

2012-01-01

# Constrained Optimization Schemes for Geophysical Inversion of Seismic Data

Uram Anibal Sosa Aguirre

*University of Texas at El Paso, uasosa@icesi.edu.co*

Follow this and additional works at: [https://digitalcommons.utep.edu/open\\_etd](https://digitalcommons.utep.edu/open_etd)



Part of the [Applied Mathematics Commons](#), and the [Geophysics and Seismology Commons](#)

---

## Recommended Citation

Sosa Aguirre, Uram Anibal, "Constrained Optimization Schemes for Geophysical Inversion of Seismic Data" (2012). *Open Access Theses & Dissertations*. 2196.

[https://digitalcommons.utep.edu/open\\_etd/2196](https://digitalcommons.utep.edu/open_etd/2196)

This is brought to you for free and open access by DigitalCommons@UTEP. It has been accepted for inclusion in Open Access Theses & Dissertations by an authorized administrator of DigitalCommons@UTEP. For more information, please contact [lweber@utep.edu](mailto:lweber@utep.edu).

CONSTRAINED OPTIMIZATION SCHEMES FOR GEOPHYSICAL INVERSION  
OF SEISMIC DATA

URAM ANIBAL SOSA AGUIRRE

Program in Computational Science

APPROVED:

---

Aaron Velasco, Chair, Ph.D.

---

Vladik Kreinovich, Ph.D.

---

Rodrigo Romero, Ph.D.

---

Carsten Burstedde, Ph.D.

---

Benjamin Flores, Ph.D.  
Dean of the Graduate School

*A mi madre y mi hermana*

*A mi amorcito y*

*A mi Padre*

*que la fuerza siga con nosotros.*

CONSTRAINED OPTIMIZATION SCHEMES FOR GEOPHYSICAL INVERSION  
OF SEISMIC DATA

by

URAM ANIBAL SOSA AGUIRRE

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at El Paso

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

Program in Computational Science

THE UNIVERSITY OF TEXAS AT EL PASO

December 2012

# Acknowledgements

I would like to express my sincere gratitude to my advisor Dr. Aaron Velasco, an excellent human being and mentor, for guiding me through this path towards the obtention of my Doctorate, for his unvaluable advice, for his constant support, for his neat language and remarkable sense of humor.

I also wish to thank the other members of my committee Dr. Vladik Kreinovich, an incredible mathematician and researcher. His inexhaustible source of energy allows him to know almost about everything in science, and know it well. Dr. Rodrigo Romero is the most skilled computer scientist I have ever met. As former Associate Director of the Cyber-ShARE center he helped me grow with it as an interdisciplinary scientist, he was always willing to share his wisdom and advice. Finally, to Dr. Carsten Burstedde a leader researcher in computational science, an outstanding scientist, and a remarkable friend. Their suggestions, comments and additional guidance were definitive for the completion of this work. I sincerely would not be at this stage of my professional life without their contribution in all fronts to my personal and professional life.

I specially want to thank Paula Gonzalez, the love of my life and future wife, a woman every man would dream to have by his side. Her support, love, and delicious meals became the fuel that I needed to finish my PhD. As important as Paula was the incredible effort of my parents Clara and Efrain, who always encouraged me to pursuit a higher degree, to do science and to become a better man day by day; without them, would not be possible for me to reach this stage of my life.

This work would not have been possible without the scientific collaborations of Dr. Rodrigo Romero, Dr. Carsten Burstedde, Dr. Miguel Arguez and Dr. Leticia Velazquez. I would like to thank all the people from the UTEP Program in Computational Science and from the Cyber-ShARE Center. As a former mathematician, my involvement in the geoscience research sub-project at the center has been a life-changing experience. Also no less important was the financial support of the NSF grants of Crest Cyber-ShARE HRD-0734825 and NEESR-

SG CMMI-0619078, and from Universidad Icesi (Cali, Colombia). Chapter 2 is based on a paper co-authored with Dr. Miguel Arguez, Dr. Leticia Velazquez, and Dr. Rodrigo Romero. Chapter 3 is based on a paper co-authored with Lennox Thompson and Dr. Rodrigo Romero. Chapter 4 is based on a paper co-authored with Dr. Carsten Burstedde, Dr. Miguel Arguez, Dr. Leticia Velazquez and Dr. Omar Ghattas.

In the dissertation's acknowledgements of my advisor Dr. Aaron Velasco, it is mentioned that he felt motivated to excel beyond his own expectations. I recognize, without any doubt, that he has excelled since then in his field and as a mentor. I have not seen a Professor more devoted to the excellence of his students, to their support, and to making them grow. In the name of all of his students, I thank him and hope to use similar skills in the future, with those who will become my students.

# Abstract

Many experimental techniques in geophysics advance the understanding of Earth processes by estimating and interpreting Earth structure (e.g., velocity and/or density structure). These techniques use different types of geophysical data which can be collected and analyzed separately, sometimes resulting in inconsistent models of the Earth depending on data quality, methods and assumptions made. This dissertation presents two approaches for geophysical inversion of seismic data based on constrained optimization. In one approach we expand a one dimensional (1-D) joint inversion least-squares (LSQ) algorithm by introducing a constrained optimization methodology. Then we use the 1-D inversion results to produce 3-D Earth velocity structure models. In the second approach, we provide a unified constrained optimization framework for solving a 1-D inverse wave propagation problem.

In Chapter 2 we present a constrained optimization framework for joint inversion. This framework characterizes 1-D Earth's structure by using seismic shear wave velocities as a model parameter. We create two geophysical synthetic data sets sensitive to shear velocities, namely receiver function and surface wave dispersion. We validate our approach by comparing our numerical results with a traditional unconstrained method, and also we test our approach robustness in the presence of noise. Chapter 3 extends this framework to include an interpolation technique for creating 3-D Earth velocity structure models of the Rio Grande Rift region. Chapter 5 introduces the joint inversion of multiple data sets by adding delay travel times information in a synthetic setup, and leave the possibility to include more data sets. Finally, in Chapter 4 we pose a 1-D inverse full-waveform propagation problem as a PDE-constrained optimization program, where we invert for the material properties in terms of shear wave velocities throughout the physical domain. We facilitate the implementation and comparison of different constrained optimization methods, through a unified affine invariant approach that incorporates inequality constraints for solving the inverse problem. We expect to contribute in broadening the use of constrained optimization algorithms to solve geophysical inverse problems.

# Table of Contents

	Page
Acknowledgements . . . . .	iv
Abstract . . . . .	vi
Table of Contents . . . . .	vii
List of Figures . . . . .	x
<b>Chapter</b>	
1 Introduction . . . . .	1
2 Constrained Optimization Framework for Joint Inversion of Geophysical Data Sets .	3
2.1 Introduction . . . . .	3
2.2 Joint inversion formulation . . . . .	5
2.3 Methodology . . . . .	7
2.3.1 Truncated Singular Value Decomposition (TSVD) . . . . .	8
2.3.2 Primal-Dual Interior-Point (PDIP) Method . . . . .	8
2.4 Numerical Experimentation . . . . .	11
2.4.1 Receiver Function . . . . .	12
2.4.2 Surface Wave Dispersion . . . . .	13
2.4.3 Synthetic Models . . . . .	14
2.5 Test Results . . . . .	14
2.5.1 Velocity Model Tests . . . . .	17
2.5.2 Parameter Selection Tests . . . . .	17
2.5.3 Noise Synthetic Tests . . . . .	23
2.6 Discussion . . . . .	24
2.7 Conclusions . . . . .	28
3 3-D Structure of the Southern Rio Grande Rift from 1-D Constrained Joint Inversion of Receiver Functions and Surface Waves Dispersion . . . . .	29
3.1 Introduction . . . . .	29

3.2	Tectonic Setting . . . . .	31
3.3	Receiver Functions . . . . .	32
3.4	Surface Wave Dispersion . . . . .	33
3.5	1-D Constrained Joint Inversion . . . . .	34
3.6	Joint Inversion Results . . . . .	36
3.7	Kriging interpolation from 1-D velocity profiles . . . . .	37
3.8	Discussion . . . . .	43
3.8.1	PDIP methodology . . . . .	43
3.8.2	3-D velocity structure from 1-D velocity profiles . . . . .	44
3.8.3	Interpretation: Basin and Range, Colorado Plateau, RGR, and Great Plains . . . . .	45
3.8.4	Socorro Magma Body . . . . .	46
3.9	Conclusions . . . . .	47
4	Constrained Optimization Schemes for a 1D Inverse Wave Propagation Problem . .	49
4.1	Introduction . . . . .	49
4.1.1	Notation . . . . .	51
4.2	Formulation of the inverse problem . . . . .	52
4.2.1	Problem statement . . . . .	53
4.2.2	Inexact Newton-Krylov algorithm . . . . .	54
4.2.3	Affine invariance . . . . .	56
4.3	Three constrained optimization schemes . . . . .	56
4.3.1	Logarithmic-barrier method . . . . .	57
4.3.2	Primal-dual interior point method . . . . .	59
4.3.3	Primal-dual active set method . . . . .	61
4.4	Algorithms for constrained optimization . . . . .	62
4.4.1	Generic Newton-Krylov method . . . . .	63
4.4.2	Force positivity . . . . .	64
4.4.3	Sufficient decrease . . . . .	65
4.4.4	Unified algorithm framework . . . . .	67

4.5	Numerical results for an inverse wave propagation problem . . . . .	68
4.5.1	A one-dimensional wave propagation example . . . . .	68
4.5.2	Error measures, convergence and affine invariance . . . . .	70
4.5.3	Basic tests . . . . .	71
4.5.4	Experiment A . . . . .	76
4.5.5	Experiment B . . . . .	81
4.6	Summary and concluding remarks . . . . .	86
5	Joint Inversion of Multiple Geophysical Data Sets . . . . .	88
5.1	Seismic travel times . . . . .	89
5.1.1	Travel times inversion . . . . .	90
5.2	Joint inversion of receiver functions, surface waves and travel times data . . .	92
6	Concluding Remarks . . . . .	98
	References . . . . .	99
	Curriculum Vitae . . . . .	108

# List of Figures

2.1	(Left) Waveform of a Receiver Function. The halfspace $h$ represents usually the Moho. (Right) The Moho converted phase $Ps$ and the multiples $PpPs$ , $PpSs$ , and $PsPs$ are labeled, and their ray paths. With the exception of the first arrival, lowercase letters denote upgoing travel paths, uppercase letters denote downgoing travel paths [69]. . . . .	13
2.2	(Left) $S$ -wave synthetic velocity model (Rift) (Laske et al. 2000). (Right: top) Receiver functions for three different ray parameters $p$ , and (Right: bottom) surface wave dispersion curves (Love and Rayleigh) computed for the velocity model on the left. . . . .	15
2.3	The single inversion of receiver functions (left top) identifies velocity contrasts, while single inversion of surface waves (left bottom) gives information on the average velocities at different depths. The joint inversion (right) of these two data sets combines all this information and provides a substantial improvement in the final estimated model. . . . .	16
2.4	Archean model. Notice that the residual errors are about the same at each iteration, with faster convergence for TSVD method. This means that the TSVD method reaches minimal RMS error faster than the PDIP method. . . . .	18
2.5	Continental model. The relative RMS and residual errors are close at the last two iterations for both methods, but PDIP minimize better the RMS and converges faster than TSVD. Both average velocities come closer to the true average velocity at the end of the inversion. . . . .	18
2.6	Orogen model. The relative error is smaller for PDIP at all stages of the algorithm, while the residual error is similar for both methods. The TSVD estimated velocity is closer to the true average velocity. . . . .	19

2.7	Rift model. The relative error and the residual error are within the same order of accuracy for both methods. The average is better resolved for PDIP until the 3rd iteration, where the smaller relative RMS error is reached. . . . .	19
2.8	Logarithmic scale plot of the relative RMS and residuals errors associated to different choices of the regularization parameter $\lambda$ . We applied the joint inversion algorithm over the Rift model when solving the inverse problem by using the TSVD method. . . . .	21
2.9	Numerical results obtained for the TSVD method when we use truncation without regularization. . . . .	22
2.10	Relative RMS errors associated to different levels of Gaussian noise added to the data. As we will see in Figure 2.13, PDIP offers a steady behavior of the quality in the inverted models, i.e., they seem to be less affected by noise as it increases. . . . .	24
2.11	Relative errors associated to different levels of Gaussian noise added to the data. Both methods give comparable results, being PDIP more accurate to fit the data.	24
2.12	Rift velocity model approximation obtained for several levels of noise. On top: a) represents 5% noise and b) 10% noise added to the data. Both methods reach about the same level of convergence. The plots in the middle corresponds to c) 15% and d) 20% noise. The effect of such level of noise shows up considerably in the lower layers, still the interfaces are well characterized. Finally in the bottom with e) 25% and f) 30% Both methods fail to identify properly the lower layers, being TSVD more affected by the noise. . . . .	25
2.13	Receiver function and surface waves results for increasing noise percentage. From top 10%, 20% in the middle and bottom 30% over the Rift model. The fitness of both methods is better for the RF synthetics, while for SW we notice how the approximations begin to degrade considerably more for the TSVD estimated synthetics than for the PDIP ones. . . . .	26

3.1	Regional topography map of the SRGR and plot of all the seismicity from 1975-2012. The white circles signify the various sizes of magnitude from 1-5. Three different profiles has been traced to study the upper mantle structure of the provinces described in Figure 3.2. . . . .	30
3.2	Regional topography of the SRGR and sketch of its boundaries. The white triangles mark the stations which data was used for this study. . . . .	32
3.3	1-D joint inversion results for station 426A within the RGR. (Top) The final S-wave velocity profile, in red, provided by our joint inversion approach, initialized with the standard model ak135 in black. (Bottom) We show the acceptable fit to the RF observations (left) in black from our synthetic approximation in red for a ray parameter $p = 0.0446$ and in green for $p = 0.0546$ , and to Love and Rayleigh wave group dispersion observations (right). . . . .	38
3.4	Cross-section A-A' at latitude $34^\circ$ shows a clear distinction between the Colorado Plateau (CP), Socorro Magma Body (SMB) at the center of the RGR, and the Great Plains (GP). We find that near the CP, there is a low velocity lid with high velocities beneath this province, consistent with other studies. Anomalous high velocities begin to appear right below the RGR and continue east of the Great Plains between the depths of 200 - 300 km . . . . .	40
3.5	Cross-section B-B' coincides with longitude $107^\circ$ and intersects cross section A-A' and passes through the SMB. The SMB seems to have slow velocities between the Colorado Plateau and Great Plains. . . . .	40
3.6	Cross-section C-C' coincides with the southern part of the LA RISTRA passive experiment. Seismically fast mantle underlies the RGR and relatively slow mantle is seen beneath the Socorro Magma Body and Colorado Plateau. . . .	41
3.7	Full 3-D velocity structure perspective of the RGR view from N-W. We identify low velocities associated to the southernmost part of the RGR. The upper 200 – 300 km of mantle beneath the magmatically and tectonically active RGR and Basin & Range (B & R) is seismically distinct from the mantle beneath the stable Colorado Plateau and Great Plains. . . . .	41

3.8	Cross-section D-D' at latitude 32° covering the southern-most part of the RGR region. We notice the transition between the Basin & Range province and the Colorado Plateau to the north and the Great Plains to the west. Thin crust is identified in the B & R and thicker crust in the other two provinces. We image a low velocity zone that begins to appear beneath the RGR extending to the west below the B & R and Colorado Plateau. . . . .	42
4.1	Numerical results for problem A1. Top: No regularization. Bottom: Regularization applied with $\beta = 10^{-6}$ . In this problem, the target model parameter $\mu^*$ , (in red), always lies within the bounds. The level of discretization for $\mu$ corresponds to $j = 3, 4$ , and 5 where we allow 5, Newton-Krylov iterations for the coarser grids and 15 for the final approximation. We display the final approximations for $\mu$ with colors that vary according to each method. . . . .	78
4.2	Numerical results for problem A2. Top: No regularization. Bottom: Regularization applied with $\beta = 10^{-5}$ . In Problem A2, the target model parameter $\mu^*$ has some components lying on the boundary. The level of discretization for $\mu$ and the number of Newton-Krylov iterations is the same as for Problem A1. We show the target model $\mu^*$ in red for the last stage of the algorithm, with the same color scheme for the final approximations of $\mu$ in Figure 4.1. . . . .	79
4.3	Point-wise boundaries for PDAS (top) and PDIP (bottom) with no regularization applied to Problem A1. The lines with the markers represent the point-wise boundaries. Notice how the target $\mu$ in red and all the iterates satisfy the inequality $a_i \leq (\mu_k)_i \leq b_i$ for all $k$ . . . . .	81
4.4	Final stage of Experiment B. As before, we present the optimal $\mu$ for each method and true model parameter $\mu^*$ . No regularization is applied. For this experiment, we set $T = 50$ in (4.62), which is twice the value of $T$ in the previous experiment. . . . .	83
4.5	Final stage of Experiment B. We present the optimal $\mu$ for each method and true model parameter $\mu^*$ . Regularization applied with $\beta = 10^{-6}$ . . . . .	84

4.6	Point-wise boundaries for PDAS with $v = 0.1$ (top) and $v = 0.2$ (bottom) with no regularization applied to problem B. The lines with the markers represent the point-wise boundaries. Notice how all the iterates satisfy $a_i \leq (\mu_k)_i \leq b_i$ for all $k$ , however the target $\mu$ in red fails to be inside for the highest peaks at depth for $v = 0.1$ . . . . .	85
4.7	Point-wise boundaries for PDIP with $v = 0.1$ (top) and $v = 0.2$ (bottom) with no regularization applied to problem B. The lines with the markers represent the point-wise boundaries. Notice how the target $\mu$ in red and all the iterates hold $a_i \leq (\mu_k)_i \leq b_i$ for all $k$ . . . . .	87
5.1	(Left) Three different ray paths corresponding to the phases $sP$ , $pP$ and $P$ for an earthquake at a teleseismic distance of the receiver on a spherical Earth geometry (adapted from [69]). (Right) Flat layered model beneath the receiver with velocities $x_0, x_1, \dots, x_n$ . . . . .	91
5.2	a) Two rays showing the relationship between the angle of incidence $i$ , ray parameter, and the slope of the travel time curve. b) Ray geometry for a reflection flat-layered medium. Layer thicknesses are $h_j$ , horizontal distances traveled are $l_j$ , and one-way travel times spent in the layers are $\Delta T_j$ (adapted from [69]). . . . .	92
5.3	Single inversion of RF (left) compared to RF and TT joint inversion (right). The bottom of the model is being better resolved by the joint inversion. . . . .	93
5.4	Single inversion of SW (left) compared to SW and TT joint inversion (right). The average resolving power of surface waves has been preserved but better constrained in depth. . . . .	94
5.5	When compared the joint inversion of RF, SW and TT inversion to that of RF and SW, we identified a slight improvement in the resolution of the model with depth. . . . .	95
5.6	Numerical comparison between three different level of data set influence over the inverted models in station 223A. The results appear to be very similar for all levels of influence, but when less influence is given to receiver functions, slight differences can be seen after 200 km. . . . .	96

5.7 Numerical comparison between three different level of data set influence over the inverted models in station 219A. Allowing the influence of receiver function data to be only 20% seems to reduce rapid variations in velocities, and identify a low velocity zone at 250 km. We need more information to decide if this is an artifact or not. . . . . 97

# Chapter 1

## Introduction

The determination of one-dimensional (1-D) to three-dimensional (3-D) seismic structure of the Earth has been an area of major interest in geophysical research in the last decades. Geophysical inversion can be used to model the Earth structure for solving either geological or exploration problems. If only limited knowledge about the subsurface is required, inferences drawn directly from the data can be sufficient. This approach is generally a linear framework in which data is collected, processed, plotted, and interpreted.

However, when more detailed information about the subsurface is required, quantitative models of the Earth need to be estimated. Providing this type of models is the objective of geophysical inversion. The basic ingredients of this type of inversion can be summarized as measured data, prior information, inversion algorithms, and a physical properties distribution of the target region. Since we have access to all of these ingredients, we have been motivated to address two key problems in geophysics. First, how to integrate multiple geophysical datasets with different sensitivity and resolution domains to characterize Earth structure, and second how to identify robust schemes that incorporate inequality constraints for solving one dimensional full-waveform propagation problems. We based our motivation on the current and novel computational optimization techniques for solving inverse problems [23, 57], on the advantages of integrating multiple datasets [34, 35, 52, 56, 55] and on the resolving power of full-waveform seismic inversion to reconstruct Earth models [12].

In this dissertation, we present alternative schemes to address the above-mentioned questions, by creating generic constrained optimization frameworks for joint inversion and for full-waveform inversion. In both cases we focus on a one-dimensional problem to show that our approaches are competitive with respect to traditional strategies for solving the inverse problems. Furthermore, for our joint inversion approach we combine the 1-D inverted models with

a robust Bayesian interpolation scheme to produce 3-D Earth structure models.

Chapters 2, 3, and 5 present our contributions to expand a joint inversion algorithm [34] with a constrained optimization approach and create full 3-D Earth structure models from 1-D velocity profiles. In Chapter 2, we present our 1-D constrained optimization approach for joint inversion. We compare our approach with the traditional methodology, to demonstrate that our approach can be a robust alternative when compared to the usual unconstrained formulation. In Chapter 3 we integrate this approach with a Bayesian kriging interpolation scheme, in order to combine 1-D velocity profiles associated with 147 seismic stations deployed in the Rio Grande Rift (RGR). The interpolated results allow us to produce 3-D velocity models of the crust and upper mantle in the RGR. Finally, in Chapter 5, we add a new data set, i.e., delay travel times, to our joint inversion algorithm, to test its compatibility and benefits respect to joint inversion of only surface waves and receiver functions. To the best of our knowledge, these three data sets have not been used for joint inversion. For full-waveform inversion, we present in Chapter 4 a generic algorithmic framework that accomodates three different constrained optimization strategies for solving the inverse problem. We also make our generic framework affine invariant, we propose an improvement on line search step computations, and test the use of point-wise boundaries over the model parameter. In this case, we do not know of any attempt to use this sort of constraints for this type of inversion algorithms.

# Chapter 2

## Constrained Optimization Framework for Joint Inversion of Geophysical Data Sets

### 2.1 Introduction

The search of meaningful Earth (velocity and density) models requires efficient and robust computational techniques, especially in the context of increasing data collection of multiple types of data sets. Motivated by the current and novel computational optimization techniques for solving inverse problems [23, 57] and by the advantages of integrating multiple data sets [26, 29, 34, 35, 52, 56, 55], joint inversion of multiple types of data sources can better determine physical properties of the Earth.

Joint inversion in geophysics means simultaneous optimization of several objective functions, such as  $\ell_2$ -misfit functions, allows for estimating a model that explains all data sets at once. Since the objective function is expected to be less subject to local minima, this approach reduces intrinsic non-uniqueness of the problem ([15]). The simultaneous inversion of multiple data sets is not a new concept, and different approaches have been developed as joint inversion of independent data sets. For example, cooperative inversion [49] manually adjusts values to improve the fit to the data, yet the approach becomes usually biased to the input model. Also, a weighted scheme involving all the data sets simultaneously can be used for seismic travel times and gravity data [46], for DC resistivity and seismic data [26], receiver functions and surface wave dispersion [34], and receiver functions, surface wave dispersion, and magnetotelluric data [56, 55]. For both approaches (cooperative and weighted), the main assumption in most of the cases is that the data sets used in the inversion have similar geological boundaries.

Joint inversion can be considered successful if the following conditions hold [34]: each data

set samples the same propagating medium (consistency), and the combination of the data sets increases the resolution of the inverted model (complementarity). The success of this type of inversions, assuming that the above conditions hold, relies on the complementarity between the data sets and imposes better physical constraints, and this complementarity increases the resolution of the final model. However, difficulties arise for highly nonlinear misfit functions and large-dimensional model spaces, which include: 1) the identification of the appropriate weights and the level of influence of each data set over the final inverted model, 2) the lack of compatibility among the data sets, 3) the presence of (spurious) solutions which are not geophysically meaningful, and 4) the inherent ill-conditioning of the inverse problem, which necessitates an appropriate choice of regularization and smoothing constraints. Our goal in this work is to address the last two difficulties.

Application of equality and inequality constraints is a difficult task that existent algorithms usually do not address. As indicated in [6], the standard approach for this type of problems uses a Lagrangian method. Some previous work uses a quadratic programming approach based on stable simplex-type solvers, which is usually computationally expensive [27]. [56] attempts to keep the model parameters feasible by means of a genetic algorithm as a constrained optimization method, an inherently computational intense implementation that allows only a few parameters to be present in the inversion. [47] uses interior-point methods for single inversions in an unconstrained formulation, however primal-dual interior point (PDIP) methods have not been implemented widely for joint inversion of geophysical data.

Joint inversion optimization overcomes one major difficulty in the geosciences: to integrate multiple data sets that have different sensitivities and resolutions. Complementary information that constrains the physics of the problem can be incorporated into the inverse process, and can help us to resolve ambiguities. In this chapter, we implement a constrained optimization approach to expand a previously developed joint inversion algorithm. We apply the algorithm to two geophysical data sets: teleseismic P-wave receiver functions and surface wave dispersion velocities, to find a mutually consistent estimate of one-dimensional (1-D) Earth structure (e.g. [34, 35]). To this end, we characterize the Earth as a layered structure using seismic shear velocities as a model parameter, while we focus on a 1-D problem to establish the capabilities

of our optimization approach when compared to traditional strategies for solving the inverse problem. We test our method using synthetic cases to show that our constrained optimization approach to perform the inversion not only can be as accurate as the traditional unconstrained formulation of the inverse problem but also avoid the computation of regularization parameters, plus we find that our results both satisfy the geophysical constraints and are also less affected by noise. We conclude with a discussion on the relevance of the parameters required by the methods with respect to the inversion performance, possible disadvantages of both approaches, and the potential applications of our constrained optimization approach.

## 2.2 Joint inversion formulation

In order to characterize the Earth's structure, we estimate the distribution of a physical property that affects seismic wave propagation, such as shear wave (S-wave) velocities, which we use for our forward computation. We can compute other properties like density, compressional (P-wave) velocity, and layer thickness by using empirical or analytical relationships. To study our joint inversion optimization approach, we use the following forward and inverse formulations.

From a given experiment that provides a layered shear velocity distribution  $x = v_s$ , we evaluate a forward linear or nonlinear operator  $F$  at the given velocity  $x$  to predict the Earth's response. Here  $F$  relates the data and the model space as follows:

$$F(x) = (F_1(x), \dots, F_m(x)) \in \mathbb{R}^m, \quad x = (x_1, \dots, x_n) \in \mathbb{R}^n \quad (m \gg n), \quad (2.1)$$

where  $m$  is the number of measurements or observations and  $n$  is the number of fixed thickness plane layers. Thus, if we know the velocity model in advance, we can predict the Earth's response corresponding to that model, by using the forward operator  $F$ . Given an observed data vector,  $y \in \mathbb{R}^m$ , the inverse problem consists of finding the unknown velocity model,  $x$ , such that  $F(x)$  approximates  $y$  as much as possible, i.e.,

$$\min_x \|F(x) - y\|^2 = \min_x \sum_{i=1}^m (F_i(x) - y_i)^2 \quad (2.2)$$

The inverse problem is generally posed as an unconstrained weighted nonlinear least squares (LSQ) problem. In our case, we expect to match simultaneously data from different geophysical

domains: receiver functions (RF) and surface wave dispersion velocities (SW). Therefore we reformulate (2.2) as

$$\min_x \frac{1}{2} \|F^{SW}(x) - y^{SW}\|^2 + \frac{1}{2} \|F^{RF}(x) - y^{RF}\|^2 + \frac{\lambda}{2} \|Lx\|^2, \quad (2.3)$$

where the first two terms are nonlinear  $l_2$ -misfit functions between the forward operators  $(F^{RF}, F^{SW})$  and the observations  $(y^{RF}, y^{SW})$ , corresponding to RF and SW respectively. The last term represents a regularization with an a priori smoothness parameter  $\lambda$  and a discrete derivative operator matrix  $L$  included to avoid sharp velocity changes in adjacent layers and to smooth velocity variations. For receiver functions,  $F^{RF}$  represents the numerical computation of synthetic waveforms [2], and for surface waves,  $F^{SW}$  represents the numerical evaluation of dispersion velocities [52].

For simplicity, we introduce:

$$F(x) = W \begin{bmatrix} F^{SW}(x) \\ F^{RF}(x) \end{bmatrix} \in \mathbb{R}^m, \quad y = W \begin{bmatrix} y^{SW} \\ y^{RF} \end{bmatrix} \in \mathbb{R}^m, \quad (2.4)$$

where we have a weighted diagonal matrix  $W$ ,

$$W = \text{diag}(w_i), \quad w_i = \sqrt{\frac{\eta}{\sigma_i^2 p}}, \quad i = 1, \dots, p, \quad w_i = \sqrt{\frac{1 - \eta}{\sigma_i^2 q}}, \quad i = p + 1, \dots, m, \quad (m = p + q),$$

$\text{diag}(w_i)$  denotes a diagonal matrix used to equalize the contribution of each data set with respect to physical units and number of data points,  $\eta \in [0, 1]$  is a data set influence parameter that measures the reliability of each data set,  $\sigma_i^2$  is the approximate standard deviation of each point, and  $p$  and  $q$  are the number of RF and SW observations respectively. In this work, we assume a typical uncertainty value  $\sigma_i^2$  of 0.05 (km/s) for SW and 0.01 (s) for RF observations [34]. Hence, we use (2.4) to rewrite (2.3) as follows,

$$\min_x \|F(x) - y\|^2 + \frac{\lambda}{2} \|Lx\|^2, \quad (2.5)$$

The addition of a priori information into the regularization term is a well-known technique to improve the condition of the inverse problem (e.g., [73]), sometimes mentioned as a LSQ method with damping [53]. However, choosing the best value for the parameter  $\lambda$  remains as an

open question. [30] and [75] refer to different techniques that can be used to properly select this parameter. In some cases, we require more than just the parameter  $\lambda$  to get reliable solutions; therefore the need of a stronger regularization term may arise. The discrete derivative operator matrices  $L_i$  as the ones below, allow us to introduce such type of regularization,

$$L_1 = \begin{bmatrix} -1 & 1 & & \\ & \ddots & \ddots & \\ & & -1 & 1 \end{bmatrix} \in \mathbb{R}^{(n-1) \times n}, \quad L_2 = \begin{bmatrix} 1 & -2 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & -2 & 1 \end{bmatrix} \in \mathbb{R}^{(n-2) \times n}.$$

These matrices constrain further velocities in adjacent layers, therefore introducing a priori information about model structure into the problem, by minimizing its roughness or as a tradeoff between data misfit and roughness [16]. The influence parameter,  $\eta$  can be used to test the compatibility of the data sets comprised in the inversion [34]. For instance, we show in Section 4 how different choices of this parameter affect the final inversion results. In this work, we use to perform single inversions; in other words, we run the algorithm in absence of one data set. For example, if  $\eta = 0$  we use SW dispersion velocities only. In the next subsections, we present two different methodologies to solve the inverse problem by using seismic shear velocities as the model parameter  $x$ .

## 2.3 Methodology

The inverse problem is usually posed in an unconstrained form as presented in (2.5). However, due to the usual highly nonlinear behavior of the operator  $F$ , the numerical complexity associated to the computation of higher order derivatives represents a major difficulty. Hence, we solve (2.5) iteratively as a linearized LSQ. To this end, we use a first order Taylor approximation of the operator  $F$  around some suitable model,  $x_k$ , this is  $F(x) \approx F(x_k) + F'(x_k)\Delta x_k = F(x_k) + F'(x_k)(x - x_k)$ , where  $F'(x_k)$  is the matrix with the partial derivatives of  $F$ . Therefore, we rewrite problem (2.5) as

$$\min_x \frac{1}{2} \|F'(x_k)x + r(x_k)\|^2 + \frac{\lambda}{2} \|Lx\|^2, \quad (2.6)$$

where  $r(x_k) = F(x_k) - y - F'(x_k)x_k$ . Even if  $F'(x_k)$  is not full column rank, for a given regularization term, there is a unique least squares solution that solves the symmetric positive

definite linear system:

$$\underbrace{(F'(x_k)^T F'(x_k) + \lambda L^T L)}_H x = -F'^T(x_k) r(x_k), \quad (2.7)$$

which are called the *normal equations*. From the well known algorithms to solve these equations [77], we use the *Truncated Singular Value Decomposition* (TSVD) of the system matrix  $H$ , which is also a method for regularization [30].

### 2.3.1 Truncated Singular Value Decomposition (TSVD)

The expression obtained using a TSVD to compute  $x$  can be written as

$$x = \sum_{i=1}^{\tau} \left( \frac{u_i^T r(x_k)}{\sigma_i} \right) v_i,$$

where  $U^T \Sigma V = U^T \text{diag}(\sigma_1, \dots, \sigma_{\tau}, \dots) V$  is the SVD factorization of the system matrix  $H$ ,  $U$  and  $V$  are orthogonal matrices of sizes  $m \times m$  and  $n \times n$ , respectively, and  $\Sigma = \text{diag}(\sigma_i)$  is a  $m \times n$  diagonal matrix with all singular values  $\sigma_i$  of the matrix  $H$ . Here,  $\tau$  denotes the numerical rank of the matrix  $H$  that leads us to a truncated regularized solution whenever a good choice of  $\tau$  improves the condition of the inverse problem, in other words, the ratio  $\sigma_1/\sigma_{\tau}$  becomes moderate. This approach is computationally intensive for large-scale systems, and lacks a specific algorithm to compute both the regularization parameter  $\lambda$  and the truncation factor  $\tau$ . The parameter  $\lambda$  usually represents the tradeoff between resolution and stability, by selecting the value that shows an optimal balance.

