



HAL
open science

3D Multi-parameters Full Waveform Inversion for challenging 3D elastic land targets

Phuong-Thu Trinh

► **To cite this version:**

Phuong-Thu Trinh. 3D Multi-parameters Full Waveform Inversion for challenging 3D elastic land targets. Earth Sciences. Université Grenoble Alpes, 2018. English. NNT : 2018GREAU033 . tel-01998056

HAL Id: tel-01998056

<https://theses.hal.science/tel-01998056v1>

Submitted on 29 Jan 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE

Pour obtenir le grade de

DOCTEUR DE LA

COMMUNAUTÉ UNIVERSITÉ GRENOBLE ALPES

Spécialité : **Sciences de la Terre, de l'Univers et de l'Environnement (CESTUE)**

Arrêté ministériel du 25 Mai 2016

Préparée au sein de **de l'Institut des Sciences de la Terre**
et de l'école doctorale **l'Ecole Doctorale Terre Univers Environnement**

3D Multi-parameters Full Waveform Inversion for challenging land targets

Présentée par

Phuong-Thu TRINH

Thèse dirigée par **Romain Brossier, Ludovic Métivier et Jean Virieux**

Thèse soutenue publiquement le 24 September 2018 ,
devant le jury composé de :

M, Dimitri Komatitsch	DR CNRS	LMA Marseille	Rapporteur
M, René-Edouard Plessix	SR	Shell Technology Center	Rapporteur
M, Thomas Bohlen	Prof.	Karlsruher Institut für Technologie	Examineur
M, Yann Capdeville	DR CNRS	LPG Nantes	Président
M, Bertrand Duquet	SR	Total Exploration & Production	Invité
M, Gilles Lambaré	DR	CGG	Invité
M, Romain Brossier	MCF	Université Grenoble Alpes	Directeur de thèse
M, Ludovic Métivier	CR CNRS	Université Grenoble Alpes	Directeur de thèse

To my brother, my parents and my supervisors

Acknowledgments

The works presented in this PhD manuscript could have not been achieved without the help of my supervisors, my colleges, my friends and my family. I would like to acknowledge many people who have accompanied me during this 3-year journey.

My deepest thanks are to my supervisors, Romain Brossier, Ludovic Métivier, and Jean Virieux, for their great support both technically and personally. I had all the freedom to grow up, but always with careful orientations. Standing on their shoulders, I have never been afraid of any difficulty. After 3 years, I could feel a “new me”, technically stronger but also better personality.

I would like to thank Total for supporting my entire study: engineering, master and PhD degrees. My deep sense of gratitude is to Philippe Montagnier, who gave me the first opportunity to work with Total despite my linguistic barrier. This first step then led to my other successful projects with Total, including this PhD. I always keep his orientation deep inside my heart.

I express my sincere gratitude to my coordinators and mentors in Total, Florent Bertini and Bertrand Duquet. Without their help, I could have not got access to real data and computer resources in Total. I would like to thank Mathieu Appé, François Audebert, Laurent Lemaistre and other colleagues for helping me to set up my work at Total.

I appreciate Uwe Albertin, René-Edouard Plessix, Anusha Sekar, other academical and industrial contacts for their great technical and career advice. I would like to thank Stéphane Operto for sharing his knowledge during video-conferences and conferences.

I appreciate very much Marwan Irnaka Theodosius, Philippe Le Bouteiller, Julien Thurin, Chi-Cong Vu and Wei Zhou for being my great friends and office mates. Many thanks to great collaborators, Weiguang He, Pengliang Yang, and Paul Wellington, for fruitful ideas and discussions.

Special thanks to Sandrine Nadau, the superwomen of SEISCOPE, who helped me to sort out all my complex missions without any hesitation. I thank other members of SEISCOPE consortium, people and friends at ISTerre for their friendship and for great moments and happy time at Grenoble.

I thank my family for their consistent support. My parents never went to higher educations, never left Vietnam but their strong love accompanies me through all the continents that I have been. The source of my energy is my little brother, who never asks for anything but chocolates and nice foods from France. Each time when I feel tired, thinking about him gives me the power to continue the long roads ahead.

Many thanks to my dear friend, Hugo Scherrier, for supporting my variable emotion willingly. Great thanks to Layla and Andre Scherrier for giving me a second home to come back.

In the end, please accept my apologies if I forget someone.

Résumé

L'imagerie sismique du sous-sol à partir de données terrestres est très difficile à effectuer due à la complexité 3D de la proche surface. Dans cette zone, les ondes sismiques sous forme d'un paquet compact de phases souvent imbriquées sont dominées par des effets élastiques et viscoélastiques, couplés aux effets dus à la surface libre qui génèrent des ondes de surface de grande amplitude et dispersives. L'interaction des ondes sismiques avec une topographie plus ou moins complexe dans un contexte de fortes hétérogénéités de la proche surface induit d'importantes conversions des ondes avec de fortes dispersions d'énergie. Il est donc nécessaire de prendre en compte à la fois une représentation tridimensionnelle précise de la topographie et une physique correcte qui rend compte de la propagation du champ d'onde dans le sous-sol au niveau de précision réclamé par l'imagerie sismique. Dans ce manuscrit, nous présentons une stratégie d'inversion des formes d'onde complètes (FWI en anglais) efficace, autonome et donc flexible, pour la construction de modèles de vitesse à partir de données sismiques terrestres, plus particulièrement dans les environnements dits de chevauchements d'arrière pays (foothills en anglais) aux variations de vitesse importantes.

Nous proposons une formulation efficace de cette problématique basée sur une méthode d'éléments spectraux en domaine temporel sur une grille cartésienne déformée, dans laquelle les variations de topographie sont représentées par une description détaillée de sa géométrie via une interpolation d'ordre élevé. La propagation du champ d'onde est caractérisée par une élasticité linéaire anisotrope et par une atténuation isotrope du milieu: cette deuxième approximation semble suffisante pour l'imagerie crustale considérée dans ce travail. L'implémentation numérique du problème direct inclut des produits matrice-vecteur efficaces pour résoudre des équations élastodynamiques composant un système différentiel hyperbolique du second ordre, pour les géométries tridimensionnelles rencontrées dans l'exploration sismique.

Les expressions explicites des gradients de la fonction écart entre les données et les prédictions sont fournies et inclut les contributions de la densité, des paramètres élastiques et des coefficients d'atténuation. Ces expressions réclament le champ incident venant de la source au même temps de propagation que le champ adjoint. Pour ce faire, lors du calcul du champ adjoint à partir de l'instant final, le champ incident est recalculé au vol à partir de son état final, de conditions aux bords préalablement sauvegardées et de certains états intermédiaires sans stockage sur disques durs. Le gradient est donc estimé à partir de quantités sauvegardées en mémoire vive. Deux niveaux de parallélisme sont implémentés, l'un sur les sources et l'autre sur la décomposition du domaine pour chaque source, ce qui est nécessaire pour aborder des configurations tridimensionnelles réalistes. Le préconditionnement de ce gradient est réalisé par un filtre dit de Bessel, utilisant une implémentation différentielle efficace fondée sur la même discrétisation de l'espace du problème direct et formulée par une approche d'éléments spectraux composant un système linéaire symétrique résolu par une technique itérative de gradient conjugué. De plus, une contrainte non-linéaire sur le rapport des vitesses de compression et de cisaillement est introduite dans le processus d'optimisation sans coût supplémentaire: cette introduction

s'avère nécessaire pour traiter les données en présence de faibles valeurs de vitesse proche de la surface libre.

L'inversion élastique multi-paramètres en contexte de chevauchement est illustrée à travers des exemples de données synthétiques dans un premier temps, ce qui met en évidence les difficultés d'une telle reconstruction. Des données synthétiques simulées et extraites du réseau d'acquisition en surface du modèle de chevauchement de la deuxième phase (phase II) du projet SEAM (SEG Advanced Modeling) avec quatre lignes de vingt sources permettent une illumination tridimensionnelle complète de la sub-surface. L'inversion est plus sensible aux ondes de cisaillement car les données sont majoritairement composées d'ondes de surface. Nous proposons une stratégie de fenêtrage des données en deux étapes, en se focalisant d'abord sur les ondes de volume en supprimant brutalement les ondes de surface, puis sur l'ensemble des données incluant les ondes de surface. L'utilisation de cette hiérarchie dans la présentation des données conjointement avec la mise en œuvre du préconditionnement de Bessel, fondé sur la structure géométrique de l'image attendue, rend la reconstruction des modèles de vitesse des ondes P et des ondes S possible. Cette stratégie de fenêtrage temporel est de plus couplée avec une hiérarchie de contenu fréquentiel allant des basses fréquences aux hautes fréquences. Elle est appliquée avec succès à l'étude dans une configuration tridimensionnelle d'une ligne d'acquisition extraite des données originales de la phase SEAM II sur les chevauchements. Avec l'éclairage limité d'une acquisition ligne, la contrainte du modèle sur le rapport des vitesses d'ondes P et S joue un rôle important pour atténuer le caractère mal posé de l'inversion multi-paramètres. En considérant également les ondes de surface, nous pouvons exploiter au maximum les informations des données observées afin d'obtenir des modèles de vitesse précis, aussi bien pour la proche surface que pour des profondeurs plus importantes.

Cette méthodologie de l'inversion des formes d'onde complètes développée au cours de ce travail est appliquée à un jeu de données réelles en contexte de chevauchement. Sa mise en œuvre est délicate, principalement en raison d'une acquisition comprenant peu de sources, de la présence d'un bruit important, et d'une structure du sous-sol très complexe. On peut s'attendre à ce que des données supplémentaires telles que des diagraphies puissent aider l'inversion à mieux contraindre le modèle pour expliquer les données observées. Les résultats préliminaires, en ne considérant que les ondes de volume, améliorent la cinématique et sont cohérents avec les interprétations géologiques préalablement effectuées sur cette cible. Des contrôles sur la qualité du modèle via une analyse de l'ajustement des données permettent de mieux cerner les zones fiables des modèles reconstruits.

Abstract

Seismic imaging of onshore targets is very challenging due to the 3D complex near-surface-related effects. In such areas, the seismic wavefield is dominated by elastic and visco-elastic effects such as highly energetic and dispersive surface waves. The interaction of elastic waves with the rough topography and shallow heterogeneities leads to significant converted and scattering energies, implying that both accurate 3D geometry representation and correct physics of the wave propagation are required for a reliable structural imaging. In this manuscript, we present an efficient and flexible full waveform inversion (FWI) strategy for velocity model building in land, specifically in foothill areas.

We propose an efficient formulation based on a time-domain spectral element method (SEM) on a flexible Cartesian-based mesh, in which the topography variation is represented by an accurate high-order geometry interpolation. The wave propagation is described by the anisotropic elasticity and isotropic attenuation physics. The numerical implementation of the forward problem includes efficient matrix-vector products for solving second-order elastodynamic equations, even for completely deformed 3D geometries. Complete misfit gradient expressions, including attenuation contribution spread into density, elastic parameters and attenuation factors, are given in a consistent way. Combined adjoint and incident fields recomputation, from the final state and previously saved boundary values, allows the estimation of gradients with no I/O effort. A two-level parallelism is implemented over sources and domain decomposition, which is necessary for 3D realistic configurations. The gradient preconditioning is performed by a so-called Bessel filter using an efficient differential implementation based on the SEM discretization on the forward mesh. A non-linear model constraint on the ratio of compressional and shear velocities is introduced into the optimization process at no extra cost.

The challenges of the elastic multi-parameter FWI in complex land areas are highlighted through synthetic and real data applications. A 3D synthetic illustration is considered on a subset of the SEAM phase II Foothills model with 4 lines of 20 sources, providing a complete 3D illumination. As the data is dominated by surface waves, it is mainly sensitive to the S-wave velocity. We propose a two-step data-windowing strategy, focusing on the early body waves before considering the entire wavefield, including surface waves. The use of this data hierarchy together with the structurally-based Bessel preconditioning make possible to reconstruct accurately both P- and S-wavespeeds. The designed inversion strategy is combined with a low-to-high frequency hierarchy, successfully applied to a pseudo-2D dip-line survey of the SEAM II Foothill dataset. Under the limited illumination of a 2D acquisition, the model constraint on the ratio of P- and S-wavespeeds plays an important role to mitigate the ill-posedness of the multi-parameter inversion process. By also considering surface waves, we manage to exploit the maximum amount of information in the observed data to get a reliable model parameters estimation, both in the near-surface and in deeper parts.

The developed FWI algorithm and workflow are finally applied on a real foothill dataset. The application is challenging due to a sparse acquisition design, the presence of high levels of noise, and

complex underneath structures. Additional prior information such as the logs data is considered to assist the FWI design. The preliminary results, only relying on body waves, are shown to improve the kinematic fit and follow the expected geological interpretation. Model quality controls through data-fit analysis and uncertainty studies help to identify artifacts in the inverted models.

Contents

General Introduction	15
I Theory and Algorithms	21
1 Introduction	25
1.1 FWI in brief	25
1.2 Challenges of FWI for foothill targets	27
1.3 Spectral element method: Why?	28
1.4 Summary	30
2 3D time-domain elastic and visco-elastic modeling in SEM	33
2.1 Elastodynamics equations	34
2.1.1 Elastic modeling	35
2.1.2 Visco-elastic modeling	35
2.2 Weak-form development of the visco-elastic wave equation	38
2.2.1 Key elements in spectral element method	39
2.2.2 Some common operations related to the basis functions	41
2.2.3 From the integral form to the discrete form of the wave equation	43
2.2.4 Global system of the weak formulation	45
2.3 Optimized and low-memory visco-elastic modeling kernel	46
2.3.1 Optimized modeling kernel	47
2.3.2 Low-memory visco-elastic modeling kernel	48
2.4 Boundary conditions and source implementation	49
2.4.1 Boundary conditions	49
2.4.2 Source implementation	51
2.5 Mesh design	52
2.5.1 Cartesian-based mesh with variable element-size	53
2.5.2 Complex topography representation	54
2.6 Comparison with semi-analytical solutions - LOH benchmarks	56
2.6.1 LOH1 comparison	57
2.6.2 LOH3 comparison	58
2.7 Conclusion	59
2.A Stress and strain relationship in the time domain	61
2.B Complex P and S wave velocities in AXITRA	62
3 Inverse problem	63

CONTENTS

3.1	Gradient computation by the adjoint-state approach	64
3.1.1	Lagrangian definition & Adjoint system	64
3.1.2	Gradient expressions	67
3.1.3	Influence of the attenuation on the velocity gradient estimation	69
3.2	Solving the constrained local minimization problem	70
3.2.1	Step length computation	71
3.2.2	Descent direction computation	71
3.2.3	Preconditioning for optimization process	73
3.2.4	Non-linear model constraints on the relationship between different parameters .	74
3.3	Source estimation	76
3.4	Conclusion	77
3.A	Gradients estimation	78
4	HPC implementation in SEM46 code	81
4.1	Two-level parallelization	81
4.1.1	Scalability	82
4.1.2	Modeling efficiency compared with other modeling tools	83
4.2	Balancing memory requirement & computational cost for gradient building	84
4.3	Conclusion	86
5	Structure-oriented Bessel gradient preconditioning	87
5.1	Introduction	89
5.2	Methodology	90
5.2.1	Definition of Bessel filters and their sparse inverse operators	91
5.2.2	Approximation of the Laplace filter by the Bessel operator	92
5.3	Weak formulation in finite element methods	93
5.3.1	Weak formulation of the filter for homogeneous coherent lengths without dip and azimuth	94
5.3.2	Variable coherent lengths, dip and azimuth angles	96
5.4	Numerical illustrations	99
5.4.1	Spike test	99
5.4.2	Random noise tests	101
5.4.3	FWI gradient smoothing	105
5.5	Numerical implementation in a parallel spectral-element method scheme	106
5.5.1	Linear system construction and matrix-vector product evaluation	107
5.5.2	Parallel conjugate gradient iterative solver	108
5.6	Conclusion and discussion	112
5.7	Additional numerical investigations	113
5.7.1	Gradient smoothing in pseudo-3D Marmousi model	114
5.7.2	Influence of the preconditioning operator on inversion results	115
5.A	Modified Bessel functions and their integral representation	117
5.B	Definition of Bessel filters and their normalized factors	117
5.C	Weak formulation development	120
5.D	Matrix-free finite-difference implementation scheme	123

II Applications	125
6 Synthetic examples	129
6.1 Elastic 3D configuration	130
6.1.1 Model description & 3D acquisition design	130
6.1.2 Inversion setup	131
6.1.3 Inversion results	133
6.1.4 Discussions	135
6.2 SEAM II Foothill dataset	139
6.2.1 SEAM II Foothill Dip-line configuration	139
6.2.2 Inversion setup	140
6.2.3 Inversion results & Discussions	144
6.2.4 Perspectives	146
6.3 Conclusion	147
7 Real data application	149
7.1 Background information	149
7.1.1 Exploration context	149
7.1.2 Seismic data	151
7.1.3 Acquisition uncertainties	152
7.2 Inversion setup	153
7.2.1 Models building	155
7.2.2 Source estimation	158
7.2.3 Preconditioning and model constraint design	159
7.2.4 FWI workflow	161
7.3 Inversion results & Discussions	161
7.3.1 Model quality control through data fitting	162
7.3.2 Influence of acquisition mis-positioning on inversion results	163
7.4 Conclusion & Perspectives	165
Conclusion & Perspectives	171
References	176

General Introduction

Context

With the growing need for energy resources and raw materials, imaging the Earth's crust has been a topic of enormous interest over the last century. However, in practice, the observation of the geological structure below the surface is difficult, as it typically involves well drilling, cores analysis and petrophysical studies. For example, drilling costs alone can range from several to hundreds million dollars, depending on the geological condition and the investigation depth, leading to rather sparse spatial well sampling (Kaiser, 2007). Consequently, direct knowledge about the subsurface structures remains very limited. Up to now, subsurface images mainly rely on indirect interpretations from physical measurements at the free surface, shallow depth and through remote sensing techniques.

In the case of the oil & gas industry, seismic and non-seismic geophysical imaging techniques are already intensely used to guide the search for new resources or monitor existing reservoirs. Non-seismic methods, for instance, the gravity and magnetic measurements, can be used to determine the geometry and depth of particular sedimentary basins, where oil and gas are typically located. Due to their comparatively low-resolution, they serve generally as complementary information for seismic methods, which are the most widespread. Seismic surveying uses mechanical waves to map the underground rock structures over several kilometers depth. Depending on the acquisition design, the method can provide a 2D or 3D high-definition picture of the subsurface geology.

The evolution of this market will continue to require more and more from seismic imaging in the near future. Hydrocarbon fuels are expected to remain as a substantial part of the global energy mix (Abas Kalair et al., 2015). However, in the past few years, the global conventional hydrocarbon discovery continuously decreased, according to the study published by Rystad Energy (Figure 1)¹. The reserves are getting smaller, under more complex geological structures such as deep offshore, sub-salt, heavily faulted or thrust zones, which requires a breakthrough in exploration technology². Moreover, due to the decline of the oil price, it is vital to maximizing the recovery from existing reserves. Under these conditions, clear images of the subsurface play an important role to de-risk the structure interpretation, especially in challenging locations.

Besides the conventional sources, we are seeing an spectacular growth in the production of unconventional fuels, notably shale gas, tight oil and oil sands, with enormous estimated resources in the US and Canada (Erbach, 2014). Unlike conventional gas, which resides in porous and permeable reservoirs and can be extracted by standard vertical wells, shale gas remains trapped in its original source rock³. Their extraction requires the horizontal drilling and fracking along the gas-containing rock. The exploration for unconventional oil and gas accelerates the demand for more information to find the sweet spots to drill and fracture⁴. In this case, understanding the rock and natural fracture properties in the reservoir is the game changer. This is the area where high resolution, broadband and wide-azimuth 3D multicomponent seismic and advanced imaging techniques can play a key role in providing information about the rock types, anisotropy, and fracture orientations. Finally, the development of seismic imaging is also increasing in other sectors which require a good understanding of the subsurface structures, for instance, mineral exploration industry (Malehmir et al., 2017).

In summary, we believe that seismic imaging, especially toward high-resolution medium charac-

¹All-time low for discovered resources in 2017: Around 7 billion barrels of oil equivalent was discovered. Rystad Energy UCube and Rystad Energy research and analysis, December 21, 2017.

²Geophysical Imaging brochure, Shell Global, 2014.

³Shale Gas and Other Unconventional Sources of Natural Gas. Union of Concerned scientists, April 3, 2015.

⁴Proving the Role of Seismic in Unconventional Plays. ION Geophysical, Official show daily publisher of the 75th EAGE Conference and Exhibition, E&P DailyNews, June, 2013.

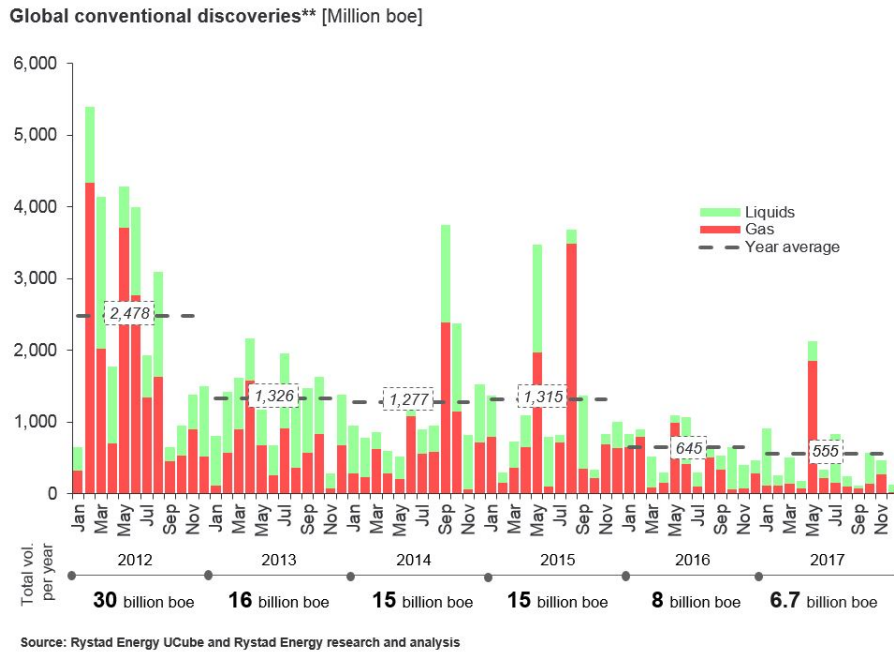


Figure 1: Global conventional discoveries. Source: Rystad Energy UCube and Rystad Energy research and analysis.

terization, remains a key indicator, not only for future hydrocarbon exploration but also for reservoir monitoring, in both conventional and unconventional resources. With that motivation, this manuscript focuses on the use of seismic waves in complex onshore targets, especially in foothill environments.

Exploration seismic acquisition

Data acquisition is the first step in any geophysical method. Acquisition design and the quality of the recorded signal are crucial for the structural image of the subsurface, directly used for prospect mapping. The design of the surface acquisition or vertical seismic profile is defined before the deployment, mainly based on the geological setting of the target, acquisition condition in the field and available infrastructures.

Marine seismic acquisition is generally accomplished using one or multiple air-gun arrays for sources. Air-guns are deployed behind the seismic vessel and generate a seismic signal by releasing highly pressured air into the water. In standard acquisitions, receivers are towed behind the ship in long streamers at several kilometers length: for instance, a 3D streamer project in offshore Gabon considers from 6 to 12 km streamers to cover a deep sub-salt target (Andrew et al., 2013). Marine receivers are mainly hydrophones, which respond to changes in water pressure. More advanced technologies, such as the dual sensor (hydrophone-vertical) and multicomponent (streamers hydrophones, and particle acceleration in x , y , and z directions) streamers are shown to be helpful in deghosting in the receiver side, thus providing broadband temporal seismic data (Carlson et al., 2007; Robertsson et al., 2008; Moldoveanu et al., 2012). Due to the water column, which acts as a natural filter of shear movement, the streamers can only record compressional wavefields. In complex structure areas, for instance, salt bodies, more advanced technologies can be considered to enhance the image quality such

as wide-azimuth, broadband and coil shooting. These techniques aim at gaining precise knowledge of specific targets, in particular at the top of the reservoir: For example, the Heidrun coil shooting survey in the Norwegian Sea with a radius of 5.6km, and 10 streamers, each 4.5 km long, separated by 75 m, provides a clearer image of the fault definition in the reservoir⁵.

Another advanced technique of marine seismic is the ocean bottom setting, which is target-oriented with a fixed-spread acquisition. The receivers can be cables (ocean bottom cables) or independent nodes (ocean bottom nodes), laying out on the sea floor. They can have up to three component geophones and an additional hydrophone. The use of ocean bottom recording has several advantages over conventional towed streamer technology such as a more flexible acquisition design, ranging from narrow to wide azimuth. It also gives the possibility of direct measurement of S-wave data in addition to P-wave data through the geophones components. Most of the systems offered today are used for reservoir monitoring, with shallow to medium water depth, limited at around 3000 m (Hovland, 2016).

Unlike marine data, land acquisition geometry is more flexible, and the shot points and the receivers can be re-arranged in many ways. Acquisition flexibility is further enhanced by the development of wireless devices, for instance, the RT3 system provided by WirelessSeismic⁶. The receivers can be single or multiple components geophones, which record both P- and S-wavefields. In 2D land acquisition, the shot points are gradually moved along a line of geophones. For 3D acquisition, an orthogonal design is normally considered, with receiver lines laid out perpendicular to the source lines, leading to a natural wide azimuth aperture. Explosive and vibroseis sources are commonly used in exploration scale acquisitions. In explosive source acquisition, a charge of dynamite is placed at several meters below ground, in a drilled hole: For example, the 3D acquisition in Jungar Basin, northwest China, uses dynamite packages from 16 to 20 kg (Lv et al., 2015). The explosion of the dynamite charge provides a broadband impulsive source signature. As opposed to the impulsive response of dynamite source, vibroseis is a truck-mounted system that uses a large oscillating mass to generate a limited band energy signals over an extended period. The recorded data then must be correlated to convert the extended source signal into an impulse. In both ocean bottom and land acquisitions, the wavefield recorded through geophones placed at the free-surface or sea-floor contains significant elastic effects such as Rayleigh waves and converted energies from P to S and vice-versa.

Foothill acquisition

Foothill conditions are generally associated with a compressive tectonic regime and significant topography variation. The underneath structures normally contain thrust faults, multi-folds, and steeply dipping formations. The lithology varies quickly in both lateral and vertical directions. The alternative of fast and slow velocity layers in depth is also found in these over-thrust areas (Ravaut et al., 2004).

Figure 2 provides an example of the extreme acquisition condition in Llanos Foothills with steep topography and surface water springs (Checa et al., 2009). Such hard-to-access areas are inaccessible by vibroseis trucks. Only explosive sources can thus be used. However, heavy dynamite charge is abandoned inside the environmental reserve because of the environmental constraint, leading to various data quality and rather weak signal-to-noise ratio. Seismic data is rarely shot in a straight line, under rather sparse sources and receivers distribution due to the presence of natural obstructions. Issues related to the positioning of receivers or shots during acquisition also affect the later processing and imaging steps. Besides the technical issues, human safety also requires special consideration and investigation, through intensive training for unskilled workers who had little or no previous exposure to such difficult acquisitions.

⁵Shooting Seismic in Linked Circles. Lasse Amundsen and Martin Landrø, GEOExPro magazine, Vol. 6, No. 1, 2009.

⁶RT3 system, WirelessSeismic.

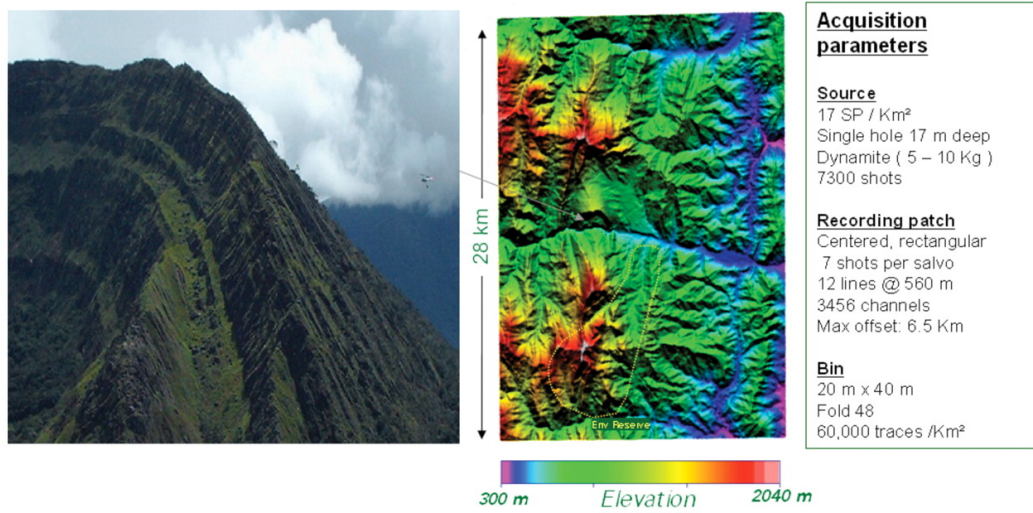


Figure 2: Example of acquisition parameters and terrain condition in Llanos foothills in Colombia, extracted from Checa et al. (2009).

More advanced acquisition trends consider the wireless receivers deployment by drones, for instance the METIS project developed by TOTAL⁷. The application of this technique in foothill condition could possibly improve the seismic traces density, and allow a more flexible acquisition geometry with larger data coverage.

Depth imaging

Detailed images of the subsurfaces need to be constructed from the recorded wavefield so that the structural interpretation could be made feasible. For exploration scale target, the structural image is generally produced in two steps: macro-models construction and data migration. The macro-models are some representative physical parameters of the sub-surface, for instance, compressional and shear wave speeds, density, anisotropy, and attenuation. Based on the low-wavenumber macro-model properties, the migration techniques construct the high-wavenumber reflectivity image of the sub-surface (Yilmaz, 2001). The migrated images then allow localizing medium interfaces, i.e. reflectors, for structural interpretation purpose.

One of the early migration techniques is based on solving the acoustic one-wave equation from source and receiver positions. The migration image is built from the coincident-time principle, so-called the imaging condition: when the source and receiver wavefields meet inside the earth (Claerbout, 1970, 1971). In a different formulation, Schneider (1978) considers summing the data along the diffraction curve to produce a focused image at each spatial location, namely Kirchhoff migration. These methods rely on the single-scattering assumption and assume rather smooth macro-velocity models, generally built from stacking-velocity estimation. They often only consider compressional waves and could efficiently produce good images for simple geology with limited heterogeneity (Gray, 2016).

As computing power was increasing, more sophisticated forms of wave-equation migration have

⁷METIS Research & Development project, an Integrated Geophysical Acquisition System for Quality, Real-Time Imaging of Complex, Hard-to-Access Onshore Areas, Florent Bertini, TOTAL Exploration & Development.

been developed such as the reverse-time migration techniques which consider the full two-way acoustic wave equation (Whitmore, 1983; McMechan, 1983), and the least-square techniques to handle variations in illumination (Nemeth et al., 1999; Valenciano et al., 2009). The 2D and 3D pre-stack migration both in time and depth becomes more popular. These advanced techniques provide better-migrated image quality, however, they also require better accuracy and precision in the velocity model: for instance depth migration is very sensitive to velocity errors but the technique is critical in complex targets such as flattening the below-salt horizons (Etgen et al., 2009; Jones, 2015; Gray, 2016). In more complex setting such as sub-salt, thrust and faulting structures, the construction of macro-model is challenging, leading to a rather noisy migration image, eventually uncertain structural interpretation.

Based on the picked travel-time or reflectors from seismic data, seismic tomography techniques aim at constructing velocity models which could match or flatten the picked events (Bishop et al., 1985; Stork, 1992; Lambaré et al., 2004). The resolving power of conventional ray-based tomography techniques is limited at several times the wavelengths of the recorded waves (Jones, 2015). Their capacity is also limited at shadow zones when the rays cannot be traced. Due to their numerical efficiency, ray-based methods still remain very popular. Full waveform inversion (FWI) or waveform tomography has the potential to generate more accurate velocity models by considering not only the arrival times but also amplitude and waveform (Lailly, 1983; Tarantola, 1987; Virieux and Operto, 2009; Virieux et al., 2017). The method iteratively improves the current velocity model by matching the predicted with the recorded waveform. It can deliver velocity structure at a scale length similar to the wavelengths of the seismic data (Devaney, 1984; Jones, 2015).

Land imaging

Seismic imaging in land is generally more complicated than marine data due to many factors which can distort the wavefield. The near-surface migration image is very poor, due to the fact that the near-surface heterogeneity is generally below the data sampling at migration frequencies. The elastic effect is pronounced in the land data, which raises the question about elastic imaging tools instead of the current acoustic approximation.

Onshore targets normally contain various near-surface issues such as topography variation, near-surface low-velocity thickness, and weathering velocity. Therefore, static corrections at both source and receiver positions are required, before any further migration techniques could be applied. The objective is to determine the reflection arrival times from a flat datum without any weathering or low-velocity materials, in order to “shift” all recorded traces into the same defined datum. One of the most common methods for these corrections is an uphole-based measurement of vertical travel times from a buried source (Cox, 1999).

The near-offset reflection data, required for migration sequences, is hidden by highly energetic surface waves. It is difficult to suppress surface waves without knowing the correct velocity variation. One current interesting direction is to build the near-surface model from surface waves, which can later be used for surface wave attenuation or static correction (Socco et al., 2010).

Accurate velocity estimation is thus critical for accurate structural mapping in onshore targets. Note that, the described challenges also occur in ocean bottom data, in which the elastic and topography effects are pronounced.

Foothill challenges

On top of the onshore imaging challenges, foothill targets are particularly difficult due to brutal topography variation, as shown in Figure 2, and geological subsurface complexity. The recorded data has a low signal-to-noise ratio due to acquisition condition, and complex underneath structures. Hard

rocks at the surface also result in poor energy transmission into the ground. Moreover, the interaction with the rough surface leads to significant converted and scattered energies, also broken and highly dispersive surface waves, making the removal of these “coherent noises” even more challenging.

Careful geometry quality control, complex statics correction, and data processing sequences need to be considered before feeding any migration tools. Velocity analysis in these areas is difficult and highly uncertain due to the complex geology and low signal-to-noise ratios (Checa et al., 2009). All of these obstacles make the migrated image very noisy.

In these areas, anomalously low-velocity layers in depth, coming from the over-thrust activities, and brutal lateral lithology variation warp ray-paths of traditional ray-based tomography methods (Ravaut et al., 2004). Also, the elastic and viscoelastic effect in the wavefield requires the beyond-acoustic-assumption methods, in which FWI is an appealing candidate for velocity model building, both at the near-surface and deeper parts.

Motivation & PhD outline

As we have shown, the demand for accurate structural images is significantly increasing in modern exploration market. In complex geology, for instance, foothill environments, this task requires non-negligible efforts in acquisition, processing and imaging sequences. In these areas, the estimated velocity models by conventional techniques are often incorrect, which eventually degrades the quality of the produced images and therefore affect the exploration decisions (Plessix et al., 2012). The velocity estimation process is generally complex, time-consuming and frequent bottleneck in seismic imaging projects. This manuscript aims at establishing an FWI strategy for velocity model building in land, specifically in foothill areas, where the conventional techniques often fail. Also, the surface waves can be considered to improve the near-surface velocity model estimation. An accurate near-surface model will be very useful in any static correction techniques or surface waves removal, for instance through adaptive subtraction (Cary and Zhang, 2009), to provide a cleaner dataset for later migration steps.

This manuscript involves two main objectives: building an accurate and efficient wave simulation tool in complex geometry and designing flexible workflow for medium parameters reconstruction. The thesis is thus divided into two main parts.

In the **first part**, the challenges associated with FWI problem in foothill targets are identified in Chapter (1), which highlights the importance of accurate physical and 3D geometry representation. This leads to the choice of the selected numerical method: spectral element approach. The theoretical development and numerical implementation of forward and inversion problems, associated with the selected numerical discretization, are provided in Chapters (2) and (3). Exploration scale applications also require optimized parallelization implementation as described in Chapter (4). The use of various preconditioning and model constraint strategies to stabilize the ill-posed inversion process, also to mitigate the cross-talk between multiple estimated parameters are addressed in Chapters (3) and (5).

In the **second part**, the presented FWI algorithm is applied to several synthetic benchmarks in Chapter (6) to illustrate the importance of each element that we develop. These applications highlight the challenges of the elastic multi-parameter FWI in complex land areas: Besides the use of a sophisticated numerical tool, an appropriate workflow is mandatory. The developed workflow in synthetic data is then applied on a real foothill dataset in Chapter (7) with realistic challenges as sparse and uncertain acquisition design, noisy recording and complex underneath structures.

Conclusion and perspectives are provided in the last section of this manuscript.

Part I

Theory and Algorithms

This part aims at developing an efficient, flexible, and accurate FWI methodology, dedicated to exploration-scale challenging onshore targets. The chapters included in this part are based on the following extended abstracts and papers

-
- *Trinh, P. T., Brossier, R., Métivier, L., Tavard, L., and Virieux, J. (2017a). Efficient 3D elastic FWI using a spectral-element method. In 87th SEG Conference and Exhibition 2017, Houston, pages 1533–1538*
- *Trinh, P. T., Brossier, R., Métivier, L., Virieux, J., and Wellington, P. (2017b). Bessel smoothing filter for spectral element mesh. Geophysical Journal International, 209(3):1489–1512*
- *Trinh, P. T., Brossier, R., Métivier, L., Virieux, J., and Wellington, P. (2017c). Structure-smoothing Bessel filter for finite element mesh: Application on 3D elastic FWI. In 79th EAGE Conference and Exhibition 2017, Paris*

The numerical development presented in this part is associated with the SEM46 code development: Spectral Element Method For Seismic Imaging at eXploration scale

- *Brossier, R. and Trinh, P. T. (2017). SEM46 Manual Version 2.1. SEISCOPE Consortium*

Chapter 1

Introduction

Contents

1.1	FWI in brief	25
1.2	Challenges of FWI for foothill targets	27
1.3	Spectral element method: Why?	28
1.4	Summary	30

In oil & gas industry, FWI technique appears as a powerful technique to extract high-resolution quantitative multiparameter models of the subsurface, down to half of the local wavelength (Lailly, 1983; Tarantola, 1984; Virieux and Operto, 2009; Virieux et al., 2017). Since the pioneering 2D FWI applications with surface seismic data (Ravaut et al., 2004; Operto et al., 2006; Brenders and Pratt, 2007), most of FWI applications at the crustal scale have been performed under the acoustic approximation, generally for marine environments (Plessix and Perkins, 2010; Sirgue et al., 2010; Warner et al., 2013; Vigh et al., 2013; Operto et al., 2015). In this PhD work, we are interested in applying the FWI technique onto challenging onshore targets, such as foothill environments.

1.1 FWI in brief

In majority of the Earth imaging applications, the physics of wave propagation can be described through partial differential equations (PDEs). The wave equation in time or frequency domain can thus be generally written as

$$F(\mathbf{m}, \mathcal{W}(\mathbf{m})) = \mathbf{S}, \quad (1.1)$$

which implies that the seismic wavefield \mathcal{W} is calculated by a forward propagator F on the model parameters \mathbf{m} , with the source term \mathbf{S} . Assuming that our numerical propagator could honor the physics of the wave propagation, the model parameters \mathbf{m} , related to the physical and mechanical parameters of the sub-surface, can be adjusted through comparisons between the recorded and the numerically simulated data. Under the assumption that a good data prediction implies a good model estimation, FWI technique improves the model parameters estimation by minimizing the misfit between the observed data \mathbf{d}_{obs} and the calculated data $\mathbf{d}_{cal}(\mathbf{m})$ at receiver positions. The calculated data is the computed wavefield \mathcal{W} at receivers location

$$\mathbf{d}_{cal}(\mathbf{m}) = R\mathcal{W}(\mathbf{m}), \quad (1.2)$$

where R is the extraction operator. An often used misfit function is the least-squares distance

$$\chi(\mathbf{m}) = \frac{1}{2} \|\mathbf{d}_{cal}(\mathbf{m}) - \mathbf{d}_{obs}\|^2. \quad (1.3)$$

The formulation implies a summation over all sources and receivers and an integral over the recording time or the considered frequency band. The minimization of the misfit function (1.3) can be addressed by two main approaches:

- Global and semi-global methods, see for instance, Metropolis and Ulam (1949); Kirkpatrick et al. (1983); Sen and Stoffa (1995): Without any significant prior knowledge about the medium, these methods aim at searching for the global minimum of the misfit function by relatively dense sampling of the model space. Therefore, a large number of forward simulations are generally required, which limits the application of these approaches to problems with a small number of degrees of freedom (10 to 1000) and/or with a cheap forward simulation.
- Local optimization methods (Nocedal and Wright, 2006): For non-convex misfit functions, these approaches require a relatively good initial model. If the starting point stays within the valley of the global minimum, these methods make possible to iteratively converge towards the global minimum. Compared with the global or semi-global approaches, the local optimization strategies could consider a larger number of degrees of freedom, even with a complex forward propagator.

Due to a large number of degrees of freedom and fairly expensive forward simulation, FWI mostly relies on local optimization strategies.

The iterative minimization process can be formulated following the perturbation theory. At each iteration k , we start from the model \mathbf{m}^k and the misfit function $\chi(\mathbf{m}^k)$. Within a vicinity of the starting point, we look for an updated model \mathbf{m}^{k+1} which decreases the value of the misfit function. The updated model can be written as the sum of the initial model and a perturbation $\Delta\mathbf{m}^k$ as

$$\mathbf{m}^{k+1} = \mathbf{m}^k + \Delta\mathbf{m}^k \quad (1.4)$$

We assume that the misfit function is twice continuously differentiable, i.e. the first-order and second-order derivatives of the misfit function exist and are themselves finite and continuous functions. Under the assumption of small perturbation model $\Delta\mathbf{m}$, the Taylor expansion of the first-order derivatives of the misfit function can be written as

$$\begin{aligned} \nabla\chi(\mathbf{m}^{k+1}) &= \nabla\chi(\mathbf{m}^k + \Delta\mathbf{m}^k) \\ &= \nabla\chi(\mathbf{m}^k) + \underbrace{\nabla^2\chi|_{\mathbf{m}^k}}_{\text{Hessian}} \Delta\mathbf{m}^k + \mathcal{O}(\|\Delta\mathbf{m}^k\|^2), \end{aligned} \quad (1.5)$$

in which the notation “ $\bullet|_{\mathbf{m}^k}$ ” implies the value of a function estimated at the parameter \mathbf{m}^k and the infinitesimal asymptotic is denoted by “ $\mathcal{O}(\bullet)$ ”.

Under the linearization of the inverse problem, the misfit function can be locally approximated by a quadratic function. The minimum of this quadratic approximation is reached when the first-order derivative of the misfit function vanishes, i.e. $\nabla\chi(\mathbf{m}^{k+1}) = 0$, giving the perturbation model

$$\Delta\mathbf{m}^k = -[\nabla^2\chi|_{\mathbf{m}^k}]^{-1} \nabla\chi(\mathbf{m}^k). \quad (1.6)$$

If the misfit function is quadratic, Equation (1.6) provides the global minimum of the misfit function after one iteration, which is equivalent to a linear inverse problem. However, the FWI problem is

non-linear, leading to an iterative minimization process of the misfit function, where each iteration is approximated by a linear problem. In Equation (1.6), the model perturbation is the product of the inverse of the Hessian matrix $[\nabla^2\chi|_{\mathbf{m}^k}]^{-1}$ with the gradient vector of the misfit function $\nabla\chi(\mathbf{m}^k)$. In practical applications, the size of the Hessian matrix is extremely big with $n_{mod} \times n_{mod}$ elements, where n_{mod} is the size of the model vector \mathbf{m} . The Hessian matrix cannot thus be explicitly computed or stored, implying that the perturbation model cannot be estimated directly from Equation (1.6). Section (3.2) will provide an overview about various local optimization approaches which are used to estimate the model perturbation $\Delta\mathbf{m}^k$ and the effect of the Hessian operator. They are all approximated versions of Equation (1.6).

1.2 Challenges of FWI for foothill targets

Standard FWI schemes require an accurate and efficient forward propagator for wavefield simulation. When considering on-shore targets, especially foothill environments, free-surface-related effects can be quite important in the wavefield. For instance, significant elastic and visco-elastic effects can be observed in the presence of weathering and unconsolidated near-surface, complex topographies, strong geological interfaces (shale-carbonate for example) (Aki and Richards, 2002; Yilmaz, 2016). An example of a real foothill seismic dataset is shown in Figure 1.1, in which the wavefield is dominated by strong amplitude and highly dispersive surface waves (or Rayleigh waves). Significant converted energies can be detected in this dataset due to the interaction with the rough topography. These complex effects cannot be fully removed or compensated by data pre-processing, such as attenuation compensation or ground-roll removal, implying that a correct description of the physics is strongly advisable for accurate model parameters estimation. Moreover, considering the complete physical phenomena of the wave propagation would allow taking benefit of each piece of recorded data, for expected more accurate results and higher imaging resolution.

Regarding anelastic media, attenuation affects both amplitudes and phases of the seismic wavefield. Its effect thus needs to be considered precisely during the modeling and imaging processes. In active seismic applications, the approximation of constant Q -value over a frequency band is widely accepted. Viscous effects can be easily considered in acoustic-based frequency-domain modeling and FWI, at least as passive parameters during inversion, thanks to the straightforward and free implementation through complex-valued velocity. Impressive 3D visco-acoustic illustrations have been shown in Operto et al. (2015); Amestoy et al. (2016), for velocity model building by a mono-parameter frequency-domain FWI based on sparse direct solver, or more recently with a joint reconstruction of the Q model (Operto and Miniussi, 2017).

When moving toward elastic and/or visco-elastic approximation, especially for 3D land data, frequency domain approaches with direct solvers are out of reach for actual computer hardware technology, as mentioned by Gosselin-Cliche and Giroux (2014). Alternative iterative solvers would require efficient and robust preconditioning for the modeling part (Li et al., 2014; Kostin et al., 2016). In addition, dispersive and complex wavefields would require considering a significant number of discrete frequencies, an additional argument reducing the attractability of frequency-domain approaches (Sirgue and Pratt, 2004; Brossier et al., 2009a). The time-domain formulation is thus the currently selected approach when performing FWI for elastic media at various scales (Tape et al., 2010; Peter et al., 2011; Fichtner et al., 2013; Vigh et al., 2014; Borisov et al., 2015; Zhu et al., 2015). The time-domain formulation also facilitates the application of time-windowing, wavefield separation, and various signal processing techniques, which favors strategies based on data-windowing hierarchy in FWI. It should be

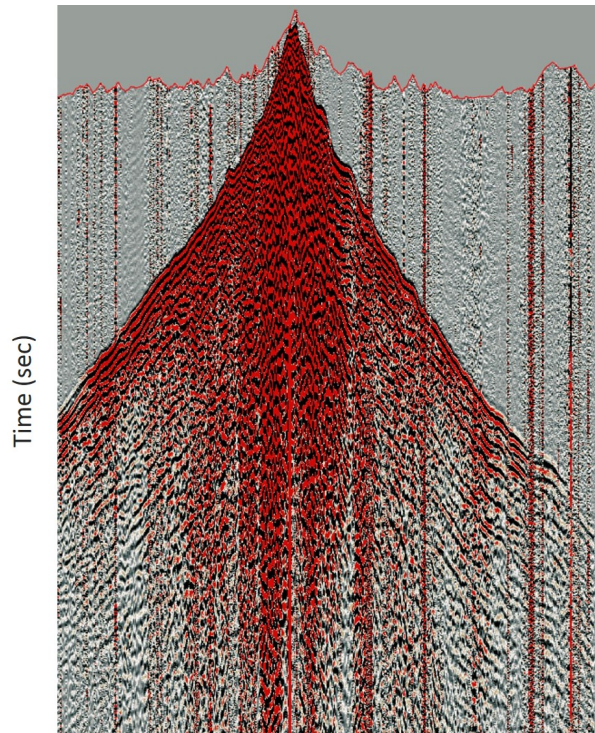


Figure 1.1: A raw 19-km seismic profile acquired from a challenging foothill area. The topography variation is converted into a time muting (red line), placed on top of the shot gather. The total variation of the topography in vertical direction is about 1 km, inducing significant imprints into the seismic data.

noted that these strategies are difficult to achieve in the frequency-domain with few discrete frequencies, aside from the product by exponential decay of the Laplace transform proposed by Shin and Cha (2008).

Figure 1.2 shows an example of a realistic topography extracted from the SEAM Phase II foothill benchmark, created by the SEAM consortium to reproduce the imaging challenges in mountain regions (Oristaglio, 2012, 2016; Regone et al., 2017). Such rapid topography variation will introduce significant footprints into the wavefield such as scattering energies when the waves hit the steep slopes at the surface. Figure 1.1 also highlights that the high-amplitude topography deforms the first-arrivals, implying that a correct 3D geometry representation is critical for accurate wavefield prediction. It should be noted that standard static corrections with respect to a pre-defined datum might correct the travel time but not the full waveform, which is generally complex due to elastic interaction between the seismic waves and the 3D geometries (Yilmaz, 2001; Jones, 2015).

1.3 Spectral element method: Why?

Up to now, most of FWI applications at the crustal scale have been performed in marine environments. In such flat-surface condition, numerical methods designed with finite-difference (FD) discretization appears to be a reasonable choice due to their numerical efficiency, their relatively simple implementation, and the fact that optimized modeling kernels developed for reverse-time migration can be shared.

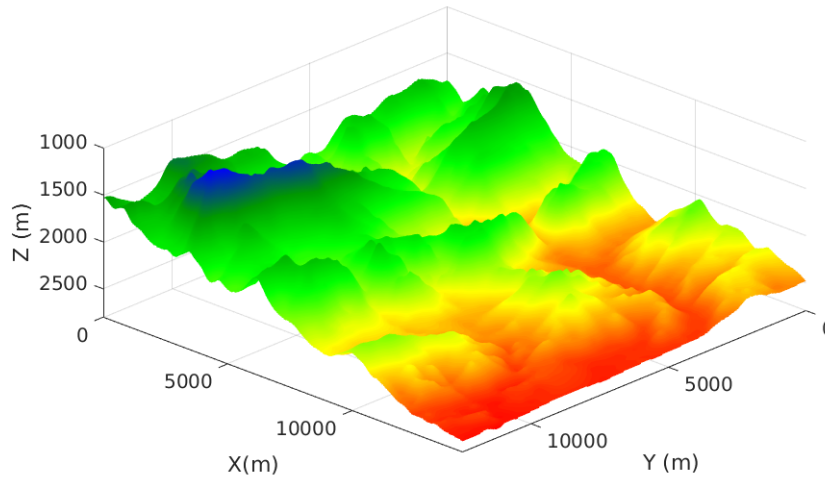


Figure 1.2: Topography of the SEAM Phase II foothill benchmark.

When considering the elastic wave-equation, fluid/solid contrasts can also be implicitly and accurately considered for representing simple bathymetry with FD methods. However, standard FD formulations are often limited to a regular-spacing grid, although there are interesting extensions for deformed grids (Moczo, 1989; Tarrass et al., 2011; Petersson and Sjögreen, 2015; Shragge, 2016). In addition, significant extra efforts are needed when considering surface topography or important geological interfaces (Robertsson, 1996; Bohlen and Saenger, 2006). Extensions, such as the summation-by-parts approaches (Petersson and Sjögreen, 2012), immersed-boundaries implementation (LeVeque, 2007; Lombard et al., 2008; Gao et al., 2015; Huiskes et al., 2017) or hybrid techniques (Moczo et al., 1997) have been proposed. Up to now, they have not displayed superior efficiency compared with finite-element methods (FEM), as described below.

FEM are based on a variational formulation or weak form of the PDEs which need to be discretized. The physical domain is decomposed into a set of elements, leading to a so-call finite element (FE) mesh. The solution is represented over a set of basis functions (often polynomials), defined within each element (Zienkiewicz et al., 2005). FEM have been used in solving the elastodynamic equations by different authors (Marfurt, 1984; Bielak et al., 2003; Koketsu et al., 2004). The free-surface boundary condition is naturally taken into account by zeroing a boundary term coming from the integration by part. The deformed mesh can be easily considered, allowing to represent complex geometries such as the topography variation. The second-order wave equation is generally considered in FE implementations, resulting in a large global mass matrix associated with the acceleration term. The size of this matrix is controlled by the square of the number of degrees of freedom in the model. Its inverse, required for any standard time integration scheme, is not trivial to compute.

Among the wide variety of FE methods, two have been particularly studied within seismic imaging and seismology: Discontinuous Galerkin (DG) and Spectral Element Methods (SEM). Their formulations naturally yields either a block diagonal or diagonal mass matrix, respectively. The inverse of the global mass matrix can be thus cheaply performed, making the use of these FE methods at a reasonable computational cost. The DG method was used to solve the electromagnetic equation in Remaki (2000), then has been intensively investigated in seismic simulation due to several interesting properties (Käser

and Dumbser, 2006; Ben Jemaa et al., 2009; Brossier et al., 2010a): Tetrahedral elements can be used in the mesh creation, allowing to map any specific interfaces or 3D geometries. Test functions are local within each element and they can be different from one element to the others. Therefore, the DG method can account for discontinuities in the wavefield and the model. The integrals in the weak form of wave equation are separately computed at each element, leading to the block diagonal global mass matrix (Cockburn et al., 2000). However, the DG strategy relies on duplicating each degree of freedom at the interface between elements, which increases significantly the computational cost. Many works have been focusing on the numerical cost reduction, for instance by considering the unstructured h -adaptive mesh combined with p -adaptive interpolations (Etienne et al., 2010; Brossier, 2011). The design of the so-called hybridizable DG method is an alternative approach which has recently attracted the attention of the applied mathematics community (Li et al., 2017; Bonnasse-Gahot et al., 2018). Also, considering a local time-stepping per element might help to improve the global cost of the explicit time integration scheme (Diaz and Grote, 2009).

On the other hand, SEM are limited to the used of quadrangle or hexahedral elements, which can complicate the mesh design if one is willing to follow specific interfaces. It is also a continuous FE method, which does not allow for discontinuities of the model and the wavefield. However, the specific choice of Lagrange polynomials as basis functions and Gauss-Lobatto-Legendre (GLL) points as integration points makes the global mass matrix naturally diagonal. This results in an efficient and accurate explicit time-marching scheme, and provides spectral convergence properties. Therefore, the method is particularly attractive for simulations where we can assume the continuity of the solution and medium properties. This is the case for regional and global scale problems where the majority applications of SEM has been performed (Faccioli et al., 1997; Komatitsch and Vilotte, 1998; Capdeville et al., 2003, 2005; Vilotte et al., 2005). The only issue could be related to the specific interfaces meshing. Many realistic geometries, such as geological basin structures or even the approximated ellipsoidal shape of the Earth are shown to be adequately described using hexahedral elements in unstructured meshes (Komatitsch and Tromp, 1999). However, this mesh build-up can require significant human efforts (Peter et al., 2011).

In the exploration scale elastic FWI, we work at low to medium frequencies for which the fields can be assumed to be continuous. Also, in this frequency band, the waves see the subsurface as a reasonably smooth medium (Capdeville et al., 2010; Capdeville and Cance, 2015). Moreover, the theoretical resolution of standard FWI is limited at a half of the local wavelength, implying rather smooth inversion results (Virieux and Operto, 2009). Therefore, SEM appears as the optimum choice. Moreover, the method can be easily adapted to parallel architecture with interesting scaling behavior. The specific issue related to the mesh generation can be mitigated by a high order geometry interpolation, as later described in the manuscript.

1.4 Summary

FWI for foothill targets is recognized as a challenging task which requires a complete physics of the wave propagation and an accurate 3D medium representation. In this Theory and Algorithms part of the manuscript, we build an efficient FWI formulation based on a time-domain SEM on a flexible Cartesian-based mesh. Our strategy focuses on the algorithmic balance between numerical efficiency, memory requirement and simulation accuracy, especially for 3D visco-elastic problems.

Chapter (2) develops the forward problem in SEM perspectives. The chapter reviews the second-order visco-elastic and elastic wave equation for the displacement field. The efficient SEM-based im-

plementation is also described, including flexible mesh design and optimized modeling kernel.

In Chapter (3), the associated FWI problem is formulated. The FWI gradient expressions are derived for anisotropic elastic parameters and isotropic attenuation coefficients, based on a least-squares misfit function. No implicit approximation is considered on attenuation mechanism, aside from the standard-linear-solid (SLS) description. As an illustration of these exact expressions, the impact of the attenuation when recovering elastic parameters will be pointed out in a toy configuration. When inverting for multiple parameters (e.g. compressional and shear wave speeds), we introduce a non-linear constraint on the relationship between inverting parameters, which is useful when dealing with complex heterogeneities or poor illumination.

Chapter (4) focuses on the high performance computing aspect. The scaling behavior of our modeling strategy is evaluated. Similar computational behaviors are observed compared with the open-source FD modeling package SW4 V1.1 (Petersson and Sjögreen, 2013) and SEM-based tool SPECFEM V2.0 (Peter et al., 2011). The chapter also illustrates our strategy to mitigate the memory expense related to the visco-elastic FWI problem.

Since the SEM mesh is considered for both modeling and inversion parts, it is critical to develop an efficient gradient preconditioning strategy to stabilize the ill-posed inversion process. Chapter (5) describes a structured-oriented Bessel smoothing strategy, directly implemented by a SEM formulation on the same mesh as the wavefield modeling (Trinh et al., 2017b). The geological prior knowledge can be incorporated into the smoothing process.

Chapter 2

3D time-domain elastic and visco-elastic modeling in SEM

Contents

2.1	Elastodynamics equations	34
2.1.1	Elastic modeling	35
2.1.2	Visco-elastic modeling	35
2.2	Weak-form development of the visco-elastic wave equation	38
2.2.1	Key elements in spectral element method	39
2.2.2	Some common operations related to the basis functions	41
2.2.3	From the integral form to the discrete form of the wave equation	43
2.2.4	Global system of the weak formulation	45
2.3	Optimized and low-memory visco-elastic modeling kernel	46
2.3.1	Optimized modeling kernel	47
2.3.2	Low-memory visco-elastic modeling kernel	48
2.4	Boundary conditions and source implementation	49
2.4.1	Boundary conditions	49
2.4.2	Source implementation	51
2.5	Mesh design	52
2.5.1	Cartesian-based mesh with variable element-size	53
2.5.2	Complex topography representation	54
2.6	Comparison with semi-analytical solutions - LOH benchmarks	56
2.6.1	LOH1 comparison	57
2.6.2	LOH3 comparison	58
2.7	Conclusion	59
2.A	Stress and strain relationship in the time domain	61
2.B	Complex P and S wave velocities in AXITRA	62

This chapter focuses on the theoretical development and numerical implementation of the forward problem in SEM framework, including a flexible mesh design. It provides a comprehensive review of the second-order visco-elastic and elastic wave equations for the displacement field. The formulation emphasizes on the coupling between the isotropic attenuation mechanism and anisotropic elasticity, in the linear mechanical regime. The weak-form of the wave equation is then derived step-by-step.

We describe a visco-elastic version of the second-order Newmark scheme to propagate the obtained system in time and preserve the second-order accuracy. The source injections for single point force and moment tensor are carefully addressed in order to obtain true-amplitude seismogram computation. The numerical implementation aims at balancing the numerical efficiency and the memory requirement, which are generally considered as bottle-necks for visco-elastic simulation. A combination of sponge layers and radiative boundary condition is used to absorb the outgoing wavefield at the boundaries of the numerical model. The comparison with semi-analytical solutions on the layered-structure LOH benchmarks (a reference model of the Southern California Earthquake Center, SCEC project (Day et al., 2003)) illustrates both the simulation quality of our modeling tools and the efficiency of the boundary absorbing implementation.

2.1 Elastodynamics equations

The elastodynamics equations in this section are developed in the small deformation regime, within the linear Hooke's law approximation. The visco-elastic wave propagation can be written in complete form as

$$\begin{cases} \rho(\mathbf{x})\partial_{tt}u_i(\mathbf{x}, t) &= \partial_j\sigma_{ij}(\mathbf{x}, t) + f_i(\mathbf{x}, t), \\ \sigma_{ij}(\mathbf{x}, t) &= M_{ijkl}(\mathbf{x}, t) *_t \varepsilon_{kl}(\mathbf{x}, t) + \mathcal{T}_{ij}(\mathbf{x}, t), \end{cases} \quad (2.1)$$

where the density is denoted by ρ , the displacement by u , the second-order stress and strain tensors respectively by σ and ε . The external force is denoted by the vector f , and the tensor \mathcal{T} is the possible stress failure. The attenuation effect is described by the relaxation rate M_{ijkl} , and the symbol “ $*_t$ ” denotes the convolution in time-domain. These equations use the Einstein convention (summation over repeated indices).

In the small deformation regime, the strain is defined through the first-order linear approximation of the displacement field, related to Hooke's law, as

$$\varepsilon_{kl} = \frac{1}{2}(\partial_k u_l + \partial_l u_k). \quad (2.2)$$

The Voigt indexing

$$11 \rightarrow 1, 22 \rightarrow 2, 33 \rightarrow 3, 23 \text{ or } 32 \rightarrow 4, 13 \text{ or } 31 \rightarrow 5, 12 \text{ or } 21 \rightarrow 6; \quad (2.3)$$

is used to define the stress and the strain in the vector notation as

$$\boldsymbol{\sigma} = (\sigma_{11}, \sigma_{22}, \sigma_{33}, \sigma_{23}, \sigma_{13}, \sigma_{12})^T = (\sigma_1, \sigma_2, \sigma_3, \sigma_4, \sigma_5, \sigma_6)^T, \quad (2.4)$$

$$\boldsymbol{\varepsilon} = (\varepsilon_{11}, \varepsilon_{22}, \varepsilon_{33}, 2\varepsilon_{23}, 2\varepsilon_{13}, 2\varepsilon_{12})^T = (\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4, \varepsilon_5, \varepsilon_6)^T, \quad (2.5)$$

where the transposed operator is denoted by “ \bullet^T ”. The spatial derivative operator D in the Cartesian space

$$D = \begin{pmatrix} \partial_1 & 0 & 0 & 0 & \partial_3 & \partial_2 \\ 0 & \partial_2 & 0 & \partial_3 & 0 & \partial_1 \\ 0 & 0 & \partial_3 & \partial_2 & \partial_1 & 0 \end{pmatrix} \quad (2.6)$$

can then be used to describe the relationship between the strain ε and the displacement field \mathbf{u} in Equation (2.2) as

$$\varepsilon = D^T \mathbf{u}. \quad (2.7)$$

2.1.1 Elastic modeling

The pure elastic wave equation can be seen as a particular case of the visco-elastic equation with the specific relaxation rate, again within the small deformation regime, as

$$M_{ijkl}(\mathbf{x}, t) = c_{ijkl}(\mathbf{x})\delta(t), \quad (2.8)$$

where c_{ijkl} is the elastic (or *unrelaxed*) stiffness coefficient and $\delta(t)$ is a Dirac delta function, resulting in the elastic wave equation

$$\begin{cases} \rho(\mathbf{x})\partial_{tt}u_i(\mathbf{x}, t) &= \partial_j\sigma_{ij}(\mathbf{x}, t) + f_i(\mathbf{x}, t), \\ \sigma_{ij}(\mathbf{x}, t) &= c_{ijkl}(\mathbf{x})\varepsilon_{kl}(\mathbf{x}, t) + \mathcal{T}_{ij}(\mathbf{x}, t). \end{cases} \quad (2.9)$$

Following the Voigt indexing and matrix notation, the second-order elastic wave equation of displacement field \mathbf{u} can be written as

$$\rho\partial_{tt}\mathbf{u} = DC D^T \mathbf{u} + \mathbf{S}, \quad (2.10)$$

where the elastic stiffness tensor is denoted by C , and the source term by \mathbf{S} . The elastic equation is self-adjoint, implying that we can develop an adjoint system similar to the forward problem. This property makes possible to use the same numerical scheme for incident and adjoint fields propagation (useful for gradient building, see Section (4.2)).

2.1.2 Visco-elastic modeling

The attenuation effect in seismic data is characterized by the energy loss per cycle of the phase, measured by the quality factor. This parameter is either understood as an observable coming from the data or as parametric tensorial description. The seismic attenuation is often described by the quality-factor tensor $Q_{ijkl}(\mathbf{x})$ (Emmerich and Korn, 1987; Carcione et al., 1988). The coefficients of this tensor are the model parameters that we consider for describing the attenuation. In the frequency domain, the inverse of these parameters are defined as the ratio of imaginary and real parts of the complex relaxation rate $\widehat{M}_{ijkl}(\mathbf{x}, \omega)$ as

$$\widehat{Q}_{ijkl}^{-1}(\mathbf{x}, \omega) = \frac{\Im[\widehat{M}_{ijkl}(\mathbf{x}, \omega)]}{\Re[\widehat{M}_{ijkl}(\mathbf{x}, \omega)]}, \quad (2.11)$$

in which the “ $\widehat{\cdot}$ ” notation denotes the Fourier transform.

2.1.2.1 Constant- Q approximation

For realistic crustal scale application, the approximation of a constant Q over the considered frequency band $[\omega_{\min}, \omega_{\max}]$ is usually accepted. Assuming that L standard-linear-solid systems (SLS), at specific reference frequencies $\omega_\nu \in [\omega_{\min}, \omega_{\max}]$, are used to fit a constant Q parameter over this frequency band, one may introduce anelastic coefficients Y_ν^{ijkl} , $\nu = 1, \dots, L$, in the definition of the complex relaxation rate $\widehat{M}_{ijkl}(\mathbf{x}, \omega)$ as

$$\widehat{M}_{ijkl}(\mathbf{x}, \omega) = c_{ijkl}(\mathbf{x}) \left(1 - \sum_{\nu=1}^L Y_\nu^{ijkl}(\mathbf{x}) \frac{\omega_\nu}{\omega_\nu + i\omega} \right), \quad (2.12)$$

(Emmerich and Korn, 1987; Blanch et al., 1995; Moczo and Kristek, 2005; van Driel and Nissen-Meyer, 2014). According to the definition of the complex relaxation rate in Equation (2.11), the resulting quality factor can be written as

$$\widehat{Q}_{ijkl}^{-1}(\mathbf{x}, \omega) = \frac{\sum_{\nu=1}^L Y_\nu^{ijkl}(\mathbf{x}) \frac{\omega_\nu \omega}{\omega_\nu^2 + \omega^2}}{1 - \sum_{\nu=1}^L Y_\nu^{ijkl}(\mathbf{x}) \frac{\omega_\nu \omega}{\omega_\nu^2 + \omega^2}}. \quad (2.13)$$

The assumption of $Q \gg 1$ in realistic attenuative media leads to the following approximation of the quality factor

$$\widehat{Q}_{ijkl}^{-1}(\mathbf{x}, \omega) \approx \sum_{\nu=1}^L Y_\nu^{ijkl}(\mathbf{x}) \frac{\omega_\nu \omega}{\omega_\nu^2 + \omega^2}. \quad (2.14)$$

This equation implies that a set of L anelastic coefficients $Y_\nu^{ijkl}(\mathbf{x})$ is required for each value of the inverse quality factor \widehat{Q}_{ijkl}^{-1} at the spatial location \mathbf{x} (Komatitsch and Tromp, 1999).

Instead of computing L anelastic coefficients $Y_\nu^{ijkl}(\mathbf{x})$ for each spatial location \mathbf{x} , Yang et al. (2016a) proposes the approximation

$$Y_\nu^{ijkl}(\mathbf{x}) \approx y_\nu Q_{ijkl}^{-1}(\mathbf{x}) \quad (2.15)$$

for the entire medium, in which the scalars y_ν being L dimensionless anelastic coefficients. By considering the expression (2.15), the approximation (2.14) of the inverse of the quality factor becomes

$$\widehat{Q}_{ijkl}^{-1}(\mathbf{x}, \omega) \approx \sum_{\nu=1}^L Q_{ijkl}^{-1}(\mathbf{x}) y_\nu \frac{\omega_\nu \omega}{\omega_\nu^2 + \omega^2}. \quad (2.16)$$

We thus aim at searching for the coefficients y_ν , which minimizes the distance between the frequency independent quality factor $Q_{ijkl}^{-1}(\mathbf{x})$ with its approximation $\widehat{Q}_{ijkl}^{-1}(\mathbf{x}, \omega)$ over the considered frequency band. Following this idea, the coefficients y_ν are estimated from the following generalized least-squares minimization problem

$$\min_{y_\nu} \left\{ \int_{\omega_{\min}}^{\omega_{\max}} \left[Q_{ref}^{-1} \left(y_\nu \frac{\omega_\nu \omega}{\omega_\nu^2 + \omega^2} - 1 \right) \right]^2 \right\}, \quad (2.17)$$

which minimizes the distance between a given constant-value Q_{ref}^{-1} with its approximation over the frequency band $[\omega_{\min}, \omega_{\max}]$. The value of this constant is usually chosen such that

$$Q_{ref} \approx \sqrt{\min_{ijkl, \mathbf{x}} (Q_{ijkl}(\mathbf{x})) \times \max_{ijkl, \mathbf{x}} (Q_{ijkl}(\mathbf{x}))}. \quad (2.18)$$

2.1.2.2 Visco-elastic equation under SLS approximation

By considering the approximation (2.15), the complex relaxation rate in Equation (2.12) can now be written as

$$\widehat{M}_{ijkl}(\mathbf{x}, \omega) = c_{ijkl}(\mathbf{x}) - \underbrace{c_{ijkl}(\mathbf{x}) Q_{ijkl}^{-1}(\mathbf{x})}_{\text{attenuation mechanisms}} \sum_{\nu=1}^L y_{\nu} \frac{\omega_{\nu}}{\omega_{\nu} + i\omega}. \quad (2.19)$$

In Equation (2.19), the product between the two fourth-order tensors c_{ijkl} and Q_{ijkl}^{-1} is a term by term product. Equations (2.19) and (2.17) incorporate explicitly the attenuation parameters $Q_{ijkl}^{-1}(\mathbf{x})$ into the wave equation, following a cheap-memory strategy: instead of storing the $L \times 21$ anelastic coefficients $Y_{\nu}^{ijkl}(\mathbf{x})$ associated with 21 coefficients $M_{ijkl}(\mathbf{x})$ at each spatial location (Komatitsch and Tromp, 1999), we only need to store L scalar y_{ν} for the entire medium and the heterogeneous attenuation parameters over the medium. We also store L reference frequencies ω_{ν} associated with L SLSs.

