Lecture notes: High resolution geophysical imaging using full waveform modeling and inversion

L. Métivier

November 20, 2025

Contents

Foreword

The aim of this course is to provide a first understanding of geophysical imaging and the mathematical concepts behind. The core of the course is the full waveform inversion method, a high resolution seismic imaging method which has been developed since the beginning of the 1980s and is now applied routinely in the industry for oil & gas exploration and in the academy for global and regional scale imaging. Despite full waveform inversion is relatively widespread and used in diverse communities, a significant research work is still ongoing around this method, both in terms of methodology and applications. The aim of this course is also to give an insight of some of these research directions.

Chapter 1

Introduction

In the introduction, we first review in which domains obtaining accurate and quantitative estimates of the subsurface mechanical properties is crucial. We then explain how this can be based on the propagation of seismic waves in the Earth's interior, and how this can be formulated as a typical inverse problem. Finally, we show an example of synthetic data recorded in a typical controlled source experiment. We explain the main principle of tomography methods, which exploits only the first arrival travel-times of the waves and hence have a low resolution power. This is the motivation for full waveform inversion, a high resolution seismic imaging technique which aims at exploiting all the data.

1.1 High resolution imaging of the subsurface: why do we need it?

Let us start by reviewing some applications where a high resolution information on the Earth's structure is crucial.

1.1.1 Earth's structure at global and continental scales

Understanding the global structure of the Earth is still at the heart of active research (Fig.1.1). Questions encompass the early formation of planets related to the accretion mechanism. Having more precise ideas of the core structure is also crucial to better understand and interpret the magnetic field. The magnetic field is generated by rotation of the liquid iron around a solid grain in the core. Delineating the core/mantle boundary is also a salient question for seismologists, as it likely plays an important role in geodynamic and geothermal processes. For instance it is said in Lowrie and Ficthner (2020) that

> "The layer at the core mantle boundary may serve a the source of material for mantle plumes that give rise to hot spots, which are important in plate tectonics. The thermal properties of this layer might also influence the outward transport of heat from the Earth's core; in turn this could affect the intricate processes that generate the Earth's magnetic field."

A high resolution information at the global scale is also crucial for the study of earthquakes. Locating hypocenters of earthquakes is a major issue and it requires an accurate knowledge of the seismic wave velocity at global and continental scales.

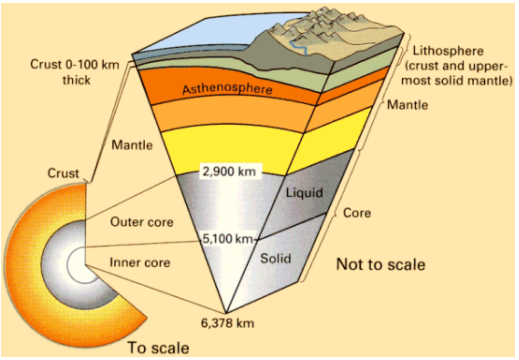

Figure 1.1: Sketch of the internal structure of the Earth at the global scale. Inner (solid) and outer (liquid) core, mantle, and crust.

1.1.2 Imaging active zones: crustal scale imaging

Better imaging the structure of particular active zones such as subduction zones, volcanic areas, faults, is also crucial to better understand their mechanisms, and to prevent the potential hazards related to their activity. An example of such zone is the Nankai trough near Japan (Fig.1.2). The underlying fault is the origin of Nankai megathrust earthquakes which devastate Japan regularly. The Eastern segment of the Nankai Trough is known to have been left without rupture during the most recent earthquakes dating back to 1944 and 1946.

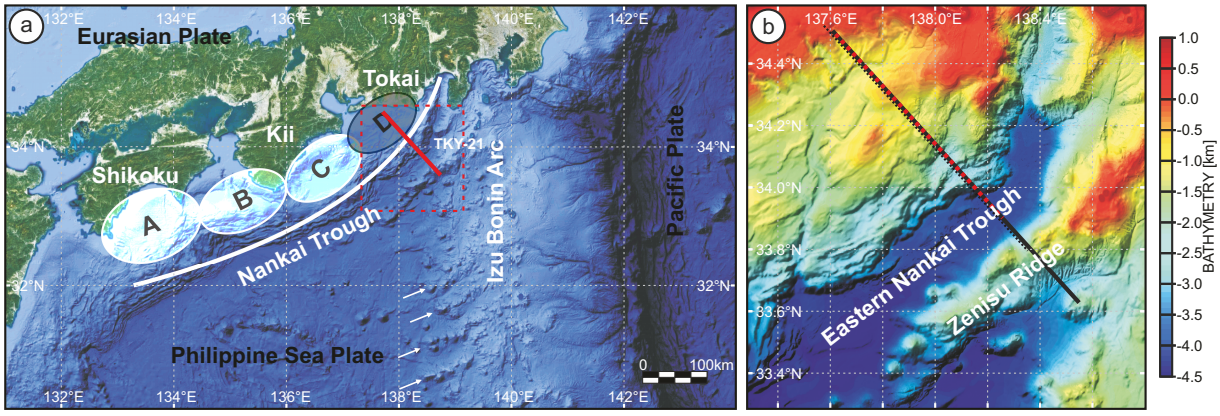

Figure 1.2: Partitioning of the Nankai Trough into four segments as described by Ando (1975). Region D was left unruptured during the most recent sequence of two large earthquakes (1944 Tonankai and 1946 Nankaido). The solid red line represents the seismic profile of the SFJ-OBS experiment. White arrows mark the cyclic volcanic ridges developing on the flank of Izu Bonin Arc. (b) Zoomed view of the survey area, overlaid with the bathymetry variations. The solid grey line and the dashed red line represent the OBS shot and receiver profile respectively, while the dashed black line corresponds to the MCS profile (Górszczyk et al., 2017). Figure taken from Gorszczyk et al. (2019).

A better information on the subsurface structure in this area is crucial to help geophysicists interpret the geology and try to infer the risk related to an earthquake in this zone (next occurrence localization, strength, effect etc...). Recent imaging results using

full waveform inversion (FWI) and their interpretation are presented in Figure 1.3. Note that the scale here is 20 km depth for 100 km long 2D model. This means that the target is deep, but still within the crust: we speak of deep crustal scale imaging. This example is taken from Gorszczyk et al. (2019).

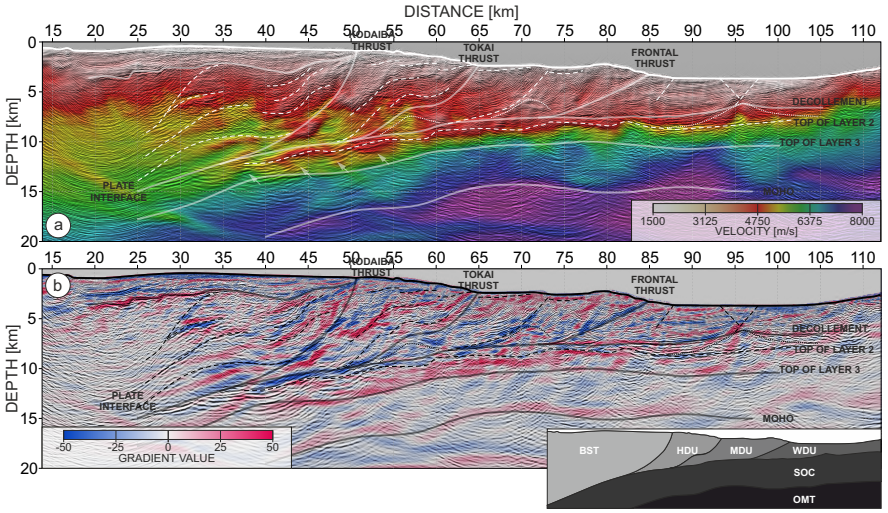

Figure 1.3: Imaging results and structural interpretation. Migrated section superimposed on: (a) FWI velocity model blended with average vertical and horizontal velocity gradient; (b) Average vertical and horizontal velocity gradient of FWI model. Inset presents different geological block-segments: **OMT** - oceanic mantle; **SOC** - subducting oceanic crust; **WDU** - weakly deformed unit; **MDU** - moderately deformed unit; **HDU** - heavily deformed unit; **BST** - backstop. Figure taken from Gorszczyk et al. (2019).

1.1.3 Hydrocarbon resources: crustal scale imaging

Precise, accurate, high resolution information on the structure is crucial for oil & gas exploration. Locating hydrocarbon resources can request resolution information down to few meters for targets few kilometers deep in area covering hundreds of square kilometers. Drilling operations are extremely complicated (and expensive) despite progresses in the technology. Uncertainty in the structure can lead to dry wells and/or hazards, in particular in presence of gas.

In Figure 1.4 we present an example of 3D imaging from North Sea data (Valhall oil field) where a gas cloud, trapped in sediment layers, make the imaging of the reservoir below challenging. High resolution imaging makes it possible to correctly image the gas cloud at 1 km depth (gray depth slices). Shallower, glacial sand channels deposits at are made visible at $z = 175$ m, a low-velocity zone intersected by scrapes left by drifting icebergs on the paleo-seafloor is made apparent at $z = 500$ m depth. This example is taken from Operto et al. (2015).

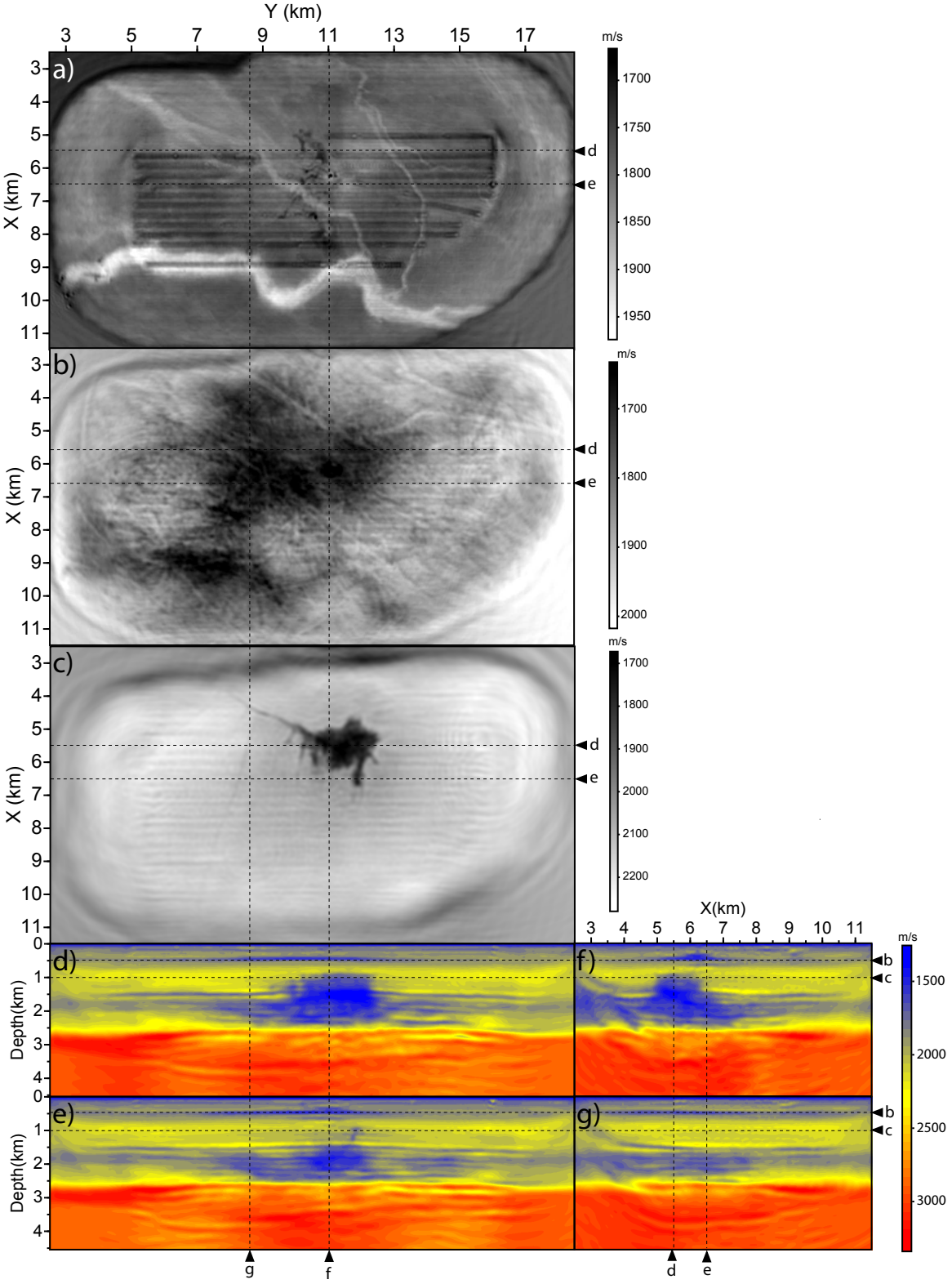

Figure 1.4: Slices of the 10 Hz FWI model.) (a-c) Horizontal slices at (a) 175m depth, (b) 500m depth, (c) 1km depth across the gas cloud. (d-e) Inline vertical slices (d) passing through the gas cloud ($X = 5.6km$) and (e) near its periphery ($X = 6.25km$). (f-g) Cross-line vertical slices at (f) $Y = 8.6km$ and (g) $Y = 11km$. Figure taken from Operto et al. (2015).

1.1.4 CO_2 storage

CO_2 storage is an important challenge industry is facing in the frame of the energy transition. Reducing greenhouse gas emissions is now accepted as one of the main issue for modern societies. Part of the solution is sought in the ability to capt directly CO2 at the source (from industrial plants), and inject it in the subsurface in a controlled environment. This requires again precise, accurate information on the subsurface structure, to evaluate permeability of the layers in which the pressured CO_2 will be trapped. This also requires the ability to monitor the state of the storage along time. This yields a specific problematic of monitoring, also known as 4D imaging. Using a permanent device of acquisition, 4D imaging consists in generating 3D images of the subsurface along time to control the evolution of the subsurface and detect potential leakage. Note that part of the industrial tools required to develop and pilot CO_2 storage are in the hand of the oil & gas industry, as they have developed tools to inject fluids in the subsurface to enhance oil extraction and to monitor reservoir production through permanent seismic acquisition devices.

We present in Figure 1.5 a schematic view of the CO_2 storage injection system which has been installed in Sleipner (Norway) since 1996. Sleipner is a pilot site, which has made possible to test and evaluate the interest of CO_2 storage technology. A review of the experiments and research led in this pilot site is proposed in Eiken (2019)

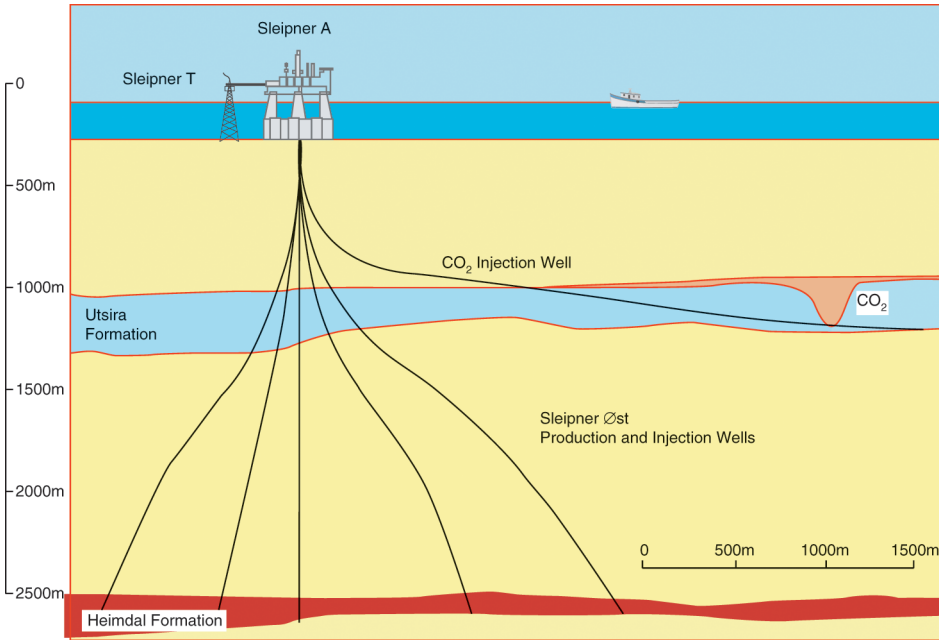

Figure 1.5: Schematic view of the CO_2 storage injection in Sleipner (Norway).

1.1.5 Geothermal energy

Geothermal energy is another technology recently developed to lessen our dependence to hydrocarbon resources and reduce greenhouse gas emissions. Shallow geothermal resources (down to 200 m depth) can be exploited to monitor the buildings temperature (both for individuals and industry), or to produce hot water. Deep geothermal resources (below 200 m depth) can be used for the same purpose or to produce electricity. In both cases (shallow and deep geothermal resources exploitation) an accurate knowledge of the subsurface

structure is required, to detect proper aquifers, and to conduct the drilling and installation of pumps and/or electricity generators. A monitoring of the installation is also required, leading, as for CO_2 storage, to the problematic of 4D imaging.

1.1.6 Geotechnical engineering

From wikipedia `https://en.wikipedia.org/wiki/Geotechnical_engineering`, the definition of geotechnical engineering (or geotechnics)

> Geotechnical engineering, also known as geotechnics, is the application of scientific methods and engineering principles to the acquisition, interpretation, and use of knowledge of materials of the Earth's crust and earth materials for the solution of engineering problems and the design of engineering works. It is the applied science of predicting the behavior of the Earth, its various materials and processes towards making the Earth more suitable for human activities and development (...).
>
> Examples of the application of geotechnics include: the prediction, prevention or mitigation of damages caused by natural hazards such as avalanches, mud flows, landslides, rockslides, sinkholes, and volcanic eruptions; the application of soil, rock and groundwater mechanics to the design and predicted performance of earthen structures such as dams; the design and performance prediction of the foundations of bridges, buildings, and other man-made structures in terms of the underlying soil and/or rock; and flood control and prediction.

For such applications, a precise, accurate, high resolution knowledge of the subsurface structure is also naturally required.

1.1.7 Archaeology

Finally, recent applications have shown how shallow subsurface imaging can be used in the frame of archaeological studies. Depth of investigation do not exceed few meters in this case. Preparing archaeological campaign requires accurate information on the subsurface structure. Conventional imaging methods based on electromagnetic measurements such as ground penetrating radar (GPR) can be limited by the penetration depth of these waves. In (Köhn et al., 2021), a medieval canal named "Fossa Carolina" located in southern Germany is investigated. Full waveform inversion of elastic wave propagation recordings is used to image the structure prior to excavation, and is compared with a more conventional tomography strategy. After excavation, it is possible to assess the quality of the images. Modern techniques such as full waveform inversion, which will be at the heart of this course, makes it possible to retrieve higher resolution images than the conventional tomography technique based on first-arrival travel time.

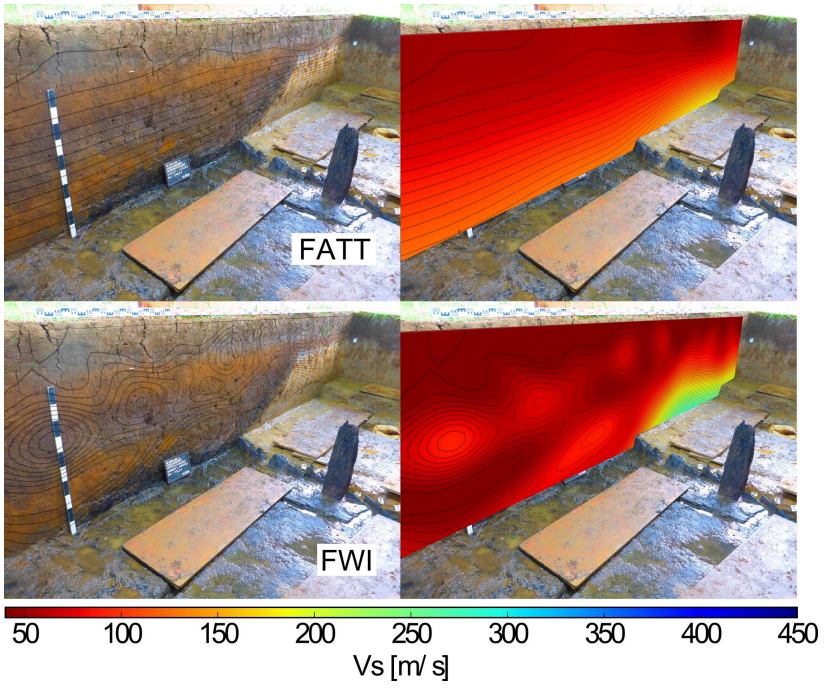

Figure 1.6: Figure extracted from Köhn et al. (2021). FWI stands for Full Waveform Inversion and FATT for First Arrival travel-time tomography. The latter technique produce a low resolution estimate of the shear wave velocity. FWI provides a higher resolution estimate. Thanks to the excavation, the results can be confronted to the true subsurface rheology. The comparison between the photography and the FWI result illustrates how the higher resolution details conform with the reality. In particular, the target (the "Fossa Carolina" canal) is much more accurately delineated in the bottom right corner of the 2D shear-wave velocity map.

1.2 How to obtain information on subsurface material properties?

Here we review the main principle of seismic imaging and its general formulation as an inverse problem. We will make a little digression around the linear case for inverse problems as it illustrates first important concepts in the frame of the solution of inverse problems.

1.2.1 Drilling is not an option

First and foremost, one could wonder why the investigation of the subsurface could not rely only on direct measurements. That is, in a pragmatic way, drilling. There are several arguments against this.

- Some applications require an indirect measurement of the subsurface properties simply because a direct measurement would destroy the target. Archaeology and geotechnical engineering belong to this category of applications.
- At larger scales, drilling provides a very local information. The subsurface, especially the crust, can not be accurately represented as a layered medium. Therefore, the direct measurement provided by a (vertical) well do not provide information on potential lateral variations of the subsurface.
- Drilling is a highly technical and complex operation. Therefore it is expensive.
- For regional scale and global scale imaging, the depth of investigation of a drilling operation is far from being sufficient.

Regarding the last point, it is interesting to know that the deepest drilling operation performed on Earth was achieved by Russians. The operation last for 19 years from 1970 to 1989, on the Kola Peninsula (Fig.1.7). They reached the depth of 12,262 km. At this depth, they had to face pressure and temperature conditions that made drilling deeper impossible, the rock behaving more as a plastic than a solid. Given the Earth's radius is approximately 6400 km, the deepest drilling ever achieved reaches less than 0.2% of the Earth's radius. To give another order of idea, the crust has an average thickness of 33 km, but shows strong variations, measuring as little as 5 km below oceans and as much as 60-80 km under some mountain ranges.

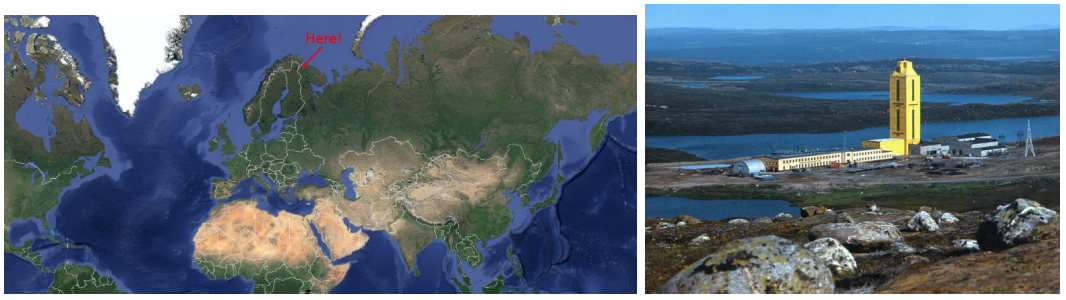

Figure 1.7: Localization of the Kola peninsula on Google Earth (left). Picture of the Kola Superdeep Borehole drilling site (right).

1.2.2 Remote sensing as a remedy

As direct measurement is not an option, we need to rely on indirect measurements. This already means that instead of having an accurate knowledge on the subsurface rheology (type of rock, permeability, water saturation), we will recover averaged properties of the subsurface.

The starting point of geophysical imaging is the following: **subsurface rheology has an impact on the propagation of waves**.

- These can be electromagnetic waves. In this case, the subsurface structure/rheology variations affect the mean permittivity and conductivity of the subsurface, which have an effect on the propagation of electromagnetic waves.

- These can be mechanical (elastic) waves. In this case, the subsurface structure/rheology variations affect the velocity, density, anisotropy, and attenuation of the subsurface, which have an effect on the propagation of mechanical (elastic) waves.

From the observation of mechanical or electromagnetic waves, we will infer information on these "average" subsurface properties. In this course, we will focus on mechanical waves and leave aside electromagnetic waves. The latter are however used for shallow applications, such as archeology and geotechnical engineering, to complement or replace seismic imaging methods. They are also used at the exploration scale, using controlled-source electo-magnetic methods (CSEM), to retrieve the subsurface resistivity down to several kilometers depth. The electrical resistivity can help to discriminate between different rock types. The resolution of such information is however relatively low.

The main idea behind geophysical imaging using mechanical waves is thus that from the recording of these waves, we can infer information about the subsurface mechanical properties (macro-scale parameters). From them, more precise information on the rheology of the subsurface can be obtained, if required, in a second stage (micro-scale parameter). This second stage is referred to as downscaling in the literature. Geophysical imaging is focused on the first stage: from seismic waves observation, recover information on the subsurface mechanical properties.

As such, geophysical imaging belongs to the class of inverse problems. The computation of seismic waves given the subsurface mechanical properties corresponds to a forward problem. It can be formalized (as described in the sequel of the course) as the solution of partial differential equations describing the propagation of elastic waves: the elastodynamics equations. The computation of the subsurface mechanical properties given the local observation of seismic waves is an inverse problem.

1.3 Seismic data

We have said that we would image Earth's structure through remote sensing/use of seismic waves. What is actually the data that we have? How does it look like?

1.3.1 Type of recordings

The data we have correspond to the recording of seismic waves by seismic sensors. These sensors can be located on ground or in the water. On ground, they are geophones. These devices record the amplitude of the local *velocity displacement* of the ground. In the water, they are hydrophones. These devices record the amplitude of the local *pressure field*.

Sensors can be deployed permanently (or for a long period of time, from months to several years). This makes it possible to record seismic activity related to earthquakes but also ambient noise due to human activities and also ocean or atmospheric forcing. Ambient noise can be treated so as to yield useful information on the subsurface. This is a whole research topic which will not be developed in this course. Interested readers might refer to Nakata et al. (2019) and references therein for an overview.

In seismology, signal from earthquakes is used to image the Earth at global or continental scales. A sketch of such acquisition is given in Figure 1.8. In exploration, sensors are deployed temporarily over a given target. Human controlled seismic sources are used to generate the seismic signal. In the water, pressure airguns are deployed (Fig. 1.9 left). On ground, vibrating trucks or explosive sources are used (Fig. 1.9 right).

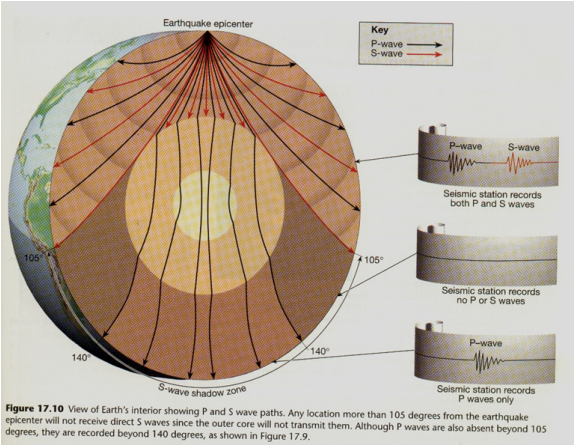

Figure 1.8: Sketch of earthquake recording with sensors deployed all around the Earth.

In terms of mathematics, the seismic data is thus a collection of time functions $d(t)$ associated with a source s and a receiver r. We will denote it as

$$d_{r,s}(t), \tag{1.1}$$

in the following, or equivalently

$$d(x_s, x_r, t), \tag{1.2}$$

or

$$d_s(x_r, t), \tag{1.3}$$

depending on the context. In these notations x_r and x_s denote the spatial position of the receiver r and the source s respectively. A single function $d_{r,s}(t)$ will be referred to as a *seismic trace* in the following.

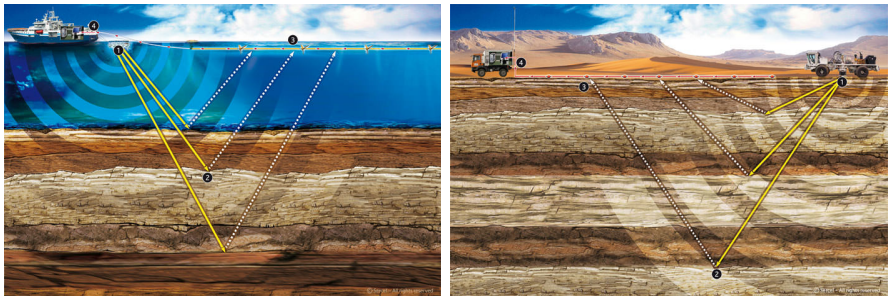

Figure 1.9: Sketch of a controlled source seismic acquisition with airgun sources in the water (left), vibrating truck on ground (right).

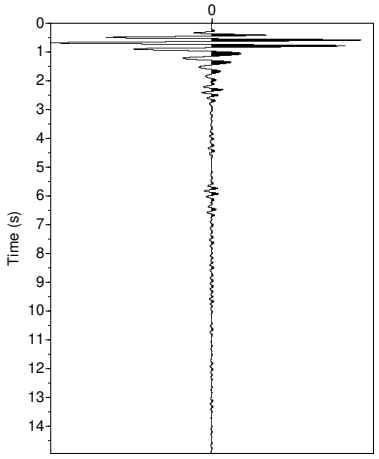

Figure 1.10: Typical seismic trace $d(t)$ as a function of time.

1.3.2 Representation and analysis of seismic data

A typical example of a seismic trace is presented in Figure 1.10. The time axis is presented vertically. The horizontal axis corresponds to the amplitude of the recorded signal (here a pressure recorded by a hydrophone, we will give the detail to which seismic experiment it corresponds in the sequel). We see different seismic arrivals with different amplitude. The first arrival is prominent in terms of energy, but later arrivals can still be clearly identified.

Instead of analyzing the data trace by trace, it can be useful to look simultaneously at several traces. The typical geophysicists' representation is a juxtaposition of traces depending on the receiver/source distance, also referred to as *offset* in the following. Such a collection of 20 traces is presented in Figure 1.11

When the number of traces is even larger, such a representation of the seismic amplitude becomes problematic. Instead of using a 2D plot for each trace, a black & white colorscale is introduced to depict the amplitude. White corresponds to negative values, black to positive values, while gray corresponds to 0. This yields the typical seismogram representation, widely used in exploration geophysics, presented in Figure 1.12. Note that juxtaposing traces depending on the offset makes it possible to visualize laterally coherent events which can be further interpreted in terms of wave propagation.

The seismogram in Figure 1.12 is synthetic (as opposed to field data): it has been computed on a computer. The model and acquisition which have served to this compu-

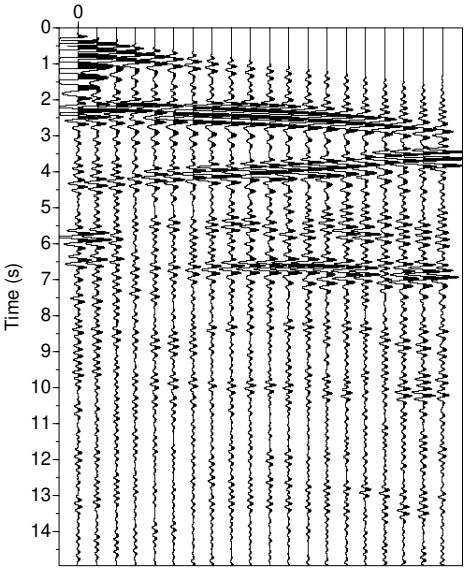

Figure 1.11: 20 seismic traces $d_r(t)$ as a function of time.

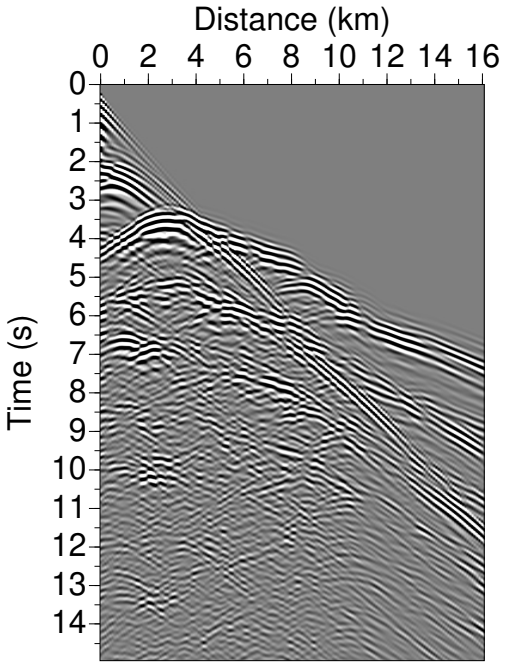

Figure 1.12: A typical seismogram in black and white representation. 161 traces spanning 16 km are used here.

tation are presented in Figure 1.13. A pressure (explosive) source is modeled at the top left corner of the model. The subsurface model which is presented here corresponds to a 2D map of the pressure wave velocity. Blue values mean low velocity values (close to 1500 $\mathrm{m.s}^{-1}$, the pressure wave velocity in the water), red values mean large velocity values (as large as 4500 $\mathrm{m.s}^{-1}$, the pressure wave velocity in rigid bodies such as salt bodies). The receivers are located all along the surface on top of the model. They record the pressure

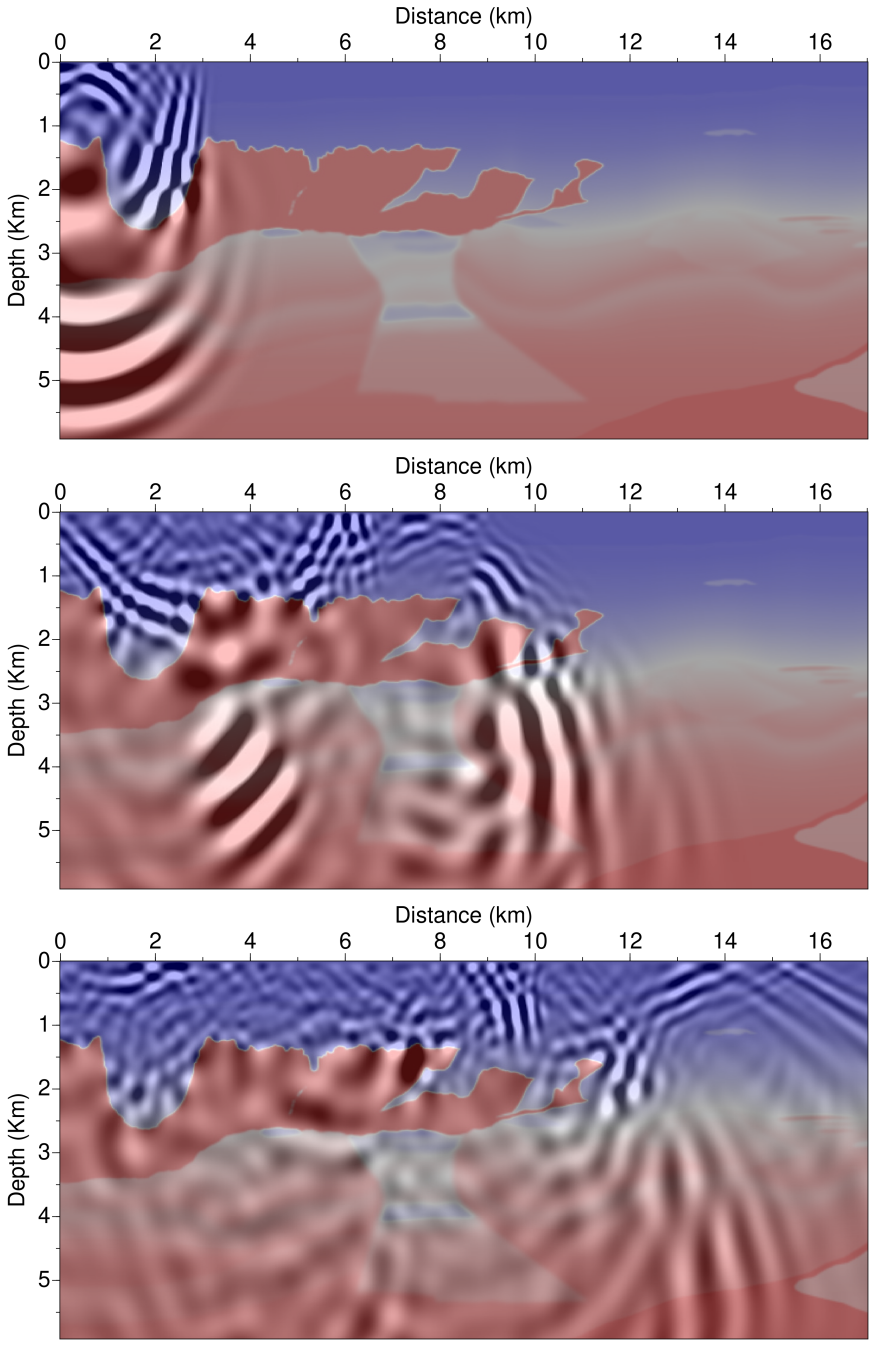

Figure 1.13: Pressure wavefield at $t = 2.3$ s (top), $t = 4.6$ s, (middle) and $t = 6.9$ s (bottom) superimposed on the velocity model used to compute the synthetic data presented in Figure 1.12.

wavefield propagating along time. The pressure field has been computed here by solving 2D acoustic wave propagation equations (to be developed in Chapter 2).

Already at this stage it is interesting to correlate the complexity of the recorded signal with the complexity/heterogeneity of the subsurface model. High contrasts between low velocity and high velocity zones generate strong reflections. The boundary condition implemented on top of the model is a free surface boundary condition which represents the

subsurface/air interface. Waves reaching this boundary are reflected back to the medium. Higher velocity values have also a visible impact on the wavelength of the propagated signal, which becomes larger.

1.4 A first glance at seismic inversion methods

Now that we have seen a first instance of seismic data, we introduce with more details seismic inversion methods based on the interpretation of such data.

1.4.1 Tomography methods: exploiting only the first arrival time

The first method which has been developed in the frame of seismic imaging to exploit seismic data and infer mechanical properties of the Earth is tomography. This methods relies on the exploitation of the first arrival travel-times between sources and receivers. The principle is relatively simple: pick the first-arrival travel-times on the seismograms by detecting the "first break" (the moment where the sensor starts recording the wave traveling the fastest). Manual or automatic picking can used. A rough example of what would be picked on the previous seismogram is presented in Figure 1.14 where the red line indicates picked travel-times.