### 2.3.2 Primal-Dual Interior-Point (PDIP) Method

We propose a constrained optimization strategy to solve (2.6), the linearized version of the inverse problem (2.5). We remove the regularization term in (2.6), i.e., we set  $\lambda = 0$ , and instead we add appropriate physical bound constraints over the model parameter  $x$  in the inverse problem:

$$\begin{aligned} \min_x \quad & \frac{1}{2} \|F'(x_k)x + r(x_k)\|^2 \\ \text{s. t.} \quad & g(x) \geq 0, \end{aligned} \quad (2.8)$$

where

$$g(x) = \begin{bmatrix} x - c_{min} \\ c_{max} - x \end{bmatrix} \in \mathbb{R}^{2n},$$

This allows us to add physical bounds corresponding to a certain minimum and maximum velocities, i.e.  $c_{min} \leq x \leq c_{max}$ . We rewrite problem (2.8) in a standard nonlinear programming form as follows:

$$\begin{aligned} \min_x \quad & \frac{1}{2} \|F'(x_k)x + r(x_k)\|^2 \\ \text{s. t.} \quad & g(x) - s = 0 \\ & s \geq 0, \end{aligned} \tag{2.9}$$

where  $s \in \mathbb{R}_+^{2n}$  is a slack variable. Since we use primal-dual interior-point methods to solve (2.9), we need to define the Lagrangian function associated to (2.9):

$$\ell(x, z, s) = \frac{1}{2} \|F'(x_k)x + r(x_k)\|^2 - (g(x) - s)^T z, \quad (s, z) \geq 0 \tag{2.10}$$

with the  $z \in \mathbb{R}^{2n}$ , is the Lagrange multiplier associated to the inequality constraints. In these methods we use a perturbation parameter  $\mu > 0$ , to define the perturbed Karush-Kuhn-Tucker (KKT) (necessary conditions) as follows,

$$\begin{bmatrix} \nabla_x L(x, s, z) \\ \nabla_x L(x, s, z) \\ SZe - \mu e \end{bmatrix} = \begin{bmatrix} F'(x_k)^T (F'(x_k)x + r(x_k)) - \nabla g^T(x)z \\ g(x) - s \\ SZe - \mu e \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \tag{2.11}$$

where  $S = \text{diag}(s_1, \dots, s_{2n})$ ,  $Z = \text{diag}(z_1, \dots, z_{2n})$  and  $\nabla$  denotes the gradient operator. Next we define the following Newton system associated to (2.11),

$$\begin{bmatrix} F'(x_k)^T F'(x_k) & -\nabla g^T(x) & 0_{n \times m} \\ \nabla g(x) & 0_{m \times m} & -I_{m \times m} \\ 0_{m \times n} & S & Z \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta z \\ \Delta s \end{bmatrix} = - \begin{bmatrix} \nabla_x \ell(x, z, s) \\ g(x) - s \\ SZe - \mu e \end{bmatrix}. \tag{2.12}$$

Here the system's matrix is called the *Jacobian*, and the last block of equations in this linear system is known as the *complementarity conditions*, a fundamental ingredient of primal-dual interior-point methods [23]. We use a methodology known as *path-following strategy* to

solve (2.12), as defined in [51]. In this methodology, there are two main components: 1) forcing positivity of the iterates  $x_k$  to keep them in the interior, and 2) a globalization strategy to guarantee descent directions and progress towards the constraints. For a perturbation parameter  $\mu > 0$ , and working from the interior, i.e., keeping the condition  $(g(x), z, s) > 0$ , we apply a linesearch Newton's method [57] to the perturbed KKT conditions (2.11), which leads us to the reduced linear system:

$$\begin{bmatrix} -F'(x_k)^T F'(x_k) & \nabla g^T(x) \\ \nabla g(x) & Z^{-1}S \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta z \end{bmatrix} = \begin{bmatrix} \nabla_x \ell(x, z, s) \\ Z^{-1}\mu e - g(x) \end{bmatrix}. \quad (2.13)$$

We solve for  $\Delta v = (\Delta x, \Delta z)$ , and then we repeat the process until an optimal solution  $x$  is obtained as the perturbation parameter  $\mu$  goes to zero. Note that unlike the non-symmetric and usually highly indefinite system (2.12), the reduced system (2.13) is symmetric with its size reduced considerably, and more likely to be better conditioned. For large-scale problems, the use of direct methods like TSVD to solve (2.13) or (2.7) can be computationally intensive. Instead, we can exploit the structure of (2.13) to solve it iteratively using a Krylov space method, e.g., implementing a variant of the conjugate gradient (CG) algorithm [63].

The PDIP algorithm is initialized with an (interior) velocity model  $x_0$ , i.e.  $(g(x), z, s) > 0$ . Then we begin an iterative process until the  $\ell_2$ -norm of the right hand side in (2.13) becomes smaller than certain tolerance value  $\epsilon > 0$ ; this is  $\|(\nabla_x \ell(x, z, s), Z^{-1}\mu e - g(x))\| \leq \epsilon$ . At each iteration, we compute the perturbation parameter  $\mu_k = \frac{s_k^T z_k}{2n}$ , and then solve for  $\Delta v_k = (\Delta x_k, \Delta z_k)$  from the system (2.13) to set  $\Delta z_k$ . To enforce positivity, once we compute the Newton step from (2.13), we correct it with a factor  $\tilde{\alpha}_k \in (0, 1]$  to guarantee that the inequalities are strictly positive; we accomplished this task by using the following formula [23, 57]:

$$\hat{\alpha} = 0.9995 \min \left\{ \frac{-1}{\min\{\frac{\Delta x}{g(x)}, -1\}}, \frac{-1}{\min\{\frac{\Delta z}{z}, -1\}} \right\}.$$

Once we have computed the enforce positivity factor  $\hat{\alpha}$ , in order to monitor progress to an optimal solution, we use a linesearch globalization strategy that requires a merit function  $M$ . This function has as role to determine a step length  $\alpha = (0, \hat{\alpha}]$  that provides sufficient decrease of the objective function, while satisfying the inequality constraints. This is possible whenever

the Newton direction results to be a descent direction for the merit function  $M$  at the current interior point  $v_k = (x_k, z_k)$ . We address this issue by implementing the Armijo condition as shown in [57], where the step length,  $\alpha$ , can be chosen such that:

$$M(v_k + \alpha \Delta v_k) \leq M(v_k) + 10^{-4} \alpha \nabla M(v_k)^T \Delta v_k,$$

where  $\Delta v_k = (\Delta x_k, \Delta z_k)$ . In the simplest case we can select the objective function as the merit function  $M$ , however we choose the modified augmented Lagrangian function introduced by [51] as a merit function due to its robustness:

$$M(x, z, s; \theta) = \ell(x, z, s) + \theta \phi(x, z),$$

where  $\theta$  is a nonnegative penalty parameter, and the second term is the penalty function to the complementarity condition,

$$\phi(x, z) = g(x)^T z - \mu \sum_{i=1}^{2n} \log(GZe). \quad (G = \text{diag}(g(x)))$$

It can be shown that the Newton direction is a descent direction for the penalty function, therefore we can find a  $\theta$ , such that the Newton direction is a descent direction for the merit function, i.e.,

$$M(v_k)^T \Delta v_k < 0.$$

Finally, we compute our update  $v_k = v_k + \alpha \Delta v_k$  and repeat the process again until we either reach convergence, e.g., until there is no significant difference between the iterates  $\|x_{k+1} - x_k\| / \|x_k\| \leq 10^{-5}$ .

## 2.4 Numerical Experimentation

We briefly describe the two synthetic data sets used in the inversion: receiver functions and surface waves. In general, receiver functions can be used to resolve discontinuities (impedance contrasts) in seismic velocities, and provide good measurement of crustal thickness, without providing a good average of shear wave velocity. Surface (Love and Rayleigh) waves provide a good average of absolute shear wave velocity, without good shear wave velocity contrasts in

layered structures (e.g., [34]). Therefore, these two data sets can be considered as complementary and consistent, as long as we sample the same medium while assuming similar geological boundaries. Since both data sets are mainly sensitive to shear wave velocity structure [34], we can assume that the forward operator  $F$  depends nonlinearly in our model parameter: the shear velocities  $x_i$ ,  $i = 1, \dots, n$ , that represent the different velocities of a halfspace with  $n$  horizontal layers.

### 2.4.1 Receiver Function

Receiver functions give a good measurement of depth velocity discontinuities and crustal thickness. Different techniques can be used to compute receiver functions, for instance the spectral water level deconvolution technique [44] and the time domain iterative deconvolution technique [48], which we implemented in this work. The resulting receiver function is a time series that can be viewed as a linear combination of delta functions, in which major negative (or positive) spike amplitudes correspond to a decrease (or increase) of seismic velocity, respectively. The time separation  $t_{Ps}$  between  $Ps$  and  $P$  phases can be used to estimate crustal thickness  $h$  from the average crustal velocities [83], as

$$h = \frac{t_{Ps}}{\sqrt{\frac{1}{v_s^2} - p^2} - \sqrt{\frac{1}{v_p^2} - p^2}},$$

where  $p = \frac{1}{v_p}$  is the ray parameter or horizontal slowness of the incident wave, and  $(v_p, v_s)$  are compressional and shear velocities respectively. Since  $t_{Ps}$  represents the differential travel time of S-waves with respect to P-waves in the crust, the dependence of  $h$  on  $v_p$  is not as strong as on our model parameter  $x = v_s$  (or, more precisely, on the ratio  $v_p/v_s$ ). Therefore, we consider receiver functions to be more sensitive to shear wave velocity contrasts. On the other hand, crustal thickness estimated only from the delay time of the Moho- $Ps$  converted phase tradeoff strongly with the  $v_p/v_s$  ratio. Since the average velocity cannot be resolved clearly, we need information from the surface waves to better constrain the shear velocities.

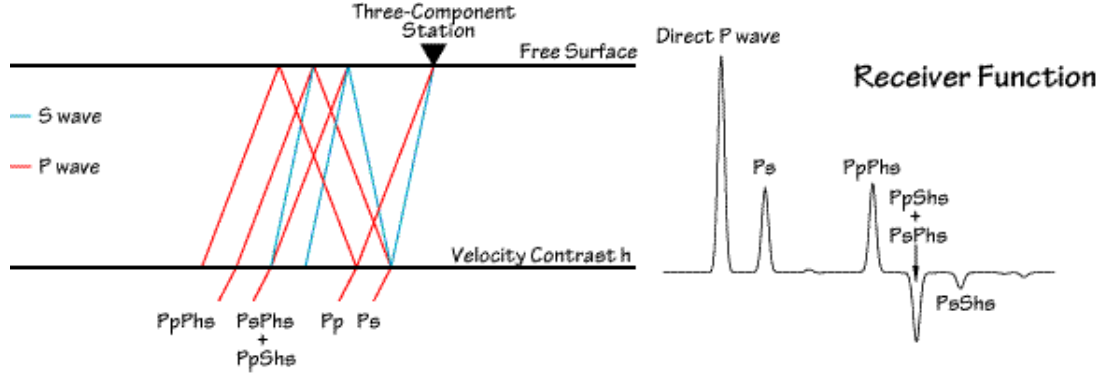


Figure 2.1: (Left) Waveform of a Receiver Function. The halfspace  $h$  represents usually the Moho. (Right) The Moho converted phase  $Ps$  and the multiples  $PpPs$ ,  $PpSs$ , and  $PsPs$  are labeled, and their ray paths. With the exception of the first arrival, lowercase letters denote upgoing travelpaths, uppercase letters denote downgoing travel paths [69].

## 2.4.2 Surface Wave Dispersion

In general, surface waves dominate seismograms as the largest amplitude waves from an earthquake, and are observed at lower frequencies than the body waves. Furthermore, surface wave velocities vary depending on the depth range sampled by each period, making the surface wave dispersion valuable for studying Earth's structure. In general, dispersion curves are extracted from three component seismograms recorded at the same station which correspond to different frequencies and distances, by using reduction algorithms that rely on spectral analysis techniques (see, e.g., [52]). Based on Rayleigh's principle, surface wave velocities are more sensitive to S-wave velocity, although they are also theoretically sensitive to P-wave velocity and density [34]. This principle can be used to show that the phase velocity perturbation  $\delta c/c$  can be viewed as a function of the sensitivity coefficients for P-wave velocity, S-wave velocity, and density. By investigating sensitivity function variation with respect to depth, the relative contribution of P-wave velocity and density to dispersion can be shown to be smaller than the one for S-wave velocity [34]. Therefore, surface wave dispersion is more also sensitive with respect to S-wave velocity.

### 2.4.3 Synthetic Models

We implemented the joint inversion algorithm for four different Earth’s crustal synthetic models. We divided each model in layers of 1 km thickness. For each synthetic velocity model, we compute two different dispersion curves, Rayleigh and Love, for either phase or group velocities. For each curve we sample 35 points or periods from 10 to 70 seconds, therefore our velocity model needs to have at least 70 km in depth. Simultaneously, we compute receiver functions for three different ray parameters  $p$ , with 2048 data points each. We create both data sets synthetically and noise free, by using standard algorithms (Hermann, 2002), based on the crustal models obtained from CRUST 2.0 (Laske et al., 2000). Figure 2.2 shows a synthetic crustal (Rift) model used to create three receiver functions corresponding to three different ray parameters ( $p = 0.03, 0.05, 0.07$  s/km), and the surface wave dispersion curves corresponding to group and phase dispersion curves, for both Rayleigh and Love waves.

## 2.5 Test Results

We solved the inversion problems (2.6) and (2.8) by using the TSVD and PDIP method respectively, when inverting for the shear wave velocity  $x = v_s$  (km/s) as the model parameter. For both data sets  $P$ –wave velocity can be inferred by assuming a constant value of the ratio  $v_p/v_s$ , as  $v_p = (v_p/v_s)v_s = (v_p/v_s)x$ . We estimate the density  $\rho$  (g/cm<sup>3</sup>) from the resulting  $v_p$  velocity through some empirical relation, for example  $\rho = 0.32v_p + 0.77$  introduced by Birch (1961). In this work, we do not attempt to quantify uncertainty. Thus, we assume certain typical uncertainty values  $\sigma_i^2$  for both data sets: 0.05 (km/s) for SW and 0.01 (s<sup>-1</sup>) for RF observations.

To illustrate how receiver functions and surface wave dispersion velocities complement each other, we present the inversion results for the data sets created from the rift velocity model (Figure 2.3). In each case, the two figures on the left represent a single inversion, this is, an inversion involving only one data set ( $\eta = 0$  or  $\eta = 1$ ). We performed these inversions to show the resolution that each data set can provide for a particular target velocity model, and to expose non-uniqueness of the inverted model. The joint inversion of both data sets (right)

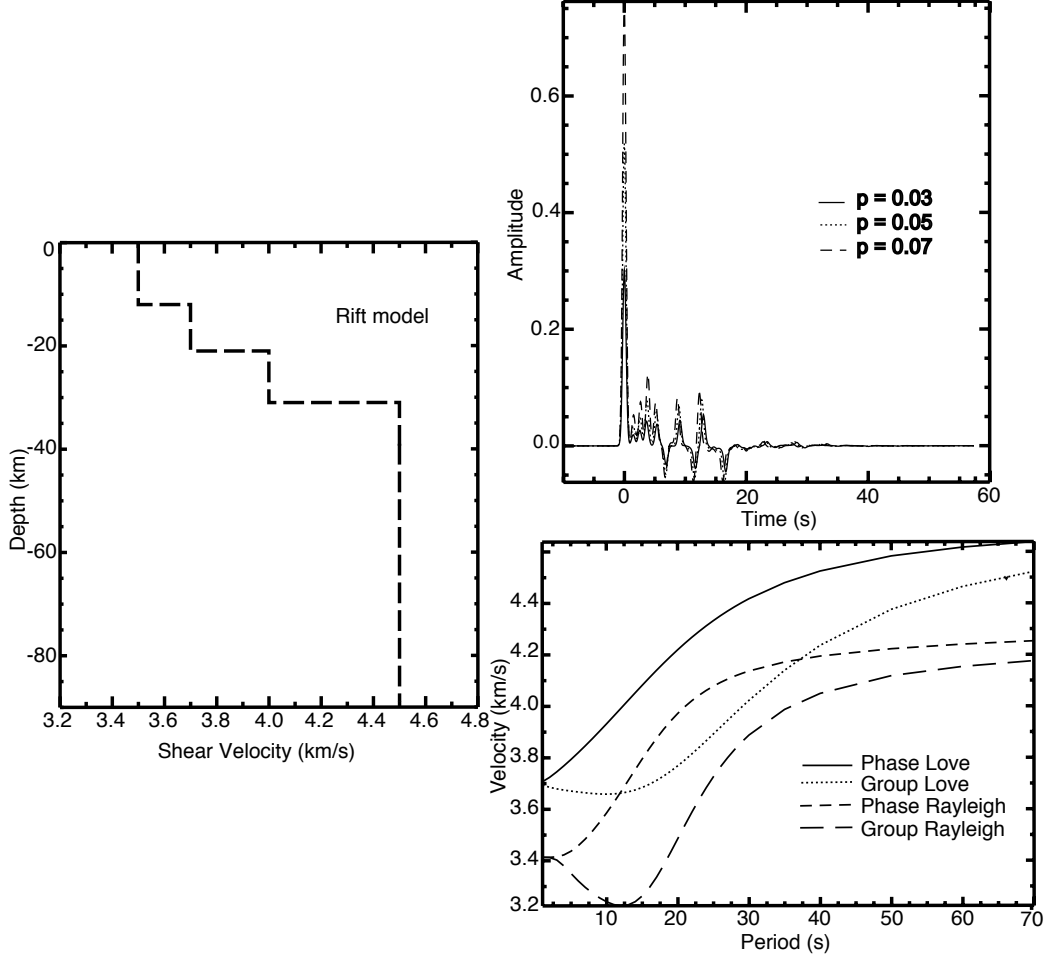


Figure 2.2: (Left)  $S$ -wave synthetic velocity model (Rift) (Laske et al. 2000). (Right: top) Receiver functions for three different ray parameters  $p$ , and (Right: bottom) surface wave dispersion curves (Love and Rayleigh) computed for the velocity model on the left.

provides a better approximation to the target model as expected.

For each 1-D velocity model, left sides of Figures 2.4 – 2.7, show from top to bottom, show the relative error (RMS) with respect to the target model  $\frac{\|x_k - x^*\|}{\|x^*\|}$ , the residual error between the predictions and the observations  $\frac{\|F(x_k) - y\|}{\|y\|}$ , and the average shear velocity (the black line represents average velocity of the model). We use the average shear velocity as a measure of how close the estimated models are to the true average velocity. This test can be used to check for spurious solutions. On the right of each figure, we show the target, initial,

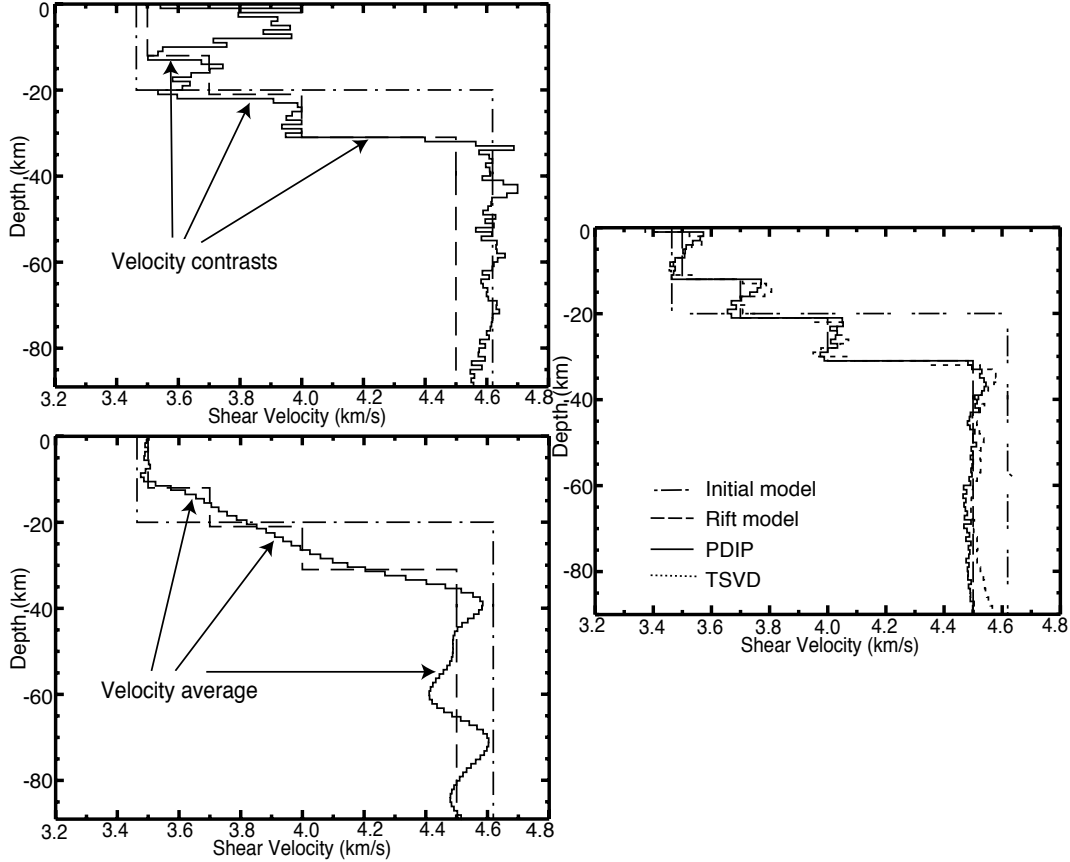


Figure 2.3: The single inversion of receiver functions (left top) identifies velocity contrasts, while single inversion of surface waves (left bottom) gives information on the average velocities at different depths. The joint inversion (right) of these two data sets combines all this information and provides a substantial improvement in the final estimated model.

and estimated models given by each method: TSVD estimated models in dashed lines, and PDIP estimated models in solid lines. We show the best velocity model approximation for both methods with all models. When using the PDIP method, the upper and lower bounds over all the models  $x$  are defined as  $c_{\min} = 3.2$  (km/s) and  $c_{\max} = 4.8$  (km/s). Recall that the objective of the numerical experimentation is to have the smaller relative error (RMS), which means that our approximation gets closer to the target model. We set the maximum number of iterations to six for all the inversions and for both methods. As we show below, the number of iterations needed for both methods to converge is sometimes smaller than six; however, we

continue the inversion until the sixth iteration, to be able to compare between the methods.

### 2.5.1 Velocity Model Tests

We begin with an initial guess of the velocity model  $x_0$  which is usually based on a priori geophysical knowledge. In our case, since we know the synthetic target models  $x^*$  we have the freedom to select an appropriate initial model; this is often possible. We use the first synthetic model (Archean) as the initial guess for the next two models, Continental and Oro-gen, and a two half-space model for the Archean and the Rift models. Once we compute the observation vectors  $y^{SW}$  and  $y^{RF}$ , we iterate by evaluating the forward problem for the different approximations  $x_k$ , and solve the inverse problem either with the TSVD method for the unconstrained formulation (2.6) or the primal-dual interior-point method for the constrained formulation (2.8). Once the selected method returns the updated model  $x_{k+1}$ , we check for convergence by requiring the residual error to be smaller than a certain tolerance  $\epsilon > 0$ , i. e.,  $\frac{\|F(x_k) - y\|}{\|y\|} \leq \epsilon$ . If it is not smaller we continue the iteration until the process either converges or reaches the maximum number of iterations. In our case, six iterations are enough to provide an acceptable reduction in the RMS value (as usual with most inversion methods that use an optimization framework).

We plot results for the different models in Figures 2.4 – 2.7. We observe that for all the models except for the Archean (Figure 2.4), the relative error and the number of joint inversion iterations required for convergence results is smaller for PDIP than for TSVD. However, the accuracy obtained by the approximated Archean model with PDIP methods is still comparable with the one provided by TSVD at the final iteration. Therefore, we see our approach for solving the inverse problem with the PDIP method to be as good as the traditional unconstrained regularized inversion that uses TSVD.

### 2.5.2 Parameter Selection Tests

We performed a variety of experiments to test the best possible selection of parameters for both types of inversion methods. We use the previously presented Rift model as the test model

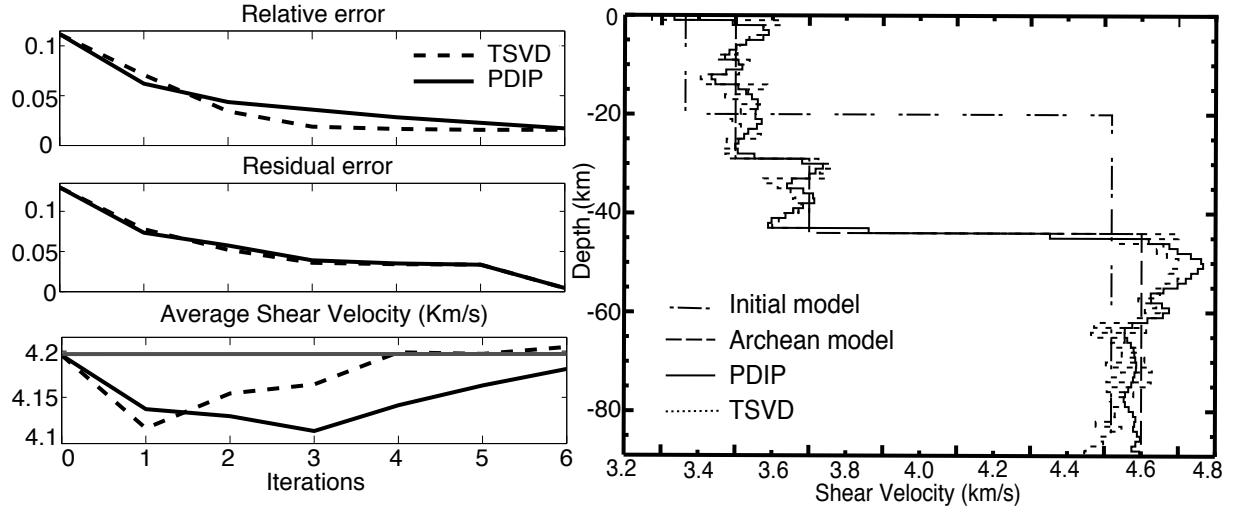


Figure 2.4: Archean model. Notice that the residual errors are about the same at each iteration, with faster convergence for TSVD method. This means that the TSVD method reaches minimal RMS error faster than the PDIP method.

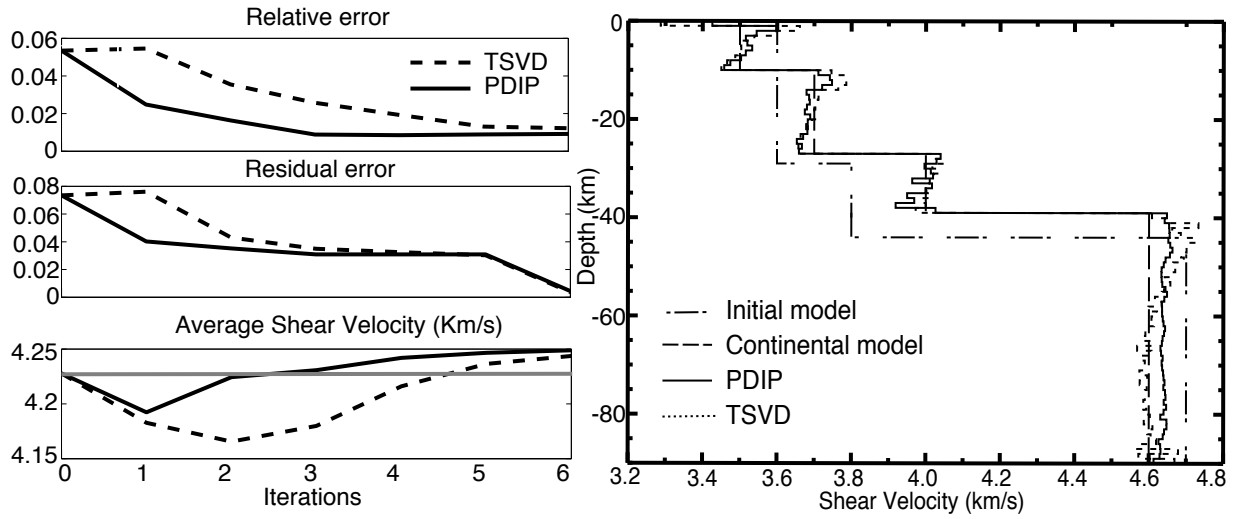


Figure 2.5: Continental model. The relative RMS and residual errors are close at the last two iterations for both methods, but PDIP minimize better the RMS and converges faster than TSVD. Both average velocities come closer to the true average velocity at the end of the inversion.

for all of the experiments in this subsection. Different choices of the influence parameter  $\eta$  can help us to improve resolution in the final model. However, this process usually becomes

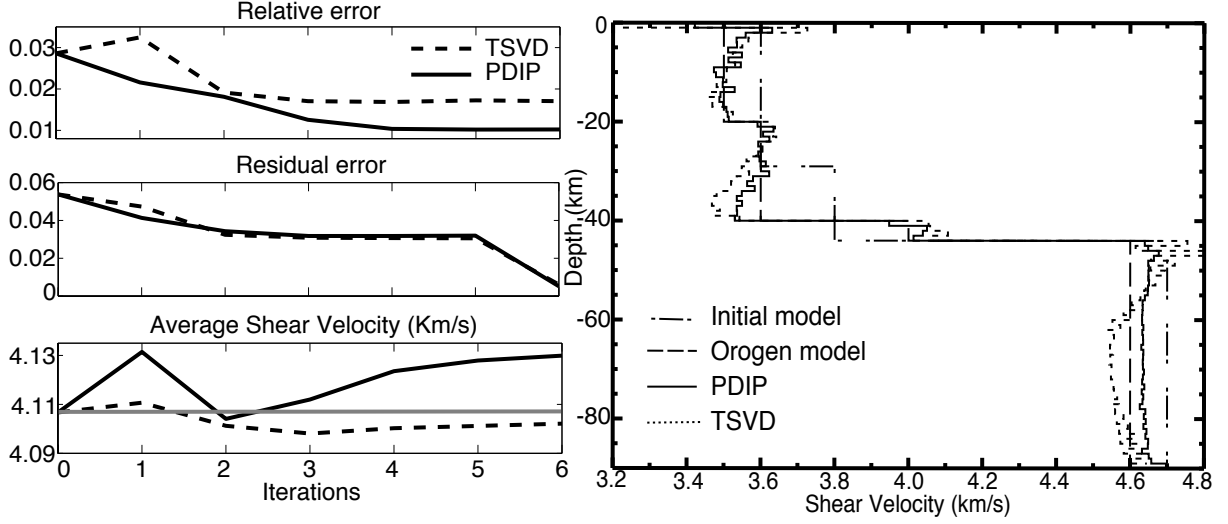


Figure 2.6: Orogen model. The relative error is smaller for PDIP at all stages of the algorithm, while the residual error is similar for both methods. The TSVD estimated velocity is closer to the true average velocity.

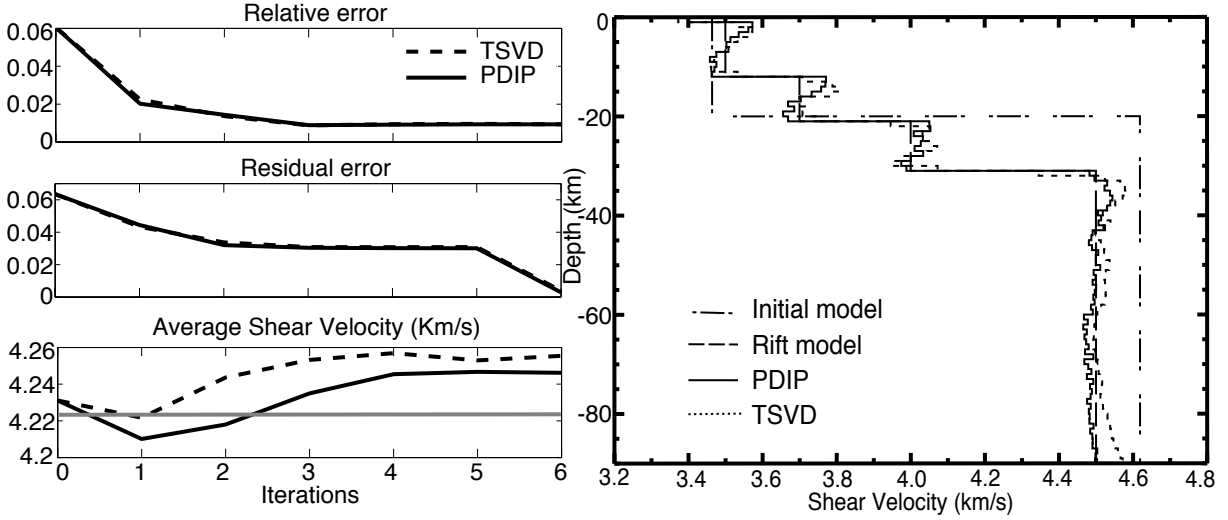


Figure 2.7: Rift model. The relative error and the residual error are within the same order of accuracy for both methods. The average is better resolved for PDIP until the 3rd iteration, where the smaller relative RMS error is reached.

expensive, since it requires several suites of inversions, and an automatic procedure to obtain such parameter does not currently exist. Thus, we restrict the discussion to a combination of

the two data sets when  $\eta = 0.5$ ; in other words we assume that both data sets are equally reliable for the inversion. In Table 2.1, we show how the relative and residual error change as we select different values for  $\eta$ .

Table 2.1: Relative RMS and residuals errors associated to different choices of the influence parameter  $\eta$  for the joint inversion. Our inverted Rift model was obtained by using PDIP as the inversion method. Note that the best choice for the influence parameter is  $\eta = 0.75$ .

$\eta$	RMS	Residual Error
0	$3.2670 \times 10^{-2}$	$2.3046 \times 10^{-1}$
0.25	$7.9577 \times 10^{-3}$	$1.1330 \times 10^{-2}$
0.5	$7.4820 \times 10^{-3}$	$3.0111 \times 10^{-3}$
0.75	$6.6809 \times 10^{-3}$	$1.6802 \times 10^{-3}$
1.0	$1.6308 \times 10^{-2}$	$7.6348 \times 10^{-4}$

## TSVD Parameters

We test different values of the regularization parameter  $\lambda$ . As  $\lambda$  moves away from zero, the regularization becomes too strong, thus introducing excessive smoothing to the final approximated model. The results appear to be similar to inverting the surface wave data alone (see Figure 2.2). The excessive regularization may blur the capabilities of the receiver function to identify large velocity contrasts. When  $\lambda$  approaches zero, however, the algorithm becomes unstable; hence, we may not achieve convergence to a meaningful 1-D velocity model. In all the tests, we set the truncation parameter to  $\tau = 0.0$ , since as we show in the next section, this choice appears to be the best for our experiments. The best choice we obtained from the suite of inversions for  $\lambda \in 0.001, 0.01, 0.1, 0.5, 1, 10$  consisted on setting  $\lambda = 0.1$ ; for this value we obtain the smaller RMS error - which is the inversion's goal - while having a good fit to the data. We present our results for relative and residual error in Figure 2.8.