In order to separate the elastic rheology (represented by the *unrelaxed stiffness coefficients* $c_{ijkl}(\mathbf{x})$ in Equation (2.19)) with the attenuation mechanisms, which are frequency-dependent, we introduce *attenuative stiffness coefficients* c_{ijkl}^a . In the case of fully anisotropic attenuation, these parameters can be linked to the unrelaxed quantities through

$$c_{ijkl}^a(\mathbf{x}) = c_{ijkl}(\mathbf{x}) Q_{ijkl}^{-1}(\mathbf{x}), \quad (2.20)$$

which is a term-by-term product. The convolution relationship between stress and strain in Equation (2.1) then becomes

$$\sigma_{ij}(\mathbf{x}, t) = c_{ijkl}(\mathbf{x}) \varepsilon_{kl}(\mathbf{x}, t) - c_{ijkl}^a(\mathbf{x}) \sum_{\nu=1}^L \psi_{\nu;kl}(\mathbf{x}, t) + \mathcal{T}_{ij}(\mathbf{x}, t), \quad (2.21)$$

where the memory variable $\psi_{\nu;kl}(\mathbf{x}, t)$ satisfies the first-order ordinary differential equation (ODE)

$$\partial_t \psi_{\nu;kl}(\mathbf{x}, t) + \omega_{\nu} \psi_{\nu;kl}(\mathbf{x}, t) = \omega_{\nu} y_{\nu} \varepsilon_{kl}(\mathbf{x}, t). \quad (2.22)$$

The mathematical demonstration is given in Appendix 2.A, which is coherent with Moczo and Kristek (2005) and Yang et al. (2016a). Similar to the elastic case (Eqn. 2.10), following the Voigt indexing, the second-order visco-elastic wave equation can thus be written as

$$\begin{cases} \rho \partial_{tt} \mathbf{u} = DC D^T \mathbf{u} - DC^a \sum_{\nu=1}^L \boldsymbol{\psi}_{\nu} + \mathbf{S}, \\ \partial_t \boldsymbol{\psi}_{\nu} + \omega_{\nu} \boldsymbol{\psi}_{\nu} = y_{\nu} \omega_{\nu} D^T \mathbf{u}, \quad \nu = 1, \dots, L, \end{cases} \quad (2.23)$$

where $\boldsymbol{\psi}_{\nu}$ is the memory-variable vector associated with each SLS

$$\boldsymbol{\psi}_{\nu} = (\psi_{\nu;1}, \psi_{\nu;2}, \psi_{\nu;3}, \psi_{\nu;4}, \psi_{\nu;5}, \psi_{\nu;6})^T. \quad (2.24)$$

The matrix $C^a = (C_{IJ}^a)_{6 \times 6}$ contains the attenuative stiffness coefficients, with 21 independent components in the case of fully anisotropic attenuation. Due to the memory variables, the visco-elastic equation (2.23) is not self-adjoint. This property is related to the energy dissipation; however, we shall show that the corresponding adjoint system can be transformed into a similar structure as for the forward problem.

It should be noted that the memory variables can act either on the stress fields as it is the case here (Eqn. 2.21), or on the displacement fields (Petersson and Sjögreen, 2012). The former strategy requires $L \times 6$ variables and 2 spatial derivatives per forward problem, whereas $L \times 3$ incident memory variables with at least 3 spatial derivatives are needed in the latter approach: 2 derivatives for the displacement fields and at least one for the memory variables. Therefore, considering displacement memory variables is less memory demanding but more numerically expensive. Since the estimation of the spatial derivatives is the most expensive operation in our SEM-based modeling kernel, we decide to inject the memory variables into the stress fields as shown in Equation (2.21) to optimize the computational cost.

2.1.2.3 Isotropic attenuation

Practical applications are generally limited to isotropic attenuation because constraining anisotropic attenuation parameters is difficult for surface seismic acquisition. It is important to highlight that the isotropic attenuation regime implies that the *attenuative stiffness tensor* $(C_{IJ}^a)_{6 \times 6}$ is isotropic but not the inverse quality factor matrix $(Q_{IJ}^{-1})_{6 \times 6}$ (Moczo et al., 1997). Hence, the Voigt-noted matrix can be written in the form

$$C^a = \begin{bmatrix} \lambda^a + 2\mu^a & \lambda^a & \lambda^a & 0 & 0 & 0 \\ \lambda^a & \lambda^a + 2\mu^a & \lambda^a & 0 & 0 & 0 \\ \lambda^a & \lambda^a & \lambda^a + 2\mu^a & 0 & 0 & 0 \\ 0 & 0 & 0 & \mu^a & 0 & 0 \\ 0 & 0 & 0 & 0 & \mu^a & 0 \\ 0 & 0 & 0 & 0 & 0 & \mu^a \end{bmatrix}, \quad (2.25)$$

with attenuative Lamé coefficients λ^a and μ^a . We thus consider the effective mechanisms with

$$\lambda^a + 2\mu^a = \frac{1}{3}Q_p^{-1}(C_{11} + C_{22} + C_{33}), \quad \mu^a = \frac{1}{3}Q_s^{-1}(C_{44} + C_{55} + C_{66}), \quad (2.26)$$

where the heterogeneous model parameters Q_p and Q_s are the attenuation parameters associated with compressional and shear wave speeds (V_p, V_s). In the case of isotropic-attenuation mechanism & isotropic-elastic rheology, the proposed solution is consistent with the development suggested by Moczo et al. (1997). Equation (2.26) also implies that isotropic attenuation has isotropic impact on compressional and shear components even when considering anisotropic elasticity, which is physically meaningful. By doing so, the attenuation parameters Q_p and Q_s are explicitly incorporated in the wave equation, even for anisotropic elasticity, and therefore can be naturally considered in the FWI framework. Let us mention that anisotropy feature is important for the characterization of the Earth, which might come from intrinsic and/or structure-induced anisotropy such as layering or intensive faulting structures. Therefore, considering isotropic attenuation while taking into account elastic anisotropy is a good approximation for describing seismic wave propagation, especially for surface acquisitions with limited illumination.

2.2 Weak-form development of the visco-elastic wave equation

As mentioned in Section (1.3), one of the main advantages of SEM formulation compared with other FEM is the diagonal global mass matrix by construction. This comes from the combination of the GLL quadrature for numerical integration with the use of Lagrange basis function. This section will provide a step-by-step development of the weak form of the wave equation to achieve a such formulation.

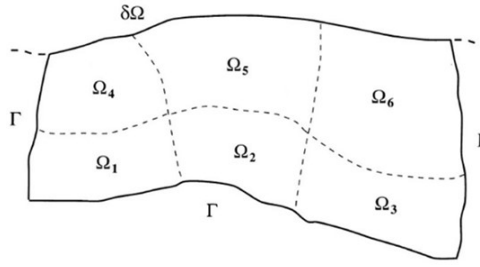


Figure 2.1: The physical model Ω is divided into a set of tensorial elements $\Omega_e (e = 1, \dots, 6)$. Figure is extracted from Komatitsch and Tromp (1999).

We firstly derive some key SEM elements and operations related to the basis functions, then moving gradually from the continuous integral form of the wave equation to the discrete form.

2.2.1 Key elements in spectral element method

To avoid any misunderstanding, we firstly define the *physical space* $\{\mathbf{x}; (z, x, y) \in \mathbb{R}^3\}$, in which the geological structures are defined according to Cartesian coordinates. In standard SEM approaches, the physical domain $\Omega \subset \mathbb{R}^3$ is decomposed into a set of non-overlapping hexahedral elements Ω_e , as illustrated by the 2D example in Figure 2.1. Each hexahedral element Ω_e can be mapped to the *reference cube* $\{\boldsymbol{\xi}; (\xi, \eta, \zeta) \in [-1, 1] \otimes [-1, 1] \otimes [-1, 1]\}$.

2.2.1.1 Mapping from the reference to the physical space

In order to correctly honor geological structures, SEM might need to employ deformed elements in the physical space. The shape of each hexahedral element is defined by a set of n_a *control points* $\mathbf{x}_a = \mathbf{x}(\xi_a, \eta_a, \zeta_a)$; $a = 1, \dots, n_a$ and a set of n_a *shape functions* $N_a(\xi, \eta, \zeta)$ (Komatitsch and Tromp, 1999). The general one-to-one mapping between the reference cube and the hexahedral element ($\boldsymbol{\xi} \mapsto \mathbf{x}$) can be thus written as follows:

$$\mathbf{x}(\xi, \eta, \zeta) = \sum_{a=1}^{n_a} N_a(\xi, \eta, \zeta) \mathbf{x}_a. \quad (2.27)$$

In standard meshing, eight corners of the element are used as control points, the shape function $N_a(\xi, \eta, \zeta)$ are thus triple product of Lagrange polynomials of degrees 1. When mid-size and center nodes are also taken into account, shape functions are triple products of Lagrange polynomials of degrees 2.

At each spatial position, the Jacobian matrix $\mathbf{J}(\boldsymbol{\xi})$ associated with the mapping in Equation (2.27) is defined by

$$\mathbf{J}(\boldsymbol{\xi}) = \begin{bmatrix} \partial_{\xi} z & \partial_{\eta} z & \partial_{\zeta} z \\ \partial_{\xi} x & \partial_{\eta} x & \partial_{\zeta} x \\ \partial_{\xi} y & \partial_{\eta} y & \partial_{\zeta} y \end{bmatrix}. \quad (2.28)$$

An element of volume $d\mathbf{x}$ in the physical space is related to an element of volume $d\boldsymbol{\xi}$ in the reference cube through

$$dzdxdy = J_e d\xi d\eta d\zeta, \quad (2.29)$$

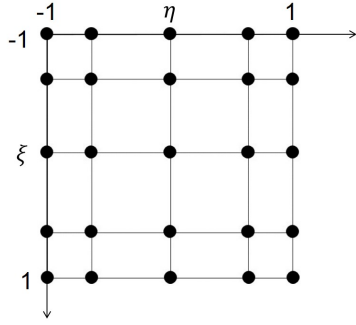


Figure 2.2: Example of GLL points in a 2D element with interpolation order $N = 4$.

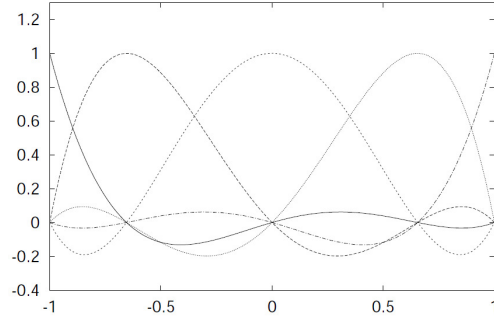


Figure 2.3: 1D Lagrange polynomials defined on GLL points of order $N = 4$.

in which the Jacobian determinant $J_e(\boldsymbol{\xi})$ is the determinant of the Jacobian matrix $\mathbf{J}(\boldsymbol{\xi})$ in Equation (2.28). The uniqueness of the projection between the reference and physical space is characterized by positive values of the Jacobian determinant $J_e(\boldsymbol{\xi})$ everywhere on the reference cube. For this purpose, the geometrical shape of each element in the physical space should satisfy the condition that all of their vertex angles should be bounded away from 0 and π (Deville et al., 2002).

When considering a *fully deformed element* in the physical space, the Jacobian matrix $\mathbf{J}(\boldsymbol{\xi})$ is filled of non-zero terms. However, this matrix becomes diagonal in the case of *regular element*, i.e. rectangular parallelepiped shape. This property simplifies the numerical implementation in the regular SEM mesh, leading to a better computational efficiency compared with the fully deformed mesh. In practical applications, it is thus strongly advised to combine deformed elements with regular elements in the case of weak topography variation.

2.2.1.2 Integration points & Gauss-Lobatto-Legendre quadrature

SEM employs Gauss-Lobatto-Legendre (GLL) integration points, which are defined in the reference cube. In 1D, the $(N + 1)$ GLL points are roots of the polynomial

$$(1 - \xi^2)L'_N(\xi) = 0, \quad (2.30)$$

where L_N is Legendre polynomial¹ of degree N , and N refers to the *interpolation order*. The GLL points in 2D or 3D are simply defined based on tensorial products of GLL points in 1D thanks to the tensorial property of quadrangular and hexahedral elements. Example of GLL point in a 2D element at interpolation order $N = 4$ is provided in Figure 2.2.

In 3D, the reference cube is discretized into a set of $(N + 1)^3$ Gauss-Lobatto-Legendre (GLL) points

$$\boldsymbol{\xi}_{\hat{k}} = (\xi_{k_1}, \eta_{k_2}, \zeta_{k_3}) \quad \hat{k} \text{ stands for the triple indexes } \{k_1, k_2, k_3 = 0, \dots, N\}, \quad (2.32)$$

¹Legendre polynomial can be defined by the following recursion relation with $L_0(x) = 1$ and $L_1(x) = x$ (Canuto et al., 2006):

$$L_{k+1}(x) = \frac{2k+1}{k+1}xL_k(x) - \frac{k}{k+1}L_{k-1}(x) \quad (2.31)$$

where k_1, k_2, k_3 are the indexing of GLL points in z, x and y dimensions, respectively. These collocation points are used to define $(N + 1)^3$ basis functions $\ell_{\hat{k}}(\boldsymbol{\xi})$

$$\ell_{\hat{k}}(\boldsymbol{\xi}) = \ell_{k_1}(\xi) \ell_{k_2}(\eta) \ell_{k_3}(\zeta); \quad (2.33)$$

each function is a triple product of Lagrange polynomials of degree N

$$\ell_j(\xi) = \prod_{\substack{i=0 \\ i \neq j}}^N \frac{\xi - \xi_i}{\xi_j - \xi_i}. \quad (2.34)$$

Lagrange polynomials have the interesting property that its values at GLL nodes are equal either 0 or 1

$$\ell_j(\xi_i) = \delta_{ji} = \begin{cases} 0 & \text{if } i \neq j, \\ 1 & \text{if } i = j \end{cases}, \quad (2.35)$$

where δ_{ji} is the Kronecker notation. The 1D Lagrange polynomials defined on GLL points of order $N = 4$ are presented in Figure 2.3.

In SEM, the same collocation points are considered for spatial interpolation and numerical integration. The approximation of a vector $\mathbf{v}(\boldsymbol{\xi}) = (v_1(\boldsymbol{\xi}), v_2(\boldsymbol{\xi}), v_3(\boldsymbol{\xi}))^T$ over the reference cube can thus be written as follows

$$\mathbf{v}(\boldsymbol{\xi}) \approx \sum_{h_1=0}^N \sum_{h_2=0}^N \sum_{h_3=0}^N \mathbf{v}_{\hat{h}} \ell_{h_1}(\xi) \ell_{h_2}(\eta) \ell_{h_3}(\zeta) = \sum_{\hat{h}=0}^N \mathbf{v}_{\hat{h}} \ell_{\hat{h}}(\boldsymbol{\xi}), \quad (2.36)$$

where $\mathbf{v}_{\hat{h}} = \mathbf{v}(\xi_{h_1}, \eta_{h_2}, \zeta_{h_3})$. Equation (2.36) implies that heterogeneous medium properties inside each element could be easily considered. This property is widely used in practical SEM implementation, especially at high interpolation order N .

The integral of any continuous function $f(\boldsymbol{\xi})$ over the reference cube can be approximated by the *Gauss-Lobatto-Legendre quadrature* as

$$\iiint_{-1}^1 f(\boldsymbol{\xi}) d\boldsymbol{\xi} \approx \sum_{\hat{q}=0}^N w_{\hat{q}} f(\boldsymbol{\xi}_{\hat{q}}), \quad (2.37)$$

where $w_{\hat{q}}$ stands for the product of quadrature weights² $w_{q_1}, w_{q_2}, w_{q_3}$, associated with the GLL integration points $\boldsymbol{\xi}_{\hat{q}}$. The convergence characteristics of SEM are guaranteed for polynomial functions up to the degree $(2N + 1)$, in which the expression (2.37) becomes exact.

2.2.2 Some common operations related to the basis functions

Since the basis function $\ell_{\hat{k}}(\mathbf{x})$ is a single-dimensional function of multiple variables $\ell_{\hat{k}}(\mathbf{x}) : \mathbb{R}^3 \mapsto \mathbb{R}$, its product with a multiple-dimensional field $\mathbf{v}(\mathbf{x}) : \mathbb{R}^3 \mapsto \mathbb{R}^d$ can be written as

$$\ell_{\hat{k}}(\mathbf{x}) \mathbf{v}(\mathbf{x}) = \ell_{\hat{k}}(\mathbf{x}) \begin{bmatrix} v_1(\mathbf{x}) \\ \dots \\ v_d(\mathbf{x}) \end{bmatrix} = \begin{bmatrix} \ell_{\hat{k}}(\mathbf{x}) v_1(\mathbf{x}) \\ \dots \\ \ell_{\hat{k}}(\mathbf{x}) v_d(\mathbf{x}) \end{bmatrix}, \quad (2.39)$$

²In each dimension, the quadrature weights associated with the GLL points are defined as

$$w_j = \frac{2}{N(N+1)} \frac{1}{[L_N(\xi_j)]^2} \quad (\forall \xi_j \neq \pm 1); \quad w_j = \frac{2}{N(N+1)} \quad (\xi_j = \pm 1) \quad (2.38)$$

where d is the dimension of the field $\mathbf{v}(\mathbf{x})$. However, we do not have similar property with respect to the spatial derivatives operator D , in Equation (2.6).

When developing the weak form of the wave equation, we encounter frequently the integral form of the spatial derivatives of a vector, for instance the stress $\boldsymbol{\sigma}$, which is a 6-dimensional field. Let's consider the integral form of the spatial derivatives of a 6-dimensional field $\mathbf{v}(\mathbf{x}) : \mathbb{R}^3 \mapsto \mathbb{R}^6$ over the model volume Ω . By integrating by parts over the model volume Ω we have

$$\int_{\Omega} \ell_{\hat{k}}(\mathbf{x}) D\mathbf{v}(\mathbf{x}) d\mathbf{x} = - \int_{\Omega} [D\ell_{\hat{k}}(\mathbf{x})] \mathbf{v}(\mathbf{x}) d\mathbf{x} + \underbrace{\int_{\Gamma_{\Omega}} [I_D \ell_{\hat{k}}(\mathbf{x})] \mathbf{v}(\mathbf{x}) ds_{\mathbf{x}}}_{\text{Boundary term}}, \quad (2.40)$$

where in the first term $[D\ell_{\hat{k}}(\mathbf{x})]$ stands for

$$[D\ell_{\hat{k}}(\mathbf{x})] = \begin{pmatrix} \partial_1 \ell_{\hat{k}}(\mathbf{x}) & 0 & 0 & 0 & \partial_3 \ell_{\hat{k}}(\mathbf{x}) & \partial_2 \ell_{\hat{k}}(\mathbf{x}) \\ 0 & \partial_2 \ell_{\hat{k}}(\mathbf{x}) & 0 & \partial_3 \ell_{\hat{k}}(\mathbf{x}) & 0 & \partial_1 \ell_{\hat{k}}(\mathbf{x}) \\ 0 & 0 & \partial_3 \ell_{\hat{k}}(\mathbf{x}) & \partial_2 \ell_{\hat{k}}(\mathbf{x}) & \partial_1 \ell_{\hat{k}}(\mathbf{x}) & 0 \end{pmatrix}, \quad (2.41)$$

and in the boundary term the expression $[I_D \ell_{\hat{k}}(\mathbf{x})]$ stands for

$$[I_D \ell_{\hat{k}}(\mathbf{x})] = \begin{pmatrix} \ell_{\hat{k}}(\mathbf{x}) & 0 & 0 & 0 & \ell_{\hat{k}}(\mathbf{x}) & \ell_{\hat{k}}(\mathbf{x}) \\ 0 & \ell_{\hat{k}}(\mathbf{x}) & 0 & \ell_{\hat{k}}(\mathbf{x}) & 0 & \ell_{\hat{k}}(\mathbf{x}) \\ 0 & 0 & \ell_{\hat{k}}(\mathbf{x}) & \ell_{\hat{k}}(\mathbf{x}) & \ell_{\hat{k}}(\mathbf{x}) & 0 \end{pmatrix}. \quad (2.42)$$

The boundary term in Equation (2.40) is the integral over the surface boundaries Γ_{Ω} of the model volume where $ds_{\mathbf{x}}$ is the surface unitary in the physical space. At the free-surface, this boundary term vanishes.

When moving from the physical space \mathbf{x} to the reference space $\boldsymbol{\xi}$ in ignoring the boundary term, the volumetric Jacobian determinant $J_e(\boldsymbol{\xi})$ and surface determinant $J_e^s(\boldsymbol{\xi})$ are introduced into the integral (2.40) as

$$\int_{\Omega} \ell_{\hat{k}}(\mathbf{x}) D\mathbf{v}(\mathbf{x}) d\mathbf{x} = - \int_{\Omega} [D\ell_{\hat{k}}(\boldsymbol{\xi})] \mathbf{v}(\boldsymbol{\xi}) J_e(\boldsymbol{\xi}) d\boldsymbol{\xi} + \int_{\Gamma_{\Omega}} [I_D \ell_{\hat{k}}(\boldsymbol{\xi})] \mathbf{v}(\boldsymbol{\xi}) J_e^s(\boldsymbol{\xi}) ds_{\boldsymbol{\xi}}, \quad (2.43)$$

where $ds_{\boldsymbol{\xi}}$ is the surface unitary in the reference space. To simplify the mathematical expressions, we introduce the following notations of the coordinates of the dimensionless physical space and the reference space

$$(p_1, p_2, p_3) := (z, x, y) = \mathbf{x} \quad \text{and} \quad (r_1, r_2, r_3) := (\xi, \eta, \zeta) = \boldsymbol{\xi}. \quad (2.44)$$

The spatial derivatives in Equation (2.43) are also modified according to the chain rules with

$$\partial_1 = \frac{\partial}{\partial p_1} = \sum_{j=1}^3 \frac{\partial}{\partial r_j} \frac{\partial r_j}{\partial p_1}, \quad (2.45)$$

$$\partial_2 = \frac{\partial}{\partial p_2} = \sum_{j=1}^3 \frac{\partial}{\partial r_j} \frac{\partial r_j}{\partial p_2}, \quad (2.46)$$

$$\partial_3 = \frac{\partial}{\partial p_3} = \sum_{j=1}^3 \frac{\partial}{\partial r_j} \frac{\partial r_j}{\partial p_3}. \quad (2.47)$$

Following this development, we have

$$[D\ell_{\hat{k}}(\boldsymbol{\xi})] = \sum_{j=1}^3 \frac{\partial \ell_{\hat{k}}(\boldsymbol{\xi})}{\partial r_j} \begin{pmatrix} \frac{\partial r_j}{\partial p_1} & 0 & 0 & 0 & \frac{\partial r_j}{\partial p_3} & \frac{\partial r_j}{\partial p_2} \\ 0 & \frac{\partial r_j}{\partial p_2} & 0 & \frac{\partial r_j}{\partial p_3} & 0 & \frac{\partial r_j}{\partial p_1} \\ 0 & 0 & \frac{\partial r_j}{\partial p_3} & \frac{\partial r_j}{\partial p_2} & \frac{\partial r_j}{\partial p_1} & 0 \end{pmatrix}. \quad (2.48)$$

We introduce the geometric matrix, associated with the projection of spatial derivatives between the Cartesian coordinates and the reference coordinates as

$$G_j(\boldsymbol{\xi}) = \begin{pmatrix} \frac{\partial r_j}{\partial p_1} & 0 & 0 & 0 & \frac{\partial r_j}{\partial p_3} & \frac{\partial r_j}{\partial p_2} \\ 0 & \frac{\partial r_j}{\partial p_2} & 0 & \frac{\partial r_j}{\partial p_3} & 0 & \frac{\partial r_j}{\partial p_1} \\ 0 & 0 & \frac{\partial r_j}{\partial p_3} & \frac{\partial r_j}{\partial p_2} & \frac{\partial r_j}{\partial p_1} & 0 \end{pmatrix}. \quad (2.49)$$

Hence,

$$[D\ell_{\hat{k}}(\boldsymbol{\xi})] = \sum_{j=1}^3 \frac{\partial \ell_{\hat{k}}(\boldsymbol{\xi})}{\partial r_j} G_j(\boldsymbol{\xi}). \quad (2.50)$$

By considering the compact form (2.50) of the spatial derivative operator when moving from the physical space \mathbf{x} to the reference space $\boldsymbol{\xi}$, the integral (2.43) becomes

$$\int_{\Omega} \ell_{\hat{k}}(\mathbf{x}) D\mathbf{v}(\mathbf{x}) d\mathbf{x} = - \int_{\Omega} \sum_{j=1}^3 \frac{\partial \ell_{\hat{k}}(\boldsymbol{\xi})}{\partial r_j} G_j(\boldsymbol{\xi}) \mathbf{v}(\boldsymbol{\xi}) J_e(\boldsymbol{\xi}) d\boldsymbol{\xi} + \int_{\Gamma_{\Omega}} [I_D \ell_{\hat{k}}(\boldsymbol{\xi})] \mathbf{v}(\boldsymbol{\xi}) J_e^s(\boldsymbol{\xi}) ds_{\boldsymbol{\xi}}. \quad (2.51)$$

The relationship (2.51) will be frequently used in the development of the weak form of the wave equation.

2.2.3 From the integral form to the discrete form of the wave equation

Considering the choice of Lagrange basis functions $\ell_{\hat{k}}(\mathbf{x})$, the integral form of the visco-elastic wave equation (2.23) over a model volume Ω can be written as

$$\left\{ \begin{aligned} \int_{\Omega} \ell_{\hat{k}}(\mathbf{x}) \rho(\mathbf{x}) \partial_{tt} \mathbf{u}(\mathbf{x}) d\mathbf{x} &= \int_{\Omega} \ell_{\hat{k}}(\mathbf{x}) D C D^T \mathbf{u}(\mathbf{x}) d\mathbf{x} - \int_{\Omega} \ell_{\hat{k}}(\mathbf{x}) D C^a \sum_{\nu=1}^L \psi_{\nu}(\mathbf{x}) d\mathbf{x} + \int_{\Omega} \ell_{\hat{k}}(\mathbf{x}) \mathbf{S}(\mathbf{x}) d\mathbf{x}, \\ \int_{\Omega} \ell_{\hat{k}}(\mathbf{x}) \partial_t \psi_{\nu}(\mathbf{x}) d\mathbf{x} + \omega_{\nu} \int_{\Omega} \ell_{\hat{k}}(\mathbf{x}) \psi_{\nu}(\mathbf{x}) d\mathbf{x} &= y_{\nu} \omega_{\nu} \int_{\Omega} \ell_{\hat{k}}(\mathbf{x}) D^T \mathbf{u}(\mathbf{x}) d\mathbf{x}. \end{aligned} \right. \quad (2.52)$$

Applying the integration by part described in Equation (2.40), we obtain

$$\left\{ \begin{aligned} \int_{\Omega} \ell_{\hat{k}}(\mathbf{x}) \rho(\mathbf{x}) \partial_{tt} \mathbf{u}(\mathbf{x}) d\mathbf{x} &= - \int_{\Omega} [D\ell_{\hat{k}}(\mathbf{x})] C D^T \mathbf{u}(\mathbf{x}) d\mathbf{x} + \int_{\Omega} [D\ell_{\hat{k}}(\mathbf{x})] C^a \sum_{\nu=1}^L \psi_{\nu}(\mathbf{x}) d\mathbf{x} \\ &\quad + \underbrace{\int_{\Gamma_{\Omega}} [I_D \ell_{\hat{k}}(\mathbf{x})] \boldsymbol{\tau}_B(\mathbf{x}) ds_{\mathbf{x}}}_{\text{Boundary term B}} + \int_{\Omega} \ell_{\hat{k}}(\mathbf{x}) \mathbf{S}(\mathbf{x}) d\mathbf{x}, \\ \int_{\Omega} \ell_{\hat{k}}(\mathbf{x}) \partial_t \psi_{\nu}(\mathbf{x}) d\mathbf{x} + \omega_{\nu} \int_{\Omega} \ell_{\hat{k}}(\mathbf{x}) \psi_{\nu}(\mathbf{x}) d\mathbf{x} &= y_{\nu} \omega_{\nu} \int_{\Omega} \ell_{\hat{k}}(\mathbf{x}) D^T \mathbf{u}(\mathbf{x}) d\mathbf{x}. \end{aligned} \right. \quad (2.53)$$

At the boundaries of the numerical model, the attenuation of the outgoing wavefield is applied through the traction $\boldsymbol{\tau}_B$ which will be further described in Section (2.4). For the sake of simplicity, we ignore temporarily the *boundary term* \mathbf{B} in the next steps.

Moving from the physical space \mathbf{x} to the reference space $\boldsymbol{\xi}$ and considering the transformation (2.51) provide

$$\left\{ \begin{aligned} \int_{\Omega} \ell_{\hat{k}}(\boldsymbol{\xi}) \rho(\boldsymbol{\xi}) \partial_{tt} \mathbf{u}(\boldsymbol{\xi}) J_e(\boldsymbol{\xi}) d\boldsymbol{\xi} &= - \int_{\Omega} \sum_{j=1}^3 \frac{\partial \ell_{\hat{k}}(\boldsymbol{\xi})}{\partial r_j} G_j(\boldsymbol{\xi}) C D^T \mathbf{u}(\boldsymbol{\xi}) J_e(\boldsymbol{\xi}) d\boldsymbol{\xi} \\ &+ \int_{\Omega} \sum_{j=1}^3 \frac{\partial \ell_{\hat{k}}(\boldsymbol{\xi})}{\partial r_j} G_j(\boldsymbol{\xi}) C^a \sum_{\nu=1}^L \psi_{\nu}(\boldsymbol{\xi}) J_e(\boldsymbol{\xi}) d\boldsymbol{\xi} + \int_{\Omega} \ell_{\hat{k}}(\boldsymbol{\xi}) \mathbf{S}(\boldsymbol{\xi}) J_e(\boldsymbol{\xi}) d\boldsymbol{\xi}, \\ \int_{\Omega} \ell_{\hat{k}}(\boldsymbol{\xi}) \partial_t \psi_{\nu}(\boldsymbol{\xi}) J_e(\boldsymbol{\xi}) d\boldsymbol{\xi} + \omega_{\nu} \int_{\Omega} \ell_{\hat{k}}(\boldsymbol{\xi}) \psi_{\nu}(\boldsymbol{\xi}) J_e(\boldsymbol{\xi}) d\boldsymbol{\xi} &= y_{\nu} \omega_{\nu} \int_{\Omega} \ell_{\hat{k}}(\boldsymbol{\xi}) D^T \mathbf{u}(\boldsymbol{\xi}) J_e(\boldsymbol{\xi}) d\boldsymbol{\xi}. \end{aligned} \right. \quad (2.54)$$

In the next step, we approximate the physical fields, \mathbf{u} , \mathbf{S} and ψ_{ν} , by the spatial interpolation over the basis function (2.36) and obtain

$$\left\{ \begin{aligned} \sum_{\hat{h}=1}^N \int_{\Omega} \ell_{\hat{k}}(\boldsymbol{\xi}) \rho(\boldsymbol{\xi}) \ell_{\hat{h}}(\boldsymbol{\xi}) \partial_{tt} \mathbf{u}_{\hat{h}} J_e(\boldsymbol{\xi}) d\boldsymbol{\xi} &= - \int_{\Omega} \sum_{j=1}^3 \frac{\partial \ell_{\hat{k}}(\boldsymbol{\xi})}{\partial r_j} G_j(\boldsymbol{\xi}) C \sum_{\hat{h}=1}^N [D^T \ell_{\hat{h}}(\boldsymbol{\xi})] \mathbf{u}_{\hat{h}} J_e(\boldsymbol{\xi}) d\boldsymbol{\xi} \\ &+ \int_{\Omega} \sum_{j=1}^3 \frac{\partial \ell_{\hat{k}}(\boldsymbol{\xi})}{\partial r_j} G_j(\boldsymbol{\xi}) C^a \sum_{\hat{h}=1}^N \ell_{\hat{h}}(\boldsymbol{\xi}) \sum_{\nu=1}^L \psi_{\nu; \hat{h}} J_e(\boldsymbol{\xi}) d\boldsymbol{\xi} + \sum_{\hat{h}=1}^N \int_{\Omega} \ell_{\hat{k}}(\boldsymbol{\xi}) \ell_{\hat{h}}(\boldsymbol{\xi}) \mathbf{S}_{\hat{h}} J_e(\boldsymbol{\xi}) d\boldsymbol{\xi}, \\ \sum_{\hat{h}=1}^N \int_{\Omega} \ell_{\hat{k}}(\boldsymbol{\xi}) \ell_{\hat{h}}(\boldsymbol{\xi}) \partial_t \psi_{\nu; \hat{h}} J_e(\boldsymbol{\xi}) d\boldsymbol{\xi} + \omega_{\nu} \sum_{\hat{h}=1}^N \int_{\Omega} \ell_{\hat{k}}(\boldsymbol{\xi}) \ell_{\hat{h}}(\boldsymbol{\xi}) \psi_{\nu; \hat{h}} J_e(\boldsymbol{\xi}) d\boldsymbol{\xi} \\ &= y_{\nu} \omega_{\nu} \int_{\Omega} \ell_{\hat{k}}(\boldsymbol{\xi}) \sum_{\hat{h}=1}^N [D^T \ell_{\hat{h}}(\boldsymbol{\xi})] \mathbf{u}_{\hat{h}} J_e(\boldsymbol{\xi}) d\boldsymbol{\xi}. \end{aligned} \right. \quad (2.55)$$

Similar to the compact form (2.50), the operator $[D^T \ell_{\hat{h}}(\boldsymbol{\xi})]$ can be written as

$$[D^T \ell_{\hat{h}}(\boldsymbol{\xi})] = \sum_{i=1}^3 G_i^T(\boldsymbol{\xi}) \frac{\partial \ell_{\hat{h}}(\boldsymbol{\xi})}{\partial r_i}, \quad (2.56)$$

which transform the system (2.55) into

$$\left\{ \begin{aligned} \sum_{\hat{h}=1}^N \int_{\Omega} \ell_{\hat{k}}(\boldsymbol{\xi}) \rho(\boldsymbol{\xi}) \ell_{\hat{h}}(\boldsymbol{\xi}) \partial_{tt} \mathbf{u}_{\hat{h}} J_e(\boldsymbol{\xi}) d\boldsymbol{\xi} &= - \int_{\Omega} \sum_{j=1}^3 \frac{\partial \ell_{\hat{k}}(\boldsymbol{\xi})}{\partial r_j} G_j(\boldsymbol{\xi}) C \sum_{\hat{h}=1}^N \sum_{i=1}^3 G_i^T(\boldsymbol{\xi}) \frac{\partial \ell_{\hat{h}}(\boldsymbol{\xi})}{\partial r_i} \mathbf{u}_{\hat{h}} J_e(\boldsymbol{\xi}) d\boldsymbol{\xi} \\ &+ \int_{\Omega} \sum_{j=1}^3 \frac{\partial \ell_{\hat{k}}(\boldsymbol{\xi})}{\partial r_j} G_j(\boldsymbol{\xi}) C^a \sum_{\hat{h}=1}^N \ell_{\hat{h}}(\boldsymbol{\xi}) \sum_{\nu=1}^L \psi_{\nu; \hat{h}} J_e(\boldsymbol{\xi}) d\boldsymbol{\xi} + \sum_{\hat{h}=1}^N \int_{\Omega} \ell_{\hat{k}}(\boldsymbol{\xi}) \ell_{\hat{h}}(\boldsymbol{\xi}) \mathbf{S}_{\hat{h}} J_e(\boldsymbol{\xi}) d\boldsymbol{\xi}, \\ \sum_{\hat{h}=1}^N \int_{\Omega} \ell_{\hat{k}}(\boldsymbol{\xi}) \ell_{\hat{h}}(\boldsymbol{\xi}) \partial_t \psi_{\nu; \hat{h}} J_e(\boldsymbol{\xi}) d\boldsymbol{\xi} + \omega_{\nu} \sum_{\hat{h}=1}^N \int_{\Omega} \ell_{\hat{k}}(\boldsymbol{\xi}) \ell_{\hat{h}}(\boldsymbol{\xi}) \psi_{\nu; \hat{h}} J_e(\boldsymbol{\xi}) d\boldsymbol{\xi} \\ &= y_{\nu} \omega_{\nu} \int_{\Omega} \ell_{\hat{k}}(\boldsymbol{\xi}) \sum_{\hat{h}=1}^N \sum_{i=1}^3 G_i^T(\boldsymbol{\xi}) \frac{\partial \ell_{\hat{h}}(\boldsymbol{\xi})}{\partial r_i} \mathbf{u}_{\hat{h}} J_e(\boldsymbol{\xi}) d\boldsymbol{\xi}. \end{aligned} \right. \quad (2.57)$$

The last step involves the application of GLL quadrature mentioned in Equation (2.37) and using the property of Lagrange polynomials described in Equation (2.35), the system (2.57) becomes

$$\left\{ \begin{aligned} & \sum_{\hat{h}=1}^N \sum_{\hat{q}=1}^N w_{\hat{q}} \rho_{\hat{q}} \delta_{\hat{k}\hat{q}} \delta_{\hat{h}\hat{q}} \partial_{tt} \mathbf{u}_{\hat{h}} J_e(\boldsymbol{\xi}_{\hat{q}}) = - \sum_{\hat{q}=1}^N w_{\hat{q}} \sum_{j=1}^3 \frac{\partial \ell_{\hat{k}}(\boldsymbol{\xi}_{\hat{q}})}{\partial r_j} G_j(\boldsymbol{\xi}_{\hat{q}}) C \sum_{\hat{h}=1}^N \sum_{i=1}^3 G_i^T(\boldsymbol{\xi}_{\hat{q}}) \frac{\partial \ell_{\hat{h}}(\boldsymbol{\xi}_{\hat{q}})}{\partial r_i} \mathbf{u}_{\hat{h}} J_e(\boldsymbol{\xi}_{\hat{q}}) \\ & \quad + \sum_{\hat{q}=1}^N w_{\hat{q}} \sum_{j=1}^3 \frac{\partial \ell_{\hat{k}}(\boldsymbol{\xi}_{\hat{q}})}{\partial r_j} G_j(\boldsymbol{\xi}_{\hat{q}}) C^a \sum_{\hat{h}=1}^N \delta_{\hat{h}\hat{q}} \sum_{\nu=1}^L \boldsymbol{\psi}_{\nu;\hat{h}} J_e(\boldsymbol{\xi}_{\hat{q}}) + \sum_{\hat{h}=1}^N \sum_{\hat{q}=1}^N w_{\hat{q}} \delta_{\hat{k}\hat{q}} \delta_{\hat{h}\hat{q}} \mathbf{S}_{\hat{h}} J_e(\boldsymbol{\xi}_{\hat{q}}), \\ & \sum_{\hat{h}=1}^N \sum_{\hat{q}=1}^N w_{\hat{q}} \delta_{\hat{k}\hat{q}} \delta_{\hat{h}\hat{q}} \partial_t \boldsymbol{\psi}_{\nu;\hat{h}} J_e(\boldsymbol{\xi}_{\hat{q}}) + \omega_{\nu} \sum_{\hat{h}=1}^N \sum_{\hat{q}=1}^N w_{\hat{q}} \delta_{\hat{k}\hat{q}} \delta_{\hat{h}\hat{q}} \boldsymbol{\psi}_{\nu;\hat{h}} J_e(\boldsymbol{\xi}_{\hat{q}}) \\ & \quad = y_{\nu} \omega_{\nu} \sum_{\hat{q}=1}^N w_{\hat{q}} \delta_{\hat{k}\hat{q}} \sum_{\hat{h}=1}^N \sum_{i=1}^3 G_i^T(\boldsymbol{\xi}_{\hat{q}}) \frac{\partial \ell_{\hat{h}}(\boldsymbol{\xi}_{\hat{q}})}{\partial r_i} \mathbf{u}_{\hat{h}} J_e(\boldsymbol{\xi}_{\hat{q}}). \end{aligned} \right. \quad (2.58)$$

By simplifying the Kronecker delta functions, we obtain a cleaner system

$$\left\{ \begin{aligned} & w_{\hat{k}} \rho_{\hat{k}} J_e(\boldsymbol{\xi}_{\hat{k}}) \partial_{tt} \mathbf{u}_{\hat{k}} = - \sum_{\hat{q}=1}^N w_{\hat{q}} J_e(\boldsymbol{\xi}_{\hat{q}}) \sum_{j=1}^3 \frac{\partial \ell_{\hat{k}}(\boldsymbol{\xi}_{\hat{q}})}{\partial r_j} G_j(\boldsymbol{\xi}_{\hat{q}}) C \sum_{\hat{h}=1}^N \sum_{i=1}^3 G_i^T(\boldsymbol{\xi}_{\hat{q}}) \frac{\partial \ell_{\hat{h}}(\boldsymbol{\xi}_{\hat{q}})}{\partial r_i} \mathbf{u}_{\hat{h}} \\ & \quad + \sum_{\hat{q}=1}^N w_{\hat{q}} J_e(\boldsymbol{\xi}_{\hat{q}}) \sum_{j=1}^3 \frac{\partial \ell_{\hat{k}}(\boldsymbol{\xi}_{\hat{q}})}{\partial r_j} G_j(\boldsymbol{\xi}_{\hat{q}}) C^a \sum_{\hat{q}=1}^N \sum_{\nu=1}^L \boldsymbol{\psi}_{\nu;\hat{q}} + w_{\hat{k}} J_e(\boldsymbol{\xi}_{\hat{k}}) \mathbf{S}_{\hat{k}}, \\ & w_{\hat{k}} J_e(\boldsymbol{\xi}_{\hat{k}}) \partial_t \boldsymbol{\psi}_{\nu;\hat{k}} + \omega_{\nu} w_{\hat{k}} J_e(\boldsymbol{\xi}_{\hat{k}}) \boldsymbol{\psi}_{\nu;\hat{k}} = y_{\nu} \omega_{\nu} w_{\hat{k}} J_e(\boldsymbol{\xi}_{\hat{k}}) \sum_{\hat{h}=1}^N \sum_{i=1}^3 G_i^T(\boldsymbol{\xi}_{\hat{k}}) \frac{\partial \ell_{\hat{h}}(\boldsymbol{\xi}_{\hat{k}})}{\partial r_i} \mathbf{u}_{\hat{h}}. \end{aligned} \right. \quad (2.59)$$

By removing $w_{\hat{k}} J_e(\boldsymbol{\xi}_{\hat{k}})$ from both sides of the second equation and changing the index \hat{k} to \hat{q} we have

$$\left\{ \begin{aligned} & \underbrace{w_{\hat{k}} \rho_{\hat{k}} J_e(\boldsymbol{\xi}_{\hat{k}})}_{M_{\hat{k}\hat{k}}} \partial_{tt} \mathbf{u}_{\hat{k}} = - \overbrace{\sum_{\hat{q}=1}^N w_{\hat{q}} J_e(\boldsymbol{\xi}_{\hat{q}}) \sum_{j=1}^3 \frac{\partial \ell_{\hat{k}}(\boldsymbol{\xi}_{\hat{q}})}{\partial r_j} G_j(\boldsymbol{\xi}_{\hat{q}}) C}^{\mathcal{D}_{\hat{k}\hat{q}}} \sum_{\hat{h}=1}^N \sum_{i=1}^3 G_i^T(\boldsymbol{\xi}_{\hat{q}}) \frac{\partial \ell_{\hat{h}}(\boldsymbol{\xi}_{\hat{q}})}{\partial r_i} \mathbf{u}_{\hat{h}} \\ & \quad + \sum_{\hat{q}=1}^N w_{\hat{q}} J_e(\boldsymbol{\xi}_{\hat{q}}) \sum_{j=1}^3 \frac{\partial \ell_{\hat{k}}(\boldsymbol{\xi}_{\hat{q}})}{\partial r_j} G_j(\boldsymbol{\xi}_{\hat{q}}) C^a \sum_{\hat{q}=1}^N \sum_{\nu=1}^L \boldsymbol{\psi}_{\nu;\hat{q}} + \underbrace{w_{\hat{k}} J_e(\boldsymbol{\xi}_{\hat{k}}) \mathbf{S}_{\hat{k}}}_{F_{\hat{k}\hat{k}}}, \\ & \partial_t \boldsymbol{\psi}_{\nu;\hat{q}} + \omega_{\nu} \boldsymbol{\psi}_{\nu;\hat{q}} = y_{\nu} \omega_{\nu} \sum_{\hat{h}=1}^N \sum_{i=1}^3 G_i^T(\boldsymbol{\xi}_{\hat{q}}) \frac{\partial \ell_{\hat{h}}(\boldsymbol{\xi}_{\hat{q}})}{\partial r_i} \mathbf{u}_{\hat{h}}. \end{aligned} \right. \quad (2.60)$$

2.2.4 Global system of the weak formulation

We can now introduce the global mass matrix \mathbf{M} which is diagonal by construction with the diagonal terms

$$M_{\hat{k}\hat{k}} = w_{\hat{k}} \rho_{\hat{k}}(\boldsymbol{\xi}_{\hat{k}}) J_e(\boldsymbol{\xi}_{\hat{k}}). \quad (2.61)$$

The weighted spatial derivative matrix \mathcal{D}^w and the spatial derivative matrix \mathcal{D} are respectively written as

$$\mathcal{D}_{\hat{k}\hat{q}}^w = \sum_{\hat{q}=1}^N w_{\hat{q}} J_e(\boldsymbol{\xi}_{\hat{q}}) \sum_{j=1}^3 \frac{\partial \ell_{\hat{k}}(\boldsymbol{\xi}_{\hat{q}})}{\partial r_j} G_j(\boldsymbol{\xi}_{\hat{q}}), \quad (2.62)$$

$$\mathcal{D}_{\hat{q}\hat{h}} = \sum_{\hat{h}=1}^N \sum_{i=1}^3 G_i^t(\boldsymbol{\xi}_{\hat{q}}) \frac{\partial \ell_{\hat{h}}(\boldsymbol{\xi}_{\hat{q}})}{\partial r_i}. \quad (2.63)$$

It should be noted that the operator \mathcal{D} estimates the spatial derivatives of a vector in the Cartesian space. The operator \mathcal{D}^w is equivalent to the spatial derivatives operator weighted by GLL weights $w_{\hat{q}}$ and the local Jacobian determinant $J_e(\boldsymbol{\xi}_{\hat{q}})$. The stiffness matrix is defined through the spatial derivative matrices as

$$\mathbf{K} = \mathcal{D}^w C \mathcal{D}. \quad (2.64)$$

The source term is embedded inside the matrix \mathbf{F} with

$$F_{\hat{k}\hat{k}} = w_{\hat{k}} J_e(\boldsymbol{\xi}_{\hat{k}}) \mathbf{S}_{\hat{k}}. \quad (2.65)$$

A similar development can be considered for the boundary term \mathbf{B} in Equation (2.53), which consists of the surface integral over boundaries of the numerical model

$$\mathbf{B} = \int_{\Gamma_{\Omega}} [I_D \ell_{\hat{k}}(\mathbf{x})] \boldsymbol{\tau}_B(\mathbf{x}) ds_{\mathbf{x}}. \quad (2.66)$$

It should be noted that the boundary term consists of the surface integral over different faces of a given 3D model volume. After discretization in the reference space, an example of the boundary term for the points located at one face of the model can be provided for instance as

$$B_{\hat{k}\hat{k}} = w_{k_1} w_{k_2} J_e^s(\boldsymbol{\xi}_{\hat{k}}) \boldsymbol{\tau}_B(\boldsymbol{\xi}_{\hat{k}}). \quad (2.67)$$

Under these discrete notations, the weak form of the second-order visco-elastic wave equation (2.23) of the displacement field \mathbf{u} can be written as

$$\begin{cases} \mathbf{M} \partial_{tt} \mathbf{u} = -\mathbf{K} \mathbf{u} + \mathcal{D}^w C^a \sum_{\nu=1}^L \psi_{\nu} + \mathbf{F} + \mathbf{B}, \\ \partial_t \psi_{\nu} + \omega_{\nu} \psi_{\nu} = \omega_{\nu} \gamma_{\nu} \mathcal{D} \mathbf{u} \end{cases} \quad (2.68)$$

In elastic medium, the attenuation contribution $\mathcal{D}^w C^a \sum_{\nu=1}^L \psi_{\nu}$ vanishes, leading to the second-order elastic system as developed by Komatitsch and Tromp (1999). The free-surface condition is naturally taken into account when the boundary terms appearing during the integration-by-parts step is ignored (Equation 2.40). The outgoing wavefield at the model boundaries is absorbed by a combination of sponge layers (Cerjan et al., 1985) and radiative boundary condition (Lysmer and Kuhlemeyer, 1969; Kouroussis et al., 2011), which are further described in Section (2.4).

2.3 Optimized and low-memory visco-elastic modeling kernel

As developed in the previous Section (2.2), the global mass matrix is diagonal by construction. Its inverse can thus be obtained without any extra cost. The visco-elastic system in Equation (2.68) without

Algorithm 1: Visco-elastic Newmark scheme

```

1  $\mathbf{u}^0 = 0; \mathbf{v}^0 = 0;$ 
2 for  $it = 1, \dots, nt$  do
3     Prediction phase
4      $\mathbf{u}^{it} = \mathbf{u}^{it-1} + \Delta t \mathbf{v}^{it-1} + \frac{\Delta t^2}{2} \mathbf{a}^{it-1};$ 
5      $\mathbf{v}^{it-1/2} = \mathbf{v}^{it-1} + \frac{\Delta t}{2} \mathbf{a}^{it-1};$ 
6     Resolution phase
7      $\boldsymbol{\varepsilon}^{it} = \mathcal{D} \mathbf{u}^{it};$ 
8      $\boldsymbol{\psi}_\nu^{it+1/2} = \gamma_\nu \boldsymbol{\psi}_\nu^{it-1/2} + \beta_\nu \boldsymbol{\varepsilon}^{it};$ 
9      $\boldsymbol{\sigma}^{it} = C \boldsymbol{\varepsilon}^{it} - C^a \sum_{\nu=1}^L \frac{\boldsymbol{\psi}_\nu^{it+1/2} + \boldsymbol{\psi}_\nu^{it-1/2}}{2} + \mathcal{T};$ 
10     $\mathbf{a}^{it} = -\mathbf{M}^{-1} (\mathcal{D}^w \boldsymbol{\sigma}^{it} + \mathbf{F}^{it});$ 
11    Correction phase
12     $\mathbf{v}^{it} = \mathbf{v}^{it-1/2} + \frac{\Delta t}{2} \mathbf{a}^{it};$ 
13 end
    
```

the boundary term can then be written as

$$\begin{cases} \partial_{tt} \mathbf{u} = -\mathbf{M}^{-1} (\mathbf{K} \mathbf{u} + \mathcal{D}^w C^a \sum_{\nu=1}^L \boldsymbol{\psi}_\nu + \mathbf{F}), \\ \partial_t \boldsymbol{\psi}_\nu + \omega_\nu \boldsymbol{\psi}_\nu = \omega_\nu y_\nu \mathcal{D} \mathbf{u} \end{cases} \quad (2.69)$$

A second-order explicit visco-elastic Newmark scheme is implemented for the time integration to compute the displacement field at each time-step as shown in Algorithm (1). This visco-elastic scheme is developed based on the Crank-Nicholson scheme (Robertsson et al., 1994) and elastic Newmark scheme described in Komatitsch (1997).

The SEM implementation used in our workflow is based on limited interpolation orders for test functions with $N = 4$ or 5 . It has been shown that these orders provide a good compromise between the numerical accuracy and the constraint on the Courant-Friedrichs-Lewy (CFL) stability condition (Komatitsch and Vilotte, 1998; Komatitsch and Tromp, 1999).

2.3.1 Optimized modeling kernel

At each time-step, the most computationally intensive part of the modeling kernel is the computation of the acceleration \mathbf{a} , which is equivalent to the second-order time-derivative of the displacement field $\partial_{tt} \mathbf{u}$. According to the definition of the stiffness matrix \mathbf{K} in Equation (2.64), the stiffness-displacement vector product can be factorized as

$$\mathbf{K} \mathbf{u} = \mathcal{D}^w C \mathcal{D} \mathbf{u}. \quad (2.70)$$

The right-hand-side of the first equation in the visco-elastic system (2.69) can then be written as

$$\mathbf{a} := \partial_{tt}\mathbf{u} = -\mathbf{M}^{-1} \left[\underbrace{\mathcal{D}^w \left(C \overbrace{\mathcal{D}\mathbf{u}}^{\boldsymbol{\varepsilon}} + C^a \sum_{\nu=1}^L \psi_\nu \right)}_{\boldsymbol{\sigma}} + \mathbf{F} \right]. \quad (2.71)$$

The numerical computation of the acceleration field is based on this factorization through a decomposition into three steps as described in the *resolution phase* in Algorithm (1):

- Line (7) - The estimation of the spatial derivatives of the displacement field $\mathcal{D}\mathbf{u}$, leading to the strain component $\boldsymbol{\varepsilon}$.
- Line (9) - The collection of the visco-elastic stress $\boldsymbol{\sigma}$, which involves two terms: The first term is related to the elastic rheology, computed through the product of the stiffness coefficients with the strain component $C\boldsymbol{\varepsilon}$. The second-term $C^a \sum_{\nu=1}^L \psi_\nu$ is only linked to the attenuation mechanism, its numerical discretization is described in Equation (2.75). If the wave-propagation system is excited by an external stress failure \mathcal{T} , it can also be included.
- Line (10) - The estimation of the weighted spatial derivatives of the stress field $\mathcal{D}^w\boldsymbol{\sigma}$.

Each spatial-derivative calculation is matrix-free. They are implemented following the Deville et al. (2002) approach, taking benefit from the tensorial properties of hexahedral elements, optimizing the loop vectorization and cache usage by smart management of fast and slow dimensions in tables, combined with manual loop unrolling for inner loops.

2.3.2 Low-memory visco-elastic modeling kernel

As described by Equation (2.71), the displacement field can be propagated under a second-order accurate time scheme. The first-order ODE governing the memory variables (Equation. 2.69)

$$\partial_t \psi_\nu + \omega_\nu \psi_\nu = \omega_\nu y_\nu \mathcal{D}\mathbf{u} \quad (2.72)$$

needs to be incorporated into the time-scheme without degrading this second-order accuracy. To do so, the memory variables ψ_ν are updated at half time-steps compared to the displacement field \mathbf{u} . To be more precise, the ODE (2.72) is discretized following the centered scheme

$$\frac{\psi_\nu^{it+1/2} - \psi_\nu^{it-1/2}}{\Delta t} + \omega_\nu \frac{\psi_\nu^{it+1/2} + \psi_\nu^{it-1/2}}{2} + \mathcal{O}(\Delta t^2) = \omega_\nu y_\nu \mathcal{D}\mathbf{u}^{it}, \quad (2.73)$$

in which it indicates the current time-step and Δt is the time-stepping. Therefore, the memory variables at the time step $(it + 1/2)$ can be explicitly updated by

$$\psi_\nu^{it+1/2} \approx \underbrace{\frac{2 - \omega_\nu \Delta t}{2 + \omega_\nu \Delta t}}_{\gamma_\nu} \psi_\nu^{it-1/2} + \omega_\nu y_\nu \underbrace{\frac{2\Delta t}{2 + \omega_\nu \Delta t}}_{\beta_\nu} \mathcal{D}\mathbf{u}^{it} + \mathcal{O}(\Delta t^2), \quad (2.74)$$

as indicated in Line (8) in the visco-elastic Newmark scheme in Algorithm 1 for $\nu = 1, \dots, L$. Following this expression, the memory variables at the next step is explicitly computed from the previous step and the already known displacement field. Therefore, we only need to store a single entity of the ψ_ν fields,

i.e. $L \times 6$ tables of memory variables, per time-step. The attenuation contribution in the stress field $C^a \sum_{\nu=1}^L \psi_{\nu}$ can be estimated following the same strategy

$$\boldsymbol{\sigma}^{it} = C\boldsymbol{\varepsilon}^{it} - C^a \sum_{\nu=1}^L \psi_{\nu}^{it} \approx C\boldsymbol{\varepsilon}^{it} - C^a \sum_{\nu=1}^L \frac{\psi_{\nu}^{it+1/2} + \psi_{\nu}^{it-1/2}}{2} + \mathcal{O}(\Delta t^2). \quad (2.75)$$

Equation (2.75) is considered for updating the stress field in Line (9) in Algorithm 1. In addition, as highlighted by Equations (2.17) and (2.19), we only store L scalar anelastic coefficients y_{ν} to reduce the memory requirement related to SLS attenuation mechanism (Yang et al., 2016a). The combination of these two strategies allows to reduce the memory footprint of visco-elastic simulation while preserving the accuracy.

The simulation quality of our approach is illustrated through the LOH benchmarks (Day et al., 2003) as shown in Section (2.6), where we obtain an excellent agreement with both elastic and visco-elastic semi-analytical solutions on layered medium, calculated from the boundary integral method (Coutant, 1989).

2.4 Boundary conditions and source implementation

In practical applications, the seismic waves are propagated from a single point source (e.g. vibroseis or explosive source). A couple of Dirichlet and Neumann conditions are then considered for initial condition in time: Zero displacement field and its temporal derivative, the velocity field, at the starting time $\mathbf{u}(t=0) = 0$ and $\mathbf{v}(t=0) = 0$.

The numerical simulation of the wave propagation can only be performed in a finite volume. Specifying boundary conditions are thus required to attenuate the outgoing wavefield. As mentioned in Section (2.2), the free-surface condition is naturally considered by the SEM weak form development of the wave equation. This section then only focus on the wavefield absorbing at the boundaries of the model and the source injection into the SEM mesh.

2.4.1 Boundary conditions

The absorbing conditions are required at the boundaries of the numerical model to stimulate the simulation in an infinite system. The outgoing wavefield is absorbed by a combination of sponge layers (Cerjan et al., 1985) and radiative boundary condition (Lysmer and Kuhlemeyer, 1969) due to its efficiency even for anisotropic materials.

The sponge consists of several layers of elements at the boundaries of the physical medium as illustrated in Figure 2.4. These layers gradually reduce the amplitude of the wavefield coming out from the physical domain. At the end of the calculation at each time-step, the wavefield propagating inside the sponge layers are damped by the factor

$$G(c, x) = e^{-[c(x/L)]^2}, \quad (2.76)$$

where c is a damping coefficient, L is the sponge thickness and x is the distance from the physical domain as described in Figure 2.5. At the interface between the physical domain and the sponge layers, the factor $G(c, x=0)$ is equal to 1, implying no amplitude attenuation inside or at the boundaries of the physical domain.

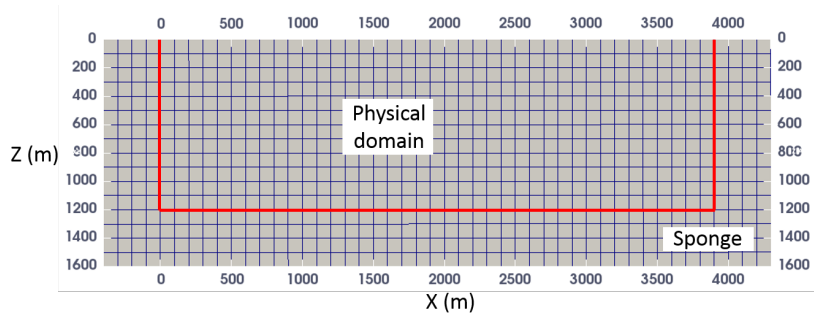


Figure 2.4: An example of the mesh design where sponge layers contain 4 elements.

In standard implementation, the damping coefficient is normally chosen as $c = 0.3$, which provides the factor $G(c = 0.3, x = 0) = 0.92$ at the outer boundaries of the sponge layers. In practice, the sponge thickness L should be around 4 times larger than the wavelength λ that we want to attenuate. The method appears to be robust and can be applied to different wave modeling problems, even when the anisotropy is considered. The effectiveness of this boundary condition does not decrease for shallow angles of incident. However, when the dataset has low frequency component, a thick sponge design is required for an efficient wavefield absorbing at the boundaries.

To improve the accuracy of the sponge technique, we combine it with a first-order radiative absorbing condition. The damping coefficient inside Equation (2.76) has been reduced to $c = 0.06$ to weaken the amplitude attenuation inside the sponge. A boundary traction τ_B is injected into the boundary term of the wave equation (2.53), at the end of the sponge, to further attenuate the outgoing wavefield. We use a classic formulation proposed by Lysmer and Kuhlemeyer (1969); Kouroussis et al. (2011)

$$\tau_B = \rho V_p \mathbf{v}_n + \rho V_s \mathbf{v}_t, \quad (2.77)$$

where ρ is the density, V_p and V_s are compressional and shear wave speeds, respectively. This expression also requires normal and tangential particles velocities \mathbf{v}_n and \mathbf{v}_t on the domain borders. It should be noted that the radiative boundary condition can be applied at no extra-cost and it works for all frequency range. However, one of the main drawbacks of this approach is that the wavefield absorbing efficiency is incident-angle dependent: The method is efficient for normal incident waves but less for other angles. By combining the sponge technique and the radiative boundary condition, both short and long wavelengths of the wavefield at the boundary can be efficiently attenuated, also at a wider range of incident angles. Figure 2.10 presents an example of a forward modeling in the LOH1 benchmark described in Section (2.6), in which the boundary effect is invisible.

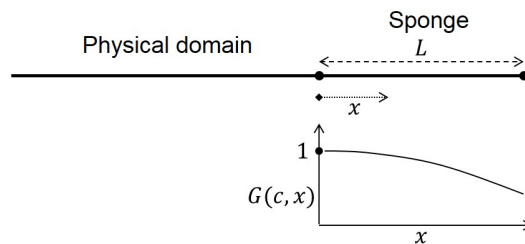


Figure 2.5: Amplitude reduction inside the sponge layers.

Aside from the described combination of the sponge and radiative boundary condition techniques, the perfectly matched layer (PML) absorbing boundary condition has been proven to be very efficient for elastic wave equation to absorb both body and surface waves. The main issues of the standard implementations are the incident-angle dependent efficiency and intrinsic instability for anisotropic materials (Bérenger, 1994). The unsplit convolution technique has been proposed to improve the absorbing efficiency at grazing incidences. However, in anisotropic media, it is well known that PML exhibits numerical instabilities: The outgoing wavefield can be amplified instead of being decreased. Correction terms can be considered to stabilize the simulation (Komatitsch and Martin, 2007; Martin et al., 2008, 2010). However, we have not investigated on these approaches yet and we keep a simple sponge and absorbing boundary approach.

2.4.2 Source implementation

In land seismic application, the source can be either vibroseis or explosive. The vibroseis source is equivalent to an external directional point force $\mathbf{f}(\mathbf{x}, t)$ in Equation (2.1). The explosive source is directly linked to moment tensor source, which can be applied through external stress failure $\mathcal{T}(\mathbf{x}, t)$ in Equation (2.1). In a surface seismic acquisition, the vibroseis source is applied at the free-surface, i.e. at the boundary of the simulation model, while the explosive source is buried at several-meters depth.

2.4.2.1 Point-force source

Each point source is consider as a product of a Dirac delta function in physical space at the source position \mathbf{x}_s with a source-time function $\mathbf{f}(t)$ as

$$\mathbf{f}(\mathbf{x}, t) = \delta(\mathbf{x} - \mathbf{x}_s)\mathbf{f}(t). \quad (2.78)$$

We choose to inject the source directly into the displacement field. Following the same workflow as the weak form development in Section (2.2), the integral form of the source term is

$$\int_{\Omega} \ell_{\hat{k}}(\mathbf{x})\delta(\mathbf{x} - \mathbf{x}_s)\mathbf{f}(t)d\mathbf{x}. \quad (2.79)$$

It should be noted that when moving from the physical space to the reference space, the delta function $\delta(\mathbf{x} - \mathbf{x}_s)$ should be normalized by the Jacobian determinant

$$\delta(\mathbf{x} - \mathbf{x}_s) = \frac{1}{J_e(\boldsymbol{\xi})}\delta(\boldsymbol{\xi} - \boldsymbol{\xi}_s), \quad (2.80)$$

leading to the source term in the reference space as

$$\int_{\Omega} \ell_{\hat{k}}(\boldsymbol{\xi})\delta(\boldsymbol{\xi} - \boldsymbol{\xi}_s)\mathbf{f}(t)d\boldsymbol{\xi}, \quad (2.81)$$

where $\boldsymbol{\xi}_s$ is the associated position of the source point in the reference space. The diagonal terms of the source matrix Equation (2.65) can then be written as

$$F_{\hat{k}\hat{k}} = \ell_{\hat{k}}(\boldsymbol{\xi}_s)\mathbf{f}(t). \quad (2.82)$$

The implementation can be simply interpreted as a spatial interpolation of the source point position over the GLL point, scaled by the source-time function $\mathbf{f}(t)$.

2.4.2.2 Moment tensor source

In petroleum geophysics, an ideal explosive source is described by a diagonal stress failure

$$\mathcal{T}_{11}(\mathbf{x}, t) = \mathcal{T}_{33}(\mathbf{x}, t) = \mathcal{T}_{33}(\mathbf{x}, t). \quad (2.83)$$

The stress failure can be injected directly into the stress calculation in Line (9) in Algorithm 1.

We implement the explosive source through the *moment-tensor* source \mathcal{M}_{ij} , which is a physical notation widely used in seismology. The moment-tensor source $\mathcal{M}(\mathbf{x}, t)$ is related to the stress failure $\mathcal{T}(\mathbf{x}, t)$ through a volumetric integral over the physical domain Ω (Komatitsch and Tromp, 1999; Aki and Richards, 2002)

$$\mathcal{M}(\mathbf{x}, t) = \int_{\Omega} \mathcal{T}(\mathbf{x}, t) d\mathbf{x}. \quad (2.84)$$

The application of an external point-moment at the position \mathbf{x}_s can be described as

$$\mathcal{M}(\mathbf{x}, t) = \mathcal{M}(t)\delta(\mathbf{x} - \mathbf{x}_s). \quad (2.85)$$

By consequence, the stress failure can be written as

$$\mathcal{T}(\mathbf{x}, t) = \mathcal{T}(t)\delta(\mathbf{x} - \mathbf{x}_s), \quad (2.86)$$

even though the notion of a point-stress source does not exist physically. Therefore, we seek a meaningful relationship between the stress time-function $\mathcal{T}(t)$ and the moment time-function $\mathcal{M}(t)$. By substituting the definitions of the external moment and stress failure into the relationship (2.84) we obtain

$$\mathcal{M}(t)\delta(\mathbf{x} - \mathbf{x}_s) = \mathcal{T}(t) \int_{\Omega} \delta(\mathbf{x} - \mathbf{x}_s) d\mathbf{x}, \quad (2.87)$$

$$\approx \mathcal{T}(t) \sum_{\hat{k}=1}^N \ell_{\hat{k}}(\mathbf{x}_s) \ell_{\hat{k}}(\mathbf{x}). \quad (2.88)$$

The spatial integral of the Dirac function $\delta(\mathbf{x} - \mathbf{x}_s)$ in Equation (2.87) provides an unitary amplitude at the source position \mathbf{x}_s , which can be approximated by the spatial interpolation over the GLL points and Lagrange basis function as indicated in Equation (2.88). By considering a volumetric integral over the physical medium Ω , the stress time-function can now be computed as

$$\mathcal{T}(t) = \frac{\mathcal{M}(t)}{\int_{\Omega} \sum_{\hat{k}=1}^N \ell_{\hat{k}}(\mathbf{x}_s) \ell_{\hat{k}}(\mathbf{x}) d\mathbf{x}} = \frac{\mathcal{M}(t)}{\sum_{\hat{h}} w_{\hat{h}} \ell_{\hat{h}}(\boldsymbol{\xi}_s) \ell_{\hat{h}}(\boldsymbol{\xi}_{\hat{h}}) J_e(\boldsymbol{\xi}_{\hat{h}})}. \quad (2.89)$$

This normalization term is important to achieve a “true amplitude” seismogram estimation. The amplitude consistency of the computed seismogram has been verified for the same physical setting on different mesh designs and different relative source positions inside a element.

