

# Important Notice

This copy may be used only for the purposes of research and private study, and any use of the copy for a purpose other than research or private study may require the authorization of the copyright owner of the work in question. Responsibility regarding questions of copyright that may arise in the use of this copy is assumed by the recipient.

UNIVERSITY OF CALGARY

Crosstalk suppression in multiparameter full waveform inversion  
through parameter decorrelation

by

Mariana Carolina Lume Larez

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES  
IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE  
DEGREE OF MASTER OF SCIENCE

GRADUATE PROGRAM IN GEOSCIENCE

CALGARY, ALBERTA

SEPTEMBER, 2023

© Mariana Carolina Lume Larez 2023

# Abstract

Multiparameter full waveform inversion (FWI) is a promising technique to estimate the elastic properties of the subsurface, but it is commonly affected by crosstalk between parameters of different nature, which impacts the convergence of the local optimization algorithms and introduces artifacts and uncertainty to the results. Hence, reducing these effects is essential to increase confidence in the estimates. This thesis is focused on proposing strategies to treat these artifacts in the model space of the density ( $\rho$ ), P-wave velocity ( $V_P$ ), and S-wave velocity ( $V_S$ ) while considering seismic surface experiments. The crosstalk effects are described and corrected by the Hessian operator, which also affects the shape of the iso-surfaces of the objective function. Therefore, two methodologies were developed based on performing constrained re-parameterizations, aiming to find an intermediate model space with decorrelated parameter classes, i.e., where the Hessian is the identity matrix, to reach convergence to an accurate minimum point, and later transform it into the original model space. The difference between both strategies lies in the type and size of the Hessian matrix used, i.e., point-wise Hessians and point-probes Hessians, as well as in the numerical approach employed to compute the transformation matrices necessary to map between model spaces. In both strategies, the estimates of  $V_S$  were relatively accurate but the results of  $V_P$  and  $\rho$  were strongly impacted by crosstalk effects in comparison to those obtained with FWI approaches that were not re-parameterized; thus, the sought intermediate model space was not properly mapped. The decorrelation ideas were successful in certain aspects, but the challenges were related to the limitations brought by the amount of crosstalk information considered through the Hessians, the selected numerical approaches, and the type of transformation matrix computed, which was able to do a good job in some locations of the model grid or for some parameter classes, but was not general enough to produce the expected transformation in a large scale.

# Acknowledgements

I would like to start by expressing my sincere gratitude to my supervisor Dr. Kris Innanen for encouraging me to work on this research topic, as well as for giving me valuable explanations, feedback, support, and understanding throughout the development of my thesis and, in general, of my graduate studies.

Likewise, I would like to thank Dr. Daniel Trad for allowing me to be one of his teaching assistants, helping me with side projects, offering rewarding suggestions as well as showing me different perspectives about the research topic of this thesis. Moreover, I would like to thank Dr. Brandon Karchewski for being part of my thesis defense as a committee member.

On the other hand, I am tremendously grateful to Dr. Scott Keating for kindly and patiently guiding me with the incorporation of the decorrelation ideas to the elastic FWI process and helping me with the formulation of the Simulated Annealing technique in the FWI problem treated in this thesis. Moreover, I am profoundly grateful to Dr. Iván Sánchez for his motivational words as well as for his suggestions and guidance in the implementation of the workflow to extract empirical radiation patterns.

I am also thankful to the NSERC and the sponsors of the CREWES project for the financial support that was provided to me and the students of the research group.

I would like to thank Kevin Hall and Kevin Bertram for helping me solve technical issues in my workstation, for the nice conversations in the CREWES office, and for making the CREWES annual meetings smooth and easier for everyone.

Likewise, I am grateful to all my CREWES research fellows and in particular to Ziguang Su, Ninoska Amundaray, Shang Huang, Tianze Zhang, Lukas Sadownyk, Qi Hu, Chioma Chineke, Luping Qu, and Xin Fu, for some of you being study partners, others being my TAs, for offer-

ing advice and suggestions or just listening to my research ideas, for attending to my technical presentations, and for the good conversations that felt like fresh air after long periods of work.

I would like to express my heartfelt gratitude to my husband for being a fundamental part of my success by staying late working with me, helping me in different stages of my graduate studies, listening to my presentation practices, being my emotional support, and for always being loving and patient with me. In addition, I would like to deeply thank my parents and grandma for encouraging me to follow my dreams, work hard, and be kind, and for supporting me beyond the distance. Last but not least, I am grateful to God for all the opportunities that I have received in my life and for giving me the strength and determination to reach my goals and overcome the challenges that appear in my path.

# Dedication

*A mis queridos padres, abuelita y esposo. Gracias por todo el esfuerzo que han hecho por mí y por el apoyo que me han dado para alcanzar mis metas y sueños. Los amo mucho.*

# Table of Contents

<b>Abstract</b>	<b>ii</b>
<b>Acknowledgements</b>	<b>iii</b>
<b>Dedication</b>	<b>v</b>
<b>Table of Contents</b>	<b>vi</b>
<b>List of Figures and Illustrations</b>	<b>ix</b>
<b>List of Tables</b>	<b>xiv</b>
<b>List of Symbols, Abbreviations and Nomenclature</b>	<b>xv</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Background . . . . .	1
1.2 Full waveform inversion . . . . .	2
1.3 Challenges in the application of full waveform inversion . . . . .	4
1.4 Thesis objectives and organization . . . . .	9
<b>2 Assessment of crosstalk under scenarios of homogeneous and heterogeneous back-grounds</b>	<b>12</b>
2.1 Introduction . . . . .	13
2.2 Scattering theory in the $\rho$ , $V_p$ , and $V_s$ model space . . . . .	14
2.3 Numerical experiments . . . . .	18
2.3.1 Extraction of empirical radiation patterns . . . . .	18

2.3.2	Radiation patterns under heterogeneous reference media . . . . .	20
2.4	Conclusions . . . . .	24
<b>3</b>	<b>Decorrelation of parameter classes and crosstalk minimization with point-wise Hessians</b>	<b>27</b>
3.1	Introduction . . . . .	28
3.2	Theory and Methods . . . . .	29
3.2.1	Full waveform inversion . . . . .	29
3.2.2	Objective function and forward modeling . . . . .	30
3.2.3	Re-parameterization of the FWI problem . . . . .	33
3.2.4	Optimization algorithms and operators . . . . .	34
3.2.4.1	The gradient . . . . .	35
3.2.4.2	The Hessian . . . . .	37
3.2.4.3	The adjoint-state method for the gradient and the Hessian . . . . .	38
3.2.4.4	Steepest descent optimization method . . . . .	41
3.2.4.5	Gauss-Newton optimization method . . . . .	42
3.2.4.6	The line search . . . . .	43
3.2.4.7	Impact of the Hessian in the convergence of the algorithms . . . . .	44
3.2.5	Numerical procedure to compute the transformation matrix . . . . .	45
3.3	Workflow and its adaptation for computational feasibility . . . . .	48
3.4	Numerical experiments . . . . .	50
3.5	Conclusions . . . . .	57
<b>4</b>	<b>Using Simulated Annealing to optimize an AVO nonlinear problem: Gaining valuable insights</b>	<b>59</b>
4.1	Introduction . . . . .	60
4.2	Forward Modeling . . . . .	62
4.3	Iterative nonlinear AVO simultaneous inversion . . . . .	64

4.4	Simulated Annealing . . . . .	66
4.5	Numerical experiments . . . . .	70
4.5.1	Broadband reflectivities . . . . .	74
4.5.2	Band-limited reflectivities . . . . .	76
4.5.2.1	Increasing the bounds in SA . . . . .	79
4.5.3	Insights into integrating SA to other geophysical seismic problems . . . . .	83
4.6	Conclusions . . . . .	89
<b>5</b>	<b>Decorrelation of parameter classes and crosstalk minimization with point-probes</b>	
	<b>Hessians</b>	<b>91</b>
5.1	Introduction . . . . .	92
5.2	Design of an objective function to be minimized with SA . . . . .	93
5.2.1	Workflow and evaluation of the objective function . . . . .	95
5.3	Pros and cons of SA for the current problem . . . . .	96
5.4	Numerical experiments . . . . .	99
5.4.1	Analysis . . . . .	104
5.5	Conclusions . . . . .	105
<b>6</b>	<b>Conclusions</b>	<b>110</b>
6.1	Summary . . . . .	110
6.2	Future work . . . . .	113
	<b>References</b>	<b>115</b>

# List of Figures and Illustrations

2.1	Representation of a localized inhomogeneity of size $L$ from which the scattered wavefield is generated. $r$ , $\psi$ , and $\zeta$ are spherical coordinates and $\mathbf{e}_r$ , $\mathbf{e}_\psi$ and $\mathbf{e}_\zeta$ are the unit base vectors in that coordinate system (Modified from Sato et al. (2012)). . . . .	15
2.2	Mathematical procedure to compute the scattered wavefield caused by a perturbation of one parameter class. The blue dot corresponds to the scatter point, while the yellow star represents the explosive source. . . . .	19
2.3	Proposed workflow to extract the empirical radiation patterns from simulated wavefields.(a) Extraction of patterns from the P-P wavefield. (b) Extraction of patterns from the P-S wavefield. $R$ corresponds to the radius indicated by the user. . . . .	20
2.4	Comparison between the normalized analytic and empirical radiation patterns of P-P and P-S wavefields. Values were normalized. . . . .	21
2.5	Different heterogeneous reference media used in this study. (a) Case 1, (b) Case 2, and (c) Case 3. . . . .	21
2.6	Overlapped radiation patterns of analytic, homogeneous, and heterogeneous cases extracted from the P-P wavefield. Values were normalized. . . . .	22
2.7	Overlapped radiation patterns of analytic, homogeneous, and heterogeneous cases extracted from the P-S wavefield. Values were normalized. . . . .	22
2.8	Crosstalk assessment with radiation patterns extracted from P-P wavefield. Values were not normalized to study the amplitude variations.(a) Analytic expressions, (b) Case 1, (c) Case 2, and (d) Case 3. . . . .	25

2.9	Crosstalk assessment with radiation patterns extracted from P-S wavefield. Values were not normalized to study the amplitude variations. (a) Analytic expressions, (b) Case 1, (c) Case 2, and (d) Case 3. . . . .	26
3.1	Illustration of a full Hessian, modified from (Métivier et al., 2015), and how a point-wise Hessian can be constructed from it. . . . .	39
3.2	Illustration of a full Hessian, modified from (Métivier et al., 2015), and how a point-probes Hessian can be constructed from it. . . . .	39
3.3	Two-variable model of descent-based optimization. a) the Hessian in the quadratic objective function is not the identity matrix; b) the Hessian in the quadratic objective function is the identity matrix. SD means "steepest descent". . . . .	45
3.4	Re-parameterization workflow followed in this study. . . . .	49
3.5	True $\rho$ , $V_P$ and $V_S$ models. . . . .	51
3.6	Models estimated with a baseline FWI, i.e., without re-parameterization. . . . .	51
3.7	Point-probes Hessian computed with baseline estimates, after perturbing parameters at location $x=50$ and $z=20$ . The $3 \times 3$ matrix is the local Hessian at the same location of the perturbation. . . . .	52
3.8	Point-wise Hessians computed at different locations of the baseline estimates. . . . .	52
3.9	Normalized crosstalk metrics computed with baseline estimates after perturbing at different locations to compute the associated point-probes Hessians. . . . .	53
3.10	Models estimated with a re-parameterized FWI after selecting location $x=50$ and $z=20$ to compute $\mathbf{T}$ . . . . .	53
3.11	Normalized crosstalk metric computed with estimates obtained after selecting the location $x=50$ and $z=20$ to compute $\mathbf{T}$ ( $r$ model space). Parameter classes were perturbed at location $x=50$ and $z=20$ to compute the associated point-probes Hessians. . . . .	54
3.12	Point-probes Hessian computed with estimates obtained from a re-parameterized FWI after selecting the location $x=50$ and $z=20$ to calculate $\mathbf{T}$ ( $r$ model space). The $3 \times 3$ matrix is the local Hessian at the same location of the perturbation. . . . .	54

3.13	Models estimated with a re-parameterized FWI using different grid locations to compute $\mathbf{T}$ : (a) $x=20$ and $z=20$ , (b) $x=50$ and $z=50$ , (c) $x=80$ and $z=80$ . . . . .	55
3.14	Horizontal and vertical profiles extracted from the estimates obtained with a re-parameterized FWI using different locations to compute $\mathbf{T}$ . . . . .	56
3.15	Point-wise Hessian matrices at different locations of the estimates in the $r$ system after inverting with different locations to compute $\mathbf{T}$ . . . . .	57
4.1	Representation of the P-P and P-S incidence angles implemented in this study (modified from Stewart et al. (1999). “MP” stands for midpoint and “CP” stands for conversion point. . . . .	64
4.2	Pseudo-code of the Metropolis Algorithm applied in this study. . . . .	69
4.3	True (continuous lines) and initial model (dashed lines) used for the nonlinear AVO inversion. . . . .	71
4.4	P-P and P-S reflection coefficients of the studied two-layer model. . . . .	71
4.5	Evolution of the estimated model parameters obtained from broadband data with different optimization methods. Continuous lines represent the estimates and dashed lines the true models. . . . .	75
4.6	Evolution of the objective function produced from inverting broadband data with different optimization methods. . . . .	76
4.7	Evolution of the estimated $V_P$ and $V_S$ values for the first layer obtained from broadband data with case 3 of SA. Continuous lines represent the estimates and dashed lines the true models. . . . .	77
4.8	Percentage of new configurations accepted per temperature while inverting broadband data for the three cases of SA. . . . .	78
4.9	Accuracy tests performed with the final estimates obtained from broadband data using different optimization methods. . . . .	79
4.10	Ormsby wavelet used to filter the P-P dataset a) in time and b) in frequency. Ormsby wavelet used to filter the P-S dataset c) in time and d) in frequency. . . . .	80

4.11	Syntethic seismograms for the a) P-P dataset and b) P-S dataset. . . . .	80
4.12	Evolution of the estimated model parameters obtained from band-limited data with different optimization methods. Continuous lines represent the estimates and dashed lines the true models. . . . .	81
4.13	Evolution of the objective function produced from inverting band-limited data with different optimization methods. . . . .	82
4.14	Percentage of new configurations accepted per temperature while inverting band-limited data for the three cases of SA. . . . .	83
4.15	Accuracy tests performed with the final estimates obtained from band-limited data using different optimization methods. . . . .	84
4.16	Evolution of the estimated $V_P$ and $V_S$ values for the first layer obtained from band-limited data with case 3 of SA. Continuous lines represent the estimates and dashed lines the true model values. . . . .	84
4.17	Evolution of the estimated model parameters obtained from band-limited data with different optimization methods. In these experiments, the bounds of the SA were larger. Continuous lines represent the estimates and dashed lines the true models. . . . .	85
4.18	Evolution of the objective function produced from inverting band-limited data with different optimization methods. In these experiments, the bounds of the SA were larger. . . . .	86
4.19	Percentage of new configurations accepted per temperature while inverting band-limited data for the three cases of SA and using larger bounds. . . . .	87
4.20	Accuracy tests performed with the final estimates obtained from band-limited data with the applied optimization methods. In these experiments, the bounds of the SA were larger. . . . .	88
4.21	Evolution of the estimated $V_P$ and $V_S$ values for the first layer obtained from band-limited data with case 3 of SA and using larger bounds. Continuous lines represent the estimates and dashed lines the true models. . . . .	88

5.1	a) Representation of the point-probes Hessian, in the $r$ model space, and the nomenclature used for each block in the objective function. b) Structure of the point-wise Hessians that is expected to be produced. . . . .	94
5.2	Re-parameterization workflow followed in this study. . . . .	95
5.3	a) True $\rho$ , $V_P$ , and $V_S$ models. b) Models estimated with a baseline FWI, i.e., without re-parameterization. c) Models estimated with a re-parameterized FWI, using the point-wise Hessian, after selecting location $x=50$ and $z=20$ to compute $\mathbf{T}$ . d) Models estimated with a re-parameterized FWI, using the point-probes Hessian, after selecting location $x=50$ and $z=20$ to compute $\mathbf{T}$ . . . . .	101
5.4	Vertical and horizontal profiles extracted from the true and estimated models in Figure 5.3. The grey lines on the true models correspond to the selected profiles. . . . .	102
5.5	Curves to evaluate the performance of SA at some frequency bands of the FWI. a) Frequency band 2, b) frequency band 4, c) frequency band 6, and d) frequency band 8. . . . .	103
5.6	Summarized versions of the point-probes Hessians in the $s$ and the $r$ model spaces at different frequency bands of the FWI process. a) Frequency band 2, b) frequency band 4, c) frequency band 6, and d) frequency band 8. . . . .	107
5.7	Summary of the point-probes Hessian obtained in the $r$ model space and at the frequency band 8, as well as different extracted point-wise Hessians. . . . .	108
5.8	Representation of the transformation of different point-wise Hessians into the $r$ model space using a unique $3 \times 3$ matrix $\mathbf{T}$ for the frequency band 8. The values of the local Hessians are not normalized. . . . .	109

# List of Tables

4.1	Annealing schedules applied to the SA algorithm when inverting broadband data. .	73
4.2	Annealing schedules applied to the SA algorithm when inverting band-limited data.	73
4.3	Annealing schedules applied to the SA algorithm using larger bounds, when inverting band-limited data. . . . .	73
5.1	Annealing schedules applied to the SA algorithm at each frequency band of the FWI.	99

# List of Symbols, Abbreviations and Nomenclature

Symbol or Abbreviation	Definition
1C	One component
2D	Two dimensions in space
3D	Three dimensions in space
AVO	Amplitude Variation with Offset
CO <sub>2</sub>	Carbon dioxide
CP	Conversion point
FWI	Full Waveform Inversion
L-BFGS	Limited-memory Broyden-Fletcher-Goldfarb-Shanno
MP	Midpoint
P-P	P to P mode energy conversion. P wave impinges the reflector and P wave reflects
P-S	P to S mode energy conversion. P wave impinges the reflector and S wave reflects
PML	Perfectly Matched Layers

PSDM	Pre-Stack Depth Migration
PSF	Point Spread Functions
SA	Simulated Annealing
$\Delta\rho/\rho$	Fractional density
$\Delta I/I$	Fractional acoustic impedance
$\Delta J/J$	Fractional shear impedance
<b>A</b>	Wavefield operator
<b>b</b>	Vector of constant terms in Zoeppritz equations
<b>C</b>	A constraint for an optimization problem
<b>d<sub>pred</sub></b>	Predicted data using estimated models
<b>d</b>	A set of recorded or measured data
<b>e<sub>1</sub>, e<sub>2</sub>, e<sub>3</sub></b>	Unit base vectors in cartesian coordinates
<b>e<sub>r</sub>, e<sub>ψ</sub>, e<sub>ζ</sub></b>	Unit base vectors in spherical coordinates
<b>f</b>	A set of seismic source terms
<b>F<sup>PP</sup></b>	Vector of amplitudes of a P scattered wavefield produced by a incident P-wave
<b>F<sup>PS</sup></b>	Vector of amplitudes of a S scattered wavefield produced by a incident P-wave
<b>g</b>	The gradient
<b>H</b>	The Hessian matrix

<b>I</b>	Identity matrix
<b>J</b>	The Jacobian matrix
<b>m</b>	A subsurface model in the AVO problem
<b>n, k</b>	Vectors in one approach to obtain <b>T</b>
<b>P</b>	Matrix of coefficients in Zoeppritz equations
<b>p</b>	Search direction
<b>q</b>	Coefficient of the linear term of a quadratic equation
<b>R</b>	Receiver sampling matrix
<b>S</b>	Sampling operator
<b>s</b>	A subsurface model in the original model space
<b>T</b>	Transformation matrix
<b><math>\mathbf{u}^{0P}</math></b>	An incident P wavefield
<b><math>\mathbf{u}^0</math></b>	An incident wavefield
<b><math>\mathbf{u}^1</math></b>	An scattered wavefield
<b><math>\mathbf{u}</math></b>	A total wavefield
<b><math>\mathbf{u}^{1PP}</math></b>	A P scattered wavefield produced by a incident P-wave
<b><math>\mathbf{u}^{1PS}</math></b>	A S scattered wavefield produced by a incident P-wave
<b>w, v</b>	Arbitrary vectors
<b>x</b>	Vector of positions
<b>X, Y, Z, <math>\mathbf{T}'</math>, <math>\tilde{\mathbf{T}}</math></b>	Tensors in one numerical approach to compute <b>T</b>

$A, B, C, D$	Contrasts of elastic properties in the AVO problem
$C_1$	Constant value for the Armijo condition
$C_2$	Constant value for the curvature condition
$c_{11}, c_{44}$	Isotropic-elastic stiffness coefficients
$E$	Energy of a system
$e$	Cooling factor in Simulated Annealing Algorithms
$f_x$	x-component of a set of seismic sources
$f_z$	z-component of a set of seismic sources
$h_{i,j}$	Elements of the matrix $\mathbf{H}$
$I$	Acoustic Impedance
$J$	Shear Impedance
$k_0$	Wavenumber of a P-wave
$k_B$	Boltzmann's constant
$L$	Lagrangian
$l_0$	Wavenumber of an S-wave
$M$	Number of dimensions
$N$	Total number of incidence angles
$N_f$	Number of discrete frequencies
$N_s$	Number of discrete sources
$n_x$	Number of samples in the horizontal direction of the model grid

$n_z$	Number of samples in the vertical direction of the model grid
$P$	Boltzmann's distribution
$R$	Radius of the masks
$r$	Intermediate model space or system
$r_1, r_2, r_3$	Model parameters in the intermediate model space $r$
$R_a$	Total number of atomic positions
$R_{PP}, R_{PS}$	Reflection coefficients of Zoeppritz equations
$s$	Original model space or system
$s_\mu, s_{i,j}$	The $\mu$ th or equivalently the $i, j$ element of $\mathbf{s}$
$T$	Temperature in the simulated Annealing algorithm
$t$	Time
$t_\mu^\lambda, t_\nu^\sigma$	Elements of the matrix $\mathbf{T}$
$T_{PP}, T_{PS}$	Transmission coefficients of Zoeppritz equations
$u_x$	Horizontal displacement wavefield
$u_z$	Vertical displacement wavefield
$V_P$	P-wave velocity
$V_S$	S-wave velocity
$v_\mu$	The $\mu$ th element of $\mathbf{v}$
$V_{P0}$	P-wave velocity of the background medium
$V_{S0}$	S-wave velocity of the background medium

$X$	Sine of an angle
$x$	Position in the horizontal direction
$Z$	Normalizing constant of the Boltzmann's distribution
$z$	Position in the vertical direction
$\alpha$	Step-length in an optimization algorithm
$\chi$	Fractional fluctuation of $\mu$
$\varepsilon$	Random number used in Simulated Annealing
$\eta, \beta, \zeta$	Coefficients of the terms in quadratic equation to obtain $\mathbf{T}$
$\gamma_0$	$V_P/V_S$ ratio of the background media
$\kappa, \xi$	Unconstrained Lagrange multipliers
$\lambda, \mu$	Lamé coefficients
$\lambda_0, \mu_0$	Lamé coefficients of the background medium
$\omega$	Angular frequency
$\bar{\kappa}, \bar{\xi}$	A special value of the Lagrange multiplier $\kappa, \xi$
$\phi$	An objective function
$\phi_D$	Data fitting term of the objective function for FWI
$\phi_P$	Prior fitting term of the objective function for FWI
$\rho$	Density
$\rho_0$	Density of the background medium
$\sigma_{ij}$	Stress Tensor

$\theta_j$	Element of the incidence angle vector
$v$	Fractional fluctuation of the density
$\varsigma$	Constant term in quadratic equation
$\xi$	Fractional fluctuation of $\lambda$
$r, \psi, \text{ and } \zeta$	Spherical coordinates
$\langle \cdot, \cdot \rangle$	Inner product
$\dagger$	The Hermitian adjoint operator
$\Delta \mathbf{d}$	Residual or misfit between the observed and the estimated data
$\Delta \mathbf{s}$	Model update
$\Delta E$	Variation of the energy of a system
$\delta \lambda, \delta \mu$	Perturbation of the Lamé coefficients
$\delta \rho$	Perturbation of the density
$\delta_{\mu\nu}$	Kronecker delta
$\frac{\delta \bar{\rho}}{\rho_0}, \frac{\delta \bar{V}_P}{V_{P0}}, \frac{\delta \bar{V}_S}{V_{S0}}$	Fourier transforms of the parameter fractional fluctuations
$\ \cdot\ _2^2$	L2 norm
$\nabla$	Spacial derivative operator in matrix form
$\partial_i, \partial_j$	Spacial derivative operators in indicial notation
$\top$	Transpose operator

# Chapter 1

## Introduction

### 1.1 Background

The seismic method consists in generating artificial seismic waves, transmitting them from the surface to the interior of the Earth, and recording the information of the waves that return to the surface. Depending on the type of seismic survey carried out (onshore or offshore), these waves are produced by dynamite explosions, hydraulic vibrators or air guns. Once in the subsurface, the seismic waves interact with different geological layers and structures that cause their diffraction. Hence, part of the energy is reflected or returns to the surface, where the amplitude and arrival times of these diffracted waves are recorded in receiver stations and represented as seismic traces. This experiment is used in exploration seismology to deduce information beneath the surface of the Earth. Generally, the objective of its application is to encounter economic oil, gas, and mineral deposits and appropriate rocks for CO<sub>2</sub> injection and monitoring, as well as for microseismic, engineering, and archeological studies (Berkhout, 1986; Keys, 1986; Talwani and Kessinger, 2003).

Direct measurements of the rock properties of interest are often impossible to obtain with geophysical methods; thus, seismic inversion plays an important role to understand the spatial distribution of these physical quantities or model parameters in the subsurface. Inversion problems aim to estimate model parameters by trying to match the model response or “synthetic data” produced by a set of idealized parameters with the measurements of the wavefields that were recorded at the surface, namely “observed data”. To generate successful inversion results, it is essential to incor-

porate in the algorithms the physics of the phenomenon that produces the observed data (forward modeling of wave propagation). However, seismic data are insufficient and contain components, such as noise, that are not always possible to explain through the selected models of the physical process, meaning that the solution of the inversion is non-unique (Keys, 1986; Lines and Treitel, 1984). In this sense, often the inversion is performed with appropriate regularization terms to make it better posed (Virieux and Operto, 2009).

Additionally, several inversion techniques consider different parts of the recorded seismic data when estimating the desired parameters, but the incomplete use of the wavefield as well as the assumptions necessary to develop the expressions for the wave propagation model can produce limitations in the results. For instance, the traveltimes of the picked first arrivals are used in traditional traveltime tomography to estimate long-wavelength velocity structures (Bording et al., 1987; Bregman et al., 1989); AVO inversion takes advantage of the amplitudes of primary reflections and, in simultaneous cases, the amplitudes of secondary reflections to estimate parameter contrasts at different layer boundaries (Aki and Richards, 2002; Fatti et al., 1994; Shuey, 1985; Smith and Gidlow, 1987); and migration is based on transforming reflection events into a scaled version of the true geometry of the subsurface, focusing and positioning the events according to a velocity macromodel and allowing the proper location of events, the collapse of diffractors, reproduction of wavelet character and other effects (Sheriff, 2002; Virieux and Operto, 2009).

## **1.2 Full waveform inversion**

Full waveform inversion (FWI) is a powerful approach that considers all the recorded information (e.g., traveltime, phase, amplitude, time-frequency content), including signals that are typically neglected during standard processing workflows, such as low frequency effects, post-critical and multiple events, diving and converted waves, among others. This technique has been useful to generate intermediate-wavelength velocity models which are later used as starting models in PSDM applications and has also proved to produce high-resolution estimates of elastic and

anisotropic properties of the rocks (Fichtner et al., 2006; Fichtner, 2010; Operto et al., 2004, 2013; Pratt, 1990; Sirgue and Pratt, 2004; Virieux and Operto, 2009). Since the full information content is used, the approach is necessarily nonlinear, i.e., the data are nonlinear functions of the model parameters (Keys, 1986; Lines and Treitel, 1984; Tarantola, 1984; Virieux and Operto, 2009). Therefore, the use of direct expressions of least-squares solutions is not adequate to estimate the subsurface features, and local optimization methods that minimize an objective function involving the difference between the observed and synthetic data are classically preferred to iteratively produce the estimations.

Lailly (1983) and Tarantola (1984) were the first authors to address the FWI problem through local optimization using the generalized nonlinear least-squares approach, considering an acoustic approximation of the elastic wave equation, but also an elastic scenario in the time domain, respectively. In these publications, the solution is produced by methods related to the reverse-time migration of unstacked data, i.e., for each iteration of the algorithm and at each point in space, the gradient is obtained with the cross-correlation between the forward propagation of the actual sources and the backpropagation in time of the data residuals, generating the updates of the model parameters. For the first iteration, these model updates or perturbations are added to a provided initial model, and then the result is used as the starting model for the following iterations using a descent-based method (Mora, 1987; Pica et al., 1990; Tarantola, 1986). Moreover, different authors such as Shin (1988), Pratt (1990) and Song and Williamson (1995a,b) applied the FWI technique in the frequency domain, but Pratt (1990) expressed the forward and inverse problems in matrix form through the application of finite-differences to the wave equation in a crosshole tomographic experiment, being a reference for the computational adaptation of FWI algorithms in the frequency domain.

### **1.3 Challenges in the application of full waveform inversion**

Full waveform inversion is based on the formulation of a local least-squares optimization. This type of inversion is characterized by a series of challenges derived from the local optimization nature to solve the problem and the selected wave propagation model that describes the observed data, usually hindering the successful application of the method. Some of these issues are (1) the high computational cost as a result of the large number of variables to be estimated, (2) the lack of enough estimations to perform uncertainty analysis of the results, (3) the convergence towards local minimum points rather than the global minimum along with the phenomenon of cycle-skipping, (4) the slow convergence depending on the applied optimization algorithm, and (5) the coupled effects or crosstalk between parameters of different classes in multiparameter FWI.

After discretization of the seismic signals, the seismic wavefield as well as the model parameters correspond to high dimensional vectors in FWI problems. Moreover, the objective function is constrained by a wave equation per each seismic source located on the surface; thus, the performance of the forward modeling to simulate the data which would be produced by a particular set of parameters is a resource-intensive process. Hence, for iterative methods, extensive simulations of the data are required, making the FWI calculations computationally expensive for large 2D and 3D models. On the other hand, large-scale problems are most generally solved with local optimization algorithms, avoiding stochastic approaches such as Monte Carlo methods. The requirement of extensively solving the forward problem makes the process not feasible to perform in practice, implying much more computational cost and time. Since there is insufficient information, it is not possible to map out the likelihood of the desired physical properties, hampering the estimation of uncertainty (Mora, 1987) and the evaluation of confidence in the results with these approaches. However, research has been done in the matter of null-space shuttling, evaluating the uncertainty in specific model directions by determining how much a final estimated model could be altered without affecting the associated data residuals and with that understanding the ambiguity introduced by the non-uniqueness of the problem, i.e., by other models that also satisfy the objective function

(Deal and Nolet, 1996; Fichtner and Zunino, 2019; Liu and Peter, 2020). Moreover, the implementation of null-space shuttle approaches for targeted uncertainty analysis restricts the exploration of all the possible directions to only focus on a subset of model-space directions that is informative for the current problem by using designed hypotheses (Keating and Innanen, 2021), proving to be a beneficial method to distinguish estimated anomalies that contribute to the generation of the data from fictitious anomalies, such as crosstalk effects.

Typically, the inversion is linearized in local optimization methods by considering a quadratic objective function, even though no practical FWI problem behaves that way. Consequently, a parabola is fit in the initial point or starting model, its minimum is found, and this value is used as a new point to fix another parabola, repeating the process iteratively until converging to a value, hopefully close to the global minimum. However, convergence towards local minima is a massive challenge in FWI because the problem is nonlinear and the final estimates depend on how accurate the starting model is, the level of noise, the low frequencies available in the data, and the simplicity of the wave theory applied. Additionally, one of the most common types of local minimum is produced by the cycle-skipping artifacts, which occur when the error between the observed data and the data simulated using the starting model is larger than half the period (Beydoun and Taramola, 1988; Virieux and Operto, 2009). Nevertheless, both types of data could match within half a wave-cycle when low frequencies are considered. In this sense, Bunks et al. (1995) described the multigrid method, or multiscale approach, and applied it to a seismic inversion problem using the Marmousi model, demonstrating that the cycle-skipping artifacts can be mitigated and the performance of the iterative inversion can be improved by decomposing the problem by scales, increasing the frequencies as the iterations progress. Low frequencies recover long-wavelength components of the model and the number of local minima is reduced, being easier to find a solution that is in the neighborhood of the global minimum, no matter how close the starting model is. Thus, as the frequency values increase, the scale becomes shorter and the solution is refined.

On the other hand, large-scale inverse problems are conveniently solved with gradient-based methods, such as steepest descent, because these algorithms imply less computational cost, but

at the expense of suffering from slow convergence and in some cases failing to produce an acceptable solution, which is often attributed to the estimation of an accurate step length to descent. Conversely, the search direction of these methods, and thus convergence speed, is significantly improved by pre-conditioning the gradient with the inverse of the second-order information of the objective function, namely the Hessian operator, i.e., by minimizing the objective function with the Newton or Gauss-Newton algorithm (Pratt et al., 1998; Santosa and Symes, 1988; Virieux and Operto, 2009). Although the Hessian contains information that is essential for the proper reconstruction of the model, it is impractically expensive to explicitly calculate this matrix for large-scale problems. Therefore, the inverse of the Hessian operator is commonly approximated through quasi-Newton methods, such as the Limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) algorithm or the minimization is done by applying Hessian-free methods, such as the Truncated Newton algorithm.

Additionally, in multiparameter FWI, i.e., when not only the P-wave velocity is estimated but also other elastic parameters, attenuation, and/or anisotropy, new issues appear. For instance, the ill-posedness of the problem increases because the increment of parameters in the chosen wave equation comes with an increase of degrees of freedom, and there is also sensitivity or resolving differences between parameter classes during the inversion process, leading to inter-parameter coupled effects, also named trade-off or crosstalk (Operto et al., 2013; Virieux and Operto, 2009). Crosstalk occurs when different physical properties are confused in the inversion because a model parameter error produces data residuals that are later attributed to a different model parameter, yielding poorly accurate results and producing convergence slowness (Keating and Innanen, 2019). The mitigation of crosstalk in multiparameter FWI is essential to increase confidence in the estimation of the model parameters. The analysis of the radiation patterns that result when an incident wave interacts with a scatter point embedded in a homogenous medium are also named sensitivity analysis. These are typically performed to identify which patterns overlap for a range of angles, and thus for which of the involved physical properties the gradient update will be similar, giving insights about the leakage of information that would occur during the inversion (Keating and

Innanen, 2019; Métivier et al., 2015).

Tarantola (1986) indicated that a good set of parameters, for the execution of the FWI, would produce radiation patterns as different as possible, allowing to resolve each property easily and independently and with high resolution. Although for some acquisition geometries and parameterizations this could be achieved, for certain cases, some properties cannot be independently estimated. For instance, in seismic surface experiments, using sources dominated by P-waves, the parameterization formed by the density, P-wave impedance, and S-wave impedance is the best choice, but the impedances are better resolved than the density. Therefore, to mitigate the inter-parameter coupled effects, FWI workflows are designed to include parameterizations based on a minor correlation of their scattering patterns and, in some cases, considering which is the dominant model parameter of a particular set to perform a sequential inversion or hierarchical strategy (Kamei and Pratt, 2013; Sears et al., 2008; Tarantola, 1986).

In the literature, we can find additional methods that have been applied in FWI to reduce inter-parameter coupled effects considering different procedures. For instance, the subspace optimization method is based on adjusting the update direction, in the optimization process, according to the contribution of each parameter class (Baumstein, 2014; Kennett et al., 1988). Additionally, the mode decomposition method is based on pre-conditioning the gradients by decoupling them or isolating their P and S wave components (Wang and Cheng, 2017; Xu et al., 2017). Moreover, Pan et al. (2018) developed an inversion strategy using isotropic elastic parameterizations by considering approximate contamination kernels (products of Hessian off-diagonal blocks with vectors of model perturbations), which quantify the parameter trade-off characteristics among different model parameters in a local or wide fashion within the volume.

In general, the selection of an adequate parameterization and the minimization of the objective function with an optimization approach that includes the Hessian matrix or a good approximation of it allows to mitigate crosstalk effects (Hu and Innanen, 2019; Innanen, 2014; Keating and Innanen, 2018; Pratt et al., 1998; Wang et al., 2016; Xing and Zhu, 2020; Yang et al., 2016). The inverse of the Hessian operator acts as a pre-conditioner of the gradient, enhancing the model updates,

since it not only compensates for the geometrical spreading effects and band-limitations through their main-diagonal blocks, but it also helps to quantify and correct crosstalk effects through their off-diagonal blocks (Operto et al., 2013; Pratt et al., 1998). Hence, trade-off suppression could also be achieved by performing an appropriate manipulation of the Hessian within the optimization scheme, since no parameter correlation would exist if the terms in the off-diagonal blocks were set to zero (Métivier et al., 2015; Operto et al., 2013).

Innanen (2020a,b,c,e) explained that (1) re-parameterizing seismic inversion problems is equivalent to performing coordinate transforms between two systems, and that we can use transformation matrices that are constructed to satisfy constraints imposed by the Hessian of the new model space, such as changing the shape of the iso-surfaces of the objective function in that system; (2) when the iso-surfaces of the objective function are spherically symmetric, i.e., when the Hessian is the Kronecker delta or an identity matrix, the Gauss-Newton directions are parallel to those of steepest descent, producing favorable convergence properties in problems solved with the latter optimization strategy, which is convenient to avoid the computational cost of the Hessian; and (3) re-parameterizing to a model space characterized by a Hessian with an identity matrix structure promises the minimization of crosstalk between parameters of different classes, since the values in the off-diagonal blocks are zero.

On the other hand, additional research has been done where pre-conditioning ideas were applied in different contexts but maintaining similitudes to the coordinate transform theory between model spaces and the use of appropriate Hessians to obtain satisfactory solutions. For instance, in a variety of cases, non-stationary Point Spread Functions (PSF) were used for migration purposes related to the correction of amplitude effects; these local point spread functions were explained by Schuster and Hu (2000) while developing the 2D and 3D migration Green's function formulas in the far-field approximation. Moreover, Guitton (2004) used a flat-layer model with a complex velocity field and the Marmousi model to estimate migrated images by approximating the inverse of the Hessian or the inverse of the PSF with a bank of non-stationary matching filters between two migrated images. Finally, Valenciano (2008) used the least-squares wave-equation inversion to

reduce illumination problems under salt bodies while considering least-squares Hessian matrices defined from the modeling/migration operators and that the information about the PSF is provided by the off-diagonal elements of the Hessian operator, measuring the quality of the imaging system, aiming to reverse the spreading of the energy with the inversion and recovering the amplitudes.

## 1.4 Thesis objectives and organization

The main objective of this thesis was to reduce the crosstalk effects that would be produced in multiparameter full waveform inversion in the  $\rho$ ,  $V_P$ , and  $V_S$  model space while considering a seismic surface experiment. The strategies to minimize these inter-parameter coupled effects were developed by applying the decorrelation ideas explained by Operto et al. (2013), Métivier et al. (2015), and Innanen (2020a,b,c,e). In this sense, the target was to transform the original model space ( $\rho$ ,  $V_P$ , and  $V_S$ ) to a root model space for FWI ( $\rho$ ,  $c_{11}$ , and  $c_{44}$ ), called  $s$  in this thesis, and then to a new or intermediate model space, called  $r$ , where the parameters of different classes or nature would be already decorrelated, i.e., where the Hessian operator would be an identity matrix or a multiple of it. This new model space would be convenient to perform an FWI where convergence could be reached towards a more accurate minimum point with minimization of coupled effects and (1) since mapping between model spaces is equivalent to a coordinate transform between two systems and (2) the value of the objective function is a scalar that keeps its magnitude in both model spaces, finding the best solution in the  $r$  model space is equivalent to find it in the original model space. To achieve this goal, appropriate transformation rules between the Hessian of both systems were necessary to obtain transformation matrices to later apply transformation rules or linear relationships to map between model spaces. However, some of the challenges in the application of these ideas lie in the type of re-parameterization done in the model grid as well as in the type of Hessian matrix used, since working with the Full Hessian would imply a significant computational cost. Within this framework, the thesis is organized as follows:

**Chapter 2** contains the study of the analytic and empirical radiation patterns produced by

scatter points or model perturbations in the  $\rho$ ,  $V_P$ , and  $V_S$  model space. The goal was to understand the type of crosstalk that would be produced in an FWI process while working with this re-parameterization of the elastic wave equation, and how these coupled effects can change if more realistic scenarios are considered, such as heterogeneous reference media or backgrounds, which are typically not contemplated in the development of the analytic expressions of the scattering patterns. In this context, the development and assumptions of the analytic expressions of the radiation patterns, in this model space, were explained, but also a workflow was introduced to extract the displacement components of the P-P and P-S wavefields generated under homogeneous and heterogeneous reference media by using radius-dependent masks, and angle sweeps were executed over them to compute the empirical radiation patterns.

