Finite Element Methods for Seismic Imaging

Cost Reduction through mass matrix preconditioning by defect correction

Shamasundar, Ranjani

DOI
10.4233/uuid:a6dbbf0f-de8b-4c39-a13f-4904f206ace6

Publication date
2019

FINITE ELEMENT METHODS FOR SEISMIC MODELLING

COST REDUCTION THROUGH MASS MATRIX PRECONDITIONING BY DEFECT CORRECTION

RANJANI SHAMASUNDAR

Propositions

accompanying the dissertation

FINITE ELEMENT METHODS FOR SEISMIC MODELLING

DEFECT CORRECTION FOR COMPUTE COST REDUCTION USING MASS MATRIX PRECONDITIONING

by

Ranjani SHAMASUNDAR

1. The computational cost of finite element methods cannot compete with that of finite differences; however, with suitable cost reducing methods, their additional expense can pay off when high accuracy is required. (This thesis)

2. When representing the wave equation with separate motion and material equations, shorter wavelengths lead to null vectors, which means that the second-order representation performs better. (This thesis)

3. For the acoustic wave equation, assembling the mass matrices locally is more efficient than global assembly for finite element basis functions of degree greater than one. (This thesis)

4. Hermite elements with a bubble function offer good dispersion properties for the first-order wave equation, but need to be modified to include material inhomogeneities and continuity of the tangential component of the pressure gradient across edges of elements. (This thesis)

5. Proclamations about the end of oil industry should be dismissed—the industry will survive as long as population grows.

6. Things we own beyond basic needs of food, water and shelter (and WiFi in 2019) are an imposition on someone else's basic needs.

7. Being born in a rich country is a privilege, yet many people fail to realise this and live lives of resentment because of closed-mindedness.

8. The role of family in traditional eastern societies is taken over by the government in the welfare societies of the west.

9. Cultural stereotypes are not wrong in themselves, they are often statistically true. However, they are dangerous because they become sole representations of individuals.

10. Modern-day slavery comes in the form of mental shackles instead of physical labour.

These propositions are regarded as opposable and defendable, and have been approved as such by the promotor Prof. dr. W.A. Mulder.

Finite Element Methods for Seismic Modelling

Defect correction for compute cost reduction using mass matrix preconditioning

Finite Element Methods for Seismic Modelling

Defect correction for compute cost reduction using mass matrix preconditioning

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology,
by the authority of Rector Magnificus Prof. dr. ir. T.H.J.J. van der Hagen,
chair of the Board of Doctorates,
to be defended publicly on Monday 1 July 2019 at 12.30 hours

by

Ranjani Shamasundar

Master of Science, Mechanical Engineering,
Indian Institute of Science, Bangalore, India
Born in Mysore, India

This dissertation has been approved by the

promotor: Prof. dr. W.A. Mulder

Composition of the doctoral committee:

Rector Magnificus	Chairman
Prof. dr. W. A. Mulder	Delft University of Technology
Independent members:	
Prof. dr. ir. E. C. Slob	Delft University of Technology
Prof. dr. ir. C. W. Oosterlee	Delft University of Technology Centrum Wiskunde en Informatica
Prof. dr. D. G. Simons	Delft University of Technology
Dr. R.-É. Plessix	Shell Global Solutions International Institut Physique du Globe de Paris
Prof. dr. J. Bruining	Technical University of Delft
Prof. dr. K. J. Batenburg	Leiden University

Keywords: Finite Element Method, Acoustic Modeling, Forward Modeling, Subsurface Imaging

Printed by: Ipskamp Printing

Front: The cover design depicts the author's abstraction of mountains, triangular finite-elements and material contrasts.

ISBN 978-94-6366-185-0

An electronic version of this dissertation is available at
http://repository.tudelft.nl/.

Asatoma Sadgamaya
Tamasoma Jyotirgamaya
–
From untruth to truth
From darkness to light

Brihadaranyaka Upanishad

Contents

Summary

Demand for hydrocarbon fuel is predicted to keep increasing in the coming decades in spite of easily accessible alternative fuels due to shifting geopolitical and economic situations. In order to find new hydrocarbon pockets, we need sharper images of earth's subsurface. Also, the exploration of other sources of energy like geothermal will benefit from better models of the what lies underneath the surface. One way to obtain better images is to use superior numerical methods for forward modelling - Finite-element methods (FEM) are one such method, but their accuracy comes at the cost of increased compute expense. This thesis explores means to reduce this cost and adapt FEM to large-scale problems in geophysics.

The Finite Difference (FD) method is the most popular numerical approximation scheme used in subsurface imaging problems. Representing the wave as the solution of individual motion and material equations is advantageous in terms of accuracy and stability and leads to the natural inclusion of density variations in the medium. This representation is referred to as the first-order formulation of the wave equation in this document. Finite Element (FE) methods are commonly derived for second-order equations because of the nature of variational formulation.

Finite-element discretisations of the acoustic wave equation in the time domain often employ mass lumping to avoid the cost of inverting a large sparse mass matrix. Unfortunately, for a first-order system of equations, mass lumping destroys the superconvergence of numerical dispersion for odd-degree polynomials. In chapter 3 of this thesis, we consider defect correction as a means to restore the convergence. We adapt the defect correction method to FEM by solving the consistent mass matrix with the lumped one as preconditioner. For the lowest-degree element, fourth-order accuracy in 1D can be obtained with just a single iteration of defect correction. In this chapter, we analyse the behaviour of the error in eigenvectors as a function of the normalized wavenumber in the form of leading terms in its series expansion and find that this error exceeds the dispersion error, except for the lowest degree where the eigenvector error is zero. We also present results of numerical experiments that confirm this analysis.

Chapter 3 concluded that defect correction can improve the convergence property of finite-elements in the first-order system of acoustic equations in 1D; the inexpensive linear elements showed the same performance as a fourth-order scheme. However, for realistic problems we need to ensure that the same improvement holds in higher dimensions. Based on the results of the earlier chapter, we conjecture that defect correction should work for 2D problems. In the first half of chapter 4, we analyze the 2-D case. Theoretical results imply that the lowest-degree polynomial provides fourth-order accuracy with defect correction, if the grid of squares or triangles is highly regular and material properties constant. But numerical results converge more slowly than theoretical predictions. Further investigation demonstrates that this is due to the activation of error-inducing wavenumbers in the delta-source representation. In the second half of the chapter, we provide a solution to this problem in the form of a tapered-sinc source function.

In chapter 5, we consider isotropic elastic wave propagation with continuous mass-lumped finite elements on tetrahedra with explicit time stepping. These elements require higher-order polynomials in their interior to preserve accuracy after mass lumping and are recently discovered up to degree 4. Global assembly of the symmetric stiffness matrix is a natural approach but requires large memory. Local assembly on the fly, in the form of matrix-vector products per element at each time step, has a much smaller memory footprint. With dedicated expressions for local assembly, our code ran about 1.3 times faster for degree 2 and 1.9 times for degree 3 on a simple homogeneous test problem, using 24 cores. This is similar to the acoustic case. For a more realistic problem, the gain in efficiency was a factor 2.5 for degree 2 and 3 for degree 3. For the lowest degree, the linear element, the expressions for both the global and local assembly can be further simplified. In that case, global assembly is more efficient than local assembly. Among the three degrees, the element of degree 3 is the most efficient in terms of accuracy at a given cost.

In chapter 6, we consider cubic Hermite elements as interpolants in place of Legendre polynomials. By nature of their C^1 continuity, they might offer a solution to the problems of ‘spurious’ wavenumbers seen in earlier chapters with conventional interpolation schemes. Results show acceptable convergence properties on homogeneous media, but the representation needs to be altered to suit discontinuities in density, which makes interesting future work.

Samenvatting

Ondanks gemakkelijk toegankelijke alternatieve brandstoffen wordt verwacht dat de vraag naar koolwaterstofbrandstof in de komende decennia zal blijven toenemen als gevolg van veranderende geopolitieke en economische situaties. Om nieuwe koolwaterstofvoorkomens te vinden, hebben we scherpere afbeeldingen van de ondergrond van de aarde nodig. Ook de verkenning van andere energiebronnen zoals geothermie zal profiteren van betere modellen van wat zich onder het oppervlak bevindt. Eén manier om betere afbeeldingen te verkrijgen is om superieure numerieke methoden te gebruiken voor voorwaarts modelleren. Eindige-elementenmethoden (EEM) zijn een voorbeeld, maar hun nauwkeurigheid gaat ten koste van een hogere rekeninspanning. Dit proefschrift onderzoekt middelen om deze kosten te verminderen en zo de EEM beter geschikt te maken voor grootschalige geofysische problemen.

De eindige-differentiemethode (EDM) is het meest populaire numerieke benaderingsschema dat wordt gebruikt voor ondergrondse beeldvormingsproblemen. Om een golf voor te stellen als de oplossing van individuele bewegings- en materiaalvergelijkingen, heeft voordelen in termen van nauwkeurigheid en stabiliteit en staat op natuurlijke wijze het meenemen van dichtheidsvariaties in het medium toe. Deze representatie wordt de eerste-ordeformulering van de golfvergelijking genoemd in dit proefschrift. Eindige-elementenmethoden worden gewoonlijk afgeleid voor vergelijkingen van de tweede orde omdat dit het meest voor de hand ligt in de zwakke formulering.

De eindige-elementendiscretisatie van de akoestische golfvergelijking in het tijdsdomein maakt vaak gebruik van massaklontering om de kosten van het inverteren van een grote ijle massamatrix te vermijden. Jammer genoeg vernietigt massaklontering, voor een eerste-ordesysteem van vergelijkingen, de superconvergentie van de numerieke dispersie voor polynomen van oneven graad. In hoofdstuk 3 van dit proefschrift beschouwen we defectcorrectie als een manier om de convergentie te herstellen. We passen defectcorrectie toe op de EEM door de vergelijkingen voor de consistente massamatrix op te lossen met de geklonterde massamatrix als preconditioner. Voor het element met de laagste graad kan een nauwkeurigheid van de vierde orde in één dimensie worden verkregen door slechts een enkele iteratie met defectcorrectie. In dit hoofdstuk analyseren we het gedrag van de fout in de eigenvectoren als een functie van het genormaliseerd golfgetal in de vorm van de dominante termen in de reeksontwikkeling en vinden dat deze fout de dispersiefout overschrijdt, behalve voor de laagste orde waar de eigenvectorfout nul is. We presenteren ook resultaten van numerieke experimenten die deze analyse bevestigen.

Hoofdstuk 3 concludeert dat, voor het eerste orde systeem van akoestische vergelijkingen in 1D, defectcorrectie de convergentie van eindige elementen kan verbeteren. De goedkope lineaire elementen vertonen dezelfde prestaties als een vierde-ordeschema. Echter, voor realistische problemen moeten we ervoor zorgen dat dezelfde verbetering geldt in hogere dimensies. Gebaseerd op de resultaten van het eerdere hoofdstuk, vermoeden we dat defectcorrectie zou moeten werken voor 2-D problemen. In de eerste helft van hoofdstuk 4 analyseren we het 2-D geval. Theoretische resultaten laten zien dat het polynoom van de laagste graad vierde-orde nauwkeurigheid oplevert met defectcorrectie, als het rekenrooster van vierkanten of driehoeken zeer regelmatig is en de materiaaleigenschappen constant. Maar numerieke resultaten convergeren langzamer dan de theoretische voorspellingen. Nader onderzoek toont aan dat dit te wijten is aan de activering van fouten veroorzakende golfgetallen ten gevolge van de representatie van de puntbron als deltafunctie. In de tweede helft van het hoofdstuk bieden we een oplossing voor dit probleem in de vorm van een 'tapered-sinc' als bron.

In hoofdstuk 5 beschouwen we isotrope elastische golfvoortplanting met continue massageklonterde eindige elementen op tetraëders met expliciete tijdstappen. Deze elementen vereisen polynomen van hogere graad in hun binnenste om de nauwkeurigheid te behouden na massaklontering en waren bij het schrijven bekend tot en met graad 3. Onlangs zijn er nieuwe gevonden tot en met graad 4. Globale assemblage van de symmetrische stijfheidsmatrix is een natuurlijke benadering, maar vereist wel veel computergeheugen. Assemblage ter plekke, in de vorm van matrix-vectorproducten per element in elke tijdstap, vraagt veel minder geheugen. Met speciale uitdrukkingen voor lokale assemblage liep onze code rond 1,3 keer sneller voor graad 2 en 1,9 keer voor graad 3 op een eenvoudig homogeen testprobleem, gebruikmakend van 24 rekenkernen. Dit is vergelijkbaar met het akoestisch geval. Voor een realistischer probleem was de winst in efficiëntie een factor 2,5 voor graad 2 en 3 voor graad 3. Voor de laagste graad, het lineaire element, kunnen de uitdrukkingen voor zowel de globale als de lokale assemblage verder worden vereenvoudigd. In dat geval is globale assemblage efficiënter dan lokale. Van de drie polynoomgraden is het element van graad 3 het meest efficiënt in termen van nauwkeurigheid voor gegeven rekenkosten.

In hoofdstuk 6 beschouwen we kubieke Hermite polynomen als interpolanten in plaats van Legendre veeltermen. Door hun C^1 continuïteit kunnen zij een oplossing bieden voor de problemen van 'valse' golfgetallen die we in eerdere hoofdstukken hebben gezien bij conventionele interpolatieschema's. Resultaten laten acceptabel convergentiegedrag zien voor homogene media, maar de representatie schiet tekort bij discontinuïteiten in dichtheid die samenvallen met de zijden van de driehoeken. Aanpassing van de methode daarvoor is interessant toekomstig werk.

1

Introduction

With the still high demand for oil and gas, new reserves need to be found. In order to find new hydrocarbon pockets, we need sharper images of the subsurface. One way to do this is to use better numerical methods for forward modelling. Finite-element methods (FEM) are one such method, but their accuracy comes at the cost of increased compute expense. How is it possible to reduce this cost and adapt FEM to large scale problems of geophysics?

What will the energy scenario of the world look like, 100 years from now? Or, even in the year 2050? With increasing population that can afford amenities that require fuel, the demand for hydrocarbons is going to keep rising. Experts predict that despite the rise of renewable energies, hydrocarbon-based fuels are going to constitute 40-60 % of the energy supply until the year 2060, based on various geopolitical scenarios (Shell-scenarios, 2017).

However, most of the hydrocarbon reserves that can be discovered using existing methods have been found. The next step in exploration geophysics towards finding new reserves is to obtain sharper images of the subsurface of our planet. The process of subsurface imaging with reflection seismology uses seismic waves that travel through the earth. Thus, a numerical method that accurately approximates the equation that governs the physical motion of waves becomes important.

Seismic waves can be generated in the subsurface by an explosive device such as dynamite or a vibroseis on land, or an air gun when the experiment is being conducted in a marine environment as sketched in figure 1.1. As the signal travels through the earth, each layer reflects a portion of the energy and transmits and/or refracts the remaining portion of the signal. The signal is then picked up by a series of receivers, known as geophones or hydrophones depending on whether the experiment is being conducted on land or water. These receivers translate the energy from the seismic signal into an electric signal, which is then recorded as a seismic trace.

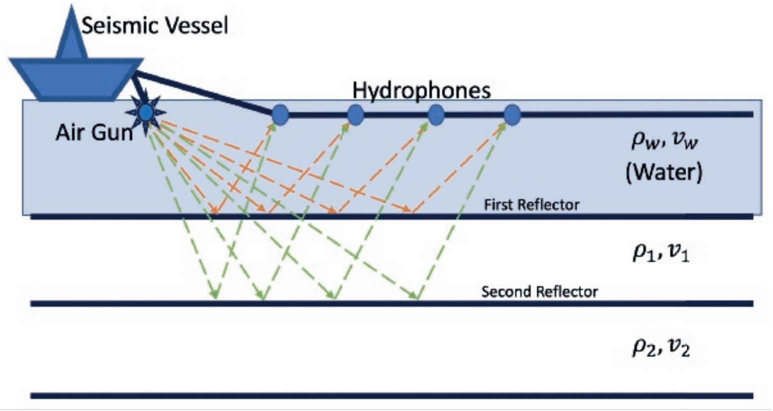


Figure 1.1: Seismic acquisition in a marine environment. Mechanical properties of the layers in the subsurface vary with depth. In this example, ρ is the property that leads to different velocities of the layers.

To convert the traces into information that can be used by interpreters who can identify the location of hydrocarbon reserves, mechanical properties such as elasticity and velocity have to be gleaned from the available experimental outputs. In order to do this, first an approximate velocity is made on which a numerical representation of the physical experiment is reproduced. The velocity model is then continuously updated until the difference between the outputs of the physical experiment and the numerical experiment are reduced to an acceptable tolerance limit. The process of seismic inversion has been schematically represented in figure 1.2

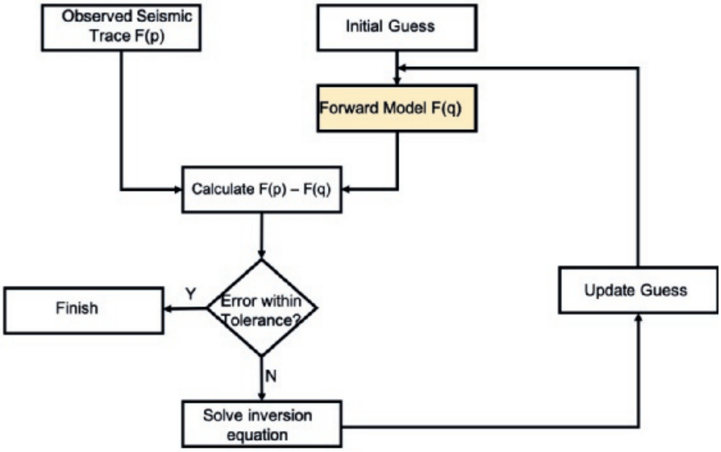


Figure 1.2: Process of seismic inversion. Forward modelling is the focus of this thesis.

Today, the finite-difference method is most popularly used in the industry because of ease of coding and avalability of legacy codes in the industry. This method is relatively easy to implement and parallelize. High-order differencing is often used to improve both computational and memory efficiency. For problems with sharp velocity contrasts, however, the finite-difference method is less attractive, because the solution is not sufficiently smooth across these contrasts and sharp interfaces between different materials cannot be easily represented on a finite-difference grid. In numerical simulations of wave propagation, this produces stair-casing, as for instance shown in figure 1 of (Mulder, 1996), reproduced here as figure 1.3. This may be a serious drawback for seismic applications in complex geologies (Zhebel et al., 2014).

Finite-element methods offer a remedy to these problems, since they can follow the boundary of sharply contrasting geological features (Marfurt, 1984). However, they are

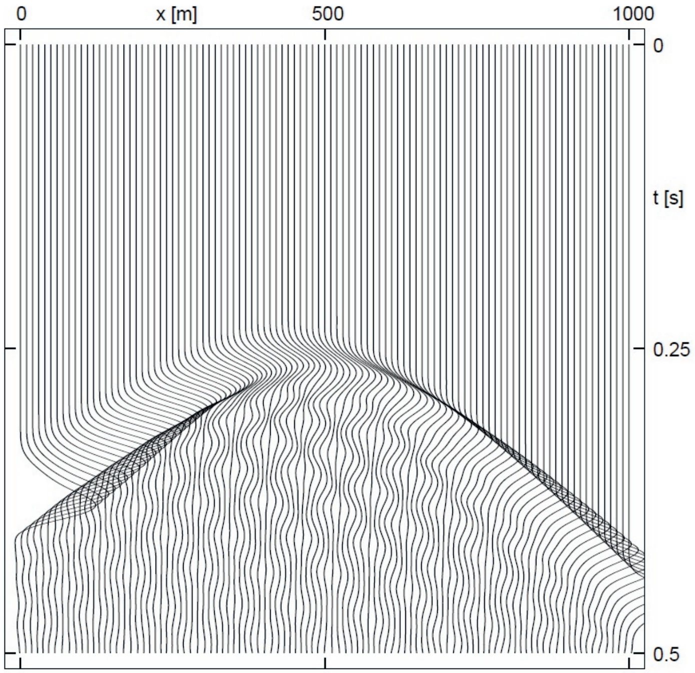


Figure 1.3: Seismic reflection traces for an interface dipped at 10°. The representation of the contrast in sound speed in a finite-difference code generates a crosshatches pattern.

not in popular use because of the additional cost of compute, and complications in parallellizing the code. This thesis focuses on trying to reduce the cost of finite-element methods so that they may applied to large scale problems of geophysical imaging.

Earlier works in this regard have looked at cost reduction by using efficient algorithms (Babuska et al., 1991; Farhat and Roux, 1991) and by using methods to reduce grid dispersion (De Basabe and Sen, 2007; Yue and Guddati, 2005). Quicker and automated mesh generation would also be one approach to make FE more suitable for seismic imaging, since the mesh may have to be updated in the inversion loop to follow the iterative updates of the model. Cost-efficient auto-mesh generation has been suggested by (Loge et al., 2007), although for the case of metals. Alternative finite-element methods such as discontinuous Galerkin (Favorskaya et al., 2016; Marcus J. Grote and Schotzau, 2006), XFEM (Julien Yvonnet, 2008; Wang et al., 2017; Yazid et al., 2009) and spectral finite-element methods (Patera, 1984; Seriani and Priolo, 1994) have been proposed for other applications. However, these were mainly designed for smaller problems in different fields and may not scale up properly for the larger problems of geophysics. The spectral-element method with Legendre-Gauss-Lobatto nodes on hexahedra does scale up and has become popular in the seismological community (Komatitsch and Tromp, 1999). Tetrahedra offer better meshing flexibility than hexahedra, but the construction of spectral elements is not as straightforward as for hexahedra. Elements of this type are considered in Chapter 5.

To avoid the cost of inverting the mass matrix and enable explicit time stepping, mass lumping is a common practice in FEM. This is done by performing row-summing operations that convert the fully populated mass matrix into a diagonal form. This operation is schematically depicted for a simple example that uses two triangular elements in figure 1.4. It has been proven that mass lumping with conventional FEM methods does not reduce the accuracy; it may even behoove the method for certain applications. In geophysical imaging, under certain circumstances, the first-order form of the wave equation

$$\rho^{-1}c^{-2}\partial_t p = \partial_x v_x + \partial_z v_z + f, \quad \rho\partial_t v_x = \partial_x p, \quad \rho\partial_t v_z = \partial_z p,$$

is more advantageous compute-wise, than the commonly used second-order formulation

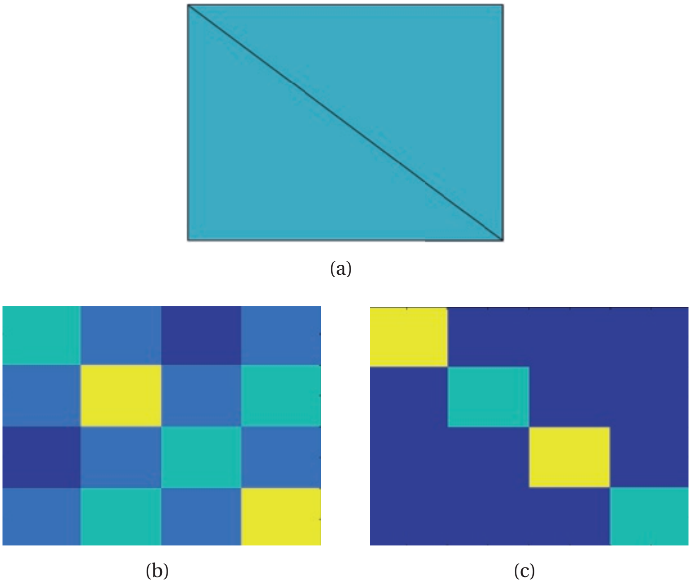
$$\rho^{-1}c^{-2}\partial_{tt} p = \partial_x(\rho^{-1}\partial_x p) + \partial_z(\rho^{-1}\partial_z p) + f'.$$

In the above equations, $\rho(x,z)$ and $c(x,z)$ are density and speed of sound and define the material at a position (x,z). At a given instance, the waveform can be described by particle velocities in the x- and z-directions, $v_x(t,x,z)$ and $v_z(t,x,z)$, and pressure $p(t,x,z)$ at time t. The source term is denoted by $f(t,x,z)$ or its time derivative $f'(t,x,z)$.

The first-order form has been explored for the finite-difference case by (Virieux, 1986) and (Virieux, 1984), and examined in further detail for FEM by (Ainsworth, 2014a,b). However, the first-order system of equations present a drawback: mass lumping cannot be used without a loss of accuracy.

In this thesis we explore different means by which this loss of accuracy can be addressed. The organisation of the thesis is as follows.

Chapter 2 gives an overview of the methods used in this thesis, specifically defect correction. It gives the generic form of how defect correction can be adapted as a preconditioner and how approximation errors are calculated.



(a)

(b)

(c)

Figure 1.4: Mass lumping is used to reduce a fully populated mass matrix into a diagonal matrix in order to reduce the cost of inversion, which is a computationally expensive process. (a) Two triangular elements (b) The assembled mass matrix of these two elements. In this figure, each cell represents one entry in the assembled mass matrix (c) The elements in (b) are collapsed into the diagonal by using the row-summing method.

Chapter 3 presents a dispersion and eigenvector analysis of various 1-D schemes. The numerical dispersion curve describes the error in the eigenvalues of the discrete set of equations. However, the error in the eigenvectors also play a role. For polynomial degrees above one and when considering a 1-D mesh with constant element size and constant material properties, a number of modes, equal to the maximum polynomial degree, are coupled. One of these is the correct physical mode that should approximate the true eigenfunction of the operator, the other are spurious and should have a small amplitude when the true eigenfunction is projected onto them. We analyze the behaviour of this error as a function of the normalized wave number in the form of the leading terms in its series expansion and find that this error exceeds the dispersion error, except for the lowest degree where the eigenvector error is zero.

The main observation in this chapter is that the simplest linear element in the first-order form has a fourth-order error. If this would generalize to 2 and 3 dimensions, it could potentially lead to a much faster modelling scheme. The price paid is that this accuracy is reduced to only second order after mass lumping. To avoid the cost of inverting the mass matrix but still preserve accuracy, defect correction is explored as a means to restore the accuracy. Then, the consistent mass matrix is approximately inverted with the lumped one as preconditioner. For the lowest-degree element on a uniform mesh, fourth-order accuracy in 1D can indeed be obtained with just a single iteration of defect correction. This doubling of the computational cost per time step is partly compensated by the larger allowable time step. Numerical 1-D tests confirm this behaviour. We briefly analyze the 2-D case, where the lowest-degree polynomial also appears to provide fourth-order accuracy with defect correction, if the mesh is structured in the form of squares divided into pairs of triangles and if material properties are constant.

Chapter 4 investigates the method in 2 space dimensions. In spite of the theoretical estimate of fourth-order accuracy with linear elements in first-order form, the practical implementation brought forth certain issues: the method did not show the expected convergence rates. The solution was extremely noisy when the conventional delta function was used to represent a point source. Masking the dispersive wavelengths is explored as a possible solution. First, a Gaussian function is attempted as the spatial smearing function for the source. This masks the unwanted wavelengths but also increases the error. Next, a tapered-sinc forcing function is designed as an alternative for the Gaussian mask. This reduces the noise in the solution while keeping the increase of the error small. Although the theoretical estimates of fourth-order are not achieved, convergence is recovered to higher levels than the mass lumped scheme. Unfortunately, the improved accuracy and larger allowable time step are not sufficient to compensate the additional cost when comparing to existing continuous mass-lumped finite elements in second-order form.

Chapter 5 considers continuous mass-lumped finite elements on tetrahedra with explicit time stepping for simulating isotropic elastic wave propagation. These elements require higher-order polynomials in their interior to preserve accuracy after mass lumping and were only known up to degree 3 at the time. Only recently have they been extended to degree 4 as well as simplified (Geevers et al., 2018). Global assembly of the symmetric stiffness matrix is a natural approach but requires large memory. Local assembly on the fly, in the form of matrix-vector products per element at each time step, has a much smaller memory footprint. In this chapter, the computational efficiency of the two approaches is compared.

In Chapter 6 we employ Hermite shape functions as a substitute for the hitherto-used Lagrangian basis functions. They strongly resemble the first-order form but avoid their instability in more than one space dimension by having an additional bubble function. They are tested on 1-D and 2-D problems. It is surmised that the higher-order restriction imposed by the C^1 continuity of the Hermite shape functions will improve the convergence properties. Whereas the method can deal with discontinuous material properties in 1D, assuming that the discontinuities occur at the element vertices, this is not true in more than one space dimension. In 2D, for instance, a discontinuity in density across the edge of a triangular element will cause the tangential velocity component to be discontinuous whereas the cubic Hermite elements will impose their continuity. This will cause some loss of accuracy, although this still may be acceptable in practice.

The thesis concludes with the final chapter, giving an outlook into the future direction this research might take.

REFERENCES

Ainsworth, M., 2014a. Dispersive behaviour of high order finite element schemes for the one-way wave equation. Journal of Computational Physics 259, 1–10.

Ainsworth, M., 2014b. Dispersive behaviour of high order finite element schemes for the one-way wave equation. Journal of Computational Physics 259, 1–10.

Babuska, I., Craig, A., Mandel, J., Pitkaranta, J., 1991. Efficient preconditioning for the p-version finite element method in two dimensions. SIAM Journal on Numerical Analysis 28 (3), 624–661.

De Basabe, J. D., Sen, M. K., 2007. Grid dispersion and stability criteria of some common finite-element methods for acoustic and elastic wave equations. Geophysics 72 (6), T81–T95.

Farhat, C., Roux, F.-X., 1991. A method of finite element tearing and interconnecting and its parallel solution algorithm. International Journal for Numerical Methods in Engineering 32 (6), 1205–1227.

Favorskaya, A., Petrov, I., Khokhlov, N., 2016. Numerical modeling of wave processes during shelf seismic exploration. Procedia Computer Science 96, 920 – 929.

Geevers, S., Mulder, W., van der Vegt, J., 2018. New higher-order mass-lumped tetrahedral elements for wave propagation modelling. SIAM Journal on Scientific Computing 40 (5), A2830–A2857.

Julien Yvonnet, H. Le Quang, Q.-C. H., 2008. An xfem/level set approach to modelling surface/interface effects and to computing the size-dependent effective properties of nanocomposites. Computational Mechanics, Springer Verlag 42, 704–712.

Komatitsch, D., Tromp, J., 1999. Introduction to the spectral-element method for 3-D seismic wave propagation. Geophysical Journal International 139 (3), 806–822.

Loge, R. E., Beringhier, M., Chastel, Y., Delannay, L., 2007. Reducing computational cost and allowing automatic remeshing in fem models of metal forming coupled with polycrystal plasticity. Vol. 908. pp. 387–392.

Marcus J. Grote, A. S., Schotzau, D., 2006. Discontinuous Galerkin finite element method for the wave equation. SIAM J. Numer. Anal., 44 (6), 2408–2431.

Marfurt, K. J., 1984. accuracy of finite element and finite difference modeling of the elastic wave equation. Geophysics 49 (5), 533–549.

Mulder, W. A., 1996. A comparison between higher-order finite elements and finite differences for solving the wave equation. In: Désidéri, J.-A., LeTalleca, P., Oñate, E., Périaux, J., Stein, E. (Eds.), Proceedings of the Second ECCOMAS Conference on Numerical Methods in Engineering, Paris, Sept. 9–13, 1996. John Wiley & Sons, Chichester, pp. 344–350.

Patera, A. T., 1984. A spectral element method for fluid dynamics: Laminar flow in a channel expansion. Journal of Computational Physics 54 (3), 468 – 488.

Seriani, G., Priolo, E., 1994. Spectral element method for acoustic wave simulation in heterogeneous media. Finite elements in analysis and design 16 (3), 337–348.

Shell-scenarios, 2017. New lens scenarios. http://www.shell.com/energy-and-innovation/the-energy-future/scenarios/new-lenses-on-the-future.html, accessed: 2017.

Virieux, J., 1984. SH-wave propagation in heterogeneous media: Velocity-stress finite-difference method. Geophysics 49, 1933–1942.

Virieux, J., 1986. P-SV wave propagation in heterogeneous media: Velocity-stress finite-difference method. Geophysics 51.

Wang, B., Liu, J., Gu, S., 2017. An xfem/level set strategy for simulating the piezoelectric spring-type interfaces with apparent physical background. Finite Elements in Analysis and Design 133, 62 – 75.

Yazid, A., Abdelkader, N., Abdelmadjid, H., 2009. A state-of-the-art review of the x-fem for computational fracture mechanics. Applied Mathematical Modelling 33 (12), 4269 – 4282.

Yue, B., Guddati, M. N., 2005. Dispersion-reducing finite elements for transient acoustics. The Journal of the Acoustical Society of America 118 (4), 2132–2141.

Zhebel, E., Minisini, S., Kononov, A., Mulder, W. A., 2014. A comparison of continuous mass-lumped finite elements with finite differences for 3-D wave propagation. Geophysical Prospecting 62 (5), 1111–1125.

2

Introduction to Methods

In this chapter, existing methods of mathematically approximating the wave equation are briefly reviewed, and concepts used in the thesis are explained in greater detail.

What is the 'defect' we are correcting in defect correction? How can we use Fourier analysis to derive the error behaviour of an approximation? These are some questions that will be answered in this chapter. It is written to aid a better understanding of the forthcoming chapters.

The finite-difference (FD) method is the most popular numerical approximation scheme used in subsurface imaging problems. They are relatively easy to code up and parallelize and provide fairly accurate results. Over the last decade, they have become the main tool for seismic imaging in complex geological models, in spite of their substantial computational cost.

Finite-elements are computationally more costly but their superior accuracy in the presence of complex topography can make them more efficient than finite differences. Finite-element (FE) methods are commonly derived for wave equation in second-order form, involving a mass matrix and stiffness matrix. In some cases, the first-order form may be more advantageous in terms of accuracy or in directly providing observable quantities.

Both the second- and first-order form lead to a mass matrix that has to be inverted at each time step. To avoid that cost, it can be replace by a diagonal mass matrix, with entries proportional to numerical quadrature weights. If this leads to an unacceptable loss of accuracy, an iterative method can be considered. The defect-correction principle shows that one iteration with a lumped mass matrix as preconditioner may suffice in some cases.

Dispersion analysis is a common tool for estimating numerical phase errors in finite-difference codes, but can also be used for finite elements to enable a quick comparison between the expected performance of various schemes.

FINITE-ELEMENT METHODS

Finite-elements methods employ the weak form of the differential equation. To focus the discussion, consider the acoustic wave equation in one space dimension:

$$\frac{1}{\rho c^2}\frac{\partial^2 p}{\partial t^2} = \frac{\partial}{\partial x}\left(\frac{1}{\rho}\frac{\partial p}{\partial x}\right), \tag{2.1}$$

with sound speed $c(x)$ and density $\rho(x)$. The pressure $p(t,x)$ is a function of time $t \in (0,T)$, between zero and maximum time T, and position $x \in \Omega$, in a given domain Ω with suitable conditions on the boundary $\partial\Omega$, for instance, zero Dirichlet boundary conditions $p(x) = 0$ for $x \in \partial\Omega$.

For the weak formulation, we assume that p belongs to the usual function space $U =$

$H_0^1(\Omega)$, the Sobolev space of functions with square-integrable derivatives and zero on $\partial\Omega$, and multiply equation (2.1) by a test function $q(x)$ belonging to a function space $V = L^2(\Omega)$, the space of square-integrable functions on Ω. The test function is assumed to vanish on the boundary $\partial\Omega$. We ignore the descretization in time and for brevity also drop the dependence on t. Multiplying the wave equation by q and integrating by parts then produces the weak form

$$\int_\Omega \frac{1}{\rho c^2} q \frac{\partial^2 p}{\partial t^2}\,\mathrm{d}x = -\int_\Omega \left(\frac{\partial q}{\partial x}\right)\frac{1}{\rho}\left(\frac{\partial p}{\partial x}\right)\mathrm{d}x, \tag{2.2}$$

to be satisfied for all $q(x) \in L^2(\Omega)$.

Equation (2.1) is called the second-order form of the wave equation. Its first-order form is

$$\frac{1}{\rho c^2}\frac{\partial p}{\partial t} = \frac{\partial v}{\partial x}, \quad \rho\frac{\partial v}{\partial t} = \frac{\partial p}{\partial x}, \tag{2.3}$$

with velocity $v(t,x)$. Although the second- and first-order form are the same, this is no longer true after discretization (Brezzi and Fortin, 1991; Joly, 2003). A weak form (2.3) is

$$\int_\Omega \frac{1}{\rho c^2} q \frac{\partial p}{\partial t}\,\mathrm{d}x = -\int_\Omega v\frac{\partial q}{\partial x}\,\mathrm{d}x, \quad \int_\Omega \rho u \frac{\partial v}{\partial t} = \int_\Omega u\frac{\partial p}{\partial x}\,\mathrm{d}x,$$

with $p \in U = H_0^1(\Omega)$ and $v \in V = L^2(\Omega)$. and test functions $q(x) \in U$ and $u(x) \in V$. This is the primal formulation. The dual formulation

$$\int_\Omega \frac{1}{\rho c^2} q \frac{\partial p}{\partial t}\,\mathrm{d}x = \int_\Omega q\frac{\partial v}{\partial x}\,\mathrm{d}x, \quad \int_\Omega \rho u \frac{\partial v}{\partial t} = -\int_\Omega p\frac{\partial u}{\partial x}\,\mathrm{d}x,$$

basically swaps the spaces U and V and involves different regularity requirements.

The finite-element discretization proceeds with a discrete approximation U^h of U and also V^h of V for the first-order form. The domain Ω is partitioned into elements, intervals of finite length in the 1-D case. Common choices are triangles or quadrilaterals in two dimensions and tetrahedra or hexahedra in three dimensions. The simplexes offer more meshing flexibility than the blocks. Piecewise polynomials are often chosen for the basis and test functions. The Galerkin approach uses the same for both but they can be different, as in the Petrov-Galerkin method required for the first-order formulation.