2.5 Mesh design

In a completely unstructured FE mesh, a relatively expensive set-up stage is required at the beginning of any simulation, where the mesh is generated and the associated indices or any neighborhood look-up tables are pre-computed. We simplify the mesh design by considering a Cartesian-based deformed

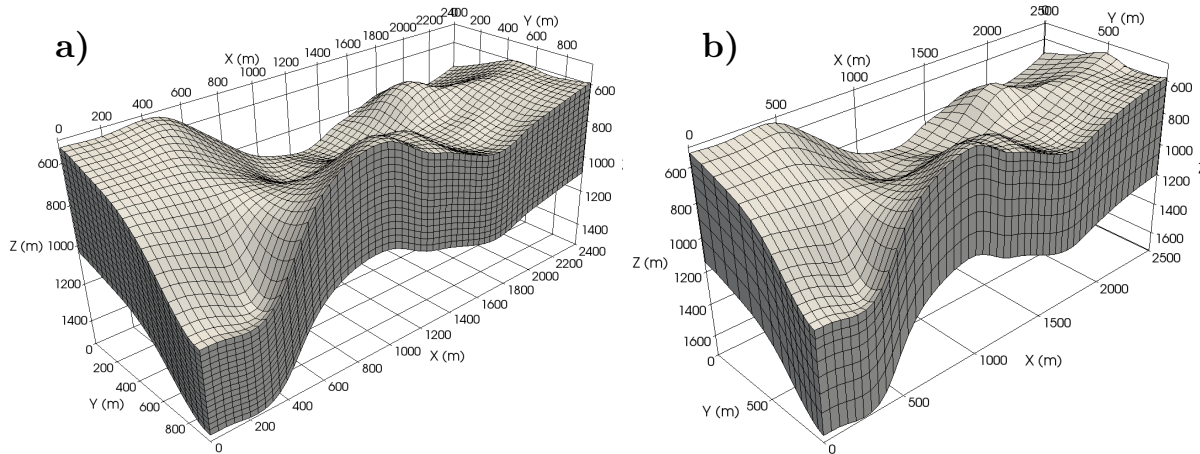


Figure 2.6: 3D mesh generation on a realistic example (4 km in z direction, 8 km in x direction and 2 km in y direction): (a) A subset of constant element-size mesh with $dz = dx = dy = 40$ m. (b) A subset of variable element-size mesh, where the number of elements is reduced 3.03 times in z direction, 1.57 times in x direction and 1.47 times in y direction: $dz \in [49 - 142]$ m, $dx \in [49 - 139]$ m, $dy \in [49 - 100]$ m.

mesh to combine the accurate representation of the topography allowed by the FE meshes and the convenience of an implementation on a FD grid (Cupillard et al., 2011; Gokhberg and Fichtner, 2016). The spatial position of each element can be obtained directly from its x, y, z indices, without using any neighborhood look-up tables or searching steps over the global mesh. The topography variation is represented by vertically deformed elements as shown in Figure 2.6(a). The numbers of elements in x, y and z directions are constant, although the size of elements can vary. For the basis functions at interpolation order $N = 4$ or 5 , SEM allows to accurately model elastic waves propagation with around 6 GLL nodes per shortest wavelength (Komatitsch, 1997). This condition is referred as the *volume condition*.

2.5.1 Cartesian-based mesh with variable element-size

In most geophysical targets, the velocity is spatially varying, leading to spatially varying wavelengths. The element-size should thus be adapted to the variation of the local shortest wavelength. Figure 2.6(b) shows an example of such a flexible mesh design for a 3D target, where the element-size varies from 49 m to 142 m in three directions. The mesh is built upon the condition that 6 GLL points per shortest wavelength are required. Under the same constraint, Figure 2.6(a) illustrates the constant element-size mesh, where the element-size is 40 m in each direction. In this cross-section, by using a variable element-size mesh, the number of elements is reduced by a factor of 3.03 times in z direction, 1.57 times in x direction and 1.47 times in y direction: almost a factor 7 in total, while keeping the Cartesian-based topology of the mesh. The reduction of the number of elements systematically decreases the numerical cost by the same factor, or even more if it induces the relaxation of the CFL stability condition of the explicit time-marching scheme.

2.5.2 Complex topography representation

When considering significant topography variations, hexahedral elements can be vertically deformed as illustrated in Figure 2.6. As described in Section (2.2.1.1), the shape of each element can be defined by a set of $(n + 1)$ *control points* in each direction, leading to a total $n_a = (n + 1)^3$ control points and associated *shape functions* in 3D geometry. For instance, $n = 1$ means only eight corners of the elements are used as control points, while $n = 2$ considers also the mid-size and center nodes.

The spatial position of GLL points inside the element is computed from these control points and shape functions through

$$\mathbf{x}(\xi, \eta, \zeta) = \sum_{k_1=1}^{n+1} \sum_{k_2=1}^{n+1} \sum_{k_3=1}^{n+1} \ell_{\hat{k}}(\xi, \eta, \zeta) \mathbf{x}_{\hat{k}}, \quad (2.90)$$

where \hat{k} again stands for the triple indexes k_1, k_2, k_3 . The associated shape functions are a triple product of Lagrange polynomials of degree n : $\ell_{\hat{k}}(\xi, \eta, \zeta)$. Since shape functions are products of Lagrange polynomials, their partial derivatives are analytically determined even for high order n , implying that the Jacobian matrix $\mathbf{J}(\boldsymbol{\xi})$ in Equation (2.28) is also analytically determined. The number of control points and shape functions in each direction $(n + 1)$ is not related to the interpolation order N of the test functions needed for solving the wave equation PDE. However, we normally consider $n \leq N$ to avoid any possible artifact coming from the geometry representation.

Figure 2.7(a) and (b) show the relative position of GLL points inside a deformed elements under P_1 (using eight corners of each element as control points and linear shape functions) and P_4 representation ($(4 + 1)^3$ GLL points are used as control points and order-4 shape functions). The P_1 representation of the surface cannot honor sharp spatial variation of the free surface, as shown in Figure 2.7(c), which might affect the accuracy of the simulation due to the interaction between elastic waves and the complex surface. Decreasing the element size is one way for following the rapid variation of the topography, namely the *surface condition*: At least two spatial sampling points per shortest topographic wavelength are required to honor the interaction at the surface. This criterion might be stricter than the *volume condition*, and would significantly increase the computational cost. The surface condition limitation can be overcome by using higher-order P_n representation, such as $n = 4$ (Figure 2.7d) where all GLL points inside the element are considered as control points.

Figure 2.7(d) further confirms that the P_4 representation provides a better description of the complex topography since $(4 + 1)^3$ GLL points are used in each element to capture the topography map, instead of $(1 + 1)^3$ points for the P_1 case. This thus leads to a more accurate spatial positioning of GLL points inside the element. To evaluate the influence of the topography representation on the accuracy of the wavefield simulation, we use a homogeneous model with complex surface extracted from the SEAM Phase II Foothills benchmark as shown in Figure 2.8(a). The medium is meshed at 100 m element-size, under P_1 and P_4 topography representations. A single-valued envelop misfit estimated at each receiver location (Kristeková et al., 2006) is used to compare the observed seismograms with the reference solution, computed from a 25 m mesh under P_1 topography representation. The comparison is shown in Figure 2.8(b): For the same number of elements, thus identical numerical cost, the P_4 surface representation provides a more accurate estimation of the wavefield. Moreover, the simulation error of the P_1 (100 m) mesh is accumulated with offset, which might damage the information at far offsets, thus the velocity estimation of the near surface during the inversion. Trade-offs between topography representation and velocity parameters in acoustic land-FWI have been investigated numerically by Huiskes et al. (2017). When considering surface-waves, the influence of topography representation on elastic FWI is expected to be even more important.

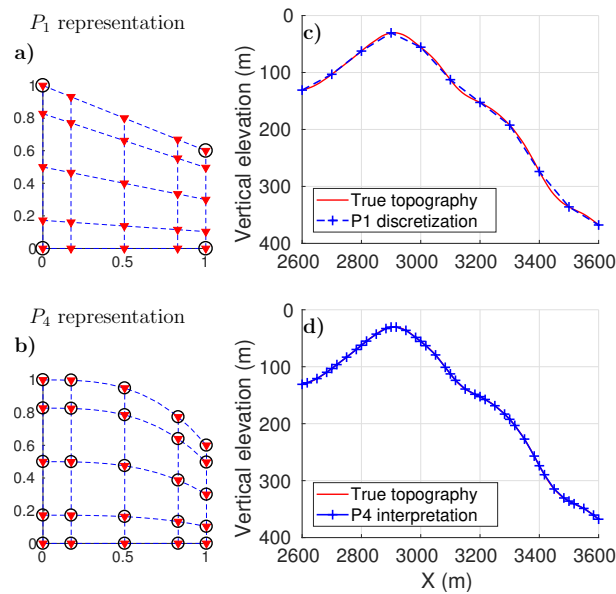


Figure 2.7: Left - Position of GLL points (red triangles) of a cross-section inside a 3D element for interpolation order $N = 4$ using (a) the eight corners of each element and linear shape function and (b) the $(4 + 1)^3$ GLL control points associated with P_4 shape functions. The control points are highlighted by black circles. Right - The associated with topography description (for a cross-section extracted from SEAM II Foothills model) using (c) P_1 and (d) P_4 representation. The position of control points are marked by the sign '+', showing that the P_4 representation provides a better representation of the complex topography. The element size is 100 m for both case.

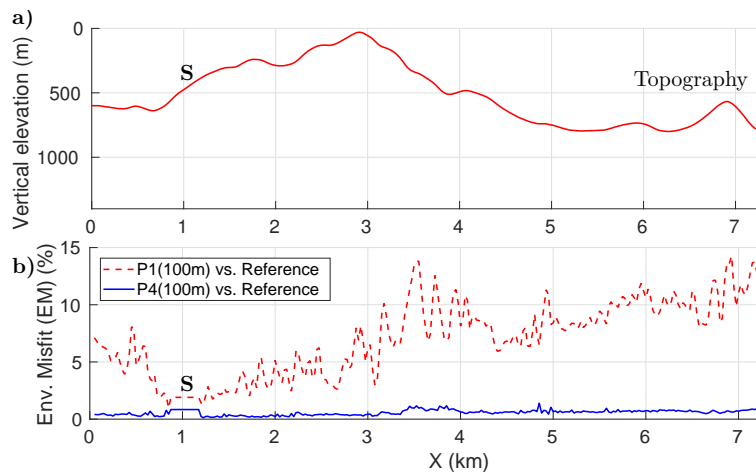


Figure 2.8: (a) A complex topography extracted from the SEAM II Foothills model. (b) Envelop misfit comparison: Red - Between P_1 (100 m) mesh with the reference solution, Blue - Between P_4 (100 m) mesh with the reference solution, which shows significant improvement in the numerical accuracy. The reference solution is obtained from the simulation on P_1 (25 m) mesh.

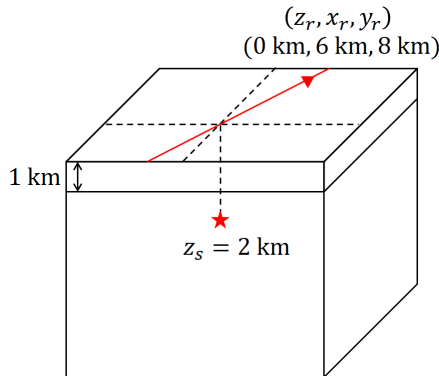


Figure 2.9: LOH layered model. The source is placed in the lower half-space ($z_s = 2$ km, $x_s = y_s = 15$ km) and the receiver is placed on the free surface ($z_r = 0$ km, $x_r = 21$ km, $y_r = 23$ km).

It should be noticed that only the Jacobian matrix $\mathbf{J}(\boldsymbol{\xi})$ and the Jacobian determinant $J_e(\boldsymbol{\xi})$ associated with the mapping from the reference space to the physical space are required for the wave propagation. Using high-order geometry representation (\mathbf{P}_n) therefore only affects the mesh building step and the computation of the Jacobian terms, which are computed only once in the FWI workflow. Moreover, similar optimized strategy as the matrix-free spatial-derivative estimation described in Section (2.3) can be applied for the computation of the volumetric Jacobian. The computational cost of the wavefield modeling is thus unaltered, while the simulation accuracy related to the complex wave-phenomena at the free-surface is significantly improved.

The combination of the variable element-size mesh design with high-order topography representation makes possible using as coarse elements as possible without degrading the medium geometries representation, then providing a good balance between computational cost and numerical accuracy. We are aware that such high-order shape functions might affect the SEM convergence properties, because the polynomial order inside the integrals in Equation (2.57) becomes superior than $2n + 1$. However, our numerical experiments in Figure 2.8 confirm a rather low simulation error compared with a reference solution. Also, it is widely known that the convergence characteristics of SEM are already lost when solving the wave equation in a heterogeneous medium inside each element, which is the cases for a majority practical applications.

2.6 Comparison with semi-analytical solutions - LOH benchmarks

The simulation accuracy of our elastic and visco-elastic forward modeling, embedded inside the SEM46 code, is illustrated through the comparison with the boundary integral method (Coutant, 1989). Following this approach, the AXITRA open-source package offers highly accurate semi-analytical solution on layered medium. For visco-elastic solution, only the Kjartansson's constant Q -model is originally available in AXITRA (Kjartansson, 1979). We thus introduce the approximation by 3 pre-defined SLS mechanisms over the modeling frequency range of the benchmark into AXITRA.

The comparison is performed on the LOH benchmark, which has a layered material model, where the top 1000 m have different properties than the rest of the domain (cf. Table 2.1 and Figure 2.9). The problem is driven by a single point moment source $\mathcal{M}_{xy} = 10^{18}$ N.m, positioned in the lower half-space at depth $z_s = 2$ km. The selected receiver is located at the free-surface with the distance

	V_p [m/s]	V_s [m/s]	ρ [m/s]	Q_p	Q_s
Top layer	4000	2000	2600	120	40
Half space	6000	3464	2700	155.9	69.3

Table 2.1: Medium properties of the LOH benchmark.

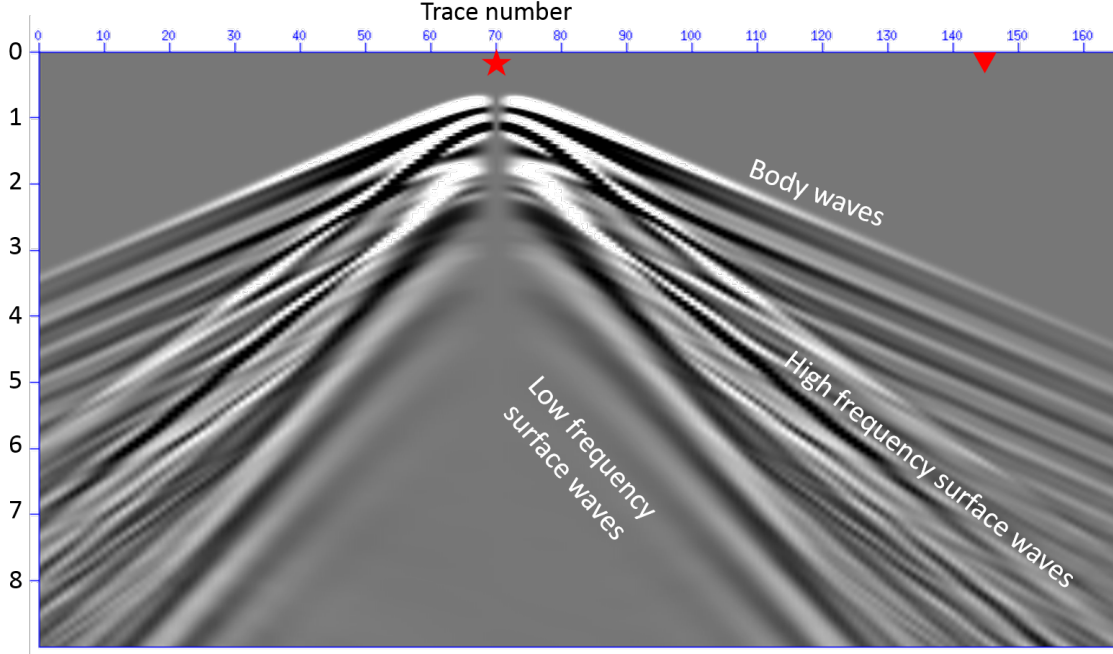


Figure 2.10: Wavefield recorded in LOH1 benchmark, showing dispersive surface waves and multiple body-wave-reflections between the free surface and the sharp contrast. The interested receiver position (z_r, x_r, y_r) is highlighted by red triangular. The wavefield at the boundary is efficiently attenuated thanks to the combination of the sponge layers and the radiative boundary condition.

about 10 km from the source ($\Delta x_{sr} = 6$ km, $\Delta y_{sr} = 8$ km). We also deploy a surface receiver array across the source and selected receiver positions as indicated by the red line in Figure 2.9.

2.6.1 LOH1 comparison

LOH1 solutions come from an elastic simulation with a Gaussian source time function centered at 2.65 Hz. The distance source-receiver is equivalent to about 14 times the dominant surface wavelength (703 m). SEM mesh consists of a set of regular cubic elements, with element size calculated from the volume condition: $dx = dy = 160$ m. The element-size in the vertical direction is a factor of the thickness of the top-layer, $dz = 100$ m, to better represent the sharp interface. A sponge layer of 8 elements is considered at each model boundary, equivalent to 1280 m thickness in x and y directions and 800 m at the bottom of the model.

Figure 2.10 shows the recorded wavefield at the free-surface with multiple body-wave reflections between the free-surface and the sharp-contrast at 1 km depth. Complex and dispersive surface energy can be clearly identified in the recorded wavefield. Moreover, the residual wavefield at the boundary is insignificant thanks to the efficiency of our wavefield absorbing strategy. Figure 2.11 shows excellent

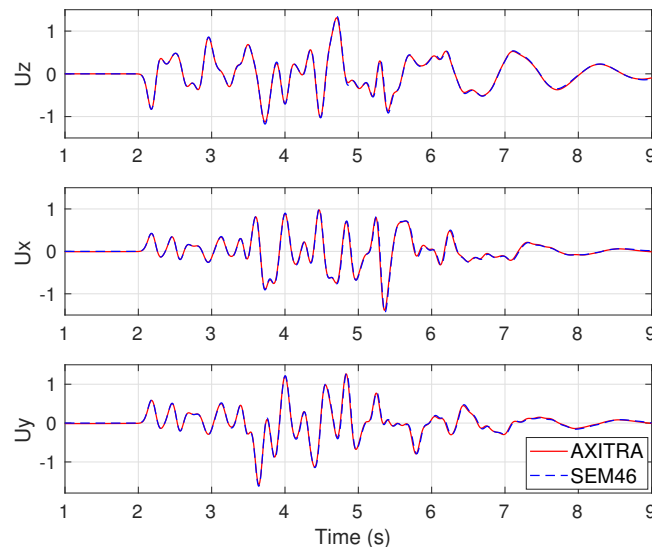


Figure 2.11: Excellent agreement between the elastic SEM46 simulation result and the semi-analytic solution obtained from AXITRA on LOH1 configuration. Blue - SEM46 solution. Red - AXITRA solution.

agreement between the elastic SEM46 simulation result and the semi-analytic solution obtained from AXITRA. Since SEM is a continuous FE method, it has some intrinsic limitation in the representation of the sharp contrast. We thus do not push the comparison further with any time-frequency analysis.

2.6.2 LOH3 comparison

LOH3 solutions come from a visco-elastic simulation with a Gaussian source time function at 3.18 Hz center frequency. The medium is meshed with regular cubic elements $dx = dy = 160$ m and $dz = 100$ m to better represent the sharp-contrast at $z = 1$ km. SEM46 uses 3 SLSs to approximate the constant quality factors. We perform two separated visco-elastic comparisons in the LOH3 setting:

- (1) Similar SLS attenuation mechanism, different wavefield computation method: By using the same SLS mechanisms (similar dimensionless anelastic coefficients y_ν and reference frequencies ω_ν), we modify the complex relaxed P- and S-waves speeds inside the AXITRA following Equation (2.105) in Appendix (2.B). It should be noted that the computation kernel inside AXITRA is kept untouched, only the estimation of the complex wave speeds are modified. The additional SLS complex velocities implementation provides an excellent agreement between the SEM46 and modified-AXITRA solutions in Figure 2.12(a).
- (2) Difference attenuation mechanism, different wavefield computation method: A very good fit between SEM46 simulation and AXITRA solution obtained with Kjartansson attenuation model is also obtained in Figure 2.12(b). In this comparison, a reference frequency at 100 Hz is considered for Kjartansson model as described in Appendix (2.B).

One should notice the use of a “tuning” frequency for matching the solutions obtained by different attenuation models. This comes from the fact that SLS model has a finite phase velocity at infinite

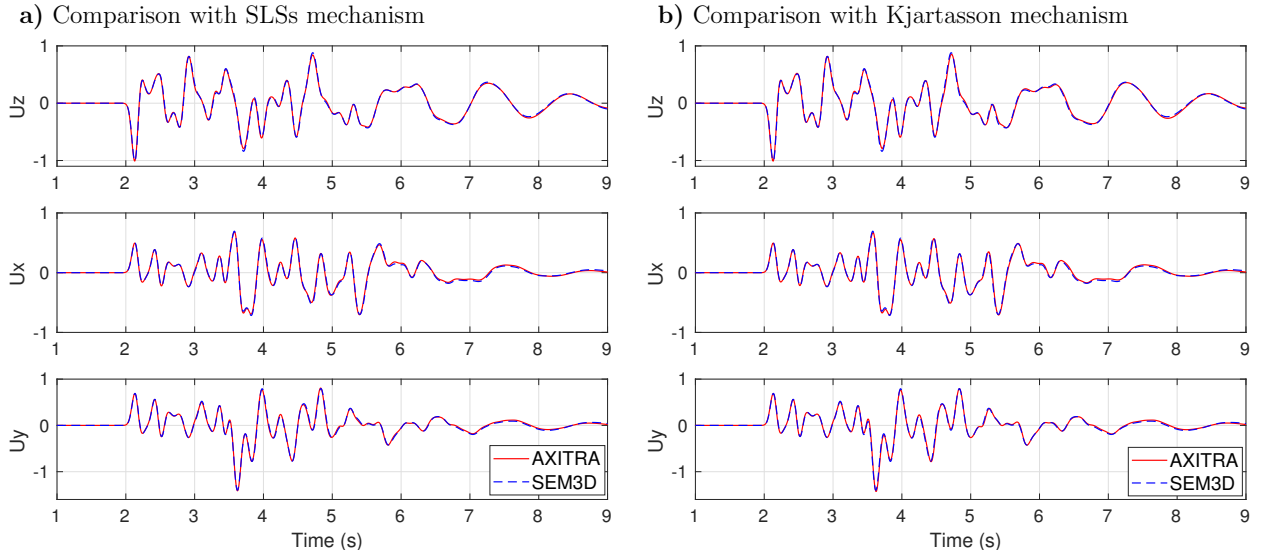


Figure 2.12: Excellent agreement between the visco-elastic SEM46 simulation result and the semi-analytic solution obtained from AXITRA on LOH3 configuration: (a) Using 3 SLSs mechanism, (b) Kjartansson model with 100 Hz reference frequency. Blue - SEM46 solution. Red - AXITRA solution.

frequency ($V_\infty < \infty$), while Kjartansson's constant- Q model assumes infinite velocity at infinite frequency $V_\infty = \infty$ (Wang, 2009). Therefore, a finite reference frequency at 100 Hz is required to make the phase velocities in SLS and Kjartansson models compatible. It should be remarked that different attenuation models provide different expressions of the medium parameters, leading to difficulties for assessing the validation of the attenuation propagation.

Aside from the comparison with the semi-analytical solutions, our simulation also shows excellent waveform and true-amplitude agreement compared with a time-domain FD implementation in SW4 V1.1 (Sjögreen and Petersson, 2012). Their implementation also describes the attenuation effect by constant- Q quality factors, approximated by a set of SLSs over the considered frequency band. However, the so-called memory variables act on the displacement fields, which is different compared with our strategy described in Section (2.1.2.2).

2.7 Conclusion

This chapter provides a review of the 3D modeling in visco-elastic medium, using SLS mechanisms. We separate purposely the elasticity with the attenuation mechanisms, which offers an easy and explicit coupling between the anisotropic elasticity and isotropic attenuation. The attenuation parameters are explicitly incorporated in the wave equations, and therefore can be naturally reconstructed in the FWI framework as discussed in the next Chapter (3).

We present a step-by-step weak-formulation development for the visco-elastic system which has a similar form as the second-order elastic weak form presented by Komatitsch and Tromp (1999). A visco-elastic Newmark scheme is then considered to propagate the system in time. The first-order ODE governing the memory variables has been integrated into this time-marching scheme following a cheap-memory implementation strategy, while still preserving the second-order accuracy of the time scheme.

The outgoing wavefield at the model boundaries is absorbed by a combination of sponge layers (Cerjan et al., 1985) and radiative boundary condition (Lysmer and Kuhlemeyer, 1969; Clayton and Engquist, 1977; Kouroussis et al., 2011). The combination efficiently attenuates both short and long wavelength at a wide range of incident angles.

The simulation accuracy is ensured by the use of a Cartesian-based deformed mesh with high-order geometry interpolation to capture rough topography variations. This high-order geometry representation is shown to improve the simulation quality while keeping a coarse mesh design (?). The element-size can be adapted to the variation of the local shortest wavelength to reduce the numerical cost, especially for inversion applications. The comparison with semi-analytical solutions computed from the boundary integral method confirms the accuracy of our elastic and visco-elastic implementations (Coutant, 1989).

APPENDICES

2.A Stress and strain relationship in the time domain

According to Equation (2.1), in anelastic medium, the stress σ_{ij} is the convolution of the relaxation rate M_{ijkl} with the strain ε_{kl} as

$$\sigma_{ij}(\mathbf{x}, t) = M_{ijkl}(\mathbf{x}, t) *_t \varepsilon_{kl}(\mathbf{x}, t) + \mathcal{T}_{ij}(\mathbf{x}, t). \quad (2.91)$$

In the frequency-domain, this relationship becomes

$$\hat{\sigma}_{ij}(\mathbf{x}, \omega) = \widehat{M}_{ijkl}(\mathbf{x}, \omega) \widehat{\varepsilon}_{kl}(\mathbf{x}, \omega) + \widehat{\mathcal{T}}_{ij}(\mathbf{x}, \omega). \quad (2.92)$$

Under the assumption that L SLSs at specific reference frequencies ω_ν are used to fit a constant Q parameter over the frequency band $[\omega_{\min}, \omega_{\max}]$, the associated expression for the complex relaxation rate $\widehat{M}_{ijkl}(\mathbf{x}, \omega)$ is represented in Equation (2.19). Equation (2.92) now becomes

$$\hat{\sigma}_{ij}(\mathbf{x}, \omega) = \left[c_{ijkl}(\mathbf{x}) - c_{ijkl}(\mathbf{x}) Q_{ijkl}^{-1}(\mathbf{x}) \sum_{\nu=1}^L y_\nu \frac{\omega_\nu}{\omega_\nu + i\omega} \right] \widehat{\varepsilon}_{kl}(\mathbf{x}, \omega) + \widehat{\mathcal{T}}_{ij}(\mathbf{x}, \omega). \quad (2.93)$$

By using the definition of the attenuative stiffness coefficients c_{ijkl}^a given in Equation (2.20), we can rewrite Equation (2.93) as

$$\hat{\sigma}_{ij}(\mathbf{x}, \omega) = c_{ijkl}(\mathbf{x}) \widehat{\varepsilon}_{kl}(\mathbf{x}, \omega) - c_{ijkl}^a(\mathbf{x}) \sum_{\nu=1}^L y_\nu \frac{\omega_\nu}{\omega_\nu + i\omega} \widehat{\varepsilon}_{kl}(\mathbf{x}, \omega) + \widehat{\mathcal{T}}_{ij}(\mathbf{x}, \omega). \quad (2.94)$$

Introducing the complex memory variable $\widehat{\psi}_{\nu;kl}(\mathbf{x}, \omega)$ such that

$$\widehat{\psi}_{\nu;kl}(\mathbf{x}, \omega) = y_\nu \frac{\omega_\nu}{\omega_\nu + i\omega} \widehat{\varepsilon}_{kl}(\mathbf{x}, \omega), \quad (2.95)$$

leads to the simplification of Equation (2.94) into

$$\hat{\sigma}_{ij}(\mathbf{x}, \omega) = c_{ijkl}(\mathbf{x}) \widehat{\varepsilon}_{kl}(\mathbf{x}, \omega) - c_{ijkl}^a(\mathbf{x}) \sum_{\nu=1}^L \widehat{\psi}_{\nu;kl}(\mathbf{x}, \omega) + \widehat{\mathcal{T}}_{ij}(\mathbf{x}, \omega). \quad (2.96)$$

In time domain, Equation (2.96) is equivalent to the relationship between stress and strain mentioned in Equation (2.21).

Equation (2.95) can be developed as

$$(i\omega + \omega_\nu) \widehat{\psi}_{\nu;kl}(\mathbf{x}, \omega) = \omega_\nu y_\nu \widehat{\varepsilon}_{kl}(\mathbf{x}, \omega). \quad (2.97)$$

By applying the invers Fourier transform into this equation, we obtain

$$\partial_t \psi_{\nu;kl}(\mathbf{x}, t) + \omega_\nu \psi_{\nu;kl}(\mathbf{x}, t) = \omega_\nu y_\nu \varepsilon_{kl}(\mathbf{x}, t), \quad (2.98)$$

which is the ODE governing the propagation of the memory variables, mentioned in Equation (2.22).

2.B Complex P and S wave velocities in AXITRA

Kjartansson model

The Kjartansson's constant Q model (Kjartansson, 1979) originally implemented in AXITRA (Coutant, 1989) is

$$\widehat{V}_p(\mathbf{x}, \omega)_{relaxed} = V_p(\mathbf{x})_{unrelaxed} \left(\frac{\omega}{\omega_{ref}} \right)^\gamma (1 - iQ_p^{-1}), \quad (2.99)$$

$$\widehat{V}_\nu(\mathbf{x}, \omega)_{relaxed} = V_\nu(\mathbf{x})_{unrelaxed} \left(\frac{\omega}{\omega_{ref}} \right)^\gamma (1 - iQ_\nu^{-1}), \quad (2.100)$$

where the coefficient γ is set as

$$\gamma = \frac{\arctan(Q^{-1})}{\pi}. \quad (2.101)$$

The computation of these complex velocities requires a reference frequency ω_{ref} .

SLS approximation

Following the Voigt indexing, the complex relaxation rate in Equation (2.19) can be written as

$$\widehat{M}_{IJ}(\mathbf{x}, \omega) = C_{IJ}(\mathbf{x}) \left(1 - Q_{IJ}^{-1}(\mathbf{x}) \sum_{\nu=1}^L y_\nu \frac{\omega_\nu^2 - i\omega\omega_\nu}{\omega_\nu^2 + \omega^2} \right). \quad (2.102)$$

For isotropic attenuation mechanism, the complex compressional and shear velocities can then be written as

$$\widehat{V}_p^2(\mathbf{x}, \omega) = \frac{\widehat{M}_{11}(\mathbf{x}, \omega)}{\rho(\mathbf{x})} = \frac{C_{11}(\mathbf{x})}{\rho(\mathbf{x})} \left(1 - Q_p^{-1}(\mathbf{x}) \sum_{s=1}^L y_\nu \frac{\omega_\nu^2 - i\omega\omega_\nu}{\omega_\nu^2 + \omega^2} \right), \quad (2.103)$$

$$\widehat{V}_\nu^2(\mathbf{x}, \omega) = \frac{\widehat{M}_{44}(\mathbf{x}, \omega)}{\rho(\mathbf{x})} = \frac{C_{44}(\mathbf{x})}{\rho(\mathbf{x})} \left(1 - Q_s^{-1}(\mathbf{x}) \sum_{s=1}^L y_\nu \frac{\omega_\nu^2 - i\omega\omega_\nu}{\omega_\nu^2 + \omega^2} \right). \quad (2.104)$$

It should be remarked that the unrelaxed P-wave speed is $C_{11}(\mathbf{x})/\rho(\mathbf{x})$, and S-wave speed is $C_{44}(\mathbf{x})/\rho(\mathbf{x})$, implying that the complex P- and S-wave speeds can be directly estimated from the unrelaxed quantities as

$$\begin{aligned} (\widehat{V}_p^2(\mathbf{x}, \omega))_{relaxed} &= (V_p^2(\mathbf{x}))_{unrelaxed} \left(1 - Q_p^{-1}(\mathbf{x}) \sum_{s=1}^L y_\nu \frac{\omega_\nu^2 - i\omega\omega_\nu}{\omega_\nu^2 + \omega^2} \right), \\ (\widehat{V}_\nu^2(\mathbf{x}, \omega))_{relaxed} &= (V_\nu^2(\mathbf{x}))_{unrelaxed} \left(1 - Q_s^{-1}(\mathbf{x}) \sum_{s=1}^L y_\nu \frac{\omega_\nu^2 - i\omega\omega_\nu}{\omega_\nu^2 + \omega^2} \right). \end{aligned} \quad (2.105)$$

These complex velocities have been added in the AXITRA code.

Chapter 3

Inverse problem

Contents

3.1 Gradient computation by the adjoint-state approach	64
3.1.1 Lagrangian definition & Adjoint system	64
3.1.2 Gradient expressions	67
3.1.3 Influence of the attenuation on the velocity gradient estimation	69
3.2 Solving the constrained local minimization problem	70
3.2.1 Step length computation	71
3.2.2 Descent direction computation	71
3.2.3 Preconditioning for optimization process	73
3.2.4 Non-linear model constraints on the relationship between different parameters	74
3.3 Source estimation	76
3.4 Conclusion	77
3.A Gradients estimation	78

Standard FWI is a local optimization method, which attempts to reduce iteratively the data misfit between the calculated \mathbf{d}_{cal} and the observed data \mathbf{d}_{obs} at receiver locations, for instance through the least-squares distance in Equation (1.3). The minimization problem is solved under some constraints related to the wave equation, in which the descent direction relies on the gradient of the misfit function. The Lagrangian strategy is used to develop the gradient expressions.

The visco-elastic gradient for a set of model parameters $(\rho, V_p, V_s, Q_p^{-1}, Q_s^{-1})$ can be computed by the adjoint-state method (Plessix, 2006). The impact of attenuation when recovering elastic parameters will be pointed out in a toy configuration. A review of various local optimization approaches for model perturbation estimation is also provided. They are all approximated versions of the model perturbation estimation through the inverse Hessian and gradient in Equation (1.6). When inverting for multiple parameters, for example compressional and shear wave speeds, the use of an additional non-linear model constraint on the relationship between inverting parameters becomes attractive to reduce the possible model-search space. This could lead to a more reliable model parameters estimation as pointed out in the application chapters.

3.1 Gradient computation by the adjoint-state approach

We define the incident vector from the displacement \mathbf{u} and memory variables fields ψ_ν associated with L SLS mechanisms as

$$\mathcal{W} = \underbrace{(u_1, u_2, u_3)}_{\mathbf{u}}, \underbrace{(\psi_{1;1}, \dots, \psi_{1;6})}_{\psi_1}, \dots, \underbrace{(\psi_{L;1}, \dots, \psi_{L;6})}_{\psi_L}^T, \quad (3.1)$$

which is the full wavefield estimated from the forward problem. We recast the visco-elastic system in Equation (2.23) into a forward problem with variable \mathcal{W} . According to the definition of the incident vector, the forward problem (2.23) can now be written as

$$\begin{aligned} & \underbrace{\begin{pmatrix} \rho I_3 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 \end{pmatrix}}_{\mathbf{B}_2(\mathbf{m})} \partial_{tt} \mathcal{W} + \underbrace{\begin{pmatrix} 0 & 0 & \dots & 0 \\ 0 & I_6 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & I_6 \end{pmatrix}}_{\mathbf{B}_1} \partial_t \mathcal{W} \\ & + \underbrace{\begin{pmatrix} -DCD^T & DC^a & \dots & DC^a \\ -y_1 \omega_1 D^T & \omega_1 I_6 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ -y_L \omega_L D^T & 0 & \dots & \omega_L I_6 \end{pmatrix}}_{\mathbf{B}_0(\mathbf{m})} \mathcal{W} = \mathbf{S}, \end{aligned} \quad (3.2)$$

or in a more compact form with the introduction of matrices $\mathbf{B}_2(\mathbf{m})$, \mathbf{B}_1 and $\mathbf{B}_0(\mathbf{m})$ as

$$F(\mathbf{m}, \mathcal{W}) = \mathbf{B}_2(\mathbf{m}) \partial_{tt} \mathcal{W} + \mathbf{B}_1 \partial_t \mathcal{W} + \mathbf{B}_0(\mathbf{m}) \mathcal{W} - \mathbf{S} = 0. \quad (3.3)$$

The optimal choice of model parameters \mathbf{m} , especially in attenuating medium, is not obvious and is expected to be dependent on the medium complexity, acquisition design and radiation pattern. As a reference and natural parametrization, we compute the gradient for the density ρ , the independent components of the unrelaxed stiffness components (C_{IJ}) and the inverse of the quality factors Q_p and Q_s

$$\mathbf{m} = (\rho, C_{IJ;(1 \leq I \leq J \leq 6)}, Q_p^{-1}, Q_s^{-1}). \quad (3.4)$$

3.1.1 Lagrangian definition & Adjoint system

By considering the incident vector \mathcal{W} , the least-square misfit function in Equation (1.3) between the calculated \mathbf{d}_{cal} and the observed data \mathbf{d}_{obs} can be written as

$$\chi(\mathbf{m}) = \frac{1}{2} \|R\mathcal{W}(\mathbf{m}) - \mathbf{d}_{obs}\|^2, \quad (3.5)$$

where R is a restriction operator extracting the full wavefield \mathcal{W} at receiver positions for each source. We introduce the Lagrangian function with the Lagrangian multiplier vector $\overline{\mathcal{W}}$, subjected to the wave equation constraint as

$$L(\mathbf{m}, \mathcal{W}, \overline{\mathcal{W}}) = \frac{1}{2} \|R\mathcal{W} - \mathbf{d}_{obs}\|^2 + \left(\overline{\mathcal{W}}, F(\mathbf{m}, \mathcal{W}) \right). \quad (3.6)$$

The inner product is defined over the space and time domain $\Omega \times [0 - t]$. It should be noted that the definition of the Lagrangian multiplier is not unique, and not the physical field. We can choose the adjoint fields to suit our numerical purpose as described in the latter steps.

Let $\mathcal{W}(\mathbf{m})$ be the solution of the forward problem (3.3) for the model parameter \mathbf{m} . Then $F(\mathbf{m}, \mathcal{W}(\mathbf{m})) = 0$, the second term of the Lagrangian function will be thus canceled

$$L(\mathbf{m}, \mathcal{W}, \overline{\mathcal{W}}) = \frac{1}{2} \|R\mathcal{W} - \mathbf{d}_{obs}\|^2 = \chi(\mathbf{m}), \quad (3.7)$$

giving the misfit function $\chi(\mathbf{m})$. Thus, the gradient of the misfit function can be computed from the derivative of the Lagrangian function with respect to model parameters \mathbf{m}

$$\frac{\partial \chi(\mathbf{m})}{\partial \mathbf{m}} = \frac{\partial L(\mathbf{m}, \mathcal{W}, \overline{\mathcal{W}})}{\partial \mathbf{m}}. \quad (3.8)$$

Using the chain rule, this equation is equivalent to

$$\frac{\partial \chi(\mathbf{m})}{\partial \mathbf{m}} = \left(\overline{\mathcal{W}}, \frac{\partial F(\mathbf{m}, \mathcal{W})}{\partial \mathbf{m}} \right) + \frac{\partial L(\mathbf{m}, \mathcal{W}, \overline{\mathcal{W}})}{\partial \mathcal{W}} \frac{\partial \mathcal{W}(\mathbf{m})}{\partial \mathbf{m}}, \quad (3.9)$$

where the adjoint fields are considered as independent with the model parameters (i.e. $\partial \overline{\mathcal{W}} / \partial \mathbf{m} = 0$).

Similar to the construction of the incident full wavefield \mathcal{W} , we define the Lagrangian multiplier from adjoint displacement $\overline{\mathbf{u}}$ and memory variable fields $\overline{\phi}_\nu$ as

$$\overline{\mathcal{W}} = (\overline{\mathbf{u}}, \overline{\phi}_1, \dots, \overline{\phi}_L)^T, \quad (3.10)$$

which satisfies

$$\frac{\partial L(\mathbf{m}, \mathcal{W}, \overline{\mathcal{W}})}{\partial \mathcal{W}} = 0. \quad (3.11)$$

Our later development will show that this condition is equivalent to a new PDE to be solved.

The gradient of the misfit function on model parameters \mathbf{m} in Equation (3.9) can then be simplified as

$$\frac{\partial \chi(\mathbf{m})}{\partial \mathbf{m}} = \left(\overline{\mathcal{W}}, \frac{\partial F(\mathbf{m}, \mathcal{W})}{\partial \mathbf{m}} \right). \quad (3.12)$$

Adjoint fields construction

We would like to remind some properties of adjoint operators “ \bullet^\dagger ” with respect to time and spatial derivatives that we use in the next developments

$$D^\dagger = -D^T \quad , \quad (\partial_t)^\dagger = -\partial_t, \quad (3.13)$$

where D is the first-order spatial derivative operator described in Equation (2.6). According to the definition of the Lagrangian multiplier defined in Equation (3.11), we can compute the adjoint fields $\overline{\mathcal{W}}$ through

$$0 = \frac{\partial L(\mathbf{m}, \mathcal{W}, \overline{\mathcal{W}})}{\partial \mathcal{W}} = R^\dagger (R\mathcal{W} - \mathbf{d}_{obs}) + \left(\frac{\partial F(\mathbf{m}, \mathcal{W})}{\partial \mathcal{W}} \right)^\dagger \overline{\mathcal{W}}. \quad (3.14)$$

By considering the properties of adjoint operator in Equation (3.13), we obtain

$$\mathbf{B}_2^\dagger(\mathbf{m}) \partial_{tt} \overline{\mathcal{W}} - \mathbf{B}_1^\dagger \partial_t \overline{\mathcal{W}} + \mathbf{B}_0(\mathbf{m})^\dagger \overline{\mathcal{W}} = -R^\dagger (R\mathcal{W} - \mathbf{d}_{obs}). \quad (3.15)$$

The data residual is mainly driven by the displacement residual receiver positions $\Delta d_{\mathbf{u}}$ and no contribution is coming from the memory variable residuals Δd_{ϕ_ν} as

$$R\mathcal{W} - \mathbf{d}_{obs} = [\Delta d_{\mathbf{u}}, \Delta d_{\phi_1}, \dots, \Delta d_{\phi_L}]^T := [\Delta d_{\mathbf{u}}, 0, \dots, 0]^T. \quad (3.16)$$

The system (3.15) governing the Lagrangian multiplier can now be expanded as

$$\begin{aligned} & \underbrace{\begin{pmatrix} \rho I_3 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 \end{pmatrix}}_{\mathbf{B}_2^\dagger(\mathbf{m})} \partial_{tt} \begin{bmatrix} \bar{\mathbf{u}} \\ \bar{\phi}_1 \\ \dots \\ \bar{\phi}_L \end{bmatrix} - \underbrace{\begin{pmatrix} 0 & 0 & \dots & 0 \\ 0 & I_6 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & I_6 \end{pmatrix}}_{\mathbf{B}_1^\dagger} \partial_t \begin{bmatrix} \bar{\mathbf{u}} \\ \bar{\phi}_1 \\ \dots \\ \bar{\phi}_L \end{bmatrix} + \\ & + \underbrace{\begin{pmatrix} -DCD^T & y_1\omega_1 D & \dots & y_L\omega_L D \\ -C^a D^T & \omega_1 I_6 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ -C^a D^T & 0 & \dots & \omega_L I_6 \end{pmatrix}}_{\mathbf{B}_0^\dagger} \begin{bmatrix} \bar{\mathbf{u}} \\ \bar{\phi}_1 \\ \dots \\ \bar{\phi}_L \end{bmatrix} = \begin{bmatrix} -R^\dagger \Delta d_{\mathbf{u}} \\ 0 \\ \dots \\ 0 \end{bmatrix}, \quad (3.17) \end{aligned}$$

resulting in the second-order adjoint system

$$\begin{cases} \rho \partial_{tt} \bar{\mathbf{u}} = DCD^T \bar{\mathbf{u}} - D \sum_{\nu=1}^L y_\nu \omega_\nu \bar{\phi}_\nu - R^\dagger \Delta d_{\mathbf{u}}, \\ \partial_t \bar{\phi}_\nu - \omega_\nu \bar{\phi}_\nu = -C^a D^T \bar{\mathbf{u}}. \end{cases} \quad (3.18)$$

In the adjoint system, the displacement misfit at receiver positions $R^\dagger \Delta d_{\mathbf{u}}$, i.e. the residues which are not explained in the current model description, acts as the source term. We introduce the modified adjoint memory variables $\bar{\psi}_\nu$ such that

$$y_\nu \omega_\nu \bar{\phi}_\nu = C^a \bar{\psi}_\nu, \quad (3.19)$$

which transforms the adjoint system (3.18) into

$$\begin{cases} \rho \partial_{tt} \bar{\mathbf{u}} = DCD^T \bar{\mathbf{u}} - DC^a \sum_{\nu=1}^L \bar{\psi}_\nu - R^\dagger \Delta d_{\mathbf{u}}, \\ \partial_t \bar{\psi}_\nu - \omega_\nu \bar{\psi}_\nu = -y_\nu \omega_\nu D^T \bar{\mathbf{u}}. \end{cases} \quad (3.20)$$

The obtained system now has nearly identical equations as the forward problem in Equation (2.23), except the sign “-” in the memory variable ODE and the adjoint source.

It should be noted that the back-propagation of the adjoint wavefield comes from its zero final condition, which naturally appears through the Lagrangian formalism when the zero initial condition of the incident wavefield is included as a constraint. Here, for the sake of simplicity, we have only defined the Lagrangian through the volumetric constraints on the incident wavefield without introducing initial and boundary conditions, and we refer to more mathematical oriented studies for a complete derivation of the gradient (see for instance (Plessix, 2006) and references therein).

Regarding the stability of the adjoint wavefield back-propagation, the change of sign in the ODEs defining the relaxation mechanisms would induce an exponential growth of the energy forward in time.

However, as the adjoint wavefield is back-propagated in time, the relaxation mechanism decreases the energy of the adjoint wavefield backward in time. Also, the set of coupled PDE/ODE defining the incident and adjoint wavefields solely differs from the change of sign in the ODE part. This significantly simplifies the implementation: a significant part of the code designed for the incident wavefield can be re-used for the adjoint wavefield.

This section constructs the Lagrangian function and adjoint fields associated with the least-squares misfit function, but similar workflow can be applied for other misfit functions. Changing the misfit function only affects the adjoint source term (Brossier et al., 2010b).

3.1.2 Gradient expressions

The section presents the gradient computations for the set of model parameters \mathbf{m} in Equation (3.4), where the mathematical development is presented in Appendix (3.A).

For both elastic and visco-elastic problems, the gradient on density ρ is the zero-lag cross-correlation of the adjoint displacement field and the incident acceleration field

$$\frac{\partial \chi(\mathbf{m})}{\partial \rho} = (\bar{\mathbf{u}}, \partial_{tt} \mathbf{u})_{\Omega, t}. \quad (3.21)$$

The gradient on attenuation parameters Q_p^{-1} and Q_s^{-1} can be estimated as

$$\frac{\partial \chi(\mathbf{m})}{\partial Q_{p,s}^{-1}} = - \left(\bar{\boldsymbol{\varepsilon}}, \sum_{\nu=1}^L \frac{\partial C^a}{\partial Q_{p,s}^{-1}} \boldsymbol{\psi}_\nu \right)_{\Omega, t}, \quad (3.22)$$

which is the zero-lag cross-correlation of the adjoint strain field $\bar{\boldsymbol{\varepsilon}}$, and the incident memory variable field $\boldsymbol{\psi}_\nu$. The obtained expression (3.22) is coherent with the Q -gradient proposed by Fichtner and van Driel (2014). Under the isotropic attenuation regime, our approximation in Equations (2.25) and (2.26) provides simple expressions of the derivatives of the attenuative stiffness tensor C^a with respect to attenuation parameters Q_p^{-1} and Q_s^{-1} as

$$\frac{\partial C^a}{\partial Q_p^{-1}} = \frac{1}{3} \left(\sum_{I=1}^3 C_{II} \right) \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad (3.23)$$

$$\frac{\partial C^a}{\partial Q_s^{-1}} = \frac{1}{3} \left(\sum_{I=4}^6 C_{II} \right) \begin{bmatrix} 0 & -2 & -2 & 0 & 0 & 0 \\ -2 & 0 & -2 & 0 & 0 & 0 \\ -2 & -2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \quad (3.24)$$

Following the separation of the elastic rheology and attenuation mechanism in Equation (2.19), the C_{IJ} elementary gradient has two terms

$$\frac{\partial \chi(\mathbf{m})}{\partial C_{IJ}} = \left(\bar{\boldsymbol{\varepsilon}}, \frac{\partial C}{\partial C_{IJ}} \boldsymbol{\varepsilon} \right)_{\Omega, t} - \left(\bar{\boldsymbol{\varepsilon}}, \sum_{\nu=1}^L \frac{\partial C^a}{\partial C_{IJ}} \boldsymbol{\psi}_\nu \right)_{\Omega, t}. \quad (3.25)$$

The first term has identical expression as the elastic gradients proposed by Tromp et al. (2005); Virieux et al. (2017), which is the zero-lag cross-correlation of the adjoint $\bar{\varepsilon}$ and the incident strain fields ε , weighted by the spatial radiation pattern $\partial C/\partial C_{IJ}$. Both adjoint and incident strain fields are calculated in the anelastic medium. The second term is related to the memory variables ψ_ν . Under our approximation of the isotropic attenuation in Equations (2.25) and (2.26), the derivatives of the attenuative stiffness tensor C^a with respect to elastic stiffness coefficients can be simplified as

$$\left. \frac{\partial C^a}{\partial C_{IJ}} \right|_{I \neq J} = 0, \quad (3.26)$$

$$\left. \frac{\partial C^a}{\partial C_{II}} \right|_{I=1,2,3} = \frac{1}{3} Q_p^{-1} \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad (3.27)$$

$$\left. \frac{\partial C^a}{\partial C_{II}} \right|_{I=4,5,6} = \frac{1}{3} Q_s^{-1} \begin{bmatrix} 0 & -2 & -2 & 0 & 0 & 0 \\ -2 & 0 & -2 & 0 & 0 & 0 \\ -2 & -2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \quad (3.28)$$

The expression (3.27) confirms that the attenuation coefficient Q_p has an isotropic impact on compressional components, thus having isotropic impact on V_p estimation (similar interpretation for the attenuation coefficient Q_s through Equation (3.28)). The development of $\partial C^a/\partial C_{IJ}$ also highlights that, in a weakly attenuative medium, the second term in the visco-elastic C_{IJ} -gradient (Eqn. 3.25) can be neglected and tends to zero in purely elastic medium. Since the memory variables ψ_ν obey the first-order ODE (2.22) with the incident strain field ε as the source, the magnitude of the second term is generally small compared with the first term. However, we shall expect an accumulative influence on the amplitude of the model parameter reconstruction. Numerical illustration of this effect on a simple inclusion example is presented in Section (3.1.3).

This interpretation is also consistent with other theoretical investigations in the literature. Tarantola (1988) showed that the gradient with respect to the general relaxation rate is the cross-correlation of the adjoint and the incident strain fields. By considering the chain rule with respect to the stiffness coefficients, where the relaxation function M_{ijkl} is given in Equation (2.19), a two-term expression for the C_{IJ} -gradient, as for our formulation (3.25), shall be obtained. Liu and Tromp (2008) also expressed the gradient of the misfit function through the cross-correlation of the adjoint and incident strain fields, scaled by the perturbation of the relaxation rate δM . The time-dependency of this perturbation is related to the second term in our gradient expression. For numerical implementation, Liu and Tromp (2008); Komatitsch et al. (2016) assumed that δM_{ijkl} is time-independent, resulting in a simplification of the gradient kernels, explaining why they obtained similar expression as for the elastic case.

Furthermore, the gradient for any parameter α (seismic velocity, anisotropic parameter, impedance etc.) can be computed by chain rule using the density ρ , C_{IJ} and $Q_{p,s}^{-1}$ elementary gradients

$$\frac{\partial \chi(\mathbf{m})}{\partial \alpha} = \sum_{i=1}^6 \sum_{j=i}^6 \frac{\partial \chi}{\partial C_{IJ}} \frac{\partial C_{IJ}}{\partial \alpha} + \frac{\partial \chi}{\partial \rho} \frac{\partial \rho}{\partial \alpha} + \frac{\partial \chi}{\partial Q_p^{-1}} \frac{\partial Q_p^{-1}}{\partial \alpha} + \frac{\partial \chi}{\partial Q_s^{-1}} \frac{\partial Q_s^{-1}}{\partial \alpha}. \quad (3.29)$$

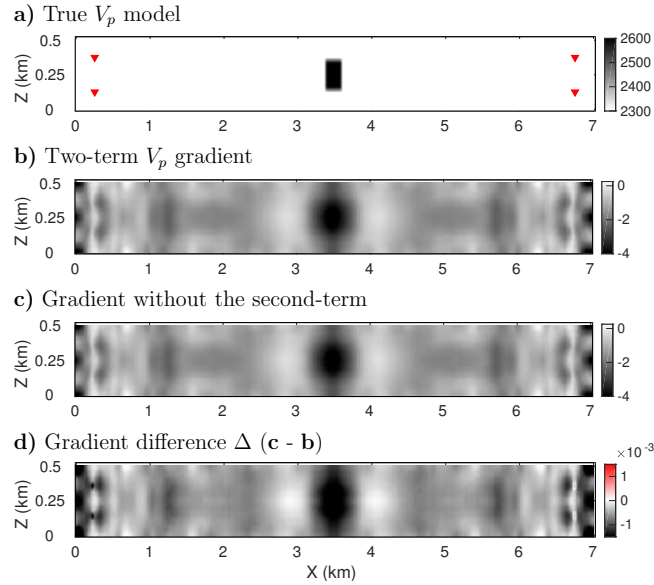


Figure 3.1: (a) Homogeneous background with a V_p inclusion in the center of the model. Source positions are highlighted by red triangles. (b) First visco-elastic FWI V_p -gradient computed from elementary C_{IJ} -gradients in Equation (3.25). (c) First FWI gradient without considering the second term. (d) Difference between gradient (c) and (b), which highlights contribution related to the second term gradient.

3.1.3 Influence of the attenuation on the velocity gradient estimation

The second term in Equation (3.25) has never been explicitly reported before, although correctly handled by Fichtner and van Driel (2014). It is generally ignored in practical applications (Liu and Tromp, 2008; Komatitsch et al., 2016). Its magnitude is small in weakly attenuating medium compared with the first-term contribution. However, its influence should increase with the attenuation, especially at far offsets. Therefore, in highly attenuating medium, ignoring the second-term might affect the model parameter estimation.

The contribution of the second term in Equation (3.25) is illustrated on a simple toy example. We use a simple setting with homogeneous background $V_p = 2300$ m/s, $V_s = 1500$ m/s, $\rho = 1000$ kg/m³ and a V_p inclusion in the center of the model as shown in Figure 3.1(a). The medium is strongly attenuating with $Q_p = Q_s = 40$.

In Figure 3.1(a), the medium is illuminated by 4 sources, highlighted by red triangles. For each source, a plan containing 6241 receivers is located at the opposite side of the inclusion: for instance when the source position is $x_s = 250$ m, the receiver plan is at $x_r = 6750$ m. Free-surface conditions are applied at all faces of the model to increase the illumination through the reflection at the boundary of the model. The source time function is a Ricker wavelet, centered at 10 Hz. With this setting, the distance between sources and receivers is about 30 dominant P-wavelengths. Since a very sparse acquisition is used in this toy example, we use long recording time (9 sec) so that the medium can be sampled by multiple reflected wavefield. We use a homogeneous initial model $V_p = 2300$ m/s. Figure 3.1(b) shows an estimation of the first V_p -gradient computed from elementary C_{IJ} -gradients in Equation (3.25). The inversion problem is ill-constrained leading to significant artifacts in the gradi-

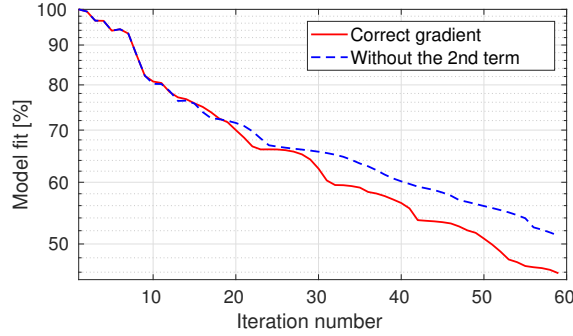


Figure 3.2: An example of the model-fit at the inclusion position between the true V_p and the inverted model using the correct gradient estimation (red) and without considering the 2^{nd} term (blue), which highlights the accumulative effect on model parameter estimation.

ents. However, we only apply a minimal Bessel smoothing to remove the numerical artifacts above the Nyquist wave-number associated with the underlying mesh. The gradient smoothing operator is discussed in Chapter (5).

Figure 3.1(c) shows the V_p -gradient estimation when ignoring the second term in the elementary C_{IJ} -gradients. The difference between the gradients (b) and (c) in Figure 3.1(d) highlights the small amplitude modification at the inclusion position and the places intensively sampled by the wavefield. It should be noted that without the second term, the gradient (c) is under-estimated at the inclusion position, implying a non-correct estimation of the model update. On this setting, we run a FWI test with 60 iterations where the model-fit at the inclusion position is shown in Figure 3.2: The model-fit decreases faster when considering the correct gradient. The comparison effectively shows that ignoring the second term in the gradient estimation might affect the reconstruction of the model parameter. The effect manifests at later stages, when the shape of the target has been found, and the inversion attempts to recover the correct amplitude of model parameter.

In practical applications, these effects might be negligible compared with other sources of uncertainties. However, the correct gradient can be collected without any extra cost since the incident fields ψ_ν are available during the simultaneous forward and backward propagations of the adjoint and the incident fields (Algorithm (5) in Section (4.2)).

3.2 Solving the constrained local minimization problem

As described by Equation (1.6), in the Newton optimization framework, the opposite of the gradient weighted by the inverse of the Hessian provides the descent direction of the misfit function. The use of Hessian also offers the possibility to mitigate the cross-talk between different parameters in multiple-parameters FWI. However, the explicit computation of the Hessian is unfeasible. Some rather simple approximations of the Hessian are normally considered in practical applications.

Our inversion optimization kernel relies on the SEISCOPE optimization toolbox (Métivier and Brossier, 2016). The library embeds various non-linear optimization methods, which have been fully included in the inversion part of the SEM46 code, without any upper-level wrapper. This section provides a quick review of the numerical methods for solving constrained local optimization problem,

available in this optimization toolbox. At each iteration, all algorithms aim at estimating the updated model as

$$\mathbf{m}^{k+1} = \mathbf{m}^k + \alpha_k \Delta \mathbf{m}^k, \quad (3.30)$$

where $\Delta \mathbf{m}^k$ now refers to a *descent direction* and α_k is a *step length*.

When the full Hessian is not available, the introduction of some additional prior information into the inversion process is necessary to reduce the searching model space in multiple parameters FWI. The section will also discuss the use of optimization preconditioning and the integration of multiple non-linear constraints on the inversion problem.

3.2.1 Step length computation

At each iteration, the step length value controls how far to move along the descent direction $\Delta \mathbf{m}^k$. The step-length α_k is computed through a line-search process, which satisfies the Wolfe conditions (Wolfe, 1969). The condition involves two criteria (Nocedal and Wright, 2006) :

- *Sufficient decrease*: For a given size of step-length, the misfit function reduction should be proportional to both the step length and the directional derivative

$$\chi(\mathbf{m}^k + \alpha_k \Delta \mathbf{m}^k) \leq \chi(\mathbf{m}^k) + b_1 \alpha_k \nabla \chi|_{\mathbf{m}^k}^T \Delta \mathbf{m}^k. \quad (3.31)$$

In practice, the constant b_1 is chosen to be quite small as $b_1 = 10^{-4}$.

- *Curvature condition*: This condition guarantees not too small steps which would result in slow convergence through

$$\nabla \chi(\mathbf{m}^k + \alpha_k \Delta \mathbf{m}^k)^T \Delta \mathbf{m}^k \geq b_2 \nabla \chi(\mathbf{m}^k)^T \Delta \mathbf{m}^k. \quad (3.32)$$

It provides an indication that we reduce the misfit function by moving further along the chosen descent direction $\Delta \mathbf{m}^k$. In practice, the constant b_2 is chosen to be close to 1 as $b_2 = 0.9$.

The Wolfe condition ensures the convergence towards the nearest “local” minimum from an arbitrary initial guess, under the condition that the misfit function is bounded and twice continuously differentiable.

3.2.2 Descent direction computation

3.2.2.1 Steepest descent direction

The steepest descent approach simply follows the negative of the first derivative of the misfit function (i.e. the gradient)

$$\Delta \mathbf{m}^k := -\nabla \chi|_{\mathbf{m}^k}. \quad (3.33)$$

This approach has been proved to have a slow convergence (Nocedal and Wright, 2006). Moreover, the convergence might not be uniform for the ensemble of parameters. For example, in a standard surface-acquisition, the medium is not uniformly illuminated. The areas suffering from poor illumination will be poorly reconstructed due to the unbalance energy in the gradient.

3.2.2.2 Non-linear conjugate gradient method

The conjugate gradient method has been shown to be effective for convex quadratic system. For the non-linear FWI problem, we consider the non-linear conjugate gradient approach. At each iteration, the descent direction is a combination of the current gradient $\nabla\chi|_{\mathbf{m}^k}$ with the previous descent direction $\Delta\mathbf{m}^{k-1}$ as

$$\Delta\mathbf{m}^k = -\nabla\chi|_{\mathbf{m}^k} + \beta_k\Delta\mathbf{m}^{k-1}, \quad (3.34)$$

where β_k is computed following the formula of Dai and Yuan (1999). By doing so, the method accelerates the convergence compared with the steepest descent method. However, this acceleration is not predictable: non-linear conjugate gradient can exhibit varying performance.

3.2.2.3 Truncated Newton method

The full Newton method compute the descent direction based on the inverse of the Hessian matrix as described in Equation (1.6) as

$$\Delta\mathbf{m}^k = -[\nabla^2\chi|_{\mathbf{m}^k}]^{-1}\nabla\chi(\mathbf{m}^k). \quad (3.35)$$

The consideration of the Hessian could significantly improve the convergence rate of the iterative FWI process and is powerful in multi-parameter inversion. However, we do not compute explicitly the forward or inverse Hessian due to the significant cost. This leads to the truncated Newton method which estimates the descent direction based on solving an approximated linear system of the Newton equation (Métivier et al., 2013, 2014; Castellanos et al., 2015)

$$[\nabla^2\chi|_{\mathbf{m}^k}]\Delta\mathbf{m}^k = -\nabla\chi(\mathbf{m}^k). \quad (3.36)$$

Even when the linear system can be efficiently solved by using a matrix-free conjugate gradient algorithm, the overall cost remains significant, especially for SEM implementation of elastic and visco-elastic problems.

3.2.2.4 Quasi-Newton l -BFGS method

The method allows taking benefit from the information inside the Hessian without paying the full computational cost of the full Newton or truncated Newton methods. A limited number of gradients and models from the previous iterations are stored in the memory and then used to estimate an approximation of the inverse Hessian (Byrd et al., 1995; Nocedal and Wright, 2006). This information is then included inside the descent direction as

$$\Delta\mathbf{m}^k = -Q_k\nabla\chi|_{\mathbf{m}^k}, \quad (3.37)$$

where Q_k is the l -BFGS approximation of the inverse Hessian operator. The number of required previous gradients is relatively small, between 5 and 20, but the method significantly improve the inversion results. The numerical implementation follows a too-loop recursion strategy, which avoids any explicit estimation of the inverse Hessian. The approximation of the Hessian matrix-vector product only involves basis arithmetic manipulations between vectors, leading to a very attractive efficiency (Brossier et al., 2009b).

All our synthetic and real data applications rely on the l -BGFS optimization method, with and without preconditioning as introduced in the next section.

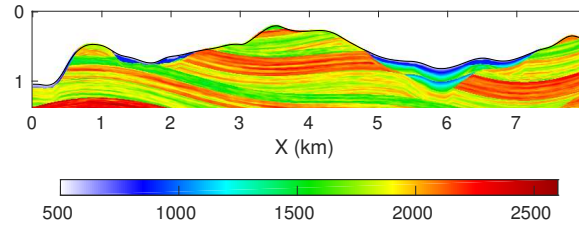


Figure 3.3: An example of the complex near-surface extracted from the SEAM Phase II Foothill benchmark. The alluvial deposits at the free-surface have low shear velocity, at about 550 m/s.

3.2.3 Preconditioning for optimization process

The use of an appropriate optimization preconditioning P_k might accelerate the convergence of the FWI problem. In the preconditioned l -BFGS method, the preconditioning P_k serves as an initial estimation of the approximation of the Hessian \tilde{Q}_k . The descent direction now can be estimated as

$$\Delta \mathbf{m}^k = -\tilde{Q}_k \nabla \chi|_{\mathbf{m}^k}. \quad (3.38)$$

In practice, depth preconditioning is widely used to enhance the model estimation with depth, following the z direction as

$$P_k(z, x, y) = (\beta + z)^\alpha, \quad (3.39)$$

where α and β are pre-defined constants. In 3D surface-acquisition application, the power α is usually chosen as 2. In marine applications, the scalar β can be set as 0 since the zero depth is located inside the water column. However, in land applications, the tuning of both α and β coefficients is required. The depth preconditioning appears to be robust and efficient in flat-topography condition. However, the simple preconditioning as defined in Equation (3.39) might have significant footprints from a rough topography variation in foothills environment.

Since the Hessian has a diagonal-dominant structure, the approximations of its diagonal are widely used to precondition the inverse problem (Pratt et al., 1998; Operto et al., 2006). Following the development from Shin et al. (2001), we consider the illumination preconditioning based on the displacement field \mathbf{u} as

$$P_k(\mathbf{x}) = \frac{1}{\|\mathbf{u}(\mathbf{x})\|^\alpha + \mu}, \quad (3.40)$$

where α is the power that we apply to the norm of the wavefield and μ is a stabilizing factor to avoid the division by zero. Since the preconditioning is directly computed from the wavefield, it naturally compensates for the amplitude loss due to the 3D propagation and complex heterogeneity. Practical applications show that $\alpha = 1$ provides the best energy balancing between the shallow and deep targets. It should be noted that surface energies are also included in the construction of the preconditioning, implying a high α value might artificially amplify the model estimation at greater depth due to high amplitude surface waves.

Our synthetic and real data applications only consider the illumination preconditioning for the l -BFGS optimization process, as presented in Equation (3.40) with $\alpha = 1$.

Algorithm 2: Standard POC algorithm

```

1  $x_0 = \mathbf{m}^k + \alpha_k \Delta \mathbf{m}^k$ 
2 while (Not convergence .and.  $k < n_{max}$ ) do
3   |  $y_k = \mathcal{P}_{\mathcal{C}_1}(x_k)$ ;
4   |  $x_{k+1} = \mathcal{P}_{\mathcal{C}_2}(y_k)$ ;
5 end

```

3.2.4 Non-linear model constraints on the relationship between different parameters

The multi-parameter FWI results have significant artifacts in the model estimation due to the cross-talk between parameters and the unequal sensitivity of the inversion process with respect to different parameters. For instance elastic FWI applications with the least-squares misfit function is generally more sensitive to V_s , due to strong amplitude shear events, leading to insignificant V_p estimation when inverting simultaneously for V_p and V_s . When considering significant geological heterogeneities such as alluvial deposits in the near-surface in Figure 3.3(a), the inversion process will be dominated by the V_s update in those low-velocity zones. This might lead to *unrealistic updates* where $V_s > V_p$ or too small value of V_s . Therefore, additional prior information should be incorporated through preconditioning and model regularization or constraints in addition to the standard gradient smoothing strategy.

To mitigate this issue, we introduce a non-linear constraint on the relationship between V_p and V_s . This constraint is applied in addition to the (hard) bounds constraint on the range of the inverting parameters, widely used in standard FWI applications. The inverse problem is then described as a multiple-constrained non-linear minimization problem

$$\min_{\mathbf{m}} \chi(\mathbf{m}) \quad \text{where} \quad \mathbf{m} \in \underbrace{(\mathcal{C}_1 \cap \mathcal{C}_2)}_{\Omega}. \quad (3.41)$$

The misfit function $\chi(\mathbf{m})$ is minimized over the restricted model space Ω , where the model \mathbf{m} should simultaneously satisfy two sets of constraints:

- Bounds constraint \mathcal{C}_1 : V_p and V_s should vary within pre-defined ranges.
- Ratio constraint \mathcal{C}_2 : The ratio V_p/V_s should vary within a pre-defined range as

$$r_1(\mathbf{x}) \leq \frac{V_p(\mathbf{x})}{V_s(\mathbf{x})} \leq r_2(\mathbf{x}). \quad (3.42)$$

These constraints can be designed based on well data or simple geological knowledge about the structure.

At each iteration, after the step length and descent direction computation, if the estimated model does not satisfy the set of constraints $\mathcal{C}_1, \mathcal{C}_2$, it needs to be projected back to the restricted space Ω , giving a more general expression of Equation (3.30) as

$$\mathbf{m}^{k+1} = \mathcal{P}_{\Omega}(\mathbf{m}^k + \alpha_k \Delta \mathbf{m}^k). \quad (3.43)$$

Under the assumption that all sets of constraints are convex, the projection-onto-convex-sets (POCS) strategy can be used to find a point in their intersection (Boyd et al., 2010; Baumstein, 2013). The algorithm performs a sequential projections onto these convex sets: the input of the current projection onto

Algorithm 3: Standard Dykstra algorithm

```

1  $x_0 = \mathbf{m}^k + \alpha_k \Delta \mathbf{m}^k$  and  $p_0 = 0, q_0 = 0$ 
2 while (Not convergence .and.  $k < n_{max}$ ) do
3    $y_k = \mathcal{P}_{C_1}(x_k + p_k);$ 
4    $x_{k+1} = \mathcal{P}_{C_2}(y_k + q_k);$ 
5    $p_{k+1} = x_k + p_k - y_k;$ 
6    $q_{k+1} = y_k + q_k - x_{k+1};$ 
7 end
    
```

the set of one constraint \mathcal{P}_{C_2} is the output of the previous projection onto the set of another constraint \mathcal{P}_{C_1} , as shown in Algorithm (2). If the applications of two constraints on the current state x_k are not commutative

$$\mathcal{P}_{C_1}(\mathcal{P}_{C_2}(x_k)) \neq \mathcal{P}_{C_2}(\mathcal{P}_{C_1}(x_k)), \quad (3.44)$$

the output of the POCS algorithm will depend on the order of the constraints application. It should be noted that the projection onto the set of bound-constraint \mathcal{P}_{C_1} is trivial: At the pixels where the values of the estimated model parameters are outside of the pre-defined range, they are replaced by the boundary values. However, designing a projection onto the set of ratio-constraint \mathcal{P}_{C_2} which is commutative with \mathcal{P}_{C_1} is not obvious. We thus decide to consider other strategies.