**Chapter 3** entails the theory and optimization methods behind the application of full waveform inversion, which are required to understand the workflows proposed in this thesis and employed in the different numerical experiments. Additionally, this chapter includes the first numerical experiments done to prove the decorrelation ideas between parameters of different classes, by applying local-wise re-parameterizations of the model grid and using local  $3 \times 3$  Hessians or point-wise Hessians to obtain transformation matrices of the same size that allow the mapping between the original model space  $s$  and the new or intermediate model space  $r$ . To obtain the transformation matrices, the numerical procedure reported by Innanen (2020d) was applied. The model parameters estimated with this strategy were compared with the results obtained with a baseline or reference FWI, i.e., without applying any transformation rule, and the aspect of the resulting Hessians from both scenarios was evaluated to understand the outcomes and the modifications that could be done to the workflow.

**Chapter 4** includes the Simulated Annealing theory, as well as the development of the forward problem and inverse problem operators necessary to solve a one-dimensional AVO nonlinear problem based on the re-parameterization of the Zoeppritz equations and using three different optimization methods, i.e., Gauss-Newton, steepest descent, and Simulated Annealing, while inverting broadband and band-limited synthetic data. The objective of performing these experiments was

to demonstrate and understand the general behavior in addition to the advantages and disadvantages of the Simulated Annealing algorithm when estimating physical properties from a simpler and smaller geophysical seismic problem than the FWI problem treated in this thesis, and by doing so, acquiring valuable insights to extend its application to other geophysical seismic problems.

**Chapter 5** comprises an alternative to the type of Hessian matrix used in Chapter 3, to consider crosstalk information from multiple locations of the model grid instead of only one. The strategy proposed in this chapter includes numerical experiments that required the transformation of a point-probes Hessian into an almost block-wise diagonal matrix to find a transformation matrix of size  $3 \times 3$  that still allows the local-wise mapping between the model spaces. To execute this idea, a different numerical procedure, from the one used in Chapter 3, was necessary to compute the transformation matrix, choosing the Simulated Annealing technique due to its advantages over other methods. Therefore, in this chapter, it is explained the advantages and disadvantages of the Simulated Annealing for the current problem, and the objective function employed for the new purposes. The model parameters estimated with this new strategy were compared with the results obtained in Chapter 3. The section finalizes with the analysis of the successes and the possible factors that could have affected the behavior of the Simulated Annealing technique.

Finally, **Chapter 6** consists of the conclusions of the thesis, summarizing the purposes of each chapter and encompassing the main breakthroughs obtained with each of the numerical experiments, as well as a discussion of successful aspects and areas for enhancement in future investigations related to this topic.

# Chapter 2

## Assessment of crosstalk under scenarios of homogeneous and heterogeneous backgrounds

### Summary

Analyzing radiation patterns is the method most commonly used to evaluate crosstalk between parameters of different classes in multiparameter full waveform inversion problems. Typically, these patterns are constructed from analytic expressions subject to not-so-realistic assumptions, such as considering homogeneous reference media. This chapter focuses on introducing a workflow to extract the empirical radiation patterns from simulated scattered wavefields to understand the crosstalk effects that exist in the  $\rho$ ,  $V_p$ , and  $V_S$  model space, and on making use of the technique to assess the scattering patterns generated from wavefields produced under different heterogeneous backgrounds. To achieve this, radius-dependent masks were used to isolate the horizontal and vertical components of the wavefields of interest as well as angle sweeps. The proposed workflow is beneficial since it allows to extract accurate empirical patterns and to demonstrate that, under heterogeneous scenarios, the shape of the radiation patterns slightly changes from what is theoretically expected, and thus different crosstalk regions might occur from those indicated by the analytic expressions, as well as sensitivity variations.

## 2.1 Introduction

Multiparameter full waveform inversion (FWI) is a challenging problem because inter-parameter coupled effects can be introduced to the seismic response in a propagation regime, depending on the selected parameterization of the physical properties of the rocks. These effects generate poor estimations of the model parameters and produce slow convergence of the optimization algorithms chosen to solve the inversion problem.

These coupled or trade-off effects, namely crosstalk, are a phenomenon that occurs when the error associated with a particular model parameter generates data residuals that are attributed to another model parameter (Keating, 2020), producing leakage of information, i.e., parameters of different nature are confused during the inversion process. The most common strategy to mitigate these effects in an FWI workflow is to analyze the radiation patterns of scattered wavefields associated with a diffractor point or model perturbation positioned in the model grid and later choose a suitable parameterization to describe the subsurface that guarantees their minimal overlap over most of the scattering angles that are in the scope of the acquisition geometry of the seismic experiment (Keating and Innanen, 2019; Operto et al., 2013).

Typically, before solving the FWI problem, analytic expressions of radiation patterns are calculated to perform crosstalk analysis for a particular parameterization and acquisition geometry. Generally, these expressions have been derived and published for different sets of known parameters and wave equations (e.g., scalar acoustic, elastic, viscoelastic, among other wave equations) using a point scatterer model of a localized heterogeneity which is embedded in a homogeneous medium, assuming plane waves, and working on the basis of the Born approximation to find the solution of the scattered wavefields (Kamath and Tsvankin, 2016; Moradi and Innanen, 2019; Sato, 1984; Sato et al., 2012; Wu and Aki, 1985). In this sense, the parameter perturbations that produce radiation patterns with a variation of amplitude that is proportional over a range of scattering angles will be easily confused between each other in the inversion process, as these have similar effects on the recorded wavefields and it will not be straightforward to distinguish them (Keating,

2020).

Therefore, the objective of this chapter was to introduce a workflow to extract the empirical radiation patterns from simulated scattered P-P and P-S wavefields generated from a virtual source or scatter point of individual model perturbations, regardless of the parameterization selected and if its analytic expressions are known or not. The workflow was tested by comparing the empirical and analytic scattering patterns produced by perturbations of  $\rho$ ,  $V_p$ , and  $V_s$  embedded in homogeneous media, and the independence from analytic expressions was exploited by studying the radiation patterns produced under scenarios of heterogeneous reference media, to understand how these patterns change from the theoretical expectations and how crosstalk might vary in this model space when considering more realistic configurations, which are not addressed in the analytic equations.

## 2.2 Scattering theory in the $\rho$ , $V_p$ , and $V_s$ model space

The scattering theory explains that the actual medium that represents the subsurface and where a wave propagates results from the sum of an homogeneous background (reference medium) and small perturbations in the model properties (Moradi and Innanen, 2015). For an isotropic elastic scenario, characterized by the density and the Lamé coefficients ( $\lambda$  and  $\mu$ ), the following relationships hold:

$$\lambda(\mathbf{x}) = \lambda_0 + \delta\lambda(\mathbf{x}) = \lambda_0 [1 + \xi(\mathbf{x})] \quad (2.1)$$

$$\mu(\mathbf{x}) = \mu_0 + \delta\mu(\mathbf{x}) = \mu_0 [1 + \chi(\mathbf{x})] \quad (2.2)$$

$$\rho(\mathbf{x}) = \rho_0 + \delta\rho(\mathbf{x}) = \rho_0 [1 + \nu(\mathbf{x})] \quad (2.3)$$

where  $\lambda_0$ ,  $\mu_0$ ,  $\rho_0$  are the properties of the background and  $\lambda$ ,  $\mu$ ,  $\rho$  are the properties of the actual medium. Additionally, the fractional fluctuations correspond to a localized heterogeneity (Figure 2.1) and these must meet  $|\xi|, |\chi|, |\nu| \ll 1$ .

To study a scattered wavefield and extract its analytic radiation patterns, we can analyze a

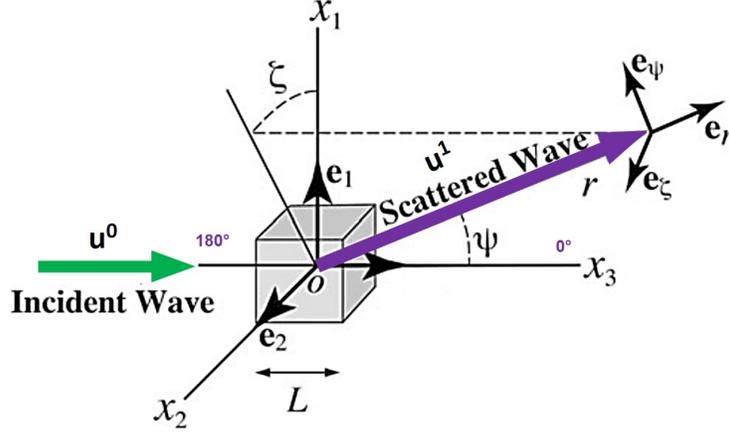


Figure 2.1: Representation of a localized inhomogeneity of size  $L$  from which the scattered wavefield is generated.  $r$ ,  $\psi$ , and  $\zeta$  are spherical coordinates and  $\mathbf{e}_r$ ,  $\mathbf{e}_\psi$  and  $\mathbf{e}_\zeta$  are the unit base vectors in that coordinate system (Modified from Sato et al. (2012)).

plane wave that interacts with a localized inhomogeneity or scatter point, using the first-order perturbation method or Born approximation in the stationary state (Sato et al., 2012). To obtain the P-P and P-S analytic radiation patterns we must consider that:

$$\mathbf{u} = \mathbf{u}^0 + \mathbf{u}^1 \quad (2.4)$$

where  $\mathbf{u}$  is the total wavefield,  $\mathbf{u}^0$  is the incident wavefield and  $\mathbf{u}^1$  is the scattered wavefield, and  $|\mathbf{u}^1| \ll |\mathbf{u}^0|$ . Overall, the scattered wavefield results after summing the response from all single scatter points. However, radiation patterns are not generally studied for all the locations of the model grid, but a representative pattern per perturbed parameter class is investigated. This means that the radiation pattern analysis does not provide information about crosstalk at different spatial locations, and thus this procedure does not completely characterize the crosstalk effects produced in the inversion (Keating and Innanen, 2019).

On the other hand, the isotropic elastic wave equation in the time domain for the full displacement wavefield  $\mathbf{u}(\mathbf{x}, t)$  is:

$$\rho(\mathbf{x})\ddot{u}_i(\mathbf{x}, t) - \partial_j \sigma_{ij}(\lambda, \mu; u_k) = 0 \quad (2.5)$$

with:

$$\sigma_{ij}(\lambda, \mu; u_k) = \lambda(\mathbf{x})\delta_{ij}\partial_l u_l(\mathbf{x}, t) + \mu(x)[\partial_i u_j(\mathbf{x}, t) + \partial_j u_i(\mathbf{x}, t)] \quad (2.6)$$

where the two overdots in  $u$  mean second time derivative and  $i = j = k = l = 1, 2, 3$ . Equation 2.5 is written for the incident wave as:

$$\rho_0 \ddot{u}_i^0 - \partial_j \sigma_{ij}(\lambda_0, \mu_0; u_k^0) = 0 \quad (2.7)$$

We can find the scattered wave equation by substituting Equation 2.4 in Equation 2.5 and using Equation 2.7. If we focus only on first order perturbations, the following equation is obtained:

$$\rho_0 \ddot{u}_i^1 - \partial_j \sigma_{ij}(\lambda_0, \mu_0; u_k^1) = \delta f_i(\mathbf{x}, t) \quad (2.8)$$

The term  $\delta f_i(\mathbf{x}, t)$  in Equation 2.8 is a mathematical expression of the interaction between the incident wave and the scatter point (equivalent body force):

$$\delta f_i(\mathbf{x}, t) = -\delta \rho \ddot{u}_i^0 + \partial_i \delta \lambda \partial_j u_j^0 + \partial_j \delta \mu (\partial_i u_j^0 + \partial_j u_i^0) + \delta \lambda \partial_i \partial_j u_j^0 + \delta \mu \partial_j (\partial_i u_j^0 + \partial_j u_i^0) \quad (2.9)$$

To solve Equation 2.8, the equivalent body force needs to be defined according to the type of incident wave that we desire to simulate. For a P-wave:

$$\mathbf{u}^{0P} = \mathbf{e}_3 e^{i(k_0 \mathbf{e}_3 \mathbf{x} - \omega t)} \quad \text{where} \quad k_0 = \frac{\omega}{V_{P0}} \quad (2.10)$$

Equation 2.10 must be used for every  $u^0$  or  $\ddot{u}^0$  in Equation 2.9. Later, the Green's function is applied to solve for the desired type of scattered waves, in the case of this study, the P-P and P-S wavefields. The derivation of  $u_i^{1PP}$  and  $u_i^{1PS}$  was developed in detail in Sato et al. (2012), producing

the following expressions:

$$u_i^{1PP}(\mathbf{x}, t) = \frac{e^{i(k_0 r - \omega t)}}{r} F_i^{PP} \quad (2.11)$$

$$u_i^{1PS}(\mathbf{x}, t) = \frac{e^{i(l_0 r - \omega t)}}{r} F_i^{PS} \quad (2.12)$$

which result after convolving the far-field component of the Green's function with the equivalent body force. In this equation,  $r$  corresponds to the observation distance for the Green's function,  $l_0 = \omega/V_{S0}$ , and  $F_i^{PP}$  and  $F_i^{PS}$  are the scattering amplitudes, which can be written in spherical coordinates as:

$$\mathbf{F}^{PP} = \sum_{i=1}^3 F_i^{PP} \mathbf{e}_i = F_r^{PP} \mathbf{e}_r + F_\psi^{PP} \mathbf{e}_\psi + F_\zeta^{PP} \mathbf{e}_\zeta \quad (2.13)$$

$$\mathbf{F}^{PS} = \sum_{i=1}^3 F_i^{PS} \mathbf{e}_i = F_r^{PS} \mathbf{e}_r + F_\psi^{PS} \mathbf{e}_\psi + F_\zeta^{PS} \mathbf{e}_\zeta \quad (2.14)$$

$$\begin{aligned} F_r^{PP} = \frac{l_0^2}{4\pi} & \left[ \left( -\frac{1}{\gamma_0^2} + \frac{\cos \psi}{\gamma_0^2} + \frac{2}{\gamma_0^4} \sin^2 \psi \right) \frac{\delta \tilde{\rho}(k_0 \mathbf{e}_r - k_0 \mathbf{e}_3)}{\rho_0} \right. \\ & - \left( \frac{2}{\gamma_0^2} \right) \frac{\delta \tilde{V}_P(k_0 \mathbf{e}_r - k_0 \mathbf{e}_3)}{V_{P0}} \\ & \left. + \left( \frac{4}{\gamma_0^4} \sin^2 \psi \right) \frac{\delta \tilde{V}_S(k_0 \mathbf{e}_r - k_0 \mathbf{e}_3)}{V_{S0}} \right] \end{aligned} \quad (2.15)$$

with  $F_\psi^{PP} = F_\zeta^{PP} = 0$

$$\begin{aligned} F_\psi^{PS} = \frac{l_0^2}{4\pi} & \left[ \left( -\sin \psi + \frac{2}{\gamma_0} \cos \psi \sin \psi \right) \frac{\delta \tilde{\rho}(l_0 \mathbf{e}_r - k_0 \mathbf{e}_3)}{\rho_0} \right. \\ & \left. + \left( \frac{4}{\gamma_0} \cos \psi \sin \psi \right) \frac{\delta \tilde{V}_S(l_0 \mathbf{e}_r - k_0 \mathbf{e}_3)}{V_{S0}} \right] \end{aligned} \quad (2.16)$$

with  $F_r^{PS} = F_\zeta^{PS} = 0$

In Equation 2.15 and 2.16,  $\gamma_0 = V_{P0}/V_{S0}$ . Additionally, each scattering amplitude contains the exchanged wavenumber vector, which is the argument in the Fourier transforms of the fractional

fluctuations  $\frac{\delta\bar{\rho}}{\rho_0}$ ,  $\frac{\delta\bar{V}_P}{V_{P0}}$ , and  $\frac{\delta\bar{V}_S}{V_{S0}}$ , i.e., the difference between the scattered wavenumber vector and the incident wavenumber vector for an angular frequency. In both equations, the terms grouped with parenthesis are the mathematical expressions of the analytic radiation patterns produced by its associated parameter perturbations.

## 2.3 Numerical experiments

### 2.3.1 Extraction of empirical radiation patterns

To examine the empirical radiation patterns produced by perturbations in  $\rho$ ,  $V_P$ , and  $V_S$ , a homogeneous reference medium with 200 samples in both the  $x$  and  $z$  directions was considered for each parameter and with values of  $V_P=3000$  m/s,  $V_S=1800$  m/s, and  $\rho=1400$  kg/m<sup>3</sup>. Additionally, the perturbed media were constructed by positioning a localized perturbation (scatter point) in the middle of the model grid for one of the three parameters. This perturbation represented an increment of 10% of the respective background value.

The isotropic elastic wave equation re-parameterized in the  $\rho$ ,  $V_P$ ,  $V_S$  model space and in the frequency domain was selected as the wave propagation model. This wave equation is explained in Section 3.2.2 of Chapter 3 with more detail. Moreover, an explosive source, represented by a Ricker wavelet with 25Hz of dominant frequency, was located in the middle of the surface of the grid and 1C vertically and horizontally-oriented geophones were placed at each grid point, responding at discrete frequencies and recording the simulated vertical ( $u_z$ ) and horizontal ( $u_x$ ) displacements of the wavefields. Each displacement component was computed twice: (1) simulating that the incident wave traveled only in the reference media (with no scatter points) and (2) simulating that it traveled through the medium perturbed in only one parameter class, but with no perturbations of the other parameter classes. The inverse Fourier Transform was applied to each displacement component and, after transforming them to the time domain, the subtraction of the former to the latter displacement was done, corresponding to the scattered wavefield generated by a change in a particular model property (Figure 2.2), as indicated by Equation 2.4.

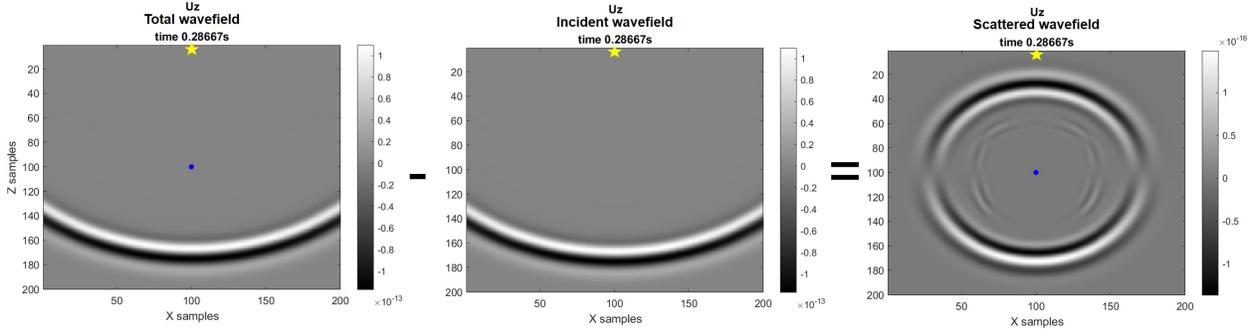


Figure 2.2: Mathematical procedure to compute the scattered wavefield caused by a perturbation of one parameter class. The blue dot corresponds to the scatter point, while the yellow star represents the explosive source.

With all the information in the time domain, the scattered wavefield was plotted at an appropriate time that allowed the visual differentiation of the P-P and P-S wavefields. For this fixed time, both types of wavefields were separated using a radius-dependent mask of the size of the model grid. Favorable radii were input to isolate both wavefields, considering the values out of the radius when working with the P-P energy and within the radius when working with the P-S energy (Figure 2.3).

Within the isolated regions of the grid, a sweep of angles was performed by fixing an angle and selecting a range of 10 discrete upper and lower degrees. The total amplitude per fixed angle was computed with the square root of the sum of the energies associated with  $u_x$  and  $u_z$  within the current group of angles, i.e.,  $|\mathbf{u}^1| = \sqrt{u_x^2 + u_z^2}$ , and the process was repeated for 360 degrees. Finally, polar plots were constructed to illustrate the empirical radiation patterns and these were compared with those obtained from the corresponding analytic expressions of Equations 2.15 and 2.16. It is observed in Figure 2.4 that the shape and values of the analytical and empirical radiation patterns were almost identical for every perturbed parameter class and for each type of energy conversion mode, which means that effectively this workflow recovered accurate patterns and can be used in situations where the analytic expressions are unknown.

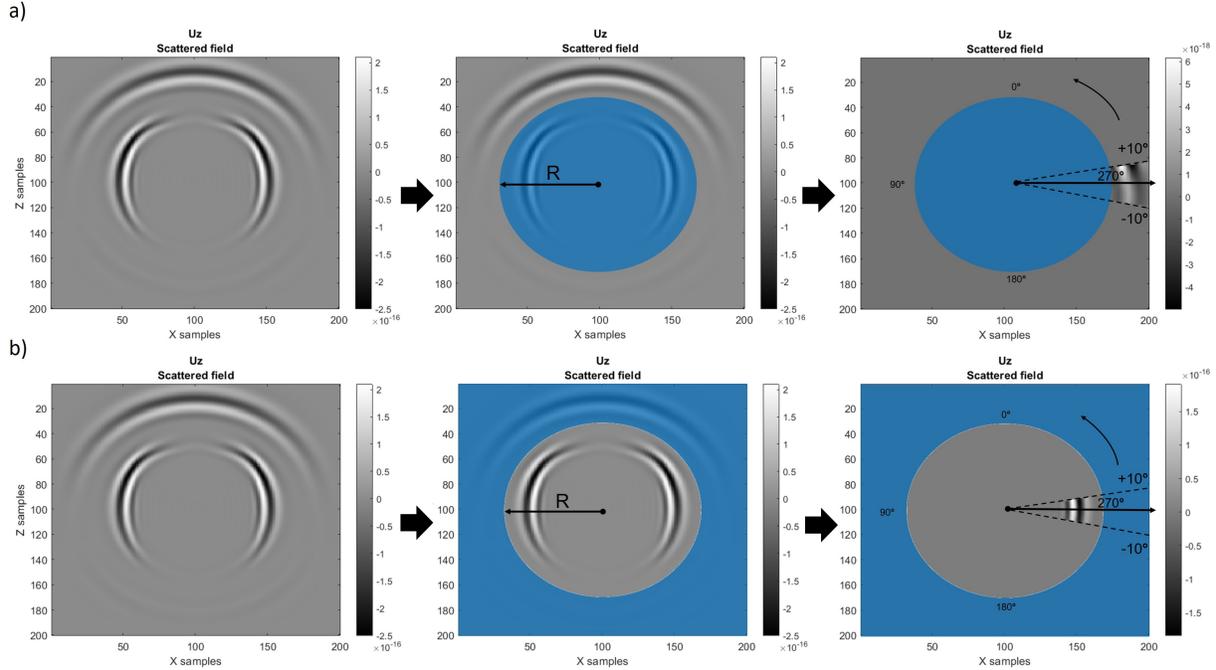


Figure 2.3: Proposed workflow to extract the empirical radiation patterns from simulated wavefields.(a) Extraction of patterns from the P-P wavefield. (b) Extraction of patterns from the P-S wavefield. R corresponds to the radius indicated by the user.

### 2.3.2 Radiation patterns under heterogeneous reference media

Three heterogeneous background media with varying complexities were studied using the procedure described in the previous section. The first case consisted in reference media for  $\rho$ ,  $V_P$ , and  $V_S$  with a linear increment of the values with depth; the second case was similar to the first one but with an increment of 25% of the slopes of the linear increment; and the third case was a modified smoothed Marmousi model (the original  $V_P$  Marmousi model was divided by 2 to obtain  $V_S$  and by 1.5 to obtain  $\rho$ ). Figure 2.5 allows to illustrate the perturbed  $V_P$  medium in each case. For each of these scenarios, scatter points were placed in the middle of the grid with values of 10% increase with respect to the reference media.

For each case, the model size was modified to avoid grid dispersion effects. In Figures 2.6 and 2.7 a comparison is shown between the analytic radiation patterns, the patterns obtained from a homogeneous background, and those obtained after considering the heterogeneous reference

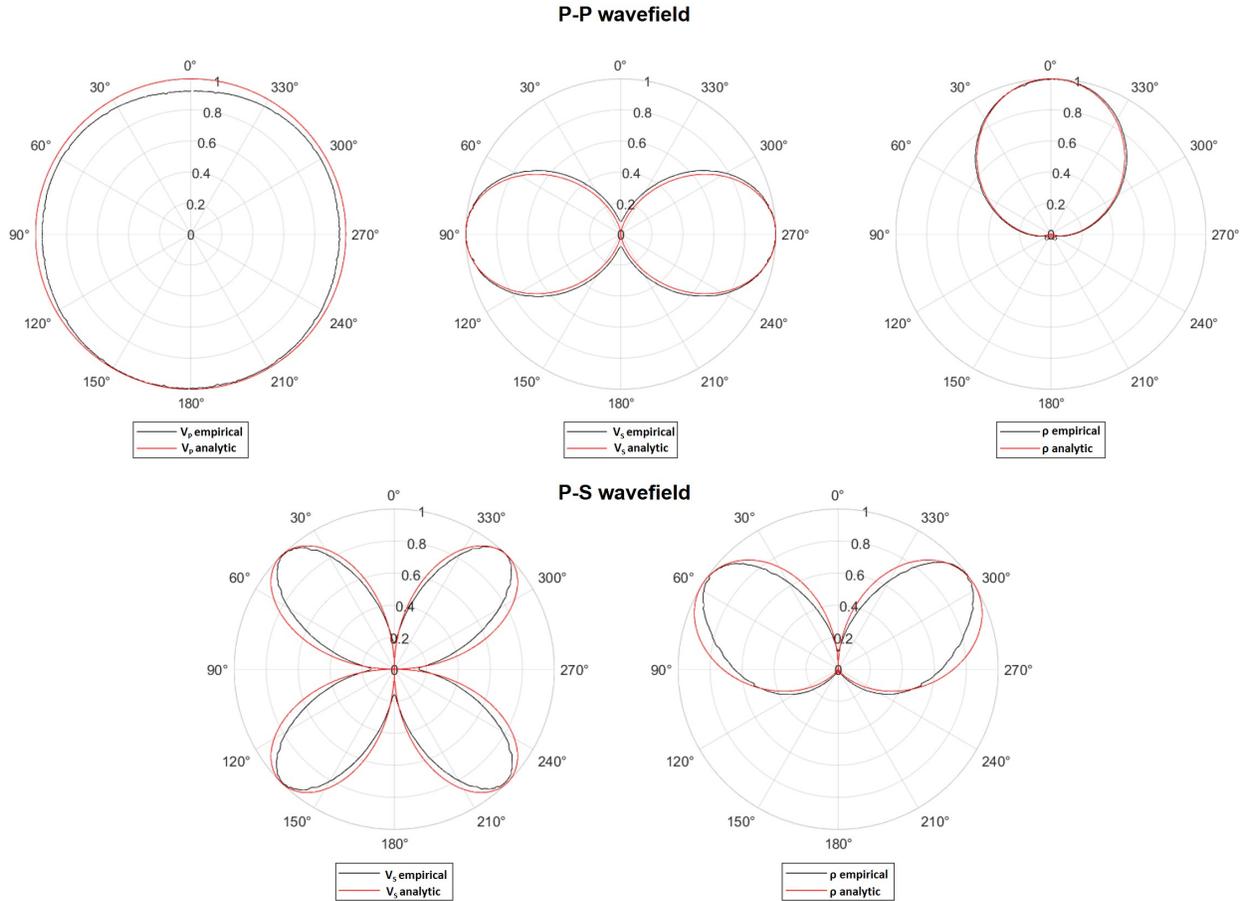


Figure 2.4: Comparison between the normalized analytic and empirical radiation patterns of P-P and P-S wavefields. Values were normalized.

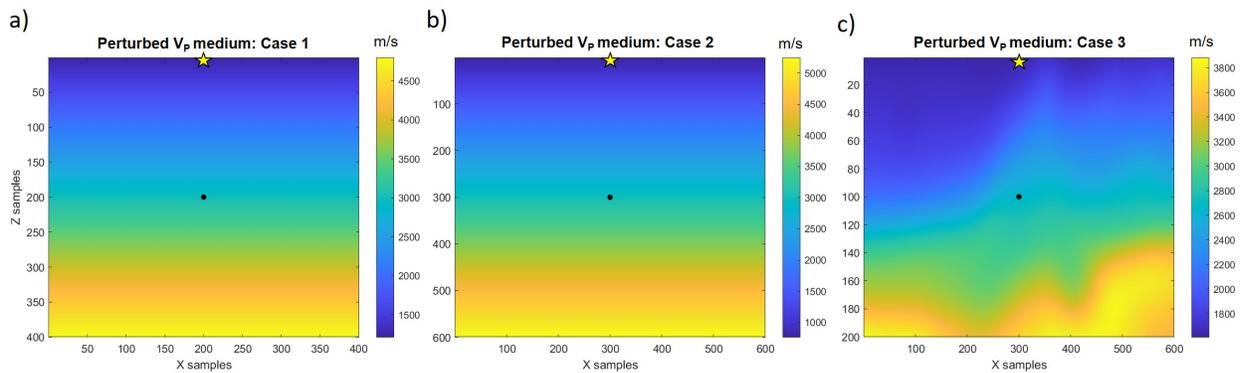


Figure 2.5: Different heterogeneous reference media used in this study. (a) Case 1, (b) Case 2, and (c) Case 3.

media proposed in this section. For the cases of heterogeneous background, the general shape of the scattering patterns was kept, but some changes appeared in terms of rotation of the lobes and the amplitude ratio.

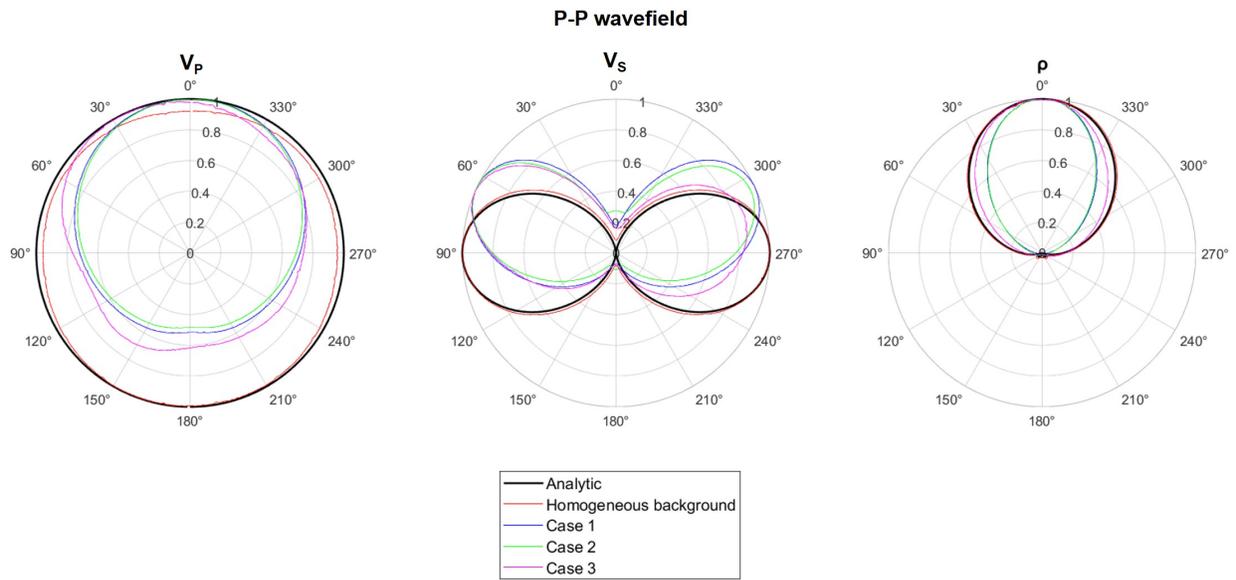


Figure 2.6: Overlapped radiation patterns of analytic, homogeneous, and heterogeneous cases extracted from the P-P wavefield. Values were normalized.

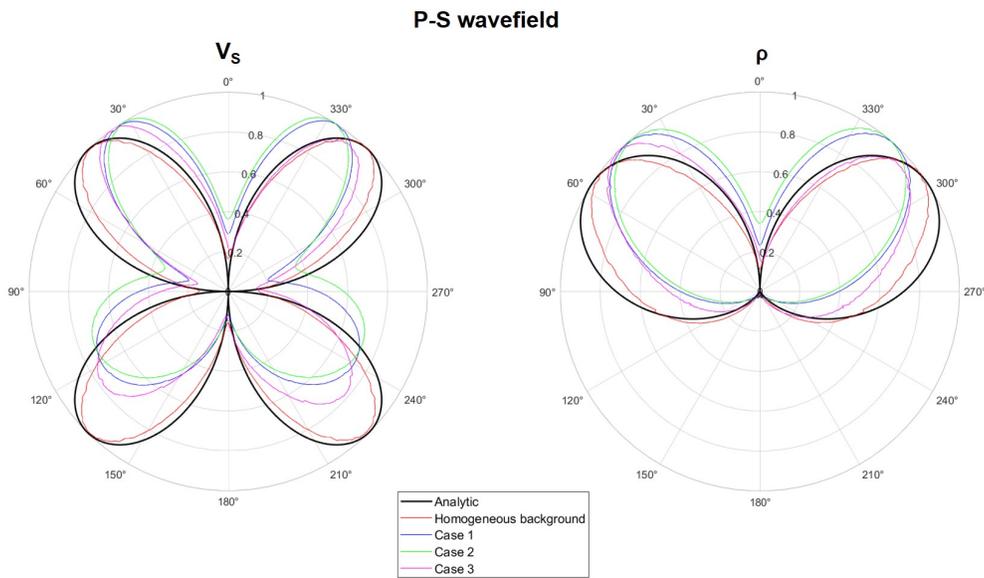


Figure 2.7: Overlapped radiation patterns of analytic, homogeneous, and heterogeneous cases extracted from the P-S wavefield. Values were normalized.

When the values of the reference media increased linearly with depth (Cases 1 and 2), the amplitudes were more focused around smaller angles than those described by the analytic responses. For instance, there was a slight upward rotation of the radiation lobes associated with changes in  $\rho$  for the P-S wavefield and in  $V_S$  for the P-P and P-S wavefields. Additionally, even though the  $V_P$  scattering pattern (P-P wavefield) was close to circular, it did not have the analytic isotropic behavior anymore, but higher amplitudes were scattered towards small and intermediate angles. Moreover, the radiation pattern caused by  $\rho$  perturbations (P-P wavefield) shrank towards smaller angles.

On the other hand, the modified Marmousi model (Case 3) exhibited radiation patterns with behaviors in between the homogeneous case and Case 1-Case 2, and the shape of the patterns differed from the analytic shapes for some angles. In this case, rotation of lobes occurred similarly to the previous heterogeneous cases, but especially for the P-P wavefield and for the  $V_P$  and  $\rho$  patterns, the amplitude values were closer to those described by the analytic expressions. Therefore, the shrinkage of the  $\rho$  radiation pattern was more subtle and the  $V_P$  radiation pattern had higher amplitudes scattered from small to large angles, but losing its circular isotropic shape.

Figure 2.8 and 2.9 contain the overlapping of the scattering patterns of each model perturbation, helping with the identification of the extent of crosstalk that would occur by performing an FWI in this model space. When considering a seismic surface experiment and the P-P wavefield, the analytic patterns of  $V_P$  and  $\rho$  indicated crosstalk in small angles ( $0-6^\circ$ ), and for the heterogeneous cases, the crosstalk occurred in slightly smaller angles, with a lack of symmetry for the Marmousi reference media; moreover, according to the analytic expressions, no crosstalk should appear between the parameters  $V_S$  and  $\rho$ , but in Case 3 a small overlap existed between  $80$  and  $90^\circ$ ; regarding  $V_P$  and  $V_S$ , crosstalk in Case 2 was the least similar to what was expected from the analytic patterns, exhibiting crosstalk in small angles. On the other hand, when considering the P-S wavefield, the analytic expressions of  $V_S$  and  $\rho$  suggested crosstalk from  $0$  to almost  $30^\circ$  and its counterpart; however, Cases 1 and 3 showed overlap of the radiation patterns from  $0$  to  $20^\circ$  and for Case 2, from  $0$  to  $15^\circ$ .

Therefore, subtle differences existed between the crosstalk effects indicated by the analytic and the empirical scattering patterns from heterogeneous media, since depending on the scenario, new small crosstalk regions appeared or the scope of the coupled effects was reduced, i.e., it occurred for smaller range of angles. Additionally, the overlapping of the radiation patterns, either in the analytic or empirical cases, gave insights into the strong variations of amplitude values that would be present in a case by case basis, which is related to the nature or class of the model parameter selected to be the perturbation or scatter point. This suggests that these sensitivity effects might be extended to the inversion process.

## 2.4 Conclusions

The use of simulated scattered wavefields allowed to execute a workflow to isolate the P-P and P-S wavefields and extract the empirical radiation patterns of the  $\rho$ ,  $V_P$ , and  $V_S$  model space with high accuracy regarding the patterns that result from the corresponding analytic expressions encountered in the literature. This workflow was beneficial to perform crosstalk analysis not only when the considered reference media was homogeneous, but also in situations where the analytic expressions are not as explicit, such as the cases of heterogeneous reference media studied in this chapter. However, the workflow could also be used to study the radiation patterns produced by model perturbations in a re-parameterized model space which require analytic expressions that are somehow unknown. This method helped to demonstrate that the shape of the radiation patterns generated from heterogeneous reference media was close to the indicated by the analytic expressions, which depend on homogeneous reference media, but in some cases, rotation of lobes, decrease of amplitude values, loss of symmetry and irregularities occurred. Moreover, it allowed to understand the extent of crosstalk that could be produced between parameters of different classes when performing an FWI in the  $\rho$ ,  $V_P$ , and  $V_S$  model space and in particular for a surface seismic experiment. Overall, for seismic surface experiments, there were crosstalk effects mostly between  $V_P$  and  $\rho$  as well as between  $V_S$  and  $\rho$ , and the overlap of these heterogeneous scattering patterns

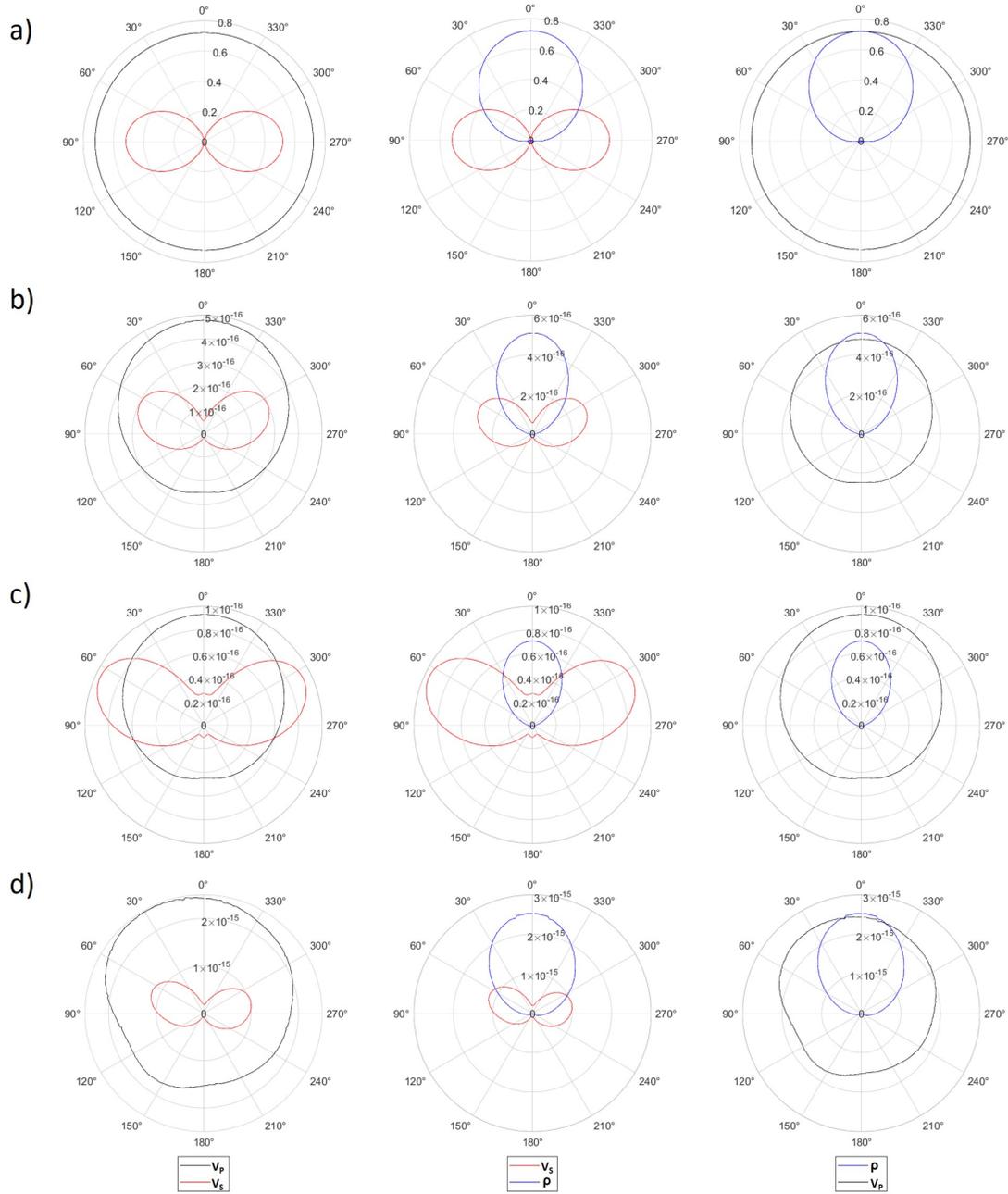


Figure 2.8: Crosstalk assessment with radiation patterns extracted from P-P wavefield. Values were not normalized to study the amplitude variations.(a) Analytic expressions, (b) Case 1, (c) Case 2, and (d) Case 3.

enabled to perform a more accurate crosstalk analysis under more realistic scenarios, since the coupled effects between parameters of different classes slightly changed with respect to those indicated by the analytic expressions and different sensitivities between radiation patterns were also

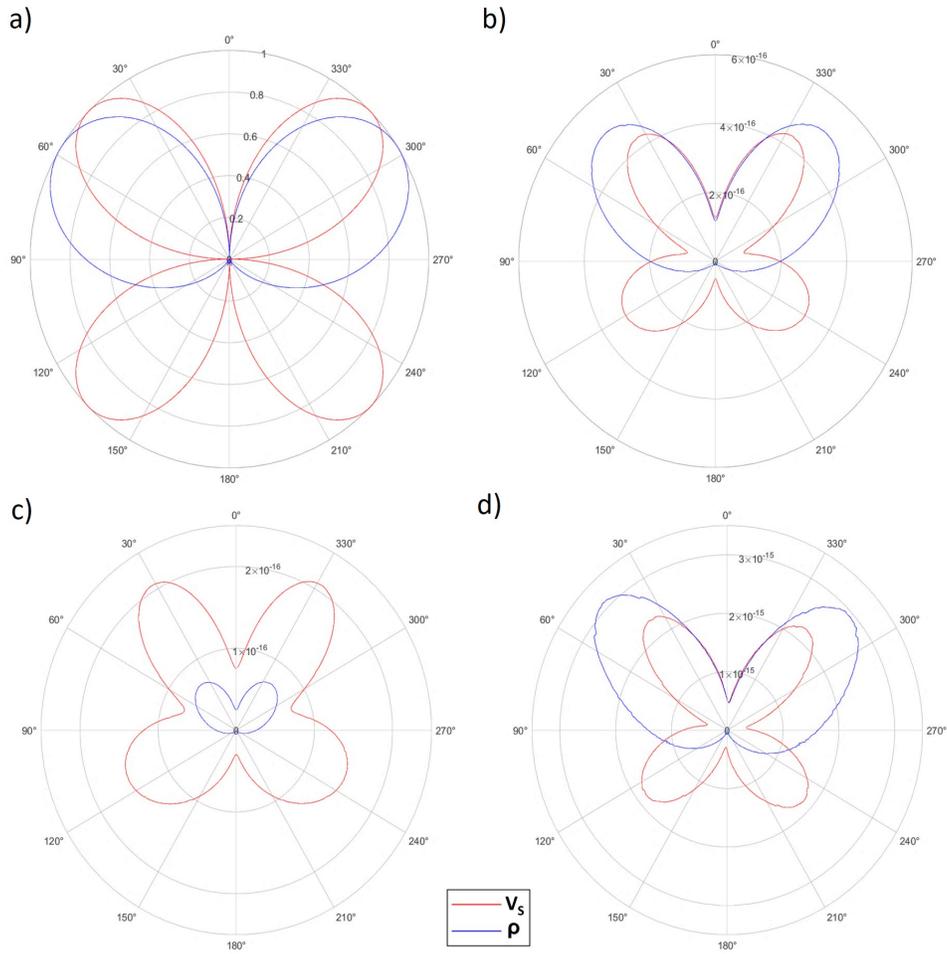


Figure 2.9: Crosstalk assessment with radiation patterns extracted from P-S wavefield. Values were not normalized to study the amplitude variations. (a) Analytic expressions, (b) Case 1, (c) Case 2, and (d) Case 3.

present between each pair of model parameters, suggesting that this characteristic would impact on the inversion of the observed data.

