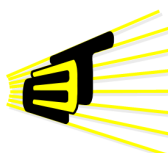


UNCERTAINTY ESTIMATION OF 2D SUBSURFACE'S VELOCITY MODELS OBTAINED BY THE FULL-WAVEFORM INVERSION IN THE TIME DOMAIN

Juan Sebastian López Guerra



Escuela de Ingenierías
Eléctrica, Electrónica
y de Telecomunicaciones



Universidad Industrial de Santander
Facultad de Ingenierías Físico-Mecánicas
Escuela de Ingeniería Eléctrica, Electrónica y de Telecomunicaciones
Maestría en Ingeniería Electrónica
Bucaramanga
2017

**UNCERTAINTY ESTIMATION OF 2D SUBSURFACE'S VELOCITY
MODELS OBTAINED BY THE FULL-WAVEFORM INVERSION IN THE
TIME DOMAIN**

Juan Sebastian López Guerra

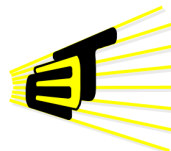
Trabajo de investigación presentado como requerimiento parcial para optar por el título de
Magister en Ingeniería Electrónica

Director:

PhD(c). Sergio Alberto Abreo Carrillo

Co-Director:

PhD. Ana Beatriz Ramírez Silva



Escuela de Ingenierías
Eléctrica, Electrónica
y de Telecomunicaciones



Universidad Industrial de Santander
Facultad de Ingenierías Físico-Mecánicas
Escuela de Ingeniería Eléctrica, Electrónica y de Telecomunicaciones
Maestría en Ingeniería Electrónica
Bucaramanga
2017

Contents

	Pag.
INTRODUCTION	13
1 Exploratory Geophysics	16
1.1 The Earth's Subsurface	16
1.2 The Seismic Experiment	17
1.3 Synthetic Velocity Models	18
2 Full Waveform Inversion	22
2.1 The Forward Problem	22
2.1.1 The Wave Equation	23
2.1.2 The Finite Difference Time Domain Method	24
2.1.3 Wave Propagation Through the SVM's	34
2.2 The Inverse Problem	37
2.2.1 Computing $\alpha_k \Delta \mathbf{m}$: the L-BFGS method	38
2.2.2 Solutions to the Inverse Problem of the SVM's	41
2.3 The Adjoint-State Method	48
2.3.1 First Order ASM: The Gradient	48
2.3.2 Second Order ASM: The Hessian	55
2.4 Discussion	63
2.5 Conclusions	66

3	Uncertainty Quantification	67
3.1	Relationship between the Hessian and the covariance matrices	68
3.2	The <i>a posteriori</i> covariance matrix	70
3.2.1	Estimating posterior uncertainties	70
3.2.2	Results of the SVM's	71
3.3	Uncertainty Appraisal	75
3.3.1	Uncertainty Bars	75
3.3.2	UQ Factor	79
3.3.3	Resolution Operator	81
3.4	Discussion	84
3.5	Conclusions	87
3.6	Future Work	88
	References	90
	Bibliography	94

List of Figures

	Pag.
1 Two images of the layers of the Earth	16
2 On-shore Seismic Survey process [1].	18
3 Seismic Trace [2].	18
4 Square Diffractor Model.	19
5 Marmousi original model [3].	20
6 Sigsbee 2A original model [4].	21
7 Vector $f(x)$	28
8 Values of a_x (blue) y b_x (red), with $Nx=100$, $CPML_{limit}=20$, $dt=4e-3$	28
9 Values of d_x	29
10 Numerical Dispersion of 1% for the square diffractor model for which $k_x\Delta x = 1.25$	33
11 Wave propagation over the square diffractor model (before diffractor).	34
12 Wave propagation over the square diffractor model (after diffractor.	35
13 Shot Gather of the square diffractor model.	36
14 Wave propagation over the Marmousi model.	36
15 Wave propagation over the Sigsbee 2A model.	37
16 Schematic of the local-minima problem in FWI. Tarantola [5].	39
17 Objective function $\Phi(\mathbf{m})$ for the square diffractor model at a frequency of 3Hz. After 40 iterations, the frequency of the source was increased for the next 40 iterations.	41
18 Initial model for the square Diffractor Model.	43
19 FWI solution of the square Diffractor Model.	44
20 Initial model for the Marmousi model.	45
21 FWI solution of the Marmousi model.	45
22 Initial model for the Sigsbee 2A model.	46

23	FWI solution of the Sigsbee 2A model.	47
24	Gradient for the first iteration of the square Diffractor Model at 3Hz.	52
25	Gradient for the first iteration of the Marmousi model at 3Hz.	53
26	Gradient for the first iteration of the Marmousi model at 9Hz.	54
27	Gradient for the first iteration of the Sigsbee 2A model at 3Hz.	54
28	Hessian matrix of the square Diffractor Model.	58
29	A column of the Hessian matrix (reshaped as matrix) of the square Diffractor Model. . .	60
30	Diagonal of the Hessian matrix (reshaped as matrix) of the square Diffractor Model. . . .	61
31	Diagonal of the Hessian matrix (reshaped as matrix) of the Marmousi Model at 3Hz. . . .	61
32	Diagonal of the Hessian matrix (reshaped as matrix) of the Marmousi Model at 9Hz. . . .	62
33	Diagonal of the Hessian matrix (reshaped as matrix) of the Sigsbee Model at 6Hz.	63
34	Posterior variance matrix of the square diffractor model.	72
35	Posterior variance matrix of the Marmousi model.	73
36	Posterior variance matrix of the Sigsbee model.	74
37	Negative values in the posterior variance matrix.	75
38	Example distributions for a same pixel before (left) and after (right) FWI. The distributions share the same μ approximately but the standard deviation is reduced from the prior to the posterior distribution.	76
39	Prior and posterior distributions for the central pixel of the diffractor model.	77
40	Uncertainty profile for the diffractor model at a location of $x = 2.625\text{km}$	78
41	Uncertainty profile for the Marmousi model at a location of $x = 5.72\text{km}$	78
42	Uncertainty Reduction Factor (UQ Factor) of the square diffractor model.	80
43	Uncertainty Reduction Factor (UQ Factor) of the Marmousi model.	80
44	Uncertainty Reduction Factor (UQ Factor) of the Sigsbee model.	81
45	Resolution operator of the square diffractor model at 12 Hz.	82
46	Resolution operator of the Marmousi model at 9 Hz.	83
47	Resolution operator of the Sigsbee model at 6 Hz.	83

List of Tables

	Pag.
1 Results for approximately 1% of numerical dispersion in the three models.	33
2 Source locations in the SVM's.	34
3 Frequencies used for the multi-scale FWI for each model.	42
4 Parameters used for each model's inversion.	42
5 Dimensions of the Hessian matrices.	62
6 Resources used by the GPU when computing Hessian matrices.	65
7 Conditioning standard deviation (σ_{cond}) for the three SVM with a source frequency of 3Hz.	71
8 Time spent computing $\mathbf{H}$ and $\tilde{\mathbf{S}}$	87

Resumen del trabajo

Título:	UNCERTAINTY ESTIMATION OF 2D SUBSURFACE'S VELOCITY MODELS OBTAINED BY THE FULL-WAVEFORM INVERSION IN THE TIME DOMAIN
Director:	PhD(c).Sergio Alberto Abreo Carrillo
Co-Director:	PhD. Ana Beatriz Ramírez Silva
Autor:	Juan Sebastian López Guerra
Tipo de Investigación:	Aplicada
Costo Total:	\$200.428.600
Plazo:	12 meses
Línea de aporte al desarrollo nacional:	Sistemas Embebidos e Investigaciones en Energía y Minería
Programa Nacional de Ciencia y Tecnología:	Electrónica, Telecomunicaciones e informática
Posibles Entidades Interesadas en los Resultados:	Escuela de Ingenierías, Electrónica y Telecomunicaciones- E^3T . Universidad Industrial de Santander- UIS. Instituto Colombiano del Petróleo ICP-ECOPETROL. Ministerio de Minas y Energía. Centros de Investigación dedicados a generar desarrollo en plataformas de alto desempeño basadas en GPU.

Reseña

En el contexto de las técnicas de exploración sísmica, como la inversión de onda completa (FWI por sus siglas en inglés), estimar la incertidumbre asociada a los resultados obtenidos es de gran utilidad para que un intérprete pueda dar un mejor análisis de la imagen sísmica. Por esta razón, encontrar una medida de incertidumbre representa una ventaja competitiva al disponer de una herramienta que contribuya con la interpretación de imágenes sísmicas de alta calidad.

Se ha encontrado que bajo condiciones especiales, una posible medida de incertidumbre se puede asociar a la matriz Hessiana de la función objetivo de la FWI y debido a esto es que surge en este proyecto el interés por encontrarla, siendo los dos grandes retos: primero, el problema de costo computacional que va asociado a las grandes dimensiones de dicha matriz y segundo, su utilización e interpretación como medida de incertidumbre. Siendo este último reto el enfoque principal del proyecto.

RESUMEN

TÍTULO:

ESTIMACIÓN DE LA INCERTIDUMBRE DE MODELOS DE VELOCIDADES DEL SUBSUELO 2D
OBTENIDOS MEDIANTE LA INVERSIÓN DE ONDA COMPLETA EN EL DOMINIO DEL TIEMPO*

Autor: Juan Sebastián López Guerra **

Palabras Clave: Modelado Sísmico, Inversión sísmica, Cuantificación de la Incertidumbre.

DESCRIPCIÓN:

En el contexto de las técnicas de exploración sísmica, como la inversión de onda completa (FWI por sus siglas en inglés), estimar la incertidumbre asociada a los resultados obtenidos es de gran utilidad para que un intérprete pueda dar un mejor análisis de la imagen sísmica.

Por esta razón, encontrar una medida de incertidumbre representa una ventaja competitiva al disponer de una herramienta que contribuya con la interpretación de imágenes sísmicas de alta calidad.

Se ha encontrado que, bajo condiciones especiales, una posible medida de incertidumbre se puede asociar a la matriz Hessiana de la función objetivo de la FWI y debido a esto es que surge en este proyecto el interés por encontrarla, siendo los dos grandes retos: primero, el problema de costo computacional que va asociado a las grandes dimensiones de dicha matriz y segundo, su utilización e interpretación como medida de incertidumbre. Este último reto fue el enfoque principal del proyecto.

La metodología trabajada consistió en hacer una implementación de la ecuación de onda acústica, isotópica y de densidad constante en dos dimensiones mediante la técnica de diferencias finitas FDTD de octavo orden espacial y segundo orden temporal. Posteriormente, se utilizó el método del estado adjunto de primer y segundo orden para resolver el problema inverso y obtener la Hessiana, respectivamente. Con esta matriz Hessiana y su inversa es que se realizaron las técnicas de cuantificación de la incertidumbre.

* Tesis de Maestría

** Facultad de Ingenierías Físico Mecánicas. Escuela de Ingenierías Eléctrica, Electrónica y Telecomunicaciones. Director: PhD(c). Sergio A. Abreo Carrillo. Co-director: PhD. Ana Beatriz Ramírez Silva

ABSTRACT

TITLE:

UNCERTAINTY ESTIMATION OF 2D SUBSURFACE'S VELOCITY MODELS OBTAINED BY THE FULL-WAVEFORM INVERSION IN THE TIME DOMAIN*

AUTHOR: Juan Sebastián López Guerra **

KEYWORDS: Seismic Modeling, Seismic Inversion, Uncertainty Quantification.

DESCRIPTION:

In the seismic explorations framework, estimating uncertainties for techniques such as the full waveform inversion (FWI) turns out to be quite helpful for an interpreter to better analyze a seismic image.

For this reason, finding a measure of uncertainty represents a competitive advantage by having available a tool that contributes to high quality seismic image interpretations.

It has been found that, under certain conditions, a possible measure of uncertainty can be related to the Hessian matrix of the objective function of the FWI, and this is the main object of study in this research: to find this measurement. The two biggest challenges are: first, the computational cost, produced by the dimensions of such a matrix and second, its use and interpretation as a measure of uncertainty. Being the latter challenge the main focus of the research.

The work-flow consisted in making an implementation of the acoustic and isotropic wave equation with constant density in two dimensions using the FDTD method with eight order in the space dimensions and second order in the time dimension. After that, the first and second order adjoint state methods were used to solve the inverse problem and obtain the Hessian matrix, respectively. With this Hessian matrix and its inverse, the uncertainty quantification was performed.

* Master Thesis

** Faculty of Physics Mechanics Engineering. Electrical, Electronics Engineering and Telecommunications School. Advisor: PhD(c). Sergio A. Abreo Carrillo. Co-advisor: PhD. Ana Beatriz Ramírez Silva

Introduction

Background and motivation

Ever since the industrial revolution, fossil fuels have been used as the main source of energy to develop our civilization and improve the life quality of human beings. According to Lindtjorn et al (2011) *“Over the last 150 years, about 1 trillion barrels of oil equivalent have been consumed, and over the next couple of decades, as much oil will have to be found and produced”* [6]. Hydrocarbons are, therefore, a big deal in global economics given their non-renewability and the necessity of a continuous availability. The current need for fossil fuels is forcing mankind to look for new places to find them. However, some of these places have not been dug or even studied because of its natural resistance to being explored. Colombia is a country known for its immense wealth in mineral resources and for the great variety of terrains over its lithosphere and there is a high interest in the country to explore all its extension.

Recent efforts of the oil industry are turning towards the exploration of complex-zones, since not complex zones have already been explored and exploited. The seismic exploration of complex-zones require the use of modern methods in order to get better resolved images of the subsurface with high confidence about the location of the gas or oil reservoirs.

The first issue to overcome in finding highly-resolved images of complex-zones is the inversion of subsurface parameters. The seismic inversion of the 2-D velocity model can be done by using the 2-D Full Waveform Inversion (FWI). The resulting seismic velocity model is of major importance because a wrong velocity model leads to a wrong subsurface image. Additional to finding the velocity model, it is important to know how likely the velocity model fits the real data. For this reason, it is important to have a figure of merit to assess the results of FWI.

Nowadays the only appraisal that is performed is done by a geological interpreter who is bestowed with the task of deciding where to drill. He can benefit from this figure of merit in order to make a quantitative assessment instead of a qualitative one. This field has been growing in the last years because of the special attention of the scientific community. However, further research is necessary for this area if it is desired to obtain reliable information from complex geological structures.

Contributions of this work

According to the motivation above, the main focus of this project is to address the problem of estimating the uncertainty associated with a given velocity model obtained with the FWI process. This problem is an appraisal problem that will contribute to improve the interpretation of the images of the ground, hence, helping make better decisions.

The main objective is to estimate a degree of uncertainty of the seismic velocity model obtained with the 2-D Full Waveform Inversion (FWI) in the time domain. This degree is a figure of merit that provides information about the reliability of the information.

The importance of this figure of merit is not only the figure itself, but the discussion that can arise from it, and the analyses that can be performed. There are both geophysical and computational aspects to be discussed: the information that this figure provides, the conditions under which this figure of merit is valid and can be obtained, the cost of obtaining it and the limitations of it. All those aspects define the applicability and ultimately frame the uses that can be given to this quantitative assessment.

Outline of the book

This book is organized as follows:

In Chapter 1, the topic of exploratory geophysics is handled, a brief description of the Earth's subsurface and the seismic experiment is presented. In this chapter are also showcased the selection of synthetic velocity models (SVM) that are going to be used through all following chapters.

In Chapter 2, the Full Waveform Inversion (FWI) is addressed. Here the discussion focuses on the forward and inverse problems, necessary to solve the seismic inversion. It also includes the Adjoint-State Method (ASM), used to compute both the gradient and the Hessian matrices. Propagation and inversion results are presented alongside gradients and Hessian matrices of the selected SVM.

Chapter 3 presents the relationship between the Hessian and the posterior covariance matrices, and how to use the covariance matrix as an uncertainty estimator. At the end, the uncertainty reduction factor is presented to showcase how much is the a-priori uncertainty reduced. As expected, results of the covariance matrix and the uncertainty reduction are presented for the selected SVM. At the end, the final discussion of the results is presented to the reader.

Exploratory Geophysics

1.1 The Earth's Subsurface

The subsurface of the Earth is composed of several layers that can be defined by their chemical and their rheological properties. This structure can be seen in figure 1.



(a) Glacier National Park, Montana [7]



(b) Folds in sedimentary rocks, Australia [8].

Figure 1: Two images of the layers of the Earth

As seen in the figures above, the layers can have varied geometries depending on the physical conditions (such as pressure) they are exposed to. In figure 1a the layers are arranged in a quite plain geometry, while in figure 1b the layers are formed in a complex geometry. These two examples correspond to a sedimentary rock in the Glacier National Park in Montana [7] and a folded sedimentary rock in Australia [8] respectively.

1.2 The Seismic Experiment

Geophysics studies the Earth's properties such as its structure and lithology, based on physical principles. Those geophysical methods range from seismic methods to methods based on gravity, electricity or magnetism and their principal use nowadays is to obtain accurate images of the Earth's subsurface (sedimentary basins).

From the methods mentioned above, the seismic is the main working method [1]. It consists of deploying a seismic source to generate acoustic waves that propagate through the medium. If the seismic experiment is done on-shore (on land), those sources can be either vibrators or explosives. On the other hand, if the experiment is done off-shore (at sea) the source is a compressed-air gun.

Waves traveling through the medium reflect at interfaces between different geological layers in the subsurface and are recorded as seismic traces by an array of sensors on the surface. Again, for an on-shore experiment those sensors are geophones while off-shore experiments use hydrophones. The image of the subsurface is obtained from the seismic data recorded at the surface. This image can be two, three or four dimensional, depending on the kind of acquisition performed and the reconstruction method used.

Figure 2 depicts the seismic experiment performed on land, where:

1. Seismic source (seismic vibrator truck) emits energy.
2. Seismic energy propagates through the medium and reflects at each geological interface.
3. Reflected energy is recorded at surface by the sensors (geophones).
4. Data is collected for processing.

Data collected in form of seismic traces (figure 3) are processed after the survey has been done. With the traces and all *a priori* information (if any) of the zone, seismic images are created after the processing is done. Those images are interpreted by experts, they are the ones who assess the results.

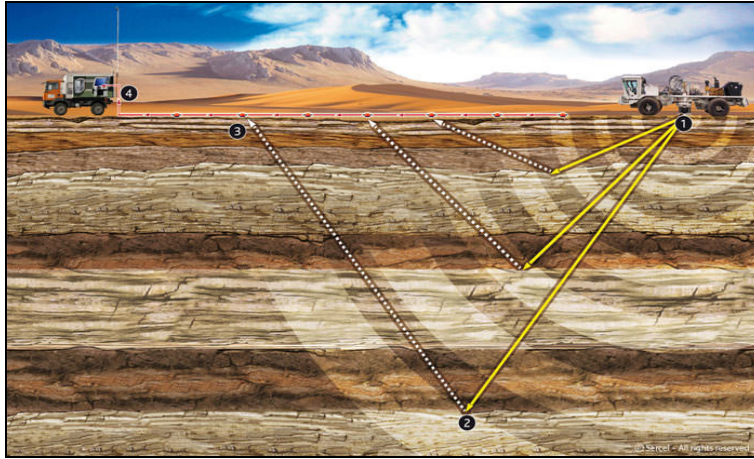


Figure 2: On-shore Seismic Survey process [1].

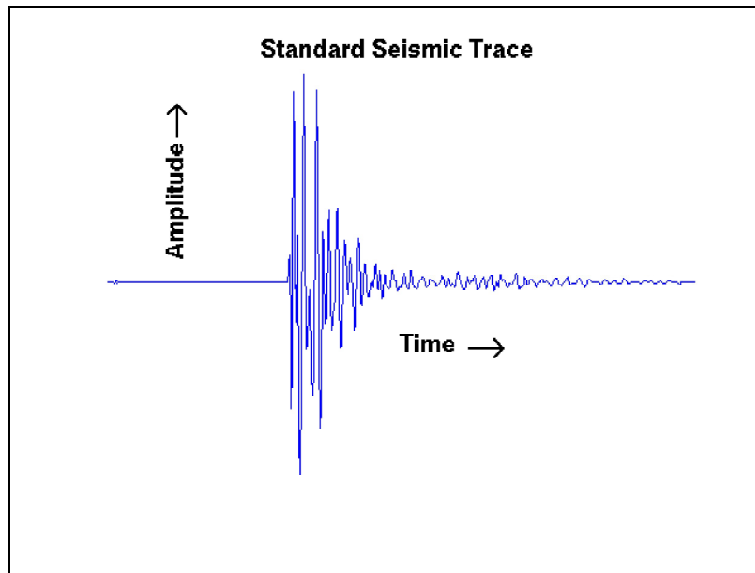


Figure 3: Seismic Trace [2].

1.3 Synthetic Velocity Models

With the intention of developing techniques and algorithms for exploring the subsurface of the Earth, Synthetic Velocity Models (SVM) were created. These allow the simulation of several of the properties in a controlled fashion. For this reason, and given the natural complexity of the Earth's subsurface, the SVM are used. Their complexity depends on the characteristics that one wants to test.

Below, there are three velocity models used throughout this work.

The first model (figure 4) corresponds to a design made by the CPS research group at UIS. Similar models have been used in works such as [9]. This model presents a background constant velocity with a high velocity intrusion in the middle of it. The selection for the velocities obeys to the interest of being able to find the location and value of the intrusion. This problem resembles the interest of finding oil and gas bodies. The sharp edges of the intrusion present an additional challenge as the corners and edges have an important effect on the wave propagations.

The specifications for this model are as follows: a model of 5 km of extension with a depth of 1.7 km distributed in a grid of 211 by 68 points respectively (uniform grid for both dimensions of 25 m) with a background velocity of 2000 m/s and a diffraction square whose center is located at the position 2.625 km with depth of 0.825km and continues four points in all four directions with a velocity of 2500 m/s.

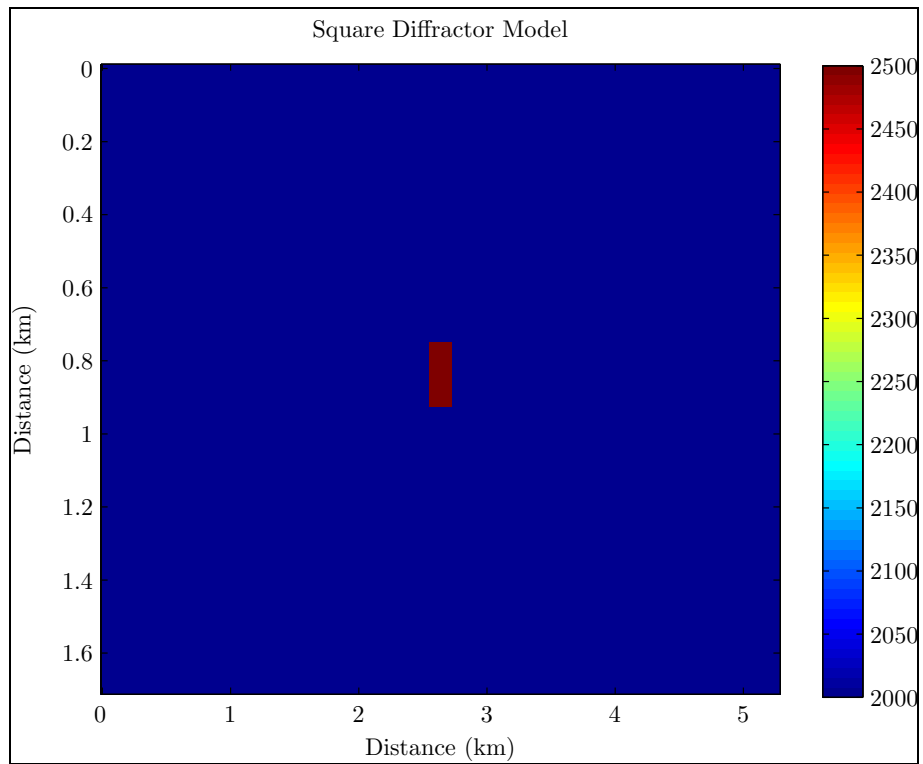


Figure 4: Square Diffractor Model.

The second model (figure 5) is a quite renowned one. The Marmousi model (Institut Français du Pétrole, 1988) is based on the geology of the Cuanza basin in Luanda, Angola [3]. The main challenges of this model are the variety of faults and the geometry of the layers itself, also, at the lowest part of it, a high velocity layer adds up to increase its complexity.

This model has a extension of 12 km with a depth of 3 km distributed in a equally spaced grid of 497 by 121 points with 25 m of separation between points for both dimensions. The velocities range from 1650 to 6000 m/s for the variety of layers that it possess.

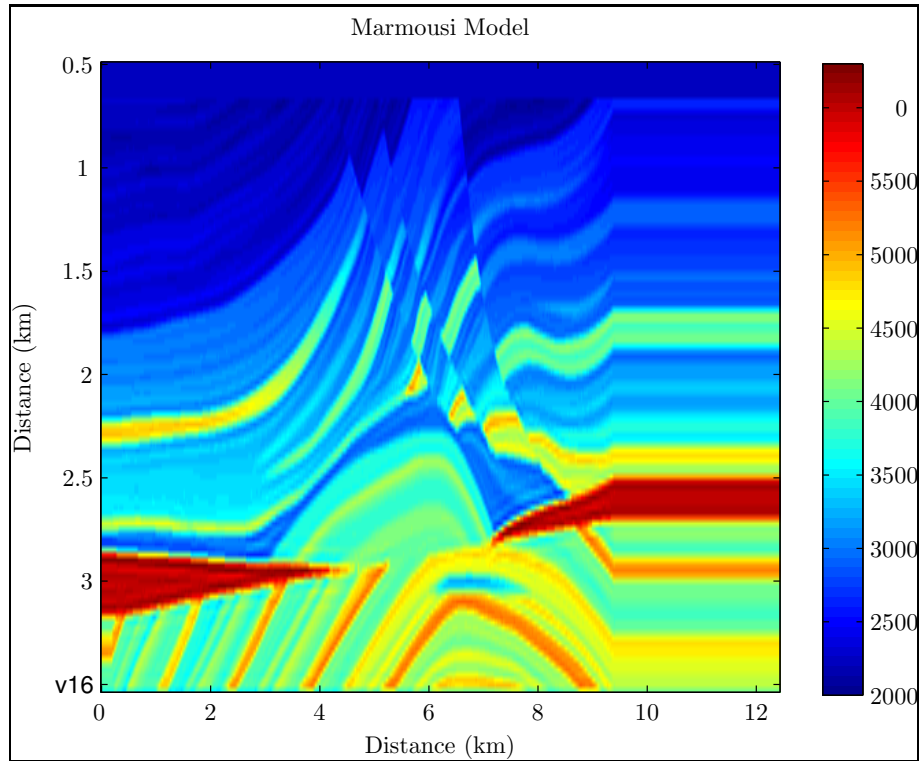


Figure 5: Marmousi original model [3].

The third and final model (figure 6) is the Sigsbee model, a complex model that contains a big salt flank. The shape of the salt flank is quite irregular, and the discontinuities at the salt boundaries are very strong. The big challenge of this model is that it presents illumination problems given the properties of the salt structure within the model, which yields to problems in the processing of the data.

The geometry for this model is defined with an extension from 3 to 27 km (around 24 km) in longitude and 9 km of depth. This model was originally spaced with 11.43 m in the horizontal axis and 7.6 m in the vertical axis but has been re-sampled to have a grid of 65 m in both the horizontal and the vertical axes the reason behind this adjustment will be discussed in chapter 2.

For this model, the velocities in the background are layered as a gradient that increases the velocity with depth whose velocities range from 1500 to 3700 m/s, but there is also a body salt which presents a high velocity of 4511 m/s.

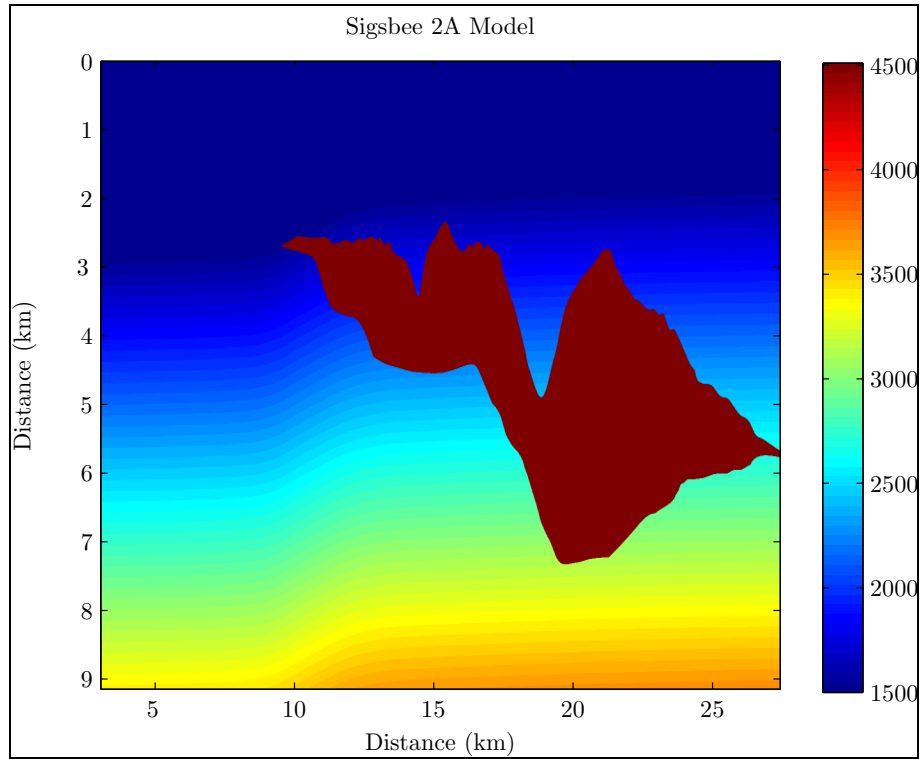


Figure 6: Sigsbee 2A original model [4].

These three SVM are the ones that will be used throughout the extent of the book, each of them offer a different challenge as the reader will see and results for each step of the processing will be shown for each one of them and every difficulty will be discussed and explain how those difficulties were (if possible) addressed.