The value of  $\tau$  is generally selected as a threshold to exclude those singular values close to

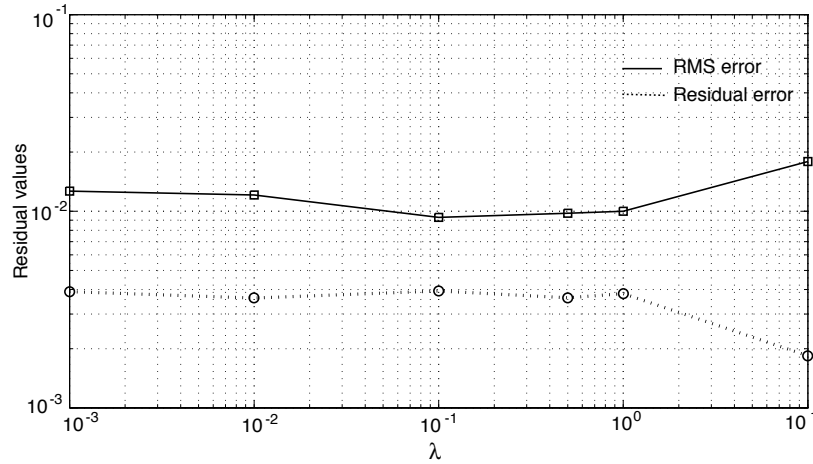


Figure 2.8: Logarithmic scale plot of the relative RMS and residuals errors associated to different choices of the regularization parameter  $\lambda$ . We applied the joint inversion algorithm over the Rift model when solving the inverse problem by using the TSVD method.

machine precision. To test the effect of the truncation parameter over the inversion results, we use a similar strategy by defining  $\lambda = 0.1$ , our best value for  $\lambda$ , to be used for all the experiments and choosing  $\tau$  from the set  $\{0.0, 10^{-5}, 10^{-3}, 10^{-2}, 10^{-1}\}$ . The choice of  $\tau = 0.0$  works best and similar results were obtained up to  $\tau = 10^{-2}$ . However if  $\tau > 10^{-2}$  then we do not have convergence. We summarize our numerical findings in Table 2.2.

Table 2.2: Relative RMS and residuals errors associated to different choices of the truncation parameter  $\tau$  for the joint inversion over the Rift model when solving the inverse problem by using the TSVD method.

$\tau$	RMS	Residual Error
0.0	$9.3036 \times 10^{-3}$	$3.9254 \times 10^{-3}$
$10^{-5} \leq \tau \leq 10^{-2}$	$\approx 9.3036 \times 10^{-3}$	$\approx 3.9254 \times 10^{-3}$
$5 \times 10^{-2}$	$2.0526 \times 10^{-1}$	$3.5542 \times 10^{-1}$
$10^{-1}$	--	--

In the case of implementing regularization with no truncation, i.e., for values of the trunca-

tion parameter  $\tau > 10^{-2}$ , we cannot achieve convergence. The high frequency components of the singular values introduce strong variations into the calculated receiver function waveform after the third iteration. As a result, the receiver function is distorted, and the surface waves deviates considerably from the first approximations, no longer matching the true synthetics. With truncation applied but no regularization, we still can achieve convergence, but with less accuracy in relation to the true Earth model (Figure 2.9). Hence, we realize the relevance of the truncation parameter  $\tau$  to achieve convergence compared to the regularization parameter  $\lambda$  mainly intended to smooth out the final approximations. When neither truncation nor regularization are applied, we do not have stability or convergence of the algorithm.

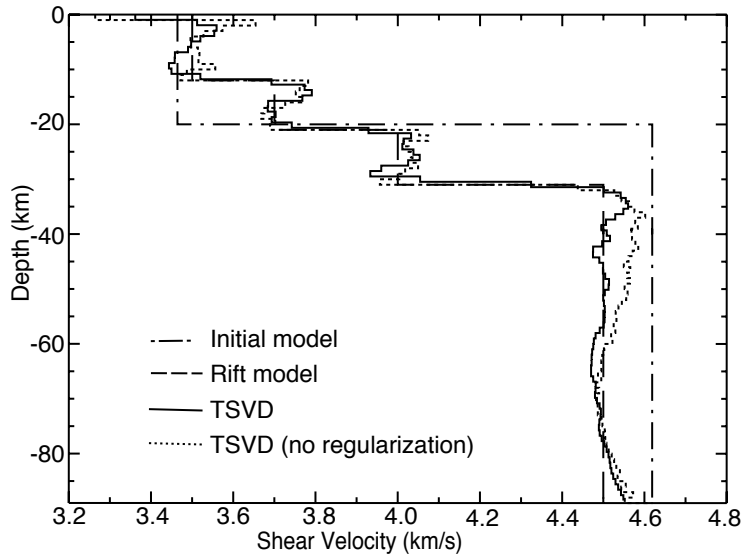


Figure 2.9: Numerical results obtained for the TSVD method when we use truncation without regularization.

## PDIP Parameters

In Section 2.3.2, we showed how to compute the perturbation parameter  $\mu$ , and how to compute the Lagrangian multipliers  $(z, s)$  associated to the PDIP method. We also wish to test the impact that different choices of PDIP bounds  $c_{\min}$  and  $c_{\max}$  may have over the performance of the method. We use the following pair choices  $(c_{\min}, c_{\max}) \in \{(0, 10), (3, 5), (3.4, 4.8)\}$ . We realized that as the lower and upper bound approach the true maximum and minimum values

of our test rift Earth model, the performance of the method improves considerably, as presented in Table 2.3. The improvement is achieved by reducing the number of iterations to converge, and by diminishing the relative RMS and residuals error.

Table 2.3: Relative RMS and residuals errors associated to different choices of the lower an upper velocity bounds over our model  $x$ . The joint inversion for the Rift model was solved by using the PDIP method. The last row shows the results when adding regularization with  $\lambda = 0.1$ .

$c_{\min}$	$c_{\max}$	RMS	Residual Error
0.0	100.0	1.9327	$9.6538 \times 10^{-2}$
3.0	5.0	$1.2789 \times 10^{-2}$	$3.9357 \times 10^{-3}$
3.4	4.8	$9.3143 \times 10^{-3}$	$2.8394 \times 10^{-3}$
3.4	4.8	$9.5301 \times 10^{-3}$	$2.8411 \times 10^{-3}$

We tested also the effect of the regularization term as presented for the unconstrained case function. This implies an additional matrix added to the first block of the reduced system matrix in equation (2.13). The impact of this regularization over the inversion results does not represent an improvement with respect to the original results obtained for PDIP without that term (see last row in Table 2.3).

### 2.5.3 Noise Synthetic Tests

We added random Gaussian noise to the data vector  $b$  in order to check robustness of the methods. We increase the level of noise to detect where the methods break down. To this end, we use the following percentages of noise: 5%, 10%, 15%, 20%, 25%, and 30% noise. In all the cases we use the best parameter set up for both methods, i.e.,  $\lambda = 0.1$  and  $\tau = 0.0$  for TSVD, and  $c_{\min} = 3.4$  and  $c_{\max} = 4.8$  for PDIP.

We display the numerical results in Figures 2.10 – 2.13.

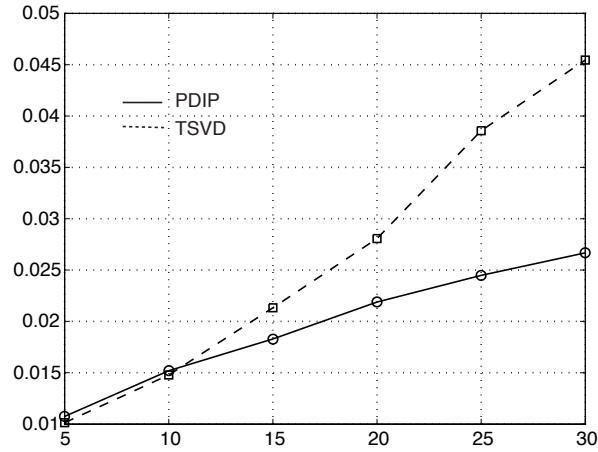


Figure 2.10: Relative RMS errors associated to different levels of Gaussian noise added to the data. As we will see in Figure 2.13, PDIP offers a steady behavior of the quality in the inverted models, i.e., they seem to be less affected by noise as it increases.

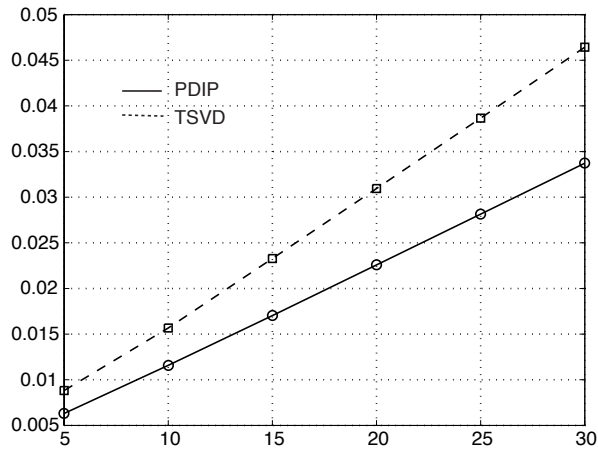


Figure 2.11: Relative errors associated to different levels of Gaussian noise added to the data. Both methods give comparable results, being PDIP more accurate to fit the data.

## 2.6 Discussion

We expanded a joint inversion algorithm [34] to implement the truncated SVD (TSVD) or the primal-dual interior-point (PDIP) method for solving an unconstrained or a constrained formulation of the inverse problem (2.3), respectively. We compared these two methodologies

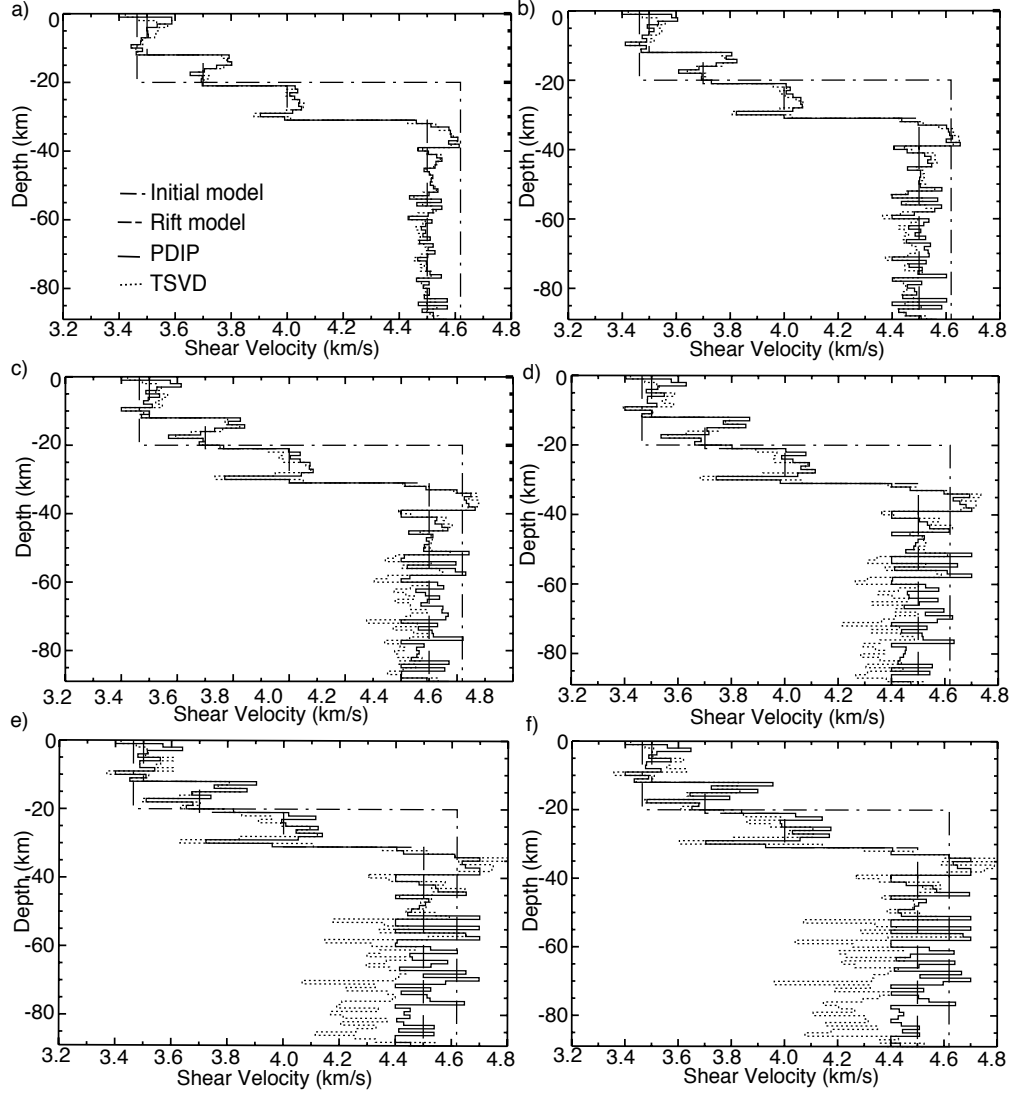


Figure 2.12: Rift velocity model approximation obtained for several levels of noise. On top: a) represents 5% noise and b) 10% noise added to the data. Both methods reach about the same level of convergence. The plots in the middle corresponds to c) 15% and d) 20% noise. The effect of such level of noise shows up considerably in the lower layers, still the interfaces are well characterized. Finally in the bottom with e) 25% and f) 30% Both methods fail to identify properly the lower layers, being TSVD more affected by the noise.

by running the joint inversion algorithm for four synthetic velocity models. As expected, the geophysical information provided by receiver function and surface waves dispersion data

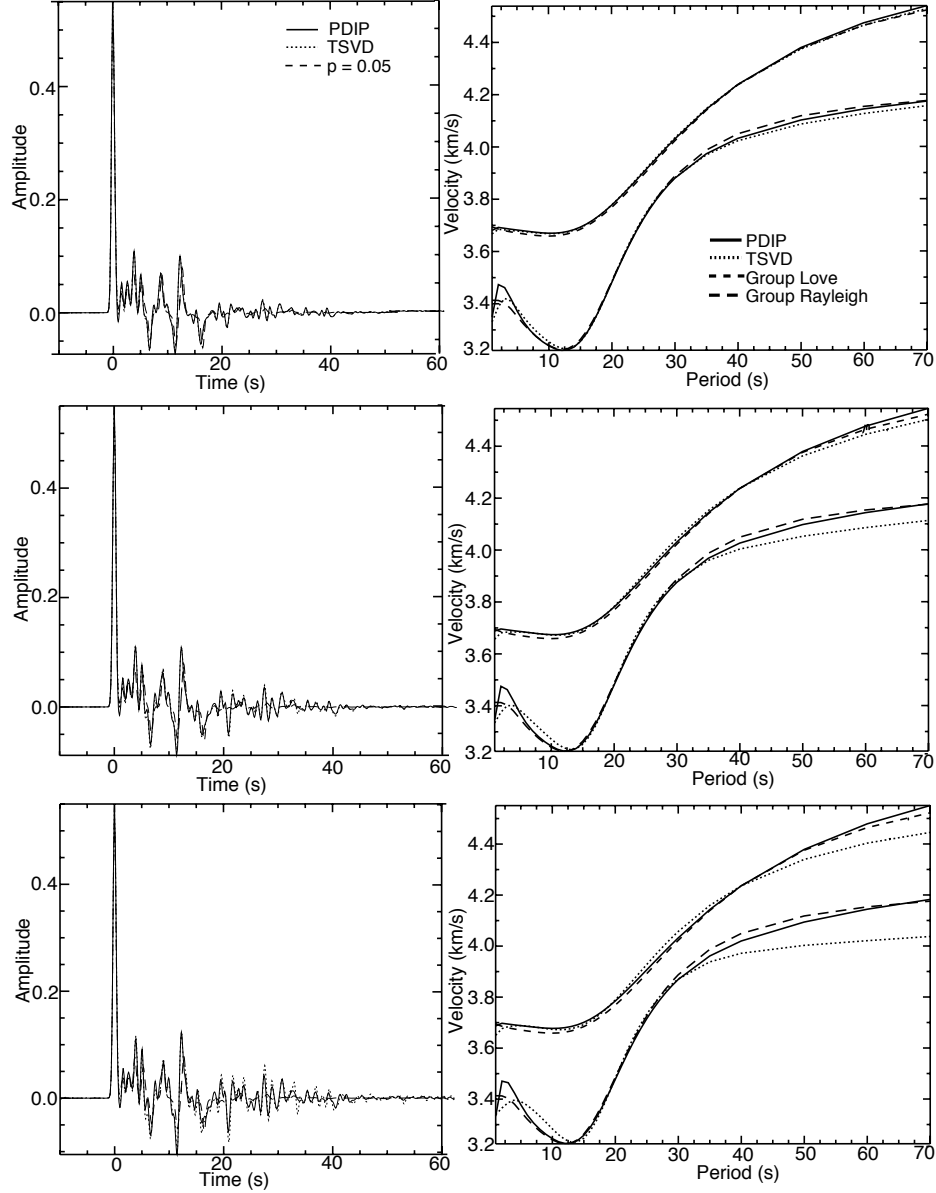


Figure 2.13: Receiver function and surface waves results for increasing noise percentage. From top 10%, 20% in the middle and bottom 30% over the Rift model. The fitness of both methods is better for the RF synthetics, while for SW we notice how the approximations begin to degrade considerably more for the TSVD estimated synthetics than for the PDIP ones.

complement each other. Therefore, we obtained an improvement in the final approximation to the target model, in comparison to single inversions (see Figure 2.3).

The PDIP method does not require the use of a regularization parameter  $\lambda$  or a truncation parameter  $\tau$ . Instead, we include a priori information by adding physical bounds over the model parameter  $x$  as an explicit constraint to reduce the model space, thus avoiding spurious solutions. Also, these constraints make the algorithm more robust to address the noisy components that may be associated to the data. We believe this is an advantage, and a major difference with respect to traditional unconstrained formulations [29, 34, 35, 52, 55, 15] in which the constraints are all included into the objective function, and also from the constrained formulation of [26], [6], and [49], which relies on a quadratic programming approach with expensive simplex-type schemes or a Lagrange multipliers method without any globalization strategy [54].

Our rationale proceeds from a possible general formulation in the unconstrained case,

$$\min_x f(x) + \sum_i \lambda_i R_i(x), \quad (2.14)$$

in which  $f$  is usually a misfit function, e.g.  $f(x) = \|F'(x_k)x + r(x_k)\|$ : this formulation relies in the capability of the terms  $\lambda_i R_i(x)$  to introduce a priori information to regularize the objective function. In our approach, we attempt to use an explicit formulation of the constraints as in (2.8); we thus avoid the expensive computation of the regularization parameters  $\lambda_i$ .

The computational complexity in terms of floating point operations (flops) is improved using PDIP. For TSVD, the computational complexity is  $O(n^3)$ , while for PDIP it is  $O(n^2)$  at each Newton step when using the Krylov method. Since we do not exceed 15 Newton iterations at any stage of the joint inversion, the TSVD method may not be suitable for large-scale problems that arise when we increase the size of the problem by requiring much bigger data sets or higher-dimensional Earth models. However, for PDIP methods, this setup only requires exploration for appropriate velocity bounds, which might need to be changed with respect to depth, and a new set of initial values for the Lagrangian multipliers.

Future work includes having an uncertainty quantification study to ensure the quality of our results when using real data. Also, it is important to try a joint inversion of real geophysical data recorded from the Rio Grande Rift region, and the inclusion of other data sets that have

shown to be compatible, i.e., magnetic and gravity data [52, 56]. If the data is insufficient to overcome the ill-posedness of the inverse problem, or is not compatible [55], we expect that by adding explicit constraints directly into the inverse problem formulation, we will enforce structural similarity [29, 26] to further stabilize our joint inversion at no extra cost of computing the regularization parameters  $\lambda_1$ .

## 2.7 Conclusions

We present a constrained formulation for joint inversion of two synthetic geophysical data sets solved by using primal-dual interior-point (PDIP) methods, and compare its performance with respect to the traditional TSVD method for regularization used for solving the unconstrained formulation of the inverse problem. We show that the addition of explicit inequality constraints instead of regularization terms into the objective function, results effective to handle the introduction of a priori information, and to regularize the inversion. This feature helps us to reduce ambiguities and the influence of noise over the inversion results. Our work explores an approach well known by the optimization community but not usually implemented by the geophysicists to solve this type of inverse problems. We believe that our approach stands as a good alternative when looking for a robust inversion method, since it presents a savings in computational costs, no need for searching for the best regularization parameters, (as required, for example, by the TSVD method), and also the smallest effect of noise in the data.

# Chapter 3

## 3-D Structure of the Southern Rio Grande Rift from 1-D Constrained Joint Inversion of Receiver Functions and Surface Waves Dispersion

### 3.1 Introduction

The Rio Grande Rift (RGR) extends approximately 1000 km from central Colorado to El Paso, Texas (Figure 3.1). Many studies have focused on the rift system; see, e.g., [39, 76, 80, 61]. The southern terminus of RGR appears poorly defined [37], and few seismic studies provide information on the deeper rift structure [39, 5]. The RGR has recent volcanism, fault scarps, and seismicity (Figure 3.1) and is widening at a small rate of about 0.5 mm/yr or less [81, 43]. Important questions about the Rift evolution remain unresolved: is it actively deforming along its southern extent [37]?, what is the role of mantle convection in the formation of the Rift?, does it propagate southwest?, and how does it influence the evolution of adjacent areas within the North American Plate [61]? Previous studies in the RGR present several possible earth models and interpretations: the difference is to the diversity of the methodologies implemented and the location of the studies. To address the abovementioned questions in the southern RGR, we need reliable information and also a robust inversion/imaging method for an integrated analysis to develop three-dimensional (3-D) models of Earth structure.

In this Chapter, we define the crustal and upper mantle structure of the southern RGR using the recently available EarthScope Transportable Array (USArray) data set

([www.earthscope.org](http://www.earthscope.org)) along with other collected data sets. In particular, we apply a constrained optimization 1-D joint inversion approach [68] using receiver functions for 147 USArray and the LA RISTRA stations [80]; [71], and a high quality surface waves dispersion data set provided in ([http://www.eas.slu.edu/eqc/eqc\\_research/NATOMO](http://www.eas.slu.edu/eqc/eqc_research/NATOMO)). We then interpolate the 1-D models into a 3-D volume, and find previously known structures beneath the region while giving a different perspective to the interpretation of some anomalies.

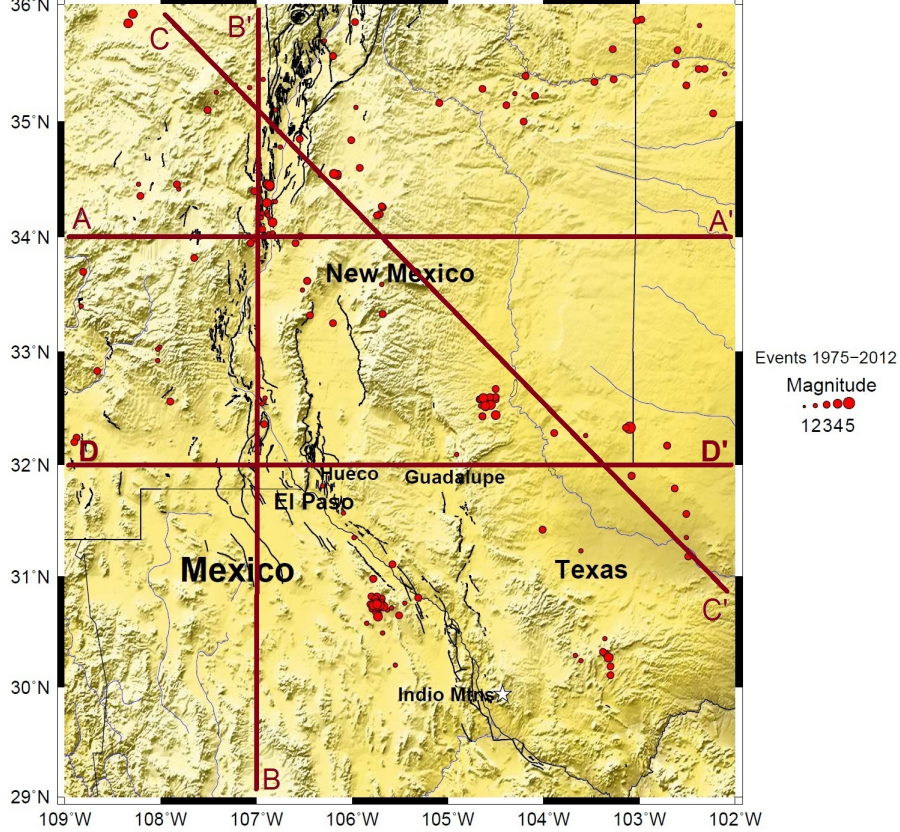


Figure 3.1: Regional topography map of the SRGR and plot of all the seismicity from 1975-2012. The white circles signify the various sizes of magnitude from 1-5. Three different profiles has been traced to study the upper mantle structure of the provinces described in Figure 3.2.

In this work, we start with an overview of the regional tectonic setting to provide a background for the data and the interpretation of our findings. We describe briefly the data sets comprised in our joint inversion, and present a constrained optimization framework that connects a joint inversion algorithm with a Bayesian interpolation scheme for high resolution

imaging of Earth structure. We describe the interpolation scheme used to smooth out and incorporate uncertainty to our independent velocity models. We show results from 1-D to 3-D Earth S-wave velocity structure of the crust and upper mantle along several profiles that cross the southern end of the RGR. Finally, based on our findings we present a discussion and draw conclusions.

## 3.2 Tectonic Setting

The RGR, a major continental rift, has existed since the late Cenozoic and separates the Proterozoic continental lithospheres of the western Great Plains and the Colorado Plateau [67, 37, 8]. Rifting and extension began in the late Oligocene or early Miocene [17]. An initial stage of extension began at 30-20 Ma, with an association of low-angle faulting and crustal doming. The second phase of extension occurred at 3-10 Ma, which involved 10 Ma of extension trending in the E-W direction [39, 78]. [78] hypothesize that this extension resulted from an upper mantle asthenosphere upwelling and thermal lithosphere erosion. Extension along the western interior portion of North America stimulated the formation of the RGR while exempting the Colorado Plateau from being deformed during this regional extension, see e.g., [78]. [37] considers that the southern RGR has experienced more extension than the northern section of the rift, while [43] did not find significant extension across most of the RGR but the southernmost part. Volcanism was prominent in the rift area during the Pliocene and Quaternary [17].

Recent work from the Colorado Plateau-Rio Grande Rift-Great Plains Seismic Transect (LA RISTRA) passive experiment has shown that the center of the RGR has a low velocity zone [80, 76], suggesting that there could be molten material or that the crust is thinning beneath the center of the RGR. The zone of crustal thinning widens southward as does the physiographic expression of the Rift. In Southern New Mexico, the Rift seems to have its maximum effect from a geophysical perspective creating the thinnest crust (less than 30 km) with very high heat flow [38]. The upper mantle structure beneath the eastern Colorado Plateau and Rio Grande Rift has been studied by integrating seismic velocities, gravity, and

xenolith data, to explore temperature and compositional variations together with partial melt content. The paper [61] interprets this variations and melting results as a product of modified and/or thinned lithosphere. Moreover, [61] argues that the RGR and southeastern Colorado Plateau underlay by a low-density upper mantle province, which does not trend along upper crustal tectonic boundaries, but correlates with regions of late Tertiary magmatism. On the other hand, [7] suggests that the low velocity mantle beneath the RGR indicates that some removal of the lithosphere has occurred.

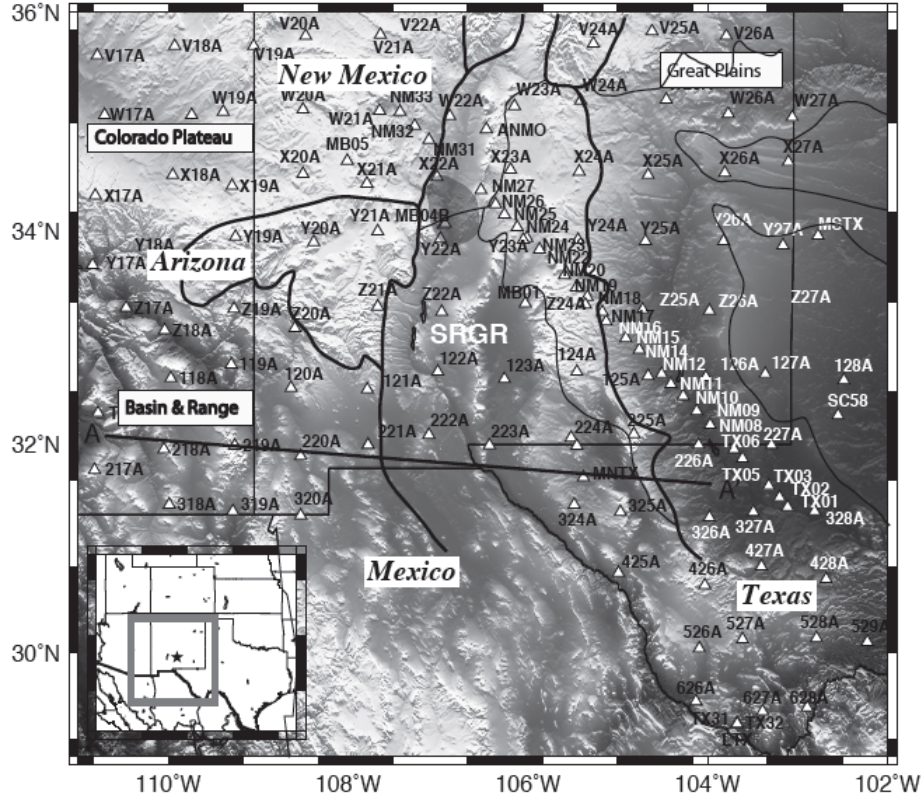


Figure 3.2: Regional topography of the SRGR and sketch of its boundaries. The white triangles mark the stations which data was used for this study.

### 3.3 Receiver Functions

A receiver function maps the seismic response of the earth beneath a seismic station to an incoming P wave. It can be calculated by deconvolving the vertical component of a teleseismic

earthquake seismogram from the radial component [45]. The resulting receiver function allows for the identification of converted phases derived from strong impedance contrasts (e.g., the crustal-mantle boundary). Moreover, since we use teleseismic events that arrive at the stations with near-vertical incidence, these events can be used for deep structure imaging of the subsurface [64]. In particular, they appear to provide valuable information for investigating magma lenses within the crust, determining the Moho depth, and other upper-mantle discontinuities [50], structure and evolution of the crust [8], and rifting extension and magmatism, see e.g., [21].

For this study, we collected three-component seismic data for 147 stations from the LA RISTRA portable seismic experiment, and from EarthScope transportable array (USArray) [71] within the area of latitudes between 29 to 36 N and longitudes between  $-111$  to  $-102$  E (Figure 3.2). The nominal spacing between USArray stations is about 70 km, which is suitable for both lower crustal and upper mantle seismic studies [64]. The LA RISTRA experiment recorded data for a year and a half since August 1999. [71] compared their findings with the EarthScope Automated Receiver Survey (EARS) website results (<http://www.seis.sc.edu/ears/>), and found inconsistencies that are likely the result of differing quality control parameters, and loss of high frequency information [79]; [64]. We utilize the receiver function data set provided by [71], which includes 434 receiver functions stacked in ray parameter bins derived from 1464 teleseismic seismic events with a minimum moment magnitude of 5.5 and occurring from January 2000 to December 2009. This data set focuses in the southern Rio Grande Rift, and the details of the stacking procedure can be found in [71].

### 3.4 Surface Wave Dispersion

In general, surface waves dominate seismograms as the largest amplitude waves from an earthquake, and with observed lower frequencies than the body waves. Furthermore, surface waves are dispersive with velocities that vary depending on the depth range sampled by each period, thus they provide valuable information for studying Earth’s deep structure [58]. In particular, Love and Rayleigh wave group dispersion observations generally account for average velocity

structure as a function of depth [34], [52].

As part of the systematic determination of earthquake moment tensors for North American earthquakes by Saint Louis University, multiple filter analysis is applied to obtain the fundamental mode Love and Rayleigh wave spectral amplitudes and group velocities to provide the data set for one type of source inversion. As of September 04, 2012, there are over 1,329,290 dispersion measurements available for use. To obtain the group velocities, a tomography technique, developed by Ammon (personal communication) and outlined in Cho et al., (2006), was used to obtain tomographic images of group velocity dispersion for North America with emphasis on the contiguous 48 states of the United States. An advantage of using dispersion measurements from regional earthquakes is the ability to obtain the dispersion at periods less than 15 seconds, which is difficult to obtain from teleseismic data. We use the dispersion data set ([http://www.eas.slu.edu/eqc/eqc\\_research/NATOMO](http://www.eas.slu.edu/eqc/eqc_research/NATOMO)) for our analysis of the Rio Grande Rift.

### 3.5 1-D Constrained Joint Inversion

Joint inversion has become a common practice for 1-D to 3-D geophysical inversion problems. In general it involves the simultaneous optimization of several objective functions such as  $\ell_2$ -norm data misfit functions. Since the combined objective function is expected to have fewer local minima, this approach reduces intrinsic non-uniqueness of the problem [15]. Some examples are: cooperative inversion [49], weighted schemes for inverting seismic travel times and gravity data [46], for DC resistivity and seismic data [26], receiver functions and surface wave dispersion [34], [68], surface wave velocity and gravity observations [52], and receiver functions, surface wave dispersion, and magnetotelluric data [56], [55]. In all these cases, the main assumption is that the data sets comprised in the inversion complement each other and sample similar geological boundaries.

In this Chapter, we apply a 1-D constrained optimization approach for joint inversion of two complementary data sets, receiver functions and surface waves group dispersion [34], by using Primal-Dual Interior Point methods as a solver [68]. Our approach addresses some of

the well known numerical difficulties that arise for highly nonlinear misfit functions and large-dimensional model spaces, by using inequality constraints to incorporate a priori information and constrain further the geophysical inversion [68]. Moreover, the presence of regularization terms usually added to the objective function seems not to be crucial for most of the inverted velocity profiles.