# Chapter 3

## Decorrelation of parameter classes and crosstalk minimization with point-wise Hessians

### Summary

Multiparameter full waveform inversion is commonly affected by crosstalk between parameters of different classes. The reduction of inter-parameter coupled effects is an important task that must be performed to increase confidence in the estimations generated with FWI. These effects are described and corrected by the Hessian, which also impacts the shape of the objective function iso-surfaces and the convergence of the optimization algorithms. This chapter focuses on finding an intermediate model space where the parameter classes are decorrelated, i.e., where the Hessian is an identity matrix, to minimize crosstalk and reach convergence to an accurate minimum point that could be transformed to the  $\rho$ ,  $V_P$ , and  $V_S$  model space. Transformation rules between model spaces were applied in an FWI workflow, using transformation matrices ( $\mathbf{T}$ ) based on a single point in space and which were constructed to satisfy constraints imposed by the Hessian of the intermediate system. Overall, this FWI method produced relatively accurate  $V_S$  estimations, but did not overcome the results of  $V_P$  and  $\rho$  obtained with a reference FWI, since more crosstalk was introduced. However, improvements in the structure of the Hessians concerning those from

the reference inversion were brought for some areas of the model grid, which makes the main decorrelation ideas promising to minimize these coupled effects. The drawbacks were related to the localized approach that was applied to compute  $\mathbf{T}$ , which might need to include, in future work, crosstalk contributions of multiple locations of the model grid.

### 3.1 Introduction

In multiparameter FWI, crosstalk between parameters of different classes means that the physical properties of the rocks are confused during the inversion process, yielding poorly accurate results and producing convergence slowness (Keating and Innanen, 2019). A common strategy to mitigate these coupled effects involves the analysis of radiation patterns, allowing to identify which patterns overlap for a range of angles, and thus for which of the involved physical properties the gradient update will be similar, giving insights about the leakage that will occur (Keating and Innanen, 2019; Métivier et al., 2015). In this sense, FWI workflows are designed considering parameterizations that produce the minimal correlation of their scattering patterns and, in some cases, the dominant parameter class. On the other hand, most local optimization algorithms require a Hessian operator that pre-conditions the gradient by correcting a variety of effects, such as crosstalk through its off-diagonal blocks. Hence, trade-off suppression could be achieved through the manipulation of the Hessian matrix, since no parameter correlation would exist if the corresponding off-diagonal blocks were zero (Métivier et al., 2015; Operto et al., 2013).

Moreover, Innanen (2020a,b,c,d,e) published a series of reports where it is explained (1) how re-parameterizing seismic inversion problems is equivalent to performing a coordinate transform between a cartesian and an oblique system (2) how Gauss-Newton and steepest descent directions are parallel, if the Hessian is an identity matrix, producing favorable convergence properties, and (3) how a model space characterized by a Hessian with the identity matrix structure promises the minimization of crosstalk effects.

In this study, an isotropic elastic full waveform inversion was performed in the frequency do-

main, with the main purpose of producing crosstalk corrected values of  $\rho$ ,  $V_P$ , and  $V_S$  (original model space) by initially inverting for an intermediate set of parameters, that ideally should not contain any leakage, since its Hessians are expected to approximate the identity matrix, and then transforming the final estimates back to the original model space. To achieve this goal, the transformation rules and the numerical procedure proposed by Innanen (2020a,b,c,d,e) were incorporated, i.e., the intermediate re-parameterization ( $r_1$ ,  $r_2$  and  $r_3$ ) was found after solving linear relationships between the parameters of a root model space or  $s$  system ( $\rho$ ,  $c_{11}$  and  $c_{44}$ ) and a transformation matrix. The crosstalk correction would be included within the transformation matrix, since it was constructed in a way that it would allow converting a local Hessian (extracted from a fixed point of the model grid in the  $s$  system) into an identity matrix in the intermediate system. In this sense, it was also aimed to understand how the selected point in space controls this re-parameterization and the scope of accuracy expected in the inversion.

## 3.2 Theory and Methods

### 3.2.1 Full waveform inversion

Full waveform inversion is a technique formulated through local optimization strategies by iteratively trying to match the recorded data or observed data with the synthetic data that would be produced by a set of idealized model parameters (Bunks et al., 1995; Tarantola, 1984; Virieux and Operto, 2009). When both types of data are in close agreement, the idealized model is expected to be a good approximation of the true subsurface characteristics. In this sense, an objective function ( $\phi$ ) is employed as a metric to evaluate how close is the synthetic data to the observed data, and it is subject to a constraint  $\mathbf{C}$  given by the wave equation selected to represent the recorded data (Bunks et al., 1995). Thus, the general form of this constrained optimization problem is:

$$\hat{\mathbf{s}} = \min_s \phi(\mathbf{u}, \mathbf{s}, \mathbf{d}) \quad \text{subject to} \quad \mathbf{C}(\mathbf{u}, \mathbf{s}) = 0 \quad (3.1)$$

where  $\hat{\mathbf{s}}$  corresponds to the inversion result, i.e., the model estimated with the optimization algorithm,  $\mathbf{s}$  is the subsurface model used to compute the synthetic data,  $\mathbf{u}$  is the simulated wavefield, and  $\mathbf{d}$  is the measured or observed data. Equation 3.1 indicates that, to perform an FWI experiment, we need to frame it by deciding the type of wave propagation model we want to use to produce synthetic data and thus how complex would this modeling be. Moreover, other choices are the appropriate parameterization of the earth, the domain we want to work in, the objective function that is suitable for our data, and the type of optimization method that will allow us to find a minimum point closer to the global solution (Eaid, 2021; Keating, 2020). All the decisions that must be made previous to the implementation of FWI will have an important impact on the quality of the obtained results.

Although the inversion strategies applied in the chapters of this thesis could be adapted to different FWI formulations, these investigations were restricted to the isotropic elastic scenario, since crosstalk effects are produced in seismic surface experiments with these model spaces, as observed in the results shown in the previous chapter. As a result, certain wave phenomena, such as dispersion and attenuation cannot be modeled. In other words, a wave propagation model is only an approximation of the physics governing the seismic waves due to the involved simplifications and assumptions about their propagation in the subsurface. The remaining decisions to execute the FWI experiments of this thesis are indicated throughout the different sections of this and the following chapters.

### 3.2.2 Objective function and forward modeling

The objective function in Equation 3.1 can penalize the difference between the observed and the synthetic data that is generated by the current idealized model parameters, but also produce results that are in agreement with some prior information about the Earth, i.e., the objective function can be separated in two parts: a wavefield-dependent data fitting ( $\phi_D$ ) and a model-dependent prior fitting ( $\phi_P$ ), taking the form:

$$\phi(\mathbf{u}, \mathbf{s}) = \phi_D(\mathbf{u}) + \phi_P(\mathbf{s}) \quad (3.2)$$

If no prior information about the subsurface is available, the objective function is only formed by  $\phi_D$ . When cycle-skipping effects have been treated, for instance by selecting a multiscale approach, the most common wavefield-dependent data fitting expression to minimize in FWI problems is the L2 norm or least-squares (Tarantola, 1984), producing a fast convergence of the optimization algorithms and taking the following form in the frequency domain:

$$\phi = \sum_{n=1}^{N_f} \sum_{m=1}^{N_s} \frac{1}{2} \|\mathbf{R}\mathbf{u}_{nm} - \mathbf{d}_{nm}\|_2^2 \quad (3.3)$$

where  $N_f$  is the number of discrete frequencies,  $N_s$  is the number of sources,  $\mathbf{d}$  is the measured data,  $\mathbf{u}$  is the predicted wavefield and  $\mathbf{R}$  is the receiver sampling matrix.

The objective function of equation 3.3 was implemented in the investigations of this thesis without adding prior information of the model parameters. In addition, the frequency domain was chosen to frame the partial derivative equations of the selected model of wave propagation because this domain offers benefits to the application of the forward modeling algorithms, such as the reduction of the forward solutions due to the limited number of frequency components used to invert the data, which is not possible in the time domain (Pratt, 1990). Additionally, the frequency domain is suitable to apply multiscale approaches using single or multiple frequencies at a time, starting with the inversion of the low-frequency components of the seismic data and progressively increasing the frequency values, allowing to mitigate cycle-skipping effects and reducing nonlinearities during the inversion (Bunks et al., 1995; Pratt, 1990; Virieux and Operto, 2009). Moreover, the chosen domain has the advantage of being computationally efficient for smaller (two-dimensional) FWI problems (Pratt et al., 1998), which is the type of problem that is treated in this research.

In the isotropic elastic approximation of the wave propagation, the medium is characterized by the density  $\rho(x,z)$  and the Lamé coefficients  $\lambda(x,z)$  and  $\mu(x,z)$ . In this scenario, the horizontal and

vertical components of the synthetic data are computed through the wave equation described by Pratt (1990):

$$\omega^2 \rho u_x + \frac{\partial}{\partial x} \left[ \lambda \left( \frac{\partial u_x}{\partial x} + \frac{\partial u_z}{\partial z} \right) + 2\mu \frac{\partial u_x}{\partial x} \right] + \frac{\partial}{\partial z} \left[ \mu \left( \frac{\partial u_z}{\partial x} + \frac{\partial u_x}{\partial z} \right) \right] + f_x = 0 \quad (3.4)$$

$$\omega^2 \rho u_z + \frac{\partial}{\partial z} \left[ \lambda \left( \frac{\partial u_x}{\partial x} + \frac{\partial u_z}{\partial z} \right) + 2\mu \frac{\partial u_z}{\partial z} \right] + \frac{\partial}{\partial x} \left[ \mu \left( \frac{\partial u_z}{\partial x} + \frac{\partial u_x}{\partial z} \right) \right] + f_z = 0 \quad (3.5)$$

where  $\omega$  is the angular frequency,  $u_x$  and  $u_z$  are the horizontal and vertical particle displacements, respectively, and  $f_x$  and  $f_z$  are the horizontal and vertical source terms. This process of computing the expected observations or geophysical data from a model or subsurface picture is called “forward modeling” or “forward problem” (Sheriff, 2002; Tarantola, 2005). Equations 3.4 and 3.5 can be solved applying the second order centered finite difference approach for the spatial derivatives, as detailed by Pratt (1990). The finite difference procedure allows to write the isotropic elastic wave equation in matrix form as:

$$\rho \omega^2 \mathbf{u} + c_{11} \nabla (\nabla \cdot \mathbf{u}) - c_{44} \nabla \times (\nabla \times \mathbf{u}) + \nabla (c_{11} - 2c_{44}) (\nabla \cdot \mathbf{u}) + \nabla c_{44} (\nabla \mathbf{u} + \nabla \mathbf{u}^T) + \mathbf{f} = 0 \quad (3.6)$$

where  $c_{11}$  and  $c_{44}$  are elements of the stiffness tensor that relate to the Lamé coefficients through the expressions  $c_{11} = \lambda + 2\mu$  and  $c_{44} = \mu$ . Hence, Equation 3.6 can be structured as:

$$\mathbf{A}(\mathbf{s})\mathbf{u} - \mathbf{f} = 0 \quad (3.7)$$

considering that  $\mathbf{A}$  is the wavefield operator or Helmholtz matrix. Equation 3.7 is equivalent to  $\mathbf{C}(\mathbf{u}, \mathbf{s})=0$  in equation 3.1, imposing that Equation 3.7 must be satisfied when an isotropic elastic wave propagation model is selected. Additionally, to guarantee the appropriate evaluation of Equation 3.7 in the edges of the models treated in this thesis, a Perfectly Matched Layer (PML) region

(Berenger, 1994) was added to the boundaries of the model grids, simulating an infinite spatial extension and avoiding boundary reflections.

### 3.2.3 Re-parameterization of the FWI problem

The model space of  $\rho$ ,  $c_{11}$ , and  $c_{44}$  can be considered the root parameterization of the FWI problem in the elastic scenario when using Equation 3.6. However, this wave equation can adopt any other 3 elastic parameters that are somehow related to this basic parameterization. For instance, in the case of the  $\rho$ ,  $V_P$  and  $V_S$  model space, these relationships are:

$$c_{11} = V_P^2 \rho \quad (3.8)$$

$$c_{44} = V_S^2 \rho \quad (3.9)$$

It is significant to acknowledge that the type of acquisition geometry of the problem and the considered model space to be inverted affect the quality of inversion outputs, introducing a certain level of coupled effects between parameters of different nature, which is case dependent (Pan et al., 2016, 2018, 2019; Tarantola, 1986). Moreover, in this thesis, the parameterization of the model grid was done by assigning the corresponding three elastic parameters to each grid cell of the mesh, representing the physical properties at each location. This approach is commonly applied in FWI problems, and although its use implies the introduction of a vast number of unknowns, it produces large spatial resolution in the results (Keating, 2020).

On the other hand, Innanen (2020a,b,c,d) explains how re-parameterizing seismic inversion problems, such as AVO and FWI, is equivalent to applying a transformation between cartesian and oblique coordinate systems. To change to a different model space it must be considered that the objective function is a scalar quantity, meaning that it is invariant under the transformation to a different coordinate system. However, the model update is a vector, expressed in its contravariant components, while the gradient and the Hessian operator are tensors, expressed in their covariant components; thus, these quantities do change under transformations. Hence, to map a contravariant

vector from an initial system or model space  $s$  to a new system  $r$  (intermediate model space) and backward, some rules are necessary:

$$s^v = t_\mu^v r^\mu \quad (3.10)$$

$$r^v = (t^{-1})_\mu^v s^\mu \quad (3.11)$$

For example, the linear relationships of Equations 3.10 and 3.11 can be expressed in matrix form considering a set of 2 model parameters, i.e.,  $\mu = v = 2$ , represented as a vector  $\mathbf{a}$ :

$$\begin{bmatrix} a^1(s) \\ a^2(s) \end{bmatrix} = \begin{bmatrix} \frac{\partial s_1}{\partial r_1} & \frac{\partial s_1}{\partial r_2} \\ \frac{\partial s_2}{\partial r_1} & \frac{\partial s_2}{\partial r_2} \end{bmatrix} \begin{bmatrix} a^1(r) \\ a^2(r) \end{bmatrix} \quad (3.12)$$

$$\begin{bmatrix} a^1(r) \\ a^2(r) \end{bmatrix} = \begin{bmatrix} \frac{\partial r_1}{\partial s_1} & \frac{\partial r_1}{\partial s_2} \\ \frac{\partial r_2}{\partial s_1} & \frac{\partial r_2}{\partial s_2} \end{bmatrix} \begin{bmatrix} a^1(s) \\ a^2(s) \end{bmatrix} \quad (3.13)$$

All these transformation rules include a transformation matrix  $\mathbf{T}$  that could be constructed to satisfy certain constraints in the problem, such as producing iso-surfaces of the objective function with a particular shape in the transformed model space. Therefore, since scalar quantities do not change under the transformation of the coordinate systems, the minimization of  $\phi$  in the new system will output  $r$  models limited by a range of  $s_\mu$  vectors that produce the same value of the objective function and, because there is only one set of parameters associated to the minimum cost, finding the minimizer of  $\phi$  in the  $r$  model space implies that the minimum point is also found in the  $s$  coordinate system.

### 3.2.4 Optimization algorithms and operators

Solving nonlinear problems with local optimization methods entails the calculation of updates of the model parameters in an iterative fashion. These updates are added to the previous solution, until converging, most generally, to a local minimizer, representing the model that produces the

smallest value of the objective function regarding the neighboring points. Hence, after computing the model updates, the values of the model parameters are modified per iteration ( $k$ ) according to:

$$\mathbf{s}_{k+1} = \mathbf{s}_k + \Delta \mathbf{s} \quad (3.14)$$

where  $\mathbf{s}_{k+1}$  is the updated model,  $\mathbf{s}_k$  is the current subsurface model, and  $\Delta \mathbf{s}$  is the model update. Equation 3.14 is performed until a termination criterium is met, which could be that no more progress is done, the solution already has sufficient accuracy or the number of pre-defined iterations has finished. In addition,  $\Delta \mathbf{s}$  is formed by the search direction, given by the negative of the gradient and, in some cases, the Hessian operator, and a step length ( $\alpha$ ) that controls the distance the algorithm moves towards that direction before updating the model and changing to another direction. The methods chosen to solve the problem impact the accuracy of the obtained results, as well as other aspects of the inversion process such as the computational time and cost. In this section, I describe the gradient and the Hessian computation as well as the optimization strategies that are necessary to understand the FWI workflows proposed in this thesis.

### 3.2.4.1 The gradient

Local optimization algorithms require gradients to minimize the objective function and find descent directions towards the minimum point that produces the best match with the recorded data. The gradient is a vector obtained by computing the first derivative of the objective function with respect to the model parameters,  $s_{\mu}=s_{i,j}$ , i.e., its components indicate the rate of change of the objective function with respect to each parameter class ( $j$ ) independently, per location  $i$  of the mesh. Hence, the gradient of the objective function in Equation 3.3 constrained to Equation 3.7 takes the following matrix form:

$$\mathbf{g} = \frac{\partial \phi}{\partial \mathbf{s}} = \mathbf{R}\mathbf{u}^{\top} \left( \frac{\partial \mathbf{A}}{\partial \mathbf{s}} \right)^{\top} \mathbf{A}^{-1} (\mathbf{R}\mathbf{u}^* - \mathbf{d}^*) \quad (3.15)$$

or, similarly:

$$\mathbf{g} = \frac{\partial \phi}{\partial \mathbf{s}} = -\mathbf{R} \left( \frac{\partial \mathbf{u}}{\partial \mathbf{s}} \right)^\top \Delta \mathbf{d}^* \quad (3.16)$$

where  $\frac{\partial \mathbf{A}}{\partial \mathbf{s}}$  corresponds to the derivative of the wave equation operator with respect to the model parameters. Moreover, the gradient and the model updates ( $\Delta \mathbf{s}$ ) are affected by the selected parameterization of the wave equation; thus, performing the FWI using the root parameterization ( $c_{11}$ ,  $c_{44}$ , and  $\rho$ ) is beneficial to compute the term  $\frac{\partial \mathbf{A}}{\partial \mathbf{s}}$  in Equation 3.15, using the chain rule (Eaid, 2021). For instance, if we are interested in the model space of  $\rho$ ,  $V_P$ , and  $V_S$ , the derivative of  $\mathbf{A}$  with respect to  $V_P$  is:

$$\frac{\partial \mathbf{A}}{\partial V_P} = \frac{\partial \mathbf{A}}{\partial c_{11}} \frac{\partial c_{11}}{\partial V_P} + \frac{\partial \mathbf{A}}{\partial c_{44}} \frac{\partial c_{44}}{\partial V_P} + \frac{\partial \mathbf{A}}{\partial \rho} \frac{\partial \rho}{\partial V_P} = 2V_P \rho \frac{\partial \mathbf{A}}{\partial c_{11}} \quad (3.17)$$

On the other hand, Equation 3.16 consists in the zero-lag cross-correlation of the data residuals with the partial derivatives of the modeled wavefield, i.e., the signals scattered by a localized parameter perturbation, while keeping the rest of the parameters fixed. This cross-correlation picks in  $\Delta \mathbf{d}^*$  the information that is transformed into a model update for a particular position of the model grid. Moreover, the derivative of the objective function of Equation 3.3 with respect to one parameter class is a weighted sum of the model perturbations related to each parameter class. Hence, as explained in Chapter 2, when the radiation patterns of two parameters of different classes overlap over a range of scattering angles, they have the same gradient update, meaning that they impact similarly on the data and it is challenging to properly separate them. This trade-off phenomenon is called crosstalk and is present in multiparameter FWI problems (Métivier et al., 2015; Operto et al., 2013). In other words, crosstalk effects occur because errors in an estimated model parameter class produce data residuals that are ascribed to a different parameter class, resulting in the update of a model parameter in response to a residual that corresponds to another model parameter (Innanen, 2014; Keating, 2020).

### 3.2.4.2 The Hessian

The gradient itself does not provide the correct model update towards the minimum point because it is affected by a series of factors that are inherent to the seismic experiment (Operto et al., 2013). Different optimization strategies require second-order information which is included through a Hessian operator, allowing the convergence of the algorithms to a more accurate minimum point due to the effects that this matrix corrects in the gradient. The linear form of the Gauss-Newton approximation of the Hessian is:

$$H_{(i,j),(k,l)}(s) = R_{I,i} \left( \frac{\partial u_i}{\partial s_{i,j}} \right) \left( \frac{\partial u_k}{\partial s_{k,l}} \right)^* R_{K,k} \quad (3.18)$$

where  $j$  and  $l$  refer to the parameter class and  $i$  and  $k$  to the position. Equation 3.18 expresses the correlation between the scattered wavefield that is caused by a diffractor point on the parameter  $s_{i,j}$  and the scattered wavefield caused by a diffractor point on  $s_{k,l}$  (Operto et al., 2013). Nevertheless, there are various optimization methods that do not explicitly construct this operator, such as the Truncated Newton and the quasi-Newton L-BFGS. Most generally, a Hessian-vector product is used, since storing and inverting the Hessian is computationally expensive. The Hessian-vector product can be written as  $\nabla g_i(s)^\top \Delta s$ , representing the rate of change of the gradient  $g_i(s)$  produced by the rate of change of  $s$ , i.e., a model perturbation.

Operto et al. (2013) and Métivier et al. (2015) illustrate how this full Hessian is organized in  $3 \times 3$  blocks, each of size  $(n_z \times n_x) \times (n_z \times n_x)$ , with  $n_z$  being the number of samples in the vertical direction and  $n_x$  the number of samples in the horizontal direction of the model grid. The elements related to parameter classes and locations of the model grid that are not the same, i.e., when  $j = l$  and  $i = k$ , correct from the gradient the wave-amplitude effects, such as geometrical spreading. In addition, the elements where the parameter classes are of the same nature but the locations of the model grid are not the same, i.e.,  $j = l$  and  $i \neq k$ , correct from the gradient the limited bandwidth effects generated by the directivity of the virtual sources, the source-receiver geometry

and the band-limitations of the sources, which reduce the resolution of the outcomes. Moreover, the elements where the parameter classes are not the same ( $j \neq l$ ), i.e., in the off-diagonal blocks of the full Hessian, describe and correct the existing trade-off between parameters of different nature. Therefore, using second-order information of the objective function is essential to reduce the crosstalk effects in the minimization process and no coupled effects between parameters of different classes would exist if the values within these blocks were zero (Métivier et al., 2015; Operto et al., 2013; Pratt et al., 1998), meaning also that there would be a minimal correlation of the involved scattering patterns.

Figure 3.1 allows to demonstrate that if the values of the full Hessian are extracted at a fixed position of the mesh, a local  $3 \times 3$  matrix can be constructed, corresponding to a point-wise Hessian and characterizing the crosstalk between parameters of different classes only at that location of the model grid. Moreover, in Figure 3.2 it is illustrated that if a perturbation of the parameter classes is made at one fixed position ( $\Delta s$ ) and we want to study the resulting change of the gradient in all locations and for all parameter classes ( $g_i(s)$ ), vertical profiles across the perturbed location can be extracted and, after reshaping them per block of the full Hessian, a point-probes Hessian can be constructed. This type of Hessian is also organized in  $3 \times 3$  blocks and, although it is larger than the local  $3 \times 3$  Hessian, it is still computable. These figures also allow to demonstrate that the point-wise Hessian can be extracted from the point-probes Hessian. Given that in this thesis three parameter classes are considered, the construction of the point-probes Hessian was done by calculating three Hessian-vector products, using a spike as the model perturbation ( $\Delta s$ ) and placing it at a particular location of the model grid.

### 3.2.4.3 The adjoint-state method for the gradient and the Hessian

The need for the partial derivative wavefield with respect to the model parameters in the inner operations of the gradient and the Hessian using Equations 3.16 and 3.18, respectively, implies a high computational cost in the FWI process. Keating (2020), Eaid (2021), and Amundaray (2023) explain thoroughly the development, assumptions, and considerations that must be made

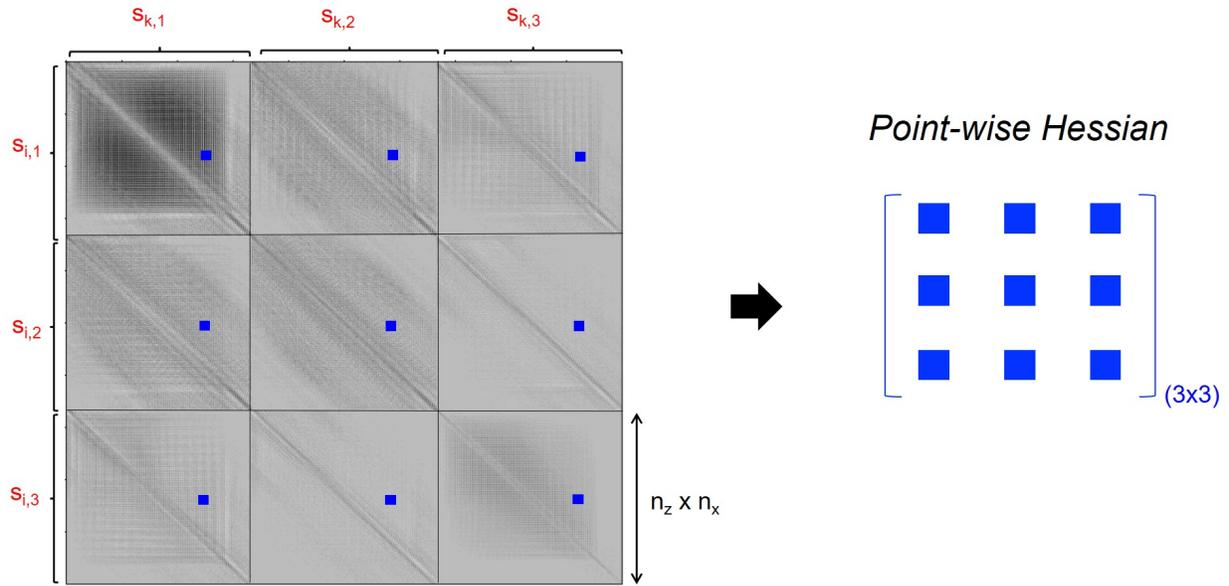


Figure 3.1: Illustration of a full Hessian, modified from (Métivier et al., 2015), and how a point-wise Hessian can be constructed from it.

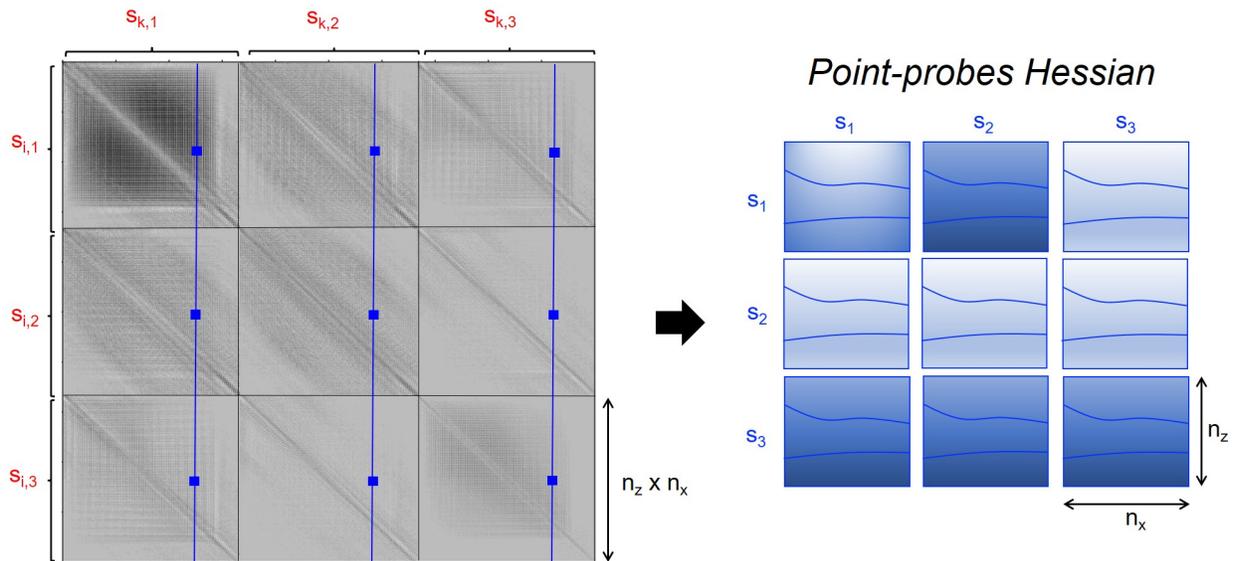


Figure 3.2: Illustration of a full Hessian, modified from (Métivier et al., 2015), and how a point-probes Hessian can be constructed from it.

to produce expressions for the gradient and the Hessian-vector product, which do not require the direct computation of  $\frac{\partial \mathbf{u}}{\partial \mathbf{s}}$ , by following the adjoint-state method (Métivier et al., 2013). In this

section of the chapter, I summarize the main mathematical expression for both tensors, as well as all the inner equations that are necessary for its proper implementation.

To obtain the gradient, the optimization in Equation 3.3 constrained by Equation 3.7, produces a Lagrangian of the form:

$$L = \frac{1}{2} \|\mathbf{R}\mathbf{u}(\mathbf{s}) - \mathbf{d}\|_2^2 + \langle \mathbf{A}(\mathbf{s})\mathbf{u} - \mathbf{f}, \boldsymbol{\kappa} \rangle \quad (3.19)$$

where  $\boldsymbol{\kappa}$  is an unconstrained Lagrange multiplier and  $\langle \cdot, \cdot \rangle$  is an inner product, such that  $\langle a, b \rangle = \sum_{k=1}^{N_s} a_k^\dagger b_k$ . In this sense, after finding the stationary points of the Lagrangian with respect to  $\mathbf{s}$ ,  $\mathbf{u}$ , and  $\boldsymbol{\kappa}$ , the expression of the gradient reduces to:

$$\frac{\partial L}{\partial \mathbf{s}} = \langle \frac{\partial \mathbf{A}}{\partial \mathbf{s}} \bar{\mathbf{u}}, \bar{\boldsymbol{\kappa}} \rangle \quad (3.20)$$

here,  $\bar{\mathbf{u}} = \mathbf{u}$ , satisfying  $\mathbf{A}(\mathbf{s})\bar{\mathbf{u}} - \mathbf{f} = 0$  to find  $\frac{\partial L}{\partial \boldsymbol{\kappa}} = 0$ , and where  $\bar{\boldsymbol{\kappa}}$  is chosen to cancel the derivative of the objective function with respect to the wavefield. Hence,  $\bar{\boldsymbol{\kappa}}$  is obtained from:

$$\frac{\partial L}{\partial \bar{\mathbf{u}}} = \mathbf{R}^\top (\mathbf{R}\bar{\mathbf{u}}(\mathbf{s}) - \mathbf{d}) + \mathbf{A}^\dagger \bar{\boldsymbol{\kappa}} = 0 \quad (3.21)$$

Equation 3.21 is equivalent to the adjoint wave equation. It is composed by the sum of two terms;  $\bar{\boldsymbol{\kappa}}$ , in the second term, corresponds to the adjoint wavefield and the first term corresponds to the source term; the calculation of the Hermitian adjoint operator of  $\mathbf{A}$ , i.e.,  $\mathbf{A}^\dagger$  indicates that the data residuals are back-propagated from the receiver locations.

For the Hessian-vector product, certain procedures are followed similarly to those used to generate the expression of the gradient, but the Lagrangian to be considered is different, taking the form:

$$L = \langle \mathbf{u}(\mathbf{s}), \mathbf{w} \rangle + \langle \mathbf{A}(\mathbf{s})\mathbf{u} - \mathbf{f}, \boldsymbol{\xi} \rangle \quad (3.22)$$

where  $\mathbf{w}$  is an arbitrary vector. In this sense, after finding the stationary points of the Lagrangian

with respect to  $\mathbf{s}$ ,  $\mathbf{u}$ , and  $\xi$ , the expression of the Gauss-Newton Hessian-vector reduces to:

$$\frac{\partial L}{\partial \mathbf{s}} = \left\langle \frac{\partial \mathbf{A}}{\partial \mathbf{s}} \bar{\mathbf{u}}, \bar{\xi} \right\rangle \quad (3.23)$$

here,  $\bar{\mathbf{u}} = \mathbf{u}$ , satisfying  $\mathbf{A}(\mathbf{s})\bar{\mathbf{u}} - \mathbf{f} = 0$  to find  $\frac{\partial L}{\partial \xi} = 0$ , and where  $\bar{\xi}$  is chosen to remove the derivative of the wavefield with respect to the model parameters and find a stationary point. Hence,  $\bar{\xi}$  is obtained from:

$$\mathbf{A}^\dagger \bar{\xi} = -\mathbf{w} \quad (3.24)$$

Equation 3.24 requires knowing the arbitrary vector  $\mathbf{w}$ , which is stated as:

$$\mathbf{w} = \mathbf{R}^T \mathbf{R} \mathbf{J} \mathbf{v} \quad (3.25)$$

In Equation 3.25,  $\mathbf{J}$  is the Jacobian matrix and  $\mathbf{v}$  another arbitrary vector. Both  $\mathbf{J}$  and  $\mathbf{v}$  are not calculated separately because it is not practical. Thus, the product of both tensors is obtained using the derivative of the forward problem (Equation 3.7) with respect to the model parameters ( $s_\mu$ ) multiplied by the elements of the vector  $\mathbf{v}$ , that is:

$$\mathbf{A}(\mathbf{J}\mathbf{v}) = -\mathbf{u} \sum \left( \frac{\partial \mathbf{A}}{\partial s_\mu} \right) v_\mu \quad (3.26)$$

#### 3.2.4.4 Steepest descent optimization method

The steepest descent algorithm is the simplest approach to minimizing Equation 3.1. This method produces model updates that are only influenced by a step length and the gradient, allowing movements along directions that are orthogonal to the iso-surfaces of the objective function and which produce the most rapid decrease. In this sense, the model updates take the form:

$$\Delta \mathbf{s} = -\alpha \mathbf{g} \quad (3.27)$$

These methods always guarantee a descent direction, linear convergence rate with zigzag movements, and robustness. However, although they are computationally inexpensive, their convergence is slow especially on difficult problems (Eriksson, 1996; Nocedal and Wright, 2006; Zou, 2020). Moreover, these algorithms are susceptible to inter-parameter crosstalk effects since the second-derivative information of the objective function is missing, being inappropriate for multiparameter FWI.

### 3.2.4.5 Gauss-Newton optimization method

The Gauss-Newton algorithm produces model updates that depend on a step length, the gradient, and the inverse of the Hessian. The inverse of the second-derivative information of the objective function pre-conditions or corrects the direction indicated by the gradient, resulting in a more accurate search direction. In this sense, crosstalk effects are mitigated since the inversion anticipates with more accuracy how changing one variable causes changes in another (Innanen, 2014). Hence, the model updates can be computed with the expression:

$$\Delta \mathbf{s} = -\alpha \mathbf{H}^{-1} \mathbf{g} \quad (3.28)$$

The convergence rate of this technique is fast and typically quadratic, but it is less robust in the presence of noisy or ill-conditioned data as well as the Hessian could be indefinite far away from the solution. In addition, the explicit calculation of the inverse of the Hessian can be error-prone and involves high computational storage and cost (Eriksson, 1996; Nocedal and Wright, 2006; Zou, 2020). This explains why the inverse of the Hessian operator is commonly approximated through quasi-Newton methods, such as the Limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) algorithm or by applying Hessian-free methods, e.g., the Truncated Newton algorithm. Since these methods approximate the Newton step, by using them, part of the Hessian information is missing and the crosstalk is only partially mitigated.

### 3.2.4.6 The line search

The optimization algorithm searches iteratively a new model that produces a lower value of the objective function along the direction indicated by the gradient (optionally corrected by the Hessian) and the distance the algorithm moves, namely step length ( $\alpha$ ), which can be found after solving:

$$\min_{\alpha > 0} \phi(\mathbf{s}_k + \alpha \mathbf{p}_k) \quad (3.29)$$

where  $\mathbf{p}_k$  is the search direction of the current  $k$ th iteration. Equation 3.29 represents an exact minimization that would output the maximum benefit for the current search direction. However, this is an expensive process, and a one-dimensional inexact minimization is typically carried out, sacrificing accuracy, but conserving computational time (Luenberger and Ye, 2008). To achieve this, several candidate step lengths are generated until finding one that approximates the minimum of Equation 3.29, while enforcing certain conditions. For the experiments of this thesis, the sufficient decrease and curvature conditions, i.e., the two Wolf conditions, were satisfied to find the best approximation of  $\alpha$  per iteration.

Nocedal and Wright (2006) explain thoroughly how both conditions are formulated. The sufficient decrease or Armijo condition ensures sufficient reduction of the objective function during each iteration and prevents overly large steps that could overshoot the minimum. It is measured by the following expression:

$$\phi(\mathbf{s}_k + \alpha \mathbf{p}_k) \leq \phi(\mathbf{s}_k) + C_1 \alpha \nabla \phi_k^T \mathbf{p}_k \quad (3.30)$$

where  $C_1 \in (0, 1)$ . In this thesis, the value  $1 \times 10^{-4}$  was chosen for  $C_1$ . Hence, this condition states that the value of the objective function at the new point must be smaller than or equal to the objective function at the current point plus a reduction factor times the dot product of the gradient and the search direction. On the other hand, the curvature condition ensures reaching the minimum point while avoiding excessively small step lengths and improving the convergence rate.

This condition is described by the following inequality:

$$\nabla\phi(\mathbf{s}_k + \alpha\mathbf{p}_k)^\top\mathbf{p}_k \geq C_2\nabla\phi_k^\top\mathbf{p}_k \quad (3.31)$$

where  $C_2 \in (C_1, 1)$ . Typically,  $C_2 = 0.9$ , and this was the value used throughout this thesis. Hence, this condition requires that the gradient of the objective function at the new point is not too large in relation to the gradient of the objective function before taking a step in the new search direction.