Full Waveform Inversion

Physical phenomena in the nature are usually represented or modeled through their laws, in this cases the model parameters are known and what is modeled is the response of this model subject to a source of perturbation. However, sometimes the subject of interest is to determine the model parameters given determined (or observed) responses. These two scenarios are know as the forward and the inverse problems, respectively.

In this chapter both problems are addressed in the context of the seismic experiment. Also, both the mathematical and computational frameworks are presented as the tools to solve both the aforementioned problems.

2.1 The Forward Problem

The first of the problems, it predicts the error-free values of the observable parameters $\mathbf{d}$ that would correspond to a given model $\mathbf{m}$. Equation 2.1 shows that in order to go from the model space to the data space, an operator $G(\cdot)$ is applied. That is,

$$\mathbf{m} \rightarrow \mathbf{d} = G(\mathbf{m}), \quad (2.1)$$

where,

$\mathbf{m}$: represents the model parameters. In our case, the velocities of the subsurface of the Earth,

$\mathbf{d}$: represents the observable parameters. This observable parameters are the data that, for the seismic experiment presented, are the seismic traces collected at the surface.

$G(\cdot)$: represents the forward operator. In other words, the mathematical model of the physical system.

2.1.1 The Wave Equation

Continuing with the description of the forward problem, one imagines that a physical phenomenon of energy propagating as waves takes place. This thought frames the context in which all subsequent approaches will focus.

The wave phenomena can be described through partial differential equations. The level of complexity depends on the level of accuracy that one is interested in. For the seismic problem, the wave equation describes the energy propagated from a source through the space-time continuum.

The Earth's subsurface is an anisotropic and viscoelastic medium in which energy propagates in all three dimensions. This model of the Earth however, is very complex. For this reason one has to add constraints to the equation that simulates the phenomenon.

The question that arises is: *what are the constraints with which such an equation will be posed?*

This issue is of key importance as it deals with the compromise of how well one wants to represent the physical phenomenon and how much is one willing to spend in terms of dimensional, mathematical and computational complexity, for such a representation.

The first constrain to select is the dimensionality of the problem. As we deal with two-dimensional models of the subsurface of the Earth, the natural choice is to model the two-dimensional wave equation.

The second constrain deals with the phenomena that will be represented, for this selection one has to determine which waves are of interest, such as P waves, S waves, Love waves, etc. This frames the type of medium with the options of acoustic, elastic or viscoelastic medium. Since we are dealing with single component acquisitions (only the vertical disturbance is measured) and seismic reflection experiments commonly use the P -wave information, the medium will be modeled as an acoustic one.

A third constrain takes into account the direction in which waves propagate, for the sake of simplicity, we will model the propagation of the waves equally in every direction, this means that, for a material underground, waves will propagate equally upwards, downwards and in horizontal direction. This modeling corresponds to an isotropic model of the subsurface.

Finally, the energy propagation in the space-time continuum relate to the impedance of the subsurface of the Earth. The impedance is the product between the velocity of the model and its density. The impedance (or velocity and density) is (are) our model parameter(s). Our constrain is to leave the density constant and work with the velocity alone.

Simple choices were made as the aim of this book is not to test the several combinations of the selected constrains. Future work could benefit from improving the accuracy of the mathematical model, however take into account that the computational load increases exponentially.

In this book, the acoustic and isotropic 2D-wave equation with constant density will be used to model the forward problem, as given by

$$\frac{\partial^2 P}{\partial x^2} + \frac{\partial^2 P}{\partial z^2} = \frac{1}{\mathbf{m}^2} \frac{\partial^2 P}{\partial t^2}, \quad (2.2)$$

where P represents the wave's pressure intensity in time (t) and space (x, z). Also $\mathbf{m}$ represents the velocity of the medium (those velocities are our model parameters).

Now, with the mathematical model in hand, the next step is to implement it computationally.

2.1.2 The Finite Difference Time Domain Method

There are several alternatives to solve the wave equation 2.2, for example analytically, this approach however becomes tedious as the models increase their complexity. For this reason a computational alternative is considered. In electro-magnetics, the Finite Difference Time Domain method (FDTD) was described by Yee [10] for solving Maxwell's curl equations on grids staggered in space and time. This method belongs in the general class of grid-based differential numerical modeling methods and can be applied to solve our wave equation.

Since it is a time-domain method, FDTD solutions can cover a wide frequency range, one of its drawbacks however is the computing time which is reduced when using for example frequency-domain methods.

Although there are several methods for solving PDE's, the advantages of the FDTD method over other methods, such as finite elements, include: short development time, ease of understanding and explicit nature.

Equation 2.2 can be rewritten in terms of finite differences [11] by expanding into its second order Taylor's series in its temporal domain and eighth order in its spatial domain, as follows:

For the temporal derivative,

$$\frac{\partial^2 P}{\partial t^2} \simeq \frac{P_{i,j}^{n+1} - 2P_{i,j}^n + P_{i,j}^{n-1}}{\Delta t^2} \quad (2.3)$$

and for each the spatial derivative,

$$\frac{\partial^2 P}{\partial x^2} \simeq \frac{-\frac{1}{560}P_{i+4,j}^n + \frac{8}{315}P_{i+3,j}^n - \frac{1}{5}P_{i+2,j}^n + \frac{8}{5}P_{i+1,j}^n - \frac{205}{72}P_{i,j}^n + \frac{8}{5}P_{i-1,j}^n - \frac{1}{5}P_{i-2,j}^n + \frac{8}{315}P_{i-3,j}^n - \frac{1}{560}P_{i-4,j}^n}{\Delta x^2} \quad (2.4)$$

The same expansion is done for the z derivative. Replacing the terms above in the wave equation 2.2 yields to

$$\begin{aligned} \frac{1}{v^2} \frac{P_{i,j}^{n+1} - 2P_{i,j}^n + P_{i,j}^{n-1}}{\Delta t^2} = & \frac{1}{\Delta x^2} \left(-\frac{1}{560}P_{i+4,j}^n + \frac{8}{315}P_{i+3,j}^n - \frac{1}{5}P_{i+2,j}^n + \frac{8}{5}P_{i+1,j}^n \right. \\ & \left. - \frac{205}{72}P_{i,j}^n + \frac{8}{5}P_{i-1,j}^n - \frac{1}{5}P_{i-2,j}^n + \frac{8}{315}P_{i-3,j}^n - \frac{1}{560}P_{i-4,j}^n \right) \\ & + \frac{1}{\Delta z^2} \left(-\frac{1}{560}P_{i,j+4}^n + \frac{8}{315}P_{i,j+3}^n - \frac{1}{5}P_{i,j+2}^n + \frac{8}{5}P_{i,j+1}^n \right. \\ & \left. - \frac{205}{72}P_{i,j}^n + \frac{8}{5}P_{i,j-1}^n - \frac{1}{5}P_{i,j-2}^n + \frac{8}{315}P_{i,j-3}^n - \frac{1}{560}P_{i,j-4}^n \right) \end{aligned} \quad (2.5)$$

where i, j and n represent the discretization variables for x, z and t respectively; and Δx , Δz represent the step in the spatial directions, and Δt the time-step.

Solving for the wave field $P_{i,j}^{n+1}$, with $\Delta x = \Delta z = \Delta h$, the equation 2.6 can be rewritten as:

$$\begin{aligned} P_{i,j}^{n+1} = & 2P_{i,j}^n - P_{i,j}^{n-1} + \frac{v^2 \Delta t^2}{\Delta h^2} \left(-\frac{1}{560}P_{i+4,j}^n + \frac{8}{315}P_{i+3,j}^n - \frac{1}{5}P_{i+2,j}^n + \frac{8}{5}P_{i+1,j}^n - \frac{205}{72}P_{i,j}^n \right. \\ & + \frac{8}{5}P_{i-1,j}^n - \frac{1}{5}P_{i-2,j}^n + \frac{8}{315}P_{i-3,j}^n - \frac{1}{560}P_{i-4,j}^n - \frac{1}{560}P_{i,j+4}^n + \frac{8}{315}P_{i,j+3}^n - \frac{1}{5}P_{i,j+2}^n \\ & \left. + \frac{8}{5}P_{i,j+1}^n - \frac{205}{72}P_{i,j}^n + \frac{8}{5}P_{i,j-1}^n - \frac{1}{5}P_{i,j-2}^n + \frac{8}{315}P_{i,j-3}^n - \frac{1}{560}P_{i,j-4}^n \right). \end{aligned} \quad (2.6)$$

Equation 2.6 was implemented in cuda-C to take advantage of parallel computing for solving the wave equation and hence, the forward problem.

Looking at equation 2.6 one can notice that in order to obtain the wave field P^{n+1} at a point i and j , one requires the current and previous values of $P_{i,j}$ and also four points in each direction of the field P^n . This is what is known as stencil, and depending on the order of the approximation it has more or less terms.

Boundaries, numerical dispersion and stability

Every time we simulate a physical phenomena through numerical methods there are artificial (computational) effects not produced by the physical phenomena itself: artificial reflections, numerical dispersion and the possibility of having divergent results (instability).

Artificial reflections are produced when the wavefront reaches the end of the computational extent of the model (read: the edges). The solution to this problem is found in the definition of Perfectly Matched Layers that absorb the energy of the wavefront avoiding reflections to happen. Stability is a desired behavior for the simulation to have. If instability occurs, the values in the simulation will tend to infinity and beyond! [12]. This happens when the values of the simulation become complex and moreover exponentially increasing. The final effect, dispersion, is almost always present but can be minimized and accounted for. It highly depends on the time and space steps, the discretization method and the frequency of the seismic source. Instability and artificial reflections are effects that have to be avoided, while dispersion is an effect that have to be minimized.

The Convolutional Perfectly Matched Layer (CPML) boundary condition is proposed in [13] to deal with the artificial reflections. This method requires two auxiliary variables for each spatial dimension. This turns equation 2.2 into

$$\frac{1}{\mathbf{m}^2} \frac{\partial^2 P}{\partial t^2} = \frac{\partial^2 P}{\partial x^2} + \frac{\partial^2 P}{\partial z^2} + \frac{\partial \psi_x}{\partial x} + \frac{\partial \psi_z}{\partial z} + \zeta_x + \zeta_z. \quad (2.7)$$

According to [13], the auxiliary variables that minimize artificial reflections are

$$\psi_q^n = b_q \psi_q^{n-1} + a_q \frac{\partial P}{\partial q} \quad (2.8)$$

$$\zeta_q^n = b_i \zeta_q^{n-1} + a_q \left[\left(\frac{\partial^2 P}{\partial q^2} \right)^n + \left(\frac{\partial \psi_q}{\partial q} \right)^n \right] \quad (2.9)$$

where q can be either x or z , ψ_q^0 y ζ_q^0 are zero at the first iteration and a_q and b_q can be computed (for the x dimension) using the equations presented in [14] as

$$a_x = \frac{d_x}{d_x - \alpha_x}(b_x - 1) \quad (2.10)$$

$$b_x = e^{-(d_x + \alpha_x)dt}, \quad (2.11)$$

where α_x and d_x are

$$d_x = d_0 V_{max} \left(\frac{f(x)}{L_x} \right)^2, \quad (2.12)$$

$$\alpha_x = \pi f \left(\frac{L_x - f(x)}{L_x} \right). \quad (2.13)$$

Here L_x and d_0 are defined as

$$L_x = \text{CPML}_{\text{limit}} \times \Delta x \quad (2.14)$$

$$d_0 = \frac{-3}{2L_x} \log(R), \quad (2.15)$$

with $\text{CPML}_{\text{limit}}$ being the amount of steps over the attenuation will take place and $R = 0.001$. The function $f(x)$ is defined as

$$f(x) = \begin{cases} L_x : \Delta h : 0 & x \in (0, \text{CPML}] \\ 0 & x \in (\text{CPML}, Nx - \text{CPML}) \\ 0 : \Delta h : L_x & x \in (Nx - \text{CPML}, Nx) \end{cases} \quad (2.16)$$

and a plot of it can be seen in 7.

Figures 8 and 9 show parameters a_x , b_x and d_x as functions of x . Just as $f(x)$, these parameters will be zero for regions where there is not attenuation making the boundary conditions valid just in the region where the CPML takes place. In this example, when $x < 20$ and $x > 80$ [15]. This means that for $20 < x < 80$, ψ_x^n and ζ_x^n will also be zero and hence, the update of the model is given by the wave equation 2.2.

The same analysis is done for the z dimension.

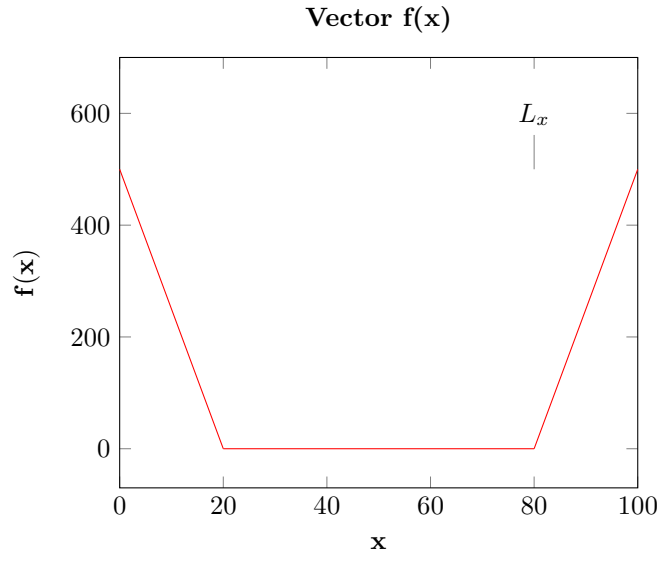


Figure 7: The $f(x)$ function, with $Nx=100$, $\text{CPML}_{\text{limit}}=20$, $\Delta h=25$, $L_x=500$, $f=3Hz$

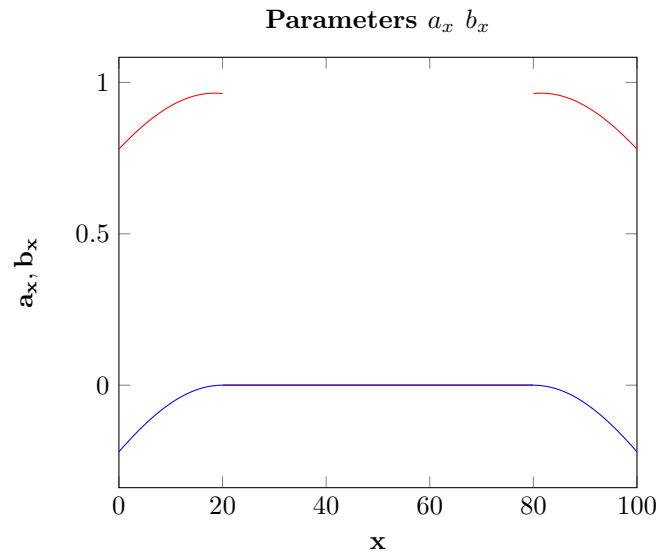


Figure 8: Values of a_x (blue) y b_x (red), with $Nx=100$, $\text{CPML}_{\text{limit}}=20$, $dt=4e-3$.

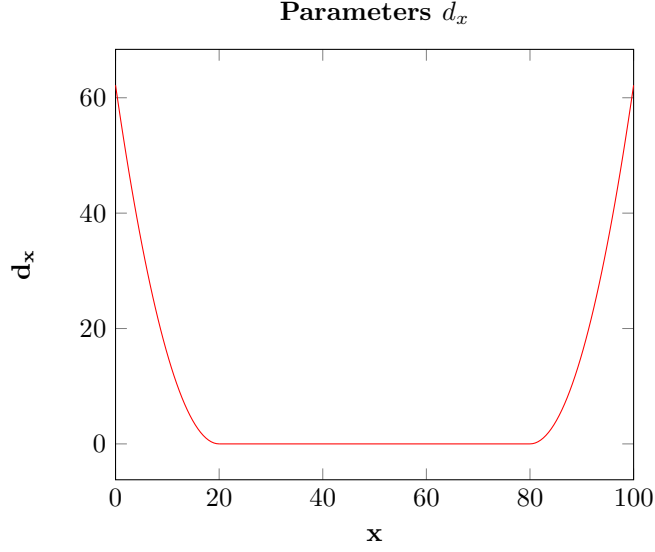


Figure 9: Values of d_x , with $Nx=100$, $\text{CPML}_{\text{limit}}=20$, $dt=4\text{e-}3$.

To determine the condition that has to be met in order to guarantee stability and account for numerical dispersion we will introduce a plane wave $U(t, x, z)$ as

$$U(t, x, z) = U_0 e^{i(n\omega\Delta t - lk_x\Delta x - mk_z\Delta z)}, \quad (2.17)$$

and we put it through our finite differences equation scheme (equation 2.5). The time derivative is

$$\begin{aligned} \frac{\partial^2 U}{\partial t^2} \simeq \frac{1}{\Delta t} & \left(U_0 e^{i(n\omega\Delta t + \omega\Delta t - lk_x\Delta x - mk_z\Delta z)} \right. \\ & - 2U_0 e^{i(n\omega\Delta t - lk_x\Delta x - mk_z\Delta z)} \\ & \left. + U_0 e^{i(n\omega\Delta t - \omega\Delta t - lk_x\Delta x - mk_z\Delta z)} \right). \end{aligned} \quad (2.18)$$

Now let us name $A = n\omega\Delta t - lk_x\Delta x - mk_z\Delta z$, and obtain the spatial derivative (one dimension) as

$$\begin{aligned} \frac{\partial^2 U}{\partial x^2} \simeq \frac{U_0}{\Delta t} & \left(\frac{-1}{560} e^{i(A-4k_x\Delta x)} + \frac{8}{315} e^{i(A-3k_x\Delta x)} \right. \\ & - \frac{1}{5} e^{i(A-2k_x\Delta x)} + \frac{8}{5} e^{i(A-k_x\Delta x)} + \frac{205}{72} e^{i(A)} \\ & - \frac{1}{560} e^{i(A+k_x\Delta x)} + \frac{8}{315} e^{i(A+2k_x\Delta x)} \\ & \left. - \frac{1}{5} e^{i(A+3k_x\Delta x)} + \frac{8}{5} e^{i(A+4k_x\Delta x)} \right), \end{aligned} \quad (2.19)$$

the term $\frac{\partial^2 U}{\partial z^2}$ is the same as the one in equation 2.19 but for z .

It is evident that the term $U_0 e^{i(A)}$ appears on both sides of the equation and therefore can be canceled, leaving us with

$$\begin{aligned} \frac{1}{\Delta t} \left(e^{i(\omega \Delta t)} - 2 + e^{-i(\omega \Delta t)} \right) &= \frac{\mathbf{m}^2}{\Delta x^2} \left(\frac{-1}{560} (e^{i(4k_x \Delta x)} + e^{-i(4k_x \Delta x)}) + \frac{8}{315} (e^{i(3k_x \Delta x)} + e^{-i(3k_x \Delta x)}) \right. \\ &\quad \left. - \frac{1}{5} (e^{i(2k_x \Delta x)} + e^{-i(2k_x \Delta x)}) + \frac{8}{5} (e^{i(k_x \Delta x)} + e^{-i(k_x \Delta x)}) - \frac{205}{72} \right) \\ &\quad + \frac{\mathbf{m}^2}{\Delta z^2} \left(\frac{-1}{560} (e^{i(4k_z \Delta z)} + e^{-i(4k_z \Delta z)}) + \frac{8}{315} (e^{i(3k_z \Delta z)} + e^{-i(3k_z \Delta z)}) \right. \\ &\quad \left. - \frac{1}{5} (e^{i(2k_z \Delta z)} + e^{-i(2k_z \Delta z)}) + \frac{8}{5} (e^{i(k_z \Delta z)} + e^{-i(k_z \Delta z)}) - \frac{205}{72} \right), \end{aligned} \quad (2.20)$$

where exponentials can be replaced by cosines leaving us with

$$\begin{aligned} (2 \cos(\omega \Delta t^2) - 2) &= \frac{\Delta t \mathbf{m}^2}{\Delta x^2} \left(\frac{-2}{560} \cos(4k_x \Delta x) + \frac{16}{315} \cos(3k_x \Delta x) - \frac{2}{5} \cos(2k_x \Delta x) + \frac{16}{5} \cos(k_x \Delta x) - \frac{205}{72} \right) \\ &\quad + \frac{\Delta t \mathbf{m}^2}{\Delta z^2} \left(\frac{-2}{560} \cos(4k_z \Delta z) + \frac{16}{315} \cos(3k_z \Delta z) - \frac{2}{5} \cos(2k_z \Delta z) + \frac{16}{5} \cos(k_z \Delta z) - \frac{205}{72} \right). \end{aligned} \quad (2.21)$$

Equation 2.21 can in turn, be converted from cosines to square sines using the double angle identity. It is worth noting that the constant terms cancel, leaving just the sines.

$$\begin{aligned} \sin^2 \left(\frac{\omega \Delta t}{2} \right) &= \frac{\Delta t^2 \mathbf{m}^2}{\Delta x^2} \left[\frac{-1}{560} \sin^2 \left(\frac{4k_x \Delta x}{2} \right) + \frac{8}{315} \sin^2 \left(\frac{3k_x \Delta x}{2} \right) - \frac{1}{5} \sin^2 \left(\frac{2k_x \Delta x}{2} \right) + \frac{8}{5} \sin^2 \left(\frac{k_x \Delta x}{2} \right) \right] \\ &\quad + \frac{\Delta t^2 \mathbf{m}^2}{\Delta z^2} \left[\frac{-1}{560} \sin^2 \left(\frac{4k_z \Delta z}{2} \right) + \frac{8}{315} \sin^2 \left(\frac{3k_z \Delta z}{2} \right) - \frac{1}{5} \sin^2 \left(\frac{2k_z \Delta z}{2} \right) + \frac{8}{5} \sin^2 \left(\frac{k_z \Delta z}{2} \right) \right] \end{aligned} \quad (2.22)$$

The goal is to solve equation 2.22 for ω . Here we assume (as in our examples) that $\Delta x = \Delta z$ and define the space step Δh . We also define $\beta = \frac{\Delta t \mathbf{m}}{\Delta x}$. Finally, we rename the terms in brackets as $B_x = \left[\frac{-1}{560} \sin^2 \left(\frac{4k_x \Delta x}{2} \right) + \frac{8}{315} \sin^2 \left(\frac{3k_x \Delta x}{2} \right) - \frac{1}{5} \sin^2 \left(\frac{2k_x \Delta x}{2} \right) + \frac{8}{5} \sin^2 \left(\frac{k_x \Delta x}{2} \right) \right]$ and B_z , respectively.

Solving for ω results in

$$\omega = \frac{2}{\Delta t} \arcsin \left(\beta \sqrt{B_x + B_z} \right), \quad (2.23)$$

which is an essential equation as it will help determine the stability condition and account for the numerical dispersion.

Given the $\arcsin(\cdot)$ operator, we know that $-1 \leq \beta\sqrt{B_x + B_z} \leq 1$, but as β is the result of a product between positive numbers, it will always be positive and also the terms inside the square root will always be positive. This gives the domain $0 \leq \beta\sqrt{B_x + B_z} \leq 1$ for ω to be real.

The analysis becomes a comparison between the maximum velocity in the model (which affects β) and the maximum value of the function inside the square root.

For our case, the stability criterion is, in terms of the grid steps and the velocity,

$$\frac{\mathbf{m}_{\max}\Delta t}{\Delta h} \leq \frac{1}{\sqrt{\max(B_x) + \max(B_z)}} = 0.5546, \quad (2.24)$$

which is the value that limits the relationship between spatial and temporal grids alongside the velocity model.

This limit is also known as the Courant's stability criterion and according to Slawinski [16], can be obtained as

$$\frac{\mathbf{m}_{\max}\Delta t}{\Delta h} \leq \frac{\sqrt{a_1}}{\sqrt{a_2}} \quad (2.25)$$

where, $\mathbf{m}_{\max}$ is the maximum velocity in the model, Δh and Δt are the time and space steps respectively, and the terms a_1 and a_2 are the sum of the temporal and spatial coefficients of the expansion, respectively.

$$a_1 = \sum |w_i| = 1 + 2 + 1 = 4 \quad (2.26)$$

$$a_2 = \sum |w_i| = 2 * \left(2 * \left(\frac{1}{560} + \frac{8}{315} + \frac{1}{5} + \frac{8}{5} \right) + \frac{205}{72} \right) = 13.0032 \quad (2.27)$$

$$\frac{\mathbf{m}_{\max}\Delta t}{\Delta h} \leq \frac{\sqrt{4}}{\sqrt{13.0032}} = 0.5546 \quad (2.28)$$

The results in equations 2.24 and 2.28 show that both approaches are possible, with the first one being more analytical than the second, but the second being more practical. Those in-equations define the stability of the solution, when the in-equation does not hold, there has to be an adjustment in the temporal and/or the spatial grid.

Numerical dispersion can be accounted by considering the normalized phase velocity $\frac{\mathbf{M}_p}{\mathbf{m}}$ as a function of frequency,

$$\frac{\mathbf{M}_p}{\mathbf{m}} = \frac{\omega}{KV}, \quad (2.29)$$

with ω obtained from equation 2.23 and we neglect the effect of multiple components of spatial frequency K to make the approximation of $K \simeq k_x$ and take only into account the dispersion for one direction. This results in

$$\frac{\mathbf{M}_p}{\mathbf{m}} = \frac{2}{k_x \Delta t \mathbf{m}} \arcsin \left(\beta \sqrt{B_x} \right), \quad (2.30)$$

where we can use the definition of β to change the term $\Delta t \mathbf{m}$ for $\beta \Delta x$ and replace B_x for its definition too, to obtain equation 2.30 in terms of β and $k_x \Delta x$ as

$$\frac{\mathbf{M}_p}{\mathbf{m}} = \frac{2}{\beta k_x \Delta x} \arcsin \left(\beta \sqrt{\frac{-1}{560} \sin^2 \left(\frac{4k_x \Delta x}{2} \right) + \frac{8}{315} \sin^2 \left(\frac{3k_x \Delta x}{2} \right) - \frac{1}{5} \sin^2 \left(\frac{2k_x \Delta x}{2} \right) + \frac{8}{5} \sin^2 \left(\frac{k_x \Delta x}{2} \right)} \right). \quad (2.31)$$

The next step is to define the range of possible $\beta's$, to do this we use its definition replacing the velocity values with the maximum and minimum velocity values of the model as

$$\beta_{\min} = \frac{\Delta t}{\Delta x} \mathbf{m}_{\min} \quad (2.32)$$

and,

$$\beta_{\max} = \frac{\Delta t}{\Delta x} \mathbf{m}_{\max}. \quad (2.33)$$

From equation 2.31 we can plot two functions, one for each β , in which the independent variable is the product $k_x \Delta x$ and the dependent variable is $\frac{\mathbf{M}_p}{\mathbf{m}}$. Figure 10 depicts the numerical dispersion plots for the square diffractor model. In the figure, there is a margin of acceptance of around 1% of numerical dispersion, and with it we define the value of $k_x \Delta x$.

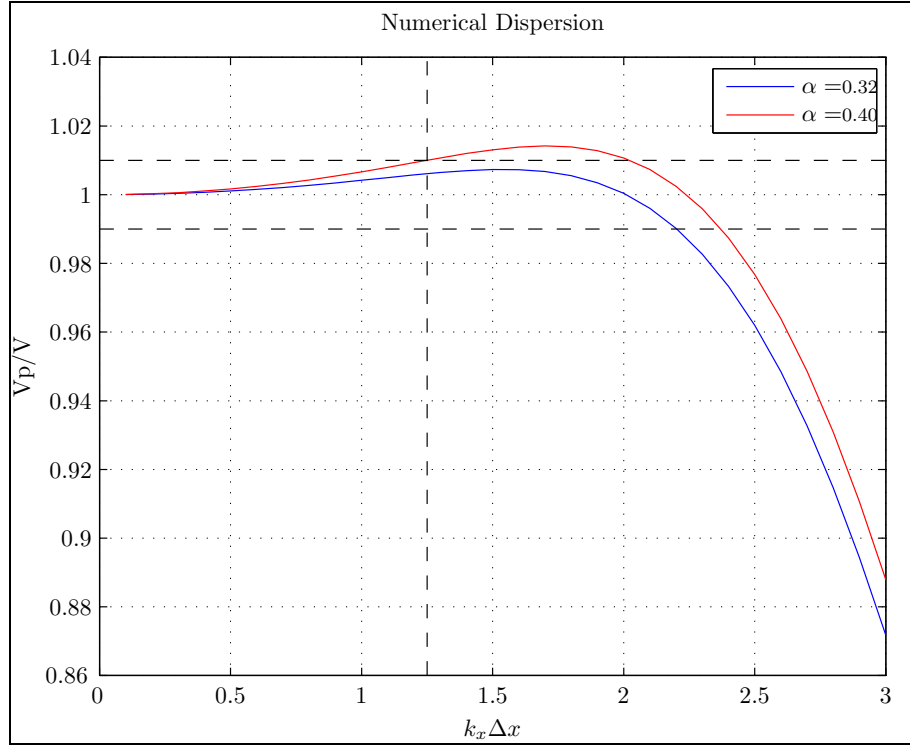


Figure 10: Numerical Dispersion of 1% for the square diffractor model for which $k_x \Delta x = 1.25$

The importance of defining the $k_x \Delta x$ comes with the selection of acceptable frequencies for the seismic sources. Knowing $k_x \Delta x$ and Δx we can get k_x and with it and equation 2.34 obtain the maximum frequency of the sources as

$$f_{\max} = \frac{k_x \mathbf{m}_{\min}}{2\pi} \quad (2.34)$$

We repeated this process for and results are presented in table 1 with the values of $\beta_{\max}$, $\beta_{\min}$ and the maximum frequency $f_{\max}$.

Model	$\beta_{\min}$	$\beta_{\max}$	$k_x \Delta x$	$f_{\max}$ [Hz]
Diffractor	0.32	0.40	1.25	15
Marmousi	0.12	0.47	1.05	10
Sigsbee	0.18	0.54	1.85	6

Table 1: Results for approximately 1% of numerical dispersion in the three models.