POCS is a special case of the Dykstra algorithm (Algorithm 3), when the additional auxiliary variables p_k and q_k are set as zero. These parameters are updated during the Dykstra's iterations and the sequence (x_k) is demonstrated to converge toward an unique solution of the original problem (3.43), independently of the order of the constraints application (Boyle and Dykstra, 1986). Figure 3.4 provides an illustration of the Dykstra projection algorithm. In general, the alternating Dykstra projection method might have a slow convergence. However, it is useful when the projection on each set is efficient, such as an analytical formula, for carrying out the projection (Boyd et al., 2010).

Following this remark, we design the projection onto the set of ratio-constraint \mathcal{P}_{C_2} such that an analytical solution of the projection exists. At each spatial position \mathbf{x} , the projection \mathcal{P}_{C_2} is equivalent

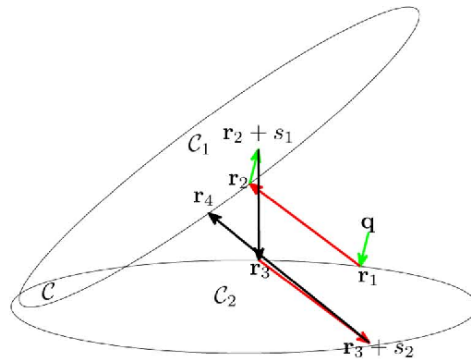


Figure 3.4: Illustration of the Dykstra algorithm. Figure extracted from Walsh and Regalia (2010)

to solving a minimization problem

$$\text{minimize}_{\alpha, \beta} f(\alpha, \beta) = \frac{(V_p - \alpha)^2}{V_p^2} + \frac{(V_s - \beta)^2}{V_s^2} \quad \text{subject to} \quad r_1 \leq \frac{\alpha}{\beta} \leq r_2. \quad (3.45)$$

If the estimated (V_p, V_s) are not satisfied the constraint (3.42), they will be replaced by the closest pair of values (α, β) , which belong to the set of ratio-constraint \mathcal{C}_2 . The analytical solution of the problem (3.45) exists

- If $V_p/V_s > r_2$, the minimization of the function $f(\alpha, \beta)$ is achieved when

$$\alpha = r_2\beta \quad \text{and} \quad \beta = \frac{V_p V_s (V_p + r_2 V_s)}{V_p^2 + r_2^2 V_s^2}. \quad (3.46)$$

- If $V_p/V_s < r_1$, the minimization of the function $f(\alpha, \beta)$ is achieved when

$$\alpha = r_1\beta \quad \text{and} \quad \beta = \frac{V_p V_s (V_p + r_1 V_s)}{V_p^2 + r_1^2 V_s^2}. \quad (3.47)$$

The uniqueness of the solution also confirms the convexity of the set of model parameters which satisfy the ratio constraint. For our problem, the Dykstra iterative process converges quickly, generally after 4 iterations. Moreover, since the analytical solutions of the projection exist, the constraint on the ratio V_p/V_s can be integrated into the inversion process without any extra cost. In this implementation, we consider a quadratic objective function (3.45), but other continuous differential functions could be easily considered.

Compared with standard model regularization strategies, the use of hard constraints is free of trade-off parameter and the inverted models are guaranteed to satisfy simultaneously all the imposed constraints (Peters and Herrmann, 2017). This section only describes the constraint on the ratio V_p and V_s , but the approach can be extended for multiple constraints consideration, even for more sophisticated relationships between parameters.

In seismic tomography, the set of model parameters $(V_p/V_s, V_p)$ is widely used, instead of (V_p, V_s) (Guoqing et al., 2007; Marjanovic et al., 2018). The bound constraint is applied on this ratio parameter to mitigate the issues related to V_s inversion. In our strategy, due to the constraint on V_p/V_s , we are able to invert simultaneously for V_p and V_s . However, it would be interesting to see the radiation patterns analysis with respect to this V_p/V_s parameter choice.

3.3 Source estimation

As described in the previous chapter, the source time function and the source physics (either point-force or point-moment) are required for the wavefield modeling. For real data applications, the source time signature is generally unknown, especially in the explosive source dataset. Pratt (1999) formulates the source inversion as a linear problem in the frequency domain, in which the source signature at each discrete frequency can be computed through

$$s(\omega) = \frac{\mathbf{d}_{mod}^\dagger(\mathbf{m}, \omega) \mathbf{d}_{obs}(\mathbf{m}, \omega)}{\mathbf{d}_{mod}^\dagger(\mathbf{m}, \omega) \mathbf{d}_{mod}(\mathbf{m}, \omega)}, \quad (3.48)$$

where \mathbf{d}_{mod} is the simulated data on the model \mathbf{m} with a delta source time function. The simulated and observed data are transformed from the time to the frequency domain by a standard Fourier transform. The source function is then converted back to the time domain via the inverse of the Fourier transform. The obtained source function can be seen as a matching filter which aims at matching the convolution $\mathbf{d}_{mod}(\mathbf{m}, \omega)s(\omega)$ with the observed data $\mathbf{d}_{obs}(\mathbf{m}, \omega)$, with respect to a least-square distance.

The source signature can be estimated for each shot individually or the ensemble of all shots. It should be noted that the obtained source function (3.48) depends on the accuracy of the background model. Since it is a deconvolution-like process, for inaccurate model parameters, a part of the model inaccuracy will be hidden inside the estimated source. Therefore, it is recommended to re-estimate the source after a significant change in the inverting model parameters (Plessix and Cao, 2011; He et al., 2018).

3.4 Conclusion

Standard FWI is a local optimization method, in which the descent direction relies on the gradient of the misfit function. This chapter presents the gradient computation following the adjoint-state approach (Plessix, 2006). Even if the second-order visco-elastic wave equation is not self-adjoint, we can develop the adjoint system with similar equations as for the incident fields, implying that they can be propagated under the same numerical scheme. Following the forward and adjoint formulation using SLS mechanisms, the gradient expressions for density, attenuation parameters and stiffness coefficients C_{IJ} are simply zero-lag cross-correlations in time between incident and adjoint fields, weighted by the associated spatial radiation pattern. The gradient for any other parameter can be computed by chain rule from these elements.

We report a two-term expression for the C_{IJ} -gradient, which is equivalent to theoretical investigations of Tarantola (1988); Liu and Tromp (2008). The first-term, related to the zero-lag cross-correlation of the adjoint and incident strain fields, is commonly used in practical application (Komatitsch et al., 2016). The second term can be interpreted as the time-dependency of the perturbation of the relaxation-rate, which has small amplitude compared with the first-term contribution. However, the influence of the second term is shown to have an accumulative effect on the amplitude of the model parameter estimation, depending on the physical domain and acquisition settings.

The chapter also provides a review of the numerical methods for solving the constrained local optimization FWI problem, available in the SEISCOPE optimization toolbox (Métivier and Brossier, 2016). Due to the limited computational resource, the computation of the full Hessian is unreachable, especially for 3D finite element implementation. To help the inversion process, we suggest to build the optimization preconditioning operator based on the simulated wavefield to compensate for the amplitude loss due to the 3D propagation. This is particularly useful in foothill environment where the topography footprint might disturb the standard depth precondition strategy. To partially mitigate the cross-talk between multiple parameters inversion, we introduce a non-linear model constraint on the relationship between inverting parameters. These preconditioning and constraint strategies are applied at no extra cost but very helpful for reliable model parameters estimation.

APPENDICES

3.A Gradients estimation

Gradient on density

The gradient on density ρ is given by

$$\frac{\partial \chi(\mathbf{m})}{\partial \rho} = \left(\overline{\mathbf{W}}, \frac{F(\mathbf{m}, \mathbf{W})}{\partial \rho} \right). \quad (3.49)$$

The derivative of the forward problem (3.3) on density is given by

$$\frac{\partial F(\mathbf{m}, \mathbf{W})}{\partial \rho} = \overbrace{\begin{pmatrix} I_3 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 \end{pmatrix}}^{\partial \mathbf{B}_2(\mathbf{m})/\partial \rho} \partial_{tt} \begin{bmatrix} \mathbf{u} \\ \psi_1 \\ \dots \\ \psi_L \end{bmatrix} = \begin{bmatrix} \partial_{tt} \mathbf{u} \\ 0 \\ \dots \\ 0 \end{bmatrix}. \quad (3.50)$$

By considering the Lagrangian multiplier vector $\overline{\mathbf{W}}$ in Equation (3.10), the gradient on density in Equation (3.49) becomes

$$\frac{\partial \chi(\mathbf{m})}{\partial \rho} = \left(\begin{bmatrix} \overline{\mathbf{u}} \\ \overline{\phi}_1 \\ \dots \\ \overline{\phi}_L \end{bmatrix}, \begin{bmatrix} \partial_{tt} \mathbf{u} \\ 0 \\ \dots \\ 0 \end{bmatrix} \right) = (\overline{\mathbf{u}}, \partial_{tt} \mathbf{u}). \quad (3.51)$$

When propagating the adjoint problem with a simultaneous reconstruction of the incident field, both adjoint displacement field $\overline{\mathbf{u}}$ and the incident acceleration field $\partial_{tt} \mathbf{u}$ are available. The gradient on density can thus be collected without any extra cost.

Gradient on attenuation parameters

Since the development of the gradient on attenuation parameters Q_p^{-1} and Q_s^{-1} are similar, we then consider $Q_{p,s}^{-1}$ standing for either Q_p^{-1} or Q_s^{-1} . The gradient on attenuation parameters is given by

$$\frac{\partial \chi(\mathbf{m})}{\partial Q_{p,s}^{-1}} = \left(\overline{\mathbf{W}}, \frac{F(\mathbf{m}, \mathbf{W})}{\partial Q_{p,s}^{-1}} \right). \quad (3.52)$$

The derivative of the forward problem (3.3) on the inverse of quality factors is given by

$$\frac{\partial F(\mathbf{m}, \mathbf{W})}{\partial Q_{p,s}^{-1}} = \overbrace{\begin{pmatrix} 0 & D \frac{\partial C^a}{\partial Q_{p,s}^{-1}} & \dots & D \frac{\partial C^a}{\partial Q_{p,s}^{-1}} \\ 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 \end{pmatrix}}^{\partial \mathbf{B}_0(\mathbf{m})/\partial Q_{p,s}^{-1}} \begin{bmatrix} \mathbf{u} \\ \psi_1 \\ \dots \\ \psi_L \end{bmatrix} = \begin{bmatrix} \sum_{s=1}^L D \frac{\partial C^a}{\partial Q_{p,s}^{-1}} \psi_\nu \\ 0 \\ \dots \\ 0 \end{bmatrix}. \quad (3.53)$$

By considering the Lagrangian multiplier vector $\overline{\mathbf{W}}$ in Equation (3.10), the gradient on inverse quality factors in Equation (3.52) becomes

$$\frac{\partial \chi(\mathbf{m})}{\partial Q_{p,s}^{-1}} = \left(\begin{bmatrix} \overline{\mathbf{u}} \\ \overline{\phi}_1 \\ \dots \\ \overline{\phi}_L \end{bmatrix}, \begin{bmatrix} \sum_{\nu=1}^L D \frac{\partial C^a}{\partial Q_{p,s}^{-1}} \psi_\nu \\ 0 \\ \dots \\ 0 \end{bmatrix} \right) \quad (3.54)$$

$$= \left(\overline{\mathbf{u}}, \sum_{\nu=1}^L D \frac{\partial C^a}{\partial Q_{p,s}^{-1}} \psi_\nu \right). \quad (3.55)$$

By considering the properties of adjoint operator in Equation (3.13), we obtain

$$\frac{\partial \chi(\mathbf{m})}{\partial Q_{p,s}^{-1}} = - \left(D^t \overline{\mathbf{u}}, \sum_{\nu=1}^L \frac{\partial C^a}{\partial Q_{p,s}^{-1}} \psi_\nu \right). \quad (3.56)$$

We introduce the adjoint strain $\overline{\boldsymbol{\varepsilon}} = D^t \overline{\mathbf{u}}$, the gradient on attenuation parameters can be written in a cleaner form as

$$\frac{\partial \chi(\mathbf{m})}{\partial Q_{p,s}^{-1}} = - \left(\overline{\boldsymbol{\varepsilon}}, \sum_{\nu=1}^L \frac{\partial C^a}{\partial Q_{p,s}^{-1}} \psi_\nu \right). \quad (3.57)$$

Gradient on stiffness coefficients

The gradient on stiffness coefficients C_{IJ} is given by

$$\frac{\partial \chi(\mathbf{m})}{\partial C_{IJ}} = \left(\overline{\mathbf{W}}, \frac{F(\mathbf{m}, \mathbf{W})}{\partial C_{IJ}} \right). \quad (3.58)$$

The derivative of the forward problem (3.3) on C_{IJ} coefficients is given by

$$\frac{\partial F(\mathbf{m}, \mathbf{W})}{\partial C_{IJ}} = \overbrace{\begin{pmatrix} -D \frac{\partial C}{\partial C_{IJ}} D^t & D \frac{\partial C^a}{\partial C_{IJ}} & \dots & D \frac{\partial C^a}{\partial C_{IJ}} \\ 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 \end{pmatrix}}^{\partial \mathbf{B}_0(\mathbf{m}) / \partial C_{IJ}} \begin{bmatrix} \mathbf{u} \\ \psi_1 \\ \dots \\ \psi_L \end{bmatrix} \quad (3.59)$$

$$= \begin{bmatrix} -D \frac{\partial C}{\partial C_{IJ}} D^t \mathbf{u} + \sum_{\nu=1}^L D \frac{\partial C^a}{\partial C_{IJ}} \psi_\nu \\ 0 \\ \dots \\ 0 \end{bmatrix}. \quad (3.60)$$

By considering the Lagrangian multiplier vector $\overline{\mathbf{W}}$ in Equation (3.10), the elementary C_{IJ} -gradient in Equation (3.58) becomes

$$\frac{\partial \chi(\mathbf{m})}{\partial C_{IJ}} = \left(\begin{bmatrix} \overline{\mathbf{u}} \\ \overline{\phi}_1 \\ \dots \\ \overline{\phi}_L \end{bmatrix}, \begin{bmatrix} -D \frac{\partial C}{\partial C_{IJ}} D^t \mathbf{u} + \sum_{\nu=1}^L D \frac{\partial C^a}{\partial C_{IJ}} \psi_\nu \\ 0 \\ \dots \\ 0 \end{bmatrix} \right), \quad (3.61)$$

$$= \left(\bar{\mathbf{u}}, -D \frac{\partial C}{\partial C_{IJ}} D^t \mathbf{u} \right) + \left(\bar{\mathbf{u}}, \sum_{\nu=1}^L D \frac{\partial C^a}{\partial C_{IJ}} \psi_{\nu} \right). \quad (3.62)$$

By considering the properties of adjoint operator in Equation (3.13), we obtain

$$\frac{\partial \chi(\mathbf{m})}{\partial C_{IJ}} = \left(D^t \bar{\mathbf{u}}, \frac{\partial C}{\partial C_{IJ}} D^t \mathbf{u} \right) - \left(D^t \bar{\mathbf{u}}, \sum_{\nu=1}^L \frac{\partial C^a}{\partial C_{IJ}} \psi_{\nu} \right). \quad (3.63)$$

We simplify this expression by considering the adjoint $\bar{\varepsilon} = D^t \bar{\mathbf{u}}$ and the forward strain $\varepsilon = D^t \mathbf{u}$, the gradient on stiffness coefficients C_{IJ} can be written as

$$\frac{\partial \chi(\mathbf{m})}{\partial C_{IJ}} = \left(\bar{\varepsilon}, \frac{\partial C}{\partial C_{IJ}} \varepsilon \right) - \left(\bar{\varepsilon}, \sum_{\nu=1}^L \frac{\partial C^a}{\partial C_{IJ}} \psi_{\nu} \right). \quad (3.64)$$

In the gradient expression (3.64), we separate the elastic rheology (the 1st term) with the attenuation mechanism (the 2nd term). The adjoint $\bar{\varepsilon}$ and incident fields ε, ψ_{ν} are available at each time-step, the gradients can thus be collected without any extra cost.

Without attenuation contribution, the second term in Equation (3.64) vanishes, leading to the elastic C_{IJ} gradient

$$\left. \frac{\partial \chi(\mathbf{m})}{\partial C_{IJ}} \right|_{elastic} = \left(\bar{\varepsilon}, \frac{\partial C}{\partial C_{IJ}} \varepsilon \right), \quad (3.65)$$

which is coherent with the gradient kernels proposed by Liu and Tromp (2006) for 2nd-order elastic system. Similar conclusion is draw for density gradient. However, it should be noted that our adjoint source is $-R^\dagger(R\mathcal{W} - d_{obs})$, which has opposite sign compared with the one in Liu and Tromp (2006), thus resulting in the opposite sign for the gradient expressions. The gradients on stiffness-coefficients C_{IJ} and inverse quality factors Q_p^{-1} and Q_s^{-1} are also consistent with the development for the 1st-order system by Yang et al. (2016a).

Chapter 4

HPC implementation in SEM46 code

Contents

4.1	Two-level parallelization	81
4.1.1	Scalability	82
4.1.2	Modeling efficiency compared with other modeling tools	83
4.2	Balancing memory requirement & computational cost for gradient building	84
4.3	Conclusion	86

Exploration scale applications generally consist of multiple shots and covering a large area. This chapter focuses on our two-level parallelization implementation for both modeling and inversion parts to handle such targets. We present the visco-elastic gradient estimation following the CARFS strategy (Yang et al., 2016b), relying on the trade-off between the the computation cost and memory requirement.

4.1 Two-level parallelization

Our implementation relies on a two-level MPI-based parallelization. The inner level is designed on a Cartesian-based domain decomposition as shown in Figure 4.1, which ensures an efficient load-balancing and an easy implementation. In particular, this avoids the use of a third-party mesh-partitioner,

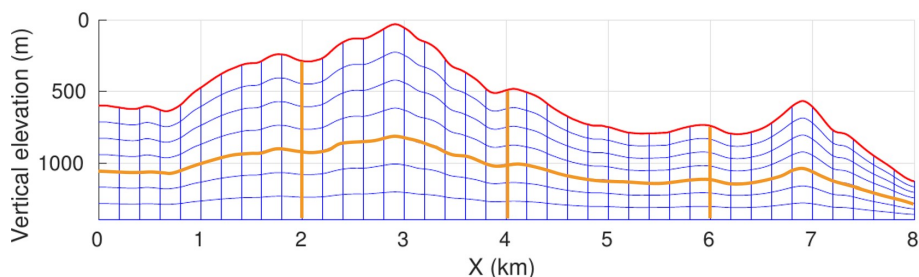


Figure 4.1: Example of a Cartesian-based domain decomposition where the boundaries between sub-domains are highlighted by orange lines.

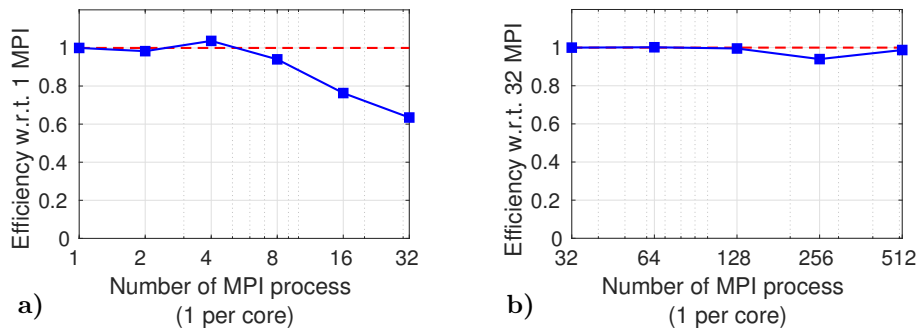


Figure 4.2: Strong scaling of the domain decomposition MPI-level for a $192 \times 192 \times 192$ elements application, on a Intel Omni-Path interconnect & Intel Xeon E5-2697 v4 node (36 cores/node) architecture: a) Intra-node efficiency, b) Inter-node efficiency, showing stable and good scaling property.

even if the number of possible sub-domains should be constrained by the mesh splitting in each direction. This parallel level is quite standard and efficient in SEM as gathering information through MPI communications is equivalent to the assembly procedure of FE schemes. This communication involves a single depth “layer” of degree of freedom at each domain interface whatever is the SEM interpolation order, different from FD schemes.

The outer MPI-level is performed over seismic shots managed in parallel. This level is embarrassingly parallel as the communications steps only involve the summation of misfit function and gradient contributions of each seismic shot, and the scattering of physical parameters models and descent directions. Practical applications show that this second MPI-based level generally satisfies the perfect theoretical efficiency. It has to be noted that, compared to other SEM implementations relying on a single shot framework, our frame generally involves a large number of shots in parallel. The domain decomposition per shot thus contains few tens of CPU-cores, leading to a strong memory and computational load per core. In practical exploration FWI-liked applications, the communication between sub-domains is negligible compared with the computation on each domain. Asynchronous communication would thus only provide marginal savings.

4.1.1 Scalability

The parallel efficiency of our implementation has been assessed on different hardware architectures (Intel CPU-based) with different interconnect technologies, showing very good properties. An example of strong scaling for the domain decomposition level is conducted on a benchmark with $192 \times 192 \times 192$ elements (4.5×10^8 degrees of freedom (dof)). The efficiency is assessed from an architecture embedding Intel Omni-Path interconnect & Intel Xeon E5-2697 v4 nodes (36 cores/node). Figure 4.2(a) shows the efficiency for intra-node parallelization. As usual for PDE-based numerical schemes, the memory bandwidth appears to be a strong bottleneck: for low number of MPI process, we benefit from the whole memory bandwidth leading to artificially good performances (scalability larger than 90% up to 8 process). When the number of MPI process is increased, memory-bandwidth-bound effects appear, leading to an efficiency decrease (down to 60% for 32 process). This effect is directly associated with the memory-bandwidth-bound, and when the hardware is fully used as in production, only the 32 processes number should be considered as a reference. In Figure 4.2(b), the strong scaling is pushed up to 512 MPI process, showing stable and good scaling property, above 94%.

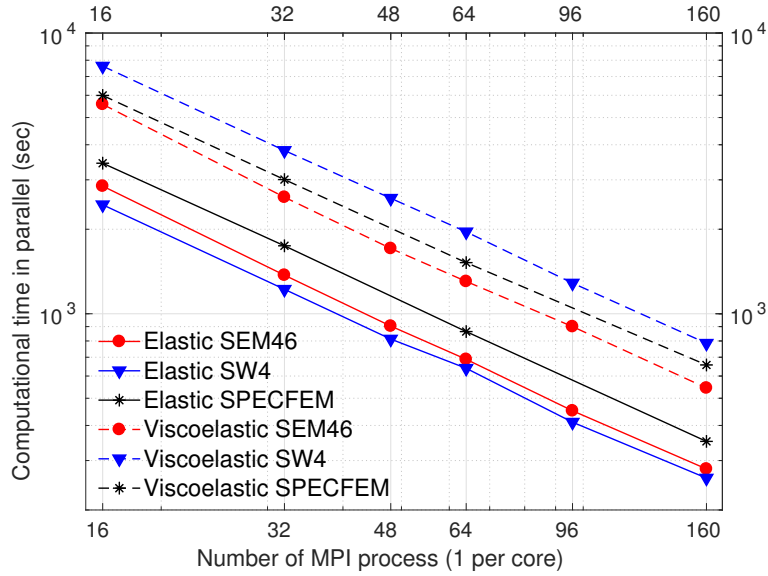


Figure 4.3: Comparable behavior of numerical efficiency for different implementations of elastic and visco-elastic forward modeling over different domain decomposition: SEM46, SW4 V1.1 and SPECFEM V2.0 (CPU-based) packages.

4.1.2 Modeling efficiency compared with other modeling tools

We compare the overall behavior of computational efficiency for 3D elastic and visco-elastic seismic modeling of SEM46 with two open-source programs: SW4 V1.1 and SPECFEM V2.0 (CPU version). SW4 uses a 4th-order FD approach (Sjögreen and Petersson, 2012) while SPECFEM relies on a SEM approach (Peter et al., 2011).

We use a homogeneous model for the comparison. For visco-elastic comparison, in each code, quality factors are set as $Q_p = Q_s = 40$, described by 3 SLSs. In most practical applications, SW4 recommends to use at least 6 points per shortest wavelength for an acceptable accuracy, leading to the grid sampling at 25 m (Petersson and Sjögreen, 2013). Since around 6 GLL points per shortest wavelength allows to accurately model elastic waves, we decide to consider the same number of dof (1.8×10^8) for both FD and SEM meshes. In this case, the SEM mesh is filled with 100 m elements and the interpolation order $N = 4$ is used for both SEM46 and SPECFEM.

The total recording time is 6 sec, leading to 908 time-steps in SW4 and 4000 time-steps in SEM46, due to the CFL condition. A similar configuration is considered for SPECFEM with identical number of dof and time-steps as for SEM46. Figure 4.3 compares the computational cost of SEM46 with the one of SW4 and SPECFEM for parallelism over different domain decomposition settings. Even with much larger number of time-steps, SEM implementations exhibit comparable computational cost as SW4 for both elastic and visco-elastic modeling.

In this comparison, SEM46 uses a triclinic elasticity coupled with an isotropic attenuation implementation, while isotropic elasticity and isotropic attenuation is considered in SPECFEM. The anelastic effect is described by independent Q_p and Q_s quality factors in SEM46 as described in Section (2.1.2.3). Only isotropic elasticity and shear quality factor Q_s is taken into account by the considered version V2.0 of SPECFEM, which somehow simplifies the numerical implementation. Even though,

our SEM46 implementation exhibits comparable computational performances as the ones of SPEC-FEM for the forward 3D elastic and visco-elastic simulations, under the same computer and compilation environment.

4.2 Balancing memory requirement & computational cost for gradient building

FWI gradient estimation, required as the input of the optimization process, involves the zero-lag cross-correlation in time between the incident and adjoint wavefields, weighted by the spatial radiation pattern (Equations 3.21, 3.22, 3.25). The adjoint-state approach is thus considered for the gradient estimation, which requires the access to the incident and adjoint wavefields at the same time (Plessix, 2006). We might store the incident wavefield in the core-memory or out-of-core memory at the expense of the I/O cost (Boehm et al., 2016). To avoid heavy memory requirement, the incident field can be re-computed from the last snapshot and the boundary wavefield simultaneously with the propagation of the adjoint field. For elastic medium, the incident field can be perfectly reconstructed by the reverse propagation in time from the last snapshot and the stored wavefield at the boundaries, synchronously with the propagation of the adjoint field backward in time. Gradients are directly accumulated during this process, resulting in a cheap operation with limited I/O operations (Dussaud et al., 2008).

Typical SEM simulation requires a number of time-steps to the order of 10^4 , which might lead to significant memory requirement for storing the boundary wavefields. Our implementation relies on compressed boundary storage to mitigate the memory cost (Yang et al., 2016c). The time-stepping given by the CFL condition is generally much smaller than the Nyquist requirement, and compression technique can be quite efficient: during the forward propagation, the boundary wavefield is saved at every r time-step ($r \gg 1$), where r is the *decimation ratio*. When reconstructing the incident field, at each time-step, the boundary wavefield is interpolated from the decimated time series through Lagrange polynomial interpolation. Numerical experiments show that only the displacement field \mathbf{u} (and not its spatial derivatives, nor time derivatives) needs to be saved at boundaries, after the prediction phase in the visco-elastic Newmark time-marching scheme (Algorithm 1), for an accurate reconstruction of the incident displacement \mathbf{u} and velocity \mathbf{v} fields.

In anelastic medium, the propagation of the adjoint field backward in time (Equation 3.20) is numerically stable, as the forward propagation of the incident fields (Equation 2.23), because they share the same equations. However, the irreversibility of the visco-elastic wave equation makes the reverse propagation of the incident field unstable, which requires specific strategies for efficient gradient accumulation (Tarantola, 1988; Griewank and Walther, 2000; Liu and Tromp, 2008; Komatitsch et al.,

Algorithm 4: Forward propagation

```

1 for  $it = 1, \dots, nt$  do
2   Forward propagation of the incident field:  $F^{it} : \mathcal{W}^{it-1} \rightarrow \mathcal{W}^{it}$ ;
3   Record the reference strain-energy  $E_{ref}^{it}$  for  $\mathcal{W}^{it}$ ;
4   if (At checkpoint position) then Store the snapshot  $\mathbf{u}^{it}, \mathbf{v}^{it}, \psi_\nu^{it-1/2}$ ;
5   if (At decimated location:  $mod(it,r)=0$ ) then Store the boundary of the displacement field
       $\mathbf{u}^{it}$ ;
6 end

```

Algorithm 5: Gradient estimation

```

1 for  $it = nt - 1, \dots, 1$  do
2   Adjoint time-step:  $at = nt - it$  ;
3   Forward propagation of the adjoint field:  $F^{at} : \bar{\mathcal{W}}^{at-1} \rightarrow \bar{\mathcal{W}}^{at}$  ;
4   if (At checkpoint position) then
5     Read the snapshot  $\mathbf{u}^{it}, \mathbf{v}^{it}, \psi_\nu^{it-1/2}$  ;
6   else
7     Interpolate the boundary of the displacement field  $\mathbf{u}^{it}$  ;
8     Backward propagation of the incident field:  $(F^{it})^{-1} : \mathcal{W}^{it+1} \rightarrow \mathcal{W}^{it}$  ;
9     Measure the energy of the reconstructed wavefield  $E_{rec}^{it}$ ;
10    if (Instability:  $|E_{ref}^{it} - E_{rec}^{it}| > tol \cdot E_{ref}^{it}$ ) then
11      Forward propagation from the closest checkpoint prior to the current time-level  $it$  ;
12      Replace the used checkpoints by new snapshot positions;
13    end
14  end
15  Gradient collection (Equation 3.21, 3.22, 3.25);
16 end

```

2016). To overcome this issue, we implement a visco-elastic version of the checkpointing-assisted reverse-forward simulation (CARFS) algorithm (Yang et al., 2016b), which makes a smart decision between the reverse modeling and the forward modeling using checkpoints based on the strain-energy measure. Similar to standard checkpointing techniques, the checkpoints distribution is computed from the binomial law, the optimal number of checkpoint being related to the total number of time-steps nt by $\log_2(nt)$. For each checkpoint, we store the displacement, velocity and memory variable fields $\mathbf{u}^{it}, \mathbf{v}^{it}, \psi_\nu^{it-1/2}$. During the propagation of the incident field (Algorithm 4), at each time-step, a reference global strain energy is also recorded

$$E_{ref}^{it} = \frac{1}{2} \left(\boldsymbol{\sigma}^{it} \boldsymbol{\varepsilon}^{it} \right)_\Omega. \quad (4.1)$$

This global measure is used for monitoring the stability of the reverse simulation of the incident wavefield based on a pre-defined energy tolerance: For instance, $tol = 0.1$ is used in our applications as shown in Algorithm 5. Above this tolerance, the nearest earlier in time checkpoint is considered. A forward simulation from this checkpoint until the current time level is performed. During this forward simulation, the checkpoints are also redistributed. Once done, the reverse simulation of the incident wavefield together with the propagation of the adjoint wavefield backward in time can be continued.

The combination of this compression technique through decimation and interpolation with the CARFS approach provides a good balance between the memory requirement, simulation accuracy and the computational cost, which is directly linked to the recomputation ratio of the incident field. Additional compression strategies could also be used to further reduce the memory requirement and I/O requests (Boehm et al., 2016).

4.3 Conclusion

Our integrated SEM-based workflow is capable of efficiently performing 3D elastic and visco-elastic modeling and inversion in the time-domain. Two MPI-based parallelism levels are considered for tackling large-scale and multiple-shots experiments. Thanks to the use of a Cartesian-based mesh, the domain decomposition and communication between sub-domains are straightforward. The strong scaling tests show stable and good behavior, above 94%. For the same physical model and similar numerical settings, we achieve comparable computational cost as two open-source reference codes: SW4 V1.1 and SPECFEM V2.0 (CPU version) for both elastic and visco-elastic simulations, for different domain-decomposition configuration.

For the gradient calculation, the incident field is reconstructed from the stored boundary wavefield in the memory, synchronously with the propagation of the adjoint field. The gradient is directly accumulated during this process, resulting in a cheap operation (Dussaud et al., 2008). To reduce the memory requirement, the boundary wavefield is saved at decimated positions in time and re-interpolated when needed (Yang et al., 2016c). In visco-elastic medium, the instability in the reconstruction of the incident wavefield can be mitigated by using the elastic version of CARFS (Yang et al., 2016b). This strategy is critical in realistic applications, normally associated with a large number of degrees of freedom and time-steps, in which massive storage and intensive I/O might not be the optimal choice.

Future works include considering GPU implementation, which is expected to drastically improve the computational efficiency (Komatitsch et al., 2010; Michéa and Komatitsch, 2010).

Chapter 5

Structure-oriented Bessel gradient preconditioning

Contents

5.1	Introduction	89
5.2	Methodology	90
5.2.1	Definition of Bessel filters and their sparse inverse operators	91
5.2.2	Approximation of the Laplace filter by the Bessel operator	92
5.3	Weak formulation in finite element methods	93
5.3.1	Weak formulation of the filter for homogeneous coherent lengths without dip and azimuth	94
5.3.2	Variable coherent lengths, dip and azimuth angles	96
5.4	Numerical illustrations	99
5.4.1	Spike test	99
5.4.2	Random noise tests	101
5.4.3	FWI gradient smoothing	105
5.5	Numerical implementation in a parallel spectral-element method scheme	106
5.5.1	Linear system construction and matrix-vector product evaluation	107
5.5.2	Parallel conjugate gradient iterative solver	108
5.6	Conclusion and discussion	112
5.7	Additional numerical investigations	113
5.7.1	Gradient smoothing in pseudo-3D Marmousi model	114
5.7.2	Influence of the preconditioning operator on inversion results	115
5.A	Modified Bessel functions and their integral representation	117
5.B	Definition of Bessel filters and their normalized factors	117
5.C	Weak formulation development	120
5.D	Matrix-free finite-difference implementation scheme	123

The FWI problem under standard surface acquisition is mathematically ill-posed and poorly constrained, leading to significant artifacts (generally high wavenumber) in the estimated perturbation model. As specified in Section (3.2), the perturbation estimation directly relies on the gradient of the misfit function. This chapter will describe both the theoretical development and numerical implementation of a gradient preconditioning strategy, based on the Bessel function, in the form of a published paper in the journal *Geophysical Journal International* (Trinh et al., 2017b). The described smoothing operator is directly implemented on the SEM mesh in which the local geological prior information can be incorporated through the 3D anisotropic filter shape and local rotation. Numerical illustrations on the influence of the preconditioning operator on gradient and inversion results are extracted from *extended EAGE* and *SEG abstracts* (Trinh et al., 2017a,c).

The Bessel gradient preconditioning described in this chapter is not related to the optimization preconditioning in Section (3.2.3), which is an approximation of the Hessian to accelerate the convergence of the iterative inversion process. Both strategies should be considered together to speed up the convergence and also to guide the inversion toward the desired solutions.

Bessel smoothing filter for spectral-element mesh

P.T. Trinh, R. Brossier, L. Métivier, J. Virieux, P. Wellington

Geophysical Journal International, **209**, 1489-1512

Abstract

Smoothing filters are extremely important tools in seismic imaging and inversion, such as for travel-time tomography, migration and waveform inversion. For efficiency, and as they can be used a number of times during inversion, it is important that these filters can easily incorporate prior information on the geological structure of the investigated medium, through variable coherent lengths and orientation. In this study, we promote the use of the Bessel filter to achieve these purposes. Instead of considering the direct application of the filter, we demonstrate that we can rely on the equation associated with its inverse filter, which amounts to the solution of an elliptic partial differential equation. This enhances the efficiency of the filter application, and also its flexibility. We apply this strategy within a spectral-element-based elastic full waveform inversion framework. Taking advantage of this formulation, we apply the Bessel filter by solving the associated partial differential equation directly on the spectral-element mesh through the standard weak formulation. This avoids cumbersome projection operators between the spectral-element mesh and a regular Cartesian grid, or expensive explicit windowed convolution on the finite-element mesh, which is often used for applying smoothing operators. The associated linear system is solved efficiently through a parallel conjugate gradient algorithm, in which the matrix vector product is factorized and highly optimized with vectorized computation. Significant scaling behavior is obtained when comparing this strategy with the explicit convolution method. The theoretical numerical complexity of this approach increases linearly with the coherent length, whereas a sub-linear relationship is observed practically. Numerical illustrations are provided here for schematic examples, and for a more realistic elastic full waveform inversion gradient smoothing on the SEAM II benchmark model. These examples illustrate well the efficiency and flexibility of the approach proposed.

5.1 Introduction

Full waveform inversion (FWI) offers the possibility to extract high-resolution quantitative multi-parameters of the subsurface from seismic data. The majority of these applications are carried out under finite-difference (FD) approximation, due to the numerical efficiency of this method and its ease of implementation. Standard formulations of this approach are, however, limited on regular grids, which require significant extra effort in terms of design and computational cost in the presence of surface topography or important geological interfaces (Robertsson, 1996; Bohlen and Saenger, 2006; Huiskes et al., 2016; Fuji et al., 2016). Finite-element (FE) methods have become popular for regional and global problems, especially spectral-element methods (SEM), where complex geometry can be handled with accurate numerical calculation of wavefields (Komatitsch and Tromp, 1999). In dealing with 3-D elastic FWI, we would like to develop a complete inversion numerical workflow using a spectral FE scheme while including an efficient smoothing filter implemented directly on the FE mesh.

In most geophysical applications, FWI is introduced as an iterative local optimization problem that attempts to minimize the least-squares residuals between the observed and the calculated data at the receiver location. This inverse problem is mathematically ill-posed, which thus leads to the non-uniqueness of the solution. For seismic imaging using either travel times or wavefields, the inversion often needs to be stabilized by applying regularization. This can be performed through model-driven regularization (Menke, 1984; Tarantola, 2005) or by preconditioning the gradient through filtering operators (Guitton et al., 2012). This data-driven strategy is sometimes called model preconditioning (Fomel and Claerbout, 2003). These regularization techniques normally assume the particular properties or structure of the model, such as smoothness or geological features, to guide the problem toward the desired solution. In 3-D elastic FWI, the computational cost remains one of the main challenges (Virieux and Operto, 2009), therefore decimation of the data volume by reducing the number of computed shots, random subsampling strategies, and/or source-encoding becomes mandatory (Capdeville et al., 2005; Krebs et al., 2009; Herrmann et al., 2009; Warner et al., 2013; Castellanos et al., 2015). This implies a decrease in the data coverage and the presence of cross-talk and artifacts, which makes the inversion problem even more ill-posed. In this context, the role of regularization at each iteration becomes crucial and needs to be performed efficiently.

Dealing with complex structures or geological heterogeneity might also require adequate regularization strategies. These issues can be addressed by attenuation of the wave-number content in a particular direction, such as plane-wave destruction filters (Fomel, 2002; Claerbout, 1992), or by imposition of expected structures through directional Laplacian preconditioning in the model space (Hale, 2007; Guitton et al., 2012). This latter highlights the importance of non-stationary filters in complex geology; for example, the Laplace filter can smooth or drastically reduce local planar events according to a local dip field.

These nonstationary preconditioning filters are expressed as convolution operators, such as Gaussian or Laplacian filters, through

$$\mathbf{s}(\mathbf{x}) = \int C_{nD}(\mathbf{x} - \mathbf{x}')\mathbf{m}(\mathbf{x}')d\mathbf{x}'. \quad (5.1)$$

The vector $\mathbf{m}(\mathbf{x})$ is transformed into a new vector $\mathbf{s}(\mathbf{x})$ through the convolution operator C_{nD} , where n is the dimension of the problem ($n = 2$ or $n = 3$). In the FD grid, the Gaussian filter (without local rotation) can be applied efficiently to any model or gradient vector, due to the tensorial property of the function. For long filters, recursive implementation can be used to improve the computing performance (Deriche, 1992; Van Vliet et al., 1998). These approaches can be extended to FE methods by including

a projection between the Cartesian mesh, where the parameters to be reconstructed are located, and the FE mesh. However, this can be limited by the accuracy of the back and forth projections. Filtering can also be applied as a windowed convolution of the filter and the vector (Tape et al., 2010; Peter et al., 2011), as for the SPECFEM open-source package

$$\mathbf{s}(\mathbf{x}) \approx \int_{-\alpha L}^{\alpha L} C_{nD}(\mathbf{x} - \mathbf{x}') \mathbf{m}(\mathbf{x}') d\mathbf{x}', \quad (5.2)$$

where L is the coherent length associated with the filter: as the kernel of the filter is decaying, the integration can be limited over the finite domain expressed by the effective radius αL from the position \mathbf{x} . This convolution approach can be relatively computer demanding, because for each input point to be filtered, contributions of other points of the medium are required in the surrounding volume, which leads to significant computer manipulation, especially for functions with long tails, such as the Laplace filter.

In this report, we highlight that the inverse of given operators C_{nD} can be relatively sparse. Therefore, it is interesting to consider them to obtain the contribution of the smoothing operator (Wellington, 2016). Instead of performing the convolution in Equation (5.1), we can consider solving the following Equation (5.3), relying on the inverse operator

$$\int C_{nD}^{-1}(\mathbf{x} - \mathbf{x}') \mathbf{s}(\mathbf{x}') d\mathbf{x}' = \mathbf{m}(\mathbf{x}), \quad (5.3)$$

which requires knowledge of the inverse operator C_{nD}^{-1} . This leads to the definition of Bessel filters, which are defined by the modified Bessel functions (Abramowitz and Stegun, 1972) and their sparse inverse operators. As the filter is defined through a nonhomogeneous elliptic partial differential equation (PDE) with delta source function, its inverse operator can be expressed as a distribution function. The integral Equation (5.3) can then be efficiently solved using any FD or FE method. In 3-D, applying the Bessel filter twice provides an excellent approximation of the Laplace operator, with negligible mismatch at the origin. This approximation implies that the Bessel filter can be applied either once as a smoothing filter, or twice to reproduce the decay of the Laplace filter.

This paper is organized as follows. In section 2, we first define the Bessel filter through a PDE with specific boundary conditions, and the approximation of the Laplace filter by Bessel functions in 2-D and 3-D geometries. In section 3, the application of the Bessel sparse inverse operator on a vector is described by a system of PDEs. The weak formulation of these equations yields a sparse linear system, which is symmetric positive even for variable coherent lengths, dip and azimuth angles. In section 4, several numerical illustrations and an example from the synthetic 3-D SEAM Phase II foothills model are provided of such smoothing processes. In section 5, efficient numerical schemes in a high-performance computer environment are presented. We show how a parallel conjugate gradient (CG) iterative solver can be implemented in a matrix-free fashion. The convergence of this CG iterative solver is analysed with respect to several parameters. Significant scaling behavior is also obtained when this strategy is compared with the explicit convolution method. Conclusions and perspectives are given in section 6. We highlight that the choice of applying the Bessel filter once or twice depends on the specific application, and this decision will be case dependent.

5.2 Methodology

Before moving to the mathematical development, we would like to specify some definitions regarding forward and inverse filters. Let us consider the kernel $C(\mathbf{x})$ as a 2-D or 3-D smoothing filter. The

application of this filter to a function $g(\mathbf{x})$ is defined by the convolution operator, which is denoted by the symbol “*”, through

$$f(\mathbf{x}) = C(\mathbf{x}) * g(\mathbf{x}), \quad \text{which is equivalent to} \quad f(\mathbf{x}) = \int C(\mathbf{x} - \mathbf{x}')g(\mathbf{x}')d\mathbf{x}'. \quad (5.4)$$

The inverse filter of $C(\mathbf{x})$, namely $C(\mathbf{x})^{-1}$, is defined through the relation

$$C^{-1}(\mathbf{x}) * C(\mathbf{x}) = \delta(\mathbf{x}). \quad (5.5)$$

5.2.1 Definition of Bessel filters and their sparse inverse operators

Depending on the space dimension, we introduce the normalized Bessel filters

$$B_{3D}(z, x, y) = \frac{1}{(2\pi)^{3/2}L_zL_xL_y}r^{-1/2}K_{1/2}(r) \quad \text{where} \quad r = \sqrt{\frac{z^2}{L_z^2} + \frac{x^2}{L_x^2} + \frac{y^2}{L_y^2}}, \quad (5.6)$$

$$B_{2D}(z, x) = \frac{1}{2\pi L_zL_x}K_0\left(\sqrt{\frac{z^2}{L_z^2} + \frac{x^2}{L_x^2}}\right), \quad (5.7)$$

where L_x , L_y and L_z are coherent lengths in the x , y and z directions, and K_ν is the modified Bessel function of the second kind, as presented in Appendix 5.A. While the 2-D Bessel filter is the normalized modified Bessel function $K_\nu(r)$ with $\nu = 0$, the 3-D Bessel filter is the modified spherical Bessel function $r^{-1/2}K_{\nu+1/2}(r)$ with $\nu = 0$ (Abramowitz and Stegun, 1972). When coherent lengths are uniform over space, these Bessel filters are unique solutions of the following PDEs

$$B_{3D}(z, x, y) - \left(L_z^2 \frac{\partial^2}{\partial z^2} + L_x^2 \frac{\partial^2}{\partial x^2} + L_y^2 \frac{\partial^2}{\partial y^2}\right)B_{3D}(z, x, y) = \delta(z, x, y), \quad (5.8)$$

$$B_{2D}(z, x) - \left(L_z^2 \frac{\partial^2}{\partial z^2} + L_x^2 \frac{\partial^2}{\partial x^2}\right)B_{2D}(z, x) = \delta(z, x), \quad (5.9)$$

with radiative boundary conditions $B_{nD}(\mathbf{x}) \rightarrow 0$ as $\|\mathbf{x}\| \rightarrow \infty$, and the additional bounded constraint on the integral through $0 < \int B_{nD}(\mathbf{x})d\mathbf{x} < \infty$. The construction of these normalized Bessel filters is developed in Appendix 5.B.

According to eqs (5.8) and (5.9) and the definition of the inverse filter in Equation (5.5), the associated sparse inverse of the Bessel filters can be defined as

$$B_{3D}^{-1}(z, x, y) = \delta^0(z, x, y) - \left(L_z^2 \delta^2(z)\delta^0(x)\delta^0(y) + L_x^2 \delta^2(x)\delta^0(y)\delta^0(z) + L_y^2 \delta^2(y)\delta^0(x)\delta^0(z)\right) \quad (5.10)$$

$$B_{2D}^{-1}(z, x) = \delta^0(z, x) - \left(L_z^2 \delta^2(z)\delta^0(x) + L_x^2 \delta^2(x)\delta^0(z)\right), \quad (5.11)$$

where δ^n is defined as

$$\int \delta^n(x - x')f(x')dx' = (-1)^n \frac{d^n f(x)}{dx^n}. \quad (5.12)$$

It is important to note that Bessel filters only depend on the radial part (Appendix 5.B), which favors the transformation from Cartesian coordinates to polar coordinates (r, θ) in 2-D, or to spherical coordinates (r, θ, ϕ) in 3-D

$$(z, x) \xrightarrow{2D} \sqrt{\frac{z^2}{L_z^2} + \frac{x^2}{L_x^2}}; \quad (z, x, y) \xrightarrow{3D} \sqrt{\frac{z^2}{L_z^2} + \frac{x^2}{L_x^2} + \frac{y^2}{L_y^2}}. \quad (5.13)$$

Let us consider an original vector \mathbf{m} in the model space Ω that we want to smooth. The vector \mathbf{s} will be the smoothed vector obtained by applying the Bessel filter $B_{3D}(z, x, y)$. We can equivalently consider the application of the inverse Bessel filter, which leads to the choice between the solutions of the following two problems:

$$\mathbf{s} = B_{3D} * \mathbf{m} \quad \text{or} \quad \mathbf{m} = B_{3D}^{-1} * \mathbf{s}. \quad (5.14)$$

Both systems are convolution over space. The first one embeds a relatively broad kernel shape, where the computation can be demanding. Therefore, we are interested in the second equation. According to the definition of the inverse operator B_{3D}^{-1} (Equation (5.10)), this convolution can be translated into the following PDE over the domain Ω

$$\mathbf{s}(z, x, y) - \left(L_z^2 \frac{\partial^2}{\partial z^2} + L_x^2 \frac{\partial^2}{\partial x^2} + L_y^2 \frac{\partial^2}{\partial y^2} \right) \mathbf{s}(z, x, y) = \mathbf{m}(z, x, y), \quad (5.15)$$

in which the vector \mathbf{m} appears on the right-hand side. Once discretized by any FD or FE method, this system leads to a sparse operator and can be solved efficiently, to obtain the smoothed vector \mathbf{s} . Our approach is similar in some ways to the structure-oriented smoothing filter proposed by Williamson et al. (2011), which was derived from the diffusion equation as proposed by Fehmers and Höcker (2003). This smoothing process is controlled by a symmetric diffusion tensor, without knowing the kernel shape of the forward filter.

5.2.2 Approximation of the Laplace filter by the Bessel operator

The convolution of two Bessel filters can provide an excellent approximation of the Laplace filter in 2-D and 3-D. We first illustrate this for the 3-D case, as this is our main interest. The double operator $B_{3D}^{-1} * B_{3D}^{-1}$ is the inverse filter of the operator $B_{3D} * B_{3D}$, which is the solution of the following equation

$$[B_{3D}^{-1} * B_{3D}^{-1}(z, x, y)] * f(z, x, y) = \delta(z, x, y), \quad (5.16)$$

where the function f satisfies the following PDE

$$\left[1 - \left(L_z^2 \frac{\partial^2}{\partial z^2} + L_x^2 \frac{\partial^2}{\partial x^2} + L_y^2 \frac{\partial^2}{\partial y^2} \right) \right]^2 f(z, x, y) = \delta(z, x, y). \quad (5.17)$$

Changing parameters from Cartesian coordinates to spherical coordinates, where

$$r = \sqrt{\frac{z^2}{L_z^2} + \frac{x^2}{L_x^2} + \frac{y^2}{L_y^2}}, \quad (5.18)$$

and knowing that the expected solution, the filter $B_{3D} * B_{3D}$, only depends on the radial part, Equation (5.17) now becomes

$$\left[1 - \frac{\partial^2}{\partial r^2} \right] f(r) = \frac{\delta(r)}{4\pi r^2} + \underbrace{\left[\frac{\partial^2}{\partial r^2} + \frac{4}{r} \frac{\partial}{\partial r} - \frac{\partial^4}{\partial r^4} - \frac{4}{r} \frac{\partial^3}{\partial r^3} - \frac{4}{r^2} \frac{\partial^2}{\partial r^2} \right]}_{M_{3D}[f(r)]} f(r). \quad (5.19)$$

The normalized 3-D Laplace filter $L_{3D}(z, x, y)$ is given by

$$L_{3D}(z, x, y) = \frac{1}{8\pi L_z L_x L_y} e^{-\sqrt{\frac{z^2}{L_z^2} + \frac{x^2}{L_x^2} + \frac{y^2}{L_y^2}}} \quad \text{or} \quad L_{3D}(r) = \frac{1}{8\pi L_z L_x L_y} e^{-r}. \quad (5.20)$$

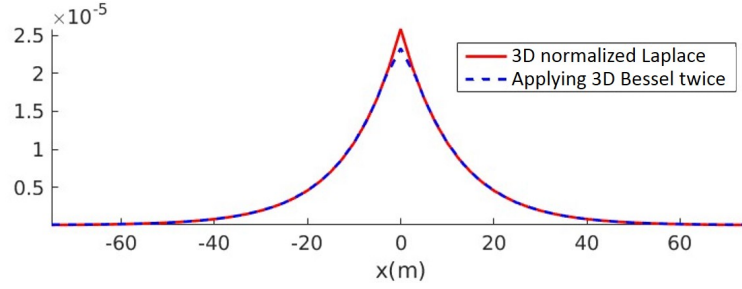


Figure 5.1: Comparison between the behavior of the 3-D normalized Laplace filter defined in Equation (5.20) and the convolution of two 3-D Bessel filters for $L_x = L_y = L_z = 20$ m, obtained from the iterative solving of Equation (5.8), which defines the 3-D Bessel filter.

This is the solution of Equation (5.19), for $r > 0$, provided the residual term $M_{3D}[L_{3D}(r)]$ can be ignored, as

$$L_{3D}(r) - \frac{\partial^2}{\partial r^2} L_{3D}(r) = 0. \quad (5.21)$$

The residual term M_{3D} for the 3-D Laplace filter satisfies

$$\left\| \frac{M_{3D}[L_{3D}(r)]}{L_{3D}} \right\| = \frac{4}{r^2}, \quad (5.22)$$

which implies that the residual term becomes smaller when the distance r increases. Figure 5.1 presents an excellent match between the 3-D normalized Laplace filter defined in Equation (5.20) and the convolution of the two 3-D Bessel filters for $L_x = L_y = L_z = 20$ m, obtained from solving the PDE defining the 3-D Bessel filter. The mismatch at zero origin comes from the residual terms $M_{3D}[L_{3D}(r)]$, which are more important at small distances r . This analysis is also coherent with the 3-D inverse operator of the Laplace filter proposed by Tarantola (2005).

In 2-D, by applying the same workflow, we can find a similar approximation. However, we would like to highlight the flexibility of this approximation by introducing a scaling parameter a into the definition of the 2-D normalized Laplace filter

$$L_{2D}(z, x, a) = \frac{1}{2\pi a^2 L_x L_z} e^{-\sqrt{\frac{z^2}{a^2 L_z^2} + \frac{x^2}{a^2 L_x^2}}} \quad \text{or} \quad L_{2D}(r) = \frac{1}{2\pi a^2 L_x L_z} e^{-r}. \quad (5.23)$$

This 2-D Laplace operator can be well approximated by the application of two 2-D Bessel filters, as shown in Figure 5.2, except at the origin where there is a singularity. By using different values of the scaling factor a , we have different approximations of the Laplace filter, which mitigates the discrepancies for both the amplitude and the decay.

5.3 Weak formulation in finite element methods

Following the weak formulation, as usual for FE methods, we can discretize Equation (5.15) and deduce a sparse linear system that can be efficiently solved using an iterative linear solver. First, we start by introduction of the weak formulation in SEM for homogeneous coherent lengths without dip and azimuth, which naturally leads to a symmetric linear system. We then introduce an efficient method to incorporate variable coherent lengths, and dip and azimuth angles into our linear system, while

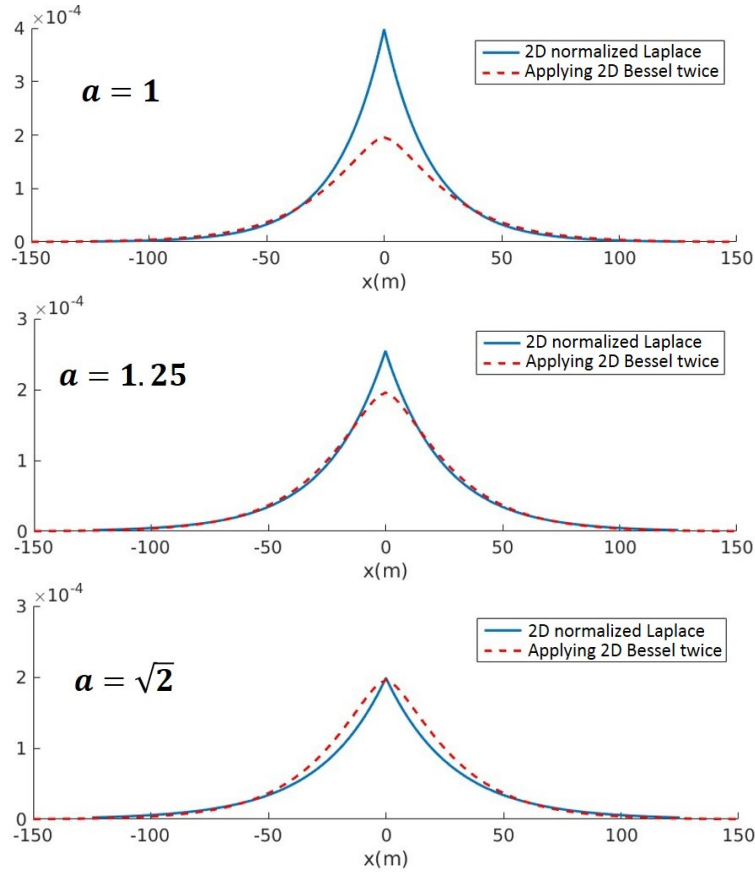


Figure 5.2: Comparison between the 2-D normalized Laplace filter defined in Equation (5.23) and the application of two 2-D Bessel filters for $L_x = L_z = 20$ m at different values of the scaling parameter a . The Bessel filters are obtained from the iterative solving of Equation (5.9), which defines the 2-D Bessel filter.

preserving the symmetry of the left-hand-side matrix. Although the following was developed for SEM, it can be extended to other FE formulations, and also for FD methods (Wellington, 2016).

5.3.1 Weak formulation of the filter for homogeneous coherent lengths without dip and azimuth

We do not develop the weak formulation of Equation (5.15) in Cartesian space, but in the dimensionless coordinates in the domain $\tilde{\Omega}$, as defined by the following expressions

$$\tilde{z} = \frac{z}{L_z}, \quad \tilde{x} = \frac{x}{L_x}, \quad \tilde{y} = \frac{y}{L_y}. \quad (5.24)$$

As coherent lengths L_z , L_x and L_y are homogeneous, the relationship of Equation (5.24) is an one-to-one projection from (z, x, y) to $(\tilde{z}, \tilde{x}, \tilde{y})$, which implies that solving Equation (5.15) in dimensionless or Cartesian coordinates should provide identical results. In the dimensionless coordinates, Equation (5.15) can be written in its weak form as

$$\iiint_{\tilde{\Omega}} [\mathbf{s} - \nabla_{\tilde{z}, \tilde{x}, \tilde{y}}^2 \mathbf{s}] v \, d\tilde{z} d\tilde{x} d\tilde{y} = \iiint_{\tilde{\Omega}} \mathbf{m} v \, d\tilde{z} d\tilde{x} d\tilde{y}, \quad (5.25)$$

leading to the weak formulation

$$\iiint_{\tilde{\Omega}} \mathbf{s}v \, d\tilde{z}d\tilde{x}d\tilde{y} + \iiint_{\tilde{\Omega}} \nabla_{\tilde{z},\tilde{x},\tilde{y}} \mathbf{s} \nabla_{\tilde{z},\tilde{x},\tilde{y}} v \, d\tilde{z}d\tilde{x}d\tilde{y} = \iiint_{\tilde{\Omega}} \mathbf{m}v \, d\tilde{z}d\tilde{x}d\tilde{y}, \quad (5.26)$$

where the test function $v(\tilde{z}, \tilde{x}, \tilde{y})$ is chosen such that it satisfies the homogeneous Dirichlet boundary conditions, which implies that the boundary terms have vanished.

We introduce the notation of the division of two vectors $(x_1, x_2, x_3)^T$ and $(y_1, y_2, y_3)^T$ (where “ \bullet^T ” stands for the transpose operator)

$$\mathbf{B} = \frac{(y_1, y_2, y_3)^T}{(x_1, x_2, x_3)^T}, \quad (5.27)$$

such that \mathbf{B} satisfies the relationship

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \mathbf{B} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}. \quad (5.28)$$

This matrix \mathbf{B} is actually the transformation matrix when the coordinate transformation is performed from the system (x_1, x_2, x_3) to (y_1, y_2, y_3) .

In SEM, the physical domain is decomposed into a set of nonoverlapping hexahedral elements. Each element is further discretized by $(N + 1)^3$ Gauss-Lobatto-Legendre (GLL) points $(\xi_{k_1}, \eta_{k_2}, \zeta_{k_3})$ in the reference space, $k_1, k_2, k_3 = 0, \dots, N$. The mapping from the reference space (ξ, η, ζ) to the Cartesian space (z, x, y) is expressed by the Jacobi matrix, according to the notation in Equation (5.27),

$$\mathbf{J} = \frac{(\partial z, \partial x, \partial y)^T}{(\partial \xi, \partial \eta, \partial \zeta)^T} \quad \text{and} \quad J_e = \det(\mathbf{J}). \quad (5.29)$$

Similarly, the relationship between the reference space and the dimensionless coordinate is defined through

$$\tilde{\mathbf{J}} = \frac{(\partial \tilde{z}, \partial \tilde{x}, \partial \tilde{y})^T}{(\partial \xi, \partial \eta, \partial \zeta)^T} \quad \text{and} \quad \tilde{J}_e = \det(\tilde{\mathbf{J}}). \quad (5.30)$$

In the reference space, the basis functions are defined as Lagrange polynomials over the GLL points. The integrals are then numerically approximated by GLL quadrature (more details on SEM are given in Appendix 5.C). Due to these ingredients and the property of the Lagrange polynomials that have values at the GLL nodes of either 0 or 1, the weak formulation in Equation (5.26) can be written as

$$\underbrace{(\mathbf{M} + \mathbf{K})}_{\mathbf{A}} \mathbf{s} = \mathbf{M} \mathbf{m}, \quad (5.31)$$

in which the mass matrix \mathbf{M} is diagonal, and the stiffness matrix \mathbf{K} is symmetric. The impedance matrix $\mathbf{A} = \mathbf{M} + \mathbf{K}$ is then symmetric. For the mass matrix, its components are given by

$$M_{\hat{k}\hat{h}} = w_{k_1} w_{k_2} w_{k_3} \tilde{J}_e(\xi_{k_1}, \eta_{k_2}, \zeta_{k_3}) \quad \hat{k} \text{ stands for the triple of indexes } \{k_1, k_2, k_3\}, \quad (5.32)$$

and, for the stiffness matrix, by

$$K_{\hat{k}\hat{h}} = \sum_{\hat{q}=0}^N w_{q_1} w_{q_2} w_{q_3} \tilde{J}_e(\xi_{q_1}, \eta_{q_2}, \zeta_{q_3}) \left[\sum_{i=1}^3 \sum_{j=1}^3 \frac{\partial v_{\hat{k}}}{\partial r_i} \left(\sum_{d=1}^3 \frac{\partial r_i}{\partial p_d} \frac{\partial r_j}{\partial p_d} \right) \frac{\partial v_{\hat{h}}}{\partial r_j} \right], \quad (5.33)$$

where

$$(p_1, p_2, p_3) := (\tilde{z}, \tilde{x}, \tilde{y}) \quad \text{and} \quad (r_1, r_2, r_3) := (\xi, \eta, \zeta).$$

We introduce here the geometric factors G_{ij} associated with the projection between the dimensionless coordinates and the reference coordinates

$$G_{ij}(\xi_{q_1}, \eta_{q_2}, \zeta_{q_3}) = \sum_{d=1}^3 \frac{\partial r_i}{\partial p_d} \frac{\partial r_j}{\partial p_d} \tilde{J}_e(\xi_{q_1}, \eta_{q_2}, \zeta_{q_3}), \quad (5.34)$$

which simplifies the expression of the stiffness matrix to

$$K_{\hat{k}\hat{h}} = \sum_{\hat{q}=0}^N w_{q_1} w_{q_2} w_{q_3} \left[\sum_{i=1}^3 \sum_{j=1}^3 \frac{\partial v_{\hat{k}}}{\partial r_i} G_{ij} \frac{\partial v_{\hat{h}}}{\partial r_j} \right]. \quad (5.35)$$

These expressions of eqs (5.32), (5.34), and (5.35) are used for the implementation of the matrices \mathbf{M} and to compute the product of the matrix \mathbf{A} with a given vector in the dimensionless coordinates. To do so, the evaluation of the geometric factors G_{ij} is critical. Note that the linear system of Equation (5.31) is constructed and solved in the dimensionless coordinates and not in the Cartesian coordinates. The results are not affected by this coordinate transformation, due to the bijective correspondence between the two coordinates systems (Equation (5.24)).

When wave-propagation simulation is performed by SEM, all of the elements of the inverse Jacobi matrix \mathbf{J}^{-1} and the volumetric Jacobian J_e associated to this projection are available at no extra cost. It is of great interest to incorporate these ingredients into the construction of matrices \mathbf{M} and geometric factors G_{ij} . According to the definition of the relationship between dimensionless coordinates and Cartesian coordinates (Equation (5.24)), the determinant \tilde{J}_e of the Jacobi matrix $\tilde{\mathbf{J}}$ can be estimated from the volumetric Jacobian J_e of the Jacobi matrix \mathbf{J} through

$$\tilde{J}_e(\xi, \eta, \zeta) = \left| \frac{(\partial \tilde{z}, \partial \tilde{x}, \partial \tilde{y})^T}{(\partial \xi, \partial \eta, \partial \zeta)^T} \right| = \frac{1}{L_z L_x L_y} \left| \frac{(\partial z, \partial x, \partial y)^T}{(\partial \xi, \partial \eta, \partial \zeta)^T} \right| = \frac{1}{L_z L_x L_y} J_e(\xi, \eta, \zeta). \quad (5.36)$$

The inverse matrix $\tilde{\mathbf{J}}^{-1}$ can be deduced from the elements of the inverse Jacobi matrix \mathbf{J}^{-1} through the identity

$$\begin{bmatrix} \partial_z \xi & \partial_x \xi & \partial_y \xi \\ \partial_z \eta & \partial_x \eta & \partial_y \eta \\ \partial_z \zeta & \partial_x \zeta & \partial_y \zeta \end{bmatrix} = \begin{bmatrix} L_z \partial_z \xi & L_x \partial_x \xi & L_y \partial_y \xi \\ L_z \partial_z \eta & L_x \partial_x \eta & L_y \partial_y \eta \\ L_z \partial_z \zeta & L_x \partial_x \zeta & L_y \partial_y \zeta \end{bmatrix}. \quad (5.37)$$

Following this framework, the application of the Bessel filter to a model vector, through the solving of the linear system (Equation (5.31)), can be efficiently achieved in the dimensionless coordinate system, whereas the wave-propagation simulation is still performed in Cartesian coordinates.

The symmetric matrix \mathbf{A} is referred to as the discrete Helmholtz operator, which has been shown to be positive definite (Deville et al., 2002). The CG method is thus the method of choice to iteratively solve the linear system. In these numerical experiments, the condition number of the matrix \mathbf{A} ranges from 10^2 to 10^4 , which depends on the value of the coherent lengths. Therefore, the iterative solver converges rapidly toward the desired solution, as we will see with the numerical implementation.

5.3.2 Variable coherent lengths, dip and azimuth angles

In the previous section, the weak formulation of the Bessel filter in a dimensionless coordinate system was shown to naturally yield a symmetric stiffness matrix. When introducing variable parameters—coherent lengths, dip and azimuth angle—we want to preserve the symmetry of the matrix \mathbf{A} , and

therefore we develop the weak formulation in a rotated dimensionless coordinate system $(\tilde{v}, \tilde{u}, \tilde{w})$. Before defining these coordinates, let us provide the definition of the azimuth θ and dip φ angles for a given vector. The azimuth is the horizontal angle measured from the North, and the dip is the angle that the vector has with the horizontal (Sheriff, 2002). Their values range such that $\theta \in [-\pi, \pi]$ and $\varphi \in [-\pi/2, \pi/2]$. In 3-D space, a rotation R_{3D} with azimuth θ and dip φ transforms the Cartesian coordinates (z, x, y) into the rotated coordinates (v, u, w) , where v is vertical direction, perpendicular to the bedding planes, and the two horizontal directions u and w define the plane of the geological structure. Considering the previous definitions, we have the following rotational operator

$$R_{3D}(\theta, \varphi) = \frac{(v, u, w)^T}{(z, x, y)^T} = \begin{bmatrix} \cos \varphi & -\cos \theta \sin \varphi & \sin \theta \sin \varphi \\ \sin \varphi & \cos \theta \cos \varphi & -\sin \theta \cos \varphi \\ 0 & \sin \theta & \cos \theta \end{bmatrix}. \quad (5.38)$$

Similar to the last section, we would like to develop the weak formulation of the PDEs associated with the application of the sparse inverse Bessel filter in the rotated dimensionless coordinates system $(\tilde{v}, \tilde{u}, \tilde{w})$, defined by the following relationships

$$\tilde{v} = \frac{v}{L_v(v, u, w)}; \quad \tilde{u} = \frac{u}{L_u(v, u, w)}; \quad \tilde{w} = \frac{w}{L_w(v, u, w)}, \quad (5.39)$$

in which L_v, L_u and L_w are coherent lengths in the v, u and w directions, respectively. In this rotated dimensionless coordinate system, the Bessel filter is defined by the PDE

$$B(\tilde{v}, \tilde{u}, \tilde{w}) - \left(\frac{\partial^2}{\partial \tilde{v}^2} + \frac{\partial^2}{\partial \tilde{u}^2} + \frac{\partial^2}{\partial \tilde{w}^2} \right) B(\tilde{v}, \tilde{u}, \tilde{w}) = \delta(\tilde{v}, \tilde{u}, \tilde{w}). \quad (5.40)$$

The rotated dimensionless coordinates are mapped to the Cartesian coordinates through the relationship

$$\frac{(\tilde{v}, \tilde{u}, \tilde{w})^T}{(z, x, y)^T} = \frac{(\tilde{v}, \tilde{u}, \tilde{w})^T}{(v, u, w)^T} \times \frac{(v, u, w)^T}{(z, x, y)^T} = \begin{bmatrix} \frac{1}{L_v} & 0 & 0 \\ 0 & \frac{1}{L_u} & 0 \\ 0 & 0 & \frac{1}{L_w} \end{bmatrix} \times R_{3D}(\theta, \varphi). \quad (5.41)$$

As coherent lengths, dip and azimuth (i.e., the parameters of the filter) are nonstationary, Equation (5.40) does not provide the same structure as Equation (5.8) when expressed in the Cartesian coordinate system. The chain rules for derivative estimation introduce spatial derivative terms that are related to the parameter variations in the PDE, when moving from one to the other set of coordinates, through the relationship of Equation (5.41). However, if the filter parameters vary slowly in space, their spatial derivatives can be ignored, which leads to the following approximation

$$\frac{(\partial \tilde{v}, \partial \tilde{u}, \partial \tilde{w})^T}{(\partial z, \partial x, \partial y)^T} \approx \frac{(\tilde{v}, \tilde{u}, \tilde{w})^T}{(z, x, y)^T} = \begin{bmatrix} \frac{1}{L_v} & 0 & 0 \\ 0 & \frac{1}{L_u} & 0 \\ 0 & 0 & \frac{1}{L_w} \end{bmatrix} \times R_{3D}(\theta, \varphi). \quad (5.42)$$

Within this approximation, the Bessel filters defined in the rotated dimensionless coordinates and in the Cartesian coordinates are almost identical. In other words, performing the smoothing operation in the dimensionless system or the Cartesian system should provide approximately the same results. This argument has an important role in this implementation, because it allows the weak formulation to be developed in dimensionless coordinates, to maintain the symmetry of matrices \mathbf{K} and \mathbf{A} ; this is a key point for numerical efficiency. It should be noted that this approximation will systematically introduce an error into the amplitude of the filtering operator when rapid variations of the filter parameters occur. The associated error analysis will be discussed in section 5.4.1.