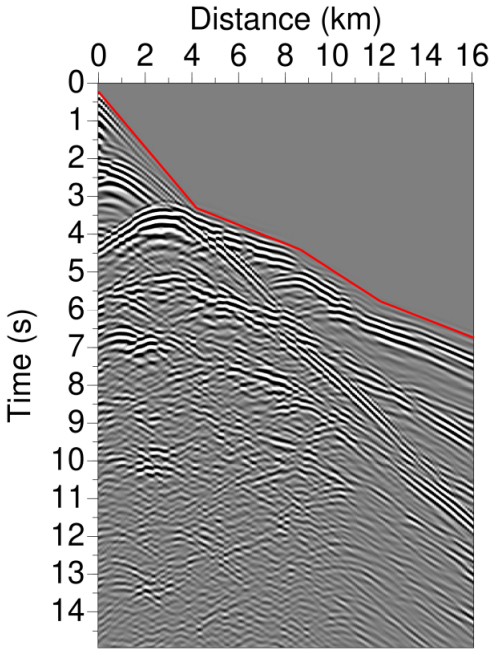

Figure 1.14: Same seismogram as in Figure 1.12 with first-arrival travel time denoted by the red line.

From these picked travel-times, an inverse tomography problem is solved. This inverse tomography problem follows the general formulation of inverse problems we have introduced previously with

$$d_{obs} = t_{obs}(x_s, x_r), \quad m = v_P \tag{1.4}$$

where $t_{obs}(x_s, x_r)$ denotes the picked travel time from source s to receiver r, and v_P is the pressure wave velocity.

The forward operator $g(m)$ consists in this case in solving an eikonal equation: computation of the first-arrival travel-times between sources and receivers. This can be done through ray tracing techniques (see Červený, 2001, for a review) or Eulerian based approach. There is a wide mathematical and geophysical literature on this topic: solving

eikonal equations is not a trivial problem, as it is a nonlinear PDE. Again, it will not be developed in this course. The interested reader might be redirected to (Nolet, 2008) and references therein for a first view.

In the following formulation, we will denote the solution of the eikonal equation by $t_{cal}(x_s, x_r)$. In the notation of the general inverse problem we have introduced, we thus have for the tomography problem

$$d_{obs} = t_{obs}(x_s, x_r), \quad m = v_P, \quad g(v_P) = t_{cal}(x_s, x_r) \tag{1.5}$$

where $t_{cal}(x_s, x_r)$ is computed through the solution of an eikonal equation.

The corresponding least-squares first-arrival tomography problem can thus be written as

$$\min_{v_P} \frac{1}{2}\|t_{cal} - t_{obs}\|^2 + \eta R(v_P), \quad t_{cal} = g(v_P). \tag{1.6}$$

where $R(v_P)$ is a regularization term, for instance

$$R(v_P) = \frac{1}{2}\|v_P - v_{P,p}\|^2 \tag{1.7}$$

where $v_{P,p}$ is a given prior P-wave velocity model.

In (1.6), the norm $\|.\|$ means summation over sources and receivers, that is

$$\|t_{cal} - t_{obs}\|^2 = \sum_{s,r} |t_{cal}(x_s, x_r) - t_{obs}(x_s, x_r)|^2 \tag{1.8}$$

A typical tomography result obtained from 3D field data acquired in the Valhall field (North Sea) is presented in Figure 1.15. Here the method employed is "reflection travel-time" tomography, hence the picked times correspond to reflected waves instead of transmitted waves as is the case with usual first-arrival travel-time tomography. Nonetheless these results illustrate a striking feature of tomography methods: the resolution of the subsurface mechanical parameters estimation is relatively poor. Only smooth variations are recovered. Comparing with results presented in Figure 1.4 one can see that indeed the resolution in the latter image is much higher. What can explain such a difference?

1.4.2 Full waveform inversion: exploiting all the data

Tomography is based on the exploitation of travel-times between sources and receivers (eventually including reflectors). Looking at picture 1.14 makes clear that this operation results in disregarding more than 99% of the data. Based on this observation, researchers at the beginning of the 80s (Lailly, 1983; Tarantola, 1984) have proposed to try to exploit not only the arrival time of selected phases (waves), but the full information, including all the phases, and their relative amplitude. By doing so, much higher resolution estimates can be obtained, as the one presented in Figure 1.4. This high resolution seismic imaging method has been introduced as full waveform inversion (FWI).

The purpose of this course is to detail the mathematical foundation of this seismic imaging technique, as well as to provide a physical understanding of the underlying concepts on which it is based. The basic idea, compared with tomography, is to replace the forward modeling operator $g(m)$ by a full wave modeling solver, and to compare the resulting synthetic data to the full observed data $d_{obs}(x_s, x_r, t)$. The FWI problem is thus formulated as

$$\min_{m} \frac{1}{2}\|d_{cal}[m] - d_{obs}\|^2 + \eta R(m), \quad d_{cal}[m] = g(m). \tag{1.9}$$

A full wave modeling solver will require to solve partial differential equations representing the wave propagation within the subsurface. The solution of the associated nonlinear least-squares problem will require the usage of local optimization strategies and efficient numerical methods to build the gradient. The implementation of the method on field data will require a physical understanding of the imaging concepts underlaid by the method. The purpose of this course is to give an overview of all these elements, as complete as possible, and also to give perspectives on current scientific questions around FWI which still stimulate active research.

The remainder of the course is decomposed in two main chapters. The first is devoted to full wave modeling: how to represent accurately the propagation of seismic waves numerically? The second is devoted to full wave inversion: how to design efficient numerical methods to solve the inverse problem? How to implement it in practice to obtain meaningful results?

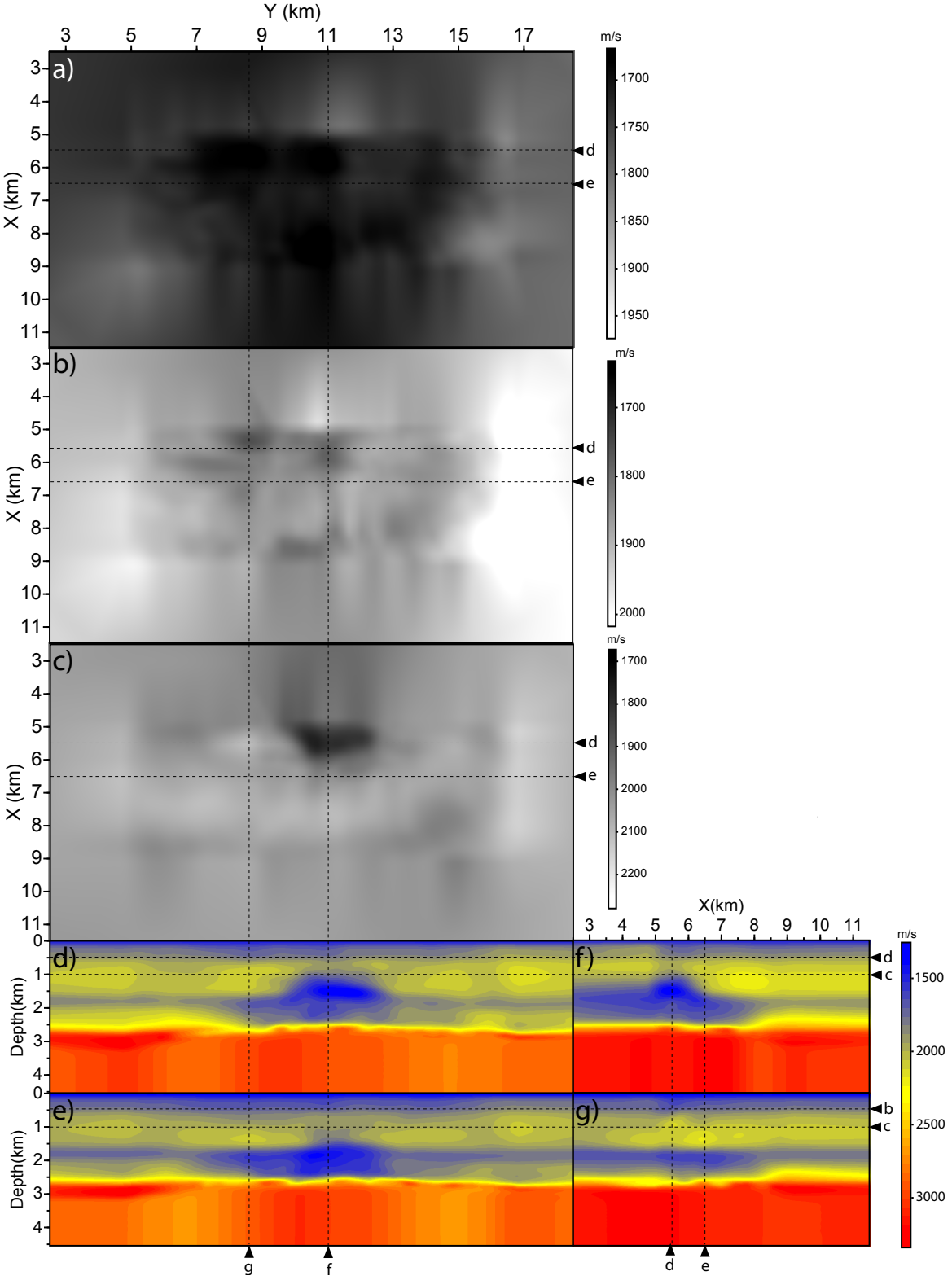

Figure 1.15: Slices of the initial model (vertical wavespeed) built by reflection traveltime tomography (courtesy of BP) (a-c) Horizontal slices at (a) 175m depth, (b) 500m depth, (c) 1km depth across the gas cloud. (d-e) Inline vertical slices (d) passing through the gas cloud ($X = 5.6km$) and (e) near its periphery ($X = 6.25km$). (f-g) Cross-line vertical slices at (f) $Y = 8.6km$ and (g) $Y = 11km$.

Chapter 2

Full waveform modeling

2.1 Derivation of elastodynamics equations

The mathematical model we choose to represent the propagation of seismic waves within the subsurface is the elastodynamics model. In this approximation, the Earth is seen as an elastic medium. Two main principles will be used to derive the elastodynamics equations. First, the Newton law relating forces and acceleration of the medium. Second, the stress strain relationship. The latter will be derived under the frame of small deformation, leading to consider only the linear part of the relation between stress and strain, neglecting all higher order effects. This leads to a linear system of partial differential equations which we now introduce. For a more complete overview on the theory of linear elasticity I suggest the books from Gould (2013), Shearer (2009) and course Aki and Richards (2002). The presentation of the stress and strain tensors and the way the elastodynamics equations are assembled in the following are mainly extracted from Shearer (2009).

2.1.1 The Cauchy stress tensor

Definition

We first introduce the Cauchy stress tensor. Let us consider an infinitesimal plane in an arbitrary direction, of an elastic medium in static equilibrium. The orientation of this plane is given by the unit vector normal to it $\mathbf{n}$. We denote the traction force per unit area exerted on this plane in the direction of $\mathbf{n}$ by $\mathbf{t}(\mathbf{n})$ (Fig. 2.1). Because of the static

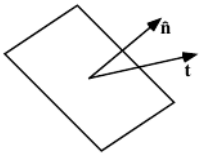

Figure 2.1: Infinitesimal plane with normal $\mathbf{n}$ and traction $\mathbf{t}(\mathbf{n})$ (from Shearer, 2009).

equilibrium we have

$$\mathbf{t}(-\mathbf{n}) = -\mathbf{t}(\mathbf{n}). \tag{2.1}$$

The part of $\mathbf{t}$ which is normal to the plane (aligned with $\mathbf{n}$) is called the normal stress. The part of the traction which is parallel to the plane is the shear stress. In the case of a

fluid, there is no shear stress and we can write

$$\mathbf{t}(\mathbf{n}) = -P\mathbf{n}, \tag{2.2}$$

where P is the pressure (scalar value) and the minus sign is by convention.

The stress tensor σ, in a Cartesian coordinate system, is defined by the traction across the yz, xz, and xy planes of an infinitesimal cube. It writes

$$\begin{pmatrix} t_x(\mathbf{x}) & t_x(\mathbf{y}) & t_x(\mathbf{z}) \\ t_y(\mathbf{x}) & t_y(\mathbf{y}) & t_y(\mathbf{z}) \\ t_z(\mathbf{x}) & t_z(\mathbf{y}) & t_z(\mathbf{z}) \end{pmatrix} = \begin{pmatrix} \sigma_{xx} & \sigma_{xy} & \sigma_{xz} \\ \sigma_{yx} & \sigma_{yy} & \sigma_{yz} \\ \sigma_{zx} & \sigma_{zy} & \sigma_{zz} \end{pmatrix} \tag{2.3}$$

We consider the solid at equilibrium, therefore there cannot be net rotation from the

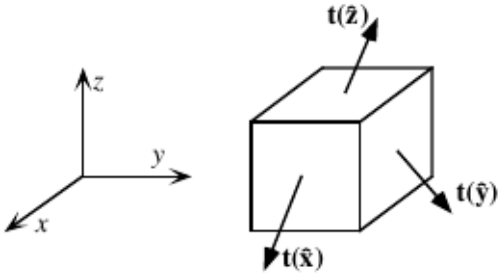

Figure 2.2: The traction vectors $\mathbf{t}(\mathbf{x})$, $\mathbf{t}(\mathbf{y})$ and $\mathbf{t}(\mathbf{z})$ describe the forces on the faces of an infinitesimal cube in a Cartesian coordinate system (from Shearer, 2009).

shear stresses, therefore we have

$$\sigma_{xz} = \sigma_{zx}, \;\; \sigma_{xy} = \sigma_{yx}, \;\; \sigma_{yz} = \sigma_{zy}. \tag{2.4}$$

As a result the stress tensor is symmetric. The traction across any arbitrary plane of orientation $\mathbf{n}$ is obtained using the matrix vector product

$$\mathbf{t}(\mathbf{n}) = \sigma\mathbf{n} = \begin{pmatrix} \sigma_{xx} & \sigma_{xy} & \sigma_{xz} \\ \sigma_{yx} & \sigma_{yy} & \sigma_{yz} \\ \sigma_{zx} & \sigma_{zy} & \sigma_{zz} \end{pmatrix} \begin{pmatrix} n_x \\ n_y \\ n_z \end{pmatrix} = \begin{pmatrix} \sigma_{xx}n_x + \sigma_{xy}n_y + \sigma_{xz}n_z \\ \sigma_{xy}n_x + \sigma_{yy}n_y + \sigma_{yz}n_z \\ \sigma_{xz}n_x + \sigma_{yz}n_y + \sigma_{zz}n_z \end{pmatrix} \tag{2.5}$$

Principal axes and principal stresses

The stress tensor, as a symmetric matrix, can be diagonalized in a proper basis of eigenvectors. This can be interpreted as finding directions $\mathbf{n}$ such that there are no shear stresses across the plane normal to $\mathbf{n}$. In this case, $\mathbf{t}(\mathbf{n})$ is aligned with the direction $\mathbf{n}$, which translates in

$$\mathbf{t}(\mathbf{n}) = \sigma\mathbf{n} = \lambda\mathbf{n}, \tag{2.6}$$

which is equivalent to

$$(\sigma - \lambda\mathbf{I}_3)\,\mathbf{n} = 0, \tag{2.7}$$

where $\mathbf{I}_3$ is the rank-3 identity matrix. We recognize here an eigenvalue problem, which has non-trivial solutions if and only if

$$\det(\sigma - \lambda\mathbf{I}_3) = 0 \tag{2.8}$$

This is a cubic equation with three solutions λ_1, λ_2 and λ_3 which are the eigenvalues of σ which we sort in decreasing order: $\lambda_1 \geq \lambda_2 \geq \lambda_3$. Finding a basis in which the stress tensor is symmetric thus corresponds to finding a coordinate system in which there is no shear stress, only normal stress. The directions of this coordinate system are referred to as principal axes. The eigenvalues $\lambda_i,\ \ i = 1\ldots, 3$ are the principal stresses.

Maximum shear stress

Based on the principal axes, it is possible to express the maximum shear stress depending on the principal stresses. Maximum shear stress appear in planes at 45° to the maximum and minimum principal stress axis. In this system of coordinates, one of these planes has normal vector $\mathbf{n} = \left(1/\sqrt{2}, 0, 1/\sqrt{2}\right)$. We can calculate the traction vector $\mathbf{t}$ across this plane following the same strategy as before, as

$$\mathbf{t} = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \end{pmatrix} = \begin{pmatrix} \lambda_1/\sqrt{2} \\ 0 \\ \lambda_3/\sqrt{2} \end{pmatrix}. \tag{2.9}$$

The traction vector can then be decomposed into normal t_N and shear stresses t_S on this plane by taking a scalar product with $\mathbf{n}$ and a unit vector in the shear stress plane $\mathbf{n}_S = \left(1/\sqrt{2}, 0, -1/\sqrt{2}\right)$. This gives

$$t_N = \frac{\lambda_1 + \lambda_3}{2},\ \ t_S = \frac{\lambda_1 - \lambda_3}{2}. \tag{2.10}$$

The latter value gives us the maximum shear stress. We see from this that if the three eigenvalues are equal $\lambda_1 = \lambda_2 = \lambda_3$ then there is no shear stress and we are in the case of a fluid.

Deviatoric stress

Because in deep Earth, stress is dominated by large compressive stress from hydrostatic pressure, seismologists are used to consider what they call deviatoric stress. This is computed by subtracting the mean normal stress σ_m, given by the average of the principal stresses

$$\sigma_m = \frac{1}{3}\left(\lambda_1 + \lambda_2 + \lambda_3\right) \tag{2.11}$$

Thanks to the trace invariance

$$Tr(P^{-1}AP) = Tr(A),\ \ \forall A \in M_n(\mathbf{R}),\ \ P \in M_n(\mathbf{R})\text{invertible} \tag{2.12}$$

the trace of the stress tensor expressed in any system of coordinates is always equal to the sum of the principal stresses. The deviatoric stress tensor is

$$\sigma_D = \begin{pmatrix} \sigma_{xx} - \sigma_m & \sigma_{xy} & \sigma_{xz} \\ \sigma_{yx} & \sigma_{yy} - \sigma_m & \sigma_{yz} \\ \sigma_{zx} & \sigma_{zy} & \sigma_{zz} - \sigma_m \end{pmatrix} \tag{2.13}$$

and any stress tensor can be decomposed as the sum of its corresponding hydrostatic stress and deviatoric stress tensors

$$\sigma = \sigma_m I + \sigma_D \tag{2.14}$$

For isotropic materials, the hydrostatic stress is responsible for volume changes without changing the shape of the object. For such materials, only the deviatoric stress can cause shape changes.

2.1.2 The strain tensor

Particle displacement, particle velocity, particle acceleration

We consider a reference point $\mathbf{x^0} = (x_1^0, x_2^0, x_3^0) \in \mathbb{R}^3$. We consider a Lagrangian approach and describe how a particle moves from this reference point at time t_0 to another position $\mathbf{x} = (x_1, x_2, x_3) \in \mathbb{R}^3$ at time $t \in \mathbb{R}^+$. This makes it possible to define for this particle, its displacement, velocity, and acceleration vectors as

$$\begin{cases} \mathbf{u}(\mathbf{x},t) = (\mathbf{x}(t) - \mathbf{x}_0) \in \mathbb{R}^3 \\ \mathbf{v}(\mathbf{x},t) = \dfrac{\partial \mathbf{u}}{\partial t}(\mathbf{x},t) \in \mathbb{R}^3 \\ \mathbf{a}(\mathbf{x},t) = \dfrac{\partial \mathbf{v}}{\partial t}(\mathbf{x},t) \in \mathbb{R}^3. \end{cases} \tag{2.15}$$

Definition of the strain tensor

We can expand the displacement field $\mathbf{u}$ by its Taylor expansion, as

$$\mathbf{u}(\mathbf{x},t) = \mathbf{u}(\mathbf{x}_0,t) + \mathbf{J}(\mathbf{x},t)(\mathbf{x}-\mathbf{x}_0,t) + O\|(\mathbf{x}-\mathbf{x}_0)\|^2 \tag{2.16}$$

or in developed form

$$\underbrace{\begin{pmatrix} u_x(\mathbf{x},t) \\ u_y(\mathbf{x},t) \\ u_z(\mathbf{x},t) \end{pmatrix}}_{\mathbf{u}(\mathbf{x},t)} = \underbrace{\begin{pmatrix} \frac{\partial u_x}{\partial x} & \frac{\partial u_x}{\partial y} & \frac{\partial u_x}{\partial z} \\ \frac{\partial u_y}{\partial x} & \frac{\partial u_y}{\partial y} & \frac{\partial u_y}{\partial z} \\ \frac{\partial u_z}{\partial x} & \frac{\partial u_z}{\partial y} & \frac{\partial u_z}{\partial z} \end{pmatrix}}_{\mathbf{J}(\mathbf{x},t)} \underbrace{\begin{pmatrix} x-x_0 \\ y-y_0 \\ z-z_0 \end{pmatrix}}_{\mathbf{x}-\mathbf{x}_0,t} + O\|(\mathbf{x}-\mathbf{x}_0)\|^2 \tag{2.17}$$

In the approximation of infinitesimal strain theory, which is used to derive the theory of linear elasticity we explore here, we keep only the first-order term in this Taylor expansion. This approximation is valid in seismology and seismic imaging to study waves propagating within the Earth. Notable exceptions exist however, for instance to study large magnitude earthquakes in the vicinity of their sources.

The Jacobian matrix J can be decomposed into its symmetric and anti-symmetric part. We have

$$\mathbf{J} = \frac{1}{2}\left(\mathbf{J} + \mathbf{J}^T\right) + \frac{1}{2}\left(\mathbf{J} - \mathbf{J}^T\right) = \epsilon + \mathbf{\Omega} \tag{2.18}$$

with

$$\epsilon = \left(\mathbf{J} + \mathbf{J}^T\right) = \begin{pmatrix} \frac{\partial u_x}{\partial x} & \frac{1}{2}\left(\frac{\partial u_x}{\partial y} + \frac{\partial u_y}{\partial x}\right) & \frac{1}{2}\left(\frac{\partial u_x}{\partial z} + \frac{\partial u_z}{\partial x}\right) \\ \frac{1}{2}\left(\frac{\partial u_y}{\partial x} + \frac{\partial u_x}{\partial y}\right) & \frac{\partial u_y}{\partial y} & \frac{1}{2}\left(\frac{\partial u_y}{\partial z} + \frac{\partial u_z}{\partial y}\right) \\ \frac{1}{2}\left(\frac{\partial u_z}{\partial x} + \frac{\partial u_x}{\partial z}\right) & \frac{1}{2}\left(\frac{\partial u_z}{\partial y} + \frac{\partial u_y}{\partial z}\right) & \frac{\partial u_z}{\partial z} \end{pmatrix} \tag{2.19}$$

and

$$\mathbf{\Omega} = \left(\mathbf{J} - \mathbf{J}^T\right) = \begin{pmatrix} 0 & \frac{1}{2}\left(\frac{\partial u_x}{\partial y} - \frac{\partial u_y}{\partial x}\right) & \frac{1}{2}\left(\frac{\partial u_x}{\partial z} - \frac{\partial u_z}{\partial x}\right) \\ \frac{1}{2}\left(\frac{\partial u_y}{\partial x} - \frac{\partial u_x}{\partial y}\right) & 0 & \frac{1}{2}\left(\frac{\partial u_y}{\partial z} - \frac{\partial u_z}{\partial y}\right) \\ \frac{1}{2}\left(\frac{\partial u_z}{\partial x} - \frac{\partial u_x}{\partial z}\right) & \frac{1}{2}\left(\frac{\partial u_z}{\partial y} - \frac{\partial u_y}{\partial z}\right) & 0 \end{pmatrix} \tag{2.20}$$

The strain tensor ϵ is defined as the symmetric part of the Jacobian matrix J.

Interpretation

The effect of ϵ and $\mathbf{\Omega}$ might be illustrated by considering what happens to an infinitesimal cube. The off-diagonal elements of ϵ cause shear strain. Indeed, consider for instance a 2D case described in (x, z) coordinates. Assume $\mathbf{\Omega} = 0$. This implies $\frac{\partial u_z}{\partial x} = \frac{\partial u_x}{\partial z}$. Assume in addition that $\frac{\partial u_x}{\partial x} = 0$ and $\frac{\partial u_z}{\partial z} = 0$. In this case

$$\mathbf{J} = \epsilon = \begin{pmatrix} 0 & \frac{\partial u_x}{\partial z} \\ \frac{\partial u_z}{\partial x} & 0 \end{pmatrix} = \begin{pmatrix} 0 & \theta \\ \theta & 0 \end{pmatrix} \tag{2.21}$$

where we introduce $\theta = \frac{\partial u_x}{\partial z}$. One can verify on Figure 2.3 that this implies a shear deformation of a square. In contrast, the matrix $\mathbf{\Omega}$ causes a rigid rotation. For instance, if $\epsilon = 0$, we have $\frac{\partial u_z}{\partial x} = -\frac{\partial u_x}{\partial z}$ and

$$\mathbf{J} = \mathbf{\Omega} = \begin{pmatrix} 0 & \frac{\partial u_x}{\partial z} \\ -\frac{\partial u_z}{\partial x} & 0 \end{pmatrix} = \begin{pmatrix} 0 & \theta \\ -\theta & 0 \end{pmatrix} \tag{2.22}$$

The interpretation as a rigid rotation can again be seen in Figure 2.3. In both examples,

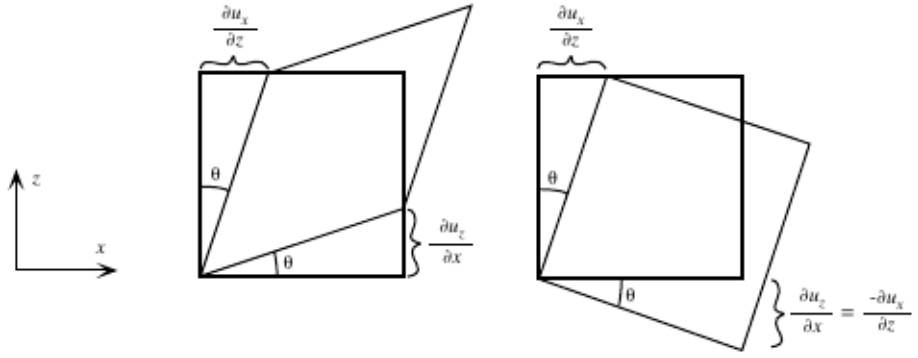

Figure 2.3: The different effects of the strain tensor ϵ and the rotation tensor are illustrated by the deformation of a square in the $x - z$ plane. The off-diagonal component ϵ cause shear deformation (left square), whereas $\mathbf{\Omega}$ causes rigid rotation (right square). Picture from Shearer (2009).

the volume of the considered infinitesimal volume (or square in 2D) does not change. The relative volume increase, or dilatation, $\Delta = (V - V_0)/V_0$ is given by the sum of the extensions in all 3 directions x, y and z

$$\Delta = \frac{\partial u_x}{\partial x} + \frac{\partial u_y}{\partial y} + \frac{\partial u_z}{\partial z} = Tr(\epsilon) = \text{div } \mathbf{u} = \nabla.\mathbf{u} \tag{2.23}$$

where we introduce the divergence operator div as the scalar product with the differential operator ∇ such that

$$\nabla = \begin{pmatrix} \partial/\partial x \\ \partial/\partial y \\ \partial/\partial z \end{pmatrix} \tag{2.24}$$

Finally if we consider the curl of the displacement, which we denote using $\nabla\times$ notation, we have

$$\nabla \times \mathbf{u} = \frac{1}{2}\left(\frac{\partial u_z}{\partial y} - \frac{\partial u_y}{\partial z}\right)\mathbf{x} + \frac{1}{2}\left(\frac{\partial u_x}{\partial z} - \frac{\partial u_z}{\partial x}\right)\mathbf{y} + \frac{1}{2}\left(\frac{\partial u_y}{\partial x} - \frac{\partial u_x}{\partial y}\right)\mathbf{z} \tag{2.25}$$

where $\mathbf{x}, \mathbf{y}, \mathbf{z}$ are the unit vectors in direction respectively x, y and z. We observe that $\nabla \times \mathbf{u} = 0$ if and only if $\mathbf{\Omega} = 0$ which means that no rigid rotations occur in the displacement field.

Of note, the same eigenvalue decomposition as for the stress tensor can be made for the strain tensor. Except in the case where the principal strains are all equal, any deformation will induce shear deformation. An example of this is provided in 2D in Figure 2.4 where an extension occurs only in the x direction, which could be represented as the diagonal strain tensor

$$\epsilon = \begin{pmatrix} \frac{\partial u_x}{\partial x} & 0 \\ 0 & 0 \end{pmatrix} \tag{2.26}$$

In this case shear strain occurs, as internal angles are not preserved (Fig. 2.4).

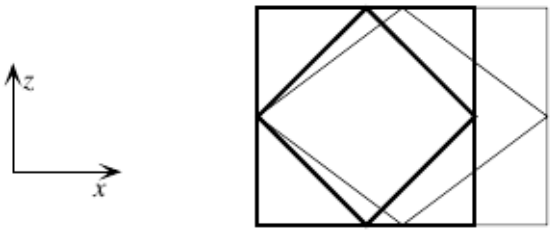

Figure 2.4: Simple extensional strain in the x direction results in shear strain; internal angles are not preserved. Picture from Shearer (2009).

2.1.3 Stress-strain relation: the Hooke's law

Definition, symmetries

The Hooke's law describes the rheology of the elastic medium, by relating the stress and strain tensors. The rheology can be elastic, plastic, visco-plastic, etc. In general, a Taylor development of this law is written as

$$\sigma_{ij} = \sigma^0_{ij} + \sum_{kl} c_{ijkl}\epsilon_{kl} + \sum_{kl}\sum_{mn} d_{ijklmn}\epsilon_{kl}\epsilon_{mn} + O(\|\epsilon\|^3). \tag{2.27}$$

Here, the *small strain assumption* states that the second order part of the rheology law can be neglected, such that we keep only the linear part and the fourth-order tensor c_{ijkl}. We also neglect the zero-order term, which corresponds to the case of "pre-stressed" medium: we will consider that the medium is at rest at time $t = 0$. We thus have

$$\sigma_{ij} = \sum_{kl} c_{ijkl}\epsilon_{kl}. \tag{2.28}$$

As we have seen before, the stress and strain tensors are symmetric: the strain tensor is symmetric by definition and the stress tensor is symmetric in the assumption of mechanic equilibrium.

The symmetry of the stress tensor

$$\sigma_{ij} = \sigma_{ji}, \tag{2.29}$$

together with Hooke's law 2.55 implies that

$$c_{ijkl} = c_{jikl}. \tag{2.30}$$

Similarly, the symmetry of the strain tensor

$$\epsilon_{kl} = \epsilon_{lk}, \tag{2.31}$$

also implies that

$$c_{ijkl} = c_{ijlk}. \tag{2.32}$$

These two symmetries are called the minor symmetries. Also, it can be shown that the stress-strain relation can be derived from an energy functional (Gould, 2013)

$$\sigma_{ij} = \frac{\partial U}{\partial \epsilon_{ij}}. \tag{2.33}$$

Therefore

$$\sum_{kl} c_{ijkl}\epsilon_{kl} = \frac{\partial U}{\partial \epsilon_{ij}}, \tag{2.34}$$

then

$$c_{ijkl} = \frac{\partial^2 U}{\partial \epsilon_{ij}\partial \epsilon_{kl}}. \tag{2.35}$$

The invariance regarding the order of derivation implies the major symmetry

$$c_{ijkl} = c_{klij}. \tag{2.36}$$

Using minor and major symmetries, the number of independent coefficients in the stiffness tensor c_{ijkl} reduces to 21. Such kind of stiffness tensor describe generic, anisotropic media, as is discussed later.

Isotropic media and mechanical parameters

In the case of isotropic media, the stiffness tensor c_{ijkl} can be reduced to an even simpler form as

$$c_{ijkl} = \lambda\delta_{ij}\delta_{kl} + \mu\left(\delta_{il}\delta_{jk} + \delta_{ik}\delta_{jl}\right) \tag{2.37}$$

where δ_{ij} is the Kronecker delta function defined by

$$\delta_{ij} = \begin{cases} 1 & \text{if} \quad i = j \\ 0 & \text{else,} \end{cases} \tag{2.38}$$

and λ and μ are the Lamé parameters. If we plug the expression (2.37) in the stress-strain relationship, we have

$$\sigma_{ij} = \lambda\delta_{ij}\sum_{kl}\delta_{kl}\epsilon_{kl} + \mu\sum_{kl}\delta_{il}\delta_{jk}\epsilon_{kl} + \mu\sum_{kl}\delta_{ik}\delta_{jl}\epsilon_{kl} \tag{2.39}$$

which simplifies as

$$\sigma_{ij} = \lambda\delta_{ij}\sum_{k}\epsilon_{kk} + 2\mu\epsilon_{ij} \tag{2.40}$$

or

$$\sigma = \begin{pmatrix} \lambda Tr\epsilon + 2\mu\epsilon_{11} & 2\mu\epsilon_{12} & 2\mu\epsilon_{13} \\ 2\mu\epsilon_{12} & \lambda Tr\epsilon + 2\mu\epsilon_{22} & 2\mu\epsilon_{23} \\ 2\mu\epsilon_{13} & 2\mu\epsilon_{23} & \lambda Tr\epsilon + 2\mu\epsilon_{33} \end{pmatrix} \tag{2.41}$$

In this isotropic case, the two Lamé parameters describe completely the elastic properties. The term μ is the shear modulus and is a measure of the resistance of the material to

shearing. Its value is given by half of the ratio between the applied shear stress and the resulting shear strain

$$\mu = \frac{\sigma_{xy}}{2\epsilon_{xy}} \tag{2.42}$$

The λ parameter has no simple physical explanation. Other sets of parameters are often considered, in mechanics and seismology communities. The couple Young modulus and shear modulus (E, μ) can be considered for instance. In this case

$$E = \frac{(3\lambda + 2\mu)\mu}{\lambda + \mu} \tag{2.43}$$

represents the ratio of extensional stress to the resulting extensional strain for a cylinder being pulled on both ends. Another choice is considering bulk modulus and shear modulus (κ, μ) where

$$\kappa = \lambda + \frac{2}{3}\mu \tag{2.44}$$

The bulk modulus is the ratio between the hydrostatic pressure to the resulting volume change, thereby a measure of material incompressibility. Another set of parameters is Poisson's ration and shear modulus (σ_P, μ) where

$$\sigma_P = \frac{\lambda}{2(\lambda + \mu)} \tag{2.45}$$

The Poisson's ratio represents the ratio of the lateral contraction of a cylinder (being pulled on both ends) to its longitudinal extension. Finally, in seismology, a usual set of parameters is P-wave and S-wave velocities, defined by

$$V_P = \sqrt{\frac{\lambda + 2\mu}{\rho}}, \quad V_S = \sqrt{\frac{\mu}{\rho}}, \tag{2.46}$$

where ρ is the density of the material.

The Poisson's ratio is often used as a measure of the relative size of P- and S-wave velocities. It can be shown that

$$\sigma_P = \frac{V_P^2 - 2V_S^2}{2(V_P^2 - V_S^2)} = \frac{(V_P/V_S)^2 - 2}{2(V_P/V_S)^2 - 2} = \tag{2.47}$$

Of note σ_P is dimensionless and varies between 0 and 0.5. The upper limit represents a fluid ($\mu = 0$). From the definition of P- and S-wave velocities, one can see that P-waves are always faster than S-waves. The minimum ratio between P- and S-wave velocities is $\sqrt{2}$ which occur at the limit when $\lambda = \sigma_P = 0$. A Poisson's solid is defined as having $\lambda = \mu$, therefore $\sigma = 0.25$ and $V_P/V_S = \sqrt{3}$. Most crustal rocks have a Poisson's ratio between 0.25 and 0.30. Finally, the Poisson's ratio is very useful to detect the presence of fluid, in which case the Poisson's ratio increases.