Given an expansion of the solution into basis functions,

$$p(x) = \sum_{j=1}^{N} p_j \phi_j(x),$$

and a similar expansion for the test functions

$$q(x) = \sum_{j=1}^{N} q_j \phi_j(x),$$

substitution into equation (2.2) and stationarity for all q_j leads to

$$\mathbf{M}\frac{\partial^2 \mathbf{p}}{\partial t^2} = -\mathbf{K}\mathbf{p},$$

with a vector $\mathbf{p}$ containing the degree of freedom p_j and with mass matrix $\mathbf{M}$ and stiffness matrix $\mathbf{K}$ having entries

$$M_{i,j} = \int_\Omega \frac{1}{\rho c^2}\phi_i(x)\phi_j(x)\,\mathrm{d}x, \quad K_{i,j} = \int_\Omega \frac{1}{\rho}\left(\frac{\partial \phi_i(x)}{\partial x}\right)\left(\frac{\partial \phi_j(x)}{\partial x}\right)\mathrm{d}x.$$

Likewise, for the first-order form, we can consider expansions into basis functions

$$p(x) = \sum_{j=1}^{N_p} p_j \phi_j(x), \quad v(x) = \sum_{j=1}^{N_v} v_j \psi_j(x),$$

where $\phi_j(x)$ and $\psi_j(x)$ are generally different. The latter are vectors in more than one dimension. With similar expansions

$$q(x) = \sum_{j=1}^{N_p} q_j \phi_j(x), \quad u(x) = \sum_{j=1}^{N_v} u_j \psi_j(x),$$

for the test functions, we obtain for the primal mixed formulation

$$\mathbf{M}^p \frac{\partial \mathbf{p}}{\partial t} = -\left(\mathbf{D}^p\right)^\top \mathbf{v}, \quad \mathbf{M}^v \frac{\partial \mathbf{v}}{\partial t} = \mathbf{D}^p\,\mathbf{p}.$$

The two mass matrices have elements

$$M^p_{i,j} = \int_\Omega \frac{1}{\rho c^2}\phi_i(x)\phi_j(x)\,\mathrm{d}x, \quad M^v_{i,j} = \int_\Omega \rho\psi_i(x)\psi_j(x)\,\mathrm{d}x.$$

The derivative operator has entries

$$D^p_{i,j} = \int_\Omega \psi_i(x) \frac{\partial}{\partial x} \phi_j(x)\, \mathrm{d}x,$$

and the superscript $(\cdot)^\mathsf{T}$ denotes the transpose. A natural choice for the space V containing ψ is the space of derivatives of U.

In general, the first-order form provides a discretization that is different from the one obtained with the second-order form. An interesting exception is the spectral-element method for polynomial basis functions up to degree p with mass lumping on Legendre-Gauss-Lobatto nodes. The latter make the diagonal values of the lumped mass matrix agree with numerical quadrature weights, up to a constant factor. If the first-order form has U as the space of piecewise continuous polynomials of degree p and V the space of piecewise *discontinuous* polynomials of the same degree, then the discrete versions for the first- and second-order form are numerically the same (Cohen, 2002). This is a simple consequence from the fact that the numerical quadrature weights are exact for polynomials up to degree $2p-1$. Numerical quadrature for the derivative matrix is therefore exact and involves basis functions evaluated at nodes. If Lagrange interpolants are used, they are either one or zero over there and only the derivatives at the nodes remain. The lumped mass matrix for the velocities is decoupled from the neighbouring elements by assumption and readily inverted. Elimination from the velocities in an element leads to an expression that equals the result of numerical quadrature applied to the contribution to the stiffness matrix of that element and the latter is exact. Therefore, the first- and second-order form are numerically the same. This result also holds in more than one space dimension on rectangular elements if the Cartesian product of the 1-D polynomials is used for the basis functions. It also is valid for the degree-1 mass-lumped finite element on triangles and tetrahedra.

An example of the dual mixed formulation is the lowest-order Raviart-Thomas element RT_0 (Raviart and Thomas, 1977) on the triangle for the fluxes across the edges, which are the normal components of the velocity $\mathbf{v}$ in our case, and the piecewise constant element P_0 for the pressure p. The velocities are piecewise linear per triangle, constant on the edges and continuous across the edges. This follows the natural choice for the dual formulation of $\phi \in V$ and $\boldsymbol{\psi} \in U$ with V the space of derivatives of U.

In this thesis, a primal formulation will be investigated with a deliberately wrong choice for the space V. When only the spatial part is considered in more than one space dimension, this results in an unstable discretization of the corresponding elliptic operator. Also, the scheme imposes continuity of the tangential velocity component across elements, which should not hold if the density is discontinuous across an element edge.

The motivation lies in the observation that the lowest-degree element with continuous piecewise linear elements for both pressure and velocity has fourth-order accuracy in 1D, as shown in Chapter 3. This would potentially offer a far less costly alternative to the existing fourth-order mass-lumped scheme on tetrahedra (Chin-Joe-Kong et al., 1999), also considered in Chapter 5. One could argue that a similar violation of (dis)continuity requirements is made in the higher-order finite-difference methods that are widely used for seismic modelling and could be acceptable if it pays off in terms of overall accuracy at a given compute cost. Unfortunately, that is not case, as turns out in Chapter 4.

Stable versions of this element are the MINI element (Arnold et al., 1984), with adds bubble functions to the velocity components, and the cubic Hermite element, which is the subject of Chapter 6.

Mass lumping and numerical quadrature

Piecewise polynomials are a common choice for finite-element basis functions. Mass lumping, in which the full mass matrix is replaced by a diagonal matrix obtained from its row sums, is attractive for explicit time stepping schemes, as it avoids the inversion of a large sparse matrix. The Legendre-Gauss-Lobatto nodes lead to a mass matrix proportional to numerical quadrature weights. For polynomials of degree p, these are exact up to degree $2p-1$, whereas only $2p-2$ is required in the second-order form (Ciarlet, 1978). Cartesian products can be used for block-type elements in more than one space dimension. This forms the basis of the spectral-element method, which has found widespread use in the seismological community for the modelling and inversion of seismic waves (Komatitsch and Tromp, 1999). Another option are the Chebyshev-Gauss-Lobatto nodes without (Patera, 1984; Seriani et al., 1992) or with a weighted scalar product, further examined in Chapter 3.

Mass lumping for triangles and tetrahedra is less straightforward. To avoid zero or negative weights, which will lead to an unstable time stepping scheme, polynomials of higher degrees have to be added (Cohen et al., 2001, 1995; Tordjman, 1995) to the interior. For triangles, elements up to degree 9 have been found, for tetrahedra up to degree 4 (Chin-Joe-Kong et al., 1999; Cohen et al., 1995; Cui et al., 2017; Geevers et al., 2018a, 2019, 2018b; Liu et al., 2017; Mulder, 1996, 2001, 2013). The disadvantage of this approach is that no general recipe for the construction of elements of arbitrary degrees has been found. Xu (2011), building on earlier work of Helenbrook (2009), actually established the non-existence of a strict Gauss-Lobatto cubature rule for the unit triangle and also that a minimum number of nodes are required for such a rule.

A systematic approach does exist for Discontinuous Galerkin methods. These naturally lead to a block-diagonal mass matrix that is easily inverted. There computational cost, however, is similar to the older mass-lumped elements on tetrahedra and far higher than the newer ones (Geevers et al., 2018a, 2019).

Another route to a systematic construction of arbitrary-order spectral elements on the triangle, is to map the rectangle, on which such elements are known, to the triangle (Dubiner, 1991; Koornwinder, 1975; Samson et al., 2012). Additional steps need to be introduced to get rid of certain artefacts arising from the use of tensorial Legendre-Gauss-Lobatto points on a square. Similar work has been done in 3D by Li and Wang (2010) where they used a collapsed coordinate transform between a cube and a tetrahederon. They use tensor products of 1D polynomials on the cube. To get around the singularity in this transform, they use Gauss-Lobatto in one direction and Gauss-Radau in the other two directions. It remains to be seen if these methods can be used for the efficient simulation of 3-D wave propagation.

An alternative is the work of Li et al. (2008), who study problems of trigonometric approximation on a hexagon and a triangle using the discrete Fourier transform and orthogonal polynomials of two variables. They deduce the analysis on a triangle based on the discrete Fourier analysis of a regular hexagon. Interestingly, a trigonometric Lagrange interpolation on a triangle is shown to satisfy an explicit compact formula, which is equivalent to the polynomial interpolation on a planer region bounded by a Steiner hypocycloid or deltoid, i.e., similar to a higher-order interpolation. They also derive a Gauss cubature on the deltoid, using the first two Chebyshev polynomials as the orthogonal bases. Building on these results, Munthe-Kaas (2006); Ryland and Munthe-Kaas (2011) and Munthe-Kaas et al. (2012) explore the use of multivariate Chebyshev polynomials for spectral elements on triangles, but they end up with deltoids rather than triangles. These can be deformed to triangles by a non-linear map. Alternatively, one can keep the deltoids an patch them together with an overlap, using the edges of the inscribed triangle to connect neighbouring elements. The generalization to spectral methods on tetrahedra should be feasible, but would require a generalization of the fast discrete triangle transform (Püschel and Rötteler, 2004) to the tetrahedron.

DEFECT CORRECTION

In some cases, mass lumping leads to a less accurate method than with the full or consistent mass matrix. Then, an iterative approach for the inversion of the mass matrix can be considered, preconditioned with its mass-lumped version, as will be done in Chapters 3 and 4. The number of iterations can be small, as shown by the defect correction principle (Stetter, 1978) reviewed next.

Consider the linear problem $Lu = f$, with a linear operator L acting on the solution u and producing a right-hand side f. An approximation $\tilde{L}$ of L is assumed to be more easy to invert and provides an approximate solution $u_0 = \tilde{L}^{-1} f$. Defect correction defines the defect $d_0 = Lu_0 - f$ as the error in the equation. The corresponding error in the solution can be estimated from $d_0 = Lu_0 - f = L(u_0 - u)$ by

$$u_0 - u = L^{-1} d_0 \simeq \tilde{L}^{-1} d_0.$$

The approximate correction for the defect leads to the solution $u_1 = u_0 - \tilde{L}^{-1} d_0$. Iterative application of the procedure gives

$$u_{i+1} = u_i - \tilde{L}^{-1} d_i, \quad i = 0, 1, \dots,$$

with defect $d_i = Lu_i - f$. We can rewrite the expression as

$$u_{i+1} = (I - \tilde{L}^{-1} L) u_i + \tilde{L}^{-1} f = G u_i + u_0 = \sum_{k=0}^{i+1} G^k u_0,$$

where $G = I - \tilde{L}^{-1} L$. Then,

$$\lim_{i \to \infty} u_i = (I - G)^{-1} u_0 = L^{-1} \tilde{L} \tilde{L}^{-1} f = L^{-1} f = u,$$

if the spectral radius of G is less than one, demonstrating convergence.

The method is the same as the classic iterative scheme

$$u_{i+1} = u_i + \tilde{L}^{-1} r_i, \quad i = -1, 0, 1, \ldots$$

with residuals $r_i = f - Lu_i$ and $u_{-1} = 0$. Note that $r_i = -d_i$ for $i \geq 0$. From this point of view, there seems to be no reason to consider defect correction as a separate notion. That changes if the operators $\tilde{L}^h$ and L^h are lower- and higher-order discrete approximations of the same operator L.

Let L^h produce a solution u^h of order p, with an error of $O(h^p)$ in a discretization parameter h, for instance, a time step in a discretization of an ordinary differential equation or a grid spacing in a partial differential equation such as the wave equation in the frequency domain. The discrete operator $\tilde{L}^h$ has a lower order $\tilde{p} < p$. Then, Stetter (1978) shows that the error in the iterates behaves as

$$\| u_i^h - I^h u \| = O\left(h^{\min((i+1)\tilde{p}, p)} \right).$$

Here, I^h projects the continuum solution u on to the discrete solution u^h. This expression provides the number of iterations required to obtain a solution with an iteration error of the order of the discretisation error.

First of all the discretised problem is defined as $L^h u^h = f^h$. Next, functions I^h and I_h map the continuous domain to the discretised and vice-versa respectively. Then, the defect has to be defined in the $_h$ domain. This can be done as: $d^h = I^h \tilde{L} I_h f^h$. in other words, $d^h = \tilde{L}^h f^h$. It is then possible to proceed with iterations in the discretised domain similar to the continuous domain as described before.

DISPERSION ANALYSIS

The dispersion caused by a numerical scheme can be quantified by the errors in eigenvalues. (Lele, 1992) evaluates the performance of the scheme in terms of resolving efficiency. On a periodic grid defined as $x = x_0 + jh$, the Fourier symbol of the nodal variable can be written as

$$f(x) = \sum_{k=-N/2+1}^{k=N/2} \hat{f}_k \exp\left(\frac{2\pi i k x}{L} \right). \tag{2.4}$$

It is convenient to define a scaled wavenumber $w = 2\pi kh/L = 2\pi k/N$ with domain $[-\pi,\pi]$, which would scale the coordinate as $s = x/h$. The exact first derivative of eq. 2.4 with respect to the scaled coordinates would be $\hat{f}'_k = iw\hat{f}_k$. This can be compared against the Fourier coefficients from the numerical method. For example, a central difference scheme would have a Fourier symbol $(\hat{f}'_k)_{fd} = iw'\hat{f}_k$. The range of wavenumbers over which the modified wavenumber approximates the exact differentiation within an acceptable limit of tolerance will be the set of well-resolved waves. The shortest resolved wave w_f is sensitive to the scheme. The factor $r_1 = 1 - w_f/\pi$ represents the fraction of poorly resolved waves; $e_1 = 1 - r_1$ can be regarded as the resolving efficiency.

While finite differencing techniques and their errors have been studied extensively, the dispersion behaviour of finite-element approximations has received less attention (Ainsworth and Wajid, 2009; Marfurt, 1990; Mulder, 1999). Dispersion analysis shows that the dispersion in a mass lumped FE scheme leads to a phase lag, as opposed to a full mass matrix that leads to a phase lead. (Ainsworth, 2014) compare the propagation of physical and various discrete waves for one-way wave equation. Results show that the spectral, Galerkin and discrete Galerkin (DG) methods all have spurious modes. Unlike the second-order case, the spectral-element method performs worse than Galerkin methods. There is a two order difference in accuracy between Galerkin methods in comparison to DG — it is better for odd and worse for even orders of interpolation.

However, (Mulder, 1999) shows that the proper way to evaluate errors in higher-order schemes should not only include eigenvalues, but also take into account the eigenvectors associated with each mode. Waves that are generally viewed as spurious or non-physical modes in higher-order approximations are given a new perspective. It is shown that for elements of degree M, each block of M modes has M eigenvectors, leading to NM distinct wavenumbers for N element), and each eigenpair can be matched to one eigenpair of the exact operator. This is important in order to properly understand the arrangement of eigenvalues to avoid "branches" in the dispersion curves. Some of these results are used for comparison in Chapter 3, where dispersion as well as eigenvector errors for the first-order form are studied, and in Chapter 6.

CONCLUSIONS

Numerical Green functions based on the acoustic or elastic wave equation have become a standard tool in industrial seismic processing and imaging, in spite of their cost. Finite-difference methods are the most common in exploration seismics, whereas the spectral-element method has gained wide popularity in the seismological community, not in the least because of freely available software. The finite-element method is less common in exploration seismics, because if problems with automatic meshing of complex geological structures and because of its perceived computational cost. That motivates the search for more efficient finite-element schemes.

REFERENCES

Ainsworth, M., 2014. Dispersive behaviour of high order finite element schemes for the one-way wave equation. Journal of Computational Physics 259, 1–10.

Ainsworth, M., Wajid, H., 2009. Dispersive and dissipative behavior of the spectral element method. SIAM Journal on Numerical Analysis 47 (5), 3910–3937.

Arnold, D. N., Brezzi, F., Fortin, M., 1984. A stable finite element for the Stokes equations. Calcolo 21, 337–344.

Brezzi, F., Fortin, M., 1991. Mixed and Hybrid Finite Element Methods. Springer Series in Computational Mathematics, 15. Springer.

Chin-Joe-Kong, M. J. S., Mulder, W. A., van Veldhuizen, M., 1999. Higher-order triangular and tetrahedral finite elements with mass lumping for solving the wave equation. Journal of Engineering Mathematics 35, 405–426.

Ciarlet, P. G., 1978. The finite element method for elliptic problems. North-Holland.

Cohen, G., Joly, P., Roberts, J. E., Tordjman, N., 2001. Higher order triangular finite elements with mass lumping for the wave equation. SIAM Journal on Numerical Analysis 38 (6), 2047–2078.

Cohen, G., Joly, P., Tordjman, N., 1995. Higher order triangular finite elements with mass lumping for the wave equation. In: Cohen, G., Bécache, E., Joly, P., Roberts, J. E. (Eds.), Proceedings of the Third International Conference on Mathematical and Numerical Aspects of Wave Propagation. SIAM, Philadelphia, pp. 270–279.

Cohen, G. C., 2002. Higher-Order Numerical Methods for Transient Wave Equations. Springer.

Cui, T., Leng, W., Lin, D., Ma, S., Zhang, L., 2017. High order mass-lumping finite elements on simplexes. Numerical Mathematics: Theory, Methods and Applications 10 (2), 331–350.

Dubiner, M., 1991. Spectral methods on triangles and other domains. Journal of Scientific Computing 6 (4), 345–390.

Geevers, S., Mulder, W., van der Vegt, J., 2018a. New higher-order mass-lumped tetrahedral elements for wave propagation modelling. SIAM Journal on Scientific Computing 40 (5), A2830–A2857.

Geevers, S., Mulder, W., van der Vegt, J., 2019. Efficient quadrature rules for computing the stiffness matrices of mass-lumped tetrahedral elements for linear wave problems. SIAM Journal on Scientific Computing 41 (2), A1041–A1065.

Geevers, S., Mulder, W. A., van der Vegt, J. J. W., 2018b. Dispersion properties of explicit finite element methods for wave propagation modelling on tetrahedral meshes. Journal of Scientific Computing 77 (1), 372–396.

Helenbrook, B., 2009. On the existence of explicit \$hp\$-finite element methods using Gauss-Lobatto integration on the triangle. SIAM Journal on Numerical Analysis 47 (2), 1304–1318.

Joly, P., 2003. Variational methods for time dependent wave propagation problems. In: Topics in Computational Wave Propagation: Direct and Inverse Problems. Lecture Notes in Computational Science and Engineering, Volume 31. Springer Berlin Heidelberg, pp. 201–264.

Komatitsch, D., Tromp, J., 1999. Introduction to the spectral-element method for 3-D seismic wave propagation. Geophysical Journal International 139 (3), 806–822.

Koornwinder, T., 1975. Two-variable analogues of the classical orthogonal polynomials. In: Askey, R. A. (Ed.), Theory and application of special functions. Academic Press, New York, pp. 435–495.

Lele, S. K., 1992. Compact finite difference schemes with spectral-like resolution. Journal of Computational Physics 103 (1), 16–42.

Li, H., Sun, J., Xu, Y., 2008. Discrete Fourier analysis, cubature, and interpolation on a hexagon and a triangle. SIAM Journal on Numerical Analysis 46 (4), 1653–1681.

Li, H., Wang, L.-L., 2010. A spectral method on tetrahedra using rational basis functions. International Journal of Numerical Analysis and Modeling 7 (2), 330–355.

Liu, Y., Teng, J., Xu, T., Badal, J., 2017. Higher-order triangular spectral element method with optimized cubature points for seismic wavefield modeling. Journal of Computational Physics 336, 458–480.

Marfurt, K. J., 1990. Analysis of higher order finite-element methods. In: Kelly, K. R., Marfurt, K. J. (Eds.), Numerical Modeling of Seismic Wave Propagation, Geophysics Reprint Series No. 13. Society of Exploration Geophysicists, pp. 516–520.

Mulder, W., 1999. Spurious modes in finite-element discretizations of the wave equation may not be all that bad. Applied Numerical Mathematics 30 (4), 425–445.

Mulder, W. A., 1996. A comparison between higher-order finite elements and finite differences for solving the wave equation. In: Désidéri, J.-A., LeTallec, P., Oñate, E., Périaux, J., Stein, E. (Eds.), Proceedings of the Second ECCOMAS Conference on Numerical Methods in Engineering. John Wiley & Sons, Chichester, pp. 344–350.

Mulder, W. A., 2001. Higher-order mass-lumped finite elements for the wave equation. Journal of Computational Acoustics 9 (2), 671–680.

Mulder, W. A., 2013. New triangular mass-lumped finite elements of degree six for wave propagation. Progress In Electromagnetics Research 141, 671–692.

Munthe-Kaas, H. Z., 2006. On group Fourier analysis and symmetry preserving discretizations of PDEs. Journal of Physics A: Mathematical and General 39 (19).

Munthe-Kaas, H. Z., Nome, M., Ryland, B. N., 2012. Through the kaleidoscope: Symmetries, groups and Chebyshev approximations from a computational point of view. In: Cucker, F., Krick, T., Pinkus, A., Szanto, A. (Eds.), Foundations of Computational Mathematics, vol. 403. Cambridge University Press.

Patera, A. T., 1984. A spectral element method for fluid dynamics: Laminar flow in a channel expansion. Journal of Computational Physics 54 (3), 468–488.

Püschel, M., Rötteler, M., 2004. Cooley-Tukey FFT like algorithm for the discrete triangle transform. In: Proceeding of Digital Signal Processing Workshop and the 3rd IEEE Signal Processing Education Workshop. pp. 158–162.

Raviart, P. A., Thomas, J. M., 1977. A mixed finite element method for second order elliptic problems. In: Galligani, I., Magenes, E. (Eds.), Mathematical Aspects of Finite Element Methods, Lecture Notes in Mathematics, vol. 606. Springer, pp. 292–315.

Ryland, B. N., Munthe-Kaas, H. Z., 2011. On multivariate Chebyshev polynomials and spectral approximations on triangles. In: Hesthaven, J., Rønquist, E. (Eds.), Spectral and High Order Methods for Partial Differential Equations. Lecture Notes in Computational Science and Engineering, vol. 76. Springer, Berlin, Heidelberg.

Samson, M. D., Li, H., Wang, L., 2012. A new triangular spectral element method I: implementation and analysis on a triangle. Numerical Algorithms 64 (3), 519–547.

Seriani, G., Priolo, E., Carcione, J., Padovani, E., 1992. High-order spectral element method for elastic wave modeling. SEG Technical Program Expanded Abstracts 11, 1285–1288.

Stetter, H. J., 1978. The defect correction principle and discretization methods. Numerische Mathematik 29 (4), 425–443.

Tordjman, N., 1995. Élements finis d'order élevé avec condensation de masse pour l'equation des ondes. Ph.D. thesis, L'Université Paris IX Dauphine.

Xu, Y., 2011. On Gauss-Lobatto integration on the triangle. SIAM Journal on Numerical Analysis 49 (2), 541–548.

3

Defect Correction and FEM : Improved accuracy with reduced cost in 1D

Parts of this chapter have been published in Journal of Computational Physics, 322, pp.689-707 (Shamasundar and Mulder, 2016)

As explained briefly in the introduction, finite-element discretisations of the acoustic wave equation in the time domain often employ mass lumping to avoid the cost of inverting a large sparse mass matrix. Unfortunately, for a first-order system of equations, mass lumping destroys the super-convergence of numerical dispersion for odd-degree polynomials. In this chapter, we consider defect correction as a means to restore the accuracy. We adapt the defect correction method to FEM by solving the consistent mass matrix with the lumped one as preconditioner. For the lowest-degree element, fourth-order accuracy in 1D can be obtained with just a single iteration of defect correction. In this chapter, we analyse the behaviour of the error in eigenvectors as a function of the normalized wavenumber in the form of leading terms in its series expansion and find that this error exceeds the dispersion error, except for the lowest degree where the eigenvector error is zero. We also present results of numerical experiments that confirm this analysis.

INTRODUCTION

Numerical simulation of the wave equation in the time domain can be accomplished by a suitable finite-difference method. This method is relatively easy to implement and parallelize. High-order differencing is often used to improve both computational and memory efficiency. For problems with sharp velocity contrasts, however, the finite-difference method is less attractive, because the solution is not sufficiently smooth across these contrasts and sharp interfaces between different materials cannot be easily represented on a finite-difference grid. In numerical simulations of wave propagation, this produces stair-casing, as shown in figure 1 of (Mulder, 1996). This may be a serious drawback for seismic applications in complex geologies (Zhebel et al., 2014).

The finite-element method can, in principle, overcome these difficulties if element faces follow sharp contrasts. Mass lumping is usually applied to avoid the cost of inverting a large sparse consistent mass matrix. However, mass lumping may cause a loss of spatial accuracy. This is not true for the second-order formulation of the wave equation. The choice of Legendre polynomials and Gauss-Lobatto points actually leads to better accuracy after mass lumping, as proven in the Appendix of (Mulder, 1999). These results were confirmed later in (Ainsworth, 2004) and (Ainsworth and Wajid, 2009).

For variable-density acoustics as well as the elastic system of wave equations, a first-order formulation can sometimes be more convenient. In the 1-D acoustic case, this provides a pair of equations in the pressure and in the particle velocity. The usual finite-element discretization involves different spaces for each, for instance, H^1 and L^2. If the solution is represented by polynomials with and without continuity across elements, the first-order formulation can be made identical to the second-order one (Cohen, 2002, section 13.4.2). Here, we adopt the naive approach of discretizing each of the pair of first-order equations for pressure and velocity with the same spectral-element method.

Unfortunately, the application of mass lumping to first-order differentiation with Legendre-Gauss-Lobatto (LGL) points leads to a decrease of accuracy (Ainsworth, 2014). In this paper, we propose to use defect correction (Stetter, 1978) to compensate for this loss of accuracy. Defect correction employs a lower-order discretization of a problem as a preconditioner for a higher-order discretization. The gain in accuracy per iteration is the same as that of the lower order (Stetter, 1978, section 7). If, for instance, an operator with fourth-order accuracy is preconditioned by one with second-order accuracy, the first step provides an approximate solution with second-order accuracy. One additional iteration already leads to fourth-order accuracy if the numerical solution is sufficiently well resolved by the discretization to lie in the asymptotic regime where it converges.

In the work of (Wathen, 1987), the diagonal of the mass matrix was used as a preconditioner to the consistent mass matrix. Here, we will show that method to be less effective.

To investigate the properties of the proposed scheme, we perform the same type of dispersion analysis as in (Mulder, 1999), but now on a discrete operator that represents the first instead of the second derivative in space. If the polynomial basis has degree M, a discrete Fourier transform of the discrete operator results in a matrix with small $M \times M$ blocks, for which eigenvalues and eigenvectors can be determined, numerically or symbolically or as a series approximation for small wavenumbers. Each of the M eigenmodes deals with one separate point on the dispersion curve. Their interaction can be characterized as 'spurious' and was quantified in (Mulder, 1999) by considering the eigenvector errors. An alternative approach was followed by (Ainsworth, 2004; Cohen, 2002; Thompson and Pinsky, 1994), where the eigenvectors were constructed directly and then the eigenvalues that constitute the dispersion curve were determined.

We examined the numerical dispersion curves and error behaviour for four schemes with polynomial basis functions: the standard elements with equidistant nodes (EQUI), the Legendre-Gauss-Lobatto points (LGL), the Chebyshev-Gauss-Lobatto nodes without a weighting function (Patera, 1984) (CGL) and with (CGLw). Section 5.2 describes the various discretizations and how we apply defect correction and analyze the numerical dispersion. Section 5.3 lists the leading error terms in the dispersion curves for the consistent mass matrix, for the lumped one, and after one iteration of defect correction. It includes estimates of the error in the eigenvectors. Numerical experiments for simple differentiation as well as for 1-D wave propagation on a periodic mesh are included. In Section 3.4, we apply Fourier analysis on a periodic grid to obtain error estimates for the 2-D case, both for square bilinear elements and for squares cut onto half to obtain a regular mesh of triangles. Section 3.5 summarizes our findings.

METHOD

ELEMENTS

A first-order formulation of the acoustic wave equation is

$$\rho\frac{\partial v}{\partial t}=\frac{\partial p}{\partial x},\quad \frac{1}{\rho c^2}\frac{\partial p}{\partial t}=\frac{\partial v}{\partial x},$$

with particle velocity $v(t,x)$ and pressure $p(x,t)$, here without the usual minus sign, as a function of time t and position x. The density $\rho(x)$ and sound speed $c(x)$ will be taken as constant for the purpose of analysis. We will not consider time stepping errors and only concentrate on the spatial discretization. Consider N elements bounded by positions $x_j = x_0 + jh_j$, $j = 0,\dots,N$. Each element has $M+1$ nodes at relative positions ζ_k, $k = 0,\dots,M$, with $\zeta_0 = -1$ and $\zeta_M = 1$. Their corresponding global positions are $x_{j,k} = x_j + \frac{1}{2}(\zeta_k+1)jh_j$. In the periodic case, the solution on x_N is the same as on x_0. The number of degrees of freedom is $N_{\mathrm{dof}} = MN$ on a periodic grid both for the particle velocity and pressure.

For the finite-element basis functions $\phi_k(\zeta)$, we take the Lagrange interpolating polynomials of degree M relative to the nodes, so $\phi_k(\zeta_l) = \delta_{k,l}$, the Kronecker delta. In each element, we have a local mass matrix A and first-derivative matrix D, each with entries

$$A_{k,l}=\int_{-1}^{1}\omega(\zeta)\phi_k(\zeta)\phi_l(\zeta)\,\mathrm{d}\zeta,\quad D_{k,l}=\int_{-1}^{1}\omega(\zeta)\phi_k(\zeta)\frac{\mathrm{d}}{\mathrm{d}\zeta}\phi_l(\zeta)\,\mathrm{d}\zeta.$$

The local lumped mass matrix, $A^{\mathrm{L}}_{k,l} = \delta_{k,l}\sum_{l=0}^{M} A_{k,l}$ is a diagonal matrix with values proportional to quadrature weights. We consider four choices for the nodes: the standard element with equidistant nodes $x_k = k/M$, $k = 0,1,\dots,M$ (EQUI); the Legendre-Gauss-Lobatto points (LGL) that are the zeros of $(1-\zeta^2)P'_M(\zeta)$, the Chebyshev-Gauss-Lobatto points $\zeta_k = -\cos(\pi k/M)$ with an unweighted scalar product (CGL) and with the weighting function $\omega(\zeta) = 1/\sqrt{1-\zeta^2}$ (CGLw). Except for CGLw, the weighting function $\omega(\zeta) = 1$. Numerical quadrature with weights $A^{\mathrm{L}}_{k,k}/\sum_{k=0}^{M} A^{\mathrm{L}}_{k,k}$ is exact for polynomials up to degree $q = 1+2\,\mathrm{floor}\{M/2\}$ for CGL and EQUI and degree $q = 2M-1$ for LGL and CGLw.

MASS MATRIX AND DEFECT CORRECTION

With the local mass and first-derivative matrices A and D defined in the previous subsection, we can assemble the global mass matrix $\mathscr{M}$ and derivative matrix $\mathscr{D}$. Using these in the acoustic wave equation (described in the beginning of section 3.2.1), a leap-frog time discretization of with time step Δt is

$$\frac{1}{\Delta t}\mathscr{M}_v(\mathbf{v}^{n+1}-\mathbf{v}^n)=\mathscr{D}_p\mathbf{p}^{n+1/2},\quad \frac{1}{\Delta t}\mathscr{M}_p(\mathbf{p}^{n+3/2}-\mathbf{p}^{n+1/2})=\mathscr{D}_v\mathbf{v}^{n+1}. \tag{3.1}$$

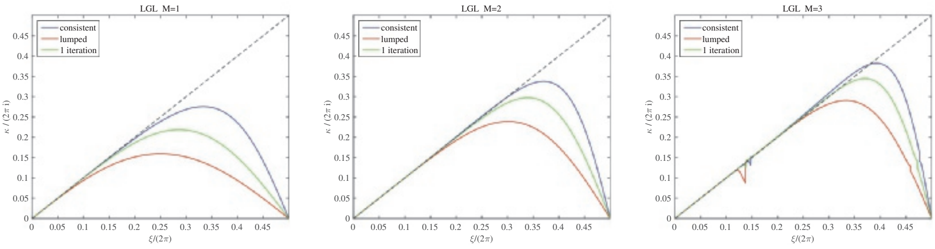
Here, the material properties are absorbed into the mass matrices and the superscript n denotes the solution at time $t^n = t_0 + n\Delta t$. Note that we have made a distinction between the first-derivative operators $\mathscr{D}_p$ and $\mathscr{D}_v$, but for the periodic problems considered later on in the analysis and numerical tests, they will be taken the same. As shown in **??**, the time-stepping stability limit for a leap-frog scheme is given by the CFL number $2\rho^{-1/2}(\tilde{\mathscr{L}})$, with $\tilde{\mathscr{L}} = -\mathscr{M}_p^{-1}\mathscr{D}_v\mathscr{M}_v^{-1}\mathscr{D}_p$ and where $\rho(\cdot)$ now denotes the spectral radius. For time stepping, we want to avoid the cost of inverting the consistent mass matrix and replace it by its lumped version. Depending on the choice of nodes, this may or may not harm the spatial accuracy. Formally, the lumped version should be exact for numerical quadrature of polynomials up to a degree of at least $2M-2$ for the second-order form of the wave equation and $2M-1$ for the first-order form. If its accuracy is less, we can iterate with the lumped mass matrix as preconditioner. This approach resembles defect correction (Stetter, 1978), which has the following convenient property. Consider two operators $\mathscr{L}_1$ and $\mathscr{L}_2$ where $\mathscr{L}_k$ has an order of accuracy p_k $(k=1,2)$ and $p_1 > p_2$. We can try to solve $\mathscr{L}_1\mathbf{u}=\mathbf{f}$ with the iterative scheme $\mathbf{u}^{-1}=0$, $\mathbf{u}^{j+1}=\mathbf{u}^j+\mathscr{L}_2^{-1}(\mathbf{f}-\mathscr{L}_1\mathbf{u}^j)$, where $j=0,1,\ldots$ denotes the iteration count, not the time step. Convergence is obtained if the operator $\mathscr{G}=\mathscr{I}-\mathscr{L}_2^{-1}\mathscr{L}_1$ has a spectral radius $\rho(\mathscr{G})<1$. In a finite-difference context, the order of accuracy of $\mathbf{u}^j$ is $\min(p_2,(j+1)p_1)$, which suggests that a few iterations will often suffice to get a sufficiently accurate though not necessarily fully converged result (Stetter, 1978). In our case, we can take the lumped mass matrix for $\mathscr{L}_2=\mathscr{M}^{\mathrm{L}}$ and the consistent mass matrix as $\mathscr{L}_1=\mathscr{M}$.

DISPERSION

The numerical dispersion of the finite-element scheme can be analyzed by considering the eigenvalues of the first-order operator $\mathscr{M}^{-1}\mathscr{D}$ or $(\mathscr{M}^{\mathrm{L}})^{-1}\mathscr{D}$ when discretized on a sufficiently fine periodic mesh with constant material properties and a constant element size h. Alternatively, we can use the fact that the elements are translation-invariant for constant material properties and element size and perform a Fourier transform on the solution. We then have to take the M degrees of freedom inside an element as a vector and do a transform on each component over the N elements. This results in a small $M \times M$ matrix in the Fourier domain. However, we can go one step further and also involve the M individual components. These are aliased but still can be considered separately by looking at the eigenvalues of the $M \times M$ block and unwrapping the result (Mulder, 1999). This produces a discrete approximation $\mathrm{i}\kappa$ to the exact operator $\mathrm{i}\xi$, where $\xi = k(x_N - x_0)/(NM) = kh/M \in [-\pi, \pi]$ is scaled version of the wavenumber k. The relative dispersion error can than be characterized by $\kappa/\xi - 1$. Note that the error in the dispersion curve does not tell the full story, because errors in the eigenvectors also play a role.

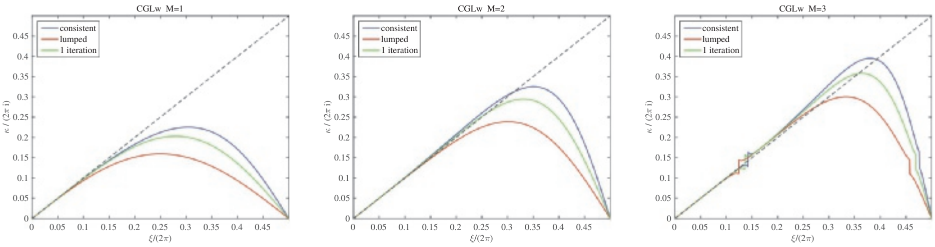
Table 3.1: Leading error terms in the dispersion curves for a polynomial basis of degree M and various sets of nodes, using the consistent or lumped mass matrix or lumped with one iteration based on $\mathscr{G}$. Its spectral radius $\rho(\mathscr{G})$ is given, as well as the CFL number without and with mass lumping.