In the rotated dimensionless coordinate system, the differential relationship between the original vector \mathbf{m} and the smoothed vector \mathbf{s} that was obtained by filtering with the Bessel filter $B_{3D}(\tilde{v}, \tilde{u}, \tilde{w})$ can be written as

$$\mathbf{s}(\tilde{v}, \tilde{u}, \tilde{w}) - \nabla_{\tilde{v}, \tilde{u}, \tilde{w}}^2 \mathbf{s}(\tilde{v}, \tilde{u}, \tilde{w}) = \mathbf{m}(\tilde{v}, \tilde{u}, \tilde{w}). \quad (5.43)$$

Ignoring the derivatives related to variations in the filter parameters allows the same workflow to be applied as for homogeneous parameters. Let us consider the mapping from reference space (ξ, η, ζ) to the rotated dimensionless coordinates $(\tilde{v}, \tilde{u}, \tilde{w})$, defined by the Jacobi matrix $\tilde{\mathbf{J}}^{\text{rot}}$

$$\tilde{\mathbf{J}}^{\text{rot}} = \frac{(\partial \tilde{v}, \partial \tilde{u}, \partial \tilde{w})^T}{(\partial \xi, \partial \eta, \partial \zeta)^T}, \quad (5.44)$$

and the volumetric Jacobian \tilde{J}_e^{rot}

$$\tilde{J}_e^{\text{rot}} = \det(\tilde{\mathbf{J}}^{\text{rot}}). \quad (5.45)$$

The weak form of Equation (5.43) can again be discretized into the linear system of Equation (5.31), where the mass matrix \mathbf{M} is diagonal and the stiffness matrix \mathbf{K} remains symmetric through relations

$$M_{\hat{k}\hat{k}} = w_{k_1} w_{k_2} w_{k_3} \tilde{J}_e^{\text{rot}}(\xi_{k_1}, \eta_{k_2}, \zeta_{k_3}) \quad \text{where } \hat{k} \text{ stands for the triple of indexes } \{k_1, k_2, k_3\}, \quad (5.46)$$

and

$$K_{\hat{k}\hat{h}} = \sum_{\hat{q}=0}^N w_{q_1} w_{q_2} w_{q_3} \left[\sum_{i=1}^3 \sum_{j=1}^3 \frac{\partial v_{\hat{k}}}{\partial r_i} G_{ij} \frac{\partial v_{\hat{h}}}{\partial r_j} \right], \quad (5.47)$$

where

$$(p_1, p_2, p_3) := (\tilde{v}, \tilde{u}, \tilde{w}) \quad \text{and} \quad (r_1, r_2, r_3) := (\xi, \eta, \zeta).$$

The geometric factors G_{ij} associated with the projection between the rotated dimensionless coordinates and the reference coordinates becomes

$$G_{ij}(\xi_{q_1}, \eta_{q_2}, \zeta_{q_3}) = \sum_{d=1}^3 \frac{\partial r_i}{\partial p_d} \frac{\partial r_j}{\partial p_d} \tilde{J}_e^{\text{rot}}(\xi_{q_1}, \eta_{q_2}, \zeta_{q_3}). \quad (5.48)$$

The numerical implementation is identical to the case of the homogeneous parameters in eqs (5.32), (5.34), and (5.35), where the linear system of Equation (5.31) is again constructed and solved in the reduced coordinates system. However, we need new expressions of the volumetric Jacobian \tilde{J}_e^{rot} , and the elements of the inverse Jacobi matrix $(\tilde{\mathbf{J}}^{\text{rot}})^{-1}$. These quantities can be computed from elements in the mapping between the reference space and the Cartesian space, knowing that

$$\tilde{\mathbf{J}}^{\text{rot}} = \frac{(\partial \tilde{v}, \partial \tilde{u}, \partial \tilde{w})^T}{(\partial \xi, \partial \eta, \partial \zeta)^T} = \frac{(\partial \tilde{v}, \partial \tilde{u}, \partial \tilde{w})^T}{(\partial z, \partial x, \partial y)^T} \times \frac{(\partial z, \partial x, \partial y)^T}{(\partial \xi, \partial \eta, \partial \zeta)^T}. \quad (5.49)$$

Combining Equation (5.49) with the approximation of Equation (5.42), we have

$$\tilde{\mathbf{J}}^{\text{rot}} \approx \begin{bmatrix} \frac{1}{L_v} & 0 & 0 \\ 0 & \frac{1}{L_u} & 0 \\ 0 & 0 & \frac{1}{L_w} \end{bmatrix} \times R_{3D}(\theta, \varphi) \times \mathbf{J}. \quad (5.50)$$

As $\det[R_{3D}(\theta, \varphi)] = 1$, the determinant of the Jacobi matrix $\tilde{\mathbf{J}}^{\text{rot}}$ is

$$\tilde{J}_e^{\text{rot}} = \frac{\det(\mathbf{J})}{L_v L_u L_w} = \frac{J_e}{L_v L_u L_w}. \quad (5.51)$$

The inverse of the Jacobi matrix \mathbf{J}^{rot} is

$$(\mathbf{J}^{\text{rot}})^{-1} = \mathbf{J}^{-1} \times R_{3D}(\theta, \varphi)^{-1} \times \begin{bmatrix} L_v & 0 & 0 \\ 0 & L_u & 0 \\ 0 & 0 & L_w \end{bmatrix}, \quad (5.52)$$

or

$$(\mathbf{J}^{\text{rot}})^{-1} = \mathbf{J}^{-1} \times \begin{bmatrix} L_v \cos \varphi & L_u \sin \varphi & 0 \\ -L_v \cos \theta \sin \varphi & L_u \cos \theta \cos \varphi & L_w \sin \theta \\ L_v \sin \theta \sin \varphi & -L_u \sin \theta \cos \varphi & L_w \cos \theta \end{bmatrix}. \quad (5.53)$$

The volumetric Jacobian \tilde{J}_e^{rot} and elements in the inverse Jacobian matrix $\tilde{\mathbf{J}}^{\text{rot}}$ can be computed from eqs (5.51) and (5.53), which completes the construction of matrix \mathbf{M} and geometric factors G_{ij} in the dimensionless coordinates. If the azimuth and dip are zero, and if the coherent lengths are homogeneous, eqs (5.51) and (5.53) are identical to eqs (5.36) and (5.37). Finally, we end up with the linear system

$$\mathbf{A} \mathbf{s} = \mathbf{M} \mathbf{m}. \quad (5.54)$$

5.4 Numerical illustrations

Various 3-D examples that illustrate the numerical efficiency of the workflow are proposed in this section. We will show that the decay of Laplace filters can be mimicked by using Bessel operators. However, it should be emphasized that there is no obligation to use a Laplace filter; for example, the software SPECFEM3D uses a 3-D Gaussian smoothing filter (Peter et al., 2011). We will also quantify the numerical approximation for variable coherent length, dip and azimuth. We first consider an estimation of the Bessel kernel by considering an original vector \mathbf{m} as a delta function at the origin. Then, we will consider the impact of the filter when the input vector only contains random noise, without any predefined structures. Finally, we show the application of the filter to a gradient obtained from a subset of the synthetic 3-D SEAM Phase II foothills model (Oristaglio, 2012).

5.4.1 Spike test

When the source term, that is, the original vector \mathbf{m}_0 , is set as a delta function at location $(0, 0, 0)$, the linear system

$$\mathbf{A} \mathbf{s} = \mathbf{M} \mathbf{m}_0, \quad (5.55)$$

should provide a solution that is identical to the Bessel filter (Equation (5.6)). When we solve the linear system of Equation (5.55) twice through the following sequential system

$$\begin{aligned} \mathbf{A} \mathbf{s}^* &= \mathbf{M} \mathbf{m}_0 \\ \mathbf{A} \mathbf{s} &= \mathbf{M} \mathbf{s}^*, \end{aligned} \quad (5.56)$$

we should obtain a response similar to the Laplace filter. Let us consider a model of size $1.5 \text{ km} \times 1.5 \text{ km} \times 1.5 \text{ km}$. A delta function is located at the centre of the model, as shown in Figure 5.3.

In this example, the coherent lengths in the x and y directions are homogeneous $L_x = L_y = 80 \text{ m}$ and the dip and azimuth angle are set to zero, as there is no specific orientation. The coherent length L_z is either constant or variable in the z direction, as indicated in Figure 5.3A(a), B(a), and C(a). In

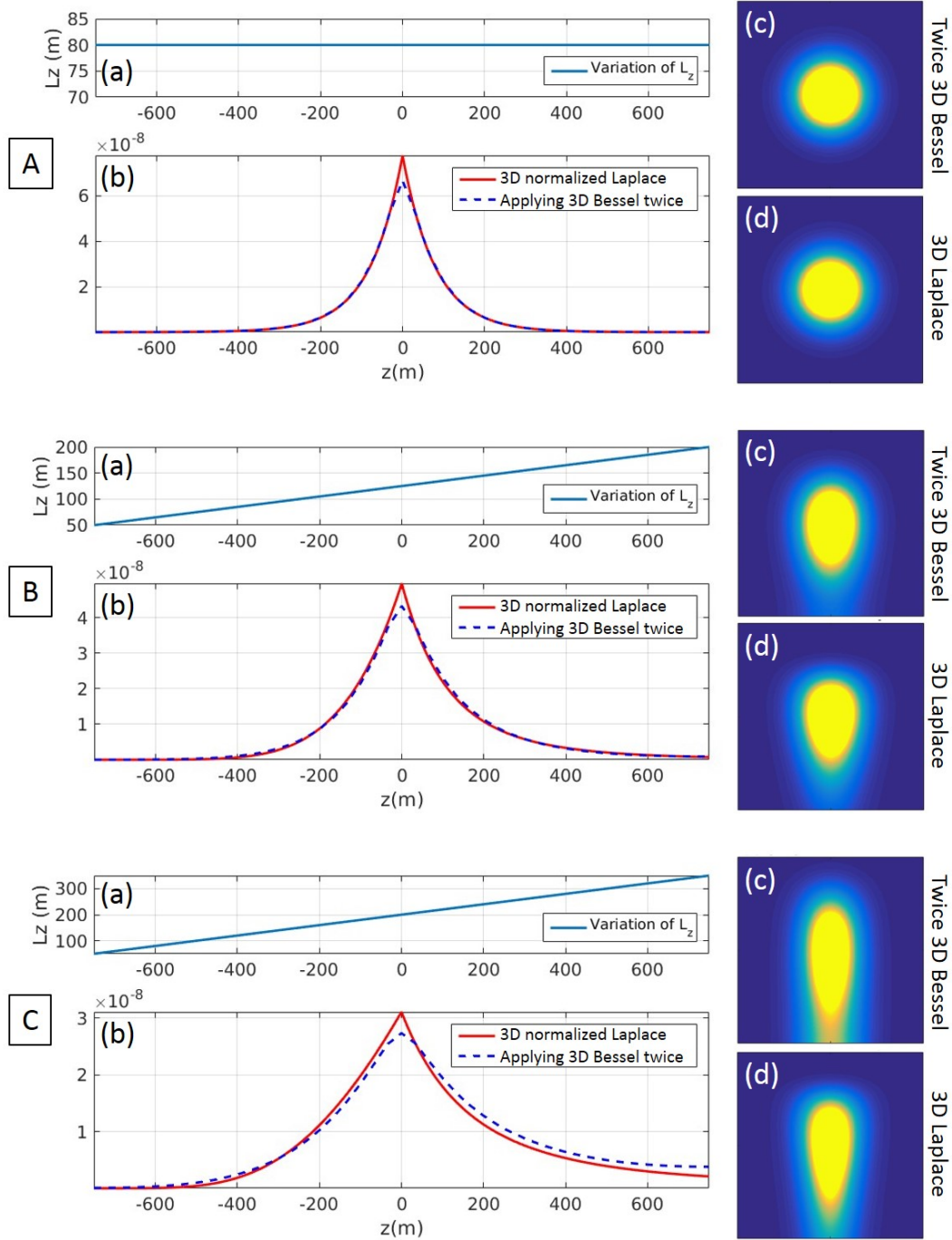


Figure 5.3: Comparison of the normalized Laplace filter with the application of two Bessel filters through the spike tests, with homogeneous coherent lengths in the x and y directions $L_x = L_y = 80$ m. Coherent length in the z direction: (A) Constant L_z ; (B) L_z varies from 50 to 200 m; (C) L_z varies from 50 to 350 m. In each panel, (a) shows the variation of L_z in the z direction, and (b) compares the output of applying the 3-D Bessel filter twice (blue dashed line), with the theoretical Laplace filter (red line). The respective zx cross-section of these filters are shown in (c) and (d).

Figure 5.3A(b), B(b), and C(b), the output of the linear system (blue dashed line) and the theoretical Laplace function (red line) in the z directions are superimposed. Figure 5.3A(b) illustrates that the double application of the Bessel filter provides an excellent approximation of the Laplace filter for homogeneous coherent lengths, although still with the negligible singularity at the origin. Under the slow variation of the coherent lengths, $\partial_z L_z = 0.1$ in Figure 5.3B(b), this method can correctly follow the shape of the Laplace filter, which implies that ignoring the spatial derivatives of the coherent lengths is acceptable. This conclusion is further supported by the identical shapes of the zx cross-sections of the 3-D Laplace filter and the output of the linear system of Equation (5.55) in Figure 5.3B(c) and B(d). When this variation becomes more important, as $\partial_z L_z = 0.2$, this approximation cannot exactly reproduce the amplitude of the theoretical filter, but the influence of this approximation still appears acceptable for the smoothing effect of the model vector \mathbf{m} that we consider.

The implementation of constant dip and azimuth is illustrated by the spike test in Figure 5.4. The delta function is again located at the centre of the model. Homogeneous different coherent lengths are used in all directions, but the filters are highly anisotropic, with $L_v = 50$ m, $L_u = 100$ m, and $L_w = 200$ m. Due to this design, the 3-D filter is a tri-axial ellipsoid with distinct semi-axis length, as indicated in the zx , xy , and zy cross-sections in Figure 5.4A. After applying the rotation with azimuth $\theta = 60^\circ$ and dip $\varphi = 30^\circ$, the comparison between Figure 5.4B and Figure 5.4A shows that the ellipsoid is tilted 30° from the x axis in the zx section, and it is also rotated 60° from the y axis in the xy cross-section.

When variable dip and azimuth are introduced into the filter, similar amplitude errors as for the variable-coherent-lengths study in Figure 5.3C are expected. Furthermore, the amplitude error introduced by the 3-D rotation can be mitigated by careful design of the coherent lengths, as the rotation has no impact in an isotropic filter. For FWI applications, which are generally limited at low frequency, the approximation of the slow variation of the filter parameters still appears acceptable for the smoothing effect.

5.4.2 Random noise tests

The filtering operator normally assumes the particular properties of the structures, which implies that the shape of the smoothed gradient/ model is driven by the imposed variation of coherent lengths, dip, and azimuth. This argument is illustrated in the following examples, when the input vector \mathbf{m} only contains random noise, with no pre-defined structures. Figure 5.5 focuses on homogeneous coherent lengths, dip, and azimuth filters on a model of size $1.5 \text{ km} \times 1.5 \text{ km} \times 1.5 \text{ km}$. The zx , xy , and zy cross-sections of the input vector are shown in Figure 5.5A, which only contains high frequency variations of random noise. Figure 5.5B shows the smoothed vector obtained with coherent lengths $L_z = L_v = 25$ m, $L_x = L_u = 100$ m, and $L_y = L_w = 25$ m, without dip and azimuth. As the filter is strongly anisotropic, with the longest semi-axis length in the x direction, the zx and xy sections both contain features aligned in the x direction. The patterns in the zy section have no specific orientation as $L_z = L_y$, that is, the 3-D filter is isotropic in the zy plane. In Figure 5.5C, after applying a rotation with 45° dip, the aligned features are tilted 45° from the x axis in the zx section. Note that the apparent coherent length L_x is now almost identical to L_y , with no specific trend in the xy cross-section. Meanwhile, the apparent L_z becomes greater than L_y , which induces the alignment along the z direction in the zy cross-section. A similar interpretation can be applied to Figure 5.5D, which shows the application of a filter with azimuth 45° to smooth the initial vector. Compared to Figure 5.5B, the features in the xy section are rotated 45° from the y axis, and the patterns presented in the zx cross-section are lost due to the apparent coherent length in the x direction.

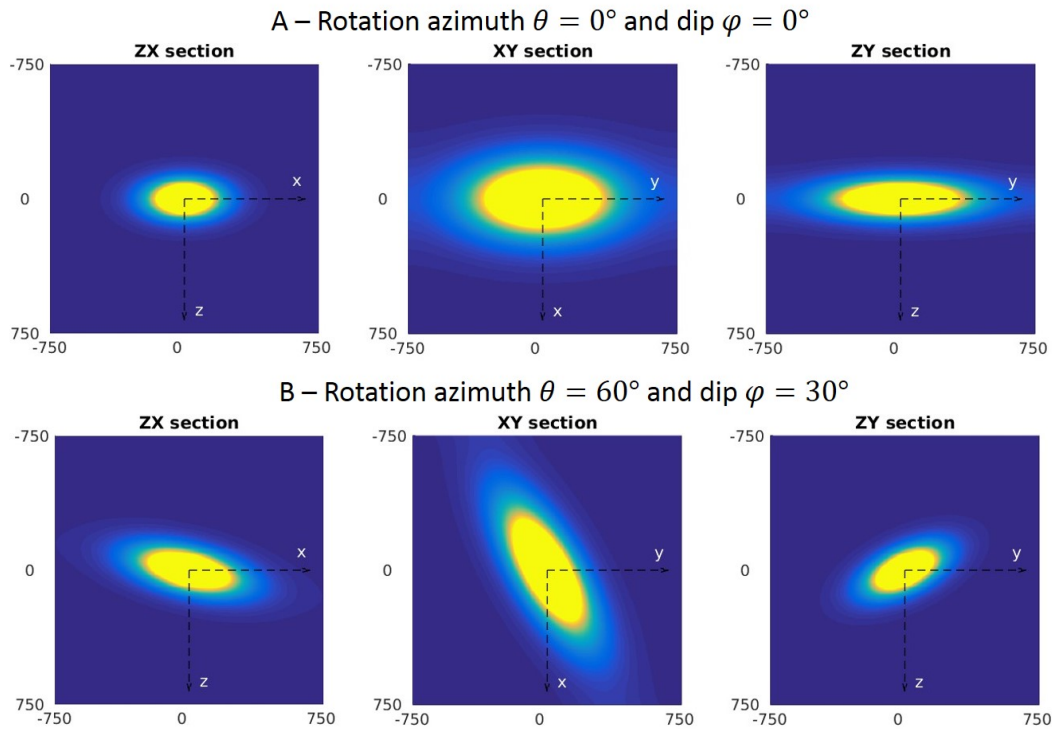


Figure 5.4: Spike test to illustrate the kernel of the 3-D Bessel filter with homogeneous coherent lengths $L_v = L_z = 50$ m, $L_x = L_u = 100$ m, and $L_y = L_w = 200$ m. (A) zx , xy and zy cross-sections of the 3-D Bessel filter without dip and spike. (B) The filter consists of a 3-D rotation with 60° azimuth and 30° dip.

The example shown in Figure 5.6 is a longer gradient/ model with 3 km length in the x direction that is used to illustrate the implementation of variable coherent lengths, dip and azimuth. We use simple sine variation for these geological properties. Again, the input vector (Figure 5.6A) contains random noise without any predefined structure. In Figure 5.6B, the coherent lengths in the z and y directions are homogeneous $L_z = L_y = 25$ m, whereas L_x varies as a sine function in the x direction, from 25 m to 85 m. The zx section of the smoothed vector (Figure 5.6B, left) correctly follows the variation of the coherent length L_x (Figure 5.6B, right). Figure 5.6C shows the zx and xy cross-sections of the smoothed vector with homogeneous coherent lengths $L_z = L_v = 25$ m, $L_x = L_u = 250$ m, and $L_y = L_w = 25$ m, and no dip and azimuth. An extremely long L_u was intentionally designed so that the smoothed vector has a layered structure, in parallel with the x direction. Figure 5.6C will be used as the reference to compare this with the variable dip and azimuth filters in the next examples. In Figure 5.6D, similar values of homogeneous coherent lengths were use as those for Figure 5.6C. The dip (Figure 5.6D, right) varies as a sine function in the x direction from -45° to 45° , which leads to folding structures in the zx section (Figure 5.6D, left). Similar variation in the xy cross-section in Figure 5.6E is obtained when the azimuth varies as a sine function in the x direction, from -60° to 60° .

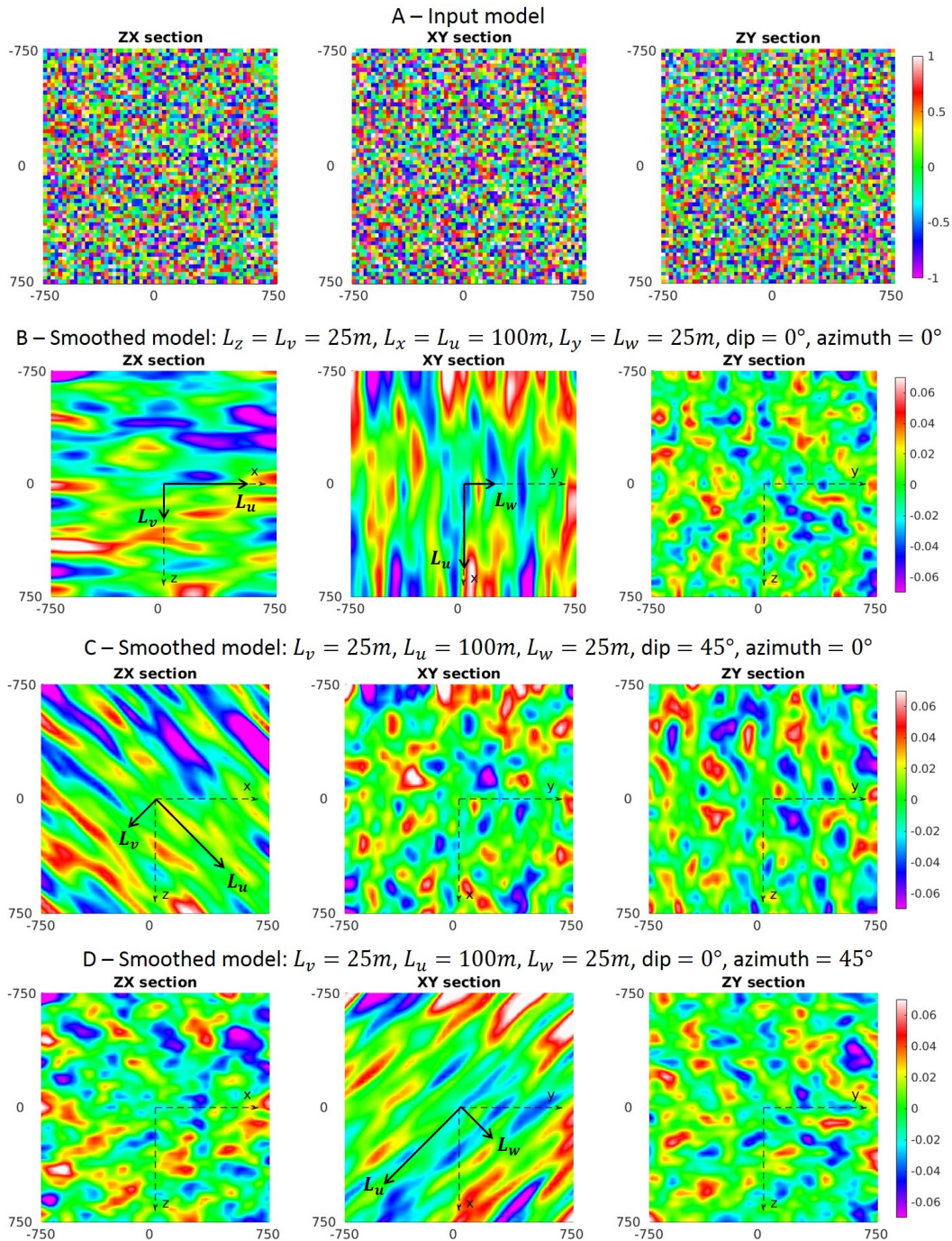


Figure 5.5: Random noise test with a stationary filter to illustrate the implementation of homogeneous coherent lengths, dip and azimuth. (A) zx , xy and zy cross-sections of the input model. (B) Smoothed model obtained from applying an anisotropic filter, $L_v = 25\text{ m}$, $L_u = 100\text{ m}$, and $L_w = 25\text{ m}$, without dip and azimuth. (C) Smoothed model when 45° dip and 0° azimuth are introduced into the filter. (D) Smoothed model when 0° dip and 45° azimuth are introduced into the filter. Model size, $1.5\text{ km} \times 1.5\text{ km}$.

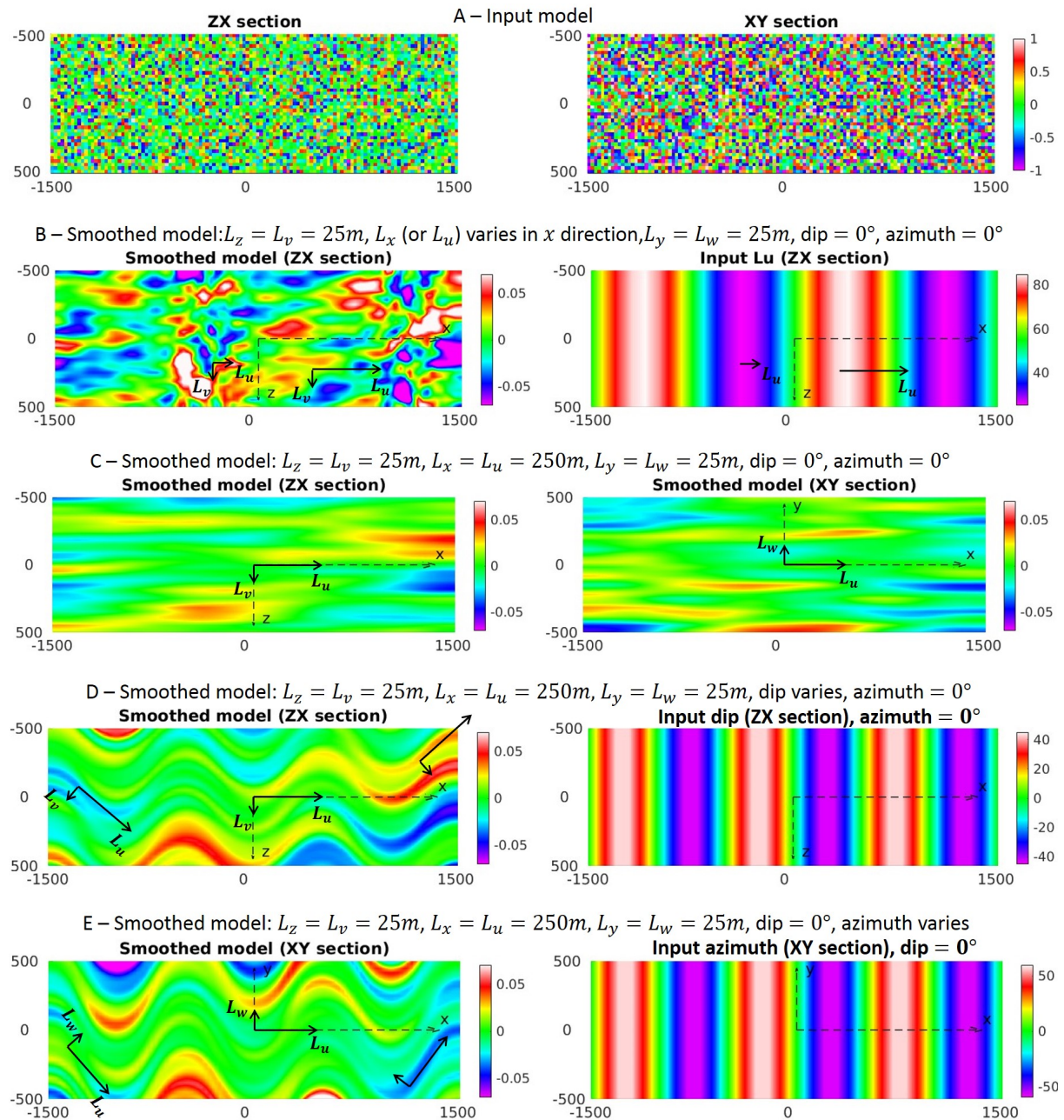
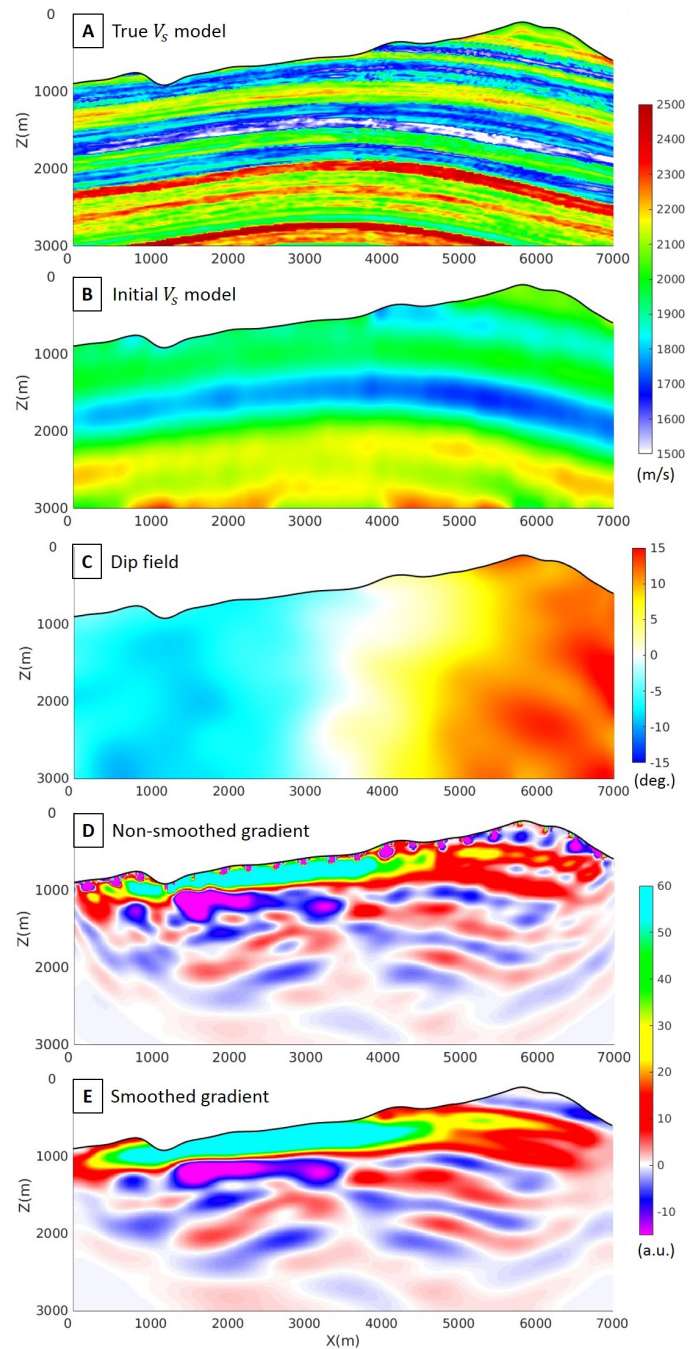


Figure 5.6: Random noise test with nonstationary filter to illustrate the implementation of variable coherent lengths, dip and azimuth. The size of the model is 1 km in the z and y directions, and 3 km in the x direction. (A) zx and xy cross-section of the input model. (B) Smoothed model (left) obtained from a filter with variable coherent length L_x (right). (C) zx and xy cross-section of the smoothed model obtained from a highly anisotropic stationary filter. (D) Smoothed model (left) obtained from a filter with variable dip (right). (E) Smoothed model (left) obtained from a filter with variable azimuth (right).

Figure 5.7: Example of the filtering operator on the FWI gradient, from a subset of the 3-D SEAM Phase II foothills model.

- (A) True shear-wave velocity model.
 (B) Initial shear-wave velocity model.
 (C) Dip field.
 (D) Original scaled gradient without any smoothing.
 (E) Smoothed gradient, with an anisotropic nonstationary Laplace filter (approximated by the application of two Bessel filters) for the dip field as presented in (C): $L_z = 0.05\lambda_s$ and $L_x = L_y = 0.15\lambda_s$, where λ_s is the local shear wavelength.



5.4.3 FWI gradient smoothing

This section illustrates the application of a nonstationary Laplace filter to a realistic gradient vector from FWI, as obtained from a subset of the 3-D SEAM Phase II foothills model (Oristaglio, 2012). Surface acquisition is used with a line of 20 sources, with 350 m between adjacent sources. The receivers are located in the whole surface, with 12.5 m between receivers. A Ricker wavelet centred at 3 Hz is used as the source signal. The 2-D cross-section of the shear-wave velocity (V_s) model underneath the source line is shown in Figure 5.7A.

The topography variation is significant in this model, with maximal vertical elevation of 900 m. SEM is used for both forward and inversion problems. The initial V_s model is shown in Figure 5.7B, which is a smoothed version of the true model. Compared with the true model in Figure 5.7A, this initial V_s is overestimated at the near surface and underestimated at greater depths. The dip field is extracted from the true velocity model by manual picking, as illustrated in Figure 5.7C. Figure 5.7D shows the first scaled gradient without any smoothing filter, which contains a significant acquisition footprint at the near surface. The horizontal oscillation of the features at greater depths might come from high wavenumber artifacts.

An anisotropic nonstationary Laplace filter with coherent lengths $L_z = 0.05\lambda_s$ and $L_x = L_y = 0.15\lambda_s$ is applied, where λ_s is the shear wavelength at each spatial position. The true dip field, as shown in Figure 5.7C, and the zero azimuth angle are used for 3-D rotation. It should be noted again here that the Laplace filter is efficiently approximated by application of two Bessel filters. The filters help to remove the near-surface artifacts due to the acquisition footprint, without degrading the deeper structures. The continuity of the features at greater depths is actually enhanced because the horizontal-oscillation artifacts are attenuated. In addition, the gradient correctly determines the update direction of the model: it reduces the V_s at the near surface and increases it at greater depths (knowing that FWI updates the model following the negative gradient).

In summary, we have illustrated the property of the Bessel filter through various numerical examples in spike and random-noise tests. The approximation of the Laplace filter by Bessel operators is shown, which indicates prospective applications of this filter, either as a smoothing filter, or to efficiently mimic the decay of the Laplace filter. We also illustrate the robustness of this method for variable coherent lengths, dip and azimuth. The numerical example on a realistic gradient highlights the potential application of the filter for FWI.

5.5 Numerical implementation in a parallel spectral-element method scheme

In FWI, the gradient vector of the misfit function is computed from the correlation of the forward and backward propagation wavefields (as one per source) at each iteration, due to the adjoint-state method (Plessix, 2006). However, the modelling mesh can sometimes be quite dense compared to the resolution that can be expected from the inversion, which leads to high wave-number noise in many applications. Consequently, the gradient/ model vector must be smoothed or regularized on this forward-modelling mesh. By doing so, the filtering operation (i.e., the Bessel operator here) has to be directly and efficiently implemented on the modelling mesh, which can be described by a domain decomposition for parallel computation. As the application of the Bessel filter is related to a PDE, this can proceed in a similar way as for the wave-propagation simulation.

Following an analogue of the framework to that designed for the SEM for wave simulation, the linear system associated with the Bessel filters is constructed and solved by a parallel CG iterative solver

$$\underbrace{(\mathbf{M} + \mathbf{K})}_{\mathbf{A}} \mathbf{s} = \mathbf{M} \mathbf{m}. \quad (5.57)$$

According to Saad (2003), only the matrix-vector product and the inner product of two vectors are required in the CG method. Thus, the most expensive operator is the product of the matrix \mathbf{A} with a given vector. Each subdomain computes its part of the product in parallel, and the communications between subdomains during the CG iterations share the same strategies as that of wave simulation, which requires no extra effort to manage the parallelism. This section presents this implementation and

highlights the efficiency of this approach, compared with the standard 3-D convolution method. The convergence rate of the CG solver for several parameters are then analysed.

5.5.1 Linear system construction and matrix-vector product evaluation

The mass matrix \mathbf{M} is diagonal, and the number of nonzero elements (NNZs) of this matrix is identical to the size of the input gradient/ model vector,

$$\text{NNZ}(\mathbf{M}) = \text{SIZE}(\mathbf{m}). \quad (5.58)$$

It is then stored in the same form as the input vector. The dimensions of the matrices \mathbf{K} and \mathbf{A} are $[\text{SIZE}(\mathbf{m})]^2$; thus the full storage is not reasonable for realistic application. Assuming that the same order of interpolation N is used in each direction, and the inverse of Jacobi matrix $\tilde{\mathbf{J}}^{\text{rot}}$ is full of nonzero elements, the NNZs in the matrix \mathbf{A} can be estimated from the size of the vector \mathbf{m} , as

$$\frac{\text{NNZ}(\mathbf{A})}{\text{SIZE}(\mathbf{m})} \rightarrow (3(N+1)^2 + 3(N+1) + 1). \quad (5.59)$$

For example, when 4th order interpolation is used (i.e., each hexahedral element is discretized by $5 \times 5 \times 5$ GLL points)

$$\text{NNZ}(\mathbf{A}) \approx 91 \times \text{SIZE}(\mathbf{m}). \quad (5.60)$$

This estimation illustrates that the stiffness matrix is extremely sparse. Therefore, the nonzero elements of this matrix can either be stored by some efficient compressed storage technique or the matrix-vector product can be directly evaluated without storing the matrix. These two implementation approaches are discussed in this section.

5.5.1.1 Coordinate list storage of nonzero elements of matrix \mathbf{A}

As the impedance matrix \mathbf{A} is sparse, only the nonzero elements of this matrix need to be computed and stored. At each CG iteration, these nonzero elements are used to evaluate the product of matrix \mathbf{A} with a given vector.

The matrix \mathbf{A} is computed from the stiffness matrix \mathbf{K} . According to Equation (5.47), the element at row \hat{k}^{th} , column \hat{h}^{th} of matrix \mathbf{K} can be computed through a triple loop over all of the degrees of freedom (q_1, q_2, q_3) inside the current element, which is computationally expensive. This component can be further developed as a sum of six terms, as shown in Equation (5.121) in Appendix 5.C. This development allows the computation of the left-hand-side matrix \mathbf{A} , while minimizing the loops over all of the degrees of freedom inside the current element.

The matrix \mathbf{A} can thus be stored using the compressed storage techniques coordinate list (COO) format (Golub, 1996), where a list of row and column indices and the associated values of the nonzero elements are stored. In addition, as the matrix \mathbf{A} is symmetric, only the upper triangular part has to be stored; that is, the elements A_{ji} where $i \geq j$. We only store half of the diagonal terms $\frac{1}{2}A_{jj}$. This helps to simplify the matrix-vector product in the CG algorithm. By applying this triangular storage strategy, and based on several tests carried out with this implementation on realistic size problems, the required memory for storage of matrix \mathbf{A} is reduced by 49%, and the computation time of the CG by 33%, .

5.5.1.2 Matrix-free implementation of the matrix-vector product $\mathbf{A}\mathbf{u}$

Instead of building explicitly the impedance matrix \mathbf{A} by storing its nonzero elements, its product with a given vector \mathbf{u} can be directly estimated inside each CG iteration in a matrix-free fashion. As the product of a vector with the diagonal mass matrix \mathbf{M} is trivial, the challenge is the stiffness matrix-vector product $\mathbf{K}\mathbf{u}$. According to the development of the weak formulation of Equation (5.26) in Appendix 5.C and the definition of the geometric factors G_{ij} in Equation (5.48), this matrix-vector product can be written as

$$\sum_{\hat{h}} K_{\hat{k}\hat{h}} u_{\hat{h}} = \sum_{\hat{q}=0}^N w_{q_1} w_{q_2} w_{q_3} \sum_{\hat{h}=0}^N \left[\sum_{i=1}^3 \sum_{j=1}^3 \frac{\partial v_{\hat{k}}}{\partial r_i} G_{ij} \frac{\partial v_{\hat{h}}}{\partial r_j} \right] u_{\hat{h}}, \quad (5.61)$$

or under the factorized version, as

$$\sum_{\hat{h}} K_{\hat{k}\hat{h}} u_{\hat{h}} = \underbrace{\sum_{i=1}^3 \left\{ \sum_{\hat{q}=0}^N w_{q_1} w_{q_2} w_{q_3} \frac{\partial v_{\hat{k}}}{\partial r_i} \left[\sum_{j=1}^3 G_{ij} \underbrace{\left(\sum_{\hat{h}=0}^N \frac{\partial v_{\hat{h}}}{\partial r_j} u_{\hat{h}} \right)}_{\mathbf{D}\mathbf{u}} \right] \right\}}_{\mathbf{D}_w \mathbf{G} \mathbf{D}\mathbf{u}}. \quad (5.62)$$

Equation (5.62) implies that the product stiffness matrix-vector can be factorized as

$$\mathbf{K}\mathbf{u} = \mathbf{D}_w \mathbf{G} \mathbf{D}\mathbf{u}, \quad (5.63)$$

where the operator \mathbf{D} evaluates the spatial derivatives of the vector \mathbf{u} in the reference space, \mathbf{G} is the geometric matrix, and \mathbf{D}_w is equivalent to a weighted spatial derivatives operator.

The derivatives operators $\mathbf{D}\mathbf{u}$ and the weighted spatial derivatives $\mathbf{D}_w(\mathbf{G}\mathbf{D}\mathbf{u})$ can be efficiently estimated using the highly efficient algorithms developed by Deville et al. (2002), which benefit from the tensorial properties of SEM and the optimization of cache usage by manual loop unrolling. For each degree of freedom, the geometric matrix \mathbf{G} is symmetric and only depends on the projection between the rotated dimensionless coordinates and the reference coordinates (Equation (5.48)). Therefore, only six vectors, as $G_{11}, G_{12}, G_{13}, G_{22}, G_{23}, G_{33}$, are stored in the memory, and they are computed outside the CG loops. Due to the factorization in Equation (5.62) and the vectorized computation for the stiffness/ vector product, each CG iteration costs about 0.46 times a time-step of wave propagation for deformed elements. Compared with the explicit matrix-vector product with the COO format of the previous section, this matrix-free implementation decreases the computation cost by a factor of around 7, and reduces the memory requirement by >20 -fold. These numbers have been estimated on the same high-performance computing architecture for the two implementations, on a model with 33×10^6 degrees of freedom, with decomposition on 64 subdomains (64 computing cores with infiniband interconnections).

5.5.2 Parallel conjugate gradient iterative solver

The parallel CG solver introduced in this section is similar to the standard algorithm, except that the matrix-vector and vector-vector operators are computed in parallel. As mentioned before, the matrices \mathbf{M} and \mathbf{A} are evaluated in parallel without assembly over the boundaries between subdomains. After

each matrix-vector product, some point-to-point communications are required to assemble the values on the shared degrees of freedom.

The modulus of the relative numerical error

$$\eta = \frac{\|\mathbf{M}\cdot\mathbf{m} - \mathbf{A}\cdot\mathbf{s}^k\|}{\|\mathbf{M}\cdot\mathbf{m}\|} \quad (5.64)$$

is used for the convergence criteria, where \mathbf{s}^k is the smoothed model obtained at the k^{th} iteration. In the next paragraphs, we study the properties of the linear system (Equation (5.57)), as well as the convergence rate based on the values of the error modulus η .

5.5.2.1 Linear system behavior

First, we build a reference system with smoothed vector \mathbf{s}_{ref} and input \mathbf{m}_{ref} , which is computed as

$$\mathbf{m}_{\text{ref}} = \mathbf{M}^{-1} \cdot \mathbf{A} \cdot \mathbf{s}_{\text{ref}}. \quad (5.65)$$

The linear system

$$\mathbf{A} \cdot \mathbf{s} = \mathbf{M} \cdot \mathbf{m}_{\text{ref}} \quad (5.66)$$

is then solved iteratively by the CG solver. At each iteration, the modulus of the relative model error γ is computed

$$\gamma = \frac{\|\mathbf{s}_{\text{ref}} - \mathbf{s}^k\|}{\|\mathbf{s}_{\text{ref}}\|}. \quad (5.67)$$

The numerical error η is defined in Equation (5.64), and it is compared to this model error γ to characterize the behavior of the linear system of Equation (5.57). An example of this comparison for a model vector that contains 30×10^6 degrees of freedom is shown in Figure 5.8. The filter is defined with homogeneous coherent lengths $L_z = L_x = L_y = 80$ m, without dip and azimuth. It should be noted that the numerical and model error modulus both decrease with the same speed, which implies that the linear system associated with the Bessel filter is well conditioned. The system converges relatively rapidly, and attains the numerical limit of 10^{-6} after 110 iterations, using FORTRAN single-precision arithmetics. Beyond this limit, the relative model accuracy remains unchanged, so the stopping criteria

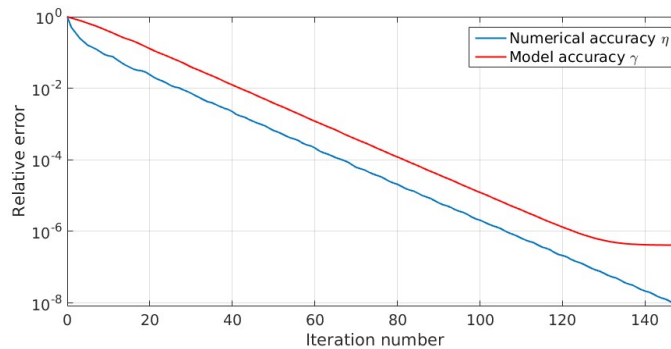


Figure 5.8: Relative numerical error modulus η and model error modulus γ for smoothing a model vector containing $80 \times 80 \times 80$ elements; that is, 30×10^6 degrees of freedom. The coherent lengths are homogeneous $L_z = L_x = L_y = 80$ m, without dip and azimuth.

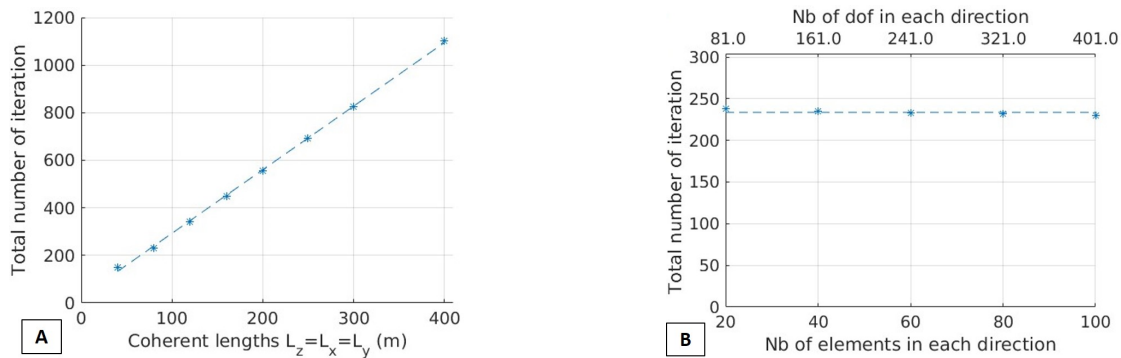


Figure 5.9: (A) Dependence of the total number of iterations (obtained from consecutively solving two linear systems associated with the Bessel filter) on coherent length. The model contains $80 \times 80 \times 80$ elements; i.e., 30×10^6 degrees of freedom. (B) Dependence of the total number of iterations on the size of the problem. For both images, the stopping criterion is $\eta = 10^{-6}$.

can be set as $\eta = 10^{-6}$. In FWI applications, to optimize the numerical cost, the CG solver could be stopped before the numerical limit is attained, because the model parameters will be repetitively modified. Evaluation of this tolerance η should be carefully studied, according to the objective of the application.

5.5.2.2 Convergence properties

After fixing the stopping criteria as $\eta = 10^{-6}$, we investigate the dependence of the convergence rate of the CG solver on several parameters, including coherent length, size of model vector, and presence of dip and azimuth. As we are approximating the Laplace filter by the Bessel operators, the linear system associated with the Bessel filter is solved twice to obtain the smoothed vector. The number of iterations shown in Figure 5.9 is then estimated by consecutively solving two linear systems. Obviously, the same conclusions should be drawn when solving one linear system, which corresponds to the direct application of the Bessel filter.

Figure 5.9A indicates that for a given model the number of iterations required for solving two linear systems increases linearly with the values of the coherent length. This plot is obtained from a model of size $80 \times 80 \times 80$ elements; that is, 30×10^6 degrees of freedom. Each point in the plot is obtained from filtering with the isotropic Bessel operator, where the coherent lengths are homogeneous. We carry out a similar study when the coherent lengths are fixed, and the dip and azimuth angles are introduced into the filter. These parameters have no influence on the computation time. Therefore, these data are not shown here. In Figure 5.9B, the filter shape is kept unchanged, with $L_z = L_x = L_y = 80$ m, but the size of the model vector increases gradually from $20 \times 20 \times 20$ elements to $100 \times 100 \times 100$ elements. The corresponding number of degrees of freedom in each direction is indicated by the axis at the top of Figure 5.9B. For a pre-defined filter, the total number of iterations required for smoothing a model vector is independent of the model size.

5.5.2.3 Numerical performance *versus* explicit 3-D convolution filtering

As Bessel filters can be efficiently used to approximate the Laplace filter, the efficiency of this method needs to be evaluated with respect to the 3-D explicit convolution approach. We limit this study to an

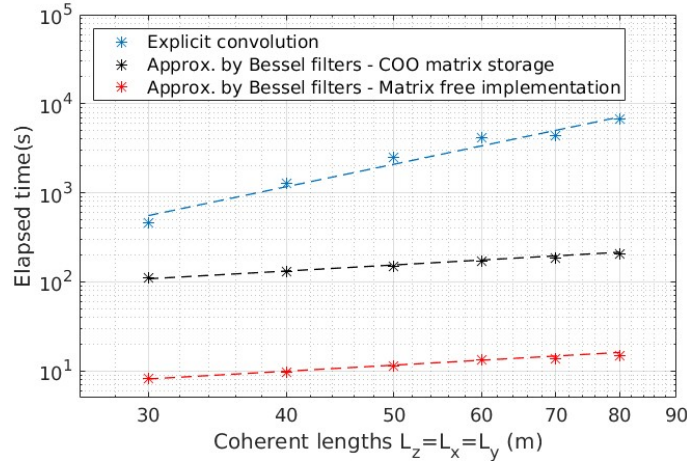


Figure 5.10: Comparison of the numerical performance of this method. The COO storage of nonzero elements in matrix \mathbf{A} (black) and the matrix-free implementation of the product matrix-vector $\mathbf{A}\mathbf{u}$ (red) are compared to the truncated explicit 3-D convolution of the Laplace filter (blue) for a model with 1.77×10^6 degrees of freedom, with decomposition for two subdomains.

isotropic and homogeneous Laplace filter

$$L_{3D}(z, x, y) = \frac{1}{8\pi L^3} e^{-\sqrt{\frac{x^2+y^2+z^2}{L^2}}}. \quad (5.68)$$

The filter theoretically has infinite tail, but it can be implemented as a truncated 3-D convolution of the filter and the model/ gradient vector. We then define a sphere $\Omega_{\alpha L}$ centered at the origin, with radius αL , such that

$$\iiint_{\Omega_{\alpha L}} L_{3D}(z, x, y) dz dx dy = 0.99, \quad (5.69)$$

which provides $\alpha \approx 8.4$ (Figure 5.11). This condition is equivalent to 99% energy of the filter. Figure 5.10 compares the computation cost of our approach with the 3-D explicit convolution in a model with 1.77×10^6 degrees of freedom and where the coherent length increases from 30 m to 80 m. The 3-D explicit convolution is the most expensive method, but the comparison is partially biased as our approaches are highly optimized. However, it should be noted that the numerical cost of the 3-D explicit convolution method is driven by the number of points within the effective sphere $\Omega_{\alpha L}$, thus with the complexity $\mathcal{O}(L^3)$. Whereas, for a given model size the computational complexity of the approximation of the Laplace filter by Bessel operators only depends on the number of CG iterations, which increases linearly with the coherent length ($\propto \mathcal{O}(L^1)$), as shown in Figure 5.9A. In Figure 5.10, the observed numerical complexity of the explicit convolution method *versus* the increase in the coherent length is $\mathcal{O}(L^{2.6})$ due to the limited model size. We obtain a sublinear relationship between the numerical cost of our approximation and the coherent length ($\mathcal{O}(L^{0.7})$), for the two implementation approaches mentioned in section 5.5.1: COO-storage and matrix-free implementation. This is due to some calculations that are independent of the coherent length. This property is encouraging for the application of this approach to large problems.

5.6 Conclusion and discussion

The application of a filter on a given vector is usually described through convolution of the vector by a filter that is a linear operation. Smoothing will require filters with decaying tails. Thus, the convolution operator might be expensive if it is computed explicitly. We promote the idea of applying the filter through its inverse operator, which is expected to be a sparse operator. The smooth output is obtained as the solution of a new linear system, in which the sparse inverse of the filter has been constructed. We introduce the Bessel filters and their corresponding diffusion-like PDEs to be solved numerically in 2-D and 3-D geometries. We analytically explain why the iterative application of two Bessel filters provides an excellent approximation of the Laplace operator for smoothing the model, which offers the choice of either using Bessel filters as smoothing operators or to mimic the decay of the Laplace filter. This argument is well supported by the numerical illustration of spike tests, where the PDE defining the Bessel filter is solved with a delta source function.

These PDEs can be discretized through a weak formulation that involves SEM, which leads to a symmetric linear system that can be efficiently solved through a CG method. There are no specific requirements for the mesh, which can be chosen in a suitable way for the application being considered.

We present a robust and efficient approach to incorporate variable coherent lengths, dip and azimuth into the Bessel filter, without losing the symmetry of the left-hand-side term of the linear system. To do so, we develop the weak formulation of the PDE associated with the sparse inverse filter in the rotated dimensionless coordinate system. The construction of this system naturally takes advantage of the available ingredients in the underlying mesh that are associated to the study problem. All of the information about the geological variations of the medium is preserved in the Jacobi matrix and its inverse, which does not break the symmetry of the stiffness matrix. The well-behaving performance of this approximation is illustrated by spike tests, which closely reproduce the shape of analytical Laplace filter when assuming slow variations of these geological properties. In more complex tests, when random noise is used as the input model, the output vector correctly follows the imposed shape that is described by the filter parameters. When the non-stationary filter is applied to a realistic FWI gradient, it effectively removes near-surface artifacts due to the acquisition footprint, and further enhances the continuity of the deeper structures.

We propose efficient implementation of the application of the Bessel filter, where the linear system is solved with a parallel CG following the same domain decomposition as the wave simulation. This parallel computation of the matrix-vector product at each CG iteration can be achieved either by using explicit building of the local matrix, or in a matrix-free fashion. For the matrix-free implementation, the stiffness matrix-vector product can be factorized and highly optimized, due to the vectorized computation and the optimizing cache usage developed by Deville et al. (2002). This strategy significantly reduces the memory requirement and computation time for the CG solver. After each matrix-vector or vector-vector product, the vectors obtained are spatially communicated to assemble the values on the sub-domain boundaries. Consequently, any projection back and forth from the modelling mesh and the inversion mesh is avoided, as well as any explicit windowed convolution.

The sparse inverse linear system associated with the Bessel filter is well conditioned, due to the structure of the PDE. The CG solver applied to this system converges relatively rapidly. The convergence depends on many factors, but it increases linearly with the coherent length. Other factors, such as the number of elements in the medium and the presence of dip and azimuth angles, appear to have no influence on the number of iterations required for the system to converge. The approximation of the Laplace filter by the Bessel filter proposed here is also compared with the windowed explicit 3-D

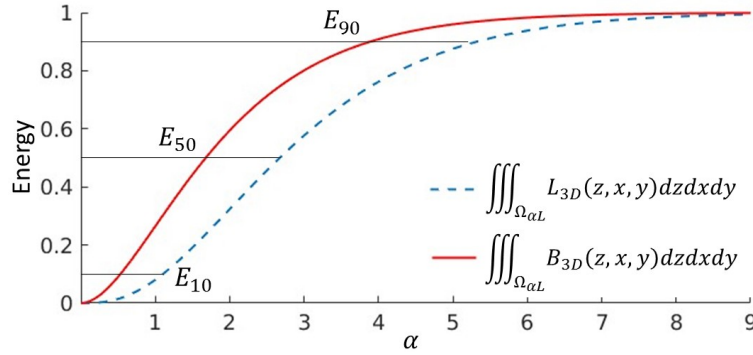


Figure 5.11: Energy analysis for isotropic Laplace and Bessel filters within the sphere $\Omega_{\alpha L}$ centered at the origin with the radius αL (L is the coherent length in the z , x and y directions). E_{10} , E_{50} and E_{90} indicate the 10%, 50% and 90% energy.

convolution method. The approach here is significantly faster, with a sublinear relationship between the numerical cost and the coherent length, which promises general applications to various large problems.

One question might be whether we should apply one Bessel filter with longer coherent length or two Bessel filters, to reproduce the decay of the Laplace filter. Choosing between Bessel and Laplace filters is case dependent, and this should be decided based on the energy distribution of the filters and the numerical cost. Knowing that the energy distribution of these filters are different (as shown in Figure 5.11), they should result in different smoothing effects. To have the same energy as the Laplace filter at 90%, 50% or 10% energy, the coherent length in the Bessel filter should be increased by a factor of 1.4, 1.6, or 2.1, respectively. These different scaling factors also imply that it is not possible to mimic the Laplace filter by simply modifying the coherent length in the Bessel filter. For the same coherent length, the contributions of the neighbors surrounding the origin are more important in the Bessel kernel, which is indicated by the 10% energy level. It should also be noted that increasing the coherent length will eventually increase the numerical cost of the iterative solving of the linear system associated with the Bessel filter.

Only the numerical implementation for SEM is presented here, but the method can be extended to other FE techniques, and also to FD techniques (Wellington, 2016). This study illustrates the potential applications for model preconditioning, but this approach of applying a filter (or a covariance matrix) can be considered for regularization and for prior-model strategies. As the filter has a robust shape, even with dip, azimuth and coherent length variations, it can be used to amplify expected geological features, while attenuating features that come from numerical artifacts in other areas. In this way, the null-space dimension is reduced, to converge to more meaningful models while honoring the data fit.

Appendix (5.D) shows an example of a possible 3D matrix-free implementation scheme for FD technique, which is an on-going collaboration with Wellington (2016).

5.7 Additional numerical investigations

The global equation governing the Bessel smoothing process can be formulated as

$$\mathbf{s}(\mathbf{x}) - \nabla_{z,x,y}^T \mathbf{P}(\mathbf{x}) \mathbf{P}^T(\mathbf{x}) \nabla_{z,x,y} \mathbf{s}(\mathbf{x}) = \mathbf{g}(\mathbf{x}), \quad (5.70)$$

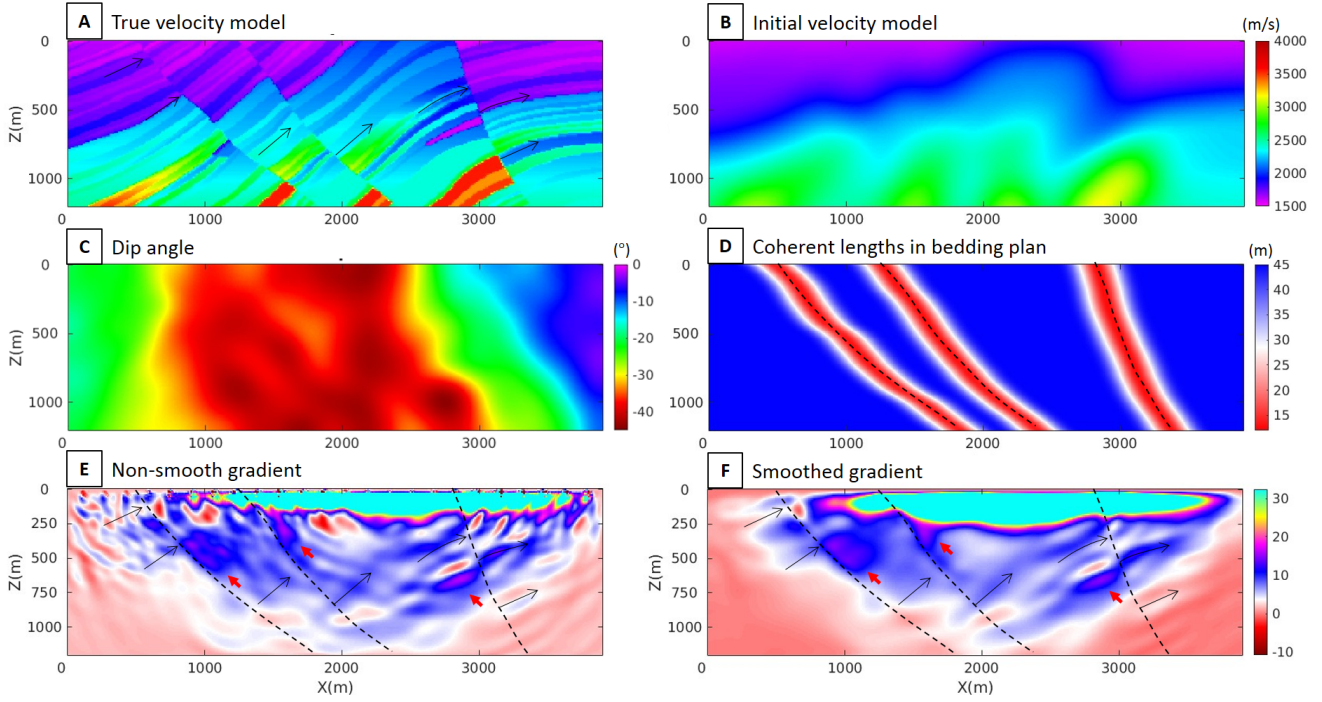


Figure 5.12: Example of the nonstationary filtering operator on FWI gradient, from pseudo-3D Marmousi model. (A) True velocity model. (B) Initial velocity model. (C) Dip field, estimated from the true velocity model. (D) Coherent lengths in bedding plan (L_u and L_w), which vary from 12 m at fault location to 45 m at other places. (E) Original scaled gradient without any smoothing. (F) Smoothed gradient with anisotropic nonstationary Laplace filter (approximated by the application of two Bessel filters): dip field as presented in Figure C, $L_u = L_w$ as presented in Figure D, and $L_v = 12$ m (≈ 0.15 of the shortest wavelength). Some interesting features are highlighted by black and red arrows, and faults are indicated by black dash-lines. Figure extracted from Trinh et al. (2017a)

where the raw gradient $\mathbf{g}(\mathbf{x})$ is transformed into the smoothed vector $\mathbf{s}(\mathbf{x})$. The information related to the geological variation of the medium (i.e. filter parameters) are embedded in the *projection matrix*

$$\mathbf{P}(\mathbf{x}) = \begin{bmatrix} L_v \cos \varphi & L_u \sin \varphi & 0 \\ -L_v \cos \theta \sin \varphi & L_u \cos \theta \cos \varphi & L_w \sin \theta \\ L_v \sin \theta \sin \varphi & -L_u \sin \theta \cos \varphi & L_w \cos \theta \end{bmatrix}, \quad (5.71)$$

appearing in the construction of the rotated dimensionless Jacobian matrix (5.53). The projection matrix describes the projection between the Cartesian space and the locally rotated dimensionless coordinates system.

5.7.1 Gradient smoothing in pseudo-3D Marmousi model

Figure 5.12 illustrates the application of an anisotropic nonstationary Laplace filter (approximated by double application of Bessel filters) on a gradient computed in the Marmousi benchmark. The 2D Marmousi V_p model in Figure 5.12A has been extended to an elastic 3D volume for this test. A surface acquisition is used with a line of 24 sources, with distance 160 m between adjacent sources. The receivers are located on the whole surface, with 12.5 m between receivers. A Ricker wavelet centered

at 8 Hz is used as the source signal. The initial model is a smooth version of the true model as shown in Figure 5.12B.

The 2D cross-section of the V_p gradient without any smoothing filter underneath the source line is shown in Figure 5.12E. The gradient contains significant acquisition footprint at the near-surface and high wavenumber artifacts in the deeper part. To remove these high wavenumber artifacts, we design an anisotropic nonstationary Bessel-based preconditioning with parameters as:

- Azimuth field θ is set as zero due to the pseudo-3D configuration. The smooth dip field φ is estimated from the true velocity model, as shown in Figure 5.12C.
- The vertical coherent length L_v is set as 12 m, which is about 0.15 times of the shortest wavelength. The horizontal coherent length L_u and L_w vary from 12 m near the fault locations to 45 m at other places as shown in Figure 5.12D. This coherent length parameters design is based on the assumption that we know the relative positions of the main faults.

The output smoothed gradient is shown in Figure 5.12F, in which the near-surface acquisition footprint is effectively removed. The continuity of the features at greater depths is enhanced, because the horizontal oscillation artifacts are attenuated, as indicated by the thin black arrows. Due to the design of the coherent lengths, the energy is not smeared out across the faults, indicated by the thick red arrows in Figures 5.12E and F. In this example, the smoothing process costs about 2.2% running time comparing to the forward problem.

5.7.2 Influence of the preconditioning operator on inversion results

Following the setting in Section (5.4.3) and Figure 5.7, we run several FWI tests, each consists of 60 iterations with l -BFGS optimization method. We invert for compressional and shear velocities but only the results for shear wave are presented here. The density is as smooth as the initial V_s model and it is kept unchanged. Figure 5.13C and D are inversion results obtained with the body waves, which arrive before the surface waves. Figure 5.13E and F are obtained respectively from the starting models 5.13C and D, after enlarging the time-window to incorporate surface waves and underlying body wave. Only a simple medium filter ($3 \times 3 \times 3$ pixels) is used for gradient smoothing at each iteration in Figure 5.13C and E. The results 5.13D and F are obtained with a fully nonstationary and anisotropic Bessel-based gradient precondition as described in Figure 5.7.

The inversion result in Figure 5.13C is a clear example for the discussion in Section (5.4.3), where the gradient artifacts in Figure 5.7D leave strong imprints in the final estimated model. In Figure 5.13C, the near-surface is highly broken due to the sparse acquisition footprint. The high wavenumber horizontal-oscillation in the gradient due to poor illumination also damages the continuity of the events at greater depth. When adding more data into the inversion process (from Figure 5.13C to E), the poor illumination issue is partially resolved and the near-surface is improved but the structure discontinuity still pronounces. These results might leads to wrong geological interpretation about medium heterogeneities.