### 3.2.4.7 Impact of the Hessian in the convergence of the algorithms

When the steepest descent approach is selected to solve an optimization problem described by an objective function that exhibits ellipsoidal iso-surfaces with pronounced eccentricities and misalignments, the parameter information is mixed due to problems encountered by the algorithm to reach the global minimum. For instance, for a scalar quadratic objective function with form:  $\phi = \mathbf{s}^\top\mathbf{H}\mathbf{s} + \mathbf{s}^\top\mathbf{q} + \zeta$ , considering a 2-variable model, the term  $\mathbf{H}$  is equivalent to the Hessian; if this Hessian is different from the identity matrix, the iso-surfaces of  $\phi$  are not symmetric and the steepest descent method fails in converging to an accurate result, contrary to the Gauss-Newton method, as observed in Figure 3.3a. Conversely, Innanen (2020c) mathematically proved that, for this quadratic objective function, the steepest descent and the Gauss-Newton updates or directions are parallel, if the Hessian is an identity matrix or a multiple of it, as observed in Figure 3.3b, meaning that a more accurate local minimum is reached, producing reduction of crosstalk and improving the convergence of the algorithms (Innanen, 2020a).

Hence, to map between model spaces, we can design transformation matrices that could be favorable to produce spherically symmetric iso-surfaces of the objective function in the transformed or intermediate model space, by solving the following system of equations, which is the transformation rule for the Hessian between the  $s$  and the  $r$  system (Innanen, 2020c):

$$t_\mu^\lambda H_{\lambda\sigma}(s) t_\nu^\sigma = H_{\mu\nu}(r) = \delta_{\mu\nu} \quad (3.32)$$

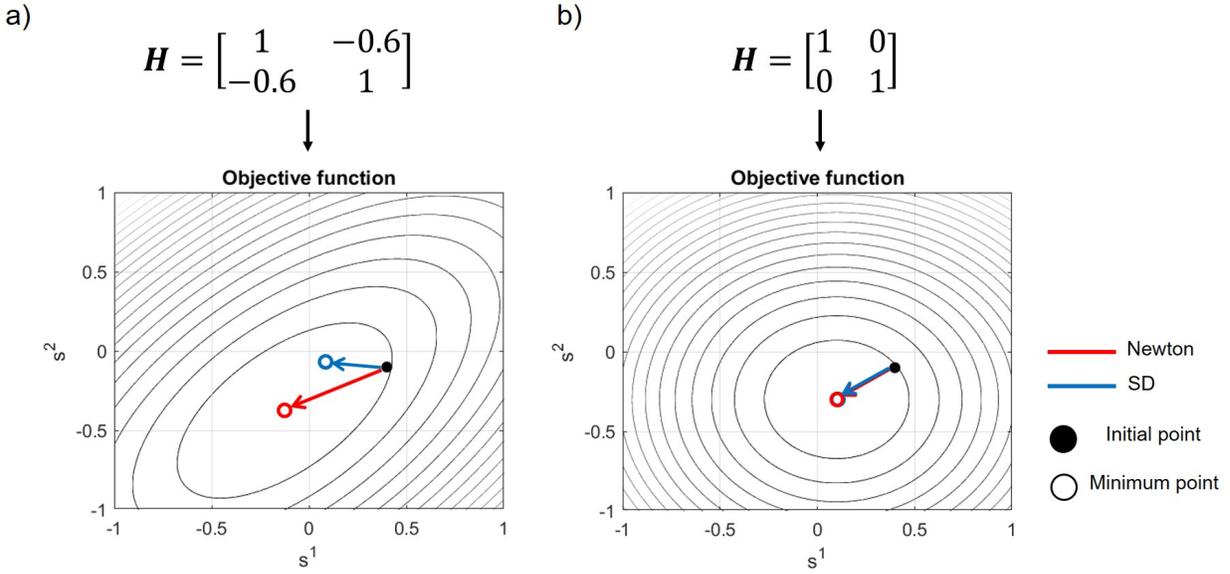


Figure 3.3: Two-variable model of descent-based optimization. a) the Hessian in the quadratic objective function is not the identity matrix; b) the Hessian in the quadratic objective function is the identity matrix. SD means "steepest descent".

with:

$$\mathbf{T} = \begin{bmatrix} t_1^1 & t_2^1 & t_3^1 \\ t_1^{2*} & t_2^2 & t_3^2 \\ t_1^{3*} & t_2^{3*} & t_3^3 \end{bmatrix} \quad (3.33)$$

after considering a three-dimensional problem, such as the one treated in this thesis.

### 3.2.5 Numerical procedure to compute the transformation matrix

Innanen (2020d) explains the development of the numerical procedure used to solve Equation 3.32 in this chapter, knowing the Hessians in the  $s$  and the  $r$  model space. Equation 3.32 can be written in matrix form as:

$$\mathbf{TH}(\mathbf{s})\mathbf{T}^T = \mathbf{H}(\mathbf{r}) = \mathbf{I} \quad (3.34)$$

where  $\mathbf{I}$  is the identity matrix. Considering an  $M$ -dimensional problem for all the involved matrices,  $\mathbf{T}$  has  $M(M - 1)/2$  degrees of freedom, meaning that the problem is non-unique and there are infinite combinations of the lower triangular elements of  $\mathbf{T}$  that allow the transformation of the Hessian into an identity matrix. Hence, for the dimensionality of  $\mathbf{T}$  treated in this thesis, the 3 elements lying below the diagonal in Equation 3.33 (marked with \*) have to be pre-selected and fixed. These numerical procedures were adapted to be performed with the Hessian matrix, but these are most generally referred to as whitening transform algorithms in the literature, based on the eigen-decomposition of the Hessian.

The procedure consists in solving for the unknown elements of  $\mathbf{T}$  in a column-by-column fashion and from left to right. For the first column, there is only one unknown, i.e., the first diagonal element. This element is obtained with a quadratic equation using coefficients that depend on the values of  $\mathbf{H}$  in the  $s$  system, that is:

$$\eta t_{1,1}^2 + \beta t_{1,1} + \zeta = 0 \quad (3.35)$$

where  $t_{1,1}$  is the element of the first row and first column in  $\mathbf{T}$  and:

$$\eta = \mathbf{n}^T \mathbf{H} \mathbf{n} \quad ; \quad \beta = 2\mathbf{n}^T \mathbf{H} \mathbf{k} \quad ; \quad \zeta = \mathbf{k}^T \mathbf{H} \mathbf{k} - 1 \quad (3.36)$$

with:

$$\mathbf{n} = [1, 0, 0, \dots, 0]^T \quad (3.37)$$

$$\mathbf{k} = [0, t_{2,1}^*, t_{3,1}^*, \dots, t_{N,1}^*]^T \quad (3.38)$$

In the following list, I summarize a series of steps that are followed for the calculation of the next  $j$  column of  $\mathbf{T}$ :

1. A matrix  $\mathbf{X}$  is computed as  $\mathbf{T}'^T \mathbf{H}$ , with  $\mathbf{T}'$  being a matrix that collects all the known element

of  $\mathbf{T}$ .

2. A vector  $\mathbf{Y}$  is computed as  $\mathbf{X}\tilde{\mathbf{T}}$ , with  $\tilde{\mathbf{T}}$  being a vector formed by the fixed elements below the diagonal at the current column.
3. An altered version of  $\mathbf{X}$  is created. This is formed with the first  $j$  columns of  $\mathbf{X}$  and appending the vector  $\mathbf{Y}$  to its right-hand side.
4. The altered version of  $\mathbf{X}$  goes through a partial elimination/backsubstitution process, stopping when each row has 3 non-zero elements.
5. A matrix  $\mathbf{Z}$  of size  $(j-1) \times 3$  is formed with the non-zero elements of the matrix resulting from step 4.
6. The following two  $N \times 1$  vectors are constructed:

$$\mathbf{n} = \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_{j-1} \\ 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad \mathbf{k} = \begin{bmatrix} g_1 \\ g_2 \\ \vdots \\ g_{j-1} \\ 0 \\ t_{j+1,j}^* \\ t_{j+2,j}^* \\ \vdots \\ t_{N,j}^* \end{bmatrix} \quad (3.39)$$

where:

$$f_k = -Z_{k,2}/Z_{k,1} \quad g_k = -Z_{k,3}/Z_{k,1} \quad (3.40)$$

7. The diagonal element of  $\mathbf{T}$  in the current column can be found with Equation 3.35, with the

coefficients given by the expressions in Equation 3.36 and using the vectors  $\mathbf{n}$  and  $\mathbf{k}$  from step 6.

8. The remaining upper elements in the column are linearly solved with this equation:

$$\begin{bmatrix} t_{1,j} \\ t_{2,j} \\ \vdots \\ t_{j-1,j} \end{bmatrix} = \begin{bmatrix} f_1 & g_1 \\ f_2 & g_2 \\ \vdots & \vdots \\ f_{j-1} & g_{j-1} \end{bmatrix} \begin{bmatrix} t_{j,j} \\ 1 \end{bmatrix} \quad (3.41)$$

9. Steps 1 to 8 are repeated for the remaining columns in  $\mathbf{T}$ .

### 3.3 Workflow and its adaptation for computational feasibility

A model space where the Hessian is the identity matrix was searched through the construction of  $\mathbf{T}$  with Equation 3.32, and the application of the transformation rule of Equation 3.11. Once every location was converted to the intermediate system ( $r$  model space), an optimization strategy was performed to find the minimum point and the obtained solution was transformed to the original system ( $s$  model space) using Equation 3.10 and later to  $\rho$ ,  $V_P$ , and  $V_S$ . The numerical experiments of this chapter were performed using the steepest descent method because it does not require the calculations and storage of the inverse of the Hessian operator, or its approximations; moreover, because it is the best approach to show the effectiveness of the decorrelation ideas when treating crosstalk effects, since if the model space that produces spherically-symmetric objective functions is found, this algorithm should be able to generate results with levels of accuracy similar to those produced by Hessian-based approaches.

Figure 3.4 contains the FWI workflow that was applied in this chapter. A multiscale approach was executed with 8 frequency bands, each with 4 frequencies; each band started at 1Hz and had equally spaced values, with a maximum frequency of 2Hz for the first frequency band and increasing to 20Hz until the last one. In addition, the initial models were set as the homogeneous

backgrounds around the modeled true heterogeneities, and 20 iterations of the steepest descent algorithm were performed per frequency band.

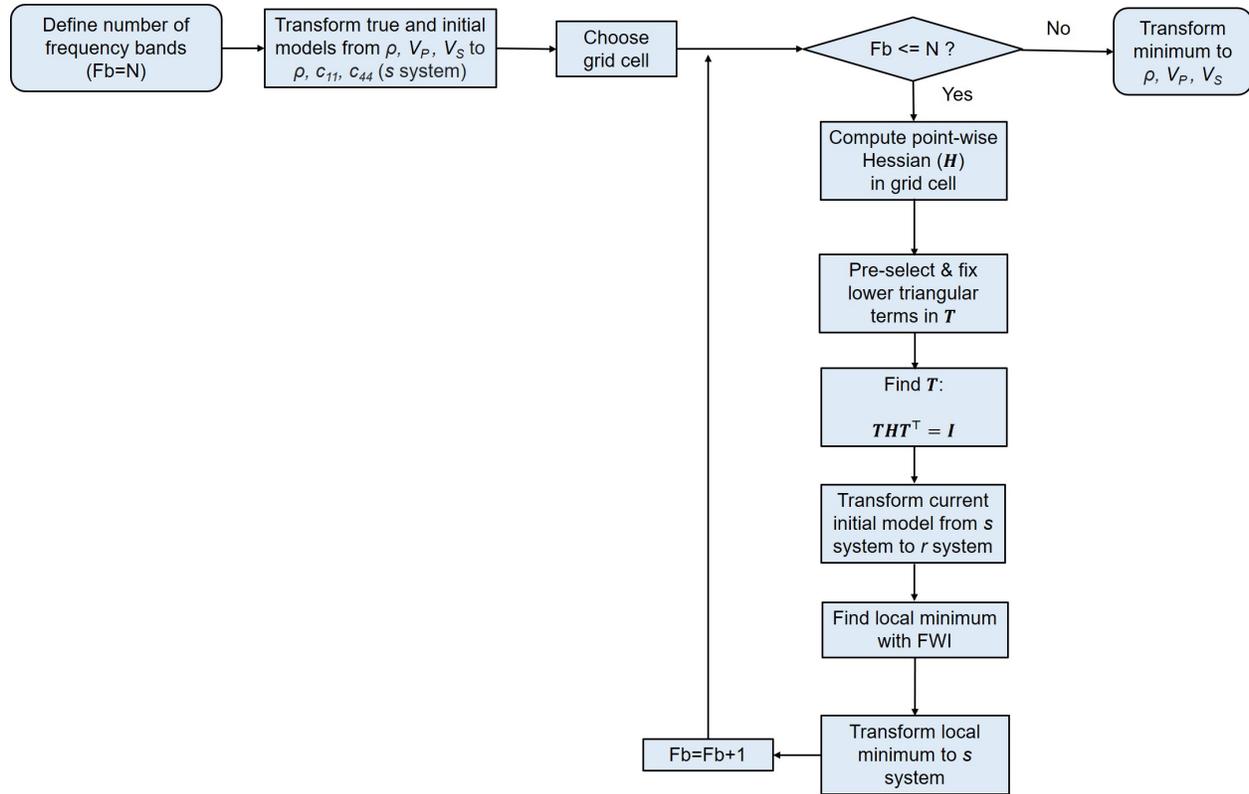


Figure 3.4: Re-parameterization workflow followed in this study.

The main assumption in the numerical experiments of this chapter was to work with a point-wise Hessian of the  $s$  model space and find the matrix  $\mathbf{T}$  that transforms this  $3 \times 3$  Hessian to the identity matrix in the  $r$  model space. This was a first attempt to make the process computationally feasible, since working with the full Hessian would imply a large computational cost. Hence, since parameterizations would be devised based on a single point, the inversion results could be affected by the location selected to compute the transformation matrix. On the other hand, to remove degrees of freedom, the lower triangular elements in  $\mathbf{T}$  were set to zero.

Moreover, a comparison was done between the true models, the estimates from the re-parameterized FWI using the proposed workflow, and from a reference or baseline FWI, i.e.,

without applying transformation rules. Additionally, the point-wise and point-probes Hessians computed with the final estimates were analyzed to determine how close they were to the sought identity matrix. For this purpose, visual inspection of the point-probes Hessians and an appropriate evaluation metric were used. The evaluation metric was a  $3 \times 3$  matrix that results after calculating the norm of each block of the corresponding point-probes Hessian. Its purpose was to capture and summarize the existent crosstalk between parameters of different classes in all the grid cells and not only at the location selected to compute  $\mathbf{T}$ . This metric indicates that the closer to zero the off-diagonal elements, the less crosstalk between parameters of different nature is in all locations of the model grid. Additionally, since some blocks of the point-probes Hessian exhibit sensitivities with varying orders of magnitude, because of the parameter class perturbed to produce the corresponding block, the crosstalk metric and the local  $3 \times 3$  Hessians were normalized, applying the following expression:

$$h_{i,j_s} = \frac{h_{i,j}}{\sqrt{h_{i,i}}\sqrt{h_{j,j}}} \quad \text{with } i = j = 1, 2, 3 \quad (3.42)$$

where  $h_{i,j}$  corresponds to the elements of the Hessian ( $\mathbf{H}$ ) to be normalized; here,  $i$  and  $j$  are the row and column of the matrix, respectively.

### 3.4 Numerical experiments

Twenty-five sources were placed at the top of the model grid, and 98 receivers were placed at the top and bottom, to enhance the illumination of the heterogeneities. Figure 3.5 shows the selected acquisition geometry as well as the dimension of the model grid and true values of  $\rho$ ,  $V_P$ , and  $V_S$ . The reference FWI was performed with the same frequency bands, initial models, optimization strategy, and the number of iterations proposed for the re-parameterized approach. Figure 3.6 contains the results obtained without applying transformation rules. Crosstalk effects are observed around the  $\rho$  heterogeneity and subtle trade-off effects are seen mostly below the  $V_P$  anomalies.

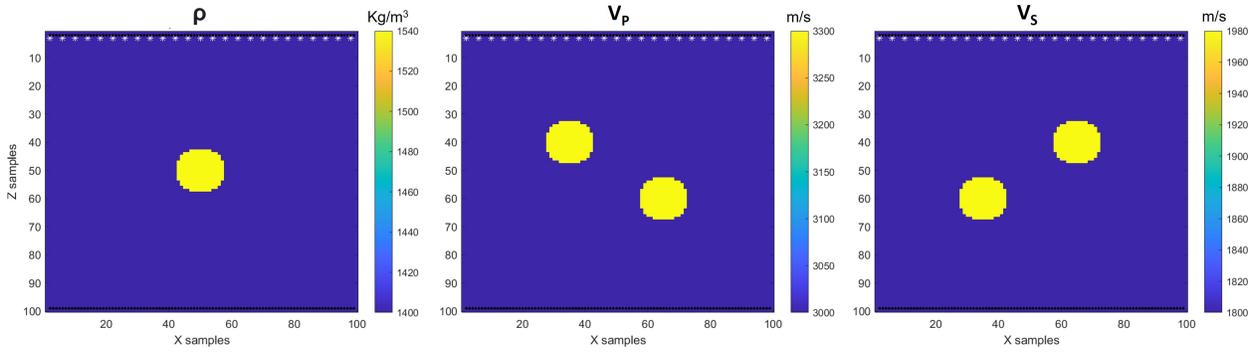


Figure 3.5: True  $\rho$ ,  $V_P$  and  $V_S$  models.

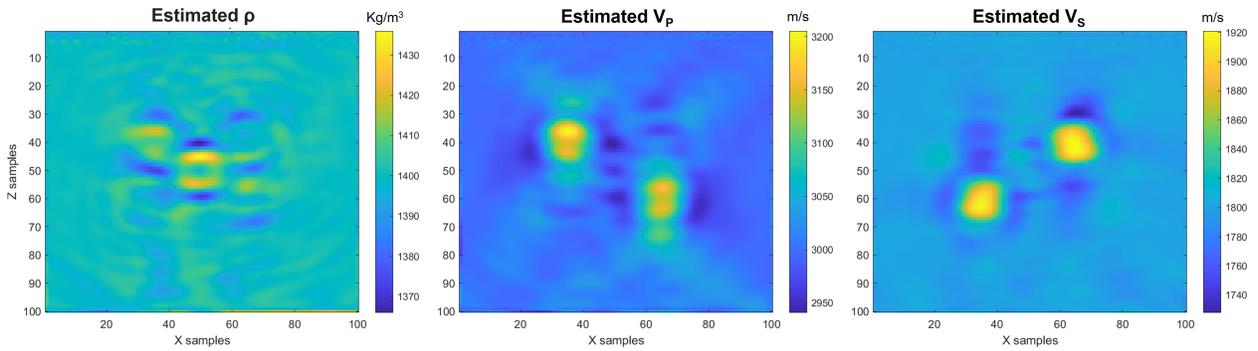


Figure 3.6: Models estimated with a baseline FWI, i.e., without re-parameterization.

Figure 3.7 contains the point-probes Hessian calculated with the estimates from the reference inversion. The point-wise Hessians computed at different locations of the model grid are shown in Figure 3.8, while Figure 3.9 represents the normalized crosstalk metric after perturbing the parameter classes at different locations. Overall, at each tested location, the local Hessians and the crosstalk metrics had an arrangement of values different from the identity matrix. Moreover, the crosstalk metrics indicated strong crosstalk between  $\rho$  and  $V_P$  and between  $\rho$  and  $V_S$ , but much less between  $V_P$  and  $V_S$ , in all locations of the mesh.

The models estimated with the re-parameterized FWI, choosing the location  $x=50$   $z=20$  to compute  $\mathbf{T}$ , are shown in Figure 3.10. This time, much more crosstalk was produced around the  $\rho$  heterogeneity, as well as at the top and bottom of the  $V_P$  anomalies, in comparison to the results from the baseline FWI. Figure 3.11 contains the results of the normalized crosstalk metric. Large

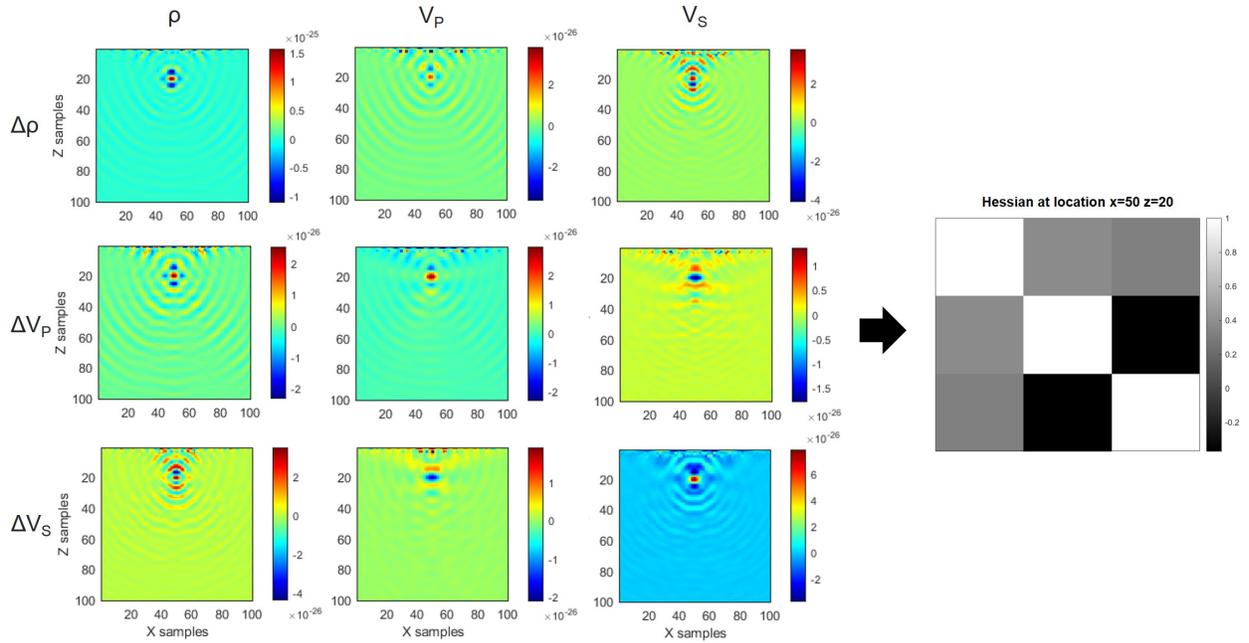


Figure 3.7: Point-probes Hessian computed with baseline estimates, after perturbing parameters at location  $x=50$  and  $z=20$ . The  $3 \times 3$  matrix is the local Hessian at the same location of the perturbation.

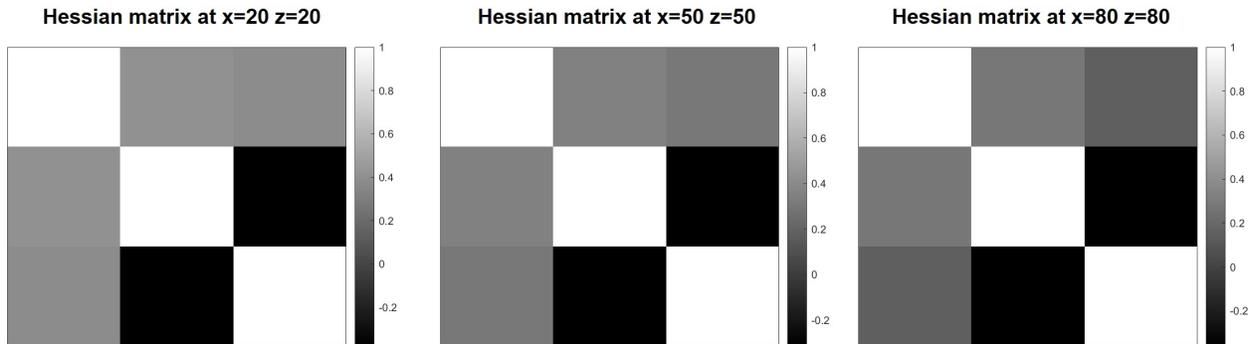


Figure 3.8: Point-wise Hessians computed at different locations of the baseline estimates.

crosstalk was produced between  $r_2$  and  $r_1$  as well as between  $r_3$  and  $r_1$ . However, there were fewer trade-off effects between  $r_2$  and  $r_3$ , in all locations of the model grid. These results might suggest that the  $V_S$  estimations had much more contribution from the parameter class  $r_2$ , while  $\rho$  and  $V_P$  were more influenced by the parameter classes  $r_1$  and  $r_3$ , bringing their coupled effects after transforming from the  $r$  model space to the  $s$  model space.

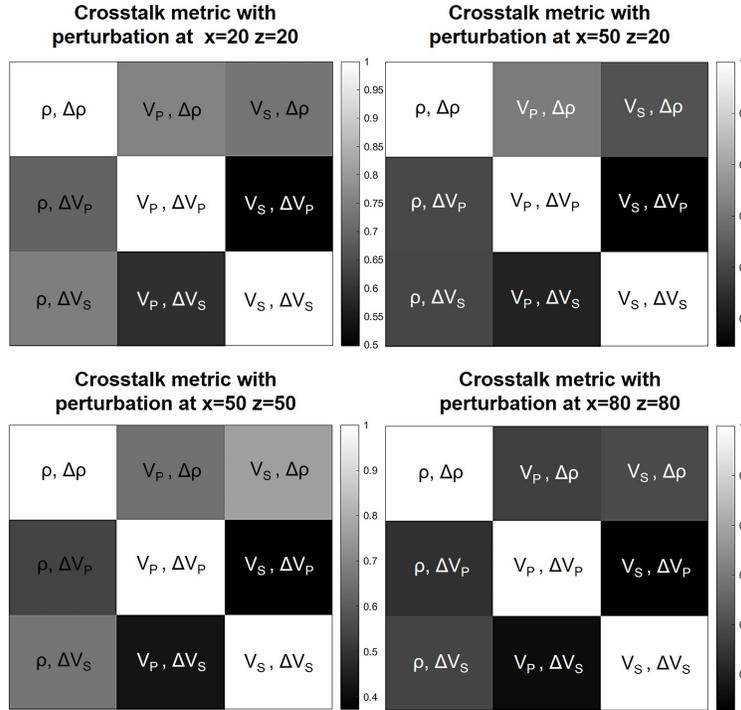


Figure 3.9: Normalized crosstalk metrics computed with baseline estimates after perturbing at different locations to compute the associated point-probes Hessians.

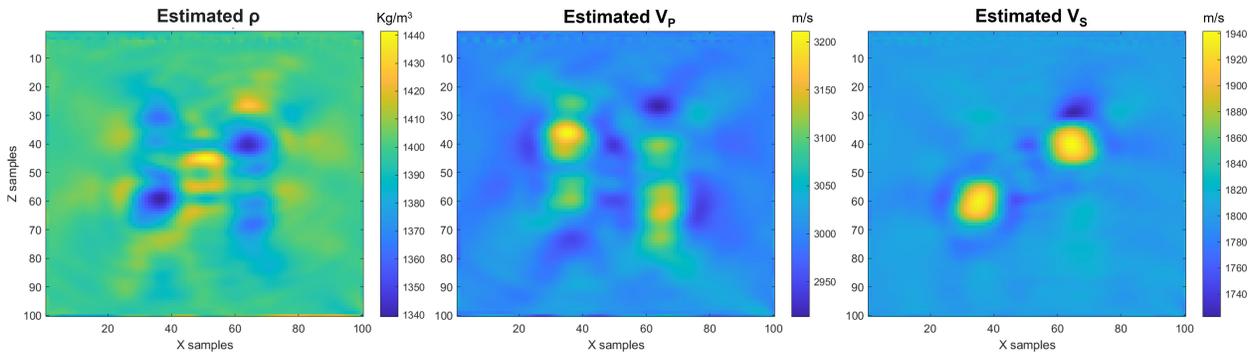


Figure 3.10: Models estimated with a re-parameterized FWI after selecting location  $x=50$  and  $z=20$  to compute  $\mathbf{T}$ .

A point-probes Hessian associated with the final estimates, in the  $r$  model space, and with Figure 3.11 is shown in Figure 3.12. The structure of the identity matrix was observed only at and close to the grid cell chosen to compute  $\mathbf{T}$ , but not in the entire model grid, as would be ideally preferred to minimize crosstalk in a large scale. Outside this small area, different correlation

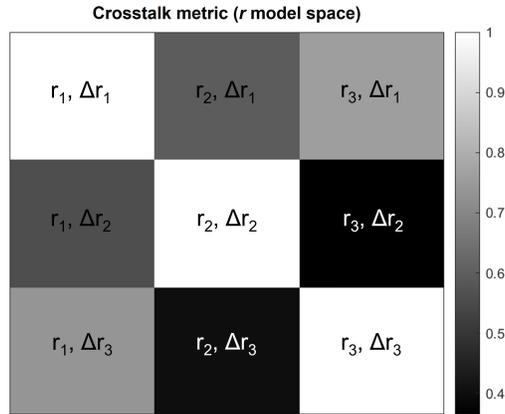


Figure 3.11: Normalized crosstalk metric computed with estimates obtained after selecting the location  $x=50$  and  $z=20$  to compute  $\mathbf{T}$  ( $r$  model space). Parameter classes were perturbed at location  $x=50$  and  $z=20$  to compute the associated point-probes Hessians.

patterns appeared in the blocks of the point-probes Hessian. Hence, this explains why the crosstalk metric summarized a noticeable trade-off between some model parameters in all the locations.

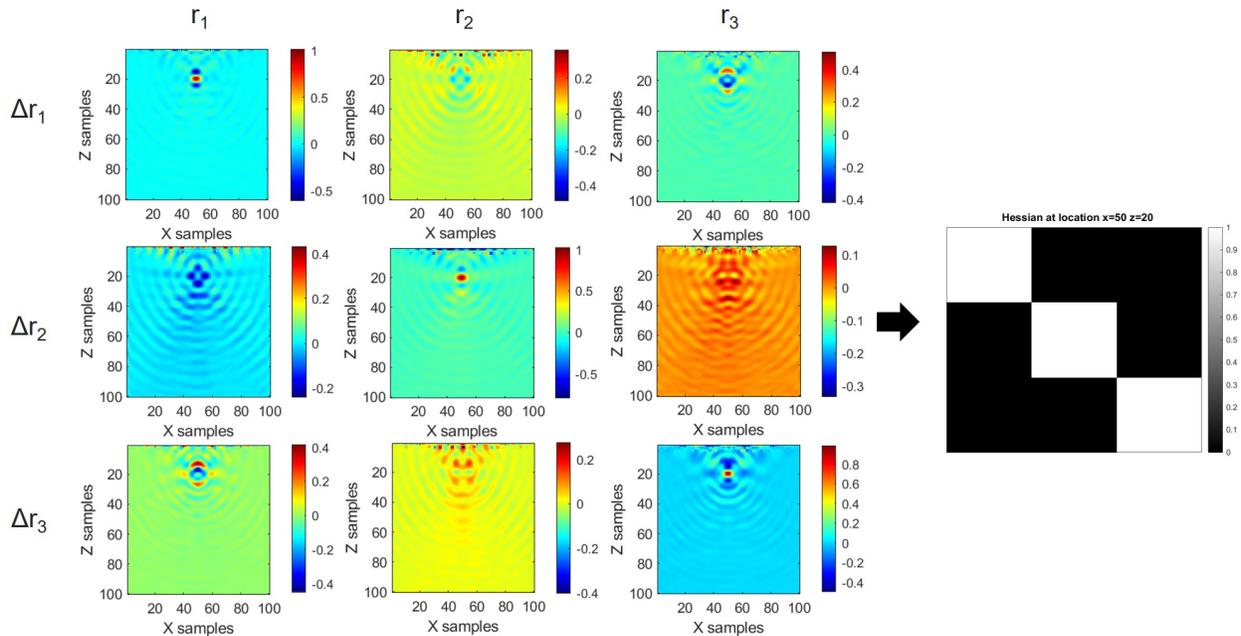


Figure 3.12: Point-probes Hessian computed with estimates obtained from a re-parameterized FWI after selecting the location  $x=50$  and  $z=20$  to calculate  $\mathbf{T}$  ( $r$  model space). The  $3 \times 3$  matrix is the local Hessian at the same location of the perturbation.

Additionally, the re-parameterized FWI was performed by selecting different grid cells to com-

pute the transformation matrix. Figure 3.13 contains the estimated model parameters for three of the tested locations. Moreover, it is demonstrated in Figure 3.14 that regardless the grid cell selected to compute  $\mathbf{T}$ , the optimization algorithm always reached convergence towards the same local minimum, producing results that most generally did not overcome the baseline inversion. Only the estimation of the  $V_S$  heterogeneity was more accurately done with the re-parameterized FWI than with the reference inversion, but for  $V_P$  and  $\rho$  the estimated values of the anomalies were close to those produced by the reference FWI and much more crosstalk was introduced rather than minimized.

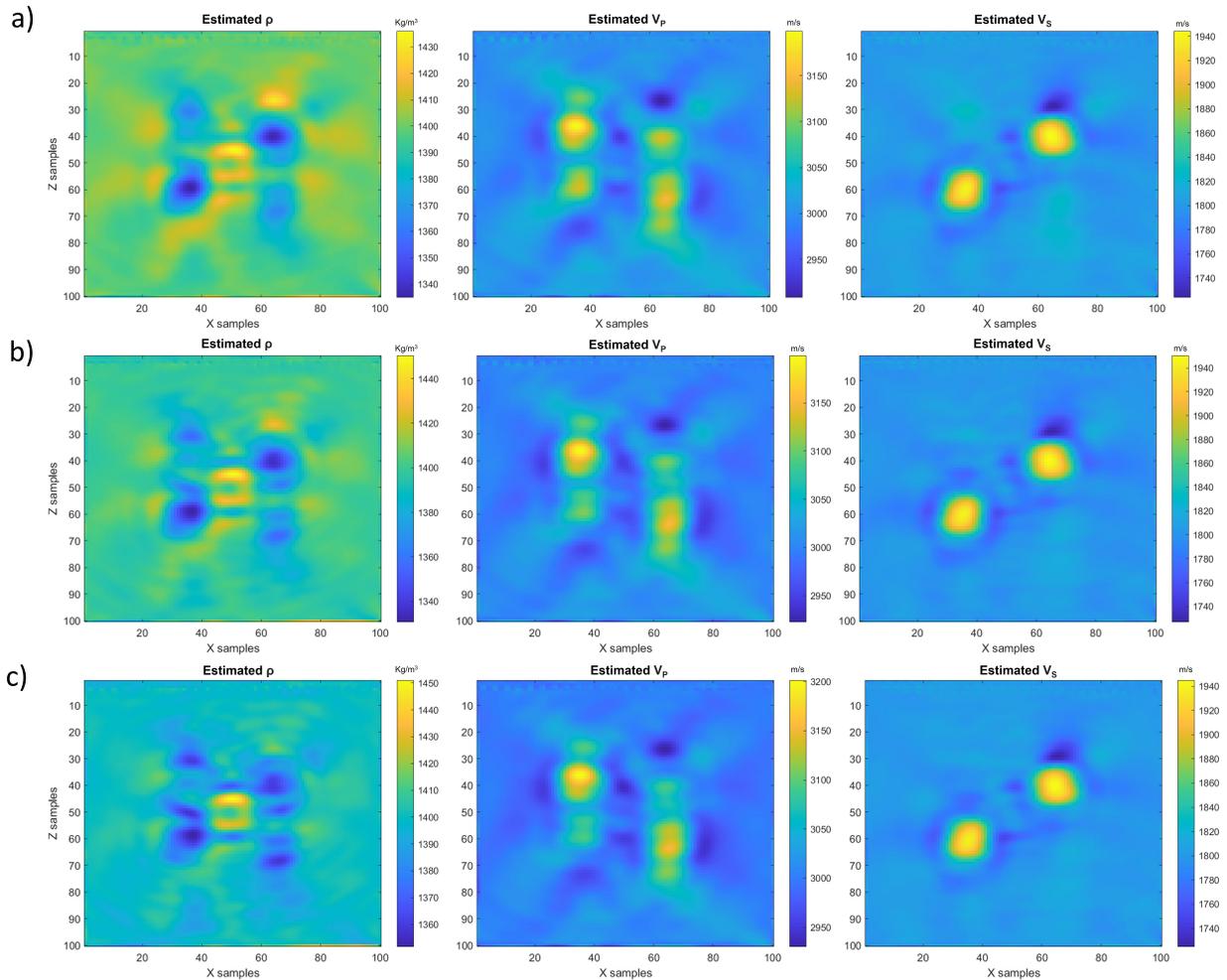


Figure 3.13: Models estimated with a re-parameterized FWI using different grid locations to compute  $\mathbf{T}$ : (a)  $x=20$  and  $z=20$ , (b)  $x=50$  and  $z=50$ , (c)  $x=80$  and  $z=80$ .

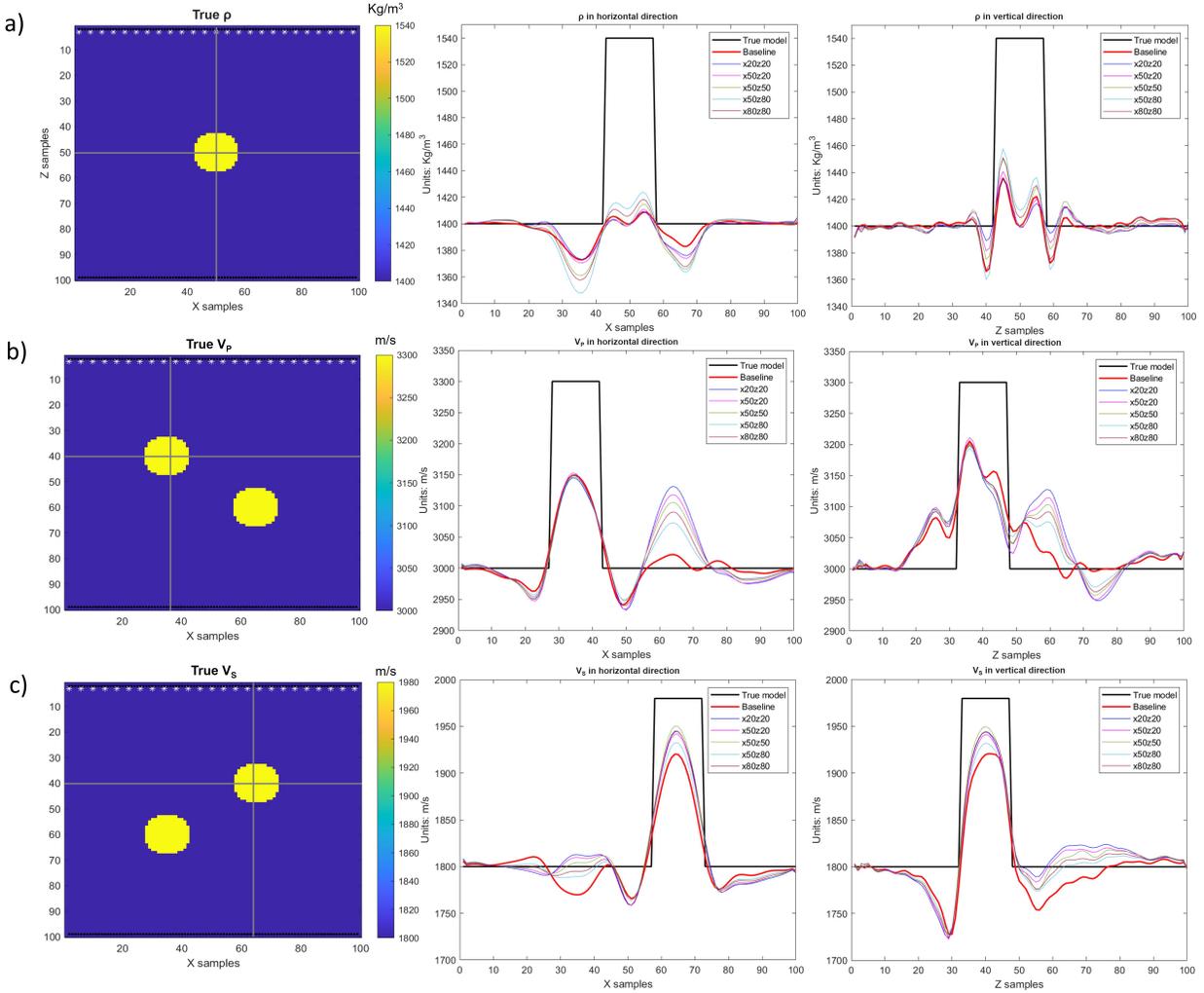


Figure 3.14: Horizontal and vertical profiles extracted from the estimates obtained with a re-parameterized FWI using different locations to compute  $\mathbf{T}$ .

Figure 3.15 allows to demonstrate that when selecting grid cells to compute  $\mathbf{T}$  close or far away from the sources, the identity matrix was only generated around the perturbed location, and although for some  $V_p$  locations the matrix was close to identity, most of the time it was not; thus, it was not enough to make the iso-surfaces of the objective function spherically symmetric and crosstalk effects were introduced to the results.