M	nodes	consistent	lumped	1 iteration	$\rho(\mathscr{G})$	CFL (consist.)	CFL (lumped)	CFL (1 iter.)
1	LGL	$-\frac{1}{180}\xi^4$	$-\frac{1}{6}\xi^2$	$-\frac{1}{30}\xi^4$	2/3	$2/\sqrt{3} = 1.155$	2	1.457
2		$\frac{1}{270}\xi^4$	$-\frac{4}{270}\xi^4$	$-\frac{4}{945}\xi^4$	3/5	$\sqrt{2}/3 = 0.471$	2/3 = 0.667	0.535
3		$-\frac{81}{39200}\xi^8$	$-\frac{27}{2800}\xi^6$	$-\frac{3}{1400}\xi^6$	4/7	0.278	0.365	0.308
4		$\frac{128}{496125}\xi^8$	$-\frac{1024}{496125}\xi^8$	$-\frac{4096}{6449625}\xi^8$	5/9	0.188	0.239	0.208
5		$\frac{-9765625}{19179224064}\xi^{12}$	$\frac{-78125}{67060224}\xi^{10}$	$-\frac{15625}{50295168}\xi^{10}$	6/11	0.138	0.171	0.151
3	CGL	see LGL	$-\frac{333}{10240}\xi^4$	$-\frac{21}{1460}\xi^2$	3/5	see LGL	0.311	0.342
4			$\frac{8}{1395}\xi^4$	$-\frac{1042}{35397}\xi^4$	5/7		0.198	0.247
5			$-\frac{231125}{134217728}\xi^4$	$\frac{5115}{4502764}\xi^2$	0.966		0.132	0.203
1	CGLw	$-\frac{1}{24}\xi^2$	$-\frac{1}{6}\xi^2$	$-\frac{1}{24}\xi^2$	1/2	1.414	2	1.570
2	CGLw	$\frac{1}{30}\xi^2$	$-\frac{2}{135}\xi^4$	$\frac{1}{48}\xi^2$	1/2	0.426	2/3 = 0.667	0.541
3	CGLw	$\frac{9}{1280}\xi^4$	$-\frac{9}{320}\xi^4$	$-\frac{9}{5120}\xi^4$	1/2	0.213	0.354	0.297
4	CGLw	$-\frac{1}{405}\xi^4$	$-\frac{32}{4725}\xi^6$	$-\frac{1}{630}\xi^4$	1/2	0.132	0.224	0.192
5	CGLw	$-\frac{625}{344064}\xi^6$	$\frac{625}{258048}\xi^6$	$\frac{-625}{1032192}\xi^6$	1/2	0.0909	0.155	0.135
3	EQUI	see LGL	$-\frac{61}{1080}\xi^4$	$-\frac{42}{295}\xi^2$	0.651	see LGL	0.369	0.329
4			$\frac{40}{1137}\xi^4$	$\frac{56825}{157068}\xi^4$	(1.72)		0.184	(0.173)
5			$\frac{-92807}{312500}\xi^4$	$\frac{33740850}{26406233}\xi^2$	(1.96)		0.125	(0.117)



(a) $M = 1$ (b) $M = 2$ (c) $M = 3$

Figure 3.1: Dispersion curves for Legendre-Gauss-Lobatto points without and with mass lumping and after one iteration, for degree $M = 1$ (a), 2 (b), and 3 (c). The blue curve corresponds to a consistent mass matrix, the red to a lumped one, and the green is the result after one defect-correction step.



(a) $M = 1$ (b) $M = 2$ (c) $M = 3$

Figure 3.2: Dispersion curves for CGLw without and with mass lumping and after one iteration, for degree $M = 1$ (a), 2 (b), and 3 (c).

RESULTS

DISPERSION ANALYSIS

We compared the various spatial discretizations in terms of their dispersion curves, obtained by Fourier analysis, as well by set of numerical experiments. As an example, figure 3.1 shows dispersion curves for polynomials of degrees 1 to 3 on Legendre-Gauss-Lobatto points (LGL). Each graph shows the result without and with mass lumping as well as with 1 iteration of defect correction. The jumps in figure 3.1c are caused by the fact that in the Fourier analysis, M modes are considered simultaneously. Each of them corresponds to a particular root of the eigenvalue equation and can be assigned to a different wavenumber in the spectrum, according to how well the corresponding eigenvector matches the Fourier mode for that wavenumber (Mulder, 1999).

With lumping, the deviation from the exact dispersion curve, the straight line, increases, but not so much at the smaller values of ξ. With one iteration of $\mathscr{G} = \mathscr{I} -$

$(\mathscr{M}^{\mathrm{L}})^{-1}\mathscr{M}$, the result is improved. For the smaller wavenumbers, we have analytically determined the asymptotic error behaviour by taking the leading term in the series expansion of $\kappa/\xi - 1$ for the eigenvalue that is valid at small ξ. The results are listed in table 3.1 for various cases. For degree $M = 1$ and $M = 2$, the standard element (EQUI), the Legendre-Gauss-Lobatto points (LGL) and the unweighted Chebyshev-Gauss-Lobatto (CGL) points lead to the same discretization and, therefore, all provide the same results. The same is true when the consistent mass matrix is used. Then, the choice of nodes does not matter. The exception is the weighted scheme with Chebyshev-Gauss-Lobatto nodes (CGLw), where the weighting functions changes the outcome. Note that for the latter, the error analysis did *not* involve a weighted norm. Figure 3.2 show dispersion curves for degrees up to 3.

Interestingly, the LGL scheme without mass lumping has a fourth-order error instead of the usual second-order. The same behaviour is known in the finite-difference world (Lele, 1992). Without lumping and just a single step of defect correction, this fourth-order behaviour is recovered, albeit with a larger error constant.

With LGL and higher but odd degrees, 1 iteration reduces the size of the error but does not suffice to recover the super-convergence obtained with a consistent mass matrix. This appears to contradict the expected behaviour of the defect correction method. An explanation might be that for $M > 1$, there are M coupled modes, each representing a different point on the dispersion curve. This coupling is responsible for what are known as 'spurious' modes and could have a negative effect on the performance of the defect correction method.

For even degrees, the error constant changes after lumping but not the exponent. The error can be reduced by one or more iterations. Appendix A shows that the spectral radius of the iteration matrix obeys $\rho(\mathscr{G}) = (M+1)/(2M+1)$.

The CFL number that dictates the maximum allowable time step is listed in the last two columns. For degree 1, it is nearly twice as large after lumping. This will amply offset the cost of one iteration if the time stepping error does not dominate the problem. For higher degrees, the increase in CFL is not as dramatic.

A closed-form expression for the leading dispersion error with the consistent mass matrix and LGL points was found by (Ainsworth, 2014) and is quoted in Appendix B. A conjecture for the lumped case is included. For odd M, the error is completely due to the

Table 3.2: Exponents of the leading error in the dispersion curve and in the eigenvectors with LGL points and polynomials up to degree 5. The first of each pair corresponds to the relative error in the eigenvalue iκ for the first-order formulation or in the square root of the eigenvalue κ^2 for the second-order formulation. The second corresponds to the exponent of ξ in the leading error of the matrix S describing the eigenvector errors. This error is zero for $M=1$. The last column shows expressions for the trend for $M>1$, suggested by these results, where $p(M) = 2\,\text{floor}\{(M+1)/2\}$, that is, $p(M)=M$ if M is even and $p(M)=M+1$ if M is odd.

order	mass matrix	$M=1$	2	3	4	5	trend ($M>1$)
1	consistent	4, –	4, 2	8, 4	8, 4	12, 6	$2p(M), p(M)$
	lumped	2, –	4, 2	6, 4	8, 4	10, 6	$2M, p(M)$
2	consistent	2, –	4, 4	6, 5	8, 6	10, 7	$2M, M+2$
	lumped	2, –	4, 4	6, 5	8, 6	10, 7	$2M, M+2$

mass lumping and the related expression for the leading error can be found in (Mulder, 1999).

With Patera's scheme (CGL), we do expect the mass lumping to lower the accuracy, as the choice of nodes for the unweighted case is not related to any type of accurate numerical quadrature. The application of a single iteration may completely ruin the formal accuracy and more iterations are required to repair the harm. The same happens in the standard case (EQUI).

The behaviour of CGLw follows a regular pattern. Note that the weighted norm was not used in the analysis. Overall, errors are larger than with LGL. If M is odd, the lumping increases the error, but if M is even, lumping improves it and iterations will only increase the error. The spectral radius of the defect correction matrix does not depend on the degree of the element: $\rho(\mathscr{G}) = 1/2$, as shown in Appendix A.

One may wonder if diagonal preconditioning (Wathen, 1987, e.g.) would perform in a similar way. As an example, we consider LGL for degree $M=3$ and let $\mathscr{H} = I - (\text{diag}\{\mathscr{M}^{\text{L}}\})^{-1}\mathscr{M}$. In the Fourier domain, we obtain eigenvalues between $-\frac{1}{6}$ and $\frac{1}{2}$. After one iteration, the dispersion curve for small ξ behaves as $\xi(1-\frac{1}{36}-\frac{9}{1120}\xi^8)$. The term with $\frac{1}{36}$ actually destroys the formal accuracy, which needs to be repaired with subsequent iterations. We therefore expect diagonal preconditioning to be far less efficient than preconditioning with the mass-lumped mass matrix.

ERROR IN THE EIGENVECTORS

The dispersion curves describe the errors in the eigenvalues. For $M>1$, the error in eigenvectors also plays a role. To obtain that error, we compare to the exact eigenfunction $\overline{\mathbf{q}}$, which is of the form $\overline{q}_j = e^{2\pi \mathrm{i} m x_j}$, with x_j the node positions as defined above. The discrete problem has eigenvectors $\mathbf{q}_l$. We can express $\overline{\mathbf{q}}$ as a the unique linear combinations of these eigenvectors by $\overline{\mathbf{q}} = \sum_{l=0}^{M-1} \phi_l \mathbf{q}_l$. The error in the eigenvectors is given by the vectors $\mathbf{r}_l = \phi_l \mathbf{q}_l - \overline{\mathbf{q}} \delta_{l=l_{\mathrm{ref}}}$, $l = 0, 1, \dots, M-1$. Here, $\delta_{l=l_{\mathrm{ref}}}$ is the Kronecker delta, which is zero except for $l = l_{\mathrm{ref}}$, the index that corresponds to the 'physical' eigenvalue that approximates $\mathrm{i}M\xi$. The other indices correspond to the 'spurious' modes. Instead of an absolute error, we can determine a relative error by dividing each vector $\mathbf{r}_l$ elementwise by $\overline{\mathbf{q}}$ to obtain $\tilde{\mathbf{r}}_l$ with $\tilde{r}_{l,j} = r_{l,j}/\overline{q}_j$. The vectors $\tilde{\mathbf{r}}_l$ can be combined into a matrix S, which has them as columns. This matrix describes the error in approximating the exact eigenfunction as well as the energy that is leaked into the 'spurious' modes.

In (Mulder, 1999), the matrix S was determined in the Fourier domain, followed by an inverse Fourier transform. We can obtain the same results by working in the spatial domain, using the eigenvectors obtained by static condensation. Given the fact that these vectors are completely defined by their first M values for $j_1 = 0, 1, \dots, M-1$ at $j_0 = 0$, the matrix S will have size $M \times M$.

In Appendix B.1, we have listed the eigenvalue and eigenvector errors for polynomials up to degree $M = 5$ and LGL points, both for the first-order formulation that is the subject of this paper and for the second-order formulation discussed elsewhere (Mulder, 1999).

Table 3.2 summarizes the exponents of the leading errors in the eigenvalues and eigenvectors. The last column contains the suggested trends for $M > 1$, where it should be noted that exponents for the dispersion error in the second-order case were proven in (Mulder, 1999) and later also in (Ainsworth, 2004) and (Ainsworth and Wajid, 2009). For the first-order case with a consistent mass matrix, a proof can be found in (Ainsworth, 2014).

Numerical experiments

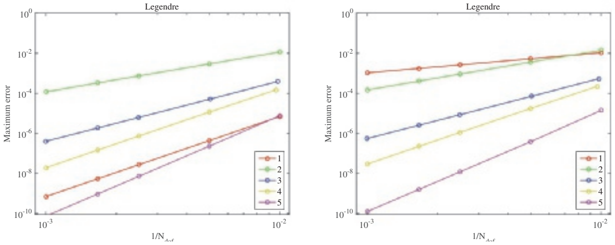
Before turning to the first-order formulation of the wave equation, we consider simple differentiation with the consistent mass matrix to verify the eigenvalue and eigenvector estimates. We consider the function $p(x) = \frac{1}{2\pi m}\sin(2\pi m x)$ with $m = 3$ on the periodic interval $\xi \in [0, 1)$. The mesh is either uniform with constant $h = 1/N$ for N elements or with two different spacings h_L and h_R. In the last case, we set $h_j = h_L$ for $j = 0, \dots, \frac{1}{2}N-1$

Table 3.3: Numerical results for the L_∞- and L_2-errors when taking the first derivative using Legendre polynomials and a consistent mass matrix. Listed are the exponents p of a power-law fit of the form ch^p, where $h \propto 1/N_{\text{dof}}$, to the L_∞- or L_2-errors shown in figure 3.3. The second and third column were obtained for a uniform grid. The fourth and fifth columns were obtained for a mesh with an abrupt jump in mesh size halfway the domain. Columns six to ten show similar results, but with projection instead of sampling of the initial data and the exact solution. The sixth column, for L_∞ on a uniform mesh, now agrees with the first row of results in table 3.2. On the non-uniform mesh, the convergence rates are worse.

	sampling				projection			
mesh	uniform		non-uniform		uniform		non-uniform	
M	L_∞	L_2	L_∞	L_2	L_∞	L_2	L_∞	L_2
1	4.0	4.5	1.0	2.0	4.0	4.5	1.0	2.0
2	2.0	2.5	2.0	2.5	2.0	2.5	1.9	2.5
3	3.0	3.5	3.0	3.5	4.0	4.6	3.0	4.2
4	4.0	4.5	3.9	4.5	3.9	4.4	3.9	4.4
5	5.0	5.5	5.0	5.5	6.1	6.6	5.1	6.3

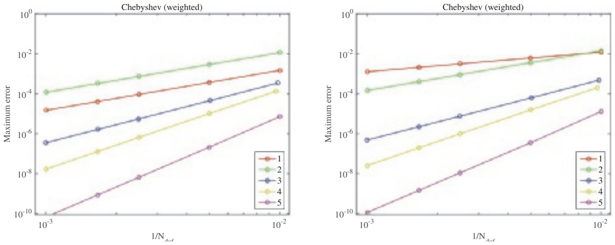
Table 3.4: As table 3.3, but for the weighted Chebyshev polynomials. See also figure 3.4.

	sampling				projection			
mesh	uniform		non-uniform		uniform		non-uniform	
M	L_∞	L_2	L_∞	L_2	L_∞	L_2	L_∞	L_2
1	2.0	2.5	1.0	2.1	2.0	2.5	1.0	2.1
2	2.0	2.5	2.0	2.5	2.0	2.5	2.0	2.5
3	3.0	3.5	3.0	3.5	4.0	4.5	2.9	4.3
4	4.0	4.5	4.0	4.5	3.9	4.5	3.9	4.4
5	5.0	5.5	5.0	5.5	6.0	6.5	5.1	6.3



(a) (b)

Figure 3.3: Maximum differentiation error for a simple test problem using Legendre polynomials as a function of the number of degrees of freedom, N_{dof}, for polynomial degrees 1 to 5. The grid spacing is either constant (a) or has an abrupt jump halfway the periodic domain (b).



(a) (b)

Figure 3.4: As figure 3.3, but for weighted Chebyshev polynomials (CGLw).

and $h_j = h_R$ for $j = \frac{1}{2}N, \ldots, N-1$, with N chosen even and $h_L = 0.8h_R$. Figure 3.3 shows the maximum error as a function of the reciprocal of the number of degrees of freedom, N_{dof}, for polynomial degrees 1 to 5. Power-law fits to the results provide the powers listed in table 3.3. With point-wise sampling of the input function and the exact solution, the error behaviour is worse than the estimates of table 3.2. With a proper projection on the basis function and a uniform mesh, the same powers are found for the L_∞ estimates. With the non-uniform mesh, the maximum error appears to behave as h^M and error cancellation and super-convergence are lost.

Similar results with weighted Chebyshev polynomials (CGLw) are shown in figure 3.4 and table 3.4. Again, the odd degrees lead to a better performance.

These numerical results confirm that dispersion error analysis by itself is insufficient and that the eigenvector errors have to be included as well.

In addition to the above dispersion-curve analysis, we have performed a set of numerical experiments on the first-order formulation of the acoustic wave equation. We consider a Ricker pulse, the second time derivative of a Gaussian, travelling around once on a periodic domain.

We ran at a fraction of 10^{-3} times the maximum time step dictated by the CFL condition to avoid too much imprint of the time stepping error. A less costly alternative would be to perform higher-order time stepping (Dablain, 1986; Gilbert and Joly, 2008; Lax and Wendroff, 1960; Shubin and Bell, 1987) or dispersion correction (Anderson et al., 2015; Stork, 2013; Wang and Xu, 2015).

As before, we used two difference spacing h_L and h_R. The standard deviation of the Ricker pulse was 0.0375 times the length of the domain. The initial and final position of its centre was at 0.74 of the length of the domain, in the part to the right that has the larger spacing.

Figure 3.5 a–c plot the maximum errors in the particle velocity $v(t_{\max}, x)$ after one round trip for a varying number of degrees of freedom without and with mass lumping and with one extra iteration for polynomial degrees $M = 1$ to 5. One iteration clearly pays off for the lowest degree, $M = 1$, and also for the higher degrees when the number of degrees of freedom is small and the error large. Overall, the effect of the eigenvector errors, summarized in table 3.2, dominates the results for degrees larger than one. The improvement with defect correction is the largest for the lowest degree, $M = 1$. Although the fourth-order super-convergence for this degree is lost on a non-uniform mesh, the accuracy after 1 iteration is still considerably better than with just mass lumping.

In addition to the above runs, a few additional experiments were conducted to investigate how a larger number of iterations affect the result and if a diagonal matrix would be a better preconditioner, as suggested by (Wathen, 1987). Figure 3.5 e–f show the result of increasing the number of iterations with the operator $\mathscr{G}$, without attempting to obtain some acceleration with the conjugate gradient method. We observe a slight improvement, but the increase in computational costs hardly pays off.

Figure 3.6 shows results after using the diagonal of the mass matrix instead of the lumped mass matrix as preconditioner. It can be seen that in order for the diag($\mathscr{M}$) to behave similar to $\mathscr{M}^{\mathrm{L}}$, at least 20 iterations are required, showing that the lumped mass matrix is superior as preconditioner.

Finally, figure 3.7 displays the error behaviour for CGLw. Note that the dispersion curves are based on the usual norm and do not involve weighting. Again, one iteration helps to improve the accuracy, as for LGL.

GENERALIZATION TO 2D

We can quickly analyze the performance in 2D by considering Fourier analysis on a periodic grid with square elements, both for bilinear elements and for linear elements on triangles.

We start with bilinear elements on squares. Let T_x denote a shift operator in the x-direction, such that $T_x p_{k,l} = p_{k+1,l}$. Here, $p_{k,l}$ denotes the discrete pressure in the point (x_k, y_l) with $x_k = x_0 + kh_x$ and $y_l = y_0 + lh_y$ and grid spacings h_x and h_y. Its Fourier symbol is $\hat{T}_x = \exp(\mathrm{i}\xi_1)$ with $|\xi_1| \le \pi$, where ξ_1 is related to the wavenumber k_x in the x-direction by $\xi_1 = k_x h_x$. Likewise, $T_y p_{k,l} = p_{k,l+1}$ with symbol $\hat{T}_y = \exp(\mathrm{i}\xi_2)$ and $|\xi_2| \le \pi$. One row of the assembled mass matrix in a single node, relative to the others, is

$$\mathscr{M} = \tfrac{1}{36}\left[16 + 4(T_x^{-1} + T_x + T_y^{-1} + T_y) + T_x^{-1}T_y^{-1} + T_x T_y^{-1} + T_x^{-1}T_y + T_x T_y\right].$$

Its symbol is

$$\hat{\mathscr{M}} = \tfrac{1}{36}(\hat{T}_x^{-1} + 4 + \hat{T}_x)(\hat{T}_y^{-1} + 4 + \hat{T}_y) = \tfrac{1}{9}(2 + \cos\xi_1)(2 + \cos\xi_2).$$

One row of the derivative matrix in x is

$$\mathscr{D}^{(1)} = \tfrac{1}{12}(T_x - T_x^{-1})(T_y^{-1} + 4 + T_y),$$

with symbol

$$\hat{\mathscr{D}}^{(1)} = \tfrac{2}{3}\mathrm{i}(2 + \cos\xi_2)\sin\xi_1.$$

For $\mathscr{D}^{(2)}$, we can swap ξ_1 and ξ_2. Then,

$$\hat{\mathscr{M}}^{-1}\hat{\mathscr{D}}^{(1)} = \frac{3\mathrm{i}\sin\xi_1}{2 + \cos\xi_1} \simeq \mathrm{i}\xi_1(1 - \tfrac{1}{180}\xi_1^4),$$

showing that we have fourth-order accuracy with bilinear elements and a consistent mass matrix. With mass lumping, the result has only second-order accuracy:

$$\hat{\mathscr{M}}^{\mathrm{L}^{-1}}\hat{\mathscr{D}}^{(1)} = \tfrac{1}{3}\mathrm{i}(2 + \cos\xi_1)\sin\xi_1 \simeq \mathrm{i}\xi_1\left[1 - \tfrac{1}{6}(\xi_1^2 + \xi_2^2)\right].$$

The expressions can be used to estimate the eigenvalues of G by noting that

$$\hat{G} = 1 - \tfrac{1}{9}(2+\cos\xi_1)(2+\cos\xi_2) \in [0, \tfrac{8}{9}].$$

After one iteration with $\hat{G}$, the error becomes

$$-\tfrac{1}{180}\left(6\xi_1^4 + 10\xi_1^2\xi_2^2 + 5\xi_2^4\right),$$

restoring the fourth-order accuracy.

We can repeat this analysis for linear elements on triangles and a regular mesh consisting of squares cut in half across the diagonal, from the left upper to the right lower corner. With unit spacing, the first triangle has vertices $(0,0)$, $(1,0)$, $(0,1)$ with basis functions $\{1-x-y, x, y\}$ and the second has $(1,1)$, $(1,0)$, $(0,1)$ with basis functions $\{-(1-x-y), 1-y, 1-x\}$. For the Fourier analysis, we select 8 triangles contained inside the 4 squares surrounding one node and assemble the matrices. Then, one row of the mass matrix is given by

$$\mathscr{M} = \tfrac{1}{12}(6 + T_x^{-1} + T_x + T_y^{-1} + T_y + T_x T_y^{-1} + T_x^{-1} T_y),$$

with corresponding symbol

$$\hat{\mathscr{M}} = \tfrac{1}{6}(3 + \cos\xi_1 + \cos\xi_2 + \cos(\xi_1 - \xi_2)).$$

A row of the x-derivative matrix is

$$\mathscr{D}^{(1)} = \tfrac{1}{6}\left[2(T_x - T_x^{-1}) + T_y(1 - T_x^{-1}) + T_y^{-1}(1 - T_x)\right],$$

with symbol

$$\hat{\mathscr{D}}^{(1)} = \tfrac{1}{3}\mathrm{i}[2\sin\xi_1 + \sin\xi_2 + \sin(\xi_1 - \xi_2)].$$

Now,

$$\hat{\mathscr{M}}^{-1}\hat{\mathscr{D}}^{(1)} \simeq \mathrm{i}\xi_1\left[1 - \tfrac{1}{360}\xi_1^2\left\{2\xi_1^2 - 5\xi_2(\xi_1 - \xi_2)\right\}\right],$$

revealing fourth-order behaviour of the error. The results for the derivative in the y-direction are the same after swapping T_x and T_y or ξ_1 and ξ_2. With mass lumping, the operator becomes

$$\left(\hat{\mathscr{M}}^{\mathrm{L}}\right)^{-1}\hat{\mathscr{D}}^{(1)} = \hat{\mathscr{D}}^{(1)} \simeq \mathrm{i}\xi_1\left[1-\tfrac{1}{6}(\xi_1^2+\xi_2^2-\xi_1\xi_2)\right],$$

providing only second-order accuracy. These expressions also provide an estimate of the eigenvalue range of G:

$$\hat{G} = \tfrac{1}{6}\,[3-\cos\xi_1-\cos\xi_2-\cos(\xi_1-\xi_2)] \in [0,\tfrac{3}{4}].$$

One iteration with $\hat{G}$ reduces the relative error to

$$-\tfrac{1}{360}\left(12\xi_1^4-25\xi_1^3\xi_2+35\xi_1^2\xi_2^2-20\xi_1\xi_2^3+10\xi_2^4\right),$$

again restoring the fourth-order accuracy.

It remains to be seen if this accuracy can actually be obtained in numerical experiments. A practical problem in seismic applications is the need to sample the wave field in arbitrary points of the computational domain. To reach a sufficiently high interpolation degree, the polynomials that represent the solution are not suited. Essentially non-oscillatory interpolation may provide a solution in that case (Harten et al., 1987; Putti et al., 1990).

Conclusions

We have compared four finite-element schemes with polynomial basis functions for the first-order formulation of the acoustic wave equation, using Legendre-Gauss-Lobatto nodes, Chebyshev-Gauss-Lobatto without and with weighting function or the standard element. Mass lumping, desired for numerical efficiency since it allows for explicit time stepping, tends to decrease the spatial accuracy. The remaining accuracy in the numerical dispersion is best for the Legendre-Gauss-Lobatto nodes and, for polynomials of odd degrees, exceeds that that of the second-order formulation of the wave equation. In some cases, the accuracy can be improved by applying one iteration on the consistent mass matrix, preconditioned by its lumped version. For polynomials of degree one, this improves the accuracy from second to fourth order in the element size. In other cases, the improvement in accuracy is less dramatic.

The error in the eigenvectors for the first-order formulation, however, is worse than obtained for the second-order formulation, without and with mass lumping. Because the eigenvector error is zero for the lowest-degree scheme, with linear polynomials, our iterative approach appears to be most attractive for just that case.

Fourier analysis in two space dimensions suggests that the fourth-order error behaviour should be obtained for the lowest-order scheme, either with bilinear elements on quadrilaterals or with linear elements on triangles, at least on very regular meshes and with constant material properties. Whether or not this still holds on general unstructured meshes remains to be seen.

ACKNOWLEDGEMENTS

This work is part of the research programme of the Foundation for Fundamental Research on Matter (FOM), which is part of the Netherlands Organisation for Scientific Research (NWO). The authors are indebted to Martin van Gijzen for helpful suggestions.

REFERENCES

Ainsworth, M., 2004. Discrete dispersion relation for hp-version finite element approximation at high wave number. SIAM Journal on Numerical Analysis 42 (2), 553–575.

Ainsworth, M., 2014. Dispersive behaviour of high order finite element schemes for the one-way wave equation. Journal of Computational Physics 259, 1–10.

Ainsworth, M., Wajid, H., 2009. Dispersive and dissipative behavior of the spectral element method. SIAM Journal on Numerical Analysis 47 (5), 3910–3937.

Anderson, J. E., Brytik, V., Ayeni, G., 2015. Numerical temporal dispersion corrections for broadband temporal simulation, rtm and fwi. In: SEG Technical Program Expanded Abstracts. pp. 1096–1100.

Cohen, G., 2002. Higher-Order Numerical Methods for Transient Wave Equations. Springer.

Dablain, M. A., 1986. The application of high-order differencing to the scalar wave equation. Geophysics 51 (1), 54–66.

Gilbert, J. C., Joly, P., 2008. Higher order time stepping for second order hyperbolic problems and optimal CFL conditions. Computational Methods in Applied Sciences 16. Springer, Berlin, pp. 67–93.

Harten, A., Engquist, B., Osher, S., Chakravarthy, S. R., 1987. Uniformly high order accurate essentially non-oscillatory schemes, iii. Journal of Computational Physics 71 (2), 231–303.

Lax, P., Wendroff, B., 1960. Systems of conservation laws. Communications on Pure and Applied Mathematics 31 (2), 217–237.

Lele, S. K., 1992. Compact finite difference schemes with spectral-like resolution. Journal of Computational Physics 103 (1), 16–42.

Mulder, W., 1999. Spurious modes in finite-element discretizations of the wave equation may not be all that bad. Applied Numerical Mathematics 30 (4), 425–445.

Mulder, W. A., 1996. A comparison between higher-order finite elements and finite differences for solving the wave equation. In: Désidéri, J.-A., LeTalleca, P., Oñate, E., Périaux, J., Stein, E. (Eds.), Proceedings of the Second ECCOMAS Conference on Numerical Methods in Engineering, Paris, Sept. 9–13, 1996. John Wiley & Sons, Chichester, pp. 344–350.

Patera, A. T., 1984. A spectral element method for fluid dynamics: Laminar flow in a channel expansion. Journal of Computational Physics 54 (3), 468–488.

Putti, M., Yeh, W. W.-G., Mulder, W. A., 1990. A triangular finite volume approach with high resolution upwind terms for the solution of groundwater transport equations. Water Resources Research 26 (12), 2865–2880.

Shamasundar, R., Mulder, W. A., 2016. Improving the accuracy of mass-lumped finite-elements in the first-order formulation of the wave equation by defect correction. Journal of Computational Physics 322, 689–707.

Shubin, G. R., Bell, J. B., 1987. A modified equation approach to constructing fourth order methods for acoustic wave propagation. SIAM Journal on Scientific and Statistical Computing 8 (2), 135–151.

Stetter, H. J., 1978. The defect correction principle and discretization methods. Numerische Mathematik 29 (4), 425–443.

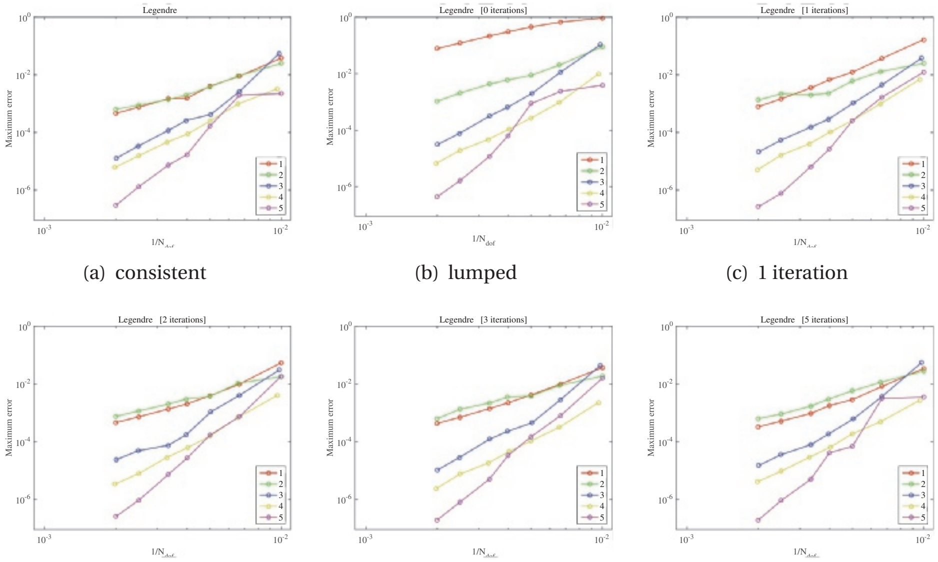
Stork, C., 2013. Eliminating nearly all dispersion error from fd modeling and rtm with minimal cost increase. In: 75th EAGE Conference & Exhibition incorporating SPE EUROPEC, Extended Abstract.

Thompson, L. L., Pinsky, P. M., 1994. Complex wavenumber Fourier analysis of the p-version finite element method. Computational Mechanics 13 (4), 255–275.

Wang, M., Xu, S., 2015. Time dispersion transforms in finite difference of wave propagation. In: 77th EAGE Conference & Exhibition, Extended Abstract.

Wathen, A. J., 1987. Realistic eigenvalue bounds for the Galerkin mass matrix. IMA Journal of Numerical Analysis 7 (4), 449–457.

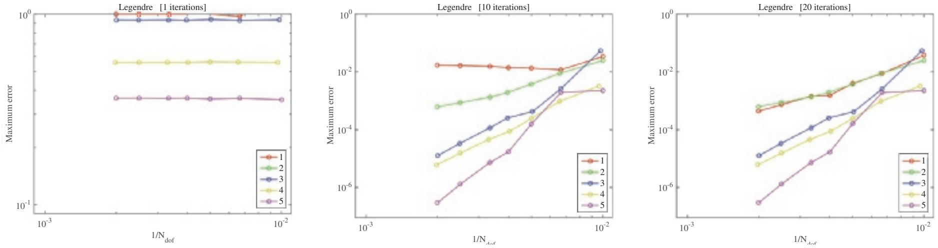
Zhebel, E., Minisini, S., Kononov, A., Mulder, W. A., 2014. A comparison of continuous mass-lumped finite elements with finite differences for 3-D wave propagation. Geophysical Prospecting 62 (5), 1111–1125.



(a) consistent (b) lumped (c) 1 iteration

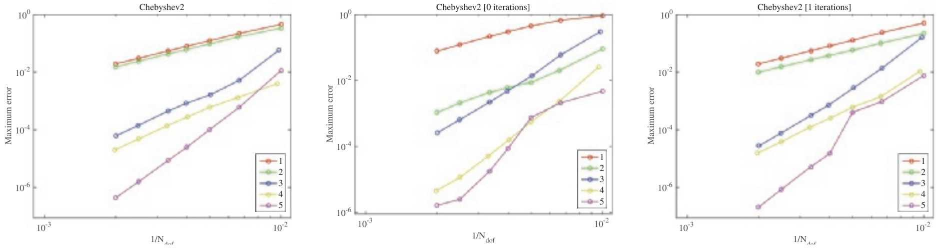
(d) 2 iterations (e) 3 iterations (f) 5 iterations

Figure 3.5: Maximum error in the particle velocity, v, as function of the inverse number of degree of freedom, $1/N_{\text{dof}}$, for the Legendre-Gauss-Lobatto nodes (LGL) with (a) the consistent mass matrix, (b) the lumped mass matrix, and after 1 (c), 2 (d), 3 (e) or 5 (f) iterations with the defect correction operator $\mathscr{G}$.



(a) 1 iteration (b) 10 iterations (c) 20 iterations

Figure 3.6: Maximum error in the particle velocity, v, as function of the inverse number of degree of freedom, $1/N_{\text{dof}}$, for the Legendre-Gauss-Lobatto nodes (LGL) using the diagonal of the mass matrix as preconditioner, after 1 (a), 10 (b), or 20 (c) iterations.



(a) consistent

(b) lumped

(c) 1 iteration

Figure 3.7: Maximum error in the particle velocity, v, as function of the inverse number of degree of freedom, $1/N_{\text{dof}}$, for the Chebyshev-Gauss-Lobatto nodes with weighting (CGLw) with the consistent mass matrix (a), its lumped version (b), or with one iteration (c).

4

2D ANALYSIS AND CUSTOMISING THE SOURCE FUNCTION

Parts of this chapter have been published in Geophysical Journal International, 215(2), pp.1231-1240 (Shamasundar and Mulder, 2018)

Chapter 3 concluded that defect correction can improve the convergence property of finite-elements in the first-order system of acoustic equations in 1D; the inexpensive linear elements showed the same performance as a fourth-order scheme. However, for real world problems we need to ensure that the same improvement holds in higher dimensions. Based on the results of the earlier chapter, we conjecture that defect correction should work for 2D problems. In the first half of this chapter, we analyze the 2-D case. Theoretical results imply that the lowest-degree polynomial provides fourth-order accuracy with defect correction, if the grid of squares or triangles is highly regular and material properties constant. But numerical results converge more slowly than theoretical predictions. Further investigation demonstrates that this is due to the activation of error-inducing wavenumbers in the delta-source representation. In the second half of the chapter, we provide a solution to this problem in the form of a tapered-sinc source function.

INTRODUCTION

Modelling of seismic data requires substantial computational resources. The finite-difference method is widely used in the oil industry because it is relatively easy to code up and optimize. The finite-element method is computationally more demanding but may offer better accuracy at a given cost in the presence of topography and large impedance contrast, but only if the mesh follows the interfaces between different rock types (Kononov et al., 2012; Zhebel et al., 2014).

A typical finite-element discretization of the wave equation in its second-order form involves a stiffness matrix, related to the spatial derivatives, and a mass matrix, related to the second derivatives in time. Because inverting the large sparse mass matrix at each time step is costly, it is replaced by its mass-lumped version, a diagonal matrix obtained by taking its row sums. The resulting weights are equivalent to those of a numerical quadrature rule. For rectangular types of elements, quadrangles in 2D and hexahedra in 3D, Legendre-Gauss-Lobatto quadrature produces the well-known spectral elements (Komatitsch and Tromp, 1999).

Spectral elements for simplicial elements, triangles in 2D and tetrahedra in 3D, are more difficult to construct. Mass lumping results in a loss of spatial accuracy, which can be recovered by augmenting the basis function with higher-degree polynomials that are the product of a bubble function and a polynomial (Fried and Malkus, 1975). A bubble function is a polynomial that vanishes on all the edges of the triangle. At present, triangular elements are known up to degree 9 (Chin-Joe-Kong et al., 1999; Cohen et al., 2001, 1995; Cui et al., 2017; Fried and Malkus, 1975; Liu et al., 2017; Mulder, 1996, 2013). In 3D on tetrahedra, two kinds of bubble functions are required: face bubbles that vanish on the edges of the faces and interior bubbles that are zero on all edges and faces of the tetrahedron Mulder (1996). Tetrahedral elements are known up to degree 3 Chin-Joe-Kong et al. (1999). Mulder and Shamasundar (2016) considered their performance for elastic wave propagation.

Discontinuous Galerkin methods offer an alternative to diagonal mass lumping by giving up conformity and restoring it by penalty terms leading to additional fluxes in the discretization Basabe and Sen (2007); Diaz and Grote (2009); Grote et al. (2006); Käser and Dumbser (2006); Riviere and Wheeler (2003). The resulting mass matrix is block diagonal and easy to invert.

Finite-element schemes for the acoustic and elastic wave equation are commonly based on the second-order form of the partial differential equations. Dispersion analysis by Ainsworth (2014) showed that the first-order form provides an accuracy that is better by 2 orders for the odd-degree Legendre-Gauss-Lobatto elements. However, if the error in the eigenvectors is included, as it should Mulder (1999), the full error only shows this improvement for the lowest-degree elements Shamasundar and Mulder (2016): a first-order formulation with linear elements in 1D has fourth-order spatial accuracy, but this requires a consistent, or full, mass matrix. With mass lumping, required to avoid the inversion of the mass matrix, the accuracy drops to second order. However, by invoking the defect-correction principle Stetter (1978), we could show that iterative inversion of the mass matrix requires only one iteration when using the lumped mass matrix as preconditioner, at least on equidistant grids. This result motivated us to consider the first-order formulation of the wave equation with continuous linear elements in 2D. Note that the first-order formulation with the discontinuous Galerkin method is less uncommon (Chung and Engquist, 2009; Delcourte et al., 2009; Etienne et al., 2010; Hesthaven and Warburton, 2002, 2007; Modave et al., 2015; Wilcox et al., 2010, e.g.).