2.1.3 Wave Propagation Through the SVM's

To conclude this chapter, the results of the implementation over the selected SVM's are shown. For each simulation a wave propagation was performed, it was used only one source (one shot). The location of the sources are presented in table 2 and represent the coordinates at the center of each of the models with frequency of 5 Hz.

Model	Horizontal Position [km]	Depth [km]
Diffractor	2.6	0.1
Marmousi	6.2	0.1
Sigsbee	15.2	0.26

Table 2: Source locations in the SVM's.

Figures 11 and 12 show two snapshots of the propagation for the square diffractor example while figure 13 show the recorded data after the reflection takes place.

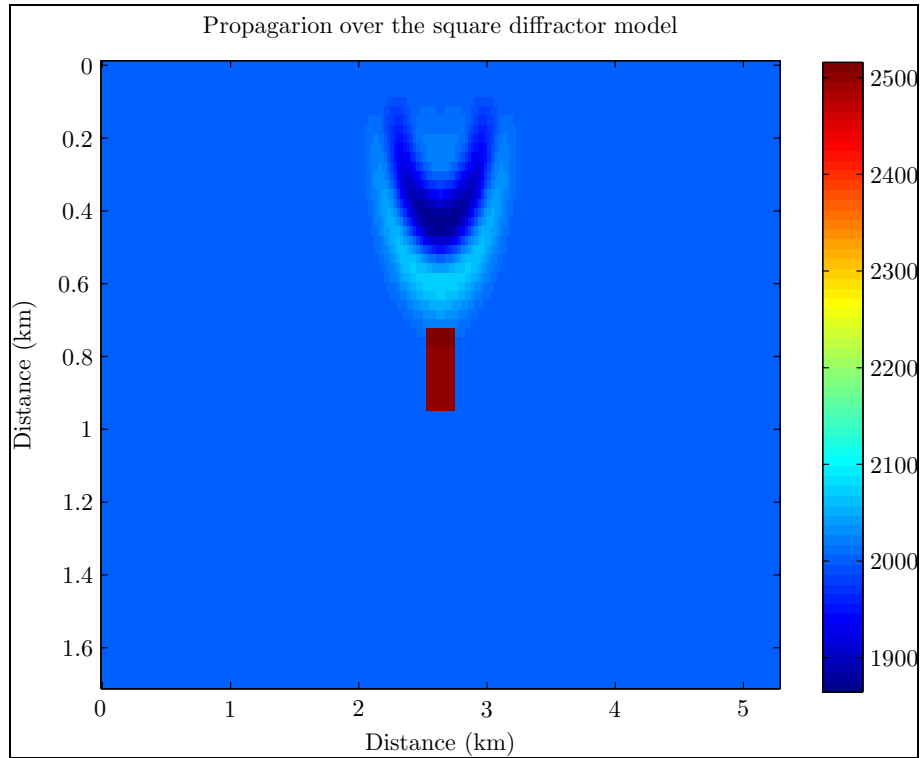


Figure 11: Wave propagation over the square diffractor model (before diffractor).

As it can be seen in figure 12, part of the energy is reflected to the surface, this energy is what is measured, as seen in figure 13.

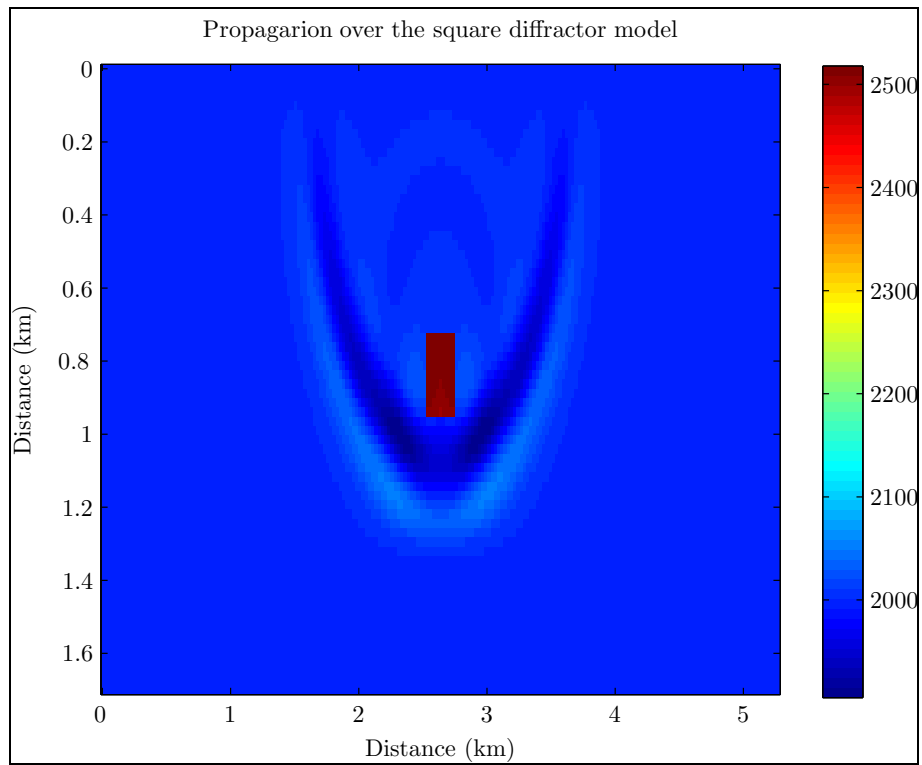


Figure 12: Wave propagation over the square diffractor model (after diffractor).

Similar to the example above, the figures 14 and 15 present the propagation through the Marmousi and the Sigsbee models.

From these images it can be seen how the complexity in the geometry (of Marmousi for example) influences the propagation and also how the high velocity absorbs the wavefront (Sigsbee example).

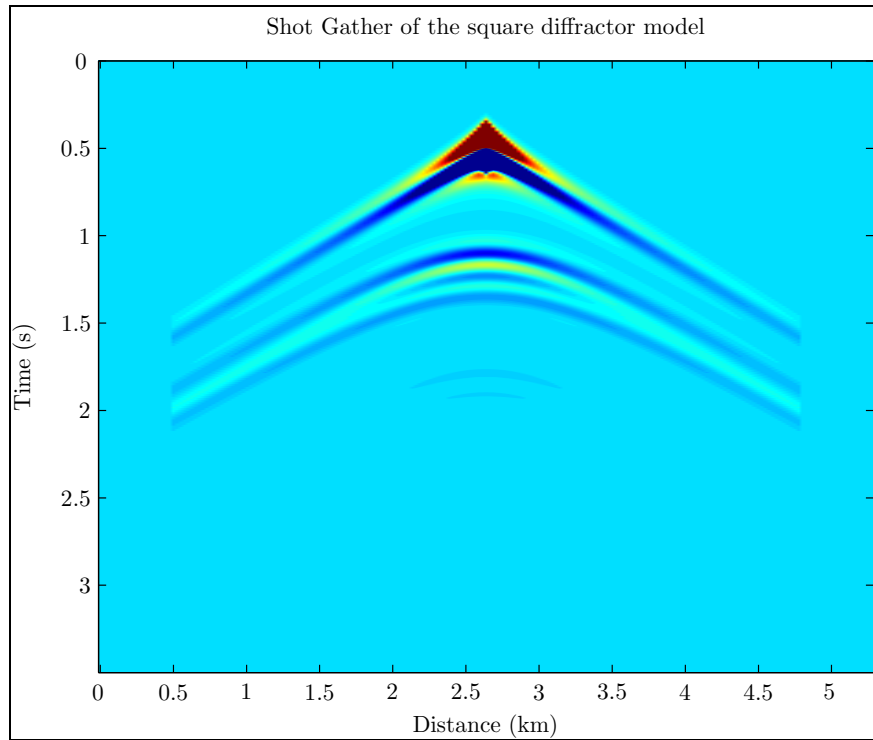


Figure 13: Shot Gather of the square diffractor model.

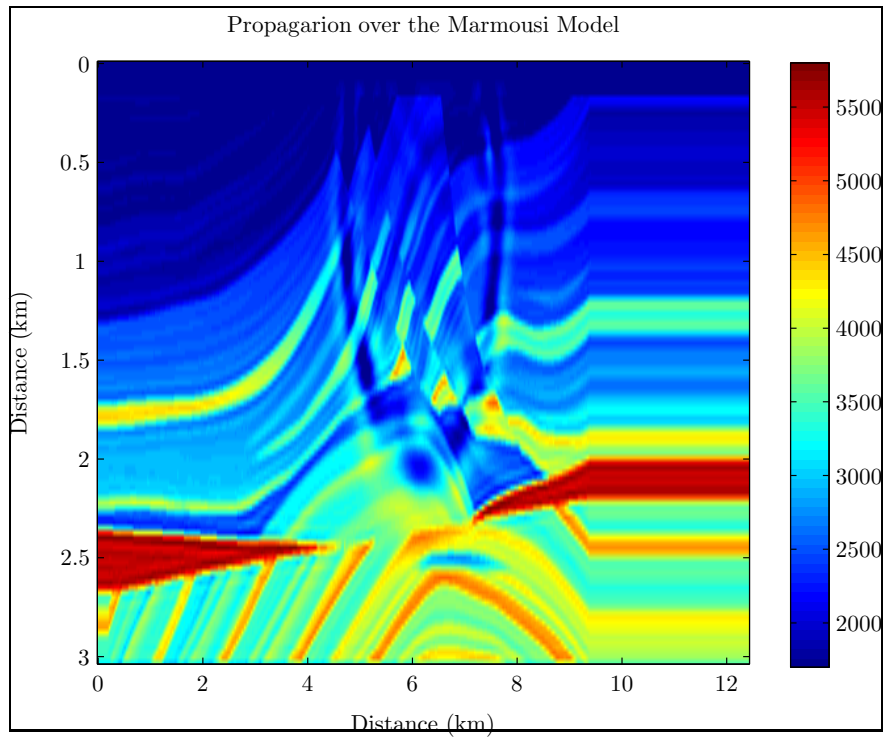


Figure 14: Wave propagation over the Marmousi model.

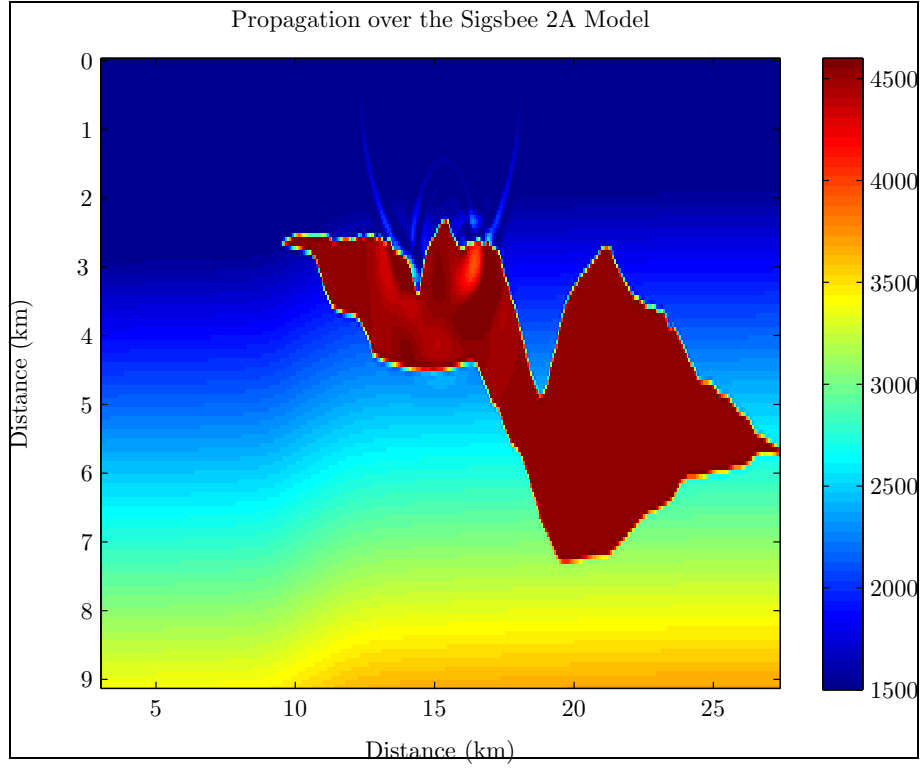


Figure 15: Wave propagation over the Sigsbee 2A model.

2.2 The Inverse Problem

Contrary to the forward problem, the inverse problem deals with the estimation of the model parameters given an initial observation of the phenomenon. The technique that will be used to solve the inverse problem is the full waveform inversion (FWI) proposed by Lailly in 1983 [17] and Tarantola in 1984 [5].

The basic principle behind the FWI involves the usage of acquired seismic data to estimate a subsurface velocity model $\mathbf{m}$, minimizing the difference between the observed $\mathbf{d}$ and modeled data $G(\mathbf{m})$.

FWI has proven to get good results in terms of resolution when compared with techniques such as travel time tomography [18].

The forward problem is presented in equation 2.1, and one could think of its counterpart (the inverse problem) as

$$\mathbf{m} = G^{-1}(\mathbf{d}). \quad (2.35)$$

The use of equation 2.35 has proven to obtain good results given that the solution model: exists, is unique and stable. This however, may not be the case given the ill-posedness of the problem[19]. The result is that the function that describes the phenomenon is non-convex (read: has multiple minima). In consequence, it is necessary to implement a numerical solution to the problem.

The solution to the inverse problem is formulated as an optimization problem specifically: minimum least squares. Thereby the cost function minimizes the misfit with the ℓ_2^2 -norm as shown in equation 2.36.

$$\Phi(\mathbf{m}) = \frac{1}{2} \|\mathbf{d} - G(\mathbf{m})\|_2^2. \quad (2.36)$$

Equation 2.36 is solved iteratively by updating the model ($\mathbf{m}$) as showed in equation 2.37:

$$\mathbf{m}_{k+1} = \mathbf{m}_k + \alpha_k \Delta \mathbf{m}, \quad (2.37)$$

where an initial guess ($\mathbf{m}_0$) is needed and the future model ($\mathbf{m}_{k+1}$) is obtained by adding a variation ($\Delta \mathbf{m}$) to the current model ($\mathbf{m}$) and the term α_k is a scale factor that improves the advancement towards the direction of the solution.

2.2.1 Computing $\alpha_k \Delta \mathbf{m}$: the L-BFGS method

After obtaining equation 2.37, the question is how to determine the terms α_k and $\Delta \mathbf{m}$. The first step is to find an expression for $\Delta \mathbf{m}$, for this we use the Newton's method [20].

The process can be listed as follows:

1. Represent $\Phi(\mathbf{m})$ through Taylor series.
2. Derive $\Phi(\mathbf{m})$ with respect to $\mathbf{m}$.
3. Equal the derivative of $\Phi(\mathbf{m})$ to zero.
4. Solve the equation for $\Delta \mathbf{m}$.

The Taylor's series as presented in equation 2.38.

$$\Phi(\mathbf{m}_{k+1}) = \Phi(\mathbf{m}_k) + \Delta \mathbf{m}^T \mathbf{g}(\mathbf{m}_k) + \frac{1}{2} \Delta \mathbf{m}^T \mathbf{H}(\mathbf{m}_k) \Delta \mathbf{m} + \mathbf{O}(\Delta \mathbf{m}^3), \quad (2.38)$$

where, $\mathbf{g}(\mathbf{m}_k) = \frac{\partial \Phi(\mathbf{m}_k)}{\partial \mathbf{m}}$ and $\mathbf{H}(\mathbf{m}_k) = \frac{\partial^2 \Phi(\mathbf{m}_k)}{\partial \mathbf{m}^2}$ represent the gradient and the Hessian of the cost function respectively.

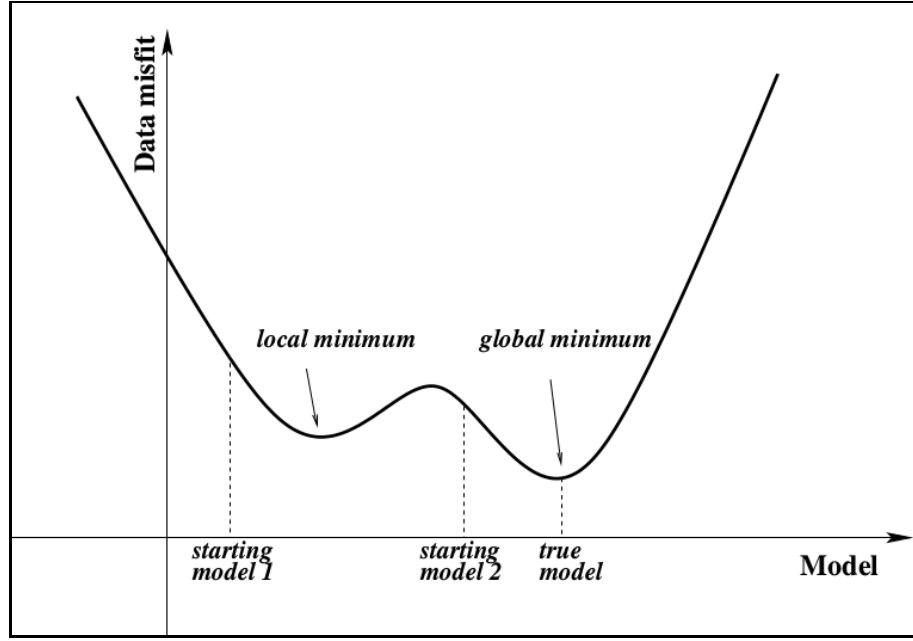


Figure 16: Schematic of the local-minima problem in FWI. Tarantola [5].

From equation 2.38, it can be obtained the model update $\Delta \mathbf{m}$ as showed in equation 2.39.

$$\Delta \mathbf{m} = -[\mathbf{H}(\mathbf{m}_k)]^{-1} \mathbf{g}(\mathbf{m}_k). \quad (2.39)$$

Replacing 2.39 into 2.37, the latter can be rewritten as equation 2.40.

$$\mathbf{m}_{k+1} = \mathbf{m}_k - \alpha_k [\mathbf{H}(\mathbf{m}_k)]^{-1} \mathbf{g}(\mathbf{m}_k). \quad (2.40)$$

This approach has one key issue. That is, the starting point ($\mathbf{m}_0$), as it plays a fundamental role in the solution process. As seen in figure 16, selecting the starting point close to a local minimum will yield to never reaching the global minimum (a.k.a. the solution) because this method is a local optimization method. The $\mathbf{m}_0$ can be provided either by: experts, prior information of the zone or by global optimization methods.

Nevertheless, we move forward with an expression for $\Delta \mathbf{m}$. Now, it must be clarified how the update value is computed. Equation 2.40 refers to the update of the model. There are three terms to be found in that update term: the step-length related to the value of α_k , the step-curvature related to the inverse of $\mathbf{H}(\mathbf{m}_k)$ and the update direction related to $\mathbf{g}(\mathbf{m}_k)$. The selection of α_k becomes a problem of adjustment while the computation of $[\mathbf{H}(\mathbf{m}_k)]^{-1} \mathbf{g}(\mathbf{m}_k)$ is performed using a quasi-Newton method.

The most popular quasi-Newton algorithm is the BFGS method, named after its developers Broyden, Fletcher, Goldfarb and Shanno. We however used the limited memory version of it given that the Hessian matrix cannot be computed at a reasonable cost. It would bottleneck the whole inversion process as it has to be computed at each iteration, for each frequency.

The L-BFGS method [21] uses the gradients and models of the m previous iterations (with $m \leq 10$) for computing an approximation of the projections of the inverse of the Hessian matrix over the gradient $([\mathbf{H}(\mathbf{m}_k)]^{-1} \mathbf{g}(\mathbf{m}_k))$ recursively, as presented as in the algorithm 1

Algorithm 1 : L-BFGS Method

```

1:  $q \leftarrow \mathbf{g}(\mathbf{m}_k)$ ;
2: for  $i = k - 1 : -1 : k - m$  do
3:    $\epsilon_i \leftarrow \sigma_i s_i^T q$ ;
4:    $q \leftarrow q - \epsilon_i y_i$ ;
5: end for
6:  $r \leftarrow D_k^0 q$ ;
7: for  $i = k - m : 1 : k - 1$  do
8:    $\beta \leftarrow \sigma_i y_i^T r$ ;
9:    $r \leftarrow s_i(\epsilon_i - \beta)$ ;
10: end for
```

where, $s_k = \mathbf{m}_{k+1} - \mathbf{m}_k$ is the variation in velocity models, while $y_k = \mathbf{g}(\mathbf{m}_{k+1}) - \mathbf{g}(\mathbf{m}_k)$ is the variation in the gradients and

$$\sigma_k = \frac{1}{y_k^T s_k}, \quad (2.41)$$

$$D_k^0 = \gamma_k \mathbf{I}, \quad (2.42)$$

$$\gamma_k = \frac{s_{k-1}^T y_{k-1}}{y_{k-1}^T y_{k-1}}, \quad (2.43)$$

Finally, $r \simeq -[\mathbf{H}(\mathbf{m}_k)]^{-1} \mathbf{g}(\mathbf{m}_k)$ and the step-length term α_k is set to fulfill the condition

$$\Phi(\mathbf{m}_k - \alpha_k [\mathbf{H}(\mathbf{m}_k)]^{-1} \mathbf{g}(\mathbf{m}_k)) < \Phi(\mathbf{m}_k) \quad (2.44)$$

starting with $\alpha_k = 1$ and scaling it with a constant $c_1 \in (0, 1)$ until the condition is satisfied. Other criteria such as Wolfe conditions (presented in [21]) could be implemented, but given the increment in the computational cost and that our objective function $\Phi(\mathbf{m})$ is always positive, such conditions will be satisfied just by satisfying condition 2.44.

One advantage of this method is that it accomplishes convergence with less than 40 iterations for each frequency, tests for convergence showed that performing further iterations does not add up to the solution, in fact, as seen in figure 17, after 20 iterations the decrease rate is almost null.

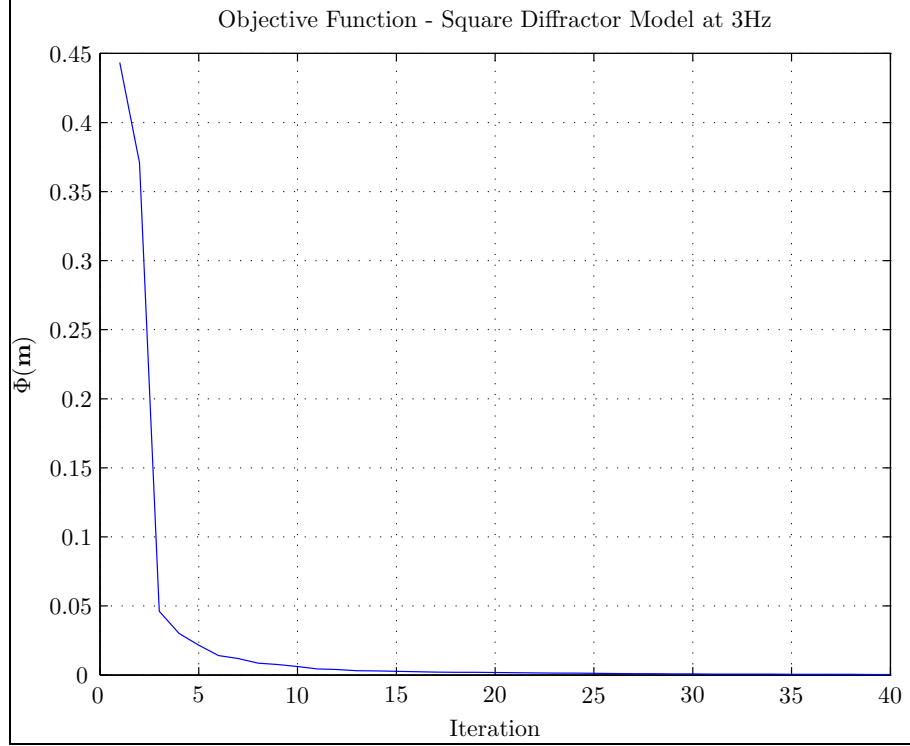


Figure 17: Objective function $\Phi(\mathbf{m})$ for the square diffractor model at a frequency of 3Hz. After 40 iterations, the frequency of the source was increased for the next 40 iterations.

2.2.2 Solutions to the Inverse Problem of the SVM's

As it was remarked, the initial model from which the FWI will start is of vital importance. Therefore, below there are not only the results but also the starting velocity models from which the inversion started.

Before showing the results, another concept has to be presented. That is, the multi-scale FWI. This approach takes advantage of the iterative process to conduct several FWI's one after the other using the latest result as initial model for the following process [22].

Given a number of iteration, the question is, what varies from experiment to experiment? The answer: *the source!* Specifically, the frequency of the source.

The method is simple, design the experiment: that is, number of sources and receivers, time of acquisition and finally the frequency of the source. The trick is to start from a low frequency and increase it every time the FWI finishes using the final model as initial model of the following FWI.

For our experiments the description of the source is

$$s = -2\pi^2 f^2 (t - t_0) e^{-(\pi^2 f^2)(t-t_0)^2}, \quad (2.45)$$

where t is the recording time, t_0 is when the recording starts and f is the frequency of the source limited by the numerical dispersion (see table 1). Table 3 shows which frequencies were used for each model.

Model	f_1 [Hz]	f_2 [Hz]	f_3 [Hz]	f_4 [Hz]
Diffractor	3	6	9	12
Marmousi	3	6	9	-
Sigsbee	3	6	-	-

Table 3: Frequencies used for the multi-scale FWI for each model.

As it turns out, four steps were done for the diffractor model, three for the Marmousi model, and just two for the Sigsbee model. Each step at 3Hz over the previous step.

The description of each experiment is presented in table 4 where all parameters for the inversion are defined. For all experiments, a CPML boundary of 20 points is added to the model and the total number of sources used is 21. These sources are horizontally distributed through the non-CPML zone from the first point until the last one with a stepsize of the distance between them divided by the number of sources minus one with a 5-points depth for each model.

Parameter	Diffractor	Marmousi	Sigsbee
Horizontal points (Nx)	211	497	375
Vertical points (Nz)	68	121	141
Spatial step [m] (Δh)	25	25	65
Temporal step[ms] (Δt)	4	2	4
Recording time[s]	3.5	4	14

Table 4: Parameters used for each model's inversion.

Continuing with the results of the inversion, our first case as always, is the square diffractor model.

In figure 18 it can be seen that the starting model has no information whatsoever about the high velocity square, it is a plain model of velocity 2000 m/s.

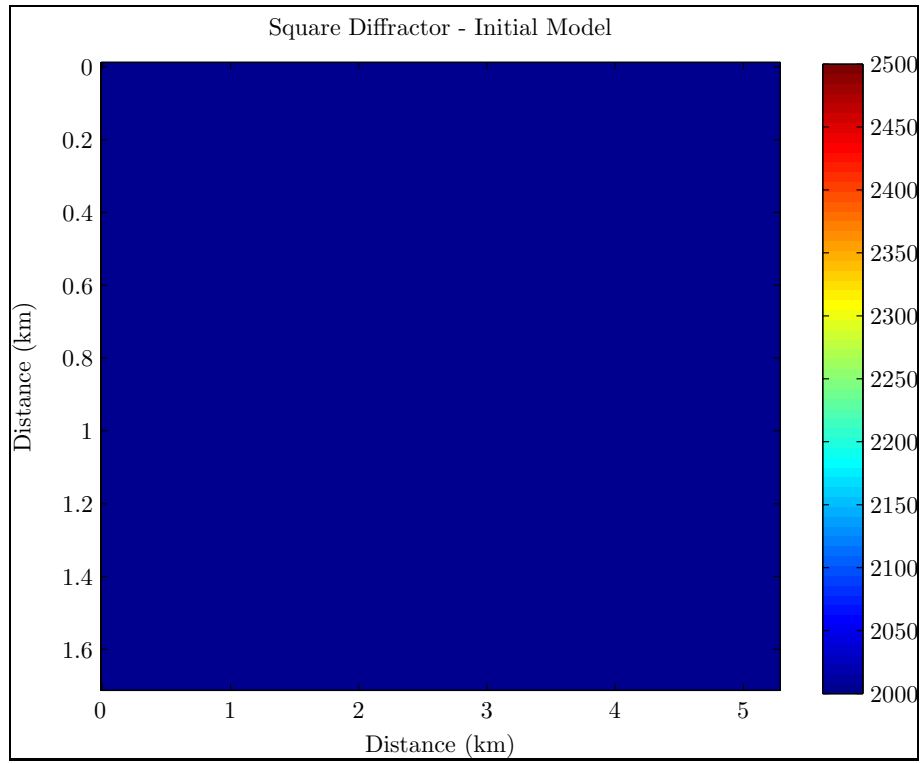


Figure 18: Initial model for the square Diffractor Model.

The result after the multi-scaled FWI can be seen in figure 19 where it is noticeable how the high velocity intrusion appears in the final model.

Even though the values are not exactly the same, it is a worthy result as there was not a single indication in the initial model that this high velocity zone was present. The highest values were present at second top layer of the intrusion with a maximum value of 2504 m/s. The lowest value inside the square was 2255 m/s, at its bottom corners.

Again, this result is quite good as it includes the intrusion on its own. The cost however is damaging the real values around the intrusion causing slight variations (some noticeable in the image, some not).