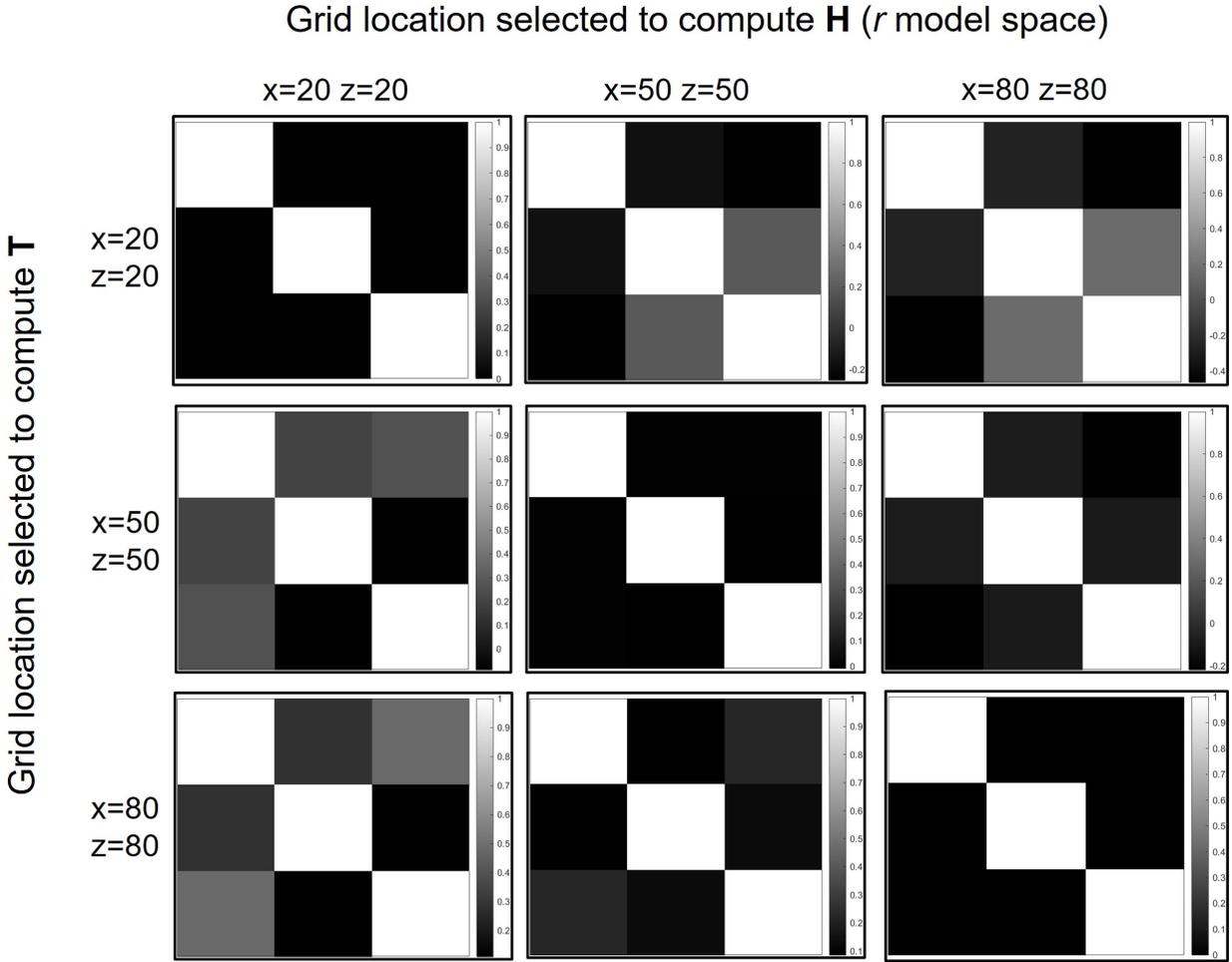


Figure 3.15: Point-wise Hessian matrices at different locations of the estimates in the  $r$  system after inverting with different locations to compute  $\mathbf{T}$ .

### 3.5 Conclusions

The use of the transformation rules and numerical procedures published by Innanen (2020a,b,c,d,e) into an FWI workflow allowed to find a model space where the Hessian was the identity matrix, but only in locations close to the grid cell chosen to compute the transformation matrices ( $\mathbf{T}$ ), generally losing this identity matrix structure at points distant from the selected location, producing different correlation patterns outside this small area, and thus introducing coupled effects between parameters of different classes. Moreover, the steepest descent algorithm reached convergence towards the same local minimum when selecting different locations to

compute  $\mathbf{T}$ ; hence, all models were resolved similarly, generating more accurate  $V_S$  estimates, but not better  $V_P$  and  $\rho$  results than those from the reference inversion, since more crosstalk was introduced. In this sense, the crosstalk metric showed that the trade-off effects between  $V_P$  and  $\rho$  were mostly influenced by existent coupled effects between the intermediate parameter classes  $r_1$  and  $r_3$ . Finally, although working with a transformation matrix based on a single point did not overcome the results of a baseline inversion, the estimates indicated that the decorrelation ideas are promising and that a different numerical procedure to compute  $\mathbf{T}$  should be investigated, aiming to produce a more constant identity matrix structure at all locations of the mesh by considering the contribution of crosstalk in multiple locations and not only at a fixed point.

# Chapter 4

## Using Simulated Annealing to optimize an AVO nonlinear problem: Gaining valuable insights

### Summary

Linearized AVO inversion approaches are based on approximations of the Zoeppritz equations subject to several assumptions, including the limitation of the incidence angles to 35-40°. Thus, in long-offset acquisitions, these approaches fail. In this chapter, an AVO nonlinear inversion is developed so it could be appropriate under these circumstances; this inversion is based on re-parameterizing Zoeppritz equations in terms of the fractional density and compressional and shear impedances. The purpose of this study was to solve a simpler and smaller geophysical problem than the FWI experiments treated in this thesis, with the Simulated Annealing technique and compare the obtained estimations with the outcomes from the Gauss-Newton and steepest descent methods. These experiments would give insights into the performance of the Simulated Annealing algorithm while inverting synthetic seismic data and would allow to understand the feasibility of its application in the FWI problems of this thesis. To achieve this, synthetic noise-free P-P and P-S datasets were computed from a two-layer model. Overall, the Simulated Annealing demonstrated to be an efficient method to estimate the model parameters, mostly outperforming the quality of

the results produced by the local optimization methods, but at the cost of requiring ten times more iterations, computational time, trial and error tests, and a well understanding of the values that the solutions can admit. The learnings gained from this exercise were implemented in Chapter 4 in efforts of finding more appropriate transformation matrices and addressing the challenges stated in Chapter 3.

## 4.1 Introduction

The variation of reflection and transmission coefficients with the incidence angles, and thus offset, is known as Amplitude Variation with Offset (AVO) (Grossman, 2003). This interpretation tool has become important for reservoir description, providing complementary information to the conventional stacked seismic (Downton et al., 2000). In most cases, single-fold data are not pure enough to provide reliable amplitude measurements and the results are doubtful. In such cases, the AVO of the converted mode can be advantageous (Xu and Bancroft, 1997), since including the S-wave velocity and P-S reflectivity allows to augment conventional AVO analysis (Stewart, 1994). Supplementary P-S seismic data increases interpretation confidence, provides additional imaging constraints, and is useful to compute rock property estimates (Larsen, 1999). Large incidence angles for P-S converted waves are typically achieved at shorter offsets than for P-P reflections, meaning that for a given aperture, more complete AVO information is available for P-S data than for P-P data, allowing a more reliable parameter estimation (Ursenbach, 2003).

One method to invert AVO data is to seek linearized approximations of the Zoeppritz equations, and analytically solve these linear equations for the earth parameters (Grossman, 2003). For instance, one successfully applied method is the weighted stacking technique, first proposed by Smith and Gidlow (1987) and expanded by Stewart (1990) and Larsen (1999). This method is based on the minimization of the misfit between the observed data and its theoretical description (synthetic or predicted data) using the Aki-Richards approximations, and although it produces accurate estimates, several underlying assumptions must be met for its proper implementation, such

as considering weak contrasts of elastic parameters, incidence angles smaller than the critical one, incidence angles smaller than  $35^\circ$ , thus very small values of  $\Delta\rho/\rho$  are assumed, and values of the  $V_P/V_S$  ratio between 1.5 and 2 (Smith and Gidlow, 1987; Larsen, 1999).

Moreover, the traditional procedure for angles larger than the critical one is to limit them, but this compromises the reliability of the estimates since they depend on the range of angles considered (Downton and Ursenbach, 2005). Additionally, there is a growing interest in seismic acquisitions with long-offset ranges for different industrial purposes, such as studying reservoirs with interfaces of strong contrast or using streamer recordings to analyze the missing converted S-waves energy, occurring at wider angles of incidence (Skopintseva et al., 2011; Williams et al., 2001). Then, given that linearized inversions would not produce accurate estimations if the available angles are higher than  $35\text{-}40^\circ$ , a more appropriate inversion strategy is necessary to account for scenarios with long offsets/large incidence angles. Since the Zoeppritz equations are not mathematically conditioned to any particular range of angles, a different wave propagation model was developed based on the re-parameterization of the Zoeppritz equations in terms of the fractional impedances and density. However, in reality, a plane wave approximation would be made, since seismic data is not produced by plane but spherical waves.

The objective of developing this AVO inversion was to solve a nonlinear geophysical inverse problem, which would be simpler and with less dimensionality than the FWI problem treated in this thesis by using the Simulated Annealing (SA) technique. These experiments would give insights into the advantages, disadvantages, and general performance of the Simulated Annealing when inverting synthetic seismic data and would allow to understand how it could be applied for other types of geophysical seismic problems. With this purpose, synthetic noise-free broadband and band-limited P-P and P-S reflectivities were generated from a two-layer model which does not produce a critical angle, i.e., only the effects of increasing the incidence angles were considered. Moreover, due to the characteristics of the forward modeling and the Simulated Annealing algorithm, this global optimization strategy was applied considering 3 different cases and comparing the obtained estimations with the ones equivalently produced by local optimization methods.

## 4.2 Forward Modeling

For the forward problem, the plane wave Zoeppritz equations are considered:

$$\mathbf{P} \begin{bmatrix} R_{PP} \\ R_{PS} \\ T_{PP} \\ T_{PS} \end{bmatrix} = \mathbf{b} \quad (4.1)$$

where:

$$\mathbf{P} = \begin{bmatrix} -X & -\sqrt{1-B^2X^2} & CX & \sqrt{1-D^2X^2} \\ \sqrt{1-X^2} & -BX & \sqrt{1-C^2X^2} & -DX \\ 2B^2X\sqrt{1-X^2} & B(1-2B^2X^2) & 2AD^2X\sqrt{1-C^2X^2} & AD(1-2D^2X^2) \\ -(1-2B^2X^2) & 2B^2X\sqrt{1-B^2X^2} & AC(1-2D^2X^2) & -2AD^2X\sqrt{1-D^2X^2} \end{bmatrix}$$

and:

$$\mathbf{b} = \begin{bmatrix} X \\ \sqrt{1-X^2} \\ 2B^2X\sqrt{1-X^2} \\ 1-2B^2X^2 \end{bmatrix}$$

and where:

$$A = \frac{\rho_2}{\rho_1}, B = \frac{V_{S1}}{V_{P1}}, C = \frac{V_{P2}}{V_{P1}}, D = B \frac{V_{S2}}{V_{S1}}, X = \sin(\theta_j)$$

here, a two-layer model is considered; subscript 1 refers to the physical property of the first layer and subscript 2 refers to the physical property of the second layer around the reflector of inter-

est. Equation 4.1 is re-parameterized in terms of the upper medium properties and jumps in the impedances and density, which is familiar from AVO analysis. To accomplish this goal, the terms  $A$ ,  $C$ , and  $D$  are modified according to equation 4.2, which is commonly used when linearizing problems.

$$\frac{Y_2}{Y_1} = \left(1 + \frac{1}{2} \frac{\Delta Y}{Y}\right) / \left(1 - \frac{1}{2} \frac{\Delta Y}{Y}\right) \quad (4.2)$$

Since  $C$  and  $D$  depend on  $V_P$  and  $V_S$ , respectively, these expressions are altered in terms of the impedances  $I=V_P\rho$  and  $J=V_S\rho$ , obtaining the following expressions:

$$C = \frac{I_2 \rho_1}{I_1 \rho_2} \quad ; \quad D = B \frac{J_2 \rho_1}{J_1 \rho_2} \quad (4.3)$$

Applying equation 4.2 to term  $A$  and to terms  $C$  and  $D$  of equation 4.3 produces:

$$A = \left(1 + \frac{1}{2} \frac{\Delta \rho}{\rho}\right) / \left(1 - \frac{1}{2} \frac{\Delta \rho}{\rho}\right) \quad (4.4)$$

$$C = A^{-1} \left(1 + \frac{1}{2} \frac{\Delta I}{I}\right) / \left(1 - \frac{1}{2} \frac{\Delta I}{I}\right) \quad (4.5)$$

$$D = BA^{-1} \left(1 + \frac{1}{2} \frac{\Delta J}{J}\right) / \left(1 - \frac{1}{2} \frac{\Delta J}{J}\right) \quad (4.6)$$

In this sense, each element of  $\mathbf{P}$  is a nonlinear function of  $\Delta I/I$ ,  $\Delta J/J$ , and  $\Delta \rho/\rho$ . By inverting  $\mathbf{P}$  and multiplying it on both sides of equation 4.1, the solution of the four coefficients is obtained for a chosen P-wave incidence angle  $\theta_j$ , with  $j$  varying from 1 to the total number of incidence angles ( $N$ ). As part of the raytracing methodology applied in this study, to work with P-P and P-S data simultaneously, different incident angles for the P-P and P-S mode energy conversions were computed, in order each pair of plane waves could reach the same receiver (offset) (Figure 4.1). Hence, the forward problem was performed twice, once using the set of P-P incidence angles and another time using the set of P-S incidence angles. Subsequently, an appropriate vector of

coefficients  $\mathbf{u}$  was formed by considering that the first coefficient from the P-P forward problem corresponds to  $R_{PP}$ , while the second one from the P-S forward problem corresponds to  $R_{PS}$ .

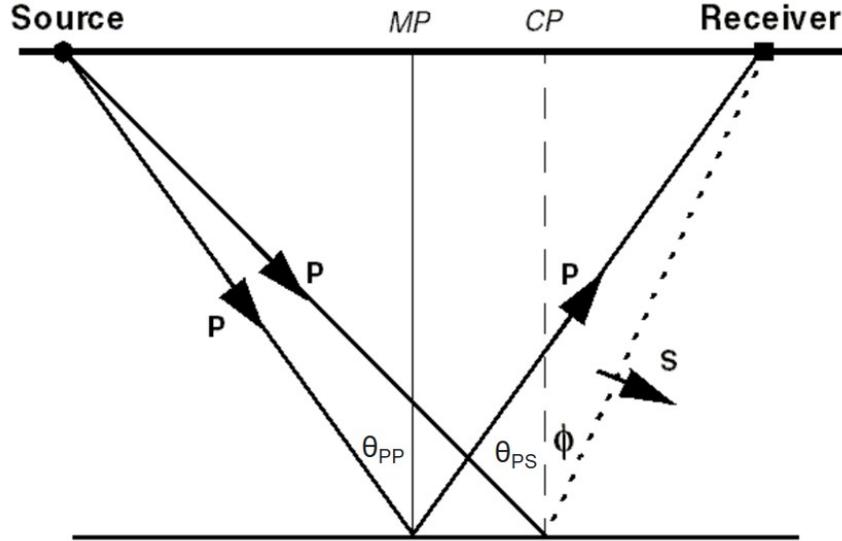


Figure 4.1: Representation of the P-P and P-S incidence angles implemented in this study (modified from Stewart et al. (1999)). “MP” stands for midpoint and “CP” stands for conversion point.

### 4.3 Iterative nonlinear AVO simultaneous inversion

The vector of coefficients  $\mathbf{u}$  contains four elements per incidence angle, but a sampling operator  $\mathbf{S}$  is used to extract the P-P and P-S reflectivities:

$$\mathbf{d}_{\text{pred}} = \mathbf{S}\mathbf{u} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} R_{PP} \\ R_{PS} \\ T_{PP} \\ T_{PS} \end{bmatrix} = \begin{bmatrix} R_{PP} \\ R_{PS} \end{bmatrix} \quad (4.7)$$

The counterpart to the predicted data  $\mathbf{d}_{\text{pred}}$  is the observed data  $\mathbf{d}$ , which are elements of the same space that can be compared through subtraction. When performing an unconstrained opti-

mization, an objective function (without restrictions on the values that the variables can take) is minimized (Nocedal and Wright, 2006). This objective function is constructed by measuring the difference between predicted and observed data vectors, i.e., using the L2 norm:

$$\phi(\mathbf{m}) = \frac{1}{2} \sum_{j=1}^N \left( \mathbf{Su}(\mathbf{m}, \theta_j) - \mathbf{d}(\theta_j) \right)^\top \left( \mathbf{Su}(\mathbf{m}, \theta_j) - \mathbf{d}(\theta_j) \right) \quad (4.8)$$

where  $\mathbf{m}$  corresponds to the vector of model parameters:

$$\mathbf{m} = \left[ \frac{\Delta I}{I} \quad \frac{\Delta J}{J} \quad \frac{\Delta \rho}{\rho} \right]^\top \quad (4.9)$$

The iterative approach requires derivatives of the objective function with respect to all the model parameters. The gradient is a vector in model space pointing in the direction of the most rapid ascent in the objective function; for the current AVO problem, it takes the form:

$$\mathbf{g} = \left[ \frac{\partial \phi(\mathbf{m})}{\partial m_1} \quad \frac{\partial \phi(\mathbf{m})}{\partial m_2} \quad \frac{\partial \phi(\mathbf{m})}{\partial m_3} \right]^\top \quad (4.10)$$

and:

$$\mathbf{g} = \sum_{j=1}^N \mathbf{J}_j^\top \left( \mathbf{Su}(\mathbf{m}, \theta_j) - \mathbf{d}(\theta_j) \right) \quad (4.11)$$

where  $\mathbf{J}$  is a  $2 \times 3$  Jacobian matrix, which is formed by the derivatives of the predicted data with respect to each model parameter. The elements of the Jacobian matrix, per incidence angle, are expressed as:

$$J_\mu^i = S_K^i \frac{\partial u^K}{\partial m^\mu}; \quad i = 1, 2; \mu = 1, 2, 3; K = 1, 2, 3, 4 \quad (4.12)$$

After differentiating both sides of equation 4.1 with respect to each model parameter and inverting the matrix  $\mathbf{P}$ , the following expression is obtained:

$$\frac{\partial u^K}{\partial m^\mu} = -(P^{-1})_L^K \frac{\partial P_M^L}{\partial m^\mu} u^M; \quad M = L = 1, 2, 3, 4 \quad (4.13)$$

When computing  $\mathbf{J}$ , the main new task is to determine all the 48 elements of  $\partial P_M^L / \partial m^\mu$  in equation 4.13, which is a straightforward, yet laborious process. Thus, the Jacobian takes the following matrix form:

$$\mathbf{J} = \begin{bmatrix} J_{PP}^{m_1} & J_{PP}^{m_2} & J_{PP}^{m_3} \\ J_{PS}^{m_1} & J_{PS}^{m_2} & J_{PS}^{m_3} \end{bmatrix} \quad (4.14)$$

However, these are some considerations to account for the selected raytracing approach in Equation 4.13: (1)  $\partial u^K / \partial m^\mu$  is a  $4 \times 3$  matrix. Its rows are formed by the derivative of a coefficient with respect to each model parameter. (2) To construct this matrix,  $\mathbf{u}$  is computed twice, one per each type of incidence angle, to use  $\mathbf{u}_{PP}$  and  $\mathbf{P}_{PP}$  to calculate the first and third rows of  $\partial u^K / \partial m^\mu$  and  $\mathbf{u}_{PS}$  and  $\mathbf{P}_{PS}$  for the second and fourth rows. (3) The construction of  $\partial u^K / \partial m^\mu$  is repeated per each pair of incidence angles, forming  $N$  Jacobian matrices.

On the other hand, the Hessian of the objective function in Equation 4.8 can be computed with the following expression:

$$\mathbf{H} = \sum_{j=1}^N \mathbf{J}_j^\top \mathbf{J}_j \quad (4.15)$$

## 4.4 Simulated Annealing

Simulated Annealing (SA) is a heuristic method to solve optimization problems and corresponds to an approximation of a global optimization algorithm. This technique was introduced by Kirkpatrick et al. (1983) and requires the implementation of the Metropolis Algorithm. Simulated Annealing is based on the analogy of statistical mechanics of carefully annealing solids, where a material is initially melted at high temperatures and then it is slowly cooled until it freezes and crystalizes, i.e., the material is taken from a high energy (poor and unordered solution) to a low energy or ground state (highly ordered material), reaching thermal equilibrium at each temperature. A system formed by a total of  $R_a$  atomic positions, i.e.,  $r = [r_i, i = 1, \dots, R_a]$ , namely

configurations, is in thermal equilibrium with a probability dictated by the Boltzmann or Gibbs distribution, i.e.:

$$P = \frac{1}{Z} \exp \left[ \frac{-E(r)}{k_B T} \right] \quad (4.16)$$

where  $E$  is the energy of the system,  $k_B$  is the Boltzmann's constant,  $T$  is the temperature and  $Z$  is a normalizing constant. As  $T$  decreases, the Boltzmann distribution collapses into the low energy state. Moreover, the form in which the system is cooled influences the type of crystal that is generated; if the system is cooled too fast, the crystal will reach a metastable state and will present many defects with no crystalline order; if the system is cooled slowly enough, the ground state will be reached and a highly ordered and defect-free crystal will be produced.

Simulated Annealing uses operations to mimic this physical process in non-physical optimization problems, by iteratively searching for a finite set of model parameters (equivalent to the atomic positions) that minimizes a real-valued objective function (equivalent to the energy of the system). This method has been used to solve combinatorial optimization problems, encompassing the search of large and discrete values, but also in problems where the model parameters can take continuous magnitudes. The strategy has been applied in a variety of areas, such as image processing, computer design, molecular physics, and chemistry (Faming et al., 2014). The physical experiment requires a temperature; in optimization problems,  $T$  is a parameter with the same units as the objective function and it plays an important role in controlling the iterative process (Kirkpatrick et al., 1983; Rutenbar, 1989). As the temperature decreases, the Metropolis Algorithm produces mainly iterative improvements of the objective function by taking downhill steps, but also sometimes incorporates uphill steps in a controlled manner according to what is dictated by the Boltzmann's distribution, reaching a momentarily worse state to jump out of local minima and potentially be able to find a more accurate solution.

Figure 4.2 is a scheme of the Metropolis Algorithm. Per each temperature of the annealing

process, several random perturbations, drawn from a selected probability distribution, are proposed for each model parameter. These perturbations allow the creation of new configurations of model parameters. Later, the change of the objective function due to this new configuration is evaluated; if this value is negative, the objective function was effectively reduced and, the new configuration of model parameters is accepted and used as the starting point for the next iteration; if  $\Delta E$  is positive, the objective function was increased, and this new configuration may be accepted with a probability indicated by  $P(\Delta E) = \exp[-\Delta E/k_B T]$ , i.e, the Boltzmann's distribution. This probability is simulated by comparing  $P(\Delta E)$  with a number ( $\epsilon$ ) randomly drawn from a uniform distribution in the interval  $[0,1]$ . If  $\epsilon < P(\Delta E)$ , the new configuration is accepted; if  $\epsilon > P(\Delta E)$  the new set of model parameters is rejected and the original configuration is used as the starting point for the next iteration.

In non-physical optimization problems,  $k_B$  is absorbed into  $T$ . At high temperatures (melting stage), SA makes a wide search in the model space, accepting almost every proposal, and the probability of accepting uphill steps is high. As  $T$  decreases, the probability of accepting uphill steps is smaller. Repeating this algorithm many times, as the temperature decreases, simulates the material achieving thermal equilibrium, and thus, the annealing process.

Typical SA algorithms start with a set of random realizations of the model parameters and new states are proposed based on perturbing one parameter at a time, drawing its value from a selected probability distribution (in all the experiments of this thesis, a Gaussian distribution was used), accepting or rejecting the current proposal according to the Boltzmann's distribution, and repeating the process by perturbing the rest of the parameters sequentially. This sequential SA usually converges to accurate minimum points, but generally implies a large computational time that is not always convenient. As a strategy to balance between execution time and the quality of the solution, the sequential procedure was replaced in this thesis by proposing new states based on perturbing all the model parameters at the same time and evaluating the changes that the objective function experiences under each proposed configuration.

An important step in the application of the Simulated Annealing is the definition of an appro-

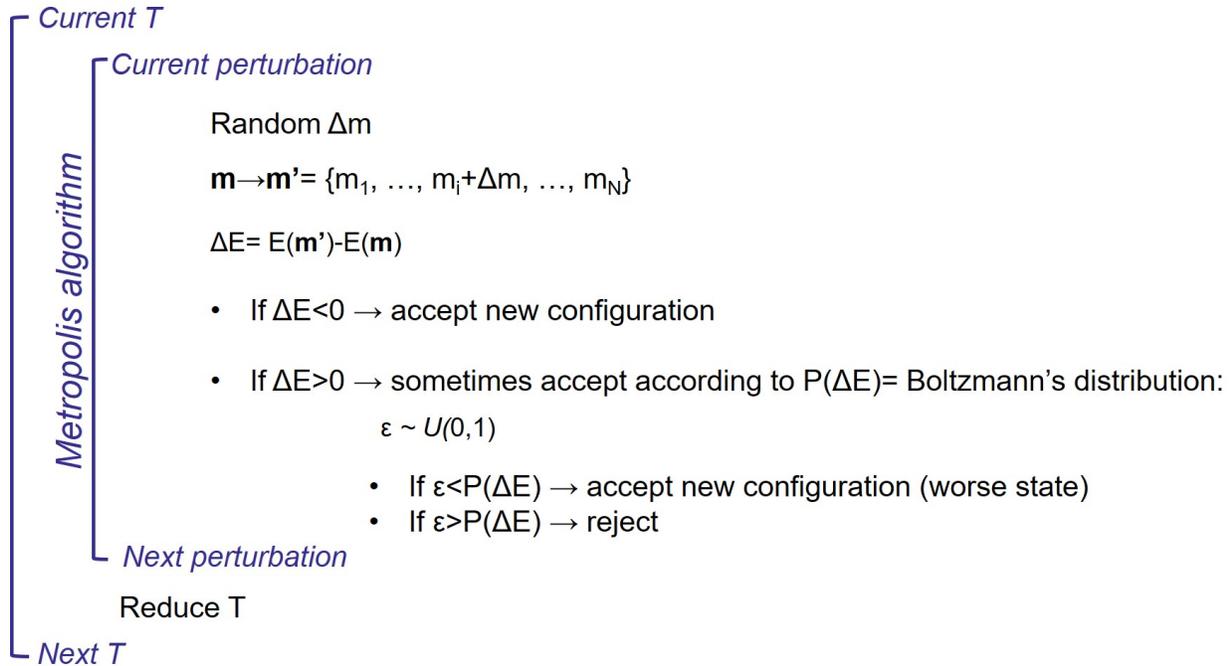


Figure 4.2: Pseudo-code of the Metropolis Algorithm applied in this study.

appropriate annealing schedule so the system can reach equilibrium at each temperature. The annealing schedule typically consists of a starting hot temperature, rules to lower the temperature indicating how much it should be decreased (cooling factor) and the termination criteria, and the number and type of perturbations for the implementation of the Metropolis Algorithm. Generally, the initial temperature is determined empirically, selecting a value by which almost all the proposed configurations are accepted. This is understood through the construction of plots showing the percentage of new configurations accepted versus temperature steps. Hence, temperatures that produce 80% or higher acceptance are considered good values to start the annealing process.

Additionally, the crudest possible manner to reduce the temperature is with the logarithmic expression:

$$T_{k+1} = e \cdot T_k \quad (4.17)$$

being  $k$  the current iteration and  $e$  the cooling factor with a value less than 1;  $e$  usually ranges

from 0.9 for “easy” problems to 0.995 for “difficult” problems, and it is also determined through experimentation. This logarithmic function is workable but seldom fast, thus many other non-increasing expressions have been proposed in the literature, balancing between computational time and quality of the solution (Aarts and Korst, 1989; Faming et al., 2014; Kirkpatrick et al., 1983; Locatelli, 2000). For all the experiments done in this thesis with SA, the temperature was reduced using Equation 4.17. On the other hand, efficient SA algorithms are applied with Gaussian or Cauchy probability distributions for the random generation of model perturbations, and the new configurations must be within user-defined bounds that are acceptable for the solutions of the problem.

## 4.5 Numerical experiments

A two-layer model of solid units in welded contact was designed with elastic properties that do not produce a critical angle (Figure 4.3). A range of offsets was assumed from 0m to 4000m with intervals of 80m and 51 incidence angles were determined through raytracing for the P-P and P-S reflectivities. In general, these incidence angles ranged from  $0^\circ$  to approximately  $53^\circ$  for the P-P dataset and from  $0^\circ$  to approximately  $65^\circ$  for the P-S dataset, as shown in the AVO curves of Figure 4.4, corresponding to values larger than the typical  $35\text{-}40^\circ$  allowed in the linearized approaches. The selected elastic true model meets most of the assumptions underlying the accurate performance of a linearized inversion, i.e., small contrasts across the interface and a value of  $V_P/V_S$  ratio of 1.9. However, in this model, the effects of transmission losses, geometric spreading, and attenuation were not included.

Two different types of one-dimensional experiments were performed, the first one consisted in inverting noise-free broadband data and the second one in using noise-free band-limited data, both being synthetically generated from the interaction of an incident P-wave with the reflector. For each type of experiment, the reflectivities were inverted with the Gauss-Newton, steepest descent, and Simulated Annealing optimization methods.

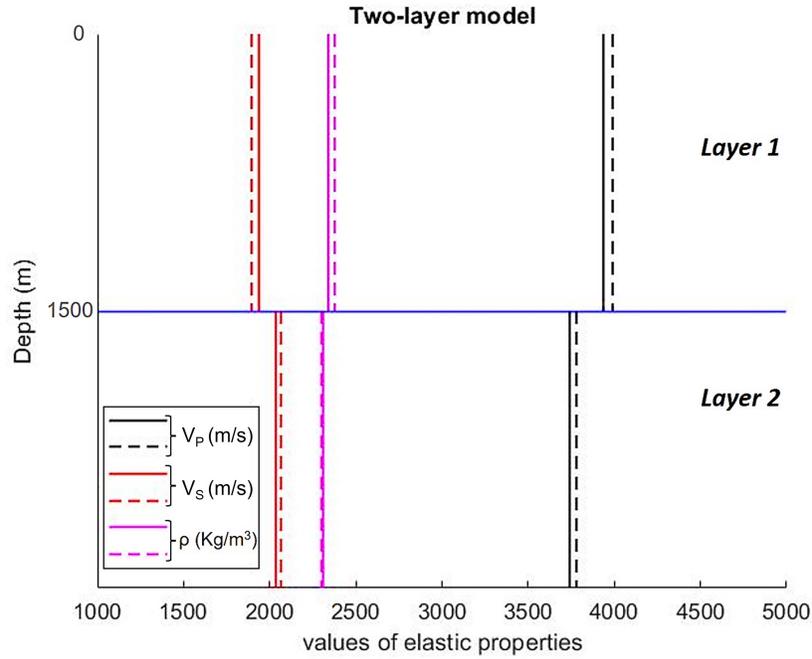


Figure 4.3: True (continuous lines) and initial model (dashed lines) used for the nonlinear AVO inversion.

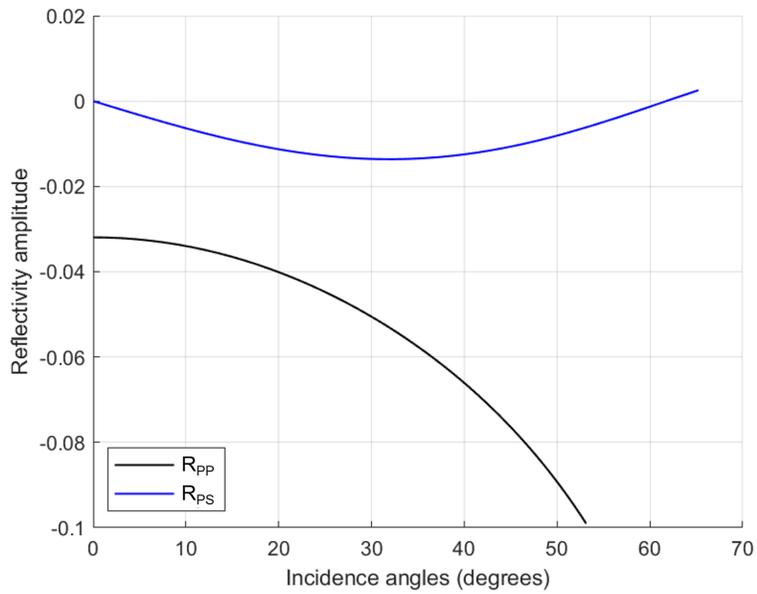


Figure 4.4: P-P and P-S reflection coefficients of the studied two-layer model.

For the local optimization algorithms, the definition of a step length per iteration was required. In the case of the Gauss-Newton optimization, a fixed step length  $\alpha=1$  was considered. Moreover,

for the steepest descent, an inexact line search, using the Wolf conditions, was applied to compute the appropriate step length per iteration, starting with the value  $\alpha=1$ , and using the following constants  $C_1 = 1 \times 10^{-4}$  and  $C_2 = 0.3$ . Additionally, an initial model was constructed with weak perturbations from the true values of  $\rho$ ,  $V_P$ , and  $V_S$ , simulating well-log measurements, as shown in Figure 4.3. Conversely, since the fractional impedances result from multiplying values of velocities and densities, the initial perturbations of the model parameters were not that small; these initial values can be studied at the first point of the plots showing the evolution of the estimated model parameters per iteration.

On the other hand, for both types of experiments, i.e., considering broadband or band-limited data, three different Simulated Annealing methods were implemented varying the type of initial model used. In this context, for this AVO nonlinear problem, after re-parameterizing Equation 4.1 by using the Equations 4.4, 4.5, and 4.6, still the matrices  $\mathbf{P}$  and  $\mathbf{b}$  depend on the term  $B$ , which only involves the values of  $V_P$  and  $V_S$  of the first layer and not the jumps of the elastic properties, thus three cases were derived from this; case 1 consisted in using the initial model of Figure 4.3 and fixing the corresponding value of  $B$  in all the iterations and inverting for  $\Delta I/I$ ,  $\Delta J/J$ , and  $\Delta\rho/\rho$  (the same procedure was used with the local optimization approaches); case 2 consisted in generating a set of random values to form the starting model and fixing the value of  $B$  of case 1 in all the iterations, while inverting for  $\Delta I/I$ ,  $\Delta J/J$ , and  $\Delta\rho/\rho$ ; finally, case 3 consisted in generating random values of all the model parameters to form the starting model, not fixing  $B$ , and inverting for  $\Delta I/I$ ,  $\Delta J/J$ ,  $\Delta\rho/\rho$ ,  $V_{P1}$ , and  $V_{S1}$ . Moreover, since the band-limited data is the closest scenario to the type of data used in the FWI experiments of this thesis, a more extended study was done, in this scenario, by increasing the user-defined bounds for each of the model parameters, i.e., introducing more difficulty to the problem because there would be more values to explore for the algorithm, and assessing the performance of the Simulated Annealing under these circumstances.

For the Simulated Annealing technique, defining a step length was not required, but several tests were performed to find the best set of tuning parameters or annealing schedules that would allow the convergence of the algorithm towards the most acceptable solution. Tables 4.1, 4.2, and

4.3 contain the best annealing schedules found for each case of SA.

<b>Case</b>	<b>Initial temperature</b>	<b>e</b>	<b>Number of temperature steps</b>	<b>Number of perturbations</b>
1	0.6	0.95	400	400
2	0.4	0.956	300	350
3	1	0.945	400	300

Table 4.1: Annealing schedules applied to the SA algorithm when inverting broadband data.

<b>Case</b>	<b>Initial temperature</b>	<b>e</b>	<b>Number of temperature steps</b>	<b>Number of perturbations</b>
1	1	0.96	400	500
2	1	0.96	400	500
3	2	0.945	400	300

Table 4.2: Annealing schedules applied to the SA algorithm when inverting band-limited data.

<b>Case</b>	<b>Initial temperature</b>	<b>e</b>	<b>Number of temperature steps</b>	<b>Number of perturbations</b>
1	15	0.95	400	300
2	10	0.96	400	500
3	15	0.965	500	400

Table 4.3: Annealing schedules applied to the SA algorithm using larger bounds, when inverting band-limited data.

The performance of the optimization algorithms was evaluated through plots of the evolution of the estimated models and the objective function values per iteration. In addition, in the case of the Simulated Annealing technique, the percentage of new configurations accepted per temperature was analyzed to understand how good was the selected initial temperature. Furthermore, all the estimations were evaluated with accuracy tests, applying the following mathematical expression:

$$\% \text{ error} = \frac{\frac{\Delta(I, J, \text{ or } \rho)}{(I, J, \text{ or } \rho)} \Big|_{\text{CALCULATED}} - \frac{\Delta(I, J, \text{ or } \rho)}{(I, J, \text{ or } \rho)} \Big|_{\text{TRUE}}}{\frac{\Delta(I, J, \text{ or } \rho)}{(I, J, \text{ or } \rho)} \Big|_{\text{TRUE}}} \times 100 \quad (4.18)$$

### 4.5.1 Broadband reflectivities

When inverting the noise-free broadband  $R_{PP}$  and  $R_{PS}$  sets, the three optimization methods reached convergence at different iteration numbers, as can be seen in Figure 4.5 and Figure 4.6 with the evolution of the model parameter values and objective function as the iterations progressed. In all these experiments, the estimated models were close to the true values and these also produced data similar to the observed one, since the values of the objective function were close to zero in the last iteration number. However, in case 3 of the SA experiments, the estimated values of  $V_P$  and  $V_S$  for the first layer were not as close to the true values, in comparison to the remaining model parameters, as observed in Figure 4.7.

The Gauss-Newton method had the fastest convergence of all the applied techniques, finding invariable values at the second iteration and taking 0.09 seconds to complete all the iterations. The steepest descent method reached convergence at the iteration 17, taking 0.2 seconds to complete all the iterations. Moreover, depending on the case, the Simulated Annealing required hundreds of iterations or, in this case, temperature steps, to converge into an acceptable value, exhibiting the slowest performance of the three optimization methods; for instance, in case 1, almost 79 seconds were required to complete all the temperature steps. On the other hand, Figure 4.8 contains the plots of the percentage of configurations accepted per temperature step, for all the cases tested with Simulated Annealing. It can be seen that at the melting stage de percentage of acceptance was above 80%, which means that the chosen initial temperature was adequate.

The estimates of  $\Delta I/I$ ,  $\Delta J/J$ , and  $\Delta \rho/\rho$  were numerically compared with the true values from the two-layer model through the accuracy tests of Equation 4.18. Figure 4.9 contains the comparison of the resulting percent errors. It is observed that the results from both local optimization algorithms got trapped into the same local minimum, producing fractional impedances that were

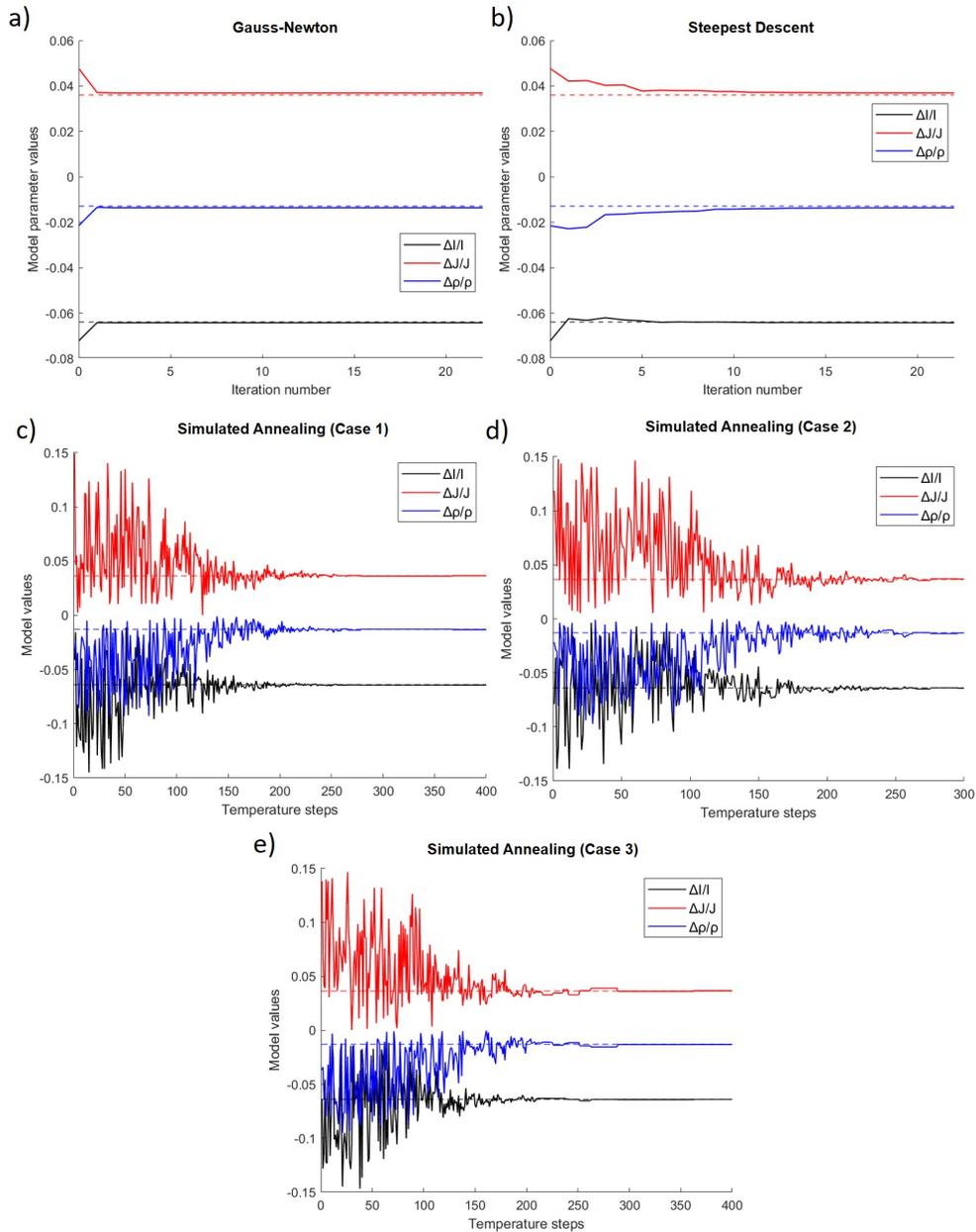


Figure 4.5: Evolution of the estimated model parameters obtained from broadband data with different optimization methods. Continuous lines represent the estimates and dashed lines the true models.

more accurate than the fractional density. Furthermore, the Simulated Annealing, with the selected annealing schedules, outperformed the local methods since the estimations of the model parameters were similar or more accurate than those from the Gauss-Newton and steepest descent techniques, especially in the case of  $\Delta J/J$  and  $\Delta \rho/\rho$ .