In seismic simulations, the source term is typically much smaller in size than a wavelength and can therefore be represented by a delta function. In the finite-element formulation of the wave equation, be it in second- or first-order form, integration of the delta function against the basis functions offers a natural way to obtain its discrete representation. Nevertheless, an imprint of the triangular shape of the element may appear in the solution and a ‘rounder’ representation might provide a better accuracy. Then, a gaussian is an option.

Another reason to choose a gaussian is the odd-even or checker-board decoupling that may occur for some discrete schemes in first-order form Brossier et al. (2008). This decoupling is related to the shortest wavelengths that happen to lie in the null-space of the discrete spatial operator. Once excited, they will not disappear if the scheme is not dissipative. A gaussian source with sufficiently large standard deviation will avoid the excitation of such waves.

An alternative to a gaussian is a tapered sinc Hicks (2002), proposed for finite-difference schemes. A sinc function is the spatial equivalent of a band-limited delta function and the tapering keeps it localized. Walden (1999) presented piecewise polynomial approximations of the delta function, both for finite-element and for finite-difference schemes. Petersson et al. (2016) applied these ideas to finite-difference scheme for the wave equation in first-order form.

We will examine the performance of three source representations, delta function, gaussian and tapered sinc, with finite elements. For testing purposes, we consider the standard second-order finite-element discretization of the acoustic wave equation with linear elements on a triangular mesh using mass lumping. We also look at the first-order formulation with linear elements and defect correction. Since for the latter, the dispersion curve in 1D returns to zero at the highest spatial frequency, we expect that odd-even decoupling will play a role, as with the scheme of Brossier et al. (2008).

The next section contains a description of the second-order and first-order formulation of the acoustic wave equation, the source term representations, and Fourier analysis of the schemes for a simple structured 2D periodic mesh, offering some insight in what to expect. The following section presents results for a series of numerical experiments that asses the performance of the various schemes. It ends with a non-trivial example. The last section summarizes the main conclusions.

METHOD

FINITE ELEMENTS

The examples further on will involve both the first- and second-order form of the 2-D acoustic wave equation. We start with the latter:

$$\frac{1}{\rho c^2}\frac{\partial^2 p}{\partial t^2} = \frac{\partial}{\partial x}\frac{1}{\rho}\frac{\partial p}{\partial x} + \frac{\partial}{\partial z}\frac{1}{\rho}\frac{\partial p}{\partial z} + f. \tag{4.1}$$

Here, $p(t,\mathbf{x})$ is the pressure as a function of time t and position $\mathbf{x} = (x,z)$, $f = w(t)s(\mathbf{x})$ is the source term with wavelet $w(t)$ and spatial distribution $s(\mathbf{x})$, typically taken as a delta function $s(\mathbf{x}) = \delta(x - x_s, z - z_s)$ for a source position (x_s, z_s). The sound speed $c(\mathbf{x})$ and density $\rho(\mathbf{x})$ characterize the material through which the waves propagate.

The first-order form is given by the system

$$\frac{1}{\rho c^2}\frac{\partial p}{\partial t} = \frac{\partial u}{\partial x} + \frac{\partial v}{\partial z} + g, \tag{4.2a}$$

$$\rho \frac{\partial u}{\partial t} = \frac{\partial p}{\partial x}, \tag{4.2b}$$

$$\rho \frac{\partial v}{\partial t} = \frac{\partial p}{\partial z}, \tag{4.2c}$$

where the particle velocity in the x-direction is denoted by u and in the z-direction by v. The source term $g = W(t)s(\mathbf{x})$, with $\frac{\partial}{\partial t}W(t) = w(t)$. An intermediate representation is

$$\frac{1}{\rho c^2}\frac{\partial^2 p}{\partial t^2} = \frac{\partial a}{\partial x} + \frac{\partial b}{\partial z} + f, \tag{4.3a}$$

$$\rho a = \frac{\partial p}{\partial x}, \tag{4.3b}$$

$$\rho b = \frac{\partial p}{\partial z}, \tag{4.3c}$$

with accelerations a and b. At the level of the partial differential equations, these are equivalent. After discretisation, they may yield solutions with different numerical errors.

For the finite-element discretization in second-order form, we consider a triangular mesh with N nodes and expand the pressure as

$$p = \sum_{j=1}^{N} p_j \phi_j(\mathbf{x}), \tag{4.4}$$

where the basis functions $\phi_j(\mathbf{x})$ are piecewise linear on those triangles that have $\mathbf{x}_j$ as one of their vertices and $\phi_j(\mathbf{x}_k) = \delta_{j,k}$ for all vertices $\mathbf{x}_k$. The mass matrix $\mathbf{M}$ and stiffness matrix $\mathbf{K}$ on the computational domain Ω have elements

$$\mathbf{M}_{j,k} = \int_\Omega \frac{1}{\rho c^2}\phi_j\phi_k \,\mathrm{d}\mathbf{x}, \quad \mathbf{K}_{j,k} = \int_\Omega \frac{1}{\rho}\left(\nabla\phi_j\right)\cdot\left(\nabla\phi_k\right)\mathrm{d}\mathbf{x}, \tag{4.5}$$

respectively. The lumped mass matrix $\mathbf{L}$ is obtained from the row sum of $\mathbf{M}$: $\mathbf{L}_{j,k} = \delta_{j,k}\sum_{k=1}^{N}\mathbf{M}_{j,k}$. The discrete scheme becomes

$$\mathbf{p}^{n+1} = 2\mathbf{p}^n - \mathbf{p}^{n-1} + (\Delta t)^2\mathbf{L}^{-1}\left(\mathbf{f}^n - \mathbf{K}\mathbf{p}^n\right), \tag{4.6}$$

where $\mathbf{p}^n$ contains the pressures p_j on the nodes at time $t_n = t_0 + n\Delta t$. The size of the time step Δt should not exceed $2/\sqrt{\lambda_{\max}\left(\mathbf{L}^{-1}\mathbf{K}\right)}$, where $\lambda_{\max}(\cdot)$ denotes the spectral radius or maximum eigenvalue. We will discuss the source term vector $\mathbf{f}^n$ later on.

The pressure should be zero at the free surface. This condition can be imposed on the mass matrix by simply eliminating the entries that correspond to the free-surface boundary. Alternatively, one can set those entries to zero in the inverse lumped mass matrix $\mathbf{L}^{-1}$. For the other boundaries, sponge boundary conditions are used Cerjan et al. (1985) together with a zero pressure on the boundary.

For the first-order formulation, we expand u and v into the same basis function as the pressure p and define derivative matrices $\mathbf{D}^{(x)}$ and $\mathbf{D}^{(z)}$ with elements

$$\mathbf{D}^{(x)}_{j,k} = \int_\Omega \phi_j \frac{\partial}{\partial x}\phi_k \,\mathrm{d}\mathbf{x}, \quad \mathbf{D}^{(z)}_{j,k} = \int_\Omega \phi_j \frac{\partial}{\partial z}\phi_k \,\mathrm{d}\mathbf{x}. \tag{4.7}$$

Now there are three mass matrices, $\mathbf{M}^{(p)}$, $\mathbf{M}^{(u)}$ and $\mathbf{M}^{(v)}$. The first is the same as in equation (4.5). The other two mass matrices have entries

$$\mathbf{M}^{(u)}_{j,k} = \mathbf{M}^{(v)}_{j,k} = \int_\Omega \rho\phi_j\phi_k \,\mathrm{d}\mathbf{x}. \tag{4.8}$$

Zero-pressure boundary values can be eliminated from $\mathbf{M}^{(p)}$. Doing the same for the differentiation matrices, we obtain non-square matrices. More precisely, we have

$$\int_\Omega \boldsymbol{\psi}\cdot\nabla\phi \,\mathrm{d}\mathbf{x} = \int_{\delta\Omega} \phi\,(\boldsymbol{\psi}\cdot\mathbf{n})\,\mathrm{d}\mathbf{x} - \int_\Omega \phi\nabla\cdot\boldsymbol{\psi}\,\mathrm{d}\mathbf{x}, \tag{4.9}$$

where $\delta\Omega$ denotes the boundary of the domain Ω and $\mathbf{n}$ the outward normal on that boundary. Here, the scalar field $\phi = \phi^{(p)}(\mathbf{x})$ and the vector $\boldsymbol{\psi} = \left(\phi^{(u)}(\mathbf{x}), \phi^{(v)}(\mathbf{x})\right)^\mathsf{T}$. If we set $\phi = 0$ everywhere on the boundary $\delta\Omega$, the first term on the right-hand side vanishes. We can let the earlier matrix $\mathbf{D}^{(x)}$ act on $\mathbf{p}$ and drop the columns that correspond to zero pressure values on the boundary and do the same with $\mathbf{D}^{(z)}$. For the velocities, minus the transpose matrices can then be used. Note that in this way, the condition of zero transverse velocity is not explicitly imposed.

With a leap-frog time stepping scheme, the discrete system becomes

$$\frac{1}{\Delta t}\mathbf{M}^{(p)}\left(\mathbf{p}^{n+1} - \mathbf{p}^n\right) = \\ \mathbf{g}^{n+1/2} - \left(\mathbf{D}^{(x)}\right)^\mathsf{T}\mathbf{u}^{n+1/2} - \left(\mathbf{D}^{(z)}\right)^\mathsf{T}\mathbf{v}^{n+1/2}, \tag{4.10a}$$

$$\frac{1}{\Delta t}\mathbf{M}^{(u)}\left(\mathbf{u}^{n+1/2}-\mathbf{u}^{n-1/2}\right)=\mathbf{D}^{(x)}\mathbf{p}^{n}, \tag{4.10b}$$

$$\frac{1}{\Delta t}\mathbf{M}^{(v)}\left(\mathbf{v}^{n+1/2}-\mathbf{v}^{n-1/2}\right)=\mathbf{D}^{(z)}\mathbf{p}^{n}. \tag{4.10c}$$

The superscripts with n denote the solution at time $t^n = t^0 + \Delta t$. The time step Δt should not exceed

$$2\lambda_{\max}^{-1/2}\Big(\left(\mathbf{M}^{(p)}\right)^{-1}\big[\left(\mathbf{D}^{(x)}\right)^{\mathsf{T}}\left(\mathbf{M}^{(u)}\right)^{-1}\mathbf{D}^{(x)}+ \\ \left(\mathbf{D}^{(z)}\right)^{\mathsf{T}}\left(\mathbf{M}^{(v)}\right)^{-1}\mathbf{D}^{(z)}\big]\Big) \tag{4.11}$$

In 2D, the inversion of the mass matrices can be accomplished by a sparse Cholesky decomposition, but is costly. One or two iterations preconditioned by the lumped mass matrix should suffice. As the lumped mass matrix provides second-order accuracy and the consistent one fourth-order, at least in 1D on a uniform mesh, the defect-correction principle states that one extra iteration on top of the initial step should suffice. On non-uniform meshes and in the presence of odd-even decoupling, we do not expect fourth-order convergence but still hope for some improvement in accuracy.

To describe the method, define iteration matrices

$$\begin{aligned}\mathbf{G}^{(p)}&=\mathbf{I}-\left(\mathbf{L}^{(p)}\right)^{-1}\mathbf{M}^{(p)},\\ \mathbf{G}^{(u)}&=\mathbf{I}-\left(\mathbf{L}^{(u)}\right)^{-1}\mathbf{M}^{(u)},\\ \mathbf{G}^{(v)}&=\mathbf{I}-\left(\mathbf{L}^{(v)}\right)^{-1}\mathbf{M}^{(v)}.\end{aligned} \tag{4.12}$$

and let

$$\mathbf{A}^{(x)}=\left(\mathbf{L}^{(u)}\right)^{-1}\mathbf{D}^{(x)},\quad \mathbf{A}^{(z)}=\left(\mathbf{L}^{(v)}\right)^{-1}\mathbf{D}^{(z)}, \tag{4.13}$$

$$\mathbf{B}^{(x)}=-\left(\mathbf{L}^{(p)}\right)^{-1}\left(\mathbf{D}^{(x)}\right)^{\mathsf{T}},\quad \mathbf{B}^{(z)}=-\left(\mathbf{L}^{(p)}\right)^{-1}\left(\mathbf{D}^{(z)}\right)^{\mathsf{T}}, \tag{4.14}$$

and

$$\overline{\mathbf{g}}=\left(\mathbf{L}^{(p)}\right)^{-1}\mathbf{g}. \tag{4.15}$$

The N_i iterations proceed as

$$\mathbf{d}_0=\mathbf{B}^{(x)}\mathbf{u}^{n-1/2}+\mathbf{B}^{(z)}\mathbf{v}^{n-1/2}+\overline{\mathbf{g}}^{n-1/2}, \tag{4.16a}$$

$$\mathbf{d}_m = \mathbf{G}^{(p)}\mathbf{d}_{m-1}, \quad m > 0, \tag{4.16b}$$

$$\mathbf{p}^n - \mathbf{p}^{n-1} = \Delta t \sum_{m=0}^{N_i} \mathbf{d}_m. \tag{4.16c}$$

Likewise, following the same pattern but written in a concise form:

$$\mathbf{u}^{n+1/2} - \mathbf{u}^{n-1/2} = \Delta t \sum_{m=0}^{N_i} \left(\mathbf{G}^{(u)}\right)^m \mathbf{A}^{(x)}\mathbf{p}^n, \tag{4.17}$$

$$\mathbf{v}^{n+1/2} - \mathbf{v}^{n-1/2} = \Delta t \sum_{m=0}^{N_i} \left(\mathbf{G}^{(v)}\right)^m \mathbf{A}^{(z)}\mathbf{p}^n. \tag{4.18}$$

The factor Δt may be absorbed into $\mathbf{A}^{(x)}$, $\mathbf{A}^{(z)}$, $\mathbf{B}^{(x)}$, $\mathbf{B}^{(z)}$ and $\overline{\mathbf{g}}$ for efficiency.

Higher-order time stepping for equation (4.6) can be accomplished by the Cauchy-Kowalevsky or Lax-Wendroff or Dablain or modified equation method Dablain (1986); Lax and Wendroff (1960); Shubin and Bell (1987); von Kowalevsky (1875), which are all the same. Higher-order time stepping for equations (4.16) to (4.18) is easier to implement for the discrete form of the intermediate representation (4.3). The second-order time stepping discretization of the latter is

$$\mathbf{a}^n = \sum_{m=0}^{N_i} \left(\mathbf{G}^{(u)}\right)^m \mathbf{A}^{(x)}\mathbf{p}^n, \quad \mathbf{b}^n = \sum_{m=0}^{N_i} \left(\mathbf{G}^{(v)}\right)^m \mathbf{A}^{(z)}\mathbf{p}^n, \tag{4.19a}$$

$$\mathbf{p}^{n+1} - 2\mathbf{p}^n + \mathbf{p}^{n-1} = (\Delta t)^2 \sum_{m=0}^{N_i} \left(\mathbf{G}^{(p)}\right)^m \left[\mathbf{B}^{(x)}\mathbf{a}^n + \mathbf{B}^{(z)}\mathbf{b}^n + \overline{\mathbf{f}}^n\right], \tag{4.19b}$$

Note that higher-order time stepping can be avoided altogether by suitable post-processing of the recorded time series at the receivers, using Stork's dispersion correction method Anderson et al. (2015); Qin et al. (2017); Stork (2013); Wang and Xu (2015).

Some elements next to the surface topography may end up with zero pressures on all three vertices. If a receiver happens to be located inside that element, linear interpolation to its position will result in a dead trace. Higher-order mass-lumped or discontinuous Galerkin finite elements do not suffer from that nuisance.

Source term representations

We consider three different ways to discretize the source term, all in weak form. The delta function is the most straightforward, leading to a source term vector

$$\mathbf{s} = \int_{\Omega} \phi(\mathbf{x})\delta(\mathbf{x}-\mathbf{x}_s)\,\mathrm{d}\mathbf{x}. \tag{4.20}$$

Almost all entries are zero except for $\mathbf{s}_j = \phi_j(\mathbf{x}_s)$ on the three vertices $\mathbf{x}_j$ of the triangle that contains $\mathbf{x}_s$.

For a gaussian with standard deviation σ, we have

$$\mathbf{s} = C_\sigma \int_{\Omega} \phi(\mathbf{x}) e^{-(\mathbf{x}-\mathbf{x}_s)\cdot(\mathbf{x}-\mathbf{x}_s)/(2\sigma^2)}\,\mathrm{d}\mathbf{x}. \tag{4.21}$$

The normalization constant C_σ ensures that $\sum_{j=1}^{N} \mathbf{s}_j = 1$ when summed over all vertices, similar to integration of the delta function over the domain.

The tapered-sinc function in 2D reads

$$s(x,z) = \frac{1}{2}\left[1+\cos\left(\frac{\pi\zeta}{n_w+1}\right)\right]\frac{\sin\pi\zeta}{\pi\zeta}, \qquad \text{for } \zeta = \frac{1}{r_s}\sqrt{x^2+z^2} \le (1+n_w), \tag{4.22}$$

and zero otherwise. The integer n_w, typically 2 or 3, controls the length of the taper in terms of a number of extra loops of the sinc function and and $(1+n_w)r_s$ defines its actual radius. The corresponding source term vector is

$$\mathbf{s} = C_s \int_{\Omega} \phi(\mathbf{x})\, s(\mathbf{x}-\mathbf{x}_s)\,\mathrm{d}\mathbf{x}, \tag{4.23}$$

with normalization constant C_s.

Coercivity and the first-order formulation

To obtain some insight in the properties of the chain of first-order operators in (4.19), we consider its Fourier representation on a simple mesh (c.f. Shamasundar and Mulder, 2016). The mesh is assumed to consist of squares with sides of length h, each one divided in two triangles with relative positions $(0,0)$, $(h,0)$ and $(0,h)$ for one and $(h,0)$, (h,h), $(0,h)$ for the other. The pressure $p_{i,j}$ is defined on vertices (ih, jh). Shift operators are defined by $T_x p_{i,j} = p_{i+1,j}$ and $T_z p_{i,j} = p_{i,j+1}$. The mass matrix M and derivative operators D_x and D_z can then be expressed as

$$M = \tfrac{h^2}{12}\left[6 + T_x + T_x^{-1} + T_z + T_z^{-1} + T_x T_z^{-1} + T_x^{-1} T_z\right], \tag{4.24a}$$

$$D_x = \tfrac{h}{6}\left[2(T_x - T_x^{-1}) + T_z - T_z^{-1} + T_z^{-1} T_x - T_x^{-1} T_z\right], \tag{4.24b}$$

$$D_z = \tfrac{h}{6}\left[2(T_z - T_z^{-1}) + T_x - T_x^{-1} + T_x^{-1} T_z - T_z^{-1} T_x\right], \tag{4.24c}$$

with Fourier symbols

$$\hat{M} = \tfrac{h^2}{6}[3 + \cos\xi + \cos\eta + \cos(\xi - \eta)], \tag{4.25a}$$

$$\hat{D}_x = \tfrac{\mathrm{i}h}{3}[2\sin\xi + \sin\eta + \sin(\xi - \eta)], \tag{4.25b}$$

$$\hat{D}_z = \tfrac{\mathrm{i}h}{3}[2\sin\eta + \sin\xi + \sin(\eta - \xi)]. \tag{4.25c}$$

The scaled wavenumbers in x and z are $\xi = k_x h_x$ and $\eta = k_z h_z$, where $h_x = h_z = h$ denote the lengths of the sides of the squares and k_x and k_z the wavenumbers.

The corresponding second-order spatial operator is

$$B = M^{-1} D_x M^{-1} D_x + M^{-1} D_z M^{-1} D_z, \tag{4.26}$$

with symbol

$$\hat{B} = \frac{4}{h^2}[3 + \cos\xi + \cos\eta + \cos(\xi - \eta)]^{-2}\big\{5(\sin^2\xi + \sin^2\eta) + \\ 8\sin\xi\sin\eta + 2\sin(\xi - \eta)[\sin\xi - \sin\eta + \sin(\xi - \eta)]\big\}. \tag{4.27}$$

Near the origin of the wavenumber domain,

$$\hat{B} \simeq \tfrac{1}{h^2}(\xi^2 + \eta^2) - \tfrac{1}{180h^2}\big[2(\xi^6 + \eta^6) \\ - 5\xi\eta(\xi - \eta)^2(\xi^2 + \xi\eta + \eta^2)\big]. \tag{4.28}$$

The first term represent the exact operator, $k_x^2 + k_z^2$, the second the discretization error, which is clearly of order four relative to the exact operator.

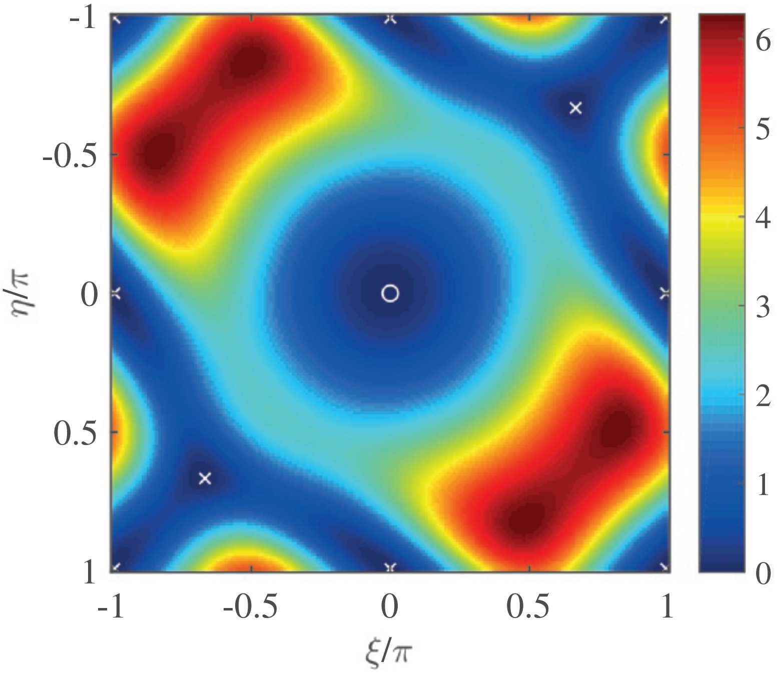


Figure 4.1: Symbol of minus the discrete laplace operator for the first-order form as a functions of the scaled horizontal and vertical wavenumber for a specific structured periodic mesh. Near the centre, the operator follows the exact one, $\xi^2+\eta^2$. Further away, the error is of order four, but still further away, the operator becomes zero in a number of points, marked by crosses. The zero at the centre, indicated by a circle, should be present, but the other cause a violation of coercivity.

Figure 4.1 displays $\hat{B}$ over the whole wavenumber domain. The bowl near the origin shows the term $\xi^2 + \eta^2$. At higher wavenumbers, the errors start to grow. Unfortunately, coercivity is not satisfied. The symbol $\hat{B}$ should vanish only at the origin, but it is also zero at the points $(\xi, \eta) = (m_1\pi, m_2\pi)$, with integer m_1 and m_2, and at $\eta = -\xi = \pm\frac{2}{3}\pi$. This means that, viewed as an elliptic operator, B is unstable. For the wave equation, there are certain waves that are not seen by the spatial operator. Once excited, they will start to live a life of their own and not disappear, except perhaps at an absorbing boundary. The net effect will be a noisy pressure wavefield. We therefore either have to abandon the first-order formulation altogether or ensure that such waves are not excited. A sufficiently band-limited source function can accomplish that. For the chosen structured periodic mesh, figure 4.1 suggests that wavenumbers for $\sqrt{\xi^2 + \eta^2} \lesssim \frac{1}{2}\pi$ or $\sqrt{k_x^2 + k_z^2} \lesssim \frac{1}{2}\pi/h$ should not be excited.

If we require the gaussian to have half its maximum amplitude in the wavenumber domain halfway the spectrum, this leads to a standard deviation $\sigma/h = (2/\pi)\sqrt{2\log 2} = 0.75$. In the weak form of equation (4.23) and with an inverse mass matrix, this is not expected to be very different.

A similar consideration can guide the choice of parameters for the tapered sinc. Figure 4.2 shows a number of dispersion curves for the first-order formulation in the 1-D case, taken from Shamasundar and Mulder (2016). For the first-derivative operator with a consistent mass matrix, the dispersion curve is given by $3\sin(\xi)/(2+\cos\xi)$ and is shown in red. With mass lumping, it is given by $\sin\xi$, shown in red, and its accuracy reduces to second order. One iteration produces the green curve, described by $\frac{1}{3}(4-\cos\xi)\sin\xi$, and restores fourth-order accuracy. To obtain the spectra for the tapered sinc, we choose a 1-D uniform periodic grid with element size h, placed a source at $0.2h$ from a vertex, evaluated equation (4.23), applied the inverse mass matrix and performed a Fourier transform. The precise position of the source inside an element does not seem to matter for the results shown in figure 4.2, obtained for $n_w = 3$ and $r_s = 2h$ or $r_s = 3h$. Larger values of n_w will make the transition from 1 to 0 steeper, at the expense of increasing the spatial source size, which will complicate matters when close to the free surface. We expect that parameters in this range will be close to optimal in the 2-D case.

In the weak form, the spatial part of the source function will be multiplied by the basis functions, after which it will be multiplied by the inverse of the mass matrix or its iterative approximation. The effect of this will be another second-order error, as can be seen easily by considering the same Fourier analysis as before. If the integration against the basis functions is denoted by a linear operator $\mathbf{\Phi}$, its Fourier symbol on the earlier structured mesh becomes

$$\hat{\Phi} = 2h^2 \frac{\sin\eta - \sin\xi + \sin(\xi-\eta)}{\xi\eta(\xi-\eta)}. \tag{4.29}$$

For small ξ,

$$h^{-2}\hat{\Phi} \simeq 1 - \tfrac{1}{12}(\xi^2+\eta^2-\xi\eta), \tag{4.30}$$

showing its second-order error. The inverse mass matrix does not compensate for that:

$$\hat{M}^{-1}\hat{\Phi} \simeq 1 + \tfrac{1}{12}(\xi^2+\eta^2-\xi\eta). \tag{4.31}$$

This suggests that we cannot obtain fourth-order convergence with the first-order formulation.

One may wonder if the second-order error term can be removed by adjusting the spatial source distribution. An attempt to recover fourth-order accuracy is presented in the appendix, for an equidistant mesh in 1D. The idea is to compensate the second-order impact of the discretization, $\hat{M}^{-1}\hat{\Phi}$, in the source function. In the 1D equidistant case, that can be accomplished easily. However, it is not clear how to generalize this idea to an unstructured 2-D mesh.

Finally, we remark that the symbol for the laplace operator in the second-order formulation on the chosen structured periodic mesh is

$$\tfrac{1}{h^2}\left[\sin^2(\xi/2)+\sin^2(\eta/2)\right] \simeq \tfrac{1}{h^2}\left[(\xi^2+\eta^2) - \tfrac{1}{12}(\xi^4+\eta^4)\right]. \tag{4.32}$$

It does not have coercivity problems but is only second-order accurate.

Results

We examine the performance of the two discretizations, in first- or second-order form, and three source terms, delta function, gaussian, or tapered sinc. The test problem is homogeneous with a constant sound speed $c = 1.5\,\text{km/s}$ and constant density $\rho = 1\,\text{g/cm}^3$.

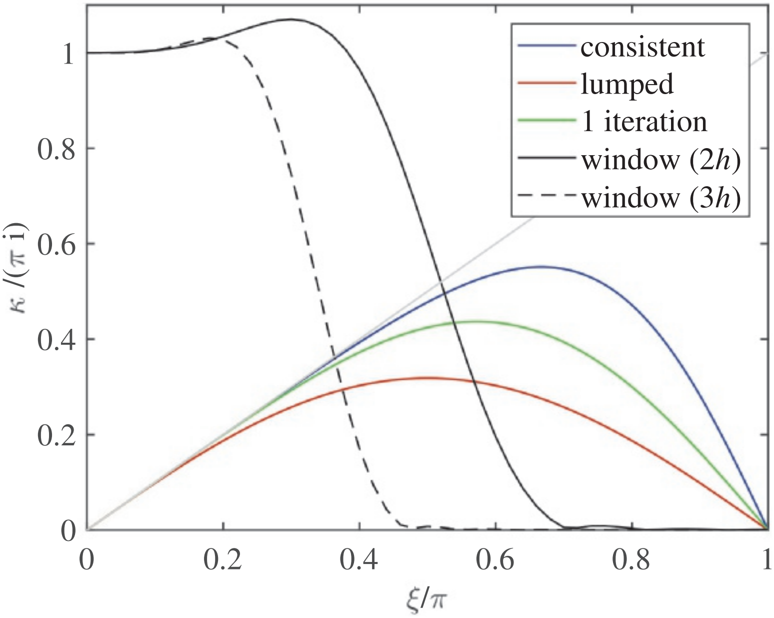


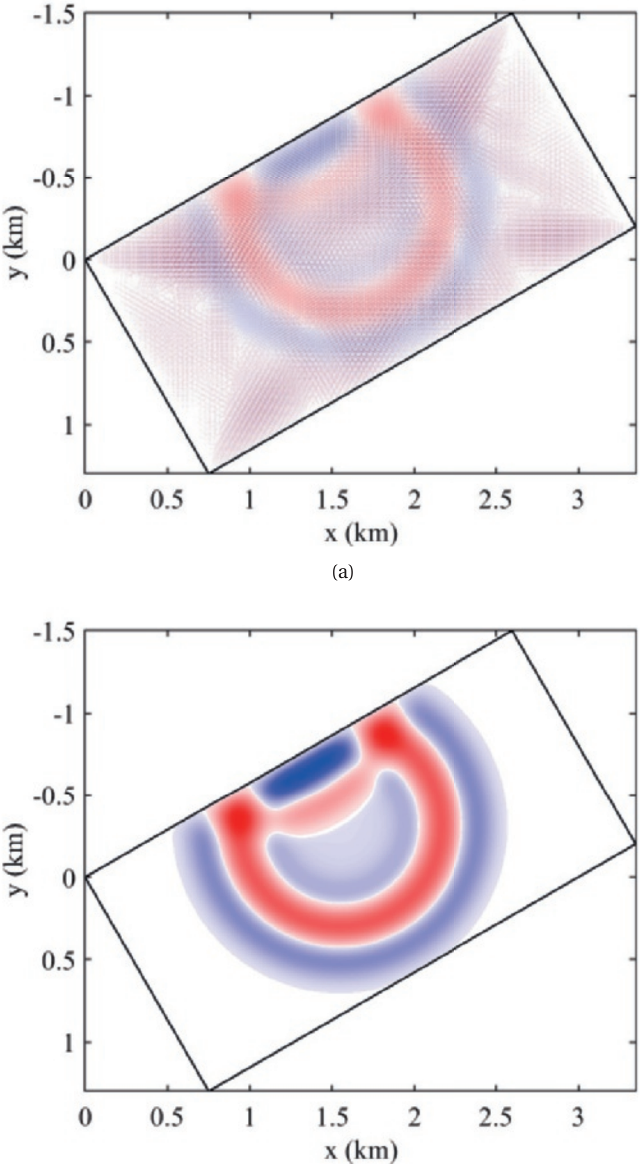
Figure 4.2: Dispersion curves for the first derivative and two tapered-sinc source window functions with $n_w = 3$ and $r_s = 2h$ or $3h$ that should suppress high wavenumbers towards the right, where the dispersion curves deviate strongly from the exact $\kappa = i\xi$ and coercivity is lost at the highest wavenumber.

The rectangular domain has a size of 3 by 1.5 km. A point source is located at $x_s = 1.5$ km and $z_s = 0.5$ km. The compactly supported wavelet is

$$w(t) = \begin{cases} -(T_w/8)^2 \frac{\mathrm{d}}{\mathrm{d}t}[1-(2t/T_w)^2]^8, & \text{if } |t| < \frac{1}{2}T_w, \\ 0, & \text{otherwise.} \end{cases} \tag{4.33}$$

The length of the wavelet, T_w, is related to peak frequency by $T_w = 0.934129/f_{\text{peak}}$ and we chose $f_{\text{peak}} = 3$ Hz. The simulations run from a time $-\frac{1}{2}T_w = -0.156$ to $t_{\max} = 0.45$ s. At that time the wave has reflected once against the free surface but has not yet reached the other boundaries, which we all take as zero dirichlet. The error in the pressure at $t_{\max}$ is measured at all vertices. The coordinates and velocities were rotated by 30° for testing purposes.

Figure 4.3a illustrates what happens if the violation of coercivity in the first-order formulation is ignored. The delta function as source generates short wavelengths that dominate the solution. With the tapered sinc, we obtain the result in figure 4.3b. For these examples, we used the consistent mass matrix and fourth-order time stepping with



(a)

(b)

Figure 4.3: Wavefield at 0.45 s for a delta function source (a) and for a tapered sinc (b). Positive pressures are red, negative blue.

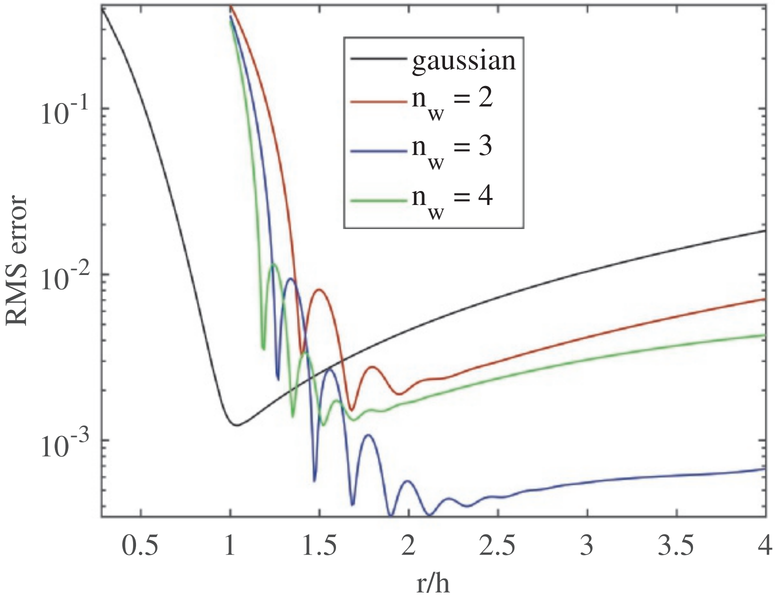


Figure 4.4: Scan over the source size r scaled by the element size h. The RMS error for a gaussian is smallest for a standard deviation around 1. For the tapered sinc, $n_w = 3$ performs best with $r/h = r_s/h$ between 2 and 3 for the current test problem.

a time step close to the stability limit. The unstructured mesh had 52412 vertices.

To find good parameters for the gaussian and tapered sinc, we computed the RMS error for the above problem over a range of source sizes, using the consistent mass matrix and a fourth-order time stepping method based on equation (4.19). The time step was chosen close to the stability limit. For the latter, we use $\Delta t \leq C \min_j (d_{\text{inner},j}/c_j)$ where d_{inner}/c is the ratio of the diameter of the inscribed circle over the sound speed and the minimum is taken over all triangles j. The constant C is estimated to be 1.36 with the consistent mass matrix in equation (4.19), $C = 2.41$ with mass lumping, $C = 1.76$ with one iteration and $C = 1.56$ with two. With fourth-order time stepping, these constants can be increased by a factor $\sqrt{3}$.

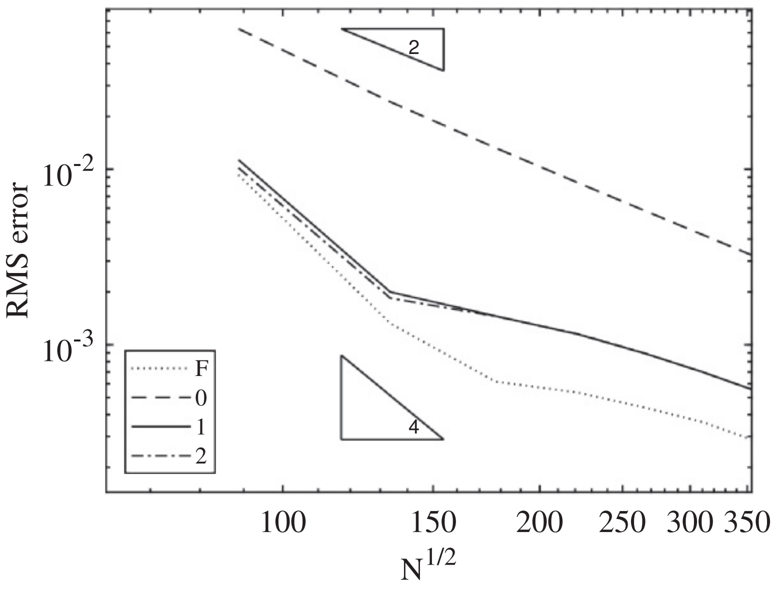
Figure 4.4 plots the RMS error as a function of the source size r scaled by the element size h, defined by the longest edge of the element that contains the source. The consistent mass matrix was used, despite its higher cost. For a gaussian, r is its standard deviation scaled by element size and the smallest error is obtained at $r/h = \sigma/h = 1.04$. The graphs for the tapered sinc are less smooth. The smallest error occurs for $n_w = 3$ and $r/h = r_s/h = 1.91$. The result is better than for a gaussian source. Here, h is the maximum edge length of the element that contains the source position.

Next, we study convergence on a range of meshes, from coarse to fine, both structured and unstructured. We use fourth-order time stepping and a source based on the tapered sinc with $n_w = 3$ and $r/h = r_s/h = 2$. Figure 4.5 shows the RMS error as a function of the square root of the number of degrees of freedom, N. The element size behaves as $N^{-1/2}$. The errors for an unstructured mesh in figure 4.5 a start out with fourth-order behaviour on the coarser grids but degrade to second-order on finer ones. Given the results of the 2-D Fourier analysis in the previous section, we cannot expect to do better than second-order. The results for the consistent or full mass matrix are included as a reference but are costly to compute. With mass lumping, the accuracy drops to second-order but one iteration provides a significant improvement in accuracy, as expected. With unstructured meshes, the errors are more erratic. Nevertheless, our defect-correction approach appears to pay off.