2.1.4 Wave equation in an elastic medium

Now that we introduced the stress and strain tensor, and their relationship using the Hooke's law, we have all what we need to derive the equations representing the propagation of waves in a solid medium. We consider an infinitesimal cube with a closed volume V around $\mathbf{x} = (x_1, x_2, x_3)$, in the Cartesian coordinate system. As we have seen before, the

forces exerted on each face of the cube are given by the product of the traction vector and the surface area. For instance, the force on the plane normal to x_1 can be represented as

$$\mathbf{F}(\mathbf{x}_1) = \mathbf{t}(\mathbf{x}_1)dx_2dx_3 = \sigma\mathbf{x}_1 = \begin{pmatrix}\sigma_{11}\\ \sigma_{21}\\ \sigma_{31}\end{pmatrix} dx_2dx_3 \tag{2.48}$$

with $\mathbf{F}$ the force, $\mathbf{t}$ the traction vector and σ the stress tensor. At equilibrium, we have seen that we have $\mathbf{F}(-\mathbf{x}_1) = -\mathbf{F}(\mathbf{x}_1)$. In this case there is no net force applied to the cube. This corresponds to the case of an homogeneous stress tensor, which does not depend on space: the same stress tensor describes what happens on left and right faces of the infinitesimal cube. If there is no net force on the cube, then it remains at equilibrium and no wave propagates. We are interested here in describing what happens away from this equilibrium. To make this analysis, we need to make the assumption that the stress tensor is no longer homogeneous. In this case, the net force applied to the cube from the planes normal to x_1 is given by

$$\mathbf{F}(\mathbf{x}_1) = \frac{\partial}{\partial x_1}\begin{pmatrix}\sigma_{11}\\ \sigma_{21}\\ \sigma_{31}\end{pmatrix} dx_1dx_2dx_3. \tag{2.49}$$

Therefore, the total force from the stress field on all the faces of the cube is given by

$$F_i = \sum_{j=1}^{3}\frac{\partial\sigma_{ij}}{\partial x_j}dx_1dx_2dx_3 \quad i = 1,\dots 3 \tag{2.50}$$

We can also consider there is a volumetric force (body force) into the considered cube, which corresponds to

$$F_i^{body} = f_i dx_1dx_2dx_3, \tag{2.51}$$

Finally, the mass m of the infinitesimal cube is given

$$m = \rho dx_1dx_2dx_3 \tag{2.52}$$

Now we apply the 2nd Newton's law, equating the sum of the forces exerted on the cube with the mass times the acceleration of the cube. We obtain

$$m\frac{\partial^2\mathbf{u}}{\partial t^2} = \mathbf{F} + \mathbf{F}^{body} \tag{2.53}$$

thus

$$\rho\frac{\partial^2 u_i}{\partial t^2} = \sum_{j=1}^{3}\frac{\partial\sigma_{ij}}{\partial x_j} + f_i, \quad i = 1,\dots 3 \tag{2.54}$$

The Newton's law relates the displacement field to the stress field. We use the formerly derived Hooke's relation to close the system with an equation relating the stress field to the displacement (through the strain field).

$$\sigma_{ij} = \sum_{k,l=1}^{3} c_{ijkl}\epsilon_{kl} = \sum_{k,l=1}^{3} c_{ijkl}\frac{1}{2}\left(\frac{\partial u_k}{\partial x_l} + \frac{\partial u_l}{\partial x_k}\right) \tag{2.55}$$

We end up with the following system of elastodynamics equations

$$\begin{cases} \rho \dfrac{\partial^2 u_i}{\partial t^2} = \displaystyle\sum_{j=1}^{3} \frac{\partial \sigma_{ij}}{\partial x_j} + f_i, \quad i = 1, \dots 3 \\ \sigma_{ij} = \displaystyle\sum_{k,l=1}^{3} c_{ijkl} \frac{1}{2} \left(\frac{\partial u_k}{\partial x_l} + \frac{\partial u_l}{\partial x_k} \right), \quad i = 1, \dots 3 \;\; j = 1, \dots 3. \end{cases} \tag{2.56}$$

This system can be rewritten as the following velocity-stress system, which will be useful to lead the analysis on hyperbolic systems in the next Section. We take a time derivative of the stress-displacement equation, and substitute the particle displacement field $\mathbf{u}(\mathbf{x}, t)$ with the particle velocity field $\mathbf{v}(\mathbf{x}, t)$ in both equations. We obtain

$$\begin{cases} \rho \dfrac{\partial v_i}{\partial t} = \displaystyle\sum_{j=1}^{3} \frac{\partial \sigma_{ij}}{\partial x_j} + f_i, \quad i = 1, \dots 3 \\ \dfrac{\sigma_{ij}}{\partial t} = \displaystyle\sum_{kl=1}^{3} c_{ijkl} \frac{1}{2} \left(\frac{\partial u_k}{\partial x_l} + \frac{\partial u_l}{\partial x_k} \right), \quad i = 1, \dots 3 \;\; j = 1, \dots 3. \end{cases} \tag{2.57}$$

2.1.5 Solutions in an isotropic medium: P-waves and S-waves

If we substitute the stress with the displacement in the displacement-stress system (2.56) we reduce the elastodynamic model to a system of 3 equations on the particle displacement

$$\rho \frac{\partial^2 u_i}{\partial t^2} = \sum_{j=1}^{3} \frac{\partial}{\partial x_j} \left(\sum_{kl=1}^{3} c_{ijkl} \frac{1}{2} \left(\frac{\partial u_k}{\partial x_l} + \frac{\partial u_l}{\partial x_k} \right) \right) + f_i \tag{2.58}$$

If we further assume the medium is isotropic and described with the Lamé parameters we obtain

$$\rho \frac{\partial^2 u_i}{\partial t^2} = \sum_{j=1}^{3} \frac{\partial}{\partial x_j} \left(\lambda \delta_{ij} \sum_{k=1}^{3} \frac{\partial u_k}{\partial x_k} + \mu \left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right) \right) \tag{2.59}$$

therefore

$$\rho \frac{\partial^2 u_i}{\partial t^2} = \frac{\partial}{\partial x_i} \left(\lambda \sum_{k=1}^{3} \frac{\partial u_k}{\partial x_k} \right) + \sum_{j=1}^{3} \frac{\partial}{\partial x_j} \left(\mu \left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right) \right) \tag{2.60}$$

and

$$\begin{aligned} \rho \frac{\partial^2 u_i}{\partial t^2} = \; & \left(\sum_{k=1}^{3} \frac{\partial u_k}{\partial x_k} \right) \frac{\partial \lambda}{\partial x_i} + \sum_{j=1}^{3} \left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right) \frac{\partial \mu}{\partial x_j} + \\ & \lambda \frac{\partial}{\partial x_i} \sum_{k=1}^{3} \frac{\partial u_k}{\partial x_k} + \mu \sum_{j=1}^{3} \frac{\partial}{\partial x_j} \left(\frac{\partial u_j}{\partial x_i} + \frac{\partial u_i}{\partial x_j} \right) \end{aligned} \tag{2.61}$$

We focus on the last term of this expression. We have

$$\mu \sum_{j=1}^{3} \frac{\partial}{\partial x_j} \left(\frac{\partial u_j}{\partial x_i} + \frac{\partial u_i}{\partial x_j} \right) = \mu \frac{\partial}{\partial x_i} \sum_{j=1}^{3} \frac{\partial u_j}{\partial x_j} + \mu \sum_{j=1}^{3} \frac{\partial^2 u_i}{\partial x_j^2} \tag{2.62}$$

therefore we have

$$\rho \frac{\partial^2 u_i}{\partial t^2} = \left(\sum_{k=1}^{3} \frac{\partial u_k}{\partial x_k} \right) \frac{\partial \lambda}{\partial x_i} + \sum_{j=1}^{3} \left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right) \frac{\partial \mu}{\partial x_j} + (\lambda + \mu) \frac{\partial}{\partial x_i} \sum_{k=1}^{3} \frac{\partial u_k}{\partial x_k} + \mu \sum_{j=1}^{3} \frac{\partial^2 u_i}{\partial x_j^2} \tag{2.63}$$

This can be rewritten using vector notations as

$$\rho\frac{\partial^2\mathbf{u}}{\partial t^2} = (\text{div }\mathbf{u})\nabla\lambda + \left(\nabla\mathbf{u} + \nabla\mathbf{u}^T\right)\nabla\mu + (\lambda+\mu)\,\nabla\,(\text{div }\mathbf{u}) + \mu\Delta\mathbf{u} \tag{2.64}$$

where Δ is the Laplacian operator, such that

$$(\Delta u)_i = \sum_{j=1}^{3}\frac{\partial^2 u_i}{\partial x_j^2} \tag{2.65}$$

We now use the vector identity

$$\text{curl curl }\mathbf{u} = \nabla\text{div }\mathbf{u} - \Delta\mathbf{u} \tag{2.66}$$

where the curl operator is defined by

$$\text{curl }\mathbf{u} = \begin{pmatrix} \dfrac{\partial u_y}{\partial u_z} - \dfrac{\partial u_z}{\partial u_y} \\ \dfrac{\partial u_x}{\partial u_z} - \dfrac{\partial u_z}{\partial u_x} \\ \dfrac{\partial u_x}{\partial u_y} - \dfrac{\partial u_y}{\partial u_x} \end{pmatrix} \tag{2.67}$$

We thus have

$$\rho\frac{\partial^2\mathbf{u}}{\partial t^2} = (\text{div }\mathbf{u})\nabla\lambda + \left(\nabla\mathbf{u} + \nabla\mathbf{u}^T\right)\nabla\mu + (\lambda+2\mu)\,\nabla\,(\text{div }\mathbf{u}) - \mu\text{curl curl }\mathbf{u} \tag{2.68}$$

The two first term on the right hand side (2.68) depends on the gradient of the medium properties, here the Lamé parameters λ and μ. We here assume that the medium is homogeneous so that these gradient vanish. In practice this will of course not be the case: the subsurface is heterogeneous. But here we consider this canonical case to identify the type of waves which propagate in an isotropic elastic medium.

Using this simplifying assumption, we obtain

$$\rho\frac{\partial^2\mathbf{u}}{\partial t^2} = (\lambda+2\mu)\,\nabla\,(\text{div }\mathbf{u}) - \mu\text{curl curl }\mathbf{u} \tag{2.69}$$

We now apply the Helmholtz-Hodge decomposition theorem, which states that for sufficiently regular function on a connex domain with sufficiently smooth boundary, we can always decompose a vector field $\mathbf{u}(\mathbf{x},t)$ into the gradient of a scalar potential field $\psi(\mathbf{x},t)$ and the curl of a vector potential field $\Psi(\mathbf{x},t)$ such that

$$\mathbf{u}(\mathbf{x},t) = \nabla\psi(\mathbf{x},t) + \text{curl }\mathbf{\Psi}(\mathbf{x},t), \quad \text{div }\mathbf{\Psi} = 0. \tag{2.70}$$

Replacing $\mathbf{u}$ with this decomposition in (2.69) yields

$$\rho\frac{\partial^2\left(\nabla\psi + \text{curl }\mathbf{\Psi}\right)}{\partial t^2} = (\lambda+2\mu)\,\nabla\left(\text{div }\nabla\psi + \text{div curl }\mathbf{\Psi}\right) - \mu\text{curl curl }\left(\nabla\psi + \text{curl }\mathbf{\Psi}\right) \tag{2.71}$$

We now use the identities

$$\text{curl}\left(\nabla\psi\right) = 0, \quad \text{div}\left(\text{curl }\psi\right) = 0, \quad \text{div }\nabla\psi = \Delta\psi. \tag{2.72}$$

We thus obtain

$$\nabla\left(\rho\frac{\partial^2\psi}{\partial t^2}\right)-\nabla\left(\left(\lambda+2\mu\right)\Delta\psi\right)=-\text{curl}\,\left(\rho\frac{\partial^2\mathbf{\Psi}}{\partial t^2}\right)-\mu\text{curl curl curl}\,\mathbf{\Psi} \tag{2.73}$$

Taking into account (2.66) we see that

$$\text{curl}\left(\text{curl curl}\,\mathbf{\Psi}\right)=\text{curl}\,\nabla\text{div}\,\mathbf{\Psi}-\text{curl}\,\Delta\mathbf{\Psi}=-\text{curl}\,\Delta\mathbf{\Psi}. \tag{2.74}$$

Therefore

$$\nabla\left(\rho\frac{\partial^2\psi}{\partial t^2}\right)-\nabla\left(\left(\lambda+2\mu\right)\Delta\psi\right)=-\text{curl}\,\left(\rho\frac{\partial^2\mathbf{\Psi}}{\partial t^2}\right)+\text{curl}\,\mu\Delta\mathbf{\Psi} \tag{2.75}$$

which we can rewrite as

$$\nabla\left(\rho\frac{\partial^2\psi}{\partial t^2}-\left(\lambda+2\mu\right)\Delta\psi\right)=-\text{curl}\,\left(\rho\frac{\partial^2\mathbf{\Psi}}{\partial t^2}-\mu\Delta\mathbf{\Psi}\right) \tag{2.76}$$

The latter equation is very important. One solution to this equation is to cancel simultaneously the left and right hand side, which yields one equation on the scalar potential ψ

$$\frac{\partial^2\psi}{\partial t^2}-\frac{\lambda+2\mu}{\rho}\Delta\psi=0, \tag{2.77}$$

and one equation on the vector potential $\mathbf{\Psi}$

$$\frac{\partial^2\mathbf{\Psi}}{\partial t^2}-\frac{\mu}{\rho}\Delta\mathbf{\Psi}. \tag{2.78}$$

Both are **wave equations**. The first describes the propagation of pressure waves at speed V_P such that

$$V_P=\sqrt{\frac{\lambda+2\mu}{\rho}} \tag{2.79}$$

The second describes the propagation of shear-waves at speed V_S such that

$$V_S=\sqrt{\frac{\mu}{\rho}} \tag{2.80}$$

From this derivation, we observe that the displacement field separates into two independent contributions:

- A scalar potential field φ, associated with dilation/compression deformations of the medium. These correspond to P-waves (primary or compression waves). In this mode, particles oscillate in the same direction as the wave propagation, leading to alternating compression and dilatation of the medium (Stein and Wysession, 2003). An illustration of the particle displacement associated with the P-wave compared to the propagation direction is given in Figure 2.5.

- A divergence-free vector potential field $\boldsymbol{\psi}$, associated with shearing deformations of the medium. These correspond to S-waves (secondary or shear waves). In this mode, particle motion is perpendicular to the propagation direction, resulting in shear strains without any volume change (Stein and Wysession, 2003). In Figure 2.5 we illustrate the particle motion associated with the S-wave. For this reason, S-waves are also called transverse waves. The plane perpendicular to the propagation direction in which S-waves oscillate determines their type. If the motion occurs in a vertical plane, the wave is called an SV-wave; if it occurs in a horizontal plane, it is referred to as an SH-wave.

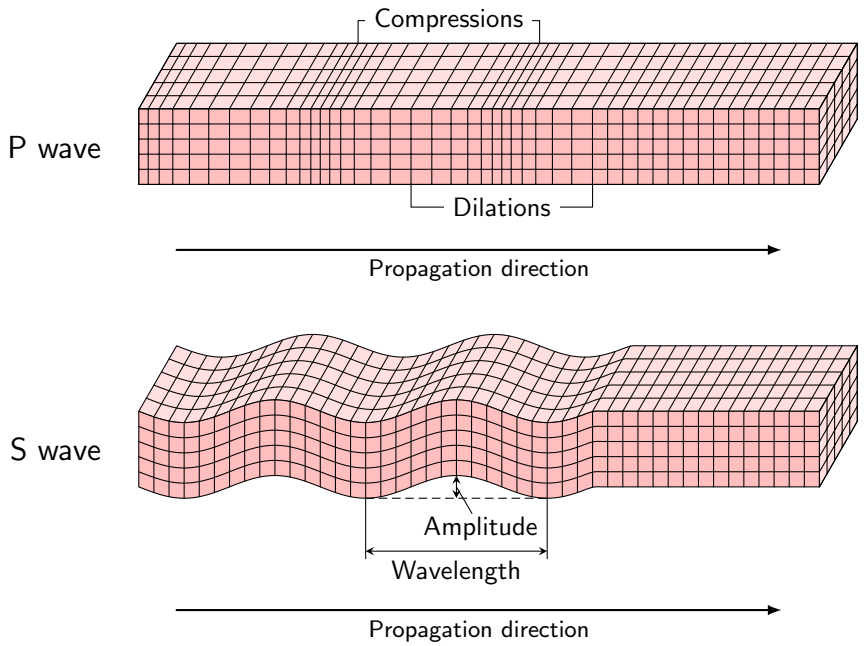

Figure 2.5: Schematic illustration of particle motion associated with seismic body waves relative to the propagation direction. P-waves (top) produce particle displacements parallel to the propagation direction, alternating between compression and dilation of the medium. S-waves (bottom) induce particle motion perpendicular to the propagation direction, either in the vertical plane (SV-waves) or in the horizontal plane (SH-waves). Figure from Grange (2025).

2.1.6 Surface waves

The analysis of the elastodynamics equations in an isotropic medium makes it possible to identify to modes of propagation: one mode aligned with the direction of propagation (P-waves), one mode where the displacement is in a plane orthogonal to the direction of propagation (S-waves).

The propagation of these two modes give birth to specific interface waves at the physical boundaries of the domain of propagation. A good illustration of such interface is the surface/subsurface interface, that is the boundary between the solid Earth and the atmosphere. A this interface, the traction vector vanishes, and the surface is considered to be able to move freely, hence the terminology of "free surface" condition. This conditions translates mathematically as

The traction vector T acting along a direction represented by a normalized three-dimensional vector n, given a particular stress tensor σ is obtained as

$$\mathbf{t} = \sigma\mathbf{n} = 0, \tag{2.81}$$

where $\mathbf{n}$ is the unit vector normal to the surface. Assuming the topography is flat (flat free surface), and the normal points upward, we have

$$n = [0\ 0\ 1]^T \tag{2.82}$$

and thus the free surface condition translates into

$$\sigma_{13} = 0, \;\; \sigma_{23} = 0, \;\; \sigma_{33} = 0. \tag{2.83}$$

We see that the free surface also follows the definition of a surface which is subject to zero normal stress and zero parallel shear stress. This boundary condition plays a central role in seismic wave propagation because it modifies the way body waves (P and S waves) interact with the surface. Instead of simply being reflected, part of their energy is converted into a distinct class of waves that are trapped near the surface, which hence are called surface waves.

The first type of surface wave is Rayleigh (1887) wave. It can be shown that the velocity V_R of Rayleigh waves is always lower than the S-wave velocity V_s (Landau and Lifchitz, 1967). In a homogeneous medium with a flat free surface, Rayleigh waves are non-dispersive, meaning that their phase velocity is independent of frequency and $V_R \approx 0.919 V_s$ in a material with a Poisson's ratio of $\nu \approx 0.25$. In contrast, if the medium exhibits near-surface heterogeneities, Rayleigh waves become dispersive. Similarly, in the case of a non-flat topography, even in a homogeneous medium, Rayleigh waves are also dispersive (Pilant, 1979). In Figure 2.6 we illustrate the particle displacement associated with this wave relative to the direction of propagation. It shows that the particle motion at the free surface is elliptical, producing a deformation of the ground. Rayleigh waves can be observed in seismic recordings on both the vertical component and the horizontal component parallel to the propagation direction.

The second type of surface wave is the Love (1911) wave. Its particle motion is transverse and confined to the horizontal plane, i.e., perpendicular to the direction of propagation, as illustrated in Figure 2.6. Consequently, Love waves are observed exclusively on the horizontal component perpendicular to the propagation direction in 3-component seismograms. Unlike Rayleigh waves, they cannot exist in a homogeneous medium, even in the presence of surface topography. They only arise in stratified media, where an upper layer with different elastic properties overlies a half-space. Furthermore, when they exist, Love waves are inherently dispersive. More in details, Love waves result from interferences between incident, reflected and refracted S-H waves in an heterogeneous zone close to the surface. These waves do not exist in homogeneous media. We have

$$V_{S,1} < V_L < V_{S,2}, \tag{2.84}$$

$V_{S,1}$ and $V_{S,2}$ being the S-wave velocities of the two layers close from the surface, and V_L being the velocity of Love waves.

Compared with P-wave and S-waves, which are volumetric waves, surface waves (Rayleigh and Love waves) only propagate along the surface. They can have catastrophic effects when earthquakes occurs: these are the most destructive waves as they shake the surface. They have a very low penetration depth: the energy remains concentrated along the surface. In terms of imaging and seismic recordings, they dominate volumetric waves in terms of amplitudes. For specific targets (in particular at exploration scale) they can be detrimental to imaging as they hide the signal associated with volumetric waves and carry information only on the near surface. In this case they are considered as noise (ground roll) and specific surface waves removal algorithm are employed to remove their imprint from the data.

2.1.7 Anisotropy

We have seen that rhe rheology of the medium is described by the stiffness tensor c_{ijkl}. To analyze in details wave propagation, we have considered for now an isotropic medium,

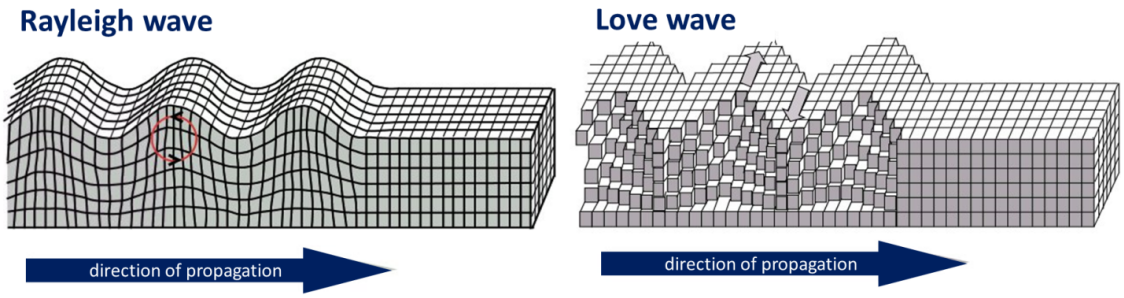

Figure 2.6: Schematic illustration of particle motion associated with each type of surface wave relative to the propagation direction. Rayleigh waves (left) produce elliptical motion in the vertical plane parallel to propagation, while Love waves (right) induce transverse motion in the horizontal plane, perpendicular to propagation. This Figure is taken from Wikipedia.

which can be described by stiffness tensor with 2 independent coefficient. In this approximation, velocity of waves does not depend on the direction of propagation.

When anisotropy is considered, the P-wave and S-wave velocities depend on the direction of propagation. This dependence is described by symmetries of increasing complexity. The first type of anisotropy, which is the one most often considered in practice in seismic exploration, is called transverse isotropy (TI). In this symmetry, there is one preferred direction, along which the velocity differs, compared to the velocity in the plane orthogonal to this direction. If this preferred direction is aligned with the vertical axis in a Cartesian coordinate system, we speak of Vertical Transverse Isotropy (VTI). If it is aligned with one of the horizontal axis (x or y) we speak of Horizontal Transvere Isotropy (HTI). Both VTI and HTI anisotropy require 5 independent coefficients in the stiffness tensor to be described. If the preferred direction is not aligned with any of the Cartesian coordinate axis, then we speak of Tilted Transverse Isotropy (TTI) and 2 additional independent coefficient (corresponding to 2 angles) are required in the stiffness tensor to describe this anisotropy.

Such kind of anisotropy is for instance observed in case of thin layering. A vertical pile of thin layers, each layer aligned with the (x, y) plane, of thickness significantly smaller than the wavelength, will produce an "equivalent" smooth VTI medium. If the layers are aligned along a (x, z) or (y, z) plane and piled in the y or x direction, this gives rise to an "equivalent" smooth HTI medium. If such layers pile with a tilt angle, this gives rise to an "equivalent" smooth TTI medium. The notion of "equivalent" smooth anisotropic medium is not developed in this course, it is associated with the theory of homogenization of elastic material (see for instance Bensoussan et al., 1978; Capdeville et al., 2020). Illustration of VTI, HTI, and TTI anisotropy is provided in Figure 2.7.

A higher level of symmetry involve orthorombic anisotropy. The orthorhombic anisotropy is defined by three mutually orthogonal planes of symmetry leading to 9 independent coefficients if these planes are aligned with the coordinate axis (Tsvankin, 2005). This is illustrated in Figure 2.8. In this symmetry, the velocities will be decomposed between one vertical velocity (aligned with z), and two horizontal velocities (aligned with x and y respectively). Such kind of anisotropy can be observed for instance in environment combining vertical piles of horizontal layers, and network of cracks or fractures aligned with one of the horizontal direction (x or y).

The most complex anisotropy type is called triclinic anisotropy, which requires 21 independent coefficient to be described. In this general case, there is no specific symmetry

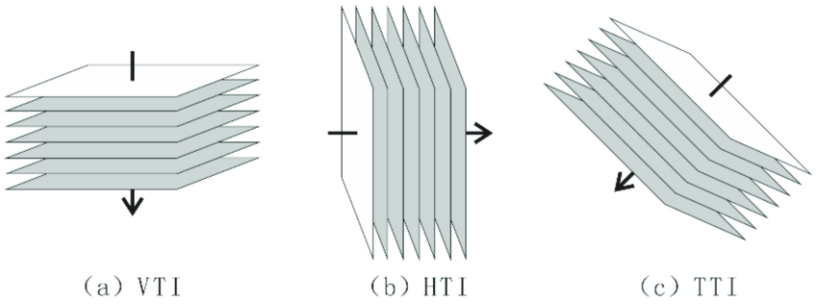

Figure 2.7: Illustration of VTI, HTI, and TTI anisotropy. Liu, Zhengping Zhu, Renfang Pan, Bole Gao and Jineng Jin - Analysis of AVAZ Seismic Forward Modeling of Fracture-Cavity Reservoirs of the Dengying Formation, Central Sichuan Basin. http://doi.org/10.3390/en15145022, CC BY 4.0, https://commons.wikimedia.org/w/index.php?curid=153532663

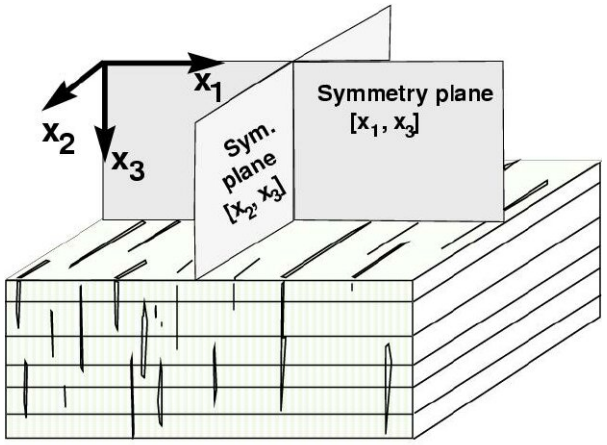

Figure 2.8: Illustration of orthoroombic anisotropy.

involved, apart the minor and major symmetries of the stiffness tensor, which come from the definition of the strain and stress tensor and energy based considerations. In triclinic anisotropy, the velocities depend on the three direction of spaces, with no specific alignment on the Cartesian coordinates.

The origin of anisotropy can be intrinsic to the medium: the mineral composition for instance makes naturally emerge preferred orientations. It can be extrinsic: thin layering for instance. Piling up thin horizontal layers, with layers thickness much smaller than the propagated wavelength, result in a medium which is "seen" as VTI by the waves propagating inside it. The theory of homogeneization tells us that for any finite-frequency propagation in an elastic medium, the "effective" medium seen by the waves propagating inside it is actually anisotropic. The Earth is such that there is a continuous wavenumber spectrum in terms of heterogeneities. Independently of the frequency band which is propagated inside it, there will alway be heterogeneities smaller than the propagated wavelenghts. These heterogeneities will be seen in practice as an anisotropy.

2.1.8 Acoustic approximation

Some applications at the exploration scale rely on the acoustic approximation, because sensors and sources are located in a water layer, where shear waves do not propagate. In this case, at first approximation, modeling only compression waves (P-waves) under the acoustic approximation is a useful pragmatic choice. Here we review how to derive acoustic equations from the elastodynamic equations.

Let us first rewrite the isotropic elastodynamics system in velocity-stress formulation

$$
\left\{
\begin{array}{rcl}
\dfrac{\partial v_x}{\partial t} &=& \dfrac{1}{\rho}\left(\dfrac{\partial \sigma_{xx}}{\partial x}+\dfrac{\partial \sigma_{xy}}{\partial y}+\dfrac{\partial \sigma_{xz}}{\partial z}\right)+f_x \\
\dfrac{\partial v_y}{\partial t} &=& \dfrac{1}{\rho}\left(\dfrac{\partial \sigma_{xy}}{\partial x}+\dfrac{\partial \sigma_{yy}}{\partial y}+\dfrac{\partial \sigma_{yz}}{\partial z}\right)+f_y \\
\dfrac{\partial v_z}{\partial t} &=& \dfrac{1}{\rho}\left(\dfrac{\partial \sigma_{xz}}{\partial x}+\dfrac{\partial \sigma_{yz}}{\partial y}+\dfrac{\partial \sigma_{zz}}{\partial z}\right)+f_z \\
\dfrac{\partial \sigma_{xx}}{\partial t} &=& (\lambda+2\mu)\dfrac{\partial v_x}{\partial x}+\lambda\dfrac{\partial v_y}{\partial y}+\lambda\dfrac{\partial v_z}{\partial z}+\dfrac{\partial \sigma_{xx_0}}{\partial t} \\
\dfrac{\partial \sigma_{yy}}{\partial t} &=& \lambda\dfrac{\partial v_x}{\partial x}+(\lambda+2\mu)\dfrac{\partial v_y}{\partial y}+\lambda\dfrac{\partial v_z}{\partial z}+\dfrac{\partial \sigma_{yy_0}}{\partial t} \\
\dfrac{\partial \sigma_{zz}}{\partial t} &=& \lambda\dfrac{\partial v_x}{\partial x}+\lambda\dfrac{\partial v_y}{\partial y}+(\lambda+2\mu)\dfrac{\partial v_z}{\partial z}+\dfrac{\partial \sigma_{zz_0}}{\partial t} \\
\dfrac{\partial \sigma_{xy}}{\partial t} &=& \mu\left(\dfrac{\partial v_x}{\partial y}+\dfrac{\partial v_y}{\partial x}\right)+\dfrac{\partial \sigma_{xy_0}}{\partial t} \\
\dfrac{\partial \sigma_{yz}}{\partial t} &=& \mu\left(\dfrac{\partial v_y}{\partial z}+\dfrac{\partial v_z}{\partial y}\right)+\dfrac{\partial \sigma_{yz_0}}{\partial t} \\
\dfrac{\partial \sigma_{xz}}{\partial t} &=& \mu\left(\dfrac{\partial v_x}{\partial z}+\dfrac{\partial v_z}{\partial x}\right)+\dfrac{\partial \sigma_{xz_0}}{\partial t}
\end{array}
\right. \tag{2.85}
$$

If we neglect all shear effects, the elastodynamics system of equations boils down to the acoustic system. Setting $\mu = 0$ in the previous system yields automatically

$$
\sigma_{yz} = 0, \quad \sigma_{xz} = 0, \quad \sigma_{xy} = 0. \tag{2.86}
$$

This yields

$$\begin{cases} \dfrac{\partial v_x}{\partial t} &= \dfrac{1}{\rho}\dfrac{\partial \sigma_{xx}}{\partial x} + f_x \\ \dfrac{\partial v_y}{\partial t} &= \dfrac{1}{\rho}\dfrac{\partial \sigma_{yy}}{\partial y} + f_y \\ \dfrac{\partial v_z}{\partial t} &= \dfrac{1}{\rho}\dfrac{\partial \sigma_{zz}}{\partial z} + f_z \\ \dfrac{\partial \sigma_{xx}}{\partial t} &= \lambda\dfrac{\partial v_x}{\partial x} + \lambda\dfrac{\partial v_y}{\partial y} + \lambda\dfrac{\partial v_z}{\partial z} + \dfrac{\partial \sigma_{xx_0}}{\partial t} \\ \dfrac{\partial \sigma_{yy}}{\partial t} &= \lambda\dfrac{\partial v_x}{\partial x} + \lambda\dfrac{\partial v_y}{\partial y} + \lambda\dfrac{\partial v_z}{\partial z} + \dfrac{\partial \sigma_{yy_0}}{\partial t} \\ \dfrac{\partial \sigma_{zz}}{\partial t} &= \lambda\dfrac{\partial v_x}{\partial x} + \lambda\dfrac{\partial v_y}{\partial y} + \lambda\dfrac{\partial v_z}{\partial z} + \dfrac{\partial \sigma_{zz_0}}{\partial t} \end{cases} \tag{2.87}$$

We introduce P as $P = \sigma_{xx} = \sigma_{yy} = \sigma_{zz}$ and obtain

$$\begin{cases} \dfrac{\partial v_x}{\partial t} &= \dfrac{1}{\rho}\dfrac{\partial \sigma_{xx}}{\partial x} + f_x \\ \dfrac{\partial v_y}{\partial t} &= \dfrac{1}{\rho}\dfrac{\partial \sigma_{yy}}{\partial y} + f_y \\ \dfrac{\partial v_z}{\partial t} &= \dfrac{1}{\rho}\dfrac{\partial \sigma_{zz}}{\partial z} + f_z \\ \dfrac{\partial P}{\partial t} &= \lambda\dfrac{\partial v_x}{\partial x} + \lambda\dfrac{\partial v_y}{\partial y} + \lambda\dfrac{\partial v_z}{\partial z} + \dfrac{\partial \sigma_{P_0}}{\partial t} \end{cases} \tag{2.88}$$

We can also rewrite λ in terms of V_P in the acoustic case as

$$\lambda = \rho V_P^2 \tag{2.89}$$

and we obtain the compact form of the acoustic equations

$$\begin{cases} \dfrac{\partial v}{\partial t} - \dfrac{1}{\rho}\nabla P = f \\ \dfrac{\partial P}{\partial t} - \rho V_P^2 \operatorname{div}(v) = \dfrac{\partial \sigma_{P_0}}{\partial t} \end{cases} \tag{2.90}$$

2.2 Considerations on first-order hyperbolic systems

We have seen that the minor and major symmetries grouped together makes the number of independent coefficients in the Hooke's tensor reduce from 81 to 21. These symmetries are exploited by the Voigt's notations to avoid manipulating 4th order tensors and recast the Hooke's law as a matrix vector product. The Voigt's notations are expressed as the following correspondance between indices:

$$11 \to 1,\ 22 \to 2,\ 33 \to 3,\ 23 \to 4,\ 13 \to 5,\ 12 \to 6. \tag{2.91}$$

This makes σ and ϵ expressible as 6 components vectors

$$\sigma = [\sigma_{11}\ \sigma_{22}\ \sigma_{33}\ \sigma_{23}\ \sigma_{13}\ \sigma_{12}]^T = [\sigma_1\ \sigma_2\ \sigma_3\ \sigma_4\ \sigma_5\ \sigma_6]^T, \tag{2.92}$$

$$\epsilon = [\epsilon_{11}\ \epsilon_{22}\ \epsilon_{33}\ 2\epsilon_{23}\ 2\epsilon_{13}\ 2\epsilon_{12}]^T = [\epsilon_1\ \epsilon_2\ \epsilon_3\ \epsilon_4\ \epsilon_5\ \epsilon_6]^T. \tag{2.93}$$

The product of the 4th order stiffness matrix c_{ijkl} and the 2nd order stress tensor σ_{ij} in the Hooke's law 2.55 can be rewritten as the matrix-vector product

$$\sigma = C\epsilon, \tag{2.94}$$

with

$$C = \begin{pmatrix} C_{11} & C_{12} & C_{13} & C_{14} & C_{15} & C_{16} \\ & C_{22} & C_{23} & C_{24} & C_{25} & C_{26} \\ & & C_{33} & C_{34} & C_{35} & C_{36} \\ & & & C_{44} & C_{45} & C_{46} \\ & & & & C_{55} & C_{56} \\ & & & & & C_{66} \end{pmatrix}, \tag{2.95}$$

and the same rule for the coefficients of C

$$C_{11} = c_{1111}, C_{12} = c_{1122}, C_{13} = c_{1133}, C_{14} = c_{1123}, C_{15} = c_{1113}, C_{16} = c_{1112}. \tag{2.96}$$

For instance

$$\sigma_{23} = \sum_{kl} c_{23kl}\epsilon_{kl} = c_{2311}\epsilon_{11} + c_{2322}\epsilon_{22} + c_{2333}\epsilon_{33} + 2\left(c_{2312}\epsilon_{12} + c_{2313}\epsilon_{13} + c_{2323}\epsilon_{23}\right), \tag{2.97}$$

yields

$$\sigma_4 = C_{41}\epsilon_1 + C_{42}\epsilon_2 + C_{43}\epsilon_3 + C_{46}\epsilon_6 + C_{45}\epsilon_5 + C_{44}\epsilon_4. \tag{2.98}$$

We introduce the matrix operator D as

$$D = \begin{pmatrix} \partial x_1 & 0 & 0 & 0 & \partial x_3 & \partial x_2 \\ 0 & \partial x_2 & 0 & \partial x_3 & 0 & \partial x_1 \\ 0 & 0 & \partial x_3 & \partial x_2 & \partial x_1 & 0 \end{pmatrix}. \tag{2.99}$$

We have

$$\epsilon = D^T u, \tag{2.100}$$

and also

$$\rho\frac{\partial^2 u}{\partial t^2} = D\sigma + F. \tag{2.101}$$

Gathering equations we obtain

$$\begin{cases} \rho\dfrac{\partial v}{\partial t} - D\sigma = F \\ \dfrac{\partial \sigma}{\partial t} - CD^T v = 0 \end{cases} \tag{2.102}$$

We first introduce the following definition.

Definition 1. *A first-order hyperbolic system is a system of partial differential equations which can be put under the form*

$$\frac{\partial w}{\partial t} + \sum_{j=1}^{d} A_j(x) \frac{\partial w}{\partial x_j} + A_0(x) w = f \tag{2.103}$$

where $w(x,t) \in \mathbb{R}^p$, $p \in \mathbb{N}$, $(x,t) \in \mathbb{R}^d \times \mathbb{R}^+$, $A_j(x) \in \mathbb{M}_p(\mathbb{R})$, $j = 0, \ldots, d$. *Here* $\mathbb{M}_p(\mathbb{R})$ *denotes the ensemble of square real matrices of size* p.

The general system of elastodynamics equations can be put under the form of a general first-order hyperbolic system with

$$\begin{cases} w(x,t) = [v(x,t) \;\; \sigma(x,t)]^T \in \mathbb{R}^9, \\ \\ \displaystyle\sum_{j=1}^{3} A_j(x) \frac{\partial w}{\partial x_j} = \begin{pmatrix} 0 & \frac{1}{\rho} D \\ C D^T & 0 \end{pmatrix} w, \\ \\ A_0(x) = 0. \end{cases} \tag{2.104}$$

We can transform this hyperbolic system in a simpler form to study its properties. We have

$$\frac{\partial w}{\partial t} - \begin{pmatrix} 0 & \frac{1}{\rho} D \\ C D^T & 0 \end{pmatrix} w = \begin{pmatrix} \frac{1}{\rho} F \\ \frac{\partial \sigma^0}{\partial t} \end{pmatrix} \tag{2.105}$$

By definition of the stiffness tensor, we know that C is symmetric positive definite. We introduce the compliance matrix $S(x) = C(x)^{-1}$. We then have

$$\begin{pmatrix} \rho I_3 & 0 \\ 0 & S \end{pmatrix} \frac{\partial w}{\partial t} - \begin{pmatrix} 0 & D \\ D^T & 0 \end{pmatrix} w = \begin{pmatrix} F \\ S \frac{\partial \sigma^0}{\partial t} \end{pmatrix} \tag{2.106}$$

which can be rewritten as

$$\widetilde{S} \frac{\partial w}{\partial t} - \sum_{j=1}^{3} X_j \frac{\partial w}{\partial x_j} = \widetilde{F}, \tag{2.107}$$

where

$$\widetilde{S} = \begin{pmatrix} \rho I_3 & 0 \\ 0 & S \end{pmatrix}, \;\; X_j = \begin{pmatrix} 0 & D_j \\ D_j^T & 0 \end{pmatrix}, \;\; \widetilde{F} = \begin{pmatrix} F \\ S \frac{\partial \sigma^0}{\partial t} \end{pmatrix} \tag{2.108}$$

and

$$D_1 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}, \; D_2 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}, \; D_3 = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}. \tag{2.109}$$

The system is interesting as we can prove easily an energy estimate. More precisely, we can prove that the norm of the solution at a time t only depends on the norm of the solution at time 0 if we consider a Cauchy problem with homogeneous Dirichlet boundary conditions

Theorem 1. *Consider the following Cauchy problem on $\Omega \times [0,T]$ with $\Omega \subset \mathbb{R}^3$ and $T \in \mathbb{R}_+^*$:*

$$\begin{cases} \widetilde{S}\dfrac{\partial w}{\partial t} - \displaystyle\sum_{j=1}^{3} X_j \frac{\partial w}{\partial x_j} = 0, \quad (x,t) \in \Omega \times [0,T], \\ \\ \forall x \in \Omega, \quad w(x,0) = w_0(x), \\ \\ \forall x \in \partial\Omega, \quad \forall t \in [0,T], \quad w(x,t) = 0. \end{cases} \tag{2.110}$$

The L^2 norm at time t of the solution w of this Cauchy problem, we denote as $\|w(.,t)\|$, is controlled by $\|w_0\|$:

$$\exists K \in \mathbb{R}_+^*, \quad \|w(.,t)\| \leq K\|w_0\| \tag{2.111}$$

Proof. We recall the definition of the L^2 norm here. For two functions defined on $L^2(\Omega)$, $u(x)$ and $v(x)$, the L^2 scalar product is defined by

$$(u,v) = \int_\Omega u(x)v(x)dx, \tag{2.112}$$

and the L^2 norm is the norm induced by this scalar product

$$\|u\|^2 = (u,u). \tag{2.113}$$

We thus have

$$\|w(.,t)\|_{L^2}^2 = \int_\Omega w^2(x,t)dx. \tag{2.114}$$

We first note that the matrix $\widetilde{S}(x)$ is symmetric positive definite by construction, as the compliance matrix $S(x)$ is symmetric positive definite.