By properly preconditioning the gradient, as shown in Figure 5.7E, we obtain a cleaner model estimation in Figure 5.13D. The model quality is obviously limited by the amount of provided information and the data-resolving capacity of our FWI technique. However, compared with Figure 5.13C, the artifacts linked to the acquisition footprint and poor illumination issues are effectively eliminated. At each iteration, the gradient smoothing requires less than 1% cost of one FWI iteration. When enlarging the

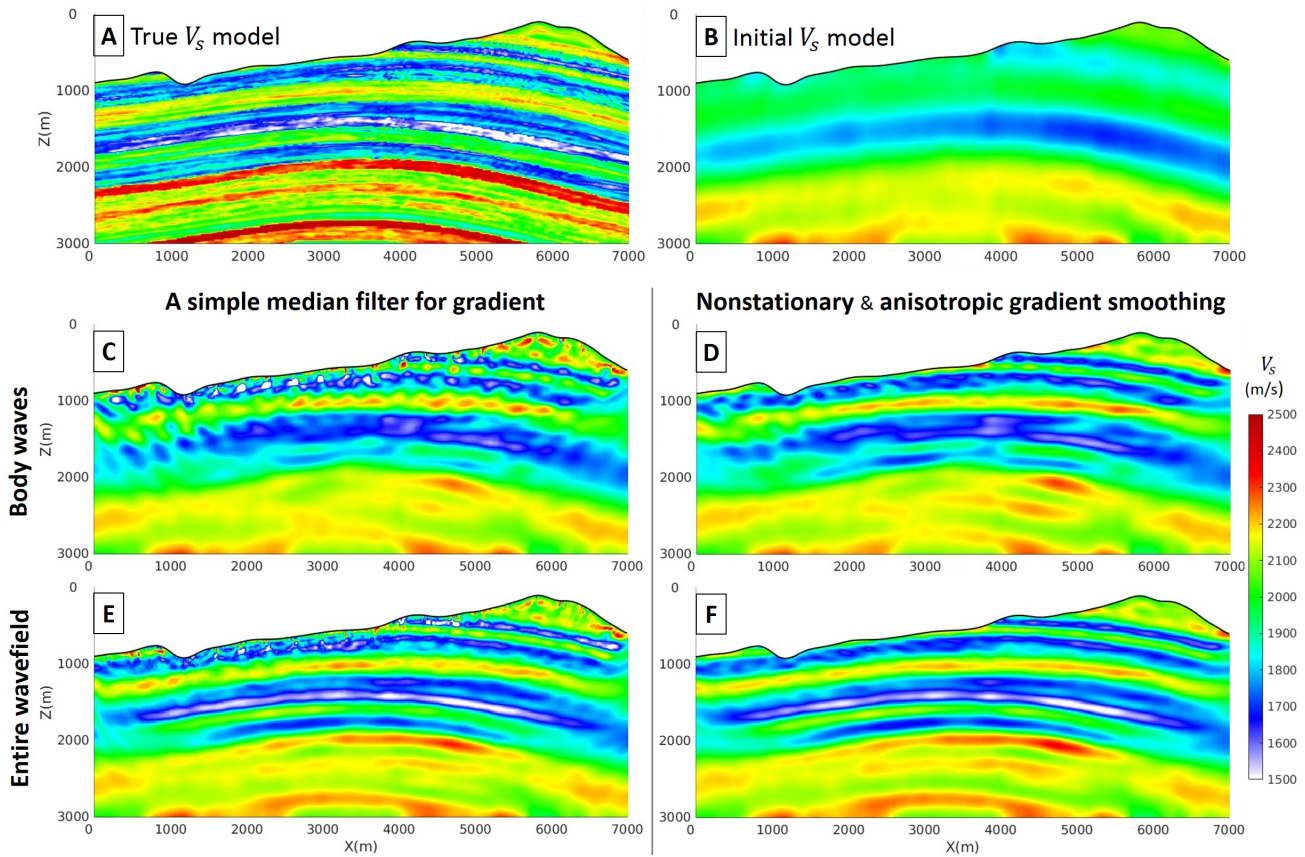


Figure 5.13: A - True V_s model. B - Initial V_s model. C - Inverted model obtained with body waves, where only a median filter over $3 \times 3 \times 3$ degrees of freedom is applied onto the gradient. D - Inverted model obtained with body waves, where the gradient is smoothed by an anisotropic non-stationary Bessel-based filter as described in Figure 5.7: $L_v = 0.05\lambda_s$ and $L_u = L_v = 0.15\lambda_s$, where λ_s is the local shear wavelength. E and F - Inverted models obtained respectively from C and D, after enlarging the time-window to incorporate surface waves and underlying body wave. Figure extracted and modified from Trinh et al. (2017c)

time-window to incorporate more data into the inversion process (Figure 5.13F), the inversion result is further enhanced, and still cleaner than a naive gradient preconditioning strategy in Figure 5.13E.

APPENDICES

5.A Modified Bessel functions and their integral representation

In this Appendix, we provide the definition of the modified Bessel function of the second kind $K_\nu(x)$ for $\nu \in \mathbf{R}$. This function is defined from the modified Bessel function of the first kind $I_\nu(x)$ (Abramowitz and Stegun, 1972), written in the power series expansion

$$I_\nu(x) = \left(\frac{x}{2}\right)^\nu \sum_{k=0}^{\infty} \frac{1}{\Gamma(k+1)\Gamma(\nu+k+1)} \left(\frac{x}{2}\right)^{2k}, \quad (5.72)$$

where the gamma function $\Gamma(x)$ is defined as

$$\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt. \quad (5.73)$$

When ν is an integer, we have the following property of the function $I_\nu(x)$

$$I_\nu(x) = I_{-\nu}(x). \quad (5.74)$$

According to the definition of the function $I_\nu(x)$ in Equation (5.72), we describe how it behaves when $x \rightarrow 0$:

$$\lim_{x \rightarrow 0} I_\nu(x) = \begin{cases} 0 & \text{Re}(\nu) > 0 \\ 1 & \nu = 0 \\ \pm\infty & \text{Re}(\nu) < 0 \end{cases}. \quad (5.75)$$

This implies that $I_\nu(x)$ and $I_{-\nu}(x)$ are linearly independent if ν is not an integer. The modified Bessel function of the second kind $K_\nu(x)$ is then defined as

$$K_\nu(x) = \frac{\pi}{2} \frac{I_{-\nu} - I_\nu}{\sin \pi \nu}, \quad (5.76)$$

for $\nu \notin \mathbf{Z}$. For $n \in \mathbf{Z}$, the function $K_n(x)$ is defined as

$$K_n(x) := \lim_{\nu \rightarrow n} K_\nu(x). \quad (5.77)$$

5.B Definition of Bessel filters and their normalized factors

In this Appendix, we demonstrate that 2-D and 3-D Bessel filters are unique solutions of the PDEs (5.8) and (5.9), respectively, when considering the radiative boundary condition

$$B_{nD}(\mathbf{x}) \rightarrow 0 \quad \text{as} \quad \|\mathbf{x}\| \rightarrow \infty \quad (5.78)$$

and the boundness property

$$0 < \int B_{nD}(\mathbf{x}) d\mathbf{x} < \infty \quad (5.79)$$

The 2-D case

We look for all possible analytical solutions of Equation (5.9):

$$C(z, x) - \left(L_z^2 \frac{\partial^2}{\partial z^2} + L_x^2 \frac{\partial^2}{\partial x^2} \right) C(z, x) = \delta(z, x). \quad (5.80)$$

Changing from Cartesian to dimensionless coordinates

$$x \rightarrow \tilde{x} = \frac{x}{L_x} \quad ; \quad z \rightarrow \tilde{z} = \frac{z}{L_z}, \quad (5.81)$$

knowing that L_x and L_z are homogeneous, we obtain

$$C(\tilde{z}, \tilde{x}) - \left(\frac{\partial^2}{\partial \tilde{x}^2} C(\tilde{z}, \tilde{x}) + \frac{\partial^2}{\partial \tilde{z}^2} C(\tilde{z}, \tilde{x}) \right) = \delta(\tilde{z}, \tilde{x}). \quad (5.82)$$

We first solve the homogeneous form of Equation (5.82) for $(\tilde{z}, \tilde{x}) \neq (0, 0)$. Again, we move from dimensionless coordinates (\tilde{z}, \tilde{x}) to polar coordinates (r, θ) , and assume that the solution is separated; that is, $C(r, \theta) = R(r)A(\theta)$. The homogeneous form of Equation (5.82) now becomes

$$R(r)A(\theta) - \left(R''(r)A(\theta) + \frac{1}{r}R'(r)A(\theta) + \frac{1}{r^2}R(r)A''(\theta) \right) = 0. \quad (5.83)$$

Dividing both sides of this equation by $R(r)A(\theta)r^2$:

$$\frac{r^2 R''(r)}{R(r)} + \frac{r R'(r)}{R(r)} - r^2 = -\frac{A''(\theta)}{A(\theta)}. \quad (5.84)$$

As $R(r)$ and $A(\theta)$ are independent, Equation (5.84) can be set as a constant V , for $r > 0$. Note that valid solutions must satisfy the periodicity in the angular domain; that is, $A(\theta) = A(\theta + 2\pi)$, which leads to a nonnegative value of V .

By considering the angular part of the solution:

$$\frac{A''(\theta)}{A(\theta)} = -V \iff A''(\theta) + V.A(\theta) = 0. \quad (5.85)$$

Two independent real solutions of this equation are

$$A_1(\theta) = \cos(\sqrt{V}\theta) \quad ; \quad A_2(\theta) = \sin(\sqrt{V}\theta). \quad (5.86)$$

As $A(\theta) = A(\theta + 2\pi)$, V must satisfy $V = \nu^2$, where ν is an integer. The radial part now becomes

$$r^2 R''(r) + r R'(r) - (r^2 + \nu^2)R(r) = 0, \quad (5.87)$$

which is actually the modified Bessel differential equation (Polyanin and Nazaikinskii, 2002; Abramowitz and Stegun, 1972). Two linearly independent solutions of this equation are the modified Bessel function of the first kind $I_\nu(r)$, singular at ∞ , and the second kind $K_\nu(r)$, singular at zero, where ν is a real number.

As we are interested in building a smoothing filter, it must vanish at infinity (a condition of Equation (5.78)):

$$R(r) \rightarrow 0 \quad \text{as} \quad r \rightarrow \infty, \quad (5.88)$$

which implies that the modified Bessel function of the first kind I_ν can be excluded, as

$$\lim_{r \rightarrow +\infty} I_\nu(r) = +\infty. \quad (5.89)$$

In the polar coordinates (r, θ) , the condition of Equation (5.79) can be re-written as

$$0 < \int_{-\pi}^{\pi} d\theta \int_0^{\infty} r dr C(r, \theta) < \infty. \quad (5.90)$$

As our solution has the form $C(r, \theta) = R(r)A(\theta)$, this condition becomes

$$0 < \int_{-\pi}^{\pi} d\theta A(\theta) \int_0^{\infty} r dr R(r) < \infty. \quad (5.91)$$

The angular part of the forward filter has the form

$$A_\nu(\theta) = c_1 \cos(\nu\theta) + c_2 \sin(\nu\theta) \quad \text{where } \nu = 0, \pm 1, \pm 2 \dots \quad (5.92)$$

and c_1, c_2 are constant, which leads to

$$\int_{-\pi}^{\pi} d\theta A_\nu(\theta) = \begin{cases} c_1 \sin(\nu\theta)|_{-\pi}^{\pi} - c_2 \cos(\nu\theta)|_{-\pi}^{\pi} = 0 & \text{if } \nu = \pm 1, \pm 2 \dots \\ 2\pi(c_1 + c_2) = \text{Cst} > 0 & \text{if } \nu = 0. \end{cases} \quad (5.93)$$

Consequently, the modified Bessel function of the second kind K_ν with $\nu > 0$ cannot be our expected filter. In addition,

$$\int_0^{\infty} r K_0(r) dr = 1, \quad (5.94)$$

which clearly satisfies the condition

$$0 < \int_{-\pi}^{\pi} d\theta A_0(\theta) \int_0^{\infty} r dr K_0(r) < \infty. \quad (5.95)$$

The boundary conditions at infinity imply that K_0 is the only possible solution of the homogeneous form of Equation (5.82). As the function $K_0(r)$ is singular at zero, the normalized function $\frac{1}{2\pi L_x L_z} K_0\left(\sqrt{\frac{x^2}{L_x^2} + \frac{z^2}{L_z^2}}\right)$ is also the solution of the non-homogeneous Equation (5.80).

The 3-D case

Similar to the 2-D case, we look for all possible solution of Equation (5.8):

$$C(z, x, y) - \left(L_z^2 \frac{\partial^2}{\partial z^2} + L_x^2 \frac{\partial^2}{\partial x^2} + L_y^2 \frac{\partial^2}{\partial y^2} \right) C(z, x, y) = \delta(z, x, y). \quad (5.96)$$

By repeating the same workflow as for the 2-D case, moving from Cartesian coordinates to spherical coordinates, and assuming again that the solution is separated $C(r, \theta, \phi) = R(r)A(\theta)B(\phi)$, where $0 < r < \infty$, $0 \leq \theta \leq \pi$, and $-\pi \leq \phi \leq \pi$, the homogeneous form of Equation (5.96) becomes

$$R(r)A(\theta)B(\phi) - \left(\frac{\partial^2 R(r)}{\partial r^2} + \frac{2}{r} \frac{\partial R(r)}{\partial r} \right) A(\theta)B(\phi) - \frac{1}{r^2} \left(\frac{\cos \theta}{\sin \theta} \frac{\partial A(\theta)}{\partial \theta} + \frac{\partial^2 A(\theta)}{\partial \theta^2} \right) R(r)B(\phi)$$

$$- \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2 B(\phi)}{\partial \phi^2} R(r) A(\theta) = 0. \quad (5.97)$$

Multiplying both sides of this equation by $\frac{r^2 \sin \theta}{R(r) A(\theta) B(\phi)}$, assuming it does not vanish, we obtain

$$\sin^2 \theta \left(r^2 - r^2 \frac{R''(r)}{R(r)} - 2r \frac{R'(r)}{R(r)} \right) - \left(\cos \theta \sin \theta \frac{A'(\theta)}{A(\theta)} + \sin^2 \theta \frac{A''(\theta)}{A(\theta)} \right) - \frac{B''(\phi)}{B(\phi)} = 0, \quad (5.98)$$

which should be valid for all different sets of values of (r, θ, ϕ) . By choosing $\theta = 0$, Equation (5.98) should hold for any values of r and ϕ , leading to

$$\frac{B''(\phi)}{B(\phi)} = 0 \implies B(\phi) = \text{Constant} \quad (5.99)$$

(because this function is periodic $B(\phi) = B(\phi + 2\pi)$). Replacing this expression back into Equation (5.98), and dividing both sides of the equation by $\sin \theta$, we obtain

$$\sin \theta \left(r^2 - r^2 \frac{R''(r)}{R(r)} - 2r \frac{R'(r)}{R(r)} \right) - \left(\cos \theta \frac{A'(\theta)}{A(\theta)} + \sin \theta \frac{A''(\theta)}{A(\theta)} \right) = 0. \quad (5.100)$$

Again, this equation should hold for any values of r and $\theta = 0$, which leads to

$$\frac{A'(\theta)}{A(\theta)} = 0 \implies A(\theta) = \text{Constant}. \quad (5.101)$$

Replacing this expression back into Equation (5.100), we get

$$r^2 R''(r) + 2r R'(r) - r^2 R(r) = 0. \quad (5.102)$$

According to Abramowitz and Stegun (1972), this is the modified spherical Bessel function, which has two particular solutions:

$$\sqrt{\frac{1}{r}} I_{\pm 1/2}(r) \quad \text{and} \quad \sqrt{\frac{1}{r}} K_{1/2}(r). \quad (5.103)$$

Again, we expect that the solution vanishes at infinity, which means that $r^{-1/2} I_{\pm 1/2}(r)$ can be excluded. The function $r^{-1/2} K_{1/2}(r)$ satisfies the normalizable condition of Equation (5.79) in 3-D as

$$\int_{-\pi}^{\pi} d\phi \int_0^{\pi} \sin \theta d\theta \int_0^{\infty} r^2 [r^{-1/2} K_{1/2}(r)] dr = (2\pi)^{3/2}. \quad (5.104)$$

Again, as the function $K_0(r)$ is singular at zero, the normalized function $\frac{1}{(2\pi)^{3/2} L_x L_y L_z} r^{-1/2} K_{1/2}(r)$ is also the solution of the non-homogeneous Equation (5.96).

5.C Weak formulation development

This Appendix clarifies the definition of the mesh used in SEM, which consists of Gauss-Lobatto-Legendre (GLL) points, defined in the reference space. The base functions are Lagrange polynomials. GLL quadrature is used to approximate analytical integrals over collocation points. Let us recall the weak formulation of Equation (5.26)

$$\iiint_{\tilde{\Omega}} \mathbf{s} v \, d\tilde{z} d\tilde{x} d\tilde{y} + \iiint_{\tilde{\Omega}} \nabla_{\tilde{z}, \tilde{x}, \tilde{y}} \mathbf{s} \cdot \nabla_{\tilde{z}, \tilde{x}, \tilde{y}} v \, d\tilde{z} d\tilde{x} d\tilde{y} = \iiint_{\tilde{\Omega}} \mathbf{m} v \, d\tilde{z} d\tilde{x} d\tilde{y}, \quad (5.105)$$

where $v(\tilde{z}, \tilde{x}, \tilde{y})$ is a test function, \mathbf{m} and \mathbf{s} are the original and the smoothed vectors, respectively.

In SEM, the physical domain Ω is decomposed into a set of nonoverlapping hexahedral elements Ω_e . This implies that the dimensionless space $\tilde{\Omega}$ is also represented by a set of elements $\tilde{\Omega}_e$, which is related to the physical element Ω_e through the one-to-one mapping (Equation (5.24)). Each reduced coordinate element $\tilde{\Omega}_e$ can be mapped to the unitary reference space of the GLL points, where the cube $[-1, 1] \otimes [-1, 1] \otimes [-1, 1]$ is discretized into $(N + 1) \times (N + 1) \times (N + 1)$ GLL points $(\xi_{k_1}, \eta_{k_2}, \zeta_{k_3})$; $k_1, k_2, k_3 = 0, \dots, N$. These collocation points define $(N + 1) \times (N + 1) \times (N + 1)$ base functions, which are triple products of Lagrange polynomials of degree N , over the element $\tilde{\Omega}_e$

$$v_{\hat{k}}(\xi, \eta, \zeta) = \ell_{k_1}(\xi)\ell_{k_2}(\eta)\ell_{k_3}(\zeta) \quad \hat{k} \text{ stands for the triple of indexes } \{k_1, k_2, k_3\}. \quad (5.106)$$

Lagrange polynomials have the interesting property that their values at GLL nodes are either 0 or 1

$$\ell_j(\xi) = \prod_{\substack{i=0 \\ i \neq j}}^N \frac{\xi - \xi_i}{\xi_j - \xi_i} \quad ; \quad \ell_j(\xi_i) = \delta_{ji}. \quad (5.107)$$

An element of volume $d\tilde{z}d\tilde{x}d\tilde{y}$ is related to an element of volume $d\xi d\eta d\zeta$ in the reference cube by

$$d\tilde{z}d\tilde{x}d\tilde{y} = \tilde{J}_e d\xi d\eta d\zeta, \quad (5.108)$$

where the volumetric Jacobian \tilde{J}_e is the determinant of the Jacobian matrix $\tilde{\mathbf{J}}$

$$\tilde{\mathbf{J}} = \begin{bmatrix} \partial_\xi \tilde{z} & \partial_\eta \tilde{z} & \partial_\zeta \tilde{z} \\ \partial_\xi \tilde{x} & \partial_\eta \tilde{x} & \partial_\zeta \tilde{x} \\ \partial_\xi \tilde{y} & \partial_\eta \tilde{y} & \partial_\zeta \tilde{y} \end{bmatrix}. \quad (5.109)$$

We introduce the following notations of the coordinates of the dimensionless physical space and the reference space

$$(p_1, p_2, p_3) := (\tilde{z}, \tilde{x}, \tilde{y}) \quad \text{and} \quad (r_1, r_2, r_3) := (\xi, \eta, \zeta), \quad (5.110)$$

to develop the weak formulation in Equation (5.105) as

$$\iiint_{\tilde{\Omega}} \mathbf{s} v_{\hat{k}} d\tilde{z}d\tilde{x}d\tilde{y} + \iiint_{\tilde{\Omega}} \sum_{d=1}^3 \frac{\partial v_{\hat{k}}}{\partial p_d} \frac{\partial \mathbf{s}}{\partial p_d} d\tilde{z}d\tilde{x}d\tilde{y} = \iiint_{\tilde{\Omega}} \mathbf{m} v_{\hat{k}} d\tilde{z}d\tilde{x}d\tilde{y}. \quad (5.111)$$

When moving to the reference space, according to the chain rule, this equation becomes

$$\begin{aligned} \iiint_{-1}^1 \mathbf{s} v_{\hat{k}} \tilde{J}_e(\xi, \eta, \zeta) d\xi d\eta d\zeta + \iiint_{-1}^1 \sum_{d=1}^3 \left[\sum_{i=1}^3 \frac{\partial v_{\hat{k}}}{\partial r_i} \frac{\partial r_i}{\partial p_d} \right] \left[\sum_{j=1}^3 \frac{\partial \mathbf{s}}{\partial r_j} \frac{\partial r_j}{\partial p_d} \right] \tilde{J}_e(\xi, \eta, \zeta) d\xi d\eta d\zeta \\ = \iiint_{-1}^1 \mathbf{m} v_{\hat{k}} \tilde{J}_e(\xi, \eta, \zeta) d\xi d\eta d\zeta. \end{aligned} \quad (5.112)$$

or

$$\iiint_{-1}^1 \mathbf{s} v_{\hat{k}} \tilde{J}_e(\xi, \eta, \zeta) d\xi d\eta d\zeta + \iiint_{-1}^1 \sum_{i=1}^3 \sum_{j=1}^3 \frac{\partial v_{\hat{k}}}{\partial r_i} \left[\sum_{d=1}^3 \frac{\partial r_i}{\partial p_d} \frac{\partial r_j}{\partial p_d} \right] \frac{\partial \mathbf{s}}{\partial r_j} \tilde{J}_e(\xi, \eta, \zeta) d\xi d\eta d\zeta$$

$$= \iiint_{-1}^1 \mathbf{m} v_{\hat{k}} \tilde{J}_e(\xi, \eta, \zeta) d\xi d\eta d\zeta. \quad (5.113)$$

In the following step, we use the approximation of a field \mathbf{u} over the element $[-1, 1] \otimes [-1, 1] \otimes [-1, 1]$, as follows

$$\mathbf{u}(\xi, \eta, \zeta) \approx \sum_{h_1=0}^N \sum_{h_2=0}^N \sum_{h_3=0}^N u_{\hat{h}} \ell_{h_1}(\xi) \ell_{h_2}(\eta) \ell_{h_3}(\zeta) =: \sum_{\hat{h}=0}^N u_{\hat{h}} v_{\hat{h}}(\xi, \eta, \zeta), \quad (5.114)$$

where $u_{\hat{h}} = \mathbf{u}(\xi_{h_1}, \eta_{h_2}, \zeta_{h_3})$ for vectors \mathbf{m} and \mathbf{s} . Therefore, Equation (5.113) becomes

$$\begin{aligned} \sum_{\hat{h}=0}^N \iiint_{-1}^1 v_{\hat{k}} v_{\hat{h}} s_{\hat{h}} \tilde{J}_e d\xi d\eta d\zeta + \sum_{\hat{h}=0}^N \iiint_{-1}^1 \sum_{i=1}^3 \sum_{j=1}^3 \frac{\partial v_{\hat{k}}}{\partial r_i} \left[\sum_{d=1}^3 \frac{\partial r_i}{\partial p_d} \frac{\partial r_j}{\partial p_d} \right] \frac{\partial v_{\hat{h}}}{\partial r_j} s_{\hat{h}} \tilde{J}_e d\xi d\eta d\zeta \\ = \sum_{\hat{h}=0}^N \iiint_{-1}^1 v_{\hat{k}} v_{\hat{h}} m_{\hat{h}} \tilde{J}_e d\xi d\eta d\zeta. \end{aligned} \quad (5.115)$$

The integral over the reference cube can be approximated by the GLL quadrature

$$\iiint_{-1}^1 f(\xi, \eta, \zeta) d\xi d\eta d\zeta \approx \sum_{\hat{q}=0}^N w_{q_1} w_{q_2} w_{q_3} f(\xi_{q_1}, \eta_{q_2}, \zeta_{q_3}), \quad (5.116)$$

where $w_{q_1}, w_{q_2}, w_{q_3}$ are quadrature weights. As the test function $v_{\hat{k}}$ is the triple product of Lagrange polynomials (Equation 5.106), which is 1 at node $(\xi_{k_1}, \eta_{k_2}, \zeta_{k_3})$, and 0 at other nodes (Equation (5.107)), the weak formulation developed in Equation (5.115) can be transformed into a linear system

$$\mathbf{M}\mathbf{s} + \mathbf{K}\mathbf{s} = \mathbf{M}\mathbf{m}. \quad (5.117)$$

The mass matrix \mathbf{M} is diagonal with

$$M_{\hat{k}\hat{k}} = w_{k_1} w_{k_2} w_{k_3} \tilde{J}_e(\xi_{k_1}, \eta_{k_2}, \zeta_{k_3}) \quad (5.118)$$

and the stiffness matrix \mathbf{K} is symmetric and sparse

$$K_{\hat{k}\hat{h}} = \sum_{\hat{q}=0}^N w_{q_1} w_{q_2} w_{q_3} \tilde{J}_e(\xi_{q_1}, \eta_{q_2}, \zeta_{q_3}) \left[\sum_{i=1}^3 \sum_{j=1}^3 \frac{\partial v_{\hat{k}}}{\partial r_i} \left(\sum_{d=1}^3 \frac{\partial r_i}{\partial p_d} \frac{\partial r_j}{\partial p_d} \right) \frac{\partial v_{\hat{h}}}{\partial r_j} \right]. \quad (5.119)$$

The property mentioned in Equation (5.107) of the test functions can be used again to simplified the spatial derivatives of the test functions as

$$\frac{\partial v_{\hat{k}}}{\partial \xi}(\xi_{q_1}, \eta_{q_2}, \zeta_{q_3}) = \ell'_{k_1}(\xi_{q_1}) \delta_{k_2, q_2} \delta_{k_3, q_3}, \quad (5.120)$$

similar expressions can be written for $\partial v_{\hat{k}}/\partial \eta$ and $\partial v_{\hat{k}}/\partial \zeta$ at $(\xi_{q_1}, \eta_{q_2}, \zeta_{q_3})$. Therefore, the element at row \hat{k}^{th} , column \hat{h}^{th} of the matrix \mathbf{K} can be explicitly computed as

$$K_{\hat{k}\hat{h}} = \sum_{q_1=0}^N w_{q_1} w_{k_2} w_{k_3} J_e(\xi_{q_1}, \eta_{k_2}, \zeta_{k_3}) \ell'_{h_1}(\xi_{q_1}) \ell'_{k_1}(\xi_{q_1}) \left[\sum_{d=1}^3 (\partial_{p_d} \xi(\xi_{q_1}, \eta_{k_2}, \zeta_{k_3}))^2 \right]$$

$$\begin{aligned}
 & + \sum_{q_2=0}^N w_{k_1} w_{q_2} w_{k_3} J_e(\xi_{k_1}, \eta_{q_2}, \zeta_{k_3}) \ell'_{h_2}(\eta_{q_2}) \ell'_{k_2}(\eta_{q_2}) \left[\sum_{d=1}^3 (\partial_{p_d} \eta(\xi_{k_1}, \eta_{q_2}, \zeta_{k_3}))^2 \right] \\
 & + \sum_{q_3=0}^N w_{k_1} w_{k_2} w_{q_3} J_e(\xi_{k_1}, \eta_{k_2}, \zeta_{q_3}) \ell'_{h_3}(\zeta_{q_3}) \ell'_{k_3}(\zeta_{q_3}) \left[\sum_{d=1}^3 (\partial_{p_d} \zeta(\xi_{k_1}, \eta_{k_2}, \zeta_{q_3}))^2 \right] \\
 & + w_{k_1} w_{h_2} w_{k_3} J_e(\xi_{k_1}, \eta_{h_2}, \zeta_{k_3}) \ell'_{h_1}(\xi_{k_1}) \ell'_{k_2}(\eta_{h_2}) \left[\sum_{d=1}^3 \partial_{p_d} \xi(\xi_{k_1}, \eta_{h_2}, \zeta_{k_3}) \partial_{p_d} \eta(\xi_{k_1}, \eta_{h_2}, \zeta_{k_3}) \right] \\
 & + w_{h_1} w_{k_2} w_{k_3} J_e(\xi_{h_1}, \eta_{k_2}, \zeta_{k_3}) \ell'_{h_2}(\eta_{k_2}) \ell'_{k_1}(\xi_{h_1}) \left[\sum_{d=1}^3 \partial_{p_d} \xi(\xi_{h_1}, \eta_{k_2}, \zeta_{k_3}) \partial_{p_d} \eta(\xi_{h_1}, \eta_{k_2}, \zeta_{k_3}) \right] \\
 & + w_{k_1} w_{k_2} w_{h_3} J_e(\xi_{k_1}, \eta_{k_2}, \zeta_{h_3}) \ell'_{h_1}(\xi_{k_1}) \ell'_{k_3}(\zeta_{h_3}) \left[\sum_{d=1}^3 \partial_{p_d} \xi(\xi_{k_1}, \eta_{k_2}, \zeta_{h_3}) \partial_{p_d} \zeta(\xi_{k_1}, \eta_{k_2}, \zeta_{h_3}) \right] \\
 & + w_{h_1} w_{k_2} w_{k_3} J_e(\xi_{h_1}, \eta_{k_2}, \zeta_{k_3}) \ell'_{h_3}(\zeta_{k_3}) \ell'_{k_1}(\xi_{h_1}) \left[\sum_{d=1}^3 \partial_{p_d} \xi(\xi_{h_1}, \eta_{k_2}, \zeta_{k_3}) \partial_{p_d} \zeta(\xi_{h_1}, \eta_{k_2}, \zeta_{k_3}) \right] \\
 & + w_{k_1} w_{k_2} w_{h_3} J_e(\xi_{k_1}, \eta_{k_2}, \zeta_{h_3}) \ell'_{h_2}(\eta_{k_2}) \ell'_{k_3}(\zeta_{h_3}) \left[\sum_{d=1}^3 \partial_{p_d} \eta(\xi_{k_1}, \eta_{k_2}, \zeta_{h_3}) \partial_{p_d} \zeta(\xi_{k_1}, \eta_{k_2}, \zeta_{h_3}) \right] \\
 & + w_{k_1} w_{h_2} w_{k_3} J_e(\xi_{k_1}, \eta_{h_2}, \zeta_{k_3}) \ell'_{h_3}(\zeta_{k_3}) \ell'_{k_2}(\eta_{h_2}) \left[\sum_{d=1}^3 \partial_{p_d} \eta(\xi_{k_1}, \eta_{h_2}, \zeta_{k_3}) \partial_{p_d} \zeta(\xi_{k_1}, \eta_{h_2}, \zeta_{k_3}) \right]. \quad (5.121)
 \end{aligned}$$

5.D Matrix-free finite-difference implementation scheme

The appendix provides an example of the finite difference implementation of Bessel smoothing strategy in a matrix-free perspective. The numerical implementation is inspired from the structure-tensor implementation in Hale (2009). Since the matrix $\mathbf{P}(\mathbf{x})\mathbf{P}^T(\mathbf{x})$ in Equation (5.70) is symmetric, only six elements of this product need to be computed. They can be either computed outside or inside the loop as

```

! Compute 6 tables of the geometric factor
G11=P11*P11+P12*P12+P13*P13
G12=P11*P21+P12*P22+P13*P23
G13=P11*P31+P12*P32+P13*P33

G22=P21*P21+P22*P22+P23*P23
G23=P21*P31+P22*P32+P23*P33

G33=P31*P31+P32*P32+P33*P33
    
```

Under any numerical discretization method, the linear system derived from the self-adjoint PDE (5.70) is symmetric and positive-definite, the conjugate gradient (CG) method is thus a nature choice for solving this system. Inside each CG iteration, the double spatial derivatives of a given vector $\mathbf{s}(\mathbf{x})$

$$\mathbf{d}(\mathbf{x}) = \nabla_{z,x,y}^T \mathbf{P}(\mathbf{x}) \mathbf{P}^T(\mathbf{x}) \nabla_{z,x,y} \mathbf{s}(\mathbf{x}), \quad (5.122)$$

can be estimated following a matrix-free fashion, at second-order spatial precision, as

STRUCTURE-ORIENTED BESSEL GRADIENT PRECONDITIONING

```
DO i3=1,n3 ! Looping in y direction
  DO i2=1,n2 ! Looping in x direction
    DO i1=1,n1 ! Looping in z direction
      ! Smoothing tensor coefficients at nodes (i1,i2,i3)
      g11=G11(i1,i2,i3); g12=G12(i1,i2,i3); g13=G13(i1,i2,i3);
      g22=G22(i1,i2,i3); g23=G23(i1,i2,i3); g33=G33(i1,i2,i3);

      ! Extract the neighbors
      r000=S(i1,i2,i3); r001=S(i1,i2,i3-1); r010=S(i1,i2-1,i3);
      r011=S(i1,i2-1,i3-1); r100=S(i1-1,i2,i3); r101=S(i1-1,i2,i3-1);
      r110=S(i1-1,i2-1,i3); r111=S(i1-1,i2-1,i3-1);

      ! Spatial derivative of the vector S in the "shifted" grid
      ! The first spatial derivative centered at (i1-1/2,i2-1/2,i3-1/2)
      ra = r000-r111; rb = r001-r110; rc = r010-r101; rd = r011-r100;

      r1= (ra+rb+rc+rd)/4/hz ! Eq(c1)
      r2= (ra+rb-rc-rd)/4/hx ! Eq(c2)
      r3= (ra-rb+rc-rd)/4/hy ! Eq(c3)

      ! Multiply by the smoothing tensor, and divide for the normalization 4*h
      s1=(g11*r1+g12*r2+g13*r3)/4/hz
      s2=(g12*r1+g22*r2+g23*r3)/4/hx
      s3=(g13*r1+g23*r2+g33*r3)/4/hy

      ! Scatter back to the "normal" grid for the estimation of vector D
      ! The second spatial derivative centered at (i1,i2,i3)
      D(i1,i2,i3)=D(i1,i2,i3)+s1+s2+s3 !r000 Eq(a1)
      D(i1,i2,i3-1)=D(i1,i2,i3-1)+s1+s2-s3 !r001 Eq(a2)
      D(i1,i2-1,i3)=D(i1,i2-1,i3)+s1-s2+s3 !r010 Eq(a3)
      D(i1,i2-1,i3-1)=D(i1,i2-1,i3-1)+s1-s2-s3 !r011 Eq(a4)
      D(i1-1,i2,i3)=D(i1-1,i2,i3)-s1+s2+s3 !r100 Eq(a5)
      D(i1-1,i2,i3-1)=D(i1-1,i2,i3-1)-s1+s2-s3 !r101 Eq(a6)
      D(i1-1,i2-1,i3)=D(i1-1,i2-1,i3)-s1-s2+s3 !r110 Eq(a7)
      D(i1-1,i2-1,i3-1)=D(i1-1,i2-1,i3-1)-s1-s2-s3 !r111 Eq(a8)
    ENDDO
  ENDDO
ENDDO
```

The work is an on-going continuation of the collaboration with Wellington (2016).

Part II

Applications

This part is based on the following extended abstracts and papers

-
- *Trinh, P. T., Brossier, R., Métivier, L., and Virieux, J. (2018). Data-oriented strategy and Vp/Vs model-constraint for simultaneous Vp and Vs reconstruction in 3D viscoelastic FWI: Application to the SEAM ii foothills dataset. In 88th SEG Conference and Exhibition 2018, Anaheim*

Chapter 6

Synthetic examples

Contents

6.1 Elastic 3D configuration	130
6.1.1 Model description & 3D acquisition design	130
6.1.2 Inversion setup	131
6.1.3 Inversion results	133
6.1.4 Discussions	135
6.2 SEAM II Foothill dataset	139
6.2.1 SEAM II Foothill Dip-line configuration	139
6.2.2 Inversion setup	140
6.2.3 Inversion results & Discussions	144
6.2.4 Perspectives	146
6.3 Conclusion	147

In this chapter, we apply our FWI algorithm on different subsets of the SEAM Phase II Foothill benchmark (Oristaglio, 2012, 2016; Regone et al., 2017) to highlight the importance of each element of the methodological workflow which has been presented in the previous chapters. We are interested in the FWI applications for simultaneous reconstruction of elastic parameter V_p and V_s . Applications on similar sub-sets of the SEAM Phase II Foothill benchmark have been successfully conducted by using an envelop-based misfit function as presented in Borisov et al. (2017).

The first example is a 3D isotropic elastic configuration in which the observed data is generated on a different mesh compared with the inversion. We consider an “easy setting” with full 3D illumination and close initial models to focus on the challenge of multi-parameters FWI. In the second example, we apply the workflow developed from the first example on the visco-elastic SEAM Phase II Foothill dataset. The data is created by the SEAM consortium to reproduce the imaging challenges in mountain regions (Regone et al., 2017). We aim at inverting the pseudo-2D dip-line visco-elastic survey, starting from relatively crude initial models. Besides the poor illumination of the pseudo-2D acquisition, the task is particularly difficult due to the strong near-surface heterogeneities and complex geometries.

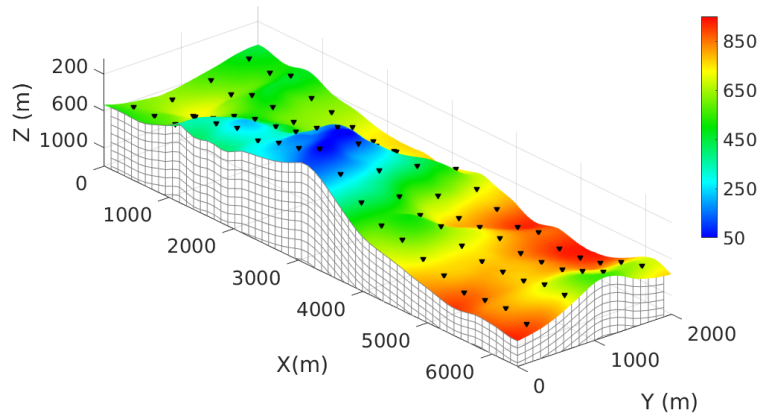


Figure 6.1: 3D acquisition design superimposed on the topography and the underlying mesh, where the colorbar illustrates the absolute depth Z_{abs} (m) from a pre-defined zero-depth: Sources positions are marked by red triangles $\Delta S_x = 320$ m, $\Delta S_y = 500$ m. For each source, 3-component receivers are deployed in the whole surface with $\Delta R = 12.5$ m.

6.1 Elastic 3D configuration

This section considers a 3D isotropic elastic subtarget of the SEAM Phase II Foothill model, in which both forward and inversion problem use SEM, but on different meshes. The observed data is generated by ourself, and the inversion is run with the true source function. The inversion starts from a relatively close initial model, in the regime of no cycle-skipping. Even in a such “easy” setting, appropriate data-windowing hierarchy and gradient preconditioning are required to reliably estimate the model parameters. The required computer resources for 3D elastic and visco-elastic FWI on this subset are also provided.

6.1.1 Model description & 3D acquisition design

The topography variation is significant for this target, with maximal vertical variation of 800 m as shown in Figure 6.1. The bedding plans are gently dipping in x -direction with folding structures, which can be visualized from the depth section in Figure 6.4(a). The model also has an unconformity at 2-3 km depth, which might be difficult to recover through FWI. Velocity attributes extracted at 20 m below the free-surface in Figure 6.6(a) show interesting features with complex geometry.

We use a 3D surface acquisition with 4 source-lines, each line including 20 sources. The source positions are indicated by red triangles in Figure 6.1, with inline and crossline source-spacing ΔS_x and ΔS_y taken at 300 m and 500 m, respectively. For each source, a grid of 3-component (3C) receivers is deployed on the whole surface, the distance between two adjacent receivers being 12.5 m. We use vertical point-force. The source-time function is a Ricker wavelet, centered at 3.5 Hz. The total recording time is equal to 6 sec.

The observed data is generated with a constant element-size mesh, globally satisfying the volume condition of around 6 points per shortest wavelength. The true compressional and shear wavespeeds for the observed data generation are presented in Figures 6.5(a). A sharp density model, as these velocity models, is considered for the observed data generation. As shown in the observed-data panel in Figure

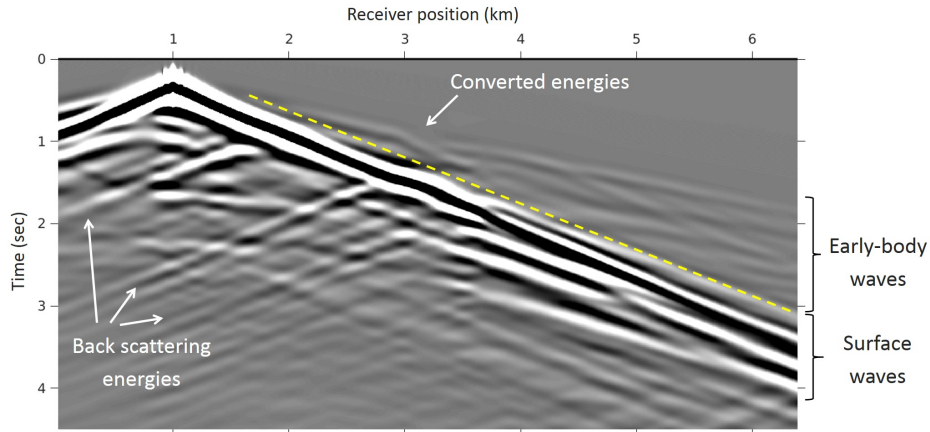


Figure 6.2: Example of an observed seismic shot: The wavefield is complex with highly energetic and dispersive surface waves. Significant converted and back-scattering energies occur when the waves hit the steep-slope topography. The separation between the early-body waves and surface waves can be done through a simple bottom mute.

6.2, the seismic wavefield is complex, including highly dispersive surface waves due to the rapidly varying topography. Significant back-scattering of body and surface waves as well as mode conversion at steep-slope surface positions can also be observed, for instance at $X = 3$ km (the associated V_s cross-section is presented in Figure 6.3(a)). In the seismogram in Figure 6.2, we distinguish between the early-body waves, which arrival before the surface waves, with the rest of the wavefield.

6.1.2 Inversion setup

The inversion problem is computed over variable element-size mesh, locally satisfying the volume condition of around 6 points per shortest wavelength. As discussed in Section (2.5.2), a high-order geometry representation, i.e. the P_4 representation in this case, should be used to describe rapid topography variations of the surface, especially for this variable element-size mesh. Since the sources and receivers are located at the free-surface, P_4 representation also helps to correctly position sources and receivers.

The initial V_p, V_s models that we use (Figures 6.5(b) and 6.6(b)) are smoothed versions of the true model. Similar smoothed model is used for density as the input of the inversion process. To investigate on different aspects of elastic FWI, such as gradient preconditioning and data-windowings hierarchy, we choose here a set of starting models compatible with the frequency-content of the data: the calculated data in the initial models is not cycle-skipped compared with the observed data, as shown in Figure 6.7(a). We invert simultaneously for V_p and V_s , and the density is kept unchanged, as a passive parameter. Each inversion sequence consists of 60 iterations of the l -BFGS optimization method. We do not apply other preconditioning or regularization than the Bessel gradient smoothing as detailed below.

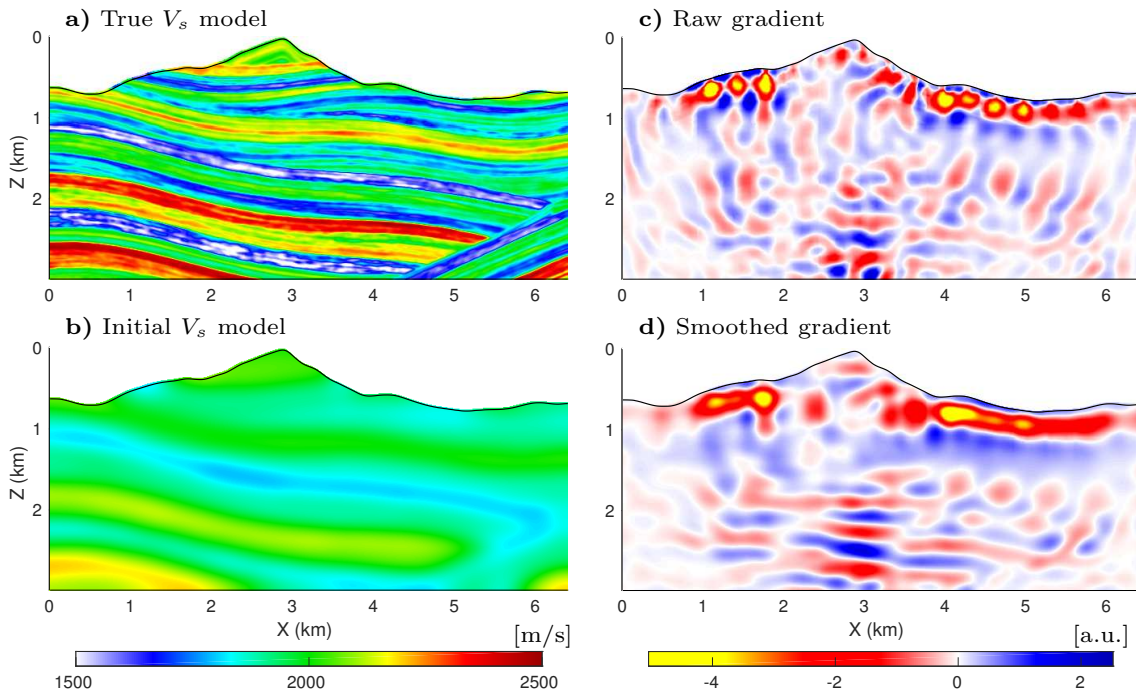


Figure 6.3: Gradient smoothing illustration underneath the source line at $Y = 750$ m: (a) True velocity model. (b) Initial model. (c) Original scaled gradient without any smoothing, showing significant acquisition footprint and artifacts due to sparse acquisition. (d) Smoothed gradient by Bessel filter where the parameters are described in Figure 6.4(b,c), which clearly show the dip & azimuth filtering effect.

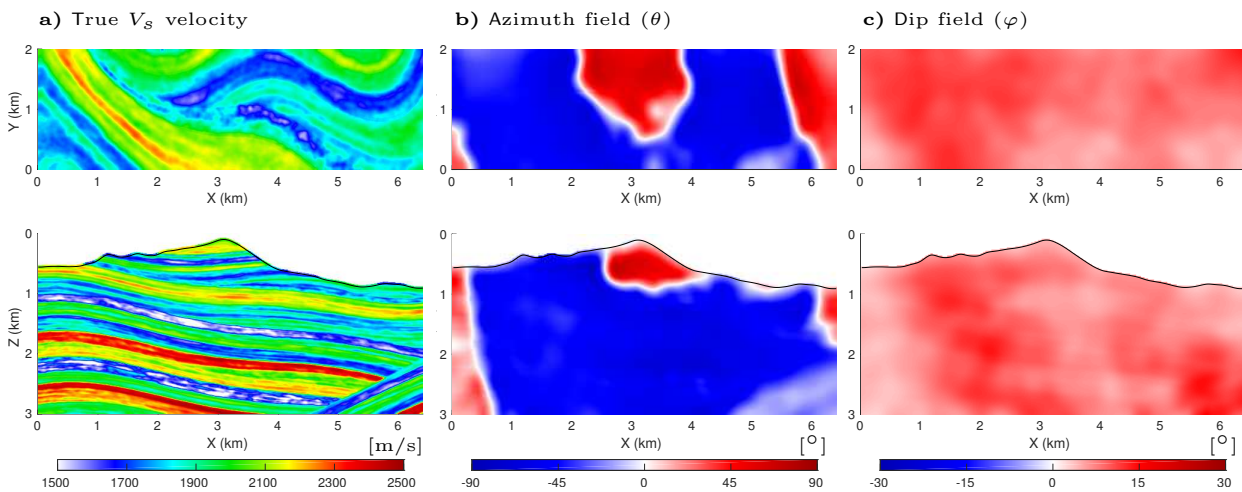


Figure 6.4: Top - Depth section (xy -view) at $z = 1$ km, bottom - Cross section (zx) at $Y = 100$ m: (a) True V_s velocity model with gentle dipping in zx -section and folding in xy -view. (b) Smooth azimuth field, which nicely follows the folding the structures in xy -view. (c) Smooth dip field, which is almost homogeneous.

6.1.2.1 Structure-oriented gradient preconditioning

The raw gradient computed on the initial model is presented in Figure 6.3(c), showing significant acquisition footprint at the near-surface. Unrealistic oscillations also occur at greater depths, due to the limited illumination coming from the sparsity of the sources. However, the targets have considerable dipping and folding structures, which can be characterized by smooth azimuth and dip fields as shown in Figures 6.4(b) and (c). Since the structures vary quicker in vertical direction than the horizontal directions, we design a highly anisotropic filter shape in which the vertical coherent length L_v is set to 25 m, much smaller than for other directions. Due to the presence of an unconformity, the horizontal coherent lengths L_u and L_w increases from 25 m near the discontinuity position to 100 m elsewhere. By doing so, the filter has an isotropic shape near the unconformity to avoid any smearing effect across the unconformity. The anisotropic coherent lengths design together with the 3D rotation result in the oblate-spheroid shape of the filter volume, with the major plan toward the dipping direction of the bedding plan. In practical applications, the filter design can be based on an interpretation of the migrated image computed from the initial models or any geological prior information about the medium such as well logs and geological description.

By considering these filter parameters in the structure-oriented Bessel preconditioning, we obtain the smoothed gradient 6.3(d): The artifacts due to the acquisition footprint and the poor illumination are effectively reduced, without degrading the deeper structures. The continuity of the features at greater depths is enhanced since the oscillations are weakened. We also obtain a correct orientation of the geological features. In this example, the smoothing process costs only 0.4% running time of one FWI iteration. The structure-oriented Bessel gradient precondition as described above is systematically applied into any inversion sequence in the next section.

6.1.2.2 Data-windowing hierarchy

Due to the rough topography, the seismic wavefield is quite complex with high-amplitude and dispersive surface waves and significant back-scattering waves as described in Figure 6.2. A two-step data-windowing hierarchy is considered to mitigate the dominance of the V_s parameter over V_p , due to the presence of surface waves. The models are first inverted with early-body waves, arriving before the surface waves. The observed models will then be used as starting models for the inversion of the entire wavefield. During the first step, a bottom mute is used to separate the early-body waves from the surface waves and underlying reflected and back-scattering waves, as illustrated in Figure 6.2.

6.1.3 Inversion results

The cross-sections of the results obtained after the first step are shown in Figure 6.5(c). The inversion successfully recovers main structures of the V_p model. The reconstruction of the V_s model is limited at 2 km depth due to the shallow penetration of the shear component. The models presented in Figures 6.5(c) and (d) are obtained after the second step, accounting for the whole dataset. Adding surface waves and other parts of the wavefield does not degrade the V_p estimation and improve significantly the V_s model. The continuity of the near-surface features is strengthened and the deeper structures are better resolved. The unconformity is well reconstructed in the V_s model at the 2 km depth (Figure 6.5(d), middle and right panels). This is in part due to the nonstationary design of the Bessel filter. When looking at the velocity attributes extracted at 20 m below the free-surface, (Figures 6.6(c) and

SYNTHETIC EXAMPLES

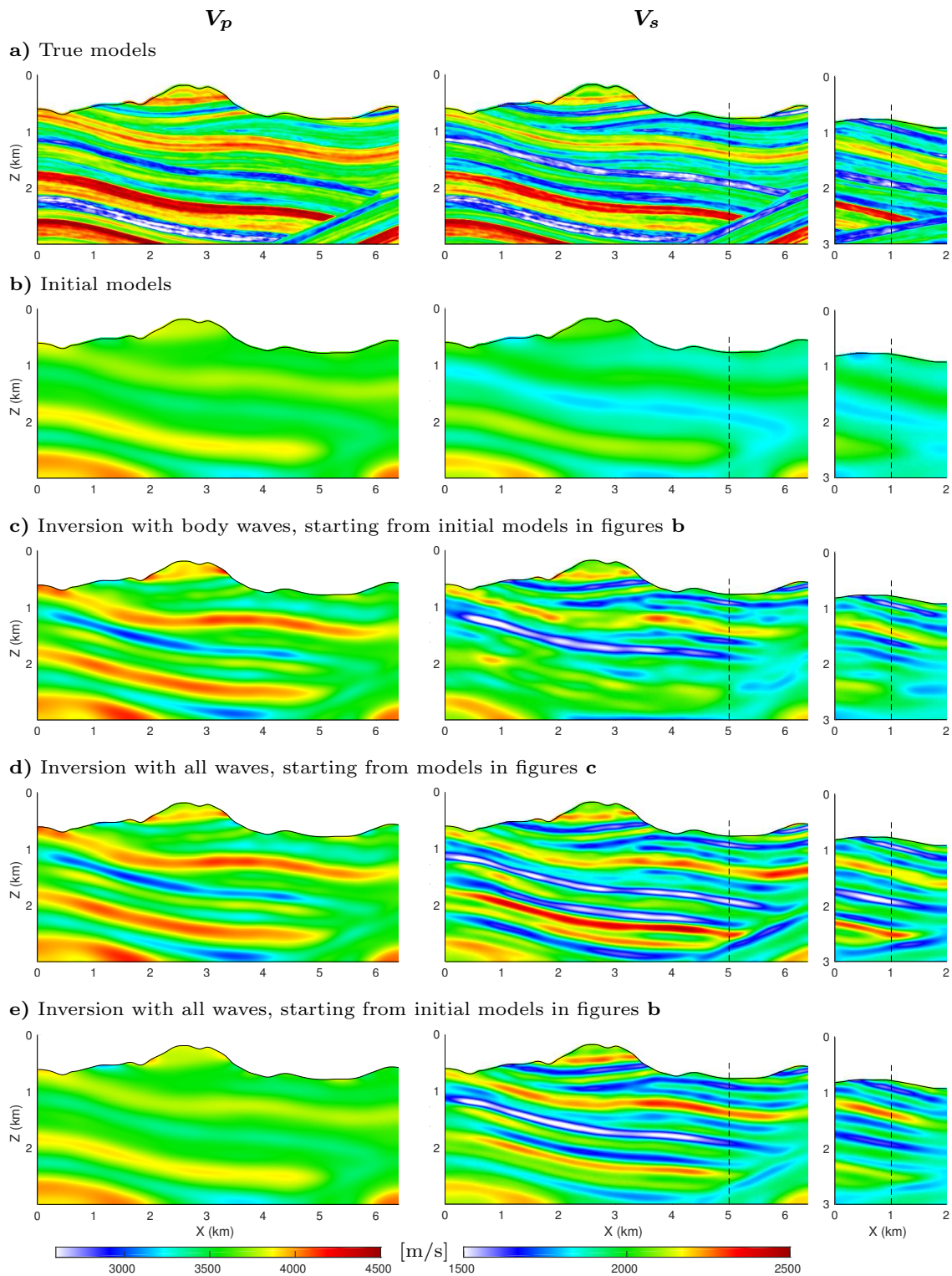


Figure 6.5: Left - V_p velocity cross-section at $Y = 1$ km, middle - V_s cross-section at $Y = 1$ km, right - V_s cross-section at $X = 5$ km: (a) True models. (b) Initial models. (c) Inversion results by only using the early-body waves, starting from the initial models. (d) Inversion results by using all the wavefield, starting from models in Figure (c), showing significant improvement for the V_s estimation. (e) Inversion results by using all the wavefield, starting from the initial models - no significant update in V_p estimation.

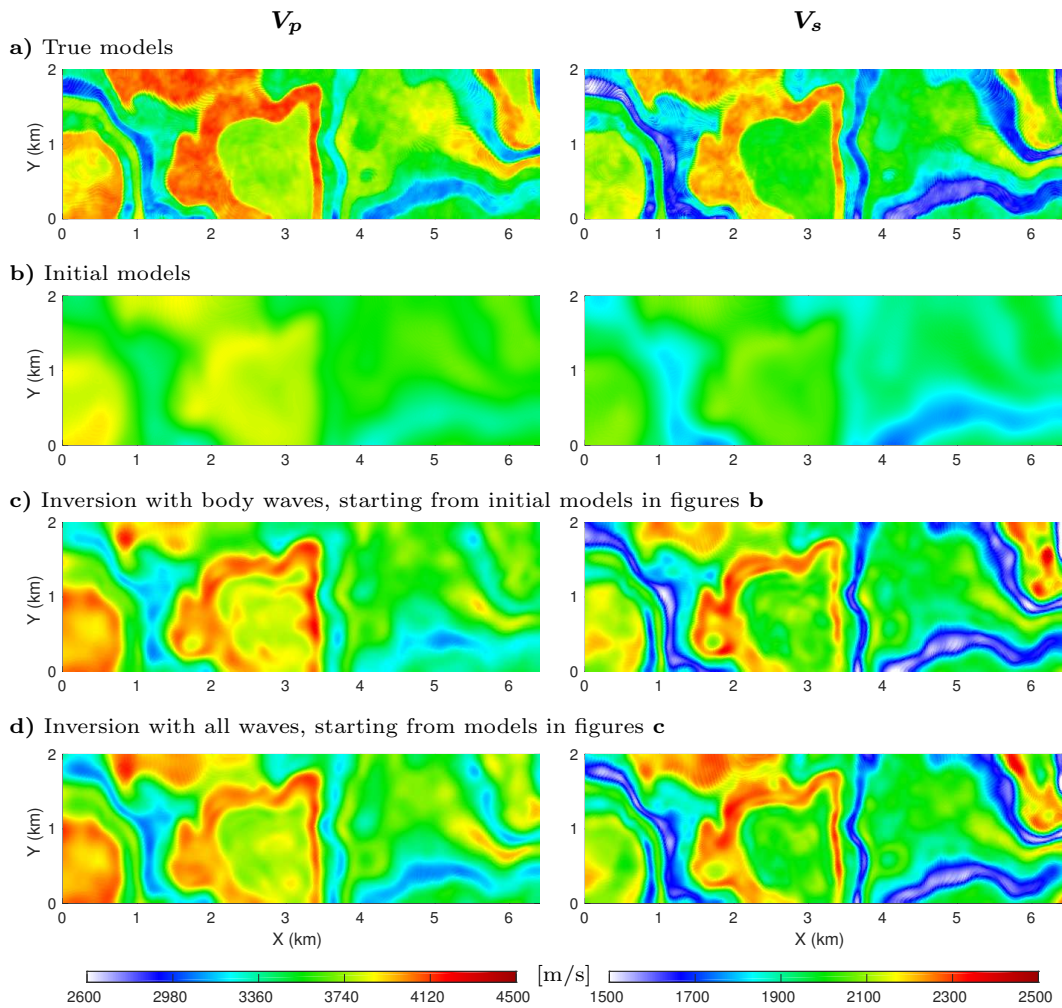


Figure 6.6: Velocity attributes extracted at the relative depth $\Delta Z = 20m$ from the free-surface. Left - V_p , right - V_s : (a) True models with interesting channels. (b) Initial models. (c) Inversion results by only using the early-body waves, starting from the initial models. (d) Inversion results by using all the wavefield, starting from models in Figure (c), showing that surface waves can bring additional information to improve inversion results at near-surface for both V_p and V_s .

(d) inverting for surface waves actually improves the thin-structure imaging and amplitude estimation for both V_p and V_s .

6.1.4 Discussions

6.1.4.1 Importance of the data-windowing hierarchy

Since the calculated data from the initial models is not cycle-skipped compared with the observed data, one can run the inversion with the entire wavefield, without any distinction between body and surface waves. The inverted results are shown in Figure 6.5(e), where the V_s is well-reconstructed. However, we observe insignificant updates of the V_p model. As the data is dominated by surface waves, the least-

squares misfit function is mainly sensitive to V_s . The data-oriented strategy we use in this 3D example is thus crucial: By focusing on early body waves before considering the whole data, our data-windowing hierarchy makes possible to better constrain the P-wave velocity.

Figure 6.7 shows the comparison of observed data with the calculated data computed from initial models and final inverted models (in Figures 6.5 and 6.6(d)). It should be remarked that compared with the observed data, the initial models produce simple surface waves estimation and weak amplitude of the first-arrival due to wrong velocities at the source location. After the two-step data-windowing hierarchy, the inversion successfully recovers details in the velocity models, resulting in an improved

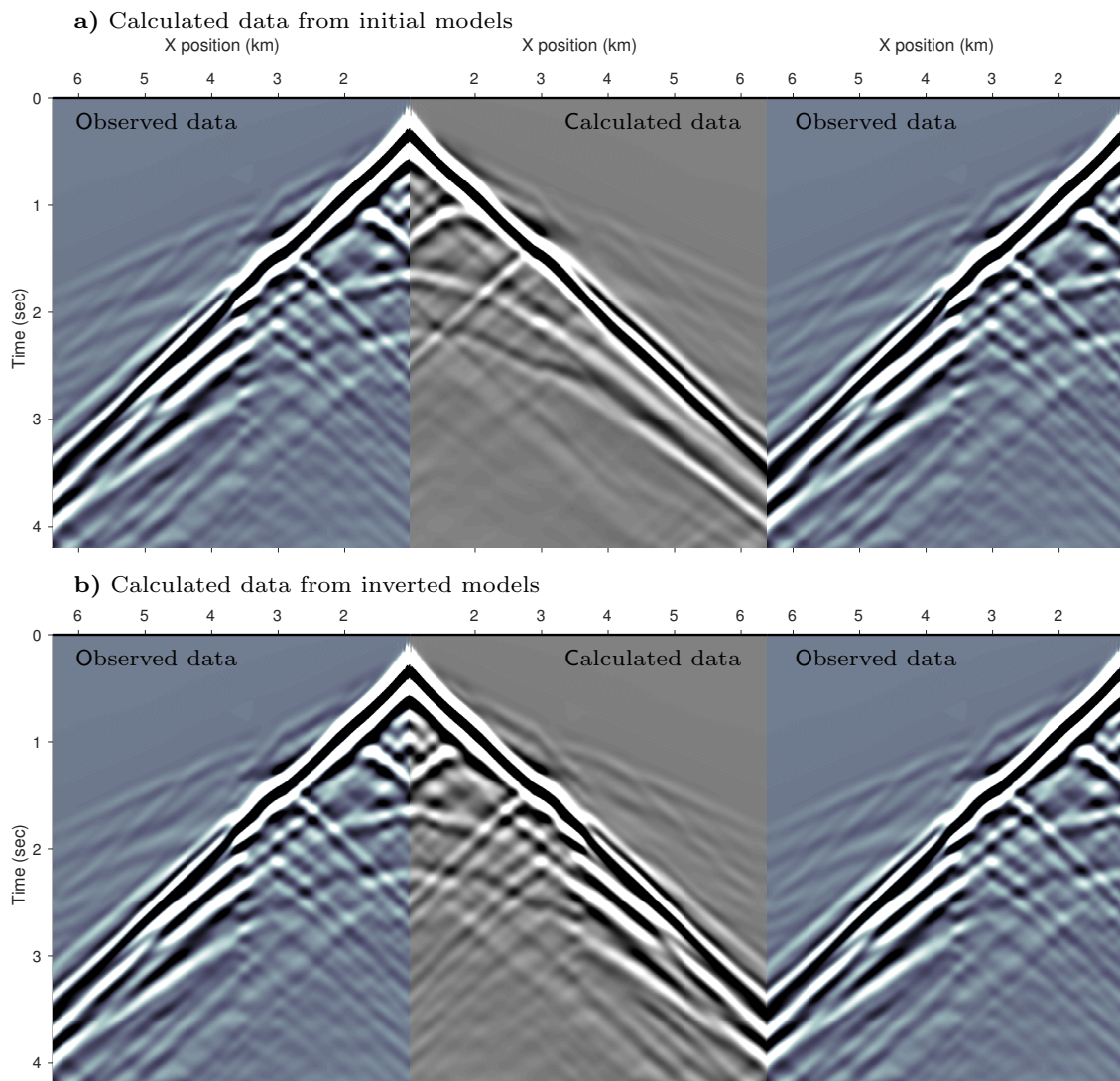


Figure 6.7: Data comparison at the shot-line $Y = 750$ m between observed data and calculated data: (a) d_{cal} estimated from initial models in Figure 6.6 and 6.5(b); (b) d_{cal} computed from inverted models in Figure 6.6 and 6.5(d). Final inverted models significantly improves the data-fit for the back-scattering, body and surface waves. The inversion also recovers the body-wave amplitude, related to the radiation efficiency of the source from model perturbation.

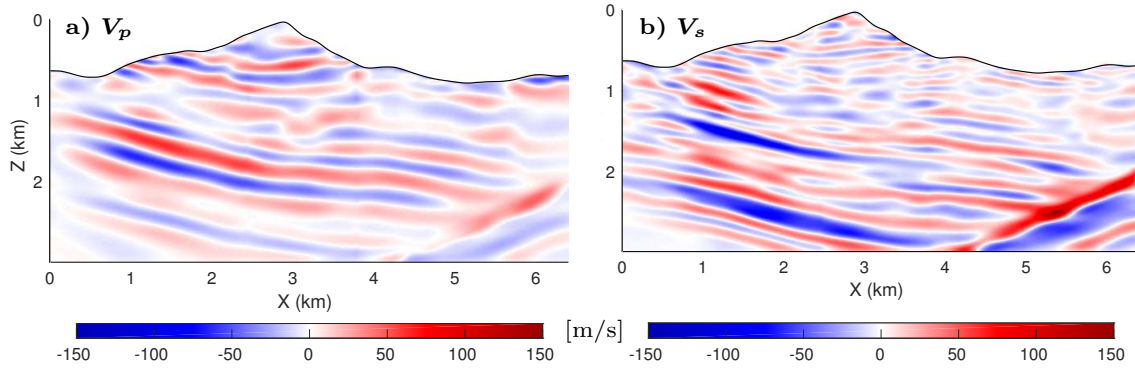


Figure 6.8: Difference between inverted results with structure-oriented preconditioning (Figure 6.5(d)) and the inversion results with stationary smoothing $L_v = 25$ m, $L_u = L_w = 100$ m, 0° dip and azimuth angles. (a) and (b) are V_p and V_s cross-sections at $Y = 750$ m. The rotational gradient preconditioning mainly affects the amplitude estimation.

agreement with the observed data for surface, body and back-scattering waves.

For complex real data applications, especially in foothill environment, this data separation between early body waves and surface waves might be difficult to achieve. In such a case, a model hierarchy could be used on top of this data-windowing strategy, where Bessel gradient preconditioning and/or model regularization could be used to monitor the wavenumber-content of the allowed model updates. At the first step, only low-wavenumber V_s updates would be allowed in order to enhance the V_p reconstruction.

6.1.4.2 Structure-oriented Bessel preconditioning enhances the model estimation

The importance of the gradient preconditioning is highlighted in Figure 6.3. If prior information about the structures such as the local 3D rotation is not available, a simple stationary anisotropic Bessel preconditioning could be enough to remove artifacts in the gradient (Trinh et al., 2017b), for example with coherent lengths $L_v = 25$ m, $L_u = L_w = 100$ m, and 0° dip and azimuth angles. It should be repeated that the anisotropic design relies on the fact that geological features vary quicker in the vertical direction than in horizontal directions.

Without the 3D rotation and variable coherent lengths in the gradient preconditioning design, the overall shape of the estimated models will remain similar as the results presented in Figure 6.5 and 6.6. These properties mainly affect the amplitude of the estimated parameters as highlighted in Figure 6.8, which shows the difference between inverted results with the nonstationary structure-oriented preconditioning and the inversion with the stationary filter, under the same inversion hierarchy and setting. The comparison shows considerable modification of the amplitude estimation for both V_p and V_s , especially following the dipping directions. In particular, notable difference can also be observed at the unconformity because the nonstationary filter is reduced to isotropic shape near this location. The model update is thus not smeared out across the unconformity, leading to a sharp interface as shown in the final V_s model in Figure 6.5(d). It should be noted that we only impose the relative position of the unconformity, and we simply reduce the filter effect at this location. By doing so, the filter does not artificially sharpen the interface but let FWI free to reconstruct it.

SYNTHETIC EXAMPLES

	Total memory per shot	1 st gradient estimation	FWI (60 it- erations)
3D Elastic FWI example ($3 \times 7 \times 2$ km) - 1600 cores: 3.98×10^6 dofs; 82.6×10^3 receivers per source; 10^4 time steps; Decimation ratio = 5; Mesh with all de- formed elements. Inverting for V_p, V_s parameters.	44 Gb	20 min	20.8 h
3D viscoelastic FWI test - 1600 cores: Quality factors $Q_p = Q_s = 50$; Incident fields re- construction: 80 checkpoints, decimation ratio = 5, energy tolerance = 5 % leading to recomputation ra- tio ≈ 3.2 .	70 Gb	1.36 h	84.8 h (es- timation)

Table 6.1: Top - Memory requirement and computational cost for the elastic FWI test on the 3D subset of the SEAM Phase II Foothills benchmark. Bottom - Estimation for 3D viscoelastic test with similar setting.

6.1.4.3 Computer memory and elapsed time

The 3D elastic inversion study has been performed on 1600 cores, where the computation for each shot being performed on 20 cores (i.e. 80 sources and only 20 sub-domains per shot). The incident fields are reconstructed from the decimated boundary wavefield with a decimation ratio equal to 5. The total memory requirement per shot is thus about 44 Gb. Each gradient computation in parallel is performed in 20 min, leading to almost 21 hours for one FWI sequence of 60 iterations, as shown in Table 6.1. It should be noted that storing the incident fields at each time-step would require approximately 889.5 Gb per shot without any compression.

On the same 3D models with similar numerical setting, we add homogeneous viscous properties with $Q_p = Q_s = 50$, which is the value of quality factors in the near-surface region of the SEAM II Foothills benchmark. To take benefit from whole memory while optimizing the elapsed time, 80 checkpoints are stored and a decimation ratio equal to 5 is used for wavefields saving at the boundaries. The energy tolerance for monitoring the reconstruction of incident fields is set to 5 % (Algorithm 5), leading to a recomputation ratio equal to 3.2. The wave propagation in anelastic medium is more expensive than for the elastic cases: one gradient costs 1.36 hours in parallel, providing an estimation of 85 hours for 60 iterations. In this setting, we obtain a factor 4 between elastic and visco-elastic inversion. It should be noted that storing the incident fields at each time-step would require 3.5 Tb per shot, due to the fact that C_{IJ} -gradient requires access to both strain and memory-variable incident fields (Eqn. 3.25).

It should be noted that Table 6.1 provides estimations for a relatively small example. In realistic applications, even with variable-element mesh design, we expect to consider 10^8 dof with longer recording time. In such applications, storing a large number of checkpoints or incident wavefields even at the Nyquist frequency in core-memory is infeasible. One might use the out-of-core memory at the expense of the I/O cost. Similar strategy could be considered for the storage of the wavefield in the boundaries. We rather focus on storing a small number of checkpoints and decimated boundary wavefield at Nyquist frequency in core-memory with limited I/O request but higher recomputation ratio of the incident fields.

