naturwissenschaftler*. Springer, 2008.

- [42] R. Dautray and J. Lions. *Analyse mathématique et calcul numérique pour les sciences et les techniques*. Masson, 3 edition, 1988.
- [43] M. de Buhan and M. Kray. A new approach to solve the inverse scattering problem for waves: combining the TRAC and the adaptive inversion methods. *Inverse Problems*, 29(8):085009, 2013.
- [44] M. de Buhan and A. Osses. Logarithmic stability in determination of a 3D viscoelastic coefficient and a numerical example. *Inverse Problems*, 26(9):095006, 2010.
- [45] M. C. Delfour and J.-P. Zolésio. *Shapes and Geometries*. Society for Industrial and Applied Mathematics, second edition, 2011.
- [46] V. Dolean, P. Hauret, F. Nataf, C. Pechstein, R. Scheichl, and N. Spillane. Abstract robust coarse spaces for systems of PDEs via generalized eigenproblems in the overlaps. *Numerische Mathematik*, 126(4):741–770, 2014.
- [47] V. Dolean, P. Jolivet, and F. Nataf. *An Introduction to Domain Decomposition Methods. Algorithms, Theory, and Parallel Implementation*. SIAM, 2015.
- [48] W. Dörfler and S. Sauter. A posteriori error estimation for highly indefinite Helmholtz problems. *Computational Methods in Applied Mathematics*, 13(3):333–347, 2013.
- [49] J. Douglas, J. E. Santos, D. Sheen, and L. Schreyer. Frequency domain treatment of one-dimensional scalar waves. *Mathematical Models and Methods in Applied Sciences*, 3(2):171–194, 1993.
- [50] Y. Du and H. Wu. Preasymptotic error analysis of higher order FEM and CIP-FEM for Helmholtz equation with high wave number. *SIAM Journal on Numerical Analysis*, 53(2):782–804, 2015.
- [51] S. C. Eisenstat and H. F. Walker. Choosing the forcing terms in an inexact Newton method. *SIAM Journal on Scientific Computing*, 17(1):16–32, 1996.
- [52] B. Engquist and A. Majda. Absorbing boundary conditions for numerical simulation of waves. *Proceedings of the National Academy of Sciences*, 74(5):1765–1766, 1977.
- [53] B. Engquist and L. Ying. Sweeping preconditioner for the Helmholtz equation: moving perfectly matched layers. *SIAM Journal on Multiscale Modeling and Simulation*, 9(2):686–710, 2011.
- [54] Y. A. Erlangga, C. W. Oosterlee, and C. Vuik. On a class of preconditioners for solving the Helmholtz equation. *Applied Numerical Mathematics*, 50(3–4):409–425, 2004.
- [55] O. G. Ernst and M. J. Gander. Why it is difficult to solve Helmholtz problems with classical iterative methods. In I. G. Graham, T. Y. Hou, O. Lakkis, and R. Scheichl, editors, *Numerical Analysis of Multiscale Problems*, pages 325–363. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.
- [56] L. C. Evans. *Partial Differential Equation*. American Mathematical Society, 2010.
- [57] L. C. Evans and R. F. Gariepy. *Measure theory and fine properties of functions*. CRC Press, 1992.

- [58] E. Fatemi, S. Osher, and L. I. Rudin. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1–4):259–268, 1992.
- [59] H. Federer. *Geometric Measure Theory*. Springer-Verlag, 1969.
- [60] R. W. Freund and N. M. Nachtigal. QMR: a quasi-minimal residual method for non-Hermitian linear systems. *Numerische Mathematik*, 60(1):315–339, 1991.
- [61] J. M. Gander and H. Zhang. A class of iterative solvers for the Helmholtz equation: factorizations, sweeping preconditioners, source transfer, single layer potentials, polarized traces, and optimized Schwarz methods. *SIAM Review*, 61(1):3–76, 2019.
- [62] M. Gander. Optimized Schwarz methods. *SIAM Journal on Numerical Analysis*, 44(2):699–731, 2006.
- [63] M. J. Gander, I. G. Graham, and E. A. Spence. Applying GMRES to the Helmholtz equation with shifted Laplacian preconditioning: what is the largest shift for which wavenumber-independent convergence is guaranteed? *Numerische Mathematik*, 131(3):567–614, 2015.
- [64] M. J. Gander and F. Kwok. Optimal interface conditions for an arbitrary decomposition into subdomains. In Y. Huang, R. Kornhuber, O. Widlund, and J. Xu, editors, *Domain Decomposition Methods in Science and Engineering XIX*, pages 101–108, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg.