Therefore, $\widetilde{S}(x)$ induces a scalar product and a norm on $L^2(\Omega)$ which we will denote as $(.,.)_{\widetilde{S}}$ and $\|.\|_{\widetilde{S}}$ respectively

$$(u,v)_{\widetilde{S}} = (\widetilde{S}u, v), \quad \|u\|_{\widetilde{S}} = (u,u)_{\widetilde{S}} = (\widetilde{S}u, u). \tag{2.115}$$

Then we consider the quantity $\|w(.,t)\|_{\widetilde{S}}^2$. We are interested in evaluating its variation with respect to time. We have

$$\frac{d}{dt}\|w(.,t)\|_{\widetilde{S}}^2 = \frac{d}{dt}(w,w)_{\widetilde{S}} = \left(\frac{\partial w}{\partial t}, w\right)_{\widetilde{S}} + \left(w, \frac{\partial w}{\partial t}\right)_{\widetilde{S}} = \left(\widetilde{S}\frac{\partial w}{\partial t}, w\right) + \left(\widetilde{S}w, \frac{\partial w}{\partial t}\right). \tag{2.116}$$

We replace $\widetilde{S}\frac{\partial w}{\partial t}$ in the first term only, using the hyperbolic system. We obtain

$$\frac{d}{dt}\|w(.,t)\|_{\widetilde{S}}^2 = \left(\sum_{j=1}^{3} X_j \frac{\partial w}{\partial x_j}, w\right) + \left(\widetilde{S}w, \frac{\partial w}{\partial t}\right). \tag{2.117}$$

The linearity of the scalar product makes possible to rewrite this equation as

$$\frac{d}{dt}\|w(.,t)\|_{\widetilde{S}}^2 = \sum_{j=1}^{3}\left(X_j \frac{\partial w}{\partial x_j}, w\right) + \left(\widetilde{S}w, \frac{\partial w}{\partial t}\right). \tag{2.118}$$

Using the symmetry of the matrices X_j we have

$$\frac{d}{dt}\|w(.,t)\|_{\widetilde{S}}^2 = \sum_{j=1}^{3}\left(\frac{\partial w}{\partial x_j}, X_j w\right) + \left(\widetilde{S}w, \frac{\partial w}{\partial t}\right). \tag{2.119}$$

Using integration by parts now, and the fact that we assume homogeneous Dirichlet boundary condition, we have

$$\frac{d}{dt}\|w(.,t)\|^2_{\widetilde{S}} = -\sum_{j=1}^{3}\left(w, \frac{\partial X_j w}{\partial x_j}\right) + \left(\widetilde{S}w, \frac{\partial w}{\partial t}\right). \tag{2.120}$$

We now use the fact that the matrices X_j do not depend on space (their coefficients are constant). Therefore we have

$$\frac{d}{dt}\|w(.,t)\|^2_{\widetilde{S}} = -\sum_{j=1}^{3}\left(w, X_j\frac{\partial w}{\partial x_j}\right) + \left(\widetilde{S}w, \frac{\partial w}{\partial t}\right). \tag{2.121}$$

Using now the symmetry of $\widetilde{S}$ and the linearity of the scalar product we obtain

$$\frac{d}{dt}\|w(.,t)\|^2_{\widetilde{S}} = \left(w, \widetilde{S}\frac{\partial w}{\partial t} - \sum_{j=1}^{3} X_j\frac{\partial w}{\partial x_j}\right). \tag{2.122}$$

As w is the solution of the Cauchy problem (2.110), we have

$$\frac{d}{dt}\|w(.,t)\|^2_{\widetilde{S}} = 0. \tag{2.123}$$

Therefore, we have conservation of the S norm of the solution along time, which means that

$$\forall t \in [0,T], \quad \|w(.,t)\|^2_{\widetilde{S}} = \|w(.,0)\|^2_{\widetilde{S}} = \|w_0\|^2_{\widetilde{S}}. \tag{2.124}$$

To end the proof, we use the following theorem on the equivalence of norm on $L^2(\Omega)$. There exists $(K_1, K_2) \in \mathbb{R}^*_+ \times \mathbb{R}^*_+$ such that

$$\forall u \in L^2(\Omega), \quad K_1\|u\| \leq \|u\|_{\widetilde{S}} \leq K_2\|u\|. \tag{2.125}$$

Therefore we have

$$\|w(.,t)\|^2 \leq \frac{1}{K_1}\|w(.,t)\|^2_{\widetilde{S}} = \frac{1}{K_1}\|w_0\|^2_{\widetilde{S}} \leq \frac{K_2}{K_1}\|w_0\|^2, \tag{2.126}$$

which proves the theorem with

$$K = \sqrt{\frac{K_2}{K_1}}. \tag{2.127}$$

□

2.3 Absorbing boundary conditions

2.3.1 Simple 1D acoustic example

The first category of methods to address the boundary conditions problem is the design of absorbing boundary conditions. The philosophy here is to prescribe specific boundary conditions on the edges of the computation domain so as to minimize the energy of reflected waves.

Let us introduce this concept in the simple frame of 1D acoustic wave equation. With these assumptions, (2.90) rewrites

$$\begin{cases} \dfrac{\partial v}{\partial t} - \dfrac{1}{\rho}\dfrac{\partial P}{\partial x} = 0 \\ \\ \dfrac{\partial P}{\partial t} - \rho c^2 \dfrac{\partial v}{\partial x} = f \end{cases} \tag{2.128}$$

We can rewrite (2.128) as a second-order in time partial differential equation on the pressure field only. We first derive the second equation in time and replace $\frac{\partial v}{\partial t}$ inside it using the first equation. This yields

$$\frac{\partial^2 P}{\partial t^2} - \rho c^2 \frac{\partial}{\partial x}\left(\frac{1}{\rho}\frac{\partial P}{\partial x}\right) = s, \tag{2.129}$$

with

$$s = \frac{\partial f}{\partial t}. \tag{2.130}$$

We will further assume the density is constant to make things even simpler. In this case the 1D acoustic equation boils down to the simple 1D scalar wave equation

$$\frac{\partial^2 P}{\partial t^2} - c^2 \frac{\partial^2 P}{\partial x^2} = s, \tag{2.131}$$

Interestingly, (2.131) can be factorized as

$$\left(\frac{\partial}{\partial t} - c\frac{\partial}{\partial x}\right)\left(\frac{\partial}{\partial t} + c\frac{\partial}{\partial x}\right) P = s. \tag{2.132}$$

Assuming the source is localized in space, we have, away from the source

$$\left(\frac{\partial}{\partial t} - c\frac{\partial}{\partial x}\right)\left(\frac{\partial}{\partial t} + c\frac{\partial}{\partial x}\right) P = 0, \tag{2.133}$$

which implies that the general solution of (2.131) is the superposition of

- 1 wave traveling in the direction $x > 0$, at speed c, solution of

$$\left(\frac{\partial}{\partial t} - c\frac{\partial}{\partial x}\right) P = 0. \tag{2.134}$$

- 1 wave traveling in the direction $x < 0$, at speed c, solution of

$$\left(\frac{\partial}{\partial t} + c\frac{\partial}{\partial x}\right) P = 0. \tag{2.135}$$

To define proper absorbing boundary conditions, it is thus sufficient to impose on each edge that the component traveling inward should be equal to 0. In this case, if we denote the 1D medium by the segment $[-L\ L] \subset \mathbb{R}$, we would impose

$$\begin{cases} \dfrac{\partial P}{\partial t}(L,t) + c\dfrac{\partial P}{\partial x}(L,t) = 0, \\ \\ \dfrac{\partial P}{\partial t}(-L,t) - c\dfrac{\partial P}{\partial x}(-L,t) = 0, \end{cases} \tag{2.136}$$

In practice, these conditions can still be imposed in multi-dimensional context. If we consider for instance the acoustic system (2.90), we can impose this type of absorbing boundary conditions on each side of the volume where we would like to mimic a transparent wall. In this context however, this type of boundary conditions will be exact only for waves arriving at normal incidence with velocity c. For waves arriving at grazing incidence, the reflection coefficient will not be negligible, resulting in spurious reflections.

2.3.2 Generalization to elastodynamics

How does this generalize to more complex elastodynamics equations? To answer this question it is enough to go back to the general hyperbolic form of the elastodynamics equations (2.103). The first step is to symmetrize the system. We introduce the following definition

Definition 2. *A hyperbolic system is symmetrizable if and only if there exists a symmetric positive definite matrix $S(x)$ such that*

$$(SA_j)^T = SA_j, \;\; j = 1, \ldots, d \tag{2.137}$$

We can first prove that $\widetilde{S}$ that we have introduced previously is a symmetrizer for our system of elastodynamics equations. Indeed we have demonstrated that we could write it as

$$\widetilde{S}\frac{\partial w}{\partial t} - \sum_{j=1}^{3} X_j \frac{\partial w}{\partial x_j} = \widetilde{F}, \tag{2.138}$$

which is equivalent to

$$\frac{\partial w}{\partial t} - \sum_{j=1}^{3} \widetilde{S}^{-1} X_j \frac{\partial w}{\partial x_j} = \widetilde{F}. \tag{2.139}$$

The latter expression is the general form with

$$A_j = \widetilde{S}^{-1} X_j \tag{2.140}$$

As X_j are symmetric matrices, $\widetilde{S}$ is a symmetrizer for our elastodynamics system.

Now consider a generic symmetrizer S, with square root B such that $S = B^T B$ for a general hyperbolic system of the form

$$\frac{\partial u}{\partial t} + \sum_{j=1}^{d} A_j \frac{\partial u}{\partial x_j} + A_0 u = f. \tag{2.141}$$

To symmetrize the general first-order hyperbolic system, we first multiply by B to obtain

$$\frac{\partial (Bu)}{\partial t} + \sum_{j=1}^{d} BA_j \frac{\partial u}{\partial x_j} + BA_0 u = Bf \tag{2.142}$$

Next we rely on the identity

$$BA_jB^{-1}\frac{\partial (Bu)}{\partial x_j} = BA_jB^{-1}B\frac{\partial u}{\partial x_j} + BA_jB^{-1}\frac{\partial B}{\partial x_j}u \tag{2.143}$$

that is

$$BA_jB^{-1}\frac{\partial (Bu)}{\partial x_j} = BA_j\frac{\partial u}{\partial x_j} + BA_jB^{-1}\frac{\partial B}{\partial x_j}u \tag{2.144}$$

Therefore

$$BA_j\frac{\partial u}{\partial x_j} = BA_jB^{-1}\frac{\partial\left(Bu\right)}{\partial x_j} - BA_jB^{-1}\frac{\partial B}{\partial x_j}u \tag{2.145}$$

and finally

$$BA_j\frac{\partial u}{\partial x_j} = BA_jB^{-1}\frac{\partial\left(Bu\right)}{\partial x_j} - BA_jB^{-1}\frac{\partial B}{\partial x_j}B^{-1}Bu \tag{2.146}$$

In the same spirit we have

$$BA_0u = BA_0B^{-1}Bu \tag{2.147}$$

and we can rewrite the system as

$$\frac{\partial\left(Bu\right)}{\partial t} + \sum_{j=1}^{d} BA_jB^{-1}\frac{\partial\left(Bu\right)}{\partial x_j} - \sum_{j=1}^{d} BA_j\frac{\partial B}{\partial x_j}B^{-1}Bu + BA_0B^{-1}Bu = Bf \tag{2.148}$$

Introducing $w = Bu$ this yields

$$\frac{\partial w}{\partial t} + \sum_{j=1}^{d} BA_jB^{-1}\frac{\partial w}{\partial x_j} - \sum_{j=1}^{d} BA_j\frac{\partial B}{\partial x_j}B^{-1}w + BA_0B^{-1}w = Bf \tag{2.149}$$

which we can rewrite as

$$\frac{\partial w}{\partial t} + \sum_{j=1}^{d} \widetilde{A_j}\frac{\partial w}{\partial x_j} + \widetilde{A}_0 = Bf \tag{2.150}$$

with

$$\widetilde{A}_j = BA_jB^{-1},\ \ \widetilde{A}_0 = BA_0B^{-1} - \sum_{j=1}^{d} BA_j(x)B^{-1}\frac{\partial B}{\partial x_j}B^{-1} \tag{2.151}$$

The matrices $\widetilde{A}_j = BA_jB^{-1}$ are symmetric.

Proof.

$$\exists C > 0,\ \ SA_j = C^TC. \tag{2.152}$$

Therefore

$$A_j = B^{-1}B^{-T}C^TC, \tag{2.153}$$

and

$$BA_jB^{-1} = B^{-T}C^TCB^{-1} = \left(CB^{-1}\right)^T CB^{-1}. \tag{2.154}$$

□

This shows that if a symmetrizer exists, it can be used to define a change of variables such that with this new variables we have a symmetric hyperbolic system. To capture the different components of the solution traveling at different speeds on a **symmetrized hyperbolic system**, it is then sufficient

- to compute the eigenvalues $\alpha_{j,k}$ of the matrices A_j: this gives the velocities of the different phases traveling in the medium
- to compute the eigenvectors associated with these eigenvalues: the projection of the solution on the eigenspace $\mathrm{Ker}(I - \alpha_{j,k}A_j)$ gives the quantity traveling at speed α_k in the direction j.

- On a specific border, we apply the same method as previously: the absorbing boundary conditions consists in canceling the components traveling inward.

We do not intend to prove formally the previous results. Alternatively, we are going to test if this strategy gives consistent results for the previous 1D acoustic system. We start with the velocity stress equations (2.128)

In this case we have

$$u = [v\ P]^T,\ \ A = \begin{pmatrix} 0 & \frac{1}{\rho} \\ \rho c^2 & 0 \end{pmatrix},\ \ A_0 = 0 \tag{2.155}$$

A symmetrizer for A is

$$S = \begin{pmatrix} \rho^2 c^2 & 0 \\ 0 & 1 \end{pmatrix}, \tag{2.156}$$

A square root B of the symmetrizer is thus

$$B = \begin{pmatrix} \rho c & 0 \\ 0 & 1 \end{pmatrix},\ \ B^{-1} = \begin{pmatrix} \frac{1}{\rho c} & 0 \\ 0 & 1 \end{pmatrix}. \tag{2.157}$$

Symmetrizing the 1D velocity-stress acoustic system thus yields

$$\frac{\partial (Bu)}{\partial t} - BAB^{-1}\frac{\partial (Bu)}{\partial x} + BAB^{-1}\frac{\partial B}{\partial x}B^{-1}(Bu) = 0 \tag{2.158}$$

We have

$$Bu = \begin{pmatrix} \rho c v \\ P \end{pmatrix},\ \ BAB^{-1} = \begin{pmatrix} 0 & c \\ c & 0 \end{pmatrix},\ \ \frac{\partial B}{\partial x} = \begin{pmatrix} \frac{\partial (\rho c)}{\partial x} & 0 \\ 0 & 0 \end{pmatrix} \tag{2.159}$$

and thus

$$BAB^{-1}\frac{\partial B}{\partial x} = \begin{pmatrix} 0 & 0 \\ c\frac{\partial (\rho c)}{\partial x} & 0 \end{pmatrix}, \tag{2.160}$$

and

$$BAB^{-1}\frac{\partial B}{\partial x}B^{-1} = \begin{pmatrix} 0 & 0 \\ \frac{1}{\rho}\frac{\partial (\rho c)}{\partial x} & 0 \end{pmatrix}, \tag{2.161}$$

If we expand the symmetric system, we have

$$\begin{cases} \dfrac{\partial\,(\rho c v)}{\partial t} - c\dfrac{\partial P}{\partial x} = 0 \\ \dfrac{\partial P}{\partial t} - c\dfrac{\partial (\rho c v)}{\partial x} + \dfrac{1}{\rho}\dfrac{\partial (\rho c)}{\partial x}(\rho c v) = 0 \end{cases} \tag{2.162}$$

This is equivalent to

$$\begin{cases} \dfrac{\partial v}{\partial t} - \dfrac{1}{\rho}\dfrac{\partial P}{\partial x} = 0 \\ \dfrac{\partial P}{\partial t} - \rho c^2\dfrac{\partial v}{\partial x} - c\dfrac{\partial (\rho c)}{\partial x}v + \dfrac{1}{\rho}\dfrac{\partial (\rho c)}{\partial x}(\rho c v) = 0 \end{cases} \tag{2.163}$$

Therefore this is equivalent to

$$\begin{cases} \dfrac{\partial v}{\partial t} - \dfrac{1}{\rho}\dfrac{\partial P}{\partial x} = 0 \\ \dfrac{\partial P}{\partial t} - \rho c^2\dfrac{\partial v}{\partial x} = 0 \end{cases} \tag{2.164}$$

We recover the previously analyzed first-order 1D acoustic velocity-stress system (2.128), which is consistent (safety check!)

Let us now analyze the eigenvalues of the symmetrized system. We can obtain the eigenvalues of BAB^{-1} by solving

$$\det(BAB^{-1} - XI_2) = 0 \tag{2.165}$$

which gives

$$\det \begin{pmatrix} -X & c \\ c & -X \end{pmatrix} = 0 \tag{2.166}$$

that is

$$X^2 - c^2 = 0 \tag{2.167}$$

Therefore BAB^{-1} has two eigenvalues: $\lambda_1 = c, \lambda_2 = -c$. This gives two velocities, one propagating to the left at speed c, the other propagating to the right at the same speed c. This is already consistent with what we observed previously with the factorization of the 1D equation as two transport equations (2.134) and (2.135).

Now we compute the corresponding eigenvectors e_1 and e_2. Let us start with e_1. We solve the linear system

$$BAB^{-1}x = cx \tag{2.168}$$

which gives

$$\begin{cases} cx_2 &= cx_1 \\ cx_1 &= cx_2 \end{cases} \tag{2.169}$$

Therefore $x_1 = x_2$ and we can define e_1 as $e_1 = [1\ 1]^T$. For e_2 we solve the linear system

$$BAB^{-1}x = -cx \tag{2.170}$$

which gives

$$\begin{cases} cx_2 &= -cx_1 \\ cx_1 &= -cx_2 \end{cases} \tag{2.171}$$

Therefore $x_1 = -x_2$ and we can define e_2 as $e_2 = [1\ -1]^T$.

The projectors on the corresponding eigenspaces are

$$P_i = e_i e_i^T, \tag{2.172}$$

which gives

$$P_1 = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \quad P_2 = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \tag{2.173}$$

The projection of the solution u on $\mathrm{Ker}(I - \alpha_1 BAB^{-1})$ is given by

$$P_1(Bu) = \begin{pmatrix} \rho cv + P \\ \rho cv + P \end{pmatrix} \tag{2.174}$$

The condition on the right boundary is thus

$$\rho cv(L,t) + P(L,t) = 0 \tag{2.175}$$

Deriving this condition in time gives

$$\rho c \frac{\partial v}{\partial t}(L,t) + \frac{\partial P}{\partial t}(L,t) = 0 \tag{2.176}$$

Replacing $\frac{\partial v}{\partial t}(L,t)$ with equation (2.128) gives

$$c\frac{\partial P}{\partial x}(L,t) + \frac{\partial P}{\partial t}(L,t) = 0 \tag{2.177}$$

which is consistent with the previously derived absorbing boundary condition for the 1D acoustic system (2.136).

If we look at the condition on the left boundary we have

$$\rho c v(-L,t) - P(-L,t) = 0 \tag{2.178}$$

Similarly

$$\rho c\frac{\partial v}{\partial t}(-L,t) - \frac{\partial P}{\partial t}(-L,t) = 0 \tag{2.179}$$

and replacing $\frac{\partial v}{\partial t}(-L,t)$ with equation (2.128) gives

$$c\frac{\partial P}{\partial x}(-L,t) - \frac{\partial P}{\partial t}(-L,t) = 0 \tag{2.180}$$

which is also consistent with the previously derived absorbing boundary condition for the 1D acoustic system (2.136).

2.3.3 Higher order absorbing boundary conditions

The absorbing boundary conditions we have presented so far are first-order absorbing boundary conditions (based on first order derivatives). The efficiency of such approach is known to be relatively limited. For waves arriving at normal incidence on the interfaces, the reflection coefficient is close to 0 (up to discretization error). However, as soon as the incidence angle vary from the normal, the reflection coefficient becomes not negligible, the worst case being waves arriving at grazing angles.

To improve the efficiency of absorbing boundary conditions, a considerable research work, mainly in the applied math community, has been devoted to the design of higher order absorbing boundary conditions (Collino, 1993). After years of investigation, it turns out that improved accuracy can be obtained, however to the cost of potentially cumbersome implementation, and with particularities associated with the wave propagation problems of interest (while we have seen that first-order absorbing boundary conditions can be generalized easily to any symmetrizable hyperbolic systems). The main issue is that the use of high order derivatives operator loses the locality property of the boundary conditions. Non-local operators yields difficulties in terms of implementation, especially when parallelism through domain decomposition is involved.

In practice, another class of strategies, named absorbing layers, has emerged. We present it in the next section.

2.3.4 Absorbing layers

Sponge layers

The leading concept behind absorbing layers is rather simple. Instead of specifying boundary conditions, a layer is added around the domain of interest in which an artificial damping coefficient is introduced. In terms of mathematics, it amounts to introduce a diagonal zero-order term in the original hyperbolic system (2.103)

$$\frac{\partial u}{\partial t} + \sum_{j=1}^{d} A_j(x)\frac{\partial u}{\partial x_j} + A_0(x)u + \sigma(x)u = f \tag{2.181}$$

The function $\sigma(x)$ is referred to as a damping function. This function should be 0 in the domain of interest, and growth smoothly in the layer. Basics formula are based on polynomial, cosine, or exponential functions.

This relatively basic strategy is known as sponge layer technique. It has been introduced by Cerjan et al. (1985) using a practitioner approach. In practice it is unconditionally stable however it generates also non negligible artificial reflections at the boundaries, where the damping function $\sigma(x)$ transition from 0 to non-zero positive values.

Perfectly Matched Layers

A more elaborate layer method, named as perfectly matched layer (PML) has been introduced by Bérenger (1994) for electromagnetic (Maxwell's equation). Contrary to sponge layers, PML have the property to yield null reflection coefficient at the interface between the domain of the interest and the absorbing layer. This technique relies on the definition of a split PDE system in the layer, coupled with the original PDE system in the domain of interest. The description of this method is beyond the scope of this course, however interested reader might refer to Collino and Tsogka (2001) for an introduction of this method in the frame of elastodynamics equations.

One difficulty with PML is their potential instability as soon as anisotropic media are considered. They are also originally designed for first-order hyperbolic systems, which are not the one numerical methods solve in practice (see next Section 2.4). The definition of PML equations for second-order in time systems is more involved in terms of implementation. For these two reasons, in practice, for elastodynamics equations, a combination of sponge layers and first-order absorbing conditions is used. The first-order absorbing boundary condition is defined at the external border of the sponge layers. This combination makes possible to efficiently damp outgoing waves using moderately thick layers. The accuracy is controlled by the thickness layer.

2.4 Computing numerical approximation of solutions of the elastodynamic equations

In general a very good book to make the first step towards implementation of numerical schemes for the solution of elastodynamics equations is the one by Heiner Igel (Igel, 2017).

2.4.1 Generalities

The elastodynamics equations which have been introduced and analyzed so far do not exhibit analytical solutions unless in very specific cases, namely in homogeneous media. As soon as the property of the medium, namely the density and the stiffness tensor coefficients, depend on space, no simple analytical solution can be found.

Solving these equations thus need to rely on numerical approximations. Four main families of numerical methods for solving partial differential equations might be identified

- the finite difference method;
- the spectral or pseudo-spectral method;
- the finite element method;
- the finite volume method.

Finite difference is the simplest numerical strategy among this four. It relies on the computation of the solution on the nodes of a predefined Cartesian grid. The partial differential operators are approximated on this grid, using formulas based on Taylor's development of the solution (therefore assuming sufficient regularity of the solution, at least up to the order of the method which is used).

The pseudo-spectral method relies on a Fourier expansion of the solution to compute its spatial derivatives, while the discretization in the time domain remains similar as for finite difference method. This makes possible to compute the spatial derivatives up to machine precision, however it requires intensive use of Fourier and inverse Fourier transform (which can be done through fast Fourier transform algorithms). The pseudo-spectral is not really used nowadays because the spatial derivative operator approximations are done globally and would require communication between all the points of the grid in a parallel environment. This non-locality makes it not possible to be used on parallel computing platforms, which is the state-of-the-art for nowadays realistic scale applications.

The finite element method relies on the decomposition of the solution on a certain basis of polynomial of arbitrary orders, and the design of a mesh on which are attached this polynomial. Depending on the type of mesh and the type of basis which are used, different flavors of finite element strategy can be derived. Finite element methods are very powerful, however less easy to implement than finite-difference methods. One key advantage is the ability to adapt the mesh to specific interfaces (free surface, internal discontinuities) which is not possible using a conventional finite difference method on a Cartesian mesh.

The finite volume method is designed for the solution of partial differential equations for which the solution is known to exhibit discontinuities. Excepted specific cases for wave propagation in the subsurface, this situation is not met and finite volume methods are not used in this frame.

In this Section we will illustrate how the finite-difference method can be used efficiently to solve acoustic equations. This method is very powerful for marine environment with flat topography where the free surface can be imposed very easily.

We will then introduce how a specific finite element methods, namely the spectral element method (not to be confused with pseudo-spectral methods), can be used to approximate the solution of general elastodynamics equations. In this case, the ability of this finite element method to account for varying topography, in conjunction with a specific choice of polynomial basis to represent the solution, makes it particularly powerful to solve elastodynamics equations. This latter method is now the state of the art in seismology to solve elastodynamics equations.

2.4.2 Finite-differences

Constant density acoustic equations

For the sake of simplicity we consider here the constant-density acoustic equations we have already introduced (2.131) in 1D. In 3D, we obtain

$$\frac{\partial^2 P}{\partial t^2} - c^2 \Delta P = s \tag{2.182}$$

where Δ is the Laplacian operator

$$\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \tag{2.183}$$

and s is a generic pressure source term.

The domain of interest is

$$\Omega \subset \mathbb{R}^3. \tag{2.184}$$

We assume zero initial conditions

$$P(x,y,z,0) = 0, \quad \frac{\partial P}{\partial t}(x,y,z,0) = 0, \tag{2.185}$$

and homogeneous Dirichlet boundary conditions

$$\forall (x,y,z) \in \partial\Omega, \quad P(x,y,z,t) = 0, \tag{2.186}$$

where $\partial\Omega$ denotes the boundary of Ω.

Second-order finite difference of second-order derivatives

We recall the definition of the notation $u(x) = \underset{h\to 0}{o}(v(x))$:

$$u(x) = \underset{x\to 0}{o}(v(x)) \Longleftrightarrow \lim_{x\to 0} \frac{u(x)}{v(x)} = 0 \tag{2.187}$$

The Taylor development formula for a given function $f(x)$ which is C^∞ is

$$\begin{aligned} f(x+h) &= f(x) + hf'(x) + h^2 \frac{f''(x)}{2!} + \cdots + h^n \frac{f^{(n)}(x)}{n!} + \underset{h\to 0}{o}(h^n) \\ f(x-h) &= f(x) - hf'(x) + h^2 \frac{f''(x)}{2!} + \cdots + (-h)^n \frac{f^{(n)}(x)}{n!} + \underset{h\to 0}{o}(h^n). \end{aligned} \tag{2.188}$$

By summation of these two developments we get

$$f(x+h)+f(x-h)=2f(x)+h^2f''(x)+\frac{h^4}{12}f^{(4)}(x)+\underset{h\to 0}{o}(h^4). \tag{2.189}$$

Therefore

$$f''(x)=\frac{f(x+h)-2f(x)+f(x-h)}{h^2}=\frac{h^2}{12}f^{(4)}(x)+\underset{h\to 0}{o}(h^2), \tag{2.190}$$

which is a second-order approximation of $f''(x)$. This is the formula we will use to approximate the solution of the acoustic equation through finite-differences.

Cartesian grid

We consider a simple 3D grid where the spatial discretization steps are such that

$$\Delta x=\Delta y=\Delta z=h\in\mathbb{R}_+^*. \tag{2.191}$$

We solve the acoustic equations on a 3D volume we denote as

$$\Omega=[0\ \ L_x]\times[0\ \ L_y]\times[0\ \ L_z]\subset\mathbb{R}^3, \tag{2.192}$$

where

$$(L_x,L_y,L_z)\in\left(\mathbb{R}_+^*\right)^3. \tag{2.193}$$

The time discretization is performed with a step

$$\Delta t\in\mathbb{R}_+^*, \tag{2.194}$$

and the time interval is

$$[0\ \ T]\subset\mathbb{R}, \tag{2.195}$$

with the total time duration of the modeling being

$$T\in\mathbb{R}_+^*. \tag{2.196}$$

The spatial discrete grid is Ω_h such that

$$\Omega_h=\left\{x_i,y_j,z_k,\ \ (i,j,k)\in\mathbb{N}^3\right\}, \tag{2.197}$$

where

$$x_i=ih,\ \ y_j=jh,\ \ z_k=kh. \tag{2.198}$$

The time discretization is performed at time t^n such that

$$t^n=(n-1)\Delta t. \tag{2.199}$$

Finite difference scheme

We introduce the conventional notations for the value of the pressure on the nodes of the finite difference grid

$$P_{ijk}^n=P(x_i,y_j,z_k,t^n) \tag{2.200}$$

We can approximate the Laplacian operator Δ using the second-order finite difference formula (2.190). With the previously introduced notations, this gives

$$\begin{aligned}&\Delta P(x_i, y_i, z_i, t^n) \simeq \\ &\frac{P^n_{i+1jk} - 2P^n_{ijk} + P^n_{i-1jk}}{h^2} + \frac{P^n_{ij+1k} - 2P^n_{ijk} + P^n_{ij-1k}}{h^2} + \frac{P^n_{ijk+1} - 2P^n_{ijk} + P^n_{ijk-1}}{h^2}\end{aligned} \tag{2.201}$$

Similarly, the second-order time derivatives is approximated as

$$\frac{\partial^2 P}{\partial t^2}(x_i, y_j, z_k, t^n) \simeq \frac{P^{n+1}_{ijk} - 2P^n_{ijk} + P^{n-1}_{ijk}}{\Delta t^2} \tag{2.202}$$

This yields the finite difference approximation

$$\begin{aligned}&\frac{P^{n+1}_{ijk} - 2P^n_{ijk} + P^{n-1}_{ijk}}{\Delta t^2} - \\ &c^2_{ijk}\left(\frac{P^n_{i+1jk} - 2P^n_{ijk} + P^n_{i-1jk}}{h^2} + \frac{P^n_{ij+1k} - 2P^n_{ijk} + P^n_{ij-1k}}{h^2} + \frac{P^n_{ijk+1} - 2P^n_{ijk} + P^n_{ijk-1}}{h^2}\right) \\ &= s^n_{ijk}\end{aligned} \tag{2.203}$$

We can rewrite things under the form of an explicit formula for updating P^{n+1}_{ijk} such that

$$\begin{aligned}P^{n+1}_{ijk} \;=\;& 2P^n_{ijk} - P^{n-1}_{ijk} + \\ &+\frac{c^2_{ijk}\Delta t^2}{h^2}\left(P^n_{i+1jk} - 2P^n_{ijk} + P^n_{i-1jk} + P^n_{ij+1k} - 2P^n_{ijk} + P^n_{ij-1k} + P^n_{ijk+1} - 2P^n_{ijk} + P^n_{ijk-1}\right) \\ &+\Delta t^2 s^n_{ijk}.\end{aligned} \tag{2.204}$$

Zero initial conditions also give

$$P^1_{ijk} = 0, \quad \frac{P^2_{ijk} - P^1_{ijk}}{\Delta t} = 0. \tag{2.205}$$

assuming a first order discretization of the time derivative at time $t = 0$. Thus

$$P^1_{ijk} = 0, \quad P^2_{ijk} = 0. \tag{2.206}$$

In addition, boundary conditions give

$$P^n_{ijk} = 0, \quad \forall\,(x_i, y_j, z_k) \in \partial\Omega_h. \tag{2.207}$$

In practice, how to choose the discretization steps h and Δt? This choice is driven by two fundamentals requirement

- stability of the scheme
- small enough numerical dispersion

In the next two paragraphs, we detail these two notions, and explain how to choose h and Δt to fulfill these requirements. We provide this analysis in the simple case of the 1D constant density constant velocity acoustic equation

$$\frac{\partial^2 P}{\partial t^2} - c^2 \frac{\partial^2 P}{\partial x^2} = s, \tag{2.208}$$

discretized using the second-order finite-difference scheme

$$\frac{P_i^{n+1} - 2P_i^n + P_i^{n-1}}{\Delta t^2} - c^2 \frac{P_{i+1}^n - 2P_i^n + P_{i-1}^n}{h^2} = s_i^n. \tag{2.209}$$

Stability analysis

We start with the stability. We follow the conventional Von Neumann analysis. First we define the truncation error. We assume that the analytic solution $P(x,t)$ satisfies the discretized wave equation with an error related to the truncation of the partial differential operator following the Taylor's development. This is the truncation error, which we define by

$$\eta_i^n = \frac{P(x_i, t^n + \Delta t) - 2P(x_i, t^n) + P(x_i, t^n - \Delta t)}{\Delta t^2} - c^2 \frac{P(x_i + h, t^n) - 2P(x_i, t^n) + P(x_i - h, t^n)}{h^2}. \tag{2.210}$$

From the definition of the second-order finite difference formulas, it is easy to show that the truncation error satisfies

$$\eta_i^n = O(\Delta t^2) + O(h^2). \tag{2.211}$$

We now define the convergence error. The convergence error is

$$\varepsilon_i^n = P(x_i, t^n) - P_i^n, \tag{2.212}$$

where $P(x,t)$ is the analytical solution and P_i^n the numerical solution provided by the finite-difference scheme. What is important to see is that the convergence error obeys the same discrete equation as the numerical solution. We have

$$\frac{\varepsilon_i^{n+1} - 2\varepsilon_i^n + \varepsilon_i^{n-1}}{\Delta t^2} = \frac{P(x_i, t^n + \Delta t) - 2P(x_i, t^n) + P(x_i, t^n - \Delta t)}{\Delta t^2} - \frac{P_i^{n+1} - 2P_i^n + P_i^{n-1}}{\Delta t^2}, \tag{2.213}$$

and

$$\frac{\varepsilon_{i+1}^n - 2\varepsilon_i^n + \varepsilon_{i-1}^n}{h^2} = \frac{P(x_i + h, t^n) - 2P(x_i, t^n + P(x_i - h, t^n)}{h^2} - \frac{P_{i+1}^n - 2P_i^n + P_{i-1}^n}{h^2}. \tag{2.214}$$

Therefore

$$\frac{\varepsilon_i^{n+1} - 2\varepsilon_i^n + \varepsilon_i^{n-1}}{\Delta t^2} - c^2 \frac{\varepsilon_{i+1}^n - 2\varepsilon_i^n + \varepsilon_{i-1}^n}{h^2} = \eta_i^n - \frac{P_i^{n+1} - 2P_i^n + P_i^{n-1}}{\Delta t^2} - c^2 \frac{P_{i+1}^n - 2P_i^n + P_{i-1}^n}{h^2}, \tag{2.215}$$

which yields by definition of the numerical scheme

$$\frac{\varepsilon_i^{n+1} - 2\varepsilon_i^n + \varepsilon_i^{n-1}}{\Delta t^2} - c^2 \frac{\varepsilon_{i+1}^n - 2\varepsilon_i^n + \varepsilon_{i-1}^n}{h^2} = \eta_i^n. \tag{2.216}$$

We now consider an expansion of the convergence error as a Fourier series of the form

$$\varepsilon(x,t) = \sum_k A_k(t) e^{ikx}. \tag{2.217}$$

We look at a single component of this Fourier expansion, at time t^n and point x_i,

$$\varepsilon_i^n = A^n e^{ikx_i}. \tag{2.218}$$

We consider the truncation error can be neglected (we are interested here only in the convergence error evolution and can look at the specific case where the truncated error is 0) and plug this component in the discrete wave equation to obtain

$$\frac{A^{n+1}e^{ikx_i} - 2A^n e^{ikx_i} + A^{n-1}e^{ikx_i}}{\Delta t^2} - c^2 \frac{e^{ik(x_i+h)} - 2e^{ikx_i} + e^{ik(x_i-h)}}{h^2} A^n = 0. \tag{2.219}$$

Dividing by e^{ikx_i} yields

$$\frac{A^{n+1} - 2A^n + A^{n-1}}{\Delta t^2} - c^2 \frac{e^{ikh} - 2 + e^{-ikh}}{h^2} A^n = 0. \tag{2.220}$$

We can rewrite (2.220) as

$$A^{n+1} = -A^{n-1} + 2BA^n, \tag{2.221}$$

where

$$B = 1 + \frac{\Delta t^2 c^2}{2h^2}(e^{ikh} - 2 + e^{-ikh}). \tag{2.222}$$

We introduce

$$\gamma = \frac{A^{n+1}}{A^n}. \tag{2.223}$$

We thus have

$$\gamma^2 - 2B\gamma + 1 = 0. \tag{2.224}$$

To ensure stability, we want to have

$$|\gamma| < 1. \tag{2.225}$$

This is to make sure that the amplitude of the Fourier component we consider is not amplified at each iteration in time, which would lead to an exponential increase of the amplitude of the solution (explosion in finite-time). The discriminant of the second-order polynomial is

$$\Delta = 4B^2 - 4. \tag{2.226}$$

The condition $\Delta \geq 0$ is equivalent to $|B| \geq 1$, and we have in this case

$$\gamma = B \pm \sqrt{B^2 - 1}, \tag{2.227}$$

therefore $|\gamma| \geq 1$. $\Delta < 0$ is equivalent to $|B| < 1$ and

$$\gamma = B \pm i\sqrt{|B^2 - 1|}. \tag{2.228}$$

In this case

$$|\gamma|^2 = B^2 + B^2 - 1 = 2B^2 - 1, \tag{2.229}$$

and we have $|\gamma| < 1$, as

$$0 < B^2 < 1, \tag{2.230}$$

then

$$0 < 2B^2 < 2, \tag{2.231}$$

and

$$-1 < 2B^2 - 1 < 1. \tag{2.232}$$

The stability condition is thus ensured by $|B| < 1$, which is equivalent to

$$-1 \leq 1 + \frac{\Delta t^2 c^2}{2h^2}(e^{ikh} - 2 + e^{-ikh}) \leq 1. \tag{2.233}$$

This can be rewritten as

$$-1 \leq 1 + \frac{\Delta t^2 c^2}{h^2}(\cos(kh) - 1) \leq 1. \tag{2.234}$$

We recall the identity

$$\left(\sin\frac{a}{2}\right)^2 = \frac{1 - \cos(a)}{2}. \tag{2.235}$$

This comes from the fact that

$$\cos(a + b) = \cos a \cos b - \sin a \sin b, \tag{2.236}$$

then

$$\cos 2a = \cos^2 a - \sin^2 a = 1 - 2\sin^2 a. \tag{2.237}$$

Hence

$$\sin^2 a = \frac{1 - \cos 2a}{2}. \tag{2.238}$$

Using this identity we obtain

$$-1 \leq 1 - \frac{2\Delta t^2 c^2}{h^2}\left(\sin\frac{kh}{2}\right)^2 \leq 1, \tag{2.239}$$

hence

$$-2 \leq -\frac{2\Delta t^2 c^2}{h^2}\left(\sin\frac{kh}{2}\right)^2 \leq 0, \tag{2.240}$$

and

$$0 \leq \frac{\Delta t^2 c^2}{h^2}\left(\sin\frac{kh}{2}\right)^2 \leq 1. \tag{2.241}$$

The latter is ensure by the choice

$$\frac{\Delta t^2 c^2}{h^2} \leq 1, \tag{2.242}$$

hence

$$\frac{\Delta t c}{h} \leq 1. \tag{2.243}$$

This corresponds to the standard CFL condition. Expressing it as

$$\Delta t \leq \frac{\Delta h}{c}, \tag{2.244}$$

makes it clear that Δt is constrained by

- the size of the mesh: the finer the mesh, the smaller the time step;
- the wave velocity: the fastest the wave speed, the smaller the time step.

This is important in terms of application and computation cost: we go back to this after the analysis of the numerical dispersion.

Numerical dispersion

For the numerical dispersion analysis, we consider a plane-wave solution of the form

$$P(x,t) = e^{i(kx-\omega t)}. \tag{2.245}$$

Plugging this ansatz in the wave equation yields

$$(-i\omega)^2 e^{i(kx-\omega t)} - c^2(ik)^2 e^{i(kx-\omega t)} = 0. \tag{2.246}$$

We thus have

$$-\omega^2 + c^2k^2 = 0, \tag{2.247}$$

which can be rewritten simply as

$$c = \frac{\omega}{k}. \tag{2.248}$$

This equation is known as the dispersion relation. The temporal and spatial oscillations are related to the velocity. In the analytic case, the velocity is the same for all oscillations k and w. The medium is not dispersive: the propagation velocity does not depend on the frequency of propagation.