We characterize the Earth's structure by using S-wave velocities as model parameter. The forward nonlinear operator  $F \in \mathbb{R}^m$  at the given velocity  $x$  corresponds to a given prediction of the Earth's response according to the data used as input. For a given observed data vector  $y \in \mathbb{R}^m$ , we pose the inverse problem as

$$\min_x \frac{1}{2} \|F^{SW}(x) - y^{SW}\|_W^2 + \frac{1}{2} \|F^{RF}(x) - y^{RF}\|_W^2, \quad (3.1)$$

where we have a weighted diagonal matrix  $W$ , used to equalize the contribution of each data set with respect to physical units and number of data points, while accounting for data set influence [68]. In our case, our forward operator  $F$  collects both the numerical computation of synthetic waveforms of receiver functions  $F^{RF}$  [2], and the numerical evaluation of surface waves dispersion velocities  $F^{SW}$  [52]. In this work, we assume a typical uncertainty value  $\sigma_i^2$  of 0.05 (km/s) for SW, 0.01 (s) for RF observations [34], and we accommodate the amount of influence for each data set according to the station data quality, that in general is set equally for most of the stations.

Instead of the standard formulation of the inverse problem as in the unconstrained (linearized) weighted nonlinear least squares (NLSQ) setting (3.1), we solve a sequence of linearized constrained LSQ where  $F'(x_k)$  is the matrix with the partial derivatives of  $F$ . We rewrite the problem (3.1) as,

$$\begin{aligned} \min_x \quad & \frac{1}{2} \|F'(x_k)x + r(x_k)\|^2 \\ \text{s. t.} \quad & g(x) - s = 0 \\ & s \geq 0, \end{aligned} \quad (3.2)$$

where

$$r(x_k) = F(x_k) - y - F'(x_k)x_k, \quad g(x) = \begin{bmatrix} x - c_{min} \\ c_{max} - x \\ \|Lx\| - \gamma \end{bmatrix} \in \mathbb{R}^{2n+1}, \quad \gamma = \frac{c_{max} - c_{min}}{n}.$$

This allows us to add appropriate bounds corresponding to a priori minimum and maximum velocities, i.e.  $c_{min} \leq x \leq c_{max}$ , while enforcing a roughness model constraint by using a first order discrete derivative operator  $L$  [14]. As before,  $s \in \mathbb{R}^{2n+1}$  is a slack variable. We apply a primal-dual interior point (PDIP) methodology to solve problem (3.2), which introduces an intrinsic regularization to the inverse problem making the joint inversion algorithm more robust [57, 68]. In this methodology, we need to define the Lagrangian function associated to (3.2):  $\ell(x, z, s) = \frac{1}{2} \|F'(x_k)x + r(x_k)\|^2 - (g(x) - s)^T z$ ,  $(s, z) \geq 0$ , where  $z \in \mathbb{R}^{2n+1}$  is the Lagrange multiplier associated to the inequality constraints. Interior point methods are based on Newton's method. In our case, the necessary perturbed Karush-Kuhn-Tucker (KKT) conditions, computed by differentiating  $\ell$  with respect to the primal variables  $x$  and  $z$ , provide the right hand side of a Newton's system. This system can be solved iteratively by using a linesearch strategy [57] while enforcing the iterates to stay in a feasible (interior) region as described in [68]. The iterative process proceeds until it is terminated either when the misfit reaches a desired level, or a maximum number of iterations is reached, or the difference between iterates fail to differ more than a certain threshold. For all the stations involved in the geophysical inversion, the initial velocity model  $x_0$  corresponds to the AK-135 model of [40], starting at 10 km depth and distributed at a 2 km interval up to 70 km depth, then at a 5 km interval up to 250 km and finally at 10 km until 420 km.

### 3.6 Joint Inversion Results

We perform 1-D joint inversions using our PDIP approach for 147 stations from USArray and LA RISTRA experiment. In general, each independent joint inversion include: at least three receiver function bins created according to an average ray parameter with a width of approximately 0.01 s/km between 0.04 s/km and 0.07 s/km. The average ray parameter was

determined by taking the mean value of the maximum and minimum ray parameter of each individual station, before being used for stacking. The number of receiver functions employed to create these stacks changes between stations, but in general we use at least 25 receiver functions per ray parameter. Each receiver function consists of 820 data points for a time range from  $-5$  to 80 seconds. Also fundamental mode Love and Rayleigh group velocities with 50 to 65 dispersion measurements, with periods between 5 to 140 seconds, were used. Since the station spacing of the USArray is about 70 km, we anticipate lateral resolution in that order for each individual 1-D inversion in the southern RGR region. We present in Figure 3.3 an example for crustal and upper mantle 1-D velocity structure computed by using the constrained joint inversion algorithm for station 426A. We show the acceptable fit to the RF observations, to the Love and Rayleigh wave group dispersion curves, and the final model approximation provided by the inversion for that particular station. Velocity values were extracted from layered models as described before.

### 3.7 Kriging interpolation from 1-D velocity profiles

Since our ultimate goal consists of creating a 3-D Earth structure of the Rio Grande Rift region, we use the 1-D S-wave velocity profiles of each station as input data for a kriging interpolation algorithm. In general, interpolation algorithms aim for estimating values of certain quantities by using a weighted sum of surrounding data values. Kriging represents an example of a computationally efficient interpolation technique that allows for incorporating uncertainty into the predicted values. In this work, we implement a Bayesian kriging approach that incorporates variable spatial damping; it is a useful tool to control the kriged solution in extrapolation zones where few or no data is available [65]. In our case, we can think of our target RGR region as a 2-D spatial grid with nodes corresponding to the stations considered in this study. With this setup and considering a vertical projection of this grid, we estimate the unknown velocities of the 2-D grid at different depths based on the known velocities at the surrounding nodes.

Initially, we need to remove an appropriate trend prior to the application of kriging [65].

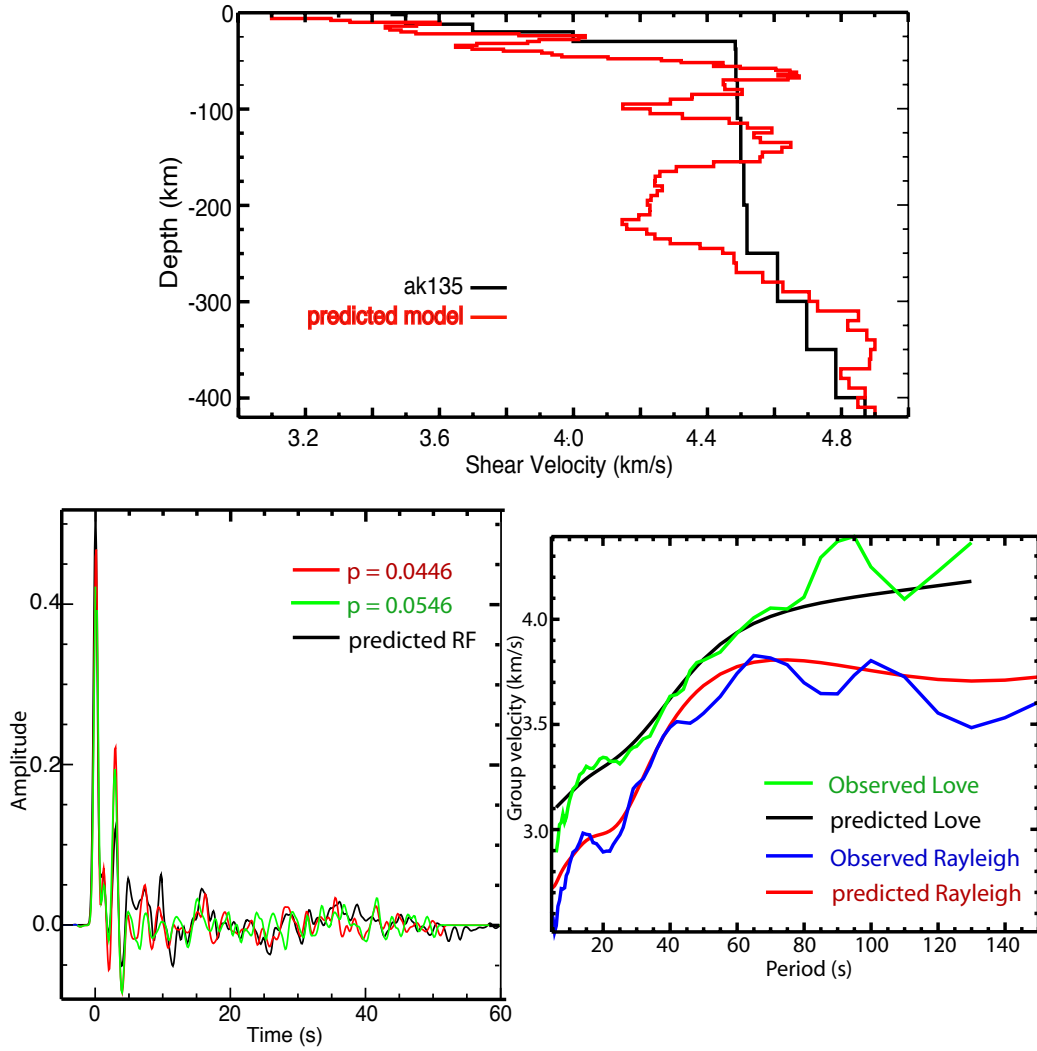


Figure 3.3: 1-D joint inversion results for station 426A within the RGR. (Top) The final S-wave velocity profile, in red, provided by our joint inversion approach, initialized with the standard model ak135 in black. (Bottom) We show the acceptable fit to the RF observations (left) in black from our synthetic approximation in red for a ray parameter  $p = 0.0446$  and in green for  $p = 0.0546$ , and to Love and Rayleigh wave group dispersion observations (right).

The trend we use corresponds to the mean of the velocities at a certain depth. Then a spatially damped kriging estimator incorporates variable damping and measurements error multiplied by a unit-normalized function, which decreases noise values to zero according to distance relative to the prediction point. As a result, we obtain a smoothly damping effect over the

predicted velocities, an effect that varies according to each node velocity and its surroundings, to constrain further the mapping of crust and upper mantle S-wave velocities. From this analysis, we choose the blending functions of  $2^\circ$  to guarantee good spatial sampling.

We expect that blending the profiles by means of the kriging interpolation scheme can help us to illuminate better the Earth structure beneath each station in the RGR. Schematically, if each station had perfect azimuthal coverage, the region below each station would have cone shaped raypaths, where at a certain depth (that depends on station spacing), the raypaths at adjacent stations begin to overlap, providing us with full subsurface structure coverage. Before this depth, we expect that the surface wave group dispersion information obtained from regional earthquakes can improve not only average crustal velocity structure but also vertical resolution [64]. In this form, we account for velocity structure resolution with no need of increasing the number of inversion, by grouping the 1-D profiles depending on azimuthal range as indicated by [7]. Generally, upper mantle of a tectonically active region is expected to exhibit 3-D heterogeneities with length scale smaller than both the lateral resolution of surface waves and the vertical resolution of receiver functions [58].

Therefore the models obtained by using these two data sets should resolve the main features beneath the region. However, since the inversion algorithms produce non-unique results, with ours not being an exception, it seems to be very helpful to start with other information such as known geological constraints. We believe that by incorporating explicit velocity bound constraints with a measure of roughness into the inversion model, and - for some of the stations - adding a regularization term, we can produce a better constrained model while having more stable inversions [68]. We now present in Figures 3.4 through 3.8 a summary of 3-D crustal and upper mantle structure images, for the four different profiles in Figure 3.1 created from kriging interpolation of 1-D velocity models. Only brief discussions over the main features of these profiles are presented. In all these Figures, we show on the left the full range of velocities computed to show the main provinces, and on the right, we constrain the velocity to 4-5 km/s to highlight features deep in the upper mantle.

Cross-section A-A' (Figure 3.4) is 740 km long, coincides with latitude  $34^\circ$ , and passes through the Colorado Plateau (CP), Socorro Magma Body (SMB) and ends at the Great

Plains (GP). Cross-section B-B' (Figure 3.5) is 1000 km long, coincides with longitude  $107^\circ$  and intersects cross sections A-A' and D-D', passing through the Socorro Magma Body. Cross-section C-C' (Figure 3.6) is 700 km long, coincides with the Southern part of the Rio Grande Rift (LA RISTRA) experiment and cross cut models A-A', B-B' and D-D'. It starts at Colorado Plateau pass through the SMB, and ends west of the southern RGR.

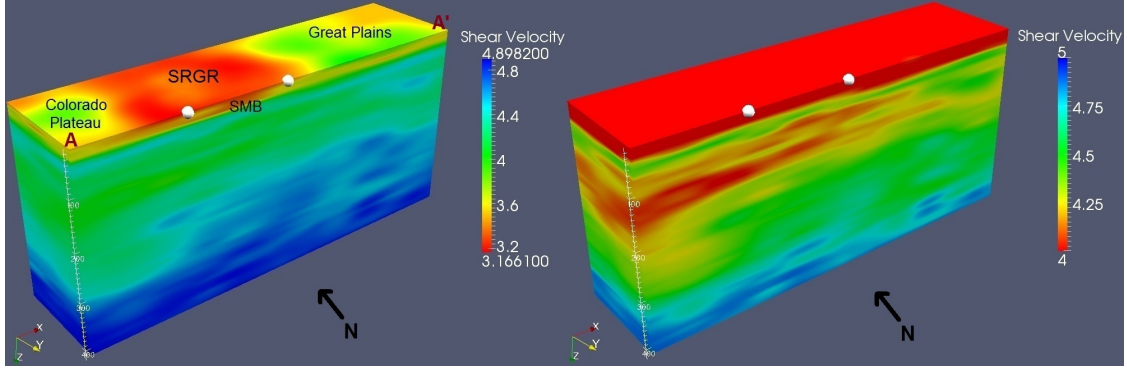


Figure 3.4: Cross-section A-A' at latitude  $34^\circ$  shows a clear distinction between the Colorado Plateau (CP), Socorro Magma Body (SMB) at the center of the RGR, and the Great Plains (GP). We find that near the CP, there is a low velocity lid with high velocities beneath this province, consistent with other studies. Anomalous high velocities begin to appear right below the RGR and continue east of the Great Plains between the depths of 200 - 300 km

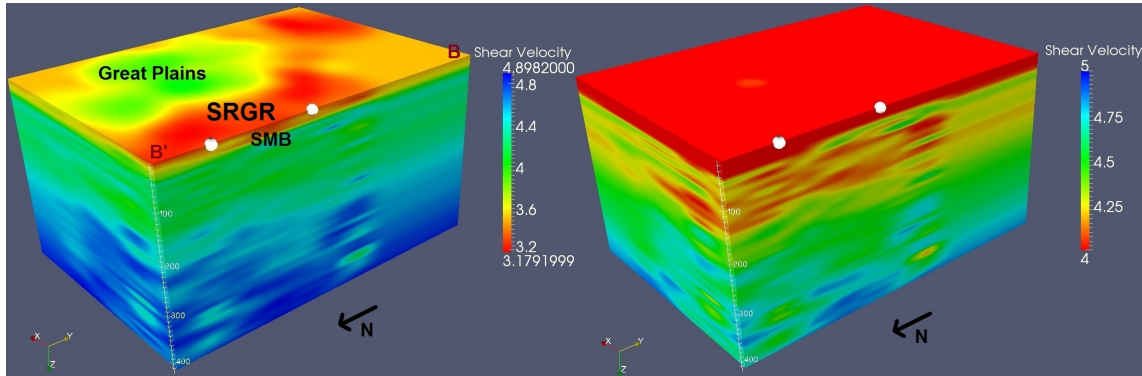


Figure 3.5: Cross-section B-B' coincides with longitude  $107^\circ$  and intersects cross section A-A' and passes through the SMB. The SMB seems to have slow velocities between the Colorado Plateau and Great Plains.

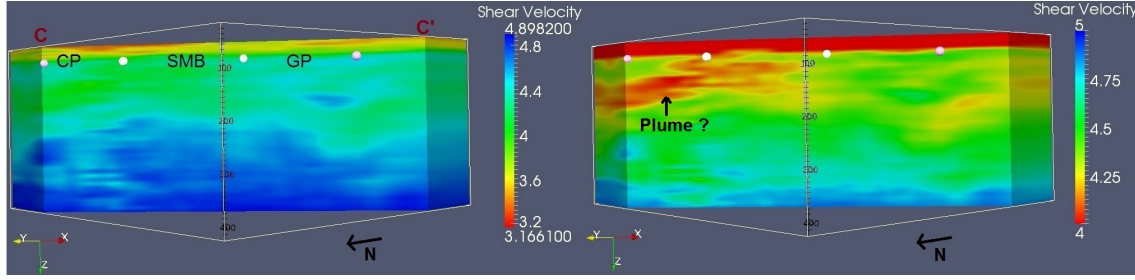


Figure 3.6: Cross-section C-C' coincides with the southern part of the LA RISTRA passive experiment. Seismically fast mantle underlies the RGR and relatively slow mantle is seen beneath the Socorro Magma Body and Colorado Plateau.

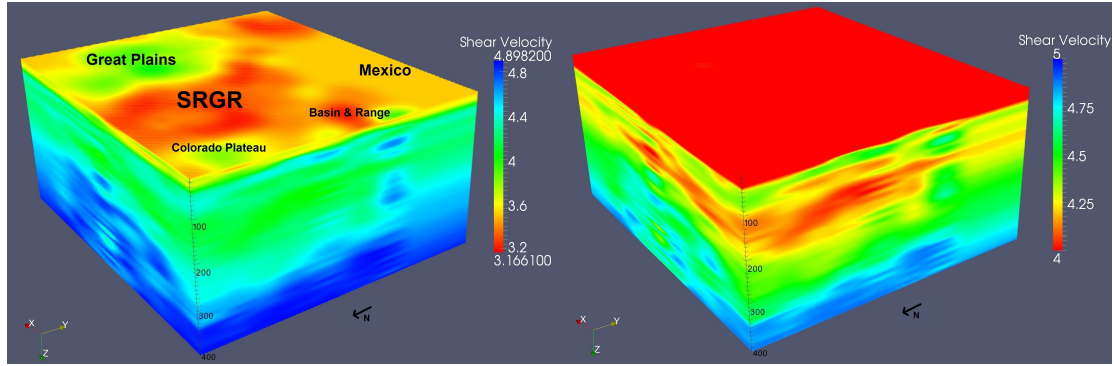


Figure 3.7: Full 3-D velocity structure perspective of the RGR view from N-W. We identify low velocities associated to the southernmost part of the RGR. The upper 200 – 300 km of mantle beneath the magmatically and tectonically active RGR and Basin & Range (B & R) is seismically distinct from the mantle beneath the stable Colorado Plateau and Great Plains.

Cross-section D-D' (Figure 3.8) is 750 km long coincide with latitude 32° and covers the southern most part of the study area.

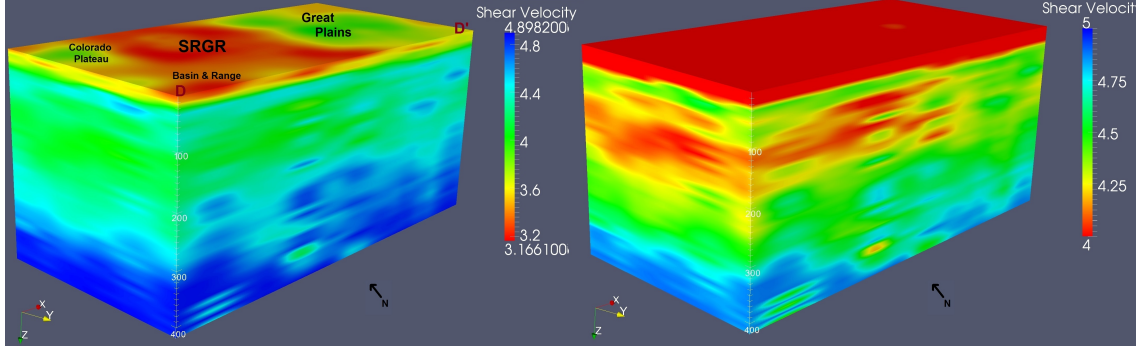


Figure 3.8: Cross-section D-D' at latitude 32° covering the southern-most part of the RGR region. We notice the transition between the Basin & Range province and the Colorado Plateau to the north and the Great Plains to the west. Thin crust is identified in the B & R and thicker crust in the other two provinces. We image a low velocity zone that begins to appear beneath the RGR extending to the west below the B & R and Colorado Plateau.

Cross-section A-A' shows slightly uplifted Moho beneath the Basin & Range province, lower and middle crust that might be related to magmatic activity in the upper mantle. This area of active upper crust extension is suggested to be primarily the product of magmatic activity in the lower crust and upper mantle (Serpa, [1990]).

As stated by [76] an asthenospheric low velocity channel underlies the region west of the Great Plains and extend to 300 km depth. This low velocity zone extends further to the Colorado Plateau at 200-300 km depth, but differing from [76], we do see it directly beneath the Rift and we do not see it forming inverted U-shape. No evidence for a deep mantle source is found under the Rift, implying that is not currently driven by deep mantle upwelling [76]. As in previous studies, we image low velocity mantle beneath the RGR [76], [61], sharp changes between the RGR and the two surrounding provinces of Basin & Range and Colorado Plateau [7].

## 3.8 Discussion

### 3.8.1 PDIP methodology

We implement a novel approach for joint inversion of receiver function and surface wave group dispersion data, based on constrained optimization [68]. We create independent 1-D Earth velocity profiles of upper mantle velocity structure along the southern RGR. In general, no smoothing or damping factors were used to stabilize the inversions. This is an advantage of our approach when compared to other standard techniques, which often require tuning of several regularization parameters. Moreover, for severely ill-posed inversions that may appear for some stations, determining the optimum regularization parameters remain a difficult (and often speculative) task. Our PDIP approach reduces the subjectivity of these selections, since the simple inclusion of a damping parameter may be sufficient to stabilize the inversion. As we pointed out in the previous section, some stations presented convergence issues for the inversion mostly due to the lack of enough RF data to identify absolute S-wave velocities and sharp discontinuities. Those stations were independently tuned by using several damping factors. We noticed that with a damping factor of 0.1 and an influence parameter equal to 0.25 higher improvements could be obtained. Moreover, we found that in some cases, just modifying the influence parameter to 0.75 already leads to good results, there is no need to use the damping factor. We discuss with detail the impact of regularization (damping/smoothing) on our constrained optimization approach in [68].

A similar approach combining results obtained by independent joint inversion of surface wave phase velocities and receiver function information to compute 1-D S-wave velocity profiles has been implemented for the Colorado Plateau [7]. This work differs from ours, since to actually create the 3-D models, the authors of [7] require to extract common features of nearby stations, which then needs to be reconciliated with observed gravity anomalies. These anomalies are established by using empirical relations with density structure, which may lead to 3-D models biased by the adjustments required to fit the gravity observations. Furthermore, we use surface waves group velocities information instead of phase velocity, since it requires less forward computations and does not need to account for uncertainties related to the source.

As stated by [7], we agree that when independent inversions that produce 1-D velocity profiles to create 3-D images of Earth structure, the results are inferior to inversion in a full 3-D parameter setup. However, we do not disregard the possibility of a full inversion simply because of the high volume of information, and the great amount of parameters involved in the joint inversion. In fact, it may be numerically tractable depending on the inversion method implemented by the user. In our case, it is well known that PDIP methods have been proven to be a succesful tool to solve large-scale problems [57]. In fact the migration of our technique to a full 3-D setting is a future avenue of research.

We can actually adapt our inversion approach by tuning the kriging interpolation blending function according to the geological province where the stations are located, (e.g., Basin and Range, RGR, Great Plains), to enhance the performance of the inversion. However our robust approach allowed us to start with a standard initial model, e.g., AK-135 [40]. As a result we produce coherent independent inversions that could be combined for a consistent 3-D structure, without weighing the contribution of each inversion separately to select a suitable initial model as in [7]. The independent joint inversions were carried out concurrently in a Linux SunOS server with 2 UltraSPARC-III+ processors running at 1 GHz and 2GB of RAM. The total time of each individual inversion per station is about 3 min, and the total time for the entire region was never greater than 4 hours. Therefore, this algorithm can be adapted easily to run in any modern platform with a Linux operating system and 2GB of RAM.

### **3.8.2 3-D velocity structure from 1-D velocity profiles**

We aim to contribute to the geological questions that remain on debate about the evolution and state of the lithosphere on the Rio Grande Rift, in particular about the role of mantle upwelling in the formation of the Rift. To this end, we facilitate a framework that connects a joint inversion algorithm with an interpolation scheme for high resolution imaging of Earth structure. This scheme efficiently provides a robust alternative to advance the understanding of Earth structure, by extending simultaneous independently created 1-D S-wave velocity models to full 3-D crustal and upper mantle velocity structure. Our joint inversion approach of receiver function stacks and surface waves group dispersion observations uses data from 147 stations of

the USArray and LA RISTRA broadband seismic experiment. The complementary properties and high quality control of these data sets, together with the robust interpolation technique implemented in this work, lead us to be confident in our results. Moreover, by performing independent inversions with our constrained optimization approach, we expect to reduce the number of spurious solutions, e.g., those that lie outside the physical bound constraints. These 1-D models, interpolated by means of a Bayesian kriging scheme, allowed us to create 3-D velocity models with smoother varying resolution than previously known, while requiring a modest amount of computational resources compared to full 3-D geophysical inversions.

### **3.8.3 Interpretation: Basin and Range, Colorado Plateau, RGR, and Great Plains**

The Rio Grande Rift system is characterized by anomalously high topography [61] and its crustal thickness is consistent in areas that should be thicker and other that should be thinner [38]. As in [7] in Figures 3.4 – 3.8 we found sharp changes in crustal thickness that show a distinction between the Colorado Plateau and surrounding provinces dominated by extension, e.g. BR, GP and RGR. [61] showed that low wave speed is broader in the north and narrow in the south. Our results seem to indicate that the opposite is occurring: the wider portion of the Rift that is located in the southernmost part. [60] concluded that analysis of the LA RISTRA profile has been related to phenomena typically associated with Rift tectonics, such as mantle convection and partial melting condition in the crust and upper mantle. We think that there is no conclusive evidence of melting but rather of cold mantle upwelling (downwelling) evidenced in Figures 3.4 and 3.6, which corresponds to latitude  $34^\circ$  and LA RISTRA profile respectively.

If the thin lithosphere comes together with high heat flow, then this may suggest that the lower crust has been removed by delamination. We image decreasing and increasing S-wave velocities at our approximation to the LA RISTRA profile, which may be associated to evidence of delamination. Initially delamination was proposed as one of the main alternative mechanisms of lithospheric recycling in continental collision areas [9]. The delamination process leads to considerable thickening of the crust in areas with weak lithosphere, which we can appreciate

in the distinction of the Great Plains and RGR provinces in Figures 3.4 and 3.8. The mafic material in the lower crust transforms to a dense eclogite, and once it reaches a critical mass, it starts to sink as an eclogitic drop [42]. After delamination, the crust above the descending drop can be associated with low seismic velocities, as those present in Figures 3.4 to 3.7. On the other hand, [8] showed that the B & R is characterized by a thin crust similar to a typical cratonic crust, suggesting that crustal thinning was a result of simple stretching of the original crust rather than delamination. We do find thin crust under this province, but we think that the evidence is not conclusive to a hypothesis that the origin of such feature is either delamination, or simple stretching of the crust.

### 3.8.4 Socorro Magma Body

The Socorro Magma Body (SMB) within central RGR is one of the largest active intrusions in the Earth’s continental crust, and is associated with a steady central uplift [59]. Also, the SMB has been linked to strong magma influence, e.g., diffusion of fluid moving upward from depth, due to an underlying low-velocity molten layer. On the one hand, [61] associated this low velocity zone to a possible combination of partial melt, temperature and compositional variations. On the other hand, [62] suggested that there is not a strong direct magmatic influence in the seismic activity of the SMB, and that this activity is more prone to be associated with characteristics of a continental rift, like preexisting highly fractured crust. We image what appears to be low velocities beneath the Socorro Magma Body between the Colorado Plateau and Great Plains. In particular, we appreciate that the upper 200 – 300 km of mantle beneath the magmatically and tectonically active Rio Grande rift and Basin and Range, are seismically distinct from the mantle beneath the stable Colorado Plateau and Great Plains. Fast mantle underlies the Rio Grande rift and Jemez lineament [7]; [61], and relatively slow mantle is seen beneath the Navajo volcanic field within the Colorado Plateau. Upwelling of that mantle material has been associated with anomalously high heat flow at the surface and magma supply to the crust. We think that the constant uplift around Socorro may not be indicative of magma injection, since the relative variations of S velocities imply that there seems to be more solid material, which could explain the cause of the relatively fast upper mantle. Sharp boundaries

in mantle seismic velocity are coincident with boundaries of Proterozoic structural trends, implying that ancient lithospheric structure exerts a control on the tectonic and magmatic activity in the region. Approximately at 300 km depth, there seems to be upper mantle upwelling, since the shear wave velocities are faster in the Socorro Magma Body and slower on the west and east flank of the RGR. We believe that this feature is produced by cold mantle upwelling, as we see in our three dimensional velocity model.

### 3.9 Conclusions

This study presents a new model of crustal and upper mantle structure beneath the southern RGR. Separate joint inversions were performed for 147 Earthscope USArray and LA RISTRA stations. We facilitate a generic framework that connects a constrained optimization joint inversion algorithm, with a Bayesian interpolation scheme for high resolution imaging of Earth structure. Furthermore, this scheme efficiently provides a robust alternative to extend simultaneous independently created 1-D S-wave velocity models; it enables us to produce 3-D images of the Earth compared to full 3-D inversions. Our framework generates a continuous and smooth 3-D velocity model of the Rift system, revealing the complexities of the southern RGR and helping us to better characterize its crustal and upper mantle velocity structure. The reliability of our models resides in our ability to resolve previously identified geological structures in the region. As a result our models provide a new perspective on the upper mantle structure of the RGR that allows us to contribute to the debate about tectonic and lithospheric activity of the Rift.

We observe the current deformation and extension of the Rift system based from the results in Figures 3.4 – 3.8. Our findings are consistent with rifting in the SRGR. The strongest variations are within the upper 300 km of the mantle, unsurprisingly, and correlate well with surface tectonic behavior. In particular, Figure 3.4, 3.5, and 3.8 show evidence of crustal thinning in the center of the Rift. We can appreciate how the thinner crust from the Basin & Range becomes thicker beneath the Great Plains after transitioning through the RGR. This seems particularly clear due to the higher S-wave velocities that may signify cold mantle

upwelling or downwelling. We focused on imaging the southern Rio Grande Rift, the widest section of the Rift, as we show in Figure 3.7, to help answer questions about how it formed and to determine whether this process is still ongoing. Our hypothesis supported the hypothesis the driving force that causes the deformation of the Rift derives from a mantle upwelling. To test this hypothesis, we mapped the velocity structure of the RGR, observing the locations of high S-wave velocities. These characteristic velocities may be explained by cold temperatures consistent with low heat flow movements, while low S-wave velocities may indicate more solid basement material. We were able to observe points of high and low velocities, along with areas of thick and thin crust, which allowed us to make a better assessment of what activity the RGR is experiencing. We find two fast seismic anomalies, beneath the central Colorado Plateau and the Great Plains respectively, and a third anomalously slow velocity within the upper mantle at the Socorro Magma Body. We identify the boundaries between the provinces of B & R, CP, GP, and RGR.

Furthermore, these boundaries between the active and stable mantle are very sharp and are located near ancient suture and shear zone boundaries, indicating that old lithospheric structure plays a key role in tectonic behavior as suggested by previous work on the RGR; see e.g., [36]. Finally, we think that the southern-most part of the RGR dies out in El Paso, but this region remains unresolved due to lack of data for the northern-most portion of Mexico.

The next phase in this geophysical inversion scheme will be to incorporate other compatible data sets into the joint inversion, e.g., gravity and delay travel time data, expecting to constrain further the inversion process and increase the resolution of the mantle. Also, we want to implement a stronger structural constraint to remove noisy components during the inversion, and a grid continuation scheme for modeling resolution.

# Chapter 4

## Constrained Optimization Schemes for a 1D Inverse Wave Propagation Problem

### 4.1 Introduction

The use of nonlinear optimization methods for inverse wave propagation problems remains a challenge due to many mathematical and numerical difficulties. Original literature on the topic goes back a long time; we can only cite some examples here [13, 70]. Several difficulties, such as the existence of multiple local minima, the ill-conditioning of total variation regularization, and achieving scalability to large parameter spaces, have been addressed previously (see e.g. [41, 11, 1, 24]). Our goal in this work is to provide effective constrained nonlinear optimization strategies to be used in conjunction with robust and efficient forward-inverse algorithms for solving a 1D full-waveform inversion problem as reported in [12]. We examine how the incorporation of inequality constraints over the inversion parameter, in this case the wave speed throughout the domain, impacts the performance of the strategies we implement. Inequality constraints are essential to ensure that the forward solver remains stable after iterative updates to the wave speed that occur during the process of finding a numerical solution of the inverse problem.

Let us outline the relation of the forward and the inverse problem. The forward problem consists of determining the dynamic response  $u(x, t)$  of the earth displacement due to a prescribed excitation, under the assumption that the source and the material properties are known. The inverse problem estimates a spatial distribution of the material properties, in our case the elastic modulus  $\mu(x)$  that is related to the wave speed  $v_0$  by  $v_0 = \sqrt{\mu/\rho}$  (the density  $\rho$  is assumed to be known), which results in a predicted earth response that most closely

matches observed ground motion measurements generated by a force term  $f(x, t)$ . Here we restrict the setup to a one-dimensional experiment,  $x \in (0, L)$ , where we aim to identify the elastic modulus along a (vertical) column.

We pose the inverse problem as a nonlinear program, where the goal is to minimize an objective functional that consists of a least squares misfit term between observed and synthetic seismic waveforms  $u^*$  and  $u$ , respectively. A regularization term, like Tikhonov or Total Variation (TV), is typically added to the objective functional to penalize variations in the unknown model parameter  $\mu$ . The problem is constrained by the partial differential equation (PDE) governing elastic wave propagation, and by lower and upper bounds over the model parameter. A considerable part of the literature that deals with the introduction of inequality constraints in PDE constrained optimization problems involves interior-point methods using logarithmic barrier formulations. As an example, coupling the inequality constraints by logarithmic barrier functions can involve a barrier parameter in the context of a multilevel predictor-corrector continuation strategy [3]. In this case, the multilevel scheme is formulated with respect to a hierarchy of discretizations. More recently, an approach that describes a practical implementation of a two step line-search interior point algorithm for large-scale nonlinear optimization presents two PDE-constrained optimization problems as numerical examples [19]. With these two approaches, the challenge remains to explore the specific structure of the ill-conditioning caused by the logarithmic barrier penalization term [23, 57]. On the other hand, [74] provides a detailed analysis of a primal-dual interior-point method for PDE-constrained optimization optimal control problems with control constraints in  $L^p$ . To the best of our knowledge, a generic and affine-invariant framework that accomodates the addition of inequality constraints over the model parameter has not been considered so far for this particular type of PDE constrained optimization problems. To solve the nonlinear programming problem, we employ matrix-free inexact Newton-Krylov iterations, and implement different (path-following) methodologies that have been extensively studied on their own [23, 57]: logarithmic-barrier and primal-dual interior-point methods as well as primal-dual active set methods [33]. By comparing these three constrained optimization schemes, we find that some algorithmic building blocks are shared, and others can be aligned by suitable changes of parameters and rules. Only

few steps remain clearly distinct between the methods. One main contribution of this article is thus to construct a unified algorithm that can accommodate all three methods by suitable substitutions. Furthermore, we introduce two technical improvements to generally well-known procedures: We improve a standard formula for the Newton step length in order to allow a full-length step in the interior, and outfit our inversion procedures with scaling factors that ensure that the iterates remain invariant under affine transformations of the objective functional. The latter feature is particularly important when considering black-box optimization for physically motivated applications due to non-dimensionalization and other commonly used transformations.