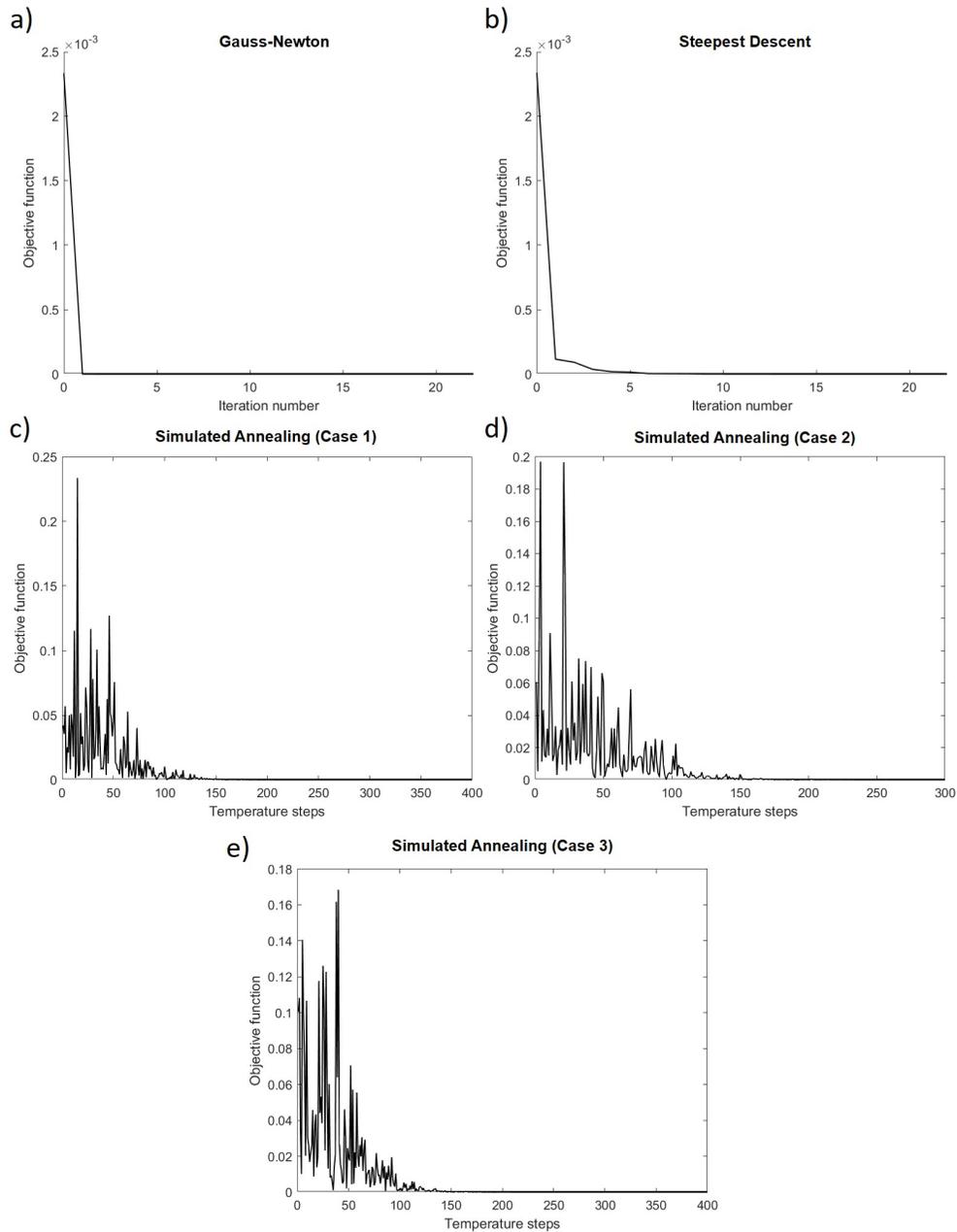


Figure 4.6: Evolution of the objective function produced from inverting broadband data with different optimization methods.

## 4.5.2 Band-limited reflectivities

To continue evaluating the effectiveness of the Simulated Annealing technique in these types of nonlinear problems, both synthetic noise-free P-P and P-S datasets were filtered using an Ormsby

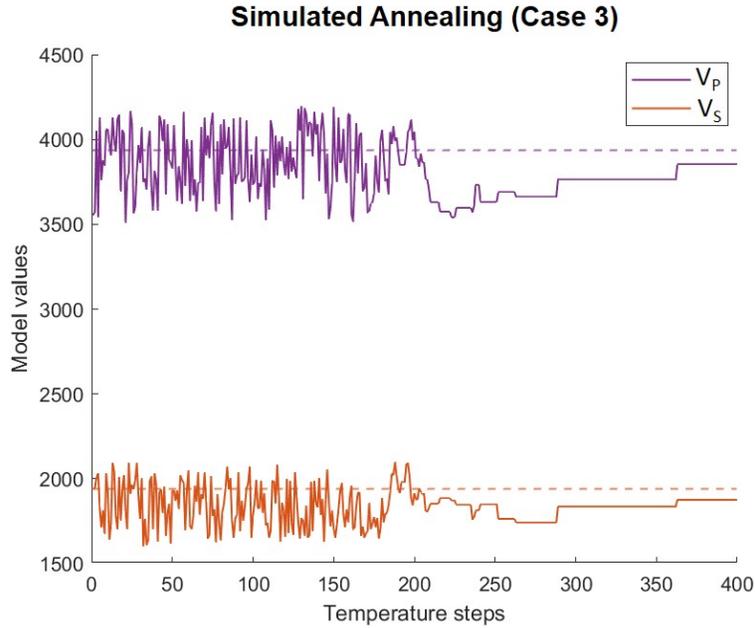


Figure 4.7: Evolution of the estimated  $V_P$  and  $V_S$  values for the first layer obtained from broadband data with case 3 of SA. Continuous lines represent the estimates and dashed lines the true models.

wavelet and applying a different range of frequencies depending on the type of data. For the P-P reflectivities, the set of frequencies used was 5-10-60-75 Hz, and for the P-S dataset, the frequencies used were 5-10-40-55 Hz. The considered wavelets are shown in Figure 4.10 and the NMO-corrected synthetic CMP and CCP gathers are shown in Figure 4.11. The missing frequencies in the measured data happen as a result of (1) the earth attenuating the high frequencies of the wavelets, (2) some seismic sources not being capable of producing low frequencies, and (3) only some special geophones reliably recording the energy at the surface. Additionally, band-pass filters are usually applied, in the processing stage, to eliminate the low-frequency ground-roll or some coherent high-frequency noise, and with this, some desired data can be removed (Mahmoudian and Margrave, 2003). The missing low frequencies (0-5 Hz) in the data are essential for creating the character or scale of the impedance log and the higher frequencies contribute to the detail of the impedance log (Lloyd and Margrave, 2011).

When the reflection coefficients are convolved with the wavelet, the inversion of such band-limited data would result in band-limited impedances (Mahmoudian and Margrave, 2003). There-

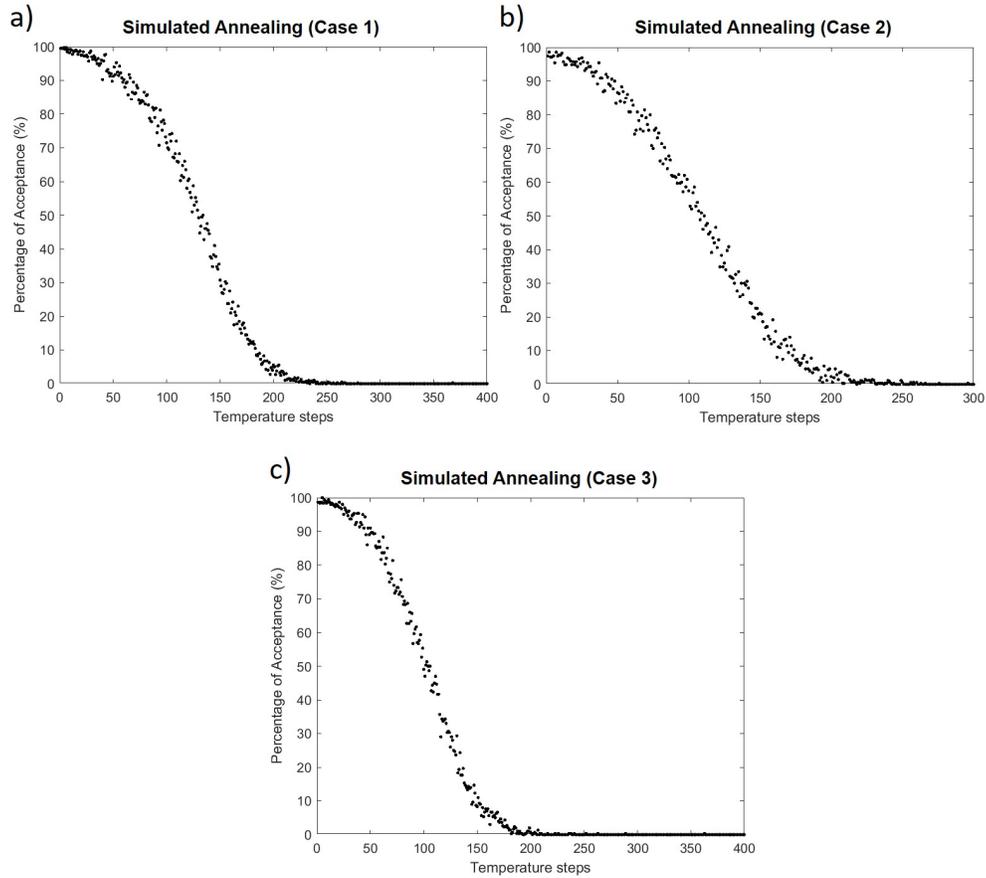


Figure 4.8: Percentage of new configurations accepted per temperature while inverting broadband data for the three cases of SA.

fore, the accuracy of the results was evaluated with Equation 4.18, comparing the estimates against the band-limited true fractional values, since there is a limit on how good is the data, and thus it is necessary to evaluate the results against a fair benchmark.

Figure 4.12 and 4.13 contain the evolution of the estimated model parameters and objective function per iteration with each algorithm, and for the Simulated Annealing, the plots of the percentage of acceptance per temperature step are shown in Figure 4.14. Moreover, in Figure 4.15 the accuracy tests are illustrated. All these plots helped to demonstrate that, once more, both local optimization algorithms got trapped into the same local minimum, and the Simulated Annealing technique produced, for the three cases, similar values of  $\Delta I/I$  and  $\Delta J/J$  with respect to those produced with the Gauss-Newton and the steepest descent method, but significantly more accurate

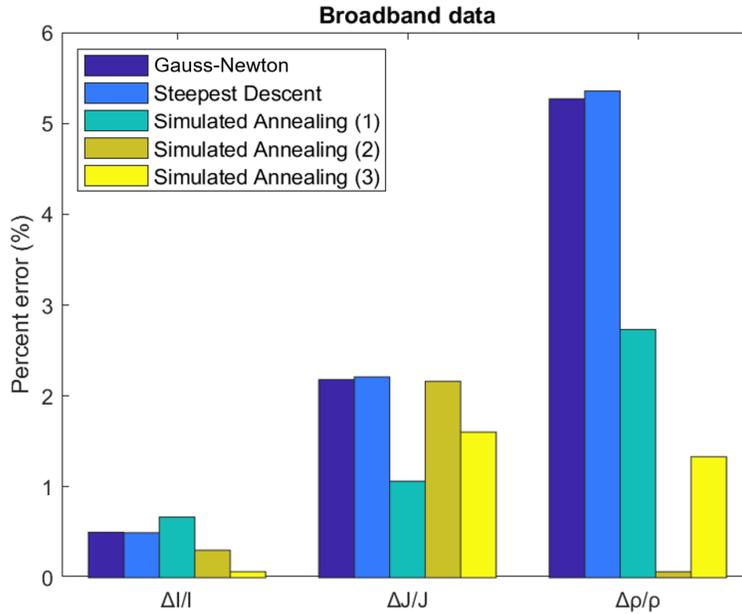


Figure 4.9: Accuracy tests performed with the final estimates obtained from broadband data using different optimization methods.

results of  $\Delta\rho/\rho$ . Moreover, in Figure 4.16 it is showed that in case 3 of the SA experiments, the estimated value of  $V_P$  for the first layer was almost identical to the true value, but the value of  $V_S$  was not as close.

#### 4.5.2.1 Increasing the bounds in SA

To understand the influence of the user-specified upper and lower bounds for the proposal of new configurations in the SA algorithm, the main experiment of this section was repeated, but increasing the width of the bounds by 30% for all the model parameters. This would help to understand how well the method convergence to an accurate solution if the bounds to draw new configurations are not so narrow and close to the right answer. Figure 4.17 and 4.18 contain the evolution of the model estimates and the minimization of the objective function as the iterations progress, while Figure 4.19 corresponds to the plots of the percentage of new configurations accepted per temperature steps for all the Simulated Annealing cases.

When increasing the bounds, the accuracy of the results obtained with Simulated Annealing

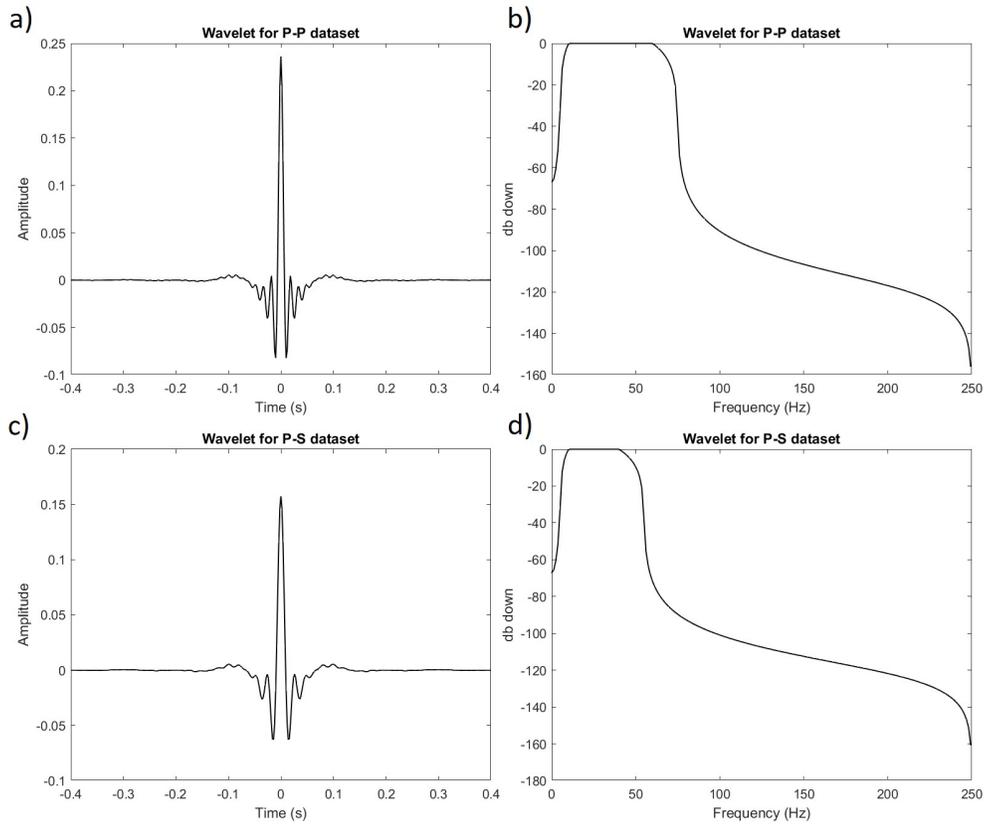


Figure 4.10: Ormsby wavelet used to filter the P-P dataset a) in time and b) in frequency. Ormsby wavelet used to filter the P-S dataset c) in time and d) in frequency.

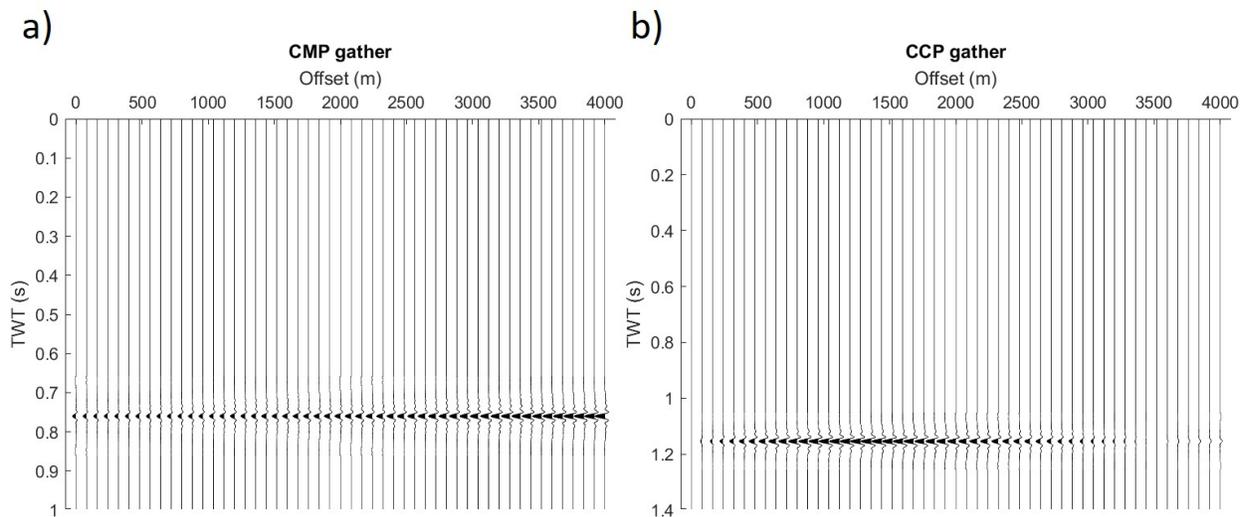


Figure 4.11: Synthetic seismograms for the a) P-P dataset and b) P-S dataset.

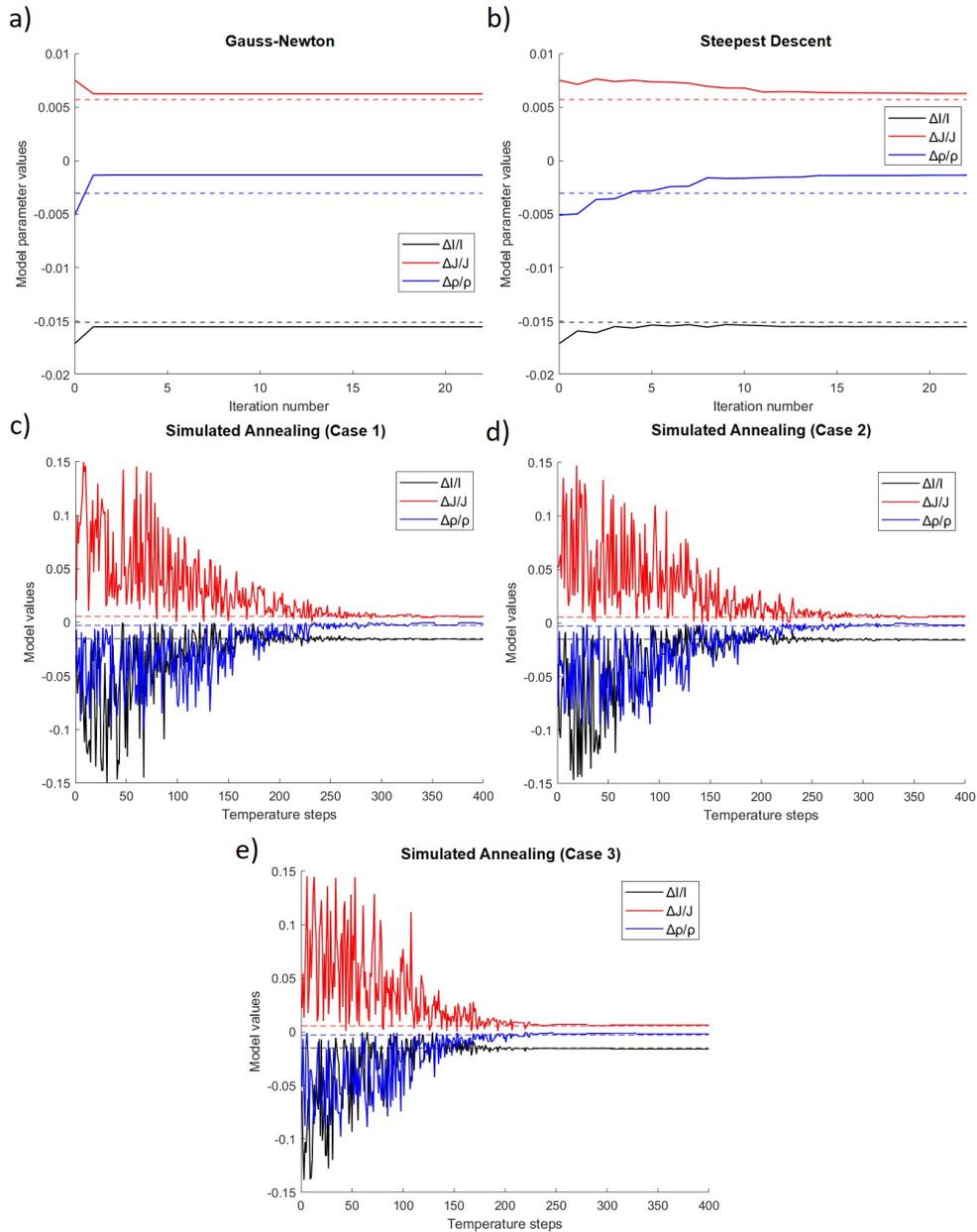


Figure 4.12: Evolution of the estimated model parameters obtained from band-limited data with different optimization methods. Continuous lines represent the estimates and dashed lines the true models.

was slightly different from the results obtained with bounds closer to the true values, as indicated in Figure 4.20. The values of  $\Delta I/I$  were similar between the 3 cases of SA and the local optimization methods, but for  $\Delta J/J$  and  $\Delta \rho/\rho$  the accuracy varied depending on the case. For instance, in case 1, SA produced a value of  $\Delta J/J$  that underperformed the local optimization algorithms, and for

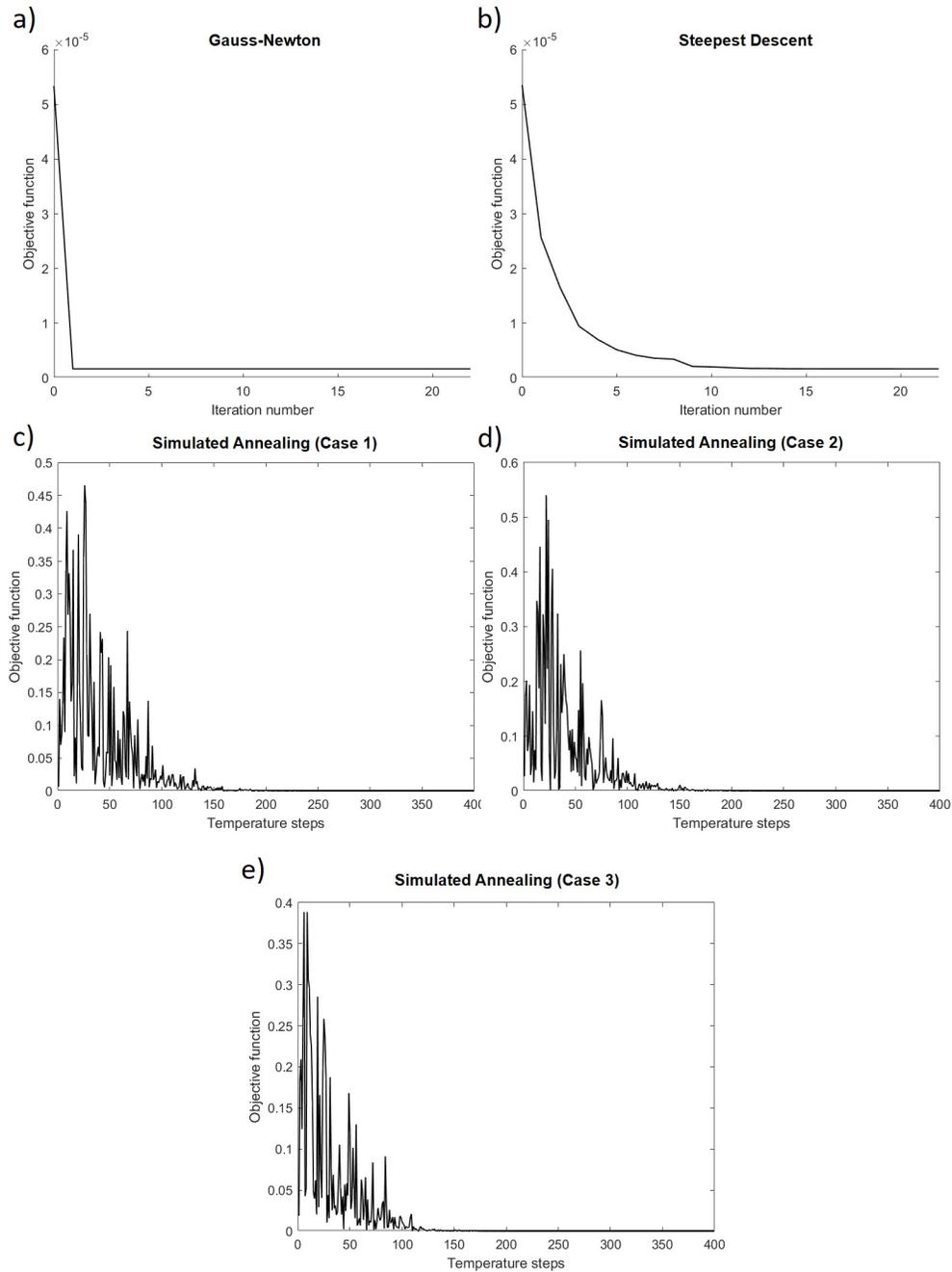


Figure 4.13: Evolution of the objective function produced from inverting band-limited data with different optimization methods.

case 3, the percent error of  $\Delta\rho/\rho$  was higher than the one produced by the Gauss-Newton and the steepest descent technique; for the remaining cases, SA had a better performance than the other optimization methods. On the other hand, in Figure 4.21 it is shown that for the case 3 of the SA,

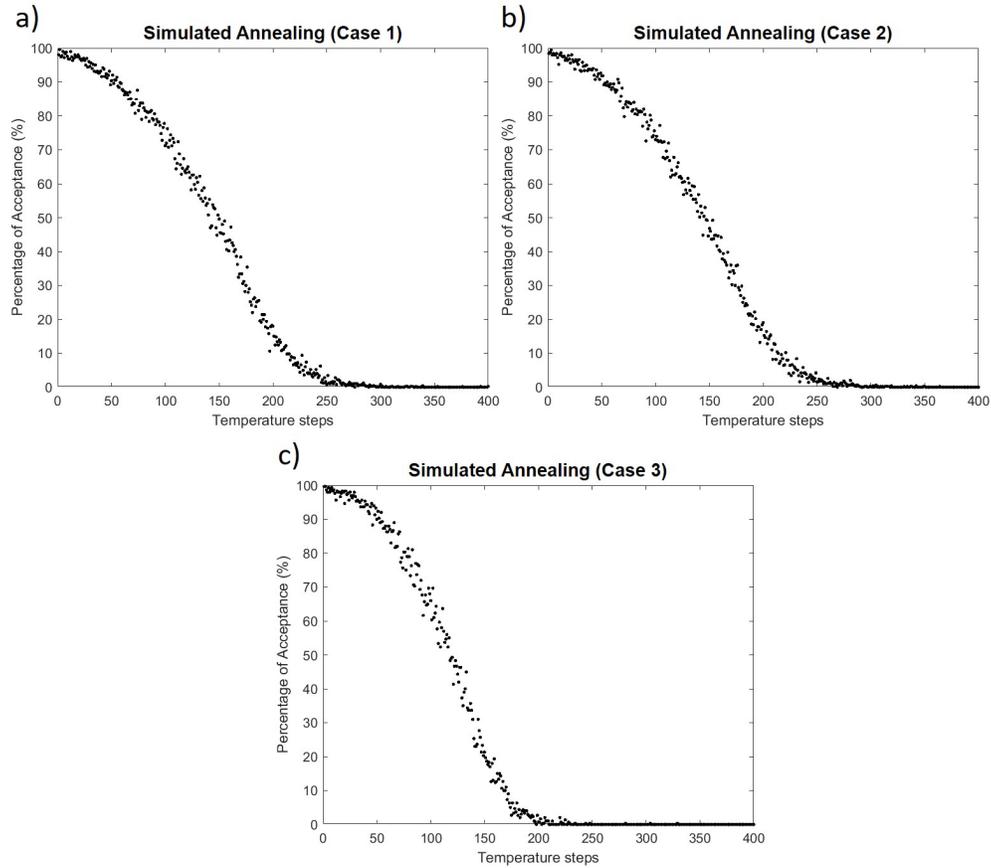


Figure 4.14: Percentage of new configurations accepted per temperature while inverting band-limited data for the three cases of SA.

the estimated value of  $V_S$  for the first layer was almost identical to the true value, but the value of  $V_P$  was not as close. Therefore, by increasing the width of the bounds, there was slightly more difficulty to reach a unified convergence to a point close to the global minimum for all the model parameters, but overall the results were still equivalent or better than the ones produced by local optimization methods.

### 4.5.3 Insights into integrating SA to other geophysical seismic problems

The Simulated Annealing technique has been implemented to solve different types of optimization problems and it could also be suitable in geophysical seismic inverse problems, such as finding the transformation matrices ( $\mathbf{T}$ ) that could produce the desired re-parameterizations of this thesis.

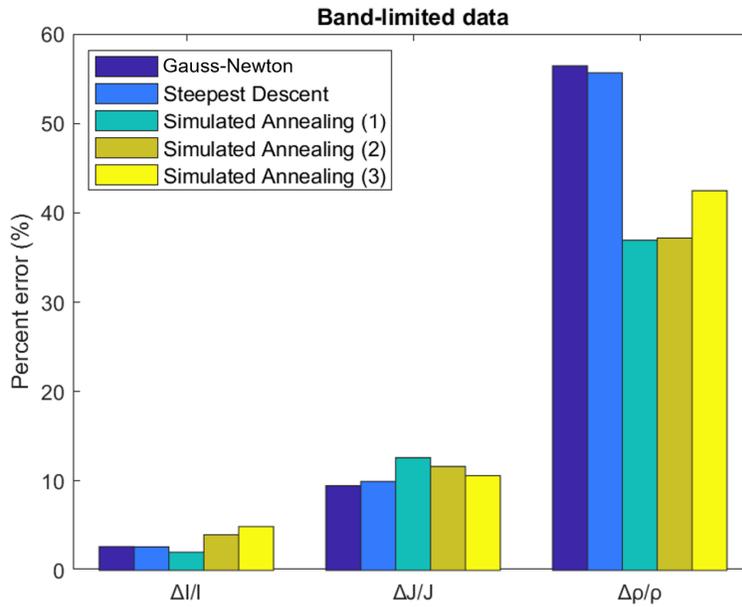


Figure 4.15: Accuracy tests performed with the final estimates obtained from band-limited data using different optimization methods.

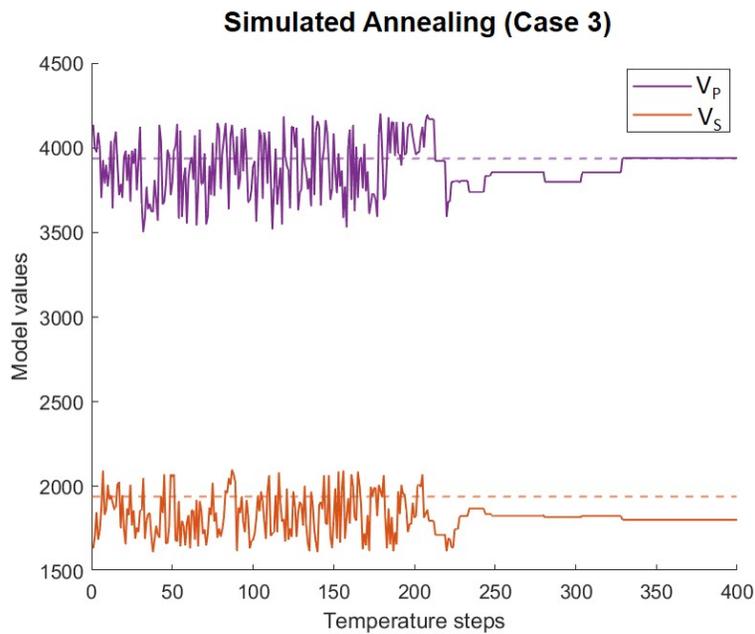


Figure 4.16: Evolution of the estimated  $V_P$  and  $V_S$  values for the first layer obtained from band-limited data with case 3 of SA. Continuous lines represent the estimates and dashed lines the true model values.

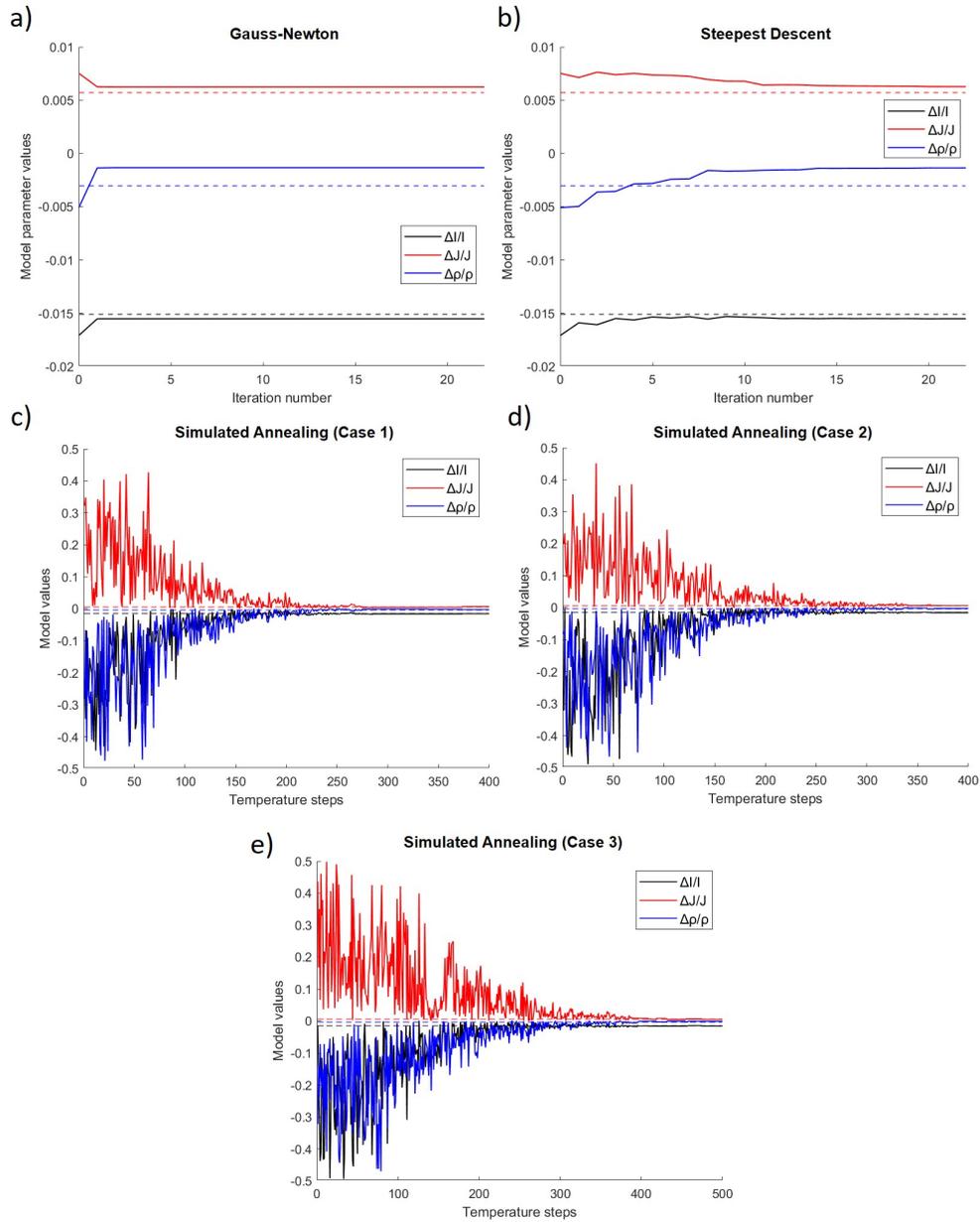


Figure 4.17: Evolution of the estimated model parameters obtained from band-limited data with different optimization methods. In these experiments, the bounds of the SA were larger. Continuous lines represent the estimates and dashed lines the true models.