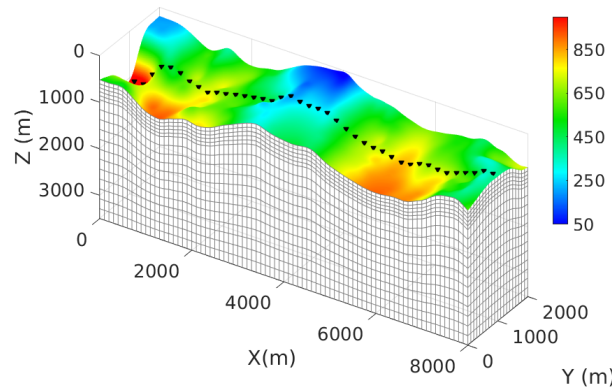


Figure 6.9: The 3D SEM mesh designed at 3 Hz maximal frequency: 100 m elements size at the near-surface and 200 m elements size at the deeper part. The topography map is put on top of the mesh, where the colorbar illustrates the absolute depth Z_{abs} (m) from a pre-defined zero-depth. The source positions of the 2D line are indicated by black triangles ($\Delta S = 200$ m).

6.2 SEAM II Foothill dataset

In this section, we apply our FWI frame and the two-step data-windowing hierarchy proposed in the previous section on the SEAM Phase II Foothill dataset. The data is computed by the SEAM consortium and we do not have information about the data generation. Starting from relatively crude initial models, we focus on the simultaneous reconstruction of compressional and shear waves velocities from the pseudo-2D dip-line visco-elastic survey. Besides the poor illumination of the pseudo-2D acquisition, the task is particularly difficult due to the strong near-surface heterogeneities and complex geometries. We also highlight the importance of the model constraint on the ratio V_p/V_s developed in Section (3.2.4) for reliable V_p and V_s reconstruction. We show that it is possible to reconstruct the shallower part from early-body waves before integrating surface waves for better constraining the construction at depth.

6.2.1 SEAM II Foothill Dip-line configuration

Our main target is the cross-section of the SEAM II Foothill model, underneath the pseudo-2D dip-line survey. The entire 2D line has 17 km length, but we limit the investigated area at 8 km in the x -direction. The associated compressional and shear velocities are presented in Figure 6.13(a). The topography variation in this area is significant, with maximal vertical variation of 900 m as shown in Figure 6.9. The structure is characterized by gentle dip in the x -direction with folding structures and an unconformity at 2-3 km depth as shown in Figure 6.13(a). The near-surface has alluvial deposits

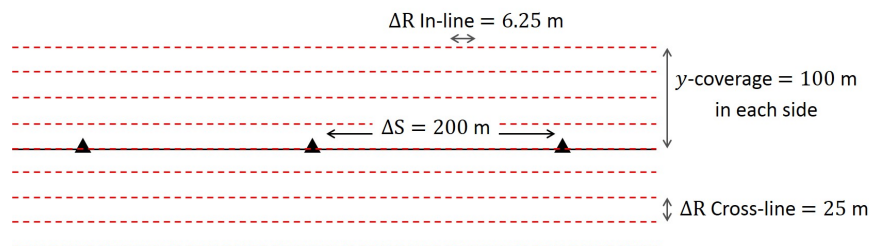
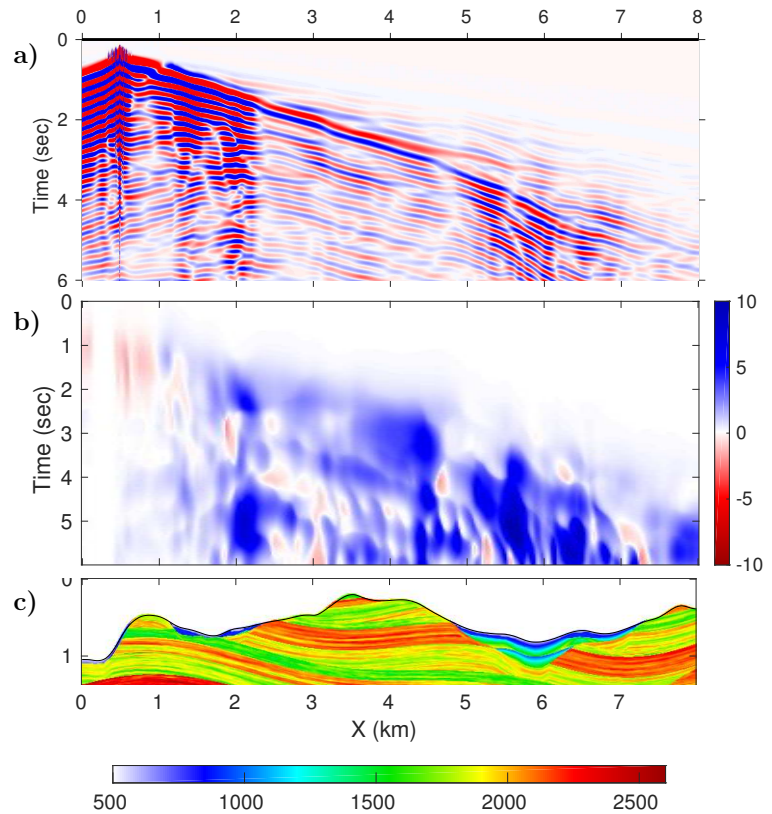


Figure 6.10: The pseudo-2D acquisition design with 1 source line (in black) and 9 receiver lines (in red dash lines). The source positions are indicated by black triangles ($\Delta S = 200$ m).

Figure 6.11: (a) The z -component of the observed visco-elastic data, low-pass filtered at 5 Hz. (b) Time envelope-misfit [%] between the simulated elastic and visco-elastic data, computed for the whole seismic gather: showing negligible difference. (c) The near-surface V_s model.



resulting from rapid erosion (Regone et al., 2017). These channels have low shear velocity, about 550 m/s at the surface, which are main obstacles for FWI. Moreover, the first 500 m below the surface has anelastic properties with $Q_p = Q_s = 50$.

The pseudo-2D line is acquired along the middle of the SEAM II Foothill model in the dip (x) direction. Due to limited computational resources, we only treat 40 shots: their positions are indicated by black triangles in Figure 6.9, with inline source-spacing being 200 m. There are 9 parallel receiver lines: the central line lies along the source line, with 4 parallel lines on either side, spaced at 25 m in the crossline (y) direction as illustrated in Figure 6.10. Receivers in the inline direction are located each 6.25 m (Oristaglio, 2016). All of the receivers are 3-components geophones and the sources are vertical point-force.

The acquired visco-elastic data is complex, including highly energetic and dispersive surface waves, as shown in Figure 6.11(a). By cross-comparing with the near-surface model in Figure 6.11(c), we observe significant back-scattering and converted waves at the steep-slope positions and at the strong-contrast interfaces between the alluvial deposits and the background medium, for example at the intervals (1 – 2 km) and (5 – 7 km) in x direction. The reverberations inside the low-velocity channels also dominate the wavefield amplitude, which are the most challenging parts for data fitting.

6.2.2 Inversion setup

The data considered for our inversion are the ones generated by the SEAM consortium, for which we do not have information about the numerical method, the mesh design and the source wavelet. The application is therefore not in the inverse-crime configuration.

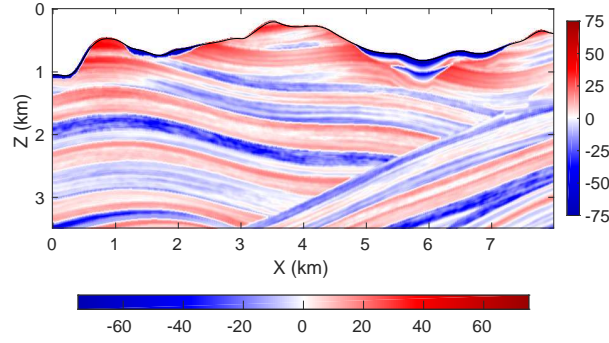


Figure 6.12: Relative model error of initial V_s model compared with the true model, computed from Equation (6.1).

6.2.2.1 Initial models & Mesh design

The initial V_p and V_s models are built in a realistic way, from 5 logs at positions 0, 2, 4, 6 and 8 km in x -direction. The models are obtained from a standard interpolation between upscaling logs data, as shown in Figure 6.13(b). We assume that no precise information about the alluvial deposits is available by removing all the V_s value less than 1000 m/s, leading to a wrong near-surface description in the starting velocity models. The inaccuracy of the initial model is also highlighted by Figure 6.12, which shows the relative model error between the initial and the true V_s model computed through

$$\text{Error}_{rel} = \frac{V_s^{true} - V_s^{initial}}{V_s^{true}}. \quad (6.1)$$

This estimation indicates a significant relative error at the near-surface, with more than 200 % inside the low velocity channels. The density model is calculated from V_p through a linear regression obtained from wells data as:

$$\rho \text{ [kg/m}^3\text{]} = 0.177 \times V_p \text{ [m/s]} + 1663 \quad (6.2)$$

To account for the 3D propagation effects, we consider a narrow 3D model with 2 km in y -direction. Since we limit at 8 km offset in x -direction, we do not expect any model update below 3.5 km in depth. The 3D initial models are built from these 2D sections assuming invariance along the y -direction.

Different SEM meshes are designed for different inversion frequency bands (i.e. 0-3 Hz and 0-5 Hz). Since we expect significant heterogeneities at the near-surface, the first 500 m in z -direction is filled with small elements. The element sizes are computed from the expected shear wavelength. An example of the mesh design for the maximal frequency at 3 Hz is shown in Figure 6.9: the near surface contains 100 m elements and the deeper part is filled with 200 m elements size. Each element is deformed by the 4th-order shape functions to describe the sharp topography (?).

6.2.2.2 3D elastic inversion setup: Data-windowing hierarchy

We apply our FWI frame on SEAM II Foothill dataset, low-pass filtered in the [0-5 Hz] frequency band. The time-frequency envelope misfit (Figure 6.11(b)) is used to compare the synthetic elastic and visco-elastic seismograms, showing negligible difference: less than 10% misfit (Kristeková et al., 2006). Based on this observation, we decide to use elastic engine to invert for the visco-elastic dataset, at least at low frequency. However, it is interesting to note that the main differences occur at the areas

associated with the trapped energies inside the alluvial deposits, between 1-2 km and 5-7 km as cross-compared with Figure 6.11(c).

We invert simultaneously for V_p and V_s . To mitigate the dominance of the V_s parameter over V_p , due to the presence of surface waves, we use a two-step data-windowing hierarchy: (1) The early-body waves, arriving before the surface waves, are first considered for the inversion. (2) The reconstructed models will then be used as starting models for the inversion with the whole wavefield. We simply use a bottom mute to separate the early-body waves with other parts of the wavefield.

Each step contains different FWI sequences as detailed in the next sections. Each sequence contains 30 iterations, where the density and the source wavelet are kept unchanged. After each inversion sequence, the source wavelet is re-estimated as described in Section (3.3) (Pratt, 1999), and the density is computed from the obtained V_p model according to the empirical relationship (6.2). We consider one single wavelet estimation for the entire dataset, under the assumption that the source signature is coherent from shot to shot.

We do not apply other preconditioning than the Bessel-based gradient smoothing (Trinh et al., 2017b). The coherent lengths design is adapted to the working frequency: $L_v = 100\text{m}$ and $L_u = L_w = 400\text{m}$ for [0-3 Hz]; $L_v = 30\text{m}$ and $L_u = L_w = 120\text{m}$ for [0-5 Hz]. We suppose that the smooth dip field is known and 0° strike as our main target is a 2D cross-section.

Inversion with early-body waves

When considering the early body-waves within the range [0-5 Hz], the calculated data computed from the initial models are cycle-skipped compared with the observed data at the far offsets. The cycle-skipping issue can be assessed by the *time-shift* attributes, as shown in Figure 6.14(a), which measures the time-lag of the maximum of the cross-correlation function between the calculated and the observed windowed seismic traces. From this calculation, we can also extract the *maximal cross-correlation* attribute to evaluate the trace-by-trace similarity between the calculated and the observed data, given in Figure 6.14(c). These attributes are computed for all traces acquired by the central receiver line. It should be noted that the synthetic data is considered as cycle-skipped when the measured time-shift with the observed data is more than 0.1 sec (half of the apparent wavelength in the observed data). A good data match is indicated by small time-shift (close to 0 sec) and high maximal correlation values (close to 1).

The time-shift and maximal cross-correlation attributes in Figure 6.14(a) and (c) are computed from the early-body waves, which highlight the cycle-skipping issue at the far-offset. Since the initial models have no information about the low-velocity alluvial deposits, significant dissimilarity is also visible for the near-offset data in these areas. This analysis is further confirmed by the data comparison between the observed data and the calculated data computed from initial models for the first shot, given in Figure 6.15A. Significant phase mismatch for early body waves at all offset ranges can be identified from this superimposed comparison.

We apply the standard low-to-high frequency strategy (Bunks et al., 1995) by considering two frequency bands: [0-3 Hz] and [0-5 Hz]. For each frequency band, we firstly inject half of the early-body waves package in time. The recovered models will then be used as starting models for the inversion will all the early-body waves. This strategy helps to enhance the V_p estimation since the early arrivals have more P-wave signature. The final models after the inversion with the early-body waves are shown in Figure 6.13(c).

Inversion with all the wavefield

Since the seismic wavefield is dominated by surface waves and back-scattering energies (Figure 6.11(a)), we would like to incorporate these information into the inversion. The comparison of the entire wavefield between the calculated and the observed data indicates that the models in Figure 6.13(c) are not good enough to avoid the cycle-skipping issue at 5 Hz (The time-shift attribute map is not shown here).

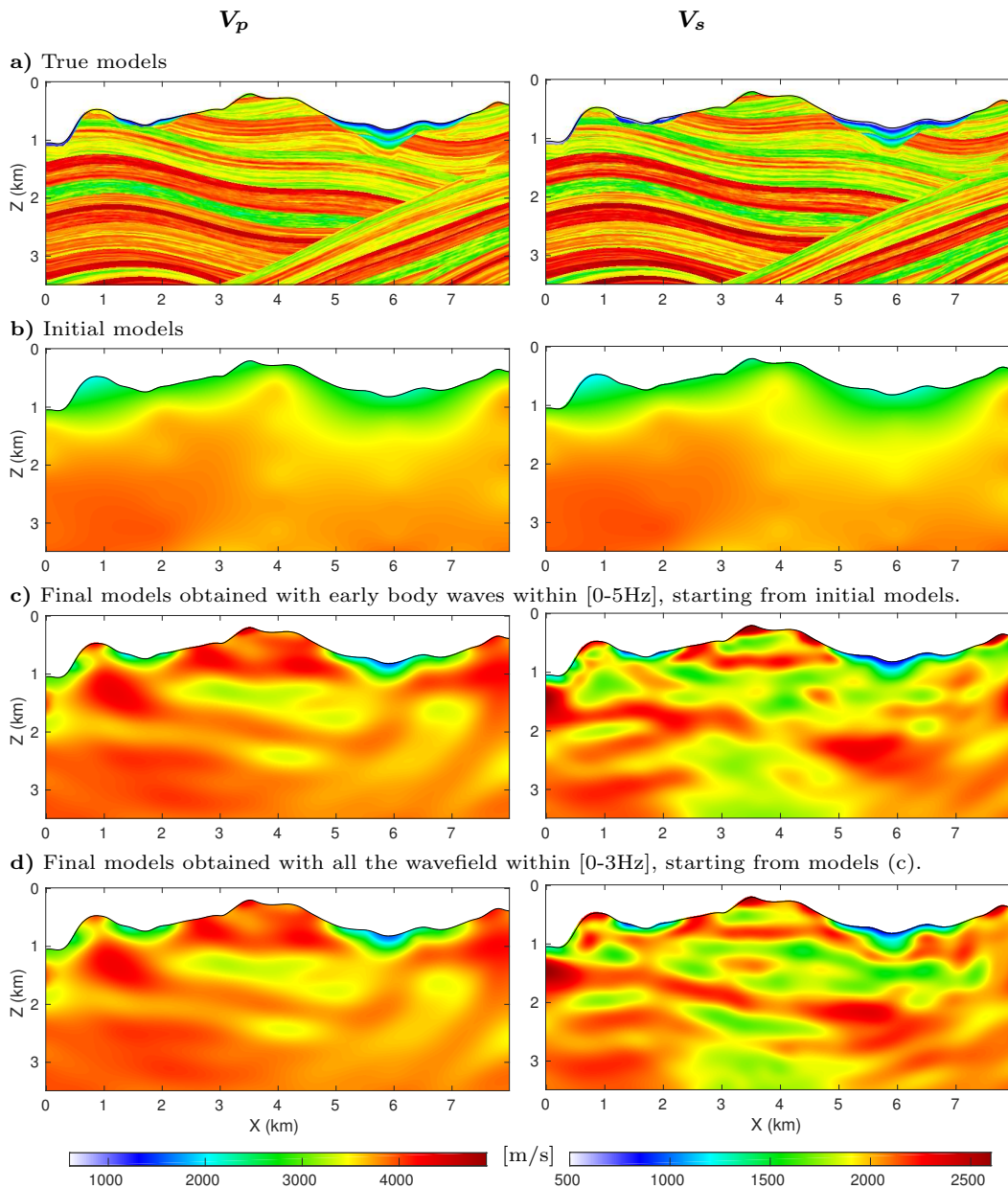


Figure 6.13: Left - V_p target cross-section, Right - V_s target cross-section: (a) True models. (b) Initial models. (c) Inversion results obtained with early-body waves within [0-5 Hz], showing interesting reconstruction for both V_p and V_s . (d) Inversion results by using all the wavefield within [0-3 Hz], showing improvement for the V_s estimation at depth.

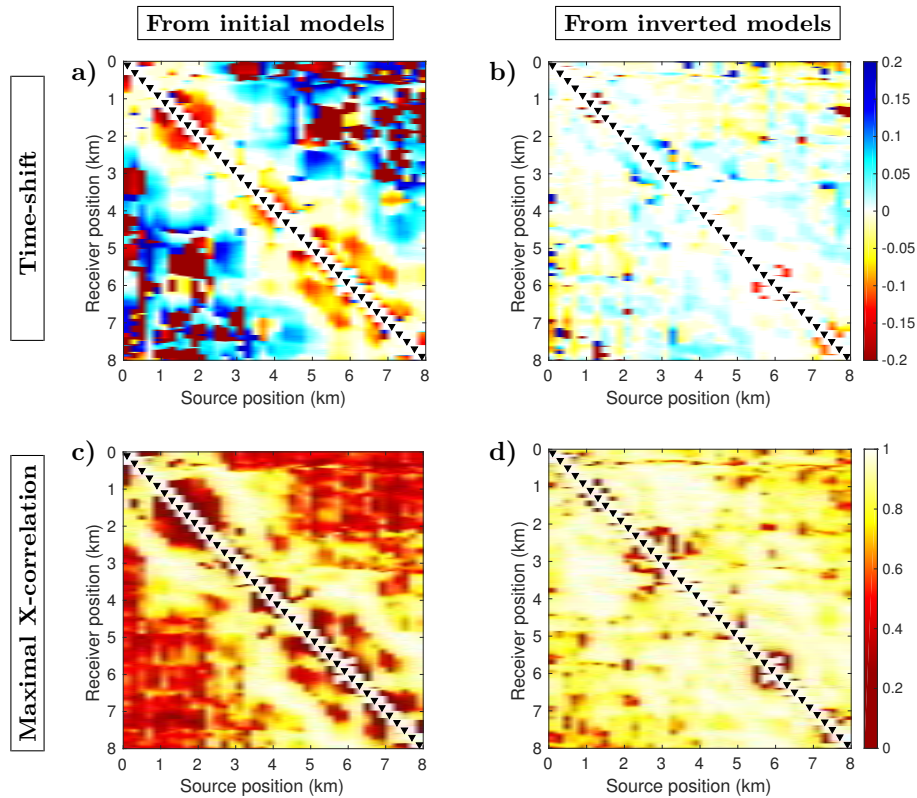


Figure 6.14: Time-shift [sec] and maximal cross-correlation maps between the calculated and the observed data. The attributes are computed for early-body waves within the range [0-5 Hz]: (a) and (c) are from initial models; (b) and (d) are from inverted models in Figure 6.13(c). Source positions are indicated by black triangles.

We then consider again the low-to-high frequency strategy by considering two frequency bands: [0-3 Hz] and [0-5 Hz].

Starting from the models 6.13(c), the V_p and V_s parameters obtained from the inversion with the entire wavefield within [0-3 Hz] are shown in Figure 6.13(d). We do not show the results obtained from the next frequency band [0-5 Hz] due to insignificant improvement.

6.2.3 Inversion results & Discussions

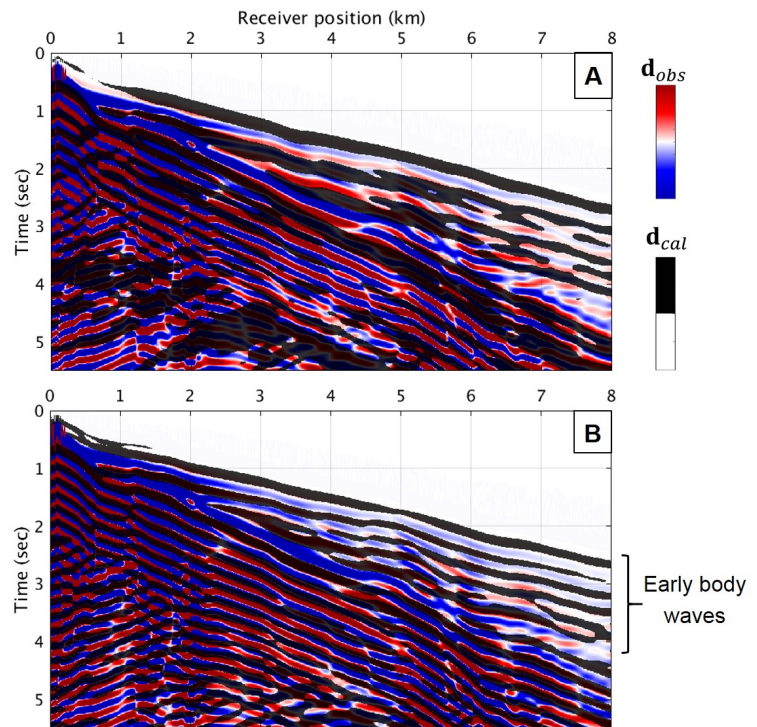
After the first step with low-to-high frequency strategy on early-body waves, the inversion successfully recovers main structures of the P- and S-velocity models (Figure 6.13(c)). The alluvial deposits and near-surface heterogeneities are well detected on both V_p and V_s . The time-shift and maximal cross-correlation comparisons between the observed and the calculated data on these models (Figure 6.14(b) and (d)) shows significant improvement compared with the measurement on the initial models (Figure 6.14(a) and (c)), both at near and far offsets. However, the V_s model contains discontinuous geological features due to shallow penetration of the shear component, the poor illumination of the pseudo-2D acquisition and the sparsity of the sources.

Considering the entire wavefield helps to improve the V_s model, without degrading the V_p estima-

Figure 6.15: Data comparison for the first shot between the observed data d_{obs} and calculated data d_{cal} : (A) d_{cal} estimated from initial models in Figure 6.13B, (B) d_{cal} estimated from inverted models in Figure 6.13D.

Observed data is displayed with red-and-blue color. Calculated data is displayed with black and transparent color.

Good data-fit is identified when the red loops are perfectly covered by the black loops.



tion, as shown in Figure 6.13(d). The continuity of the near-surface features are strengthened and the deeper structures at 2 km depth are better resolved. This is due to the fact that the inversion can exploit the shear body waves hidden by the surface waves and back-scattering energies. However, we stop the second inversion step at 3 Hz since the obtained near-surface models are not good enough to explain the surface energies at 5 Hz. It should be noted that the model resolution is limited by a half of the local wavelength. The inversion thus cannot reconstruct sharp contrasts in the obtained velocity models, leading to the limited capacity to fit the back-scattering and surface energies in the data. This is also one of the main reasons why we consider the low-to-high frequency strategy instead of other surface waves inversion strategies, such as the layer stripping technique proposed by Masoni et al. (2016).

Figure 6.15 shows the comparison of the 5Hz observed data with the calculated data computed from initial models and final inverted models (Figures 6.13(b) and (d), respectively). Compared with the observed data, the initial models produce rather incoherent body and surface waves with offset as shown in Figure 6.15(A). Significant mismatch between the observed and calculated data is also observed at all offset ranges and at all time-window levels. After the two-step data-windowing hierarchy, the early-body waves data fit has been improved as shown in Figure 6.15(B), especially for the first arrivals. However, one can detect some phase mismatch related to the velocity inaccuracy inside the low-velocity channels between 5 and 7 km for the early-body waves part. As described before, the 5 Hz surface waves are completely cycle-skipped, again confirming that this data part cannot be used by the inversion. It should be noted that the coherency of the calculated wavefield with offset is significantly improved for both body and surface waves. Even when the surface waves are cycle-skipped but the waveform looks rather similar, implying that a more cautious inversion strategy might offer the possibility to consider this surface-related energies.

In many elastic FWI applications for land problems, the least-squares misfit function is more sensitive to V_s , as shear events have strong imprint in the data. This is the case for our application where the models have significant geological heterogeneities such as low velocity alluvial deposits at the near-

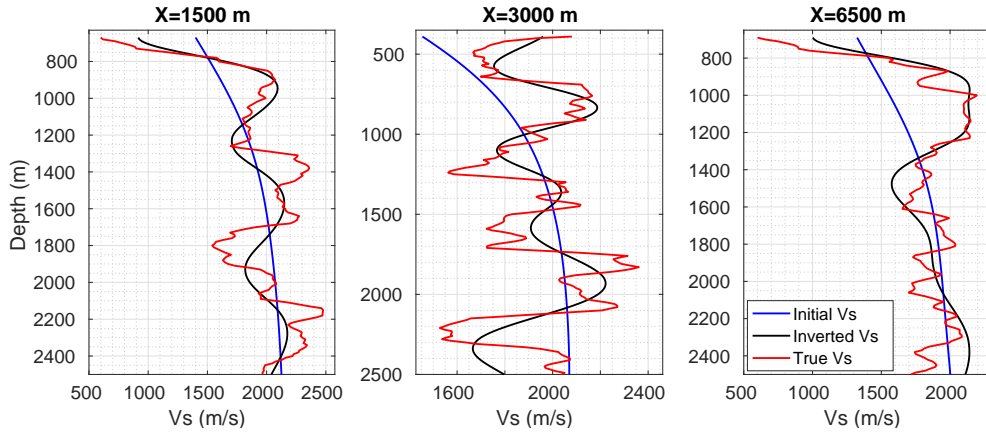


Figure 6.16: V_s logs comparisons at several position: $X = 1500$ m and 6500 m logs go through low-velocity channels, $X = 3000$ m is within the areas without alluvial deposits. Blue - initial model, black - inverted V_s and red - true model. The logs are only displayed from the free-surface level.

surface, leading to significant converted energies in the wavefield. The vertical-point-force source also generates a considerable amount of shear energy. The inversion process is then dominated by the V_s update in those low-velocity zones. This consequently produces unrealistic updates where $V_s > V_p$ or too small value of V_s . Once these issues occur the inversion process will stop after several iterations due to numerical instabilities. To overcome these issues, Section (3.2.4) introduces a non-linear constraint on the ratio V_p/V_s , which is applied simultaneously with the bound constraints on the range of V_p and V_s . For this application, we only consider homogeneous constraint on the ratio as

$$1.4 \leq V_p/V_s \leq 3.0. \quad (6.3)$$

These upper and lower bounds are simply designed based on well data. Such a simple constraint does not affect the FWI descent direction. At each iteration, it only removes unrealistic values of the estimated model parameters, leading to more reliable V_p and V_s estimation.

Since we suffer from the poor illumination of the pseudo-2D acquisition and sparse source distribution, i.e. only 40 sources are considered in this experiment, the estimated models thus contain significant artifacts. However, the inversion process in Figure 6.13 is still able to recover the main features in both V_p and V_s . The argument is further supported by some logs extractions at different locations as shown in Figure 6.16, in which the V_s amplitude for shallow targets are retrieved. The V_s estimation below 2500 m is unreliable, they are thus not shown here.

6.2.4 Perspectives

In this example, we treat the visco-elastic dataset by an elastic approach. Even when time-envelope misfit in Figure 6.11(b) shows insignificant difference between the elastic and visco-elastic data, the phase and amplitude distortion still appears in the visco-elastic dataset. Therefore, one source estimation per shot could be used to partially absorb the attenuation effect into the source wavelet. Moreover, the inversion hierarchy only consists of two frequency bands, with few time-windows per frequency band. With the current 5 Hz dataset, a more elaborated inversion strategy might help to improve the model estimations before injecting the surface energies. Higher frequency early-body waves could be

also considered to produce a more accurate near-surface models before considering the 5 Hz surface energies. In addition, considering more data, for instance more seismic shots, could also help to improve the estimated models quality.

6.3 Conclusion

The synthetic studies illustrate the complexity of the data and elastic multi-parameter FWI problem in complex land areas. We showed that the two-step data-windowing hierarchy makes possible for simultaneous estimation of V_p and V_s . If the surface waves and back-scattering energies are not cycle-skipped, we can take advantage from the hidden information below these events to get better V_p and V_s estimation, both at the near-surface and deeper parts. We treat body waves and surface energies independently, following low-to-high frequency strategies, to get a reliable model parameters estimation but also to exploit the maximum amount of information in the observed data. We also notice that body and surface waves require different frequency-hierarchy. In the second example, higher model resolution is required before the [0-5 Hz] surface and back-scattering energies can be integrated into the inversion.

We highlight the importance of additional prior information for providing reliable inverted model parameters. These information can be injected through the gradient preconditioning or model constraint. Experiences show that the inversion does not require sophisticated prior: A stationary gradient preconditioning and homogeneous model constraint are enough for stable inversion results. However, the accuracy of inverted models could be enhanced if some additional information is available, such as 3D rotation and relative faults position, which can be incorporated through the nonstationary Bessel-based preconditioning. More delicate prior, for example lithology changes, can be introduced through the constraint on the relationship between inverting parameters, such as the ratio V_p/V_s .

We promote the use of time-shift and maximal cross-correlation attributes for data quality control but also for inversion hierarchy design. These attributes provide a global indication of the data fit, which is directly associated with the quality of the inverted models.

Chapter 7

Real data application

Contents

7.1	Background information	149
7.1.1	Exploration context	149
7.1.2	Seismic data	151
7.1.3	Acquisition uncertainties	152
7.2	Inversion setup	153
7.2.1	Models building	155
7.2.2	Source estimation	158
7.2.3	Preconditioning and model constraint design	159
7.2.4	FWI workflow	161
7.3	Inversion results & Discussions	161
7.3.1	Model quality control through data fitting	162
7.3.2	Influence of acquisition mis-positioning on inversion results	163
7.4	Conclusion & Perspectives	165

In this chapter, we apply our FWI algorithm within the workflow developed on the SEAM Phase II Foothill benchmark on a real foothill dataset. This application is particularly challenging due to the sparse 2D acquisition both in sources and receivers, the presence of high levels of noise and complex underneath structure. It is difficult to progress quickly in this real data application since for instance, the process generates considerable amount of artifacts in the V_s estimation. The models presented in this chapter are thus preliminary results. To avoid any showing right issues, no structure images, migrated sections or absolute spatial locations are presented nor discussed.

7.1 Background information

7.1.1 Exploration context

We are interested in a complex foothill area with significant topography variation. The 3D volume of the entire area is shown in Figure 7.1, which is a SEM mesh generated from a given LIDAR topography

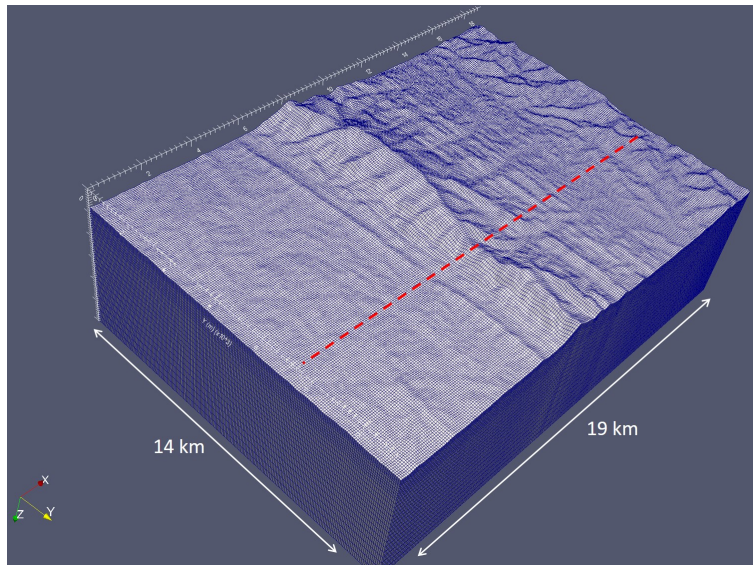


Figure 7.1: The 3D volume of the entire area. The considering 2D seismic line is indicated by the red dash line.

map. The topography has maximal vertical variation of 1000 m as shown in a subset of the LIDAR topography map in Figure 7.2. The sub-surface is characterized by verging thrusts with complex faulting and a series of anticlines and synclines, holding a large quantity of trapping hydrocarbons. Significant folding and highly dipping structures are resulted from the active tectonic activities in the past.

Some seismic surveys were acquired during the exploration phase for target imaging. Several pre- and post-stack images were generated, but they failed at imaging the top of the reservoir. This is due to significant amount of noise in the data, as shown in Figure 7.3, sparse acquisition and mainly the inaccuracy of the V_p velocity model. The rough topography variation and the excessive amount of noises make the standard velocity model building approach very difficult, for instance velocity picking and tomography techniques. The large geological uncertainties remain the most challenging factor for exploration. For this reason, multidisciplinary approaches and their integration have been considered for exploration but also during the later development phases. Different types of data, from seismic to non-seismic methods, have been acquired in this field. The strategic and business decisions mainly rely on the combination of different geophysical interpretations, driven by the appraisal wells. However, the geometry of the reservoir structure remains uncertain, especially the top of the structure.

In this chapter, we re-visit the velocity model building step with the 3D elastic FWI technique. As described in the previous chapter, our SEM46 tool is capable of correctly modeling the elastic wavefield, even under rapid topography variation. This modeling accuracy offers a possibility to improve the velocity models reconstruction through any data-fitting technique such as FWI. Among different seismic dataset, we consider a 2D seismic line which has the highest receiver density. The selected 2D line is indicated by the red line in Figure 7.1.

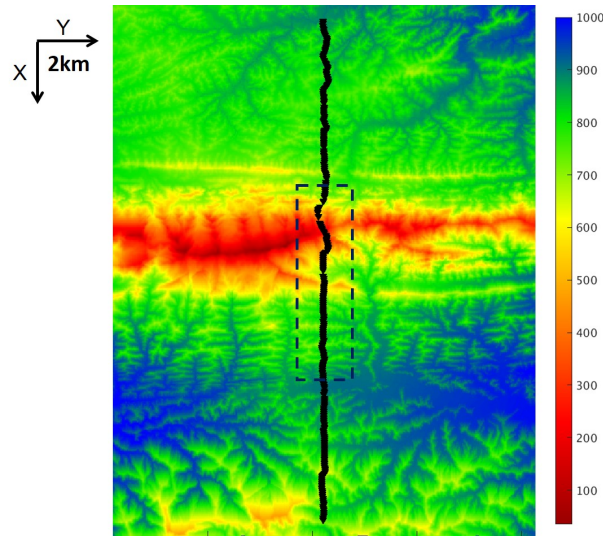


Figure 7.2: A subset of the LIDAR topography map where the colorbar illustrates the absolute depth from a pre-defined zero-depth. Sources positions are marked by black triangles, showing significant deviation in y direction. The selected sub-target is indicated by the rectangle.

7.1.2 Seismic data

The considered 2D line is located across the mountain trend with significant topography change. The entire seismic line is acquired with explosive sources, buried at 7 m depth. All receivers are single vertical (z) component, located at the free-surface. The sources and receivers follow the same line, as indicated by the black triangles in Figure 7.2. The acquisition has significant deviation in the y direction, at about 400 m, especially when crossing the mountain. This deviation implies that the 3D effect consideration is advisable for any imaging technique. The source and receiver spacings are not regular with rather sparse distribution: the average distance between adjacent receivers is approximately 37 m, and about 40 m between adjacent sources.

The raw seismic shot in Figure 7.3 shows clear topography footprint in the first arrivals. Due to the considerable amount of vertical stripping and coherent noises, the later arrivals are hardly seen. The shot gather also contains highly energetic and dispersive surface waves. At the display frequency range (between 0 and about 20 Hz), the surface waves patterns are completely broken due to the interaction with the rough free surface. Also, the surface waves is aliased at this frequency range due to rather sparse receiver distribution. We also notice remarkable converted energies, indicated by red arrows in Figure 7.3, when the waves hit the steep slopes at the near surface. The amplitude variation of early arrivals, i.e. dimming and brightening events, indicates possible lateral heterogeneities at the near-surface. The complexity in the seismic wavefield further confirms the challenging structure that we are dealing with.

The use of explosive seismic source produces un-controlled source signature but it also generates the low frequency content. The data spectrum showing in Figure 7.4 indicates the appearance of low frequency content, below 5 Hz, which is beneficial for FWI.

7.1.3 Acquisition uncertainties

As described in Chapter (3), the velocity models building relies on the comparison between the calculated and the real data. The calculated wavefield is computed on a finite element mesh, generated from a given topography map. The simulation thus requires sources and receivers staying within the SEM mesh. This condition might be tricky when we have an inaccurate topography measure and the sources and receivers positions are close to the free-surface. In practical implementation, to avoid putting sources and receivers in the air, we implement in SEM46 the relative depth from the free-surface instead of the absolute depth from a pre-defined datum.

In majority of the onshore fields, the 3D topography map of the entire field is given by LIDAR measurement, which has lower resolution compared with the sources and receivers positions recorded inside the SEG-Y header. We cross-check the acquisition information by comparing the absolute depth $z_{absolute}$ recorded inside the SEG-Y header with the computed depth $z_{computed}$ from the LIDAR measurement. For each source or receiver position, the $z_{computed}$ is the sum of the surface elevation z_{LIDAR} at the considering position with the relative depth $z_{relative}$ inside the SEG-Y header as

$$z_{computed} = z_{relative} + z_{LIDAR}. \quad (7.1)$$

The computed depth from LIDAR and the absolute depth from SEG-Y header for all sources are superimposed in Figure 7.5, which shows significant mismatch at the top of the mountains. The histograms of these difference ($z_{absolute} - z_{computed}$) for all sources and receivers are shown in Figure 7.5. There is a systematic error at (-1.64 m) mean and 8.42 m standard deviation (std) for all sources and (-1.6 m) mean and 8.5 m standard deviation for all receivers. There is a non-negligible number of sources and receivers with more than 20 m mismatch, especially at the top of the mountain with more than 40 m shift as indicated in Figure 7.5. This mis-positioning in depth might affect the inversion results since FWI heavily relies on kinematic data fitting. For instance, FWI might tend to put low velocity layers at the near-surface to absorb this error. The influence of the acquisition uncertainty on FWI process will be further investigated in the discussion Section (7.3.2).

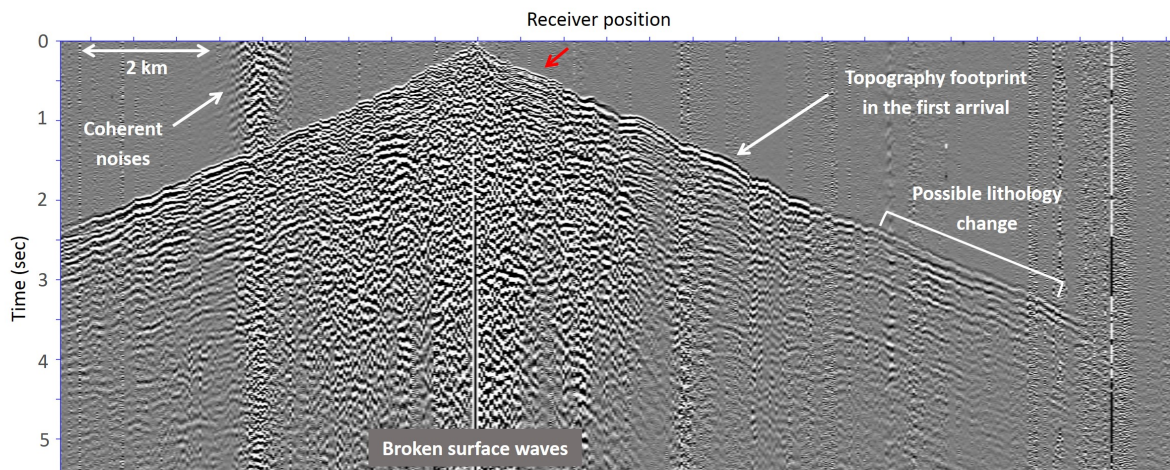


Figure 7.3: Example of a raw seismic shot: The data contains a significant amount of noises. The elastic effect is pronounced with energetic and dispersive surface waves and converted energies (indicated by red arrows). The amplitude variation of early arrivals indicates possible lateral heterogeneities at the near-surface.

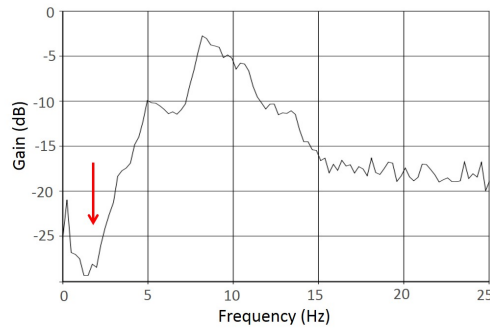


Figure 7.4: Example of the amplitude spectrum of a shot gather. The cut-off frequency is indicated by the red arrow

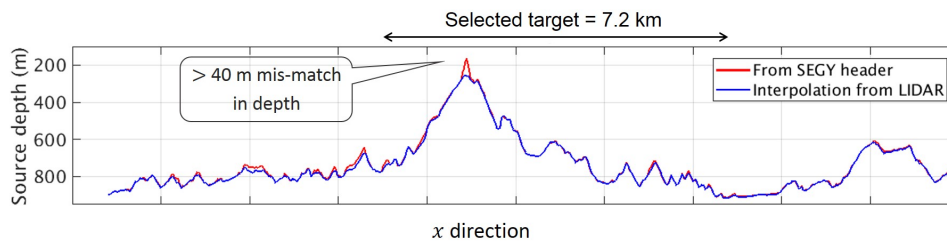


Figure 7.5: Comparison between the absolute source depth extracted from SEGY header $z_{absolute}$ and the computed source depth $z_{computed}$ from LIDAR measurement following Equation (7.1).

7.2 Inversion setup

Since our SEM46 tool can perform true-amplitude and waveform elastic waves propagation, we do not apply any sophisticated processing sequences such as static correction or surface waves removal. We only remove the low frequency noises below 2 Hz by a high-pass filter. The data within the range [2-4 Hz] is considered for the first FWI frequency band, which provides a reasonable signal to noise ratio.

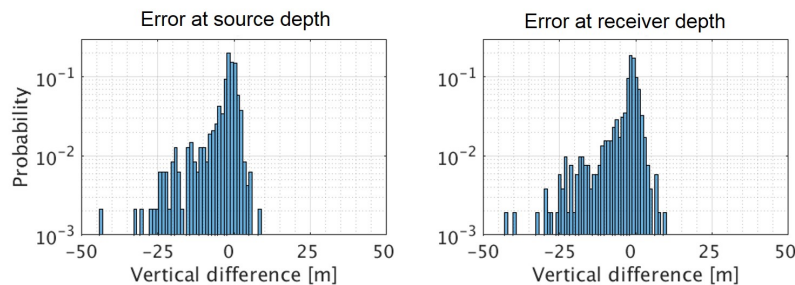


Figure 7.6: Systematic error in sources and receivers depth. In the source side, the error in depth is at median -1.64 m with standard deviation 8.42 m. In the receiver side, the error in depth is at 1.60 m median with standard deviation at 8.5 m.

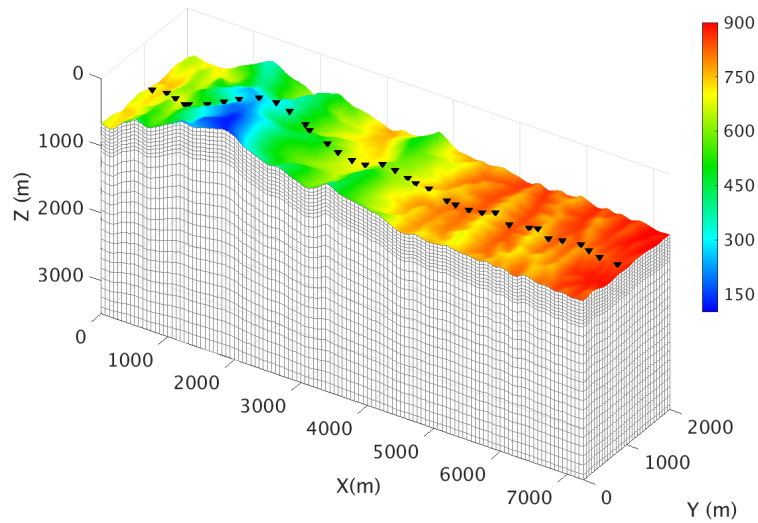


Figure 7.7: The 3D SEM mesh designed at 4 Hz maximal frequency: 62.5 m elements size at the near surface and 175 m elements size at the deeper part. The topography map is put on top of the mesh, where the colorbar illustrates the absolute depth Z_{abs} (m) from a pre-defined datum. The source positions of the 2D line are indicated by black triangles ($\Delta S = 200$ m).

The initial V_p model is provided by the data owner, which is generated from the standard picking workflow, calibrated by the walk-away vertical seismic profile. The initial model quality control shows acceptable starting data fit for the selected frequency range. We do not have any starting V_s or density models, however, several wells with compressional, shear sonic logs and density measurement are available. These wells are sparsely distributed and they do not follow the considered seismic line. The distance from the 2D line to these wells ranges from 1.7 km to 10.5 km, the spatial interpolation thus cannot be considered for initial V_s and density models building. Also, they cannot be used for quality control of the inversion results. However, they still provide useful rock physics properties. We generate the initial V_s and density from V_p through some empirical relationships computed from logs data. Other elements of the inversion process such as mesh generation and model constraint design also rely on these logs information.

Due to limited computational resources, we only treat a sub-set of 7.2 km of the given seismic line. We select the area with the most topography variation, indicated by the rectangular box in Figure 7.2. Figure 7.5 shows that the selected area contains some sources with significant depth uncertainty. For this target, only 36 sources are considered with distance between two adjacent sources being 200 m as illustrated in Figure 7.7. Since the data quality varies from shot to shot, the shot selection has to be done manually to select the 36 best shots within the considered target. Also, an amplitude calibration between shots is performed based on the root-mean-square measure to bring all the shots amplitude to the same range: one scaling constant is applied per shot. As described in Section (7.1), the 3D elastic propagation should be considered due to significant acquisition deviation, topography issue and complex underneath structure.

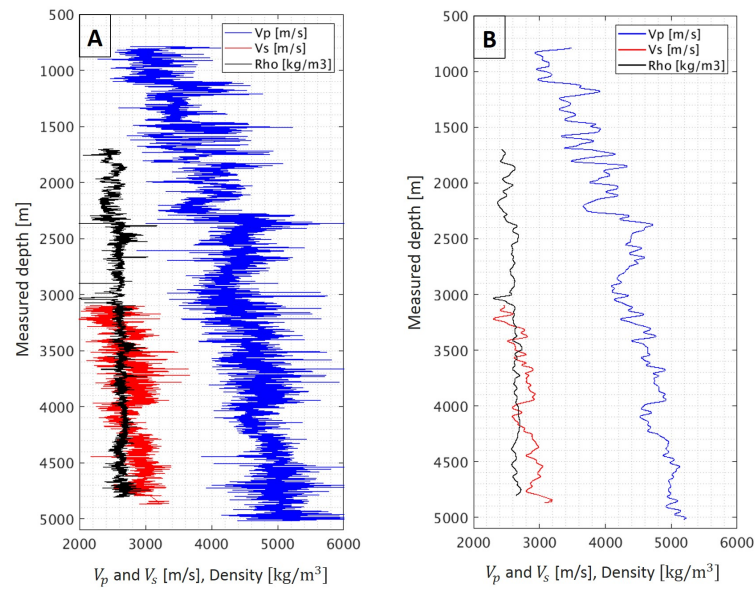


Figure 7.8: Example of logs upscaling for the well W4: A - Input logs data; B - Upscaled logs data.

7.2.1 Models building

There are five calibrated wells, named as W1 to W5, with sonic logs and density measurement which are available as the input information. Since the logs data is at much higher frequency than the seismic frequency range, they are upscaled by a Gaussian window with standard deviation $\sigma = 12.5$ m, about half of the highest expected resolution. Figure 7.8 shows an example of the logs data before and after the upscaling. Some additional corrections need to be applied to the logs such as removing the unmeasured intervals and unexpected values. For each well, different logs (V_p , V_s , ρ) are measured at different depth intervals, so the next cross-plots in this chapter are only produced from the common intervals between different logs.

7.2.1.1 Mesh creation

To account for the 3D propagation effect, we consider a narrow 3D model with 2 km in y -direction. Since the selected sub-set has 7.2 km in x -direction, the investigation depth is limited to 3.5 km. The mesh design follows the same strategy as for the synthetic benchmark in Section (6.2), in which the first 500 m in z -direction is filled with small elements. To prepare for any near-surface low-velocity anomalies, we allow the minimum shear wave speed V_s to reach 400 m/s, leading to the 62.5 m elements size at the shallow part.

The well W2 ranging from 1700 m to 2800 m depth detects possible gas-sand at the reservoir level with low V_s , about 1200 m/s. This value is thus accepted as the smallest possible V_s value for the deeper part, leading to the 175 m elements size. The mesh design for the maximal frequency at 4 Hz is shown in Figure 7.7. Each element is deformed by the 4th-order shape functions to describe the sharp topography.

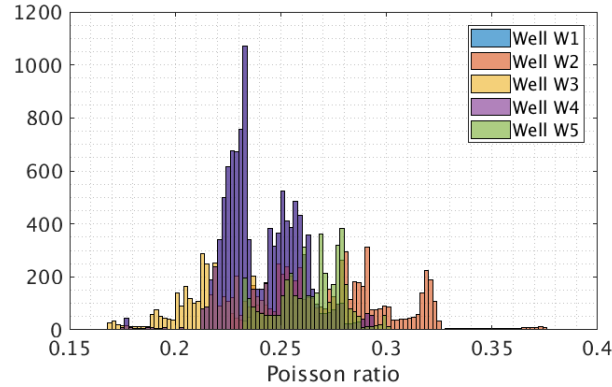


Figure 7.9: Histogram of the Poisson ratio computed from upscaling V_p and V_s wells following Equation (7.2), indicating the existence of various rock-types.

7.2.1.2 Initial V_s and density ρ models building from the given V_p

We aim at building some empirical relationships from upscaled logs data to transform the given compressional velocity V_p in Figure 7.11A to meaningful initial shear velocity V_s and density ρ models. The main challenge is that the logs are only measured at the reservoir level. The shallowest density measurement is at 1700 m measured depth as shown in the well W4 in Figure 7.8, and for V_s is also at 1700 m depth, for the well W2 that we do not show here. Any regression at the reservoir level then needs to be propagated up to the surface level, which is associated with low V_p values, at about 1000 m/s. From the upscaling V_p and V_s wells, the Poisson ratio can be computed as

$$\text{Poisson ratio} = \frac{1}{2} \frac{(V_p/V_s)^2 - 2}{(V_p/V_s)^2 - 1}. \quad (7.2)$$

The histogram of the computed Poisson ratio is shown in Figure 7.9, which is a multi-pole distribution, implying the existence of various rock-types. Therefore, an additional challenge is to build meaningful empirical relationships for the ensemble of different lithology.

Initial density model

Among five wells, only W3 and W4 have the density measurement. Following the Gardner's relationship (Gardner et al., 1974), the density can be computed from V_p as

$$\rho \text{ [kg/m}^3\text{]} = 326 \times V_p^{0.25} \text{ [m/s]}. \quad (7.3)$$

When extrapolating this regression up to the surface level with estimated V_p at 1000 m/s, we obtain rather small density value as shown in Figure 7.10A. It should be noted that, in reservoir characterization, the Gardner's relationship is considered for a short logs interval, normally associated with a specific rock type, which is not the case in our application. When considering a simple linear fitting, we obtain the relationship

$$\rho \text{ [kg/m}^3\text{]} = 0.137 \times V_p \text{ [m/s]} + 1977, \quad (7.4)$$

leading to more reliable density estimation for small V_p value. Moreover, experiences show that FWI does not require a very accurate density as a passive parameter. Therefore, we accept the linear regression (7.4), resulting in the initial density model presented in Figure 7.11B.

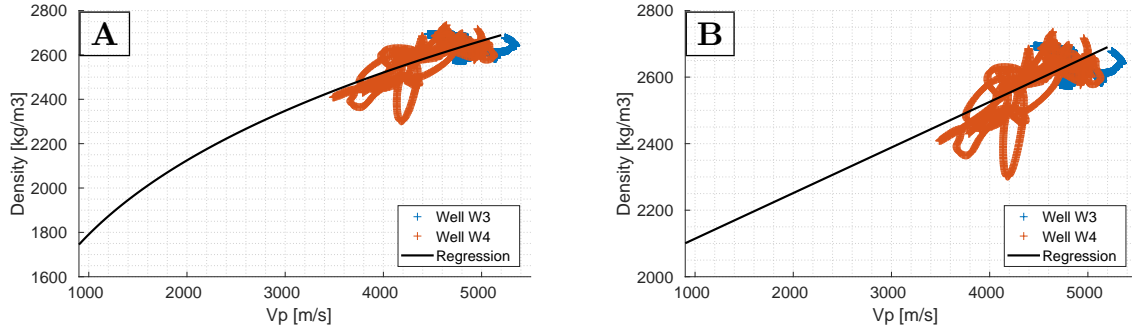


Figure 7.10: Empirical relationship between density ρ and V_p : A - Gardner's relationship in Equation (7.3); B - Linear relationship as mentioned in Equation (7.4).

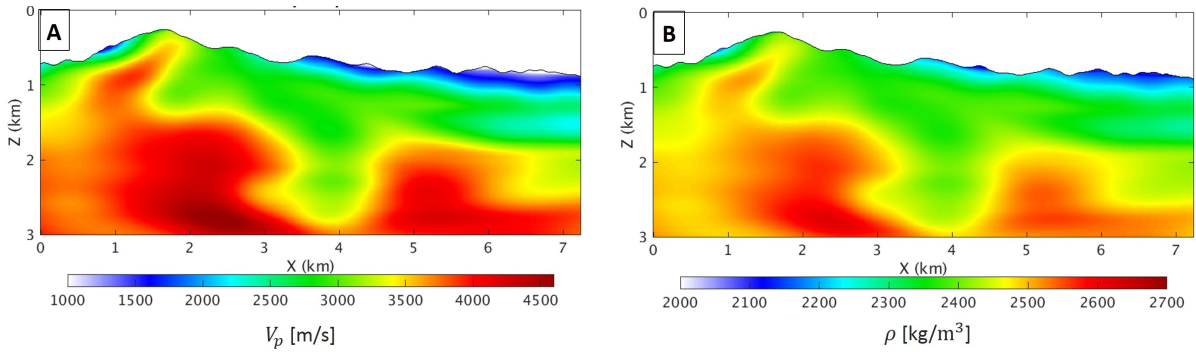


Figure 7.11: A - Initial V_p model provided by the data owner. B - Initial density model computed from the initial V_p in Figure A through the linear regression in Equation (7.4).

Initial shear wave velocity model

Similar as the density initial model building, we aim at estimating the initial V_s from the given V_p model. The linear relationship between V_s and V_p is widely considered in rock physics, which provides a good fit at the reservoir level. However, the extrapolation for small V_p values at 1000 m/s gives too small V_s values, close to zero, as shown in Figure 7.12A. The existence of various lithology, indicated by the multi-poles Poisson ratio distribution in Figure 7.9, also confirms that a simple linear fitting is not enough for our purpose. To avoid unexpected small V_s values at the near-surface, the V_s model is computed from V_p through the relationship

$$V_s = \frac{V_p}{-1e^{-4}V_p + 2.22}, \quad (7.5)$$

which is equivalent to a linear fitting between the ratio V_p/V_s and V_p at the reservoir level. The extrapolation of this regression provides the estimation $V_s \approx 500$ m/s for small V_p values as shown in Figure 7.12B. The obtained initial V_s model is presented in Figure 7.17A.

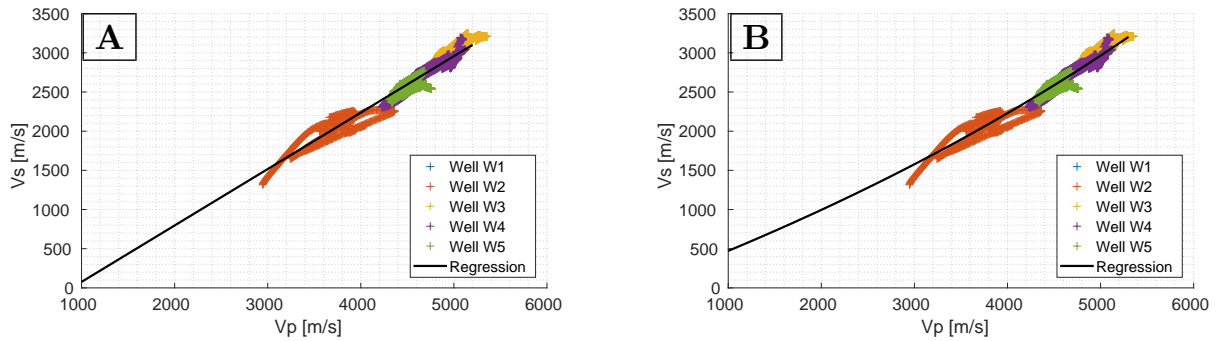


Figure 7.12: Empirical relationship between V_s and V_p : A - Linear fitting; B - Regression provided in Equation (7.5).

7.2.2 Source estimation

Since the data is acquired with explosive sources, each shot has its own source time signature, both in waveform and starting time. Therefore, one source wavelet is estimated per shot, following the strategy proposed by Pratt (1999) as described in Section (3.3). The source estimation is performed over a time-windowed dataset as shown in Figure 7.13. The very-near offset traces are removed due to their high amplitude and the unphysical wavefield simulation near the source position. A time window, based on the first break picking, is applied to the data to remove the noises before the first arrival and the data part which is not considered by the inversion.

The source estimation is a deconvolution-like process, it thus requires a sufficient quantity of observed data for a stable wavelet estimation. However, each shot gather contains less than 200 traces due to the sparse receiver distribution. We thus do not apply any other stronger muting than what has been described in Figure 7.13. The short to medium offsets of surface waves are also included, which actually provide a better source amplitude and waveform compared with the source estimation from only far offsets body waves.

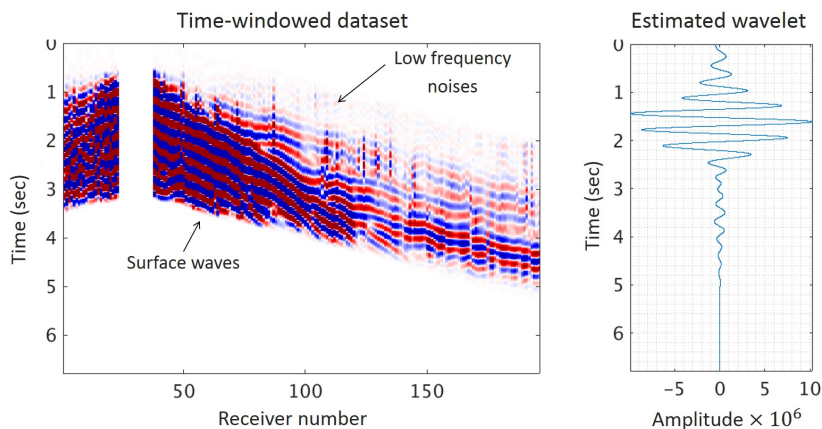


Figure 7.13: An example of the source estimation over a time-windowed dataset: The very near-offset traces are ignored and the noises before the first arrivals are muted out.

It should be noted that the seismic data contains coherent noises, arriving before the first arrivals. At some shots, it could be difficult to distinguish between the coherent noises and first arrivals. The estimated wavelet is thus anti-causal and contains artifacts from the coherent noises. To partially mitigate this issue, we perform the source estimation over different time-windows which contains different amount of noises. A source signature stacking is carried on in a second stage to only keep the coherent signatures and attenuate the artifacts. A taper is also applied at the beginning and the end of the source wavelet to make it causal and remove spurious oscillations.

7.2.3 Preconditioning and model constraint design

The importance of gradient preconditioning and model constraints has been highlighted through our synthetic applications in Chapter 6. The considered real dataset is even more challenging due to complex structures and acquisition issues. Also, the data is acquired with explosive sources, leading to limited information for V_s reconstruction. Therefore, the use of gradient preconditioning and model constraint becomes critical for reliable model parameters estimation.

7.2.3.1 Preconditioning design

To prepare for any rapid vertical velocity variation at the near-surface, the preconditioning has highly anisotropic shape at this area with vertical coherent length being $L_v = 16$ m, and horizontal coherent lengths as $L_u = 200$ m. It should be note that L_v is equivalent to a quarter of the element size at the near-surface (62.5 m as shown in Figure 7.7) and a large value of L_u is used to attenuate the acquisition footprint due the sparse source spacing. In the deeper part, we follow similar coherent lengths design as the previous applications with $L_v = 70$ m and $L_u = 150$ m. The action of the non-stationary filter is illustrated in Figure 7.14B.

The highly anisotropic smoothing at the near-surface requires an “appropriate” rotation design, otherwise the preconditioning operator might introduce thin pinch-out artifacts following unexpected direction. The initial models provides rather good velocity trends, correctly following the topography and the expected structures orientation for the first 1 km. We thus compute the near-surface dip field

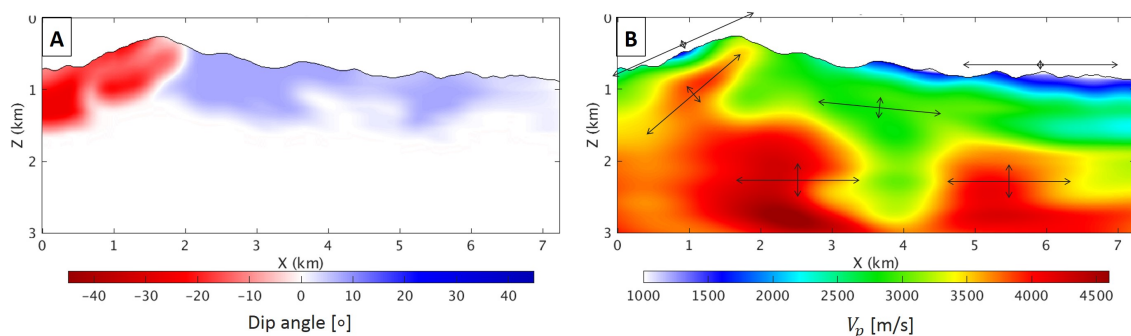


Figure 7.14: Gradient preconditioning design: A - The dip field considering by the Bessel-based gradient smoothing, in which the first 1 km below the free-surface is extracted from the initial V_p velocity model; B - The cartoon to illustrate the effect of the smoothing operator onto the gradient.

from the initial V_p and the deeper part is set to 0° as shown in Figure 7.14A. The filtering orientation is also illustrated in Figure 7.14B.

7.2.3.2 Model constraint on the ratio V_p/V_s

Section (6.2) describes the importance of the model constraint on the ratio V_p/V_s , where the use of homogeneous upper and lower bounds is enough for a stable V_s estimation. When dealing with the considered real dataset, we suffer from the instability of the V_s reconstruction meanwhile the V_p remains stable. One might start by inverting for V_p first, which could requires a better V_s model building than the empirical relation described in Section (7.2.1.2). This comes from the fact that the kinematic of the early arrivals is mainly controlled by V_p , whereas V_s affects the waveform and the lateral amplitude variation.

We believe that the V_s model computed from the given V_p through the regression (7.5) is not good enough to keep V_s as a passive parameter. We then decide to invert simultaneously for V_p and V_s and we constraint the estimated V_s through the inverting V_p following the non-stationary constraint on the ratio V_p/V_s as described in Section (3.2.4). At each spatial position \mathbf{x} , the estimation V_s should satisfy the following condition

$$r_1(\mathbf{x}) \leq \frac{V_p(\mathbf{x})}{V_s(\mathbf{x})} \leq r_2(\mathbf{x}), \quad (7.6)$$

where both upper $r_2(\mathbf{x})$ and lower bounds $r_1(\mathbf{x})$ depend on the local V_p value. A tight constraint on the ratio V_p/V_s is considered at the reservoir level, normally associated with large V_p value, whereas a less strict constraint is applied at the near-surface as shown in Figure 7.15. The upper and lower bounds are computed from V_p as

$$r_1(\mathbf{x}) = 1.6 - \frac{2000}{V_p + 6000} \quad \text{and} \quad r_2(\mathbf{x}) = \begin{cases} 2.3 + \frac{800}{V_p - 1200} & \text{for } V_p > 1500 \text{ [m/s]} \\ 5 & \text{for } V_p \leq 1500 \text{ [m/s]} \end{cases}. \quad (7.7)$$

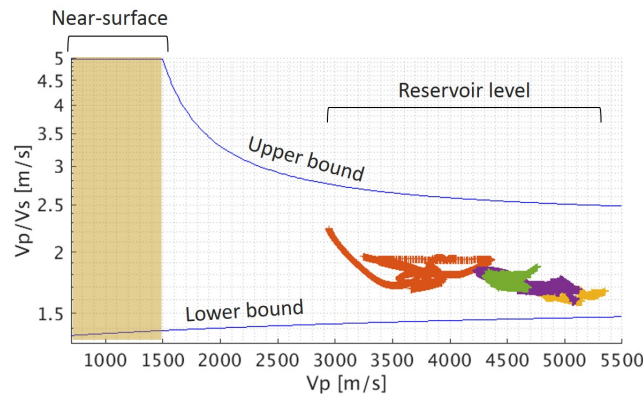


Figure 7.15: The design of model constraint on the ratio V_p/V_s based on wells data. A tight constraint on the ratio V_p/V_s is considered at the reservoir level normally associated with large V_p value, whereas a less strict constraint is applied at the near-surface.

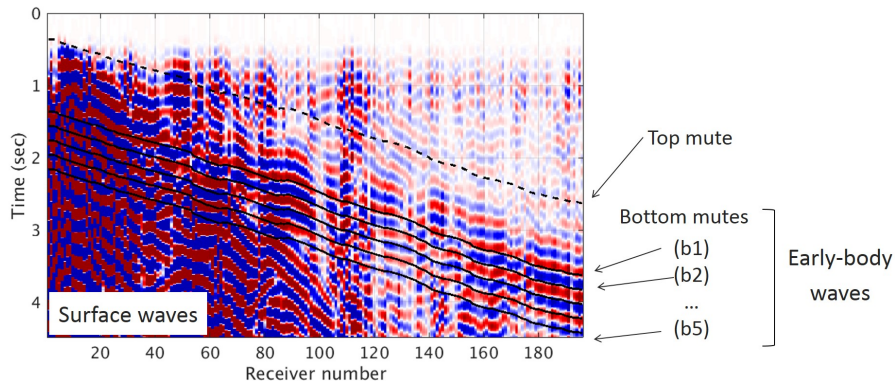


Figure 7.16: Progressive time-windows which are considered for the inversion process. Each time-window is defined by top and bottom mutes, which are designed based on the first break.

7.2.4 FWI workflow

We aim at inverting simultaneously for V_p and V_s , by considering the data-windowing hierarchy developed on synthetic benchmarks. This section only presents the inversion for the early-body waves, arriving before the surface waves as indicated in Figure 7.16. The incorporation of surface waves is under investigation. A simple bottom mute is used to separate the early body waves and surface waves.

In the body waves packages, the cycle-skipping issue occurs for the later arrivals but not for the early arrivals. We thus design a progressive time-windows to incorporate gradually the later arrivals into the inversion process. As shown in Figure 7.16, each time-window is defined by a pair of top and bottom mutes, designed from the first-break. The first-break picking is provided to us by the data owner. The top mute is important to remove the coherent noises arriving before the first arrival. The bottom mute increases progressively by a half of the apparent wavelength per time-window as indicated by the black lines (b1) to (b5) in Figure 7.16.

The inversion contains different sequence, each being performed over one time-window with 4 iterations of the preconditioned l -BFGS optimization method. Within each sequence, the density and the source wavelet are kept unchanged. We use the optimization preconditioning computed from the calculated wavefield described in Section (3.2.3). Since a relative short data interval with few traces is considered as the FWI input, we keep relatively small number of iterations per sequence. After each inversion sequence, the source wavelet is re-estimated as explained in Section (7.2.2), and the density is computed from the obtained V_p model according to the linear regression (7.4).

The gradient preconditioning and model constraint designs in Section (7.2.3) are considered at each iterations, without any modification.

7.3 Inversion results & Discussions

The inversion process makes some significant amplitude changes in both V_p and V_s models as shown in Figure 7.17B. The main fault trend which already existed in the initial model is preserved in the inverted models. The trend is further refined in the inverted V_p , and the high V_p value is propagated up to the free-surface, which is coherent with the geological observation. The inversion also remove all the low- V_p values at the near-surface, and considerably increases the V_p background.