This paper is organized as follows. In Section 4.2, we formulate a nonlinear constrained optimization problem and motivate its numerical solution by inexact Newton-Krylov iterations. In Section 4.3 we present inequality-constrained PDE optimization schemes, with a description of the logarithmic-barrier, primal-dual interior-point, and primal-dual active set methods. In the following Section 4.4 we introduce a unified algorithmic framework that covers all of these methods and implements globalization through step size control. In Section 4.5 we describe a full-waveform seismic inverse problem modeled as a partial differential equation in one dimension and conduct numerical experiments for synthetic noise-free example problems where the velocity models are derived from borehole seismic measurements. We report the performance of each method in terms of the number of conjugate gradient iterations, computational time, and accuracy. Finally, we draw conclusions in Section 4.6.

### 4.1.1 Notation

We denote matrices by capital letters and vectors by lower-case letters. Given a vector  $x = (x_1, \dots, x_n)^T$ , we write  $x > 0$  to mean that  $x_i > 0$  for all  $i$ ;  $X$  denotes the diagonal matrix  $\text{diag}(x)$  associated with  $x$ , and  $x^{-1} = X^{-1}e$  denotes the vector whose  $i$ -th component is  $x_i^{-1}$ , where  $e = (1, \dots, 1)^T \in \mathbb{R}^n$ .

The notation  $\bar{x}$ ,  $\hat{x}$  and  $\Delta x$  is used interchangeably to denote pointwise increments over the

variable  $x$ . The first and second variations of a function  $f$  at a point  $x$  are defined as:

$$\partial_x f(\bar{x}) = \lim_{h \rightarrow 0} \frac{f(x + h\bar{x}) - f(x)}{h}, \quad (4.1)$$

$$\partial_x^2 f(\bar{x}, \hat{x}) = \partial_x [\partial_x f(\bar{x})] (\hat{x}). \quad (4.2)$$

If a metric is available through an inner product  $\langle \cdot, \cdot \rangle$  we define the gradient vector  $\nabla_x f$  and the self-adjoint Hessian operator  $\nabla_x^2 f$  as follows,

$$\partial_x f(\bar{x}) = \langle \nabla_x f, \bar{x} \rangle \quad \text{for all } \bar{x}, \quad (4.3)$$

$$\partial_x^2 f(\bar{x}, \hat{x}) = \langle (\nabla_x^2 f) \bar{x}, \hat{x} \rangle \quad \text{for all } \bar{x}, \hat{x}. \quad (4.4)$$

In the following text, we drop the subscript from the gradient and simply write  $\nabla f$ .

The following identity for a linear operator  $A$  that depends on a parameter  $x$ ,

$$\partial_x (A^{-1})(\bar{x}) = -A^{-1} [\partial_x A(\bar{x})] A^{-1}, \quad (4.5)$$

is used in Section 4.2. We denote the inverse adjoint operator by  $A^{-*} = (A^*)^{-1} = (A^{-1})^*$ , with the usual definition  $\langle y, Ax \rangle = \langle A^* y, x \rangle$ .

We use the vector spaces  $V$  and  $W$  endowed with the inner product  $\langle \cdot, \cdot \rangle$  to denote the space-time information and model parameter space, respectively. These spaces can simply be  $\mathbb{R}^m$  or another finite-dimensional vector space, representing a discretization of continuous state variables, but they can also be infinite-dimensional function spaces introduced with the weak formulation of a partial differential equation (PDE).

## 4.2 Formulation of the inverse problem

In this section, we present a generic formulation of the inverse problem and derive a Newton-Krylov iterative solver. This serves as the basis for modifications that deal with inequality constraints, to be presented in Section 4.3, and implement step size control, see Section 4.4. We introduce a sufficiently abstract notation such that a particular choice of the forward problem is immaterial for the formalism, which can then be applied for inverse problems with general nonlinear forward-modeling operators.

### 4.2.1 Problem statement

We define the solution  $u$  of the forward problem by  $Au = f$ , where  $f$  is a given source term, and the linear operator  $A = A_\mu$  depends (nonlinearly) on the parameter  $\mu$ . The goal of solving the inverse problem is to infer the value of  $\mu$  from measurements; it can be formulated as the following equality constrained optimization problem:

$$\begin{aligned} & \text{minimize} \quad \mathcal{F}(u, \mu) = B(u) + \beta R(\mu) \\ & \text{subject to} \quad \langle Au - f, p \rangle = 0, \quad \text{for all } p \in V. \end{aligned} \quad (4.6)$$

Here  $B(u)$  is a weighted norm of the misfit between the observed state,  $u^* = A_{\mu^*}^{-1}f$ , which corresponds to the unknown inversion target  $\mu^*$ , and the calculated state  $u \in V$ .  $\beta R(\mu)$  is a regularization term depending on the parameter  $\mu \in W$  with  $\beta \geq 0$ . In our case, the nonlinear equality constraint corresponds to the PDE discretization of the elastic wave equation for a given source  $f = f(x, t)$ , which needs to be solved for a state  $u = u(x, t)$  that describes a time-space wavefield.

The Lagrangian functional associated with (4.6) is

$$\mathcal{L}(p, u, \mu) = \mathcal{F}(u, \mu) - \langle Au - f, p \rangle, \quad (4.7)$$

where  $p$  is the Lagrange multiplier associated to the equality constraint, referred to below as the *adjoint variable*. The Karush-Kuhn-Tucker (KKT) necessary conditions for an extremum at  $(p, u, \mu)$  are:

$$\partial_p \mathcal{L}(\bar{p}) = -\langle Au - f, \bar{p} \rangle = 0 \quad \text{for all } \bar{p} \in V, \quad (4.8)$$

$$\partial_u \mathcal{L}(\bar{u}) = \partial_u B(\bar{u}) - \langle A\bar{u}, p \rangle = 0 \quad \text{for all } \bar{u} \in V, \quad (4.9)$$

$$\partial_\mu \mathcal{L}(\bar{\mu}) = \beta \partial_\mu R(\bar{\mu}) - \langle [\partial_\mu A(\bar{\mu})] u, p \rangle = 0 \quad \text{for all } \mu \in W. \quad (4.10)$$

Our purpose is to obtain a reduced formulation for the KKT conditions where the only remaining variable is the parameter  $\mu$ . To this end, we view the variables  $u = u(\mu)$  and  $p = p(u(\mu), \mu) = p(\mu)$  as functions of  $\mu$ , which satisfy the state equation (4.8) and the adjoint equation (4.9) for all  $\bar{p}$  and  $\bar{u}$ , respectively. In other words,

$$Au = f, \quad (4.11)$$

$$A^*p = \nabla B. \quad (4.12)$$

Then our goal is to find a  $\mu$  that satisfies (4.10) given (4.11) and (4.12), which equivalently fulfills the necessary condition for the following unconstrained optimization problem

$$\text{minimize } \mathcal{J}(\mu) = \mathcal{L}(p(\mu), u(\mu), \mu). \quad (4.13)$$

The optimality condition for (4.13) is computed from (4.10) as

$$g = \beta \nabla R - E_u^* p = 0, \quad (4.14)$$

which is a square nonlinear system of equations, where  $g$  denotes the gradient  $g = \nabla \mathcal{J}(\mu)$ , and the linear operator  $E_u$  is defined as follows:

$$E_u \bar{\mu} = [\partial_\mu A(\bar{\mu})] u. \quad (4.15)$$

In the specific case of an inverse wave propagation problem, the operator  $E_u : W \rightarrow V$  transforms the parameter increment  $\bar{\mu}$  into a space-time variable in  $V$ ; consequently the adjoint  $E_u^*$  collapses space-time information into a parameter update.

Notice that the computation of the gradient  $g$  requires solving one forward and one adjoint equation for obtaining  $u$  and  $p$ , respectively. While gradient-based optimization algorithms can in principle be used to find a solution of (4.13), they generally degrade in performance with increasing dimension of the model parameter space. For this reason, we proceed with describing an inexact Newton-Krylov algorithm that incorporates information on second derivatives.

### 4.2.2 Inexact Newton-Krylov algorithm

Newton's method applied to (4.14) for the parameter  $\mu$  has the form

$$\mu_{k+1} = \mu_k + \Delta\mu = \mu_k - H^{-1}g|_{\mu_k}, \quad (4.16)$$

where  $H = \nabla^2 \mathcal{J}$  is the Hessian and  $\Delta\mu$  is the Newton step. Assuming that (4.11) and (4.12) hold, we have

$$\begin{aligned} \langle H \bar{\mu}, \hat{\mu} \rangle &= \partial_\mu^2 \mathcal{J}(\bar{\mu}, \hat{\mu}) = \beta \partial_\mu^2 R(\bar{\mu}, \hat{\mu}) - \langle [\partial_\mu^2 A(\bar{\mu}, \hat{\mu})] u, p \rangle \\ &\quad - \langle [\partial_\mu A(\bar{\mu})] \partial_\mu [A^{-1} f](\hat{\mu}), p \rangle - \langle [\partial_\mu A(\bar{\mu})] u, \partial_\mu [A^{-*} \nabla B](\hat{\mu}) \rangle. \end{aligned} \quad (4.17)$$

Using the identity (4.5) and the chain rule for  $\nabla B$  (which depends on  $u(\mu)$ ), we compute the Hessian as

$$H = \beta \nabla^2 R + E_u^* A^{-*} (\nabla^2 B) A^{-1} E_u + E_u^* A^{-*} E_p + E_p^* A^{-1} E_u - E_{up}, \quad (4.18)$$

where we define

$$E_p \bar{\mu} = [\partial_\mu A^*(\bar{\mu})] p, \quad (4.19)$$

$$\langle E_{up} \bar{\mu}, \hat{\mu} \rangle = \langle [\partial_\mu^2 A(\bar{\mu}, \hat{\mu})] u, p \rangle. \quad (4.20)$$

Note the relation  $\langle E_p, u \rangle = \langle E_u, p \rangle$ . The first two terms in (4.18) constitute the positive definite Gauss-Newton approximation and do not involve the adjoint  $p$ . The last three terms represent second order information about  $H$ , which may introduce negative eigenvalues away from the optimum. The Newton step  $\Delta\mu$  can now be obtained by solving the linear system

$$H \Delta\mu = -g, \quad (4.21)$$

for either the full Hessian  $H$  or its Gauss-Newton approximation. Since we are interested in potentially very large-dimensional parameter spaces we solve (4.21) iteratively by a Krylov method, namely by the method of conjugate gradients (CG [31]) which does not require assembly of  $H$  but only the application of  $H$  to parameter increment vectors  $\bar{\mu} \in W$ . We propose Algorithm 1 to accomplish this with only one forward and one adjoint solve. It can be simplified to perform a Gauss-Newton approximation instead by setting the adjoint  $p$  to zero (or, more efficiently, removing the terms containing  $p$ ).

---

**Algorithm 1** Compute Hessian-vector product  $H\bar{\mu}$  given  $u$ ,  $p$ , and increment  $\bar{\mu}$

---

- 1: Incremental forward solve: Compute  $\bar{u} \leftarrow A^{-1} E_u \bar{\mu}$ .
  - 2: Incremental adjoint solve: Compute  $\bar{p} \leftarrow A^{-*} [(\nabla^2 B) \bar{u} + E_p \bar{\mu}]$ .
  - 3: **return**  $H\bar{\mu} \leftarrow \beta (\nabla^2 R) \bar{\mu} + E_u^* \bar{p} + E_p^* \bar{u} - E_{up} \bar{\mu}$ .
- 

Additionally, even though the Hessian is generally indefinite away from the minimum of  $\mathcal{J}$ , it is at least positive semi-definite at the minimum; we can thus employ the CG algorithm with an additional stopping criterion to detect non-positivity [20]. Furthermore, it is not necessary

to solve (4.21) to high accuracy away from the minimum; we use an inexact Newton's method with a stopping criterion for CG of the form

$$\|H\Delta\mu + g\| \leq \epsilon \|g\|. \quad (4.22)$$

Keeping the tolerance  $\epsilon$  at a fixed value can destroy the quadratic convergence of Newton's method, so scaling  $\epsilon$  proportionally to the gradient can be appropriate [22].

### 4.2.3 Affine invariance

It is often desirable to enhance generality and robustness of the optimization algorithms with respect to the scale of the objective functional, since the scale is seldomly known a priori and depends, for example, on the particular problem. For the algorithms proposed here we thus enforce invariance of the iterates  $\mu_k$  under an affine transformation  $\mathcal{J} \rightarrow \zeta\mathcal{J} + \theta$ . The Newton equation (4.21) and the CG stopping criterion (4.22) naturally satisfy this requirement since gradient and Hessian scale with the same factor as the cost functional. When implementing progressive stopping criteria, we can maintain affine invariance by using

$$\|H\Delta\mu + g\| \leq \epsilon \|g_0\|^{-s} \|g\|^{1+s}, \quad s \geq 0. \quad (4.23)$$

Here  $\|g_0\|$  can, for example, be the norm of the initial gradient. We comment below on how we maintain affine invariance for all specializations discussed in the remainder of this paper.

## 4.3 Three constrained optimization schemes

Let us recall that our objectives are to solve problem (4.6) efficiently and accurately, and to determine an effective strategy to incorporate bounds over the parameter  $\mu$ . To this end, and based on the reduced formulation (4.13), we present the following PDE constrained optimization problem:

$$\begin{aligned} & \text{minimize} && \mathcal{J}(\mu) \\ & \text{subject to} && a \leq \mu \leq b, \end{aligned} \quad (4.24)$$

where the objective function  $\mathcal{J}$  is twice continuously differentiable,  $\mu \in W = \mathbb{R}^n$ , and  $a, b \in \mathbb{R}^n$  are the upper and lower physical bounds for  $\mu$ , respectively.

We implement three constrained optimization algorithms for solving (4.24): logarithmic-barrier, primal-dual interior point, and primal-dual active set methods. All of them have proven to be successful for large scale nonlinear programs. On the one hand, numerical experiments indicate that interior point methods are often faster than primal-only active set methods, particularly when the number of active constraints is small [57]. On the other hand, the primal-dual active set variant constitutes an improvement over the primal-only strategy [33]. In the next subsections, we describe each of these methods; among these methods, the first two are considered as path-following strategies for obtaining an optimal solution of problem (4.24). Before we discuss them, let us present a modified definition of a path-following strategy as given in [4].

(Path-following strategy) *For a fixed parameter  $\tau > 0$ , we solve a sequence of nonlinear systems by an inexact solver, damping the step to keep the iterates away from the boundaries except possibly at the solution of the problem (we stay strictly feasible with respect to the inequality constraints), and apply a globalization strategy to force progress towards the constraints and reduce the objective function. If a solution is not found, the parameter  $\tau$  is reduced and the process is repeated.*

### 4.3.1 Logarithmic-barrier method

To solve problem (4.24) we define the logarithmic-barrier subproblem for a fixed *barrier* parameter  $\tau$  as

$$\text{minimize } \mathcal{L}_B(\mu; \tau) \tag{4.25}$$

where  $\mathcal{L}_B(\mu; \tau) = \mathcal{J}(\mu) - \tau \sum_{i=1}^2 \log \left( \frac{c_i(\mu)}{\|b - a\|_\infty} \right)$  is the *barrier Lagrangian* and  $c(\mu) = [c_1(\mu), c_2(\mu)]^T = [\mu - a, b - \mu]^T \in \mathbb{R}^{2n}$ . To ensure affine invariance as outlined in Section 4.2.3, the argument of the logarithm has been scaled, and  $\tau$  can be substituted with

$$\tau = \kappa \eta, \quad \eta = \|g_0\| \|b - a\|_\infty, \tag{4.26}$$

$\kappa \ll 1$ . For  $\tau > 0$ , the necessary optimality condition for the logarithmic-barrier subproblem is given by:

$$\begin{aligned} F_B(\mu; \tau) &= \nabla_\mu \mathcal{L}_B(\mu; \tau) = \nabla \mathcal{J}(\mu) - \sum_{i=1}^2 \frac{\tau}{c_i(\mu)} \nabla c_i(\mu) \\ &= g - \tau(c_1^{-1} - c_2^{-1}) = 0, \end{aligned} \quad (4.27)$$

and the Jacobian of  $F_B$  is

$$\begin{aligned} F'_B(\mu; \tau) &= \nabla_\mu^2 \mathcal{L}_B(\mu; \tau) \\ &= \nabla^2 \mathcal{J}(\mu) - \tau \sum_{i=1}^2 \frac{\nabla^2 c_i(\mu)}{c_i(\mu)} - \tau \sum_{i=1}^2 \frac{\nabla_\mu c_i(\mu) \nabla_\mu c_i(\mu)^T}{c_i^2(\mu)} \\ &= H + \tau(C_1^{-2} + C_2^{-2}), \end{aligned} \quad (4.28)$$

where  $C_1 = \text{diag}(c_1(\mu))$  and  $C_2 = \text{diag}(c_2(\mu))$ . The path-following strategy in the logarithmic-barrier method solves problem (4.24) by generating a sequence of iterates  $\mu_k$  as solutions of logarithmic-barrier subproblems for decreasing values of the barrier parameter  $\tau = \tau_k = \kappa \frac{z^T c(\mu)}{2n}$ , where  $z$  scales with  $\|g_0\|$  and is defined as a Lagrange multiplier associated to the inequality constraints  $c(\mu) > 0$ . The Newton system for each subproblem reads

$$F'_B(\mu; \tau) \Delta\mu = -F_B(\mu; \tau),$$

or more explicitly,

$$(H + C_B) \Delta\mu = -(g + d_B), \quad (4.29)$$

where  $C_B = \tau(C_1^{-2} + C_2^{-2})$  and  $d_B = -\tau(c_1^{-1} - c_2^{-1})$ . Thus, the Hessian is still positive semi-definite at the minimum.

It is well known that the penalty term in the logarithmic-barrier subproblems (4.25) avoids the combinatorial aspect of nonlinear programs [23, 57], and that the convergence of  $\{\mu_k\}$  to a solution of (4.24) requires that  $\frac{\tau}{c(\mu)} \rightarrow z$ . However, when  $c(\mu) \rightarrow 0$ , it is easy to see that  $\frac{\tau}{c(\mu)^2} \rightarrow \infty$ , which implies that the Hessian of the logarithmic barrier function becomes very ill-conditioned. Moreover, once we introduce the Lagrange multiplier  $z$ , it can be shown that the perturbed KKT conditions for primal-dual methods are equivalent to the KKT conditions for logarithmic-barrier function problem, in the sense that they have the same solutions [23].

However, the roles that the multipliers play in the removal of the inherent ill-conditioning are quite distinct, and the primal-dual Newton interior-point method will produce different iterates compared to those produced by the Newton method applied to the KKT conditions for the logarithmic barrier subproblem (4.25) (e.g. [23, 57, 25]). To make a clear distinction between the two interior point methodologies, we decrease the value of the barrier parameter  $\tau$  defined in (4.26) iteratively,

$$\tau_{k+1} = \kappa_k \tau_k, \quad \kappa_k \in (0, 1), \quad \tau_0 = \kappa \eta, \quad (4.30)$$

as mentioned in [57].

### 4.3.2 Primal-dual interior point method

To solve problem (4.24) we define the *primal-dual Lagrangian*:

$$\mathcal{L}_P(\mu, z_1, z_2) = \mathcal{J}(\mu) - c_1(\mu)^T z_1 - c_2(\mu)^T z_2, \quad (4.31)$$

where  $c_1(\mu), c_2(\mu) > 0$ , and  $z_1, z_2 > 0$  are the Lagrange multipliers associated to the inequality constraints. For a given *perturbation parameter*  $\tau > 0$ , which can be rescaled for affine invariance as in (4.26), the perturbed KKT conditions are

$$F_P(\mu, z_1, z_2; \tau) = \begin{pmatrix} \nabla_\mu \mathcal{L}_P \\ C_1 Z_1 e - \tau e \\ C_2 Z_2 e - \tau e \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \quad (c_1, c_2, z_1, z_2) > 0, \quad (4.32)$$

and the Jacobian of  $F_P$  is

$$F'_P(\mu, z_1, z_2; \tau) = \begin{pmatrix} H & -I & I \\ Z_1 & C_1 & 0 \\ -Z_2 & 0 & C_2 \end{pmatrix}, \quad (4.33)$$

where  $\nabla_\mu \mathcal{L}_P = g - z_1 + z_2$ ,  $Z_1 = \text{diag}(z_1)$ ,  $Z_2 = \text{diag}(z_2)$ , and  $g$  and  $H$  are defined as in (4.14) and (4.18), respectively. The lower two blocks in equation (4.32) are called the *perturbed complementarity conditions*. They are a fundamental ingredient of primal-dual interior point methods.

In this method, the path-following strategy to solve problem (4.24) proceeds as follows: For a  $\tau > 0$  and an initial guess  $(\mu, z_1, z_2) > 0$  from the interior of the feasible parameter range, we apply Newton's method to the perturbed KKT conditions (4.32), which leads us to the linear system:

$$F'_P(\mu, z_1, z_2; \tau) \begin{pmatrix} \Delta\mu \\ \Delta z_1 \\ \Delta z_2 \end{pmatrix} = -F_P(\mu, z_1, z_2; \tau), \quad (4.34)$$

where the solution improves as the perturbation parameter  $\tau$  approaches zero.

Note that the Jacobian matrix  $F'_P$  occurring in system (4.34) is non-symmetric and fundamentally indefinite. Thus, to facilitate a numerical solution we decouple the system as follows: From the second and third block of equations in (4.34), we infer,

$$\Delta z_1 = -z_1 + C_1^{-1}(\tau e - Z_1 \Delta\mu), \quad (4.35)$$

$$\Delta z_2 = -z_2 + C_2^{-1}(\tau e + Z_2 \Delta\mu). \quad (4.36)$$

Then, substituting  $\Delta z_1$  and  $\Delta z_2$  into the first block of equations, we obtain

$$(H + C_P) \Delta\mu = -(g + d_P), \quad (4.37)$$

where  $C_P = C_1^{-1}Z_1 + C_2^{-1}Z_2$  and  $d_P = -\tau(c_1^{-1} - c_2^{-1})$ . Observe that once  $\Delta\mu$  is found, then  $\Delta z_1$  and  $\Delta z_2$  are calculated by using simple arithmetic operations, therefore the error when computing  $\Delta\mu$  is the same wheter we solve (4.34) or (4.37).

The new system of equations (4.37) has several advantages over system (4.34). First, the coefficient matrix in system (4.37) is symmetric, and its size is reduced considerably. Second, we increase the likelihood that the matrix  $H + C_P$  is positive semi-definite far away from the solution, since  $C_P$  is a positive-definite diagonal matrix. Therefore, we can apply Algorithm 1 for solving (4.37) with the hope of solving a better conditioned problem compared to (4.21). Also, it is important to realize that for  $\tau = 0$ , equations (4.32) are the KKT conditions, and for  $\tau \neq 0$  it can be shown that (4.32) is equivalent to (4.27) [23].

One issue that has to be addressed with primal-dual interior point methods is the choice of initial values for the dual variables  $z_1$  and  $z_2$ . To maintain affine invariance, we can initialize

them to

$$z_1 = z_2 = \kappa \|g_0\| e, \quad (4.38)$$

where  $\kappa \ll 1$  is a non-dimensional parameter (see Section 4.2.3).

### 4.3.3 Primal-dual active set method

The logarithmic barrier and primal-dual interior point methods integrate the inequality bounds for  $\mu$  (4.24) through terms that require the solution to be strictly feasible, that is, never touching the bounds. The primal-dual active set method, on the other hand, allows part or all of the solution to be equal to the bounds [33] which can be realized within the Newton step (4.16) by specifying the following subproblem [32],

$$\begin{aligned} \Delta\mu &= \arg \min K(\bar{\mu}) = \frac{1}{2}\bar{\mu}^T H \bar{\mu} + \bar{\mu}^T g \\ &\text{subject to } a \leq \mu_k + \Delta\mu \leq b, \end{aligned} \quad (4.39)$$

where the Hessian  $H$  and gradient  $g$  correspond to the current iterate  $\mu_k$ . The first-order necessary condition for a constrained minimizer  $\mu^* \in M_{\text{ad}}$  of  $K(\bar{\mu})$  can be written as

$$\partial_{\mu^*} K(\bar{\mu} - \mu^*) \geq 0 \quad \text{for all } \bar{\mu} \in M_{\text{ad}}, \quad (4.40)$$

where the admissible set is defined as  $M_{\text{ad}} = \{\bar{\mu} | a \leq \mu_k + \bar{\mu} \leq b\}$ . We can now combine (4.39) and (4.40) to obtain a necessary equation for  $\mu^*$ ,

$$\lambda^T (\bar{\mu} - \mu^*) \leq 0 \quad \text{for all } \bar{\mu} \in M_{\text{ad}} \quad \text{with} \quad \lambda = -(H\mu^* + g). \quad (4.41)$$

For each component  $i$ , of the corresponding vectors this formulation leads to different results in the following three cases,

$$\begin{aligned} -c_1(\mu_k)_i &= \mu_i^* && \Rightarrow \lambda_i \leq 0, \quad (\text{lower active set } a) \\ -c_1(\mu_k)_i < \mu_i^* < c_2(\mu_k)_i && \Rightarrow \lambda_i = 0, \quad (\text{inactive set } I) \\ \mu_i^* &= c_2(\mu_k)_i && \Rightarrow \lambda_i \geq 0, \quad (\text{upper active set } b). \end{aligned} \quad (4.42)$$

Hence, (4.41) is equivalent to the following system (called *optimality system*):

$$\begin{aligned} H\mu^* + g &= -\lambda, \\ -c_1(\mu_k) &\leq \mu^* \leq c_2(\mu_k), \\ [\lambda_i(\mu^* + c_1(\mu_k))_i = 0 \wedge \lambda_i \leq 0] &\vee [\lambda_i(\mu^* - c_2(\mu_k))_i = 0 \wedge \lambda_i \geq 0]. \end{aligned} \quad (4.43)$$

These complementarity conditions can be expressed through the single equation

$$\lambda = \min\{0, \lambda + \hat{c}(\mu^* + c_1(\mu_k))\} + \max\{0, \lambda + \hat{c}(\mu^* - c_2(\mu_k))\}, \quad (4.44)$$

with an arbitrary constant  $\hat{c} > 0$  [33]. This identity allows to choose indices  $i$  in the lower and upper active sets  $A_a$  and  $A_b$  based on whether the minimum or maximum term are nonzeros. An index belongs to the inactive set  $I$  if neither condition holds and  $\lambda_i = 0$ . After reordering the indices by the inactive and lower and upper active sets, the Newton system takes the form

$$\begin{pmatrix} H_{II} & H_{Ia} & H_{Ib} \\ H_{aI} & H_{aa} & H_{ab} \\ H_{bI} & H_{ba} & H_{bb} \end{pmatrix} \begin{pmatrix} \Delta\mu_I \\ -c_1(\mu_k)_a \\ c_2(\mu_k)_b \end{pmatrix} = - \begin{pmatrix} g_I \\ g_a \\ g_b \end{pmatrix} - \begin{pmatrix} 0_I \\ \lambda_a \\ \lambda_b \end{pmatrix}. \quad (4.45)$$

We obtain  $\Delta\mu_I$  by executing the conjugate gradient method on the inactive block  $H_{II}$ , setting all other rows and columns to zero, and then use this result to compute  $\lambda$  in a postprocessing step. Notice that for this method the Hessian and gradient are not modified, that is  $C_A = 0$  and  $d_A = 0$  in the context of (4.29) or (4.37). Algorithm 2 iterates this procedure to solve (4.39). In practice, we terminate the algorithm after a certain number of primal-dual iterations even if the active sets have not converged. The choice of the constant  $\hat{c}$  is irrelevant for the final active sets, its value is chosen here with the purpose of rendering the algorithm affine-invariant. It can be seen from this formulation that the primal-dual active set method requires neither the selection of initial values nor the tuning of tolerances, and that it does not rely on a path-following strategy. On the other hand, it may be more expensive than a primal-dual interior point method, due to the loop that aims to converge the active sets and is executed for each Newton step.

## 4.4 Algorithms for constrained optimization

All three constrained optimization schemes described above are based on Newton's method. They differ by distinct modifications to the Newton system of equations and its right hand side, and are traditionally combined with different globalization methods that take into account the inequality bounds. In this section, we propose a generic scheme that accommodates the

---

**Algorithm 2** Compute primal-dual active set Newton step  $\mu_{k+1}$  given  $\mu_k, a, b$

---

- 1: Choose  $\Delta\mu^0 = 0, \lambda^0 = 0, c = 0.01 \times \|g_0\| / \|b - a\|_\infty$
  - 2: **for**  $\ell = 0, 1, \dots$  **do**
  - 3:   Set lower active set  $A_a^\ell \leftarrow \{i | (\lambda^\ell + \hat{c}(\Delta\mu^\ell + c_1(\mu_k)))_i \leq 0\}$
  - 4:   Set upper active set  $A_b^\ell \leftarrow \{i | (\lambda^\ell + \hat{c}(\Delta\mu^\ell - c_2(\mu_k)))_i \geq 0\}$
  - 5:   **if** active sets unchanged since previous iteration, **return**  $\mu_{k+1} \leftarrow \mu_k + \Delta\mu^\ell$
  - 6:   Set inactive set  $I^\ell \leftarrow \{i\} \setminus (A_a^\ell \cup A_b^\ell)$
  - 7:   Use CG to solve (4.45) for  $\Delta\mu^{\ell+1}$  and  $\lambda^{\ell+1}$ , thus enforcing (4.42)
  - 8: **end for**
- 

specifics of each constrained optimization strategy with only minor adaptations, with the goal of facilitating the implementation of all of these different approaches.

#### 4.4.1 Generic Newton-Krylov method

Using Newton's method, the parameter values  $\mu_k$  are updated iteratively. Each of these (outer) Newton iterations determines a step  $\Delta\mu$  through the solution of the system

$$(H + C) \Delta\mu = -(g + d), \quad (4.46)$$

which is performed using an (inner) iterative method from the family of Krylov solvers, e.g. by using the method of conjugate gradients (CG). The diagonal matrix  $C$  and the vector  $d$  depend on the constrained optimization scheme selected; the corresponding expressions been derived in Section 4.3. For the logarithmic-barrier and primal-dual active set methods,  $\Delta\mu$  is just the direction we are looking for, which we denote by  $\Delta v = \Delta\mu$  for future reference. For the primal-dual interior point method, complete information on the direction  $\Delta v$  is obtained after we compute the Lagrange multipliers for the bounds by (4.35) and (4.36); in other words, in this case the complete Newton-Krylov direction is  $\Delta v = (\Delta\mu, \Delta z_1, \Delta z_2)$ . We use the unified notation  $v_k$  below for the Newton iterates (which include  $\mu_k$ ).

The step  $\Delta\mu$  produced in each Newton iteration, however, is not always suitable to determine a new parameter value  $\mu_{k+1}$  due to the following reasons:

- Inequality constraints: The proposed parameter update  $\mu_{k+1} = \mu_k + \Delta\mu$  violates the inequality constraints  $a \leq \mu_{k+1} \leq b$ .
- Nonlinearity: The proposed step either increases the value of the objective functional  $\mathcal{J}(\mu_{k+1})$  or does not decrease it sufficiently.

In the remainder of this section, we describe components of a generic optimization framework that addresses these issues.

#### 4.4.2 Force positivity

For any proposed Newton step  $\Delta\mu$  as computed from (4.46), the constraint condition

$$a \leq \mu_k + \Delta\mu \leq b, \quad (4.47)$$

may be violated for one or more vector components. A possible remedy is to compute a correction factor  $\hat{\alpha} \leq 1$  and scale the Newton step as  $\hat{\alpha}\Delta\mu$  to ensure the positivity of  $c(\mu)$  and thus (4.47). A frequently used construction is

$$\hat{\alpha}_\mu = \lambda \min \left\{ \frac{-1}{\min\{C_1^{-1}\Delta\mu, -1\}}, \frac{-1}{\min\{-C_2^{-1}\Delta\mu, -1\}} \right\}, \quad (4.48)$$

using a given parameter  $0 < \lambda \leq 1$ . Here the first term limits the step for negative components of  $\Delta\mu$  that could overstep the lower bound, and the second term limits the step for positive components and the upper bound. For the primal-dual interior point method, the step in the dual variables  $z$  also needs to be controlled to keep them positive and away from zero, which can be done by specifying

$$\hat{\alpha}_z = \lambda \min \left\{ \frac{-1}{\min\{Z_1^{-1}\Delta z_1, -1\}}, \frac{-1}{\min\{Z_2^{-1}\Delta z_2, -1\}} \right\} \quad (4.49)$$

and taking  $\hat{\alpha} = \min\{\hat{\alpha}_\mu, \hat{\alpha}_z\}$ . By construction, it holds that  $\hat{\alpha} \leq \lambda$ , where overall step length is then often chosen using a given parameter  $0 < \lambda \leq 1$ ,

$$\hat{\alpha} = \min\{\lambda\hat{\alpha}_\mu, \lambda\hat{\alpha}_z\}, \quad (4.50)$$

the parameter  $\lambda$  must be strictly less than 1 for logarithmic-barrier and interior point methods [23, 57, 82]. While this construction is generally well accepted, it does not achieve a full

correction of the step length – even when the proposed parameter value is well away from the bounds; this can potentially reduce the convergence rate of Newton’s method, which, in the case of full step correction, is usually second-order.