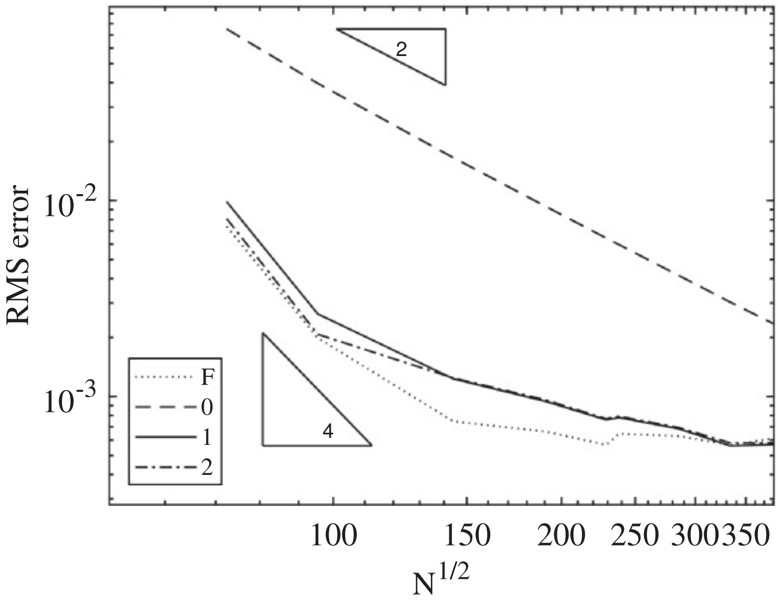
We repeated the exercise for the second-order formulation in equation (4.6), but now with second-order time stepping and mass lumping without iterations. With a gaussian source distribution, the smallest RMS error was obtained at $r/h = \sigma/h = 0.31$, but was only 4% smaller than with a delta function source. For the tapered sinc, the best result was found for $r/h = r_s/h = 0.92$ and also only 4% smaller than with a delta function source. Given the simplicity of latter, there seems to be no reason to replace it.

Comparing the errors for the second-order and first-order formulation, the latter is more accurate but requires more matrix-vector multiplications per time step. Its better accuracy and larger allowable time step are not sufficient to compensate for its higher cost, resulting in a lower efficiency than the simple second-order formulation. Although we observed this in our Matlab® implementations of both schemes, which is not really suited for measuring performance, we believe this will carry over to implementations in a compiled language like C or C++. Given the fact that higher-order mass-lumped schemes in second-order form are even more efficient Mulder (1996); Mulder and Shamasundar (2016), this makes the first-order formulation less attractive, although an acoustic fourth-order scheme in 3D requires 50 degrees of freedom per element Chin-Joe-Kong et al. (1999), considerably more than a first-order formulation with linear elements and 4 unknowns per vertex.

As an application, figure 4.6 a displays an inhomogeneous sound speed model. A



(a)



(b)

Figure 4.5: RMS error as a function of the square root of the number of degrees of freedom for the tapered sinc with $n_w = 3$ and $r/h = r_s/h = 2$ on structured (a) and unstructured (b) meshes. Results are shown for the consistent or full mass matrix (F) and for mass lumping with no (0), 1 or 2 additional iterations. The triangles indicate the slopes for second- and fourth-order convergence.

source at $x_s = 2468.36$ and $z_s = 410.351$ m with a 12-Hz Ricker wavelet generated the wavefield displayed in figure 4.6 b, using the first-order formulation with fourth-order time stepping. The tapered-sinc source had $n_w = 3$ and $r_s/h = 3$. The mesh contained 800466 elements and 401764 vertices. The computations started at −0.17 s to let the Ricker wavelet peak at zero time.

CONCLUSIONS

We have examined the performance of three source distributions, the delta function, a gaussian and a tapered sinc, in a finite-element formulation of the acoustic wave equation. In the standard second-order form, the gaussian and tapered sinc hardly improve the accuracy and a delta function appears to be the most attractive choice, given its simplicity.

Because the discrete first-order form of the wave equation is not coercive, it requires a cut-off of the short wavelengths. This disqualifies the delta function as source distribution. We have performed numerical experiments to find suitable parameters for the gaussian and for the tapered sinc. The latter provided the most accurate results.

The first-order form has a much better accuracy than the second-order form, but that does not appear sufficient to compensate for its higher cost, at least not in our 2-D Matlab® implementations.

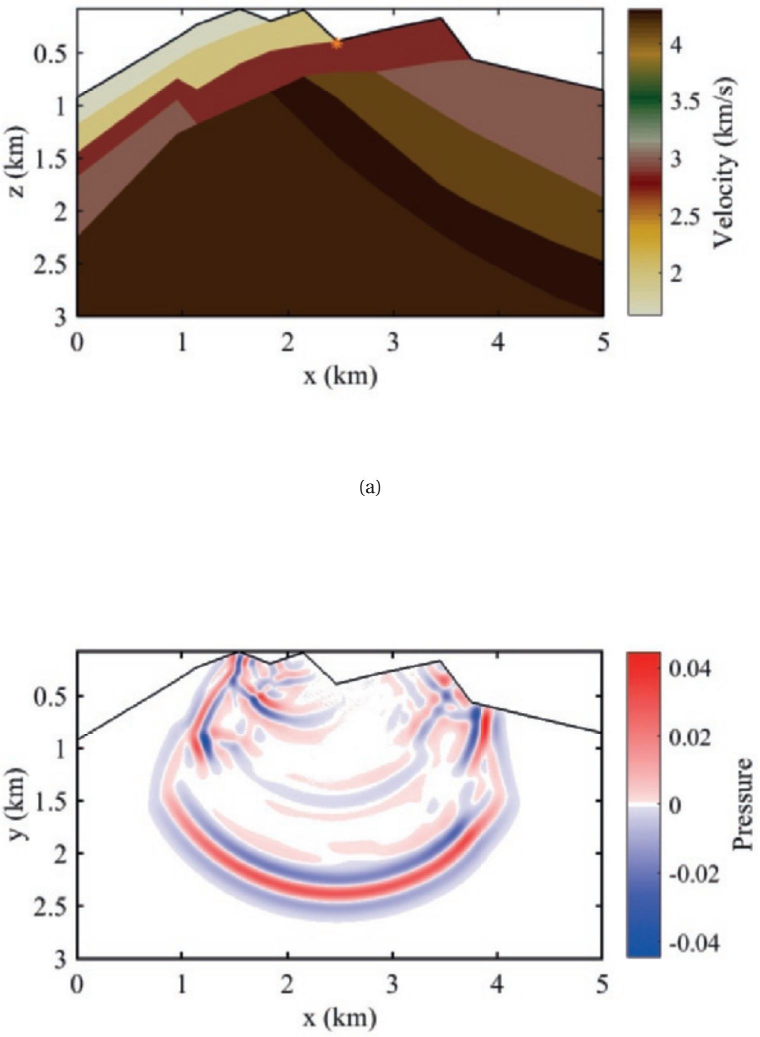
ACKNOWLEDGEMENTS

This work was partly funded by the Industrial Partnership Programme (IPP) 'Computational sciences for energy research' of the Foundation for Fundamental Research on Matter (FOM), which is part of the Netherlands Organisation for Scientific Research (NWO). This research programme is co-financed by Shell Global Solutions International B.V.

REFERENCES

Ainsworth, M., 2014. Dispersive behaviour of high order finite element schemes for the one-way wave equation. Journal of Computational Physics 259, 1–10.

Anderson, J. E., Brytik, V., Ayeni, G., 2015. Numerical temporal dispersion corrections for broadband temporal simulation, rtm and fwi. In: SEG Technical Program Expanded



(a)

(b)

Figure 4.6: (a) Velocity model for an inhomogeneous sound speed model, including topography. The orange star marks the source position. (b) Pressure wavefield at 0.5s.

Abstracts. pp. 1096–1100.

Basabe, J. D. D., Sen, M. K., 2007. Grid dispersion and stability criteria of some common finite-element methods for acoustic and elastic wave equations. Geophysics 72 (6), T81–T95.

Brossier, R., Virieux, J., Operto, S., 2008. Parsimonious finite-volume frequency-domain method for 2-d p-sv-wave modelling. Geophysical Journal International 175 (2), 541–559.

Cerjan, C., Kosloff, D., Kosloff, R., Reshef, M., 1985. A nonreflecting boundary condition for discrete acoustic and elastic wave equations. Geophysics 50 (4), 705–708.

Chin-Joe-Kong, M. J. S., Mulder, W. A., van Veldhuizen, M., 1999. Higher-order triangular and tetrahedral finite elements with mass lumping for solving the wave equation. Journal of Engineering Mathematics 35, 405–426.

Chung, E. T., Engquist, B., 2009. Optimal discontinuous Galerkin methods for the acoustic wave equation in higher dimensions. SIAM Journal on Numerical Analysis 47 (5), 3820–3848.

Cohen, G., Joly, P., Roberts, J. E., Tordjman, N., 2001. Higher order triangular finite elements with mass lumping for the wave equation. SIAM Journal on Numerical Analysis 38 (6), 2047–2078.

Cohen, G., Joly, P., Tordjman, N., 1995. Higher order triangular finite elements with mass lumping for the wave equation. In: Cohen, G., Bécache, E., Joly, P., Roberts, J. E. (Eds.), Proceedings of the Third International Conference on Mathematical and Numerical Aspects of Wave Propagation. SIAM, Philadelphia, pp. 270–279.

Cui, T., Leng, W., Lin, D., Ma, S., Zhang, L., 2017. High order mass-lumping finite elements on simplexes. Numerical Mathematics: Theory, Methods and Applications 10 (2), 331–350.

Dablain, M. A., 1986. The application of high-order differencing to the scalar wave equation. Geophysics 51 (1), 54–66.

Delcourte, S., Fezoui, L., Glinsky-Olivier, N., 2009. A high-order discontinuous Galerkin method for the seismic wave propagation 27, 70–89.

Diaz, J., Grote, M. J., 2009. Energy conserving explicit local time stepping for second-order wave equations. SIAM Journal on Scientific Computing 31 (3), 1985–2014.

Etienne, V., Chaljub, E., Virieux, J., Glinsky, N., 2010. An hp-adaptive discontinuous Galerkin finite-element method for 3-d elastic wave modelling. Geophysical Journal International 183 (2), 941–962.

Fried, I., Malkus, D. S., 1975. Finite element mass matrix lumping by numerical integration with no convergence rate loss. International Journal of Solids and Structures 11, 461–466.

Grote, M. J., Schneebeli, A., Schötzau, D., 2006. Discontinuous Galerkin finite element method for the wave equation. SIAM Journal on Numerical Analysis 44 (6), 2408–2431.

Hesthaven, J. S., Warburton, T., 2002. Nodal high-order methods on unstructured grids: I. time-domain solution of maxwell's equations. Journal of Computational Physics 181 (1), 186–221.

Hesthaven, J. S., Warburton, T., 2007. Nodal Discontinuous.

Hicks, G. J., 2002. Arbitrary source and receiver positioning in finite-difference schemes using.

Käser, M., Dumbser, M., 2006. An arbitrary high-order Discontinuous Galerkin method for elastic waves on unstructured meshes – I. The two-dimensional isotropic case with external source terms. Geophysical Journal International 166 (2), 855–877.

Komatitsch, D., Tromp, J., 1999. Introduction to the spectral-element method for 3-D seismic wave propagation. Geophysical Journal International 139 (3), 806–822.

Kononov, A., Minisini, S., Zhebel, E., Mulder, W. A., June 2012. A 3D tetrahedral mesh generator for seismic problems. In: Proceedings of the 74th EAGE Conference & Exhibition. p. B006.

Lax, P., Wendroff, B., 1960. Systems of conservation laws. Communications on Pure and Applied Mathematics 31 (2), 217–237.

Liu, Y., Teng, J., Xu, T., Badal, J., 2017. Higher-order triangular spectral element method with optimized cubature points for seismic wavefield modeling. Journal of Computational Physics 336, 458–480.

Modave, A., St-Cyr, A., Mulder, W., Warburton, T., 2015. A nodal discontinuous Galerkin method for reverse-time migration on GPU clusters. Geophysical Journal International 203 (2), 1419–1435.

Mulder, W. A., 1996. A comparison between higher-order finite elements and finite differences for solving the wave equation. In: Désidéri, J.-A., LeTallec, P., Oñate, E., Périaux, J., Stein, E. (Eds.), Proceedings of the Second ECCOMAS Conference on Numerical Methods in Engineering. John Wiley & Sons, Chichester, pp. 344–350.

Mulder, W. A., 1999. Spurious modes in finite-element discretisations of the wave equation may not be all that bad. Applied Numerical Mathematics 30, 425–445.

Mulder, W. A., 2013. New triangular mass-lumped finite elements of degree six for wave propagation. Progress In Electromagnetics Research 141, 671–692.

Mulder, W. A., Shamasundar, R., 2016. Performance of continuous mass-lumped tetrahedral elements for elastic wave propagation with and without global assembly. Geophysical Journal International 207 (1), 414–421.

Petersson, N. A., O'Reilly, O., Bj, 2016. Discretizing singular point sources in hyperbolic wave propagation problems. Journal of Computational Physics 321, 532–555.

Qin, Y., Quiring, S., Nauta, M., 2017. Temporal dispersion correction and prediction by using spectral mapping. In: 79th EAGE Conference & Exhibition, Paris, France, Extended Abstracts.

Riviere, B., Wheeler, M. F., 2003. Discontinuous finite element methods for acoustic and elastic wave problems. Vol. 329. Providence, RI: American Mathematical Society, pp. 4–6.

Shamasundar, R., Mulder, W., 2018. Numerical noise suppression for wave propagation with finite elements in first-order form by an extended source term. Geophysical Journal International 215 (2), 1231–1240.

Shamasundar, R., Mulder, W. A., 2016. Improving the accuracy of mass-lumped finite-elements in the first-order formulation of the wave equation by defect correction. Journal of Computational Physics 322, 689–707.

Shubin, G. R., Bell, J. B., 1987. A modified equation approach to constructing fourth order methods for acoustic wave propagation. SIAM Journal on Scientific and Statistical Computing 8 (2), 135–151.

Stetter, H. J., 1978. The defect correction principle and discretization methods. Numerische Mathematik 29 (4), 425–443.

Stork, C., 2013. Eliminating nearly all dispersion error from fd modeling and rtm with minimal cost increase. In: 75th EAGE Conference & Exhibition incorporating SPE EUROPEC, Extended Abstract.

von Kowalevsky, S., 1875. Zur Theorie der partiellen Differentialgleichung. Journal fur die reine und angewandte Mathematik 80, 1–32.

Walden, J., 1999. On the approximation of singular source terms in differential equations. Numerical Methods for Partial Differential Equations 15 (4), 503–520.

Wang, M., Xu, S., 2015. Time dispersion transforms in finite difference of wave propagation. In: 77th EAGE Conference & Exhibition, Extended Abstract.

Wilcox, L. C., Stadler, G., Burstedde, C., Ghattas, O., 2010. A high-order discontinuous Galerkin method for wave propagation through coupled elastic-acoustic media. Journal of Computational Physics 229 (24), 9373–9396.

Zhebel, E., Minisini, S., Kononov, A., Mulder, W. A., 2014. A comparison of continuous mass-lumped finite elements with finite differences for 3-D wave propagation. Geophysical Prospecting 62 (5), 1111–1125.

5

Performance of continuous mass-lumped tetrahedral elements for elastic wave propagation with and without global assembly

Parts of this chapter have been published in Geophysical Journal International, 207(1), pp.414-421 (Mulder and Shamasundar, 2016).

We consider isotropic elastic wave propagation with continuous mass-lumped finite elements on tetrahedra with explicit time stepping. These elements require higher-order polynomials in their interior to preserve accuracy after mass lumping and are only known up to degree 3. Global assembly of the symmetric stiffness matrix is a natural approach but requires large memory. Local assembly on the fly, in the form of matrix-vector products per element at each time step, has a much smaller memory footprint. With dedicated expressions for local assembly, our code ran about 1.3 times faster for degree 2 and 1.9 times for degree 3 on a simple homogeneous test problem, using 24 cores. This is similar to the acoustic case. For a more realistic problem, the gain in efficiency was a factor 2.5 for degree 2 and 3 for degree 3. For the lowest degree, the linear element, the expressions for both the global and local assembly can be further simplified. In that case, global assembly is more efficient than local assembly. Among the three degrees, the element of degree 3 is the most efficient in terms of accuracy at a given cost.

Introduction

Finite-difference modelling of seismic wave propagation has become the workhorse of the industry for imaging hydrocarbon reservoirs. The spectral finite-element method plays a similar rôle in seismology. Higher-order finite-difference methods have problems with sharp material contrasts and topography, because they assume differentiability where it does not hold. Modifications can alleviate the decrease in accuracy, but at a cost in terms of complexity and compute time. Finite-element methods have an inherently larger computational cost, but do not suffer from a loss of accuracy if the mesh follows the interfaces between different materials and the topography. Because of their better accuracy, they may outperform the finite-difference method in some cases (Moczo et al., 2011; Mulder, 1996; Wang et al., 2010; Zhebel et al., 2014, e.g.). However, mesh generation can sometimes be difficult.

Spectral finite elements (Komatitsch and Tromp, 1999; Maday and Ronquist, 1990; Orszag, 1980; Patera, 1984; Seriani et al., 1992) require hexahedral meshes. Tetrahedral elements offer more flexibility in gridding, for instance, near pinch-outs. Suitable schemes are discontinuous Galerkin (DG) methods (Dumbser and Käser, 2006; Etienne et al., 2010; Käser and Dumbser, 2006; Riviere and Wheeler, 2003; Wilcox et al., 2010, e.g.), rectangular spectral elements mapped to triangles or tetrahedra (Mercerat et al., 2006; Sherwin and Karniadakis, 1995), hybridized versions (Cockburn et al., 2009; Giorgiani et al., 2013), finite-volume methods (Brossier et al., 2008; Dumbser et al., 2007), mixed methods (Bécache et al., 2002; Cohen and Fauqueux, 2005) or continuous mass-lumped finite elements, which we will consider here. DG methods offer the advantage that they can mix orders and types of elements on, for instance, hexahedra, tetrahedra and prisms, and also can work on non-conforming meshes. However, the fluxes required to impose continuity increase the computational cost. Since the mass matrix is block diagonal, its inversion is not costly.

Continuous mass-lumped triangular or tetrahedral finite elements avoid the cost of inverting a large sparse mass matrix by lumping the mass matrix into a diagonal one. (Fried and Malkus, 1975) noted, however, that with quadratic 2-D triangular elements, the lumping decreases the order of accuracy. They considered the heat equation, but the same holds for the acoustic and elastic wave equations in second-order form. Augmenting the element with polynomials of a higher degree in the interior can repair this deficiency (Fried and Malkus, 1975). For the element of degree 2 in 2D, a bubble function that vanishes on the edges suffices. (Tordjman, 1995) and (Cohen et al., 1995) used this idea to construct a 2-D element of degree 3 on the edges and a bubble function times a polynomial of degree 1 in the interior, leading to an interior degree of 4. (Cohen et al., 2001) provides error estimates. (Mulder, 1996) found an element of degree 4 and interior degree 5. (Chin-Joe-Kong et al., 1999) found several elements of degree 5. The highest degree for mass-lumped triangular elements known so far is 6 (Mulder, 2013).

(Mulder, 1996) made the generalization to tetrahedral elements with an element of degree 2 on the edges, 4 on the faces as product of a cubic bubble function and polynomial of degree 1, and degree 4 in the interior as a product of a quartic bubble function and constant polynomial. (Lesage et al., 2010) and (Zhebel et al., 2011) applied that element to acoustic wave propagation modelling. (Chin-Joe-Kong et al., 1999) constructed 2 elements of degree 3. The second one allows for a larger time step than the first (Zhebel et al., 2011, 2014) and will be used in the current paper. Elements of higher degree have not been found so far. (Mulder et al., 2014) list stability estimates for the known tetrahedral lumped elements of degrees 1 to 3 as well as for the symmetric interior-penalty discontinuous Galerkin method up to degree 4.

(Bao et al., 1998) worked with the classic linear tetrahedral mass-lumped elements for elastic wave propagation modelling. Here, we will also include elements of degree 2 and 3.

With explicit time stepping, we can consider two approaches for assembling the stiffness and diagonal, lumped mass matrix: global assembly or local assembly on the fly. Global assembly is a standard approach with finite elements. The elements of the lumped mass matrix or its inverse can be represented by one value per node. For the symmetric global stiffness matrix, we store the symmetric block diagonal and the block upper triangular part separately, the latter in Block Compressed Sparse Row format. With local assembly on the fly, the contribution of each element to the solution update is treated independently. The displacement components on the nodes of one element are copied from a global vector and multiplied by precomputed stiffness matrices on the reference element, nine in total. The results are then combined by geometrical factors that handle the map from the reference element to the actual element, multiplied by the inverse mass matrix, and used to increment the global solution vector for the new time level.

One might expect global assembly to produce results quicker than local assembly, at the expense of considerably larger storage, but as it turns out, this does not appear to be the case for the acoustic wave equation. The main question we address here is if a similar results also holds in the elastic case. To obtain performance figures within the same order of magnitude, we derive dedicated expressions for the matrix-vector multiplications that are part of the local assembly on the fly.

In Section 5.2, we describe the discretization and provide expressions for global assembly and local assembly for the general case. Simpler expressions are provided for linear elements. Section 5.3 presents results for global and local assembly on 24 cores. We start with the linear element. Then, we briefly consider the acoustic case, where local assembly outperforms global assembly for degree 3, before turning towards degree 2 and 3 for the isotropic elastic case. The section ends with a slightly more realistic example. Section 3.5 summarizes the main conclusions.

METHOD

DISCRETIZATION

The elastic system of wave equations for an isotropic medium in second-order form is

$$\rho \frac{\partial^2 u_m}{\partial t^2} = \sum_{j=1}^{3} \left[\frac{\partial}{\partial x_m}\left(\lambda \frac{\partial u_j}{\partial x_j}\right) + \frac{\partial}{\partial x_j}\left\{\mu\left(\frac{\partial u_m}{\partial x_j} + \frac{\partial u_j}{\partial x_m}\right)\right\}\right] + s_m.$$

The displacement in coordinate direction x_m, $m = 1,2,3$, is $u_m(t,\mathbf{x})$ as a function of time, t, and position, $\mathbf{x}$. The material properties are density $\rho(\mathbf{x})$ and Lamé parameters $\mu(\mathbf{x}) = \rho v_s^2$ and $\lambda(\mathbf{x}) = \rho v_p^2 - 2\mu$, with P-wave velocity $v_p(\mathbf{x})$ and S-wave velocity $v_s(\mathbf{x})$. The forcing source function is typically of the form $s_m(t,\mathbf{x}) = f_m w(t)\delta(\mathbf{x}-\mathbf{x}_s)$, with wavelet $w(t)$ and force amplitude f_m at a source position $\mathbf{x}_s$. The domain consists of a subset of the Earth, bounded by a free surface. In exploration geophysics, absorbing boundaries are usually implemented on the sides where the domain is truncated.

The domain is meshed by tetrahedra, preferably such that the element size scales with the shear velocity, v_s (Kononov et al., 2012; Mulder et al., 2014). As wavelength scales with velocity, this provides a more or less uniform resolution over the entire mesh.

Here, the material parameters are assumed to be constant per element.

Next, we define the geometrical components (Zienkiewicz and Taylor, 2000, Chapter 9, e.g.). Let the four vertices of the tetrahedron be denoted by $\mathbf{x}_k$, $k = 0,1,2,3$. In terms of reference element, $\mathbf{x} = \sum_{k=0}^{3} \mathbf{x}_k \phi_k(\mathbf{x})$ with the basis functions, ϕ_k, of the linear element. The natural coordinates on the tetrahedron are $\xi_k = \phi_k$ for $k = 1,2,3$, augmented with $\phi_0 = \xi_0 = 1 - \xi_1 - \xi_2 - \xi_3$. The coordinate transformation is $\mathbf{x} = \mathbf{x}_0 + \xi_1(\mathbf{x}_1 - \mathbf{x}_0) + \xi_2(\mathbf{x}_2 - \mathbf{x}_0) + \xi_3(\mathbf{x}_3 - \mathbf{x}_0) = \sum_{k=0}^{3} \xi_k \mathbf{x}_k$ with Jacobian matrix $\mathbf{J} = \frac{\mathrm{d}\mathbf{x}}{\mathrm{d}\boldsymbol{\xi}} = (\mathbf{x}_a, \mathbf{x}_b, \mathbf{x}_c)$. It is convenient to define relative vertex positions

$$\mathbf{x}_a = \mathbf{x}_1 - \mathbf{x}_0, \quad \mathbf{x}_b = \mathbf{x}_2 - \mathbf{x}_0, \quad \mathbf{x}_c = \mathbf{x}_3 - \mathbf{x}_0,$$

and the cross products

$$\mathbf{g}_1 = \mathbf{x}_b \times \mathbf{x}_c, \quad \mathbf{g}_2 = \mathbf{x}_c \times \mathbf{x}_a, \quad \mathbf{g}_3 = \mathbf{x}_a \times \mathbf{x}_b.$$

Note that

$$\mathbf{g}_1 \times \mathbf{g}_2 = J_0 \mathbf{x}_c, \quad \mathbf{g}_2 \times \mathbf{g}_3 = J_0 \mathbf{x}_a, \quad \mathbf{g}_3 \times \mathbf{g}_1 = J_0 \mathbf{x}_b.$$

Then, $\det \mathbf{J} = J_0 = \mathbf{x}_a \cdot \mathbf{g}_1 = 6V$, with V the volume of the tetrahedron. The matrix $\mathbf{F} = J_0 \mathbf{J}^{-\mathsf{T}}$ has $\mathbf{g}_k$, $k = 1,2,3$, as columns.

The mass matrix $\mathbf{A}$ on the reference element has elements

$$A_{j,k} = \int_0^1 \mathrm{d}\xi_1 \int_0^{1-\xi_1} \mathrm{d}\xi_2 \int_0^{1-\xi_1-\xi_2} \mathrm{d}\xi_3 \;\; \phi_j(\boldsymbol{\xi}) \phi_k(\boldsymbol{\xi}),$$

for $j,k = 0,1,2,3$. Mass lumping replaces this matrix by a diagonal one with the row sums as the diagonal elements: $A^{\mathrm{L}}_{j,k} = \delta_{j,k} \sum_{k=0}^{3} A_{j,k}$.

The nine stiffness matrices $\mathbf{B}^{m,n}$ on the reference element are

$$B^{m,n}_{j,k} = \int_0^1 \mathrm{d}\xi_1 \int_0^{1-\xi_1} \mathrm{d}\xi_2 \int_0^{1-\xi_1-\xi_2} \mathrm{d}\xi_3 \;\; \frac{\partial \phi_j}{\partial \xi_m} \frac{\partial \phi_k}{\partial \xi_n}.$$

They are symmetric: $\mathbf{B}^{n,m} = (\mathbf{B}^{m,n})^{\mathsf{T}}$. For the higher order mass-lumped finite elements, the coordinate permutations listed in Appendix D can simplify the implementation. Then, the code only has to define two arrays with pre-computed values on the reference element, for instance, $\mathbf{B}^{1,1}$ and $\mathbf{B}^{1,2}$, and the other 7 follow from permutations and symmetries.

The stiffness matrix $\overline{\mathbf{B}}$ for the isotropic elastic system of equations per element can be constructed from the above $\mathbf{B}^{m,n}$. To obtain a matrix, the displacement components are taken as fastest index and the nodes as slowest. Then, the matrix elements are

$$\begin{aligned} &J_0 \overline{B}_{m+3j,n+3k} = \\ &\quad \sum_{p,q=1}^{3} F_{m,p} F_{n,q} \left(\lambda B_{j,k}^{p,q} + \mu B_{k,j}^{p,q} \right) + \\ &\qquad\qquad \mu \delta_{m,n} \sum_{p,q,r=1}^{3} F_{r,p} F_{r,q} B_{k,j}^{p,q}. \end{aligned} \tag{5.1}$$

Here m and n run over the 3 components of the displacement, whereas j and k run over the nodes of the element: $m,n = 1,2,3$ and $j,k = 0,1,\ldots,N_p - 1$. The number of nodes for the mass-lumped elements is $N_p = 4$ for degree 1, 23 for degree 2 and 50 for degree 3. The global stiffness matrix follows from the contributions of $\overline{\mathbf{B}}$ per element.

The upper triangular part of the sparse symmetric block matrix is stored in Block Compressed Sparse Row format, with 3 × 3 full blocks. The block diagonal is treated separately, as the small 3 × 3 blocks are symmetric and only 6 values need to be stored per element. Somewhat to our surprise, we found that our code, using OpenMP, outperformed the Intel® Math Kernel Library routine `mkl_cspblas_scsrsymv()` that also uses OpenMP.

With local assembly, we can exploit the fact that the stiffness matrices $\mathbf{B}^{m,n}$ on the reference element have a zero row sum and, since they are symmetric, also a zero column sum. The zero row sum implies that the application of a stiffness matrix to a constant produces zero. We therefore define

$$v_k^m = u_k^m - u_0^m, \tag{5.2}$$

for nodes $k = 1,\ldots,N_p - 1$ and components $m = 1,2,3$, subtracting the values of the displacement components at the first vertex that corresponds to $k = 0$. Note that any node of the element can be selected here, with the first or last as a convenient choice. Let $\mathbf{r} = \overline{\mathbf{B}}\mathbf{u} = \overline{\mathbf{B}}\mathbf{v}$ per element. The zero column sum of the stiffness matrix implies

$$r_0^m = -\sum_{k=1}^{N_p-1} r_k^m, \quad m = 1,2,3. \tag{5.3}$$

This means that we can drop the first three rows and columns of the local elastic stiffness matrix $\overline{\mathbf{B}}$, work with v_k^m for $k = 1, \dots, N_p - 1$ and $m = 1,2,3$, and reconstruct the first three entries of r_k^m by equation (5.3). The result has to be multiplied by the precomputed inverse of the diagonal global mass matrix and can then be used to increment the solution. Repeating this for all tetrahedra accomplishes the time step, together with the source term and interpolation to obtain the receiver traces at selected positions.

We can further simplify the evaluation of $\overline{\mathbf{B}}\mathbf{v}$. Let $\mathbf{F}^\lambda = \frac{\lambda}{J_0}\mathbf{F}$, $\mathbf{F}^\mu = \frac{\mu}{J_0}\mathbf{F}$ and define the symmetric 3×3 matrix $\mathbf{C}^\mu = \frac{\mu}{J_0}\mathbf{F}^\mathsf{T}\mathbf{F} = \mathbf{F}^\mathsf{T}\mathbf{F}^\mu$. Define a set of 9 vectors for $p = 1,2,3$ and $q = 1,2,3$ with elements

$$\sigma_j^{p,q;n} = \sum_{k=1}^{N_p-1} B_{j,k}^{p,q}(u_k^n - u_0^n) = \sum_{k=1}^{N_p-1} B_{j,k}^{p,q} v_k^n,$$

for components $n = 1,2,3$ and nodes $j = 1, \dots, N_p - 1$, ignoring node 0. Compute

$$\alpha_j^p = \sum_{n,q=1}^{3} \left(F_{n,q}^\lambda \sigma_j^{p,q;n} + F_{n,q}^\mu \sigma_j^{q,p;n} \right).$$

Then,

$$r_j^m = \sum_{p=1}^{3} F_{m,p}\alpha_j^p + \sum_{p,q=1}^{3} C_{p,q}^\mu \sigma_j^{q,p;m}.$$

for nodes $j = 1, \dots, N_p - 1$ and components $m = 1,2,3$. Finally, use equation (5.3) to obtain the values at node $j = 0$, multiply by the subset of the global inverse matrix on the element and update the solution. For degrees higher than one, the main computational effort consists in the 9 matrix-vector products between the matrices $\mathbf{B}^{p,q}$ of the reference element and the vector $\mathbf{v}$.

The standard second-order time stepping scheme reads

$$\mathbf{u}^{n+1} = 2\mathbf{u}^n - \mathbf{u}^{n-1} + (\Delta t)^2 \mathcal{M}^{-1}(\mathbf{f} - \mathcal{K}\mathbf{u}^n),$$

with global stiffness matrix $\mathcal{K}$ and diagonal global mass matrix $\mathcal{M}$. The inverse of mass matrix can in principle be avoided by considering the diagonal scaling

$$\mathcal{D} = \Delta t \mathcal{M}^{-1/2}, \quad \tilde{\mathbf{u}} = \mathcal{D}^{-1}\mathbf{u}, \quad \tilde{\mathbf{f}} = \mathcal{D}\mathbf{f},$$

and the symmetric matrix

$$\tilde{\mathcal{K}} = \mathcal{D}\mathcal{K}\mathcal{D},$$

leading to

$$\tilde{\mathbf{u}}^{n+1} = 2\tilde{\mathbf{u}}^{n} - \tilde{\mathbf{u}}^{n-1} + \tilde{\mathbf{f}} - \tilde{\mathcal{K}}\tilde{\mathbf{u}}^{n}.$$

However, we have not used this approach for the numerical experiments reported further on as it complicates reading off receiver data. The required storage then consists in the solution at two time instances, which requires 3 times the number of nodes, the inverse mass matrix multiplied by $(\Delta t)^2$, also with a size equal to the number of nodes, and either the globally assembled sparse matrix or, with local assembly, the average of λ and of μ per element.

LINEAR ELEMENT

The above expressions hold for any degree. For the linear element, we derive simpler expressions that will speed up the code. Let $\mathbf{g}_0 = -\mathbf{g}_1 - \mathbf{g}_2 - \mathbf{g}_3$ and define a linear array $g_{m+3k} = g_k^m$, with nodes $k = 0,1,2,3$ and components $m = 1,2,3$. Note that $g_k^m = F_{m,k}$ for $k = 1,2,3$. Let $\mathbf{g}^\lambda = \lambda/(6J_0)\mathbf{g}$ and $\mathbf{g}^\mu = \mu/(6J_0)\mathbf{g}$. Table 5.1 lists pseudo-code in Matlab® style for the evaluation of the local stiffness matrix, $\overline{\mathbf{B}}$. When recoded in a language like C or C++, this code is more efficient than that of (Alberty et al., 2002), which is geared towards use with Matlab® .

For local assembly, let

$$s_m = \sum_{k=1}^{3} F_{m,k} v_k^m, \quad w_{m,m} = 2\frac{\mu}{6J_0} s_m,$$

and

$$w_{m,n} = w_{n,m} = \frac{\mu}{6J_0} \sum_{k=1}^{3} \left(F_{m,k} v_k^n + F_{n,k} v_k^m\right),$$

for $m < n$. Then, a simpler expression is

$$r_k^m = F_{m,k}\left(w_{m,m} + \frac{\lambda}{6J_0}\sum_{n=1}^{3} s_n\right) + \sum_{\substack{n=1\\ n\neq m}}^{3} F_{n,k} w_{m,n},$$

for nodes $k = 1,2,3$ and components $m = 1,2,3$. Equation (5.3) provides the values at

Table 5.1: Pseudocode in Matlab® style for the evaluation of the stiffness matrix B per element for linear basis functions on a tetrahedron, with `glamba` as $\mathbf{g}^{\lambda}$ and `gmu` as $\mathbf{g}^{\mu}$, defined in the text. Unknowns are taken as triples of displacements on vertices 0 to 3.

```
glamba = (lambda/(6*J0))*g; gmu = (mu/(6*J0))*g;
B = zeros(12,12);
for k1=0:3:9,
  for k2=0:3:k1,
    s = 0;
    for m2=1:3,
      for m1=1:3,
        h1 = gla(k1+m1)*g(k2+m2);
        h2 = gmu(k1+m1)*g(k2+m2);
        B(k1+m1,k2+m2) = h1+h2;
        if(m1 == m2), s = s+h2; end
      end
    end
    for m=1:3,
      B(k1+m,k2+m) = B(k1+m,k2+m)+s;
    end
    % copy symmetric elements
    if(k2 < k1),
       for m2=1:3,
         for m1=1:3,
           B(k2+m2,k1+m1) = B(k1+m1,k2+m2);
         end
       end
    end
  end
end
```

node $k = 0$.

ACOUSTICS

For the *acoustic* case, which we will briefly consider later on, it is convenient to define symmetric matrices

$$\overline{\mathbf{C}} = J_0 \mathbf{J}^{-1} \mathbf{J}^{-\top} = J_0^{-1} \mathbf{F}^\top \mathbf{F},$$

and

$$\tilde{\mathbf{B}}^{p,q} = \mathbf{B}^{p,q} + \mathbf{B}^{q,p} = \mathbf{B}^{p,q} + (\mathbf{B}^{p,q})^\top.$$

The contribution of an element to the stiffness matrix is

$$\overline{\mathbf{B}}^{\text{acou}} = \sum_{p=1}^{3} \left(\overline{C}_{p,p} \mathbf{B}^{p,p} + \sum_{q=p+1}^{3} \overline{C}_{p,q} \tilde{\mathbf{B}}^{p,q} \right), \tag{5.4}$$

where $\mathbf{B}^{p,p}$ and $\tilde{\mathbf{B}}^{p,q}$ are symmetric matrices on the reference element, containing predetermined numerical values only, and $\overline{\mathbf{C}}$ deals with the geometry of the actual tetrahedron. For the linear element, the simplified expressions presented by (Zhebel et al., 2014) are more efficient. For degree 2 and 3 and with local assembly, the evaluation of $\overline{\mathbf{B}}^{\text{acou}}\mathbf{u}$ per element was implemented as 6 matrix-vector multiplications, namely $\mathbf{B}^{p,p}\mathbf{u}$ and $\tilde{\mathbf{B}}^{p,q}\mathbf{u}$. The vector $\mathbf{u}$ contains the pressure values on the nodes of the element. The matrices correspond to those in (5.4) and were hardcoded from numerical values computed with Mathematica® .