Let us now use the same ansatz in the discretized wave equation. We have

$$\frac{e^{i(kx-\omega(t+\Delta t))} - 2e^{i(kx-\omega t)} + e^{i(kx-\omega(t-\Delta t))}}{\Delta t^2} - c^2 \frac{e^{i(k(x+h)-\omega t)} - 2e^{i(kx-\omega t)} + e^{i(k(x-h)-\omega t)}}{h^2} = 0, \tag{2.249}$$

which gives

$$e^{-i\omega\Delta t} - 2 + e^{i\omega\Delta t} = c^2 \frac{\Delta t^2}{h^2}\left(e^{-ikh} - 2 + e^{ikh}\right), \tag{2.250}$$

by simplification by $e^{i(kx-\omega t)}$. Thus

$$\cos\omega\Delta t - 1 = c^2 \frac{\Delta t^2}{h^2}\left(\cos kh - 1\right), \tag{2.251}$$

or, using the same trigonometric identity as previously,

$$\sin^2\frac{\omega\Delta t}{2} = c^2\frac{\Delta t^2}{h^2}\sin^2\frac{kh}{2}. \tag{2.252}$$

Hence we have

$$\sin\frac{\omega\Delta t}{2} = c\frac{\Delta t}{h}\sin\frac{kh}{2}. \tag{2.253}$$

Here we can note that the latter equation has a solution if and only if the CFL condition is satisfied, otherwise the modulus of the right hand side could be strictly larger than 1 while the modulus of the left hand side is bounded by 1.

From the latter equation we have

$$\frac{\omega\Delta t}{2} = \sin^{-1}\left(c\frac{\Delta t}{h}\sin\frac{kh}{2}\right), \tag{2.254}$$

then

$$\omega = \frac{2}{\Delta t}\sin^{-1}\left(c\frac{\Delta t}{h}\sin\frac{kh}{2}\right), \tag{2.255}$$

which gives the numerical dispersion relation

$$c(k) = \frac{\omega}{k} = \frac{2}{k\Delta t} \sin^{-1}\left(c\frac{\Delta t}{h} \sin\frac{kh}{2}\right). \tag{2.256}$$

We see that in the limit $h \to 0$, we have

$$\frac{\omega}{k} \to \frac{2}{k\Delta t} \sin^{-1}\left(c\frac{\Delta t k}{2}\right). \tag{2.257}$$

Then in the limit $\Delta t \to 0$, we recover the analytical dispersion relation

$$\frac{\omega}{k} \to c. \tag{2.258}$$

However, in practice, we cannot choose arbitrarily small time steps and spatial discretization steps. Therefore it is interesting to see how the discrete propagation velocity varies with the choice of discretization setups. In practice, the choice of Δt is already governed by the previous stability condition, therefore it is common to study the evolution of the discrete propagation velocity with respect to the spatial discretization step h.

Consider n_w the number of discretization points per minimal wavelength. We have

$$n_w = \frac{\lambda_{min}}{h} \tag{2.259}$$

where

$$\lambda_{min} = \frac{c}{f_{max}} \tag{2.260}$$

We present in Figure 2.9 the numerical velocity $\frac{\omega}{k}$ for a choice of $k = 0.05 = \frac{1}{20}$ m.$^{-1}$, a velocity $c = 2000$ m.s^{-1}, and a choice of a spatial discretization h such that n_w varies from 1 to 20, considering a maximum frequency of propagation f_{max} reaching 12.5 Hz. Three curves are obtained, depending on how the time discretization step is chosen. The first choice correspond to a fixed Δt, set to satisfy exactly the CFL condition for 10 discretization points per smallest wavelength. The second choice corresponds to set Δt as

$$\Delta t = 0.9\frac{\Delta x}{c}. \tag{2.261}$$

The third choice corresponds to set Δt exactly at the CFL condition for each Δx. Interestingly in this case, the numerical dispersion is only related to the time discretization. However this situation is only possible in 1D. In higher dimension, this special case (vanishing of the numerical dispersion associated with the spatial discretization) is no longer valid.

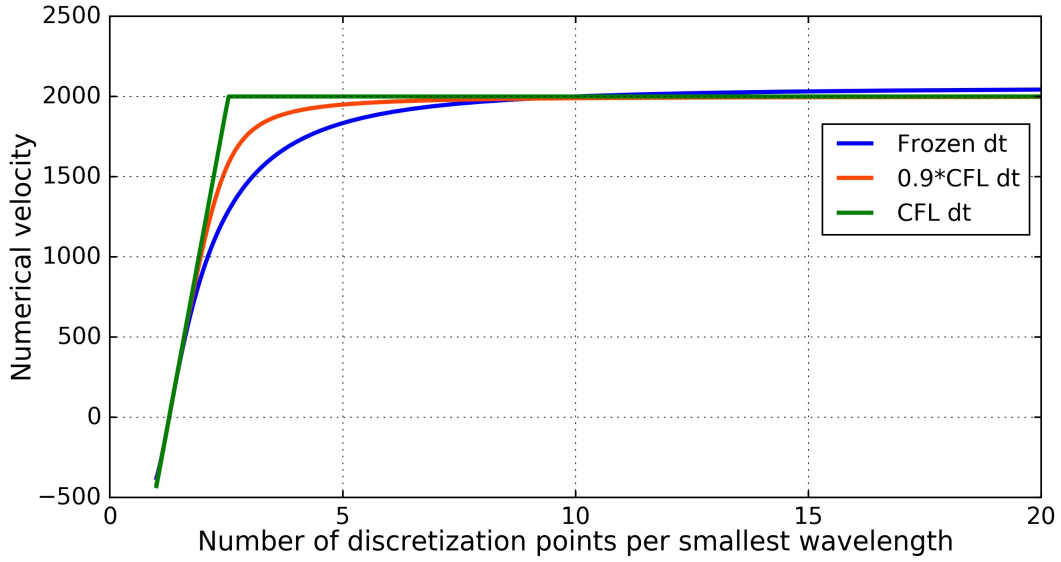

Figure 2.9: Dispersion curve for a choice of $k = 0.05 = \frac{1}{20}$ m.$^{-1}$, a velocity $c = 2000$ m.s^{-1}, and a choice of a spatial discretization varying from 1 to 20 discretization point per smallest wavelength, considering a maximum frequency of propagation reaching 12.5 Hz. Three curves are obtained, depending on how the time discretization step is chosen. The first choice correspond to a fixed Δt, set to satisfy exactly the CFL condition for 10 discretization points per smallest wavelength (blue curve). The second choice corresponds to set Δt as 0.9 time the CFL condition (red curve). The third choice consists in setting Δt exactly at the CFL condition, removing in this particular 1D case the dispersion associated with the spatial discretization (green curve).

2.4.3 Staggered grid finite difference scheme for 2D elastodynamics equation

The finite-difference technique has been popularized in the geophysics and seismology community by the work of Virieux (1986). In this study, a 2D finite-difference scheme is proposed for the 2D P-SV elastodynamics equations in the velocity-stress formulation (1st order hyperbolic system). This 2D finite-difference scheme is actually a transposition to the 2D P-SV equations of the Yee scheme (Yee, 1966) designed for the 2D Maxwell's equation (electromagnetic wave propagation).

The 2D P-SV equations, assuming a VTI anisotropy, and a force source (f_x, f_z), are

$$\begin{aligned}
&\frac{\partial v_x}{\partial t} - \frac{1}{\rho}\frac{\partial \sigma_{xx}}{\partial x} - \frac{1}{\rho}\frac{\partial \sigma_{xz}}{\partial z} = f_x \\
&\frac{\partial v_z}{\partial t} - \frac{1}{\rho}\frac{\partial \sigma_{xz}}{\partial x} - \frac{1}{\rho}\frac{\partial \sigma_{zz}}{\partial z} = f_z \\
&\frac{\partial \sigma_{xx}}{\partial t} - C_{11}\frac{\partial v_x}{\partial x} - C_{13}\frac{\partial v_z}{\partial z} = 0 \\
&\frac{\partial \sigma_{zz}}{\partial t} - C_{13}\frac{\partial v_x}{\partial x} - C_{33}\frac{\partial v_z}{\partial z} = 0 \\
&\frac{\partial \sigma_{xz}}{\partial t} - C_{44}\left(\frac{\partial v_x}{\partial z} + \frac{\partial v_z}{\partial x}\right) = 0.
\end{aligned} \tag{2.262}$$

The original idea of the staggered finite-difference scheme is to discretize each of the field $v_x, v_z, \sigma_{xx}, \sigma_{zz}, \sigma_{xz}$ using a centered (second-order) finite difference scheme on different staggered grids. An explanation is given in Figure 2.10.

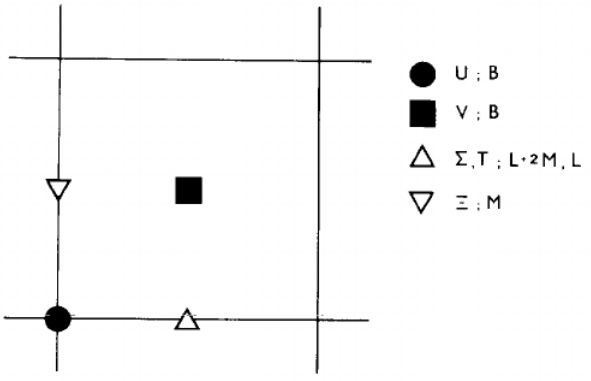

FIG. 1. Discretization of the medium on a staggered grid. Black symbols are for velocities and buoyancy at time $k\Delta t$. White symbols are for stresses and Lamé coefficients at time $(k + 1/2)\Delta t$.

Figure 2.10: Original figure from Virieux (1986)

In our notations

- B corresponds to the buoyancy, which is the inverse of the density $\frac{1}{\rho}$

- U corresponds to the horizontal velocity displacement v_x
- V corresponds to the vertical velocity displacement v_z
- Σ corresponds to the horizontal normal stress σ_{xx}
- T corresponds to the vertical normal stress σ_{zz}
- Θ corresponds to σ_{xz}

We see that in this scheme,

- v_x is centered at the (i,j) point
- v_z is at half grid point $(i+1/2, j+1/2)$ point
- σ_{xx} and σ_{zz} are at $(i+1/2, j)$ point
- σ_{xz} is at $(i, j+1/2)$ point

The corresponding finite-difference scheme is

$$\begin{aligned}
&\frac{(v_x)_{ij}^{n+1/2} - (v_x)_{ij}^{n-1/2}}{\Delta t} && - \left(\frac{1}{\rho}\right)_{ij} \frac{(\sigma_{xx})_{i+1/2,j}^{n} - (\sigma_{xx})_{i-1/2,j}^{n}}{h} \\
& && - \left(\frac{1}{\rho}\right)_{ij} \frac{(\sigma_{xz})_{i,j+1/2}^{n} - (\sigma_{xz})_{i,j-1/2}^{n}}{h} = (f_x)_{ij}^{n} \\
&\frac{(v_z)_{i+1/2j+1/2}^{n+1/2} - (v_z)_{i+1/2j+1/2}^{n-1/2}}{\Delta t} && - \left(\frac{1}{\rho}\right)_{i+1/2j+1/2} \frac{(\sigma_{xz})_{i+1,j+1/2}^{n} - (\sigma_{xz})_{i,j+1/2}^{n}}{h} \\
& && - \left(\frac{1}{\rho}\right)_{i+1/2j+1/2} \frac{(\sigma_{zz})_{i+1/2,j+1}^{n} - (\sigma_{zz})_{i+1/2,j}^{n}}{h} = (f_z)_{i+1/j+1/2}^{n} \\
&\frac{(\sigma_{xx})_{i+1/2j}^{n+1} - (\sigma_{xx})_{i+1/2j}^{n}}{\Delta t} && - (C_{11})_{i+1/2j} \frac{(v_x)_{i+1,j}^{n+1/2} - (v_x)_{ij}^{n+1/2}}{h} \\
& && - (C_{13})_{i+1/2j} \frac{(v_z)_{i+1/2,j+1/2}^{n+1/2} - (v_z)_{i+1/2j-1/2}^{n+1/2}}{h} = 0 \\
&\frac{(\sigma_{zz})_{i+1/2j}^{n+1} - (\sigma_{zz})_{i+1/2j}^{n}}{\Delta t} && - (C_{13})_{i+1/2j} \frac{(v_x)_{i+1,j}^{n+1/2} - (v_x)_{ij}^{n+1/2}}{h} \\
& && - (C_{33})_{i+1/2j} \frac{(v_z)_{i+1/2,j+1/2}^{n+1/2} - (v_z)_{i+1/2j-1/2}^{n+1/2}}{h} = 0 \\
&\frac{(\sigma_{xz})_{ij+1/2}^{n+1} - (\sigma_{xz})_{ij+1/2}^{n}}{\Delta t} && - (C_{44})_{ij+1/2} \frac{(v_x)_{ij+1}^{n+1/2} - (v_x)_{ij}^{n+1/2}}{h} \\
& && - (C_{44})_{ij+1/2} \frac{(v_z)_{i+1/2,j+1/2}^{n+1/2} - (v_z)_{i-1/2j+1/2}^{n+1/2}}{h} = 0
\end{aligned} \tag{2.263}$$

Such a scheme exploits the particular structure of the 2D elastodynamics equations, the temporal derivatives of each field are related to spatial derivatives of other fields (never the same field). This makes possible to use a centered scheme (therefore of order 2) while using a spatial discretization step of h. Full grid scheme would require the use of a spatial grid of size $2h$ to be centered and hence reach second-order accuracy

$$f'(x) =\simeq \frac{f(x+h) - f(x-h)}{2h} + O(h^2) \tag{2.264}$$

The `FORTRAN` code ELAS2DWM available on my webpage, in the directory associated with this lecture, implements this 2D staggered grid scheme. Please feel free to download it and have a look at it. The implementation is rather simple. The numerical experiments presented in the slides have been performed using this simple code. To compile it you only need a gfortran compiler (free). The output files are binary files that can be visualized easily through Octave, Matlab, or Python libraries for instance.

2.4.4 Spectral-element for elastodynamics equations

Even if the finite-difference method is very powerful and easy to implement, it has some limitations. The main limitation is related to the modeling of wave propagation in media with strong interfaces which vary in space, *i.e.* which do not conform to the Cartesian grid. The most important of these interfaces is the subsurface/air interface. In marine environment, this interface. In addition, with receivers and sources located in the water, the acoustic approximation is relatively accurate. With only hydrophones recording the pressure, the imprint of elastic effects on the data is weak. In this case, finite-difference modeling for acoustic equations is a very good option.

In land environment, the subsurface/air interface is not likely to be flat: spatial variations due to varying topography is common. In addition, receivers on land are geophones, which record vertical and horizontal displacement velocities. In these recording, the elastic imprint of the wave propagation is strong. We thus need to rely on elastodynamics equations. TO model accurately P-wave, S-wave, and surface waves, we need to take into account accurately the subsurface/air interface. This is difficult through finite-difference technique, therefore finite-element methods are preferred and shown to be more efficient.

In the sequel, we present the state-of-the-art method for numerical model of wave propagation in seismology (see SPECFEM code `https://geodynamics.org/cig/software/specfem3d/`, SALVUS code `https://cos.ethz.ch/software/production/salvus.html`, or SEM46 code we develop in SEISCOPE). It is a particular instance of finite-element strategy, making it possible to model very accurately 3D elastodynamics propagation in heterogeneous anisotropic media with irregular topography. We introduce this method and review its specificities.

Second-order displacement formulation

The starting point for the spectral element discretization of the elastodynamics equations is to rewrite the first-order hyperbolic system as a second-order in time hyperbolic system on the velocity only. This reduces the number of equations from 9 to 3. This system reads

$$\rho \frac{\partial^2 u}{\partial t^2} - DCD^T u = s \tag{2.265}$$

where the source term s is given by

$$s = F + D\sigma^0. \tag{2.266}$$

We will refer to the domain of interest as $\Omega \subset \mathbb{R}^3$ and the time interval as $[0, T]$. We want to compute an approximation of the displacement function

$$u(x,t), \quad (x,t) \in \Omega \times [0,T] \tag{2.267}$$

solution of (2.265).

We impose free surface boundary condition on the boundary of Ω which we denote as previously as $\partial\Omega$. The free surface boundary condition expresses as

$$\sigma(x,t)n(x) = 0, \quad (x,t) \in \partial\Omega \times [0 \;\; T], \tag{2.268}$$

which means in terms of displacement that

$$\left(CD^T u(x,t)\right) n(x) = 0, \quad (x,t) \in \partial\Omega \times [0 \;\; T]. \tag{2.269}$$

We also use zero initial conditions

$$\forall x \in \Omega, \;\; u(x,0) = 0, \;\; \frac{\partial u}{\partial t}(x,0) = 0. \tag{2.270}$$

Weak form

Here we introduce the L^2 scalar product on Ω as

$$(u,v) = \int_\Omega u(x,t)v(x,t)dx \tag{2.271}$$

Finite-element methods are based on the weak form of the elastodynamics equation. This is, compute for a solution u which satisfies, for any sufficiently smooth function w

$$\left(\rho\frac{\partial^2 u}{\partial t^2}, w\right) - \left(DCD^T u, w\right) = (s,w)\,. \tag{2.272}$$

Basically we require that w and its derivatives are in $\left(L^2(\Omega)\right)^3$. We keep in mind that the scalar product is in space only and not in time. The integral thus depends on time in the expression above but we do not mention it explicitly to simplify the notations.

Integrating by part the second term of the left-hand side in (2.272) yields

$$\left(\rho\frac{\partial^2 u}{\partial t^2}, w\right) - \left(CD^T u(x,t), D^T w\right) + \left(CD^T u(x,t).n, w\right)_{\partial\Omega} = (s,w) \tag{2.273}$$

The free surface condition makes the boundary term naturally vanish, yielding

$$\left(\rho\frac{\partial^2 u}{\partial t^2}, w\right) - \left(CD^T u(x,t), D^T w\right) = (s,w)\,. \tag{2.274}$$

1D elastodynamics equation

For the sake of simplicity, we will consider in the sequel only the 1D case for the propagation of shear waves. This simplifies the previous displacement system as

$$\rho\frac{\partial^2 u}{\partial t^2} - \partial x(\mu(x)\partial x)u = s \tag{2.275}$$

where $\rho(x)$ is the density and $\mu(x)$ is the shear modulus. The free surface condition is expressed as

$$\mu(x)u(x,t) = 0, \;\; (x,t) \in \partial\Omega \times [0 \;\; T]. \tag{2.276}$$

The initial conditions remain the same.

Finite element strategy

Finite element methods rely on the approximation of the weak solution as a decomposition in a basis of N_p given functions $\varphi_i(x)$,

$$u(x,t) = \sum_{i=1}^{N_p} u_i(t)\varphi_i(x). \tag{2.277}$$

This is named as the Galerkin *method*. In addition, we assume that the test functions $w(x)$ are decomposed in the same basis of functions. This assumption is referred to as the Galerkin *principle*. We thus obtain

$$\left(\rho\frac{\partial^2 u}{\partial t^2},\varphi_j\right)-\left(\mu(x)\frac{\partial u}{\partial x},\frac{\partial\varphi_j}{\partial x}\right)=\sum_{i=1}^{N_p}\left(s_i\varphi_i,\varphi_j\right),\quad j=1,\ldots,N_p. \tag{2.278}$$

Injecting (2.277) into (2.278) yields

$$\sum_{i=1}^{N_p}\frac{\partial^2 u_i}{\partial t^2}\left(\rho\varphi_i,\varphi_j\right)-\sum_{i=1}^{N_p}u_i(t)\left(\mu(x)\frac{\partial\varphi_i}{\partial x},\frac{\partial\varphi_j}{\partial x}\right)=\sum_{i=1}^{N_p}\left(s_i\varphi_i,\varphi_j\right),\quad j=1,\ldots,N_p. \tag{2.279}$$

In matrix notation, this can be rewritten as

$$M\frac{\partial^2 u}{\partial^2 t}(t)+Ku(t)=Bs \tag{2.280}$$

where

$$M_{ij}=\left(\rho\varphi_i,\varphi_j\right),\quad K_{ij}=\left(\mu(x)\frac{\partial\varphi_i}{\partial x},\frac{\partial\varphi_j}{\partial x}\right),\quad B_{ij}=\left(\varphi_i,\varphi_j\right) \tag{2.281}$$

For the time derivative, we use a second-order finite-difference scheme, such that we obtain

$$M\frac{u^{n+1}-2u^n+u^{n-1}}{\Delta t^2}-Ku^n=Bs^n \tag{2.282}$$

This gives

$$u^{n+1}=\Delta t^2M^{-1}\left(Ku^n+Bs^n\right)+2u^n-u^{n-1} \tag{2.283}$$

From here we already see that it each time step, the finite-element scheme requires to apply the inverse of the mass matrix M. For large scale problem, this can be an issue in terms of computational cost. Before moving to the specificity of the spectral element method, we localize the previous equation element by element.

Element decomposition

So far the field $u(x,t)$ and the basis functions $\varphi_i(x)$ are defined globally on Ω. We further introduce the discretization over elements by assuming that the function φ_i are attached to elements Ω_e partitioning the domain Ω. We have

$$\Omega=\bigcup_{e=1}^{N_e}\Omega_e,\quad\forall(e,e'),\quad e\neq e'\Longleftrightarrow\Omega_e\cap\Omega_{e'}=\emptyset. \tag{2.284}$$

The total number of elements is denoted by N_e. We therefore approximate $u(x,t)$ as

$$\forall x\in\Omega_e,\quad u(x,t)=\sum_{i=1}^{N_p}u_i^e(t)\varphi_i^e(x) \tag{2.285}$$

and obtain

$$\sum_{i=1}^{N_p}\frac{\partial^2 u_i^e}{\partial t^2}\left(\rho\varphi_i^e,\varphi_j^e\right)-\sum_{i=1}^{N_p}u_i^e(t)\left(\mu(x)\frac{\partial\varphi_i^e}{\partial x},\frac{\partial\varphi_j^e}{\partial x}\right),\quad=\sum_{i=1}^{N_p}\left(s_i\varphi_i^e,\varphi_j^e\right),\quad j=1,\ldots,N_p \tag{2.286}$$

element by element.

Deriving as above, we obtain the same set of equations as before except that they are now local to each element. We write it as

$$M^e \frac{\partial^2 u}{\partial^2 t}(t) + K^e u = B^e s \tag{2.287}$$

We will see after how to connect these equations globally (assembly step).

Choice of Lagrange polynomials

The spectral element method relies on two key elements. The first is the choice of basis functions φ_i as Lagrange polynomials. So far what has been derived is general and common to any finite element strategy.

In 1D, Lagrange polynomials are given by

$$\varphi_i(\xi) = l_i(\xi) = \prod_{j=1, j\neq i}^{N_p+1} \frac{\xi - \xi_i}{\xi_i - \xi_j} \tag{2.288}$$

for a given set of distinct points $\xi_j,\ j = 1, \ldots, N_p + 1$ in the reference cube $[-1, 1]^3$.

The Lagrange polynomial are defined on the domain $[-1; 1]$, named as reference element in the following. To compute the scalar product implying the polynomial l_i and l_j and their derivatives it is therefore necessary to define a mapping from the reference element to the actual element e. This mapping is denoted by F^e

$$\begin{array}{rccc} F^e: & \xi & \mapsto & F^e(\xi) = x \\ & [-1,1] & \to & \Omega_e \end{array} \tag{2.289}$$

The scalar products over Ω_e, can thus be rewritten as

$$(u, v)_{\Omega_e} = \int_{-1}^{1} u\left(F^e(\xi)\right) v\left(F^e(\xi)\right) \frac{dx}{d\xi} d\xi. \tag{2.290}$$

For the mass term, this gives

$$\sum_{i=1}^{N_p} \frac{\partial^2 u_i^e}{\partial t^2} \int_{-1}^{1} \rho\left(F^e(\xi)\right) l_i\left(F^e(\xi)\right) l_j\left(F^e(\xi)\right) \frac{dx}{d\xi} d\xi \tag{2.291}$$

which we rewrite as

$$\sum_{i=1}^{N_p} \frac{\partial^2 u_i^e}{\partial t^2} \int_{-1}^{1} \rho(\xi) l_i(\xi) l_j(\xi) \frac{dx}{d\xi}(\xi) d\xi \tag{2.292}$$

with the slight abuse of notation

$$\rho(F^e(\xi)) = \rho(\xi),\ \ l_i(F^e(\xi)) = l_i(\xi),\ \ l_j(F^e(\xi)) = l_j(\xi). \tag{2.293}$$

For the stiffness term we obtain

$$\sum_{i=1}^{N_p} \int_{-1}^{1} \mu(F^e(\xi)) \frac{\partial l_i(F^e(\xi))}{\partial x} \frac{\partial l_j(F^e(\xi))}{\partial x} \frac{dx}{d\xi} d\xi \tag{2.294}$$

which gives

$$\sum_{i=1}^{N_p} \int_{-1}^{1} \mu(\xi) \frac{\partial l_i(\xi)}{\partial \xi} \frac{\partial l_j(\xi)}{\partial \xi} \left(\frac{d\xi}{dx}\right)^2 \frac{dx}{d\xi} d\xi \tag{2.295}$$

with the same abuse of notations. The term $\left(\frac{d\xi}{dx}\right)^2$ comes from the chain rule for the derivative with respect to ξ.

Finally the source term becomes

$$\int_{-1}^{1} s(t,\xi) l_j(\xi) \frac{dx}{d\xi} d\xi \tag{2.296}$$

with again the same abuse of notations.

We end up with the following equations, element by element

$$\sum_{i=1}^{N_p} \frac{\partial^2 u_i^e}{\partial t^2} \int_{-1}^{1} \rho(\xi) l_i(\xi) l_j(\xi) \frac{dx}{d\xi}(\xi) d\xi - \sum_{i=1}^{N_p} \int_{-1}^{1} \mu(\xi) \frac{\partial l_i(\xi)}{\partial \xi} \frac{\partial l_j(\xi)}{\partial \xi} \left(\frac{d\xi}{dx}\right)^2 \frac{dx}{d\xi} d\xi = \int_{-1}^{1} s(t,\xi) l_j(\xi) \frac{dx}{d\xi} d\xi \tag{2.297}$$

Numerical integration

What remains to be done is now to evaluate numerically the integration over the reference element. In principle, we could use any set of discrete points $\xi_k, k = 1, \ldots, K$ to perform this numerical integration. By numerical integration, we mean, for a function f,

$$\int_{-1}^{1} f(\xi) d\xi \simeq \sum_{k=1}^{K} w_k f(\xi_k) \tag{2.298}$$

If we want to obtain a diagonal mass matrix, we need to obey the following rule: *choose the same integration points as the collocation points defining the Lagrange polynomial.* This will guarantee the following property

$$l_i(\xi_k) = \delta_{ik} \tag{2.299}$$

where δ_{ik} is the Kronecker product

$$\delta_{ik} = \left| \begin{array}{ll} 1 & \text{if} \qquad i = k \\ 0 & \text{else.} \end{array} \right. \tag{2.300}$$

This also defines the weights w_k as (replacing f by a polynomial l_i in the quadrature formula (2.298)

$$w_i = \int_{-1}^{1} l_i(\xi) d\xi. \tag{2.301}$$

We thus have for the mass term

$$\sum_{i=1}^{N_p} \frac{\partial^2 u_i^e}{\partial t^2} \sum_{k=1}^{K} w_k \rho(\xi_k) l_i(\xi_k) l_j(\xi_k) \frac{dx}{d\xi}(\xi_k) \tag{2.302}$$

which gives, using (2.299)

$$\sum_{i=1}^{N_p} \frac{\partial^2 u_i^e}{\partial t^2} \sum_{k=1}^{K} w_k \rho(\xi_k) \delta_{ik} \delta_{jk} \frac{dx}{d\xi}(\xi_k) \tag{2.303}$$

that is

$$\sum_{i=1}^{N_p} \frac{\partial^2 u_i^e}{\partial t^2} w_i \rho(\xi_i) \frac{dx}{d\xi}(\xi_i) \delta_{ji}. \tag{2.304}$$

We finally obtain

$$\sum_{i=1}^{N_p} M_{ij}^e \frac{\partial^2 u_i^e}{\partial t^2} - K_{ij}^e u_i(t) = \sum_{i=1}^{N_p} B_{ij}^e s_i^e, \quad j = 1, \ldots, N_p \tag{2.305}$$

with

$$\begin{cases} M_{ij}^e &= w_i \rho(\xi_i) \dfrac{dx}{d\xi}(\xi_i) \delta_{ij} \\ K_{ij}^e &= \displaystyle\sum_{k=1}^{N_p} w_k \mu(\xi_k) \frac{\partial l_i(\xi_k)}{\partial \xi} \frac{\partial l_j(\xi_k)}{\partial \xi} \left(\frac{d\xi}{dx}(\xi_k) \right)^2 \frac{dx}{d\xi}(\xi_k) \\ B_{ij}^e &= w_i \dfrac{dx}{d\xi}(\xi_i) \delta_{ij} \end{cases} \tag{2.306}$$

Choice of the collocation and integration points

So far we have not discussed how to the choose the points ξ_k. We have just mentioned that to obtain a diagonal mass matrix, we need to approximate the integral over the same set of points used to define the Lagrange polynomial.

A very interesting choice is the selection of the Gauss-Lobatto-Legendre points, defined as the roots of the polynomial

$$(1 - \xi^2) L'(\xi) \tag{2.307}$$

where $L(\xi)$ is a Legendre polynomial. An example of the repartition of such points in 1D is given in Figure 2.11.

This choice is motivated by the three following reasons:

1. The GLL points satisfy Gaussian quadrature rule: the integration of a polynomial of order up to $2N - 1$ with N GLL points is exact.

2. The GLL points are such that the boundary of the elements are used. It means that there is a degree of freedom to represent the solution at the boundary of each element. This property is very appealing, as we have to enforce the continuity of the solution on the global domain. To do so, we only need to consider each degree of freedom at the boundary of an element as unique. Therefore, in the assembly of the mass matrix and stiffness matrix, one only needs to sum up the contributions from different elements for boundary degrees of freedom. This simplifies greatly the assembly of the global matrices M and K.

3. Lagrange polynomials defined over the set of Gauss-Lobatto-Legendre points ensure the spectral convergence property. Schematically, if we denote by u^* the solution of the elastodynamics equations and u_{SEM} its approximation through the spectral-element method, we have

$$\| u^*(., t) - u_{SEM}(., t) \|_{H^1(\Omega)} \leq C N^{-s} \| u^* \|_{H^{s+1}(\Omega)} \tag{2.308}$$

where N is the order of Lagrange polynomial used to discretize the solution over the elements, and $H^s(\Omega)$ is the Sobolev space of functions over Ω (namely with s derivatives which are in $L^2(\Omega)$). The spectral convergence thus means that the convergence speed is driven by the smoothness of the solution. The smoother the solution is, the fastest the convergence is. This is reflected by the factor N^{-s} in the

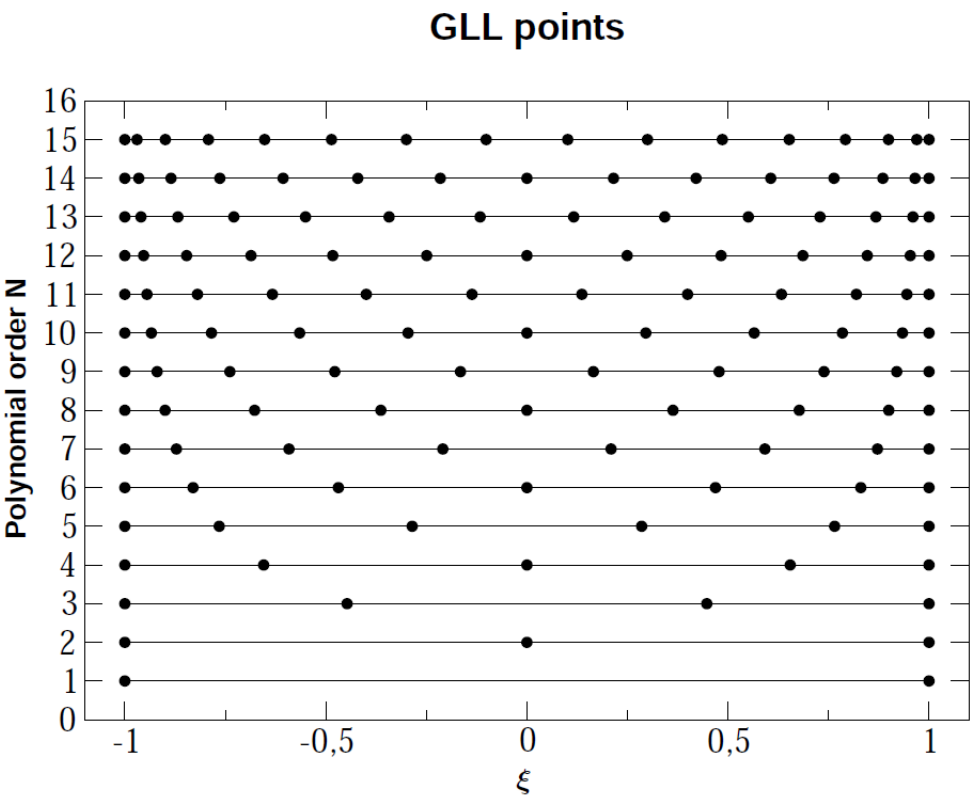

Figure 2.11: Example of Gauss-Lobatto-Legendre points repartitions in 1D for different orders of polynomials. Note that the partition is uneven. Note also that the two extremal points of the 1D segment are always included in the Gauss-Lobatto-Legendre points.

right hand side of the inequality (2.308). This is of particular interest when solving elastodynamics equations in reasonably smooth media, as it is often the case in the frame of full waveform inversion algorithms. Indeed, in this case, the variations of the medium remain relatively smooth with respect to the propagated wavelength, and the wave equation solution is smooth.

Besides, when the solution is analytic, that is infinitely differentiable and locally representable by a convergent power series, the spectral element method exhibits exponential convergence

$$\|u^*(.,t) - u_{SEM}(.,t)\|_{H^1(\Omega)} \leq Ce^{-\alpha N} \tag{2.309}$$

for some constants C, α.

The convergence formula (2.308) and (2.309) are to be compared with standard finite element convergence results in classical polynomial spaces, where we would have

$$\|u^*(.,t) - u_{SEM}(.,t)\| \leq Ch^p \tag{2.310}$$

where h is the size of the element and p the order of the chosen polynomial basis to represent the solution.

More details on this spectral element methods can be found for instance in Patera (1984); Bernardi and Maday (1992)

Chapter 3

Full waveform inversion

3.1 Generalities on inverse problem

In general mathematical notations, an inverse problem can be defined as follows. Let $d_{obs} \in \mathcal{D}$ be the data we have on a given physical system. The physical system is described by some parameters $m \in \mathcal{M}$. There exists a physical process which relies the observation and m in this physical system. This process is described by an operator g such that

$$\begin{array}{rccc} g: & m & \mapsto & g(m) \\ & \mathcal{M} & \to & \mathcal{D}. \end{array} \tag{3.1}$$

The inverse problem is posed as finding m such that

$$g(m) = d_{obs}. \tag{3.2}$$

In geophysical imaging, d_{obs} corresponds to some partial measurement of the propagating waves. The parameter m^* corresponds to the mechanical properties of the subsurface. The operator g corresponds to the wave propagation operator.

We can refer to Hadamard to have the definition of a well posed problem. Such a problem should satisfy

- existence
$$\exists m^*, \;\; g(m^*) = d_{obs} \tag{3.3}$$
- uniqueness
$$m \neq m^* \iff g(m) \neq d_{obs} \tag{3.4}$$
- continuous dependency of the solution with respect to the data
$$\forall \varepsilon, \; \exists \eta, \;\; \|d_{obs,1} - d_{obs,2}\| < \eta \Longrightarrow \|m_1^* - m_2^*\| < \varepsilon, \tag{3.5}$$
where we define m_1^* and m_2^* as
$$g(m_1^*) = d_{obs,1}, \;\; g(m_2^*) = d_{obs,2}. \tag{3.6}$$

In practice, inverse problems are often ill-posed to a certain extent. First, it is often difficult to have strict existence of a solution, simply because of the measurement noise. In practice, at best, we have

$$d_{obs} = g(m^*) + \varepsilon. \tag{3.7}$$

From this, existence and uniqueness need to be recast as an equivalence class property:

$$m \text{ solution } \iff m \in M_\varepsilon^* \subset M,\ M_\varepsilon^* = \{m \in M,\ \|g(m) - d_{obs}\| < \varepsilon\}\,. \tag{3.8}$$

If the continuous dependency of the solution can still be verified, we can qualify the inverse problem as "weakly ill-posed" (informal notion). If however, the continuous dependency of the solution is not verified, the inverse problem can be qualified as "strongly ill-posed" (informal notion too).

3.2 The linear case

3.2.1 Formulation

To have a first insight on how inverse problems behave and how we can solve them, it is interesting to consider the linear case. In this frame we will have

$$d \in \mathcal{D} \subset \mathbb{R}^N,\ \ m \in \mathcal{M} \subset \mathbb{R}^M,\ \ g(m) = Am, \tag{3.9}$$

with

$$A \in \mathcal{M}_{N,M}(\mathbb{R}), \tag{3.10}$$

where $\mathcal{M}_{N,M}(\mathbb{R})$ is the space of real matrices with N rows and M columns.

We have to solve the problem

$$g(m) = d_{obs}, \tag{3.11}$$

that is, in this case

$$Am = d_{obs}. \tag{3.12}$$

The solution of the inverse problems can thus be defined by

$$m^* = A^\dagger d_{obs}. \tag{3.13}$$

where $A^\dagger$ denotes a generalized inverse for the (rectangular) matrix A. Here we focus on inverse problems which are over-determined: we have more information than discrete unknowns, that is $M < N$. This is the situation that we have in practice for full waveform inversion. If we assume that A is full rank, that is

$$rank(A) = M \tag{3.14}$$

then we have

$$A^\dagger = (A^T A)^{-1} A^T \tag{3.15}$$

Consequently, the solution m^* is given by

$$m^* = (A^T A)^{-1} A^T d_{obs}. \tag{3.16}$$

3.2.2 Link with the least-squares solution

It is interesting to note that in this case m^* corresponds to the least-squares solution of the inverse problem. Consider the functional

$$\begin{array}{rccc} f: & m & \mapsto & \|g(m) - d_{obs}\|^2 \\ & \mathcal{M} \subset \mathbb{R}^M & \to & \mathbb{R}^+. \end{array} \tag{3.17}$$

We have

$$f(m) = \|Am - d_{obs}\|^2. \tag{3.18}$$

The least-squares problem is defined as

$$\min_m f(m). \tag{3.19}$$

A solution of this minimization problem is found by searching for the zero of the gradient of $f(m)$. To compute this gradient, we consider an increment $dm \in \mathcal{M}$. We have

$$\begin{aligned} f(m+dm) = \|A(m+dm) - d_{obs}\|^2 &= \|Am - d_{obs} + Adm\|^2 \\ &= f(m) + 2(Am - d_{obs}, Adm)_N + \|dm\|^2 \end{aligned} \tag{3.20}$$

where $(.,.)_N$ denotes the conventional Euclidean scalar product in $\mathbb{R}^N$. Therefore

$$\begin{aligned} f(m+dm) - f(m) &= 2(Am - d_{obs}, Adm)_N + \|Adm\|^2 \\ &= 2(A^T(Am - d_{obs}), dm)_M + \|Adm\|^2 \end{aligned} \tag{3.21}$$

Note here that transposing the matrix A to apply it to the left hand side of the scalar product, we move from a scalar product in $\mathbb{R}^N$ (data space) to a scalar product in $\mathbb{R}^M$ (model space).