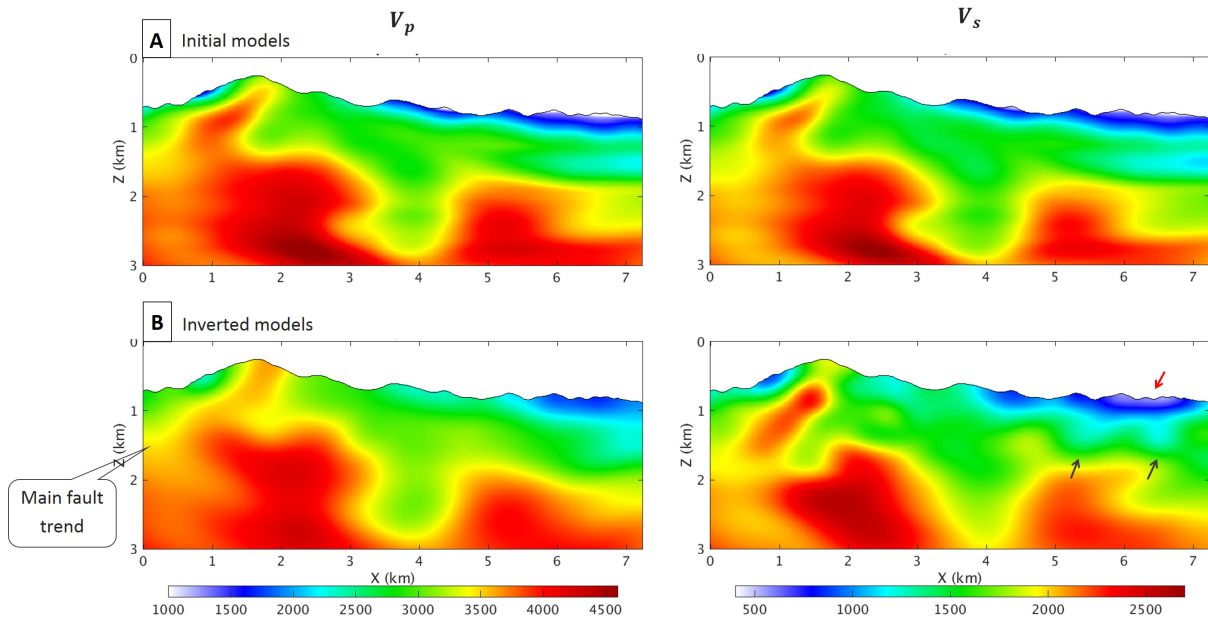


Figure 7.17: Left - V_p target cross-section; Right - V_s target cross-section. A - Initial models; B - Inverted model obtained with the workflow described in Section (7.2.4). Possible near-surface ruptures are indicated by the black arrows. Low-velocity V_s layer at the free-surface with about 100 m thickness is indicated by the red arrow.

The inverted V_s is more difficult to interpreted since it has a considerable amount of artifacts. This is due to the fact that the early body waves, acquired by the explosive sources, bring limited shear information. Even if the main fault direction is presented, the dimming of the estimated V_s at the top of the mountain, at $X \approx 1.8$ km, is not coherent with the high velocity trend in V_p . The inverted V_s detects some interesting low velocity trends at the near-surface, indicated by the black arrows in Figure 7.17. Their coherent appearance, size and amplitude in various FWI setting make us believe that they are geologically correct. The geological interpretation of the area indeed shows some possible near-surface ruptures. Since the available logs data are not located along the seismic line, they cannot be used to control the quality of inverted models.

An about 100 m thickness low-velocity layers underneath the free-surface appears in the V_s model, as indicated by the red arrow in Figure 7.17B. This low-velocity layer does not have coherent shape and has never been detected by the V_p estimation through our various FWI settings. Its appearance is not supported by any geological evidence. This inconsistency makes us believe that it is not geological correct and we are interested in which sort of uncertainty might cause this problem.

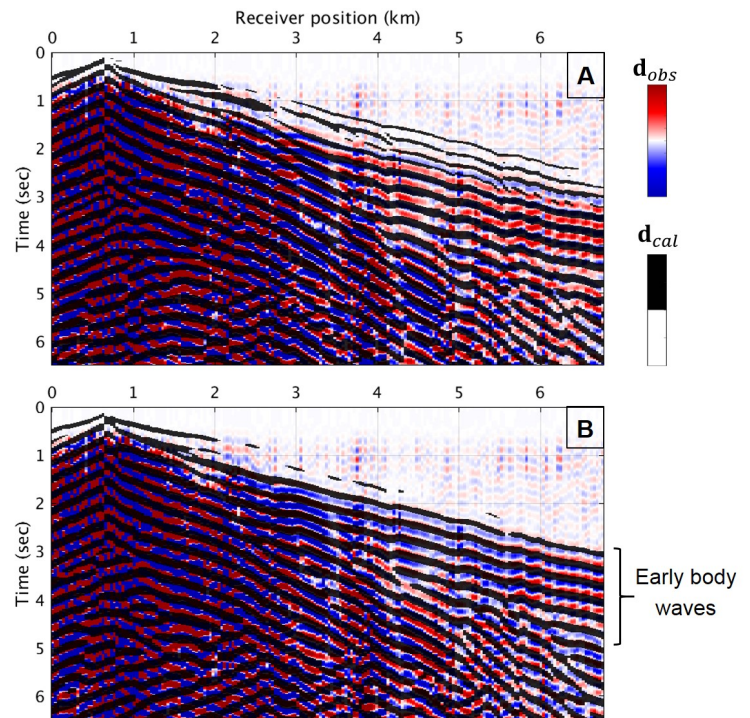
7.3.1 Model quality control through data fitting

Figure 7.18 shows the comparison of the maximal 4 Hz real data with the calculated computed from initial models and final inverted models in Figure 7.17. Compared with the real data, the initial models produce relatively good wavefield prediction, with coherent body and surface waves at all the offsets. However, the early body waves package is cycle-skipped, apart from some first arrivals, as shown in Figure 7.18A. By injecting real data progressively into the inversion process through the progressive

Figure 7.18: Data comparison for the first shot between the observed data \mathbf{d}_{obs} and calculated data \mathbf{d}_{cal} : (A) \mathbf{d}_{cal} estimated from initial models in Figure 7.17A, (B) \mathbf{d}_{cal} estimated from inverted models in Figure 7.17B.

Observed data is displayed with red-and-blue color. Calculated data is displayed with black and transparent color.

Good data-fit is identified when the red loops are perfectly covered by the black loops.



time-widows strategy, we improve the early body waves fitting as indicated by Figure 7.18B. We seem to also have a better surface waves prediction and mode convergence, for example at around 3 km position. However, the predicted data is almost cycle-skipped at the far offsets, at 6 km position. This issue is also detected in other shot gathers.

The absolute time-shift is considered for a global indication of data-fitting between the early arrivals of calculated and read data, as shown in Figure 7.19. A good data-fit is indicated by a close to 0 time-shift. Due to the significant amount of noises in the data, the calculated attribute map is noisy and difficult to interpret. This is also the reason why we consider the absolute shift instead of the normal time-shift. When comparing the time-shift maps computed from initial and inverted models in Figure 7.17A and B, a data-fit improvement can still be observed within the area [0 - 5.5 km]. However, a global degradation for the traces recorded within [5.5 - 7 km] is detected, mainly due to the low-velocity V_s layers underneath the free-surface as indicated by the red arrow in Figure 7.17B.

7.3.2 Influence of acquisition mis-positioning on inversion results

As described in Section (7.1.3), our selected target contains considerable amount of acquisition mis-positioning in depth, about 45 m at the top of the main mountain. This is equivalent to a fifth of the shortest P-wavelength in the initial models 7.17A. We thus use two synthetic benchmarks to study the influence of acquisition error on the inverted models: the first one consists of homogeneous medium properties, while the second is the initial foothill models in Figure 7.17A.

Both synthetic benchmarks have the same size and similar mesh design as the selected target in Figure 7.7. The observed data is computed from the source and receiver positions extracted from the SEG-Y header. We do not consider the deviation in the y -direction by putting all sources and receivers in the center of the model, at $Y = 1000$ m. The inversion is performed on the true models with modified acquisition that is computed from the LIDAR measurement. To simplify the numerical setting, in

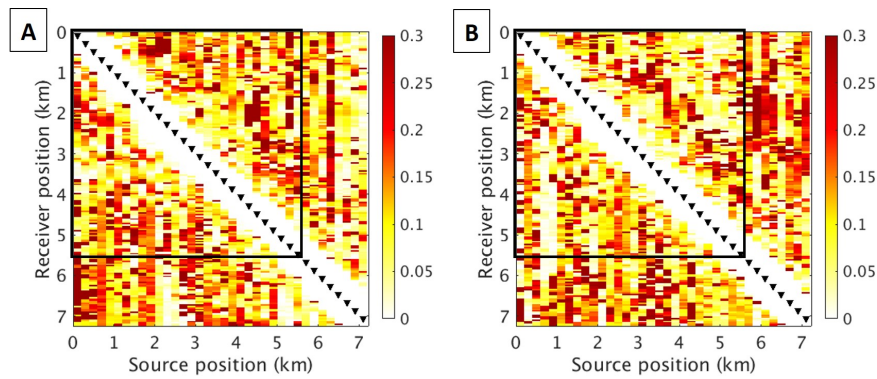


Figure 7.19: Absolute time-shift map [sec] between the calculated and the real data. The attribute is computed for the early arrivals within the range [2-4 Hz]: A - from the initial model; B - from the final models. The time-shift map is noisy due to the data quality but the improvement can be seen inside the black box.

all simulations, the relative source depths are kept as 7 m from the free-surface and all receivers are placed at the surface. Only the topography has been modified to accommodate the acquisition variation in Figure 7.5. The difference between the topography built from the absolute acquisition positions in SEG Y header and the variation computed from LIDAR measurement is presented in Figure 7.20, which is equivalent to the acquisition mis-positioning in depth. The black arrows indicate the places where the main difference occurs.

In both examples, to respect the maximal frequency at 4 Hz, a Ricker wavelet centered at 1.6 Hz is considered as the source signal. We consider the same acquisition design as the selected target with 36 explosive sources. The following inverted V_p and V_s models are obtained with one FWI iteration, by considering the entire wavefield. The gradient preconditioning is kept as minimal with vertical coherent length $L_v = 20$ m and horizontal values $L_u = L_w = 60$ m.

7.3.2.1 Homogeneous test

The reference model has homogeneous medium properties with $V_p = 2200$ m/s, $V_s = 1300$ m/s and density $\rho = 1000$ kg/m³. The shortest P- and S-wavelength are thus $\min(\lambda_p) = 550$ m and

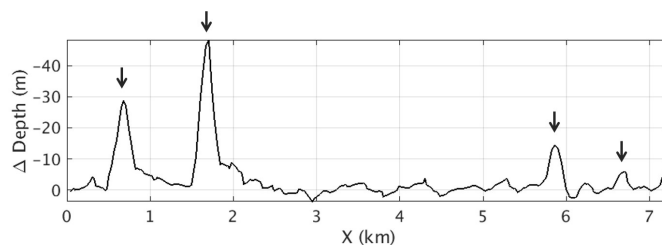


Figure 7.20: A smooth version of the difference between two topography variations in Figure 7.5, which is equivalent to the acquisition mis-positioning in depth. The peaks of the difference are indicated by black arrows.

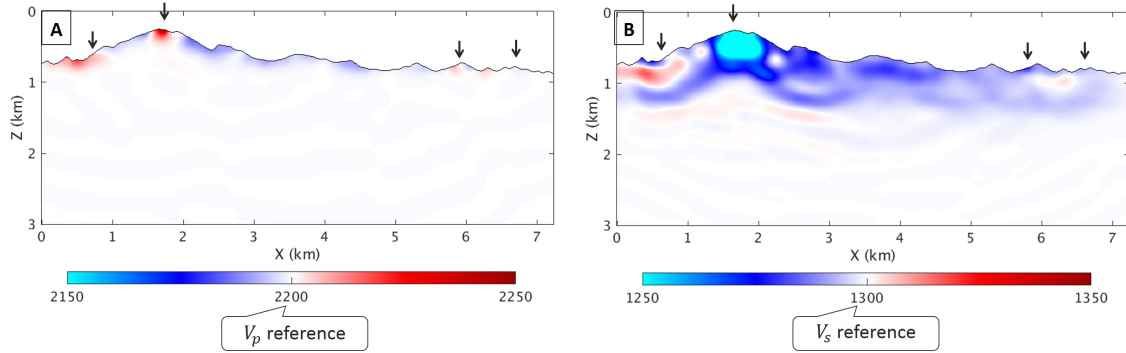


Figure 7.21: Inverted V_p (A) and V_s models (B). The main locations of acquisition mis-positioning are indicated by black arrows.

$\min(\lambda_s) = 325$ m. The acquisition mis-positioning in Figure 7.20 is relatively small compared with the P-wavelength, leading to negligible artifacts in the inverted V_p in Figure 7.21A. However, the near-surface V_s model is under-estimated, especially underneath the peak at 1.8 km, which is associated with the largest acquisition error.

Through this simple numerical setting, we confirm that the inversion tends to put low velocity anomalies at the near-surface to absorb the travel-time mismatch due to the acquisition issue. The inverted V_s is more affected because the effect depends on the size of the mis-positioning error with respect to the local wavelength.

7.3.2.2 Heterogeneous models

The reference model consists of the foothill V_p model provided by the real data owner, the computed V_s in Figure 7.17A and homogeneous density $\rho = 2000$ km/m³. The inversion is performed on the same model parameters with modified acquisition. Figure 7.22 shows the difference between the reference and the inverted shear velocity as $(V_s^{reference} - V_s^{inverted})$, in which the low velocity anomalies are highlighted by blue and purple colors. In this heterogeneous model, similar conclusion is obtained: The inversion puts low velocity layers underneath the locations of major mis-positioning in depth to compensate for the travel time error. The origin of the deeper modification is not clear to us yet.

The near-surface low velocity layers between 5.5 and 7 km are coherent with the anomaly detected by our inverted V_s model in Figure 7.17B, which might explain the origin of this artifact. Moreover, the dimming of the estimated V_s at the top of the mountain at $X \approx 1.8$ km might be caused by the significant acquisition error at this position, as shown by Figure 7.22.

7.4 Conclusion & Perspectives

The preliminary inversion results from the real data application illustrate the challenges in a real foothill environments. Apart from the data complexity and elastic multi-parameter FWI problem, we have to deal with various practical issues such as significant amount of noise in the data, sparse acquisition and topography representation uncertainty. We illustrate how additional prior information such as the logs data can be used to assist the FWI design. Even when the wells are not located along the the

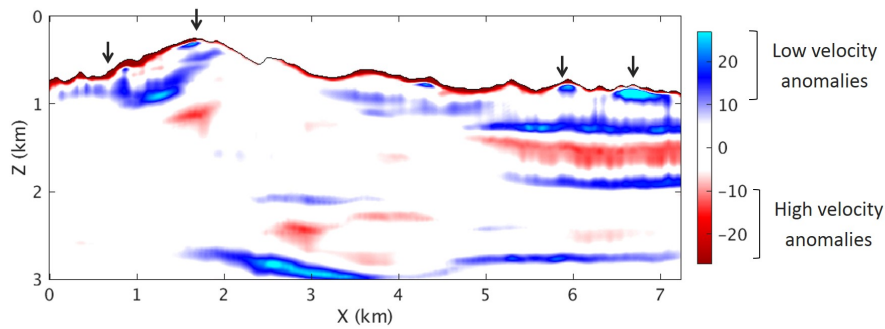


Figure 7.22: Difference between the reference and the inverted model ($V_s^{reference} - V_s^{inverted}$), where the main locations of acquisition mis-positioning are indicated by black arrows.

considered 2D seismic line to be served as final quality control tool, they participate in almost every step of the inversion process: initial shear wave speed and density models building from a given initial V_p , and mesh generation. A V_p -dependent model constraint on the ratio V_p/V_s , again relying on the logs information, is considered to constrain the V_s based on the stable V_p estimation.

At the current stage, the inversion mainly focuses on the early-body waves, which is gradually incorporated through a progressive time-windowing. The strategy helps to mitigate the cycle-skipping issue at the latter arrivals and maximize the considering data quantity. Due to the data lacking, the estimated V_s is unstable with considerable amount of artifacts. To avoid biased interpretation, we highlight the importance of inverted model quality control through data-fit analysis and uncertainty studies. The numerical tests based on the depth mis-positioning of sources and receivers indicate that the acquisition error might become one of the major challenges for the near-surface reconstruction at higher frequency. Moreover, our inversions rely on the assumption of perfect explosive sources, describing by equally diagonal terms in the moment tensor point-source. The influence of the incorrect physics of the source should also be considered.

Despite the inversion complexity, the body and surface waves prediction has coherent features as the real dataset. At this stage, the surface waves are still cycle-skipped, they cannot be considered into the inversion process yet. In the next step, to compensate for the data lacking, we aim at considering more shots, for example through the data sub-sampling technique (Ha and Shin, 2013; Warner et al., 2013), and adding the short offsets surface waves into the inversion process. However, a more recently acquired dataset or the re-processed data might help to improve the inverted models quality. Advance data processing is beyond our capacity, but we could suggest several steps such as bad traces edition and data continuity enhancing. Moreover, various non-seismic dataset and associated inverted images are available which could be used as additional data or prior information for the seismic inversion.

Effective topography representation/Acquisition positioning

The influence of acquisition error into the inversion results has been highlighted in Section 7.3.2. To reduce the mis-positioning in depth, the inversion could consider an *effective smooth topography variation* z_{eff} , lying between the LIDAR measurement z_{LIDAR} and the recorded source and receiver positions in the SEG-Y header s_{SEG-Y} . The effective topography could be estimated from an optimization process

under the following misfit function

$$\chi_{topo}(\mathbf{z}_{eff}) = \frac{1}{2} \|\mathbf{z}_{eff} - \mathbf{z}_{LIDAR}\|^2 + \frac{\mu_1}{2} \|R\mathbf{z}_{eff} - \mathbf{s}_{SEGY}\|^2 + \underbrace{\mu_2 \|\nabla \mathbf{z}_{eff}\|^2}_{\text{Smoothness}}, \quad (7.8)$$

where μ_1 and μ_2 are trade-offs parameters and R is a restriction operator at sources and receiver positions. The minimization of the least-square distance between the effective surface and the LIDAR measurement is ensured by the first term in the function (7.8). By increasing the trade-off parameter μ_1 , the effective topography is forced to go through the recorded source and receiver positions in the SEG Y header. The smoothness of the estimated surface \mathbf{z}_{eff} is guaranteed by the third Tikonov regularization term. The minimization of the misfit function (7.8) can be solved by any local optimization process starting from the initial guess \mathbf{z}_{LIDAR} , or global methods due to small number of degrees of freedom (Metropolis and Ulam, 1949; Nocedal and Wright, 2006).

Conclusion & Perspectives

Conclusion

The manuscript provides a complete review of the 3D multi-parameter FWI method in visco-elastic medium that we have developed from theoretical development to numerical implementation. The work leads to a fully integrated 3D modeling and inversion code, namely SEM46, with both elastic and viscoelastic physics. The code has been used to perform various multi-parameter FWI applications, which highlight the importance of a correct physical description, adapted inversion strategies and the introduction of geological prior information for successful applications of our method.

Theory and algorithms

The theoretical development of the wave propagation separates purposely the linear elasticity with the attenuation mechanisms, which offers an easy and explicit coupling between the anisotropic elasticity and isotropic attenuation. The attenuation parameters are explicitly incorporated in the wave equation, and therefore can be naturally reconstructed in the FWI framework. Even if the second-order visco-elastic wave equation is not self-adjoint, we can develop the adjoint system with similar equations as for the incident fields, implying that they can be propagated under the same numerical scheme. Following the forward and adjoint formulation using SLS mechanisms, the gradient expressions for density, attenuation parameters and stiffness coefficients are simply zero-lag cross-correlations in time between incident and adjoint fields, weighted by the associated spatial radiation pattern. The gradient for any other parameters can be computed by chain rule from these elementary gradients, which offers the flexibility to consider any set of parameterization. We also highlight the cumulative effect of the attenuation contribution on the recovered amplitude of elastic parameters, which was neglected in some state-of-the-art implementations (Liu and Tromp, 2008; Komatitsch et al., 2016)

We present an integrated SEM-based workflow, capable of efficiently performing 3D modeling and FWI in the time-domain for elastic and visco-elastic exploration-scale targets. Our approach relies on the trade-off between computational cost, accuracy and memory requirement. The simulation accuracy is ensured by the use of a Cartesian-based deformed mesh with high-order geometry interpolation to capture rough topography variations. The size of the elements can be adapted to the variation of the local shortest wavelength to reduce the numerical cost. The Cartesian-based mesh design also allows to access the spatial position and the neighbors of each element without any extra cost. Two MPI-based parallelism levels are considered for tackling large-scale and multiple-shots experiments, associated with an efficient computation and low-memory implementation of incident and adjoint fields. The simulation accuracy of our approach is illustrated through the comparison with elastic and visco-elastic semi-analytical solutions via the boundary integral method (Coutant, 1989). For the same physical model and similar numerical settings, we achieve comparable computational cost as two open-source reference codes: SW4 V1.1 and SPECFEM V2.0 (CPU version) for both elastic and visco-elastic simulations. The gradient is accumulated during the synchronous reconstruction of the incident field with the propagation of the adjoint field, applicable for both elastic and visco-elastic physics (Dussaud et al., 2008). For this purpose, the incident field is propagated backward in time from the last snapshot and the wavefield at the boundaries. Our low-memory implementation relies on the decimation in time of the boundaries wavefield and elastic-CARFS technique, which mitigates the instability in the reconstruction of the incident wavefield (Yang et al., 2016b,c). This strategy is critical in realistic applications, normally associated with a large number of degrees of freedom and time-steps, in which massive storage and intensive I/O might not be the optimal choice.

The structurally-based nonstationary and anisotropic Bessel smoothing is considered for an efficient gradient preconditioning. The geological prior information can be introduced through filter parameters such as 3D coherent lengths and local rotation. The smoothing process is implemented as a PDE solved with SEM, directly on the modeling mesh, following the same high-performance-computing structure as the one for the wave equation (Trinh et al., 2017b). A non-linear model constraint on the relationship between parameters is introduced to partially mitigate the cross-talks between the different classes of inverted parameters

Applications

The challenges of the elastic multi-parameter FWI in complex land areas are highlighted through different 3D sub-sets of the SEAM phase II Foothills benchmark and real data applications. Even under an ideal 3D setting with full 3D illumination and accurate initial models, the simultaneous estimation of P- and S- wavespeeds is not obvious, due to the unbalanced compressional and shear energies in the wavefield. To mitigate the unequal sensitivity of the inversion process with respect to different parameters, we propose a two-step data-windowing strategy, focusing on early body waves before considering the entire wavefield, including surface waves. The main idea is to start the inversion with the data parts which have the most P-wave signature before including more S-waves energy. The use of this data hierarchy together with the structurally-based Bessel preconditioning make possible to reconstruct accurately both compressional and shear wavespeeds.

When the data constraint is not sufficient, for instance under limited illumination in the pseudo-2D dip-line SEAM II Foothill survey, some additional constraint on model parameters are required to reduce the model space. For this purpose, the constraint on the ratio of P- and S-wavespeed, based on the rock-physics prior, could be considered to mitigate the ill-posedness of the multi-parameter inversion process. Aside from this model constraint, the two-step data-windowing strategy is combined with a low-to-high frequency hierarchy to get reliable model parameters estimation, but also to exploit the maximum amount of information in the observed data. We notice that if the surface-related energies could be accurately modeled, within half of the apparent wavelength, we can take advantage from the hidden information below these events to get better model estimation, both at the near surface and deeper parts.

The real foothill dataset brings more difficulties into our already complex elastic multi-parameter FWI problem with sparse acquisition design, noisy recordings, complex underneath structures and topography representation uncertainty. Additional prior information such as the logs data and geological interpretation has to be considered to assist the FWI process, from initial design to final model quality control. Despite the inversion complexity, the preliminary results, only relying on body waves, are shown to improve the kinematic fit and follow the expected geological interpretation. This experiment also highlights the importance of well-understanding both physics and geometry of the considered problem, to avoid any biased interpretation coming from possible artifacts in the inverted models.

Perspectives

Numerical implementation

The manuscript presents a complete numerical implementation for visco-elastic modeling and inversion. The visco-elastic modeling is 2 times more expensive than elastic modeling for regular elements,

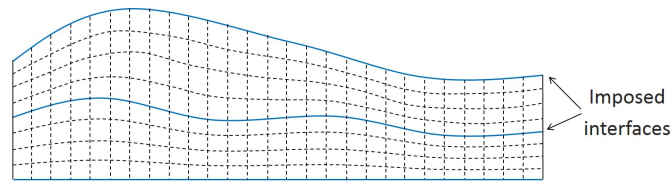


Figure 7.23: Example of mesh design with multiple imposed interfaces.

and 2.5 times more expensive for deformed elements. In addition, taking viscosity into account increases the cost of the gradient computation through the CARFS strategy, due to the instability in the reconstruction of the incident field on-the-fly. If the storage at the Nyquist frequency on the disk becomes affordable with reasonable I/O cost (using fast access SSD disk for instance), this option will significantly reduce the computational cost of the gradient calculation, making it potentially more attractive for practical applications. Additional compression strategies could also be used to further reduce the memory requirement and I/O requests (Boehm et al., 2016).

The numerical accuracy of the simulation under rapid topography variation is ensured by a high-order geometry interpolation. All presented applications are conducted with the P_4 shape function as shown in Figure 2.7(b). However, the combination of different order shape functions into a single mesh design is possible, which shall not affect the simulation cost but the implementation complexity. The presented meshes in the manuscript are designed with a single topography variation at the top the associated model. Following similar strategy, multiple interfaces could be incorporated into the future mesh design, as shown in Figure 7.23, which could be used for known multi-layers structures such as marine environment to honor the sea floor.

FWI strategies

Misfit function

In this manuscript, the surface waves are inverted with the standard least-square misfit function. In many examples, the surface waves inversion is limited by the fact that the initial near-surface models are not good enough to provide a surface waves prediction within a half of the apparent wavelength. This problem could be mitigated by considering other misfit functions with larger basins of attraction such as the envelop function (Wu et al., 2014), cross-correlation based function (van Leeuwen and Mulder, 2008), normalized deconvolution or adaptive misfit function (Warner and Guasch, 2016) and optimal transport method (Métivier et al., 2016a,b). These alternative misfit designs could be integrated in the future into the developed framework of this study.

Data consideration

Due to the limited computational resources, we normally consider a decimated number of sources. In the next step, to compensate for the data lacking, we will aim at including information from more shots, for example through the data sub-sampling technique (Ha and Shin, 2013; Warner et al., 2013).

Additional prior to assist the FWI process

Through synthetic and real data applications, the manuscript highlights that additional prior information could help for getting more reliable inversion results. Model regularization could be considered to introduce any model prior such as well data or interpreted geological structures from migrated images. Challenging exploration targets normally require multidisciplinary approaches for geological structure de-risking, implying that various seismic and non-seismic data types are available. These data and their interpreted results could be used as additional data or prior information for the seismic inversion. Since our numerical scheme relies on SEM mesh, the Bessel-liked operator could be the optimal choice for designing the spatially variant model covariance matrix.

Moreover, the model constraint on the ratio of P- and S-wavespeeds, described in Section (3.2.4), could be generalized for other set of parameters with any expected differentiable and monotonic relationship between parameters. The numerical implementation shall be similar, at insignificant computational cost.

Imaging condition and quality controls

The input for the application in Chapter (7) is a raw foothill dataset, which contains significant amount of noise. A processing sequence dedicated to foothill dataset should be established to prepare the data for the inversion, at least to remove bad traces and attenuate the low frequency noise. Any processing step should preserve and take into account the significant acquisition deviation in y -direction and vertical elevation of sources and receivers.

After obtaining reliable velocity models, migrated images are generated to get a high resolution image about the sub-surface structures. Due to the rough topography variation, the near-surface image produced by standard finite difference techniques is very noisy, also affecting the structures at greater depths. The development of a SEM-based migration tool could take benefit from our optimized modeling & inversion kernels and Bessel smoothing operator to remove artifacts in the migrated image. By considering a complete physics and accurate 3D geometry, the tool could deliver a better near-surface image, and simply some steps in the processing sequences such as static correction.

The manuscript promotes the use of time-shift and maximal cross-correlation attributes for data quality controls but also to assist the design of inversion hierarchy. The interpretation is intuitive for 2D acquisition designs but more effort would be required for 3D acquisitions. Other attributes less sensitive to noise should be considered to facilitate the interpretation of inverted results and the detection of artifacts, especially in 3D acquisitions.

Code development history

The theoretical development and numerical implementation of the SEM46 code presented in this manuscript has been carried by Romain Brossier and the main author, under the orientation of other supervisors (Brossier and Trinh, 2017). The optimized elastic modeling and inversion kernels, implemented on a linearly deformed Cartesian-based mesh, was available at the beginning of the PhD project. The two-level MPI-based parallelization and the general structure of the code has been previously established. The former version only considered vertical point-force source, and the wavefield at the boundary was absorbed by a first-order radiative absorbing boundary condition (Equation (2.77) in Section (2.4)).

During the PhD project, the main author has been making several contributions into the code development as

- **Mesh design:** The Cartesian-based deformed mesh is improved with high-order geometry representation to capture complex topography. The element sizes can vary in three directions, following the expected shortest wavelength, to reduce the numerical cost. The mesh is expanded to accommodate for the sponge layers at the model boundaries to absorb the out-going wavefield.
- **Visco-elastic modeling:** Based on the theoretical review on the first-order visco-elastic wave equation of Yang et al. (2016a), the main author works on similar formulation for second-order system of the displacement field, following SEM perspectives. The main author considers the visco-elastic Newmark-scheme for the propagation in time, which preserves the second-order accuracy of the existing elastic scheme. The moment-tensor source implementation is introduced for the numerical comparison with the semi-analytical solutions on LOH benchmarks, described in Section (2.6).
- **Adjoint system & Gradient building:** The main author has developed the adjoint system with similar equations as for the incident fields, resulting in simple visco-elastic gradients for the set of parameters $C_{IJ}, \rho, Q_p^{-1}, Q_s^{-1}$ as zero-lag cross-correlations of adjoint and incident fields, weighted by the associated spatial radiation pattern. For efficient and low-memory gradient computations, the main author integrates the existing wavefield-decimation toolbox and an elastic version of the CARFS toolbox, developed by (Yang et al., 2016b,c), into the SEM46 inversion scheme.
- **Bessel smoothing filter:** Under the original idea made by Wellington et al. (2017), the main author developed a complete theoretical base for the Bessel smoothing filter in both 2D and 3D. The 3D implementation in SEM mesh for gradient preconditioning has been introduced into the SEM46 code.
- **Non-linear constraint on the ratio between inverting parameters.**

Besides the scientific contribution, the main author completely changes the interface of the code, to ease the user experiences. The code has been distributed within ISTERRE laboratory and industrial sponsors with positive feedback in both scientific capacities and user experiences.

Bibliography

- Abas Kalair, N., Kalair, A., and Khan, N. (2015). Review of Fossil Fuels and Future Energy Technologies. *Futures*, 69.
- Abramowitz, M. and Stegun, I. (1972). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover publications, Inc., New York.
- Aki, K. and Richards, P. G. (2002). *Quantitative seismology, theory and methods, second edition*. University Science Books, Sausalito, California.
- Amestoy, P., Brossier, R., Buttari, A., L'Excellent, J.-Y., Mary, T., Métivier, L., Miniussi, A., and Operto, S. (2016). Fast 3D frequency-domain full waveform inversion with a parallel Block Low-Rank multifrontal direct solver: application to OBC data from the North Sea. *Geophysics*, 81(6):R363 – R383.
- Andrew, L., von Abendorff Eddy, Justin, N., and Alcindo, P. M. M. (2013). Simultaneous long offset (SLO) towed streamer seismic acquisition. *ASEG Extended Abstracts*.
- Baumstein, A. (2013). POCS-based geophysical constraints in multi-parameter Full Wavefield Inversion. In *EAGE, 75th EAGE Conference and Exhibition incorporating SPE EUROPEC 2013*.
- Ben Jemaa, M., Glinsky-Olivier, N., Cruz-Atienza, V. M., and Virieux, J. (2009). 3D Dynamic rupture simulations by a finite volume method. *Geophysical Journal International*, 178:541–560.
- Bérenger, J.-P. (1994). A perfectly matched layer for absorption of electromagnetic waves. *Journal of Computational Physics*, 114:185–200.
- Bielak, J., Loukakis, K., Hisada, Y., and Yoshimura, C. (2003). Domain Reduction Method for Three-Dimensional Earthquake Modeling in Localized Regions, Part I: Theory. *Bulletin of the Seismological Society of America*, 93(2):817–824.
- Bishop, T. N., Bube, K. P., Cutler, R. T., Langan, R. T., Love, P. L., Resnick, J. R., Shuey, R. T., and Spinder, D. A. (1985). Tomographic determination of velocity and depth in laterally varying media. *Geophysics*, 50:903–923.
- Blanch, J., Robertson, J. O. A., and Symes, W. W. (1995). Modeling of a constant Q: Methodology and algorithm for an efficient and optimally inexpensive viscoelastic technique. *Geophysics*, 60:176–184.
- Boehm, C., Hanzich, M., de la Puente, J., and Fichtner, A. (2016). Wavefield compression for adjoint methods in full-waveform inversion. *Geophysics*, 81(6):R385–R397.

BIBLIOGRAPHY

- Bohlen, T. and Saenger, E. H. (2006). Accuracy of heterogeneous staggered-grid finite-difference modeling of Rayleigh waves. *Geophysics*, 71:109–115.
- Bonnasse-Gahot, M., Calandra, H., Diaz, J., and Lanteri, S. (2018). Hybridizable discontinuous Galerkin method for the 2-D frequency-domain elastic wave equations. *Geophysical Journal International*, 213(1):637–659.
- Borisov, D., Dmitry, and Singh, S. C. (2015). Three-dimensional elastic full waveform inversion in a marine environment using multicomponent ocean-bottom cables: a synthetic study. *Geophysical Journal International*, 201:1215–1234.
- Borisov, D., Modrak, R., Gao, F., and Tromp, J. (2017). Spectral-element based 3D elastic full-waveform inversion of surface waves in the presence of irregular topography using an envelope-based misfit function. *Geophysics*.
- Boyd, S., Parikh, N., Chu, E., Peleato, B., and Eckstein, J. (2010). Distributed optimization and statistical learning via the alternating direction of multipliers. *Foundations and trends in machine learning*, 3(1):1–122.
- Boyle, J. P. and Dykstra, R. L. (1986). A method for finding projections onto the intersection of convex sets in Hilbert spaces. In Dykstra, R., Robertson, T., and Wright, F. T., editors, *Advances in Order Restricted Statistical Inference: Proceedings of the Symposium on Order Restricted Statistical Inference held in Iowa City, Iowa, September 11–13, 1985*, pages 28–47. Springer New York, New York, NY.
- Brenders, A. J. and Pratt, R. G. (2007). Full waveform tomography for lithospheric imaging: results from a blind test in a realistic crustal model. *Geophysical Journal International*, 168:133–151.
- Brossier, R. (2011). Two-dimensional frequency-domain visco-elastic full waveform inversion: Parallel algorithms, optimization and performance. *Computers & Geosciences*, 37(4):444 – 455.
- Brossier, R., Etienne, V., Operto, S., and Virieux, J. (2010a). Frequency-domain numerical modelling of visco-acoustic waves based on finite-difference and finite-element discontinuous galerkin methods. In Dissanayake, D. W., editor, *Acoustic Waves*, pages 125–158. SCIYO.
- Brossier, R., Operto, S., and Virieux, J. (2009a). 2D elastic frequency-domain full-waveform inversion for imaging complex onshore structures. In *Expanded Abstracts*, page U019. EAGE.
- Brossier, R., Operto, S., and Virieux, J. (2009b). Robust frequency-domain full-waveform inversion using the l_1 norm. *Geophysical Research Letters*, 36:L20310.
- Brossier, R., Operto, S., and Virieux, J. (2010b). Which data residual norm for robust elastic frequency-domain full waveform inversion? *Geophysics*, 75(3):R37–R46.
- Brossier, R. and Trinh, P. T. (2017). *SEM46 Manual Version 2.1*. SEISCOPE Consortium.
- Bunks, C., Salek, F. M., Zaleski, S., and Chavent, G. (1995). Multiscale seismic waveform inversion. *Geophysics*, 60(5):1457–1473.
- Byrd, R. H., Lu, P., and Nocedal, J. (1995). A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific and Statistical Computing*, 16:1190–1208.

- Canuto, C., Hussaini, M. Y., Quarteroni, A., and Zang, T. A. (2006). *Spectral Methods Fundamentals in Single Domains*. Springer.
- Capdeville, Y. and Cance, P. (2015). Residual homogenization for elastic wave propagation in complex media. *Geophysical Journal International*, 200(2):986.
- Capdeville, Y., Chaljub, E., Vilotte, J., and Montagner, J. (2003). Coupling the spectral element method with a modal solution for elastic wave propagation in global earth models. *Geophysical Journal International*, 152:34–67.
- Capdeville, Y., Guillot, L., and Marigo, J.-J. (2010). 2-D non-periodic homogenization to upscale elastic media for P-SV waves. *Geophysical Journal International*, 182:903–922.
- Capdeville, Y., Gung, Y., and Romanowicz, B. (2005). Towards global earth tomography using the spectral element method: a technique based on source stacking. *Geophysical Journal International*, 162:541–554.
- Carcione, J., Kosloff, D., and Kosloff, R. (1988). Wave propagation simulation in a linear viscoacoustic medium. *Geophysical Journal International*, 93(2):393–401.
- Carlson, D., Long, A., Sollner, W., Tabti, H., Tenghamn, R., and Lunde, N. (2007). Increased resolution and penetration from a towed dual-sensor streamer. *First Break*.
- Cary, P. and Zhang, C. (2009). Ground Roll Attenuation via SVD and Adaptive Substraction. In *Frontiers and Innovation, CSPG CSEG Convention*.
- Castellanos, C., Métivier, L., Operto, S., Brossier, R., and Virieux, J. (2015). Fast full waveform inversion with source encoding and second-order optimization methods. *Geophysical Journal International*, 200(2):720–744.
- Cerjan, C. D., Kosloff, D., and Reshef, M. (1985). A non-reflecting boundary condition for direct acoustic and elastic wave equations. *Geophysics*, 50:705–708.
- Checa, J., Duque, C., Alfonso, H., Chalar, E., Pastore, D., and Pedraza, C. (2009). Mountain-front seismic acquisition and processing. *The Leading Edge*, 28(6):668.
- Claerbout, J. (1971). Towards a unified theory of reflector mapping. *Geophysics*, 36:467–481.
- Claerbout, J. F. (1970). Coarse grid calculations of wave in inhomogeneous media with application to delineation of complicated seismic structure. *Geophysics*, 35(3):407–418.
- Claerbout, J. F. (1992). *Earth Sounding Analysis*. Blackwell Scientific Publications.
- Clayton, R. and Engquist, B. (1977). Absorbing boundary conditions for acoustic and elastic wave equations. *Bulletin of the Seismological Society of America*, 67:1529–1540.
- Cockburn, B., Karniadakis, G., and Shu, C. W. (2000). *Discontinuous Galerkin Method : Theory, Computation and Application*. Springer - Lecture Notes in Computational Science Engineering.
- Coutant, O. (1989). Program of Numerical Simulation AXITRA. *Research report, LGIT, Grenoble*.
- Cox, M. (1999). *Static Corrections for Seismic Reflection Surveys*. Society of Exploration Geophysicists.

- Cupillard, P., Delavaud, E., Burgos, G., Festa, G., Vilotte, J., Capdeville, Y., and Montagner, J. (2011). RegSEM: a versatile code based on the spectral element method to compute seismic wave propagation at the regional scale. *Geophysical Journal International*, 188(3):1203–1220.
- Dai, Y. and Yuan, Y. (1999). A nonlinear conjugate gradient method with a strong global convergence property. *SIAM Journal on Optimization*, 10:177–182.
- Day, S. M., Bielak, J., Dreger, D., Larsen, S., Graves, R., and and, A. P. (2003). *Tests of 3D elastodynamics codes: Final report for lifelines project IA02*. Pacific Earthquake Engineering Research Center, Berkley , California.
- Deriche, R. (1992). Recursively implementing the gaussian and its derivatives. In *Proceedings of the 2nd International Conference on Image Processing*, pages 263–267, Singapore.
- Devaney, A. (1984). Geophysical diffraction tomography. *Geoscience and Remote Sensing, IEEE Transactions on*, GE-22(1):3–13.
- Deville, M., Fischer, P., and Mund, E. (2002). *High Order Methods for Incompressible Fluid Flow*. Cambridge University Press.
- Diaz, J. and Grote, M. J. (2009). Energy conserving explicit local time stepping for second order wave equation. *SIAM Journal on Scientific Computing*, 31(3):1985–2014.
- Dussaud, E., Symes, W. W., Williamson, P., Lemaistre, L., Singer, P., Denel, B., and Cherrett, A. (2008). Computational strategies for reverse-time migration. In *Society of Exploration Geophysics technical program expanded abstracts 2008*, pages 2267–2271. Society of Exploration Geophysicists.
- Emmerich, H. and Korn, M. (1987). Incorporation of attenuation into time-domain computation of seismic wavefield. *Geophysics*, 52:1252–1264.
- Erbach, G. (2014). Unconventional gas and oil in North America: The impact of shale gas and tight oil on the US and canadian economies and on global energy flows. *European Parliamentary Research Service*.
- Etgen, J., Gray, S. H., and Zhang, Y. (2009). An overview of depth imaging in exploration geophysics. *Geophysics*, 74(6):WCA5–WCA17.
- Etienne, V., Chaljub, E., Virieux, J., and Glinsky, N. (2010). An hp-adaptive discontinuous Galerkin finite-element method for 3D elastic wave modelling. *Geophysical Journal International*, 183(2):941–962.
- Faccioli, E. F., Paolucci, R., and Quarteroni, A. (1997). 2D and 3D elastic wave propagation by a pseudo-spectral domain decomposition method. *Journal of Seismology*, 1:237–251.
- Fehmers, G. C. and Höcker, C. F. W. (2003). Fast structural interpretation with structure-oriented filtering. *Geophysics*, 68(4):1286–1293.
- Fichtner, A., Trampert, J., Cupillard, P., Saygin, E., Taymaz, T., Capdeville, Y., and nor, A. V. (2013). Multiscale full waveform inversion. *Geophysical Journal International*, 194:534–556.
- Fichtner, A. and van Driel, M. (2014). Models and Fréchet kernels for frequency-(in)dependent Q. *Geophysical Journal International*, 198(3):1878–1889.

- Fomel, S. (2002). Applications of plane-wave destruction filters. *Geophysics*, 67(6):1946–1960.
- Fomel, S. and Claerbout, J. F. (2003). Multidimensional recursive filter preconditioning in geophysical estimation problems. *Geophysics*, 68(2):1–12.
- Fuji, N., Ovcharenko, O., Martin, R., and Ovcharenko, O. (2016). Simple and accurate operators based on Taylor expansion for 2D elastic seismogram calculation under geological discontinuities with regular cartesian grids. In *Methods and Challenges of Seismic Wave Modelling for Seismic Imaging Workshop, 78th Annual EAGE Meeting (Vienna)*.
- Gao, L., Brossier, R., Pajot, B., Tago, J., and Virieux, J. (2015). An immersed free surface boundary treatment for seismic wave simulation. *Geophysics*, 80(5):T193–T209.
- Gardner, G. F., Gardner, L., and Gregory, A. (1974). Formation velocity and density—the diagnostic basics for stratigraphic traps. *Geophysics*, 39(6):770–780.
- Gokhberg, A. and Fichtner, A. (2016). Full-waveform inversion on heterogeneous HPC systems. *Computers & Geosciences*.
- Golub, G. H. (1996). *Matrix Computation, third edition*. Johns Hopkins Studies in Mathematical Sciences.
- Gosselin-Cliche, B. and Giroux, B. (2014). 3D frequency-domain finite-difference viscoelastic-wave modeling using weighted average 27-point operators with optimal coefficients. *Geophysics*, 79(3):T169–T188.
- Gray, S. H. (2016). Seismic imaging. *Encyclopedia of Exploration Geophysics*, pages S1–1–S1–16.
- Griewank, A. and Walther, A. (2000). Algorithm 799: Revolve: An implementation of checkpointing for the reverse or adjoint mode of computational differentiation. *ACM Trans. Math. Software*, 26:19–45.
- Guitton, A., Ayeni, G., and Díaz, E. (2012). Constrained full-waveform inversion by model reparameterization. *Geophysics*, 77(2):R117–R127.
- Guoqing, L., M., S. P., Egill, H., and H., T. C. (2007). A three-dimensional crustal seismic velocity model for southern California from a composite event method. *Journal of Geophysical Research: Solid Earth*, 112(B11).
- Ha, W. and Shin, C. (2013). Efficient laplace-domain full waveform inversion using a cyclic shot subsampling method. *Geophysics*, 78(2):R37–R46.
- Hale, D. (2007). Local dip filtering with directional laplacians. Technical report, Centre for Wave Phenomena, Colorado School of Mines.
- Hale, D. (2009). Structure-oriented smoothing and semblance. *CWP Report*, page 635.
- He, W., Brossier, R., Métivier, L., and Virieux, J. (2018). A practical workflow of full waveform inversion to process land seismic data: Synthetic study. In *Expanded Abstracts, 80th Annual EAGE Meeting (Copenhagen)*.
- Herrmann, F. J., Erlangga, Y. A., and Lin, T. T. Y. (2009). Compressive simultaneous full-waveform simulation. *Geophysics*, 74(4):A35–A40.

BIBLIOGRAPHY

- Hovland, V. (2016). Transforming ocean bottom seismic technology into an exploration tool. *First Break*, 34.
- Huiskes, M., Plessix, R., and Mulder, W. (2016). A fast3D free-surface topography method for acoustic full-waveform inversion. In *Methods and Challenges of Seismic Wave Modelling for Seismic Imaging Workshop, 78th Annual EAGE Meeting (Vienna)*.
- Huiskes, M., Plessix, R., and Mulder, W. (2017). Acoustic VTI Full-waveform inversion with 3-D Free-surface Topography. In *79th Annual EAGE Meeting (Paris)*.
- Jones, I. F. (2015). Estimating subsurface parameter fields for seismic migration: Velocity model building. *Encyclopedia of Exploration Geophysics*, pages U1–1–U1–24.
- Kaiser, M. J. (2007). A Survey of Drilling Cost and Complexity Estimation Models. *International Journal of Petroleum Science and Technology*, 1(1):1–22.
- Käser, M. and 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*, 166:855–877.
- Kirkpatrick, S., Gelatt, C. D., and Vecchi, M. P. (1983). Optimization by simulated annealing. *Science*, 220(4598):671–680.
- Kjartansson, E. (1979). Constant Q-wave propagation and attenuation. *Journal of Geophysical Research*, 84(B9):4737–4748.
- Koketsu, K., Fujiwara, H., and Ikegami, Y. (2004). Finite-element Simulation of Seismic Ground Motion with a Voxel Mesh. *Pure and Applied Geophysics*, 161:2183–2198.
- Komatitsch, D. (1997). *Méthodes spectrales et éléments spectraux pour l'équation de l'élastodynamique 2D et 3D en milieu hétérogène*. PhD thesis, Institut de Géophysique du Globe de Paris.
- Komatitsch, D., Erlebacher, G., Göddeke, D., and Michéa, D. (2010). High-order finite-element seismic wave propagation modeling with MPI on a large GPU cluster. *Journal of Computational Physics*, 229(20):7692–7714.
- Komatitsch, D. and Martin, R. (2007). An unsplit convolutional perfectly matched layer improved at grazing incidence for the seismic wave equation. *Geophysics*, 72(5):SM155–SM167.
- Komatitsch, D. and Tromp, J. (1999). Introduction to the spectral element method for three-dimensional seismic wave propagation. *Geophysical Journal International*, 139(3):806–822.
- Komatitsch, D. and Vilotte, J. P. (1998). The spectral element method: an efficient tool to simulate the seismic response of 2D and 3D geological structures. *Bulletin of the Seismological Society of America*, 88:368–392.
- Komatitsch, D., Xie, Z., Bozdog, E., de Andrade, E. S., Peter, D., Liu, Q., and Tromp, J. (2016). Anelastic sensitivity kernels with parsimonious storage for adjoint tomography and full waveform inversion. *Geophysical Journal International*, 206(3):1467–1478.

- Kostin, V., Neklyudov, D., Tcheverda, V., Belonosov, M., and Dmitriev, M. (2016). 3d elastic frequency-domain iterative solver for full-waveform inversion. In *Expanded Abstracts, 86th annual meeting (Dallas)*, pages 3825–3829. Society of Exploration Geophysics.
- Kouroussis, G., Verlinden, O., and Conti, C. (2011). Finite-Dynamic Model for Infinite Media: Corrected Solution of Viscous Boundary Efficiency. *Journal of Engineering Mechanics*, 137(7):509–511.
- Krebs, J., Anderson, J., Hinkley, D., Neelamani, R., Lee, S., Baumstein, A., and Lacasse, M. D. (2009). Fast full-wavefield seismic inversion using encoded sources. *Geophysics*, 74(6):WCC105–WCC116.
- Kristeková, M., Kristek, J., Moczo, P., and Day, S. (2006). Misfit Criteria for Quantitative Comparison of Seismograms. *Bulletin of the Seismological Society of America*, 96:1836–1850.
- Lailly, P. (1983). The seismic problem as a sequence of before-stack migrations. In Bednar, J., editor, *Conference on Inverse Scattering: Theory and Applications*. SIAM, Philadelphia.
- Lambaré, G., Alerini, M., Baina, R., and Podvin, P. (2004). Stereotomography: a semi-automatic approach for velocity macromodel estimation. *Geophysical Prospecting*, 52:671–681.
- LeVeque, R. J. (2007). *Finite Difference Methods for Ordinary and Partial Differential Equations, Steady State and Time Dependent Problems*. SIAM, Philadelphia, USA.
- Li, L., Lanteri, S., Mortensen, N. A., and Wubs, M. (2017). A hybridizable discontinuous Galerkin method for solving nonlocal optical response models. *Computer Physics Communications*.
- Li, Y., Métivier, L., Brossier, R., Han, B., and Virieux, J. (2014). 2D and 3D frequency-domain elastic wave modeling in complex media with a parallel iterative solver. *Geophysics*, page Accepted.
- Liu, Q. and Tromp, J. (2006). Finite-frequency kernels based on adjoint methods. *Bulletin of the Seismological Society of America*, 96(6):2383–2397.
- Liu, Q. and Tromp, J. (2008). Finite-frequency sensitivity kernels for global seismic wave propagation based upon adjoint methods. *Geophysical Journal International*, 174(1):265–286.
- Lombard, B., Piraux, J., Gelis, C., and Virieux, J. (2008). Free and smooth boundaries in 2-D finite-difference schemes for transient elastic waves. *Geophysical Journal International*, 172:252–261.
- Lv, H., Li, X., Jiang, Z., Zhang, L., Xiao, L., Wu, J., Liang, Y., Liang, B., Zhang, M., Li, R., Li, F., Yang, S., Van Baaren, P., Heesom, T., and Xiao, H. (2015). Land seismic in difficult terrain and complex geology: A case study from the jungar basin, northwest china. *First Break*, 33:67–74.
- Lysmer, J. and Kuhlemeyer, R. (1969). Finite Dynamic Model for Infinite Media. *Journal of the Engineering Mechanics Division*, 95(EM4):859–877.
- Malehmir, A., Maries, G., Backstrom, E., Schon, M., and Marsden, P. (2017). Developing cost-effective seismic mineral exploration methods using a landstreamer and a drophammer. *Scientific Reports*.
- Marfurt, K. (1984). Accuracy of finite-difference and finite-element modeling of the scalar and elastic wave equations. *Geophysics*, 49:533–549.

BIBLIOGRAPHY

- Marjanovic, M., Plessix, R., Stopin, A., and Singh, S. (2018). Acoustic Vs. Elastic 3-D Full-Waveform Inversion at the East Pacific Rise 9°50'n'. In *80th EAGE Conference and Exhibition 2018, Copenhagen*.
- Martin, R., Komatitsch, D., and Gedney, S. D. (2008). A variational formulation of a stabilized unsplit convolutional perfectly matched layer for the isotropic or anisotropic seismic wave equation. *Computer Modeling in Engineering and Sciences*, 37:274–304.
- Martin, R., Komatitsch, D., Gedney, S. D., and Bruthiaux, E. (2010). A high-order time and space formulation of the unsplit perfectly matched layer for the seismic wave equation using Auxiliary Differential Equations. *Computer Modelling in Engineering and Sciences - CMES*, 56.
- Masoni, I., Boelle, J.-L., Brossier, R., and Virieux, J. (2016). Layer stripping FWI for surface waves. In *SEG Technical Program Expanded Abstracts 2016*, pages 1369–1373.
- McMechan, G. (1983). Migration by extrapolation of time-dependent boundary values. *Geophysical Prospecting*, 31(3):413–420.
- Menke, W. (1984). *Geophysical Data Analysis: Discrete Inverse Theory*. Academic Press, Inc., Orlando, USA.
- Métivier, L., Breteau, F., Brossier, R., Operto, S., and Virieux, J. (2014). Full waveform inversion and the truncated Newton method: quantitative imaging of complex subsurface structures. *Geophysical Prospecting*, 62:1353–1375.
- Métivier, L. and Brossier, R. (2016). The seiscopes optimization toolbox: A large-scale nonlinear optimization library based on reverse communication. *Geophysics*, 81(2):F11–F25.
- Métivier, L., Brossier, R., Mérigot, Q., Oudet, E., and Virieux, J. (2016a). Measuring the misfit between seismograms using an optimal transport distance: Application to full waveform inversion. *Geophysical Journal International*, 205:345–377.
- Métivier, L., Brossier, R., Mérigot, Q., Oudet, E., and Virieux, J. (2016b). An optimal transport approach for seismic tomography: Application to 3D full waveform inversion. *Inverse Problems*, 32(11):115008.
- Métivier, L., Brossier, R., Virieux, J., and Operto, S. (2013). Full Waveform Inversion and the truncated Newton method. *SIAM Journal On Scientific Computing*, 35(2):B401–B437.
- Metropolis, N. and Ulam, S. (1949). The Monte Carlo Method. *Journal of the American Statistical Association*, 44(247):335–341.
- Michéa, D. and Komatitsch, D. (2010). Accelerating a 3D finite-difference wave propagation code using GPU graphics cards. *Geophysical Journal International*, 182(1):389–402.
- Moczo, P. (1989). Finite-difference technique for SH waves in 2-D media using irregular grids, application to the seismic response problem. *Geophysical Journal International*, 99:321–329.
- Moczo, P., Bystricky, E., Carcione, J. M., and Bouchon, M. (1997). Hybrid modeling of P-SV seismic motion at inhomogeneous viscoelastic topographic structures. *Bulletin of the Seismological Society of America*, 87:1305–1323.

- Moczo, P. and Kristek, J. (2005). On the rheological models used for time-domain methods of seismic wave propagation. *Geophysical Research Letters*, 32(1).
- Moldoveanu, N., Seymour, N., Manen, D., and Caprioli, P. (2012). Broadband Seismic Methods for Towed-streamer acquisition. In *74th Annual EAGE Meeting (Denmark)*, volume 73, pages A45–A49.
- Nemeth, T., Wu, C., and Schuster, G. T. (1999). Least-squares migration of incomplete reflection data. *Geophysics*, 64(1):208–221.
- Nocedal, J. and Wright, S. J. (2006). *Numerical Optimization*. Springer, 2nd edition.
- Operto, S. and Miniussi, A. (2017). Vertical wavespeed, density and attenuation imaging by 3D efficient frequency-domain FWI of wide-azimuth OBC data - a North Sea case study. In *79th EAGE Conference and Exhibition 2017, Paris*.
- Operto, S., Miniussi, A., Brossier, R., Combe, L., Métivier, L., Monteiller, V., Ribodetti, A., and Virieux, J. (2015). Efficient 3-D frequency-domain mono-parameter full-waveform inversion of ocean-bottom cable data: application to Valhall in the visco-acoustic vertical transverse isotropic approximation. *Geophysical Journal International*, 202(2):1362–1391.
- Operto, S., Virieux, J., Dessa, J. X., and Pascal, G. (2006). Crustal imaging from multifold ocean bottom seismometers data by frequency-domain full-waveform tomography: application to the eastern Nankai trough. *Journal of Geophysical Research*, 111(B09306):doi:10.1029/2005JB003835.
- Oristaglio, M. (2012). SEAM phase II - Land Seismic Challenges. *The Leading Edge*, 31(3):264–266.
- Oristaglio, M. (2016). Land Seismic Challenges wraps up with Foothills model. *The Leading Edge*, 35(3):292–293.
- Peter, D., Komatitsch, D., Luo, Y., Martin, R., Le Goff, N., Casarotti, E., Le Loher, P., Magnoni, F., Liu, Q., Blitz, C., Nissen-Meyer, T., Basini, P., and Tromp, J. (2011). Forward and adjoint simulations of seismic wave propagation on fully unstructured hexahedral meshes. *Geophysical Journal International*, 186(2):721–739.
- Peters, B. and Herrmann, F. J. (2017). Constraints versus penalties for edge-preserving full-waveform inversion. *The Leading Edge*, 36(1):94–100.
- Petersson, N. A. and Sjögreen, B. (2012). Stable and efficient modeling of anelastic attenuation in seismic wave propagation. *Commun. Comput. Phys.*, 12(1):193–225.
- Petersson, N. A. and Sjögreen, B. (2013). *User's Guide to SW4, version 1.1*. Lawrence Livermore National Laboratory, 7000 East Ave., Livermore, CA 94550, USA, llnl-sm-662014 edition. Computational Infrastructure for geodynamics (CIG), www.geodynamics.org.
- Petersson, N. A. and Sjögreen, B. (2015). Wave propagation in anisotropic elastic materials and curvilinear coordinates using a summation-by-parts finite difference method. *Journal of Computational Physics*, 299:820 – 841.
- Plessix, R. E. (2006). A review of the adjoint-state method for computing the gradient of a functional with geophysical applications. *Geophysical Journal International*, 167(2):495–503.

BIBLIOGRAPHY

- Plessix, R.-E., Baeten, G., de Maag, J. W., and ten Kroode, F. (2012). Full waveform inversion and distance separated simultaneous sweeping: a study with a land seismic data set. *Geophysical Prospecting*, 60:733 – 747.
- Plessix, R. E. and Cao, Q. (2011). A parametrization study for surface seismic full waveform inversion in an acoustic vertical transversely isotropic medium. *Geophysical Journal International*, 185:539–556.
- Plessix, R. E. and Perkins, C. (2010). Full waveform inversion of a deep water ocean bottom seismometer dataset. *First Break*, 28:71–78.
- Polyanin, A. D. and Nazaikinskii, V. E. (2002). *Handbook of Linear Partial Differential Equations for Engineers and Scientists*. CRC Press.
- Pratt, R. G. (1999). Seismic waveform inversion in the frequency domain, part I : theory and verification in a physical scale model. *Geophysics*, 64:888–901.
- Pratt, R. G., Shin, C., and Hicks, G. J. (1998). Gauss-Newton and full Newton methods in frequency-space seismic waveform inversion. *Geophysical Journal International*, 133:341–362.
- Ravaut, C., Operto, S., Improta, L., Virieux, J., Herrero, A., and dell’Aversana, P. (2004). Multi-scale imaging of complex structures from multi-fold wide-aperture seismic data by frequency-domain full-wavefield inversions: application to a thrust belt. *Geophysical Journal International*, 159:1032–1056.
- Regone, C., Stefani, J., Wang, P., Gereá, C., Gonzalez, G., and Oristaglio, M. (2017). Geologic model building in SEAM Phase II - Land seismic challenges. *The Leading Edge*, 36(9):738–749.
- Remaki, M. (2000). A new finite volume scheme for solving maxwell’s system. *International Journal for Computation and Mathematics in Electrical and Electronic Engineering*, 19(3):913–932.
- Robertsson, J., Blanch, J., and Symes, W. (1994). Viscoelastic finite-difference modeling. *Geophysics*, 59:1444–1456.
- Robertsson, J., Moore, I., Vassallo, M., Özdemir, K., van Manen, D. J., and Özbek, A. (2008). On the use of multicomponent streamer recordings for reconstruction of pressure wavefields in the crossline direction. *Geophysics*, 73(5):A45–A49.
- Robertsson, J. O. A. (1996). A numerical free-surface condition for elastic/viscoelastic finite-difference modeling in the presence of topography. *Geophysics*, 61:1921–1934.
- Saad, Y. (2003). *Iterative Methods for Sparse Linear Systems*. SIAM, Philadelphia.
- Schneider, W. A. (1978). Integral formulation for migration in two and three dimensions. *Geophysics*, 43:49–76.
- Sen, M. K. and Stoffa, P. L. (1995). *Global Optimization Methods in Geophysical Inversion*. Elsevier Science Publishing Co.
- Sheriff, R. E. (2002). Encyclopedic dictionary of applied geophysics. In Scherrer, E. F., editor, *geophysical References Series*. Society of Exploration Geophysicists.

- Shin, C. and Cha, Y. H. (2008). Waveform inversion in the Laplace domain. *Geophysical Journal International*, 173(3):922–931.
- Shin, C., Jang, S., and Min, D. J. (2001). Improved amplitude preservation for prestack depth migration by inverse scattering theory. *Geophysical Prospecting*, 49:592–606.
- Shragge, J. (2016). Acoustic wave propagation in tilted transversely isotropic media: Incorporating topography. *GEOPHYSICS*, 81(5):C265–C278.
- Sirgue, L., Barkved, O. I., Dellinger, J., Etgen, J., Albertin, U., and Kommedal, J. H. (2010). Full waveform inversion: the next leap forward in imaging at Valhall. *First Break*, 28:65–70.
- Sirgue, L. and Pratt, R. G. (2004). Efficient waveform inversion and imaging : a strategy for selecting temporal frequencies. *Geophysics*, 69(1):231–248.
- Sjögreen, B. and Petersson, N. A. (2012). A fourth order accurate finite difference scheme for the elastic wave equation in second order formulation. *Journal of Scientific Computing*, 52(1):17–48.
- Socco, L. V., Foti, S., and Boiero, D. (2010). Surface-wave analysis for building near-surface velocity models — established approaches and new perspectives. *Geophysics*, 75(5):75A83–75A102.
- Stork, C. (1992). Reflection tomography in the post migrated domain. *Geophysics*, 57:680–692.
- Tape, C., Liu, Q., Maggi, A., and Tromp, J. (2010). Seismic tomography of the southern California crust based on spectral-element and adjoint methods. *Geophysical Journal International*, 180:433–462.
- Tarantola, A. (1984). Inversion of seismic reflection data in the acoustic approximation. *Geophysics*, 49(8):1259–1266.
- Tarantola, A. (1987). *Inverse problem theory: methods for data fitting and model parameter estimation*. Elsevier, New York.
- Tarantola, A. (1988). Theoretical background for the inversion of seismic waveforms including elasticity and attenuation. *Pure and Applied Geophysics*, 128:365–399.
- Tarantola, A. (2005). *Inverse Problem Theory and Methods for Model Parameter Estimation*. Society for Industrial and Applied Mathematics, Philadelphia.
- Tarrass, I., Giraud, L., and Thore, P. (2011). New curvilinear scheme for elastic wave propagation in presence of curved topography. *Geophysical Prospecting*, 59(5):889–906.
- Trinh, P. T., Brossier, R., Métivier, L., Tvard, L., and Virieux, J. (2017a). Efficient 3D elastic FWI using a spectral-element method. In *87th SEG Conference and Exhibition 2017, Houston*, pages 1533–1538.
- Trinh, P. T., Brossier, R., Métivier, L., and Virieux, J. (2018). Data-oriented strategy and V_p/V_s model-constraint for simultaneous V_p and V_s reconstruction in 3D viscoelastic FWI: Application to the SEAM ii foothills dataset. In *88th SEG Conference and Exhibition 2018, Anaheim*.
- Trinh, P. T., Brossier, R., Métivier, L., Virieux, J., and Wellington, P. (2017b). Bessel smoothing filter for spectral element mesh. *Geophysical Journal International*, 209(3):1489–1512.

BIBLIOGRAPHY

- Trinh, P. T., Brossier, R., Métivier, L., Virieux, J., and Wellington, P. (2017c). Structure-smoothing Bessel filter for finite element mesh: Application on 3D elastic FWI. In *79th EAGE Conference and Exhibition 2017, Paris*.
- Tromp, J., Tape, C., and Liu, Q. (2005). Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels. *Geophysical Journal International*, 160:195–216.
- Valenciano, A. A., Biondi, B. L., and Clapp, R. G. (2009). Imaging by target-oriented wave-equation inversion. *Geophysics*, 74(6):WCA109–WCA120.
- van Driel, M. and Nissen-Meyer, T. (2014). Optimized viscoelastic wave propagation for weakly dissipative media. *Geophysical Journal International*, 199(2):1078–1093.
- van Leeuwen, T. and Mulder, W. (2008). Velocity analysis based on data correlation. *Geophysical Prospecting*, 56(6):791–803.
- Van Vliet, L. J., Young, I. T., and Verbeek, P. W. (1998). Recursive gaussian derivative filters. In *Pattern Recognition, 1998. Proceedings. Fourteenth International Conference on*, volume 1, pages 509–514. IEEE.
- Vigh, D., Jiao, K., Watts, D., and Sun, D. (2014). Elastic full-waveform inversion application using multicomponent measurements of seismic data collection. *Geophysics*, 79(2):R63–R77.
- Vigh, D., Moldoveanu, N., Jiao, K., Huang, W., and Kapoor, J. (2013). Ultralong-offset data acquisition can complement full-waveform inversion and lead to improved subsalt imaging. *The Leading Edge*, 32(9):1116–1122.
- Vilotte, J., Festa, G., and Madariaga, R. (2005). Spectral element simulations of rupture dynamics along kinked faults. In *EOS Transactions*, volume 86. American Geophysical Union, San Francisco, USA.
- Virieux, J., Asnaashari, A., Brossier, R., Métivier, L., Ribodetti, A., and Zhou, W. (2017). An introduction to Full Waveform Inversion. In Grechka, V. and Wapenaar, K., editors, *Encyclopedia of Exploration Geophysics*, pages R1–1–R1–40. Society of Exploration Geophysics.
- Virieux, J. and Operto, S. (2009). An overview of full waveform inversion in exploration geophysics. *Geophysics*, 74(6):WCC1–WCC26.
- Walsh, J. M. and Regalia, P. A. (2010). Belief Propagation, Dykstra’s Algorithm, and Iterated information Projections. *IEEE Transactions on Information Theory*, 56:4114–4128.
- Wang, Y. (2009). *Seismic inverse Q filtering*. John Wiley & Sons.
- Warner, M. and Guasch, L. (2016). Adaptive waveform inversion: Theory. *GEOPHYSICS*, 81(6):R429–R445.
- Warner, M., Ratcliffe, A., Nangoo, T., Morgan, J., Umpleby, A., Shah, N., Vinje, V., Stekl, I., Guasch, L., Win, C., Conroy, G., and Bertrand, A. (2013). Anisotropic 3D full-waveform inversion. *Geophysics*, 78(2):R59–R80.
- Wellington, P. (2016). *Efficient 1D, 2D and 3D Geostatistical constraints and their application to Full Waveform Inversion*. PhD thesis, Univ. Grenoble Alpes.

- Wellington, P., Brossier, R., Hamitou, O., Trinh, P., and Virieux, J. (2017). Efficient anisotropic dip filtering via inverse correlation functions. *Geophysics*, 82(4):A31–A35.
- Whitmore, N. (1983). Iterative depth migration by backward time propagation. In *53th Annual Meeting*. Society of Exploration Geophysics.
- Williamson, P., Atle, A., Fei, W., Hale, D., et al. (2011). Regularization of wave-equation migration velocity analysis by structure-oriented smoothing. In *2011 SEG Annual Meeting*. Society of Exploration Geophysicists.
- Wolfe, P. (1969). Convergence conditions for ascent methods. *SIAM Review*, 11.
- Wu, R.-S., Luo, J., and Wu, B. (2014). Seismic envelope inversion and modulation signal model. *Geophysics*, 79(3):WA13–WA24.
- Yang, P., Brossier, R., Métivier, L., and Virieux, J. (2016a). A review on the systematic formulation of 3D multiparameter full waveform inversion in viscoelastic medium. *Geophysical Journal International*, 207(1):129–149.
- Yang, P., Brossier, R., Métivier, L., and Virieux, J. (2016b). Wavefield reconstruction in attenuating media: A checkpointing-assisted reverse-forward simulation method. *Geophysics*, 81(6):R349–R362.
- Yang, P., Brossier, R., and Virieux, J. (2016c). Wavefield reconstruction from significantly decimated boundaries. *Geophysics*, 80(5):T197–T209.
- Yilmaz, O. (2001). *Seismic data analysis*. Society of Exploration Geophysicists: processing, inversion and interpretation of seismic data.
- Yilmaz, O. (2016). The near-surface problem in exploration seismology. In *Expanded Abstracts, 78th Annual EAGE Meeting (Vienna)*.
- Zhu, H., Bozdağ, E., and Tromp, J. (2015). Seismic structure of the European upper mantle based on adjoint tomography. *Geophysical Journal International*, 201(1):18–52.
- Zienkiewicz, O. C., Taylor, R. L., and Zhu, J. Z. (2005). *The Finite Element Method: Its Basis and Fundamentals*. Elsevier, London. 6th edition.