To address this point, we begin by postulating the following criteria for  $\hat{\alpha}_\mu$  that keep the step away from the bounds when it is too close,

$$\hat{\alpha}_\mu \Delta\mu \geq -\lambda(\mu - a), \quad \hat{\alpha}_\mu \Delta\mu \leq \lambda(b - \mu). \quad (4.51)$$

Note that (4.48) already ensures these criteria, and that they are affine-invariant by construction. The first equation is always satisfied for non-negative components of  $\Delta\mu$ , and the second equation is always satisfied for non-positive components. Furthermore, whenever  $\hat{\alpha}_\mu > 1$  satisfies (4.51), a smaller value for  $\hat{\alpha}_\mu$  can be chosen instead without violating these conditions. For the remaining cases, we can reformulate these conditions as

$$\hat{\alpha}_\mu \leq \frac{-\lambda}{C_1^{-1}\Delta\mu}, \quad \hat{\alpha}_\mu \leq \frac{-\lambda}{-C_2^{-1}\Delta\mu}, \quad \hat{\alpha}_\mu \leq \frac{-\lambda}{-\lambda}, \quad (4.52)$$

which we can combine into

$$\hat{\alpha}_\mu \leq \lambda \min \left\{ \frac{-1}{\min\{C_1^{-1}\Delta\mu, -\lambda\}}, \frac{-1}{\min\{-C_2^{-1}\Delta\mu, -\lambda\}} \right\}. \quad (4.53)$$

This result is just a slight modification to (4.48). When the step reaches the boundary, the result for  $\hat{\alpha}$  is the same for both formulas. The added benefit of (4.53) is that it permits  $\hat{\alpha}_\mu = 1$  on the interior even if  $\lambda < 1$ .

The corresponding derivation of the step length correction for  $\hat{\alpha}_z$  is similar, and the complete formula can be restated equivalently as

$$\hat{\alpha} = \lambda / \max \left\{ \lambda, -C_1^{-1}\Delta\mu, C_2^{-1}\Delta\mu, -Z_1^{-1}\Delta z_1, -Z_2^{-1}\Delta z_2 \right\}, \quad (4.54)$$

where the  $z$  variables are only present for the primal-dual interior point method.

### 4.4.3 Sufficient decrease

In order to monitor progress of the iterates  $v_k$  to the optimal solution, we use a line search strategy that can reduce the size of the Newton step when necessary. We judge the progress by

evaluating a *merit function*  $\mathcal{M} = \mathcal{M}(v)$ . The role of this function is to help determining a step length  $\alpha_k \in (0, \hat{\alpha}]$  that provides sufficient decrease of the objective function, while satisfying the inequality constraints. This is possible whenever the Newton-Krylov direction is a descent direction for the merit function  $\mathcal{M}$  at the current interior point  $v_k$ . We generally employ the merit function in the context of the *Armijo condition*, where the step length  $\alpha_k$  is chosen such that

$$\mathcal{M}(v_k + \alpha_k \Delta v_k) < \mathcal{M}(v_k) + 10^{-4} \alpha_k \nabla \mathcal{M}(v_k)^T \Delta v_k. \quad (4.55)$$

The backtracking Armijo line search procedure is presented in Algorithm 3.

---

**Algorithm 3** Backtracking line search

---

```

1: for  $t = 0, 1, \dots, l_{\max}$  do
2:   Set  $\alpha = \left(\frac{1}{2}\right)^t \hat{\alpha}$ .
3:   if  $\mathcal{M}(v_k + \alpha \Delta v_k) < \mathcal{M}(v_k) + 10^{-4} \alpha \nabla \mathcal{M}(v_k)^T \Delta v_k$ , then
4:     return  $\alpha_k = \alpha$ 
5:   end if
6: end for
```

---

The merit function we use for the logarithmic-barrier and for the active set methods is the objective function itself, i.e., we take  $\mathcal{M} = \mathcal{J}$ . In the case of primal-dual interior point methods we select the modified augmented Lagrangian function as introduced by Argaez and Tapia [51]:

$$\mathcal{M}(v) = \mathcal{M}(\mu, z; \rho) = \mathcal{L}_P(\mu, z) - \theta + \rho \varphi_\tau(\mu, z), \quad (4.56)$$

where  $\rho$  is a nonnegative penalty parameter,  $\theta$  needs to be removed to guarantee affine invariance and the last term is the penalty function to the complementarity condition

$$\varphi_\tau(\mu, z) = c_1^T z_1 - \tau \sum_{i=1}^n \log \left( \frac{C_1 Z_1}{\eta} e \right)_i + c_2^T z_2 - \tau \sum_{i=1}^n \log \left( \frac{C_2 Z_2}{\eta} e \right)_i. \quad (4.57)$$

It can be shown that the Newton-Krylov direction is a descent direction for the penalty function. This property allows us to find a  $\rho$  such that the Krylov-Newton direction is a descent direction for the merit function. In line with this objective, and to avoid the possibility of

failure in the backtracking line search Algorithm 3, we update the penalty parameter  $\rho$  in the following manner. Given  $\tilde{c} > 0$  and the current penalty parameter  $\rho_k$ , we take

$$\rho_{k+1} = \begin{cases} \tilde{\rho}_{k+1} + \tilde{c}, & \text{if } \tilde{\rho}_{k+1} + \tilde{c} > \rho_k, \\ \tilde{\rho}_{k+1} + c, & \text{otherwise,} \end{cases}$$

where  $\tilde{\rho}_{k+1} = \frac{\nabla \mathcal{L}_P(\mu, z)^T \Delta v_k}{|\nabla \varphi_\tau(\mu, z)^T \Delta v_k|}$ ,  $c = \rho_k - \tilde{\rho}_{k+1}$ , and  $c > \tilde{c}$ . In this work we use  $\tilde{c} = 2$ .

#### 4.4.4 Unified algorithm framework

We summarize our generic constrained optimization procedure in Algorithm 4. This general algorithm can accommodate the three strategies described in Section 4.3 by suitable selection of local subalgorithms. In Table 4.1, we summarize the different choices of the parameters, the merit functions, the matrices and the vectors used for each strategy. We compute the parameter  $\tau$  as presented in Sections 4.3.1 and 4.3.2, which are standard expressions for interior point methods [23, 57].

Table 4.1: Selection of parameters and subalgorithms for Algorithm 4, implementing the logarithmic barrier, primal-dual interior point, and primal-dual active set constrained optimization methods in a unified framework.

Method	Log-barrier	PD interior point	PD active set
$\tau$	$\tau_{k+1} = \kappa \tau_k, \tau_0 = \kappa \eta$	$\tau = \kappa \frac{z^T c(\mu)}{2n}$	—
$C$	$\tau (C_1^{-2} + C_2^{-2})$	$C_1^{-1} Z_1 + C_2^{-1} Z_2$	0
$d$	$\tau (c_1^{-1} - c_2^{-1})$	$\tau (c_1^{-1} - c_2^{-1})$	0
$\Delta \mu$	CG	CG	Algorithm 2
$v$	$\mu$	$(\mu, z_1, z_2)$	$\mu$
$\mathcal{M}$	$\mathcal{J}$	$\mathcal{L}_P(\mu, z) + \rho \varphi_\tau(\mu, z)$	$\mathcal{J}$

---

**Algorithm 4** Newton-Krylov constrained optimization with line search

---

- 1: Choose an initial interior point  $v_0$ .
- 2: **for**  $k = 0, 1, 2, \dots$  until convergence **do**
- 3:   (For LB and PDIP) Choose parameter  $\tau_k = \kappa \frac{c^T z}{2n}$ , see Section 4.3.1 or 4.3.2.
- 4:   (Newton-Krylov direction) Solve for  $\Delta\mu$  and build  $v$  according to Table 4.1:

$$(H + C) \Delta\mu = -(g + d).$$

- 5:   (Force positivity) Select  $\hat{\alpha}$  from (4.54).
  - 6:   (Sufficient decrease) Compute  $\alpha_k \in (0, \hat{\alpha}]$  using Algorithm 3.
  - 7:   (Newton step) Set  $v_{k+1} = v_k + \alpha_k \Delta v$ .
  - 8: **end for**
- 

## 4.5 Numerical results for an inverse wave propagation problem

In this section we apply the logarithmic-barrier, primal-dual interior point, and primal-dual active set constrained optimization strategies to an example inverse problem derived from seismic wave propagation. We briefly summarize the derivation of the forward and associated inverse operators and proceed with describing several numerical experiments. Our solvers and algorithms are implemented in MATLAB, and executed on a Dell XPS laptop.

### 4.5.1 A one-dimensional wave propagation example

The forward problem is defined by computing the wavefield  $u(x, t)$  from the equations

$$\rho(x) \frac{\partial^2}{\partial t^2} u(x, t) = \frac{\partial}{\partial x} \mu(x) \frac{\partial}{\partial x} u(x, t) + f(x, t) \quad \text{in } (0, L) \times (0, T), \quad (4.58)$$

$$u_{t=0} = 0, \quad \dot{u}_{t=0} = 0, \quad (4.59)$$

given a density  $\rho(x)$ , an elastic modulus  $\mu(x)$ , and a force term  $f(x, t) = \delta(x, 0)f_0(t)$  which uses the second derivative of the cdf of the Gaussian distribution as source time function  $f_0$ . We specify reflection boundary conditions at the top of the material column and absorbing

boundary conditions at the bottom,

$$\mu(x) \frac{\partial}{\partial x} u(x, t) = 0, \quad x = 0, \quad (4.60)$$

$$\mu(x) \frac{\partial}{\partial x} u(x, t) = -\sqrt{\rho(x)\mu(x)} \frac{\partial}{\partial t} u(x, t), \quad x = L. \quad (4.61)$$

We solve these equations numerically by discretizing all variables with piecewise linear finite elements in space and using central differences for the time derivative. Once we understand  $u$  and  $f$  as vectors with indices referring to discrete space and time points  $x_k$  and  $t_n$ , this equation complies with the abstract form  $Au = f$  (4.11) of the forward equation. The linear operator  $A$  does not need to be inverted directly since using the central difference scheme to discretize in time allows for an explicit time stepping procedure. The discretization also determines how the values  $\mu_k = \mu(x_k)$  at the grid points enter the forward operator  $A = A_\mu$  in a nonlinear way.

To compare the computed results with experimental data  $u^*$ , we use a least-squares misfit functional that models measurements at the top end of the domain,

$$B(u) = \frac{1}{2TN_u} \int_0^T (u(0, t) - u^*(0, t))^2 dt, \quad (4.62)$$

and penalize variations in the medium by a Tikhonov regularization term

$$R(\mu) = \frac{1}{2LN_\mu} \int_0^L (\nabla \mu)^2 dx, \quad (4.63)$$

both of which match the general form (4.6) after discretization. We introduce the normalization factors  $N_u$  and  $N_\mu$  to remove dependencies on the scale of the physical quantities  $u$  and  $\mu$ . Since both terms are quadratic forms, their derivatives have a simple structure. In particular, the right hand side of the adjoint equation (4.12) is given by the linear misfit between simulation and observation,

$$A^*p = \nabla B = \frac{1}{TN_u} (u(x, t) - u^*(x, t)) \delta(x, 0). \quad (4.64)$$

The adjoint equation can be derived in the abstract setting formulated in Section 4.2 or, equivalently, by substituting the definitions in this section. It follows that the adjoint variable  $p$  satisfies a wave equation analogous to (4.58) which runs backward in time. In the discretized setting the reversal of the time direction follows from the transposition of the

matrix  $A$ . Deriving the adjoint in the discrete setting is more tedious but offers the advantage that the discrete expressions for the gradient and the Hessian represent the derivatives exactly, which avoids the danger of inconsistent discretizations for the adjoint equation and operator derivatives (4.15), (4.19), and (4.20) [28]. For a full derivation of all first- and second-order derivative terms in three space dimensions we refer to [24].

In our numerical experiments, we derive the observed wavefield  $u^*$  by solving the forward equation with elasticity profiles  $\mu^*$  derived from borehole measurements - which constitute the target model. We then choose an initial guess  $\mu_0$  and apply the inversion procedures summarized earlier in Algorithm 4 to iteratively reconstruct the target profile.

### 4.5.2 Error measures, convergence and affine invariance

The strong nonconvexity of the cost functional often requires us to solve a series of inverse problems on subsequently finer grids and with increasing source frequencies [41, 66]. The number of Newton iterations on each grid can be increased as well in the process, since results corresponding to the earlier grid levels serve as initial guesses for the next and do not need to be fully converged.

The opportunity to use multiple grids, in turn, leads to the challenge of choosing grid-independent error measures and stopping criteria. A systematic way of achieving grid-independence is to use discrete norms derived from the original infinite-dimensional formulation of the problem in function space. This consideration leads to specifying the root mean square error in the observations as

$$\text{RMS} = \frac{\|u - u^*\|}{\|u^*\|} = \sqrt{\frac{B(u)}{B(0)}}, \quad (4.65)$$

where the norms  $\|\cdot\|$  denote discretizations of the  $\ell^2$  norm. In a finite element discretization, this usually amounts to using a discrete inner product weighted by a mass matrix whose entries reflect the size of the grid cells or time intervals, respectively.

Similarly, we define the relative error in the inversion variable and the relative progress per Newton iteration:

$$\text{Error} = \frac{\|\mu_k - \mu^*\|}{\|\mu^*\|}, \quad \text{Progress} = \frac{\|\mu_{k+1} - \mu_k\|}{\|\mu_{k+1}\|}, \quad (4.66)$$

respectively, again implying the use of discretized function space norms. By construction, these measures are invariant with respect to a rescaling of the wave field or inversion variable.

### 4.5.3 Basic tests

As mentioned in section Section 4.2.3, it is often desirable to guarantee invariance of the optimization algorithms with respect to the scale of the objective functional. We tested our algorithms with a simplified version of one synthetic case described below as Experiment A1. We verify that for all the methods implemented here, the history of the iterates  $\mu_k$  remains invariant when an affine transformation  $\mathcal{J} \rightarrow \zeta \mathcal{J} + \theta$  is applied with  $\zeta = 10^6$  and  $\theta = 10^5$ . We keep track of the behavior of each method in Tables 4.2–4.4, with and without the affine transformation applied over  $\mathcal{J}$  and accordingly modified derivatives  $g$  and  $H$ . We pay special attention to the quantities RMS and Error. We summarize that as the iterations proceed, roundoff error accumulates in solving (4.21), which can be noticed specially in the computation of  $\|g\|$ . In combination with condition numbers of the Hession between  $10^4$  and  $10^7$ , this restricts us from keeping all significant digits in the computation of the RMS and Error. Roundoff error can also change the stopping criteria and introduce deviations in the number of inner iterations, which leads to a quantitatively different behavior from that point on. The general convergence behavior, however, is preserved by our affine invariance variants in each method, up to the final iteration.

Table 4.2: Iteration history for the logarithmic barrier method applied to a variant of a small synthetic test problem (described below as experiment A1). The maximum number of CG iterations per Newton step allowed is 30, the regularization parameter  $\beta = 0$ . We iterate until we reach 10 Newton iterations. Top: Unscaled cost functional; note that in line 6 the cost increases which we allow if it occurs only once. Bottom: Affine invariant transformation  $\mathcal{J} \rightarrow 10^6 \mathcal{J} + 10^5$ .

Newton iter. $k$	Gradient $\ g\ $	RMS (4.65)	Error (4.66)	Cost $\mathcal{J}$
Initial	$1.84741 \times 10^{-2}$	$2.021138 \times 10^{-1}$	-	$3.78100 \times 10^{-3}$
1	$1.25581 \times 10^{-2}$	$1.058291 \times 10^{-1}$	$1.1480727 \times 10^{-1}$	$1.036633 \times 10^{-3}$
2	$2.15114 \times 10^{-3}$	$3.957972 \times 10^{-2}$	$1.128317 \times 10^{-1}$	$1.449974 \times 10^{-2}$
3	$2.80636 \times 10^{-3}$	$2.656788 \times 10^{-2}$	$1.055603 \times 10^{-1}$	$6.533241 \times 10^{-5}$
4	$3.40043 \times 10^{-4}$	$1.210047 \times 10^{-2}$	$9.982191 \times 10^{-2}$	$1.355251 \times 10^{-5}$
5	$1.98697 \times 10^{-4}$	$8.844359 \times 10^{-3}$	$8.909005 \times 10^{-2}$	$7.240150 \times 10^{-6}$
6	$1.25733 \times 10^{-3}$	$1.173039 \times 10^{-2}$	$8.909005 \times 10^{-2}$	$1.273620 \times 10^{-5}$
7	$7.48047 \times 10^{-5}$	$3.655517 \times 10^{-3}$	$8.440794 \times 10^{-2}$	$1.236837 \times 10^{-6}$
8	$2.69023 \times 10^{-4}$	$2.820189 \times 10^{-3}$	$8.028388 \times 10^{-2}$	$7.361585 \times 10^{-7}$
9	$2.45153 \times 10^{-5}$	$1.615279 \times 10^{-3}$	$7.890529 \times 10^{-2}$	$2.414961 \times 10^{-7}$
10	$6.01986 \times 10^{-5}$	$1.411458 \times 10^{-3}$	$7.945724 \times 10^{-2}$	$1.843956 \times 10^{-7}$
Initial	18473.9	$2.0211390 \times 10^{-1}$	-	$1.037810 \times 10^5$
1	12557.5	$1.058269 \times 10^{-1}$	$1.148081 \times 10^{-1}$	$1.010366 \times 10^5$
2	2151.06	$3.957902 \times 10^{-2}$	$1.128325 \times 10^{-1}$	$1.001449 \times 10^5$
3	2806.41	$2.656659 \times 10^{-2}$	$1.055604 \times 10^{-1}$	$1.000653 \times 10^5$
4	339.791	$1.209948 \times 10^{-2}$	$9.982318 \times 10^{-2}$	$1.000136 \times 10^5$
5	198.071	$8.819186 \times 10^{-2}$	$8.908113 \times 10^{-2}$	$1.000072 \times 10^5$
6	1238.69	$1.157256 \times 10^{-2}$	$8.908113 \times 10^{-2}$	$1.000124 \times 10^5$
7	74.6379	$3.640246 \times 10^{-3}$	$8.442551 \times 10^{-2}$	$1.000012 \times 10^5$
8	264.951	$2.794382 \times 10^{-3}$	$8.029006 \times 10^{-2}$	$1.000007 \times 10^5$
9	24.5322	$1.615692 \times 10^{-3}$	$7.891357 \times 10^{-2}$	$1.000002 \times 10^5$
10	59.9867	$1.411715 \times 10^{-3}$	$7.246918 \times 10^{-2}$	$1.000001 \times 10^5$

Table 4.3: Iteration history for primal-dual interior point method applied to the same synthetic test problem as before. Bottom: Using affine transformation  $\mathcal{J} \rightarrow 10^6 \mathcal{J} + 10^5$ .

Newton iter. $k$	Gradient $\ g\ $	RMS (4.65)	Error (4.66)	Cost $\mathcal{J}$
Initial	$2.45642 \times 10^{-2}$	$2.952261 \times 10^{-1}$	-	$8.067227 \times 10^{-3}$
1	$2.21802 \times 10^{-2}$	$2.589252 \times 10^{-1}$	$1.187725 \times 10^{-1}$	$6.205311 \times 10^{-3}$
2	$1.24403 \times 10^{-2}$	$1.652886 \times 10^{-1}$	$1.180138 \times 10^{-1}$	$2.528720 \times 10^{-3}$
3	$1.65121 \times 10^{-3}$	$8.228086 \times 10^{-2}$	$1.161171 \times 10^{-1}$	$6.266319 \times 10^{-4}$
4	$4.67892 \times 10^{-4}$	$4.810599 \times 10^{-2}$	$1.126292 \times 10^{-1}$	$2.141970 \times 10^{-4}$
5	$5.93259 \times 10^{-4}$	$3.796496 \times 10^{-2}$	$1.020278 \times 10^{-1}$	$1.334077 \times 10^{-4}$
6	$5.90589 \times 10^{-4}$	$3.166605 \times 10^{-2}$	$9.490178 \times 10^{-2}$	$9.281169 \times 10^{-5}$
7	$6.88751 \times 10^{-4}$	$2.036024 \times 10^{-2}$	$8.802908 \times 10^{-2}$	$3.836901 \times 10^{-5}$
8	$3.62015 \times 10^{-4}$	$9.585026 \times 10^{-3}$	$8.733388 \times 10^{-2}$	$8.503574 \times 10^{-6}$
9	$1.39444 \times 10^{-4}$	$5.357998 \times 10^{-3}$	$8.575199 \times 10^{-2}$	$2.657174 \times 10^{-6}$
10	$1.06702 \times 10^{-4}$	$3.200906 \times 10^{-3}$	$8.454366 \times 10^{-2}$	$9.483331 \times 10^{-7}$
Initial	24564.2	$2.952260 \times 10^{-1}$	-	$1.080672 \times 10^5$
1	22180.2	$2.589252 \times 10^{-1}$	$1.187725 \times 10^{-1}$	$1.062053 \times 10^5$
2	12440.3	$1.652886 \times 10^{-1}$	$1.180138 \times 10^{-1}$	$1.025287 \times 10^5$
3	1651.21	$8.228085 \times 10^{-2}$	$1.161170 \times 10^{-1}$	$1.006266 \times 10^5$
4	467.892	$4.810599 \times 10^{-2}$	$1.126292 \times 10^{-1}$	$1.002141 \times 10^5$
5	593.332	$3.796503 \times 10^{-2}$	$1.020277 \times 10^{-1}$	$1.001334 \times 10^5$
6	709.297	$2.884863 \times 10^{-2}$	$9.245712 \times 10^{-2}$	$1.000770 \times 10^5$
7	663.746	$1.591146 \times 10^{-2}$	$8.684648 \times 10^{-2}$	$1.000234 \times 10^5$
8	258.205	$8.688763 \times 10^{-3}$	$8.541545 \times 10^{-2}$	$1.000069 \times 10^5$
9	192.452	$5.176112 \times 10^{-3}$	$8.283053 \times 10^{-2}$	$1.000024 \times 10^5$
10	44.0267	$3.111875 \times 10^{-3}$	$8.177534 \times 10^{-2}$	$1.000008 \times 10^5$

Table 4.4: Iteration history for primal-dual active sets method applied to the same synthetic test problem as before. Top: No affine invariance used. Bottom: Using affine invariance with  $\mathcal{J} \rightarrow 10^6 \mathcal{J} + 10^5$ .

Newton iter. $k$	Gradient $\ g\ $	RMS (4.65)	Error (4.66)	Cost $\mathcal{J}$
Initial	$1.75702 \times 10^{-2}$	$2.044775 \times 10^{-1}$	-	$3.869955 \times 10^{-3}$
1	$1.9734 \times 10^{-2}$	$1.567412 \times 10^{-1}$	$9.377728 \times 10^{-2}$	$2.273952 \times 10^{-3}$
2	$2.92844 \times 10^{-3}$	$5.005222 \times 10^{-2}$	$9.296957 \times 10^{-2}$	$2.318791 \times 10^{-4}$
3	$7.47150 \times 10^{-4}$	$2.436368 \times 10^{-2}$	$8.853534 \times 10^{-2}$	$5.494156 \times 10^{-5}$
4	$5.21693 \times 10^{-4}$	$1.360523 \times 10^{-2}$	$8.613289 \times 10^{-2}$	$1.713275 \times 10^{-5}$
5	$3.80963 \times 10^{-4}$	$8.167504 \times 10^{-3}$	$8.289152 \times 10^{-2}$	$6.174385 \times 10^{-6}$
6	$7.77978 \times 10^{-5}$	$4.069779 \times 10^{-3}$	$8.010989 \times 10^{-2}$	$1.533051 \times 10^{-6}$
7	$1.09757 \times 10^{-4}$	$2.142604 \times 10^{-3}$	$7.859931 \times 10^{-2}$	$4.249119 \times 10^{-7}$
8	$2.07089 \times 10^{-5}$	$1.466189 \times 10^{-3}$	$7.746315 \times 10^{-2}$	$1.989734 \times 10^{-7}$
9	$6.71712 \times 10^{-5}$	$1.250000 \times 10^{-3}$	$7.146393 \times 10^{-2}$	$1.446222 \times 10^{-7}$
10	$2.77166 \times 10^{-6}$	$9.321519 \times 10^{-4}$	$6.919903 \times 10^{-2}$	$8.042449 \times 10^{-8}$
Initial	17573.1	$2.044943 \times 10^{-1}$	-	$1.038705 \times 10^5$
1	19743.3	$1.567818 \times 10^{-1}$	$9.378016 \times 10^{-2}$	$1.022751 \times 10^5$
2	2929.99	$5.006202 \times 10^{-2}$	$9.297273 \times 10^{-2}$	$1.002319 \times 10^5$
3	747.306	$2.436796 \times 10^{-2}$	$8.853972 \times 10^{-2}$	$1.000549 \times 10^5$
4	522.42	$1.361592 \times 10^{-2}$	$8.613376 \times 10^{-2}$	$1.000171 \times 10^5$
5	374.483	$8.153999 \times 10^{-3}$	$8.289077 \times 10^{-2}$	$1.000061 \times 10^5$
6	80.9737	$4.041659 \times 10^{-3}$	$8.009615 \times 10^{-2}$	$1.000015 \times 10^5$
7	102.254	$2.131471 \times 10^{-3}$	$7.862638 \times 10^{-2}$	$1.000004 \times 10^5$
8	22.6526	$1.474639 \times 10^{-3}$	$7.748873 \times 10^{-2}$	$1.000002 \times 10^5$
9	19.4105	$1.218895 \times 10^{-3}$	$7.444587 \times 10^{-2}$	$1.000001 \times 10^5$
10	9.61763	$1.127813 \times 10^{-3}$	$7.363295 \times 10^{-2}$	$1.000001 \times 10^5$

Table 4.5: Iteration history for Newton's method applied to a small test problem. The maximum number of CG iterations per Newton step allowed is 18. Here we disable the inequality constraints on the inversion parameter  $\mu$  and set the regularization parameter  $\beta = 0$ . We use  $\eta = 10^{-3}$  and  $s = 0$  in (4.23) and iterate until we reach roundoff accuracy. Top: Enforcing positivity by using formula (4.48). Bottom: Using formula (4.54) instead which allows a step length of 1.

Newton iter. $k$	Gradient $\ g\ $	RMS (4.65)	Error (4.66)	Step $\ \Delta\mu\ $	Cost $\mathcal{J}$
Initial	$6.65 \times 10^{-4}$	$2.10 \times 10^{-1}$	-	-	$3.59 \times 10^{-4}$
1	$3.25 \times 10^{-4}$	$9.63 \times 10^{-1}$	$9.00 \times 10^{-2}$	$4.49975 \times 10^{-1}$	$7.52 \times 10^{-5}$
2	$1.62 \times 10^{-4}$	$7.62 \times 10^{-2}$	$1.01 \times 10^{-1}$	$9.9995 \times 10^{-1}$	$4.71 \times 10^{-5}$
3	$1.57 \times 10^{-4}$	$4.90 \times 10^{-2}$	$3.36 \times 10^{-2}$	$2.49875 \times 10^{-1}$	$1.95 \times 10^{-5}$
4	$8.84 \times 10^{-6}$	$5.02 \times 10^{-3}$	$1.36 \times 10^{-2}$	$9.9995 \times 10^{-1}$	$2.04 \times 10^{-7}$
5	$2.89 \times 10^{-7}$	$5.48 \times 10^{-4}$	$2.11 \times 10^{-3}$	$9.9995 \times 10^{-1}$	$2.45 \times 10^{-9}$
6	$6.98 \times 10^{-9}$	$8.10 \times 10^{-6}$	$4.21 \times 10^{-5}$	$9.9995 \times 10^{-1}$	$5.33 \times 10^{-13}$
7	$5.44 \times 10^{-11}$	$1.62 \times 10^{-7}$	$4.81 \times 10^{-7}$	$9.9995 \times 10^{-1}$	$2.12 \times 10^{-16}$
8	$7.79 \times 10^{-13}$	$2.31 \times 10^{-10}$	$2.55 \times 10^{-10}$	$9.9995 \times 10^{-1}$	$4.34 \times 10^{-22}$
9	$1.56 \times 10^{-15}$	$1.62 \times 10^{-12}$	$2.64 \times 10^{-12}$	$9.9995 \times 10^{-1}$	$2.14 \times 10^{-26}$
1	$3.25 \times 10^{-4}$	$9.64 \times 10^{-2}$	$9.00 \times 10^{-2}$	$5 \times 10^{-1}$	$7.53 \times 10^{-5}$
2	$1.63 \times 10^{-4}$	$7.61 \times 10^{-2}$	$1.01 \times 10^{-1}$	1	$4.70 \times 10^{-5}$
3	$1.53 \times 10^{-4}$	$4.80 \times 10^{-2}$	$3.42 \times 10^{-2}$	$2.5 \times 10^{-1}$	$1.87 \times 10^{-5}$
4	$7.95 \times 10^{-7}$	$4.44 \times 10^{-3}$	$1.08 \times 10^{-2}$	1	$1.60 \times 10^{-7}$
5	$1.32 \times 10^{-7}$	$1.98 \times 10^{-4}$	$7.55 \times 10^{-4}$	1	$3.18 \times 10^{-10}$
6	$3.59 \times 10^{-9}$	$1.38 \times 10^{-6}$	$5.16 \times 10^{-6}$	1	$1.55 \times 10^{-14}$
7	$2.65 \times 10^{-12}$	$7.25 \times 10^{-9}$	$2.15 \times 10^{-8}$	1	$4.27 \times 10^{-19}$
8	$2.51 \times 10^{-14}$	$7.51 \times 10^{-11}$	$4.02 \times 10^{-10}$	1	$4.58 \times 10^{-23}$
9	$1.39 \times 10^{-16}$	$2.64 \times 10^{-13}$	$1.44 \times 10^{-13}$	1	$5.66 \times 10^{-28}$

#### 4.5.4 Experiment A

For this and the subsequent experiments, each Krylov-Newton iteration computes a direction with a conjugate gradient algorithm using a maximum of 30 iterations, we use the stopping criteria (4.22). For each problem, we use a grid continuation approach, where the first approximation of  $\mu$  is conducted over a coarser model space grid using a fixed number of Krylov-Newton iterations, until a final solution is obtained for the finer grid. The level of discretization for  $\mu$  is determined by the number of grid points in the  $x$ -direction, defined by  $i_x = 2^j + 1$ .

In this experiment we have two simple problems, which we denote as A1 and A2; for these problems in Tables 4.6 through 4.9 we report the numerical results obtained for each optimization scheme applied, with and without regularization. For these and all tables presented in this section, the first row labels the algorithm being used, and the first column identifies the best RMS value, the relative error in the parameter  $\mu$  given by (4.66), the total number of conjugate gradient iterations, and the total computing time (in seconds) for each method to obtain a converged solution. In this experiment, the physical bounds for problems A1 and A2 are  $a = 1$  and  $b = 6$ ; see Figures 4.1 and 4.2. In problem A1, the target model parameter  $\mu^*$  always lies within the bounds, while in problem A2, we have that  $\mu^*$  has some components lying on the boundary, as illustrated in Figure 4.2. These target models are hard to resolve for some of the optimization schemes. The level of discretization for  $\mu$  in both cases corresponds to  $j = 3, 4$ , and  $5$ , where we allow 5 Newton-Krylov iterations for the coarser grids while we use 15 iterations for the final stage.

In Tables 4.6–4.7 and Tables 4.8–4.9, we present the numerical results obtained for the first two problems at the final stage of the continuation strategy. For the last stage of the algorithm, we show in Figures 4.1–4.2 the target model  $\mu^*$  in red, and final approximations for  $\mu$  and initial guesses (computed based on the previous level results) with colors that vary according to each method.

Table 4.6: Problem A1, no regularization applied ( $\beta = 0$ ). Interior point and active set methods give similar error and RMS, while interior point methods converge slightly faster than PDAS.

	Log-Barrier	Primal-Dual interior point	Primal-Dual Active Sets
RMS	$2.9851 \times 10^{-4}$	$2.2717 \times 10^{-4}$	$1.7859 \times 10^{-4}$
Error	$7.6869 \times 10^{-2}$	$7.9913 \times 10^{-2}$	$8.2931 \times 10^{-2}$
CG iter.	479	465	485
Time (sec)	194	181	223

Table 4.7: Problem A1, regularization applied ( $\beta = 10^{-6}$ ). The regularization causes all methods converge with similar RMS, while the Error for PDIP is less affected by the smoothing effect of the regularization. The error and the number of CG iterations increase considerably for PDAS.

	Log-Barrier	Primal-Dual interior point	Primal-Dual Active Sets
RMS	$3.0217 \times 10^{-3}$	$3.0865 \times 10^{-3}$	$3.1845 \times 10^{-3}$
Error	0.1552	0.1214	0.1647
CG iter.	490	329	655
Time (sec)	197	156	298

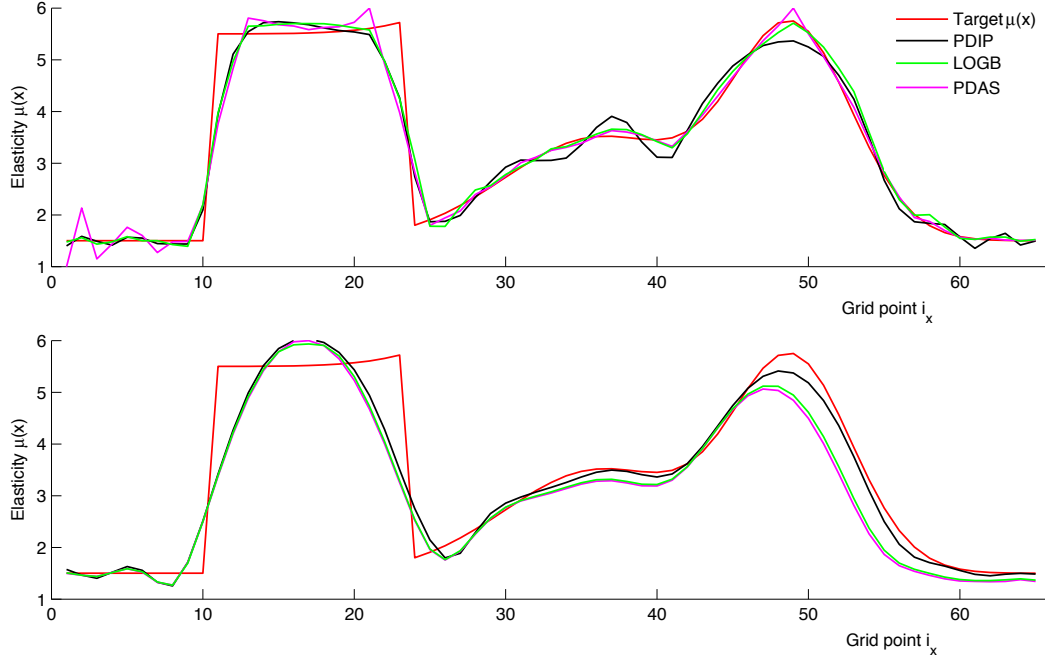


Figure 4.1: Numerical results for problem A1. Top: No regularization. Bottom: Regularization applied with  $\beta = 10^{-6}$ . In this problem, the target model parameter  $\mu^*$ , (in red), always lies within the bounds. The level of discretization for  $\mu$  corresponds to  $j = 3, 4$ , and 5 where we allow 5, Newton-Krylov iterations for the coarser grids and 15 for the final approximation. We display the final approximations for  $\mu$  with colors that vary according to each method.