Consider a first-order Taylor development of $f(m)$, we should have

$$\forall\, dm \in \mathcal{M},\quad f(m+dm) = f(m) + (\nabla f(m), dm)_M + O_{dm\to 0}(\|dm\|^2) \tag{3.22}$$

By identification of (3.21) and (3.22), we have

$$\nabla f(m) = 2A^T(Am - d_{obs}). \tag{3.23}$$

Therefore, looking for m such that $\nabla f(m) = 0$ is equivalent to solve

$$2A^T(Am - d_{obs}) = 0, \tag{3.24}$$

which gives

$$m = \left(A^T A\right)^{-1} A^T d_{obs} \tag{3.25}$$

We recover the formula (3.16) which we got from the definition of the generalized inverse matrix.

3.2.3 Stability to noise

Consider that

$$d_{obs} = Am^* + \varepsilon \tag{3.26}$$

where $\varepsilon \in \mathcal{D}$ represents measurement noise. In this case, the least-squares solution is given by

$$m_{LS} = \left(A^T A\right)^{-1} A^T(Am^* + \varepsilon) = m^* + \left(A^T A\right)^{-1} A^T \varepsilon \tag{3.27}$$

Let m_ε be the projection of the noise in the model space through the operator A^T

$$m_\varepsilon = A^T \varepsilon \tag{3.28}$$

Consider a decomposition of m_ε in the eigenvector basis of A^TA

$$m_\varepsilon = \sum_{i=1}^{M} \beta_i v_i \tag{3.29}$$

where $v_i, \quad i = 1, \ldots, M$ is a basis of eigenvectors of A^TA, associated with eigenvalues $\alpha_i, \ \ i = 1, \ldots, M$.

Then, we have

$$m_{LS} = m^* + \left(A^TA\right)^{-1} m_\varepsilon = m^* + \sum_{i=1}^{M} \beta_i \left(A^TA\right)^{-1} v_i = m^* + \sum_{i=1}^{M} \frac{\beta_i}{\alpha_i} v_i \tag{3.30}$$

If A^TA has small eigenvalues (tending to 0) then potentially the contribution from the noise can be significant in the reconstruction and the reconstruction error $\left(A^TA\right)^{-1} A^T\varepsilon$ can be big.

We see that the conditioning of the normal matrix A^TA is an indication over the sensitivity to noise of the linear inverse problem in this case. Note also the following for the remainder of the course: A^TA is the Hessian matrix associated with the functional $f(m)$, which has an important role in the analysis of the seismic inverse problem related to full waveform inversion.

3.2.4 Regularization

In practice, one way to reduce the sensitivity of the inverse problem to noise is to regularize it. A very standard example of regularization consists in using a prior model m_p. We assume we know such a prior model, from prior information on the solution. If such a model is known, we can add a penalization term in the misfit function $\frac{1}{2}\|m - m_p\|^2$, weighted by a scalar parameter $\eta \in \mathbb{R}^+$. We obtain a regularized inverse problem

$$\min_m f_\eta(m) = \frac{1}{2}\|Am - d_{obs}\|^2 + \frac{\eta}{2}\|m - m_p\|^2. \tag{3.31}$$

This regularization belongs to the class of Tikhonov regularization (Tikhonov et al., 2013).

The interpretation of this regularization term is that we are looking for a model m which should at the same time satisfy the linear system $Am = d_{obs}$ in the least-squares sense while not being too far from the prior model m_p (in the least squares sense too). It can be understood as a way to guide the inversion to look for models satisfying $Am = d_{obs}$ in the vicinity of m_p. The "size" of the vicinity depends on the parameter η.

A solution of the problem (3.31) can be found by setting the gradient of the regularized function $f_\eta(m)$ to 0 (as previously). We have

$$\nabla f_\eta(m) = A^T(Am - d_{obs}) + \eta(m - m_p). \tag{3.32}$$

Setting this gradient to 0 gives the solution $m_{LS,\eta}$

$$m_{LS,\eta} = \left(A^TA + \eta I_M\right)^{-1} \left(A^Td_{obs} + \eta m_p\right). \tag{3.33}$$

where I_M is the identity matrix of size M.

Let us denote $H_\eta = A^TA + \eta I_M$ [1]. If we assume that d_{obs} is as previously such that

$$d_{obs} = Am^* + \epsilon \tag{3.34}$$

[1] the use of the notation H will become obvious later, H is actually the Hessian matrix, the matrix of the second-order derivatives of the misfit function f_η

we have

$$m_{LS,\eta} = H_\eta^{-1} A^T A m^* + H_\eta^{-1} m_\epsilon + \eta H_\eta^{-1} m_p \tag{3.35}$$

where $m_\epsilon = A^T \epsilon$ as previously.

We decompose m^*, m_ϵ and m_p in the basis of the eigenvectors of $A^T A$, such that

$$m^* = \sum_{i=1}^{M} \beta_i^* v_i, \quad m_\epsilon = \sum_{i=1}^{M} \beta_i^\epsilon v_i, \quad m_p = \sum_{i=1}^{M} \beta_i^p v_i. \tag{3.36}$$

The eigenvalues of H_η are $\alpha_i + \eta$ with eigenvectors v_i. Indeed, since v_i are the eigenvectors of $A^T A$ with eigenvalues α_i we have

$$\forall i = 1, \ldots, M, \quad A^T A v_i = \alpha_i v_i. \tag{3.37}$$

Hence,

$$\forall i = 1, \ldots, M, \quad H_\eta v_i = (A^T A + \eta I_M) v_i = A^T A v_i + \eta v_i = (\alpha_i + \eta) v_i. \tag{3.38}$$

Thus $v_i, \quad i = 1, \ldots, M$ are the eigenvectors of H_η with eigenvalues $\alpha_i + \eta$.

Another way to see this is that we have the decomposition

$$A^T A = P^T D P \tag{3.39}$$

with $D = \operatorname{diag}(\alpha_1, \ldots, \alpha_M)$ and P an orthonormal matrix

$$P^T = P^{-1}. \tag{3.40}$$

Thus

$$H_\eta = A^T A + \eta I_M = P^T D P + \eta I_M = P^T D P + \eta P^T P = P^T (D + \eta I_M) P. \tag{3.41}$$

We analyze the three terms in the solution $m_{LS,\eta}$. We have

$$H_\eta^{-1} A^T A = \left(P^T (D + \eta I_M) P\right)^{-1} P^T D P = P^T (D + \eta I_M)^{-1} D P \tag{3.42}$$

Thus $H_\eta^{-1} A^T A$ as the same eigenvectors v_i as $A^T A$, but with eigenvalues $\frac{\alpha_i}{\alpha_i + \eta}$. Thus

$$H_\eta^{-1} A^T A m^* = \sum_{i=1}^{M} \beta_i^* \frac{\alpha_i}{\alpha_i + \eta} v_i \tag{3.43}$$

Besides

$$H_\eta^{-1} m_\epsilon = \sum_{i=1}^{M} \beta_i^\epsilon \frac{1}{\alpha_i + \eta} v_i \tag{3.44}$$

and finally

$$\eta H_\eta^{-1} m_p = \sum_{i=1}^{M} \beta_i^p \frac{\eta}{\alpha_i + \eta} v_i. \tag{3.45}$$

Hence,

$$m_{LS,\eta} = \sum_{i=1}^{M} \beta_i^* \frac{\alpha_i}{\alpha_i + \eta} v_i + \sum_{i=1}^{M} \beta_i^\epsilon \frac{1}{\alpha_i + \eta} v_i + \sum_{i=1}^{M} \beta_i^p \frac{\eta}{\alpha_i + \eta} v_i. \tag{3.46}$$

We can analyze what happens in terms of the limit of η. If η tends to 0, it means we remove the regularization. The third term cancels as we have

$$\lim_{\eta \to 0} \frac{\eta}{\alpha_i + \eta} = 0. \tag{3.47}$$

The first term converges to m^* as we have

$$\lim_{\eta \to 0} \frac{\alpha_i}{\alpha_i + \eta} = 1 \tag{3.48}$$

The second term converges to $A^T A^{-1} m_\epsilon$ as we have

$$\lim_{\eta \to 0} \frac{1}{\alpha_i + \eta} = \frac{1}{\alpha_i} \tag{3.49}$$

We are thus go back to the previous situation, without regularization, with the risk that small α_i values yield a strong imprint of the noise in the reconstruction of m.

Contrary to this situation, if η tends to $+\infty$, the first term converges to 0

$$\lim_{\eta \to +\infty} \frac{\alpha_i}{\alpha_i + \eta} = 0 \tag{3.50}$$

The second term also converges to 0, as we have

$$\lim_{\eta \to +\infty} \frac{1}{\alpha_i + \eta} = 0 \tag{3.51}$$

and the third term converges to m_p as we have

$$\lim_{\eta \to +\infty} \frac{\eta}{\alpha_i + \eta} = 1 \tag{3.52}$$

We thus have

$$\lim_{\eta \to +\infty} m_{LS,\eta} = m_p \tag{3.53}$$

This can be easily interpreted: when the regularization weight η dominates, then the data misfit term is negligible, and the optimal solution consists in producing a model as close as possible from m_p, with, in the limit case, m_p itself.

In practice we choose regularization weights "in between" these two regimes. What is interesting is to note that, to mitigate the noise, η makes the reconstruction of m^* imperfect, as it has an effect on the first term. The smallest eigenvalues α_i are indeed thresholded by η making the reconstruction of these components driven by the regularization rather than the data.

3.3 FWI: a PDE constrained least-squares optimization problem

3.3.1 Formulation

We introduce here the notations which will be used throughout this chapter. The model parameters that we will aim at recovering from the seismic data will be denoted by m. These parameters will be represented by functions of space $m(x)$ and will live in the model space $\mathcal{M}$. To link with the modeling chapter, these parameters can be the P-wave velocity

$V_P(x)$, the density $\rho(x)$, or any combination of the components $C_{IJ}(x)$ of the stiffness tensor. We will make a distinction between mono-parameter inversion, in which case $m(x)$ represents a single parameter class, and multi-parameter inversion, in which case $m(x)$ gathers several parameter classes, for instance $m(x) = [V_P(x) \;\; \rho(x)]$ for a multi-parameter inversion of P-wave velocity $V_P(x)$ and density $\rho(x)$. In any case, the model parameters $m(x)$ always depend only on space. We assume that during the seismic recording experiment, the properties of the subsurface do not change.

The seismic data associated with a source/receiver couple s and r will be denoted by $d_{obs,s}(x_r, t)$, for $t \in [0, T]$ where T is the total recording time. The corresponding calculated data will be denoted by $d_{cal,s}(x_r, t)$.

The wavefield generated by the source s will be denoted by $u_s(x, t)$. Here $x \in \Omega \subset \mathbb{R}^d$ where d is the dimension of the problem. In practice $d = 3$, and sometimes, resulting from approximations, $d = 2$ is chosen to reduce the computational cost.

The relation between the calculated data $d_{cal,s}(x_r, t)$ and the wavefield $u_s(x, t)$ will be denoted as

$$d_{cal,s} = Ru_s, \tag{3.54}$$

where R is a restriction operator extracting the values of the wavefield at the receiver location x_r. This can be denoted by the convolution with a Dirac delta function

$$(Ru_s)(x_r, t) = \int_\Omega \delta(x - x_r)u(x, t)dx. \tag{3.55}$$

For the sake of generality, the wavefield $u_s(x, t)$ will be considered as the solution of a general partial-differential equation representing the wave propagation within the subsurface, from the acoustic to the elastic approximation. We will denote this wave propagation operator by $A(m)$, such that we have

$$A(m)u_s = \varphi_s, \tag{3.56}$$

where $\varphi_s(x, t)$ is the source term corresponding to the source number s.

To distinguish between the dependency of $u_s(x, t)$ with respect to space and time (as a function of space and time) and its dependency with respect to the model parameters m through the solution of the PDE 3.56, we will sometimes denote this dependency through brackets $[m]$ such that the wavefield in the model m solution of 3.56 is denoted by

$$u_s[m](x, t). \tag{3.57}$$

The same notation can applied to the calculated data, and we denote it as

$$d_{cal,s}[m](x_r, t). \tag{3.58}$$

In this frame, the full waveform inversion problem corresponds to the following PDE-constrained optimization problem

$$\begin{aligned} &\min_m \frac{1}{2}\sum_{s=1}^{N_s} \|d_{cal,s} - d_{obs,s}\|^2, \\ \text{subject to} \quad &\forall s = 1, \dots, N_s, \;\; d_{cal,s} = Ru_s, \\ &\forall s = 1, \dots, N_s, \;\; A(m)u_s = \varphi_s, \end{aligned} \tag{3.59}$$

where N_s is the total number of seismic sources (shots), and the least-squares norm $\|.\|$ corresponds to the L^2 distance

$$\|d_1 - d_2\| = \sum_{r=1}^{N_r} \int_0^T |d_1(x_r, t) - d_2(x_r, t)|^2 dt, \tag{3.60}$$

with N_r the total number of receivers. For practical applications, this number can depend on the source s. Here we assume that the same number of receivers is attached to each source for the sake of simplicity. This has no impact on the generality of the results presented in the sequel.

3.3.2 Lagrangian operator and KKT system

The numerical solution of such constrained optimization problem requires to study the Lagrangian operator associated with such problem. Here and in the sequel, we will make the assumption that $N_s = 1$, only for the sake of concision and clarity. All that will be developed afterwards can be extended to the case of multiple sources by summation, thanks to the linearity of the different operators that we will consider with respect to the sources. For this reason, from now on, we drop the index s in the mathematical derivations.

The Lagrangian operator associated with problem (3.59) is

$$L(m, u, d_{cal}, \lambda, \mu) = \frac{1}{2}\|d_{cal} - d_{obs}\|^2 + (A(m)u - \varphi, \lambda)_{\mathcal{W}} + (d_{cal} - Ru, \mu)_{\mathcal{D}}\,. \tag{3.61}$$

In equation (3.131), we have introduced two different scalar products. The first is defined in the wavefield space, denoted by $(.,.)_{\mathcal{W}}$, corresponding to

$$(u_1, u_2)_{\mathcal{W}} = \int_0^T \int_\Omega u_1(x, t) u_2(x, t) dx dt. \tag{3.62}$$

The second is defined in the data space, denoted by $(.,.)_{\mathcal{D}}$, corresponding to

$$(d_1, d_2)_{\mathcal{D}} = \sum_{r=1}^{N_r} \int_0^T d_1(x_r, t) d_2(x_r, t) dt. \tag{3.63}$$

The solution of problem (3.59) are characterized by the zero of the first-order derivatives of the Lagrangian operator (hence the importance of this operator). More precisely, the first-order necessary conditions for a candidate $(m, u, d_{cal}, \lambda, \mu)$ to be a solution of (3.59) correspond to canceling the first-order derivatives of the Lagrangian. The corresponding system is known as the Karush-Kunh-Tucker (KKT) system. In our case, it writes as

$$\begin{cases} \dfrac{\partial L(m, u, d_{cal}, \lambda, \mu)}{\partial m} &= 0 \\ \dfrac{\partial L(m, u, d_{cal}, \lambda, \mu)}{\partial u} &= 0 \\ \dfrac{\partial L(m, u, d_{cal}, \lambda, \mu)}{\partial d_{cal}} &= 0 \\ \dfrac{\partial L(m, u, d_{cal}, \lambda, \mu)}{\partial \lambda} &= 0 \\ \dfrac{\partial L(m, u, d_{cal}, \lambda, \mu)}{\partial \mu} &= 0. \end{cases} \tag{3.64}$$

In the specific case of the Lagrangian operator we have introduced this amounts to

$$\begin{cases} \left(\dfrac{\partial A}{\partial m}u, \lambda\right)_T & = & 0 \\ \dfrac{\partial L(m,u,d_{cal},\lambda,\mu)}{\partial u} & = & 0 \\ \dfrac{\partial L(m,u,d_{cal},\lambda,\mu)}{\partial d_{cal}} & = & 0 \\ A(m)u - \varphi & = & 0 \\ d_{cal} - Ru & = & 0, \end{cases} \tag{3.65}$$

where the product $(.,.)_T$ is defined by, for $(u,v) \in \mathcal{W} \times \mathcal{W}$,

$$(u,v)_T = \int_0^T u(x,t)v(x,t)dt \tag{3.66}$$

We do not get into the details of the three first equations at this stage (this will be developed later). We see that at convergence, the optimal point in the full space $(m,u,d_{cal},\lambda,\mu)$ should at least satisfy the constraints (which is expected).

Conventional approaches working in this full space are known as primal-dual approach, and amounts to the solution of the min-max problem

$$\min_{m,u,d_{cal}} \max_{\lambda,\mu} L(m,u,d_{cal},\lambda,\mu). \tag{3.67}$$

Solving the full waveform inversion problem with this kind of strategy is never done in practice, because of the inherent size of the full space. The number of degrees of freedom to be used under the frame of such a full space approach makes it very difficult to solve unless very small 1D problem are considered.

Instead of such full space approach, reduced space approach are preferred conventionally, which we describe in the following.

3.3.3 Reduced space approach

The rationale behind the reduce space approach consists in eliminating the variables d_{cal} and u from the problem by solving the corresponding constraint equations. This leads to the reformulation

$$\min_m \frac{1}{2}\|RA(m)^{-1}\varphi - d_{obs}\|^2. \tag{3.68}$$

By doing so, the solution is searched only in the model space $\mathcal{M}$, which reduces significantly the size of the problem. The problem is also formulated as an unconstrained optimization problem.

However, it is important to keep in mind that the original full waveform inversion problem can be considered as a PDE-constrained optimization problem. In particular, it provides insight on how to compute the gradient of the misfit function appearing in the reduced space problem (3.68).

To solve numerically the problem (3.68), we employ optimization algorithms dedicated to the solution of large-scale unconstrained problems. We present on overview of these strategies in the following.

3.4 Numerical optimization

3.4.1 General frame

In this section we consider the following unconstrained optimization problem

$$\min_m f(m), \tag{3.69}$$

where

$$\begin{array}{rcl} m & \mapsto & f(m) \\ \mathcal{M} & \to & \mathbb{R}^+. \end{array} \tag{3.70}$$

After discretization we have $\mathcal{M} \subset \mathbb{R}^M$. Key quantities will be the gradient of the misfit function

$$\nabla f(m) = \begin{bmatrix} \frac{\partial f}{\partial m_1} \\ \\ \frac{\partial f}{\partial m_2} \\ \\ \vdots \\ \\ \frac{\partial f}{\partial m_M} \end{bmatrix} \in \mathbb{R}^M, \tag{3.71}$$

as well as its Hessian $H(m)$, the matrix of second-order derivatives of the function $f(m)$, such that

$$H_{ij}(m) = \frac{\partial^2 f}{\partial m_i \partial m_j}, \tag{3.72}$$

that is

$$H(m) = \begin{pmatrix} \frac{\partial^2 f}{\partial m_1^2} & \frac{\partial^2 f}{\partial m_2 \partial m_1} & \cdots & \frac{\partial^2 f}{\partial m_M \partial m_1} \\ \vdots & \vdots & & \vdots \\ \frac{\partial^2 f}{\partial m_1 \partial m_M} & \frac{\partial^2 f}{\partial m_2 \partial m_M} & \cdots & \frac{\partial^2 f}{\partial m_{M^2}} \end{pmatrix}. \tag{3.73}$$

Note: the Hessian matrix $H(m)$ is symmetric by construction, following Schwarz's theorem (the second-order partial derivatives can be switched).

In this Section we consider linesearch-based numerical optimization methods, which obey the following iteration. From an initial guess m_0, build the sequence

$$m_{k+1} = m_k + \alpha_k \Delta m_k, \tag{3.74}$$

where m_{k+1} is the new iterate towards the solution of the problem (3.69), $\alpha_k \in \mathbb{R}_*^+$ is a scalar parameter representing the length of the step taken in the direction Δm_k, referred to as the steplength in the following, and Δm_k is a descent direction. This means that this vector satisfies

$$(\Delta m_k, -\nabla f(m_k)) = -\Delta m_k^T \nabla f(m_k) > 0, \tag{3.75}$$

where $(.,.)$ is the Euclidean scalar product on $\mathbb{R}^M$. Roughly speaking, Δm_k is in a direction not too far from the opposite of the gradient, which is the local steepest descent direction.

In the sequel we first present how to compute the steplength α_k (linesearch strategy), then we discuss how the descent direction Δm_k is computed. For a wide overview of numerical optimization, the reader is referred to the excellent book from Nocedal and Wright (2006).

3.4.2 Linesearch strategy

Wolfe conditions

The exact linesearch problem, for a given descent direction Δm_k, would consist in solving the problem

$$\min_{\alpha} \phi(\alpha) = f(m_k + \alpha \Delta m_k). \tag{3.76}$$

However, in practice, this is never implemented because this would require to solve an optimization problem on its own at each iteration of the minimization algorithm. This optimization problem would require to compute several time the gradient of the misfit function $f(m)$, which we try to avoid as (as will be seen later) this represents the main computationally expensive task.

Instead of the exact linesearch strategies, standard linesearch algorithms aims at computing a steplength α satisfying the Wolfe's condition. These two Wolfe's conditions are expressed as

1. $f(m_k + \alpha \Delta m_k) \leq f(m_k) + c_1 \alpha \nabla f(m_k)^T \Delta m_k,$

2. $\nabla f(m_k + \alpha \Delta m_k)^T \Delta m_k \geq c_2 \nabla f(m_k)^T \Delta m_k,$

where $c_1 < c_2$ are user defined constant (usual choices are $c_1 = 10^{-4}$, $c_2 = 0.9$).

The condition 1 in (3.4.2) is referred to as the sufficient decrease condition. As Δm_k is a descent direction, we know that

$$\exists \varepsilon > 0, \quad \nabla f(m_k)^T \Delta m_k \leq -\varepsilon < 0. \tag{3.77}$$

Therefore, condition 1 implies that

$$f(m_k + \alpha \Delta m_k) < f(m_k), \tag{3.78}$$

and

$$f(m_k) - f(m_k + \alpha \Delta m_k) \geq c_1 \alpha \varepsilon > 0. \tag{3.79}$$

The condition 1 in (3.4.2) thus guarantees that the step taken in the direction Δm_k produces a decrease at least equal to $c_1 \alpha \varepsilon$.

The condition 2 in (3.4.2) is referred to as the sufficient curvature condition. We have

$$\phi'(\alpha) = \nabla f(m_k + \alpha \Delta m_k)^T \Delta m_k. \tag{3.80}$$

Therefore condition 2 can be interpreted as

$$\phi'(\alpha) \geq c_2 \phi'(0). \tag{3.81}$$

Condition 2 therefore implies that the slope of the function $\phi(\alpha)$ should be greater at the selected step α than at the initial point $\alpha = 0$. Indeed, if the slope is smaller, this indicates that increasing the step α is likely to produce a smaller misfit function, therefore it should be interesting to increase the step. The condition 2 serves as a safeguard towards taking too small steps, which could lead to convergence issues.

An illustration of the 2 Wolfe conditions is presented in Figure 3.1, taken from Nocedal and Wright (2006). In the notations of the document, p_k should be replaced by Δm_k (descent direction) and x_k by m_k.

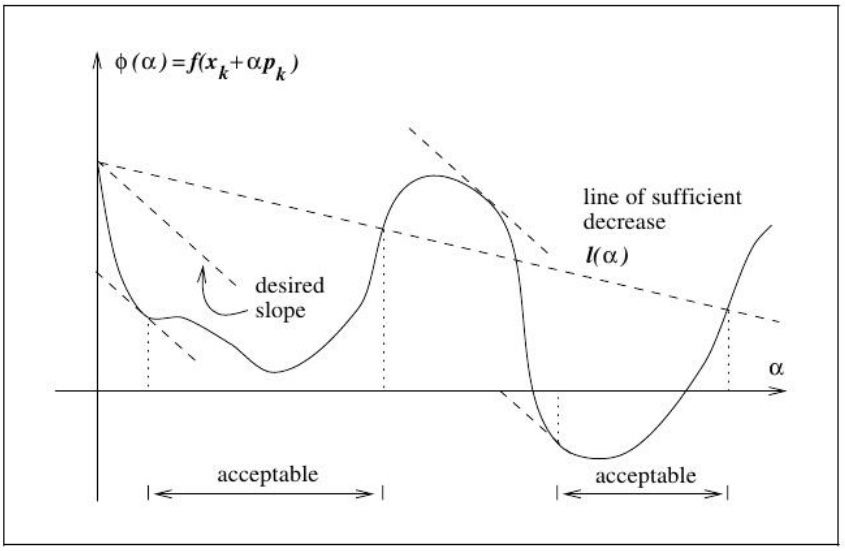

Figure 3.1: Illustration of the 2 Wolfe conditions taken from Nocedal and Wright (2006). In the notations of the document, p_k should be replaced by Δm_k (descent direction) and x_k by m_k.

The Wolfe conditions are extremely important because it can be proven that, if the descent direction Δm_k are chosen such that

$$\cos\theta_k = \frac{-\Delta m_k \nabla f(m_k)^T}{\|\Delta m_k\| \|\nabla f(m_k)\|} \geq \delta > 0 \tag{3.82}$$

(which means the angle between the descent direction and the opposite of the gradient is bounded from $\Pi/2$ *i.e.* the descent direction is never orthogonal to the steepest descent direction), then

$$\lim_{k\to\infty} \|\nabla f(m_k)\| = 0 \tag{3.83}$$

when we follow the iteration (3.74). In other words, the iterative scheme (3.74) is globally convergent as soon as the descent direction Δm_k satisfies condition (3.82) and the steplength α_k satisfies the Wolfe's conditions.

By globally convergent, it means that starting from any initial guess m_0, the iteration (3.74) will converge towards the closest local minimum of the misfit function satisfying $\|\nabla f(m_k)\| = 0$.

This property is extremely important as it guarantees the robustness of the optimization methods that we will use. Note that there exists also other conditions than the Wolfe's conditions which can provide this guarantee (Goldstein's condition for instance). More details on these aspects are to be found in chapter 3 of Nocedal and Wright (2006).

Linesearch algorithm

We have now some conditions that α should fulfill to yield a convergent minimization algorithm. In practice, how to choose/compute α? This can be done in various ways. One

efficient way is the so-called bracketing strategy. The leading idea is to increase the step length as soon as the curvature condition is not satisfied, and to decrease it if the sufficient decrease condition is not satisfied. The full algorithm is given in the sequel.

Linesearch algorithm
$\alpha_0 > 0$;
$\alpha = \alpha_0$;
$\alpha_L = 0$;
$\alpha_R = 0$;
while *Wolfe's conditions not satisfied* **do**
  if *1st Wolfe's condition not satisfied* **then**
    $\alpha_R = \alpha$;
    $\alpha = 0.5(\alpha_L + \alpha_R)$;
  end
  else if *2nd Wolfe's condition not satisfied* **then**
    $\alpha_L = \alpha$;
    if $\alpha_R == 0$ **then**
      $\alpha = 10\alpha$;
    end
    else
      $\alpha = 0.5(\alpha_L + \alpha_R)$;
    end
  end
end
Algorithm 1: Linesearch algorithm using a bracketing + dichotomy strategy.

3.4.3 Computing the descent direction

Steepest descent

The simplest way to compute the descent direction is to take it as the opposite of the gradient.

$$\Delta m_k = -\nabla f(m_k). \tag{3.84}$$

This is known as the steepest descent direction. This is an easy way to satisfy the criterion (3.82) as in this case we have systematically

$$\cos\theta_k = 1. \tag{3.85}$$

Using a linesearch algorithm which satisfies the Wolfe's condition, we thus obtain a convergent method.

However, one well known drawback with the steepest descent is a potential very slow convergence rate. This issue can be illustrated on a simple example. Consider a quadratic misfit function

$$f(x_1, x_2) = \frac{1}{2}\left(x_1^2 + \eta x_2^2\right), \tag{3.86}$$

where $(x_1, x_2) \in \mathbb{R}^2$ and $\eta \in \mathbb{R}^+$. The minimum of $f(x_1, x_2)$ is reached at point $(0, 0)$. We have

$$\nabla f(x_1, x_2) = (x_1, \eta x_2). \tag{3.87}$$

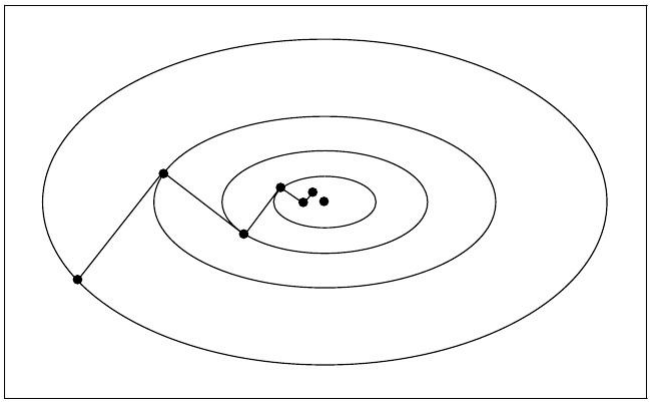

Figure 3.2: Typical oscillations in convergence when using the steepest descent algorithm for the minimization of a quadratic function of two parameters. The horizontal axis corresponds to x_2 and the vertical to x_1. The level set of the function $f(x_1, x_2)$ are represented.

Imagine we start from $(1, 1)$. We have

$$\nabla f(1, 1) = (1, \eta). \tag{3.88}$$

If η is small, we will progress only very slowly in the direction of x_2, and this will produce oscillations in the direction of x_1, crossing each time the optimal value in x_1. A graphical view of this this situation is illustrated in Figure 3.2. In this figure, x_2 corresponds to the horizontal axis and x_1 to the vertical axis. The level set of the function $f(x_1, x_2)$ are represented. We observe the typical oscillations in the convergence along the x_1 axis. A geometrical interpretation is that the gradient is always orthogonal to the level sets. When taking η smaller than 1, the level sets correspond to ellipsoids. Following the steepest descent in this case makes difficult to converge rapidly in the x_2 direction.

Newton method

Steepest descent is often referred to as a first-order method because it relies only on the gradient direction. A significant improvement of this optimization method is obtained by considering the Newton method. Initially, the Newton method is a root finding algorithm. It aims at finding the zero of nonlinear functions.

Consider a nonlinear function of a single variable $g(x)$. At each iteration, we approximate $g(x)$ with its tangent (first-order Taylor development)

$$g(x) = g(x_k) + (x - x_k)g'(x_k). \tag{3.89}$$

A new iterate is found by finding the 0 of this tangential approximation

$$g(x_k) + (x_{k+1} - x_k)g'(x_k) = 0 \implies x_{k+1} = x_k - \frac{-g(x_k)}{g'(x_k)}. \tag{3.90}$$

A graphical illustration is provided in Figure 3.3. The convergence of the Newton algorithm is guaranteed only if starting in a vicinity of the solution.

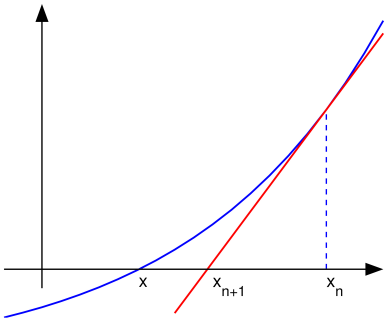

Figure 3.3: Illustration of one Newton iteration for finding the root of a real value function $f(x)$.

An extension of the Newton method to the solution of nonlinear systems under the form

$$G(x) = 0, \tag{3.91}$$

where

$$\begin{array}{rccc} G: & x & \mapsto & G(x) \\ & \mathbb{R}^M & \to & \mathbb{R}^M, \end{array} \tag{3.92}$$

is

$$x_{k+1} = x_k - J(x_k)^{-1} G(x_k), \tag{3.93}$$

where $J(x)$ is the Jacobian operator associated with the operator G such that

$$J(x) = \begin{pmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} & \cdots & \frac{\partial g_1}{\partial x_M} \\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} & \cdots & \frac{\partial g_2}{\partial x_M} \\ \vdots & \vdots & & \vdots \\ \frac{\partial g_M}{\partial x_1} & \frac{\partial g_M}{\partial x_2} & \cdots & \frac{\partial g_M}{\partial x_M} \end{pmatrix}. \tag{3.94}$$

The division by the derivative is replaced with the multiplication by the inverse of the Jacobian matrix $J(x)$ associated with $G(x)$.

The Newton method can be applied in the field of optimization by considering we are looking for stationary point of the gradient. In other words, we look for m such that

$$\nabla f(m) = 0. \tag{3.95}$$

This is a system of nonlinear equations. If we apply to it the Newton iteration, we obtain

$$m_{k+1} = m_k - H(m_k)^{-1} \nabla f(m_k). \tag{3.96}$$

We see that, compared with the steepest descent algorithm, we have replaced the descent direction by

$$\Delta m_k = -H(m_k)^{-1} \nabla f(m_k), \tag{3.97}$$

and that the steplength is equal to 1.

We can give yet another interpretation to the Newton method applied in the frame of optimization. Consider the second-order development of the misfit function $f(m)$

$$f(m) = f(m_k) + \nabla f(m_k)^T (m - m_k) + \frac{1}{2}(m - m_k)^T H(m_k)(m - m_k) + o\left(\|m - m_k\|^2\right). \tag{3.98}$$

The iterate m_{k+1} is the minimizer of this quadratic approximation of $f(m_k)$. Indeed, let $q_k(m)$ be defined by

$$q_k(m) = f(m_k) + \nabla f(m_k)^T(m - m_k) + \frac{1}{2}(m - m_k)^T H(m_k)(m - m_k). \tag{3.99}$$

We have

$$\nabla q_k(m) = \nabla f(m_k) + H(m_k)(m - m_k). \tag{3.100}$$

Finding a minimizer of $q_k(m)$ consists in zeroing its gradient. We have

$$\begin{aligned} \nabla q_k(m) = 0 \quad &\iff \quad \nabla f(m_k) + H(m_k)(m - m_k) = 0 \\ &\iff \quad m = m_k - H(m_k)^{-1}\nabla f(m_k). \end{aligned} \tag{3.101}$$

Therefore we see that the Newton iteration, in the optimization frame, amounts to find at each iteration the minimizer of a local quadratic approximation of the misfit function $f(m)$.

What is the interest of the Newton method compared with the steepest-descent strategy? The convergence rate. While the steepest-descent exhibit poor convergence (it can be proven to have first-order convergence only), Newton algorithm can be proven to have second-order convergence. We look back at the 2 parameters example with $f(x_1, x_2)$ given by

$$f(x_1, x_2) = \frac{1}{2}\left(x_1^2 + \eta x_2^2\right). \tag{3.102}$$

We have

$$\nabla f(x_1, x_2) = (x_1, \eta x_2), \tag{3.103}$$

and

$$H(x_1, x_2) = \begin{pmatrix} 1 & 0 \\ 0 & \eta \end{pmatrix}. \tag{3.104}$$

Therefore

$$H(x_1, x_2)^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{\eta} \end{pmatrix}, \tag{3.105}$$

and we have in this case a descent direction given by

$$-H(x_1, x_2)^{-1}\nabla f(x_1, x_2) = -(x_1, x_2). \tag{3.106}$$

This proves that, starting from any point, the Newton algorithm converge in a single iteration in this case, no matter the shape of the ellipsoid.

Despite this appealing second-order convergence property, Newton method is not really usable for large scale minimization problems. This comes from the requirement to deal with the Hessian matrix and its inverse. As soon as the number of parameter reaches $O(10^4)$, storing, computing, and inverting the Hessian operator reveals computationally impracticable. To have an order of magnitude, for realistic size FWI problems, we often deal with $O(10^6)$ problems in 2D and up to $O(10^9)$ problems in 3D. This makes the Newton method unreachable for FWI. More tractable methods are thus required.

Quasi-Newton methods

Quasi-Newton methods have been designed to overcome the above mentioned difficulty associated with Hessian computation, storage, and inversion. In this frame, approximations of the inverse Hessian operator is used, such that the descent direction is given by

$$\Delta m_k = -Q_k \nabla f(m_k), \tag{3.107}$$

where Q_k is an approximation of $H(m_k)^{-1}$.

A very useful approximation is the one provided by Broyden Fletcher Goldfarb and Shanno, known as the BFGS approximation. We do not enter into the details of this approximation here. We simply mention that this method has been designed in the 80s (Nocedal, 1980) and that it relies on successive rank 2 updates computed from previous gradient values of an initial guess of the inverse Hessian operator (often chosen as an identity matrix).

In practice, a limited-memory version of this algorithm has been proposed under the name of l-BFGS algorithm, which is the sate of the art method for solving large scale unconstrained nonlinear minimization problem. The philosophy is to keep in memory only the l-previous gradient $\nabla f(m_k), \nabla f(m_{k-1}, \ldots, \nabla f(m_{k-l+1})$ and iterates m_k, m_{k-1}, m_{k-l+1}. From these quantities, an approximation Q_k of $H(m_k)^{-1}$ can be obtained, following the formula

$$\begin{aligned} Q_k \;=\; & \left(V_{k-1}^T \ldots V_{k-l}^T\right) Q_k^0 \left(V_{k-l} \ldots V_{k-1}\right) \\ & +\rho_{k-l} \left(V_{k-1}^T \ldots V_{k-l+1}^T\right) s_{k-l} s_{k-l}^T \left(V_{k-l+1} \ldots V_{k-1}\right) \\ & +\rho_{k-l+1} \left(V_{k-1}^T \ldots V_{k-l+2}^T\right) s_{k-l+1} s_{k-l+1}^T \left(V_{k-l+2} \ldots V_{k-1}\right) \\ & +\ldots \\ & +\rho_{k-1} s_{k-1} s_{k-1}^T, \end{aligned} \tag{3.108}$$

where the pairs s_k, y_k are

$$s_k = m_{k+1} - m_k, \quad y_k = \nabla f(m_{k+1}) - \nabla f(m_k), \tag{3.109}$$

the scalar ρ_k are

$$\rho_k = \frac{1}{y_k^T s_k}, \tag{3.110}$$

and the matrices V_k are defined by

$$V_k = I - \rho_k y_k s_k^T. \tag{3.111}$$

In its implementation, the l-BFGS algorithm does not build explicitly the Q_k matrix, which would be computationally challenging for large scale problems. It relies on a matrix-free double recursion loop, which makes it possible to compute directly the matrix vector product $Q_k m$, for a given model m and the stored values of the gradient and the models.

More details on the l-BFGS algorithm can be found in chapter 7 of Nocedal and Wright (2006). In particular, the double recursion loop is described in the algorithm 7.4 there, p. 178.

3.5 Gradient computation in FWI

Following the l-BFGS strategy, the key quantity to compute the implement FWI and solve the minimization problem associated with the reduced space approach (3.68) is the gradient of the misfit function $f(m)$. We discuss here how we can compute it efficiently in terms of computational cost, and how we can interpret it in terms of wave propagation and seismic imaging.