RESULTS

LINEAR ELEMENT

As a test problem, we chose a homogeneous problem for which the exact solution is readily available. The constant material properties were a density $\rho = 2\,\text{g/cm}^3$, a P-wave velocity $v_p = 2\,\text{km/s}$ and an S-wave velocity $v_s = 1.2\,\text{km/s}$. The domain had a size $[-2,2] \times [-1,1] \times [0,2]\,\text{km}^3$ and was divided into cubes with an edge length of 20 m. Each cube was partitioned into 6 tetrahedra, leading to 12,000,000 tetrahedra and 2,050,401 vertices. The cube has six possible tetrahedral decompositions. We used the periodic one, with matching diagonals on opposite faces and one diagonal to the cube's centre.

A vertical force source was placed at the centre of the domain. A line of receivers

Table 5.2: Performance on linear elements with global assembly of the stiffness matrix and with local assembly. For the latter, the wall-clock time with 24 threads is doubled in this particular example but less storage is needed.

assembly	threads	assembly	stepping	total	storage
global	24	7.9 s	9.2 s	17.0 s	3.0 GByte
	12	7.9 s	10.7 s	18.6 s	
	6	8.2 s	16.8 s	25.1 s	
local	24		30.0 s	30.0 s	2.1 GByte
	12		57.8 s	57.8 s	
	6		114 s	114 s	

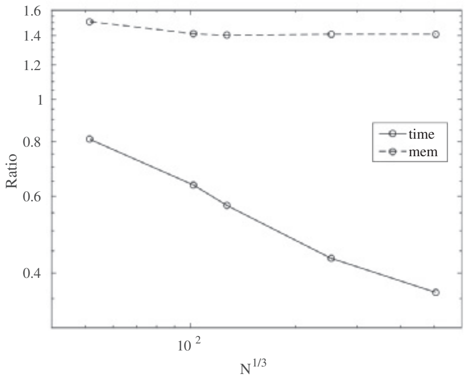


Figure 5.1: Ratios for compute time (drawn line) and storage (dashed) with linear elements as a function of $N^{1/3}$, where N is the number of nodes on the mesh. The results for the globally assembled case were divided by those for locally assembled stiffness matrices. The latter requires less storage, but is slower. The obtained reduction in storage does not seem to justify the larger compute times with linear elements.

was located at a depth of 800 m with $y = 0$ m and x between -1925 and $+1925$ m, using a 50-m interval. The time steps started at -0.3875 s to let the 3.5-Hz Ricker wavelet peak at zero time. The time step, $\Delta t = 0.003125$ s, corresponded to 0.77% of CFL-limit. Data were recorded up to 0.6 s at a 5-ms interval. We used the natural (free-surface) boundary conditions all around for simplicity.

Table 5.2 lists the timings and storage requirements using 24, 12 or 6 threads, all for the same mesh described earlier. Throughout this paper, reported timings are the average of 5 runs. The table shows that a smaller number of threads does not lead to a severe performance drop with global assembly, because memory access is the limiting factor. For local assembly on the fly, the performance is limited by the available compute power, at least up to the available 24 cores. OpenMP directives handled the multi-threading. The hardware consisted of a single board with two 12 core Intel® Xeon® CPU E5-2680 v3 processors running at 2.50 GHz and had hyper-threading disabled.

Figure 5.1 shows the ratios between the runs with global assembly and those with local assembly, in terms of the required compute time and the maximum required storage, for a range of mesh sizes. Global assembly requires about 40% more storage, but the gain in performance appears to amply justify that. Table 5.2 suggests that we could have used less than 24 threads for local assembly, as the computations are bound by memory access.

Note that the performance data should be taken as a rough indication, since the results strongly depend on code implementation, optimization and compiler. We did not put a lot of effort in code tuning for the specific compiler and hardware, but instead relied on the basic formulation of the method and the optimization capabilities of the Intel® compiler and OpenMP. The use of templated functions in terms of the number of nodes per element improved the performance of our C++ code.

THE ACOUSTIC WAVE EQUATION

Before going to the higher-order elements for elastics, we briefly review the acoustic case, which can serve as a point of reference for the elastic problem. We consider the same test problem as before for degrees $M = 1$, 2 and 3. Table 5.3 lists the ratios of the compute time and of the required storage with and without global assembly, using 24 threads. The same tetrahedral mesh, derived from cubes with an edge length of 20 m, was used for

Table 5.3: Ratio of compute time and memory with and without global assembly for the *acoustic* case on 24 cores.

M	time	storage
1	0.38	1.2
2	1.0	11
3	1.9	26

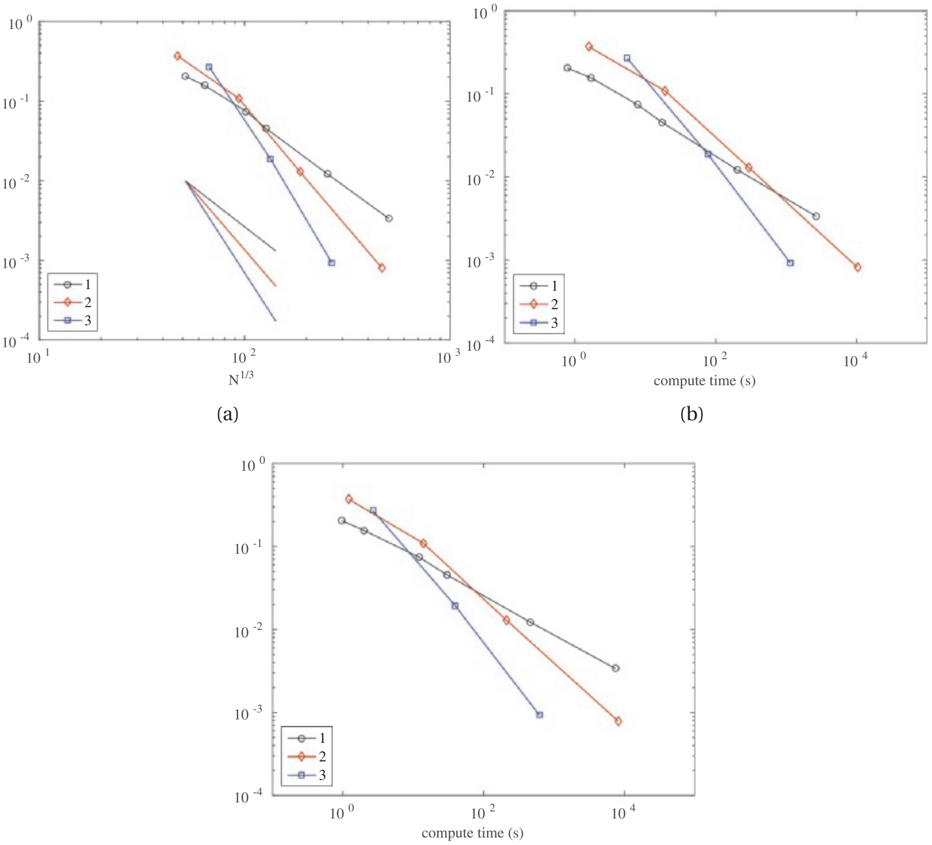
each degree. For the linear element of degree 1, assembly of the global stiffness matrix reduces the required time significantly with only a 20% increase of storage. For degree 2, there is no performance gain and the required storage is much larger. For degree 3, the scheme runs slower than with local assembly on the fly and requires a lot more memory. For that reason, (Zhebel et al., 2011, 2014) only mentioned local assembly.

HIGHER ORDERS

We now turn to the elastic case with discretizations of degree 2 and 3, using the same homogeneous problem on meshes of different size.

Figure 5.2 a plots the maximum observed error in the receiver data for the vertical displacement component, scaled by the maximum amplitude over all traces, as a function of the number of scalar degrees of freedom or number of nodes. Figure 5.2 b depicts the same errors as a function of the required compute time with 24 cores. The actual number of degrees of freedom is $3N$ and equals the size of the numerical displacement vector $\mathbf{u}$. The element size scales with $N^{-1/3}$. The error is expected to behave as $N^{-(M+1)/3}$ for degree M. The numerical experiments more or less follow the expected trend. The compute time only includes the wall-clock time for sparse matrix assembly and time stepping, not the time spent on reading and checking the mesh, setting up the nodes, the local-to-global map, and locating source and receivers on the mesh. Because the scheme for degree $M = 1$ was treated in a different way, it performs quite well even with a large number of elements. If errors around 10% are acceptable, it can be a viable alternative for the scheme of degree 3.

Figure 5.2 c is similar to 5.2 b, but with the product of the element stiffness matrix and element displacements evaluated on the fly during each time step. To better illustrate the differences in performance and memory usage, figure 5.3 plots the ratio in observed compute time as well as required storage between global assembly and local



(a)

(b)

(c)

Figure 5.2: Maximum error in the vertical displacement, scaled by the maximum amplitude, for elements of degree 1, 2 or 3, as a function of (a) $N^{1/3}$, where N is the number of degrees of scalar degrees of freedom, (b) compute time with global assembly, and (c) with local assembly. The extra set of 3 short lines in (a) depicts the theoretical asymptotic error behaviour.

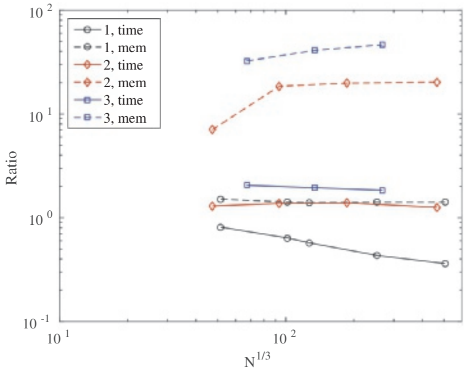


Figure 5.3: Ratios between compute time (drawn lines) and required storage (dashed lines) for global assembly and local assembly on the fly with elements of degree 1, 2 or 3. Global assembly is faster for degree 1 at 40% more storage. For degrees 2 and 3, local assembly is faster and requires substantially less storage.

assembly for elements of degree 1, 2 and 3. For degree 1, repeated from figure 5.1, the differences are not that large. Global assembly takes about 40% more storage but runs more quickly. For degree 2, local assembly is faster by a factor of about 1.3 on 24 cores. For degree 3, it is about 1.9 times as fast. The savings in storage compared to global assembly are substantial. Therefore, global assembly may only be attractive for degree 1.

A MORE REALISTIC EXAMPLE

We ran the code on the non-trivial model shown in figure 10 of (Zhebel et al., 2014), which is slightly more realistic than a homogeneous problem. The material properties are constant per layer and listed in Table 5.4. Figure 5.4 a displays a vertical cross section of the P-wave velocity. The source, indicated by a red star, is a vertical force at the surface, and has the signature of a Ricker wavelet with an 8-Hz peak frequency. The vertical displacement after 1 second in figure 5.4 b shows strong Rayleigh waves. The tetrahedral mesh has 1,528,595 vertices and 8,826,636 elements of degree 3. The time step was about 75% of the maximum value dictated by the CFL condition. Figure 5.4 c shows the vertical displacement, measured at the surface along a line corresponding to the earlier vertical cross section. The computation ran up till a time of 2 s.

Figure 5.5 plots the observed ratios between compute time and memory requirements with global and with local assembly on different meshes using 24 cores. The behaviour is similar to that of figure 5.3. Again, global assembly is only faster for the linear

Table 5.4: Isotropic elastic properties: P- and S-wave velocities and densities are constant per layer.

v_p (km/s)	v_s (km/s)	ρ (g/cm^3)
2.000	1.200	2.046
5.000	3.000	2.602
3.000	1.800	2.290
4.400	2.640	2.250
6.000	3.600	2.723
5.500	3.300	2.665

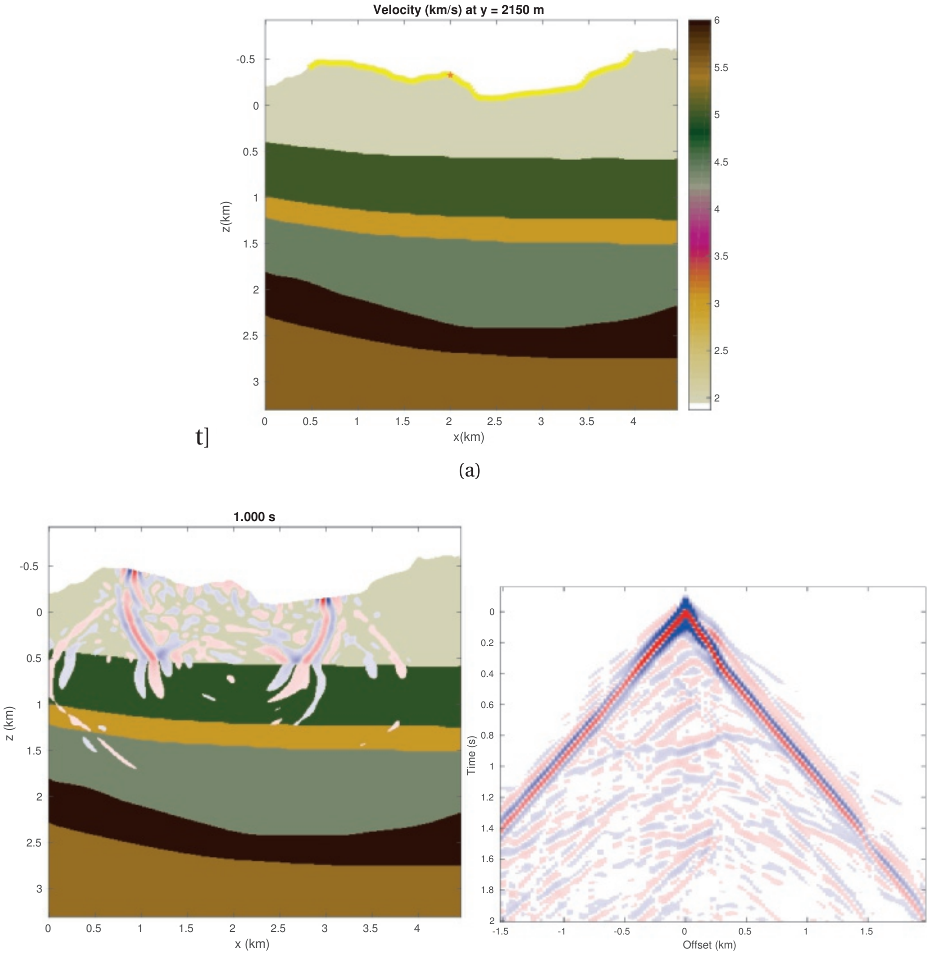
elements, whereas local assembly on the fly wins for degree 2 and 3. For the latter, the performance gain now is about 2.5 and 3 times, respectively.

CONCLUSIONS

We have compared the performance of mass-lumped tetrahedral finite elements on isotropic elastic wave propagation without and with global assembly of the stiffness matrix. To preserve their accuracy after mass lumping, the higher-order elements are augmented with higher-degree polynomials in the interior of the faces and the tetrahedron. For the lowest degree, the linear elements, this is not necessary. For that case, we simplified the expression for the stiffness matrix.

We ran performance tests on a homogeneous problem. The parallelization of the most compute intensive loops was performed by OpenMP directives. With global assembly, this involved symmetric sparse matrix assembly and the matrix-vector product during the time stepping. With assembly on the fly, the local assembly and local matrix-vector multiplication per element were parallelized in a single OpenMP loop. Further code optimizations were left to the compiler.

In the acoustic case, local assembly is more efficient than global assembly, except for the lowest-order case with linear elements. In the elastic case, the same appears to be true. For degree 1, the code with global assembly ran faster and used about 40% more storage than with local assembly. For degree 2, the numerical experiments with local assembly on the fly on 24 cores were about 1.4 times faster than with global assembly in one experiment and about 2 times in another. For degree 3, the gain was a factor 1.9 in one and 3 in the other. At the same time, the memory requirements were smaller by at least on order of magnitude for degree 2 and 3.



(a)

(b)

(c)

Figure 5.4: (a) P-wave velocities in km/s. The red star denotes the source positions and the yellow inverted triangles the receivers. (b) Vertical-displacement wavefield after 1 second. (c) Seismogram with vertical displacement.

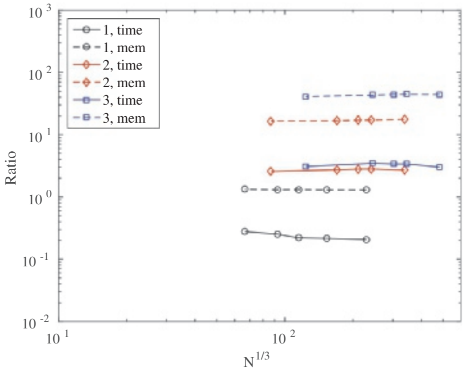


Figure 5.5: Ratios between compute time (drawn lines) and required storage (dashed lines) for global assembly and local assembly on the fly with elements of degree 1, 2 or 3. As in figure 5.3, global assembly is faster for degree 1 at 30 to 40% more storage, whereas for degrees 2 and 3, local assembly is faster and requires substantially less storage.

We observed in a simple test problem that, for high accuracy, augmented cubic elements performed best in terms of compute time for a given accuracy. For low accuracy, the linear element may still be attractive. In that case, its efficiency compensates the need for a much finer mesh.

ACKNOWLEDGEMENTS

This work was partly funded by the Industrial Partnership Programme (IPP) 'Computational sciences for energy research' of the Foundation for Fundamental Research on Matter (FOM), which is part of the Netherlands Organisation for Scientific Research (NWO). This research programme is co-financed by Shell Global Solutions International B.V.

REFERENCES

Alberty, J., Carstensen, C., Funken, S. A., Klose, R., 2002. Matlab implementation of the finite element method in elasticity. Computing 69 (3), 239–263.

Bao, H., Bielak, J., Ghattas, O., Kallivokas, L. F., O'Hallaron, D. R., Shewchuk, J. R., Xu, J., 1998. Large-scale simulation of elastic wave propagation in heterogeneous media on parallel computers. Computer Methods in Applied Mechanics and Engineering 152 (1–2), 85–102, containing papers presented at the Symposium on Advances in Computational Mechanics.

Bécache, E., Joly, P., Tsogka, C., 2002. A new family of mixed finite elements for the linear elastodynamic problem. SIAM Journal on Numerical Analysis 39 (6), 2109–2132.

Brossier, R., Virieux, J., Operto, S., 2008. Parsimonious finite-volume frequency-domain method for 2-D P-SV-wave modelling. Geophysical Journal International 175 (2), 541–559.

Chin-Joe-Kong, M. J. S., Mulder, W. A., van Veldhuizen, M., 1999. Higher-order triangular and tetrahedral finite elements with mass lumping for solving the wave equation. Journal of Engineering Mathematics 35, 405–426.

Cockburn, B., Gopalakrishnan, J., Lazarov, R., 2009. Unified hybridization of discontinuous Galerkin, mixed, and continuous Galerkin methods for second order elliptic problems. SIAM Journal on Numerical Analysis 47 (2), 1319–1365.

Cohen, G., Fauqueux, S., 2005. Mixed spectral finite elements for the linear elasticity system in unbounded domains. SIAM Journal on Scientific Computing 26 (3), 864–884.

Cohen, G., Joly, P., Roberts, J. E., Tordjman, N., 2001. Higher order triangular finite elements with mass lumping for the wave equation. SIAM Journal on Numerical Analysis 38 (6), 2047–2078.

Cohen, G., Joly, P., Tordjman, N., 1995. Higher order triangular finite elements with mass lumping for the wave equation. In: Cohen, G., Bécache, E., Joly, P., Roberts, J. E. (Eds.), Proceedings of the Third International Conference on Mathematical and Numerical Aspects of Wave Propagation. SIAM, Philadelphia, pp. 270–279.

Dumbser, M., Käser, M., 2006. An arbitrary high-order Discontinuous Galerkin method for elastic waves on unstructured meshes – II. The three-dimensional isotropic case. Geophysical Journal International 167 (1), 319–336.

Dumbser, M., Käser, M., de la Puente, J., 2007. Arbitrary high-order finite volume schemes for seismic wave propagation on unstructured meshes in 2D and 3D. Geophysical Journal International 171 (2), 665–694.

Etienne, V., Chaljub, E., Virieux, J., Glinsky, N., 2010. An hp-adaptive discontinuous Galerkin finite-element method for 3-d elastic wave modelling. Geophysical Journal International 183 (2), 941–962.

Fried, I., Malkus, D. S., 1975. Finite element mass matrix lumping by numerical integration with no convergence rate loss. International Journal of Solids and Structures 11, 461–466.

Giorgiani, G., Fernández-Méndez, S., Huerta, A., 2013. Hybridizable discontinuous Galerkin p-adaptivity for wave propagation problems. International Journal for Numerical Methods in Fluids 72 (12), 1244–1262.

Käser, M., Dumbser, M., 2006. An arbitrary high-order Discontinuous Galerkin method for elastic waves on unstructured meshes – I. The two-dimensional isotropic case with external source terms. Geophysical Journal International 166 (2), 855–877.

Komatitsch, D., Tromp, J., 1999. Introduction to the spectral-element method for 3-D seismic wave propagation. Geophysical Journal International 139 (3), 806–822.

Kononov, A., Minisini, S., Zhebel, E., Mulder, W. A., June 2012. A 3D tetrahedral mesh generator for seismic problems. In: Proceedings of the 74th EAGE Conference & Exhibition. p. B006.

Lesage, A. C., Aubry, R., Houzeaux, G., Polo, M. A., Cela, J., June 2010. 3D spectral element method combined with H-refinement. 72nd EAGE Conference & Exhibition, Barcelona, Spain, Extended Abstracts, C047.

Maday, Y., Ronquist, E. M., 1990. Optimal error analysis of spectral methods with emphasis on non-constant coefficients and deformed geometries. Computer Methods in Applied Mechanics and Engineering 80 (1–3), 91–115.

Mercerat, E. D., Vilotte, J. P., Sánchez-Sesma, F. J., 2006. Triangular Spectral Element simulation of two-dimensional elastic wave propagation using unstructured triangular grids. Geophysical Journal International 166 (2), 679–698.

Moczo, P., Kristek, J., Galis, M., Chaljub, E., Etienne, V., 2011. 3-D finite-difference, finite-element, discontinuous-Galerkin and spectral-element schemes analysed for their accuracy with respect to P-wave to S-wave speed ratio. Geophysical Journal International 187 (3), 1645–1667.

Mulder, W. A., 1996. A comparison between higher-order finite elements and finite differences for solving the wave equation. In: Désidéri, J.-A., LeTallec, P., Oñate, E., Périaux, J., Stein, E. (Eds.), Proceedings of the Second ECCOMAS Conference on Numerical Methods in Engineering. John Wiley & Sons, Chichester, pp. 344–350.

Mulder, W. A., 2013. New triangular mass-lumped finite elements of degree six for wave propagation. Progress In Electromagnetics Research 141, 671–692.

Mulder, W. A., Shamasundar, R., 2016. Performance of continuous mass-lumped tetrahedral elements for elastic wave propagation with and without global assembly. Geophysical Journal International 207 (1), 414–421.

Mulder, W. A., Zhebel, E., Minisini, S., 2014. Time-stepping stability of continuous and discontinuous finite-element methods for 3-D wave propagation. Geophysical Journal International 196 (2), 1123–1133.

Orszag, S. A., 1980. Spectral methods for problems in complex geometries. Journal of Computational Physics 37 (1), 70–92.

Patera, A. T., 1984. A spectral element method for fluid dynamics: laminar flow in a channel expansion. Journal of Computational Physics 54 (3), 468–488.

Riviere, B., Wheeler, M. F., 2003. Discontinuous finite element methods for acoustic and elastic wave problems. Vol. 329. Providence, RI: American Mathematical Society, pp. 4–6.

Seriani, G., Priolo, E., Carcione, J., Padovani, E., 1992. High-order spectral element method for elastic wave modeling. SEG Technical Program Expanded Abstracts 11, 1285–1288.

Sherwin, S. J., Karniadakis, G. E., 1995. A new triangular and tetrahedral basis for high-order (hp) finite element methods. International Journal for Numerical Methods in Engineering 38 (22), 3775–3802.

Tordjman, N., 1995. Élements finis d'order élevé avec condensation de masse pour l'equation des ondes. Ph.D. thesis, L'Université Paris IX Dauphine.

Wang, X., Symes, W. W., Warburton, T., 2010. Comparison of discontinuous Galerkin and finite difference methods for time domain acoustics. SEG Technical Program Expanded Abstracts 29 (1), 3060–3065.

Wilcox, L. C., Stadler, G., Burstedde, C., Ghattas, O., 2010. A high-order discontinuous Galerkin method for wave propagation through coupled elastic-acoustic media. Journal of Computational Physics 229 (24), 9373–9396.

Zhebel, E., Minisini, S., Kononov, A., Mulder, W. A., June 2011. Solving the 3D acoustic wave equation with higher-order mass-lumped tetrahedral finite elements. In: Proceedings of the 73rd Conference & Exhibition. p. A010.

Zhebel, E., Minisini, S., Kononov, A., Mulder, W. A., 2014. A comparison of continuous mass-lumped finite elements with finite differences for 3-D wave propagation. Geophysical Prospecting 62 (5), 1111–1125.

Zienkiewicz, O. C., Taylor, R. L., 2000. The Finite Element Method. Volume 1: The Basis. Butterworth-Heinemann, Oxford, 5th edition.

6

Performance of Hermite polynomials in finite element schemes for forward modelling

Previous chapters have used Legendre polynomials for interpolation in the finite element scheme. The performance of LGL schemes was promising for first-order problems in combination with defect correction in 1D, but fell short for 2D problems. In this chapter we look into Hermite polynomials as an alternative, since they are C1 continuous, they offer better representation of derivatives of pressure on triangles. They display fourth order accuracy in 2 dimensions, for homogenous models, but need to be modified for variable densities.

INTRODUCTION

The first-order form with pressure and velocities on the element's vertices has some resemblance to elements based on cubic Hermite interpolating polynomials (Ciarlet and Raviart, 1972). Felippa (Wall et al., 2001) presents a 1-D application of these polynomials to bending elements and includes dispersion curves that include a physical and a spurious mode (Park and Flaggs, 1984). The latter is also called 'optical' (Cottrell et al., 2006).

Here, we will present Fourier analysis for the homogeneous acoustic wave equation. We examine the errors in the eigenvalues, representing the dispersion curves, and the error in the eigenvectors, which show how much energy ends in the physical and how much in the spurious modes as a function of wavenumber.

In 1D, the cubic polynomials per element are represented by the pressure and its derivatives on the vertices or nodes. In 2D on the triangle, the element is defined by cubic polynomials with pressure and its two derivatives on the vertices. To obtain the ten degrees of freedom required to represent a cubic polynomial, a bubble function for the pressure is added to the interior of the triangle and represented by a pressure value at its centroid. In 3D on tetrahedra, the element is defined by the pressure and its three derivatives on the vertices and bubble functions for the pressure on each of the four faces, providing the 20 degrees of freedom that determine a 3-D cubic polynomial.

Unfortunately, the continuity of these elements across element boundaries makes them unsuited for problems with discontinuities across element interfaces. The reason is that the velocity, defined as the gradient of the pressure divided by the local density, has a normal component that is continuous across element boundaries, but this not true for the tangential component(s) if the density has a jump across the element boundary. This means these elements are only suited for smoothly varying media as are often used in the initial stages of seismic full-waveform inversion.

Here, we will make the additional assumption that the material parameters, density and sound speed, are homogeneous. This simplifies the finite-element method but limits its applicability. In the next section, we will consider the 1-D case. Then, a 2-D example will be presented. The last section summarizes the conclusions.

1D

FINITE-ELEMENT DISCRETIZATION FOR THE HOMOGENEOUS CASE

The acoustic wave equation in 1D reads

$$\frac{1}{\rho c^2}\frac{\partial^2 p}{\partial t^2} = \frac{\partial a}{\partial x}, \quad \rho a = \frac{\partial p}{\partial x}.$$

The pressure $p(t,x)$ and particle acceleration $a(t,x)$ are continuous. The sound speed $c(x)$ and density $\rho(x)$ depend on position. For the dispersion analysis and numerical tests, we will only consider the homogeneous problem with constant coefficients, which greatly simplifies the construction of the finite elements.

For the finite-element discretization, we choose $N+1$ vertices x_k, $k=0,\dots,N$, that define elements of size $h_\ell = x_\ell - x_{\ell-1}$, $\ell = 1,\dots,N$. The basis functions $\phi_i(\xi)$ and $\phi_i'(\xi)$, with normalized coordinate $\xi \in [0,1]$ inside the element, should obey

$$\phi_i(j) = \delta_{ij}, \quad \frac{\mathrm{d}\phi_i}{\mathrm{d}\xi}(j) = 0, \quad \text{for } i,j = 0,1,$$

and

$$\phi_i'(j) = 0, \quad h^{-1}\frac{\mathrm{d}\phi_i'}{\mathrm{d}\xi}(j) = \delta_{ij}, \quad \text{for } i,j = 0,1.$$

In the cubic case, this leads to

$$\phi_0(\xi) = (1-\xi)^2(1+2\xi), \quad \phi_0'(\xi) = h(1-\xi)^2\xi, \quad \phi_1(\xi) = \xi^2(3-2\xi), \quad \phi_1'(\xi) = -h(1-\xi)\xi^2.$$

Here, h is the length of the element. Note that $\phi_1(\xi) = \phi_0(1-\xi)$ and $\phi_1'(\xi) = -\phi_0'(1-\xi)$.

The discretization is straight-forward if the material properties are homogeneous.

The contribution to the mass matrix per element is

$$A = hQ\bar{A}Q, \quad \bar{A} = \frac{1}{420}\begin{pmatrix} 156 & 22 & 54 & 13 \\ 22 & 4 & 13 & 3 \\ 54 & 13 & 156 & 22 \\ 13 & 3 & 22 & 4 \end{pmatrix},$$

where

$$Q = \mathrm{diag}\{1, h, 1, -h\}.$$

Here, the degrees of freedom are paired as p and $\partial_x p$ on the left and on the right side of the element. Note that h may vary from element to element. The contribution to the stiffness matrix is

$$B = \frac{1}{h} Q\bar{B}Q, \quad \bar{B} = \frac{1}{30}\begin{pmatrix} 36 & 3 & -36 & -3 \\ 3 & 4 & -3 & 1 \\ -36 & -3 & 36 & 3 \\ -3 & 1 & 3 & 4 \end{pmatrix}.$$

Note that the resulting A differs from the one in equation (8) of (Wall et al., 2001) with $\mu_1 = \mu_2 = \mu_3 = 1$, but agrees if the last matrix in that equation is replaced by

$$\frac{\mu_3}{2800}\begin{pmatrix} 4 & 2 & -4 & 2 \\ 2 & 1 & -2 & 1 \\ -4 & -2 & 4 & -2 \\ 2 & 1 & -2 & 1 \end{pmatrix}.$$

The entries at positions 1,3 and 3,1 should have a minus sign. Our matrix B agrees if $\gamma_1 = \gamma_2 = 1$ in equation (13) of (Wall et al., 2001).

Dispersion analysis

For the dispersion analysis, the mesh is assumed to be equidistant and periodic. Then, the mass matrix $\mathcal{M}$ and stiffness matrix $\mathcal{K}$ become

$$\mathcal{M} = \frac{h}{420}\begin{pmatrix} 6[52+9(T+T^{-1})] & -13h(T-T^{-1}) \\ 13h(T-T^{-1}) & h^2[2+3(2-T-T^{-1})] \end{pmatrix},$$

$$\mathcal{K} = \begin{pmatrix} \frac{6}{5h}(2-T-T^{-1}) & \frac{1}{10}(T-T^{-1}) \\ -\frac{1}{10}(T-T^{-1}) & \frac{h}{30}[6+2-T-T^{-1}) \end{pmatrix}.$$

Here T is a shift operator defined by $T^n p_m = p_{m+n}$, where p_m approximates the pressure at x_m. We also have $T^m p'_n = p'_{m+n}$ for the derivative p'_n that approximates $\frac{\partial p}{\partial x}(x_m)$. The matrices operate on vectors $(p_m \;\; p'_m)^\mathsf{T}$.

The Fourier symbol of T is $\hat{T} = e^{ikh}$ for wavenumber k. The dispersion curve follows from the eigenvalues of $\hat{L} = \hat{\mathscr{M}}^{-1}\hat{\mathscr{K}}$, which is now a 2×2 system. Here, $\hat{\mathscr{M}}$ represents the Fourier symbol of the mass matrix and $\hat{\mathscr{K}}$ that of the stiffness matrix. The eigenvalues are

$$\kappa_\pm^2 = \frac{6}{h^2}\,\frac{141-4\zeta(8+\zeta)\pm\sqrt{13056+\zeta(3856+\zeta(-7524+(1656-19\zeta)\zeta))}}{65+\zeta(\zeta-36)},$$

where $\zeta = \cos(kh)$. For small k,

$$(\kappa_-/k)^2 \simeq 1 + \frac{(kh)^6}{30240},$$

demonstrating sixth-order behaviour of the dispersion error, relative to the exact wavenumber k.

Figure 6.1(a) plots the eigenvalues $\kappa_\pm$ as a function of the normalized wavenumber $\eta = kh/(2\pi)$. Note that Nyquist-Shannon sampling theorem requires $|kh| \le \pi$ in the scalar case. Here, with both p and $\frac{\partial p}{\partial x}$, we have $|kh| \le 2\pi$. The results for negative wavenumbers follow by symmetry and are not plotted.

Had we only shown the results for $|kh| \le \pi$, then one eigenvalue, κ_- would be physical and the other spurious or 'optical' (Cottrell et al., 2006; Wall et al., 2001). By enlarging the domain to $|kh| \le 2\pi$, we can unwrap the two eigenvalues: $\kappa_-/(2\pi\eta)$ for $\eta \in [0, \frac{1}{2}]$ and $\kappa^+/(2\pi\eta)$ for $\eta \in [\frac{1}{2}, 1]$. The other ones, $\kappa^+/(2\pi\eta)$ for $\eta \in [0, \frac{1}{2}]$ and $\kappa_-/(2\pi\eta)$ for $\eta \in [\frac{1}{2}, 1]$ then remain as spurious modes. Because the eigenvalues depend on ζ, there is The symmetry $\kappa_\pm^2(1-\eta) = \kappa_\pm^2(\eta)$ follows from the dependence of the eigenvalues on $\zeta = \cos(2\pi\eta)$.

Figure 6.1(b) shows the physical eigenvalues after scaling by the exact eigenvalue. The two values near the discontinuity at $\eta = 1/2$ are $\kappa h = \sqrt{168/17}$ and $\sqrt{10}$, both close to π. If all the energy could be restricted to these modes, there would be no spurious modes. For instance, if k is small, all wave energy should be confined to the eigenvector of κ_-. In practice, some energy may end up in the eigenvector of κ_+ for small k. Next, we will study this in more detail by considering the error in the eigenvectors.

To determine the error in the eigenvectors, we follow (Mulder, 1999) and express the Fourier symbol of the spatial operator as $\hat{L} = Q\Lambda Q^{-1}$, where the columns of Q are the eigenvectors of $\hat{L}$ and the diagonal matrix Λ contains the eigenvalues $\kappa_\pm^2$ on its diagonal.

The exact eigenvector corresponding to the mode e^{ikx} in the Fourier domain is $\hat{\mathbf{e}}_0 =$

$(1 \;\; \mathrm{i}k)^\mathsf{T}$. The second spatial derivative turns this into $k^2\hat{\mathbf{e}}_0$, whereas the numerical approximation produces $\hat{L}\hat{\mathbf{e}}_0$. The error in the eigenvector is then something like $k^{-2}\hat{L}\hat{\mathbf{e}}_0 - \hat{\mathbf{e}}_0$. To separate the dispersion error from the error in the eigenvectors, we can replace the numerical eigenvalues κ^2 in Λ by the exact k^2, evaluate the effect of the modified operator $\hat{L}$ on the exact eigenvector $\hat{\mathbf{e}}_0$, divide by k^2 afterwards, and compare the result to the same exact eigenvector. We can also do that for each of the eigenvectors separately by setting the eigenvalues to zero except for the one of interest. Assuming that the first eigenvector corresponds to κ_- and the second to κ_+, we can focus on κ_- for small k. We define vectors

$$\hat{\mathbf{s}}_1 = k^{-2} Q \,\mathrm{diag}\{k^2, 0\} Q^{-1} \hat{\mathbf{e}}_0$$

and

$$\hat{\mathbf{s}}_2 = k^{-2} Q \,\mathrm{diag}\{0, k^2\} Q^{-1} \hat{\mathbf{e}}_0.$$

These describe the following steps: project the exact eigenvector on the numerical ones, propagate with the exact wavenumber, project back, rescale by the squared the wavenumber, and compare to the input. The matrix $\hat{S} = (\hat{\mathbf{s}}_1 \;\; \hat{\mathbf{s}}_2)^\mathsf{T}$ has these vectors as its first and second column. Then,

$$\hat{S} \simeq \begin{pmatrix} 1 - \frac{2}{4725}(kh)^6 & \frac{2}{4725}(kh)^6 \\ \mathrm{i}k\left[1 + \frac{2}{315}(kh)^4\right] & \mathrm{i}k\left[-\frac{2}{315}(kh)^4\right] \end{pmatrix}.$$

The first column approximates the exact eigenvector $\hat{\mathbf{e}}_0$, the second column describes how much of it ends up in the other mode and should be classified as spurious energy. This column has the opposite sign of the error, $\mathbf{s}_1 - \hat{\mathbf{e}}_0$, in the first column, that is, $\mathbf{s}_1 + \mathbf{s}_2 = \hat{\mathbf{e}}_0$. The matrix shows that the first row, corresponding to p, has a sixth-order error and the second row, corresponding to the derivative of p, has a fourth-order error. The last determines the overall error behaviour of the scheme.