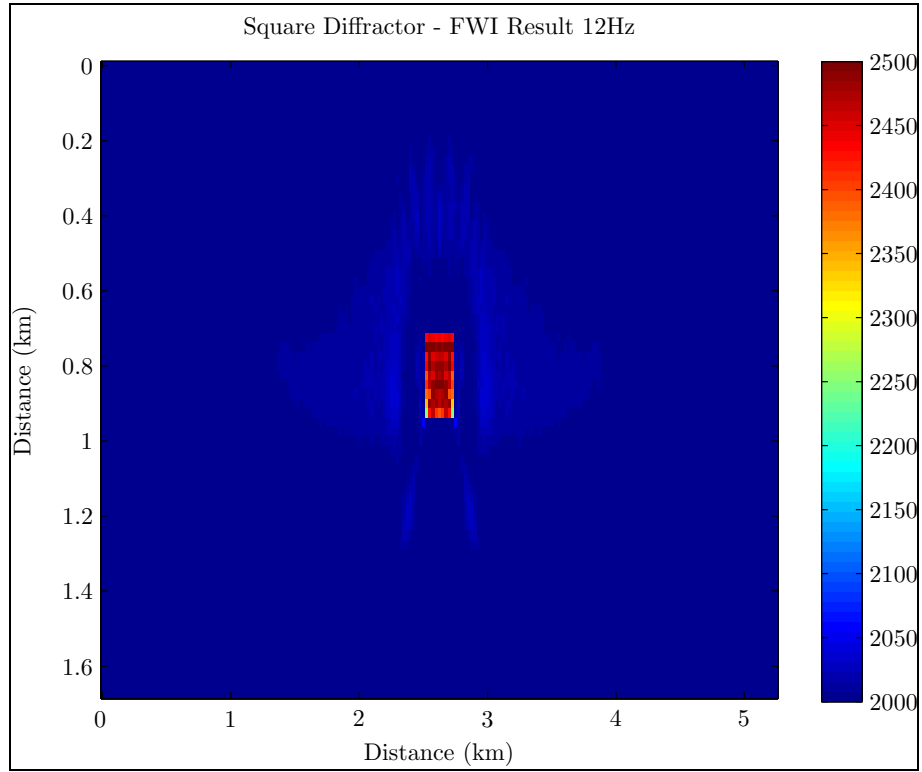


Figure 19: FWI solution of the square Diffractor Model.

The following model is the Marmousi, for this example the initial model (figure 20) was created averaging 50 times every point of the original model with up to the 5 furthest points in all directions. This was done with the purpose of having a good starting model to begin the inversion.

The result can be seen in figure 21. Due to the limit imposed by the numerical dispersion, this experiment was performed at just three scales in frequency, again with forty iterations each (after which the source frequency was changed). As the reader can observe, the solution contains the sharp definitions found in the original model (figure 5).

Again, the values are not exactly the same, but the gain in resolution turns out to be quite positive as the FWI helped define the model quite well from a visual point of view.

In contrast with the good definition of the top layers of the model, the bottom layers of the model present little improvement. This is due to the illumination of the model as it lowers close to the borders (especially at the bottom) of it while the absorbing boundaries also mitigate energy at the borders of the model.

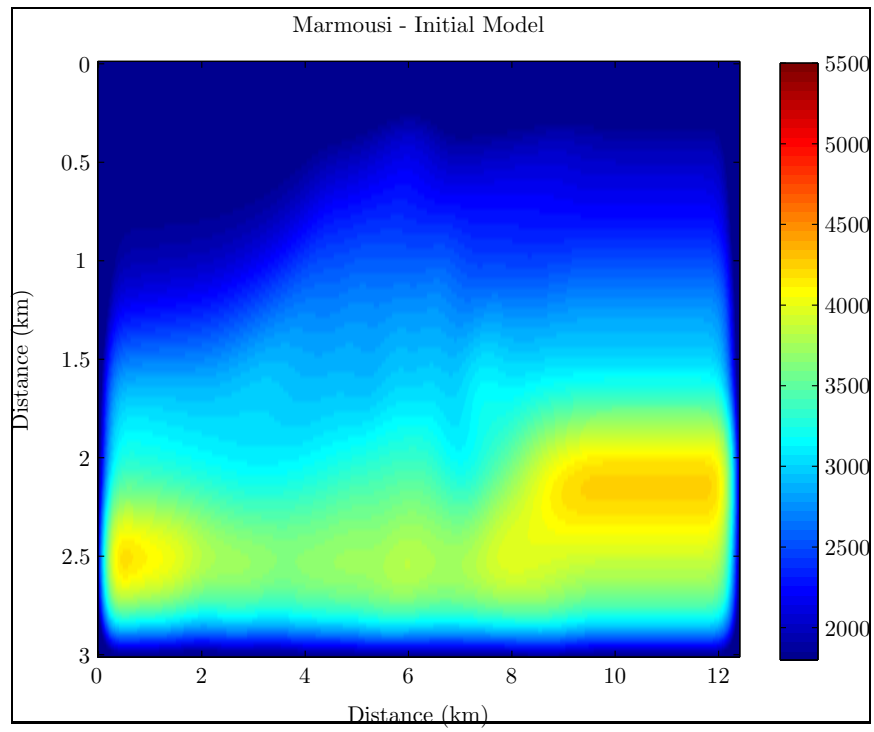


Figure 20: Initial model for the Marmousi model.

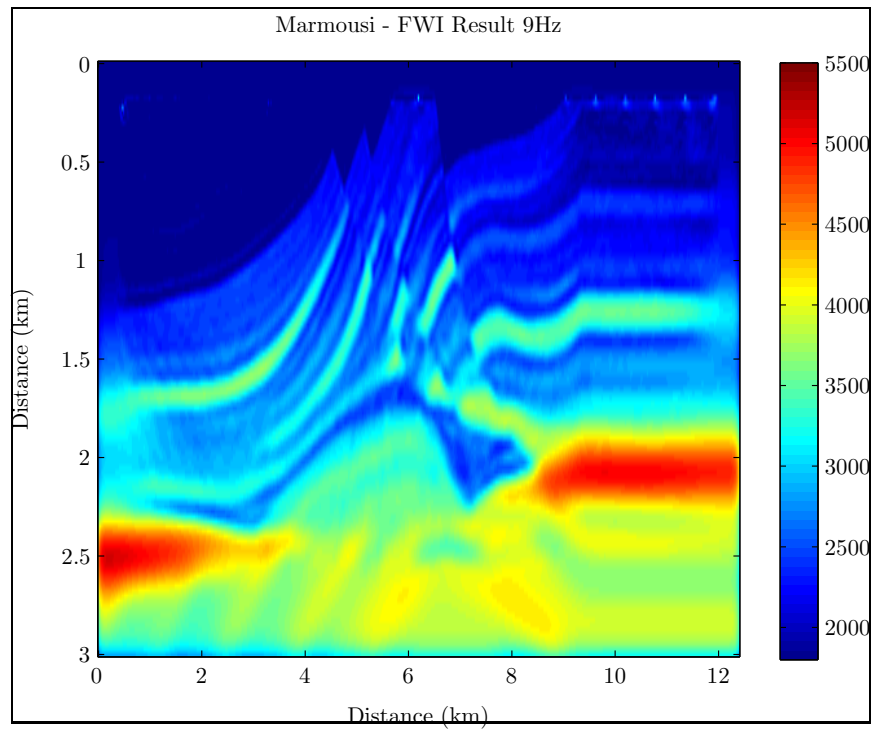


Figure 21: FWI solution of the Marmousi model.

So far we've seen how the FWI deals properly with two different models, each with its own degree of difficulty. The thing in common for these models is that we rely on the information that returns to the surface, but a question rises however when such data is not measured, that is, what happens when energy does not return to the surface?

The final test, the Sigsbee model, presented a real challenge as it did not allow for most of the energy to return to the surface, therefore the FWI did not have as much information as it had with the previous examples. The initial model was obtained using a global optimization technique, named particle swarm optimization (PSO) [23]. The task for this model was to include the intrusion in a similar fashion as with the square diffractor model. The variation in velocities this time, is much higher. The model is shown in figure 22.

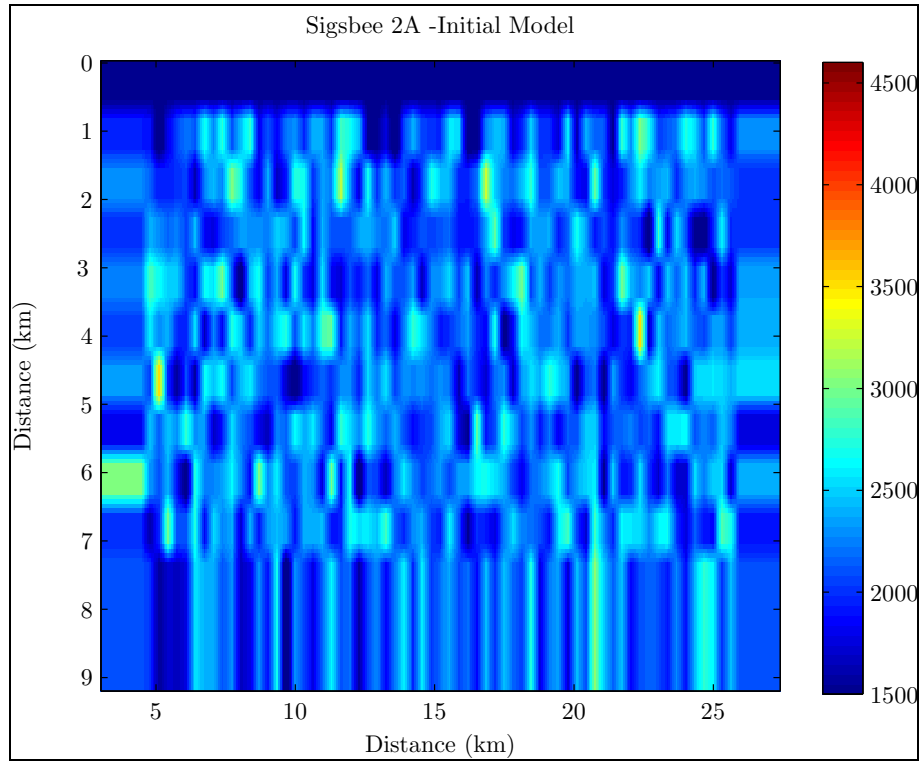


Figure 22: Initial model for the Sigsbee 2A model.

The final model is presented in figure 23, and, as stated before, is the model which the FWI had the most troubles with. The high velocity zone serves as a trap for the energy, and as the energy propagates through that red area, almost all of it is trapped, think of it as a waveguide. And as result, the information measured at the surface is not enough to update correctly the model.

It is noticeable however, how the background velocities were somewhat solved but nevertheless, the method fails to ultimately include the full high velocity intrusion. The areas below the intrusion seem unaffected even above the CPML area.

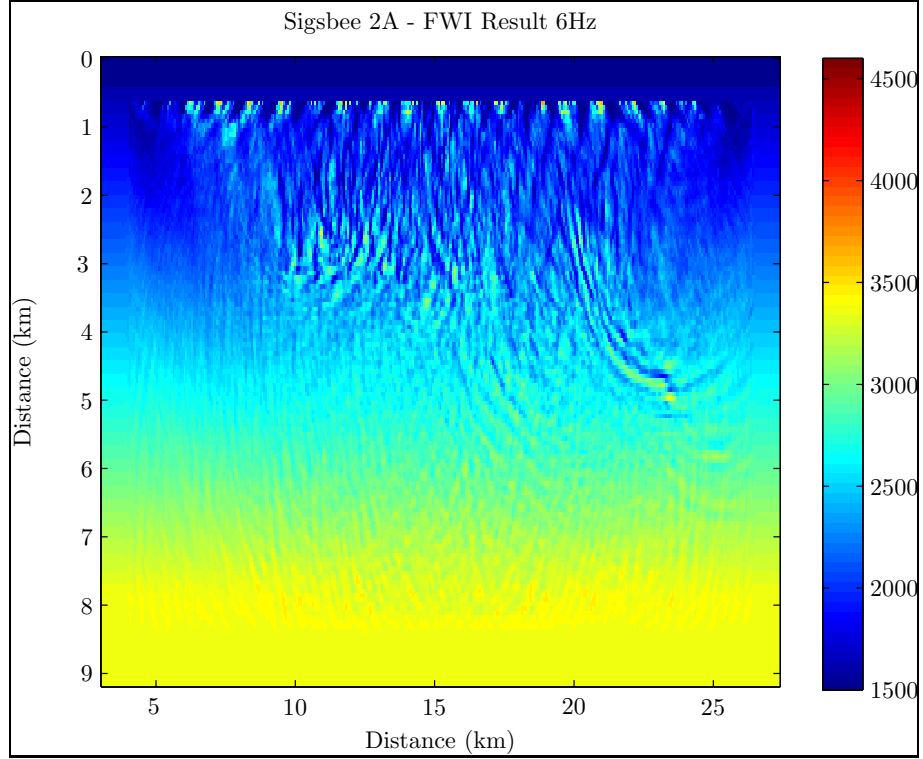


Figure 23: FWI solution of the Sigsbee 2A model.

The outcome of this model is produced by the physics of the model itself, one of the ideas tried to improve the model was to increase the source frequency, but even at 80Hz the result was practically the same.

To end this section, we have shown that the inverse problem has been solved. We have however, left a question unanswered, the question of how to compute the exact gradient and the Hessian matrices because so far we just discussed how to approximate the Hessian matrix from previous gradients. For that reason the next section is focused on that subject.

2.3 The Adjoint-State Method

The adjoint-state method (ASM) is a technique that arose under the context of optimal control, but it found its application into the world of geophysics thanks to Métivier in 2013 [24]. This method has proven useful in the context of Full Waveform Inversion (FWI) for obtaining the gradient and Hessian-vector products of the misfit function associated with the inversion problem.

The process involving the ASM can be summarized as follows:

- Take the Lagrangian of the misfit function to add constraints to the equation.
- Take the derivative of the Lagrangian.
- Define the adjoint equations and their variables.
- Solve the adjoint state equations.

The gradient is obtained by the first order ASM while the Hessian-vector products are obtained via the second order ASM.

2.3.1 First Order ASM: The Gradient

The FWI problem is formulated as

$$\Phi(\mathbf{m}) = \min_{\mathbf{m}} \frac{1}{2} \|RP(\mathbf{m}) - \mathbf{d}\|_2^2, \text{ subject to } LP(\mathbf{m}) = f_s, \quad (2.46)$$

where,

P : seismic wave field,

R : mapping of the wave field to the receiver locations,

$\mathbf{d}$: dataset associated to the source,

L : wave operator $L = \frac{1}{\mathbf{m}^2} \frac{\partial^2}{\partial t^2} - \nabla^2$,

f_s : seismic source.

In order to put the constrain in equation 2.46 we take the Lagrangian of it. The result is an extra term added to the equation in the form of an inner product of the constrain in canonical form with the lagrange multiplier λ as

$$\mathcal{L}(\mathbf{m}, P(\mathbf{m}), \lambda) = \frac{1}{2} \|RP(\mathbf{m}) - \mathbf{d}\|_2^2 + \langle LP(\mathbf{m}) - f_s, \lambda \rangle. \quad (2.47)$$

If we denote $\hat{P}(\mathbf{m})$ the solution of the forward problem, then the Lagrangian is the misfit function

$$\mathcal{L}(\mathbf{m}, \hat{P}(\mathbf{m}), \lambda) = \Phi(\mathbf{m}). \quad (2.48)$$

The following step is to take the derivatives of equation 2.48 with respect to $\mathbf{m}$, this becomes the derivative of the misfit function with respect to the model parameters, also known as *the gradient*

$$\frac{\partial}{\partial \mathbf{m}} \left(\mathcal{L}(\mathbf{m}, \hat{P}(\mathbf{m}), \lambda) \right) = \nabla \Phi(\mathbf{m}) = \mathbf{g}(\mathbf{m}). \quad (2.49)$$

The derivative can be obtained as

$$\frac{\partial}{\partial \mathbf{m}} (\mathcal{L}(\mathbf{m}, P(\mathbf{m}), \lambda)) = \frac{\partial}{\partial \mathbf{m}} \left(\frac{1}{2} \|RP(\mathbf{m}) - \mathbf{d}\|_2^2 + \langle LP(\mathbf{m}) - f_s, \lambda \rangle \right). \quad (2.50)$$

From equation 2.50, the derivative of the first term equals zero (traces are not a function of $\mathbf{m}$). This leaves us with only the second term (the inner product) to be derived as

$$\begin{aligned} \frac{\partial}{\partial \mathbf{m}} (\mathcal{L}(\mathbf{m}, P(\mathbf{m}), \lambda)) &= \frac{\partial}{\partial \mathbf{m}} (\langle LP(\mathbf{m}) - f_s, \lambda \rangle) \\ &= \left\langle \frac{\partial}{\partial \mathbf{m}} (LP(\mathbf{m})) - \frac{\partial}{\partial \mathbf{m}} (f_s), \lambda \right\rangle + \left\langle LP(\mathbf{m}) - f_s, \frac{\partial}{\partial \mathbf{m}} (\lambda) \right\rangle \\ &= \left\langle \frac{\partial}{\partial \mathbf{m}} (LP(\mathbf{m})), \lambda \right\rangle. \end{aligned} \quad (2.51)$$

Equation 2.49 can also be solved by the chain rule applied to multi-variable functions as:

$$\begin{aligned} \frac{\partial}{\partial \mathbf{m}} (\mathcal{L}(\mathbf{m}, P(\mathbf{m}), \lambda)) &= \frac{\partial}{\partial \mathbf{m}} (\mathcal{L}(\mathbf{m}, P(\mathbf{m}), \lambda)) \frac{\partial \mathbf{m}}{\partial \mathbf{m}} \\ &\quad + \frac{\partial}{\partial P(\mathbf{m})} (\mathcal{L}(\mathbf{m}, P(\mathbf{m}), \lambda)) \frac{\partial P(\mathbf{m})}{\partial \mathbf{m}} \\ &\quad + \frac{\partial}{\partial \lambda} (\mathcal{L}(\mathbf{m}, P(\mathbf{m}), \lambda)) \frac{\partial \lambda}{\partial \mathbf{m}}. \end{aligned} \quad (2.52)$$

Since λ is independent of $\mathbf{m}$, $\frac{\partial \lambda}{\partial \mathbf{m}} = 0$ and this simplifies equation 2.52 into:

$$\frac{\partial}{\partial \mathbf{m}} (\mathcal{L}(\mathbf{m}, P(\mathbf{m}), \lambda)) = \frac{\partial}{\partial \mathbf{m}} (\mathcal{L}(\mathbf{m}, P(\mathbf{m}), \lambda)) + \frac{\partial}{\partial P(\mathbf{m})} (\mathcal{L}(\mathbf{m}, P(\mathbf{m}), \lambda)) \frac{\partial P(\mathbf{m})}{\partial \mathbf{m}}. \quad (2.53)$$

This is an equation of the form $X = X + Y$, where $Y = 0$ and is precisely this second term that can be derived as

$$\frac{\partial}{\partial P(\mathbf{m})} (\mathcal{L}(\mathbf{m}, P(\mathbf{m}), \lambda)) = \frac{\partial}{\partial P(\mathbf{m})} \left(\frac{1}{2} \|RP(\mathbf{m}) - \mathbf{d}\|_2^2 + \langle LP(\mathbf{m}) - f_s, \lambda \rangle \right) = 0. \quad (2.54)$$

The derivative of the first term of equation 2.54 can be solved by using the definition of the ℓ_2^2 norm as $\|\cdot\|^2 = (\cdot)^T(\cdot)$, which leads to

$$\begin{aligned}\frac{\partial}{\partial P} \left(\frac{1}{2} \|RP(\mathbf{m}) - \mathbf{d}\|^2 \right) &= \frac{\partial}{\partial P} \left(\frac{1}{2} ((RP(\mathbf{m}) - \mathbf{d})^T (RP(\mathbf{m}) - \mathbf{d})) \right) \\ &= \frac{1}{2} \frac{\partial}{\partial P} ((P^T R^T - \mathbf{d}^T)(RP - \mathbf{d})) \\ &= \frac{1}{2} \frac{\partial}{\partial P} (P^T R^T RP - P^T R^T \mathbf{d} - \mathbf{d}^T RP + \mathbf{d}^T \mathbf{d}).\end{aligned}\tag{2.55}$$

The first three terms in equation 2.55 can be solved using the properties of matrix derivatives for two complex vectors $\mathbf{c}$ and $\mathbf{w}$ and matrix $\mathbf{Q}$ [25]:

1. For $g = \mathbf{w}^T \mathbf{Q} \mathbf{w}$, the derivative of g with respect to $\mathbf{w}$ is $\nabla_{\mathbf{w}} g = 2\mathbf{Q} \mathbf{w}$.
2. For $g = \mathbf{w}^T \mathbf{c}$, the derivative of g with respect to $\mathbf{w}$ is $\nabla_{\mathbf{w}} g = 2\mathbf{c}$.
3. For $g = \mathbf{c}^T \mathbf{w}$, the derivative of g with respect to $\mathbf{w}$ is $\nabla_{\mathbf{w}} g = \mathbf{0}$.

Using each property for each term respectively (the first term with the first property and so on) one can obtain the results of the derivatives. The last term equals zero as it does not depend on the variable P . This results in:

$$\begin{aligned}\frac{\partial}{\partial P} \left(\frac{1}{2} \|RP(\mathbf{m}) - \mathbf{d}\|^2 \right) &= \frac{1}{2} \left(\frac{\partial}{\partial P} (P^T R^T RP) - \frac{\partial}{\partial P} (P^T R^T \mathbf{d}) - \frac{\partial}{\partial P} (\mathbf{d}^T RP) + \frac{\partial}{\partial P} (\mathbf{d}^T \mathbf{d}) \right) \\ &= \frac{1}{2} (2R^T RP - 2R^T \mathbf{d} - 0 + 0) \\ &= R^T RP - R^T \mathbf{d} \\ &= R^T (RP - \mathbf{d}).\end{aligned}\tag{2.56}$$

The second term of equation 2.54 is derived as

$$\begin{aligned}\frac{\partial}{\partial P} (\langle LP(\mathbf{m}) - f_s, \lambda \rangle) &= \frac{\partial}{\partial P} (LP(\mathbf{m}) - f_s) \lambda + \frac{\partial \lambda}{\partial P} (LP(\mathbf{m}) - f_s) \\ &= \frac{\partial}{\partial P} (LP(\mathbf{m})) \lambda - \frac{\partial f_s}{\partial P} \lambda + 0 (LP(\mathbf{m}) - f_s) \\ &= \frac{\partial}{\partial P} (LP(\mathbf{m})) \lambda \\ &= \left(\frac{\partial L}{\partial P} P(\mathbf{m}) + \frac{\partial P(\mathbf{m})}{\partial P} L \right) \lambda \\ &= L \lambda.\end{aligned}\tag{2.57}$$

The results of equations 2.56 and 2.57 can be replaced into equation 2.54, and thus obtain

$$L\lambda = -R^T (RP - \mathbf{d}), \quad (2.58)$$

which has the structure of a wave equation, just as the restriction in equation 2.46. With it, we can compute the lagrange multiplier λ .

Now, from equation 2.54, we can define the state variable $\hat{\lambda}$ such that

$$\frac{\partial}{\partial P(\mathbf{m})} (\mathcal{L}(\mathbf{m}, P(\mathbf{m}), \hat{\lambda})) = 0. \quad (2.59)$$

To find $\hat{\lambda}$, we replace the solution $\hat{P}(\mathbf{m})$ into equation 2.58. The field $\hat{\lambda}$ is equivalent to what is known as the adjoint state variable. Replacing $\hat{P}(\mathbf{m})$ and $\hat{\lambda}$ into equation 2.51 gives us the gradient as presented in equation 2.49.

$$\mathbf{g}(\mathbf{m}) = \frac{\partial}{\partial \mathbf{m}} (\mathcal{L}(\mathbf{m}, \hat{P}(\mathbf{m}), \hat{\lambda})) = \left\langle \frac{\partial}{\partial \mathbf{m}} (L\hat{P}(\mathbf{m})), \hat{\lambda} \right\rangle. \quad (2.60)$$

An important remark is to take into account that the wave operator L depends on the model parameters $\mathbf{m}$, therefore, it has an internal derivative: $-2/\mathbf{m}^3$.

Gradients of the SVM's

Once the computation of the gradient is clear, the gradients of the SVM's can be shown. The gradients coming up, are the matrices representing the first derivatives of the objective function with respect to the model parameters ($\frac{\partial \Phi}{\partial \mathbf{m}}$).

In the case of the square diffractor model (figure 24), it is observed that the major amplitudes occur in the area where the high velocity square intrusion is located, however there are also variations located outside of it; this causes areas that have already the correct value (2000 m/s) to get unaccurately updated altering values of the output model. Fortunately this negative effect gets corrected by the iterative process of the inversion. This can be seen in the final velocity model (figure 19).

Is important to point out that as iterations go by and the model gets closer to the solution, the magnitude of the gradient is lower because of the reduction of the difference of the modeled and observed data ($RP - \mathbf{d}$), that is used to compute the adjoint field $\hat{\lambda}$ in equation 2.58. A lower difference produces a weaker field and the inner product that produced the gradient (equation 2.60) is therefore reduced.

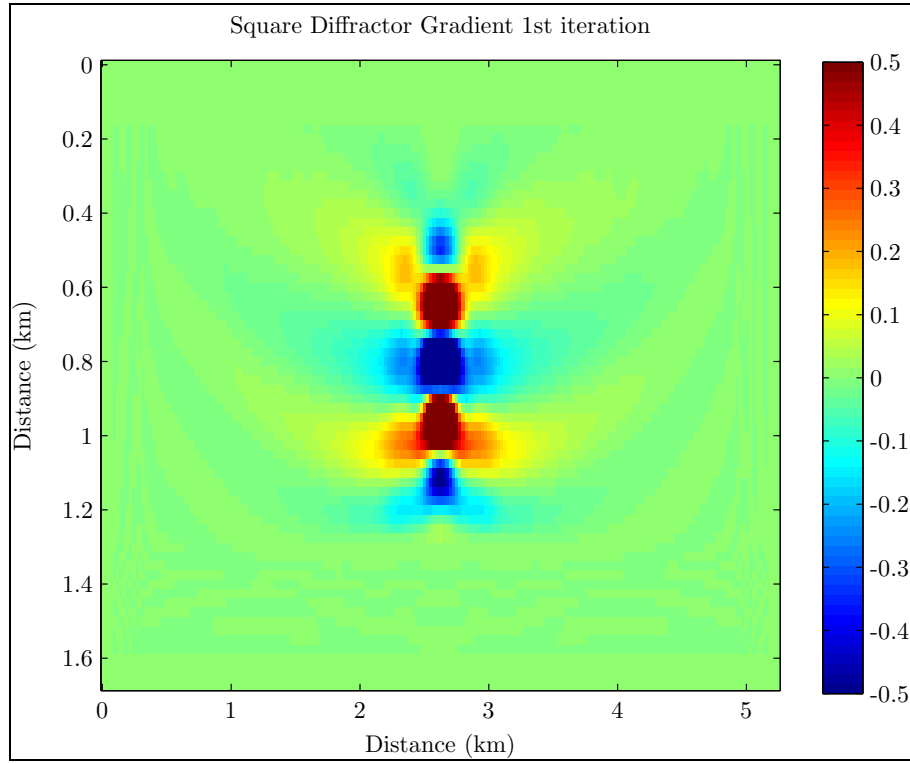


Figure 24: Gradient for the first iteration of the square Diffractor Model at 3Hz.

The effect of the multi-scale approach can be seen in the Marmousi example below. Figure 25 shows how the first stage of the FWI (3 Hz) starts solving the model without much sharpness in comparison with figure 26 that presents much sharper definition (edges).

Just as every step or iteration improved the solution, every stage of the multi-scale approach also improves the final result. This effect is produced thanks to the wavelength of the source, as its frequency rises, its wavelength lowers, making it possible to solve finer details in the model.

The effects of iterating are also present even if they are not shown directly, what occurs is the same as the previous example, the first iteration updates most of the model while the latter iterations refine more details. Overall, later iterations show lower energy than earlier iterations.

As previously seen for the Sigsbee model, the FWI dealt poorly with its update because of the lack of energy measured at the surface. As we know that what updates the model is the gradient, this “poor” gradient is what comes next.

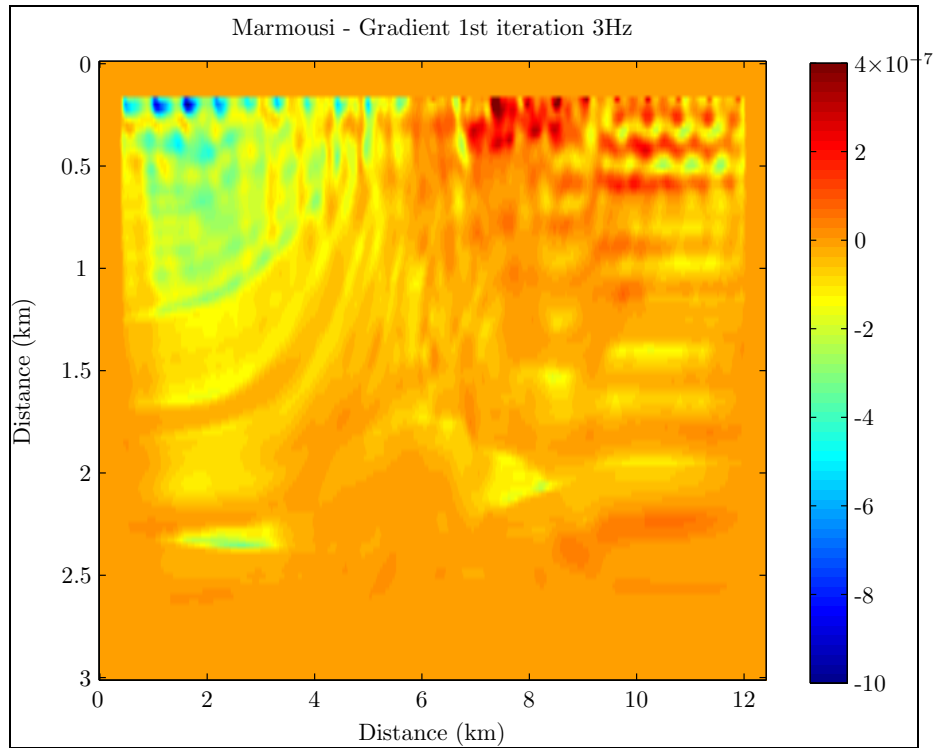


Figure 25: Gradient for the first iteration of the Marmousi model at 3Hz.

The outcome of a poorly resolved model caused by the lack of updates (read: *bad gradients*) presents a limitation not only of the FWI, but also of the seismic method altogether. Increasing the number of iterations or the frequency will not deal with the fact that there is not enough data recorded at the surface. As our main focus is reflection seismic, this is a major problem.

Figure 27 depicts the first gradient (first iteration at 3 Hz), very similar to the updated model, noticeable wave interactions at the zone where the velocity intrusion is located with the highest values close to the surface of the model.