Table 4.8: Problem A2, no regularization applied ( $\beta = 0$ ). PDAS does not have problems handling the target model lying on the bounds. Interior point methods performs similarly, with PDIP being more accurate than Log-Barrier.

	Log-Barrier	Primal-Dual interior point	Primal-Dual Active Sets
RMS	$2.3055 \times 10^{-2}$	$1.3410 \times 10^{-2}$	$1.1841 \times 10^{-2}$
Error	0.1760	0.1684	0.1424
CG iter.	439	254	443
Time (sec)	178	103	212

Table 4.9: Problem A2, regularization applied ( $\beta = 10^{-5}$ ). Again, PDAS provides the smaller error. However, the regularization forces all the methods to converge similarly, being more cost effective for interior point methods.

	Log-Barrier	Primal-Dual interior point	Primal-Dual Active Sets
RMS	$6.4954 \times 10^{-2}$	$2.2022 \times 10^{-2}$	$2.1278 \times 10^{-2}$
Error	0.1873	0.1862	0.1788
CG iter.	393	421	436
Time (sec)	159	169	217

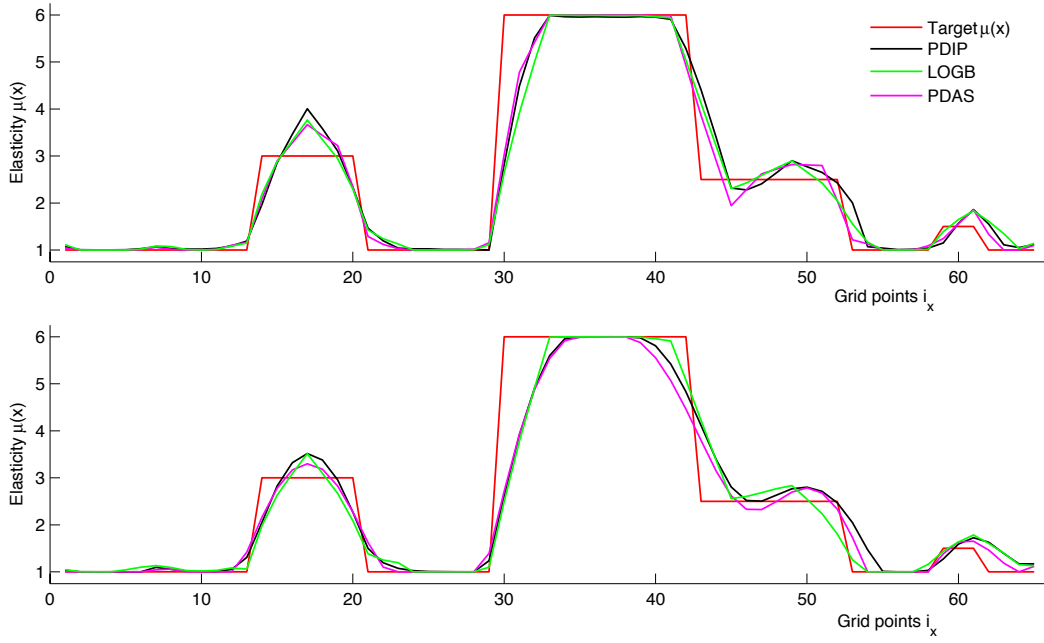


Figure 4.2: Numerical results for problem A2. Top: No regularization. Bottom: Regularization applied with  $\beta = 10^{-5}$ . In Problem A2, the target model parameter  $\mu^*$  has some components lying on the boundary. The level of discretization for  $\mu$  and the number of Newton-Krylov iterations is the same as for Problem A1. We show the target model  $\mu^*$  in red for the last stage of the algorithm, with the same color scheme for the final approximations of  $\mu$  in Figure 4.1.

For this experiment, we also implement a point-wise physical bound strategy for primal-dual methods. We define our physical bounds according to the initial guess  $\mu_0$ , which in

general can be properly selected a priori. With bounds that vary with depth based on physical observations and uncertainty we may better constrain our model space, which can be helpful as the complexity of our target model  $\mu^*$  increases. For this purpose, we define

$$a_i = \max\{a_i, (\mu_0)_i - v(b - a)_i\}, \quad b_i = \min\{b_i, (\mu_0)_i + v(b - a)_i\}, \quad (4.67)$$

where  $0 \leq v \leq 1$  represents the percentage of confidence in our initial guess. To analyze the behavior of the iterates and to demonstrate that (4.67) is honored by all the iterates  $\mu_k$ , we use Problem A1 without regularization with a value of  $v = 0.4$ . In Table 4.10 we show a comparison between the numerical results obtained for this problem, (presented in Table 4.6), and those that we obtain by using point-wise boundaries. We show, in Figure 4.3, the target model  $\mu^*$  in red, the point-wise boundaries in blue with markers, and all the iterates  $\mu_k$  for primal-dual methods in different colors.

Table 4.10: Numerical comparison with the results presented in Table 4.6 (in parenthesis) for Problem A1 when using the point-wise boundaries for primal-dual methods. PDAS converges faster while losing some accuracy and PDIP provides a smaller error and RMS with a moderate increase in compute time.

	Primal-Dual interior point	Primal-Dual Active Sets
RMS	$1.5359 \times 10^{-4}(2.2717 \times 10^{-4})$	$3.6376 \times 10^{-4}(1.7859 \times 10^{-4})$
Error	$6.9401 \times 10^{-2}(7.9913 \times 10^{-2})$	$8.3719 \times 10^{-2}(8.2931 \times 10^{-2})$
CG iter.	499(465)	370(485)
Time (sec)	199(181)	172(223)

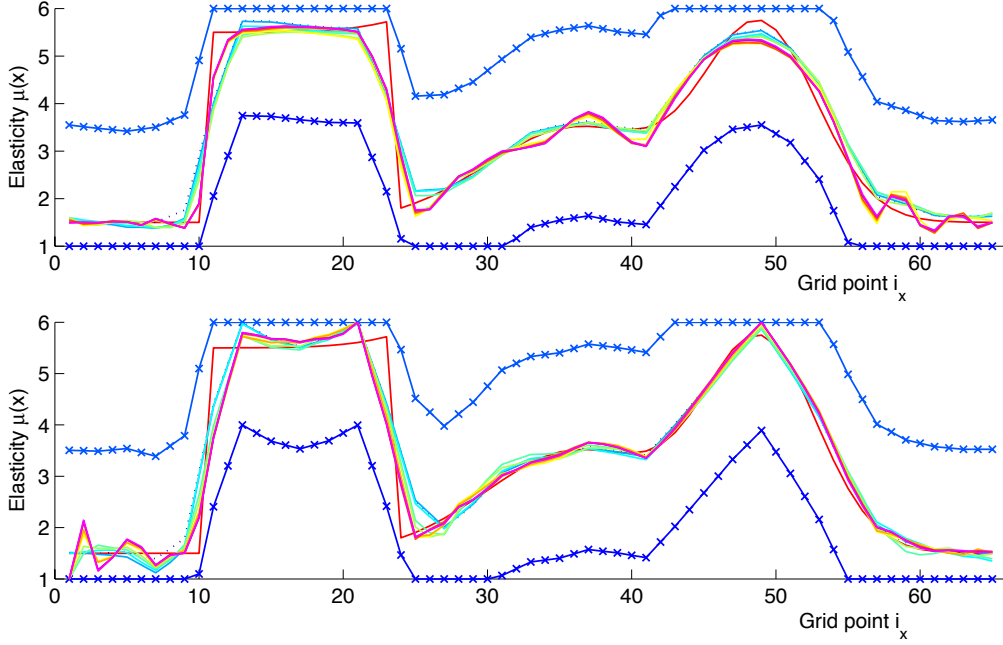


Figure 4.3: Point-wise boundaries for PDAS (top) and PDIP (bottom) with no regularization applied to Problem A1. The lines with the markers represent the point-wise boundaries. Notice how the target  $\mu$  in red and all the iterates satisfy the inequality  $a_i \leq (\mu_k)_i \leq b_i$  for all  $k$ .

In this experiment a regularization with  $\beta = 10^{-5}$  seemed to be too strong in problem A1, thus we reduced its order to be  $\beta = 10^{-6}$ . Also, in problem A2 it is clear that for active set methods the Error reduction is better than for interior point methods, since many components of the target model  $\mu^*$  lie on the boundaries.

#### 4.5.5 Experiment B

This is a more complex experiment due to its size and higher frequency components as showed in Figures 4.4 and 4.5. The physical bounds for this problem are  $a = 1.27239$  and  $b = 18.1603$ . The level of discretization for  $\mu$  in this case is set to  $j = 4, 5, 6$  and  $7$  where we allow 5 Newton-Krylov iterations at all coarser grids in the continuation scheme and 10 iterations for the finer grid. Tables 4.11 and 4.12 summarize the results for this problem. In Table 4.12, Tikhonov regularization is applied with  $\beta = 10^{-6}$ .

Table 4.11: Experiment B, no regularization applied ( $\beta = 0$ ). PDIP has better accuracy, while LOGB seems to converge faster being less accurate as in the previous problems.

	Log-Barrier	Primal-Dual interior point	Primal-Dual Active Sets
RMS	$4.34 \times 10^{-3}$	$4.22 \times 10^{-3}$	$4.07 \times 10^{-3}$
Error	$7.29 \times 10^{-2}$	$6.82 \times 10^{-2}$	$7.45 \times 10^{-2}$
CG iter.	481	527	421
Time (sec)	487	551	474

Table 4.12: Experiment B, regularization applied ( $\beta = 10^{-6}$ ). PDIP is again less affected by the smoothing effect of the regularization in terms of computational effort and error.

	Log-Barrier	Primal-Dual Interior Point	Primal-Dual Active Sets
RMS	$2.38 \times 10^{-2}$	$2.41 \times 10^{-2}$	$2.41 \times 10^{-2}$
Error	0.1184	0.1032	0.1254
CG iter.	446	394	363
Time (sec)	451	400	428

In Figure 4.4, we show the target model  $\mu^*$  and final approximations obtained by each method for the model parameter  $\mu$ . To distinguish between the different estimations, we use a similar color code as in the previous figures. The results when Tikhonov regularization is applied are given in Figure 4.5.

For this experiment, we again implemented a point-wise physical bound strategy for primal-dual methods. As before, our physical bounds are defined according to the initial guess  $\mu_0$ , but now we use the values  $v = 0.1$  and  $v = 0.2$  in (4.67) since the range  $b - a$  is four times larger than the one of Experiment A. Table 4.13 shows a comparison between the numerical results obtained for this problem, (presented in Table 4.11), and those that we get by using point-wise boundaries for  $v = 0.1$ . Table 4.14 shows a similar comparison with  $v = 0.2$ . We present, in Figures 4.6 and 4.7, the target model  $\mu^*$  in red, the point-wise boundaries in blue

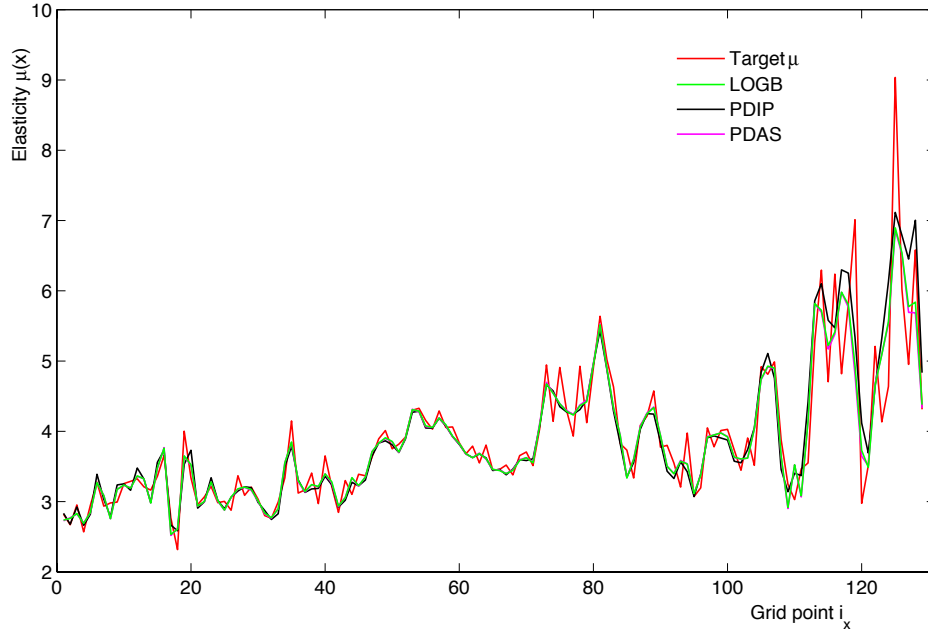


Figure 4.4: Final stage of Experiment B. As before, we present the optimal  $\mu$  for each method and true model parameter  $\mu^*$ . No regularization is applied. For this experiment, we set  $T = 50$  in (4.62), which is twice the value of  $T$  in the previous experiment.

(with markers with  $v = 0.1$  and  $v = 0.2$  for PDAS and PDIP respectively), and all the iterates  $\mu_k$  for primal-dual methods in different colors.

Table 4.13: Numerical comparison with the results presented in Table 4.11 (in parenthesis) for Problem B when using the point-wise boundaries with  $v = 0.1$  for primal-dual methods. Both methods do not improve the error neither the RMS but PDIP reduces the number of iterations.

	Primal-Dual interior point	Primal-Dual Active Sets
RMS	$4.25 \times 10^{-3}(4.22 \times 10^{-3})$	$3.73 \times 10^{-3}(4.07 \times 10^{-3})$
Error	$7.55 \times 10^{-2}(6.82 \times 10^{-2})$	$9.62 \times 10^{-2}(7.45 \times 10^{-2})$
CG iter.	472(527)	595(421)
Time (sec)	546(551)	870(474)

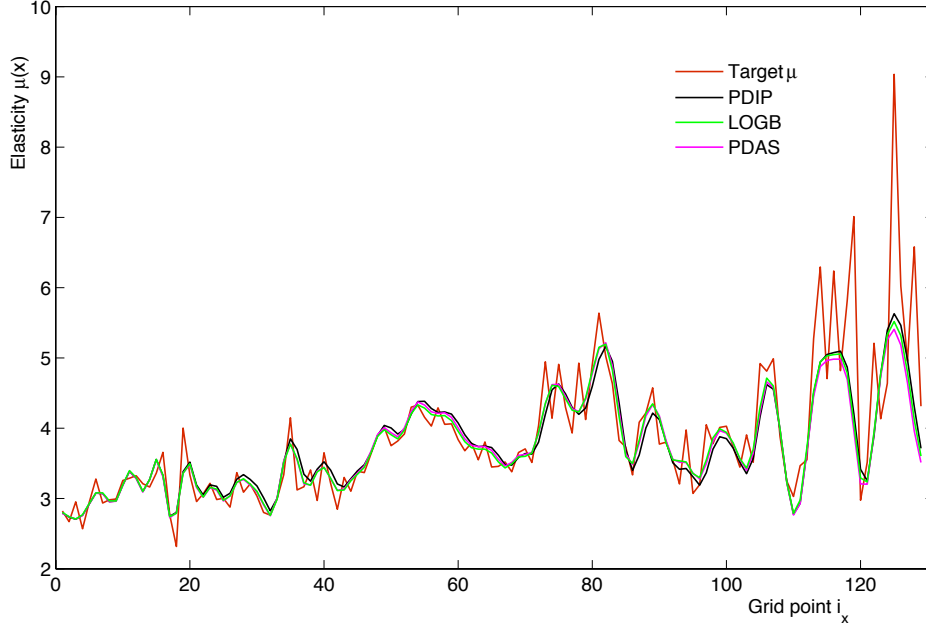


Figure 4.5: Final stage of Experiment B. We present the optimal  $\mu$  for each method and true model parameter  $\mu^*$ . Regularization applied with  $\beta = 10^{-6}$ .

Table 4.14: Numerical comparison as in Table 4.13 for Problem B with  $\nu = 0.2$  for primal-dual methods. PDAS converges slower while gaining accuracy and PDIP provides a similar error with smaller RMS but reducing the number of CG iterations.

	Primal-Dual interior point	Primal-Dual Active Sets
RMS	$3.90 \times 10^{-3}(4.22 \times 10^{-3})$	$4.15 \times 10^{-3}(4.07 \times 10^{-3})$
Error	$6.84 \times 10^{-2}(6.82 \times 10^{-2})$	$7.41 \times 10^{-2}(7.45 \times 10^{-2})$
CG iter.	509(527)	522(421)
Time (sec)	522(551)	584(474)

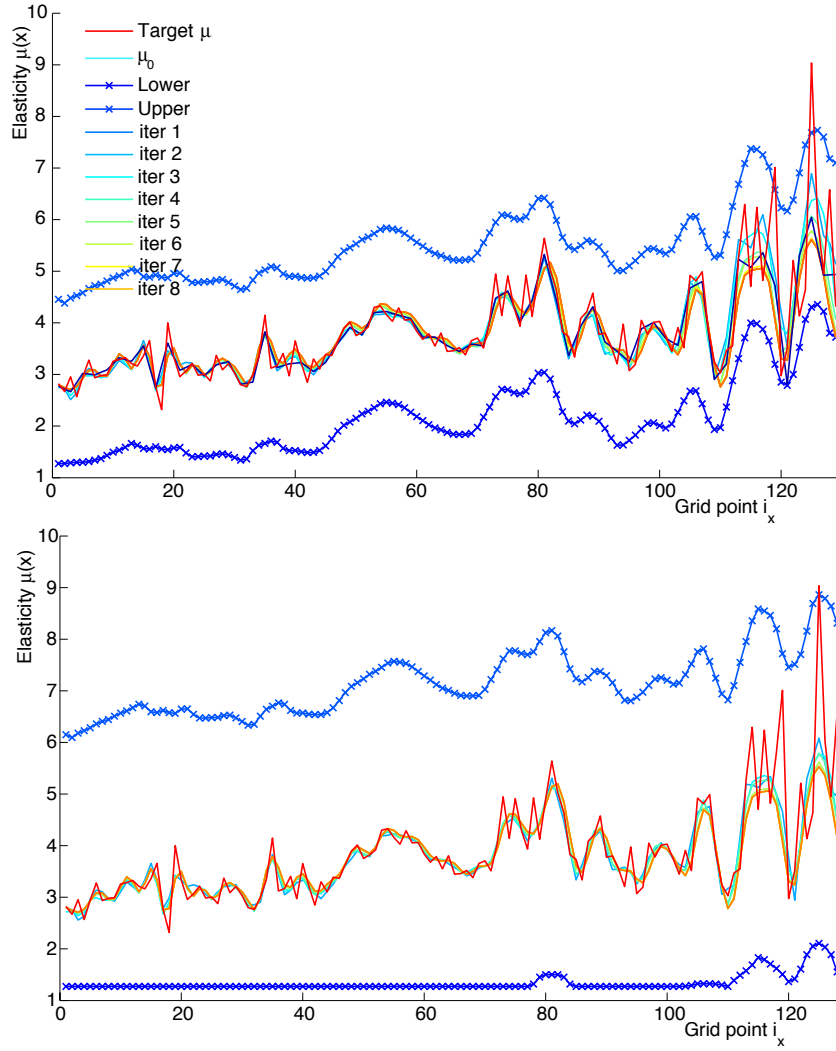


Figure 4.6: Point-wise boundaries for PDAS with  $v = 0.1$  (top) and  $v = 0.2$  (bottom) with no regularization applied to problem B. The lines with the markers represent the point-wise boundaries. Notice how all the iterates satisfy  $a_i \leq (\mu_k)_i \leq b_i$  for all  $k$ , however the target  $\mu$  in red fails to be inside for the highest peaks at depth for  $v = 0.1$ .

For Experiment B, we noticed that if we use a regularization parameter greater or equal than  $\beta = 10^{-5}$  excessive smoothness is caused, thus we set  $\beta = 10^{-6}$  in this case. From our experience with the variation of the regularization parameter  $\beta > 0$  for Experiments A and B, we notice that higher values of regularization avoid a good match of the complexity of the target model  $\mu^*$ . *[I don't understand this from here...]* We wanted to see if using fixed values of the

barrier or perturbation parameter  $\tau$ , interior point methods will perform similarly to the case when regularization is applied. This is, we study the behavior of both methods for fixed values of  $\tau$  with  $\beta = 0$ . It turns out that in the absence of regularization, neither log-barrier or primal dual methods converge for fixed values of  $\tau$ . Thus the role of the regularization parameter can not be replaced by the single use of the barrier or perturbation parameter in the inversion. [...to here] Generally speaking, the regularization attempts to make the waveform smoother, as we show in Figure 4.5, where the higher peaks appear off the final approximation given by the three methods; we think that one of the reasons[Please rephrase this sentence] for this to happen is the shrinking effect caused by the regularization applied.

## 4.6 Summary and concluding remarks

In this work, we have exercised three different strategies for solving 1D inverse seismic wave propagation posed as a constrained optimization problem. We use a robust and efficient forward-inverse algorithm—in conjunction with affine-invariant interior-point and active-set algorithms—to handle the inequality constraints over the model parameter  $\mu$ .

In general, primal-dual methods appear to be more accurate than log-barrier methods. Interior point methods seem to converge faster to a good approximation of the target parameter when few active constraints are present. For primal-dual methods, we demonstrated the benefits of our slightly improved version to enforce feasible iterates, when compared to the traditional formulation. Also, our findings from the point-wise boundaries strategy presented showed us a slight benefit to reduce the Error for interior point methods, while not for the active sets strategy since it generally increases the number of CG iterations to reach convergence.

Future research is concerned with the extension of this work to large-scale problem setting, where primal-dual methods have shown to be a robust and efficient alternative, the introduction of uncertainty quantification in our generic framework, and also to explore Total Variation as a possible regularization in our algorithms.

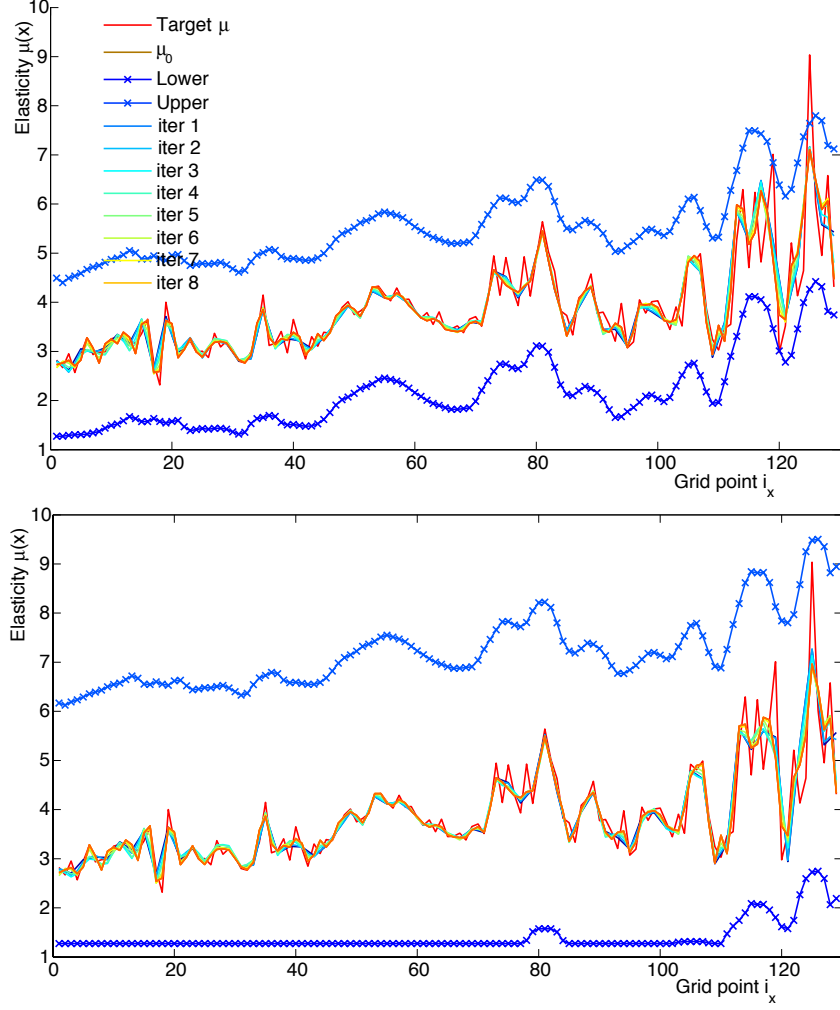


Figure 4.7: Point-wise boundaries for PDIP with  $v = 0.1$  (top) and  $v = 0.2$  (bottom) with no regularization applied to problem B. The lines with the markers represent the point-wise boundaries. Notice how the target  $\mu$  in red and all the iterates hold  $a_i \leq (\mu_k)_i \leq b_i$  for all  $k$ .

# Chapter 5

## Joint Inversion of Multiple Geophysical Data Sets

In Chapters 2 and 3 we successfully integrated a joint inversion algorithm [34] with a constrained optimization approach. This approach uses a primal-dual interior point (PDIP) method for solving the geophysical inverse problem (2.2), when surface waves and receiver functions are used as input. We also demonstrated that our approach can be considered as an efficient and robust alternative for geophysical inversion of seismic data. Nevertheless, the presence of artifacts in the numerical results, due to data quality, and noise inherently associated with the inversion process, remains a challenge to overcome. To address this issue, we expect that the inclusion of more data sets may help us to better constrain the geological structures and increase the resolution of the final inverted models. However, the addition of more data sets for joint inversion represents a difficult task. We need to have additional forward operators, partial derivative matrices, and sometimes the presence of stronger empirical links among the model parameters. In some cases, not all of these requirements can be fulfilled. In other cases, like with magnetotelluric or gravity data, this data has been shown to complement surface waves and receiver functions [56, 55], and in consequence, better inverted models were obtained. In this Chapter we have been able to add seismic travel time data,  $y^{TT}$ , to our joint inversion scheme. We want to study if this data set complements the other two and how its use will impact the inversion results. To the best of our knowledge, this data set has not been included for joint inversion of multiple data sets. We describe, in the next sections the tools used to compute the predicted travel times: a synthetic case example, and an example using real data from two stations deployed in the Rio Grande Rift Region.

## 5.1 Seismic travel times

As before, we assume the model parameter to be the same for all data sets comprised in the inversion, i.e., we assume that the forward operator depends on the S-wave velocity only. For this assumption to be valid we need to know a relationship between travel times  $T$  and velocities  $x$  in other words, we need to define a forward operator  $F^{TT}$  such that  $F^{TT} = F^{TT}(x)$ . The original formulation of the joint inversion problem (2.3), without the regularization term, will now be rewritten as

$$\min_x \frac{1}{2} \|F^{SW}(x) - y^{SW}\|_W^2 + \frac{1}{2} \|F^{RF}(x) - y^{RF}\|_W^2 + \frac{1}{2} \|F^{TT}(x) - y^{TT}\|_W^2, \quad (5.1)$$

that later can be accommodated in the linearized constrained formulation (2.9), to be solved using interior point methods as we discussed in Section 2.3.2.

The computation of seismic travel times has been developed since the early 1900s, and the literature related to this evolution process can be found in [10]. In fact, we use a program that implements the so-called tau-p ( $\tau(p)$ ) method described in [10], as the basis for the computation of travel times. Here  $\tau = \tau(p)$  is usually called the intercept-slowness function that represents the travel time curves. As before,  $p$  is the ray parameter and  $\tau$  in this chapter will represent the intercept with the travel time axis. The discrete calculations proposed in the  $\tau(p)$  method use a base (standard) Earth model (e.g., like the AK-135 model previously employed in this work), a given interpolation scheme, and appropriate discretizations for modeling depth, wave ray parameter, ray integrals, and other functions. Although less intuitive, the  $\tau(p)$  formulation is equivalent to the formulation of the forward operator that we use to compute travel times, and that is presented in equation (5.2) from the next section.

The main application of seismic travel times is usually associated with seismic tomography [72]. In general, tomography makes use of the arrival times of body waves to infer seismic velocity structure, since different distributions of velocity with depth produce characteristic travel time curves. However, first-arrival times are not useful for identifying structural discontinuities in the subsurface, so additional information is required to identify certain geological structures like magma chambers. Since receiver functions have such capability, and the sur-

face waves provides average structure information, we expect that the joint inversion helps the travel times data to overcome this issue. In the following subsections, we validate this hypothesis and investigate if this new data set contributes to increase the resolution of the inverted models.

### 5.1.1 Travel times inversion

Based on simplified assumptions, travel times may be defined as the time required for a wave front to propagate from the source to the receiver, e.g., from an earthquake to the seismic station respectively. To infer the distribution of velocity with depth, travel times can be compiled from seismograms for different sources at prescribed distances. For a given seismic station, we can collect several seismograms, to select an arrival, (or phase corresponding to a travel path), and then pick the arrival time.

There are several approaches to derive the velocity structure from travel times data. One is to use a computer program like the widely-used IASPEI `ttimes` package. The other is to solve the inverse problem directly by deriving the velocities  $x$  from the computation of the travel times curves, since we know the source-receiver distance. In both cases, these methods often fail due to the presence of low-velocity zones, e.g., magma chambers. In this work we combine a computer program, the TauP Toolkit [18], and a flat layered formulation to compute the travel times (Figure 5.1). We coupled the TauP Toolkit [18] to calculate the predicted travel times for our joint inversion algorithm. This toolkit handles many types of velocity models and it can calculate times for virtually any seismic phase. The idea with this combined formulation is to use the full power of the toolkit to introduce the spherical Earth geometry into the computation, and to use the simple flat layered approach to identify the major geological features beneath a given station. We expect to see how the travel times data set contributes to the inversion, by running some tests similar to those presented in Chapter 2.

Following [69] we now describe the formulation used to compute travel times of a flat layered Earth model. In Figure 5.2 a) we show the relation between the travel time curve and ray paths. Consider the ray paths to two points  $dl$  apart, which differ by the travel time  $\Delta T$ , then it can be shown that  $p = \frac{\sin i}{x} = \frac{\Delta T}{dl}$ . In Figure 5.2 b) we show that the ray path from the

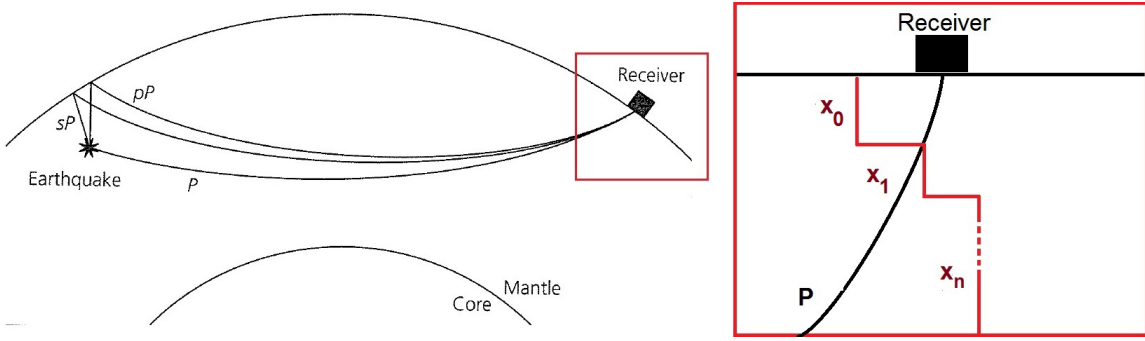


Figure 5.1: (Left) Three different ray paths corresponding to the phases  $sP$ ,  $pP$  and  $P$  for an earthquake at a teleseismic distance of the receiver on a spherical Earth geometry (adapted from [69]). (Right) Flat layered model beneath the receiver with velocities  $x_0, x_1, \dots, x_n$ .

top layer has traveled through  $n$  layers, each of thickness  $h_j$ , and velocity  $x_j$ . By Snell's law, the ray parameter  $p$  is constant along a ray, thus the incidence angles  $i_j$  can be found from the incidence angle  $i_0$  at the top layer,  $p = \frac{\sin i_j}{x_j} = \frac{\sin i_0}{x_0}$ . Therefore, a downgoing ray that travels a horizontal distance  $l_j$  in the  $j^{th}$  layer, spends a time  $\Delta T_j$  in that layer. Then to compute half of the total time traveled we have

$$T = F^{TT}(x) = \sum_{j=0}^n \Delta T_j = \sum_{j=0}^n \frac{h_j}{x_j \cos i_j}, \quad (5.2)$$

from equation (5.2), combined with Snell's law, we obtain the partial derivatives entries required to complement the Jacobian matrix in (2.12) as

$$\frac{\partial T}{\partial x_j} = -\frac{h_j}{x_j^2 \cos(\sin^{-1}(x_j p))^2} \left( \cos(\sin^{-1}(x_j p)) - \frac{(x_j p)^2}{\sqrt{1 - (x_j p)^2}} \right). \quad (5.3)$$

Equations (5.2) – (5.3) provide us with both the forward computations of travel times, and its partial derivatives. This formulation is useful when the ray paths between the source and the receiver are short enough that the Earth's curvature can be neglected. Since we are interested in studying the Earth's uppermost mantle structure and our formulation combines both models, the spherical obtained with the  $\tau(p)$  method and the flat layered geometry, this approach is still valid. Furthermore, if we do not have available S-wave arrival times, we can

modify equation (5.2) to work with a P-wave arrival times. We just need to multiply  $x$  by an appropriate  $v_p/v_s$  ratio.