- [65] M. J. Gander, F. Magoulès, and F. Nataf. Optimized Schwarz methods without overlap for the Helmholtz equation. *SIAM Journal on Scientific Computing*, 24:38–60, 2002.
- [66] O. Gauthier, A. Tarantola, and J. Virieux. Two-dimensional nonlinear inversion of seismic waveforms: numerical results. *GEOPHYSICS*, 51(7):1387–1403, 1986.
- [67] S. Geevers, W. A. Mulder, and J. J. W. van der Vegt. New higher-order mass-lumped tetrahedral elements for wave propagation modelling. *SIAM Journal on Scientific Computing*, 40(5):A2830–A2857, 2018.
- [68] R. Glowinski, S. Kähkönen, T. Rossi, and R. Mäkinen. Solution of time-periodic wave equation using mixed finite elements and controllability techniques. *Journal of Computational Acoustics*, 19(4):335–352, 2011.
- [69] R. Glowinski and S. Lapin. Solution of a wave equation by a mixed finite element – fictitious domain method. *Computational Methods in Applied Mathematics*, 4(4):431–444, 2004.
- [70] R. Glowinski and T. Rossi. A mixed formulation and exact controllability approach for the computation of the periodic solutions of the scalar wave equation. (I): Controllability problem formulation and related iterative solution. *Comptes Rendus Mathématique*, 343(7):493–498, 2006.
- [71] G. H. Golub, P. C. Hansen, and D. P. O’Leary. Tikhonov regularization and total least squares. *SIAM Journal on Matrix Analysis and Applications*, 21(1):185–194, 1999.

- [72] G. H. Golub and C. F. Van Loan. *Matrix Computations*. Johns Hopkins University Press, Baltimore, MD, USA, 3 edition, 1996.
- [73] I. Graham, O. Pembery, and E. Spence. The Helmholtz equation in heterogeneous media: a priori bounds, well-posedness, and resonances. *Journal of Differential Equations*, 226(6):2869–2923, 2019.
- [74] I. G. Graham and S. A. Sauter. Stability and error analysis for the Helmholtz equation with variable coefficients. *Mathematics of Computation*, 2018.
- [75] I. G. Graham, E. A. Spence, and E. Vainikko. Domain decomposition preconditioning for high-frequency Helmholtz problems with absorption. *Mathematics of Computation*, 86(307):2089–2127, 2017.
- [76] M. J. Grote and J. B. Keller. On nonreflecting boundary conditions. *Journal of Computational Physics*, 122(2):231–243, 1995.
- [77] M. J. Grote, M. Kray, and U. Nahum. Adaptive eigenspace method for inverse scattering problems in the frequency domain. *Inverse Problems*, 33(2):025006, 2017.
- [78] M. J. Grote, M. Mehlin, and T. Mitkova. Runge–Kutta-based explicit local time-stepping methods for wave propagation. *SIAM Journal on Scientific Computing*, 37(2):A747–A775, 2015.
- [79] M. J. Grote and U. Nahum. Adaptive eigenspace for multi-parameter inverse scattering problems. *Computers and Mathematics with Applications*, 77(12):3264–3280, 2019.
- [80] M. J. Grote, F. Nataf, J. H. Tang, and P.-H. Tournier. Parallel controllability methods for the Helmholtz equation. arXiv:1903.12522, Numerical Analysis, 2019.
- [81] M. J. Grote, D. Peter, M. Rietmann, and O. Schenk. Newmark local time stepping on high-performance computing architectures. *Journal of Computational Physics*, 334:308–326, 2017.
- [82] M. J. Grote and I. Sim. Efficient PML for the wave equation. arXiv:1001.0319v1, Numerical Analysis, 2010.
- [83] M. J. Grote and J. H. Tang. On controllability methods for the Helmholtz equation. *Journal of Computational and Applied Mathematics*, 358:306–326, 2019.
- [84] E. Haber and E. Treister. Full waveform inversion guided by travel time tomography. *SIAM Journal on Scientific Computing*, 39(5):S587–S609, 2017.
- [85] P. C. Hansen. *Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion*. Society for Industrial and Applied Mathematics, 1998.
- [86] P. C. Hansen and D. P. O’leary. The use of the L-curve in the regularization of discrete ill-posed problems. *SIAM Journal on Scientific Computing*, 14(6):1487–1503, 1993.
- [87] I. Harari and T. J. Hughes. Finite element methods for the Helmholtz equation in an exterior domain: model problems. *Computer Methods in Applied Mechanics and Engineering*, 87(1):59–96, 1991.