This last gradient shows why the final model is not as good as expected. As data are not recorded and the gradient is obtained propagating the residual of the observed and modeled data, the residual will not contain much information of the propagation, and will not add much to the gradient and hence, the update of the model.

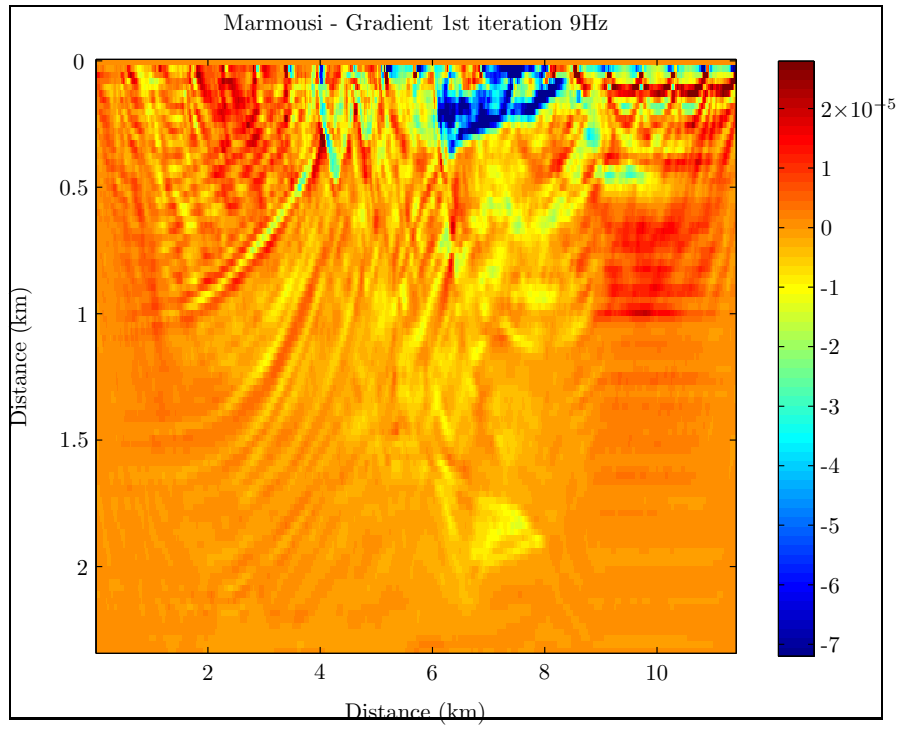


Figure 26: Gradient for the first iteration of the Marmousi model at 9Hz.

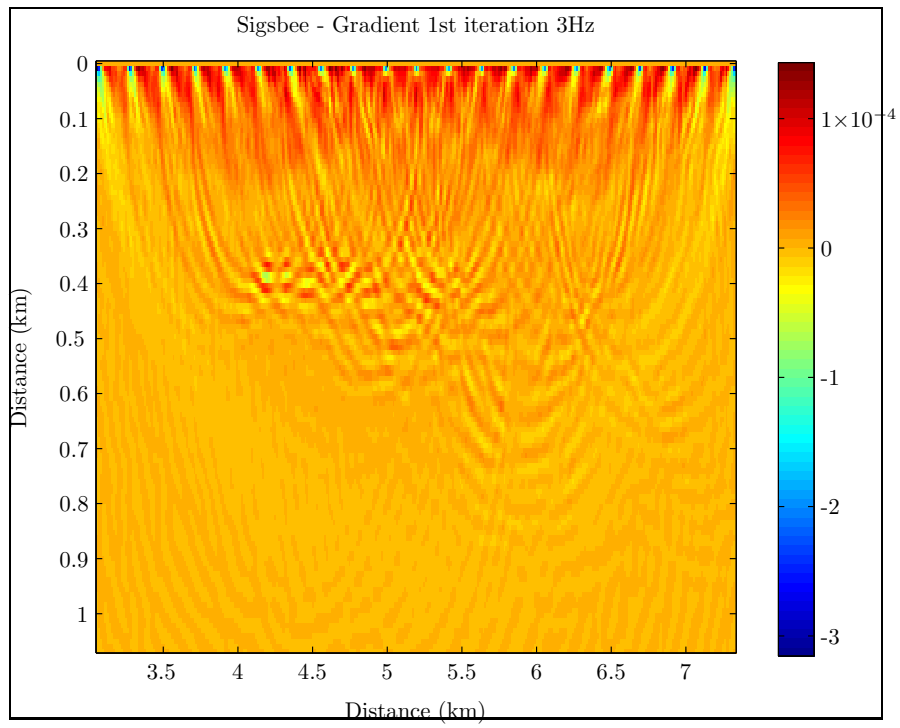


Figure 27: Gradient for the first iteration of the Sigsbee 2A model at 3Hz.

2.3.2 Second Order ASM: The Hessian

Unlike the first-order ASM, the second-order ASM does not give the full Hessian matrix, but the multiplication of the matrix by a given vector ν . For this, let h_ν be a functional defined as the inner product

$$h_\nu = \langle g(\mathbf{m}), \nu \rangle. \quad (2.61)$$

The gradient of the functional h_ν is therefore

$$\nabla h_\nu = \mathbf{H}(\mathbf{m})\nu. \quad (2.62)$$

According to [26], the functional in equation 2.61 can be rewritten as

$$h_\nu(\mathbf{m}) = \langle R^T(RP(\mathbf{m}) - \mathbf{d}), J(\mathbf{m})\nu \rangle, \quad (2.63)$$

where $J(\mathbf{m})$ is the Jacobian matrix.

Deriving the forward problem (wave equation) with respect to $\mathbf{m}$ in the directions ν_j

$$\frac{\partial L}{\partial \mathbf{m}_j}(P(\mathbf{m})\nu_j) + L \frac{\partial}{\partial \mathbf{m}_j}(P(\mathbf{m})\nu_j) = 0, \quad j = 1, 2, \dots, m, \quad (2.64)$$

and summing on j gives us

$$\sum_{j=1}^m \frac{\partial L}{\partial \mathbf{m}_j}(P(\mathbf{m})\nu_j) + \sum_{j=1}^m L \frac{\partial}{\partial \mathbf{m}_j}(P(\mathbf{m})\nu_j) = 0. \quad (2.65)$$

Equation 2.65 can be rewritten as

$$\sum_{j=1}^m L \frac{\partial}{\partial \mathbf{m}_j}(P(\mathbf{m})\nu_j) = - \sum_{j=1}^m \frac{\partial L}{\partial \mathbf{m}_j}(P(\mathbf{m})\nu_j), \quad (2.66)$$

and is equivalent to

$$L\alpha_\nu(\mathbf{m}) = \varphi_\nu, \quad (2.67)$$

which resembles a wave equation, where $\alpha_\nu(\mathbf{m}) = J(\mathbf{m})\nu = \sum_{j=1}^m \frac{\partial}{\partial \mathbf{m}_j}(P(\mathbf{m})\nu_j)$ and $\varphi_\nu = - \sum_{j=1}^m \frac{\partial L}{\partial \mathbf{m}_j}(P(\mathbf{m})\nu_j)$.

Now we may consider the constrained minimization problem:

$$\min_p \langle R^T(RP(\mathbf{m}) - \mathbf{d}), \alpha \rangle, \quad \text{subject to } LP(\mathbf{m}) = f_s \text{ and } L\alpha_\nu(\mathbf{m}) = \varphi_\nu. \quad (2.68)$$

As in equation 2.47, the Lagrangian associated with this problem is

$$\mathcal{L}_\nu(\mathbf{m}, P(\mathbf{m}), \alpha_\nu(\mathbf{m}), \lambda, \mu) = \langle R^T(RP(\mathbf{m}) - \mathbf{d}), \alpha_\nu(\mathbf{m}) \rangle + \langle LP(\mathbf{m}) - f_s, \lambda \rangle + \langle L\alpha_\nu(\mathbf{m}) - \varphi_\nu, \mu \rangle, \quad (2.69)$$

where λ and μ are the two lagrange multipliers.

Let us denote $\hat{P}(\mathbf{m})$ and $\hat{\alpha}_\nu(\mathbf{m})$ the solutions to the wave equations and we get

$$\mathcal{L}_\nu(\mathbf{m}, \hat{P}(\mathbf{m}), \hat{\alpha}_\nu(\mathbf{m}), \lambda, \mu) = h_\nu(\mathbf{m}), \quad (2.70)$$

then take the derivative with respect to $\mathbf{m}$, which is, the gradient of the functional of equation 2.62

$$\frac{\partial}{\partial \mathbf{m}} \left(\mathcal{L}_\nu(\mathbf{m}, \hat{P}(\mathbf{m}), \hat{\alpha}_\nu(\mathbf{m}), \lambda, \mu) \right) = \nabla h_\nu(\mathbf{m}) = \mathbf{H}(\mathbf{m})\nu. \quad (2.71)$$

Following a similar process of 2.50, the derivative can be found to be

$$\begin{aligned} \frac{\partial}{\partial \mathbf{m}} (\mathcal{L}(\mathbf{m}, P(\mathbf{m}), \alpha_\nu(\mathbf{m}), \lambda, \mu)) &= \left\langle \frac{\partial \mathcal{L}}{\partial \mathbf{m}} P(\mathbf{m}), \mu \right\rangle + \left\langle \frac{\partial \mathcal{L}}{\partial \mathbf{m}} \alpha, \lambda \right\rangle \\ &+ \sum_{j=1}^m \nu_j \left(\left\langle \frac{\partial}{\partial \mathbf{m}_j} \left(\frac{\partial \mathcal{L}}{\partial \mathbf{m}} P(\mathbf{m}) \right), \lambda \right\rangle \right). \end{aligned} \quad (2.72)$$

Developing equation 2.71 with the chain rule we obtain

$$\begin{aligned} \frac{\partial}{\partial \mathbf{m}} (\mathcal{L}(\mathbf{m}, P(\mathbf{m}), \alpha_\nu(\mathbf{m}), \lambda, \mu)) &= \frac{\partial}{\partial \mathbf{m}} (\mathcal{L}(\mathbf{m}, P(\mathbf{m}), \alpha_\nu(\mathbf{m}), \lambda, \mu)) \frac{\partial \mathbf{m}}{\partial \mathbf{m}} \\ &+ \frac{\partial}{\partial P(\mathbf{m})} (\mathcal{L}(\mathbf{m}, P(\mathbf{m}), \alpha_\nu(\mathbf{m}), \lambda, \mu)) \frac{\partial P(\mathbf{m})}{\partial \mathbf{m}} \\ &+ \frac{\partial}{\partial \alpha_\nu(\mathbf{m})} (\mathcal{L}(\mathbf{m}, P(\mathbf{m}), \alpha_\nu(\mathbf{m}), \lambda, \mu)) \frac{\partial \alpha_\nu(\mathbf{m})}{\partial \mathbf{m}} \\ &+ \frac{\partial}{\partial \lambda} (\mathcal{L}(\mathbf{m}, P(\mathbf{m}), \alpha_\nu(\mathbf{m}), \lambda, \mu)) \frac{\partial \lambda}{\partial \mathbf{m}} \\ &+ \frac{\partial}{\partial \mu} (\mathcal{L}(\mathbf{m}, P(\mathbf{m}), \alpha_\nu(\mathbf{m}), \lambda, \mu)) \frac{\partial \mu}{\partial \mathbf{m}}, \end{aligned} \quad (2.73)$$

where again λ and μ are independent of $\mathbf{m}$. Therefore $\frac{\partial \lambda}{\partial \mathbf{m}} = 0$ and $\frac{\partial \mu}{\partial \mathbf{m}} = 0$, yielding

$$\begin{aligned} \frac{\partial}{\partial \mathbf{m}} (\mathcal{L}(\mathbf{m}, P(\mathbf{m}), \alpha_\nu(\mathbf{m}), \lambda, \mu)) &= \frac{\partial}{\partial \mathbf{m}} (\mathcal{L}(\mathbf{m}, P(\mathbf{m}), \alpha_\nu(\mathbf{m}), \lambda, \mu)) \\ &+ \frac{\partial}{\partial P(\mathbf{m})} (\mathcal{L}(\mathbf{m}, P(\mathbf{m}), \alpha_\nu(\mathbf{m}), \lambda, \mu)) \frac{\partial P(\mathbf{m})}{\partial \mathbf{m}} \\ &+ \frac{\partial}{\partial \alpha_\nu(\mathbf{m})} (\mathcal{L}(\mathbf{m}, P(\mathbf{m}), \alpha_\nu(\mathbf{m}), \lambda, \mu)) \frac{\partial \alpha_\nu(\mathbf{m})}{\partial \mathbf{m}} \end{aligned} \quad (2.74)$$

and we get an equation of the form $X = X + Y + Z$, where $Y = 0$ and $Z = 0$.

The next step is to define the state variables $\hat{\lambda}$ and $\hat{\mu}$ such that

$$\frac{\partial}{\partial \hat{P}(\mathbf{m})} \left(\mathcal{L}_\nu(\mathbf{m}, P(\mathbf{m}), \alpha_\nu(\mathbf{m}), \hat{\lambda}, \mu) \right) = 0, \quad (2.75)$$

$$\frac{\partial}{\partial \hat{\alpha}_\nu(\mathbf{m})} (\mathcal{L}_\nu(\mathbf{m}, P(\mathbf{m}), \alpha_\nu(\mathbf{m}), \lambda, \hat{\mu})) = 0. \quad (2.76)$$

After a similar process to the first order adjoint state method (equations 2.52 through 2.58), the following equations can be obtained from 2.75 and 2.76,

$$L\lambda = -R^T(RP(\mathbf{m}) - \mathbf{d}), \quad (2.77)$$

$$L\mu = -R^T R\alpha_\nu(\mathbf{m}) - \sum_{j=1}^m \nu^T \frac{\partial L}{\partial \mathbf{m}_j} \lambda. \quad (2.78)$$

Equation 2.77 is the same equation as 2.58, while equation 2.78 turns out to resemble yet another wave equation using the terms at the right side of the equation as the source.

Then again, to find $\hat{\lambda}$ we must replace the solution $\hat{P}(\mathbf{m})$ into equation 2.77, then replace the solution $\hat{\alpha}_\nu(\mathbf{m})$ and $\hat{\lambda}$ into equation 2.78 to obtain $\hat{\mu}$. Equations 2.79 and 2.80 result from this process as

$$L\hat{\lambda} = -R^T(R\hat{P}(\mathbf{m}) - \mathbf{d}), \quad (2.79)$$

$$L\hat{\mu} = -R^T R\hat{\alpha}_\nu(\mathbf{m}) - \sum_{j=1}^m \nu^T \frac{\partial L}{\partial \mathbf{m}_j} \hat{\lambda}. \quad (2.80)$$

Finally, after solving four wave equations: two forward fields for the computation of $\hat{P}(\mathbf{m})$ and $\hat{\alpha}$ and two adjoint fields for the computation of $\hat{\lambda}$ and $\hat{\mu}$, the derivative of the lagrangian in equation 2.72 becomes gradient of the functional h_ν (equation 2.71) which is of course the Hessian-vector product we were looking for (equation 2.63). The result for the Hessian-vector product is

$$\mathbf{H}(\mathbf{m})\nu = \left\langle \frac{\partial L}{\partial \mathbf{m}} \hat{P}(\mathbf{m}), \hat{\mu} \right\rangle + \left\langle \frac{\partial L}{\partial \mathbf{m}} \hat{\alpha}_\nu(\mathbf{m}), \hat{\lambda} \right\rangle + \sum_{j=1}^m \nu_j \left(\left\langle \frac{\partial}{\partial \mathbf{m}_j} \left(\frac{\partial L}{\partial \mathbf{m}} \hat{P}(\mathbf{m}) \right), \hat{\lambda} \right\rangle \right). \quad (2.81)$$

What is important to remark here is that the ASM is proposed initially from the optimization point of view but as we developed its physical sense started to flowrish. We can get a glimpse of this physical meaning as the langrange multipliers turned into state variables, but then they became what are known as adjoint variables to finally be computed as fields obtained from (wave-like) operators equivalent to the adjoint operator in the real domain.

Hessian matrices of the SVM's

There are many things to discuss about the Hessian matrix but due to the size of the it, looking at the full matrix might not show the importance of the information contained in it. However looking at it as is, might as well reveal interesting aspects of this matrix.

As the reader must be used to by now, we will start with the Hessian matrix of our first model, the square diffractor, that can be seen in figure 28.

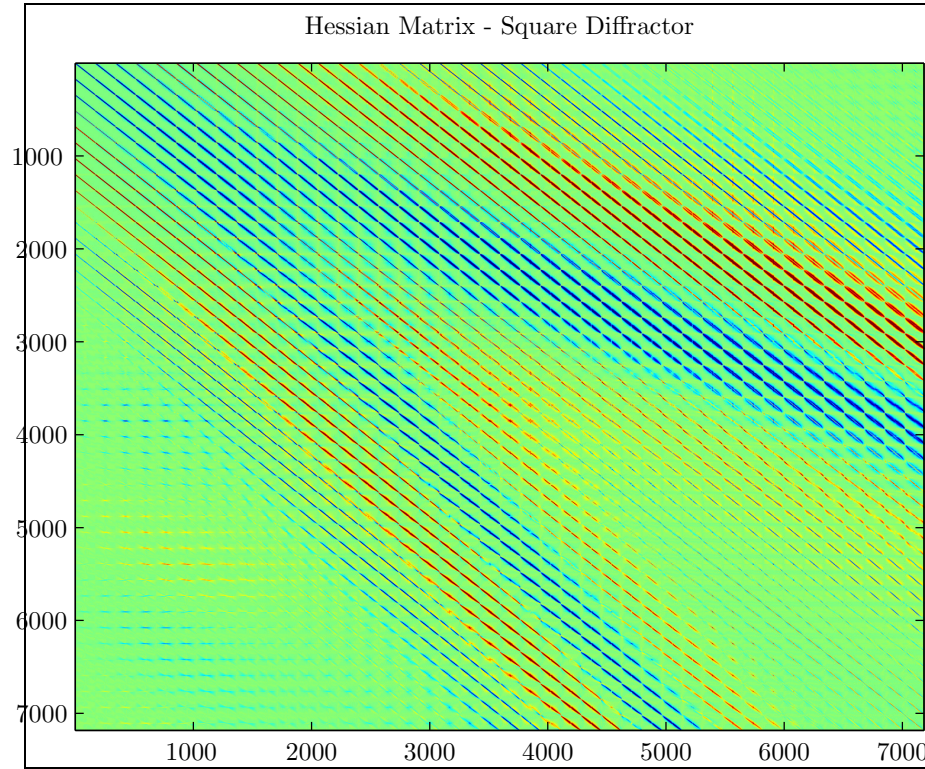


Figure 28: Hessian matrix of the square Diffractor Model.

This matrix contains the information about the second derivatives of the objective function with respect to the model parameters, but let us not dwell into that and, in turn discuss some of the properties the matrix such as: size, definiteness and symmetry.

First of all, the size of the Hessian matrix. It depends exclusively on the model, forming always a square matrix of $M \times N$ by $M \times N$ elements, given an M by N velocity model. As we are computing the Hessian matrix, one column at a time, this results in a column of $M \times N$ elements and a total of $M \times N$ columns, hence the size of the Hessian matrix.

The matrix in figure 28 has 7182×7182 elements, which correspond to a 171×42 model. The difference in size in comparison with the original size is that the absorbing boundaries have been taken off.

Another property is that the Hessian matrix of a convex function is positive semi-definite. With figure 16 it was explained that this condition is not met and therefore the resulting Hessian matrices are not positive semi-definite.

The final property is related to the differentiation operator. As we deal with mixed derivatives, we have always known that according to Schwarz's theorem [27], for continuous functions the order of differentiation does not matter. This in theory would yield to a symmetric Hessian matrix. Again, this statement falls short for two reasons: discretization and geometry.

The first reason, discretization plays a major role as we are discretizing the derivative operator into a finite difference operator, therefore, we no longer have the necessary continuity for the statement to hold.

The second reason is the geometry of the experiment itself, as seen in [28] locating both sources and receivers at the surface of the model will not produce results as good as locating sources on the top and left layers and receivers at the bottom and right layers. This practice may help the symmetry of the matrix, but as the focus of this work is the exploration of the subsurface of the Earth from the surface, we restrained ourselves of including unrealistic practices such as placing underground receivers.

Now, we may dwell into the things that the Hessian matrix tells us.

As stated above, the Hessian matrix is the matrix associated with the second derivatives of the objective function with respect to the model parameters ($\frac{\partial^2 \Phi}{\partial \mathbf{m}^2}$). The second derivatives have the information about the curvature of the objective function and for that reason is that initially the Hessian matrix is proposed as updating factor alongside the gradient (in fact, is not the Hessian matrix itself but its inverse). The problem of this approach is, as usual, the computation cost of calculating such a matrix. The large amount of terms in the matrix make it impractical to calculate the matrix at each step or iteration, that is why the FWI uses only the gradient and compensate the lack of Hessian matrix with an scale factor.

Problems aside, the Hessian matrix contains a lot of information and has other important jobs to accomplish. We will finish this chapter talking about some of the information that the matrix contains while we leave chapter 3 to talk about its applications.

Continuing with the square diffractor model, we have already seen the full Hessian matrix, which is striking in size, but does not tell that much visually. We need therefore, a way to extract that information, which is not apparent to the eye.

The reason behind the size of the matrix obeys that for each point in the velocity model, a column of the Hessian matrix is generated, and each column has the same size of the model. Knowing that, we can take every column of the Hessian matrix and associate it with a point in the model.

What we see in a column of the Hessian matrix (rearranged as a M by N matrix in figure 29) is the interaction of a perturbation at one point (i,j) with the perturbation in all the points of the model. In this case, the perturbation is located at the point $(20,30)$ or in km $(0.637,1.229)$.

A second strategy is to extract the diagonal of the Hessian matrix, and again, rearrange it in matrix form. By doing so, we are looking at the second derivatives with respect of each model parameter as depicted in figure 30.

Figure 30 visually resembles to figure 24 in terms of the locations of the highest values, the difference lies in that the former represents the second derivatives while the latter, the first derivatives. In terms of magnitude, the values in the Hessian diagonal are by far smaller than the values in the gradient (around 3 orders of magnitude lower).

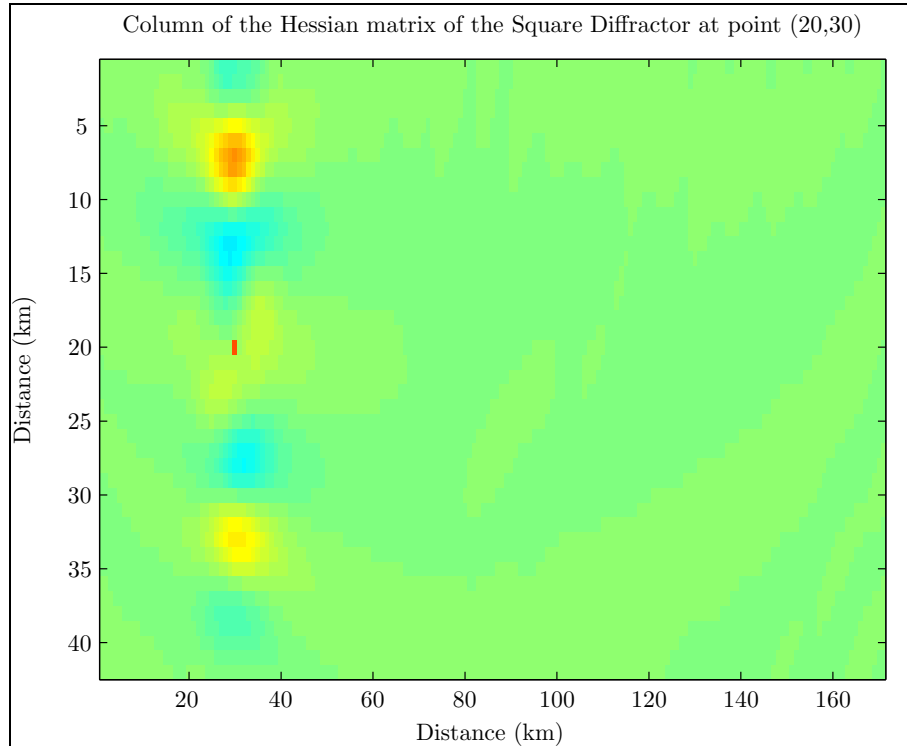


Figure 29: A column of the Hessian matrix (reshaped as matrix) of the square Diffractor Model.

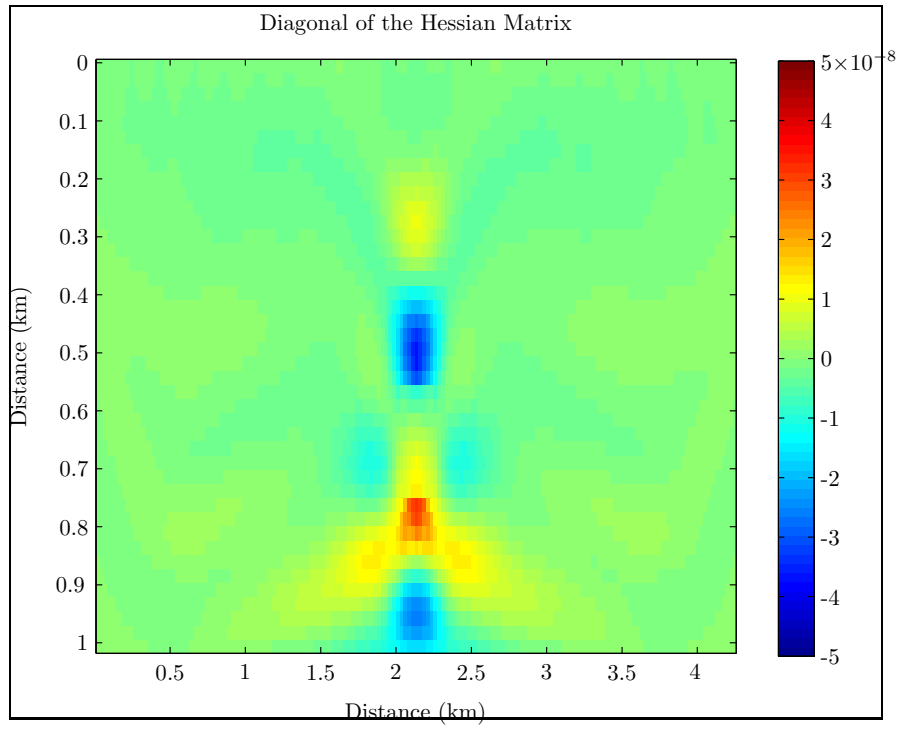


Figure 30: Diagonal of the Hessian matrix (reshaped as matrix) of the square Diffractor Model.

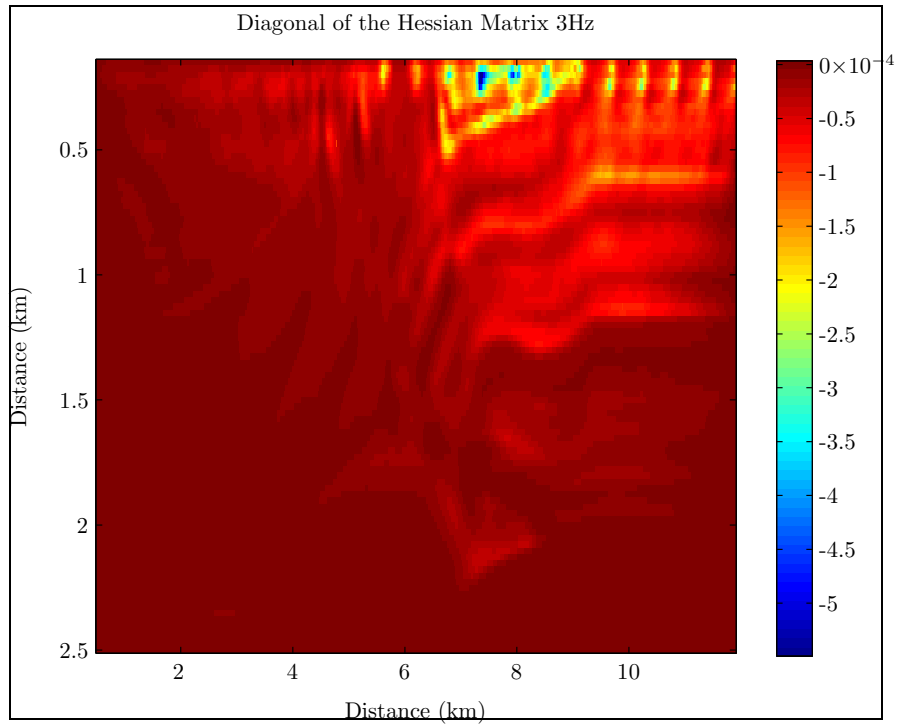


Figure 31: Diagonal of the Hessian matrix (reshaped as matrix) of the Marmousi Model at 3Hz.

In figure 31 we see the diagonal elements of the Hessian matrix of the Marmousi model at 3Hz, the full matrix is not shown because it does not tell much as an image, just like the one seen in figure 28.

Figure 32 shows the diagonal elements of the Hessian at 9 Hz. These two figures with their respective gradients represent the second derivatives with respect to each model parameter.