The experiments of this chapter demonstrated several key takeaways of the method. Firstly, to obtain satisfactory estimations it is essential to define an objective function that properly addresses the problem. Moreover, it is necessary to have a level of knowledge about the values that the model parameters can admit to define the upper and lower bounds in the Metropolis Algorithm; as the

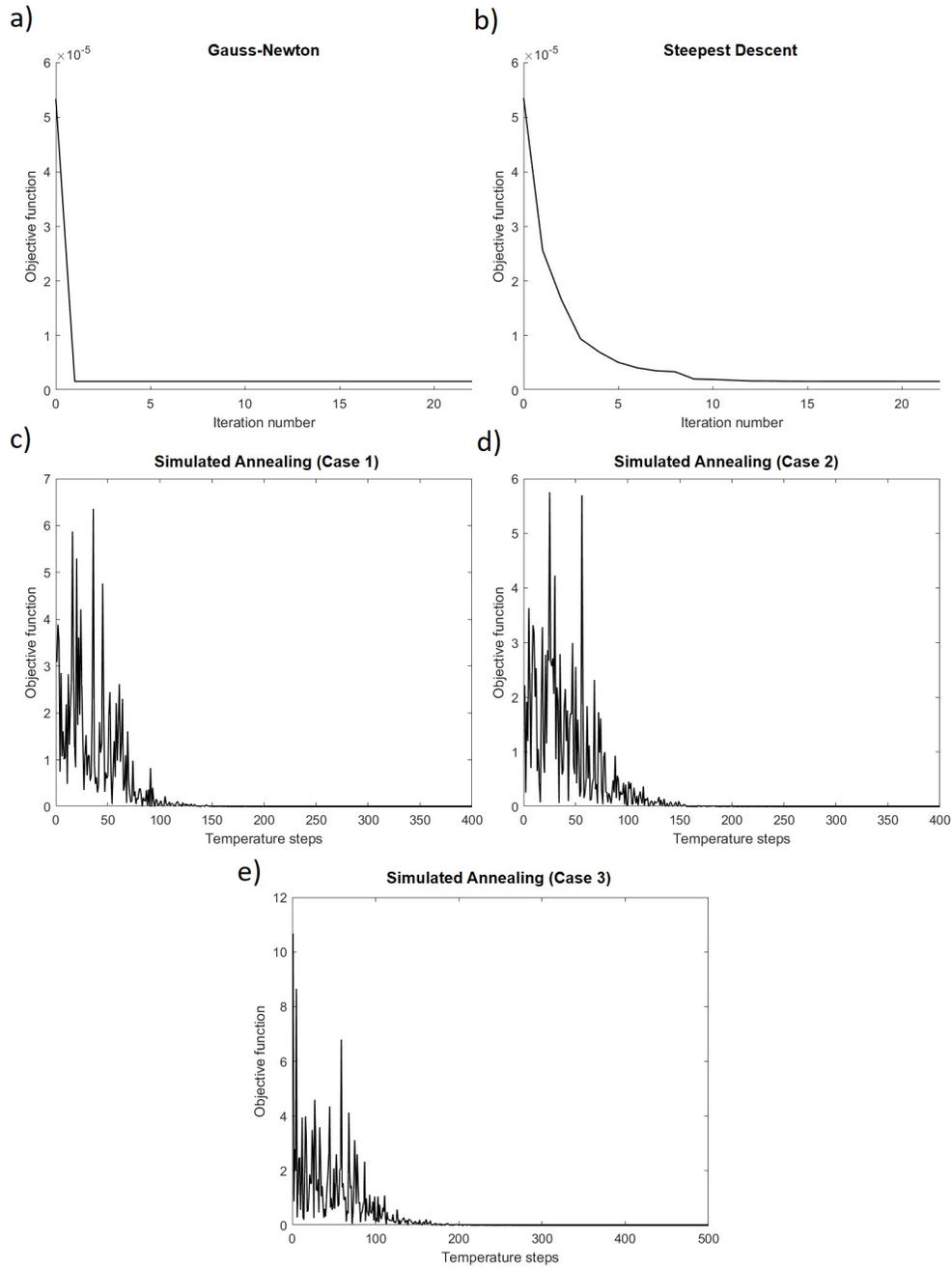


Figure 4.18: Evolution of the objective function produced from inverting band-limited data with different optimization methods. In these experiments, the bounds of the SA were larger.

knowledge about these values gets closer to the true values, it becomes easier for the algorithm to find a more accurate solution. In well-defined inverse problems, another important step for the implementation of SA is the selection of an appropriate annealing schedule or tuning parameters

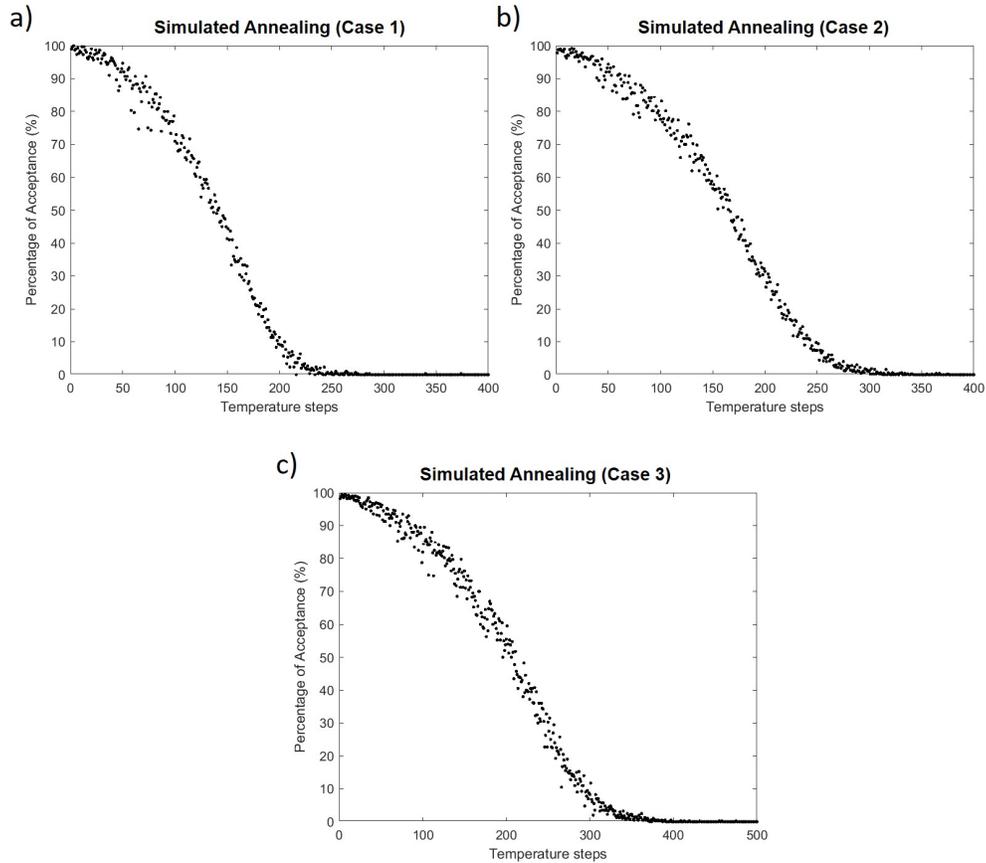


Figure 4.19: Percentage of new configurations accepted per temperature while inverting band-limited data for the three cases of SA and using larger bounds.

to properly minimize the objective function and achieve convergence towards an acceptable minimum point. However, even if the objective function is well-minimized, the model values could be not accurate enough, then it is relevant to define adequate methods to assess the quality of the estimates.

To find the best annealing schedule, several tests or runs of the inversion must be performed until finding a combination of initial temperature, cooling factor, number of perturbations, and number of temperature steps that allow the convergence of the algorithm to an accurate solution. Depending on how difficult the problem is, the execution of these trial and error tests could take a considerable time and the results could be highly affected by small variations of these tuning parameters. Moreover, this technique can produce estimates with a similar or better level of ac-

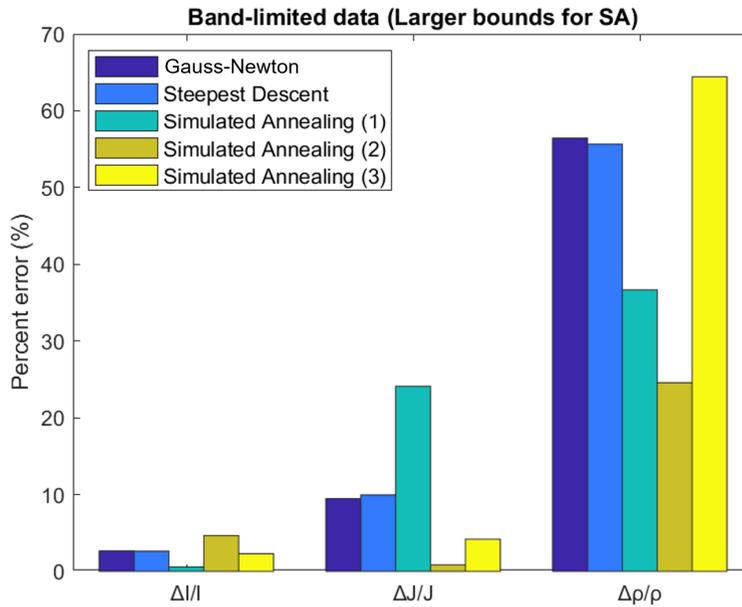


Figure 4.20: Accuracy tests performed with the final estimates obtained from band-limited data with the applied optimization methods. In these experiments, the bounds of the SA were larger.

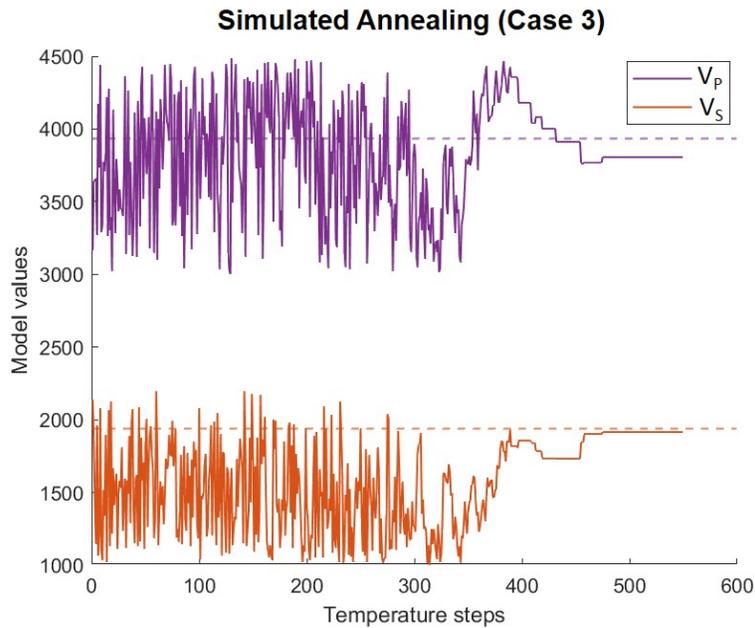


Figure 4.21: Evolution of the estimated  $V_P$  and  $V_S$  values for the first layer obtained from band-limited data with case 3 of SA and using larger bounds. Continuous lines represent the estimates and dashed lines the true models.

curacy than the local optimization methods, such as Gauss-Newton and steepest descent, but at the cost of more iterations or temperature steps and more computational time. However, intrinsic issues of applying local optimization methods for geophysical seismic inverse problems, such as not producing an accurate estimation of the density, could be overcome.

On the other hand, the experiments of inverting band-limited data demonstrated the suitability of the technique in producing accurate estimations of the model parameters when there are missing frequencies, and the tests that were done using random values as starting points (case 3) corresponded to the best practice in setting up this optimization algorithm, since, contrary to the local optimization methods, it was demonstrated that it is not adequate or necessary to introduce any known starting point to find accurate solutions.

## 4.6 Conclusions

The development and testing of an AVO nonlinear inverse problem, with synthetic broadband and band-limited seismic data, served as a starting point to demonstrate the advantages, disadvantages, and the expected performance of the Simulated Annealing strategy, with respect to local optimization methods, when solving a small geophysical nonlinear problem. Overall, although the Simulated Annealing algorithm was slower and took hundreds of iterations to reach convergence, in comparison to the faster performance of the local optimization methods, the estimated model parameters were equivalent or more accurate than those obtained with Gauss-Newton and steepest descent, after finding a good annealing schedule. In this sense, it was illustrated that SA significantly outperformed the results generated by the local optimization algorithms when estimating commonly difficult model parameters, such as  $\Delta\rho/\rho$ ; then, the technique is promising for achieving accurate estimations, since it is capable of overcoming some of the convergence challenges of the local methods. Moreover, it was demonstrated that accurate results could be achieved with SA after starting the iterative process from a random point in the model space, contrary to Gauss-Newton and steepest descent. On the other hand, distinct characteristics were recognized

about the implementation of this optimization technique, such as the requirement of the definition of an appropriate objective function, the execution of several trial and error tests to find the best annealing schedule, the understanding of the values that the solutions can admit, and the selection of appropriate evaluation metrics, to estimate model parameters with high accuracy and avoid instabilities, such as those observed when increasing the bounds to propose new configurations in the Metropolis Algorithm. These learnings were implemented in Chapter 5, by defining a new problem to compute, with the SA strategy, the transformation matrices that would allow the sought mapping between model spaces, while attempting to address the issues of Chapter 3.

# Chapter 5

## Decorrelation of parameter classes and crosstalk minimization with point-probes Hessians

### Summary

Inter-parameter coupled effects could be reduced if the values within the off-diagonal blocks of the Hessian operator were zero, i.e., if the Hessian is an identity matrix or a multiple of it, meaning that there would be no correlation between parameters of different classes. After considering point-wise Hessians in Chapter 3, transformation matrices ( $\mathbf{T}$ ) were searched to re-parameterize the original coordinate system into a new one where the associated Hessian was the identity matrix. However, this matrix structure was only observed at the location selected to compute  $\mathbf{T}$  and thus trade-off effects were introduced to the estimates. These results were the motivation to calculate transformation matrices that could consider parameter crosstalk from more locations and not only at a fixed point in space, aiming to extend the identity matrix structure to more positions. In this sense, point-probes Hessians were used to find the transformation matrices of interest with size  $3 \times 3$ , which would allow the implementation of local-wise re-parameterizations. Given that calculating  $\mathbf{T}$  with the transformation rules of the Hessian is not adequate anymore, the Simulated Annealing technique was implemented as the optimization method to find the elements of  $\mathbf{T}$  that

could convert the point-probes Hessian into a block-wise diagonal matrix in the intermediate model space. Overall, in terms of crosstalk, the estimations did not overcome those shown in Chapter 3 and although some improvements were produced for the density, these results were not better than those from a baseline FWI. Moreover, reduction of crosstalk was achieved in some off-diagonal blocks of the point-probes Hessian, but almost always one pair of blocks was still strongly affected. These drawbacks were related to the challenges experienced by SA to find a unique matrix  $\mathbf{T}$  that properly accounts for the variability of the values within the point-probes Hessian and diagonalizes this matrix, under the considered assumptions and the description of the problem.

## 5.1 Introduction

In Chapter 3, local-wise re-parameterizations were done using transformation matrices of size  $3 \times 3$  that resulted from considering Equation 3.34 with local  $3 \times 3$  Hessians or namely point-wise Hessians that only characterize the crosstalk information between parameters of different classes at a fixed position in the model grid. With this strategy, an intermediate model space was found where the Hessian was the identity matrix only at the location selected to compute  $\mathbf{T}$ , adding misalignments and eccentricities to the objective function in the farthest locations of the model grid, and thus introducing strong crosstalk effects to the final estimates of  $\rho$ ,  $V_P$ , and  $V_S$ . The objective of this chapter was to change the approach and compute transformation matrices of size  $3 \times 3$  related to Hessian operators that could contain crosstalk information from much more locations in the mesh and not only at a fixed position in space, aiming to extend the structure of the identity matrix for the Hessian to more grid cells, and as a result attempt to reduce the crosstalk effects in the final estimates of the original model space. Point-probes Hessians were considered as these are the next computationally feasible operators that could be used, but the numerical approach proposed by Innanen (2020d) was changed for the Simulated Annealing technique since the original method was not suitable anymore to continue mapping between model spaces in a local-wise fashion, and because SA presented interesting advantages regarding other optimization methods. In this sense,

the computation of the elements of the matrix  $\mathbf{T}$  was treated as an optimization problem and, for the implementation of the Simulated Annealing, an appropriate objective function was designed and the learnings gained in Chapter 4 were adopted to frame the problem.

## 5.2 Design of an objective function to be minimized with SA

After considering point-wise Hessians in Chapter 3, the next computationally feasible Hessian operator to use is the point-probes Hessian. By solving Equation 3.34, using a point-probes Hessian, and applying the numerical procedure reported by Innanen (2020d), the resulting matrix  $\mathbf{T}$  would be of the same size as the selected Hessian matrix. In this study, a point-probes Hessian is a  $300 \times 300$  matrix, thus  $\mathbf{T}$  would be of the same size. Under this scenario, applying the transformation rule between the  $s$  and the  $r$  model space would not be a local-wise operation anymore and the transformation process would be ambiguous, since an  $s$  vector of 300 elements should be considered, corresponding to a vertical or horizontal profile across the location selected to compute  $\mathbf{T}$ .

To continue performing transformations between model spaces in a local-wise fashion, as in Chapter 3, a different approach to compute  $\mathbf{T}$  was proposed and developed. The objective of the approach was to transform a point-probes Hessian of the  $s$  system into a matrix of the same size, in the  $r$  system, with zeros in the off-diagonal blocks, i.e., producing a block-wise diagonal matrix, by using a  $3 \times 3$  matrix  $\mathbf{T}$ . Figure 5.1a is a representation of the point-probes Hessian in the  $r$  system and it contains the nomenclature chosen to identify each block. Using the definition of the Hessian-vector product, for each block, the subscript letter (in red) represents the parameter class that was perturbed ( $\Delta r$ ), and the base letter (in black) corresponds to the parameter class selected to evaluate the change in the gradient ( $\Delta g_i(r)$ ) due to the associated perturbation. The 9 elements of  $\mathbf{T}$  would be found by minimizing an objective function that describes the values within the blocks of the point-probes Hessian. Hence, an objective function was designed with 6 penalty terms, encompassing the normalized off-diagonal blocks of the point-probes Hessian in the  $r$  system. The

considered objective function was:

$$\phi = \sum_n^{n_x \times n_z} \left( \frac{r_{2n\Delta r_1}^2}{\sqrt{r_{1n\Delta r_1}^2 \cdot r_{2n\Delta r_2}^2}} \right) + \left( \frac{r_{3n\Delta r_1}^2}{\sqrt{r_{1n\Delta r_1}^2 \cdot r_{3n\Delta r_3}^2}} \right) + \left( \frac{r_{3n\Delta r_2}^2}{\sqrt{r_{2n\Delta r_2}^2 \cdot r_{3n\Delta r_3}^2}} \right) + \left( \frac{r_{1n\Delta r_{2n}}^2}{\sqrt{r_{1n\Delta r_1}^2 \cdot r_{2n\Delta r_2}^2}} \right) + \left( \frac{r_{1n\Delta r_3}^2}{\sqrt{r_{1n\Delta r_1}^2 \cdot r_{3n\Delta r_3}^2}} \right) + \left( \frac{r_{2n\Delta r_3}^2}{\sqrt{r_{2n\Delta r_2}^2 \cdot r_{3n\Delta r_3}^2}} \right) \quad (5.1)$$

where  $n$  is the total number of locations in the model. With the minimization of this objective function, it is expected that, for each location of the mesh, all the point-wise Hessians would be diagonal matrices in the  $r$  system, as illustrated in Figure 5.1b.

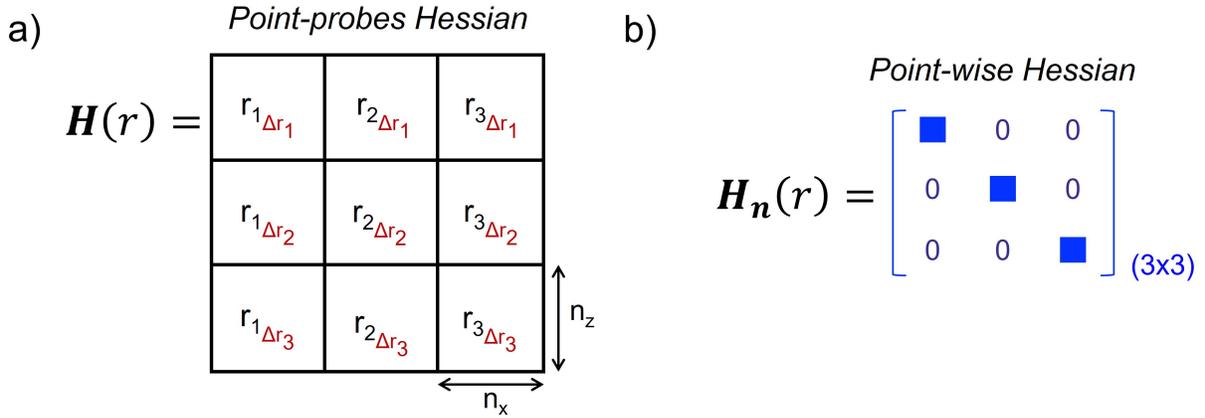


Figure 5.1: a) Representation of the point-probes Hessian, in the  $r$  model space, and the nomenclature used for each block in the objective function. b) Structure of the point-wise Hessians that is expected to be produced.

The construction of the point-probes Hessian, of the  $s$  system, was done through the computation of the Hessian-vector products, as indicated in Chapter 3, hence depending on which parameter class was perturbed, the sensitivities of the change in the gradient due to that perturbation would be different from the sensitivities of the change in the gradient if another parameter class was perturbed, i.e., the order of magnitude could change notoriously between blocks of the point-probes Hessian. Thus, the scale selected to normalize the blocks in the objective function

of Equation 5.1 guarantees the symmetry of the Hessian in the  $r$  model space, as well as allows a controlled minimization of the objective function and thus crosstalk by lowering the values of the off-diagonal blocks concerning those of the main diagonal blocks and removing the variations of sensitivities between blocks.

### 5.2.1 Workflow and evaluation of the objective function

The workflow followed in this chapter was similar to the one in Chapter 3 (Figure 3.4). However, some alterations were made in the methodology to obtain the transformation matrix, as can be seen in Figure 5.2, i.e., per frequency band of the FWI, the same grid cell was chosen to compute a point-probes Hessian in the  $s$  system and this operator was used inside the SA algorithm to calculate the associated  $3 \times 3$  matrix  $\mathbf{T}$  that transforms it into a block-wise diagonal matrix in the  $r$  model space. Hence, per frequency band, the objective function of Equation 5.1 was iteratively evaluated using the blocks of point-probes Hessians, in the  $r$  model space, which were obtained with proposals of  $3 \times 3$  matrices  $\mathbf{T}$  and the point-probes Hessian of the  $s$  model space.

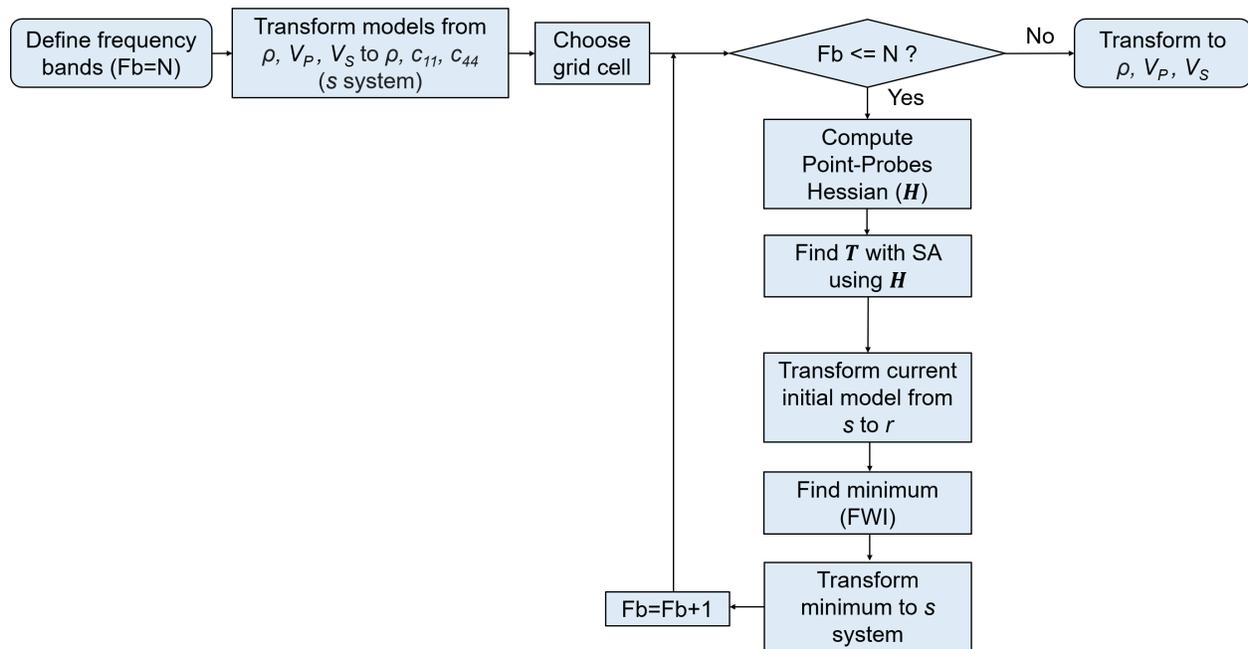


Figure 5.2: Re-parameterization workflow followed in this study.

Within the Simulated Annealing algorithm, for a particular temperature and perturbation, the evaluation of the objective function in the Metropolis Algorithm of Figure 4.2 was implemented as follows: (1) the point-probes Hessian matrix, of the  $s$  model space, with size  $300 \times 300$  was organized into 10.000 local  $3 \times 3$  Hessian matrices or point-wise Hessians, (2) the forward problem of Equation 3.32 was computed using each of the local Hessians from the previous step and a unique configuration proposed for the matrix  $\mathbf{T}$ , (3) the new local  $3 \times 3$  Hessians in the  $r$  model space were re-organized to form the corresponding point-probes Hessian of size  $300 \times 300$ , and finally (4) Equation 5.1 was assessed to verify if the proposal for  $\mathbf{T}$  should be accepted or rejected and continue with the next iteration of the SA process.

On the other hand, a sequential SA could converge to more accurate solutions, but it implies an extensive computational time that is not convenient when working with FWI problems. In this case, given that the computation of the forward problem is local-wise, applying a sequential SA would require 10.000 calculations per model parameter individually proposed and thus 90.000 operations per perturbation, since the matrix  $\mathbf{T}$  is formed by 9 elements, and this series of computations would be repeated per each temperature of the annealing process. Instead, as explained in Chapter 4, to achieve a trade-off between the computational time and the quality of the solution, the model configurations were proposed based on perturbing all the model parameters at the same time.

### **5.3 Pros and cons of SA for the current problem**

Different optimization algorithms could minimize the objective function described in Equation 5.1; however, the most straightforward methods are the grid search strategy, local optimization methods, and the Simulated Annealing technique, solving the nonlinear problems with different approaches. The grid search method can find the minimizer of an objective function by exhaustively examining a large set of user-specified trial solutions for each model parameter. In this sense, due to the dimensionality of the problem and the imprecise understanding of the potential range of values within which the solutions may exist, this may not be the most optimal strategy. Further-

more, using local optimization algorithms require (1) providing a good enough starting model, (2) solving derivatives of the objective function with respect to the model parameters to compute the gradient, the step size, and, depending on the method, the Hessian to reach an accurate solution, and (3) working with more complicated and error-prone algorithms.

Conversely, Simulated Annealing presents a set of advantageous characteristics that make it an appealing choice for computing the matrix  $\mathbf{T}$ . This technique is distinguished by its efficacy in tackling highly nonlinear problems, as demonstrated in the previous chapter with an AVO exercise, and given that it corresponds to an approximation of a global optimization approach, it is independent of the choice of the starting values for the model parameters; similarly, if the initial model influences the solution, then the tuning parameters of the algorithm must be changed to produce a more appropriate search (Smith and Wong, 2017). Hence, this method should be capable of producing acceptable solutions for the 9 elements of  $\mathbf{T}$  without providing a good enough initial model, if an appropriate objective function and annealing schedule are designed as well as adequate bounds are chosen for the proposed model configurations. In addition, the Simulated Annealing algorithm is more compact and easier to implement than the local optimization methods, not requiring the calculations of derivatives to define a search direction for the minimum point, but only the assessment of the objective function per iteration or temperature step.

Nevertheless, the SA method could present certain limitations. The transformation matrix is a mathematical tool, not a physical quantity, such as the velocities or densities of the rocks, consequently, the knowledge regarding the possible values that its elements can adopt is limited, affecting the selection of appropriate and narrow bounds. For instance, computing the matrix  $\mathbf{T}$  associated with different and spatially separated local  $3 \times 3$  Hessians by using the numerical procedure of Chapter 3, can offer insights into the plausible values of  $\mathbf{T}$ ; however, the uncertainty would persist, as a significant number of locations beyond the ones studied in the model grid would remain unexplored. However, considering the empirical basis of the Simulated Annealing strategy, it is still feasible to determine the optimal values for these bounds; the plots of the evolution of the model parameters and the objective function as the temperature cools down provide hints on how

well the algorithm is performing under the chosen bounds and the selected annealing schedule, as well as what changes need to be done to achieve better convergence.

Nonetheless, the main drawback of the method is related to the prolonged execution times. While using SA, improvements of the solution are done iteratively and the algorithm requires several temperatures, as well as multiple proposed perturbations at each temperature in order good results could be achieved; moreover, several trial and error tests need to be done to find the best annealing schedule, requiring additional time. Hence, implementing the SA technique to solve the matrix  $\mathbf{T}$  per frequency band, would considerably lengthen the computational time required by the FWI, taking more time than the inversions performed in Chapter 3. Further downsides of the method are the large number of parameters that must be tuned in the annealing schedule to find an acceptable solution; throughout the trial and error tests, diligent adjustment of the tuning parameters is necessary to understand the source of the changes in the convergence process and recognize that the precision of the chosen values affects the quality of the solution. Additionally, this method is probabilistic and generates different solutions per run of the algorithm, thus it is necessary to use “seeds” to initialize the generator of random values and produce the same answers, allowing a more cautious study of the tuning parameters that could be affecting the convergence.

An additional disadvantage pertains to the uncertainty surrounding whether any annealing schedule could yield an optimal solution for the optimization problem, as well as the ease of annealing or even accepting annealing solutions. For problems where the configuration space is not characterized by smooth and gradual variations between multidimensional hills and valleys, but narrow and deep hollows (gopher holes) are present, jumping out of the local minimum points and thus finding the global solution is a challenging task. In some of these scenarios, it is possible to converge to an optimal point, but the design of the objective function demands significant insight and judgment, or suitable constraints need to be applied to the energy function or the generation of neighboring states. Nevertheless, in some instances, the problem cannot be annealed due to an excessively irregular configuration landscape or impractical time requirements to achieve a satisfactory solution (Rutenbar, 1989).

## 5.4 Numerical experiments

To maintain consistency and facilitate fair comparisons between the current and previous experiments, the same sources and receivers of Chapter 3 were used in this study. For the FWI, the same multiscale approach, initial models, optimization method, and iterations were applied. Table 5.1 displays the annealing schedules implemented at each frequency band of the FWI. Overall, the starting temperatures were much higher than those used for the AVO problem of Chapter 4, because these values are directly related to the order of magnitude of the objective function after being evaluated with the current model parameters. Additionally, for each frequency band, the objective function was difficult to minimize, since large values of the cooling factor “ $e$ ” (Equation 4.17) were necessary to generate the solutions, contrary to the values that were found for the experiments of Chapter 4. This means that the SA algorithm required a remarkably slow cooling schedule and hence a vast number of temperature steps. Moreover, several perturbations were proposed per temperature in the annealing process, varying depending on the frequency band, which along with the number of temperature steps were translated into larger execution times that were introduced to the FWI process.

Frequency band	Initial temperature	$e$	Number of temperature steps	Number of perturbations
1	$4 \times 10^6$	0.99	1300	200
2	$6 \times 10^6$	0.993	2000	350
3	$4 \times 10^6$	0.993	1900	350
4	$4 \times 10^6$	0.993	2000	350
5	$4 \times 10^6$	0.992	1600	200
6	$3 \times 10^6$	0.989	1300	700
7	$3 \times 10^6$	0.989	1300	700
8	$2.5 \times 10^6$	0.988	1200	600

Table 5.1: Annealing schedules applied to the SA algorithm at each frequency band of the FWI.

Figure 5.3d contains the estimates obtained from the FWI workflow proposed in this study.

Stronger coupled effects are observed around all the heterogeneities, and especially around  $V_P$ , in comparison to the baseline results (Figure 5.3b) and the estimates obtained when re-parameterizing with the point-wise Hessian (Figure 5.3c). Moreover, there was a slight reduction of crosstalk effects in the results of  $\rho$  in comparison to the estimates in Figure 5.3c, but these results did not surpass those obtained with the baseline FWI. Conversely, in Figure 5.4 it was shown the overlapped vertical and horizontal profiles extracted from the estimates of  $\rho$ ,  $V_P$ , and  $V_S$ . It is observed that re-parameterizing the FWI problem with the point-wise Hessian or with the point-probes Hessian produced estimations of the anomaly values that were slightly more accurate than the estimates produced by the baseline FWI, especially in the case of  $\rho$  and  $V_S$ .

On the other hand, the performance of the Simulated Annealing was examined with Figure 5.5, for some of the frequency bands of the FWI. Overall, the curves of the evolution of the objective function and model values as the temperature decreases indicated an adequate behavior of the algorithm. Nevertheless, the curves of the percentage of new configurations acceptance had a different shape from those observed in Figures 4.8, 4.14, and 4.19 of Chapter 4, showing a more inconsistent pattern of acceptance and rejection of new proposals, especially at intermediate temperatures, i.e., the curves tended to spread for these temperatures, meaning that for a particular temperature, a large number of configurations were rejected, but for the next ones much more configurations were accepted, instead of showing a smoother pattern of rejection of new proposals as indicated in the literature.

Furthermore, although the objective function was effectively minimized at each frequency band and thus the convergence of the Simulated Annealing was achieved, and despite several efforts were made to find a suitable annealing schedule, the estimated elements of  $\mathbf{T}$  did not produce the expected diagonal point-probes Hessian in the  $r$  model space, as illustrated in Figure 5.6. The  $3 \times 3$  matrices shown in this figure are the summarized versions of the point-probes Hessians in the  $s$  and in the  $r$  model space. These summaries were obtained after calculating the sum of the absolute values of the elements in all locations within each block of the point-probes Hessian; subsequently, the resulting  $3 \times 3$  matrix was normalized or scaled, dividing each element by its

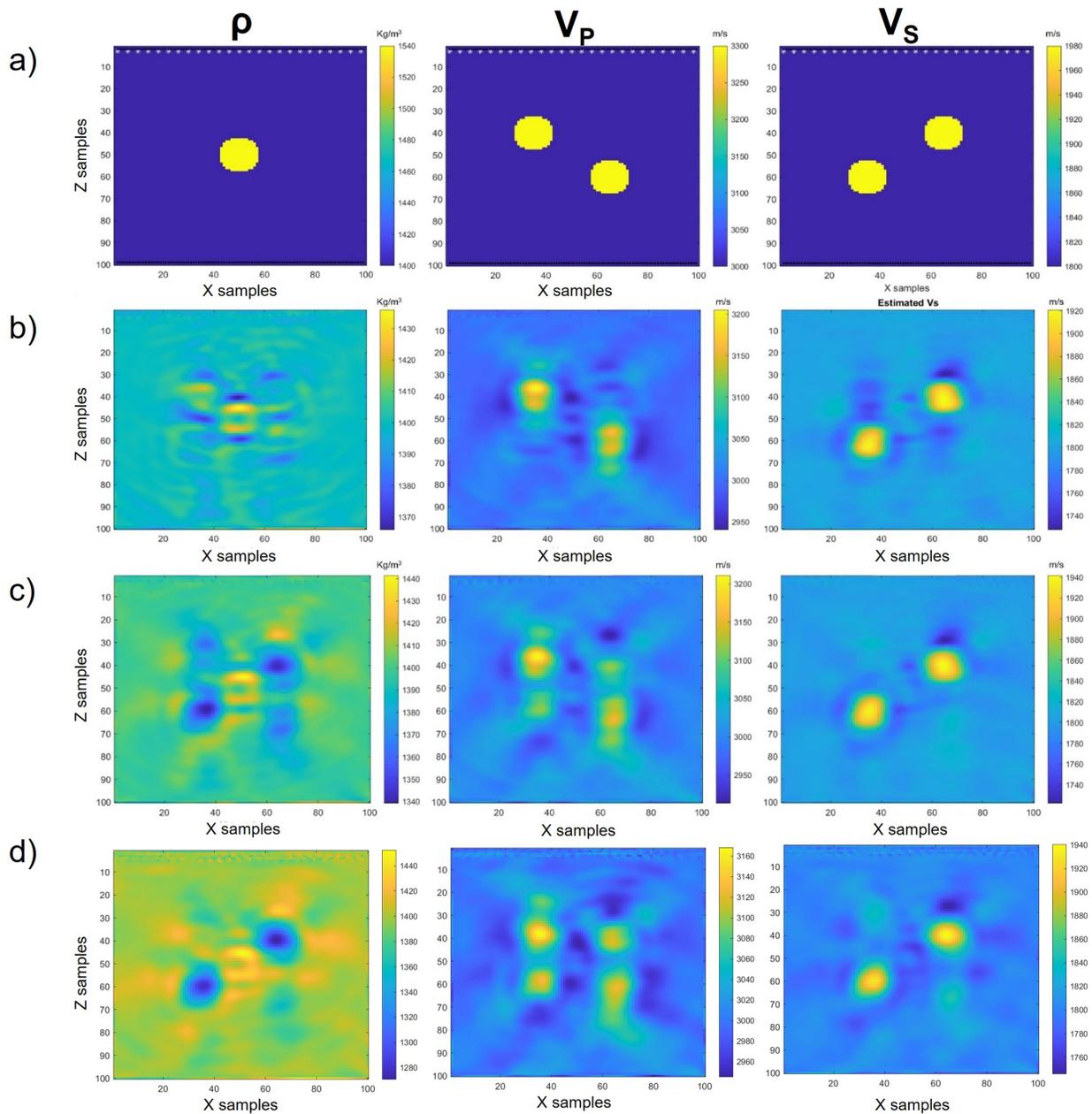


Figure 5.3: a) True  $\rho$ ,  $V_P$ , and  $V_S$  models. b) Models estimated with a baseline FWI, i.e., without re-parameterization. c) Models estimated with a re-parameterized FWI, using the point-wise Hessian, after selecting location  $x=50$  and  $z=20$  to compute  $\mathbf{T}$ . d) Models estimated with a re-parameterized FWI, using the point-probes Hessian, after selecting location  $x=50$  and  $z=20$  to compute  $\mathbf{T}$ .

maximum value. The normalization applied to these summaries was different from the one used in the crosstalk metric of Chapter 3 since the symmetry in the Hessian was already imposed with the normalization of the objective function of Equation 5.1. In this sense, a  $3 \times 3$  matrix with zeros in

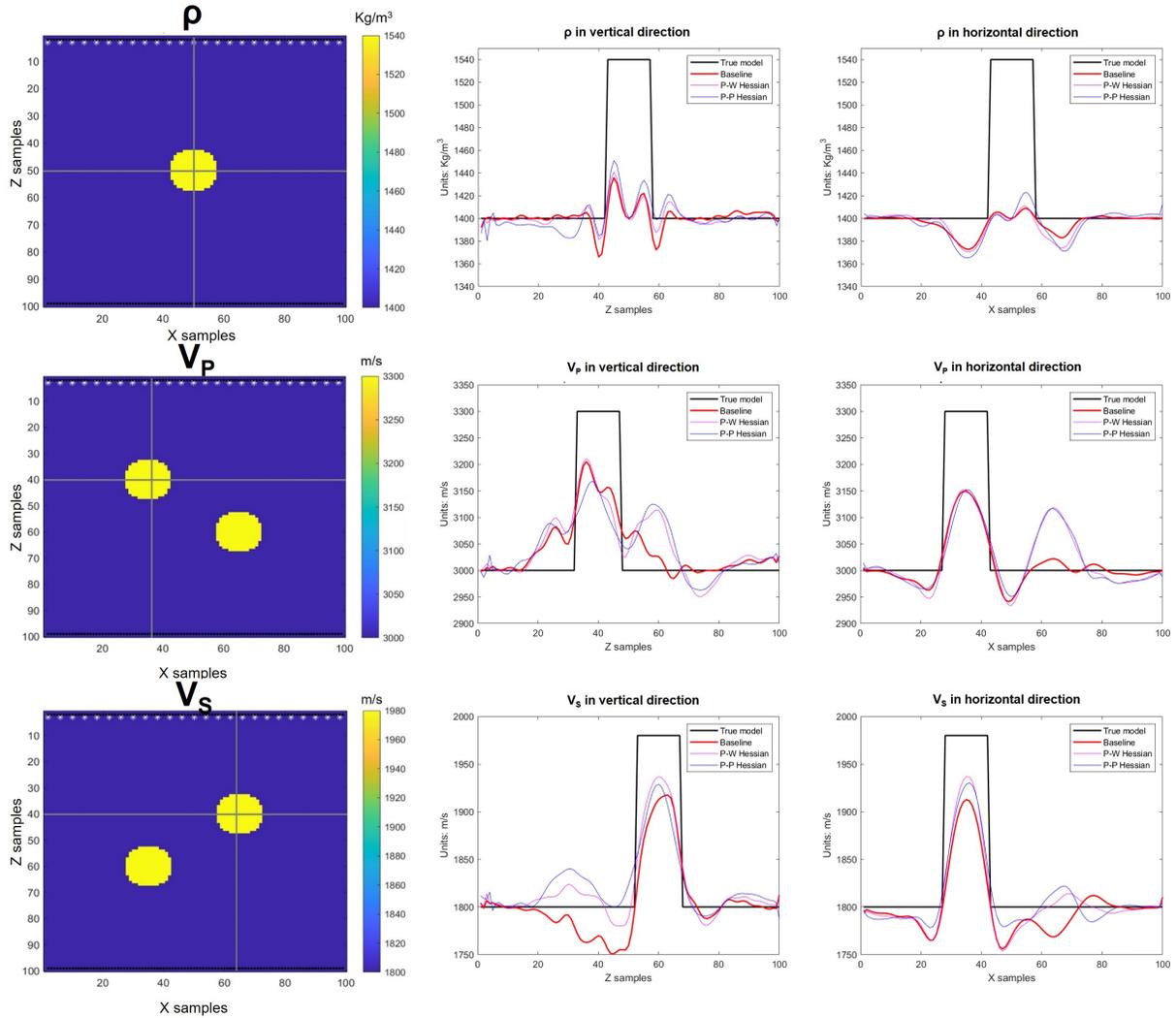


Figure 5.4: Vertical and horizontal profiles extracted from the true and estimated models in Figure 5.3. The grey lines on the true models correspond to the selected profiles.

its off-diagonal elements should be obtained, if the crosstalk effects were minimized. Overall, there were noticeable coupled effects in at least one pair of off-diagonal blocks, whereas the remaining pairs exhibited relatively lower levels of crosstalk; however, these lower values were not negligible enough to be considered close to zero, which explains the introduction of notorious coupled effects to the final estimates. Additionally, for some of the frequency bands, among the main diagonal blocks, at least one exhibited a significantly low value, which could potentially introduce additional complications during the inversion because the gradient related to the involved model parameters

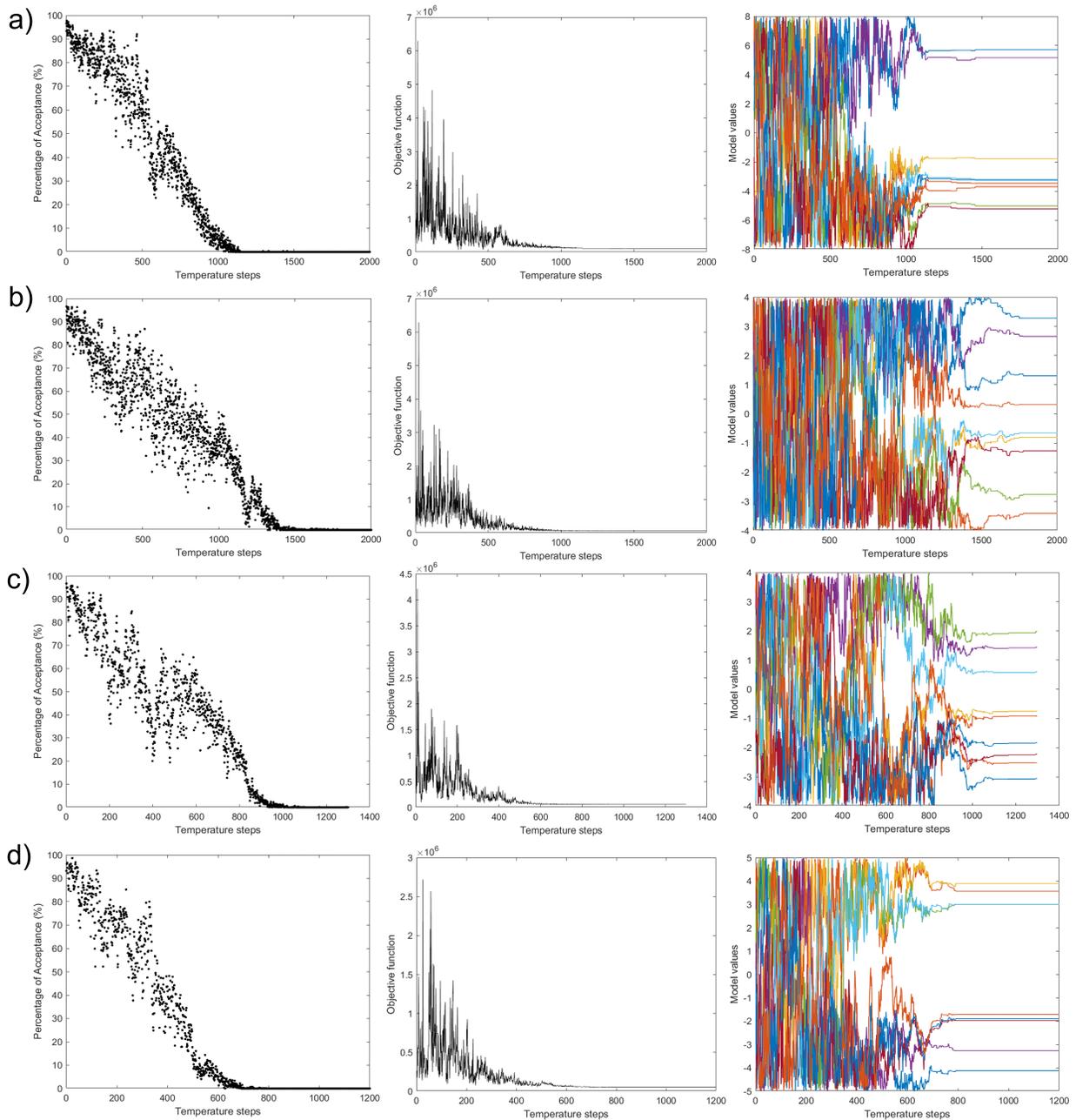


Figure 5.5: Curves to evaluate the performance of SA at some frequency bands of the FWI. a) Frequency band 2, b) frequency band 4, c) frequency band 6, and d) frequency band 8.

would not be properly corrected from geometric spreading and band-limited effects.