To study the eigenvector error over the whole domain, we first rescale the eigenvector to obtain relative errors, by dividing out the factor $\mathrm{i}k$. Let $D = \mathrm{diag}\{1, (\mathrm{i}k)^{-1}\}$. The normalized exact eigenvector becomes $\hat{\mathbf{e}}_1 = D\hat{\mathbf{e}}_0 = (1 \;\; 1)^\mathsf{T}$ and the numerical ones the columns of $\tilde{Q} = DQ$:

$$\tilde{Q} = \begin{pmatrix} \frac{a+w}{d} & \frac{a-w}{d} \\ \\ \frac{\sin\xi}{\xi} & \frac{\sin\xi}{\xi} \end{pmatrix},$$

with $\xi = kh = 2\pi\eta$, $\zeta = \cos(\xi)$, $d = 6(52-17\zeta)$, $a = 80+\zeta(52-27\zeta)$ and

$$w = \sqrt{13056+\zeta(3856+\zeta(-7524+\zeta(1656-19\zeta)))}.$$

We then consider the vectors

$$\hat{\mathbf{r}}_1 = \tilde{Q}\,\mathrm{diag}\{1,0\}\tilde{Q}^{-1}\hat{\mathbf{e}}_1,$$

$$\hat{\mathbf{r}}_2 = \tilde{Q}\,\mathrm{diag}\{0,1\}\tilde{Q}^{-1}\hat{\mathbf{e}}_1.$$

In the present example, we happen to have $\hat{\mathbf{r}}_1 + \hat{\mathbf{r}}_2 = 1$. The vectors $\mathbf{r}_1$ and $\mathbf{r}_2$ contain 4 components that describe the eigenvector error. The drawn line in figure 6.2 consists in $\hat{r}_{1,1} - 1$ for $\eta < \frac{1}{2}$ and $\hat{r}_{2,2} - 1$ for $\eta > \frac{1}{2}$, with 0 at $\eta \to \frac{1}{2}$. The dashed line follows $\hat{r}_{1,2} - 1$ for $\eta < \frac{1}{2}$ and $\hat{r}_{2,1} - 1$ for $\eta > \frac{1}{2}$, with -1 at $\eta \to \frac{1}{2}$. These represent the relative difference between the approximate and exact eigenvectors. The missing components represent the spurious modes and just have the opposite sign, because $\hat{\mathbf{r}}_1 + \hat{\mathbf{r}}_2 = 1$, and are therefore not shown.

Figure 6.2 seems to suggests that we should stay at some distance below the Nyquist limit of $\eta = \frac{1}{2}$, since one if the branches shoots off to -1 around $\eta = \frac{1}{2}$. This may be too pessimistic as around $\eta = \frac{1}{2}$, the two eigenvalues $\kappa_{\pm}^2$ are nearly equal. However, the amplitude of the dashed curve rapidly increases for η above $\frac{1}{2}$, so having $|\eta| \le \eta_{\max}$ with $\eta_{\max}$ just below $\frac{1}{2}$ is advisable.

2D

We have tested the method on a 2-D standing-wave problem in a homogeneous constant-density acoustic model. The partial differential equation is

$$\frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} = \frac{\partial a_1}{\partial x_1} + \frac{\partial a_2}{\partial x_2}, \quad a_1 = \frac{\partial p}{\partial x_1}, \quad a_2 = \frac{\partial p}{\partial x_2}.$$

The finite-element discretization starts with the reference triangle with barycentric coordinates $\xi_0 = 1 - \xi_1 - \xi_2$, ξ_1 and ξ_2. The true coordinates inside a triangle with vertices $(x_{1,k}, x_{2,k})$, $k = 0,1,2$, are $x_j = \sum_{k=0}^{2} \xi_k x_{j,k}$ for $j = 1,2$. Let $\alpha_j = x_{j,1} - x_{j,0}$ and $\beta_j = x_{j,2} - x_{j,0}$ for $j = 1,2$. The 10 degrees of freedom are the pressure values p_k and their derivatives $v_{1,k}$ and $v_{2,k}$ in x_1 and x_2, respectively, on the three vertices indexed by $k = 0,1,2$, as well as the pressure p_c at the centroid.

We order them as $\{p_0, a_{1,0}, a_{2,0}, p_1, a_{1,1}, a_{2,1}, p_2, a_{1,2}, a_{2,2}, p_c\}$ per element. The corresponding basis functions are

$$\begin{aligned}
\phi_1 &= \xi_0[(3-2\xi_0)\xi_0 - 7\xi_1\xi_2],\\
\phi_2 &= \xi_0[\alpha_1(\xi_0-\xi_2)\xi_1 + \beta_1(\xi_0-\xi_1)\xi_2],\\
\phi_3 &= \xi_0[\alpha_2(\xi_0-\xi_2)\xi_1 + \beta_2(\xi_0-\xi_1)\xi_2],\\
\phi_4 &= \xi_1[(3-2\xi_1)\xi_1 - 7\xi_0\xi_2],\\
\phi_5 &= \xi_1[\alpha_1(2\xi_2\xi_0 - \xi_1(1-\xi_1)) - \beta_1(\xi_0-\xi_1)\xi_2],\\
\phi_6 &= \xi_1[\alpha_2(2\xi_2\xi_0 - \xi_1(1-\xi_1)) - \beta_2(\xi_0-\xi_1)\xi_2],\\
\phi_7 &= \xi_2[(3-2\xi_2)\xi_2 - 7\xi_0\xi_1],\\
\phi_8 &= \xi_2[\beta_1(2\xi_1\xi_0 - \xi_2(1-\xi_2)) - \alpha_1(\xi_0-\xi_2)\xi_1],\\
\phi_9 &= \xi_2[\beta_2(2\xi_1\xi_0 - \xi_2(1-\xi_2)) - \alpha_2(\xi_0-\xi_2)\xi_1],\\
\phi_{10} &= 27\xi_0\xi_1\xi_2.
\end{aligned}$$

The last is the bubble function. The mass and stiffness matrix per element follow from exact integration over the triangle and serve as input for the global assembly. Note that α_j, β_j and $\beta_j - \alpha_j$ in the basis functions are related to projections on the edges of the acceleration vectors defined by the pressure gradient.

The time stepping scheme is

$$\mathbf{q}^{n+1} = 2\mathbf{q}^n - \mathbf{q}^{n-1} - (\Delta t)^2 \mathcal{M}^{-1}\mathcal{K}\mathbf{q}^n,$$

where the superscript denotes time $t^n = t^0 + n\Delta t$. The vector $\mathbf{q}$ denotes the degrees of freedom. The time step Δt should be chosen such that $0 \le (\Delta t)^2 L \le 4$, with $L = \mathcal{M}^{-1}\mathcal{K}$.

The domain for the test problem has a size $[0,2] \times [0,1]$ in dimensionless units. We choose a unit sound speed c and density ρ, also in dimensionless units. The exact solution is a standing wave of the form $p = \sin(\alpha_1 x)\sin(\alpha_2 z)\cos(\omega t)$, with $\omega = c(\alpha_1^2 + \alpha_2^2)^{1/2}$ and $\alpha_k = 2\pi m_k$ for $k = 1, 2$. The solution obeys zero Dirichlet boundary conditions The initial-value problem is started at time zero and runs until $t_{\text{max}} = 2\pi/\omega$. The mesh was generated by taking a uniform background mesh with squares cells, perturbing internal vertices randomly by at most 10% to make it more irregular and applying a Delaunay triangulation. Figure 6.3 shows a fairly coarse mesh and the initial pressure. For drawing purposes, the latter was interpolated from the given degrees of freedom on the mesh to a much finer Cartesian grid, using the cubic Hermite polynomial representation.

The root-mean-square (RMS) errors in the pressure p at the vertices, p_c at the element centroids, and in the horizontal and vertical velocities $a_1 = \frac{\partial p}{\partial x_1}$ and $a_2 = \frac{\partial p}{\partial x_2}$ at the vertices was measured at a time t_{max}. Figure 6.4 plots the results as a function of the square-root of the number of vertices N, which is proportional to the inverse of the average element size. Power-law fits provide an error of order 4 for p and p_c and order 3 for a_1 and a_2, as expected for a cubic-polynomial representation of the pressure. The second-order time-stepping scheme ran at about half the maximum allowable value, which appears to be small enough to prevent it from showing up in the graphs.

CONCLUSIONS

We have analyzed the dispersion properties of finite elements based on cubic Hermite polynomials applied to the constant-density acoustic wave equation with a constant velocity. The dispersion curve has a sixth-order error, whereas the eigenvector error that describes the cross talk with the spurious mode has an error of order six for the pressure and of order four for its gradient. The accuracy at higher wavenumbers is reasonable up to a value somewhat below the Nyquist limit for the scalar case. A numerical test in two space dimensions shows fourth-order accuracy for the pressure and one order lower for its gradient.

In 1D, the current approach can be easily generalised to inhomogeneous problems with piecewise constant sound speed and density per element by storing the pressure gradient divided by the density on the nodes and multiply it by the density in the element during matrix assembly.

In 2D, the bubble function generates a discontinuity in the normal component of the pressure gradient across the edges, except at the vertices, whereas the PDE requires its continuity, although in the homogeneous case, this does not pose a real problem. Unfortunately, if the solution in 2D is represented in terms of pressure and derivatives of pressure scaled by a density, the tangential component of the pressure gradient along the edge becomes discontinuous if the density is discontinous across that edge. This violates the continuity of the pressure across edges. It seems that the Hermite representation is not able to meet all these requirements.

REFERENCES

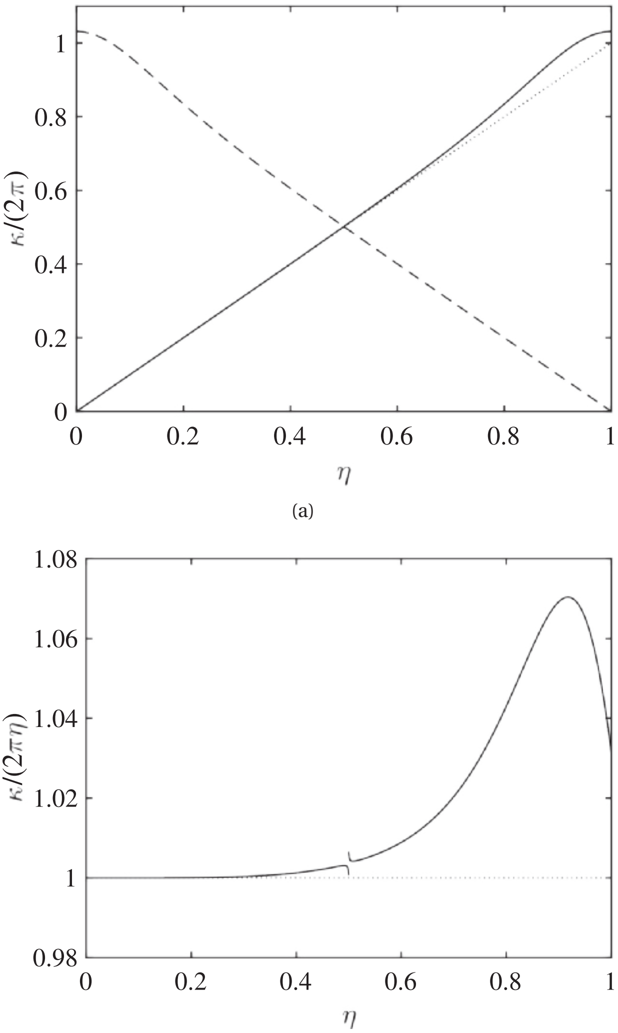
Ciarlet, P. G., Raviart, P., 1972. General lagrange and hermite interpolation in rn with applications to finite element methods. Archive for Rational Mechanics and Analysis 46 (3), 177–199.

Cottrell, J. A., Reali, A., Bazilevs, Y., Hughes, T. J. R., 2006. Isogeometric analysis of structural vibrations. Computer Methods in Applied Mechanics and Engineering 195 (41), 5257–5296, john H. Argyris Memorial Issue. Part II.

Mulder, W. A., 1999. Spurious modes in finite-element discretizations of the wave equation may not be all that bad. Applied Numerical Mathematics 30 (4), 425–445.

Park, K., Flaggs, D., 1984. A Fourier analysis of spurious mechanisms and locking in the finite element method. Computer Methods in Applied Mechanics and Engineering 46 (1), 65–81.

Wall, W. A., Bletzinger, K., Schweizerhof, K., Haase, G., Langer, U., Lindner, E., Máuhlhuber, W., 2001. Trends in computational structural mechanics. International Center for Numerical Methods in Engineering (CIMNE).



(a)

(b)

Figure 6.1: (a) The positive square-roots of the two eigenvalues, scaled by 2π, as a function of the normalized wavenumber η. The spurious modes are shown as dashed lines. (b) Unwrapped normalized dispersion curve for the 1-D element based on cubic Hermite polynomials, showing the numerical approximation κ of the wavenumber normalized by the exact one, $2\pi\eta$. The dotted line is the exact result, the drawn and dashed lines mark the two eigenvalue branches.

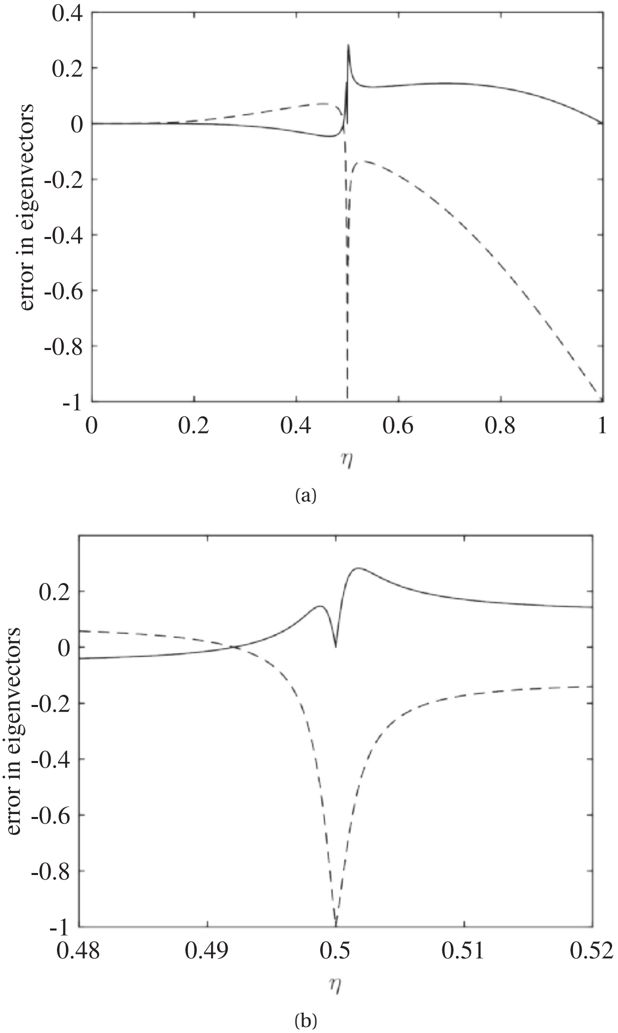


Figure 6.2: Error in the eigenvectors. Only two of the four components are shown, since the other two have just the opposite sign. (b) Detail of (a).

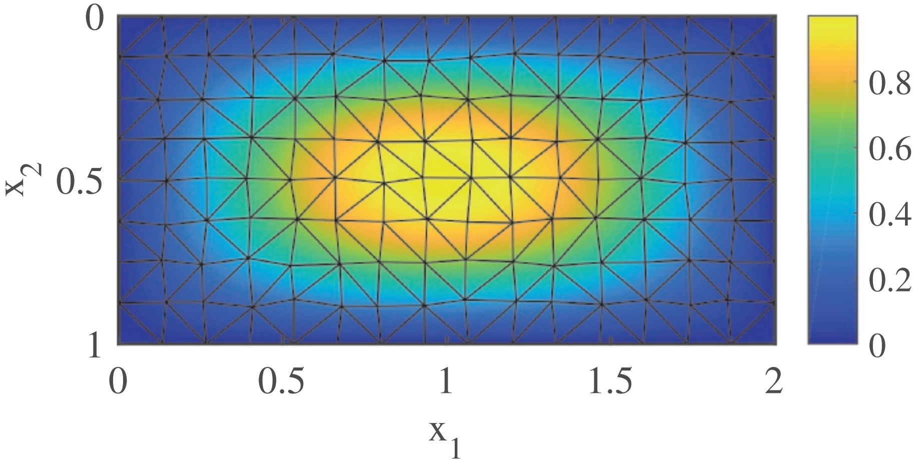


Figure 6.3: Mesh with the initial pressure as backdrop.

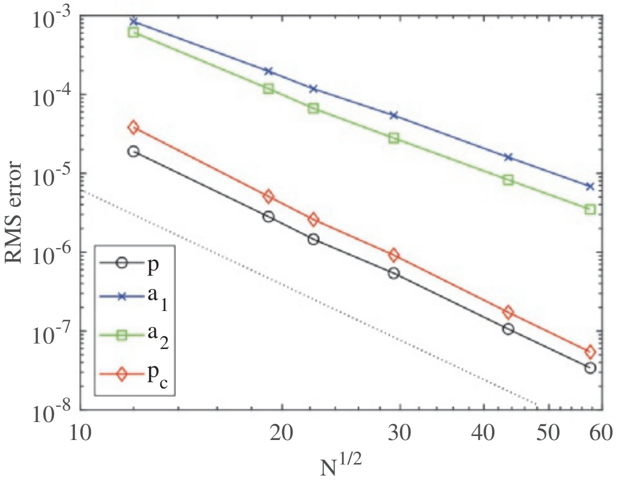


Figure 6.4: Convergence with cubic Hermite polynomials as basis functions for a 2-D test problem. The root-mean-square error as a function of the square-root of the number of vertices, N, shows fourth-order convergence, indicated by the dotted line, for the pressure p (black line) at the nodes and p_c (red line) at the centroids, whereas the horizontal velocity v_1 (blue) and vertical velocity v_2 (green) have third-order convergence.

7

CONCLUSIONS

In chapter 1 of this thesis, we identify the need for sharper imaging techniques for geophysical exploration. In chapter 2, we examine existing methods that facilitate more accurate numerical modelling of wave propagation, which will contribute to more detailed images of the earth's subsurface. In chapter 3, we have compared four finite-element schemes with polynomial basis functions for the first-order formulation of the acoustic wave equation, using Legendre-Gauss-Lobatto nodes, Chebyshev-Gauss-Lobatto without and with weighting function or the standard element. The first-order formulation of the wave equations is commonly used for finite-difference modelling of the wave equation because of its lower memory requirements. For finite-element methods, it is more common to use the second-order formulation, since it allows mass lumping with no loss in accuracy. Mass lumping avoid the inversion the mass matrix and allows for explicit time stepping. For the first-order formulation of the wave equation, however, mass lumping tends to decrease the spatial accuracy. Of the four interpolation schemes chosen, numerical dispersion is least for Legendre-Gauss-Lobatto nodes. For polynomials of odd degree, they are more accurate than the second-order formulation of the wave equation but this gain is lost after mass lumping. We have shown that accuracy can be restored by defect correction, applying one iteration on the consistent mass matrix, preconditioned by its lumped version. For polynomials of degree one, this improves the accuracy from second to fourth order in the element size. In other cases, the improvement in accuracy is less dramatic. The error in the eigenvectors for the first-order formulation, however, is worse than obtained for the second-order formulation, without and with mass lumping. Because the eigenvector error is zero for the lowest-degree scheme, with linear polynomials, our iterative approach appears to be most attractive for just that case.

Fourier analysis in two space dimensions suggests that the fourth-order error behaviour should be obtained for the lowest-order scheme, either with bilinear elements on quadrilaterals or with linear elements on triangles, at least on very regular meshes and with constant material properties. In chapter 4, we test whether this holds in unstructured meshes. It turns out that the spatial operator in the discrete first-order form of the wave equation may have short-wavelength null-vectors. The corresponding waves are therefore not seen by the spatial operator and persist on their own once excited. The result is a noisy solution that can be avoided by suppressing these short wavelengths. One approach is to replace the delta-function source in the weak form by a source of wider extent. We have performed numerical experiments to find suitable parameters for the Gaussian, for the tapered sinc and for a polynomial approximation of the delta function. The tapered sinc provided the most accurate results.

In the standard second-order form, the Gaussian and tapered sinc hardly improve the accuracy and a delta function appears to be the most attractive choice, given its simplicity. The first-order form with one iteration may have a better accuracy than the second-order form, but that does not appear sufficient to compensate for its higher cost, at least not in our 2-D Matlab implementations. The second-order form and in particular its higher-order mass-lumped versions appears to be more attractive.

In chapter 5, we have compared the performance of mass-lumped tetrahedral finite elements on isotropic elastic wave propagation without and with global assembly of the stiffness matrix. To preserve their accuracy after mass lumping, the higher-order elements are augmented with higher-degree polynomials in the interior of the faces and the tetrahedron. For the lowest degree, the linear elements, this is not necessary. For that case, we simplified the expression for the stiffness matrix.

We ran performance tests on a homogeneous problem. The parallelization of the most compute intensive loops was performed by OpenMP directives. With global assembly, this involved symmetric sparse matrix assembly and the matrix-vector product during the time stepping. With assembly on the fly, the local assembly and local matrix-vector multiplication per element were parallelized in a single OpenMP loop. Further code optimizations were left to the compiler.

In the acoustic case, local assembly is more efficient than global assembly, except for the lowest-order case with linear elements. In the elastic case, the same appears to be true. For degree 1, the code with global assembly ran faster and used about 40 per cent more storage than with local assembly. For degree 2, the numerical experiments with local assembly on the fly on 24 cores were about 1.4 times faster than with global assembly in one experiment and about two times in another. For degree 3, the gain was a factor 1.9 in one and 3 in the other. At the same time, the memory requirements were smaller by at least on order of magnitude for degrees 2 and 3.

We observed in a simple test problem that, for high accuracy, augmented cubic elements performed best in terms of compute time for a given accuracy. For low accuracy, the linear element may still be attractive. In that case, its efficiency compensates the need for a much finer mesh.

In chapter 6, we have analyzed the dispersion properties of finite elements based on cubic Hermite polynomials applied to the constant-density acoustic wave equation with a constant velocity. The dispersion curve has a sixth-order error, whereas the eigenvector error that describes the cross talk with the spurious mode has a error of order six for the pressure and of order four for its gradient. The accuracy at higher wavenumbers is reasonable up to a value somewhat below the Nyquist limit for the scalar case. A numerical test in two space dimensions shows fourth-order accuracy for the pressure and one order lower for its gradient.

The current, straightforward approach cannot be extended to problems with a discontinuous density in more than one space dimension. In the variable-density case, the pressure and the normal velocity or acceleration component should be continuous across the edges. One could think of a representation in terms of pressure and derivatives of pressure scaled by a piecewise constant density per element. But then, the tangential component of the pressure gradient along the edge may become discontinuous, which violates the continuity of the pressure across edges. Additional complications are the fact that bubble function generates a discontinuity in the normal component of the pressure gradient across the edges, except at the vertices, whereas the PDE requires its continuity, although in the homogeneous case, this does not pose a real problem.

While FEM has proven to be a great tool for seismic modeling, several areas remain to be explored by future researchers. One is to exploit the continuity of the normal velocity and not only the pressure in heterogeneous media with discontinuous material properties. Another is the use of curved elements. Automatic mesh generation for geophysical applications is highly needed. Unlike applications in manifacturing, where edges need to be represented with high accuracy, the mesh needs to follow the topography, seismic horizons and outlines of geological bodies like salt diapirs only with a fraction of a seismic wavelength.

The acoustic equation is a simplification of the elastodynamic wave equation. To capture more phenomena accurately, one include anisotropic and viscous effects. Such work has been done extensively for the static case in the mechanics of elasticity for various applications (Castaings et al., 2004; Hilton and Yi, 1993; Moresi et al., 2003; Puso and Weiss, 1998; Taylor et al., 2009). In seismics, it is less common, mainly because of the associated compute cost (Komatitsch et al., 2000). (Dumont et al., 2018) presents interesting work to reduce the cost of mesh regeneration in the form of 4D remeshing. Applying this technique in combination with the visco-elastic model for seismic problems might help to more strongly affirm a place for FEM in geophysics.

REFERENCES

Castaings, M., Bacon, C., Hosten, B., Predoi, M., 2004. Finite element predictions for the dynamic response of thermo-viscoelastic material structures. The Journal of the Acoustical Society of America 115 (3), 1125–1133.

Dumont, S., Jourdan, F., Madani, T., 2018. 4d remeshing using a space-time finite element method for elastodynamics problems. Mathematical and Computational Applications 23 (2).

Hilton, H. H., Yi, S., 1993. Anisotropic viscoelastic finite element analysis of mechanically and hygrothermally loaded composites. Composites Engineering 3 (2), 123–135.

Komatitsch, D., Barnes, C., Tromp, J., 2000. Simulation of anisotropic wave propagation based upon a spectral element method. Geophysics 65 (4), 1251–1260.

Moresi, L., Dufour, F., Mühlhaus, H.-B., 2003. A lagrangian integration point finite element method for large deformation modeling of viscoelastic geomaterials. Journal of computational physics 184 (2), 476–497.

Puso, M., Weiss, J., 1998. Finite element implementation of anisotropic quasi-linear viscoelasticity using a discrete spectrum approximation. Journal of biomechanical engineering 120 (1), 62–70.

Taylor, Z. A., Comas, O., Cheng, M., Passenger, J., Hawkes, D. J., Atkinson, D., Ourselin, S., 2009. On modelling of anisotropic viscoelasticity for soft tissue simulation: Numerical solution and gpu execution. Medical image analysis 13 (2), 234–244.

A

Spectral radius of G

The spectral radius of $\mathscr{G} = I-(\mathscr{M}^{\mathrm{L}})^{-1}\mathscr{M}$, with mass matrix $\mathscr{M}$ and its lumped version $\mathscr{M}^{\mathrm{L}}$, should be smaller than 1 for convergence. Here, we provide estimates on a periodic domain with N_x elements, each with size h_j, $j = 0,\dots,N_x-1$. The basis functions have degree M.

We start with some simple observations. The vector consisting of all ones is an eigenvector of $\mathscr{G}$ with eigenvalue 0. This follows immediately from the fact that $\mathscr{M}^{\mathrm{L}}$ is a diagonal matrix obtained from the row sums of $\mathscr{M}$. The eigenvalues of G do not change under the similarity transform $(\mathscr{M}^{\mathrm{L}})^{1/2}\mathscr{G}(\mathscr{M}^{\mathrm{L}})^{-1/2}$. Since this is a symmetric matrix, its eigenvalues should be non-negative. Note that $\mathscr{M}^{\mathrm{L}}$ has positive entries on the diagonal.

For the lowest degree, $M = 1$, the mass matrix per element is

$$A = \tfrac{1}{6}\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix},$$

and the assembled mass matrix is of the form $\mathscr{M}_{j,j-1} = \frac{1}{6}h_{j-1}$, $\mathscr{M}_{j,j} = \frac{1}{3}(h_{j-1}+h_j)$, $\mathscr{M}_{j,j+1} = \frac{1}{6}h_j$, and zero otherwise. In the periodic case, the j should be interpreted as j mod N_x. Then,

$$\mathscr{G}_{j,j-1} = -\tfrac{1}{3}\frac{h_{j-1}}{h_{j-1}+h_j}, \quad \mathscr{G}_{j,j} = \tfrac{1}{3}, \quad \mathscr{G}_{j,j+1} = -\tfrac{1}{3}\frac{h_j}{h_{j-1}+h_j},$$

and zero otherwise. In the equidistant case with constant h_j, the eigenfunctions are $\mathbf{q}_k$, $k = 0,\dots,N_x-1$, with $q_{k,l} = \exp(2\pi i k l/N_x)$, $l = 0,\dots,N_x-1$. The corresponding eigenvalues are $\frac{1}{3}[1-\cos(2\pi k/N_x)]$. Therefore, the eigenvalues of $\mathscr{G}$ lie in the interval $[0,2/3]$.

In the non-equidistant case, Gershgorin's theorem S.Gerhsgorin (1931) can be applied: $|\lambda - g_{i,i}| \le \sum_{j\neq i}|g_{i,j}|$ leads to $|\lambda-\frac{1}{3}| = \frac{1}{3}$, implying $0 \le \lambda \le 2/3$, which are the same bounds as in the equidistant case.

LEGENDRE POLYNOMIALS

We now turn to the general case, $M \ge 1$. The mass matrix for a single element in modal form is defined by $A^{\mathrm{m}}_{k,l} = \int_{-1}^{1} w(\zeta)\psi_k(\zeta)\psi_l(\zeta)$ with a weighting function $w(x)$ and model basis functions $\psi_k(\zeta)$, $k = 0,\dots,M$. The lumped mass matrix in nodal form is $2W$, where $W = \mathrm{diag}\{w_0, w_1,\dots,w_M\}$ is diagonal with $w_0 = w_M = 1/(M(M+1))$ and $w_j = 1/(M(M+

1)$P_M(x_j)^2$) for $j = 1, \dots, M-1$.

For Legendre polynomials, this results in a diagonal matrix with $A^{\mathrm{m}}_{j,j} = 1/(j + \frac{1}{2})$, $j = 0, \dots, M$. To obtain its nodal representation $A^{\mathrm{n}} = F^{\mathrm{n}} A^{\mathrm{m}}$, we take the Legendre-Gauss-Lobotto (LGL) points ζ_j that are the roots of $(1-\zeta^2)\frac{\mathrm{d}}{\mathrm{d}\zeta} P_M(\zeta) = 0$.

The modal-to-nodal map $F^{\mathrm{n}} = (F^{\mathrm{m}})^{-1}$ with $F^{\mathrm{m}}_{k,l} = \psi_k(\zeta_l)$, for $k, l = 0, \dots, M$. This can be expressed in closed form as (Teukolsky, 2015, e.g.)

$$F^{\mathrm{n}}_{j,k} = \frac{2w_j}{\gamma_k}\psi_k(\zeta_j), \quad 2w_j = A^{\mathrm{L}}_{j,j}, \quad \gamma_k = 2\sum_{j=0}^{M} w_j \psi_k^2(\zeta_j).$$

Here, $\gamma_k = 1/(k+\frac{1}{2})$ for $k = 0, \dots, M-1$ and $\gamma_M = 2/M$ with the LGL nodes. Note that the numerical quadrature weights w_j should not be confused with the weighting function $w(\zeta)$.

The nodal form of the basis functions is $\boldsymbol{\phi} = F^{\mathrm{n}}\boldsymbol{\psi}$. We have $\phi_k(\zeta_l) = \delta_{k,l}$ by definition and $\psi_k(\zeta) = \sum_{l=0}^{M} \psi_k(\zeta_l)\phi_l(\zeta)$. This is the same as the earlier $F^{\mathrm{m}}\boldsymbol{\phi}$.

The lumped version of A^{n} is A^{L}, a diagonal matrix obtained from the row sums: $A^{\mathrm{L}}_{k,k} = \sum_{l=0}^{M} A^{\mathrm{n}}_{k,l}$. The latter are proportional to the LGL quadrature weights:

$$w_k = \frac{1}{2}A^{\mathrm{L}}_{k,k} = \left[M(M+1)\,(P_M(\zeta_k))^2\right]^{-1}.$$

The difference between the mass matrices is expressed by

$$(A^{\mathrm{L}} - A^{\mathrm{n}})_{j,k} = \left(\frac{2}{\gamma_M}\right)^2 \left(\gamma_M - \frac{1}{M+\frac{1}{2}}\right) w_j w_k P_M(\zeta_j) P_M(\zeta_k),$$

where $\gamma_M = 2/M$, so

$$(A^{\mathrm{L}} - A^{\mathrm{n}})_{j,k} = \frac{2M(1+M)}{2M+1} w_j w_k P_M(\zeta_j) P_M(\zeta_k).$$

Define a vector $\mathbf{f}$ with $f_k = w_k P_M(\zeta_k)$. Then $A^{\mathrm{L}} - A^{\mathrm{n}} = \frac{2M(1+M)}{2M+1}\mathbf{f}\mathbf{f}^{\mathsf{T}}$. We immediately obtain an eigenvector $\mathbf{f}$. Since

$$\mathbf{f}\cdot\mathbf{f} = \sum_{k=0}^{M} [w_k P_M(\zeta_k)]^2 = \sum_{k=0}^{M} \left(\frac{P_M(\zeta_k)}{M(M+1)[P_M(\zeta_k)]^2}\right)^2 =$$

$$= \frac{1}{M(M+1)} \sum_{k=0}^{M} w_k = \frac{1}{M(M+1)},$$

the corresponding eigenvalue is $\frac{2}{2M+1}$. The other eigenvalues are zero because the matrix has rank 1.

Next, consider the matrix $G = (A^{\mathrm{L}})^{-1}(A^{\mathrm{L}} - A^{\mathrm{n}}) = \frac{2M(1+M)}{2M+1}(A^{\mathrm{L}})^{-1}\mathbf{f}\mathbf{f}^{\mathsf{T}}$. As

$$\mathbf{f}^{\mathsf{T}}(A^{\mathrm{L}})^{-1}\mathbf{f} = \frac{1}{2}\sum_{k=0}^{M} w_k \left(P_M(\zeta_k)\right)^2 = \frac{1}{2M},$$

the matrix has an eigenvector $\mathbf{q} = 2(A^{\mathrm{L}})^{-1}\mathbf{f}$ with entries $q_k = P_M(\zeta_k)$ and the corresponding eigenvalue is $\lambda_{\max} = (M+1)/(2M+1)$. The other eigenvalues are zero, as before.

To go from this result to the assembled case, we follow Wathen (1987). The bounds of the eigenvalues, λ, obey

$$\min_{\mathbf{x}\neq 0} \frac{\mathbf{x}^{\mathsf{T}}(\mathscr{M}^{\mathrm{L}} - \mathscr{M})\mathbf{x}}{\mathbf{x}^{\mathsf{T}}\mathscr{M}^{\mathrm{L}}\mathbf{x}} \leq \lambda \leq \max_{\mathbf{x}\neq 0} \frac{\mathbf{x}^{\mathsf{T}}(\mathscr{M}^{\mathrm{L}} - \mathscr{M})\mathbf{x}}{\mathbf{x}^{\mathsf{T}}\mathscr{M}^{\mathrm{L}}\mathbf{x}},$$

For boolean matrix L represents the local-to-global map that take $(M+1)$ unknowns on the N_x elements to the global MN_x unknows. Then,

$$\min_{\mathbf{x}\neq 0} \frac{\mathbf{x}^{\mathsf{T}}L^{\mathsf{T}}(A^{\mathrm{L}} - \mathbf{A})L\mathbf{x}}{\mathbf{x}^{\mathsf{T}}L^{\mathsf{T}}A^{\mathrm{L}}L\mathbf{x}} \leq \lambda \leq \max_{\mathbf{x}\neq 0} \frac{\mathbf{x}^{\mathsf{T}}L^{\mathsf{T}}(A^{\mathrm{L}} - \mathbf{A})L\mathbf{x}}{\mathbf{x}^{\mathsf{T}}L^{\mathsf{T}}A^{\mathrm{L}}L\mathbf{x}},$$

which after setting $\mathbf{y} = L\mathbf{x}$, results in

$$\min_{\mathbf{y}\neq 0} \frac{\mathbf{y}^{\mathsf{T}}(A^{\mathrm{L}} - \mathbf{A})\mathbf{y}}{\mathbf{y}^{\mathsf{T}}A^{\mathrm{L}}\mathbf{y}} \leq \lambda \leq \max_{\mathbf{y}\neq 0} \frac{\mathbf{y}^{\mathsf{T}}(A^{\mathrm{L}} - \mathbf{A})\mathbf{y}}{\mathbf{y}^{\mathsf{T}}A^{\mathrm{L}}\mathbf{y}}.$$

Let $\mathbf{y}' = (A^{\mathrm{L}})^{1/2}\mathbf{y}$, using the fact that A^{L} is diagonal with positive entries on the diagonal. Then,

$$\min_{\mathbf{y}'\neq 0} \frac{(\mathbf{y}')^{\mathsf{T}}(A^{\mathrm{L}})^{-1/2}(A^{\mathrm{L}} - \mathbf{A})(A^{\mathrm{L}})^{-1/2}\mathbf{y}'}{(\mathbf{y}')^{\mathsf{T}}\mathbf{y}'} \leq \lambda \leq$$

$$\max_{\mathbf{y}'\neq 0} \frac{(\mathbf{y}')^{\mathsf{T}}(A^{\mathrm{L}})^{-1/2}(A^{\mathrm{L}} - \mathbf{A})(A^{\mathrm{L}})^{-1/2}\mathbf{y}'}{(\mathbf{y}')^{\mathsf{T}}\mathbf{y}'}.$$

The bounds follow from the smallest and largest eigenvalues of $(A^{\mathrm{L}})^{-1/2}(A^{\mathrm{L}} - \mathbf{A})(A^{\mathrm{L}})^{-1/2}$, which by a similarity transform based on $(A^{\mathrm{L}})^{1/2}$ are the same as those of $(A^{\mathrm{L}})^{-1}(A^{\mathrm{L}} - \mathbf{A})$, namely zero and $\lambda_{\max} = (M+1)/(2M+1)$.

Note that $\mathscr{G}$ has $N_x(M-1)$ zero and N_x non-zero eigenvalues, reflecting the fact that the element matrix $A^{\mathrm{L}} - A^{\mathrm{n}}$ has rank 1.

For even M, the maximum eigenvalue is obtained for a vector $\mathbf{v}$ obtained from chaining the highest modal function $P_M(\zeta)$ over the nodes. Consider an indexing function $q(j,k) = (Mj+k) \bmod MN_x$ that enumerates the MN_x degrees of freedom on a periodic grid with elements $j = 0,\ldots,N_x-1$ and nodes per element $k = 0,\ldots,M$. The vector $\mathbf{v}$ has elements $v_{q(j,k)} = P_M(\zeta_k^{LGL})$, the highest degree Legendre polynomial evaluated at the LGL nodes ζ_k^{LGL}. Recall that G refers to a single element and does not contain the element size. Therefore, the subset $\mathscr{G}_{q(j,k_1),q(j,k_2)} = G_{k_1,k_2}$, corresponding to the interior nodes with $k_1 = 1,\ldots,M-1$ and $k_2 = 0,\ldots,M$, does not depend on the element size h_j. At the endpoints, we have $\mathscr{G}_{q(j,0),q(j,0)-l} = \frac{h_{j-1}}{h_{j-1}+h_j} G_{0,l}$ and $\mathscr{G}_{q(j,0),q(j,0)+l} = \frac{h_j}{h_{j-1}+h_j} G_{0,l}$ for $l = 1,\ldots,M$, whereas $\mathscr{G}_{q(j,0),q(j,0)} = G_{0,0}$. Since for even values of M, the corresponding $\mathbf{v}$ is symmetric according to $v_{q(j,l)} = v_{q(j,-l)}$, for $l = 1,\ldots,M$, we find that $\mathscr{G}\mathbf{v} = \lambda_{\max}\mathbf{v}$.