- [88] F. Hecht. New development in freefem++. *Journal of Numerical Mathematics*, 20(3–4):251–265, 2012.
- [89] F. Hecht, P. Jolivet, F. Nataf, and C. Prud’homme. Scalable domain decomposition preconditioners for heterogeneous elliptic problems. In *Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis*, SC’13, pages 1–11, New York, NY, USA, 2013. ACM.
- [90] E. Heikkola, S. Mönkölä, A. Pennanen, and T. Rossi. Controllability method for the solution of linear elastic wave equation. In C. A. Motasoaes, J. A. C. Martins, H. C. Rodrigues, J. A. C. Ambrósio, C. A. B. Pina, C. M. Motasoaes, E. B. R. Pereira, and J. Folgado, editors, *III European Conference on Computational Mechanics*, page 71. Springer Netherlands, 2006.
- [91] E. Heikkola, S. Mönkölä, A. Pennanen, and T. Rossi. Controllability method for acoustic scattering with spectral elements. *Journal of Computational and Applied Mathematics*, 204(2):344–355, 2007.
- [92] E. Heikkola, S. Mönkölä, A. Pennanen, and T. Rossi. Controllability method for the Helmholtz equation with higher-order discretizations. *Journal of Computational Physics*, 225(2):1553–1576, 2007.
- [93] U. Hetmaniuk. Stability estimates for a class of Helmholtz problems. *Communications in Mathematical Sciences*, 5(3):665–678, 2007.
- [94] G. J. Hicks, R. G. Pratt, and C. Shin. Gauss–Newton and full Newton methods in frequency–space seismic waveform inversion. *Geophysical Journal International*, 133(2):341–362, 1998.
- [95] F. Ihlenburg. *Finite Element Analysis of Acoustic Scattering*. Springer, 1998.
- [96] N. Iwasaki. Local decay of solutions for symmetric hyperbolic systems with dissipative and coercive boundary conditions in exterior domains. *Publications of the Research Institute for Mathematical Sciences*, 5(2):193–218, 1969.
- [97] E. Jenkins, B. Rivière, and M. Wheeler. A priori error estimates for mixed finite element approximations of the acoustic wave equation. *SIAM Journal on Numerical Analysis*, 40(5):1698–1715, 2002.
- [98] D. Jerison and C. E. Kenig. Unique continuation and absence of positive eigenvalues for Schrödinger operators. *Annals of Mathematics*, 121(3):463–488, 1985.
- [99] B. Kaltenbacher, M. Kaltenbacher, and I. Sim. A modified and stable version of a perfectly matched layer technique for the 3-d second order wave equation in time domain with an application to aeroacoustics. *Journal of Computational Physics*, 235:407–422, 2013.
- [100] Y. Kim and L. A. Vese. Image recovery using functions of bounded variation and Sobolev spaces of negative differentiability. *Inverse Problems and Imaging*, 3(1):43–68, 2009.
- [101] A. Kirsch and D. Colton. A simple method for solving inverse scattering problems in the resonance region. *Inverse Problems*, 12(4):383–393, 1996.
- [102] D. Koyama. A controllability method with an artificial boundary condition for the exterior Helmholtz problem. *Japan Journal of Industrial and Applied Mathematics*, 20:117–145, 2003.

- [103] P. Lailly. The seismic inverse problem as a sequence of before stack migrations. In J. B. Bednar, editor, *Conference on Inverse Scattering – Theory and Application*. SIAM, 1983.
- [104] J.-L. Lions. Exact controllability, stabilization and perturbations for distributed systems. *SIAM Review*, 30(1):1–68, 1988.
- [105] P.-L. Lions. On the Schwarz alternating method. I. In R. Glowinski, G. H. Golub, G. A. Meurant, and J. Periaux, editors, *First International Symposium on Domain Decomposition Methods for Partial Differential Equations*, pages 1–42. SIAM, 1988.
- [106] D. C. Liu and J. Nocedal. On the limited memory BFGS method for large scale optimization. *Mathematical Programming*, 45(1–3):503–528, 1989.
- [107] S. P. MacLachlan, C. W. Oosterlee, and N. Umetani. A multigrid-based shifted Laplacian preconditioner for fourth-order Helmholtz discretization. *Numerical Linear Algebra with Applications*, 16(8):603–626, 2009.
- [108] J. Málek and Z. Strakoš. *Preconditioning and the Conjugate Gradient Method in the Context of Solving PDEs*. SIAM, 2014.
- [109] J. M. Melenk. *On Generalized Finite Element Methods*. PhD thesis, University of Maryland at College Park, 1995.