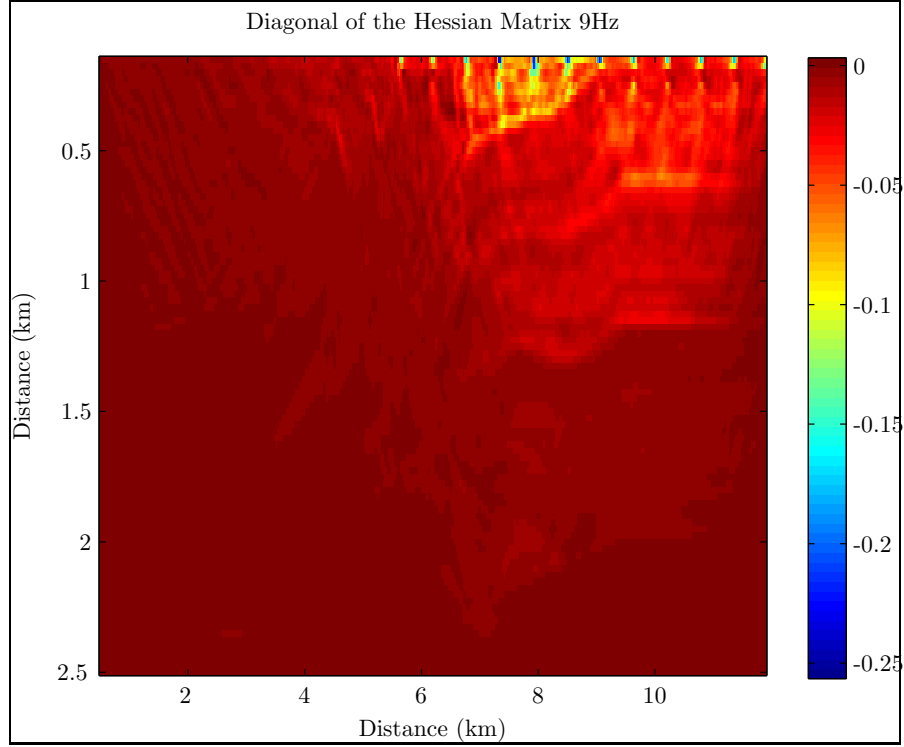


Figure 32: Diagonal of the Hessian matrix (reshaped as matrix) of the Marmousi Model at 9Hz.

Finally, the Sigsbee model, which have been testing the capabilities of the FWI from the beginning, now offers a second difficulty, its size. In its original form, the Sigsbee model is composed of a grid of 2133×1201 points, and despite not being that big, a model that size would generate a massive Hessian matrix. To help realize how big this matrix is, table 5 resumes the sizes of the models, their respective Hessian matrix and the amount of memory of the Hessian matrices.

Model	Model Size	Hessian Size	Required Memory
Diffraction	211×68	7182×7182	196.77 MB
Marmousi	497×121	43415×43415	7.02 GB
Sigsbee (Original)	2133×1201	2459275×2459275	22 TB
Sigsbee (Re-sampled)	375×141	52875×52875	5.5 GB

Table 5: Dimensions of the Hessian matrices.

The unmanageable size of the Hessian matrix of the original Sigsbee model forced us to down-sample it to a size manageable in terms of memory.

Figure 33 presents the diagonal elements of the Hessian matrix of the Sigsbee model. As expected, not much can be seen from the figure itself and we are left to wait, wheter or not this Hessian matrix will provide enough information for the quantification of the uncertainties. This will be done in chapter 3.

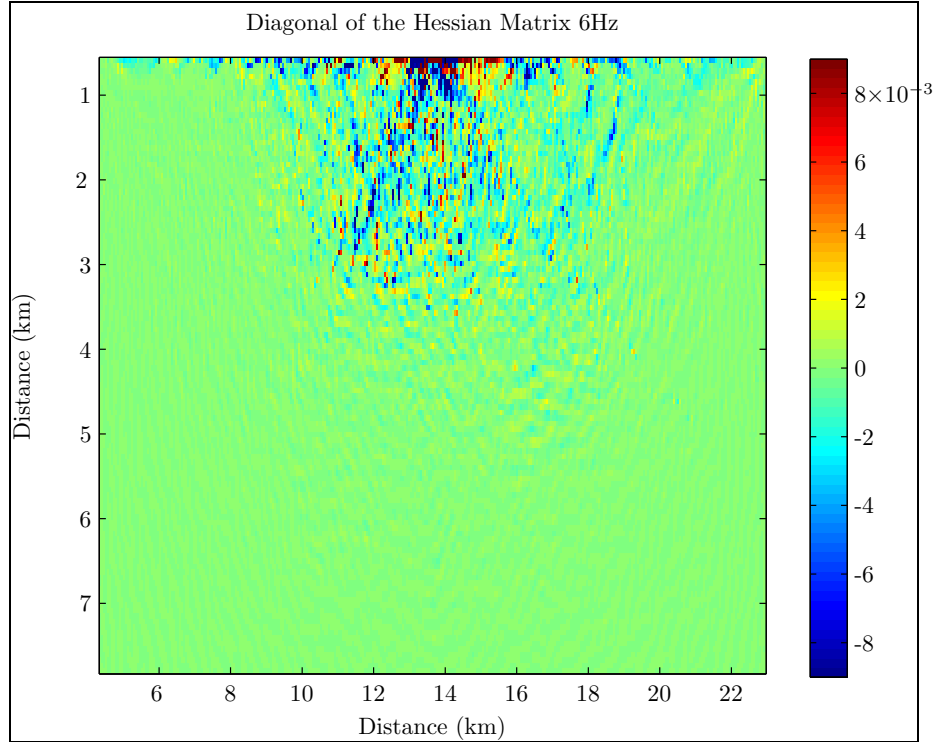


Figure 33: Diagonal of the Hessian matrix (reshaped as matrix) of the Sigsbee Model at 6Hz.

2.4 Discussion

To finish the chapter let us discuss some of the important aspects of each section, beginning with the forward problem. As we have seen, modeling the wave equation begins with the proper selection of the wave equation itself. As stated before, we have done all our tests using the acoustic and isotropic two dimensional wave equation with wave equation, the simplest case. From the physical point of view, this equation does not take into account several of the phenomena present in an actual seismic experiment but is commonly used in the Oil & Gas Industry.

Improving the equation (adding more physical phenomena) leads to an increase in the computational load, not going too far, adding the third spatial dimension turns our seismic models (planes) into cubes, increasing the size of the problem in the number of points for that discretized grid dimension which becomes not only more storage to be needed, but also more time to compute results.

The former is the easiest improvement to the equation that would in theory add only one additional term to equation 2.2, but adding for example anisotropy or elastic media requires changing drastically our equation, often turning them into a set of equations, rather than one.

One final, and plausible improvement to the equation is perhaps the addition of the density ($\rho(x, z)$) to the equation, this addition would only double our model parameters.

Our approach of *attacking* the wave equation is also subject to change, current works such as [29] and [30] does not solve the wave equation using the FDTD method, they in turn, use pseudo-spectral methods and rapid expansion methods respectively, each with its own advantages. A good reason for using FDTD is that it becomes a straightforward approach that deals quite good with the modeling of the phenomenon given that one takes into account all considerations for stability, dispersion, boundaries, etc. Dispersion becomes an important factor for dealing with real data due to the fact that (contrary to our experiments) observed data does not contain dispersion, while modeled data, almost always, contains a percentage of it and if it is too high, could yield to artifacts generated by the artificial difference in the traces.

Following the forward problems, comes the inverse problem. From it, we must point out the importance of the initial point from which the inversion starts. This local optimization problem is biased by the initial model, meaning that different initial models will yield to different final models. The question therefore, becomes: “How can I select a proper initial point to start the inversion with?”. One approach is to solve a global optimization problem (which often involves performing several forward problems) to get to the neighborhood of the global maximum, and from that point, start the local optimization problem (FWI) in order to be able to converge to that global maximum. Even though we will not dwell deep into this topic, the avid reader is recommended to read [31] and [23] for more information.

One crucial aspect to discuss of the inverse problem, that we have left out so far, is the reason behind the approximation of the inverse of the Hessian matrix using the L-BFGS method and not the second order adjoint state method. To cut the long story short: we used both methods. The former was used for updating the model as seen throughout this chapter, while the latter was used for quantifying

uncertainties in the final model after FWI as will be seen in the following chapter. It is however, very important to define why one method or the other is used for one purpose or the other.

The L-BFGS method is used for computing an approximation of the inverse of the Hessian matrix projected over the gradient. This approximation comes from the difference between gradients and models of previous iterations and its purpose is to add information about curvature of the objective function. In other words, this method, does not compute the full size matrix which saves resources and time at the cost of not getting the exact matrix. Turns out that this approximation is good enough to add the curvature information needed for the convergence of the inverse problem and therefore, is not necessary to compute the full and exact matrix. This method requires 40 iterations or less.

Contrary to the L-BFGS method, the SOASM computes the exact full Hessian matrix, which can be costly and, as seen with the Sigsbee model, sometimes implausible either by the amount of memory needed to allocate or by the amount of time needed to wait for the result to come out. Current graphic processing units (GPU's), such as the Tesla K40 (which was used for all tests), have limited resources in the same order of the size of some of our Hessian matrices (the Nvidia Tesla K40 has 12GB of RAM) making it necessary to deal with memory more efficiently than just allocating every matrix in RAM. Additionally, we know in advance that the adjoint state method computes one column of the Hessian matrix at a time, making it possible to obtain the matrix column by column without exceeding the total size of memory except when the model is too big. This approach however has to take into account that the matrix has yet to be inverted, which can become an issue.

The interest of computing such a big matrix is to account for uncertainties in the solution given by the FWI, as it will be seen in chapter 3 in order to appraise or asses solution models, the most exact Hessian possible is required. Given that the process of quantification is performed not at every iteration (as the L-BFGS method), but only at the final model, one can afford to spend the time necessary for the matrix to be obtained and inverted. This time is highly dependent on the experiment parameters (model size, time recorded, number of sources, etc.) but to give the reader an idea, table 6 presents the times used for the calculation of the Hessian matrix using the SOASM for only a single source of one frequency and the rest of the parameters of table 4.

Model	Model Size	Hessian Size	Required Memory	Required Time
Diffractor	211×68	7182×7182	196.77 MB	12 minutes
Marmousi	497×121	43415×43415	7.02 GB	8 hours
Sigsbee (Re-sampled)	375×141	52875×52875	5.5 GB	17 hours

Table 6: Resources used by the GPU when computing Hessian matrices.

Note that not always the most memory-consuming model is the most time-consuming one. In memory terms, the size of the model is what matters the most. In computing time however, the combination of spacial-step, temporal-step and recording time become a key factor.

2.5 Conclusions

Concerning the forward problem, the wave propagation, we have seen that selecting the wave phenomena and its discretization is fundamental for its proper implementation. Depending on the type of waves to be analyzed, the complexity can and will increase accordingly. A final important issue in the modeling stage is to take into account the velocity model that will be used. We have seen that with our three models, we have found three very different scenarios: a model with a small and *simple* high velocity square diffractor intrusion, a geometrically complex model with smooth velocity variations and a big complex (very) high velocity intrusion; each of them presenting its difficulties in terms of physical and computational phenomena.

Turning our attention to the inverse problem, the inversion itself, the most satisfactory results were obtained for the *simplest* (with those sharp edges is not that simple) model with a solution very close to the original. The Marmousi model also presented good results despite having such a complex geometry (such as faults), but we have to take into account that the initial model from which the FWI started was a good start point to begin with. Finally, the Sigsbee model, which turned out to be the model further from its real solution. This, due to the fact that almost all of the energy in the wavefront could not get back to the surface. In conclusion, properties such as velocity represent a much tougher challenge than complex geometries for techniques such as the FWI.

Another concern is raised in terms of the implementation. It was seen that artificial phenomena have to be either avoided (instability), suppressed (artificial reflections) or taken into account (dispersion) in the modeling or there may artifacts in the results not matching the reality.

Finally, the issue computational resources. Current platforms are capable of computing inversion processes such as FWI without much problem (with a limit however in the model size or its dimensions), but the real issue comes from the Hessian matrix, even small velocity models have Hessian matrices whose size is comparably bigger. In this chapter we have circumvented this inconvenient by computing not the full Hessian matrix, but a mere approximation of it as it was enough for updating the model. In the next (and final) chapter, we will discuss the importance of computing the full Hessian matrix and the use that it has once it is obtained for accounting for uncertainties present in the model after the FWI has been performed.

Uncertainty Quantification

Even though FWI was proposed in the late 70's and early 80's [5], it has not been used until recently due to its computational complexity. In fact, some of its applications include: local-scale engineering and exploration problems (such as near-surface seismic imaging with surface waves [15]), the study of crustal-scale deformation processes, the revelation of the detailed structures of the lower mantle and the refinement of continental-scale models for tectonic interpretations and improved tsunami warnings.

Despite all applications and problems that FWI has solved, there is a fundamental aspect of it that has not been taken into account, that is, the estimation of uncertainties and resolution [32]. In the state-of-the-art, this measure is known as Uncertainty Quantification (UQ), and it gives an estimate of how certain or uncertain is the velocity model obtained with the FWI while the latter provides information about how well resolved are the model parameters.

The importance of these two terms is that they come to deal with a fundamental shortcoming of the seismic inverse problems: the non-uniqueness of the solution. “*The non-uniqueness of the solution can be characterized by its resolution and its variance, usually represented by the resolution matrix and the (a posteriori) covariance matrix.*” [33].

Uncertainty analysis has only one problem: there is no satisfactory method to find the *a posteriori* covariance of the solution. For this reason, methods for the quantification of uncertainties in realistic applications of FWI do not exist so far [32].

As there are no quantitative means to assess both uncertainty and resolution, two qualitative measures are commonly used: visual inspection of the seismic images and analysis of the data fit. The former method is inadequate as small features can be mistaken as indicators of high resolution and the latter cannot be fully trusted either because even when there is a good fit between the observed and the calculated data, this only proves that the seismic inverse problem has been solved, not necessarily resolved [32].

This is where the role of the Hessian matrix comes into play. This matrix is closely related not only to resolution but also to the posterior covariance matrix as it will be described below and will be the focus of the rest of the chapter.

3.1 Relationship between the Hessian and the covariance matrices

Using Bayesian inference, the Hessian matrix can be related to the covariance matrix.

Let $\mathbf{m}$, an unknown solution of the inverse problem, be a random variable and, $\tilde{\mathbf{m}}$ be the current initial model for solving the inverse problem. Then, the update value $\Delta\mathbf{m}$ is

$$\Delta\mathbf{m} = \mathbf{m} - \tilde{\mathbf{m}}. \quad (3.1)$$

Since the FWI minimizes the ℓ_2^2 norm (equation 2.36), we are assuming linearity and accepting Gaussian statistics, therefore we can define the probability density function (PDF) of $\mathbf{m}$ as

$$\rho(\mathbf{m}) = \text{const } e^{-\frac{1}{2}(\mathbf{m}-\tilde{\mathbf{m}})^T \mathbf{S}(\mathbf{m})^{-1} \mathbf{m}-\tilde{\mathbf{m}}}. \quad (3.2)$$

There are three steps one must follow to find the relationship between $\mathbf{S}(\mathbf{m})$ and $\mathbf{H}(\mathbf{m})$. Those are:

- Represent the objective function $\Phi(\mathbf{m})$ through Taylor series.
- Place the expanded series into an exponential.
- Take it to the form of the posterior distribution (Gaussian).

First of all, we will expand the objective function in Taylor series

$$\Phi(\mathbf{m}) = \Phi(\tilde{\mathbf{m}}) + (\mathbf{m} - \tilde{\mathbf{m}})^T \mathbf{g}(\mathbf{m}) \Big|_{\mathbf{m}=\tilde{\mathbf{m}}} + \frac{1}{2} (\mathbf{m} - \tilde{\mathbf{m}})^T \mathbf{H}(\mathbf{m}) \Big|_{\mathbf{m}=\tilde{\mathbf{m}}} (\mathbf{m} - \tilde{\mathbf{m}}), \quad (3.3)$$

then we put it into an exponential, and take it to a Gaussian distribution form. This yields to the result of

$$\begin{aligned} \rho(\mathbf{m}) &= \text{const } e^{-\Phi(\mathbf{m})} \\ &= \text{const } e^{-\Phi(\tilde{\mathbf{m}}) - (\mathbf{m} - \tilde{\mathbf{m}})^T \mathbf{g}(\mathbf{m}) \Big|_{\mathbf{m}=\tilde{\mathbf{m}}} - \frac{1}{2} (\mathbf{m} - \tilde{\mathbf{m}})^T \mathbf{H}(\mathbf{m}) \Big|_{\mathbf{m}=\tilde{\mathbf{m}}} (\mathbf{m} - \tilde{\mathbf{m}})} \\ &= \text{const } e^{-\Phi(\tilde{\mathbf{m}})} \times e^{-(\mathbf{m} - \tilde{\mathbf{m}})^T \mathbf{g}(\mathbf{m}) \Big|_{\mathbf{m}=\tilde{\mathbf{m}}}} \times e^{-\frac{1}{2} (\mathbf{m} - \tilde{\mathbf{m}})^T \mathbf{H}(\mathbf{m}) \Big|_{\mathbf{m}=\tilde{\mathbf{m}}} (\mathbf{m} - \tilde{\mathbf{m}})}. \end{aligned} \quad (3.4)$$

There are however, two extra terms in equation 3.4.

The first term is constant

$$e^{-\Phi(\tilde{\mathbf{m}})} = \text{const} \quad (3.5)$$

and the second term consist in evaluating the gradient at the point $\tilde{\mathbf{m}}$, therefore it is also a constant

$$e^{-(\mathbf{m} - \tilde{\mathbf{m}})^T \mathbf{g}(\mathbf{m}) \Big|_{\mathbf{m}=\tilde{\mathbf{m}}}} = \text{const}. \quad (3.6)$$

The resulting equation of $\rho(\mathbf{m})$ is

$$\rho(\mathbf{m}) = \text{const } e^{-\frac{1}{2} (\mathbf{m} - \tilde{\mathbf{m}})^T \mathbf{H}(\mathbf{m}) \Big|_{\mathbf{m}=\tilde{\mathbf{m}}} (\mathbf{m} - \tilde{\mathbf{m}})}. \quad (3.7)$$

Comparing equations 3.2 and 3.7 leaves the relationship between the Hessian and the covariance matrices as

$$\mathbf{H}(\mathbf{m}) \Big|_{\mathbf{m}=\tilde{\mathbf{m}}} = \mathbf{S}(\mathbf{m})^{-1} \Big|_{\mathbf{m}=\tilde{\mathbf{m}}}. \quad (3.8)$$

Knowing this relationship, our concern is to compute, as stated in equation 3.8 the covariance matrix, via the inverse of the Hessian matrix.

As with any matrix, there are several approaches to invert the Hessian matrix [34]:

- Gaussian elimination
- Eigen decomposition
- LU decomposition

For its computational efficiency we opted for the LU decomposition [35]. The problem lies however, not in the method for inverting the matrix but in the invertibility of the matrix itself, which must be accounted for in terms of its determinant and condition number. This problem rises thanks to the ill-posedness of the FWI and can translate into a singular or close to singular Hessian matrix giving us no chance of finding such an inverse.

3.2 The *a posteriori* covariance matrix

An alternative to deal with the latent problem of singularity while also including *a priori* information to account for uncertainties, is to add a regularize our matrix by the addition of a term [36]. By doing this we can obtain what is know in Bayes theory as the posterior covariance matrix [37].

3.2.1 Estimating posterior uncertainties

The framework behind this sort of regularization lies behind the assumption that (for the linear case) there must be a point $\tilde{\mathbf{m}}$ and a covariance matrix $\tilde{\mathbf{S}}$ such that the posterior probability density can be written

$$\sigma(\mathbf{m}) = \text{const } e^{-\frac{1}{2}(\mathbf{m}-\tilde{\mathbf{m}})^T \tilde{\mathbf{S}}^{-1}(\mathbf{m}-\tilde{\mathbf{m}})}. \quad (3.9)$$

Equation 3.9 can be achieved using the Bayes' rule (as presented in [23]) where we assume $p(\mathbf{d}|\mathbf{m})$ is the conditional pdf of $\mathbf{d}$ for a given $\mathbf{m}$, $\sigma(\mathbf{m}|\mathbf{d})$ is the conditional pdf of $\mathbf{m}$ for a given $\mathbf{d}$, $p(\mathbf{d})$ is the pdf of the data and $p(\mathbf{m})$ is the pdf of the model $\mathbf{m}$ independent of $\mathbf{d}$. From the definition of conditional probabilities we have

$$\sigma(\mathbf{m}|\mathbf{d})p(\mathbf{d}) = p(\mathbf{d}|\mathbf{m})p(\mathbf{m}). \quad (3.10)$$

From equation 3.10 we can obtain the posterior probability density function $\sigma(\mathbf{m})$ knowing that the term $p(\mathbf{d})$ is, in geophysical inversion, a constant and $p(\mathbf{d}|\mathbf{m}) = \rho(\mathbf{m})$, thus giving us

$$\sigma(\mathbf{m}) \propto \rho(\mathbf{m})p(\mathbf{m}). \quad (3.11)$$

The pdf $p(\mathbf{m})$ is called the *prior distribution* and represents the information we have on the model without the knowledge of the data.

If we replace $\rho(\mathbf{m})$ from equation 3.7 into 3.11 and assume the prior distribution $p(\mathbf{m})$ to be Gaussian as

$$p(\mathbf{m}) = \text{const } e^{-\frac{1}{2}(\mathbf{m}-\tilde{\mathbf{m}})^T (\mathbf{C}_1^{-1})(\mathbf{m}-\tilde{\mathbf{m}})}, \quad (3.12)$$

we get

$$\begin{aligned} \sigma(\mathbf{m}) &= c_1 e^{-\frac{1}{2}(\mathbf{m}-\tilde{\mathbf{m}})^T \mathbf{H}(\mathbf{m}-\tilde{\mathbf{m}})} c_2 e^{-\frac{1}{2}(\mathbf{m}-\tilde{\mathbf{m}})^T \mathbf{C}_1^{-1}(\mathbf{m}-\tilde{\mathbf{m}})} \\ &= \text{const } e^{-\frac{1}{2}(\mathbf{m}-\tilde{\mathbf{m}})^T (\mathbf{H}+\mathbf{C}_1^{-1})(\mathbf{m}-\tilde{\mathbf{m}})}. \end{aligned} \quad (3.13)$$

Looking at equation 3.13, we can easily relate it to equation 3.9 where the *a posteriori* information of the parameters is centered at the final model of the FWI with posterior covariance $\tilde{\mathbf{S}}$ given by

$$\tilde{\mathbf{S}} = (\mathbf{H} + \mathbf{C_I}^{-1})^{-1}. \quad (3.14)$$

Here, $\mathbf{H}$ is the Hessian matrix and accounts for information brought by the data while $\mathbf{C_I}$ is the *a-priori* covariance matrix and accounts for *a-priori* information of the model. This of course makes the posterior distribution strongly dependent on the prior information, which can be obtained based on previous knowledge of the geological area or as newly applied methods such as meta-heuristic techniques that help the inversion by providing not only an initial model for the FWI but also a prior distribution to enhance the uncertainty quantification.

Concerning the prior covariance matrix $\mathbf{C_I}$, it must be said that it has to be constrained to help condition the Hessian matrix, but without masking the Hessian matrix.

When little or no information is known about the prior covariance matrix, one possibility is to assume an uncorrelated constant variance for every model parameter, resulting in a prior covariance matrix

$$\mathbf{C_I} = \sigma^2 * \mathbf{I}, \quad (3.15)$$

where $\mathbf{I}$ is an identity matrix of the size of $\mathbf{H}$ and standard deviation that theoretically can condition the Hessian matrix can be obtained as

$$\sigma_{\text{cond}} = \sqrt{\left| \frac{1}{\min \mathbf{H}_{\text{diag}}} \right|}. \quad (3.16)$$

Using equation 3.16, one can find a value σ_{cond} to condition the Hessian matrix and obtain the $\tilde{\mathbf{S}}$ matrix. Table 7 shows the σ_{cond} for our three models.

Model	$\sigma_{\text{cond}}[\text{m/s}]$
Diffactor	515.7
Marmousi	42.6
Sigsbee	1.34

Table 7: Conditioning standard deviation (σ_{cond}) for the three SVM with a source frequency of 3Hz.

3.2.2 Results of the SVM's

With all previous analysis we can finally obtain the posterior covariance matrices for our example models. Keeping simple the visual inspection of the images, we opted to show only the diagonal of the matrices. Such as the Hessian matrices, the size of the posterior covariance matrices makes it very difficult to assess the information embedded in these matrices. The only thing we are going to mention about the posterior covariances matrices at this point, is that their diagonal contain the variance of

each of the model parameters (σ_i^2 , where i is one of the pixels) while the off-diagonal elements contain the covariance of the model parameters (σ_{ij}^2 , where i and j represent two pixels at different positions).

Figures 34, 35 and 36 show the diagonal of the posterior covariance matrices for each of our velocity models: diffractor, Marmousi and Sigsbee, respectively. From them, the common characteristic is the resemblance between those diagonals with the diagonals of their respective Hessian matrices.

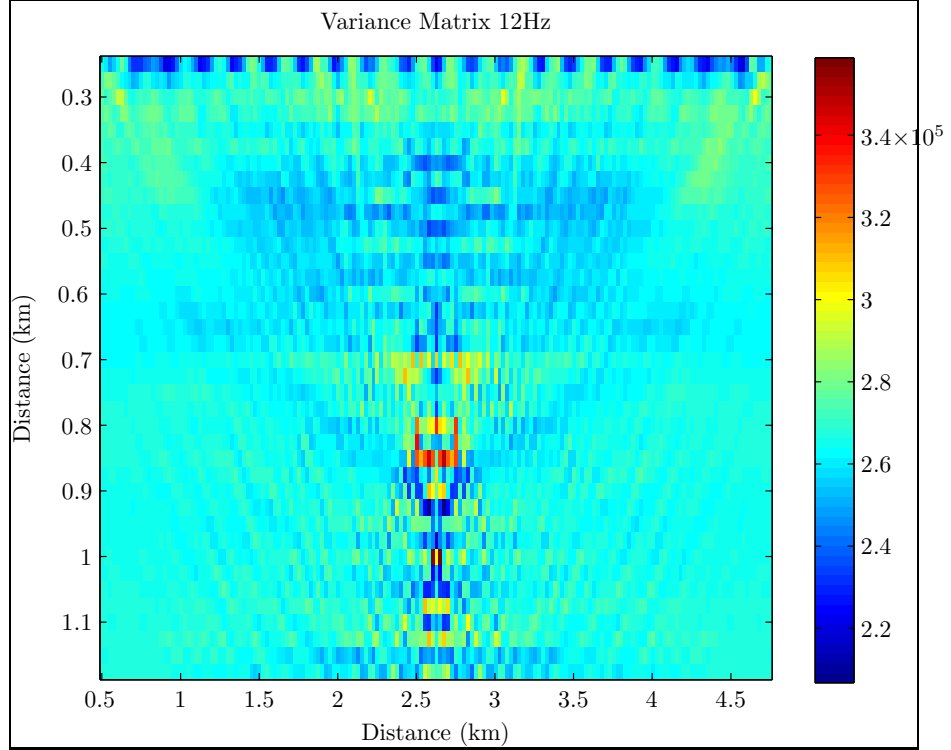


Figure 34: Posterior variance matrix of the square diffractor model.

Figure 34 shows the lowest values of variance in the area not only above the diffractor, but also around the diffractor itself. This obeys the fact that this were the most updated zones during the inversion and therefore, present the biggest reduction in variance.

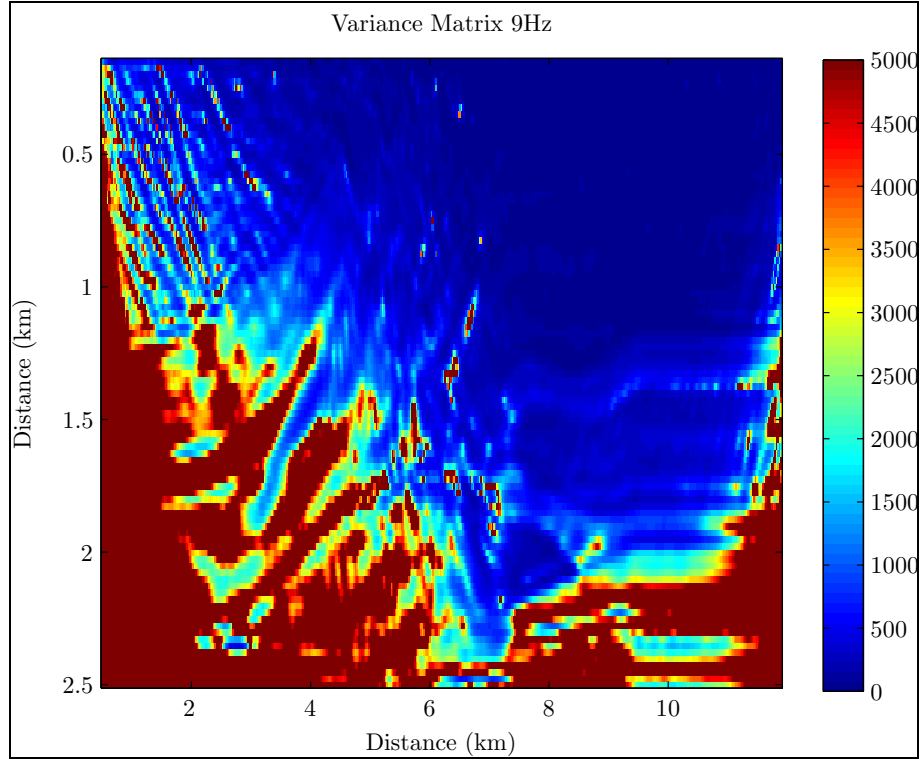


Figure 35: Posterior variance matrix of the Marmousi model.

The Marmousi posterior variance matrix presents a geometry resembling again the updated zones during the FWI. The top-right area presents the lowest values in variance, meaning that the uncertainty is higher at greater depths.

The variance matrix shown in figure 36 corresponds to the Sigsbee model. So far, this model has presented the poorest results at every stage of the processing and its result in terms of Hessian and covariance matrices are the worst yet. Just as the obtained Hessian matrix, the covariance matrix suffers from the ill-conditioning of the problem. We can see in table 7 that the σ_{cond} one can use when no information is available is very low and therefore the result was not as the previous models. For the sake of testing, we ranged the σ_{prior} used in $\mathbf{C}_I$ for up to 2000m/s, but results did not get better, in fact they got worse and worse as more and more elements of the variance matrix became negative.

Figure 37 shows the increment in negative elements when increasing the σ_{prior} . Comparing it with the percentage of negative σ_{post} from the diffractor (0.0557%) and the Marmousi model (0.2395%) it can be seen that for the Sigsbee model the amount of negative elements is quite superior (up to 53.23%).