Additionally, although the diagonalization of the Hessian operator in the  $r$  model space was not achieved, one favorable aspect of using the Simulated Annealing was that the structure in the summary of the Hessians was closely reproduced for different point-wise Hessians in the model

grid, as shown in Figure 5.7. The local  $3 \times 3$  Hessians of this figure were normalized by computing the absolute value of the 9 elements of the matrix and dividing them by the maximum value of the matrix. These results suggest that if more accurate transformation matrices are found to produce a diagonal matrix structure for the Hessian in the  $r$  model space, a heightened level of consistency could be attained across all locations of the considered point-probes Hessian.

### 5.4.1 Analysis

Several factors could explain the inability of the SA technique to generate the expected local-wise diagonal Hessians in the  $r$  model space:

1. Given that the true values of the matrix  $\mathbf{T}$  are unknown, the approach used to evaluate the quality of the estimates of  $\mathbf{T}$  relied on the performance of the forward problem. The minimization of the objective function does not necessarily guarantee the attainment of an accurate solution; consequently, further trial and error testing must be conducted to discover if a more suitable annealing schedule could lead to a more satisfactory result. Hence, there is a possibility that the obtained estimates may still lack the necessary accuracy to generate the diagonal matrix, and more improvements might need to be done to the annealing schedule.
2. The Simulated Annealing strategy is highly dependent on the objective function designed to describe the problem. In the experiments of this chapter, its minimization was produced as the temperature decreased, but the expected solution was not obtained. Therefore, it is possible that the objective function of Equation 5.1 may demand adjustment in a more resourceful way to overcome the obstacles found and achieve convergence towards an optimum solution.
3. The assumptions made regarding the matrix  $\mathbf{T}$  could have resulted in the generation of a highly irregular landscape for the model configuration. In this case, a unique transformation matrix of size  $3 \times 3$  was used to convert each point-wise Hessian of the  $s$  model space into a diagonal matrix. However, not all local Hessians, in the  $s$  system, have similar values, sensitivities, and organization, as illustrated in Figure 5.8. Hence, there is a strong likelihood

that using only one matrix  $\mathbf{T}$  would not allow to correctly solve the problem, since not enough independent equations are provided to adapt to the spatial variability of the local Hessians of the  $s$  model space and thus to map them into a diagonal matrix in the  $r$  model space. To bring more independent equations and capture all the value fluctuations of the point-probes Hessian, a larger matrix  $\mathbf{T}$  may be required, changing the current type of re-parameterization, i.e., local-wise, or the computation of more than one  $3 \times 3$  transformation matrix need to be introduced in the workflow. On the other hand, the minimization of the objective function and, simultaneously, the generation of imprecise solutions could also arise if the reason behind the incapacity of producing the diagonalization is related to the use of a unique matrix  $\mathbf{T}$ , reinforcing this possible explanation.

## 5.5 Conclusions

Using the Simulated Annealing optimization method to find a transformation matrix that (1) could consider crosstalk information from more than one location of the model grid and (2) have the adequate size to allow local-wise FWI re-parameterizations, demonstrated that it was challenging to encounter a unique matrix  $\mathbf{T}$  of size  $3 \times 3$  that transforms a point-probes Hessian from the  $s$  system into a block-wise diagonal matrix in the  $r$  model space. The attempts done to diagonalize the studied point-probes Hessians were unsuccessful, possibly due to several reasons, such as the difficulties that are intrinsic to this technique, the requirements of more independent equations to characterize the variability in all the local Hessians that form the point-probes Hessian in the  $s$  model space, and the selection of an objective function that, although was successfully minimized, did not produce the expected solution. As a result, the most successful outcome produced by implementing the SA technique was the reduction of crosstalk effects in some off-diagonal blocks, but the algorithm consistently introduced strong effects on at least one pair of blocks, generating estimates with more noticeable inter-parameter coupled effects than those exhibited by the results from the baseline FWI and from the re-parameterizations performed with point-wise Hessians. The

results of this chapter suggested that, for future work, deeper insight and judgment are required in the design of the objective function, the evaluation method for the final estimates, and the selection of the number of matrices  $\mathbf{T}$  and/or the size of the considered matrix  $\mathbf{T}$  to bring more independent equations and capture the variability of the local Hessians in space, in order the obstacles presented in this chapter could be overcome and the diagonalization could be achieved.

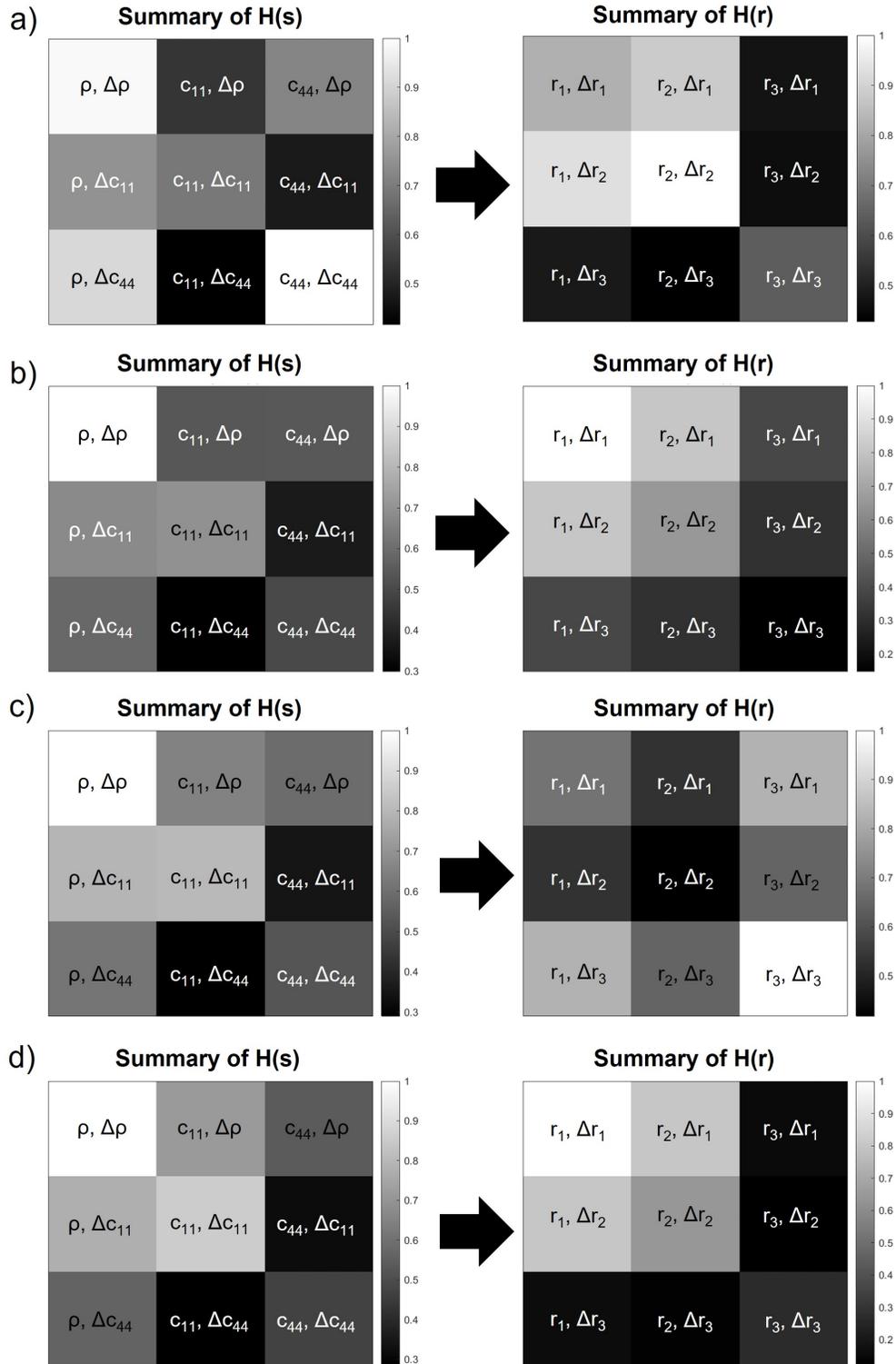


Figure 5.6: Summarized versions of the point-probes Hessians in the  $s$  and the  $r$  model spaces at different frequency bands of the FWI process. a) Frequency band 2, b) frequency band 4, c) frequency band 6, and d) frequency band 8.

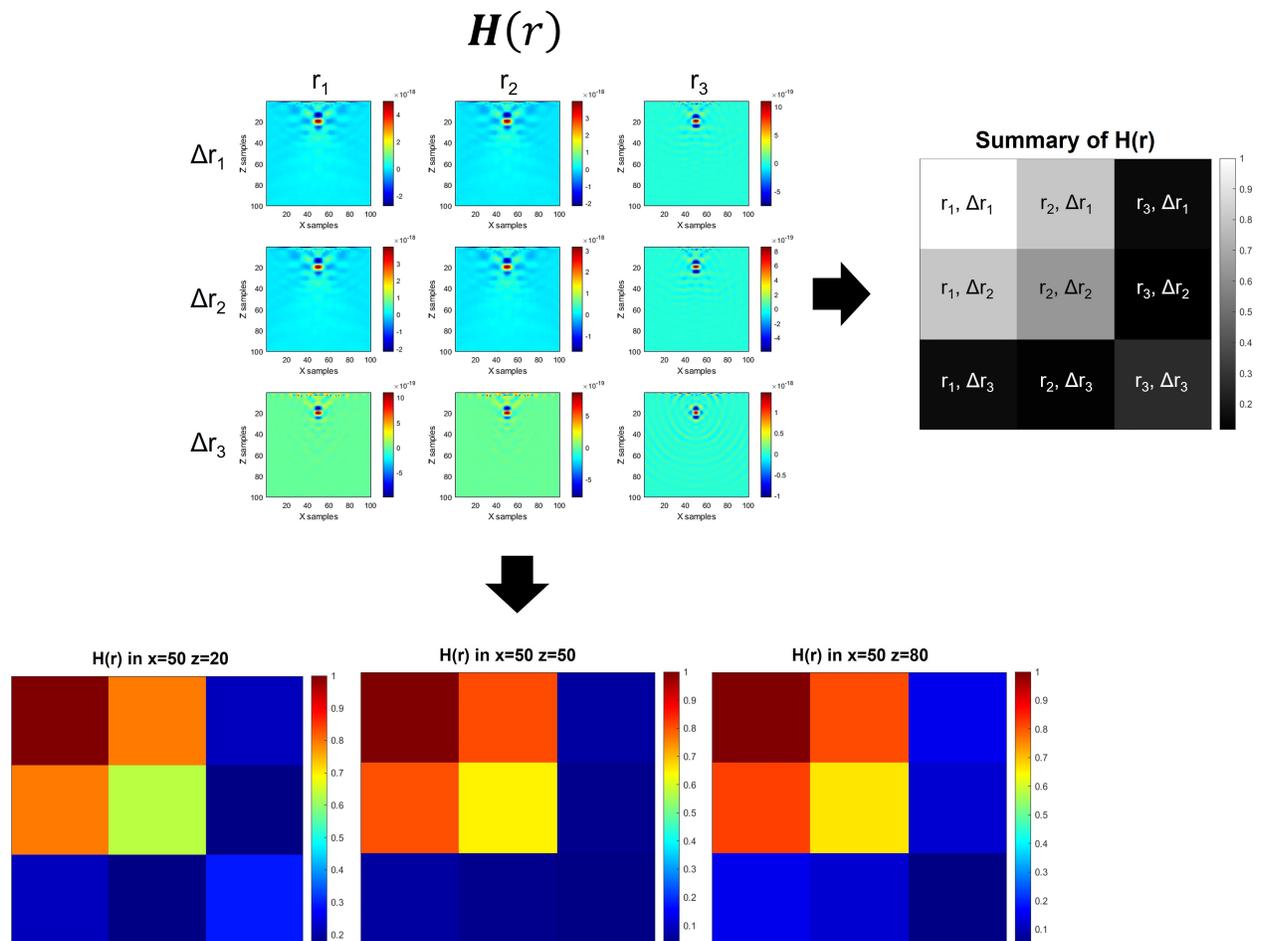


Figure 5.7: Summary of the point-probes Hessian obtained in the  $r$  model space and at the frequency band 8, as well as different extracted point-wise Hessians.

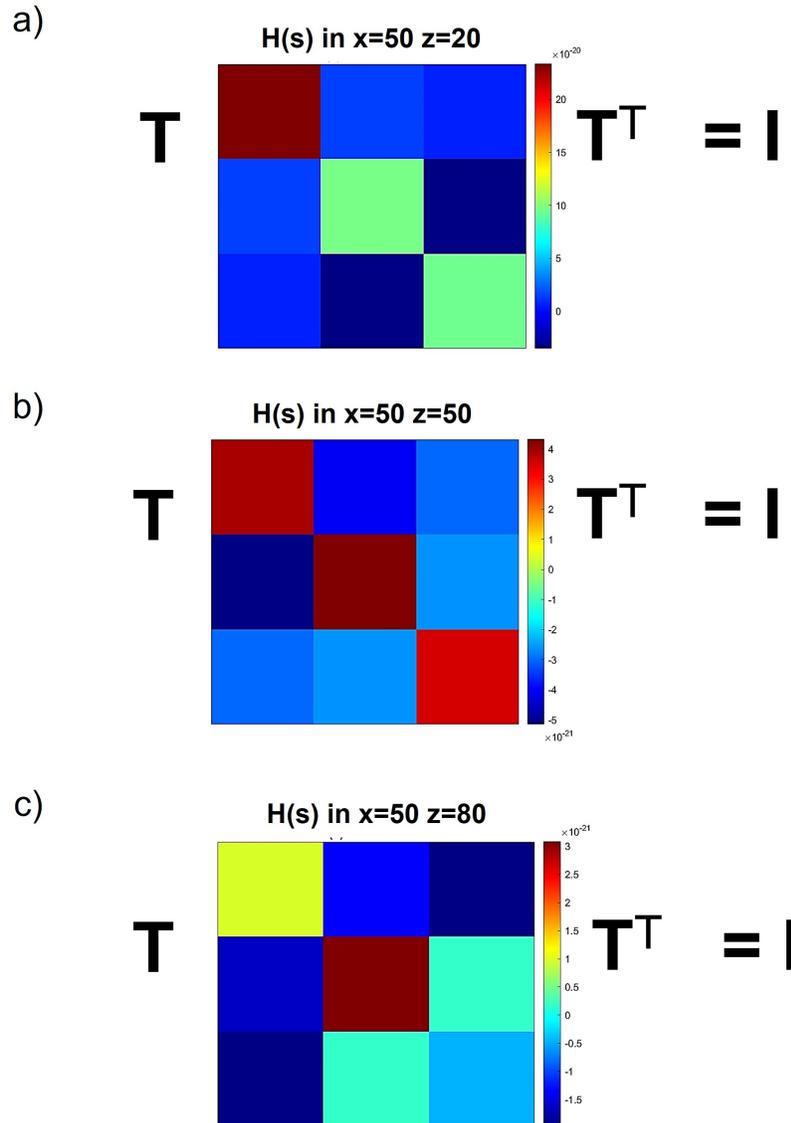


Figure 5.8: Representation of the transformation of different point-wise Hessians into the  $r$  model space using a unique  $3 \times 3$  matrix  $\mathbf{T}$  for the frequency band 8. The values of the local Hessians are not normalized.

# Chapter 6

## Conclusions

### 6.1 Summary

Full waveform inversion has been successful in estimating the physical properties of the subsurface, but it is commonly affected by difficulties that are partly related to the nature of the local optimization algorithms, which are applied to estimate the model parameters in the selected wave propagation model. In multiparameter FWI, one of these challenges is the introduction of inter-parameter coupled effects or crosstalk to the final estimates, affecting the inversion by slowing the convergence of the algorithms and producing poorly accurate results that add more uncertainty. Increasing confidence in the results is essential; thus, research has been done in the past to mitigate these effects by analyzing the radiation patterns of different model spaces, selecting appropriate parameterizations and optimization approaches, or employing other different methodologies. This thesis was focused on developing and testing strategies that could contribute to the treatment of these artifacts. In particular, two methodologies were proposed to reduce the crosstalk effects from FWI experiments with seismic surface acquisitions and in the  $\rho$ ,  $V_P$ , and  $V_S$  model space, by working with constrained re-parameterizations, which rely on the decorrelation of parameter classes linked to a convenient structure of the associated Hessian operators. This section contains a summary of the experimental breakthroughs in the chapters of this thesis.

Before beginning with the development of the re-parameterization strategies to mitigate crosstalk, the studies initiated in Chapter 2 with the examination of the coupled effects that would

arise in the  $\rho$ ,  $V_P$ , and  $V_S$  model space, considering idealized (homogeneous) and more realistic (heterogeneous) backgrounds. Empirical radiation patterns under heterogeneous reference media were extracted and it was understood that their shape was close to the indicated by the analytic expressions, which have been developed under homogeneous backgrounds, but in some cases, the lobes rotated, the amplitudes decreased, the symmetry was lost, and irregularities occurred. Moreover, in seismic surface experiments, the potential introduction of crosstalk was demonstrated most generally between  $V_P$  and  $\rho$  as well as  $V_S$  and  $\rho$ , and with much less proportion between  $V_P$  and  $V_S$ . However, in heterogeneous backgrounds, these coupled effects slightly changed regarding what was indicated by the analytic expressions, exhibiting increase, reduction, or introduction of new regions of overlapping.

In Chapter 3, efforts were done to mitigate these effects from the FWI results, since at each frequency band, the original system was attempted to be re-parameterized into an intermediate model space constrained by Hessian operators with the identity matrix structure, which would indicate that the iso-surfaces of the objective function are spherically symmetric and no correlation between parameters of different classes exist. These experiments were done with a local Hessian of size  $3 \times 3$ , namely point-wise Hessian, to compute a transformation matrix of the same size that allowed the performance of local-wise re-parameterizations. The point-wise Hessian is a small matrix extracted from a bigger matrix named point-probes Hessian, which is constructed by implementing the Hessian-vector product definition. It was shown through the numerical experiments that the proposed workflow was successful in finding an intermediate model space where the associated Hessian was the identity matrix, but only at the location selected to construct the transformation matrix, losing this sought matrix structure towards the farthest locations. As a result, although a relatively accurate estimation was done for the  $V_S$  parameter class, the estimations of  $V_P$  and  $\rho$  exhibited stronger crosstalk effects than the results from a baseline FWI.

These outcomes suggested that the decorrelation ideas were promising, but Hessians containing crosstalk information from multiple locations, rather than solely at a fixed position on the mesh, would be necessary to compute a transformation matrix that could produce a more consistent iden-

tivity matrix structure of the Hessian in space. In this sense, subsequent experiments were conducted using point-probes Hessians, but further exploration of an alternative numerical approach was required, since the one used in Chapter 3 would lead to a non  $3 \times 3$  transformation matrix, limiting the application of local-wise re-parameterizations. As a result, a different optimization problem was designed to find a transformation matrix of size  $3 \times 3$  that could transform a point-probes Hessian into a block-wise diagonal matrix, aimed to be solved with the Simulated Annealing technique.

In this sense, in Chapter 4, the Simulated Annealing strategy was tested to solve an AVO non-linear problem, which was smaller and simpler than the problem proposed to solve the challenges of Chapter 3, to gain valuable insights about the proper implementation of the method, as well as its advantages and disadvantages. The experiments demonstrated the importance of defining a good enough objective function and finding the most appropriate annealing schedule or tuning parameters to reach convergence towards a minimum point close to the global solution. It was also evidenced that although the method had a slow performance, it was capable of producing accurate results and overcoming some of the challenges in the estimation of model parameters typically experienced by local optimization methods.

Finally, these learnings were implemented in Chapter 5 when solving, per each frequency band of the FWI, the optimization problem of calculating the transformation matrix that would map the original system to an intermediate one constrained by a point-probes Hessian with a block-wise diagonal matrix structure. The estimates of  $V_P$  and  $\rho$  obtained in the numerical experiments exhibited substantially higher crosstalk effects than the results from the previous types of inversion. Moreover, it was observed through the Hessian matrices of the intermediate model space that although all the locations of the model grid exhibited the same  $3 \times 3$  matrix structure, the expected diagonal matrix was not correctly produced, bringing uncertainty to the results instead of minimizing it. Hence, three reasons were found to possibly explain the lack of success of the Simulated Annealing method in generating the expected transformation matrix; these were related to the need of finding a better annealing schedule that would lead to more accurate solutions, the definition of more appropriate objective functions, and the requirement of a larger transformation matrix or

more transformation matrices to would capture the variability of the local Hessian operators and produce a proper transformation of the point-probes Hessian.

## 6.2 Future work

In this thesis, the crosstalk that would be produced in seismic surface experiments with the  $\rho$ ,  $V_P$ , and  $V_S$  model space was investigated and two strategies were proposed to mitigate these inter-parameter coupled effects. These strategies were based on re-parameterizing the original model space to an intermediate one constrained by Hessian matrices exhibiting no correlation between model parameters. Although certain positive aspects of the decorrelation ideas were found, the re-parameterization to the expected model space was not properly achieved on a large scale for any of the two strategies. This section contains potential questions and next steps that are valuable to explore this topic in future studies.

With respect to the work discussed in Chapter 3, it was demonstrated that generalization was not achieved by implementing a unique transformation matrix to convert all the grid cells of the original system to the expected model space. Thus, it is worth contemplating if using several transformation matrices is a more appropriate approach to capture the spatial fluctuations in the values of the local Hessians, to properly transform them into the identity matrix, and produce a more consistent Hessian structure in the mesh. In this sense, efforts should be made in researching methods to compute many  $3 \times 3$  transformation matrices, but a enough quantity to not compromise the computational time and resources, as well as ideate techniques to extrapolate the information of the known transformation matrices to unexplored locations and still map each point to the sought model space.

On the other hand, considering the findings and analyses presented in Chapter 5, there are outstanding questions that could serve as the foundation for future research. Firstly, given that the Simulated Annealing technique did not produce the diagonalization of the point-probes Hessians using a unique transformation matrix, it would be interesting to investigate if using more transfor-

mation matrices could produce the expected point-probes Hessian in the intermediate model space, as proposed for the strategy of Chapter 3; in this case, a different objective function for the Simulated Annealing problem should be designed to include Hessian information produced by multiple transformation matrices. Additionally, another option entails using a unique transformation matrix of larger size, which would imply changing the type of re-parameterization conducted in this thesis (local-wise), i.e., mapping to the intermediate model space in patches of grid cells; in this sense, efforts should be made in understanding the selection of the appropriate matrix size and the optimal arrangement of grid cells. Nevertheless, in case the diagonalization of the point-probes Hessian is achieved by any of these methods, it is still important to assess if the computed transformation matrices produce a good enough generalization to properly transform other point-probes Hessians of the model grid, in order the crosstalk effects could be reduced.

Finally, should any of the proposals outlined for Chapter 3 or Chapter 5 prove successful, it would be worthwhile to perform experiments involving the modification of the optimization algorithm employed when solving the FWI. Specifically, transitioning from steepest descent to Gauss-Newton would be ideal to verify if effectively both types of estimated results are in close agreement, as indicated in the theoretical background, as well as to recognize if there are differences in computational time.

# References

- Aarts, E. and Korst, J. (1989). *Simulated Annealing and Boltzmann Machines: A Stochastic Approach to Combinatorial Optimization and Neural Computing*. John Wiley & Sons, Inc.
- Aki, K. and Richards, P. G. (2002). *Quantitative seismology*. University Science Books, 2nd edition.
- Amundaray, N. (2023). Full waveform inversion for CO2 monitoring: a feasibility study at the CMC newell county facility. Master's thesis, University of Calgary.
- Baumstein, A. (2014). Extended subspace method for attenuation of crosstalk in multi-parameter full wavefield inversion. In *SEG Technical Program Expanded Abstracts*, pages 1121–1125.
- Berenger, J. (1994). A perfectly matched layer for the absorption of electromagnetic waves. *Journal of computational physics*, 114:185–200.
- Berkhout, A. J. (1986). Seismic inversion in terms of pre-stack migration and multiple elimination. *Proceedings of the IEEE*, 74(3):415–427.
- Beydoun, W. B. and Tarantola, A. (1988). First Born and Rytov approximations: Modeling and inversion conditions in a canonical example. *The Journal of the Acoustical Society of America*, 83:1045–1055.
- Bording, R. P., Gersztenkorn, A., Lines, L. R., Scales, J. A., and Treitel, S. (1987). Applications of seismic travel-time tomography. *Geophysical Journal International*, 90:285–303.
- Bregman, N., Bailey, R., and Chapman, C. (1989). Crosshole seismic tomography. *Geophysics*, 54:200–215.

- Bunks, C., Salek, F., Zaleski, S., and Chavent, G. (1995). Multiscale seismic waveform inversion. *Geophysics*, 60:1457–1473.
- Deal, M. M. and Nolet, G. (1996). Nullspace shuttles. *Geophysical Journal International*, 124:372–380.
- Downton, J. E., Russell, B. H., and Lines, L. R. (2000). AVO for managers: pitfalls and solutions. *CREWES Research Report*, 12.
- Downton, J. E. and Ursenbach, C. (2005). Linearized AVO inversion with supercritical angles. *SEG Technical Program Expanded Abstracts 2005*.
- Eaid, M. V. (2021). *Distributed acoustic sensing: modelling, full waveform inversion, and its use in seismic monitoring*. PhD thesis, University of Calgary.
- Eriksson, J. (1996). *Optimization and Regularization of Nonlinear Least Squares Problems*. PhD thesis, Umea University.
- Faming, L., Yichen, C., and Guang, L. (2014). Simulated stochastic approximation annealing for global optimization with a square-root cooling schedule. *Journal of the American Statistical Association*, 109(506):847–863.
- Fatti, J. L., Smith, G. C., Vail, P. J., Strauss, P. J., and Levitt, P. R. (1994). Detection of gas in sandstone reservoirs using AVO analysis: A 3-D seismic case history using the geostack technique. *Geophysics*, 59:1362–1376.
- Fichtner, A. (2010). *Full Seismic Waveform Modelling and Inversion*. Springer, Heidelberg.
- Fichtner, A., Bunge, H. P., and Igel, H. (2006). The adjoint method in seismology: I. theory. *Physics of the Earth and Planetary Interiors*, 157:86–104.
- Fichtner, A. and Zunino, A. (2019). Hamiltonian nullspace shuttles. *Geophysical Research Letters*, 46:644–651.

- Grossman, J. P. (2003). AVO and AVA inversion challenges: a conceptual overview. *CREWES Research Report*, 15.
- Guitton, A. (2004). Amplitude and kinematic corrections of migrated images for nonunitary imaging operators. *Geophysics*, 69(4):1017–1024.
- Hu, Q. and Innanen, K. (2019). Study of crosstalk reduction in multiparameter acoustic FWI. *CREWES Research Report*, 31.
- Innanen, K. (2014). Seismic AVO and the inverse hessian in precritical reflection full waveform inversion. *Geophysical Journal International*, 199:717–734.
- Innanen, K. (2020a). Application of misfit-based model space coordinate system design to seismic AVO inversion. *CREWES Research Report*, 32(28).
- Innanen, K. (2020b). Application of misfit-based model space coordinate system design to seismic full waveform inversion. *CREWES Research Report*, 32(29).
- Innanen, K. (2020c). Model re-parameterization via misfit-based coordinate transforms. *CREWES Research Report*, 32(30).
- Innanen, K. (2020d). Numerical procedures for computing constrained coordinate transforms. *CREWES Research Report*, 32(31).
- Innanen, K. (2020e). A review of tensors in non-cartesian coordinate systems. *CREWES Research Report*, 32(27).
- Kamath, N. and Tsvankin, I. (2016). Elastic full-waveform inversion for VTI media: Methodology and sensitivity analysis. *Geophysics*, 81(2):C53–C68.
- Kamei, R. and Pratt, R. (2013). Inversion strategies for visco-acoustic waveform inversion. *Geophysical Journal International*, 194:859–884.

- Keating, S. (2020). *Viscoelastic full-waveform inversion: treating attenuation uncertainty, characterizing cross-talk, and quantifying confidence in inversion results*. PhD thesis, University of Calgary.
- Keating, S. and Innanen, K. (2018). Using multi-resolution truncated Newton optimization for cross-talk reduction in FWI. *CREWES Research Report*, 30.
- Keating, S. and Innanen, K. (2019). Parameter cross-talk and leakage between spatially-separated unknowns in viscoelastic full waveform inversion. *CREWES Research Report*, 31(35).
- Keating, S. D. and Innanen, K. A. (2021). Null-space shuttles for targeted uncertainty analysis in full-waveform inversion. *GEOPHYSICS*, 86:R63–R76.
- Kennett, B. L. N., Sambridge, M. S., and Williamson, P. R. (1988). Subspace methods for large inverse problems with multiple parameter classes. *Geophysical Journal-Oxford*, 94(2):237–247.
- Keys, R. G. (1986). An application of Marquardt’s procedure to the seismic inverse problem. *Proceedings of the IEEE*, 74(3):476–486.
- Kirkpatrick, S., Gelatt, C. D., and Vecchi, M. P. (1983). Optimization by simulated annealing. *Science*, 220(4598):671–680.
- Lailly, P. (1983). The seismic inverse problem as a sequence of before stack migrations. In *Conference on Inverse Scattering, Theory and Application*, pages 206–220. Society for Industrial and Applied Mathematics, Expanded Abstracts.
- Larsen, J. A. (1999). AVO analysis by simultaneous P-P and P-S weighted stacking. Master’s thesis, University of Calgary.
- Lines, L. R. and Treitel, S. (1984). Tutorial: A review of least-squares inversion and its application to geophysical problems. *Geophysical Prospecting*, 32(2):159–186.

- Liu, Q. and Peter, D. (2020). Square-root variable metric-based nullspace shuttle: A characterization of the nonuniqueness in elastic full-waveform inversion. *Journal of Geophysical Research: Solid Earth*, 125.
- Lloyd, H. J. E. and Margrave, G. F. (2011). Bandlimited impedance inversion: using well logs to fill low frequency information in a non-homogenous model. *CREWES Research Report*, 23.
- Locatelli, M. (2000). Simulated annealing algorithms for continuous global optimization: Convergence conditions. *Journal of Optimization Theory and Applications*, 104:121–133.
- Luenberger, D. G. and Ye, Y. (2008). *Linear and Nonlinear Programming*. Springer.
- Mahmoudian, F. and Margrave, G. F. (2003). AVO inversion of multicomponent data for P and S impedance. *CREWES Research Report*, 15.
- Mora, P. (1987). Nonlinear two-dimensional elastic inversion of multioffset seismic data. *Geophysics*, 52:1211–1228.
- Moradi, S. and Innanen, K. (2015). Radiation patterns associated with the scattering from viscoelastic inclusions. *CREWES Research Report*, 27(53).
- Moradi, S. and Innanen, K. (2019). Azimuthally-dependent scattering potentials and full waveform inversion sensitivities in low-loss viscoelastic orthorhombic media. *Journal of Geophysics and Engineering*, 16:367–388.
- Métivier, L., Brossier, R., Operto, S., and Virieux, J. (2015). Acoustic multi-parameter FWI for the reconstruction of P-wave velocity, density and attenuation: preconditioned truncated Newton approach. *SEG Expanded Abstracts*, page 1198–1203.
- Métivier, L., Brossier, R., Virieux, J., and Operto, S. (2013). Full waveform inversion and the Truncated Newton Method. *SIAM Journal on Scientific Computing*, 35(2):B401–B437.
- Nocedal, J. and Wright, S. J. (2006). *Numerical Optimization*. Springer, 2nd edition.

- Operto, S., Gholami, Y., Prieux, V., Ribodetti, A., Brossier, R., Métivier, L., and Virieux, J. (2013). A guided tour of multiparameter full-waveform inversion with multicomponent data: From theory to practice. *The Leading Edge*, page 1040–1054.
- Operto, S., Ravaut, C., Improta, L., Virieux, J., Herrero, A., and Dell’Aversana, P. (2004). Quantitative imaging of complex structures from dense wide-aperture seismic data by multiscale traveltimes and waveform inversions: A case study. *Geophysical Prospecting*, 52:625–651.
- Pan, W., Geng, Y., and Innanen, K. A. (2018). Interparameter trade-off quantification and reduction in isotropic-elastic full-waveform inversion: synthetic experiments and hussar land data set application. *Geophysical Journal International*, 213(2):1305–1333.
- Pan, W., Innanen, K. A., Geng, Y., and Li, J. (2019). Interparameter trade-off quantification for isotropic-elastic full-waveform inversion with various model parameterizations. *Geophysics*, 84(2):R185–R206.
- Pan, W., Innanen, K. A., Margrave, G. F., Fehler, M. C., Fang, X., and Li, J. (2016). Estimation of elastic constants for HTI media using Gauss-Newton and full-Newton multiparameter full waveform inversion. *Geophysics*, 81(5):R275–R291.
- Pica, A., Diet, J. P., and Tarantola, A. (1990). Nonlinear inversion of seismic reflection data in a laterally invariant medium. *Geophysics*, 55:284–292.
- Pratt, R. G. (1990). Frequency-domain elastic wave modeling by finite differences; a tool for crosshole seismic imaging. *Geophysics*, 55:626–632.
- Pratt, R. G., Shin, C., and Hick, G. (1998). Gauss-Newton and full Newton methods in frequency-space seismic waveform inversion. *Geophysical journal international*, 133:341–362.
- Rutenbar, R. (1989). Simulated annealing algorithms: an overview. *IEEE Circuits and Devices Magazine*, 5(1):19–26.

- Santosa, F. and Symes, W. W. (1988). Computation of the hessian for least-squares solutions of inverse problems of reflection seismology. *Inverse Problems*, 4(1):211.
- Sato, H. (1984). Attenuation and envelope formation of three-component seismograms of small local earthquakes in randomly inhomogeneous lithosphere. *J. Geophys. Res.*, 89:1221–1241.
- Sato, H., Fehler, M. C., and Maeda, T. (2012). *Seismic Wave Propagation and Scattering in the Heterogeneous Earth*. Springer, 2nd edition.
- Schuster, G. T. and Hu, J. (2000). Green's function for migration: Continuous recording geometry. *Geophysics*, 65(1):167–175.
- Sears, T. J., Singh, S., and Barton, P. (2008). Elastic full waveform inversion of multi-component OBC seismic data. *Geophysical Prospecting*, 56:843–862.
- Sheriff, R. E. (2002). *Encyclopedic Dictionary of Applied Geophysics*. Society of Exploration Geophysicists.
- Shin, C. S. (1988). *Nonlinear elastic wave inversion by blocky parameterization*. PhD thesis, University of Tulsa.
- Shuey, R. (1985). A simplification of the Zoeppritz equations. *Geophysics*, 50(4):609–614.
- Sirgue, L. and Pratt, R. G. (2004). Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies. *Geophysics*, 69:231–248.
- Skopintseva, L., Ayzenberg, M., Landrø, M., Nefedkina, T., and Aizenberg, A. M. (2011). Long-offset AVO inversion of PP reflections from plane interfaces using effective reflection coefficients. *Geophysics*, 76(6):C65–C79.
- Smith, B. and Wong, A. (2017). Simulated annealing of constrained statistical functions. In Peyvandi, H., editor, *Computational Optimization in Engineering*, chapter 4. IntechOpen, Rijeka.

- Smith, G. C. and Gidlow, P. M. (1987). Weighted stacking for rock property estimation and detection of gas. *Geophysical Prospecting*, 35:993–1014.
- Song, Z. and Williamson, P. R. (1995a). Frequency-domain acoustic-wave modeling and inversion of crosshole data: Part I—2.5-D modeling method. *Geophysics*, 60:784–795.
- Song, Z. and Williamson, P. R. (1995b). Frequency-domain acoustic-wave modeling and inversion of crosshole data: Part II—inversion method, synthetic experiments and real-data results. *Geophysics*, 60:796–809.
- Stewart, R. R. (1990). Joint P and P-SV inversion. *CREWES Research Report*, 2.
- Stewart, R. R. (1994). The present and promise of P-S seismic exploration. *CREWES Research Report*, 6.
- Stewart, R. R., Gaiser, J. E., Brown, R. J., and Lawton, D. C. (1999). Converted-wave seismic exploration: A tutorial. *CREWES Research Report*, 3(11):42.
- Talwani, M. and Kessinger, A. (2003). Exploration geophysics. In *Encyclopedia of Physical Science and Technology*, pages 709–726. Academic Press, 3rd edition.
- Tarantola, A. (1984). Inversion of seismic reflection data in the acoustic approximation. *Geophysics*, 49:1259–1266.
- Tarantola, A. (1986). A strategy for nonlinear inversion of seismic reflection data. *Geophysics*, 51:1893–1903.
- Tarantola, A. (2005). Inverse problem theory and methods for model parameter estimation. *SIAM*, 89.
- Ursenbach, C. (2003). Can multicomponent or joint AVO inversion improve impedance estimates? *SEG Technical Program Expanded Abstracts*, pages 161–164.
- Valenciano, A. A. (2008). *Imaging by wave-equation inversion*. PhD thesis, Stanford University.

- Virieux, J. and Operto, S. (2009). An overview of full-waveform inversion in exploration geophysics. *Geophysics*, 74(6):WCC1–WCC26.
- Wang, T. and Cheng, J. (2017). Elastic full waveform inversion based on mode decomposition: The approach and mechanism. *Geophysical Journal International*, 209:606–622.
- Wang, Y., Dong, L., Liu, Y., and Yang, J. (2016). 2D frequency-domain elastic full-waveform inversion using the block-diagonal pseudo-hessian approximation. *Geophysics*, 81(5):R247–R259.
- Williams, R. G., Roberts, G., and Hawkins, K. (2001). Long offset towed streamer recording - a cheaper alternative to multi-component OBC for exploration? *Exploration Geophysics*, 32:316–319.
- Wu, R. S. and Aki, K. (1985). Scattering characteristics of elastic waves by an elastic heterogeneity. *Geophysics*, 50:582–595.
- Xing, G. and Zhu, T. (2020). Hessian-based multiparameter fractional viscoacoustic full-waveform inversion. *SEG Technical Program Expanded Abstracts*, pages 895–899.
- Xu, W., Wang, T., and Cheng, J. (2017). Elastic least-squares reverse time migration based on wave-mode decomposition. *SEG Technical Program Expanded Abstracts*, pages 2497–2502.
- Xu, Y. and Bancroft, J. C. (1997). Joint AVO analysis of PP and PS seismic data. *CREWES Research Report*, 9.
- Yang, P., Brossier, R., Metivier, L., and Virieux, J. (2016). A review on the systematic formulation of 3-D multiparameter full waveform inversion in viscoelastic medium. *Geophysical Journal International*, 207:129–149.
- Zou, X. (2020). *Atmospheric Satellite Observations*. Academic Press.