For M odd but N_x even, we can do the same, but since $P_M(-1) = -1$ in that case, a minus sign needs to be applied in alternating elements: $v_{q(j,k)} = (-1)^j P_M(\zeta_k^{LGL})$. Note that application of a minus sign has the effect of reversal of the order: $P_M(\zeta_{M-k}^{LGL}) = -P_M(\zeta_k^{LGL})$ for $k = 0,\ldots,M$. With this vector, the same approach as above leads to $\mathscr{G}\mathbf{v} = \lambda_{\max}\mathbf{v}$.

Chebychev polynomials

The weighting function can be taken as $w(\zeta) = \frac{2}{\pi}(1-\zeta^2)^{-1/2}$, with an extra factor $2/\pi$ to integrate a unit constant to 2, as in the case of the Legendre polynomials. The modal basis functions are $\psi_k(\zeta) = T_k(\zeta) = \cos(k \arccos\zeta)$, $k = 0,\ldots,M$, and the Chebychev-Gauss-Lobotto (CGL) nodes $\zeta_l = -\cos(\pi l/M)$, $l = 0,\ldots,M$. The modal-to-nodal map has entries

$$F^{\mathrm{n}}_{j,k} = (-1)^k 2M w_j w_k \cos(\pi jk/M),$$

with $w_j = 1/M$, for $j = 1,\ldots,M-1$ and $w_0 = w_M = 1/(2M)$ (Funaro, 1992, eq. 3.5.6). The mass matrix in model form is $A^{\mathrm{m}} = \mathrm{diag}\{2,1,\ldots,1\}$, which represents the orthogonality of the Chebyshev polynomials. For its lumped version, we can show that $F^{\mathrm{m}} A^{\mathrm{L}} (F^{\mathrm{m}})^{\mathsf{T}} = \mathrm{diag}\{2,1,\ldots,1,2\}$. Knowing that numerical quadrature with the CGL nodes is exact for polynomials up to degree $2M-1$, we expect that the non-zero eigenvector can be represented by the modal basis function of highest degree, evaluated at the CGL nodes. If this is expressed as $\mathbf{q}$ with entries $q_j = T_M(\zeta_j) = (-1)^{M-j}$, $j = 0,\ldots,M$, then $(F^{\mathrm{n}}\mathbf{q})_j = \delta_{j,M}$.

From this, it follows that $(A^{\mathrm{L}} - A^{\mathrm{n}})\mathbf{q} = F^{\mathrm{n}}\mathrm{diag}\{0,0,\ldots,0,1\}F^{\mathrm{n}}\mathbf{q} = \frac{1}{2}F^{\mathrm{n}}\mathrm{diag}\{2,1,\ldots,1,2\}F^{\mathrm{n}}\mathbf{q} = \frac{1}{2}A^{\mathrm{L}}\mathbf{q}$. Using the same approach of Wathen (1987) as before, this implies that the eigenvalues of $\mathscr{G}$ lie between 0 and 1/2.

REFERENCES

Funaro, D., 1992. Polynomial Approximations of Differential Equations. Lecture Notes in Physics. New Series m: Monographs, Vol. 8. Springer-Verlag, Berlin.

S.Gerhsgorin, 1931. Uber die abgrenzung der eigenwerte einer matrix. Bulletin de l'Academie des Sciences de l'URSS. Classe des sciences mathématiques et naturelles 6, 749–754.

Teukolsky, S. A., 2015. Short note on the mass matrix for Gauss-Lobatto grid points. Journal of Computational Physics 283, 408–413.

Wathen, A. J., 1987. Realistic eigenvalue bounds for the Galerkin mass matrix. IMA Journal of Numerical Analysis 7 (4), 449–457.

B

Leading dispersion errors for Legendre-Gauss-Lobatto

The leading error term in the dispersion curve for the Legendre polynomials without lumping can be found in (Ainsworth, 2014, eq. (14)). In our notation and after division by $\mathrm{i}M\xi$, this provides

$$\varepsilon_{\mathrm{C}} \sim \tfrac{1}{2}(-1)^M \left(\frac{M!}{(2M+1)!}\right)^2 \begin{cases} \frac{M+1}{2M+3}(M\xi)^{2(M+1)} & \text{if } M \text{ odd,} \\ \frac{2M+1}{M+1}(M\xi)^{2M} & \text{if } M \text{ even.} \end{cases} \tag{B.1}$$

With mass lumping and the Legendre-Gauss-Lobatto (LGL) points, we conjecture that the leading error term is

$$\varepsilon_{\mathrm{L}} \sim -\frac{(M\xi)^{2M}}{2M+1}\left(\frac{M!}{(2M)!}\right)^2 \left(\frac{M}{M+1}\right)^{(-1)^M}. \tag{B.2}$$

We have verified this last result up to $M = 10$. For odd M, this matches the very last equation in Mulder (1999), which describes the error caused by replacing the consistent mass matrix by its lumped version. For even M, $\varepsilon_{\mathrm{L}} = -2M\varepsilon_{\mathrm{C}}$.

LEADING EIGENVECTOR ERRORS

In the following, we will present expressions for the discrete dispersion and for eigenvector errors. For the mass matrix, the consistent and lumped versions are considered. We only consider Legendre polynomials up to degree $M = 5$ and Legendre-Gauss-Lobatto (LGL) nodes. For reference, results for the second-order formulation of the wave equations are included, for which some can be also found elsewhere Mulder (1999). For the first-order case, the eigenvalues of the discrete operator are $\mathrm{i}M\kappa$. For the second-order case, they are κ^2 and we list only the non-negative values of κ. Because the analytic expressions rapidly become quite complicated, only results in the form of leading terms in a series representation in terms of the normalized wavenumber $\xi \in [-\pi,\pi]$ are given. Since for polynomials of degree M, M modes are coupled if elements of constant size and constant material parameters are considered, the eigenvalues come in groups of M elements and the corresponding eigenvector errors can be represented by the M columns of the matrix S, as explained in Section 3.3.2. Among the M eigenvalues, one corresponds to the 'physical' eigenvalue that approximates $\mathrm{i}M\xi$ in the first-order or $(M\xi)^2$ in the second-order formulation. The eigenvector error is absent and therefore zero for degree $M = 1$. For the higher degrees, a zero entry in the matrix should be read as $o(\xi^p)$, with p the power of ξ pulled out in front of the matrix.

$M = 1$, LGL, first-order, consistent mass matrix:

$$\kappa = \frac{3\sin\xi}{2+\cos\xi} \sim \xi\left(1 - \tfrac{1}{180}\xi^4\right), \quad S = 0.$$

$M = 1$, LGL, first-order, lumped mass matrix:

$$\kappa = \sin\xi \sim \xi\left(1 - \tfrac{1}{6}\xi^2\right), \quad S = 0.$$

$M = 1$, LGL, second-order, consistent mass matrix:

$$\kappa = \sqrt{\frac{6(1-\cos\xi)}{2+\cos\xi}} \sim \xi\left(1 + \tfrac{1}{24}\xi^2\right), \quad S = 0.$$

$M = 1$, LGL, second-order, lumped mass matrix:

$$\kappa = \sqrt{2(1-\cos\xi)} \sim \xi\left(1 - \tfrac{1}{24}\xi^2\right), \quad S = 0.$$

$M = 2$, LGL, first-order, consistent mass matrix:

$$\kappa = -\frac{\sin(\xi)\left(2\cos\xi \mp \sqrt{10-\cos^2\xi}\right)}{2-\cos^2\xi} \sim \left\{-5\xi, \xi\left(1 + \tfrac{1}{270}\xi^4\right)\right\}, \quad S = S^{\mathrm{F2C}} \sim \frac{\xi^2}{36}\begin{pmatrix} -2 & 2 \\ 1 & -1 \end{pmatrix}.$$

$M = 2$, LGL, first-order, lumped mass matrix:

$$\kappa = -\tfrac{1}{2}\sin\xi\left(\cos\xi \mp \sqrt{8+\sin^2\xi}\right) \sim \left\{-2\xi, \xi\left(1 - \tfrac{2}{135}\xi^4\right)\right\}, \quad S = S^{\mathrm{F2L}} \sim 2\,S^{\mathrm{F2C}}.$$

$M = 2$, LGL, second-order, consistent mass matrix:

$$\kappa \sim \left\{\xi\left(1 + \tfrac{1}{90}\xi^4\right), \sqrt{15}\right\}, \quad S = S^{\mathrm{S2C}} \sim -\frac{\xi^4}{360}\begin{pmatrix} -2 & 2 \\ 1 & -1 \end{pmatrix}.$$

$M = 2$, LGL, second-order, lumped mass matrix:

$$\kappa \sim \left\{\xi\left(1-\tfrac{1}{180}\xi^4\right), \sqrt{6}\right\}, \quad S = S^{\mathrm{S2L}} \sim -5\, S^{\mathrm{S2C}}.$$

$M = 3$, LGL, first-order, consistent mass matrix:

$$\kappa \sim \left\{-\sqrt{\tfrac{14}{3}}, \xi\left(1-\tfrac{81}{39200}\xi^8\right), \sqrt{\tfrac{14}{3}}\right\},$$

$$S = S^{\mathrm{F3C}} \sim \tfrac{27}{28000}\xi^4 \begin{pmatrix} 25 & -50 & 25 \\ -5-\mathrm{i}\sqrt{210} & 10 & -5+\mathrm{i}\sqrt{210} \\ -5+\mathrm{i}\sqrt{210} & 10 & -5-\mathrm{i}\sqrt{210} \end{pmatrix}.$$

$M = 3$, LGL, first-order, lumped mass matrix:

$$\kappa \sim \left\{-\sqrt{\tfrac{10}{3}}, \xi\left(1-\tfrac{27}{2800}\xi^6\right), \sqrt{\tfrac{10}{3}}\right\},$$

$$S \sim \tfrac{9}{800}\xi^4 \begin{pmatrix} 5 & -10 & 5 \\ -1-\mathrm{i}\sqrt{6} & 2 & -1+\mathrm{i}\sqrt{6} \\ -1+\mathrm{i}\sqrt{6} & 2 & -1-\mathrm{i}\sqrt{6} \end{pmatrix}.$$

$M = 3$, LGL, second-order, consistent mass matrix:

$$\kappa \sim \left\{\xi\left(1+\tfrac{81}{22400}\xi^6\right), \sqrt{\tfrac{14}{3}}, \sqrt{\tfrac{20}{3}}\right\}, \quad S = S^{\mathrm{S3C}} \sim \mathrm{i}\,\tfrac{81\sqrt{5}}{35000}\xi^5 \begin{pmatrix} 0 & 0 & 0 \\ 1 & -1 & 0 \\ -1 & 1 & 0 \end{pmatrix}.$$

$M = 3$, LGL, second-order, lumped mass matrix:

$$\kappa \sim \left\{\xi\left(1-\tfrac{27}{22400}\xi^6\right), \sqrt{\tfrac{10}{3}}, \sqrt{\tfrac{20}{3}}\right\}, \quad S = S^{\mathrm{S3L}} \sim -\tfrac{7}{3}\, S^{\mathrm{S4C}}.$$

$M = 4$, LGL, first-order, consistent mass matrix:

$$\kappa \sim \left\{-\sqrt{\tfrac{21}{8}}, -9\xi, \xi\left(1+\tfrac{128}{496125}\xi^8\right), \sqrt{\tfrac{21}{8}}\right\}, \quad S = S^{\mathrm{F4C}} \sim \tfrac{1}{3675}\xi^4 \begin{pmatrix} 0 & 56 & -56 & 0 \\ 0 & -24 & 24 & 0 \\ 0 & 21 & -21 & 0 \\ 0 & -24 & 24 & 0 \end{pmatrix}.$$

$M = 4$, LGL, first-order, lumped mass matrix:

$$\kappa \sim \left\{-\sqrt{\tfrac{21}{8}}, -4\xi, \xi(1-\tfrac{1024}{496125}\xi^8), \sqrt{\tfrac{21}{8}}\right\}, \quad S = S^{\mathrm{F4L}} \sim 2\,S^{\mathrm{F4C}}.$$

$M = 4$, LGL, second-order, consistent mass matrix:

$$\kappa \sim \left\{\xi(1+\tfrac{128}{99225}\xi^8), \tfrac{1}{4}\sqrt{210-6\sqrt{805}}, \tfrac{1}{4}\sqrt{42}, \tfrac{1}{4}\sqrt{210+6\sqrt{805}}\right\},$$

$$S \sim \frac{2\xi^6}{5325075} \begin{pmatrix} 0 & -336\sqrt{805} & 0 & 336\sqrt{805} \\ 7360 & -16\left(230+\sqrt{805}\right) & 0 & -16\left(230-\sqrt{805}\right) \\ -11270 & 7\left(805+17\sqrt{805}\right) & 0 & 7\left(805-17\sqrt{805}\right) \\ 7360 & -16\left(230+\sqrt{805}\right) & 0 & -16\left(230-\sqrt{805}\right) \end{pmatrix}.$$

$M = 4$, LGL, second-order, lumped mass matrix:

$$\kappa \sim \left\{\xi(1-\tfrac{32}{99225}\xi^8), \sqrt{\tfrac{1}{8}(55-\sqrt{1345})}, \sqrt{21/8}, \sqrt{\tfrac{1}{8}(55+\sqrt{1345})}\right\},$$

$$S \sim \frac{\xi^6}{20760075} \begin{pmatrix} 0 & 3136\sqrt{1345} & 0 & -3136\sqrt{1345} \\ -86080 & 32\left(1345+13\sqrt{1345}\right) & 0 & 32\left(1345-13\sqrt{1345}\right) \\ 131810 & -49\left(1345+31\sqrt{1345}\right) & 0 & -49\left(1345-31\sqrt{1345}\right) \\ -86080 & 32\left(1345+13\sqrt{1345}\right) & 0 & 32\left(1345-13\sqrt{1345}\right) \end{pmatrix}.$$

$M = 5$, LGL, first-order, consistent mass matrix:

$$\kappa \sim \left\{-\tfrac{2}{5}\sqrt{3(10+3\sqrt{5})}, -\tfrac{2}{5}\sqrt{3(10-3\sqrt{5})}, \xi(1-\tfrac{9765625}{19179224064}\xi^{12}), \tfrac{2}{5}\sqrt{3(10-3\sqrt{5})}, \tfrac{2}{5}\sqrt{3(10+3\sqrt{5})}\right\},$$

$$S \sim \tfrac{3125\xi^6}{133056}\begin{pmatrix} -\frac{1}{36}\left(9+5\sqrt{5}\right) & -\frac{1}{36}\left(9-5\sqrt{5}\right) & 1 & -\frac{1}{36}\left(9-5\sqrt{5}\right) & -\frac{1}{36}\left(9+5\sqrt{5}\right) \\ 0.1118+0.2522\mathrm{i} & 0.04879-0.04359\mathrm{i} & -\frac{1}{63}\left(7+5\sqrt{7}\right) & 0.04879+0.04359\mathrm{i} & 0.1118-0.2522\mathrm{i} \\ -0.008881-0.1704\mathrm{i} & -0.04055-0.04945\mathrm{i} & -\frac{1}{63}\left(7-5\sqrt{7}\right) & -0.04055+0.04945\mathrm{i} & -0.008881+0.1704\mathrm{i} \\ -0.008881+0.1704\mathrm{i} & -0.04055+0.04945\mathrm{i} & -\frac{1}{63}\left(7-5\sqrt{7}\right) & -0.04055-0.04945\mathrm{i} & -0.008881-0.1704\mathrm{i} \\ 0.1118-0.2522\mathrm{i} & 0.04879+0.04359\mathrm{i} & -\frac{1}{63}\left(7+5\sqrt{7}\right) & 0.04879-0.04359\mathrm{i} & 0.1118+0.2522\mathrm{i} \end{pmatrix}.$$

The closed-form expressions for the numerical entries are a bit lengthy. Let $S = s_{1,3}H$. Then,

$$h_{2,1} = \tfrac{1}{756}(21+8\sqrt{5}+15\sqrt{7}+\sqrt{35}) + \tfrac{\mathrm{i}}{1764}\sqrt{77(980+399\sqrt{5}+130\sqrt{7}+60\sqrt{35})},$$

$$h_{2,2} = \tfrac{1}{756}(21-8\sqrt{5}+15\sqrt{7}-\sqrt{35}) - \tfrac{\mathrm{i}}{1764}\sqrt{77(980-399\sqrt{5}+130\sqrt{7}-60\sqrt{35})},$$

and

$$h_{k,5} = h^*_{k,1}, \quad h_{k,4} = h^*_{k,2}, \quad h_{5,k} = h^*_{2,k}, \quad h_{4,k} = h^*_{3,k}, \quad k = 1,\dots,5.$$

$M = 5$, LGL, first-order, lumped mass matrix:

$$\kappa \sim \left\{-\tfrac{2}{5}\sqrt{3(7+\sqrt{14})}, -\tfrac{2}{5}\sqrt{3(7-\sqrt{14})}, \xi(1-\tfrac{78125}{67060224}\xi^{10}), \tfrac{2}{5}\sqrt{3(7-\sqrt{14})}, \tfrac{2}{5}\sqrt{3(7+\sqrt{14})}\right\},$$

$$S \sim \tfrac{625\xi^6}{12096}\begin{pmatrix} -0.5618 & 0.0618 & 1 & 0.0618 & -0.5618 \\ 0.1025+0.2115\mathrm{i} & 0.058-0.04849\mathrm{i} & -0.3211 & 0.058+0.04849\mathrm{i} & 0.1025-0.2115\mathrm{i} \\ -0.002446-0.1207\mathrm{i} & -0.04699-0.05795\mathrm{i} & 0.09887 & -0.04699+0.05795\mathrm{i} & -0.002446+0.1207\mathrm{i} \\ -0.002446+0.1207\mathrm{i} & -0.04699+0.05795\mathrm{i} & 0.09887 & -0.04699-0.05795\mathrm{i} & -0.002446-0.1207\mathrm{i} \\ 0.1025-0.2115\mathrm{i} & 0.058+0.04849\mathrm{i} & -0.3211 & 0.058-0.04849\mathrm{i} & 0.1025+0.2115\mathrm{i} \end{pmatrix}.$$

Again, with $S = s_{1,3}H$, we have

$$h_{1,1} = -\tfrac{1}{12}(3+\sqrt{14}), \quad h_{1,2} = -\tfrac{1}{12}(3-\sqrt{14}),$$

$$h_{2,1} = \tfrac{1}{504}(14+10\sqrt{7}+3\sqrt{14}) + \tfrac{\mathrm{i}}{504}\sqrt{70(59+20\sqrt{2}+10\sqrt{7}+13\sqrt{14})},$$

$$h_{2,2} = \tfrac{1}{504}(14+10\sqrt{7}-3\sqrt{14}) - \tfrac{\mathrm{i}}{504}\sqrt{70(59-20\sqrt{2}+10\sqrt{7}-13\sqrt{14})},$$

$$h_{3,1} = \tfrac{1}{504}(14-10\sqrt{7}+3\sqrt{14}) - \tfrac{\mathrm{i}}{504}\sqrt{70(59-20\sqrt{2}-10\sqrt{7}+13\sqrt{14})},$$

$$h_{3,2} = \tfrac{1}{504}(14-10\sqrt{7}-3\sqrt{14}) - \tfrac{\mathrm{i}}{504}\sqrt{70(59+20\sqrt{2}-10\sqrt{7}-13\sqrt{14})},$$

$$h_{2,3} = -(7+5\sqrt{7})/63, \quad h_{3,3} = -(7-5\sqrt{7})/63,$$

and the other entries follow the same symmetry pattern as in the previous case. $M = 5$, LGL, second-order, consistent mass matrix:

$$\kappa \sim \left\{\xi\left(1+\tfrac{390625}{804722688}\xi^{10}\right), \tfrac{2}{5}\sqrt{3(10-3\sqrt{5})}, \tfrac{1}{5}\sqrt{6(35-\sqrt{805})}, \tfrac{2}{5}\sqrt{3(10+3\sqrt{5})}, \tfrac{1}{5}\sqrt{6(35+\sqrt{805})}\right\},$$

$$S \sim \mathrm{i}\frac{15625\xi^7}{58677696}\begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ -\frac{2}{3}\sqrt{3\left(49-10\sqrt{7}\right)} & -\sqrt{\frac{1}{21}(763-210\sqrt{5}-2\sqrt{35\left(761-336\sqrt{5}\right)})} & 0 & \sqrt{\frac{1}{21}(763+210\sqrt{5}+2\sqrt{35\left(761-336\sqrt{5}\right)})} & 0 \\ \frac{2}{3}\sqrt{3\left(49+10\sqrt{7}\right)} & -\sqrt{\frac{1}{21}(763-210\sqrt{5}+2\sqrt{35\left(761-336\sqrt{5}\right)})} & 0 & -\sqrt{\frac{1}{21}(763+210\sqrt{5}-2\sqrt{35\left(761-336\sqrt{5}\right)})} & 0 \\ -\frac{2}{3}\sqrt{3\left(49+10\sqrt{7}\right)} & \sqrt{\frac{1}{21}(763-210\sqrt{5}+2\sqrt{35\left(761-336\sqrt{5}\right)})} & 0 & \sqrt{\frac{1}{21}(763+210\sqrt{5}-2\sqrt{35\left(761-336\sqrt{5}\right)})} & 0 \\ \frac{2}{3}\sqrt{3\left(49-10\sqrt{7}\right)} & \sqrt{\frac{1}{21}(763-210\sqrt{5}-2\sqrt{35\left(761-336\sqrt{5}\right)})} & 0 & -\sqrt{\frac{1}{21}(763+210\sqrt{5}+2\sqrt{35\left(761-336\sqrt{5}\right)})} & 0 \end{pmatrix}$$

$$\sim \mathrm{i}\frac{15625\xi^7}{58677696}\begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ -5.48 & -3.50 & 0 & 8.98 & 0 \\ 10.03 & -3.97 & 0 & -6.06 & 0 \\ -10.03 & 3.97 & 0 & 6.06 & 0 \\ 5.48 & 3.50 & 0 & -8.98 & 0 \end{pmatrix}.$$

$M = 5$, LGL, second-order, lumped mass matrix:

$$\kappa \sim \left\{\xi\left(1 - \tfrac{78125}{804722688}\xi^{10}\right), \tfrac{2}{5}\sqrt{3(7-\sqrt{14})}, \tfrac{1}{5}\sqrt{6(35-\sqrt{805})}, \tfrac{2}{5}\sqrt{3(7+\sqrt{14})}, \tfrac{1}{5}\sqrt{6(35+\sqrt{805})}\right\},$$

$$S \sim \mathrm{i}\tfrac{3125\xi^7}{32006016}\begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 2\sqrt{3\left(49-10\sqrt{7}\right)} & \sqrt{\tfrac{3}{2}\left(231-32\sqrt{14}-10\sqrt{7\left(23-6\sqrt{14}\right)}\right)} & 0 & -\sqrt{\tfrac{3}{2}\left(231+32\sqrt{14}+10\sqrt{7\left(23+6\sqrt{14}\right)}\right)} & 0 \\ -2\sqrt{3\left(49+10\sqrt{7}\right)} & \sqrt{\tfrac{3}{2}\left(231-32\sqrt{14}+10\sqrt{7\left(23-6\sqrt{14}\right)}\right)} & 0 & \sqrt{\tfrac{3}{2}\left(231+32\sqrt{14}-10\sqrt{7\left(23+6\sqrt{14}\right)}\right)} & 0 \\ 2\sqrt{3\left(49+10\sqrt{7}\right)} & -\sqrt{\tfrac{3}{2}\left(231-32\sqrt{14}+10\sqrt{7\left(23-6\sqrt{14}\right)}\right)} & 0 & -\sqrt{\tfrac{3}{2}\left(231+32\sqrt{14}-10\sqrt{7\left(23+6\sqrt{14}\right)}\right)} & 0 \\ -2\sqrt{3\left(49-10\sqrt{7}\right)} & -\sqrt{\tfrac{3}{2}\left(231-32\sqrt{14}-10\sqrt{7\left(23-6\sqrt{14}\right)}\right)} & 0 & \sqrt{\tfrac{3}{2}\left(231+32\sqrt{14}+10\sqrt{7\left(23+6\sqrt{14}\right)}\right)} & 0 \end{pmatrix}$$

$$\sim \mathrm{i}\frac{3125\xi^7}{32006016}\begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 16.45 & 11.72 & 0 & -28.17 & 0 \\ -30.09 & 14.01 & 0 & 16.08 & 0 \\ 30.09 & -14.01 & 0 & -16.08 & 0 \\ -16.45 & -11.72 & 0 & 28.17 & 0 \end{pmatrix}.$$

REFERENCES

Ainsworth, M., 2014. Dispersive behaviour of high order finite element schemes for the one-way wave equation. Journal of Computational Physics 259, 1–10.

Mulder, W., 1999. Spurious modes in finite-element discretizations of the wave equation may not be all that bad. Applied Numerical Mathematics 30 (4), 425–445.

C

Another window function

We examine a method that possibly may compensate for the second-order impact of the discretization. Consider a periodic equidistant 1-D mesh with $x_j = jh$, $j = 0, 1, \dots, N_x - 1$. The mass matrix has a Fourier symbol $\hat{M} = h\big[1 - \frac{2}{3}\sin^2(\xi/2)\big]$ with $\xi = kh$ and wavenumber k. Integration against the basis function has an operator symbol $\hat{\Phi} = h\big[\sin(\xi/2)/(\xi/2)\big]^2$, corresponding to the linear operator $\mathbf{\Phi}$ having $(\mathbf{\Phi}\,\mathbf{f})_j = \int_\Omega \phi_j(x) f(x)\,\mathrm{d}x$.

We choose a window function with Fourier symbol

$$\hat{w} = \left[\frac{\sin(\xi/2)}{(\xi/2)}\right]^6 \big[1 + \alpha \sin^2(\xi/2)\big]. \tag{C.1}$$

The motivation for this choice is the finite-difference fourth-order polynomial approximation of the delta function (Petersson et al., 2016, eq. (8)), given by

$$w_4(\zeta) = \begin{cases} \frac{1}{32}(16 - 4|\zeta| - 4\zeta^2 + |\zeta|^3), & |\zeta| < 2, \\ \frac{1}{96}(48 - 44|\zeta| + 12\zeta^2 - |\zeta|^3), & 2 \le |\zeta| < 4, \\ 0, & |\zeta| \ge 4, \end{cases} \tag{C.2}$$

which has a Fourier symbol $\hat{w}_4 = \big[\sin(\xi/2)/(\xi/2)\big]^4(1 + \frac{1}{6}\xi^2) \simeq 1 - \frac{11}{720}\xi^4$, revealing its fourth-order behaviour. To undo the effect of $\hat{\Phi}$, we increase the power for the sinc function from 4 to 6 to obtain $\hat{w}$. Its Fourier transform back to the spatial domain becomes simpler if ξ^2 is replaced by $\sin^2(\xi/2)$. The expansion $\hat{w} \simeq 1 + \frac{1}{4}(\alpha - 1)\xi^2 + \frac{1}{240}(7 - 20\alpha)\xi^4$, provides a fourth-order approximation for $\alpha = 1$. The inverse Fourier transform to the spatial domain leads to

$$w(\zeta) = \frac{1}{\pi}\int_0^\infty \hat{\Phi}^{-1}\hat{M}\,\hat{w}\cos(\xi\zeta)\,d\xi, \tag{C.3}$$

with $\zeta = x/h$. We have included the mass matrix and the inverse of $\mathbf{\Phi}$. The result is the compact function

$$\begin{aligned} w(\zeta) = \frac{1}{288}\Big[&114|\zeta|^3 - 70\big(|\zeta-1|^3 + |\zeta+1|^3\big) + 8\big(|\zeta-2|^3 + |\zeta+2|^3\big) + \\ &6\big(|\zeta-3|^3 + |\zeta+3|^3\big) - \big(|\zeta-4|^3 + |\zeta+4|^3\big)\Big], \end{aligned} \tag{C.4}$$

or

$$w(\zeta) = \begin{cases} \frac{1}{144}(92 - 3\zeta^2(40 - 19|\zeta|)), & |\zeta| < 1, \\ \frac{1}{144}(162 - |\zeta|(210 - |\zeta|(90 - 13|\zeta|))), & 1 \le |\zeta| < 2, \\ \frac{1}{144}(98 - |\zeta|(114 - |\zeta|(42 - 5|\zeta|))), & 2 \le |\zeta| < 3, \\ \frac{1}{144}(4 - |\zeta|)^3, & 3 \le |\zeta| < 4, \\ 0, & |\zeta| \ge 4. \end{cases} \tag{C.5}$$

REFERENCES

Petersson, N. A., O'Reilly, O., Bj, 2016. Discretizing singular point sources in hyperbolic wave propagation problems. Journal of Computational Physics 321, 532–555.

D

Permutations

Given the symmetries of the node positions, we can define various permutation arrays and their corresponding matrices. Let the N_p nodes of the element be $\mathbf{x}_k$, $k = 0,\dots,N_p-1$. The permutation array $\mathbf{p}^{2,1}$ swaps their x and y coordinates with $\mathbf{x}_{\mathbf{p}^{2,1}(k)}$ as result. Likewise, $\mathbf{p}^{3,1}$ swaps x and z and $\mathbf{p}^{3,2}$ interchanges y and z.

To these arrays correspond matrices $\mathbf{P}^{m,n}$ with elements $\mathbf{P}^{m,n}_{k,p^{m,n}_k} = 1$ and zero otherwise. The inverse and transpose of the permutation matrix equal the matrix itself:

$$\left(\mathbf{P}^{m,n}\right)^{-1} = \left(\mathbf{P}^{m,n}\right)^{\mathsf{T}} = \mathbf{P}^{m,n}.$$

With these matrices, the stiffness matrices obey

$$\mathbf{B}^{2,2} = \mathbf{P}^{2,1}\mathbf{B}^{1,1}\mathbf{P}^{2,1}, \quad \mathbf{B}^{3,3} = \mathbf{P}^{3,1}\mathbf{B}^{1,1}\mathbf{P}^{3,1},$$

$$\mathbf{B}^{1,3} = \mathbf{P}^{3,2}\mathbf{B}^{1,2}\mathbf{P}^{3,2}, \quad \mathbf{B}^{3,2} = \mathbf{P}^{3,1}\mathbf{B}^{1,2}\mathbf{P}^{3,1}.$$

Because $(\mathbf{B}^{m_1,m_2})^{\mathsf{T}} = \mathbf{B}^{m_2,m_1}$, we have

$$\mathbf{B}^{3,1} = \mathbf{P}^{3,2}\mathbf{B}^{2,1}\mathbf{P}^{3,2}, \quad \mathbf{B}^{2,3} = \mathbf{P}^{3,1}\mathbf{B}^{2,1}\mathbf{P}^{3,1}.$$

Also,

$$\mathbf{B}^{1,2} = \mathbf{P}^{2,1}\mathbf{B}^{2,1}\mathbf{P}^{2,1}, \quad \mathbf{B}^{2,1} = \mathbf{P}^{2,1}\mathbf{B}^{1,2}\mathbf{P}^{2,1},$$

$$\mathbf{B}^{1,3} = \mathbf{P}^{3,1}\mathbf{B}^{3,1}\mathbf{P}^{3,1}, \quad \mathbf{B}^{3,1} = \mathbf{P}^{3,1}\mathbf{B}^{1,3}\mathbf{P}^{3,1},$$

$$\mathbf{B}^{3,2} = \mathbf{P}^{3,2}\mathbf{B}^{2,3}\mathbf{P}^{3,2}, \quad \mathbf{B}^{2,3} = \mathbf{P}^{3,2}\mathbf{B}^{3,2}\mathbf{P}^{3,2}.$$

In summary: with 2 matrices $\mathbf{B}^{1,1}$ and $\mathbf{B}^{1,2}$, computed on the reference element, and 3 permutation vectors, $\mathbf{p}^{2,1}$, $\mathbf{p}^{3,1}$, and $\mathbf{p}^{3,2}$, all 9 element stiffness matrices $\mathbf{B}^{p,q}$ can be determined.

Curriculum Vitæ

Ranjani Shamasundar

30-09-1989	Born in Mysore, India.

Education

2003–2007	Secondary school
	S. Cadambi Vidya Kendra, Bangalore (2003–2005)
	Kendriya Vidyalaya Malleswaram, Bangalore (2005–2007)
2007–2011	Undergraduate in Mechanical Engineering
	PES Institute of Technology, Bangalore
2011–2014	Postgraduate in Mechanical Engineering
	Indian Institute of Science, Bangalore
2014–2018	PhD. Geophysics
	Delft University of Technology, The Netherlands

Awards

2018	Best Poster Award at Annual Research Symposium, Science Center, Delft
2017	SEG Student Education Program Travel Grant

List of Publications

- Shamasundar, R. & Mulder, W. A., 2018. Numerical noise suppression for wave propagation with finite elements in first-order form by an extended source term, Geophysics Journal International, 215(2), 1231–1240.

- Mulder, W. A. & Shamasundar, R., 2016. Performance of continuous mass-lumped tetrahedral elements for elastic wave propagation with and without global assembly, Geophysical Journal International, 207(1), 414 – 421.

- Shamasundar, R. & Mulder, W. A., 2016. Improving the accuracy of mass-lumped finite-elements in the first-order formulation of the wave equation by defect correction, Journal of Computational Physics, 322, 689–707.

- Shamasundar, R. and Mulder, W.A., 2017 An Improved Source Term for Finite-element Modelling with the Stress-velocity Formulation of the Wave Equation. Extended Abstract, 79th EAGE Conference & Exhibition, Paris, France, 2017.

- Shamasundar, R. and Mulder, W.A., 2016 Should we use the first- or second-order formulation with spectral elements for seismic modelling? Extended Abstract, 78th EAGE Conference & Exhibition, Vienna, Austria, 2016.

- Shamasundar, R., Khoury, R.A. and Mulder, W.A. [2015] Dispersion analysis of finite-element schemes for a first-order formulation of the wave equation. Extended Abstract, 77th EAGE Conference & Exhibition, 2015.

Acknowledgements

My time as a PhD in Delft has been memorable thanks to many people. I would like to take this opportunity to express my gratitude to them.

First and foremost, I would like to thank my supervisor Prof. Wim Mulder for his support in every aspect towards the completion of this thesis. It would have been impossible to start, continue or finish my thesis without his encouragement and backing. Besides supporting me in his capacity as an erudite scientist and a seasoned programmer, he has also helped me brace myself in stressful patches. His sincerity and perseverance kept me inspired through my PhD. Thank you, Wim

I would like to thank Jeroen Goudswaard for organising my internship at Shell, and for all the help during my stay there. Suhas Phadke taught me how to use many of the tools and Alok Soni lent me the initial velocity model, besides his valuable time and inputs during this internship. Thanks to all my colleagues from the Shell, Bangalore team for enriching my journey in geophysics.

The time at TU Delft was interesting because of many more activities besides my PhD. I enjoyed tutoring bachelors students of Applied Earth Sciences, and I thank Dominique Ngard-Tillard and Timo Heimovaara for giving me an opportunity and necessary training to teach groups of young students. I picked up a lot of essential transferable skills, and also got an opportunity to improve my Dutch doing this.

Activities for Delft Organisation for Geophysics Students were fun because of my fellow board members during two years - Matteo, Sixue, Reuben, Max, Boris - I enjoyed working together with you to organise student activities at conferences, DOGS drinks, dinners, and lectures. Gil, Diego, Joeri - I enjoyed organising the numerical methods in geophysics symposium with you.

Some of the most cherished memories at the TU come from the social activities with my colleagues. My first geo-girls dinner started with Asiya, Iris, Helena and Sixue. This blossomed into a close friendship, and the dinners continue to this day. Gradually, with more women, activities got diverse. Lisanne, Karlien, Myrna - thank you for teaching me to knit, I hope to finish a scarf one day. Lunches and coffee breaks were filled with interesting conversation, and conferences were fun thanks to my colleagues Aayush, Alex, Aparajita, Apostolos, Carlos, Chris, Florencia, Giovanni, Jan-Willem, Lele, Martha, Neils, Nicolas, Pawan, Remi, Runhai, Siddharth, Shan, and Youwei in addition to those mentioned above.

A set of friends that made Delft home away from home is my Indian community - Amy, Anand, Anjana, Arun J, Arun M, Arvind, Bharadwaj, Chandrakanta, Dharamjeet, Freddy, Gautham, Kamakshi, Karthik, Jaikishen, JK, Lakshmi Ashok, Lakshmi Nagarajan, Manu, Nirupa, Pramoda, Prashanth, Rajesh, Seshanji, Sneha, Sriram Raghav, Sriram Shankar, Shyam and Vasu. I have very special memories with each one of you, which can fill a book - so I will reserve them for regaling later.

My parents, M.R.Geethamalini and Dr. S.Shamasundar are my constants. I will forever be grateful to them for guiding me till here and always trying their best to ensure that I have the right environment. Growing up with their pursuit of spirituality taught me to find emotional strength, courage and open mindedness. Baju, my closest ally, thanks for your love and support. My in-laws, Latha, Venu and Saru, thanks for your loving care in the past four years. I have reserved a very important one for the end. My lovely husband Syam, it would have been a herculean task to finish this alone. You helped me in the most practical ways, true to your pragmatic nature; it goes without saying that there are many more things to thank you for, which I will do spending the rest of my life with you.