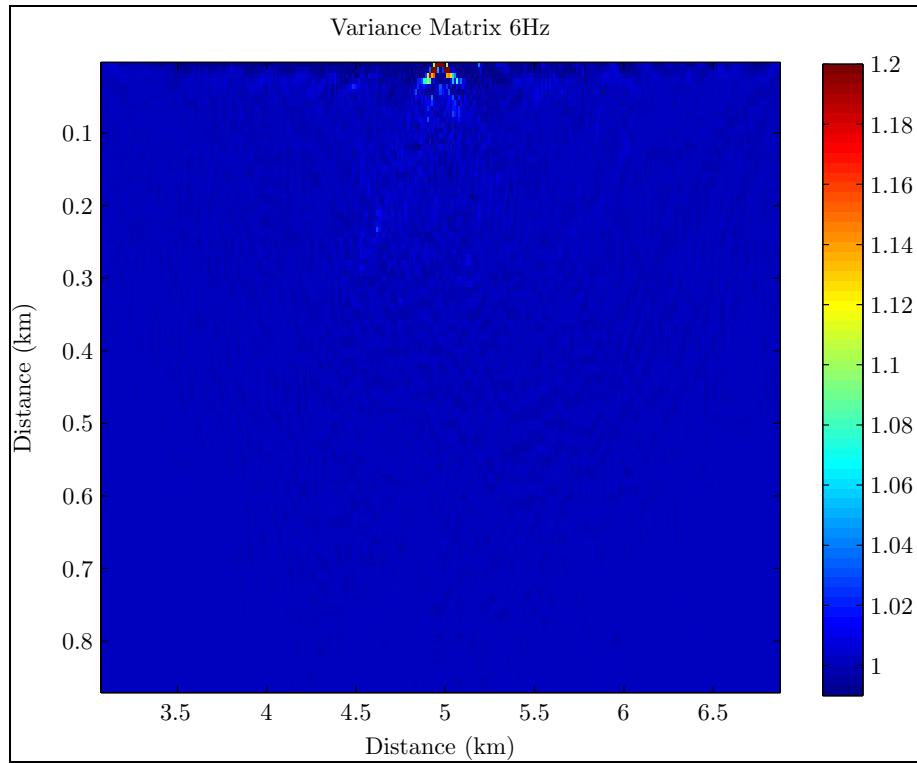


Figure 36: Posterior variance matrix of the Sigsbee model.

The issues with this model have always come from the fact that a model such as this one, with zones that prevent energy from returning to the surface, becomes a very tough challenge not only for the FWI, but also for the UQ process as they both depend on the recorded data.

Once computed, is with these posterior covariance matrices that we can perform an analysis of uncertainties, or as we like to call it: an uncertainty appraisal, which is the final purpose for this work.

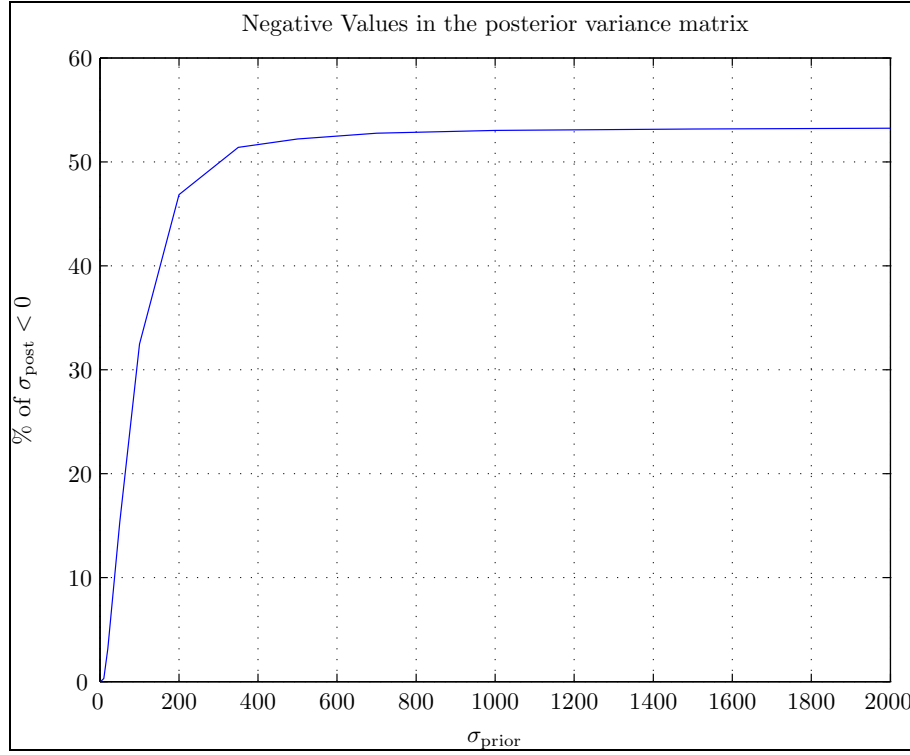


Figure 37: Negative values in the posterior variance matrix.

3.3 Uncertainty Appraisal

In the road so far, we have obtained the Hessian matrix, column by column (which involved the solution of four wave equations and the summation of three inner products per column); added to it the inverse of a prior covariance matrix; and finally, inverted that sum of matrices after correcting its weights. Now that we have the posterior covariance matrix $\tilde{\mathbf{S}}$ at hand, how can we quantify for the uncertainties with it?

To answer this question, we can look at three approaches: the percentage of reduced uncertainty (or what we called the UQ factor), the formation of uncertainty bars or intervals, and finally the resolution operator.

3.3.1 Uncertainty Bars

This is the most trivial use of the posterior covariance matrix, we can take the square roots of the diagonal elements (variances) as limits for solution intervals around the solution found by the FWI as seen in [38] and [39].

To give a better perspective of what is intended, in figure 38 there are two Gaussian distribution, a prior distribution (on the left) with μ around 2000m/s and $\sigma = 96.8\text{m/s}$ and a posterior distribution (on the right) with μ still around 2000m/s but $\sigma = 89.6\text{m/s}$. These two distributions correspond to the same pixel (each pixels represents a random variable and hence, has its own prior and posterior distributions), where it is observable that after the FWI has been done, both values (μ and σ) have been altered. The former (the mean) has moved slightly, but more importantly, the latter (the variance) has been reduced, thus reducing the range of possible solutions which can be interpreted as the reduction of the uncertainty in the solution for that pixel.

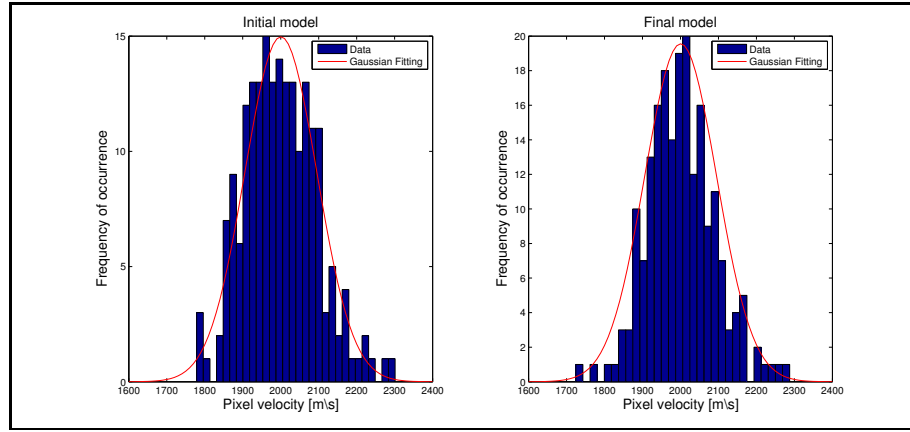


Figure 38: Example distributions for a same pixel before (left) and after (right) FWI. The distributions share the same μ approximately but the standard deviation is reduced from the prior to the posterior distribution.

The comparison of the posterior uncertainties (represented by $\tilde{\mathbf{S}}$) with the prior uncertainties (represented by $\mathbf{C_I}$) shows which parameters have been *resolved* and by how much. It is clear however that, if the posterior variance of a parameter is identical to the prior variance, no information has been brought by the data on this parameter per se.

To showcase the result of this approach, figure 39 depicts both distributions superposed for the central pixel of the diffractor model. As the initial model for the FWI was a constant layer model with velocity of 2000m/s, that is the mean of the prior distribution, while the standard deviation is the σ_{prior} selected as 515.7m/s. The posterior distribution has as its mean (the solution of the FWI for that point) 2476m/s and the standard deviation $\sigma_{\text{post}} = 500.2$. The correct solution is located at 2500m/s (the velocity of the original model for that point). This can be done for each of the points in the grid.

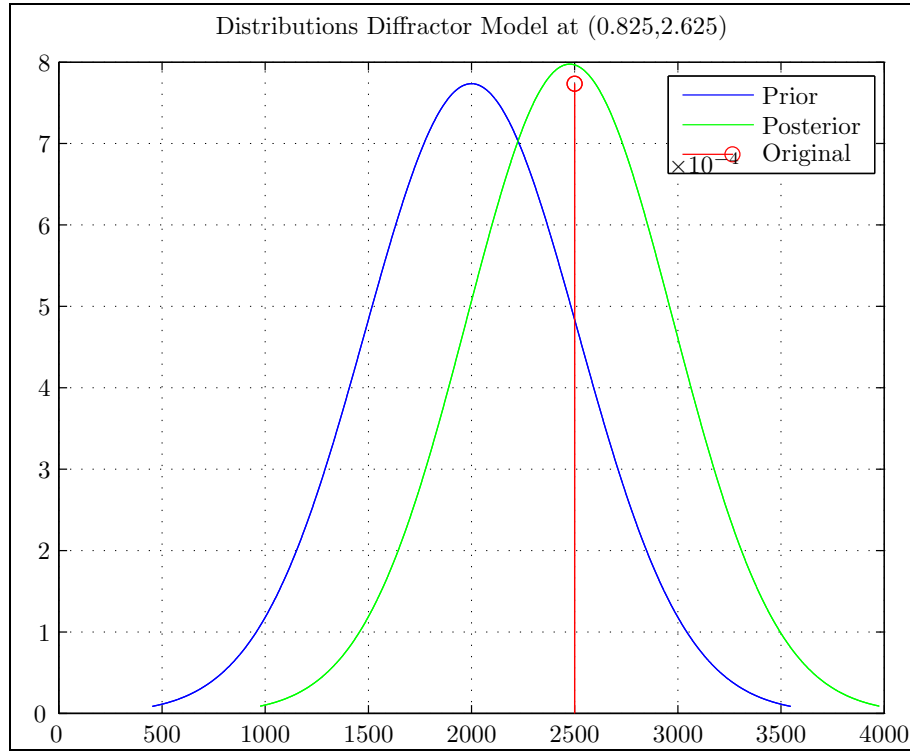


Figure 39: Prior and posterior distributions for the central pixel of the diffractor model.

An alternative to view each point separately, is to extract slices of the previous matrices, for example of the variance matrix. This example can be seen in figure 40, where we extracted the variance of the diffractor model at the middle of it and compared it against the prior variance. From the figure it can be analyzed both effects of: increment and decrease in the posterior variance matrix.

If we extend this analysis, we can examine a model distribution profile, as depicted in figure 41 where, contrary to figure 40, we extracted not the variance but the final Marmousi after FWI and used the variances to limit an interval. This interval accounts for three times the standard deviation of each point in the model shifting from a single solution to an interval of solutions.

It is noticeable how the interval is narrower at higher depths while it broadens as we move deeper and deeper. This effect is produced by the inversion, as it works its way down from the top of the model and as seen in chapter 2 solves better at shallower depths.

We have gone one step further and added also the real model profile, for academic purposes only, as with real data, not always will it be possible to have such a profile (only when well-data is available).

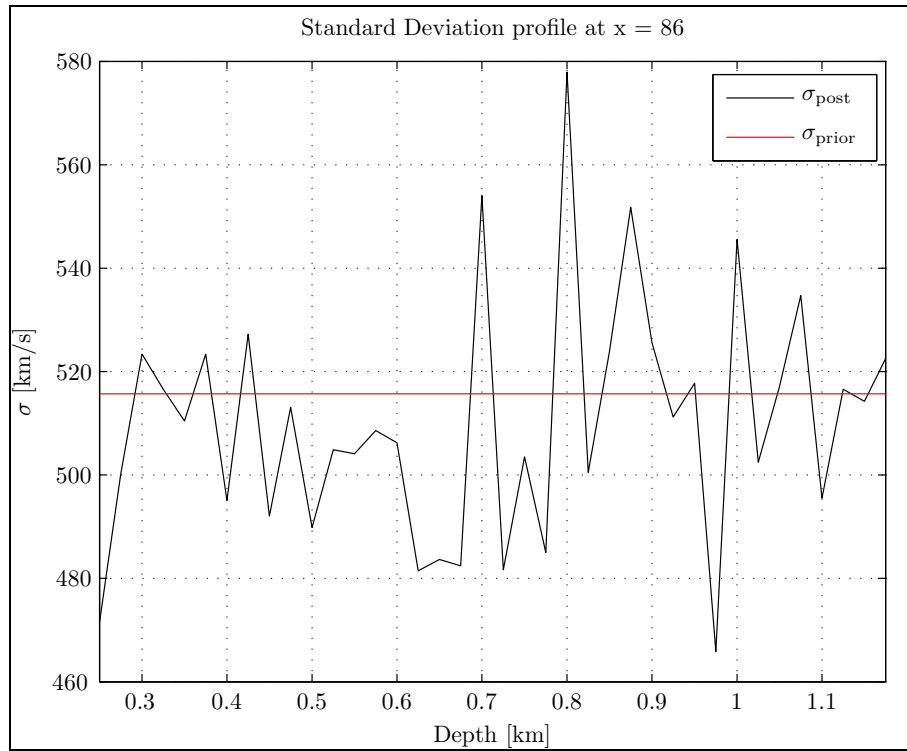


Figure 40: Uncertainty profile for the diffractor model at a location of $x = 2.625\text{km}$.

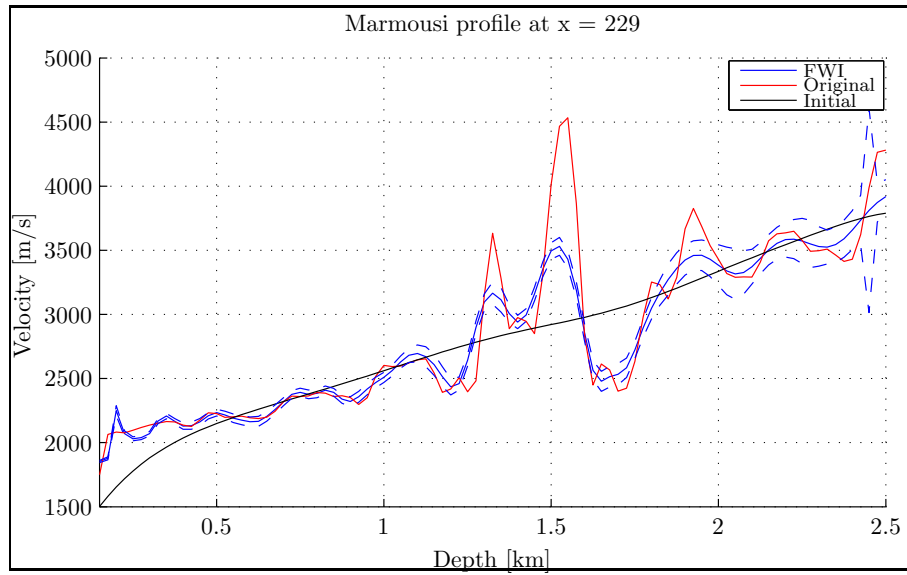


Figure 41: Uncertainty profile for the Marmousi model at a location of $x = 5.72\text{km}$.

Using the real values we can find out at which areas the velocities lie outside our solution interval, this however is not feasible when using real data.

A non-so-trivial analysis is to examine the off-diagonal elements (covariances) of the covariance matrix in order to obtain the correlation coefficients as

$$r_{ij} = \frac{\tilde{\mathbf{S}}_{ij}}{\sqrt{\tilde{\mathbf{S}}_{ii}\tilde{\mathbf{S}}_{jj}}}, \quad (3.17)$$

which hold the following property

$$-1 \leq r_{ij} \leq 1. \quad (3.18)$$

Obtaining the coefficients r_{ij} and checking if condition 3.18 is met or not, is a good way to determine if the matrix $\tilde{\mathbf{S}}$ has been calculated correctly. For example, obtaining the correlation coefficients for the Sigsbee's covariance matrix shows that 87.08% of its values lie outside of the interval. Interested readers can read [5] for more on this topic.

3.3.2 UQ Factor

Up to this point, we have visually seen how a reduction in the variance is addressed as a reduction in uncertainty, but we have yet to account for it explicitly. For this reason, the percentage of uncertainty reduction or namely, the UQ factor, addresses the question of “by how much does the uncertainty gets reduced after a process such as FWI is performed taking into account prior information on the model?” [40]. The calculation is simple enough,

$$\text{UQ} = \frac{\mathbf{C}_{\mathbf{I}} - \tilde{\mathbf{S}}}{\mathbf{C}_{\mathbf{I}}} * 100\% \quad (3.19)$$

with $\mathbf{C}_{\mathbf{I}} = \sigma_i^2 \mathbf{I}$ when having little to no prior information and assuming uncorrelated pixels in the prior distribution.

Beginning with the square diffractor model in figure 42 it is depicted the zones where uncertainty has been reduced and the color-scale shows at what percentage it has been reduced. For this model the maximum reduction was around 30% in the zone of the square diffractor, but there were zones in the background where reductions in variance were also present while at the shallow area of the model, reductions of around 15% were found. These are related to the sources.

Following the diffractor, we continue with the UQ factor for the Marmousi model. In figure 43 it is depicted the reduction of its uncertainties, where most of the reduction occurred at the top right of the model, following the same pattern of the gradient the Hessian and off course, the covariance matrices.

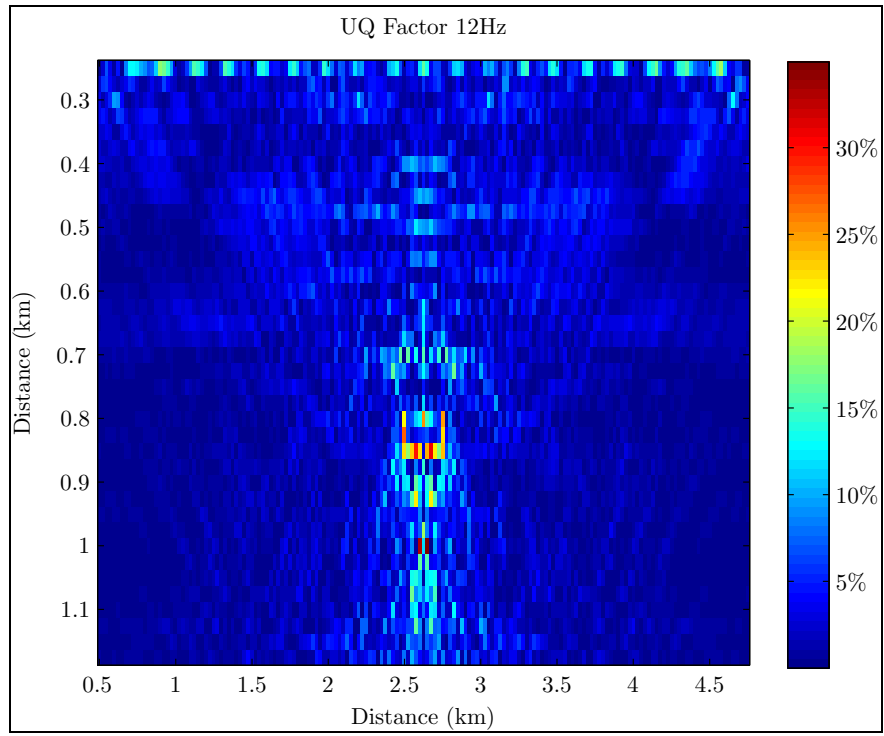


Figure 42: Uncertainty Reduction Factor (UQ Factor) of the square diffractor model.

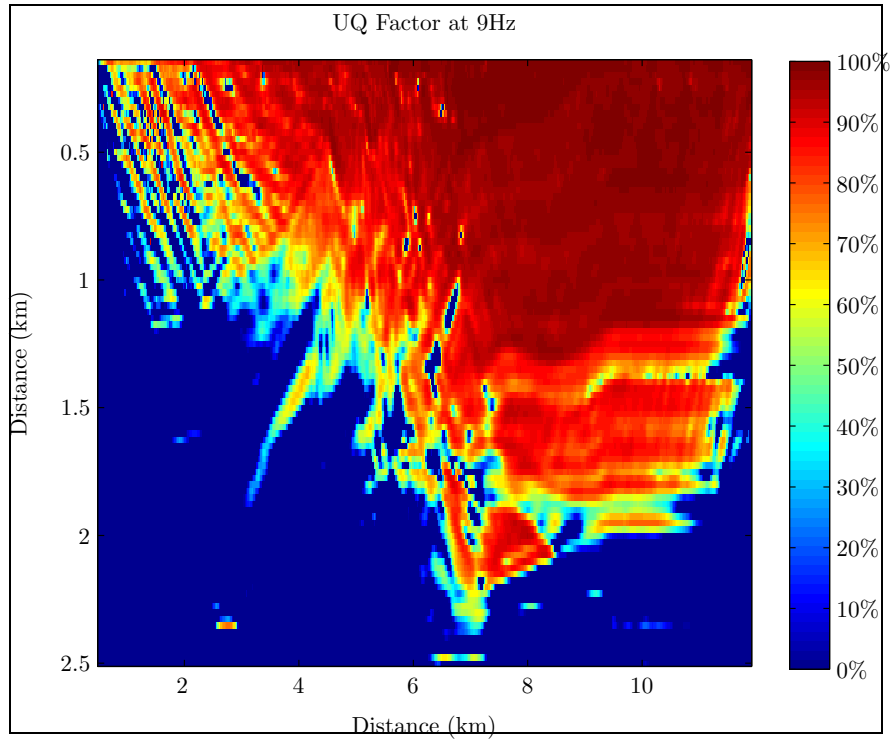


Figure 43: Uncertainty Reduction Factor (UQ Factor) of the Marmousi model.

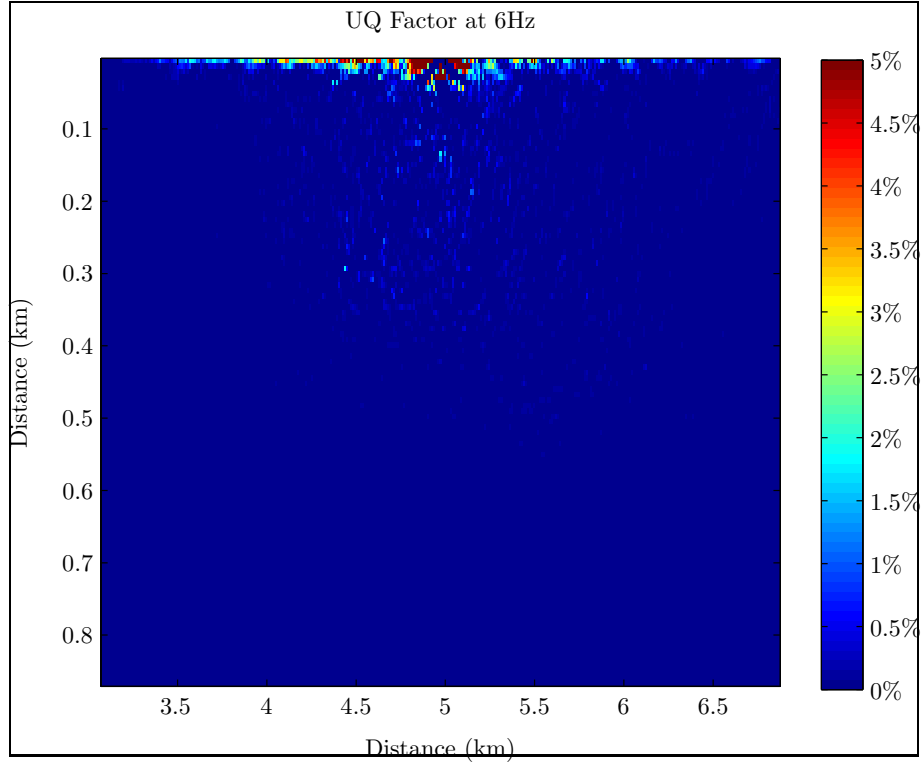


Figure 44: Uncertainty Reduction Factor (UQ Factor) of the Sigsbee model.

The big difference with the previous example is the magnitude of the reduction, which accounts for a reduction of up to 99.78%.

The final UQ factor comes from the Sigsbee model, where figure 44 shows that almost no improvement in uncertainty was achieved. Considering that the posterior covariance matrix per se does not say much, this result is not surprising. The shallow zones of the model got an UQ factor of up to 5%.

3.3.3 Resolution Operator

Another way to account for uncertainties, or better, to assess how well solved are the model parameters, is with the Resolution Operator ($\mathbf{R}$) as the difference between an identity matrix ($\mathbf{I}$) and the product of the posterior covariance matrix ($\tilde{\mathbf{S}}$) with the inverse of the prior covariance matrix ($\mathbf{C}_{\mathbf{I}}^{-1}$) as presented in [5]

$$\mathbf{R} = \mathbf{I} - \tilde{\mathbf{S}}\mathbf{C}_{\mathbf{I}}^{-1}. \quad (3.20)$$

This difference with the identity accounts for resolution, if the resolution operator were the identity operator, we would have perfectly resolved the *exact* model. In contrast, the farther the resolution operator is from the identity, the worse the resolution is [5].

Just like matrices $\mathbf{H}$ and $\tilde{\mathbf{S}}$, the diagonal of the matrix $\mathbf{R}$ can be extracted and rearranged in the same size of the model.

Figure 45 shows the resolution matrix for the square diffractor model, where once again, the best resolved parameters are found within the square diffractor and the area around it following the behavior of the covariance matrix. Additionally, at shallow depths the area near the sources finds itself being also resolved at a higher proportion.

The resolution matrix of the Marmousi model is presented in 46, and it can be seen that congruently with what has been seen previously, the top-right area of the model is what has been better resolved. This inferences could be made from the gradient and the Hessian matrices, even from the updated model. The biggest contributions of performing an analysis of this caliber is to be able to quantify the resolution rather than just visually determining where has the model been resolved.

In contrast with our previous models, the resolution matrix of the Sigsbee models, depicted in figure 47, show values that differ greatly from 1 preventing the resolution matrix from resemble an identity matrix.

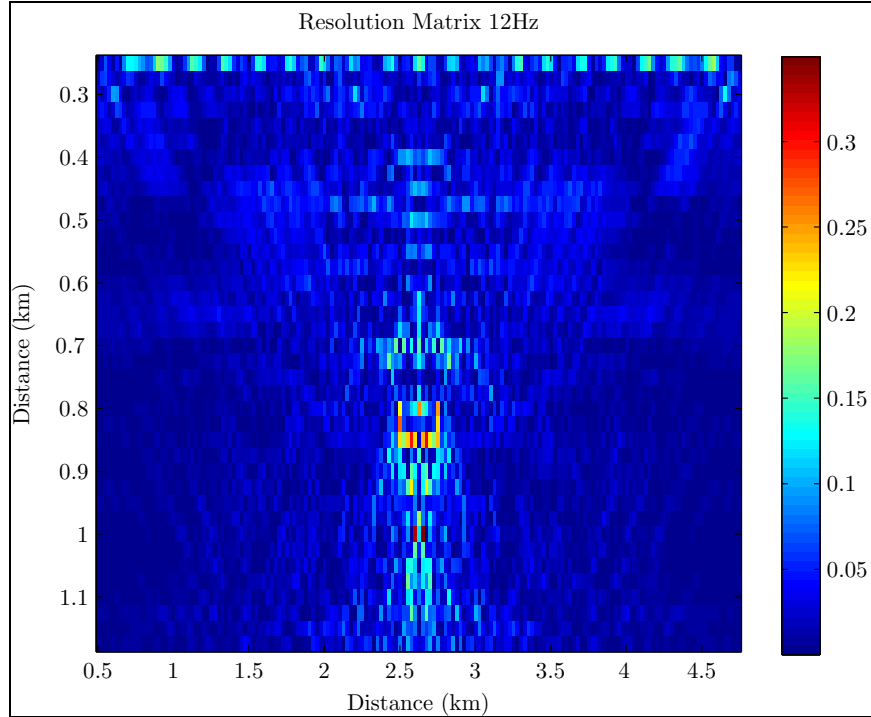


Figure 45: Resolution operator of the square diffractor model at 12 Hz.

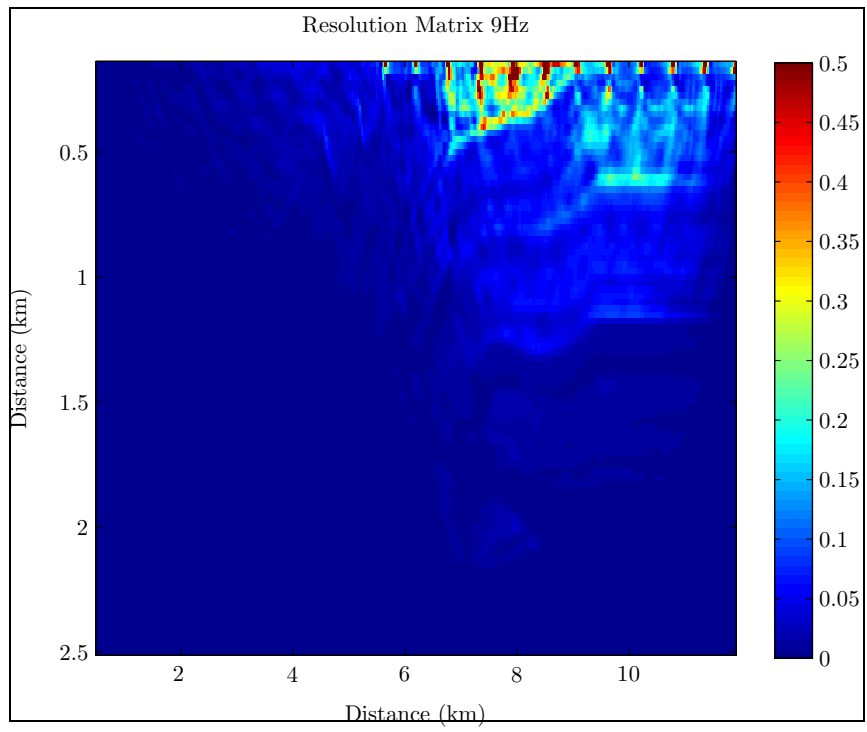


Figure 46: Resolution operator of the Marmousi model at 9 Hz.