3.5.1 Direct method: the Jacobian matrix

Derivation

We start by the straightforward formula coming from the definition of the misfit function associated with the reduced space approach. We recall it here. We have

$$f(m) = \frac{1}{2}\|d_{cal}(m) - d_{obs}\|^2 \tag{3.112}$$

where

$$d_{cal}(m) = Ru(m), \;\; A(m)u = \varphi \tag{3.113}$$

To obtain the gradient of $f(m)$, we look at its first-order Taylor expansion. Let $p \in \mathbb{R}^M$, we have

$$f(m+p) = \frac{1}{2}\|d_{cal}(m+p) - d_{obs}\|^2 = \frac{1}{2}\|d_{cal}(m) + J(m)p - d_{obs}\|^2 + O(\|p\|^2) \tag{3.114}$$

where $J(m)$ is the Jacobian operator defined by

$$J(m) = \frac{\partial d_{cal}}{\partial m} \tag{3.115}$$

In details, $J(m)$ is a rectangular matrix of N rows and M columns, such that

$$J(m) = \begin{pmatrix} \frac{\partial d_{cal,1}}{\partial m_1} & \frac{\partial d_{cal,1}}{\partial m_2} & \cdots & \frac{\partial d_{cal,1}}{\partial m_M} \\ \frac{\partial d_{cal,2}}{\partial m_1} & \frac{\partial d_{cal,2}}{\partial m_2} & \cdots & \frac{\partial d_{cal,2}}{\partial m_M} \\ \vdots & \vdots & & \vdots \\ \frac{\partial d_{cal,N}}{\partial m_1} & \frac{\partial d_{cal,N}}{\partial m_2} & \cdots & \frac{\partial d_{cal,N}}{\partial m_M} \end{pmatrix} \tag{3.116}$$

Here, M is the total number of discrete parameters describing the medium, while N is the total number of discrete data. In the assumption where $N_s = 1$, we have

$$N = N_r \times N_t \tag{3.117}$$

where N_t is the number of time steps obtained after discretization of the wave equation.

We can develop the previous expression (3.114) to obtain

$$\begin{aligned} f(m+p) &= \frac{1}{2}\|d_{cal}(m) - d_{obs}\|^2 + (J(m)p, d_{cal}(m) - d_{obs}) + O(\|p\|^2) \\ &= f(m) + \left(p, J(m)^T\left(d_{cal}(m) - d_{obs}\right)\right) + O(\|p\|^2). \end{aligned} \tag{3.118}$$

We recall here the formal first-order Taylor development of the misfit function $f(m)$

$$f(m+p) = f(m) + (p, \nabla f(m)) + O(\|p\|^2). \tag{3.119}$$

By identification between (3.118) and (3.119), we deduce that

$$\nabla f(m) = J(m)^T\left(d_{cal}(m) - d_{obs}\right) \tag{3.120}$$

Physical interpretation

How can we interpret this formula in terms of wave propagation? We need to figure out what is the physical meaning behind the Jacobian operator. Let us consider the general wave equation

$$A(m)u = \varphi \tag{3.121}$$

where $A(m)$ stands for a general wave equation operator, u is the wavefield, and φ is the seismic source. We derive this equation with respect to the model parameter $m_i \in \mathbb{R}, \quad 1 \leq i \leq M$. The model parameter m_i has to be understood as one entry of the discrete vector $m \in \mathbb{R}^M$. Therefore it corresponds to the value of m at a specific spatial position denoted by i. If the wave equation is discretized using a finite-difference method, it would correspond exactly to one node of the Cartesian grid used to discretized the wave equation.

$$\frac{\partial A}{\partial m_i}u + A(m)\frac{\partial u}{\partial m_i} = 0 \tag{3.122}$$

Therefore we have

$$A(m)\frac{\partial u}{\partial m_i} = -\frac{\partial A}{\partial m_i}u \tag{3.123}$$

and

$$\frac{\partial u}{\partial m_i} = -A(m)^{-1}\frac{\partial A}{\partial m_i}u. \tag{3.124}$$

This shows that the quantity $\frac{\partial u}{\partial m_i}$ is the solution of the wave equation with a source term $-\frac{\partial A}{\partial m_i}u$. The latter source term corresponds the incident wavefield u multiplied by the operator $\frac{\partial A}{\partial m_i}$. This operator is in general very sparse, and localized around the position i. Just for the sake of simplicity, we can thus do the assumption

$$\frac{\partial A}{\partial m_i}u \simeq u_i(t) \tag{3.125}$$

where $u_i(t)$ denotes the value of the incident wavefield at the position denoted by i, which depends on time t. We thus have

$$\frac{\partial u}{\partial m_i} \simeq -A(m)^{-1}u_i(t) \tag{3.126}$$

and

$$R\frac{\partial u}{\partial m_i} \simeq -RA(m)^{-1}u_i(t) \tag{3.127}$$

where R is the restriction operator extracting the values of the wavefield at the receiver locations. We thus obtain

$$\frac{\partial Ru}{\partial m_i} = \frac{\partial d_{cal}}{\partial m_i} \simeq -RA(m)^{-1}u_i(t) \tag{3.128}$$

We thus see that ***the column i of the Jacobian matrix $J(m)$ is nothing else than the wavefield which would have been scattered by a perturbation at position m_i during the propagation of the field u, and which would have been recorded by the receivers.***

This interpretation is crucial to understand the meaning of the gradient formula (3.120). We introduce the residuals

$$\Delta d = d_{cal}(m) - d_{obs} \tag{3.129}$$

The residuals are the difference between the calculated data and the observed data. The aim of FWI is to reduce this difference, such that $d_{cal}(m)$ is as close as possible from d_{obs}. The residuals can be seen as what remains to be interpreted through FWI by modifying the current model parameter m, assuming the wave propagation model $A(m)$.

Following the formula (3.120), we have

$$\nabla f(m)_i = \sum_{j=1}^{N} J_{ij}^T \Delta d_j = \sum_{j=1}^{N} J_{ji} \Delta d_j = \sum_{r=1}^{N_r} \int_0^T \frac{\partial d_{cal}}{\partial m_i}(x_r, t) \Delta d(x_r, t) \tag{3.130}$$

This last equality comes from the fact that the summation over the N indices representing the data space imply a summation over the receivers N_r and a summation over time time indices N_t that we have replaced here by its continuous version (integral in time) for the sake of the interpretation.

Indeed, equation (3.130) reveals that ***each component i of the gradient is built as the summation over the receivers of the zero-lag time correlation between the residuals Δd and the wavefield which would have been scattered by a perturbation m_i of the medium at the spatial position i.***

This interpretation makes clear that the gradient is a first-order scattering interpretation of the residuals. We can see the construction of the gradient as if each position of space was perturbed, producing an associated scattered wavefield following the propagation of the incident wavefield u. Each scattering response is correlated with the residuals as if to test if the scattering response correlated constructively or destructively. If the correlation is constructive, the entry in the gradient is important. If the correlation is destructive, the entry in the gradient is negligible. Following the iterative minimization scheme, these model perturbation will be added (up to a scaling and a multiplication by the inverse Hessian, which will be discussed later) to the current model estimate.

A graphical illustration of this interpretation is provided in Figure 3.5.1

While this interpretation is crucial to understand the mechanism behind FWI, in practice the formula (3.120) is never implemented. Indeed, it would require to compute the scattered wavefield in each spatial location $1 \leq i \leq M$, which is strictly impracticable. For this reason, the FWI actual implementation relies on another gradient computation strategy, known as the adjoint state strategy.

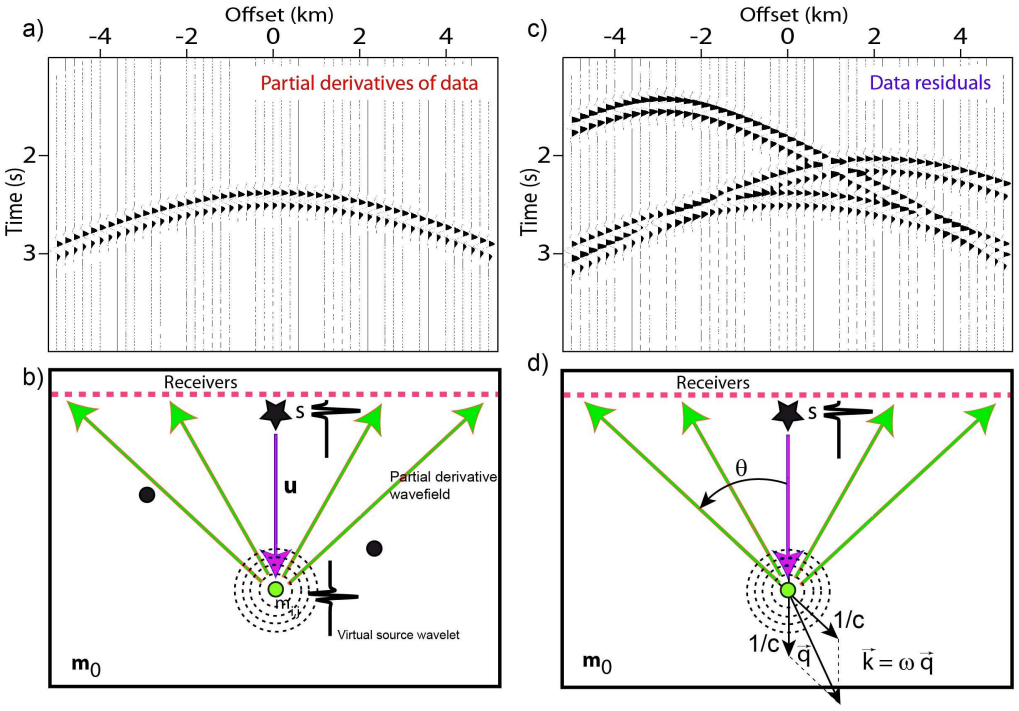

Figure 3.4: In this Figure, we consider a homogeneous background and three localized perturbations behaving as scatterers (2 black and 1 green circles). AN array of receivers is represented in dashed red line. The source is located in the middle of the receiver array. On the top left Figure, the scattering response recorded by the receivers corresponding to a perturbation located at the green circle position is presented. This correspond to one column of the Jacobian matrix, associated with this medium perturbation. On the top right Figure, we have the residuals obtained when d_{obs} is computed in the medium containing the three scatterers and the calculated data $d_{cal}(m)$ is computed in the medium without any scatterer. We see the imprint of the three missing perturbation in the data. In this case, the gradient at the position of the green circle will correlate constructively with the residuals: we see that the scattering response of the perturbation at the green circle corresponds to one of the scattering response in the residuals. Therefore the gradient will indicate to update the current model estimate m with a perturbation located at the position of the green circle.

3.5.2 Adjoint state method

The adjoint state strategy derives from control theory (Lions, 1968), and has been applied in seismic imaging first by (Chavent, 1974). A recent review of the application of the adjoint state method in geophysics is presented in (Plessix, 2006). Here we expose the main principle of the method before we present it in the particular case of 1D constant density acoustic wave equation.

General formalism

We recall the Lagrangian operator

$$L(m,u,d_{cal},\lambda,\mu) = \|d_{cal}-d_{obs}\|^2 + (A(m)u-\varphi,\lambda)_{\mathcal{W}} + (d_{cal}-Ru,\mu)_{\mathcal{D}}\,. \tag{3.131}$$

with the previous definition of the scalar products and norm.

In the case where u is the solution of the wave equation, we have

$$A(m)u = \varphi \tag{3.132}$$

We denote this solution by $\overline{u}(m)$. In the same spirit, when decal is extracted from $\overline{u}(m)$ we denote it by

$$\overline{d}_{cal}(m) = R\overline{u}(m). \tag{3.133}$$

We thus have

$$L(m,\overline{u}(m),\overline{d}_{cal}(m),\lambda,\mu) = \|\overline{d}_{cal}(m)-d_{obs}\|^2 = f(m) \tag{3.134}$$

Therefore, we can relate the gradient of $f(m)$ to the Lagrangian operator in this case, simply by deriving the latter equation with respect to the model parameters m. We obtain

$$\frac{L(m,\overline{u}(m),\overline{d}_{cal}(m),\lambda,\mu)}{\partial m} = \nabla f(m) \tag{3.135}$$

If we expand the left hand side, we obtain

$$\left(\frac{\partial A}{\partial m}u,\lambda\right)_T + \frac{L(m,\overline{u}(m),\overline{d}_{cal}(m),\lambda,\mu)}{\partial u}\frac{\partial\overline{u}(m)}{m} + \frac{L(m,\overline{u}(m),\overline{d}_{cal}(m),\lambda,\mu)}{\partial d_{cal}}\frac{\partial\overline{d}_{cal}(m)}{m} = \nabla f(m) \tag{3.136}$$

where $(.,.)_T$ is a scalar product in a space yet to be precise. We can recognize in $\frac{\partial\overline{u}(m)}{m}$ and $\frac{\partial\overline{d}_{cal}(m)}{m}$ operators close from the Jacobian operator, which we do not want to compute. We therefore look for adjoint variables λ and μ such that

$$\begin{aligned}\frac{L(m,\overline{u}(m),\overline{d}_{cal}(m),\lambda,\mu)}{\partial u} &= 0\\ \frac{L(m,\overline{u}(m),\overline{d}_{cal}(m),\lambda,\mu)}{\partial d_{cal}} &= 0\end{aligned} \tag{3.137}$$

Interestingly, we see that these two equation belong to the KKT conditions discussed in the beginning of this chapter. To compute λ and μ satisfying the two latter equations, we need to express the derivatives of the Lagrangian with respect to u and d_{cal} respectively. To this purpose, we consider $v \in \mathcal{W}$ and $d \in \mathcal{D}$ to obtain the first order Taylor developments. We have

$$L(m,u+v,d_{cal},\lambda,\mu) = \|d_{cal}-d_{obs}\|^2 + (A(m)(u+v)-\varphi,\lambda)_{\mathcal{W}} + (d_{cal}-R(u+v),\mu)_{\mathcal{D}}\,. \tag{3.138}$$

Therefore

$$L(m, u+v, d_{cal}, \lambda, \mu) = L(m, u, d_{cal}, \lambda, \mu) + (A(m)v, \lambda)_{\mathcal{W}} + (d_{cal} - Rv, \mu)_{\mathcal{D}}\,. \tag{3.139}$$

Hence

$$L(m, u+v, d_{cal}, \lambda, \mu) = L(m, u, d_{cal}, \lambda, \mu) + \left(v, A(m)^T\lambda\right)_{\mathcal{W}} + \left(v, -R^T\mu\right)_{\mathcal{W}}\,, \tag{3.140}$$

and

$$L(m, u+v, d_{cal}, \lambda, \mu) = L(m, u, d_{cal}, \lambda, \mu) + \left(v, A(m)^T\lambda - R^T\mu\right)_{\mathcal{W}}\,. \tag{3.141}$$

This shows that

$$\frac{L(m, \overline{u}(m), \overline{d}_{cal}(m), \lambda, \mu)}{\partial u} = A(m)^T\lambda - R^T\mu \tag{3.142}$$

Similarly, we have

$$L(m, u, d_{cal}+d, \lambda, \mu) = \|d_{cal} + d - d_{obs}\|^2 + (d_{cal} + d - Ru, \mu)_{\mathcal{D}}\,. \tag{3.143}$$

Therefore

$$L(m, u, d_{cal}+d, \lambda, \mu) = \|d_{cal} - d_{obs}\|^2 + (d, d_{cal} - d_{obs})_{\mathcal{D}} + O(\|d\|^2)\,(d, \mu)_{\mathcal{D}}\,.\,(d_{cal} - Ru, \mu)_{\mathcal{D}}\,, \tag{3.144}$$

and

$$L(m, u, d_{cal}+d, \lambda, \mu) = L(m, u, d_{cal}, \lambda, \mu) + (d, d_{cal} - d_{obs} + \mu)_{\mathcal{D}} + O(\|d\|^2). \tag{3.145}$$

Hence

$$\frac{L(m, \overline{u}(m), \overline{d}_{cal}(m), \lambda, \mu)}{\partial d_{cal}} = d_{cal} - d_{obs} + \mu \tag{3.146}$$

We thus obtain that system (3.137) is equivalent to

$$\begin{aligned} A(m)^T\lambda - R^T\mu &= 0 \\ d_{cal} - d_{obs} + \mu &= 0 \end{aligned} \tag{3.147}$$

which can be rewritten as

$$\begin{aligned} A(m)^T\lambda &= R^T\mu \\ \mu &= d_{obs} - d_{cal} \end{aligned} \tag{3.148}$$

The latter two equations show that to cancel the terms of the gradient expression in (3.136), the adjoint states λ and μ need to be computed such that

1. μ is equal to the opposite of the residual Δd introduced earlier;
2. λ is the solution of an adjoint equation associated with the operator $A(m)^T$ with a source term equal to $R^T\mu$ that is, following the first equation, the residuals Δd projected back into the wavefield space $\mathcal{W}$ to act as a regular wavefield source

This result is a key for an efficient implementation of FWI. The gradient $\nabla f(m)$ can be computed as

$$\nabla f(m) = \left(\frac{\partial A}{\partial m}u, \lambda\right)_T \tag{3.149}$$

where u is the incident field solution of the wave equation

$$A(m)u = \varphi \tag{3.150}$$

and λ is the adjoint field, solution of the adjoint wave equation

$$A(m)^T\lambda = -R^T\Delta d. \tag{3.151}$$

In practice, wave equation operators are self adjoint, and we are going to see that the transpose here implies that the wavefield λ is solution of the same wave equation as u, except for the source term, which is equal to the residuals backprojected from the receiver positions. In addition, the adjoint field λ is computed ***backward in time*** instead of being computed forward in time as it is the case for u. To have a better insight where this backpropagation in time comes from, we investigate in details the 1D constant density acoustic case in the following.

1D constant density acoustic approximation

Let us consider now the special case of 1D constant density acoustic approximation. In this frame, $m = V_P$ and the wave equation

$$A(m)u = \varphi \tag{3.152}$$

becomes

$$\frac{1}{V_P^2}\frac{\partial^2 u}{\partial t^2} - \frac{\partial^2 u}{\partial x^2} = \varphi \tag{3.153}$$

with initial conditions

$$u(x,0) = 0, \quad \frac{\partial u}{\partial t}(x,0) = 0 \tag{3.154}$$

and homogeneous Dirichlet boundary conditions

$$\forall x \in \partial\Omega, \quad u(x,t) = 0. \tag{3.155}$$

The Lagrangian operator is going to account for all the constraints defining the forward problem, therefore we add terms to account for the initial conditions and the boundary conditions. We introduce the adjoint variables

$$\phi_1(x), \quad \phi_2(x), \quad x \in \Omega \tag{3.156}$$

to account for the initial conditions and

$$\phi_3(x,t), \quad (x,t) \in \partial\Omega \times [0,T] \tag{3.157}$$

to account for the boundary conditions.

We also introduce the scalar product associated to the boundary functions $(.,.)_\Omega$ and $(.,.)_{\partial\Omega\times[0,T]}$ such that

$$(u,v)_\Omega = \int_\Omega u(x)v(x)dx, \tag{3.158}$$

and

$$(u,v)_{\partial\Omega\times[0,T]} = \int_0^T \int_{\partial\Omega} u(x,t)v(x,t)dxdt. \tag{3.159}$$

The Lagrangian operator becomes

$$
\begin{aligned}
&L(V_P, u, d_{cal}, \lambda, \mu, \phi_1, \phi_2, \phi_3) = \\
&\|d_{cal} - d_{obs}\|^2 + \left(\frac{1}{V_P^2}\frac{\partial^2 u}{\partial t^2} - \frac{\partial^2 u}{\partial x^2} - \varphi, \lambda\right)_{\mathcal{W}} + (d_{cal} - Ru, \mu)_{\mathcal{D}} \\
&+ (u(x,0), \phi_1)_\Omega + \left(\frac{\partial u(x,0)}{\partial t}, \phi_2\right)_\Omega + (u, \phi_3)_{\partial\Omega\times[0,T]}
\end{aligned}
\tag{3.160}
$$

To obtain the equations for the adjoint variables, we have seen in the general case that we need to cancel the first order derivatives of the Lagrangian operator with respect to u and d_{cal}. To do so, we proceed as in the general case by first computing the first-order Taylor's development with respect to these two variables to obtain the corresponding partial derivatives.

The first-order Taylor's development of the Lagrangian with respect to u is given by considering the quantity

$$
\begin{aligned}
&L(V_P, u + v, d_{cal}, \lambda, \mu) = \\
&\|d_{cal} - d_{obs}\|^2 + \left(\frac{1}{V_P^2}\frac{\partial^2 (u+v)}{\partial t^2} - \frac{\partial^2 (u+v)}{\partial x^2} - \varphi, \lambda\right)_{\mathcal{W}} + (d_{cal} - R(u+v), \mu)_{\mathcal{D}} \\
&+ (u(x,0) + v(x,0), \phi_1)_\Omega + \left(\frac{\partial u(x,0)}{\partial t} + \frac{\partial v(x,0)}{\partial t}, \phi_2\right)_\Omega + (u+v, \phi_3)_{\partial\Omega\times[0,T]}
\end{aligned}
\tag{3.161}
$$

We have

$$
\begin{aligned}
L(V_P, u + v, d_{cal}, \lambda, \mu) = \quad & L(V_P, u, d_{cal}, \lambda, \mu) \\
& + \int_\Omega \int_{t=0}^T \left(\frac{1}{V_P^2}\frac{\partial^2 v(x,t)}{\partial t^2} - \frac{\partial^2 v(x,t)}{\partial x^2}\right) \lambda(x,t) dx dt \\
& + \sum_{r=1}^{N_r} \int_{t=0}^T (Rv)(x_r, t)\mu(x_r, t) dt \\
& + \int_\Omega v(x,0)\phi_1(x) dx + \int_\Omega \frac{\partial v(x,0)}{\partial t}\phi_1(x) dx + \int_{\partial\Omega}\int_0^T v(x,t)\phi_3(x,t) dx dt
\end{aligned}
\tag{3.162}
$$

Let us consider the first term of the right hand side. We are going to use integration by parts (in time and space). A first integration by part in time gives

$$
\begin{aligned}
\int_\Omega \int_{t=0}^T \frac{1}{V_P^2}\frac{\partial^2 v(x,t)}{\partial t^2}\lambda(x,t) dx dt = \quad & - \int_\Omega \int_{t=0}^T \frac{1}{V_P^2}\frac{\partial v(x,t)}{\partial t}\frac{\partial \lambda(x,t)}{\partial t} dx dt \\
& + \int_\Omega \frac{1}{V_P^2}\frac{\partial v(x,0)}{\partial t}\lambda(x,0) dx - \int_\Omega \frac{1}{V_P^2}\frac{\partial v(x,T)}{\partial t}\lambda(x,T) dx
\end{aligned}
\tag{3.163}
$$

A second integration by part in time gives

$$\begin{aligned}\int_\Omega \int_{t=0}^{T} \frac{1}{V_P^2}\frac{\partial^2 v(x,t)}{\partial t^2}\lambda(x,t)dxdt &= \int_\Omega \int_{t=0}^{T} v(x,t)\frac{1}{V_P^2}\frac{\partial^2 \lambda(x,t)}{\partial t^2}dxdt \\ &+ \int_\Omega \frac{1}{V_P^2}\frac{\partial v(x,0)}{\partial t}\lambda(x,0)dx - \int_\Omega \frac{1}{V_P^2}\frac{\partial v(x,T)}{\partial t}\lambda(x,T)dx \\ &- \int_\Omega \frac{1}{V_P^2}v(x,0)\frac{\partial \lambda(x,0)}{\partial t}dx + \int_\Omega \frac{1}{V_P^2}v(x,T)\frac{\partial \lambda(x,T)}{\partial t}dx\end{aligned} \tag{3.164}$$

In addition, an integration by part in space gives

$$\int_\Omega \int_{t=0}^{T} \frac{\partial^2 v(x,t)}{\partial x^2}\lambda(x,t)dxdt = -\int_\Omega \int_{t=0}^{T} \frac{\partial v(x,t)}{\partial x}\frac{\partial \lambda(x,t)}{\partial x}dxdt + \int_{t=0}^{T}\int_{\partial\Omega} v(x,t)\frac{\partial \lambda(x,t)}{\partial x}dxdt \tag{3.165}$$

A second integration by part in space yields

$$\begin{aligned}\int_\Omega \int_{t=0}^{T} \frac{\partial^2 v(x,t)}{\partial x^2}\lambda(x,t)dxdt &= \int_\Omega \int_{t=0}^{T} v(x,t)\frac{\partial^2 \lambda(x,t)}{\partial x^2}dxdt + \int_{t=0}^{T}\int_{\partial\Omega} v(x,t)\frac{\partial \lambda(x,t)}{\partial x}dxdt \\ &- \int_{t=0}^{T}\int_{\partial\Omega} \frac{\partial v(x,t)}{\partial x}\lambda(x,t)dxdt\end{aligned} \tag{3.166}$$

Before assembling the results, a last technical step remains to be investigated. Based on the definition of the restriction operator, we have

$$\sum_{r=1}^{N_r}\int_{t=0}^{T}(Rv)(x_r,t)\mu(x_r,t)dt = \sum_{r=1}^{N_r}\int_{t=0}^{T}\int_\Omega \delta(x-x_r)v(x)dx\mu(x_r,t)dt \tag{3.167}$$

Therefore we have

$$\sum_{r=1}^{N_r}\int_{t=0}^{T}(Rv)(x_r,t)\mu(x_r,t)dt = \int_{t=0}^{T}\int_\Omega \sum_{r=1}^{N_r}\mu(x_r,t)\delta(x-x_r)v(x)dxdt \tag{3.168}$$

Based on all the previous integration by parts and this last rewriting, we obtain the

following identity

$$\begin{aligned} L(V_P, u+v, d_{cal}, \lambda, \mu) &= L(V_P, u, d_{cal}, \lambda, \mu) \\ &+ \int_\Omega \int_{t=0}^{T} \left(\frac{1}{V_P^2} \frac{\partial^2 \lambda(x,t)}{\partial t^2} - \frac{\partial^2 \lambda(x,t)}{\partial x^2} \right) v(x,t) dx dt \\ &+ \int_{t=0}^{T} \int_\Omega \sum_{r=1}^{N_r} \mu(x_r, t) \delta(x - x_r) v(x) dx dt \\ &+ \int_\Omega v(x,0)\phi_1(x) dx + \int_\Omega + \frac{\partial v(x,0)}{\partial t} \phi_1(x) dx + \int_{\partial\Omega} \int_0^T v(x,t)\phi_3(x,t) dx dt \\ &+ \int_\Omega \frac{1}{V_P^2} \frac{\partial v(x,0)}{\partial t} \lambda(x,0) dx - \int_\Omega \frac{1}{V_P^2} \frac{\partial v(x,T)}{\partial t} \lambda(x,T) dx \\ &- \int_\Omega \frac{1}{V_P^2} v(x,0) \frac{\partial \lambda(x,0)}{\partial t} dx + \int_\Omega \frac{1}{V_P^2} v(x,T) \frac{\partial \lambda(x,T)}{\partial t} dx \\ &+ \int_{t=0}^{T} \int_{\partial\Omega} v(x,t) \frac{\partial \lambda(x,t)}{\partial x} dx dt - \int_{t=0}^{T} \int_{\partial\Omega} \frac{\partial v(x,t)}{\partial x} \lambda(x,t) dx dt \end{aligned} \tag{3.169}$$

This makes possible the following rewriting in terms of scalar products

$$\begin{aligned} L(V_P, u+v, d_{cal}, \lambda, \mu) &= L(V_P, u, d_{cal}, \lambda, \mu) \\ &+ \left(\frac{1}{V_P^2} \frac{\partial^2 \lambda(x,t)}{\partial t^2} - \frac{\partial^2 \lambda(x,t)}{\partial x^2} + \sum_{r=1}^{N_r} \mu(x_r, t) \delta(x - x_r), v \right)_{\mathcal{W}} \\ &+ \left(\phi_1(x) + \frac{\partial \lambda(x,0)}{\partial t}, v(x,0) \right)_\Omega + \left(\phi_2(x) + \lambda(x,0), \frac{\partial v(x,0)}{\partial t} \right)_\Omega \\ &+ \left(\phi_3(x,t) + \frac{\partial \lambda(x,t)}{\partial x}, v(x,t) \right)_{\partial\Omega \times [0,T]} \\ &- \left(\frac{1}{V_P^2} \lambda(x,T), \frac{\partial v(x,T)}{\partial t} \right)_\Omega + \left(\frac{1}{V_P^2} \frac{\partial \lambda(x,T)}{\partial t}, v(x,T) \right)_\Omega \\ &- \left(\lambda(x,t), \frac{\partial v(x,t)}{\partial x} \right)_{\partial\Omega \times [0,T]} \end{aligned} \tag{3.170}$$

We thus have

$$\forall v \in \mathcal{W},\ \ \frac{\partial L(V_P, u, d_{cal}, \lambda, \mu)}{\partial u}.v = 0 \iff$$

$$\begin{cases} \dfrac{1}{V_P^2}\dfrac{\partial^2 \lambda(x,t)}{\partial t^2} - \dfrac{\partial^2 \lambda(x,t)}{\partial x^2} + \displaystyle\sum_{r=1}^{N_r} \mu(x_r,t)\delta(x-x_r) = 0 \\ \phi_1(x) + \dfrac{\partial \lambda(x,0)}{\partial t} = 0 \\ \phi_2(x) + \lambda(x,0) = 0 \\ \phi_3(x,t) + \dfrac{\partial \lambda(x,t)}{\partial x} = 0 \\ \lambda(x,T) = 0 \\ \dfrac{\partial \lambda(x,T)}{\partial t} = 0 \\ \lambda(x,t) = 0 \ \ x \in \partial\Omega \end{cases} \tag{3.171}$$

For the derivative of the Lagrangian with respect to d_{cal}, the previous derivation in the general case remains valid here, and we have

$$\frac{\partial L(V_P, u, d_{cal}, \lambda, \mu)}{\partial d_{cal}} = d_{cal} - d_{obs} + \mu \tag{3.172}$$

Putting everything together thus yields the following systems for the adjoint variable $\lambda, \mu, \phi_1, \phi_2, \phi_3$

$$\begin{cases} \dfrac{1}{V_P^2}\dfrac{\partial^2 \lambda(x,t)}{\partial t^2} - \dfrac{\partial^2 \lambda(x,t)}{\partial x^2} = -\displaystyle\sum_{r=1}^{N_r} \mu(x_r,t)\delta(x-x_r) \\ \mu(x_r,t) = d_{cal}(x_r,t) - d_{obs}(x_r,t) \\ \lambda(x,T) = 0 \\ \dfrac{\partial \lambda(x,T)}{\partial t} = 0 \\ \lambda(x,t) = 0 \ \ x \in \partial\Omega \\ \phi_1(x) = -\dfrac{\partial \lambda(x,0)}{\partial t} \\ \phi_2(x) = -\lambda(x,0) \\ \phi_3(x,t) = -\dfrac{\partial \lambda(x,t)}{\partial x} \end{cases} \tag{3.173}$$

This derivation is important as it clearly illustrates how the adjoint field is the solution of the same wave equation as the incident field, except that it is defined through ***final***

conditions and that the source term corresponds to the residuals.

Now the gradient term needs to be made explicit in the 1D acoustic approximation. We have

$$\begin{aligned}\frac{\partial L(V_P,u,d_{cal},\lambda,\mu)}{\partial V_P} &= \int_0^T \frac{\partial}{\partial V_P}\int_\Omega \left(\frac{1}{V_P^2}\frac{\partial^2 u(x,t)}{\partial t^2} - \frac{\partial^2 u(x,t)}{\partial x^2} - \varphi(x,t)\right)\lambda(x,t)dxdt \\ &= -\frac{2}{V_P^3}\int_0^T \frac{\partial^2 u(x,t)}{\partial t^2}\lambda(x,t)dt\end{aligned} \tag{3.174}$$

We recover the fact that the gradient in this case is computed as the zero-lag time correlation of the incident wavefield $u(x,t)$ and the adjoint wavefield λ, scaled by the quantity $-\frac{2}{V_P^3}$, with λ solution of the adjoint wave system (3.173).

3.5.3 Physical interpretation of the gradient in FWI

See slides of SESSION 9.

3.5.4 Gradient as a filter in the wavenumber domain

In the frame of imaging, it is important to relate the wavenumber content of the gradient to the parameters the seismic imaging experiment : acquisition, velocity, frequency content of the data. This is proposed in the papers of Wu and Toksöz (1987) and Sirgue (2003) through a plane wave analysis of the gradient.

In this frame, we consider a simple 2D case with a single source/receiver couple. We assume that the incident field (respectively the adjoint field) is a monochromatic field propagating in the direction denoted by the unit vector $\mathbf{p_s}$, from the source), (respectively $\mathbf{p_r}$, from the receiver). We have

$$u(\mathbf{x},\omega) = e^{ik_0\mathbf{p_s}.\mathbf{x}},\ \ \lambda(\mathbf{x},\omega) = e^{ik_0\mathbf{p_r}.\mathbf{x}}, \tag{3.175}$$

where $\mathbf{x} = (x,z) \in \mathbb{R}^2$ and k_0 is given by

$$k_0 = \frac{2\pi\omega}{c_0}. \tag{3.176}$$

In equations (3.175) and (3.176) ω is the frequency of the monochromatic plane waves, while c_0 is the velocity of the medium.

As a zero lag cross-correlation of the two wavefields (neglecting the weighting term), the gradient can be approximated as the product of the two monochromatic field. Indeed, following the convolution theorem, the cross-correlation in the time-domain is equivalent to a product in the frequency-domain. Therefore we have the following approximation for the gradient

$$\nabla f \simeq e^{ik_0\mathbf{p_s}.\mathbf{x}}e^{ik_0\mathbf{p_r}.\mathbf{x}} = e^{ik0(\mathbf{p_s}+\mathbf{p_r}).\mathbf{x}} \tag{3.177}$$

The situation is illustrated in Figure 3.5 where we consider a diffracting point at $\mathbf{x}$. The wavenumber content of the gradient, following equation (3.177), is

$$k = k_0\left(\mathbf{p_s} + \mathbf{p_r}\right). \tag{3.178}$$

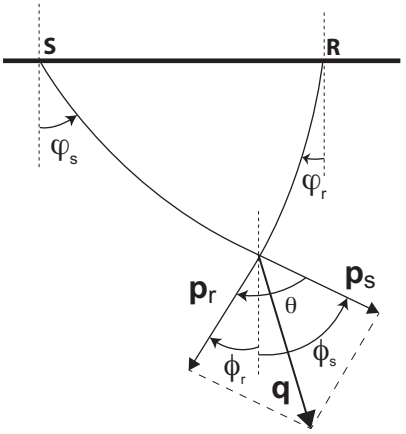

Figure 3.5: S and R denote the source and receiver position at the surface. $\mathbf{p_s}$ and $\mathbf{p_r}$ are the vectors in the direction of the rays connecting the diffraction point x to the source and receiver respectively. These rays can be computed in the asymptotic approximation. In the plane wave approximation we consider here, these rays are straight. The angle θ is the angle formed by the vector $\mathbf{p_s}$ and $\mathbf{p_r}$, while the angle ϕ is the sum of the angles ϕ_s and ϕ_r corresponding to the angle made by $\mathbf{p_s}$ and $\mathbf{p_r}$ with the vertical.

We can express it through the illumination angle θ formed by the two vectors $\mathbf{p_s}$ and $\mathbf{p_r}$ at point $\mathbf{x}$. They connect the diffracting point $\mathbf{x}$ to the source S and the receiver R. In this plane wave analysis, we consider that the ray connecting them are straight. We have

$$\|k\|^2 = k_0^2 \|\mathbf{p_s} + \mathbf{p_r}\|^2 = k_0^2 \left(\|\mathbf{p_s}\|^2 + \|\mathbf{p_r}\|^2 + 2\mathbf{p_s}.\mathbf{p_r} \right) \tag{3.179}$$

We use the fact that $\mathbf{p_s}$ and $\mathbf{p_r}$ are unit vector (their norm is equal to 1), and that the scalar product between two vectors is given by the product of their norm and the cosine of the angle between them to obtain

$$\|k\|^2 = k_0^2 \|\mathbf{p_s} + \mathbf{p_r}\|^2 = 2k_0^2 \left(1 + \cos\theta\right) \tag{3.180}$$

Here we use the following trigonometric identity

$$\cos 2x = \cos(x + x) = \cos^2 x - \sin^2 x = \cos^2 x - (1 - \cos^2 x) = 2\cos^2 x - 1. \tag{3.181}$$

Therefore

$$1 + \cos x = 2\cos^2\left(\frac{x}{2}\right). \tag{3.182}$$

Hence

$$\|k\|^2 = 4k_0^2 \cos^2\left(\frac{\theta}{2}\right), \tag{3.183}$$

and

$$\|k\| = 2k_0 \cos\left(\frac{\theta}{2}\right) \tag{3.184}$$

that is

$$k = \frac{4\pi\omega}{c_0} \cos\left(\frac{\theta}{2}\right) (\cos\phi, \sin\phi) \tag{3.185}$$

where

$$\phi = \phi_s + \phi_r \tag{3.186}$$

with ϕ_s and ϕ_r being the angle made by $\mathbf{p_s}$ and $\mathbf{p_r}$ with the vertical direction (see Fig. 3.5).

Through the formula (3.185), a point in the wavenumber space can be attached to each source/receiver pair. An example is provided in Figure 3.6, where we consider a velocity $c_0 = 2000\text{m.s}^{-1}$, a mean frequency $f_0 = 5$ Hz, a target point at x=5000 m and $z = -3000$ m, and a fixed spread acquisition with 101 sources and 101 receivers located close to the surface at z= -5 m, spanning the horizontal axis from $x = 0$ m to $x = 10000$ m, with an interval distance $\Delta x = 100$ m.

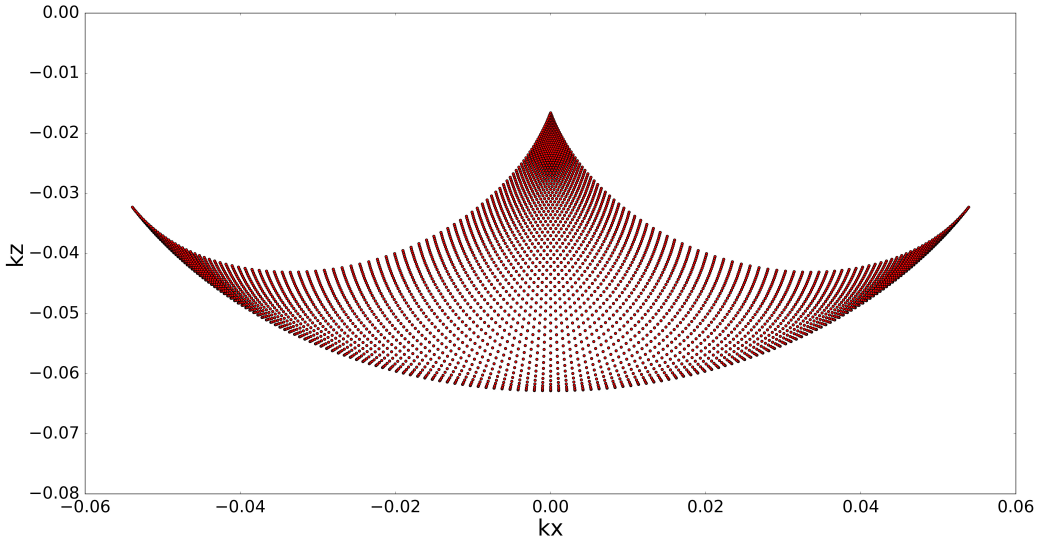

Figure 3.6: k_x and k_z denotes the horizontal and vertical wavenumber component in the 2D approximation. Each point represented here corresponds to a couple of sources and receivers. The mean frequency f_0 is equal to 5 Hz. The velocity c_0 is equal to 2000 m.s^{-1}. An ensemble of 101 sources and receivers is considered here, spanning the horizontal axis from $x = 0$ m to $x = 10000$ m, with an interval distance between each source and receiver $\Delta x = 100$ m.