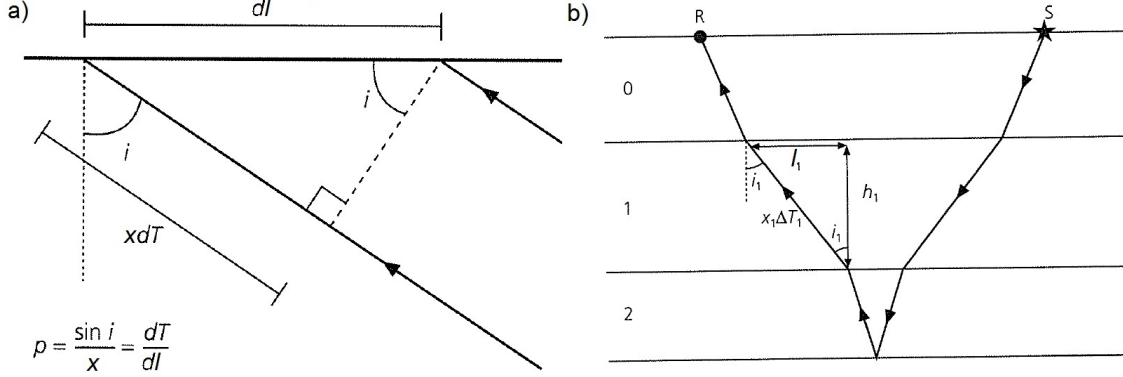


Figure 5.2: a) Two rays showing the relationship between the angle of incidence  $i$ , ray parameter, and the slope of the travel time curve. b) Ray geometry for a reflection flat-layered medium. Layer thicknesses are  $h_j$ , horizontal distances traveled are  $l_j$ , and one-way travel times spent in the layers are  $\Delta T_j$  (adapted from [69]).

## 5.2 Joint inversion of receiver functions, surface waves and travel times data

In this section, we implement our joint inversion algorithm for synthetic and real models. First we proceed as described in Chapter 2, for a Rift Earth's crustal synthetic model, divided in layers of 1 km thickness. We compute two different dispersion curves, Rayleigh and Love, for group velocities and simultaneously receiver functions for three different ray parameters  $p$ . Now we add travel times data to the inversion. Our synthetic data consists of 13 travel times computed by the IASPEI `ttimes` package for different source-receiver distances ranging from  $30^\circ$  to  $90^\circ$ , corresponding to teleseismic distances, with spacing of  $5^\circ$ . This procedure resembles the picking of arrival times on a given seismogram for a selected event.

We studied the behavior of the algorithm when using travel times data only, travel times combined with surface waves and receiver functions separately, and the three data sets together.

As a result, we find that the only resolving power of the travel times data resides in the identification of the strongest velocity contrast when used alone. This may be probably a result of the simplified formulation (5.2), in which the absence of strong velocity variations does not help the forward operator to change accordingly and resolve such variations. On the other hand, if we use travel times with either surface waves or receiver functions we noticed a remarkable improvement of the resolution in depth of the inverted models (see Figures 5.3 – 5.4).

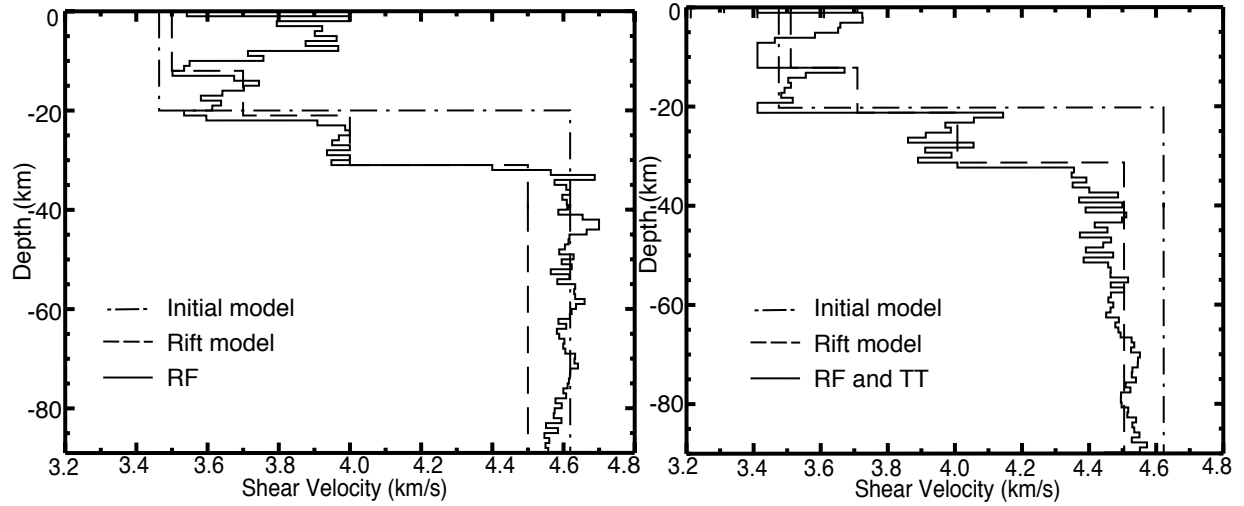


Figure 5.3: Single inversion of RF (left) compared to RF and TT joint inversion (right). The bottom of the model is being better resolved by the joint inversion.

Finally, we inverted together the multiple data sets, and compared the numerical results with those obtained using only surface waves and receiver functions. We show in Table 5.1 the numerical comparison of RMS and residual error as defined in Chapter 2, and present the inverted models for each joint inversion in Figure 5.5.

We also use seismic data from station 223A of the USArray near El Paso. We choose this station since it is located over a basin in the middle of the RGR. This type of geological structure is known to be problematic for certain techniques used to model Earth structure. In particular, for receiver functions, it is complicated to obtain a good image beneath this station, since the numerical computation of the waveform becomes usually unstable. We tested the performance

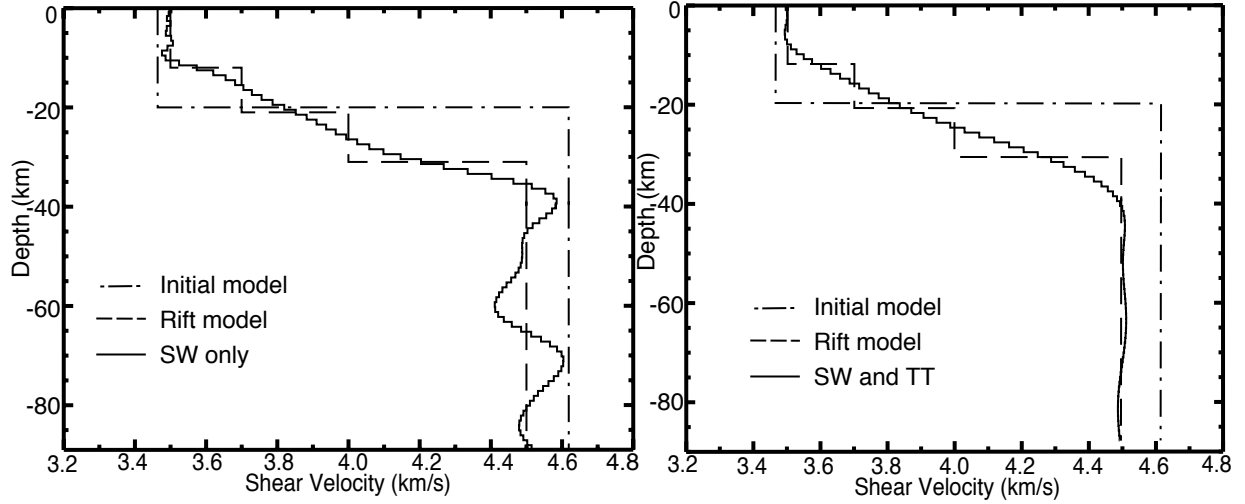


Figure 5.4: Single inversion of SW (left) compared to SW and TT joint inversion (right). The average resolving power of surface waves has been preserved but better constrained in depth.

Table 5.1: Relative RMS and residuals errors comparison for joint inversion of surface waves and receiver function when travel times are added.

Joint Inversion	RMS	Residual Error
SW & RF	0.00878717	0.00284017
SW, RF & TT	0.00739198	0.00030573

of the joint inversion when travel times are included, and found that the presence of receiver function is still required to guarantee stability of the algorithm, but with less computational effort: in other words, with smaller influence of the receiver function data sets, we obtain faster convergence. The numerical results are presented for several influence parameter choices in Figure 5.6. We investigated the performance of the joint inversion algorithm when better quality data is available. This is the case of USArray station 219A, corresponding results are presented in Figure 5.7. In Table 5.2, we present the misfit values reached by the different choices of influence parameters depicted in Figures 5.6 – 5.7.

While the numerical results appear to be quite similar for the upper crust, i.e., the region

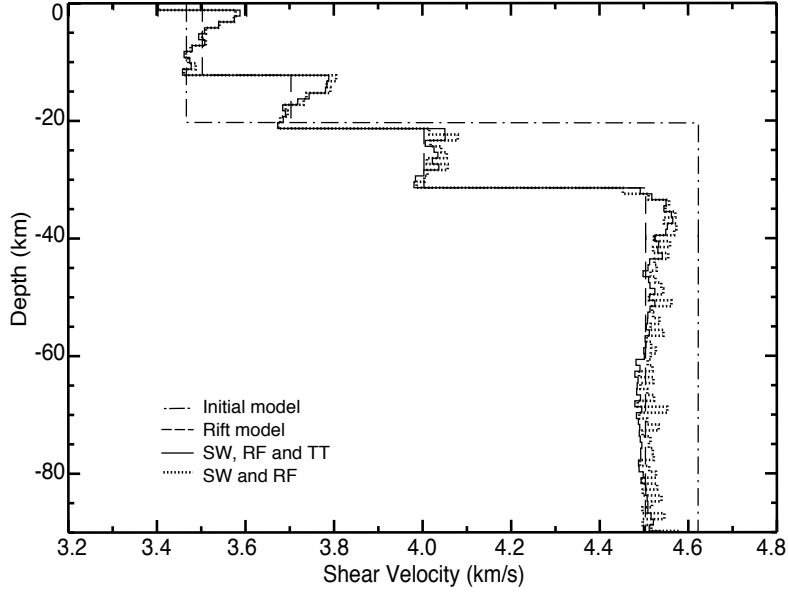


Figure 5.5: When compared the joint inversion of RF, SW and TT inversion to that of RF and SW, we identified a slight improvement in the resolution of the model with depth.

above 100 km depth, in both stations and for any choices of the influence parameters  $\eta$ ; as we go deeper into the Earth, remarkable differences begin to show up. Besides the non-uniqueness associated with any inversion process, in particular we see that when less influence is associated to receiver functions ( $\eta_2 = .2$ ), low velocity zones are shown for both stations. This is a really interesting feature, since, as we claimed before, travel times hardly detect this type of anomalies. Also, the inverted models in this case are shown to be less subject to strong variations, which means that the numerical results are more reliable. Still more work needs to be done for joint inversion of these multiple data sets, to guarantee reliable inverted results when real data is used. It is hard to find a way to validate our results, since a true model of the Earth is not available (as in the synthetic case). To this end, we plan to compare our findings with previous work on this and other regions, and probably integrate gravity data as another constraint for the inversion.

Table 5.2: Residuals errors comparison for different choices of the influence parameter  $\eta$  in the joint inversion of surface waves and receiver function when travel times are added. Here  $\eta = (\eta_1, \eta_2, \eta_3)$  represents the level of influence of each data set,  $\eta_1$  corresponding to SW,  $\eta_2$  to RF and  $\eta_3$  to TT. Let us recall that  $\eta_1 + \eta_2 + \eta_3 = 1$ .

Station	$\eta = (.4, .4, .2)$	$\eta = (.4, .2, .4)$	$\eta = (.2, .4, .4)$
219A	0.00020594	0.00016242	0.00023699
223A	0.00007066	0.00006418	0.00008621

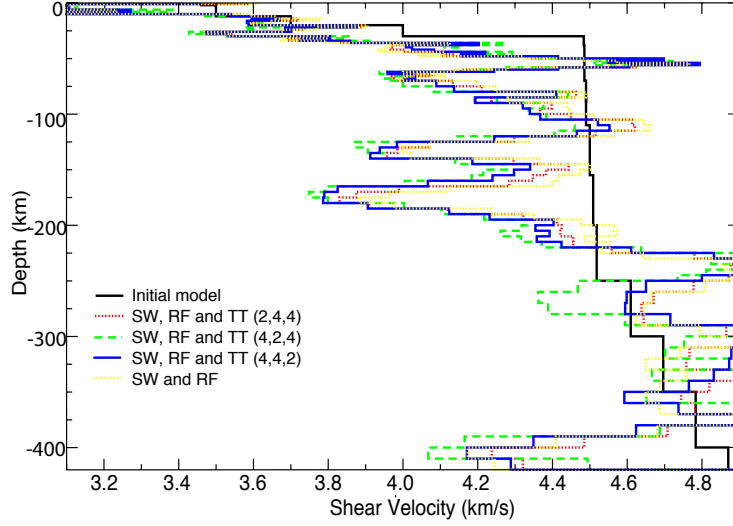


Figure 5.6: Numerical comparison between three different level of data set influence over the inverted models in station 223A. The results appear to be very similar for all levels of influence, but when less influence is given to receiver functions, slight differences can be seen after 200 km.

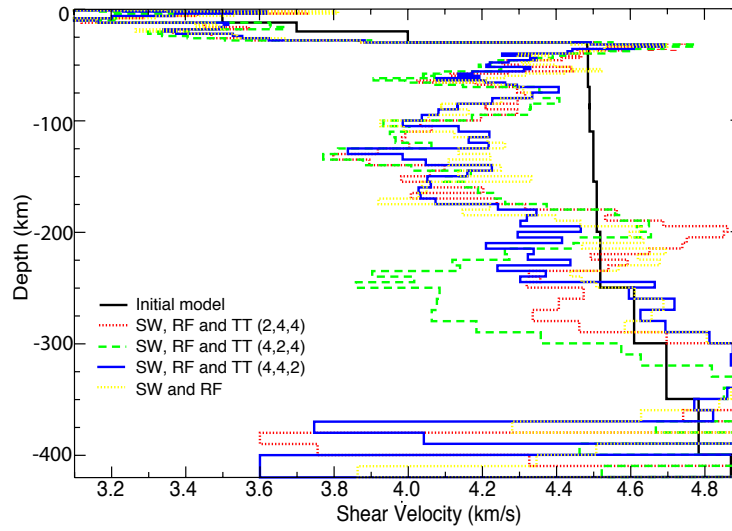


Figure 5.7: Numerical comparison between three different level of data set influence over the inverted models in station 219A. Allowing the influence of receiver function data to be only 20% seems to reduce rapid variations in velocities, and identify a low velocity zone at 250 km. We need more information to decide if this is an artifact or not.

# Chapter 6

## Concluding Remarks

This dissertation presents two approaches based on constrained optimization for geophysical inversion of seismic data. Both of these approaches highlight the benefits of including explicit bound constraints over a model parameter, when apriori information is available.

The first approach expands a 1-D joint inversion algorithm [34], to be used in conjunction with primal-dual interior point methods and a Bayesian kriging interpolation scheme [65]. In Chapters 2 and 3, we demonstrate that our approach can be considered an efficient and robust alternative for this type of geophysical inversion. By using this approach and seismic data, we present a new model of crustal and upper mantle structure beneath the southern RGR. In Chapter 5, we add travel times data to our joint inversion approach, to show the possible benefits that this new data set may bring to the inverted models.

The next phase in this geophysical inversion scheme will be to incorporate other compatible data sets into the joint inversion (e.g., gravity data), expecting to constrain further the inversion process and to increase the resolution of the models. Also, we want to implement a stronger structural constraint for removing noisy components during the inversion, and a grid continuation scheme for modeling resolution.

In Chapter 4, we present the second algorithmic approach for 1-D full-waveform propagation. We provide a generic affine invariance framework that accomodates three constrained optimization schemes and improves a standard procedure for forcing positivity. We also investigate the advantages of implementing point-wise boundaries in a grid-continuation approach. As a result, we show a slight benefit to reduce the Error for interior point methods, while not for the active sets strategy. Future research avenues are the extension of this work to large-scale problem setting, where primal-dual methods have shown to be a robust and scalable alternative, and to introduce uncertainty quantification in our generic framework.

# References

- [1] V. Akçelik, J. Bielak, G. Biros, I. Epanomeritakis, A. Fernandez, O. Ghattas, E. J. Kim, J. Lopez, D. R. O'Hallaron, T. Tu, and J. Urbanic. High resolution forward and inverse earthquake modeling on terascale computers. In *SC03: Proceedings of the International Conference for High Performance Computing, Networking, Storage, and Analysis*. ACM/IEEE, 2003. Gordon Bell Prize for Special Achievement.
- [2] C. J. Ammon. The isolation of receiver effects from teleseismic p waveforms. *Bulletin of the Seismological Society of America*, 81(6):2504–2510, 1991.
- [3] H. Antil, R. H. W. Hoppe, and C. Linsenmann. Adaptive multilevel interior-point methods in pde constrained optimization. In T. J. Barth, M. Griebel, D. E. Keyes, R. M. Nieminen, D. Roose, and T. Schlick, editors, *Domain Decomposition Methods in Science and Engineering XVIII*, volume 70 of *Lecture Notes in Computational Science and Engineering*, pages 15–26. Springer Berlin Heidelberg, 2009. 10.1007/978-3-642-02677-52.
- [4] M. Argáez, R. A. Tapia, and L. Velazquez. Numerical comparisons of path-following strategies for a primal-dual interior-point method for nonlinear programming. *Journal of Optimization Theory and Applications*, 114(2):255–272, 2002.
- [5] M. G. Averill, K. C. Miller, R. G. Keller, V. Kreinovich, R. Araiza, and S. Starks. Using expert knowledge in solving the seismic inverse problem. *International Journal of Approximate Reasoning*, 45(3):564–587, 2007.
- [6] G. E. Backus and J. F. Gilbert. Numerical applications of a formalism for geophysical inverse problems. *Geophysical Journal of the Royal Astronomical Society*, 13(1-3):247–276, 1967.
- [7] I. W. Bailey, M. S. Miller, K. Liu, and A. Levander. Vs and density structure beneath the colorado plateau constrained by gravity anomalies and joint inversions of receiver function and phase velocity data. *J. Geophys. Res.*, 117(B2):B02313, 2012.

- [8] L. Bashir, S. S. Gao, K. H. Liu, and K. Mickus. Crustal structure and evolution beneath the colorado plateau and the southern basin and range province: Results from receiver function and gravity studies. *Geochem. Geophys. Geosyst.*, 12(6):Q06008, 2011.
- [9] P. Bird. Continental delamination and the Colorado Plateau. *Journal of Geophysical Research*, 84:7561–7571, 1979.
- [10] R. Buland and C. H. Chapman. The computation of seismic travel times. *Bulletin of the Seismological Society of America*, 73(5):1271–1302, 1983.
- [11] C. Bunks, F. M. Saleck, S. Zaleski, and G. Chavent. Multiscale seismic waveform inversion. *Geophysics*, 50(5):1457–1473, 1995.
- [12] C. Burstedde and O. Ghattas. Algorithmic strategies for full waveform inversion: 1D experiments. *Geophysics*, 74(6):WCC37–WCC46, 2009.
- [13] G. Chavent. Identification of functional parameters in partial differential equations. In R. E. Goodson and M. Polis, editors, *Identification of Parameters in Distributed Systems*, pages 31–48. ASME, 1974.
- [14] H. P. Christian. *Discrete Inverse Problems*. Society for Industrial and Applied Mathematics, 2010.
- [15] D. Colombo and M. D. Stefano. Geophysical modeling via simultaneous joint inversion of seismic, gravity, and electromagnetic data: Application to prestack depth imaging. *The Leading Edge*, 26(3):326–331, 2007.
- [16] S. Constable, R. Parker, C. Constable, et al. Occam’s inversion: A practical algorithm for generating smooth models from electromagnetic sounding data. *Geophysics*, 52(3):289–300, 1987.
- [17] F. A. Cook, E. R. Decker, and S. B. Smithson. Preliminary transient heat flow model of the rio grande rift in southern new mexico. *Earth and Planetary Science Letters*, 40(3):316 – 326, 1978.

- [18] H. P. Crotwell, T. J. Owens, and J. Ritsema. The taup toolkit: Flexible seismic travel-time and ray-path utilities. *Seismological Research Letters*, 70(2):154–160, March/April 1999.
- [19] F. E. Curtis, J. Huber, O. Schenk, and A. Wächter. On the Implementation of an Interior-Point Algorithm for Nonlinear Optimization with Inexact Step Computations. *Mathematical Programming, Series B*, 2012.
- [20] R. S. Dembo and T. Steihaug. Truncated-Newton algorithms for large-scale unconstrained optimization. *Mathematical Programming*, 26(2):190–212, 1983.
- [21] M. T. Dugda, A. A. Nyblade, J. Julia, C. A. Langston, C. J. Ammon, and S. Simiyu. Crustal structure in ethiopia and kenya from receiver function analysis: Implications for rift development in eastern africa. *J. Geophys. Res.*, 110(B1):B01303, 2005.
- [22] S. C. Eisenstat and H. F. Walker. Choosing the forcing terms in an inexact Newton method. *SIAM Journal on Scientific Computing*, 17:16–32, 1996.
- [23] A. S. El-Bakry, R. A. Tapia, T. Tsuchiya, and Y. Zhang. On the formulation and theory of the primal-dual Newton interior-point method for nonlinear programming. *Journal of Optimization Theory and Applications*, 89(3):507–541, 1996.
- [24] I. Epanomeritakis, V. Akçelik, O. Ghattas, and J. Bielak. A Newton-CG method for large-scale three-dimensional elastic full-waveform seismic inversion. *Inverse Problems*, 24(3):034015 (26pp), 2008.
- [25] A. Forsgren, P. E. Gill, and M. H. Wright. Interior methods for nonlinear optimization. *SIAM Review*, 44:525–597, 2002.
- [26] L. A. Gallardo and M. A. Meju. Joint two-dimensional dc resistivity and seismic travel time inversion with cross-gradients constraints. *J. Geophys. Res.*, 109(B3):B03311, 2004.
- [27] L. A. Gallardo-Delgado, M. A. Perez-Flores, and E. Gomez-Trevino. A versatile algorithm for joint 3d inversion of gravity and magnetic data. *Geophysics*, 68(3):949–959, 2003.

- [28] M. D. Gunzburger. *Perspectives in Flow Control and Optimization*. SIAM, Philadelphia, 2003.
- [29] E. Haber and D. Oldenburg. Joint inversion: a structural approach. *Inverse Problems*, 13(1):63, 1997.
- [30] P. C. Hansen. The truncated svd as a method for regularization. *BIT Numerical Mathematics*, 27:534–553, 1987. 10.1007/BF01937276.
- [31] M. Hestenes and E. Stiefel. Methods of conjugate gradients for solving linear systems. *J. of Research National Bureau of Standards*, 49:409–436, 1952.
- [32] M. Hintermüller and M. Hinze. A SQP-semismooth Newton-type algorithm applied to control of the instationary Navier-Stokes system subject to control constraints. *SIAM Journal on Optimization*, 16(4):1177–1200, 2006.
- [33] M. Hintermüller, K. Ito, and K. Kunisch. The primal-dual active set strategy as a semi-smooth Newton method. *SIAM Journal on Optimization*, 13(3):865–888, 2003.
- [34] J. Julià, C. J. Ammon, R. B. Herrmann, and A. M. Correig. Joint inversion of receiver function and surface wave dispersion observations. *Geophysical Journal International*, 143(1):99–112, 2000.
- [35] J. Julià, C. J. Ammon, and A. A. Nyblade. Evidence for mafic lower crust in tanzania, east africa, from joint inversion of receiver functions and rayleigh wave dispersion velocities. *Geophysical Journal International*, 162(2):555–569, 2005.
- [36] K. E. Karlstrom and E. D. Humphreys. Persistent influence of proterozoic accretionary boundaries in the tectonic evolution of southwestern north america. *Rocky Mountain Geology*, 33(2):161–179, 1998.
- [37] G. Keller and W. Baldrige. The rio grande rift a geological and geophysical overview. *Rocky Mountain Geology*, 34(1):121–130, 1999.

- [38] G. R. Keller. *Geophysical constraints on the crustal structure of New Mexico. The Geology of New Mexico: A Geologic History*. 2004.
- [39] G. R. Keller, M. A. Khan, P. Morganc, R. F. Wendlandt, W. S. Baldrige, K. H. Olsen, C. Prodehl, and L. Braile. A comparative study of the rio grande and kenya rifts. *Tectonophysics*, 197(2-4):355 – 371, 1991.
- [40] B. L. N. Kennett, E. R. Engdahl, and R. Buland. Constraints on seismic velocities in the earth from traveltimes. *Geophysical Journal International*, 122(1):108–124, 1995.
- [41] P. Kolb, F. Collino, and P. Lailly. Pre-stack inversion of a 1-d medium. *Proceedings of the IEEE*, 74(3):498–508, 1986.
- [42] I. Koulakov. High-frequency p and s velocity anomalies in the upper mantle beneath asia from inversion of worldwide traveltime data. *J. Geophys. Res.*, 116(B4):B04301, 2011.
- [43] C. Kreemer, G. Blewitt, and R. A. Bennett. Present-day motion and deformation of the colorado plateau. *Geophys. Res. Lett.*, 37(10):L10311, 2010.
- [44] C. A. Langston. Structure under mount rainier, washington, inferred from teleseismic body waves. *J. Geophys. Res.*, 84(B9):4749–4762, 1979.
- [45] C. A. Langston. Evidence for the subducting lithosphere under southern vancouver island and western oregon from teleseismic p wave conversions. *J. Geophys. Res.*, 86(B5):3857–3866, 1981.
- [46] J. M. Lees and J. C. VanDecar. Seismic tomography constrained by bouguer gravity anomalies: Applications in western washington. *Pure and Applied Geophysics*, 135:31–52, 1991. 10.1007/BF00877007.
- [47] Y. Li and D. W. Oldenburg. Fast inversion of large-scale magnetic data using wavelet transforms and a logarithmic barrier method. *Geophysical Journal International*, 152(2):251–265, 2003.

- [48] J. Ligorria and C. Ammon. Iterative deconvolution and receiver-function estimation. *Bulletin of the Seismological Society of America*, 89(5):1395–1400, 1999.
- [49] L. R. Lines, A. K. Schultz, and S. Treitel. Cooperative inversion of geophysical data. *Geophysics*, 53(1):8–20, 1988.
- [50] A. Lodge and G. Helffrich. Grid search inversion of teleseismic receiver functions. *Geophysical Journal International*, 178(1):513–523, 2009.
- [51] A. M. and T. R.A. On the global convergence of a modified augmented lagrangian line-search interior-point newton method for nonlinear programming. *Journal of Optimization Theory and Applications*, 114(1):1–25, 2002.
- [52] M. Maceira and C. J. Ammon. Joint inversion of surface wave velocity and gravity observations and its application to central asian basins shear velocity structure. *J. Geophys. Res.*, 114(B2):B02314, 2009.
- [53] D. Marquardt. An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial and Applied Mathematics*, 11(2):431–441, 1963.
- [54] W. Menke. *Geophysical data analysis: Discrete inverse theory*. Academic press, 1984.
- [55] M. Moorkamp, B. Heincke, M. Jegen, A. W. Roberts, and R. W. Hobbs. A framework for 3-d joint inversion of mt, gravity and seismic refraction data. *Geophysical Journal International*, 184(1):477–493, 2011.
- [56] M. Moorkamp, A. G. Jones, and S. Fishwick. Joint inversion of receiver functions, surface wave dispersion, and magnetotelluric data. *J. Geophys. Res.*, 115(B4):B04318, 2010.
- [57] J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer Verlag, Berlin, Heidelberg, New York, second edition, 2006.
- [58] M. Obrebski, S. Kiselev, L. Vinnik, and J.-P. Montagner. Anisotropic stratification beneath africa from joint inversion of sks and p receiver functions. *J. Geophys. Res.*, 115(B9):B09313, 2010.

- [59] J. Pearse and Y. Fialko. Mechanics of active magmatic intraplate in the rio grande rift near socorro, new mexico. *J. Geophys. Res.*, 115(B7):B07413, 2010.
- [60] M. Reiter and R. Chamberlin. Alternative perspectives of crustal and upper mantle phenomena along the rio grande rift. *GSA Today*, 21(2), 2011.
- [61] M. Roy, J. MacCarthy, and J. Selverstone. Upper mantle structure beneath the eastern colorado plateau and rio grande rift revealed by bouguer gravity, seismic velocities, and xenolith data. *Geochemistry Geophysics Geosystems*, 6(10):Q10007, 2005.
- [62] C. Ruhl, S. Bilek, and J. Stankova-Pursley. Relocation and characterization of the august 2009 microearthquake swarm above the socorro magma body in the central rio grande rift. *Geophysical Research Letters*, 37(23):L23304, 2010.
- [63] Y. Saad. *Iterative methods for sparse linear systems*. Society for Industrial and Applied Mathematics, 2003.
- [64] J. Schmedes, J. Roth, R. Saltzer, and G. Leahy. Imaging the shallow crust using teleseismic tomography. *Bulletin of the Seismological Society of America*, 102(3):1276–1282, 2012.
- [65] C. Schultz, S. Myers, J. Hipp, and C. Young. Nonstationary bayesian kriging: A predictive technique to generate spatial corrections for seismic detection, location, and identification. *Bulletin of the Seismological Society of America*, 88(5):1275–1288, 1998.
- [66] L. Sirgue and R. G. Pratt. Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies. *Geophysics*, 69(1):231–248, 2004.
- [67] T. Song and D. Helmberger. A depleted, destabilized continental lithosphere near the rio grande rift. *Earth and Planetary Science Letters*, 262(1):175–184, 2007.
- [68] A. Sosa, A. A. Velasco, L. Velazquez, M. Arguez, and R. Romero. Constrained optimization framework for joint inversion of geophysical datasets. (submitted). *Geophys. J. Int.*, (Submitted), 2012.

- [69] S. Stein and M. Wyssession. *An introduction to seismology, earthquakes and Earth structure*. Blackwell Publishing Ltd, 2003.
- [70] A. Tarantola. A strategy for nonlinear elastic inversion of seismic reflection data. *Geophysics*, 51(10):1893–1903, 1986.
- [71] L. Thompson, A. A. Velasco, and M. Husein. 3-d structure of the southern rio grande rift from 1-d constrained joint inversion of receiver functions and surface wave dispersion. *Geophysical Research Letters*, (Submitted), 2012.
- [72] C. H. Thurber. Seismic tomography of the lithosphere with body waves. *Pure and Applied Geophysics*, 160:717–737, 2003. 10.1007/PL00012555.
- [73] A. Tikhonov and V. Arsenin. *Solutions of ill-posed problems*. Scripta series in mathematics. Winston, 1977.
- [74] M. Ulbrich and S. Ulbrich. Primal-dual interior-point methods for pde-constrained optimization. *Mathematical Programming*, 117:435–485, 2009. 10.1007/s10107-007-0168-7.
- [75] C. Vogel. *Computational methods for inverse problems*, volume 23. Society for Industrial Mathematics, 1987.
- [76] M. West, J. Ni, W. Baldrige, D. Wilson, R. Aster, W. Gao, and S. Grand. Crust and upper mantle shear wave structure of the southwest united states: Implications for rifting and support for high elevation. *Journal of Geophysical Research*, 109(B3):B03309, 2004.
- [77] R. A. Wiggins. The general linear inverse problem: Implication of surface waves and free oscillations for earth structure. *Reviews of Geophysics and Space Physics*, 10:251, Feb. 1972.
- [78] D. Wilson and R. Aster. Imaging crust and upper mantle seismic structure in the southwestern united states using teleseismic receiver functions. *The Leading Edge*, 22(3):232–237, 2003.

- [79] D. Wilson and R. Aster. Seismic imaging of the crust and upper mantle using regularized joint receiver functions, frequency–wave number filtering, and multimode kirchhoff migration. *Journal of geophysical research*, 110(B5):B05305, 2005.
- [80] D. Wilson, R. Aster, J. Ni, S. Grand, M. West, W. Gao, W. Baldrige, and S. Semken. Imaging the seismic structure of the crust and upper mantle beneath the great plains, rio grande rift, and colorado plateau using receiver functions. *Journal of geophysical research*, 110(B5):B05306, 2005.
- [81] L. Woodward. Rate of crustal extension across the rio grande rift near albuquerque, new mexico. *Geology*, 5(5):269–272, 1977.
- [82] S. J. Wright. *Primal-dual interior-point methods*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 1997.
- [83] L. Zhu and H. Kanamori. Moho depth variation in southern california from teleseismic receiver functions. *Journal of Geophysical Research*, 105:2969–2980, 2000.

# Curriculum Vitae

Anibal Sosa was born in Bogota, Colombia, and he is the second son of Clara Aguirre and Efrain Sosa. His interest in science, especially physics and mathematics, started in his childhood mainly motivated by his father's compulsive reading of scientific magazines and books. He attended Universidad del Valle in Cali, Colombia, majoring on Mathematics in 2000. Four years later he obtained a MSc in Applied Mathematics at the same university. Once he finished his master's, he became a full time Professor at the Universidad Icesi in Cali. As one of the organizers for the international conference APPOPT in San Andres, Colombia, he met the former director of the Program in Computational Science Dr. Leticia Velazquez, and decided to pursue a doctoral degree in that program at the University of Texas at El Paso (UTEP) starting the Fall of 2008. While pursuing his degree, he has been working as a research associate for the Cyber-ShARE center of excellence in its geosciences subproject. He is currently a member of the Society for Industrial and Applied Mathematics (SIAM), the American Geophysical Union (AGU) and the American Mathematical Society (AMS). He was one of the founder and first president of the UTEP Student Chapter of SIAM. During his academic life at UTEP he has been awarded with several travel grants from Graduate School, SACNAS, and SIAM. He was also a recipient of a grant from the National Science Foundation (NSF), which allowed him to present his research at the international IUGG Conference on Mathematical Geophysics in the United Kingdom last summer. Moreover, during 2011, he got the first place for oral presentation at the Graduate School Research Exposition in 2011 at UTEP. he has presented his research at international and national conference meetings including AGU Fall meetings, SIAM meetings, SACNAS national conferences, and the IUGG Conference on Mathematical Geophysics. He is the first author of three research papers compiled in his dissertation entitled, "Constrained Optimization Schemes for Geophysical Inversion of Seismic Data" supervised by Dr. Aaron Velasco. Upon completion of his doctoral studies, he will continue his collaboration with the Cyber-ShARE center at UTEP and continue his career as Professor in Universidad Icesi in his home town Cali, Colombia.

UTEP - Geological Sciences 124  
500 West University Avenue  
El Paso, Texas 79968-0514  
email: [usosaaguirre@miners.utep.edu](mailto:usosaaguirre@miners.utep.edu)

Home address  
200 N Festival Apt 1902  
El Paso, Texas 79912