- [110] J. M. Melenk and S. Sauter. Convergence analysis for finite element discretizations of the Helmholtz equation with Dirichlet-to-Neumann boundary conditions. *Mathematics of Computation*, 79(272):1871–1914, 2010.
- [111] J. M. Melenk and S. Sauter. Wavenumber explicit convergence analysis for Galerkin discretizations of the Helmholtz equation. *SIAM Journal on Numerical Analysis*, 49(3):1210–1243, 2011.
- [112] K. Miller. Numerical analogs to the Schwarz alternating procedure. *Numerische Mathematik*, 7, 1965.
- [113] P. Mojabi and J. LoVetri. Microwave biomedical imaging using the multiplicative regularized Gauss–Newton inversion. *IEEE Antennas and Wireless Propagation Letters*, 8:645–648, 2009.
- [114] G. Mur. The finite-element modeling of three-dimensional electromagnetic fields using edge and nodal elements. *IEEE Transactions on Antennas and Propagation*, 41(7):948–953, 1993.
- [115] U. Nahum. *Adaptive Eigenspace for Inverse Problems in the Frequency Domain*. PhD thesis, University of Basel, 2016.
- [116] J. Nečas. Elementary description of principal results. In *Direct Methods in the Theory of Elliptic Equations*, pages 1–47. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.
- [117] J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer, 2 edition, 2006.
- [118] S. Operto and J. Virieux. An overview of full-waveform inversion in exploration geophysics. *GEOPHYSICS*, 74(6):WCC1–WCC26, 2009.

- [119] R. G. Pratt. Inverse theory applied to multi-source cross-hole tomography. Part 2: Elastic wave-equation method. *Geophysical Prospecting*, 38(3):311–329, 1990.
- [120] R. G. Pratt. Seismic waveform inversion in the frequency domain, Part 1: Theory and verification in a physical scale model. *GEOPHYSICS*, 64(3):888–901, 1999.
- [121] R. G. Pratt and M. H. Worthington. Inverse theory applied to multi-source cross-hole tomography. Part 1: Acoustic waveequation method. *Geophysical Prospecting*, 38(3):287–310, 1990.
- [122] A. Quarteroni. *Numerical Models for Differential Problems*. Springer, 2014.
- [123] B. Rivière. *Discontinuous Galerkin Methods for Solving Elliptic and Parabolic Equations*. SIAM, 2008.
- [124] Y. Saad and M. H. Schultz. GMRES: a generalized minimal residual algorithm for solving nonsymmetric linear systems. *SIAM Journal on Scientific and Statistical Computing*, 7(3):856–869, 1986.
- [125] S. Sauter and C. Torres. Stability estimate for the Helmholtz equation with rapidly jumping coefficients. *Zeitschrift für angewandte Mathematik und Physik*, 69(6):139, 2018.
- [126] H. A. Schwarz. Über einen Grenzübergang durch alternierendes Verfahren. *Vierteljahrsschrift der Naturforschenden Gesellschaft in Zürich*, 15:272–286, 1870.
- [127] E. M. Stein and R. Shakarchi. *Real Analysis: Measure Theory, Integration, and Hilbert Spaces*. University Press Group Ltd, 2005.
- [128] W. W. Symes. Reverse time migration with optimal checkpointing. *GEOPHYSICS*, 72(5):SM213–SM221, 2007.
- [129] A. Tarantola. Inversion of seismic reflection data in the acoustic approximation. *GEOPHYSICS*, 49(8):1259–1266, 1984.
- [130] A. Tarantola. *Inverse Problem Theory and Methods for Model Parameter Estimation*. SIAM, 2005.
- [131] A. N. Tikhonov. On the stability of inverse problems. *Doklady Akademii Nauk SSSR*, 39(5):195–198, 1943.
- [132] F. Tröltzsch. *Optimale Steuerung partieller Differentialgleichungen*. Vieweg, 2005.
- [133] H. van der Vorst. Bi-CGSTAB: a fast and smoothly converging variant of Bi-CG for the solution of nonsymmetric linear systems. *SIAM Journal on Scientific and Statistical Computing*, 13(2):631–644, 1992.
- [134] K. Wang and Y. Wong. Pollution-free finite difference schemes for non-homogeneous Helmholtz equation. *International Journal of Numerical Analysis and Modeling*, 11:787–815, 2014.
- [135] J. Wloka. *Partial Differential Equations*. Cambridge University Press, 1987.
- [136] T. H. Wolff. A property of measures in \mathbf{R}^N and an application to unique continuation. *Geometric and Functional Analysis*, 2(2):225–284, 1992.