Formula (3.185) is very important, as it tells what is the wavenumber content injected in the model through the gradient update, depending on the illumination angle θ, the frequency content of the data ω, and the velocity c_0. As expected, for wavenumber content varies linearly the data frequency: low frequency data fills the low wavenumber content of the model, while high frequency content gives the information on higher wavenumber content (higher resolution).

Interestingly, one can also plays on the illumination angle to modify the wavenumber content injected in the model. This is important because in practice the frequency of the data is constrained by the source which is available. In seismic exploration, propagating low frequency content (lower than 3 Hz) requires very energetic source, which is not always feasible. Accessing low wavenumber content is thus an issue. This can be partly overcome by playing with the acquisition design. For illumination angles close from 180 degrees, equation (3.185) tells that the wavenumber content of the gradient tends to 0. Accessing such large illumination angles (especially at depth) requires to separate sources and re-

ceivers by a large distance (offset). The larger the offset is, the deeper balistic/transmitted energy recorded by the source/receiver couple is.

In practice, if streamer acquisition is used, the maximum offset is limited by the size of the cables towed by the acquisition ships. In modern acquisition, streamer length can reach around 10 km long. Another strategy, more and more used both in the academy and the industry, consists in using receiver spanned on the sea floor, named as Ocean Bottom Station (OBS). In this frame, the acquisition ships can shoot further aways from the OBS than in conventional streamer acquisition, and offsets can reach several tens of kilometers, provided the recording time is sufficiently long.

On the other hand, formula (3.185) also tells that higher resolution information corresponds to illumination angle close to 0. This corresponds to reflected energy, instead of balistic/transmitted energy for large illumination angles. In other words, high resolution information is contained in the reflection part of the data. Correctly interpreting the reflections will provide the highest resolution information on the subsurface model.

Finally, one can also see that the wavenumber content of the gradient is directly linked to the inverse of the velocity value. In deep part, where the velocity reaches 4000 $\mathrm{m.s}^{-1}$ and higher values, lower resolution is thus expected, compared with shallower/less consolidated zones where the P-wave velocity can be around 2000 to 3000 $\mathrm{m.s}^{-1}$. This also means that higher resolution should be expected from the reconstruction of S-wave velocity compared with P-wave velocity in the frame of elastic full waveform inversion, as S-wave velocity is always lower than the P-wave velocity.

See illustration slides of SESSION 10.

3.6 HPC aspects of FWI

3.6.1 Frequency-domain implementations

In the early implementations of FWI, instead of solving dynamic problems, the wave propagation modeling was converted to the solution of time-harmonic equations. In the elastic approximation, the time-domain partial differential equations can be written in compact form as

$$\rho \frac{\partial^2 u}{\partial t^2} - DCD^T u = s \tag{3.187}$$

Applying a Fourier transform, this leads to the solution of the elliptic problem

$$-\rho\omega^2 u - DCD^T u = s \tag{3.188}$$

where the Fourier transform operation symbol is omitted to simplify the notations, and $\omega = 2\pi f$ is the circular frequency, while f is the frequency of the monochromatic field u which is computed.

In the acoustic approximation, this comes to the solution of the following elliptic problem

$$-\omega^2 P - V_P^2 \Delta P = s \tag{3.189}$$

This elliptic problem is referred to as Helmholtz problem in the literature.

After discretization, this comes to the solution of the linear system

$$(\omega^2 I + V_P^2 D)P = -s \tag{3.190}$$

where D is the discretization of the Laplacian operator $\Delta = \frac{\partial}{\partial^2 x} + \frac{\partial}{\partial^2 y} + \frac{\partial}{\partial^2 z}$.

In terms of gradient computation, the correlation in time of the incident and adjoint field becomes simply a product between the two monochromatic fields, such that if we compute the gradient in the acoustic case we obtain

$$\nabla f(m) = -\frac{2\omega^2}{V_P^3} u(x,\omega)\lambda(x,\omega) \tag{3.191}$$

This simplifies quite significantly the gradient computation step, as we get rid of the time integral which, as will be explained in the next section, requires careful implementation when working in the time domain. If multiple frequencies are considered at the same time, the integration in time is simply replaced by a summation over frequencies, such that the gradient is given by

$$\nabla f(m) = -\sum_{\omega} \frac{2\omega^2}{V_P^3} u(x,\omega)\lambda(x,\omega) d\omega \tag{3.192}$$

In practice, two strategies exist to solve large scale linear systems

- iterative solvers
- direct solvers

Solution through iterative solvers

Iterative solvers can be good strategies when the conditioning of the linear system is good (and the spectrum as good properties, namely eigenvalues relatively well separated). Such iterative solvers are GMRES (Saad, 1986), or conjugate gradient for symmetric positive definite matrices (Hestenes and Stiefel, 1952). In practice, they do not require too much memory, because only matrix-vector products need to be computed, and the matrix of the linear system needs not be assembled explicitly. We speak of matrix-free algorithms. In addition, they can efficiently benefit from parallel computing architecture, as a matrix-vector product can be easily parallelized on multiple-core computers.

However, when the condition number of the linear system is bad, or the spectrum as clustered eigenvalues, then the convergence of iterative solvers might be dramatically slow, leading to bad computational cost/computational efficiency. The discretization of the Helmholtz problem yields such ill-conditioned linear systems. Let us see a simple example in 1D acoustic.

In this case D is a tridiagonal matrix of the form

$$D = \frac{1}{h^2} D_0 \tag{3.193}$$

where

$$D_0 = \begin{pmatrix} -2 & 1 & & & \\ 1 & -2 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & -2 & 1 \\ & & & 1 & -2 \end{pmatrix} \tag{3.194}$$

We can rewrite the Helmholtz system as

$$\left(\frac{\omega^2 h^2}{V_P^2} I + D_0\right) P = -\frac{s h^2}{V_P^2} \tag{3.195}$$

We introduce the matrix A as

$$A = \frac{\omega^2 h^2}{V_P^2} I + D_0 \tag{3.196}$$

The minimum wavelength λ_{min} is such that

$$\lambda_{min} = \frac{V_{P,min}}{f} \tag{3.197}$$

where $V_{P,min}$ is the minimum value of the P-wave velocity. To have a correct numerical dispersion with a second-order finite difference scheme, we need to ensure at least 10 discretization point per minimum wavelength, which yields

$$h = \frac{V_{P,min}}{10 f} \tag{3.198}$$

If we assume the medium is homogeneous (constant velocity V_P, the coefficient on the diagonal of A is thus

$$\frac{\omega^2 h^2}{V_P^2} - 2 = \frac{4\pi^2 f^2 V_P^2}{V_P^2 100 f^2} - 2 = \frac{\pi^2}{25} - 2 \simeq -1.6 \tag{3.199}$$

A direct observation tells that the matrix is not diagonal dominant for this value on the diagonal. In this particular case the values of the matrix remain the same but the size of the matrix increases as the frequency increases, for a given model size, following the definition of the discretization step. It can be observed numerically that the condition number of this matrix increases very fast with the size of the matrix.

Designing fast convergence iterative solvers for the Helmholtz equation or Helmholtz-type equations is a notoriously difficult problem, see for instance (Ernst and Gander, 2012). In practice, one should rely on specific preconditioning techniques. Preconditioning for a linear system

$$Ax = b \tag{3.200}$$

means considering solving a (left-) preconditioned linear system

$$PAx = Pb \tag{3.201}$$

where P is an approximate of A^{-1} such that the condition number and the spectral properties of the matrix PA are better than that of the matrix A, or a (right-) preconditioned linear system

$$APy = b, \;\; x = Py \tag{3.202}$$

where again P is an approximate of A^{-1} such that the condition number and the spectral properties of the matrix AP are better than that of the matrix A.

In practice, numerous strategies have been proposed for efficient and scalable preconditioning operator P for Helmholtz-like equation but none is fully satisfactory. All the methods finally face scalability issues when considering high frequency problems for which a large number of wavelengths needs to be propagated.

Solution through direct solvers

To avoid convergence issues related to the use iterative solvers, FWI implementation for time-harmonic equations rely on direct solvers. In this frame, a lower/upper triangular factorization of the matrix A is performed, such that

$$A = LU \tag{3.203}$$

with L and U respectively lower and upper triangular matrices. Such a decomposition is obtained through a Gaussian elimination strategy. Sophisticated parallel Gaussian elimination algorithms exist today, one instance of solvers implementing it being the MUMPS solver (MUMPS-team, 2019). MUMPS rely on a multi-frontal strategy to perform parallel LU decomposition of a given matrix (Amestoy et al., 2000; Duff et al., 1986).

One significant advantage of such direct solver strategy is to benefit from a reduced computation time for each right hand side as soon as the factorization has been computed. Indeed, once A is factorized, the problem

$$Ax = b \tag{3.204}$$

is equivalent to

$$LUx = b \tag{3.205}$$

Defining

$$y = Ux \tag{3.206}$$

one solves first

$$Ly = b \tag{3.207}$$

then

$$Ux = y \tag{3.208}$$

The solution of the linear system thus amounts to the solution of 2 triangular system. For the lower triangular system, a forward substitution algorithm employed. For the upper triangular, a backward substitution algorithm is employed. The algorithmic complexity of these two substitution is linear with the size of the linear system, making them very efficient from a computational point of view.

The main bottleneck of direct solvers is the factorization step. For large scale matrices, involving up to billions of entries, the memory requirement to store the factors, even if the matrix A is sparse, is a challenge. One has to keep in mind that, despite the sparsity of A, its LU factorization yields dense triangular matrices L and U, which is the main inconvenient of Gaussian elimination.

Time-harmonic implementation of FWI have thus rely on the use of direct solvers, up to moderate scale targets (modeling with few tens up to one hundred wavelengths in each direction in 3D). In this frame, it is possible to take into account all the sources, thanks to the good computational efficiency of the solving step once the matrix is factorized.

To end up with this time-harmonic/frequency-domain implementation, one could say that the decimation in frequency of the data is compensated by the ability to consider all the seismic sources at the same time thanks to the LU decomposition. The redundancy

of the data makes it possible to design imaging techniques with such frequency decimation.

In terms of parallel implementation, the scalability of the corresponding algorithms is limited to the scalability of the direct solvers, which is known to be not extremely satisfactory. For typical applications, the scalability is observed up to few hundred of processors.

3.6.2 Time-domain implementations

Time domain implementation is more involved than frequency-domain implementation, however it yields better scalable algorithm, less memory demanding, and therefore makes it possible to work on large scale targets. This is now the method of choice developed both in industry and academy.

Gradient implementation

We first review how to build the gradient in practice, using explicit time-stepping schemes, following the adjoint state strategy. We will again take the acoustic case as a reference to illustrate the problems and solutions.

The gradient formula in the time domain is given by

$$\nabla f(m) = -\frac{2}{V_P^3}\int_0^T \frac{\partial^2 u(x,t)}{\partial t^2}\lambda(x,t)dt \tag{3.209}$$

Let us introduce v as the second-order derivative in time of the pressure wavefield

$$v(x,t) = \frac{\partial^2 u(x,t)}{\partial t^2} \tag{3.210}$$

Therefore

$$\nabla f(m) = -\frac{2}{V_P^3}\int_0^T v(x,t)\lambda(x,t)dt \tag{3.211}$$

After time discretization, the gradient formula becomes

$$\nabla f(m) = -\frac{2}{V_P^3}\sum_{n=1}^{N} v^n(x)\lambda^n(x)\Delta t \tag{3.212}$$

where N is the total number of time steps.

The main difficulty comes from the fact that to compute the summation in time (3.212), we need to access the wavefields $v(x,t)$ and $\lambda(x,t)$ at the same time indices simultaneously. However, $v(x,t)$ is computed from an initial condition, using an explicit time-stepping scheme, such that we compute iteratively

$$v^0(x),\ \ v^1(x),\ \ v^2(x),\ \ \ldots\ \ v^N(x) \tag{3.213}$$

On the other hand, $\lambda(x,t)$ is backpropagated from a final condition, using also an explicit time-stepping scheme, such that we compute iteratively

$$\lambda^N(x),\ \ \lambda^{N-1}(x),\ \ \lambda^{N-2}(x),\ \ \ldots\ \ \lambda^0(x) \tag{3.214}$$

This is illustrated in Figure 3.7.

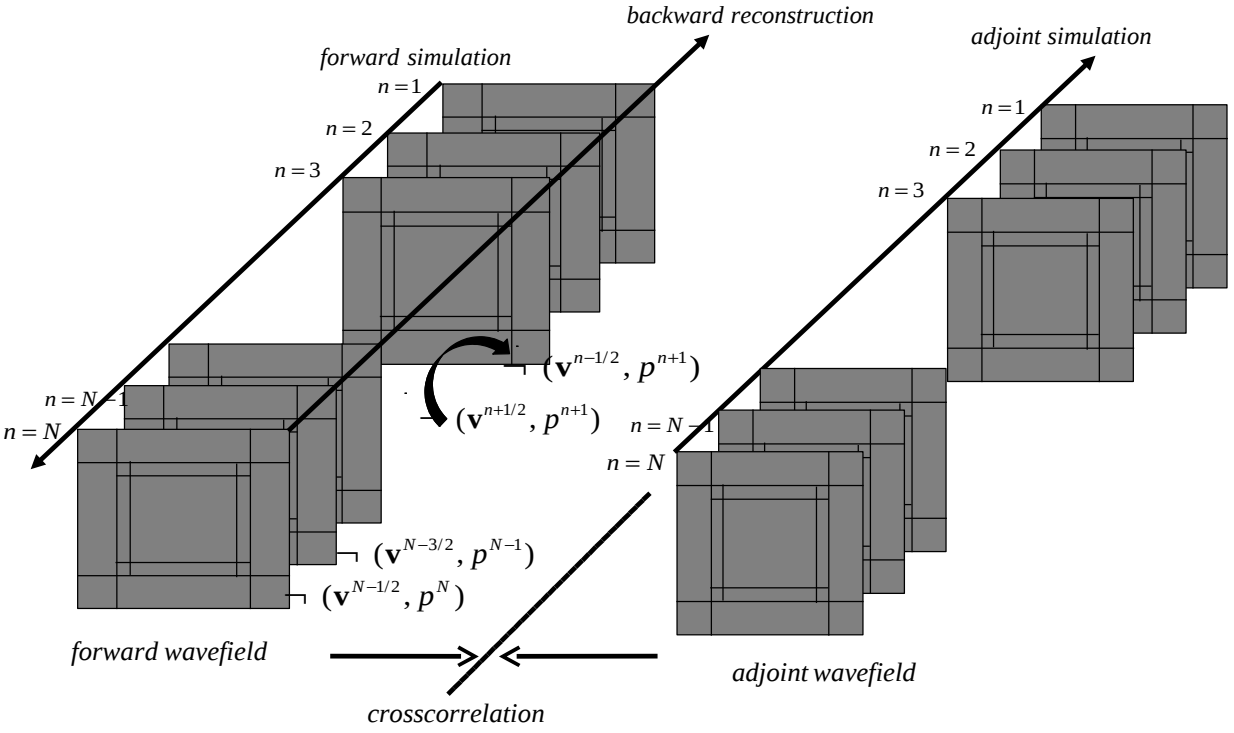

Figure 3.7: Illustration of the gradient building in the time domain following the adjoint state strategy. Picture taken courtesy of Pengliang Yang (Yang et al., 2016).

This raises questions/difficulties in terms of implementation. One obvious solution to the problem is to store in time the wavefield $v(x,t)$. It yields the following algorithm.

ALGO1: storing technique

1. Compute $v(x,t)$ from its initial condition and store it at all time steps. We keep in memory $v^0(x),\ v^1(x),\ v^2(x),\ \ldots\ v^N(x)$
2. Simultaneously backpropagate $\lambda(x,t)$ and assemble the gradient.
   For $n = N$ to $n = 0$ do :
   compute $\lambda^n(x)$ in reverse sense
   access to $v^n(x)$ in the memory
   add the contribution $\lambda^n(x)v^n(x)$ to the gradient
   End for

However, this requires potentially a heavy memory load. The number of snapshots correspond to the total number of time steps, and each snapshot can be already significant to store in a 3D volume. Consider for instance a mesh of 500^3 points, and 4000 time steps. The total volume in octets to store in memory, in single precision (32 bits precision) is

$$500 \times 500 \times 500 \times 4000 \times 4 = 500 \times 10^9 \tag{3.215}$$

which makes 500 GO of RAM. For a reference, a single node of the local HPC machine Dahu in Grenoble is 32 cores with 192 GO of memory.

Alternative algorithms have thus been designed in the frame of time-domain FWI implementation, with the objective to reduce the memory requirement. In this frame, the state-of-the-art algorithm consists in backpropagating the incident field and the adjoint field at the same time. To do so, a preliminary computation of the incident field from the initial time to the final time is performed, and the incident field is stored at the final time in the whole domain, as well as its total time history, but only on the boundaries. A first publication on this strategy is due to Dussaud et al. (2008). The corresponding algorithm is sketched below.

ALGO2: incident field backpropagation technique

1. Compute $v(x,t)$ in forward sense from its initial condition and store it at all time steps **only on the boundaries of the domain**.
   We keep in memory $v^0(x),\ v^1(x),\ v^2(x),\ \ldots\ v^{N-1}(x)$ **for** $x \in \partial\Omega$ and $v^N(x)$ **for** $x \in \Omega$

2. Simultaneously backpropagate $\lambda(x,t), v(x,t)$ and assemble the gradient.
   For $n = N$ to $n = 0$ do :

   compute $\lambda^n(x)$ in reverse sense

   recompute $v^n(x)$ in reverse sense

   add the contribution $\lambda^n(x)v^n(x)$ to the gradient

   End for

Of course the computation cost of ALGO2 is more expensive than for ALGO1 as we have to recompute the incident field from its final value and its values stored at the boundaries. In total, we have to solve 1 forward propagation problem in step 1, and 2 forward propagation problems in step 2, which leads to the solution of 3 wave propagation problems. In ALGO1, we do not have to recompute the wavefield, therefore, we solve only for 2 wave propagation problems.

On the other hand, ALGO2 is much less memory demanding. If we take the previous example, for a cubic domain of 500^3 grid points and 4000 time steps, we have to store

1. 1 final snapshot of the wavefield $v^N(x)$: $500^3 \times 4$ octets
2. the history of the wavefield on the boundaries : $6 \times 500^2 \times 4000 \times 4$ octets

The total memory request is thus

$$\left(500^3 + 6 \times 500^2 \times 4000\right) \times 4 = 24,5 \times 10^9 \tag{3.216}$$

The memory demand is thus reduced to 24.5 GO in this case.

ALGO2 is the one chosen in practice, unless very large memory machines are available, or, as is be discussed later, we have access to very large number of cores, and domain decomposition can be used in conjunction with parallelism with sources, to reduce the memory request per computational unit. This is developed in the next section on parallelization.

Before moving to this point, we need to discuss the specific case, yet very important, of attenuating media. In some applications, we use a visco-acoustic or visco-elastic modeling of the wave propagation within the subsurface. This is to represent the attenuation of the waves propagating in the subsurface in specific geologies. Earth is in practice an attenuating media (part of the energy is dissipated along the propagation, due to multiple phenomena such as thermal dissipation, cracks, anelastic behavior of the subsurface). This can be significant in fluid-filled rocks, where a network of fractures is filled with water, or gas for instance. In this case, the imprint of attenuation is important and need to be taken into account to reproduce with sufficient fidelity the observed data.

In terms of modeling, there are multiple ways to account for attenuation in the partial differential equations representing the wave propagation. The most commonly used one is Maxwell Generalized Body, or Standard Linear Solid model (these two are the same, only with different names). Mathematically, ordinary differential equations are added to the PDE's coming from elastodynamics.

In terms of numerical implementation, this has a strong consequence for the gradient building step. The backpropagation of the incident wavefield in an attenuating media is **numerically instable**. This can be understood as, in reverse computation, at each time step, the field has to be amplified rather than dissipated, and the consequence is the numerical instability. Exponential increase at each time step cannot be done with sufficient numerical accuracy, yielding the numerical error to blow up in few number of time steps. This makes ALGO2 unavailable in the frame of visco-acoustic or visco-elastic modeling.

To face this difficulty, but yet devise lighter memory technique than ALGO1, checkpointing strategies have been designed. The essence of checkpointing strategy is to do a preliminary computation of the incident wavefield in forward sense, as in ALGO2, and to store snapshot of the incident wavefield at specific time steps. For optimality, these can be distributed following a logarithmic law. Then, the adjoint field is backpropagated from its final condition, while the incident wavefield is computed in forward sense **from the closest snapshot available**. This reduces the additional computational steps, while mitigating the total memory requirement. The checkpointing algorithm is sketched in ALGO3.

ALGO3: checkpointing technique

1. Compute $v(x,t)$ from its initial condition and store it at $\log(N)$ distributed time steps.
   We keep in memory $v^{k_j}(x),\ \ j = 1\ldots,\log(N)$ **for** $x \in \Omega$.

2. Backpropagate $\lambda(x,t)$ and recompute $v(x,t)$ from the closest snapshot to assemble the gradient.
   For $n = N$ to $n = 0$ do :

   compute $\lambda^n(x)$ in reverse sense

   recompute $v^n(x)$ in forward sense from last checkpoint

   add the contribution $\lambda^n(x)v^n(x)$ to the gradient

   End for

We see that the memory demand in this case is $\log_2(N)$ snapshots of the wavefield, instead of N snapshots if the "store all" ALGO1 is used. However, in terms of computational cost, the recomputation from the closes checkpoint requires a significant number of additional computing steps. In the previously introduced example, the memory imprint of the checkpointing strategy would be

$$500^3 \times \log_2(4000) \times 4 \simeq 1.4957 \times 10^9 \tag{3.217}$$

The memory demand is thus reduced to approximately 1.5 GO in this case, which is even smaller than the revere computation strategy of ALGO2. However, here the recomputation effort is more significant.

Finally, let us mention another development proposed in Yang et al. (2016), as an improvement of conventional checkpointing techniques for gradient computation in attenuating media. The leading idea is to control the instability in the reverse propagation of the incident wavefield by measuring the total energy of the wavefield $E(t)$ defined simply by

$$E(t) = \int_\Omega \|v(x,t)\|^2 dx. \tag{3.218}$$

When performing the reverse propagation of the wavefield in the attenuating media, the numerical instability is detected as soon as the energy of the wavefield departs too far away from the one which has been recorded when computing the in the forward sense. In this case, the incident wavefield is recomputed from the closest snapshot, as in conventional checkpointing strategy. Once the stability is restored, the recomputation of the incident wavefield in backward sense can be used again, until a new instability is detected or the initial time step has been reached. This strategy makes possible to perform some backward computation of the incident wavefield, even in an attenuating media, thus reducing the number of time steps at which the incident needs to be recomputed, compared with conventional checkpointing strategies. It uses the same number of snapshots of the wavefields in memory, therefore the memory request is the same as for conventional checkpointing. This strategy is sketched in ALGO4.

ALGO4: improved checkpointing technique

1. Compute $v(x,t)$ from its initial condition and store it at $\log(N)$ distributed time steps.
   We keep in memory $v^{k_j}(x)$, $j = 1 \ldots, \log(N)$ **for** $x \in \Omega$.
   We also compute $E(t)$ and keep it in memory as $E_{ref}(t)$.

2. Backpropagate $\lambda(x,t)$ and $v(x,t)$ to assemble the gradient. If an numerical instability is detected, recompute $v(x,t)$ from the closest snapshot.
   For $n = N$ to $n = 0$ do :

   compute $\lambda^n(x)$ in reverse sense

   recompute $v^n(x)$ in reverse sense.

   If $|E(t) - E_{ref}(t)| > tol$ then recompute $v^n(x)$ from closest snapshot.

   add the contribution $\lambda^n(x)v^n(x)$ to the gradient

   End for

A comparison of the different proposed algorithms is presented in Figure 3.8. ALGO2 has been introduced as reverse propagation (RP) technique. ALGO4 has been introduced in the literature as checkpointing assisted reverse-forward simulation (CARFS).

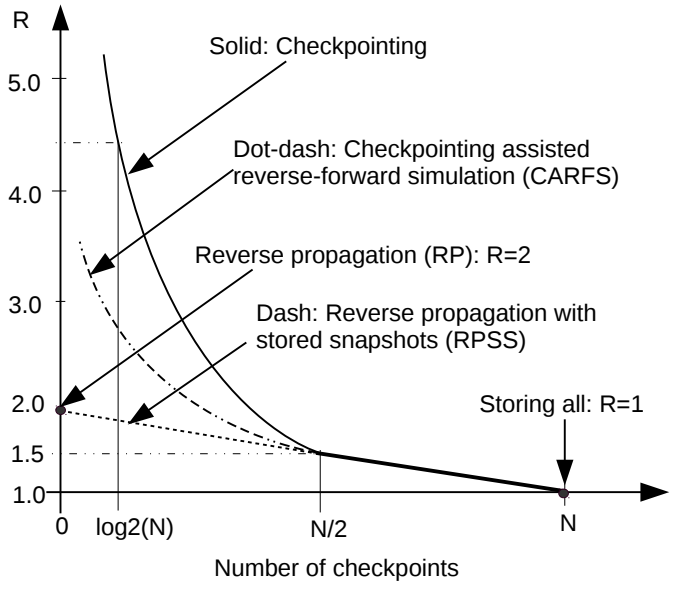

Figure 3.8: Comparison of the different algorithm in terms of memory imprint and recomputation ratio R. R is equal to 1 in the case where the incident field is stored (ALGO1). In the frame of reverse propagation (when it is stable, in non-attenuating media), R is equal to 2: the incident wavefield is computed twice, or more precisely the number of time-stepping iterations is multiplied by 2. For checkpointing algorithm (ALGO3), one can see that the recomputation ratio rapidly increases as soon as the number of stored snapshots is less than $\frac{N}{2}$ (a common situation). Finally, we see that the recomputation ratio R associated with CARFS (ALGO4) is much lower than that of checkpointing strategy for the same number of stored snapshots.

Parallelization techniques

To make FWI algorithms able to benefit from modern computational architectures, it is necessary to make them able to exploit parallelism on multi-core computers. While Moore's law has been enough to guarantee computational efficiency of sequential algorithms at the end of the XXth century, today computational efficiency is dependent on the scalability of algorithms, that is their ability to run on multi-core architectures efficiently.

For FWI, there is a first obvious level of parallelism which can be employed for time-domain algorithm. This is the parallelism on sources. Indeed, the misfit function $f(m)$ can be rewritten as a summation over sources of sub misfit function depending only on one specific seismic source, that is

$$f(m) = \sum_{s=1}^{N_s} f_s(m), \tag{3.219}$$

with

$$f_s(m) = \frac{1}{2} \sum_{r=1}^{N_r} \int_0^T |d_{cal_s}[m](x_r, t) - d_{obs,s}(x_r, t)|^2 dt. \tag{3.220}$$

In this frame, we also have

$$\nabla f(m) = \sum_{s=1}^{S} \nabla f_s(m) \tag{3.221}$$

with

$$\nabla f_s(m) = \left(\frac{\partial A}{\partial m} u_s[m], \lambda_s[m] \right), \tag{3.222}$$

or, specifically in the acoustic case

$$\nabla f_s(m) = \frac{-2}{V_P^3} \int_0^T \frac{\partial^2 u_s}{\partial t^2} \lambda_s(t) dt. \tag{3.223}$$

This means that, for assembling the gradient on a parallel computer, one can distribute the computation of the contribution associated with each source s, namely $\nabla f_s(m)$, and gathers the result as a simple summation of these contributions. Each contribution can be computed independently, and simultaneously, by different computing units. This is sketched in the algorithm below.

Gradient computation: parallelism over sources

1. For $s = 1$ to $s = N_s$ do **in parallel**

   compute $\nabla f_s(m)$ following ALGOx (x=1,2,3 or 4 depending on user choice)

   end do

2. Gather the contribution of each processor on a master processor
   $\nabla f(m) = \sum_{s=1}^{N_s} \nabla f_s(m)$

For practical applications where the number of sources can reach several thousands, this level of parallelism is very useful. Each time-domain FWI algorithm embeds this parallelism level. However, for further efficiency and to decrease the computational time, an

additional level of parallelism is very often implemented, at the level of the incident and adjoint wavefield computations.

When explicit time-stepping schemes are employed, it is relatively easy to parallelize the computation of the time-domain wavefields using **domain decomposition methods**. We will not enter into the details of these methods here, but we can mention the main properties/ideas of this strategy. The main idea is to decompose the computation domain Ω in subdomains Ω_i such that

$$\Omega = \cup_{i=1}^{N_d} \Omega_i \tag{3.224}$$

The solution of the wave propagation problem is computed at each time step in parallel on each subdomain Ω_i. Once the computation associated with the time iteration has been performed, the domains exchange information on their boundaries with the neighboring domains, so that they have access to the required information at the next time step to compute the spatial derivatives at their boundaries.

Compared with previous parallelization level over sources, the implementation of such domain decomposition methods is slightly more involved. However, in the frame of finite difference techniques with simple stencil (order 2 or order 4), the communication between domains is relatively simple. In the frame of spectral element methods, has discretization points on the border of the elements are used, the communication between domains is also relatively simple. Using a Cartesian based mesh makes things easy to locate the neighboring domains to communicate with. If not, a lookup table has to be built as a pre-processing step to register who to communicate with before entering the time-loop. A sketch of the domain decomposition algorithm for explicit time stepping schemes is presented in the algorithm below.

Wavefield computation: parallelism over subdomains

1. For $n = 1$ to $n = N$ do (time loop)

   For $i = 1$ to $i = N_d$ do **in parallel on each subdomain**

   compute $u_i^n(x), \;\; x \in \Omega_i$

   end do

   exchange information with neighboring domains

   end do

With these two levels of parallelism combined (sources + domain decomposition), we can develop time-domain FWI algorithms scalable up to tens of thousands of computing units (cores). In practice, we might not access easily to machines with so many available computing units, therefore for practical applications, another algorithm is used on top of that to reduce the overall computational cost of FWI. An illustration of such double level of parallelization is given in Figure 3.6.2.

Random source subsampling

We mentioned in the paragraph on frequency-domain/time-harmonic FWI implementation that the use of direct solvers makes possible to use the whole seismic sources at a time for a given inversion. This use of massively redundant data compensates for the severe decimation which consists in considering single or few monochromatic fields instead

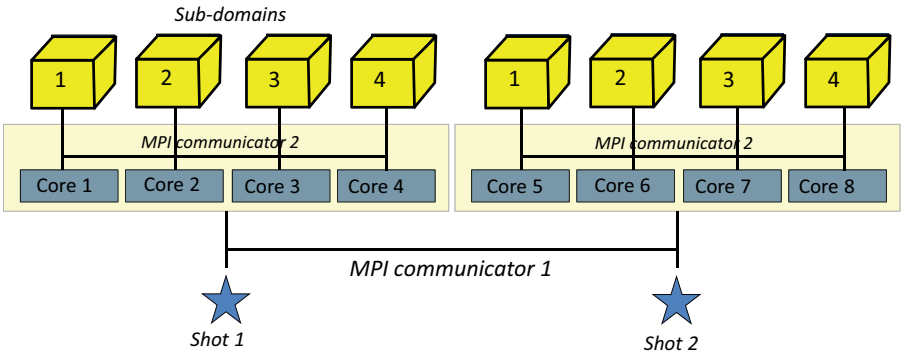

Figure 3.9: Illustration of the double level of parallelism for time-domain implementation of FWI algorithms.

of time-dependent wavefields.

Conversely for time-domain FWI, we do not consider any decimation in time, and therefore it is reasonable to decimate the data on the source side. In order not to introduce to strong artifacts in the reconstructed models yet, this needs to be done in a "reasonable" manner. One conventionally used is the choice of a random source subsampling algorithm.

The idea is the following: instead of taking into account the total number of seismic sources at the same time, we proceed to the inversion of small subsets of the datasets. The source are randomly chosen in each subset. The idea is to use non overlapping datasets, such that the pool of available sources always decrease and that each source is used at least once in the inversion.

3.7 Practical applications

See slides of SESSION 12.

Bibliography

Aki, K. and Richards, P. G. (2002). *Quantitative seismology, theory and methods, second edition*. University Science Books, Sausalito,California.

Amestoy, P., Duff, I. S., and L'Excellent, J. Y. (2000). Multifrontal parallel distributed symmetric and unsymmetric solvers. *Computer Methods in Applied Mechanics and Engineering*, 184(2-4):501–520.

Bensoussan, A., Lions, J., and Papanicolaou, G. (1978). *Asymptotic analysis of periodic structure*. North-Holland.

Bérenger, J.-P. (1994). A perfectly matched layer for absorption of electromagnetic waves. *Journal of Computational Physics*, 114:185–200.

Bernardi, C. and Maday, Y. (1992). *Approximations spectrales de problèmes aux limites elliptiques*, volume 10. Springer-Verlag Berlin Heidelberg.

Capdeville, Y., Cupillard, P., and Singh, S. (2020). An introduction to the two-scale homogenization method for seismology. In *Advances in Geophysics*, volume 61, pages 217–306. Elsevier.

Cerjan, C., Kosloff, D., Kosloff, R., and Reshef, M. (1985). A nonreflecting boundary condition for discrete acoustic and elastic wave equations. *Geophysics*, 50(4):2117–2131.

Červený, V. (2001). *Seismic Ray Theory*. Cambridge University Press, Cambridge.

Chavent, G. (1974). Identification of function parameters in partial differential equations. In Goodson, R. and Polis, M., editors, *Identification of parameter distributed systems*, pages 31–48. American Society of Mechanical Engineers, New York.

Collino, F. (1993). High order absorbing boundary conditions for wave propagation models: Straight line boundary and corner cases. In *Proceedings of the Second International Conference on Mathematical and Numerical Aspects of Wave Propagation*, pages 161–171. SIAM, Delaware.

Collino, F. and Tsogka, C. (2001). Application of the perfectly matched absorbing layer model to the linear elastodynamic problem in anisotropic heterogeneous media. *Geophysics*, 66:294–307.

Duff, I. S., Erisman, A. M., and Reid, J. K. (1986). *Direct methods for sparse matrices, second edition*. Oxford Science Publications, Oxford, U. K.

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.

Eiken, O. (2019). Twenty years of monitoring CO2 injection at Sleipner. In Davis, T. L., Landrø, M., and Wilson, M., editors, *Geophysics and Geosequestration*, pages 209–234. Cambridge University Press.

Ernst, O. G. and Gander, M. J. (2012). Why it is Difficult to Solve Helmholtz Problems with Classical Iterative Methods. In Graham, I. G., Hou, T. Y., Lakkis, O., and Scheichl, R., editors, *Numerical Analysis of Multiscale Problems*, volume 83 of *Lecture Notes in Computational Science and Engineering*, pages 325–363. Springer Berlin Heidelberg.

Gorszczyk, A., Operto, S., Schenini, L., and Yamada, Y. (2019). Crustal-scale depth imaging via joint FWI of OBS data and PSDM of MCS data: a case study from the eastern nankai trough. *Solid Earth*, 10:765–784.

Gould, P. L. (2013). *Introduction to Linear Elasticity*. Springer New York, NY, 3rd edition.

Grange, A. (2025). *Characterization of the seismicity and 3D structure of the Argentière glacier using a quantitative elastic inversion method.* PhD thesis, Université Grenoble Alpes.

Hestenes, M. R. and Stiefel, E. (1952). Methods of conjugate gradient for solving linear systems. *Journal of Research of the National Bureau of Standards*, 49:409–436.

Igel, H. (2017). *Computational seismology: a practical introduction*. Oxford University Press.

Köhn, D., Zolchow, M., Mecking, R., Wilken, D., Wunderlich, T., De Nil, D., and Rabbel, W. (2021). Seismic SH Full Waveform Inversion: A tool for high-resolution near-surface characterization. *Near Surface Geoscience Conference & Exhibition 2021; At: Bordeaux, France*, 2021(1):1–5.

Lailly, P. (1983). The seismic inverse problem as a sequence of before stack migrations. In Bednar, R. and Weglein, editors, *Conference on Inverse Scattering, Theory and application, Society for Industrial and Applied Mathematics, Philadelphia*, pages 206–220.

Landau, L. and Lifchitz, E. (1967). *Physique théorique, tome 7, Théorie de l'élasticité*. MIR, Moscou, 2nd edition edition.

Lions, J. L. (1968). *Contrôle optimal de systèmes gouvernés par des équations aux dérivées partielles*. Dunod, Paris.

Love, A. E. H. (1911). Theory of the propagation of seismic waves. *Some problemes of Geodynamics (in Geophysics reprint series (vol 24): Classics of Elastic Wave Theory)*, pages 144–179.

Lowrie, W. and Ficthner, A. (2020). *Fundamentals of Geophysics*. Cambridge University Press, 3 edition.

MUMPS-team (2019). *MUMPS - MUltifrontal Massively Parallel Solver (MUMPS 5.2.0-betapre1) Users' guide (February, 2019)*. ENSEEIHT-ENS Lyon, http://www.enseeiht.fr/apo/MUMPS/ or http://graal.ens-lyon.fr/MUMPS.

Nakata, N., Gualtieri, L., and Fichtner, A. (2019). *Seismic Ambient Noise*. Cambridge University Press.

Nocedal, J. (1980). Updating Quasi-Newton Matrices With Limited Storage. *Mathematics of Computation*, 35(151):773–782.

Nocedal, J. and Wright, S. J. (2006). *Numerical Optimization*. Springer, 2nd edition.

Nolet, G. (2008). *A Breviary of Seismic Tomography*. Cambridge University Press, Cambridge, UK.

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.

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

Pilant, W. L. (1979). *Elastic Waves in the Earth*, volume 11 of *Developments in solid earth geophysics*. Elsevier.

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.

Rayleigh, L. (1887). On waves propagated along the plane surface of an elastic solid. *Proceedings of the London Mathematical Society (in Geophysics reprint series (vol 24): Classics of Elastic Wave Theory)*, 17:4–11.

Saad, Y. (1986). GMRES: a generalized minimal residual algorithm for solving nonsymmetric linear systems. *SIAM Journal on Scientific and Statistical Computing*, 7(3):856–869.

Shearer, P. M. (2009). *Introduction to seismology*. Cambridge University Press, Cambridge, UK.

Sirgue, L. (2003). *Inversion de la forme d'onde dans le domaine fréquentiel de données sismiques grand offset*. PhD thesis, Université Paris 11, France - Queen's University, Canada.

Stein, S. and Wysession, M. (2003). *An Introduction to Seismology, Earthquakes and Earth Structure*. Blackwell Publishing.

Tarantola, A. (1984). Inversion of seismic reflection data in the acoustic approximation. *Geophysics*, 49(8):1259–1266.

Tikhonov, A., Goncharsky, A., Stepanov, V., and Yagola, A. (2013). *Numerical methods for the solution of ill-posed problems*. Springer Science & Business Media.

Tsvankin, I. (2005). Seismic signature and analysis of reflection data in anisotropic media. In Helbig, K. and Treitel, S., editors, *Seismic Exploration*, volume 29 of *Handbook of Geophysical Exploration*. Elsevier, 2 edition.

Virieux, J. (1986). P-SV wave propagation in heterogeneous media: Velocity-stress finite difference method. *Geophysics*, 51:889–901.

Wu, R. S. and Toksöz, M. N. (1987). Diffraction tomography and multisource holography applied to seismic imaging. *Geophysics*, 52:11–25.

Yang, P., Brossier, R., Métivier, L., and Virieux, J. (2016). Wavefield reconstruction in attenuating media: A checkpointing-assisted reverse-forward simulation method. *Geophysics*, 81(6):R349–R362.

Yee, K. S. (1966). Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media. *IEEE Transactions on Antennas and Propagation*, 14:302–307.