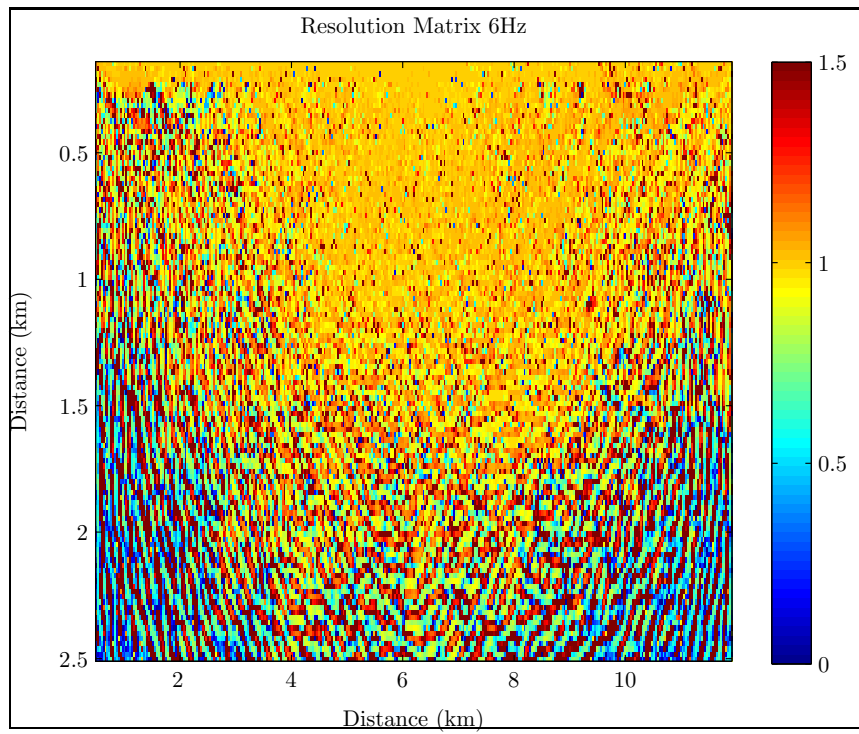


Figure 47: Resolution operator of the Sigsbee model at 6 Hz.

A closing remark for the resolution operator is presented in [5], it shows that solving equation 3.20 for $\tilde{\mathbf{S}}$, yields to the result

$$\tilde{\mathbf{S}} = (\mathbf{I} - \mathbf{R})\mathbf{C}_I^{-1}, \quad (3.21)$$

where it can be seen that if the resolution operator is close to the identity, the posterior covariance is close to zero, and the model parameters have been resolved well.

3.4 Discussion

In this chapter we have made strong assumptions, such as: Gaussian statistics and linearity. The reasoning behind these two assumptions has been stated and the consequences of them have been exploited to obtain the presented results. The problem lies on the fact that these two conditions are not always met, in fact, as stated by Tarantola [5]: if the objective function is not a linear function of the model parameters, the posterior covariance matrix is not Gaussian. Indeed, the more nonlinear the objective function is, the more remote the posterior distribution is from Gaussian. The issue is closely related to the start point of the inversion, as discussed in the previous chapter, the problem of non-uniqueness comes from the multiple minima in the objective function, which is caused by the ill-posedness of the inverse problem. Following this train of thoughts, the selection of a good start point becomes a crucial matter not only for the inversion itself but also for accurately quantifying the uncertainties.

In chapter 2 we discussed a newly applied technique of finding initial models, global optimization methods are currently used to find the solution to the inverse problem (as can be seen in [23]), but their computational load becomes their biggest drawback. The interest becomes to use these meta-heuristic techniques, such as the Particle Swarm Optimization (PSO), not to solve the whole inverse problem, but only to provide a better start point to local optimization methods such as the FWI. Referring specifically to the PSO method, in which multiple particles roam the space of solutions with an initial distribution, but more importantly, finish their minimization with also a distribution, that will be an input for both the FWI (in the form of initial model) and the UQ process (in the form of prior distribution, or more specifically, prior covariance matrix). In this manner, exploit the advantages of each type of optimization while diminishing their weak-spots.

The prior distribution is precisely what we are aiming to have in order to condition the Hessian matrix, and it is of vital importance as the inverse of the Hessian matrix can contain several *unwanted* values. These values can arrive in the form of negative values which yield to complex standard deviations. When the percentage of negative values is low (as in the cases of the diffractor and the Marmousi models) one can *bypass* this by simply taking the absolute value of the variance, but keeping track of

those variances in order to perform the uncertainty appraisal only for the originally positive values.

Another issue comes from the elements outside the main diagonal of the covariance matrix, which can produce correlation coefficients outside of the interval $[-1, 1]$, which are not correct. This can be dealt with in a similar fashion to the case of the negative variance, as one could simply limit values outside $[-1, 1]$ to 1 and -1 accordingly.

When the prior information is indeed available, either from experts, prior explorations done in the subsurface or global optimization methods; it is possible that some of the just mentioned values appear in the matrix. Depending on the case one has to analyze for example if truncating the values outside interval $[-1, 1]$ to inspect the correlation coefficients or if taking the absolute value of the diagonal of the covariance matrix can help construct the solution intervals. All these issues come from the ill-posedness of the problem and, so far, cannot be easily avoided.

Delving deeper into the main objective of this work, which was to account for uncertainties in order to assess how the FWI deals with the innate problem of non-uniqueness, it can be argued that not only the uncertainties have been accounted for after the FWI, but also that the very process of the inversion can reduce the uncertainties as seen in figures 42 and 43. This occurs as explained with figures 39 and 41 where it can be seen that the range of possible solutions has been reduced in terms of its standard deviation and also that after the FWI the solved model parameter lies closer to the real value.

As proposed by Nolet [33], we have reviewed the problem of non-uniqueness by finding the posterior covariance matrix and the resolution matrix, with them, it is possible to provide yet another tool for interpreting seismic images. Alongside with the UQ factor, these matrices become a figure of merit to assess how good (or how bad) the inversion is while also providing a figure of how much new information has been included in the solution model. With tools such as these, experts can improve their evaluations of seismic images.

The final topic to discuss is the computational cost of dealing with all these matrices. We have already shown the sizes of the Hessian matrices for each of our models in table 5. It is worth mentioning that the posterior covariance matrix and the resolution matrix have the same size that the Hessian they come from. This becomes a bigger issue for bigger models. From this point of view, there are two main limitations, as is usual with seismic problems, those are: memory and time.

The first of the issues, memory, depends exclusively on the size of the model. As we have seen the size of a Hessian matrix that comes from a $M \times N$ has a size of $MN \times MN$.

For models such as the Sigsbee, of 2133×1201 elements, a Hessian matrix of 22 TB exceeds by far the maximum amount of RAM memory currently available in most computers. The Guane supercomputer, one of the most powerful computers in Latin America, located in Bucaramanga, Colombia for example, has 16 nodes with 104 GB of RAM available [41]. Now, all computations performed in this work have been done using graphics processing units (GPU's). Currently, the maximum amount of RAM available in GPU's is around 24 GB, not enough for computing big Hessians. To give perspective to the reader, we have worked so far with synthetic models only, which are small in comparison with real data models that can easily exceed 3000×3000 points.

There are several possible strategies to overcome the memory limitation, the first of them is precisely what we did with the Sigsbee model, which is: reduce the size of the model. This can be achieved by two means, either reducing the scope of the model to invert or re-sample the model. On the one hand, reducing the scope implies keeping the original grid but reducing the objective of the model, think of it as solving a sub or mini model extracted from the full model. On the other hand, one can simply subsample the grid in order to make it feasible to compute the Hessian matrices and company. Both alternatives are widely used in the industry where from a very large acquisition no all data are used, some are skipped for the sake of computation. The cost of these approaches is that one loses the full solution when not taking into account the full model and resolution when there is more distance between nodes in the grid.

Referring specifically to the Hessian matrix, an alternative to reducing the size of the model is found in the very way that it is computed. As we have seen, with equation 2.81 we compute not the full Hessian matrix but columns of it. With this in mind one could compute not all the matrix at once, but the maximum amount of columns possible by the memory, store them in the hard drive and then continue with the computation for the next group of columns. Of course it would be necessary to have enough memory to put the pieces together. This approach becomes attractive when one has enough RAM in the computer but is using a GPU without enough memory for the full computation. The real issue comes next, when trying to invert the matrix. Fortunately, there is a way to circumvent this issue, it is called matrix partitioning [42].

The good news about the memory issues is that they are temporal as the birth of new technologies and architectures will make those computations feasible. Memories and storage have been increasing exponentially over the last 10 years [43] and is only a matter of time until we achieve greater capacities that allow us to compute bigger matrices. Time, the second big issue as seen in table 6, will be hopefully reduced with the arrival of new technologies that make CPU's and GPU's faster every passing year making possible not only the computation itself, but also making it a not so time consuming task.

The bright side in this chapter is that the computation of the inverse of the Hessian matrix does not consume as much time as the computation of it. Table 8 shows the time spent calculating and inverting a Hessian matrix for a single source, and as it can be seen, the time for inverting the matrix can be neglected as it is less than the 1% of the time spent computing the matrix. It is worth mentioning that while the number of sources increases the time for computing $\mathbf{H}$, it does not influence the time for inverting the matrix as it has the same size (the size of $\mathbf{H}$ only depends on the size $\mathbf{m}$).

Model	$\mathbf{H}$: 1 Sources	$\mathbf{H}$: 21 Sources	Inverting $\mathbf{H}$
Diffractor	12m	4h	13s
Marmousi	8h9m	7d	1h27m
Sigsbee (Re-sampled)	17h	15d	30m

Table 8: Time spent computing $\mathbf{H}$ and $\tilde{\mathbf{S}}$.

To close the discussion, it is worth pointing out that all the efforts made computing the Hessian matrix and its inverse become worthy when thinking in the advantage of computing this full Hessian matrix, as stated by [44] instead of approximate Hessians such as the ones used in [38] or even using the L-BFGS method.

3.5 Conclusions

We have computed the Hessian matrix for uncertainty quantification purposes using a parallel computing strategy with the aid of GPU's using the SOASM for three example models, however this approach can be widely applied taking into account only computational limitations. The possibility of computing these matrices opens up a wide range of tools that can work jointly with the FWI to help seismic interpretations.

Limitations of these kind of analysis come not only from computational resources (or lack thereof), but also from the very properties of the models themselves and also depend highly on the prior information. Just as the FWI relies on the initial model, the UQ relies analogically on the prior distribution for solving two issues: the bad-conditioning of the Hessian matrix and the addition of prior uncertainties.

In terms of the non-uniqueness of the solution, we have followed the work of [33] and [5] to the point of finding the posterior covariance matrix along with the resolution matrix. Additionally we have accounted for the reduction of uncertainty (performed by the FWI) presenting it as a percentage in what we defined as UQ factor. These three points of view give us the uncertainty after the inversion have been solved, the resolution of the parameters (how *well-solved* by the inversion) and also a figure of by how much the uncertainty has been reduced.

3.6 Future Work

The biggest leap from this work is the application of the exposed techniques to real data. This application is basically limited by the computational resources. The alternative to the development of new hardware is as always, optimizing the resources available. Techniques such as the SOASM have the advantage that they exploit the inherent independence of the physical event, in this case the independence of the shots made by the sources which can be computed individually. The SOASM inherits this independence and taking advantage of it can reduce both the memory needed for computing the Hessian matrix and the time needed for its computation. The other advantage is that the SOASM can compute columns one column (or a group of columns) of the Hessian at a time to be able to manage bigger Hessian coming from bigger models.

Following the real-data application, is the improvement of the covariance matrix. Even though our approach has its base, the assumptions and constrains related to it are very strong and when these conditions are poorly met, the results tend to contain, as discussed, *unwanted* values. This represents a compromise between the mathematical and computational accuracy and the interpretation of the physical phenomena. An idea of how to avoid those negative and complex values (for the variance and standard deviation) or even those correlation coefficients outside the interval is found in [45] where the authors try to find the closest correlation matrix to a matrix (in our case from $\tilde{\mathbf{S}}$) by solving an optimization problem.

Last but not least, directly from the multi-scaled FWI, comes the multi-scaled UQ. This approach resembles its FWI counterpart by solving the quantification at different frequency scales, taking the previous posterior distribution as the prior distribution of the current frequency to account for. This approach is currently in development, but unfortunately escapes the scope of this work and this book.

Hope you have enjoyed the reading of this book as much as I enjoyed the writing of it.

References

- [1] Sercel, “What on earth is geophysics?.” <http://www.sercel.com/about/Pages/what-is-geophysics.aspx>. Visited: June of 2016.
- [2] Narcowich, Francis J., “Math 414, haar wavelet analysis.” [http://www.math.tamu.edu/~sim\\$fnarc/m641/f07/m414w_ln29.html](http://www.math.tamu.edu/~sim$fnarc/m641/f07/m414w_ln29.html). Visited: June of 2016.
- [3] Irons, Trevor, “Marmousi model.” <http://www.reproducibility.org/RSF/book/data/marmousi/paper.pdf>. Visited: June of 2016.
- [4] Irons, Trevor, “Sigsbee model.” http://www.reproducibility.org/RSF/book/data/sigsbee/paper_html/paper.html. Visited: June of 2016.
- [5] A. Tarantola, “Inversion of seismic reflection data in the acoustic approximation,” *Geophysics*, vol. 49, no. 8, pp. 1259–1266, 1984.
- [6] O. Lindtjorn, R. G. Clapp, O. Pell, H. Fu, M. J. Flynn, and O. Mencer, “Beyond traditional microprocessors for geoscience high-performance computing applications,” *Ieee Micro*, vol. 31, no. 2, pp. 41–49, 2011.
- [7] Wolt, C, “Introduction to geology, sedimentary rocks.” <http://webpace.ship.edu/cjwolt/geology/slides/sed-sum.htm>. Visited: June of 2016.
- [8] DeMets, Chuck, “Lecture 3, how do rocks deform?.” [http://geoscience.wisc.edu/~sim\\$chuck/Classes/Mtn_and_Plates/rock_deformation.html](http://geoscience.wisc.edu/~sim$chuck/Classes/Mtn_and_Plates/rock_deformation.html). Visited: June of 2016.
- [9] R. G. Pratt, C. Shin, and G. Hick, “Gauss–newton and full newton methods in frequency–space seismic waveform inversion,” *Geophysical Journal International*, vol. 133, no. 2, pp. 341–362, 1998.

- [10] K. S. Yee and J. S. Chen, “The finite-difference time-domain (fdtd) and the finite-volume time-domain (fvtd) methods in solving maxwell’s equations,” *IEEE Transactions on Antennas and Propagation*, vol. 45, no. 3, pp. 354–363, 1997.
- [11] D. M. Sullivan, *Electromagnetic simulation using the FDTD method*. John Wiley & Sons, 2013.
- [12] J. Gough *et al.*, “To infinity and beyond,” *Prime Number*, vol. 30, no. 1, p. 18, 2015.
- [13] D. Pasalic, R. McGarry, *et al.*, “Convolutional perfectly matched layer for isotropic and anisotropic acoustic wave equations,” in *2010 SEG Annual Meeting*, Society of Exploration Geophysicists, 2010.
- [14] F. Collino and P. B. Monk, “Optimizing the perfectly matched layer,” *Computer methods in applied mechanics and engineering*, vol. 164, no. 1, pp. 157–171, 1998.
- [15] C. A. P. Solano, *Two-dimensional near-surface seismic imaging with surface waves: alternative methodology for waveform inversion*. PhD thesis, Ecole Nationale Supérieure des Mines de Paris, 2013.
- [16] R. Slawinski, *Seismic Wave Propagation in Fractured Media*. PhD thesis, University of Calgary, 1999.
- [17] P. Lailly, “The seismic inverse problem as a sequence of before stack migrations,” in *Conference on inverse scattering: theory and application*, pp. 206–220, Society for Industrial and Applied Mathematics, Philadelphia, PA, 1983.
- [18] C. Ravaut, S. Operto, L. Improta, J. Virieux, A. Herrero, and P. Dell’Aversana, “Multiscale imaging of complex structures from multifold wide-aperture seismic data by frequency-domain full-waveform tomography: application to a thrust belt,” *Geophysical Journal International*, vol. 159, no. 3, pp. 1032–1056, 2004.
- [19] M. S. Zhdanov, *Geophysical inverse theory and regularization problems*, vol. 36. Elsevier, 2002.
- [20] A. A. Goldstein, “On newton’s method,” *Numerische Mathematik*, vol. 7, no. 5, pp. 391–393, 1965.
- [21] J. Nocedal and S. Wright, *Numerical optimization*. Springer Science & Business Media, 2006.
- [22] C. Bunks, F. M. Saleck, S. Zaleski, and G. Chavent, “Multiscale seismic waveform inversion,” *Geophysics*, vol. 60, no. 5, pp. 1457–1473, 1995.
- [23] M. K. Sen and P. L. Stoffa, *Global optimization methods in geophysical inversion*. Cambridge University Press, 2013.

- [24] L. Métivier, R. Brossier, S. Operto, J. Virieux, *et al.*, “Second-order adjoint state methods for full waveform inversion,” in *EAGE 2012-74th European Association of Geoscientists and Engineers Conference and Exhibition*, 2012.
- [25] G. Arce, “Weiner filtering.” https://www.eecis.udel.edu/~arce/Courses_files/Chapter6_WienerFiltering.pdf. Visited: July of 2015.
- [26] L. Métivier, R. Brossier, J. Virieux, and S. Operto, “Full waveform inversion and the truncated newton method,” *SIAM Journal on Scientific Computing*, vol. 35, no. 2, pp. B401–B437, 2013.
- [27] Young, “Young’s theorem.” https://web.archive.org/web/20060518134739/http://are.berkeley.edu/courses/ARE210/fall12005/lecture_notes/Young%27s-Theorem.pdf. Visited: June of 2016.
- [28] J. Virieux and S. Operto, “An overview of full-waveform inversion in exploration geophysics,” *Geophysics*, vol. 74, no. 6, pp. WCC1–WCC26, 2009.
- [29] B. Fornberg, “The pseudospectral method: Comparisons with finite differences for the elastic wave equation,” *Geophysics*, vol. 52, no. 4, pp. 483–501, 1987.
- [30] R. C. Pestana and P. L. Stoffa, “Time evolution of the wave equation using rapid expansion method,” *Geophysics*, vol. 75, no. 4, pp. T121–T131, 2010.
- [31] J. Kennedy, “Particle swarm optimization,” in *Encyclopedia of machine learning*, pp. 760–766, Springer, 2011.
- [32] A. Fichtner and J. Trampert, “Resolution analysis in full waveform inversion,” *Geophysical Journal International*, vol. 187, no. 3, pp. 1604–1624, 2011.
- [33] G. Nolet, R. Montelli, and J. Virieux, “Explicit, approximate expressions for the resolution and a posteriori covariance of massive tomographic systems,” *Geophysical Journal International*, vol. 138, no. 1, pp. 36–44, 1999.
- [34] S. C. Chapra and R. P. Canale, *Numerical methods for engineers*, vol. 2. McGraw-Hill New York, 2012.
- [35] D. Greaves, “Lu decomposition.” <http://www.cl.cam.ac.uk/teaching/1314/NumMethods/supporting/mcmaster-kiruba-ludecomp.pdf>. Visited: May of 2016.
- [36] J. A. Scales and L. Tenorio, “Prior information and uncertainty in inverse problems,” *Geophysics*, vol. 66, no. 2, pp. 389–397, 2001.

- [37] J. A. Scales and R. Snieder, "To bayes or not to bayes?," *Geophysics*, vol. 62, no. 4, pp. 1045–1046, 1997.
- [38] Z. Fang, C. Silva, and F. Herrmann, "Fast uncertainty quantification for 2d full-waveform inversion with randomized source subsampling. 76th eage," 2014.
- [39] Z. Fang, F. J. Herrmann, and C. Y. Lee, "A stochastic quasi-newton mcmc method for uncertainty quantification of full-waveform inversion,"
- [40] A. G. Kalmikov and P. Heimbach, "A hessian-based method for uncertainty quantification in global ocean state estimation," *SIAM Journal on Scientific Computing*, vol. 36, no. 5, pp. S267–S295, 2014.
- [41] RISC, "Super computers colombia." <http://risc-project.eu/supercomputers-country/colombia/>. Visited: July of 2016.
- [42] C. R. Rao and S. K. Mitra, *Generalized inverse of matrices and its applications*, vol. 7. Wiley New York, 1971.
- [43] McCallum, John C. , "Memory prices (1957-2016)." <http://www.jcmit.com/memoryprice.htm>. Visited: September of 2016.
- [44] A. Fichtner and J. Trampert, "Hessian kernels of seismic data functionals based upon adjoint techniques," *Geophysical Journal International*, vol. 185, no. 2, pp. 775–798, 2011.
- [45] N. J. Higham, "Computing the nearest correlation matrix - a problem from finance," *IMA journal of Numerical Analysis*, vol. 22, no. 3, pp. 329–343, 2002.
- [46] A. Fournier, K. Mosegaard, H. Omre, M. Sambridge, and L. Tenorio, "Assessing uncertainty in geophysical problems-introduction," *Geophysics*, vol. 78, no. 3, pp. WB1–WB2, 2013.
- [47] R. C. Aster, B. Borchers, and C. H. Thurber, *Parameter estimation and inverse problems*. Academic Press, 2013.
- [48] W. H. K. L. Keiiti Aki, "Determination of three-dimensional velocity anomalies under a seismic array using first P arrival times from local earthquakes. 1. A homogeneous initial model," vol. 81, no. 23, 1976.

Bibliography

- A. A. Goldstein, “On newton’s method,” *Numerische Mathematik*, vol. 7, no. 5, pp. 391–393, 1965
- A. Fichtner and J. Trampert, “Hessian kernels of seismic data functionals based upon adjoint techniques,” *Geophysical Journal International*, vol. 185, no. 2, pp. 775–798, 2011
- A. Fichtner and J. Trampert, “Resolution analysis in full waveform inversion,” *Geophysical Journal International*, vol. 187, no. 3, pp. 1604–1624, 2011
- A. Fournier, K. Mosegaard, H. Omre, M. Sambridge, and L. Tenorio, “Assessing uncertainty in geophysical problems-introduction,” *Geophysics*, vol. 78, no. 3, pp. WB1–WB2, 2013
- A. G. Kalmikov and P. Heimbach, “A hessian-based method for uncertainty quantification in global ocean state estimation,” *SIAM Journal on Scientific Computing*, vol. 36, no. 5, pp. S267–S295, 2014
- A. Tarantola, “Inversion of seismic reflection data in the acoustic approximation,” *Geophysics*, vol. 49, no. 8, pp. 1259–1266, 1984
- B. Fornberg, “The pseudospectral method: Comparisons with finite differences for the elastic wave equation,” *Geophysics*, vol. 52, no. 4, pp. 483–501, 1987
- C. A. P. Solano, *Two-dimensional near-surface seismic imaging with surface waves: alternative methodology for waveform inversion*. PhD thesis, Ecole Nationale Supérieure des Mines de Paris, 2013

- C. Bunks, F. M. Saleck, S. Zaleski, and G. Chavent, “Multiscale seismic waveform inversion,” *Geophysics*, vol. 60, no. 5, pp. 1457–1473, 1995
- C. Ravaut, S. Operto, L. Imbrota, J. Virieux, A. Herrero, and P. Dell’Aversana, “Multiscale imaging of complex structures from multifold wide-aperture seismic data by frequency-domain full-waveform tomography: application to a thrust belt,” *Geophysical Journal International*, vol. 159, no. 3, pp. 1032–1056, 2004
- C. R. Rao and S. K. Mitra, *Generalized inverse of matrices and its applications*, vol. 7. Wiley New York, 1971
- DeMets, Chuck, “Lecture 3, how do rocks deform?.” [http://geoscience.wisc.edu/\\$\sim\\$sim\\$chuck/Classes/Mtn_and_Plates/rock_deformation.html](http://geoscience.wisc.edu/$\sim$sim$chuck/Classes/Mtn_and_Plates/rock_deformation.html). Visited: June of 2016
- D. Greaves, “Lu decomposition.” <http://www.cl.cam.ac.uk/teaching/1314/NumMethods/supporting/mcmaster-kiruba-ludecomp.pdf>. Visited: May of 2016
- D. M. Sullivan, *Electromagnetic simulation using the FDTD method*. John Wiley & Sons, 2013
- D. Pasalic, R. McGarry, *et al.*, “Convolutional perfectly matched layer for isotropic and anisotropic acoustic wave equations,” in *2010 SEG Annual Meeting*, Society of Exploration Geophysicists, 2010
- F. Collino and P. B. Monk, “Optimizing the perfectly matched layer,” *Computer methods in applied mechanics and engineering*, vol. 164, no. 1, pp. 157–171, 1998
- G. Arce, “Weiner filtering.” https://www.eecis.udel.edu/~arce/Courses_files/Chapter6_WienerF.pdf. Visited: July of 2015
- G. Nolet, R. Montelli, and J. Virieux, “Explicit, approximate expressions for the resolution and a posteriori covariance of massive tomographic systems,” *Geophysical Journal International*, vol. 138, no. 1, pp. 36–44, 1999
- Irons, Trevor, “Marmousi model.” <http://www.reproducibility.org/RSF/book/data/marmousi/paper.pdf>. Visited: June of 2016

Irons, Trevor, "Sigsbee model." http://www.reproducibility.org/RSF/book/data/sigsbee/paper_html/paper.html. Visited: June of 2016

J. A. Scales and L. Tenorio, "Prior information and uncertainty in inverse problems," *Geophysics*, vol. 66, no. 2, pp. 389–397, 2001

J. A. Scales and R. Snieder, "To bayes or not to bayes?," *Geophysics*, vol. 62, no. 4, pp. 1045–1046, 1997

J. Gough *et al.*, "To infinity and beyond," *Prime Number*, vol. 30, no. 1, p. 18, 2015

J. Kennedy, "Particle swarm optimization," in *Encyclopedia of machine learning*, pp. 760–766, Springer, 2011

J. Nocedal and S. Wright, *Numerical optimization*. Springer Science & Business Media, 2006

J. Virieux and S. Operto, "An overview of full-waveform inversion in exploration geophysics," *Geophysics*, vol. 74, no. 6, pp. WCC1–WCC26, 2009

K. S. Yee and J. S. Chen, "The finite-difference time-domain (fdtd) and the finite-volume time-domain (fvtd) methods in solving maxwell's equations," *IEEE Transactions on Antennas and Propagation*, vol. 45, no. 3, pp. 354–363, 1997

L. Métivier, R. Brossier, S. Operto, J. Virieux, *et al.*, "Second-order adjoint state methods for full waveform inversion," in *EAGE 2012-74th European Association of Geoscientists and Engineers Conference and Exhibition*, 2012

L. Métivier, R. Brossier, J. Virieux, and S. Operto, "Full waveform inversion and the truncated newton method," *SIAM Journal on Scientific Computing*, vol. 35, no. 2, pp. B401–B437, 2013

McCallum, John C. , "Memory prices (1957-2016)." <http://www.jcmit.com/memoryprice.htm>. Visited: September of 2016

M. K. Sen and P. L. Stoffa, *Global optimization methods in geophysical inversion*. Cambridge University Press, 2013

M. S. Zhdanov, *Geophysical inverse theory and regularization problems*, vol. 36. Elsevier, 2002

Narcowich, Francis J., "Math 414, haar wavelet analysis." [http://www.math.tamu.edu/~sim\\$fnarc/m641/f07/m414w_ln29.html](http://www.math.tamu.edu/~sim$fnarc/m641/f07/m414w_ln29.html). Visited: June of 2016

N. J. Higham, "Computing the nearest correlation matrix - a problem from finance," *IMA journal of Numerical Analysis*, vol. 22, no. 3, pp. 329–343, 2002

O. Lindtjorn, R. G. Clapp, O. Pell, H. Fu, M. J. Flynn, and O. Mencer, "Beyond traditional microprocessors for geoscience high-performance computing applications," *Ieee Micro*, vol. 31, no. 2, pp. 41–49, 2011

P. Lailly, "The seismic inverse problem as a sequence of before stack migrations," in *Conference on inverse scattering: theory and application*, pp. 206–220, Society for Industrial and Applied Mathematics, Philadelphia, PA, 1983

R. C. Aster, B. Borchers, and C. H. Thurber, *Parameter estimation and inverse problems*. Academic Press, 2013

R. C. Pestana and P. L. Stoffa, "Time evolution of the wave equation using rapid expansion method," *Geophysics*, vol. 75, no. 4, pp. T121–T131, 2010

R. G. Pratt, C. Shin, and G. Hick, "Gauss–newton and full newton methods in frequency–space seismic waveform inversion," *Geophysical Journal International*, vol. 133, no. 2, pp. 341–362, 1998

R. Slawinski, *Seismic Wave Propagation in Fractured Media*. PhD thesis, University of Calgary, 1999

S. C. Chapra and R. P. Canale, *Numerical methods for engineers*, vol. 2. McGraw-Hill New York, 2012

W. H. K. L. Keiiti Aki, "Determination of three-dimensional velocity anomalies under a seismic array using first P arrival times from local earthquakes. 1. A homogeneous initial model," vol. 81, no. 23, 1976

Wolt, C, "Introduction to geology, sedimentary rocks." <http://webpace.ship.edu/cjwolt/geology/slides/sed-sum.htm>. Visited: June of 2016

Z. Fang, C. Silva, and F. Herrmann, “Fast uncertainty quantification for 2d full-waveform inversion with randomized source subsampling. 76th eage,” 2014

Z. Fang, F. J. Herrmann, and C. Y. Lee, “A stochastic quasi-newton mcmc method for uncertainty quantification of full-waveform inversion,”